HAL
open science

Multicriterion evaluation of animal welfare : example of on-farm dairy cows

Raphaëlle Botreau

To cite this version:

Raphaëlle Botreau. Multicriterion evaluation of animal welfare : example of on-farm dairy cows. Life Sciences [q-bio]. AgroParisTech, 2008. English. NNT : 2008AGPT0008 . pastel-00004380

HAL Id: pastel-00004380

https://pastel.hal.science/pastel-00004380

Submitted on 5 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N°/__/__/__/__/__/__/__/__/__/__/

THÈSE

pour obtenir le grade de

Docteur

de

l'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech)

Spécialité : Productions Animales

présentée et soutenue publiquement

par

Raphaëlle BOTREAU

le 12 février 2008

EVALUATION MULTICRITERE DU BIEN-ETRE ANIMAL

EXEMPLE DES VACHES LAITIERES EN FERME

Directeurs de thèse : ***Isabelle VEISSIER et Patrice PERNY***

Travail réalisé dans l'Équipe Adaptation et Comportements Sociaux,
UR1213 Herbivores, INRA de Theix, F-63122 Saint-Genès-Champanelle,
dans le cadre d'une bourse CIFRE avec l'Institut de l'Elevage

Devant le jury :

M. Thierry DORE,	Professeur, **Agro Paris Tech**	**Président**
Mme Marie-Christine SALAÜN,	Ingénieur de Recherches, **INRA**	**Rapporteur**
M. Alexis TSOUKIÀS,	Directeur de Recherches, **Université Paris Dauphine**	**Rapporteur**
M. Harry BLOKHUIS,	Professeur, **Université Agronomique de Suède**	**Examinateur**
M. Patrice PERNY,	Professeur, **Université Paris 6**	**Examinateur**
Mme Isabelle VEISSIER,	Directeur de Recherches, **INRA**	**Examinateur**

L'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech) est un Grand Etablissement dépendant du Ministère de l'Agriculture et de la Pêche, composé de l'INA PG, de l'ENGREF et de l'ENSIA (décret n° 2006-1592 du 13 décembre 2006)

Tous les animaux sont égaux, mais certains sont plus égaux que d'autres

par George Orwell (*La Ferme des Animaux*)

Remerciements

Je tiens tout particulièrement à remercier mes deux directeurs de thèse, Isabelle Veissier et Patrice Perny, qui m'ont encadrée et encouragée durant mes trois années passées avec eux. Vous avez su m'apporter le soutien scientifique et humain qui m'a permis de mener ma thèse à son terme. Je vous remercie pour votre disponibilité et les discussions, toujours enrichissantes, que nous avons eues au cours des trois années qui viennent de s'écouler. Un merci spécial à Isabelle dont j'ai très souvent franchi le seuil du bureau avec plein de questions et d'interrogations...

Mon travail n'aurait pas pu se faire sans le projet européen Welfare Quality® dont il est partie intégrante. Je remercie notamment les coordinateurs du projet : Harry Blokhuis, Mara Miele, Bryan Jones et Isabelle Veissier.

Je remercie aussi l'Institut de l'Elevage, Claude Allo, Jacques Lucbert et Valérie David, sans qui mon projet, réalisé en bourse CIFRE, n'aurait peut-être pas vu le jour. Merci pour votre accueil et votre participation tout au long de mon travail !

Je suis reconnaissante à l'Unité de Recherche sur les Herbivores et à ses directeurs successifs, Jean-Baptiste Coulon et Jean-François Hocquette, de m'avoir accueillie durant ces trois ans.

Je remercie les membres de mon comité de thèse, Christine Duvaux-Ponter, Stéphane Ingrand, Jacques Capdeville, Carlos Lopez et Jacques Lucbert, pour leurs conseils et commentaires constructifs.

Je remercie également Marie-Christine Salaün et Alexis Tsoukiàs d'avoir accepté d'être les rapporteurs de mon travail, ainsi que Harry Blokhuis et Thierry Doré d'avoir bien voulu le juger.

Puisque fort heureusement les thésards ne vivent pas comme des reclus, isolés du reste du monde, je tiens également à remercier tous ceux qui m'ont soutenue, réconfortée et encouragée durant mes trois années de thèse...

Ainsi, je remercie toute l'équipe Adaptation et Comportements Sociaux pour les discussions, les conseils, le soutien et la bonne humeur de tous ses membres. Pendant ces trois années passées à Theix, j'ai découvert la définition de l'expression "esprit d'équipe". Je vous remercie tous pour votre accueil et les moments de détente que nous avons pu vivre ensemble ! Un grand merci donc à Alain, Xavier, Marie-Madeleine, Véronique, Claudia, Éric, Stéphane, Christine, Pascale, Anne-Marie, mais aussi à Gilles, Gilbert, Hervé et Mado qui ne résident désormais plus à Theix...

Je remercie aussi tous les thésards, post-docs et stagiaires qui ont partagé mon quotidien et m'ont supportée durant ma thèse : Lucile, Séverine, Angélique, Céline, Loïc, Luc, Cécile, Amandine, Nadège, Jean-Loup, Aurélie, François, Nicolas... Vous m'avez tous apporté une bouffée d'oxygène quand c'était nécessaire. Il faut paraît-il rire au moins une fois par jour, grâce à vous ce fut mission accomplie !

Je dois aussi la bonne humeur générale dans laquelle j'ai évolué pendant trois ans à tous mes collègues du troisième étage. Merci !

Je tiens également à remercier toutes les personnes qui m'ont permis d'arriver jusque là. Merci à Christine Duvaux-Ponter et Sabine Roussel pour m'avoir fait découvrir à l'INA P-G la notion de bien-être animal. Merci à Pierre Orgeur et Violaine Colson pour m'avoir permis de réaliser mon premier stage en relation avec le bien-être animal et de découvrir le monde de la recherche. Merci aussi à Anne Brulé et Jacques Capdeville pour m'avoir accompagnée et encadrée durant mon stage de fin d'études.

Un grand merci aussi à mes amis, Rémi, Coralie, Julie, William, Thomas, Marie et Johanne, pour leur soutien et avoir même fait le déplacement jusqu'en Auvergne (mention spéciale pour l'effort réalisé par les parisiens...).

Enfin, le soutien indéfectible et l'amour de mes parents, de ma sœur et de ma grand-mère m'ont permis d'arriver au bout de cette grande aventure... Merci !

Table des matières

Table des matières

Table des matières **1**

Avant-propos **5**

1 Introduction **7**

1.1 Le contexte général 7

1.1.1 La place de l'animal dans la société 7

1.1.2 La situation actuelle : entre consommateurs, associations de protection des animaux et éleveurs 10

1.1.3 Comment prendre en compte l'attente sociétale liée au bien-être animal ? 16

1.1.4 Évaluer le bien-être animal : un travail interdisciplinaire 27

1.2 Le projet Welfare Quality® 31

1.2.1 Objectifs du projet 31

1.2.2 Organisation du projet 34

1.2.3 Schéma général du système visé par Welfare Quality® 37

1.2.4 Quelle est la place de mon travail de thèse au sein du projet ? . . 39

1.3 La problématique générale de la thèse 40

1.3.1 Le bien-être animal : un concept multidimensionnel 40

1.3.2 L'évaluation du bien-être animal repose sur la réalisation de nombreuses mesures 43

1.3.3 Dans quels buts évaluer le bien-être ? 46

1.3.4 Construire une évaluation multicritère du bien-être animal 48

1.4 L'objectif de mon travail 53

2 Revue des méthodes actuellement proposées pour une évaluation globale du bien-être animal **55**

2.1 Agrégation non formalisée de plusieurs mesures de bien-être 56

2.2 Définition d'exigences minimales pour chacune des mesures effectuées . . 58
2.3 Calcul de la somme (ou moyenne) des rangs obtenus par chaque ferme sur les différentes mesures 59
2.4 Calcul de la somme (ou moyenne) des scores obtenus sur chaque mesure 61
2.5 Utilisation de règles définies au cas par cas afin de limiter les compensations entre les différents éléments 62
2.6 Conclusion 63

3 Identification des contraintes pour une évaluation globale du bien-être animal 75
3.1 Caractéristiques liées au concept de bien-être animal 75
3.2 Caractéristiques liées à l'interprétation des mesures en termes de bien-être 77
3.3 Caractéristiques liées à la collecte des données 79

4 Structure générale du modèle d'évaluation développé dans la thèse 91
4.1 Structure générale d'un processus de décision 91
4.2 Choix de la famille de critères 93
4.2.1 Les propriétés d'une famille de critères 93
4.2.2 Présentation de notre famille de critères 96
4.3 Une construction progressive pour aller des mesures à l'évaluation globale 104

5 La construction des sous-critères 111
5.1 Quelles sont les mesures servant à évaluer le bien-être dans Welfare Quality® ? 111
5.1.1 Identification des mesures servant à évaluer chacun des 12 sous-critères 111
5.1.2 Caractéristiques des mesures choisies pour évaluer le bien-être animal dans Welfare Quality® 113
5.2 Principes généraux de la construction 118
5.3 Exemples de construction 121
5.3.1 Exemple 1 : Sous-critère g_2 *"Absence de soif prolongée"* 121
5.3.2 Exemple 2 : Sous-critère g_6 *"Absence de blessures"* 124
5.3.3 Exemple 3 : Sous-critère g_3 *"Confort autour du repos"* 139
5.4 Expérimentations sur jeux de données réelles et affinage de la construction des sous-critères 145

6 La construction des critères **151**
6.1 Consultation d'experts sur des jeux de données virtuelles 152
6.2 Synthèse de l'information fournies par les sous-critères 154
6.2.1 Présentation des agrégateurs testés : *WS*, *OWA* et *Choquet* . . . 154
6.2.2 Comparaison des agrégateurs pour la construction des critères et choix de l'agrégateur le mieux adapté à notre situation 159

7 L'agrégation des critères **179**
7.1 Définition du nombre et et de la signification des catégories de bien-être 179
7.1.1 Identification des utilisations potentielles pouvant être faites de l'outil d'évaluation 180
7.1.2 Conséquences en termes de nombre et de signification des catégories de bien-être 183
7.2 Procédures d'affectation aux quatre catégories de bien-être 187
7.2.1 Présentation de la méthode ELECTRE TRI 188
7.2.2 Présentation de la méthode que nous avons adoptée 190
7.2.3 Choix des profils délimitant les catégories de bien-être 192
7.2.4 Définition des procédures d'affectation aux différentes catégories de bien-être 192
7.3 Comparaison de l'impact des quatre procédures sur l'évaluation globale . 198
7.3.1 Comparaison sur un jeu de fermes virtuelles 198
7.3.2 Comparaison sur le jeu de fermes réelles évaluées lors du WP2.4 . 198

8 Vers un outil de diagnostic et de conseil efficace **235**
8.1 Exemples de fermes affectées à la catégorie *"Non classé"* 237
8.1.1 Exemple 1 237
8.1.2 Exemple 2 239
8.1.3 Exemple 3 240
8.2 Exemples de fermes affectées à la catégorie *"Normal"* 241
8.2.1 Exemple 4 241
8.2.2 Exemple 5 242
8.2.3 Exemple 6 242
8.3 Exemples de fermes affectées à la catégorie *"Bon"* 243
8.3.1 Exemple 7 243
8.3.2 Exemple 8 244
8.3.3 Exemple 9 245
8.4 Vers une automatisation de la procédure de diagnostic-conseil 246

8.4.1 Procédure de diagnostic 246
8.4.2 Procédure de conseil 248

Discussion **251**
Un modèle "vaches laitières" présentant plusieurs atouts 252
Vers le modèle final d'évaluation du bien-être des vaches laitières en ferme . . 256
Transfert du modèle d'évaluation vers les autres couples (type d'animal, période) 264
Perspectives de recherche : vers une application du modèle pour l'évaluation d'autres notions multidimensionnelles liées à l'élevage et à l'agriculture . 265

Glossaire **269**

A Organisation du projet Welfare Quality® **271**
A.1 Sous-projet 1 : Attentes des consommateurs, des distributeurs et des éleveurs vis-à-vis du bien-être animal et de produits respectueux du bien-être des animaux dont ils sont issus 273
A.2 Sous-projet 2 : Développement d'une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir 274
A.3 Sous-projet 3 : Développement de stratégies permettant d'améliorer le bien-être des animaux d'élevage 275
A.4 Sous-projet 4 : Vers la mise en application d'une procédure standardisée d'évaluation et d'information du bien-être ainsi que vers celle de stratégies d'amélioration du bien-être 277

B Composition du Comité Conseil du projet Welfare Quality® **279**

C Description de la construction de l'ensemble des sous-critères **281**

D Démonstration de la continuité d'ordre 2 des fonctions B-splines cubiques **369**

Références bibliographiques **373**

Table des figures **387**

Liste de mes publications **391**

Avant-propos

Avant-propos

Dans un contexte où de plus en plus de citoyens se préoccupent du respect du bien-être des animaux d'élevage, la nécessité de mieux informer le citoyen-consommateur se fait ressentir. C'est ainsi que le projet Européen Welfare Quality® a été mis en place. Il vise à développer un système européen d'évaluation du bien-être animal au niveau de la ferme, utilisable en routine pour servir de base à un système européen d'information sur le bien-être animal (en particulier dans le cadre de programmes de certification des élevages). Ce travail a nécessité le développement d'une procédure formalisée d'évaluation globale du bien-être animal au niveau de la ferme afin de pouvoir standardiser la méthode.

Ma thèse a été réalisée au sein de Welfare Quality® dans le but de mettre en place un modèle d'évaluation du bien-être des vaches laitières en ferme. Ce travail constitue une application pratique de méthodes développées en recherches opérationnelles et aide à la décision multicritère.

Le manuscrit de thèse comprend huit chapitres :

Dans un premier temps nous allons introduire le contexte général dans lequel s'insère la thèse (*Chapitre 1*). Puis, nous allons rapidement présenter les principales méthodes proposées à l'heure actuelle pour produire une évaluation globale du bien-être animal à l'échelle de la ferme. Les principales qualités et défauts de ces méthodes seront alors discutées, notamment en vue d'une utilisation dans le cadre d'un système de certification (*Chapitre 2*). À l'issue de cette revue bibliographique, les caractéristiques propres à l'évaluation du bien-être animal seront identifiées. Les propriétés souhaitées pour le modèle d'évaluation seront ainsi listées (*Chapitre 3*).

Nous présenterons ensuite le modèle d'évaluation proposé pour les vaches laitières. Nous définirons alors la famille de critères (12 sous-critères, regroupés en 4 critères) sur laquelle reposera le modèle d'évaluation proposé dans la thèse (*Chapitre 4*). Le travail de construction du modèle sera présenté, en commençant bien sûr par la construction des

sous-critères à partir des mesures réalisées sur le terrain (*Chapitre 5*), avant de décrire l'agrégation des sous-critères pour évaluer les critères (*Chapitre 6*). Nous présenterons enfin la méthode choisie pour synthétiser l'information fournie par les critères en une seule évaluation globale à l'échelle du troupeau (selon une problématique de tri ordonné des fermes) (*Chapitre 7*). Le pouvoir explicatif du modèle sera ensuite présenté à l'aide d'exemples de fermes évaluées dans Welfare Quality® (*Chapitre 8*).

Dans la discussion, nous ferons le point sur le modèle proposé, sur ce qui reste à faire dans le cadre du projet Welfare Quality®, avant de conclure sur les utilisations potentielles, à plus long terme, de ce type d'application multicritère dans le domaine de l'élevage.

Les articles rédigés au cours de la thèse sont inclus dans les chapitres traitant du même sujet[1] :

- Botreau et al. (2007) *Aggregation of measures to produce an overall assessment of animal welfare : Part 1 - a review of existing methods*, à la fin du Chapitre 2.
- Botreau et al. (2007) *Aggregation of measures to produce an overall assessment of animal welfare : Part 2 - analysis of constraints*, à la fin du Chapitre 3.
- Botreau et al. (2007) *Definition of criteria for overall assessment of animal welfare*, à la fin du Chapitre 4.
- Botreau et al. (soumis) *Multicriteria evaluation of animal welfare at farm level : an application of MCDA methodologies*, à la fin du Chapitre 7. Cet article permet de résumer le contenu des Chapitres 5 à 7.

[1] Ces articles sont imprimés sur du papier de couleur "ivoire"

Chapitre 1

Introduction

Résumé de l'introduction :

Ce chapitre introductif traite successivement les quatre points suivants :

Le contexte général : présentation du contexte général dans lequel s'inscrit la thèse, en traitant de la place de l'animal en Europe, de l'émergence d'attentes sociétales quant au respect du bien-être des animaux d'élevage et des différentes solutions permettant de prendre en compte ces attentes.

Le projet Welfare Quality® : présentation du projet européen, de ses objectifs et de son organisation, puis du système d'évaluation et d'information visé par le projet, et enfin présentation de mon rôle au sein de ce projet.

La problématique générale de la thèse : définition du concept multidimensionnel de bien-être animal dont l'évaluation repose sur la réalisation de nombreuses mesures sur le terrain, puis présentation des raisons pour lesquelles il peut être nécessaire d'évaluer le bien-être, et enfin présentation de la problématique dans laquelle s'inscrit l'évaluation multicritère du bien-être animal recherchée au sein de Welfare Quality®.

L'objectif de mon travail : construire le modèle d'évaluation globale du bien-être des vaches laitières en ferme qui servira d'exemple pour la construction des modèles d'évaluation pour les autres types d'animaux concernés par Welfare Quality®.

Chapitre 1

Introduction

1.1 Le contexte général

1.1.1 La place de l'animal dans la société

Depuis le début de la domestication de l'animal par l'homme, ce dernier voit dans l'animal l'aspect utilitaire : force de travail (boeufs, chevaux...), nourriture (viande, graisse pour la conservation des aliments), habillement (fourrures, cuir), etc. L'animal d'élevage a donc une utilité économique, tout comme l'animal de compagnie a une utilité sociale et l'animal de sport une utilité ludique (Dantzer 1995).

Pendant longtemps l'animal a été considéré par certains comme une "machine", un être incapable de souffrir et pouvant donc être soumis à toutes contraintes (Bourdon 2003 ; Le Neindre 2003). Cette façon de penser l'animal se rapproche de la vision mécaniste de l'animal par Descartes conduisant à nier tout problème éthique concernant les traitements cruels dont l'animal peut être l'objet (Burgat et Dantzer 1997). Cette façon de penser l'animal a été favorisée par l'industrialisation et l'intensification de la production des animaux d'élevage, notamment au cours du 19e siècle (Verley 1997 cité par Bourdon 2003) mais aussi après la Seconde Guerre mondiale où un "pacte productiviste" est alors conclu entre le monde agricole et l'état français afin de nourrir la population (Poulain 1997). De nos jours, cette thèse n'est plus soutenue (Le Neindre 2003) du fait de l'évolution des mentalités vers le souci du bien-être animal.

Cette évolution est passée par la reconnaissance de devoirs de l'homme envers les animaux, et même pour certains par la reconnaissance de droits de l'animal. Selon Burgat et Dantzer (1997), dans la conception traditionnelle de la relation de l'homme à l'animal, l'homme a des devoirs envers les animaux, c.-à-d. qu'il ne doit pas les faire souffrir "inutilement", en leur infligeant une souffrance non requise par la finalité d'un

acte. C'est ce qu'exprime Kant en écrivant en 1797 que s'il est *"permis à l'homme de tuer des animaux de façon expéditive (sans torture), ou de leur imposer un travail à condition qu'il n'excède pas leur force, en revanche, il faut exécrer les expériences physiques au cours desquelles on les martyrise au seul profit de la spéculation, alors qu'on pourrait se passer d'elles pour atteindre le but visé"* (1986 cité par Burgat et Dantzer 1997).

Cette approche, au sens où elle autorise de faire souffrir des animaux si, et uniquement si, les actes commis ont une finalité, peut se rapprocher de la vision utilitariste. Il existe cependant une différence importante dans l'approche. En effet, alors que le kantisme vise à juger les actions sur leur conformité à certains devoirs (notamment moraux), l'utilitarisme ne se base que sur les conséquences des actes pour les juger (forme de conséquentialisme). L'utilitarisme vise en effet à produire "le plus grand bonheur du plus grand nombre"[1] en choisissant les solutions qui offrent le meilleur équilibre entre la somme des satisfactions et la somme des frustrations générées par l'action, quels que soient les moyens, du moment que le résultat est là. Certains philosophes, comme Singer (1990), ont étendu l'utilitarisme aux animaux. L'utilitarisme n'impose pas le maintien d'un bon niveau de bien-être pour tous : alors que certains animaux peuvent vivre dans des conditions de bien-être, une minorité d'entre eux peut souffrir de mal-être, tant que sur l'ensemble des animaux on juge le niveau de bien-être correct.

Au contraire, la déontologie, une approche philosophique beaucoup plus proche du Kantisme, met en avant les droits basiques de chaque individu, en particulier lorsqu'il s'agit d'êtres sensibles (voir par ex. Feinberg 1980), et ainsi déclare qu'il n'est pas possible de justifier de bons résultats si pour les obtenir on a usé de moyens violant les droits d'un seul individu (Regan 1992). Appliquée aux animaux, cette approche conduit à attribuer non seulement des devoirs aux hommes mais aussi des droits aux animaux.

Peu à peu le respect des animaux s'est introduit dans la législation française, condamnant certains actes de maltraitance commis hors du cadre prévu par la loi et prévoyant des modalités d'utilisation des animaux (Burgat et Dantzer 1997).

La loi a tout d'abord protégé les animaux utilitaires en les considérant uniquement comme des biens appartenant à leurs propriétaires. Ainsi un décret de 1791 réprimandait les mauvais traitements infligés aux animaux *d'autrui*[2]. Ce n'est que depuis 1850 que la France dispose d'un texte réglementaire, la loi Grammont, interdisant la cruauté vis-à-vis des animaux, mais seulement en public (Le Neindre 2003) : *"Seront punis d'une amende de cinq à quinze francs, et pourront l'être d'un à cinq jours de prison, ceux qui*

[1] *"the greatest good for the greatest number"* (expression de Joseph Priestley, lue par Bentham, le "père" de l'utilitarisme lors d'un discours en 1768)

[2] Ce décret a inspiré l'article 453 du Code pénal de 1810

auront exercé publiquement et abusivement des mauvais traitements envers les animaux domestiques" (extrait de la loi Grammont, cité par Dupas 2005). Ce premier pas vers la protection des animaux ne concernait pas ce qui pouvait être fait dans le domaine privé. La loi du 19 novembre 1963 (loi 63-1143, articles 453 et 454) sur la protection des animaux préconise des amendes en cas de sévices sur animaux domestiques, voire même des peines de prison en cas de sévices graves. La nécessité de tenir compte du bien-être des animaux qui vivent sous la dépendance de l'homme apparaît dans la loi sur la protection de la nature du 10 juillet 1976 : *"tout animal étant un être sensible, il doit être placé par son propriétaire dans des conditions compatibles avec les impératifs biologiques de son espèce"* (loi 76-629, article 9). Cette loi a été suivie de décrets et d'arrêtés précisant les modalités d'application pour les animaux d'élevage ou utilisés dans le cadre d'expérimentations (décrets 80-791 du 01/10/80, 87-848 du 19/10/87...). Depuis la fin des années 80, la protection des animaux a fait l'objet de règlements européens qui ont alors été transcrits en droit français (voir Section 1.1.3).

Ainsi, au cours des trente dernières années on a pu voir une nette augmentation de la préoccupation pour le bien-être animal, se traduisant par des réglementations mais aussi par une prise de conscience du grand public. En effet, l'intensification de la production conduit de plus en plus de consommateurs à critiquer les conditions dans lesquelles les animaux de rente sont élevés (Burgat et Dantzer 1997). Malheureusement, les conditions d'élevage sont pour la plupart méconnues des consommateurs. En 1930, le rapport entre la population urbaine et la population rurale s'est inversé (Poulain 1997), et depuis ne cesse d'augmenter (cf. Figure 1.1 pour voir l'évolution de 1965 à nos jours), déconnectant ainsi l'aliment acheté par le consommateur de son univers de production (Poulain 1997).

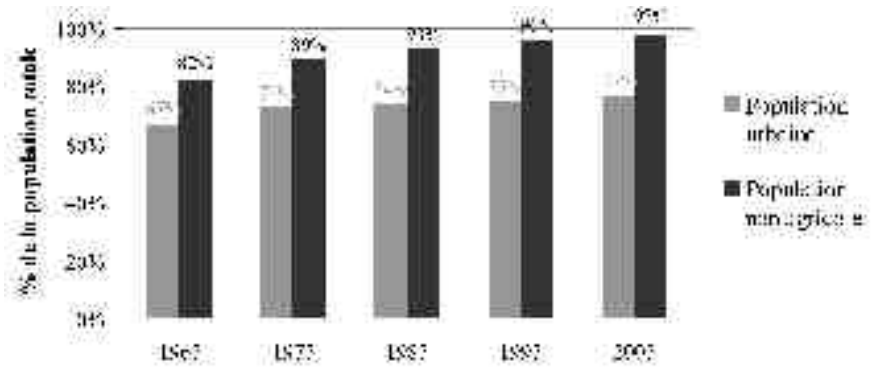


FIG. 1.1 – Évolution de la population urbaine et de la population non agricole en France, de 1965 à 2005 (Source : banque de données de la FAO[4])

1.1.2 La situation actuelle : entre consommateurs, associations de protection des animaux et éleveurs

1.1.2.1 Des consommateurs de plus en plus concernés par le bien-être

Le grand public, sensibilisé par les crises sanitaires récentes (la crise de la "vache folle", celle de la fièvre aphteuse en Angleterre et ses charniers...) et par les actions (média, communiqués...) des associations de protection des animaux, se sent de plus en plus concerné par le bien-être des animaux d'élevage. Ce phénomène est sans doute accentué par la méconnaissance par le grand public des méthodes d'élevage et des textes réglementaires. Lors de l'Eurobaromètre conduit en 2007 par l'Union Européenne sur le bien-être animal, 58% des personnes ayant répondu au questionnaire Internet souhaiteraient être mieux informées des conditions d'élevage des animaux dans leur pays (European Commission 2007b). Des distorsions apparaissent alors entre les représentations de l'élevage que les consommateurs peuvent avoir en tête et la réalité. Ainsi, lors de l'Eurobaromètre conduit en 2005, la façon dont les consommateurs ayant répondu au questionnaire Internet considèraient le niveau de bien-être en élevage des poules pondeuses et des vaches laitières était clairement influencée par leur connaissance de l'élevage (European Commission 2005). En effet, plus les consommateurs avaient eu l'occasion de visiter des élevages, plus ils considèraient que le bien-être des poules pondeuses était mauvais et que celui des vaches laitières était bon (cf. Figure 1.2).

Une enquête assez récente (citée par Le Neindre 2003) conduite sur un échantillon représentatif de la population française a montré que plus de 80% des personnes considéraient que les conditions modernes d'élevage pouvaient conduire à des altérations du bien-être des animaux. Ceci se confirme lors de l'Eurobaromètre de 2007 au cours duquel 77% des sondés ont dit penser que le bien-être des animaux d'élevage dans leur pays a besoin d'être amélioré (European Commission 2007b).

Depuis les années 50, la France, ainsi que toute l'Europe, a développé son agriculture sur le modèle du productivisme (Poulain 1997). Désormais les européens mangent à leur faim, nous sommes même depuis les années 90 dans une phase de surproduction. Les consommateurs, n'ayant plus à craindre le manque de nourriture, commencent à s'intéresser à la qualité des produits qu'on leur fournit, ainsi qu'au traitement des animaux. Dans une étude menée sur plus de mille consommateurs français, Latouche (1999) a pu mettre en évidence un grand intérêt des français pour le bien-être animal. De même au niveau européen, dans l'étude menée au sein du projet Welfare Quality® sur sept pays

[4]Organisation des Nations Unies pour l'Alimentation et l'Agriculture, sur le site Internet FAOSTAT http ://faostat.fao.org/site/550/default.aspx

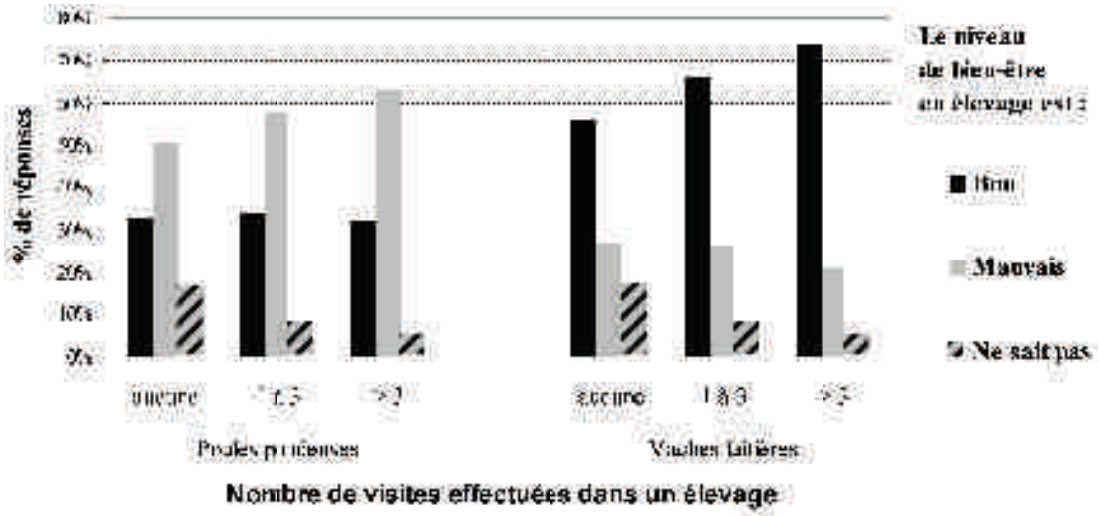


FIG. 1.2 – Réponses des consommateurs à la question *"Comment jugez-vous le niveau de bien-être animal en élevage ?"*, appliquée au cas des poules pondeuses et des vaches laitières, en fonction du nombre de fois où ils ont visité des élevages

de l'Union Européenne, 55% des personnes consultées ont déclaré considérer le bien-être animal comme étant très important (Kjaernes et al. 2007). Ainsi, quand ils achètent de la viande, 43% des personnes ayant répondu à l'Eurobaromètre de 2007 déclarent penser au moins de temps en temps au bien-être des animaux dont sont issus les produits qu'ils achètent (European Commission 2007a). Cependant, dans l'esprit de la plupart des consommateurs, la qualité sanitaire et gustative des produits qu'ils achètent est souvent reliée aux conditions d'élevage et à la qualité de vie des animaux. Ce point là a aussi pu être mis en évidence lors de l'Eurobaromètre de 2007 au cours duquel la plupart des consommateurs ont dit considérer que des produits respectueux du bien-être sont meilleurs pour leur santé et sont de meilleure qualité (European Commission 2007b) :

- 51% des sondés associent le respect du bien-être animal à des produits meilleurs pour la santé du consommateur,
- 48% à des produits de meilleure qualité,
- 43% à des animaux en meilleure santé *(seulement 43%)*,
- 34% à des produits ayant un meilleur goût,
- ...

Cette association entre qualité du produit et bien-être animal se retrouve aussi dans les études menées au sein du projet Welfare Quality®, et tout particulièrement en France (Kjærnes et al. 2007). De même, Kjærnes et al. (2007) ont constaté que les consommateurs faisaient souvent le lien entre des produits meilleurs pour la santé et

des produits plus respectueux du bien-être.

Ainsi la préoccupation concernant le bien-être des animaux de rente semble avoir des origines diverses : certaines personnes ont un réel souci de l'aspect éthique des conditions de vie des animaux, d'autres s'intéressent plus à la qualité des produits qu'elles consomment, notamment suite aux crises (ESB, fièvre aphteuse...) de ces dernières années. On peut aussi citer les personnes qui se sentent concernées de façon générale par l'environnement (Le Neindre 2003). Même si elles restent minoritaires, un nombre croissant de personnes se mettent à bannir de leur alimentation tous les produits carnés, considérant qu'il est inadmissible et incompatible avec l'éthique d'élever des animaux pour les manger dans le contexte de surabondance alimentaire des pays occidentaux (Burgat et Dantzer 1997). Burgat et Dantzer (1997) opposent alors à cette position *radicale*, une position *réformiste* visant à améliorer les modes d'élevage, de transport et d'abattage. Fischler (entretien rapporté par Autrement 1997) rappelle alors le paradoxe du consommateur qui veut manger de la viande, mais qui aimerait mieux ne pas avoir à tuer des animaux.

Enfin, on peut rapprocher le souci du bien-être animal d'un sentiment beaucoup plus général d'anxiété, d'insécurité et de méfiance dû au fait qu'on ne connaît pas l'histoire, la nature et l'identité du produit que l'on mange (Autrement 1997). Ce que produit l'agriculture étant *"transformé, puis pris en main par l'industrie, à nouveau transformé, conditionné et distribué"*, le consommateur finit par ne plus savoir ce qu'il mange (Autrement 1997).

Les attentes vis-à-vis du respect du bien-être des animaux semblent issues des questionnements plus généraux des consommateurs vis-à-vis des produits animaux, en particulier ceux qui nécessitent la mort de l'animal (la viande). Kjærnes et al. (2007) y voient la conséquence des récentes crises qui ont touché l'élevage (et qui font ainsi brusquement changer les attitudes des consommateurs, Kjærnes 2007), de l'augmentation du nombre de labels de qualité (incluant l'agriculture biologique), mais aussi du travail de communication des associations de protection animale.

1.1.2.2 Des associations de protection des animaux ayant un poids de plus en plus fort

Le rôle des associations de protection des animaux est important aussi bien auprès des citoyens qu'auprès des instances décisionnelles, au niveau des États, comme au niveau communautaire, voire même sur le plan international.

Parmi ces associations on peut citer la Société Nationale pour la Défense des Animaux (SNDA), Talis, la Fondation Ligue Française des Droits de l'Animal (LFDA), La

Protection Mondiale des Animaux de Ferme (PMAF), la Société Protectrice des Animaux (SPA), la Fondation Brigitte Bardot, l'Œuvre d'Assistance aux Bêtes d'Abattoir (OABA), la World Society for the Protection of Animals (WSPA), The Universities Federation for Animal Welfare (UFAW)....

Certaines de ces associations sont très puissantes et peuvent mener des actions d'envergure importante, comme par exemple la pétition, déposée en 2001 au Parlement européen, demandant l'amélioration du bien-être des animaux de rente et signée par plus d'un million de personnes (Le Neindre 2003) ou encore la Déclaration Universelle des Droits de l'Animal, proclamée officiellement le 15 octobre 1978 par la Ligue internationale des droits de l'animal, les ligues nationales et les personnes physiques et morales qui s'associent à elles.

En France, le monde associatif se caractérise par sa diversité, mais depuis 1970 le Conseil National de la Protection Animale (CNPA) fédère de nombreuses associations et rassemble aussi des spécialistes du bien-être animal (Fabre 1995).

La plupart des associations de protection des animaux souhaitent un abandon de l'élevage intensif, qui, selon la LFDA, engendre *"une pollution importante, une surproduction des aliments et une concentration inacceptable des animaux"* (site Internet de la LFDA), et prônent un retour à des élevages de dimensions plus modestes et plus proches de la nature. Les termes d'élevage "fermier" et "naturel" sont souvent avancés. Il s'agit de termes à forte connotation positive pour les consommateurs et dont les représentations sont souvent associées à celles de produits de qualité et respectueux de la nature. De plus ces associations souhaitent des élevages où les animaux seraient "heureux" et auraient la "joie de vivre" (Hasson 1994). Cette vision de l'élevage peut s'avérer être très éloignée des techniques d'élevages et des besoins de rentabilité économique. Toutefois, cela n'empêche pas certains éleveurs de vouloir que leurs animaux soient bien, si ce n'est "heureux" (voir plus loin).

D'autres associations sont conscientes des contraintes économiques et militent pour une amélioration des conditions de vie des animaux dans un contexte viable économiquement. C'est par exemple le cas de la PMAF, créée en 1967 par un éleveur anglais, qui estime qu'il est nécessaire d'établir un dialogue constructif entre les producteurs et les consommateurs pour pouvoir mettre en place dans les élevages des méthodes plus respectueuses du bien-être animal (Burgat et Dantzer 1997).

1.1.2.3 Des éleveurs entre contraintes et atouts

Selon M. Le Denmat, éleveur de porcs (production fortement touchée par les réglementations "bien-être"), *"l'éleveur est celui qui reste le plus proche de l'animal et qui connaît le mieux les conditions d'élevage. Son souci est d'associer productivité et amélioration du confort des animaux. Il est demandeur de critères et de règles précises qui lui permettent de progresser, donc de survivre"* (Le Denmat 1994).

Ce qui est cité par cet éleveur de porcs se retrouve en partie dans les paroles des éleveurs de bovins, de porcs et de volailles, recueillies au cours d'entretiens réalisés dans un projet ICTA[5] (Dockès 2000). En effet, ils considèrent que le métier d'éleveur revient à avoir une relation professionnelle aux animaux, à gagner sa vie avec les animaux. C'est ensuite pour eux un rapport quotidien aux animaux. 67 % des éleveurs de bovins interrogés déclarent d'ailleurs être tout à fait d'accord avec l'affirmation "je suis réellement attaché à mes animaux" et 60 % ressentent un pincement au cœur au moment du départ de leurs animaux de réforme à l'abattoir. Cependant, les éleveurs interrogés n'ont pas tous la même opinion sur les besoins des animaux. Certains leur reconnaissent des besoins comportementaux et psychologiques (absence de stress, bonne relation entre l'éleveur et ses animaux...) en plus des besoins physiologiques (bonne alimentation, bonne santé, conditions de vie...) ; d'autres reconnaissent essentiellement cette dernière catégorie. Les premiers considèrent alors comme légitimes les demandes de la société d'assurer le bien-être des animaux et souhaitent améliorer encore le confort de leurs animaux. Les seconds, par contre, estiment être agressés par les demandes de la société qu'ils jugent parfois incompatibles avec la gestion économique de leurs exploitations.

Pour la majorité des éleveurs, le principal critère d'évaluation du bien-être des animaux en élevage est le niveau des performances obtenues, selon la formule "un animal qui n'est pas malade et qui fournit de bons résultats zootechniques est un animal en situation de bien-être". Ceci tend à changer, le confort et le comportement des animaux étant de plus en plus pris en compte, notamment lorsque le bien-être de l'éleveur est lui aussi respecté (conditions de travail et de revenu satisfaisantes)(Dockès 2000). Ainsi, dans l'enquête réalisée en 2006 au sein du projet européen Welfare Quality®, les éleveurs considèrent que s'occuper des animaux en leur assurant un bon niveau de bien-être est un aspect important de leur travail (Kjærnes et al. 2007). Les éleveurs consultés insistent alors sur la nécessité de satisfaire les besoins physiologiques de leurs animaux, mais insistent aussi sur les aspects psychologiques. De même, ils mettent en avant l'observation de leurs animaux, en plus de l'étude de leurs performances zoo-

[5]Instituts et Centres Techniques Agricoles

techniques, pour évaluer leur bien-être. Cependant cette étude a mis en évidence des divergences importantes entre les pays et les secteurs d'élevage, mais aussi entre les éleveurs suivant des pratiques conventionnelles et ceux impliqués dans des démarches qualité (ces derniers étant plus soucieux du bien-être de leurs animaux).

Dans une autre étude française basée sur des entretiens avec des éleveurs, Dockès et Kling-Eveillard (2006) ont pu mettre en évidence des divergences fortes entre les éleveurs, et ce malgré des considérations communes à tous les éleveurs (ils entretiennent une relation professionnelle avec leurs animaux, au coeur de laquelle l'observation de ces derniers est une activité essentielle...). En se basant entre autres sur leurs divergences concernant leur vision éthique de leur travail, la prise en compte des besoins des animaux et la façon dont ils perçoivent leurs animaux, Dockès et Kling-Eveillard (2006) ont pu distinguer quatre catégories d'éleveurs : les éleveurs pour l'animal, les éleveurs avec l'animal, les éleveurs malgré l'animal et les éleveurs pour la technique. Cette typologie des éleveurs met clairement en évidence des différences de fond entre les éleveurs en ce qui concerne leur relation à l'animal et la vision de leur métier.

Par ailleurs, certaines mesures réglementaires, comme par exemple celles prises au niveau européen consistant à élever les truies gestantes et les veaux en cases collectives, impliquent des modifications des bâtiments et des pratiques des éleveurs. Ces modifications sont souvent coûteuses à la fois en temps et en argent, et ne s'accompagnent pas toujours de gain de productivité. Pour effectuer ces changements, les éleveurs ont alors besoin d'aides financières, via des primes compensatoires de la part de l'État. Le volet économique du bien-être animal est très loin d'être négligeable, d'autant plus que les filières animales souffrent actuellement de la crise économique.

En plus de cette pression économique et réglementaire, beaucoup d'éleveurs ressentent aussi une pression sociale (Kjærnes et al. 2007). La plupart des éleveurs souhaitent une meilleure reconnaissance de leurs savoirs professionnels et de leur engagement vis-à-vis du respect du bien-être de leurs animaux. Ils se sentent souvent accusés et stigmatisés par le reste de la société (Kjærnes et al. 2007).

Excepté les éleveurs impliqués dans des démarches qualité assurant un bon niveau de bien-être et ayant trouvé leur place au sein d'un marché de niche, les éleveurs ne voient pas qui peut payer les coûts additionnels causés par le respect des réglementations et des labels. La plupart ne font confiance ni aux transformateurs, ni aux distributeurs, ni aux consommateurs pour rémunérer leurs efforts (Kjærnes et al. 2007). Toutefois, l'amélioration du bien-être n'implique pas forcément un surcoût. Dans certains cas, elle permet au contraire d'améliorer la productivité en diminuant stress, blessures ou problèmes sanitaires (Veissier et al. 2007). De plus, les mises aux normes des élevages

peuvent s'accompagner d'une revalorisation du métier d'éleveur (cf. Dockès et Kling-Eveillard 2007 illustrant le cas des éleveurs de veaux de boucherie).

1.1.3 Comment prendre en compte l'attente sociétale liée au bien-être animal ?

Deux voies semblent possibles pour favoriser l'amélioration des conditions de vie des animaux : soit la voie réglementaire, soit celle de la labellisation et/ou de la segmentation du marché (Kjærnes et al. 2007).

1.1.3.1 La réglementation en vigueur sur le bien-être animal

L'élaboration des réglementations doit tenir compte à la fois des données scientifiques (sur les effets de différents facteurs sur le bien-être des animaux), des contraintes des éleveurs, à la fois économiques et humaines (bien-être de l'éleveur), et des demandes des citoyens et des associations de protection animale. Les règlements sont le fruit de négociations dont la dimension politique est souvent déterminante.

L'élaboration d'une réglementation se situe au niveau à la fois national, via les instances législatives françaises, et européen, via le Conseil de l'Europe et l'Union Européenne. Toute décision européenne doit être traduite en droit national afin de pouvoir être appliquée.

Le fonctionnement des instances impliquées au niveau européen dans l'élaboration des textes (qu'il s'agisse de recommandations ou de réglementations) est succinctement présenté sur la Figure 1.3.

FIG. 1.3 – Mode de fonctionnement et interactions des instances européennes en charge de la protection animale (adapté d'après Guéméné et Faure 2004)

Les conventions et recommandations du Conseil de l'Europe

Le Conseil de l'Europe, fondé en 1949 et regroupant désormais 47 états membres, constitue une force de proposition en suggérant aux états des mesures pour assurer, entre autres, le bien-être des animaux. Il a commencé à se préoccuper de la protection animale dans les années 60. Le Conseil de l'Europe élabore des *conventions*, qui devront être ratifiées puis appliquées par les états signataires. Ces conventions donnent les principes généraux à adopter, élaborés par un comité d'experts. Elles peuvent être à l'origine de *recommandations*, qui, elles, n'ont pas besoin d'être ratifiées. Il s'agit de textes plus techniques et basés sur des connaissances scientifiques. Les recommandations peuvent d'ailleurs être révisées en cas d'apport de nouvelles connaissances scientifiques et techniques.

En tout, cinq conventions européennes traitent des principes éthiques selon lesquels les animaux devraient être utilisés par les hommes (cf. Tableau 1.1). Trois d'entre elles concernent les animaux de ferme (en noir dans le tableau), alors que les deux autres traitent des animaux utilisés à des fins expérimentales et des animaux de compagnie (en gris dans le tableau).

TAB. 1.1 – Liste des cinq conventions européennes concernant la protection des animaux *(Source : site Internet du Conseil de l'Europe[7])*

Convention européenne	**Ouverture du traité**	**Entrée en vigueur**
Convention européenne sur la protection des animaux en transport international	13/12/1968	20/02/1971
Révisée	*06/11/2003*	*14/03/2006*
Convention européenne sur la protection des animaux dans les élevages	10/03/1976	10/09/1978
Convention européenne sur la protection des animaux d'abattage	10/05/1979	11/06/1982
Convention européenne sur la protection des animaux vertébrés utilisés à des fins expérimentales ou à d'autres fins scientifiques	18/03/1986	01/01/1991
Convention européenne pour la protection des animaux de compagnie	13/11/1987	01/05/1992

Parmi les conventions concernant les animaux de ferme, la première date de 1968 et traite de la protection des animaux durant les transports internationaux. Cette convention a débouché en 1987-1990 sur des recommandations pour le transport des chevaux,

[7]http ://conventions.coe.int/Treaty/Commun/ListeTraites.asp ?CM=8&CL=FRE

des porcs, des bovins, de la volaille, des moutons et des chèvres (Veissier et al. Accepté). Cette convention a été révisée en 2003.

La Convention sur la protection des animaux dans les élevages date de 1976. Elle prévoit que tout animal doit bénéficier d'un logement, d'une alimentation et de soins appropriés à ses besoins physiologiques et éthologiques, conformément à l'expérience acquise et aux connaissances scientifiques (Fabre 1995). Il en découle des recommandations concernant la plupart des espèces élevées (Veissier et al. Accepté) : bovins (1988 et 1992 pour les veaux, en cours de révision) ; chèvres et moutons (1992) ; volailles (1986 et 1995) ; porcs (1986 et 2005) ; ratites (1997) ; canards, oies et animaux à fourrure (1999) ; dindes (2001) ; poissons (2006) ; lapins (en préparation). Chacune de ces recommandations se base sur les caractéristiques biologiques de chaque espèce pour établir des exigences minimales permettant de couvrir les besoins des animaux.

Enfin, la troisième convention concerne la protection des animaux d'abattage (1979) et vise à améliorer dans les abattoirs les conditions de détention et d'abattage des équins, ruminants, porcs, lapins et volailles (Veissier et al. Accepté).

Dès qu'une convention ou une recommandation est adoptée, chaque état signataire se doit de mettre en place les mesures permettant de suivre les recommandations. En fonction du pays, les conventions européennes sont soit directement applicables, soit nécessitent d'être transcrites dans les textes de lois du pays (Veissier et al. Accepté).

Ainsi, la convention sur la protection des animaux dans les élevages a été ouverte à la signature le 10 mars 1976 et a été ratifiée par la France le 10 janvier 1978 (entrée en vigueur le 10 septembre de la même année). Cette convention a été publiée dans le Décret 78-1085[8] du 02 novembre 1978 du droit français. Ainsi, elle est désormais opposable et invocable par les justiciables français. De plus, cette convention a été traduite en droit européen par la Directive 98/58/CE.

Les directives de l'Union Européenne

Dès les années 70, l'Union Européenne (anciennement CEE) a pris des mesures pour protéger les animaux, et ce principalement afin de limiter, au sein du marché communautaire, les distorsions de concurrence qui pourraient être causées par des disparités entre des lois nationales concernant la protection des animaux (Veissier et al. Accepté).

Avec le traité d'Amsterdam (1997), les animaux sont désormais considérés au niveau européen comme des être sensibles et qui, par conséquent, doivent être protégés.

[8]Publié au Journal Officiel "Lois et Décrets" du 18/11/1978, p. 3879 (Source : site Internet www.legifrance.gouv.fr)

La Commission Européenne (pouvoir exécutif) fait office de force de proposition de lois auprès du Conseil des Ministres (pouvoir législatif). Au sein de la Commission Européenne, la Direction Générale "Santé et protection des Consommateurs" (DG-SANCO) a pour mission de contribuer à améliorer la santé, la sécurité et la confiance des citoyens européens. C'est elle qui est chargée de la protection des animaux. Quand la DG-SANCO décide de promouvoir une réglementation au sujet de la protection des animaux, elle fait appel à un comité d'experts (le *Scientific Committee on Animal Health and Animal Welfare*) appartenant à l'Autorité Européenne de Sécurité des Aliments (AESA). Ce comité scientifique crée alors un groupe *ad-hoc* d'experts qui vont produire un rapport scientifique et formuler des recommandations. La DG-SANCO se base sur les conclusions des experts pour décider ou non de proposer une directive. Le Comité Économique et Social Européen (CESE), composé de représentants des principaux acteurs socio-économiques européens, peut alors être consulté pour donner son avis sur le projet de Directive avant qu'il ne soit soumis au Conseil des Ministres. Ce n'est qu'une fois approuvé par le Conseil des Ministres que le projet de directive devient une directive européenne (Veissier et al. Accepté). Ensuite, les directives européennes se doivent d'être transcrites en droit national par tous les pays membres de l'Union Européenne.

À ce jour, des directives européennes ont été élaborées pour tous les animaux, elles concernent leurs conditions d'abattage (Directive 74/577/CE remplacée par la Directive 93/119/CEE), de transport (plusieurs directives depuis 1977) et d'élevage (directives 78/923/CE et 98/58/CE). Il existe aussi des directives spécifiques à certaines espèces, telles que les poules pondeuses (Directive 88/166/CE remplacée par la Directive 1999/74/CE), les veaux (Directive 91/629/CE amendée par la Directive 97/2/CE), ou les porcs (Directive 91/630/CE amendée par la Directive 2001/93/CE) (Veissier et al. Accepté).

Les vaches laitières en fermes ne sont protégées que par la Directive 98/58/CE[9].

Les bovins sont également protégés au moment de l'abattage et lors des transports (Directives 91/628/CEE et 95/29/CE ; règlement 411/98...).

Malgré toutes les recommandations et réglementations existantes concernant le bien-être des animaux d'élevage, il ressort de l'Eurobaromètre conduit par la Commission de l'Union Européenne en 2007 que 55% des personnes interrogées pensent que le bien-être et la protection des animaux n'est pas suffisamment pris en compte au sein de leur pays (European Commission 2007b). Les réglementations actuelles ne semblent donc pas être

[9]du Conseil du 20 juillet 1998, publiée au Journal Officiel n°L 221 du 08/08/1998 p. 0023 - 0027

perçues comme suffisantes par les citoyens, alors que bon nombre d'éleveurs se sentent sous la pression de ces mêmes réglementations.

De plus, 89% des personnes ayant répondu au questionnaire de l'Eurobaromètre de 2007 pensent que des standards de bien-être équivalents à ceux proposés en Europe devraient aussi être appliqués aux produits alimentaires importés de pays hors UE (European Commission 2007b). On touche alors le problème de la concurrence internationale. Les consommateurs ont déjà du mal à savoir comment sont élevés chez eux les animaux dont ils consomment les produits, comment savoir dans quelles conditions sont élevés les animaux dont la viande est importée en Europe ?

1.1.3.2 Les labels et certifications

Burgat et Dantzer (1997) soulignent que *"l'intensification de la production conduit une proportion croissante de consommateurs [...] à choisir des produits munis d'un label garantissant un mode d'élevage jugé plus respectueux du bien-être"*. Or, selon l'Eurobaromètre conduit en 2007 (European Commission 2007a), les consommateurs se sentent insuffisamment informés à cet égard. En effet, 51% d'entre eux considèrent que quand ils achètent des oeufs, du lait ou de la viande, il est difficile, voire impossible, d'identifier à partir de l'étiquetage actuellement fourni sur le produit si le système de production est respectueux du bien-être. Dans les conclusions de leur étude menée dans sept pays européens, Kjærnes et al. (2007) soulignent ainsi que la mise à disposition du consommateur d'une information suffisante et fiable sur les conditions d'élevage est un point crucial, pouvant même se révéler être très problématique dans plusieurs des pays étudiés.

De manière générale (c.-à-d. non spécifique au bien-être animal), pour compenser la distance croissante entre le produit et le consommateur, Fischler propose un étiquetage informatif visant soit à détailler la composition exacte du produit (comme ce qui est par exemple fait aux États-Unis) soit à communiquer sur l'origine du produit (point de vue répandu dans les pays du sud de l'Europe, tels que la France, l'Italie et l'Espagne, notamment du fait de la plus grande place accordée au plaisir et aux qualités organoleptiques du produit)(entretien rapporté dans *Autrement* 1997). Fischler note ainsi une montée en puissance des labels, ainsi qu'une politique d'unification et de garantie des signes au niveau européen. La labellisation et l'étiquetage des produits sont donc devenus des outils importants du marché européen et devraient être vus comme faisant partie intégrante de la communication entre les acteurs sociétaux, des producteurs et industriels aux consommateurs, avec ou sans intermédiaires, mais aussi entre les décideurs publics et les consommateurs (Gavinelli 2007).

La DG-SANCO de la Commission Européenne voit dans la mise en place d'un label et d'une information claire pour le marché et pour le consommateur, un moyen d'entrer dans un cycle vertueux où les consommateurs seraient de plus en plus demandeurs de produits respectueux du bien-être, demande qui serait alors transmise par la filière aux éleveurs, qui pourraient alors vendre mieux leurs produits et ainsi couvrir les coûts engendrés par l'amélioration des conditions de vie de leurs animaux (Gavinelli 2007). Il existe deux justifications principales à l'élaboration et la mise en place de labels associés à un étiquetage "bien-être" qui permettrait ainsi :

1. d'apporter une information au consommateur pour éclairer son choix ;
2. de distinguer des produits respectueux du bien-être animal, et ainsi faire face à la concurrence des produits importés de pays tiers.

Divers éléments peuvent être l'objet d'un label : l'origine du produit (avec par ex. le label *Viande bovine française*), la façon de produire (avec par ex. le label *Agriculture biologique*), la qualité (avec par ex. le *Label Rouge*), ou encore le bien-être des animaux (comme c'est par ex. le cas au Royaume-Uni avec les *Freedom Foods*). En France, le respect du bien-être animal est déjà mis en avant par des producteurs (ex. le porc *"Thierry Schweitzer"*), des transformateurs (ex. *Charal*), des distributeurs (ex. *Filière Qualité Carrefour*) et même des chaînes de restauration (ex. *McDonald's*).

Le CESE (2007) est favorable au développement d'une identification des produits offrant une garantie de bien-être animal plus élevée que la réglementation européenne. En effet, le CESE considère qu'un système basé sur le marché et sur des standards de bien-être clairement définis constitue une solution plus flexible, efficace et orientée vers le futur, que des critères imposés par la réglementation. Ainsi, pour le CESE le développement de labels, basés sur le volontariat et bénéficiant d'un mode de communication efficace (comme un logo combiné à un système de points) permettrait un meilleur développement de méthodes et systèmes d'élevage plus respectueux du bien-être des animaux que des obligations réglementaires et administratives. Au vu de l'Eurobaromètre 2007, cette opinion semble aussi partagée par les consommateurs puisque 74% des sondés considèrent qu'acheter des produits respectueux du bien-être pourrait avoir un impact positif sur le bien-être des animaux d'élevage (European Commission 2007a).

1. Des consommateurs mieux informés sont-ils prêts à payer plus cher pour des produits respectueux du bien-être ?

Fournir aux consommateurs une information additionnelle concernant le bien-être animal leur permettrait de mieux comprendre la valeur ajoutée du produit liée au respect

de standards de bien-être, et ainsi de mieux choisir les produits qu'ils vont acheter (Gavinelli 2007). Cependant, en dehors du manque d'information (auquel il est possible de remédier via la création de labels), le consommateur se retrouve face à la barrière du prix des produits (Kjærnes 2007). En effet, même s'il était intéressé par des produits plus respectueux du bien-être, quel prix serait-il prêt à y mettre ?

Lors de l'Eurobaromètre de 2007, 72% des personnes ayant répondu pensent que les éleveurs devraient être rémunérés pour les dépenses supplémentaires nécessaires pour assurer un meilleur niveau de bien-être à leurs animaux (European Commission 2007b). 0r, ceci implique un prix plus élevé pour les consommateurs à l'achat des produits.

Latouche (1999) a tenté de mesurer le consentement à payer de plus de mille français afin d'améliorer le bien-être des veaux de boucherie. Il en ressort que seulement 5% des personnes consultées n'étaient pas prêtes à payer un peu plus un produit provenant de veaux élevés dans de meilleures conditions. Il semble donc que les consommateurs soient prêts à payer un peu plus cher des produits respectueux du bien-être des animaux, ce qui a été confirmé lors de l'Eurobaromètre de 2007.

Cependant, les résultats de cet Eurobaromètre ont permis de mettre en évidence que si les consommateurs peuvent être prêts à acheter des produits respectueux du bien-être, c'est parce qu'ils les considèrent comme étant meilleurs pour leur santé et de meilleure qualité (European Commission 2007b). Ce résultat irait donc en faveur du développement de labels 'qualité' généralistes, incluant le bien-être animal. Le consommateur pourrait alors payer plus cher pour un produit dont il pourrait percevoir par lui-même qu'il est de meilleure qualité.

Kyprianou (2007), Commissaire européen pour la Santé, reste cependant optimiste et voit en Europe des opportunités pour la mise en place de systèmes de certification dans lesquels à la fois les consommateurs et les producteurs seraient gagnants.

2. Développer un label "bien-être" européen : une volonté de l'Union Européenne pour faire face à la concurrence internationale

Dans un contexte d'augmentation des échanges internationaux en produits carnés (cf. Figure 1.4), l'Union Européenne voit dans un label "bien-être" un moyen de différencier ses produits et ainsi de les protéger partiellement de la concurrence internationale. En effet, il ne faut pas oublier que les éleveurs européens doivent rester compétitifs sur le plan international, malgré des coûts de production plus élevés en partie liés au respect de standards de bien-être[10] (Kyprianou 2007).

[10]les coûts varient avec plusieurs facteurs, tels que le coût de la main-d'œuvre ou celui de la construction et de l'entretien des bâtiments

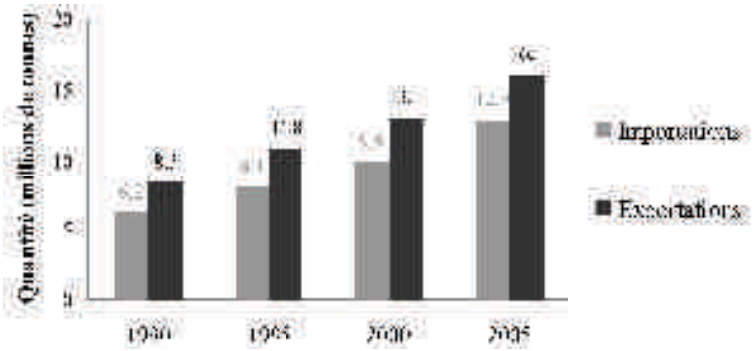


FIG. 1.4 – Évolution de la quantité de viande importée et exportée par l'Union Européenne, de 1995 à 2005 (Source : banque de données de la FAO[12])

Gavinelli (2007) considère que la création d'un label "bien-être" permettrait d'instaurer un environnement de marché procompétitif[13]. De même, Kyprianou (2007) souligne que la demande pour des produits issus de systèmes durables (incluant donc le respect du bien-être animal) ne se limite pas à l'Europe et se retrouve aussi dans des pays tels que l'Australie et les Etats-Unis. Il y voit une preuve de l'avantage compétitif de produits respectueux du bien-être au sein du secteur économique de l'alimentation. Il rappelle même que la Commission Européenne est actuellement en négociation avec l'OMC et en discussion avec l'OIE pour la mise en place d'accords bilatéraux avec des pays en voie de développement (Kyprianou 2007 ; Commission of the European Communities 2002).

À l'heure actuelle, il reste difficile d'utiliser le bien-être animal comme argument de différenciation des produits au sein de l'Organisation Mondiale du Commerce (OMC) et des accords du GATT[14]. En effet, il n'est pas possible au sein de l'OMC de se servir de mesures n'affectant pas, ou de façon négligeable, le produit final (ce qui est le cas des mesures visant à améliorer le bien-être animal) comme moyen de limiter le marché. Malgré des exceptions figurant dans les accords du GATT permettant, dans certaines circonstances, d'utiliser des mesures restrictives, et auxquelles le respect du bien-être animal peut se rattacher, Thiermann et Babcock (2005) pensent qu'il est actuellement impossible pour un gouvernement d'interdire l'importation de produits qui ne respecteraient pas les standards de bien-être de son pays.

[12]Organisation des Nations Unies pour l'Alimentation et l'Agriculture, sur le site Internet FAOSTAT http ://faostat.fao.org/site/343/default.aspx

[13]Le gain de bien-être de la collectivité est le résultat de la concurrence accrue provoquée par l'échange international. C'est l'effet procompétitif ou proconcurrentiel du commerce (définition donnée par Denis Delgay-Troïse dans l'un de ses cours à l'Université de Rennes 1 en Licence *Sciences économiques*)

[14]General Agreement on Tariffs and Trade

En revanche, Thiermann et Babcock (2005) voient le développement de traités bilatéraux ou multilatéraux et d'accords commerciaux basés sur l'utilisation de recommandations internationales en matière de bien-être animal comme une excellente alternative. Ils y voient un moyen d'obtenir des résultats similaires à ceux qui pourraient être obtenus en cas d'autorisation par l'OMC de se baser sur le bien-être animal pour restreindre le marché. Pour ce faire, il est donc nécessaire d'établir au niveau international des recommandations et/ou directives concernant le bien-être animal, tâche actuellement en cours de réalisation au sein de l'OIE.

Tout ceci va dans le sens du développement d'un label Européen (accompagné d'un logo) qui assurerait aux consommateurs que le produit qu'ils achètent est issu d'un système de production respectant les exigences européennes en matière de bien-être animal (Kyprianou 2007). Ce label facilement identifiable permettrait de différencier un produit importé issu d'un élevage ne respectant pas un niveau de bien-être aussi élevé que celui d'un produit issu d'un élevage européen (possédant le logo).

Les distributeurs : des acteurs essentiels pour la mise en place de labels

Afin de pouvoir améliorer le bien-être des animaux d'élevage à travers le marché plutôt qu'en utilisant les obligations réglementaires, la coopération des distributeurs est généralement considérée comme indispensable (Ingenbleek 2007). En effet, par le choix des produits que la grande distribution propose à ses clients, elle influence fortement le comportement des consommateurs.

Ingenbleek (2007) souligne la concentration croissante des distributeurs et par là même, une augmentation de la compétition. En parallèle, les marques de distributeurs (MDD) sont de plus en plus nombreuses (chaque distributeur pouvant en créer une ou plusieurs) et de plus en plus fortes[15] (Ingenbleek 2007). Du fait de la forte compétition existant entre les acteurs de l'agro-alimentaire (industriels et distributeurs), ils ont tendance chacun à utiliser leurs propres marques et labels de qualité (European Economic and Social Committee 2007), sur lesquels ils ont généralement une marge plus grande (Ingenbleek 2007).

Kjærnes et al. (2007) ont découvert auprès des distributeurs européens une grande diversité de stratégies commerciales faisant référence au bien-être animal. Alors qu'en Norvège et en Hongrie le bien-être animal ne semble pas être actuellement pris en compte dans de telles stratégies, il sert par contre de support à des stratégies commer-

[15] En 2006, les MDD représentaient en France 35% de parts de marché en volume et 25% en valeur ; et depuis, les MDD ont continué de prendre des parts de marché en France (Bitoun 2007)

ciales bien implantées au Royaume-Uni et aux Pays-Bas, et apparaît aussi dans des démarches qualités plus générales en France. Pour la plupart des distributeurs consultés dans le cadre de cette étude, le bien-être animal est considéré comme faisant partie d'une stratégie commerciale visant à illustrer la préoccupation du distributeur vis-à-vis de l'environnement et de la durabilité, tout en lui permettant de promouvoir son image de marque. Tout ceci s'inscrit dans le cadre d'une augmentation de la segmentation du marché visant à proposer différentes qualités et différents prix pour une même catégorie de produits.

Cependant, les distributeurs ne sont pas favorables à un label uniquement axé sur le bien-être animal en l'absence d'une forte volonté de consommation éthique et sans différence de qualité perceptible par le consommateur. Ainsi, la majorité des distributeurs considèrent que le bien-être animal doit être inclus au sein d'une démarche qualité générale du produit (avec notamment la qualité gustative du produit), ce qui est en accord avec les points de vue des consommateurs (voir p. 11).

Au cours de cette enquête auprès des distributeurs, il a aussi été constaté que le développement des marques de distributeurs a conduit à une diminution de l'utilisation sur les produits des labels standardisés. Ainsi, des standards peuvent être requis par les distributeurs quand ils achètent les produits, mais n'apparaissent pas forcément sur les produits vendus aux consommateurs.

Malgré tout, Ingenbleek (2007) voit plusieurs raisons pour lesquelles les distributeurs auraient intérêt à jouer le jeu de labels "respectueux du bien-être". En proposant de tels produits ils pourraient alors :

- attirer et fidéliser une clientèle orientée vers un développement plus éthique de l'élevage (agriculture biologique et respect du bien-être) ;
- augmenter leurs marges avec une clientèle en partie prête à payer plus cher des des produits de meilleure qualité et à valeur ajoutée "bien-être" ;
- améliorer leur image de marque en les rendant aux yeux de leurs clients socialement et éthiquement responsables. La possibilité, pour les distributeurs, d'inclure le respect du bien-être animal dans leur stratégie commerciale afin d'améliorer leur image auprès de leurs clients est également citée par Kjærnes et al. (2007).

En résumé, trois types de labels différents semblent être envisageables :

1. un label européen garantissant le respect de la législation en vigueur dans l'Union Européenne, qui pourrait alors être utilisé pour indiquer aux consommateurs que le produit européen qu'ils achètent est issu d'un système de production respectant

le niveau d'exigence communautaire en matière de bien-être animal ;

2. un label "qualité" généraliste regroupant différents attributs : le bien-être animal, la qualité gustative... De tels labels seraient alors bien en accord avec les attentes actuelles des consommateurs et des distributeurs ;
3. un label spécifique au bien-être animal assurant un excellent niveau de bien-être, un tel label correspondrait à un marché de niche constitué de consommateurs adhérant à une démarche éthique orientée vers le respect des animaux. Ce label pourrait alors être utilisé par la grande distribution pour attirer ce type de clientèle.

À terme, la mise en place de systèmes de certification basés, au moins en partie, sur le respect du bien-être animal devrait permettre d'améliorer le niveau de bien-être dans les élevages européens grâce à des stratégies déterminées par les contraintes du marché (Bock et van Leeuwen 2005).

Pour pouvoir mettre en place des systèmes de certification, il est nécessaire de pouvoir fournir une évaluation globale formalisée du bien-être des animaux (Fraser 2003 ; Gavinelli 2007). En effet, un label est censé assurer un niveau de bien-être défini. Le consommateur achetant une tel produit s'attend donc à ce que le minimum requis par le label soit atteint par la ferme dont est issu le produit en question. Des procédures de contrôle sur le terrain doivent donc être mises en place. Ces procédures reposent sur des systèmes d'évaluation du bien-être qui se doivent d'être explicites et transparents, et ce afin que le consommateur ait confiance dans le label considéré. Cette question de la confiance est un point essentiel dans la réussite commerciale d'un label. Les consommateurs doivent avoir confiance en ceux qui fournissent et contrôlent l'information qui leur est fournie (Kjaernes et al. 2007).

Par ailleurs, Gavinelli (2007) souligne l'intérêt de la création d'un "Standard de Qualité Européen" dédié au bien-être face à la multitude de labels et standards existants, parfois même concurrents et sources de confusion pour les consommateurs.

1.1.4 Évaluer le bien-être animal : un travail interdisciplinaire

Comme le soulignent Lund et al. (2006), le concept de bien-être a plusieurs dimensions à la fois scientifique, éthique, économique et politique (comme cela est illustré sur la Figure 1.5). Le Tableau 1.2 illustre la multitude d'acteurs pouvant être impliqués dans l'amélioration du bien-être animal en France. La première section de ce chapitre introductif a brossé rapidement les attentes sociétales, à la fois en termes d'éthique et

d'économie. Nous avons également abordé la prise en compte des attentes sociétales au travers de la réglementation et de la certification de produits. Tous ces aspects dépassent la simple dimension scientifique (sous-entendu : en sciences animales) de l'évaluation du bien-être. Lund et al. (2006) recommandent alors pour traiter du bien-être d'adopter une approche intégrée qui utilise des compétences conceptuelles et méthodologiques provenant de plusieurs disciplines. Les chercheurs en sciences sociales et économiques sont particulièrement impliqués dans l'identification et la compréhension des attentes sociétales et l'analyse des interactions entre acteurs. Les chercheurs en sciences animales, quant à eux, fournissent des outils pour évaluer le bien-être des animaux, comparant des conditions de vie (ou d'abattage) des animaux à l'aide de ces outils et vont souvent jusqu'à proposer des solutions permettant d'améliorer le bien-être des animaux (Veissier et al. 2007). Ils s'appuient généralement sur plusieurs disciplines : éthologie, santé animale, physiologie...

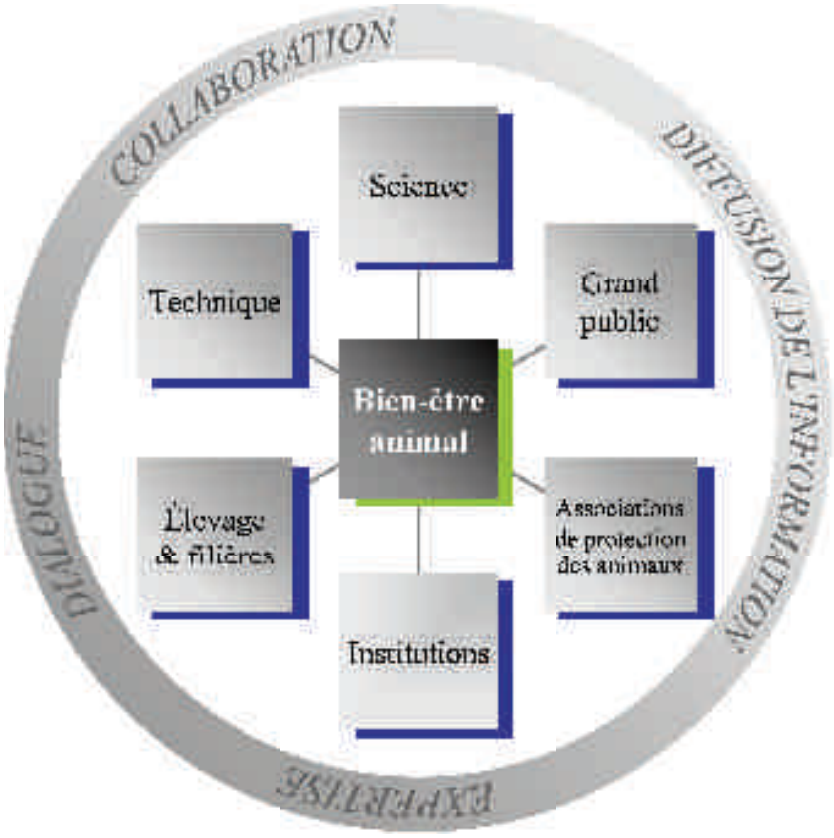


FIG. 1.5 – Le bien-être animal : au centre de nombreux questionnements

TAB. 1.2 – Qui sont les acteurs de la réflexion sur le bien-être en France ? *(Cette liste n'est pas exhaustive mais donne un idée de la diversité des acteurs pouvant être concernés)*

Domaines	**Acteurs**
Science	- Réseau AGRI Bien-Etre - INRA (avec notamment le travail des équipes ACS [Clermont-Ferrand], PRC [Tours] et SENAH [Rennes]) - Universités (notamment celles de Rennes 1, Paris XIII, Université François Rabelais de Tours...) - Écoles d'agronomie (notamment l'AgroParisTech) et de médecine vétérinaire
Technique	- Institut de l'Elevage - IFIP (Institut du Porc) - ITAVI (Institut Technique de l'Aviculture) - DSV (Direction des Services Vétérinaires) - AFSSA (Agence Française de Sécurité Sanitaire des Aliments)
Élevage & filières	- INTERBEV (Association Nationale Interprofessionnelle du Bétail et de la Viande) - OFIVAL (Office de l'Élevage) - les groupements de producteurs - les UPRA (Unités Nationales de Sélection et de Promotion des races animales) - CIV (Centre d'Information des Viandes) - Chambres Départementales et Régionales de l'Agriculture
Institutions	- Ministère de l'Agriculture et de la Pêche (avec notamment le Bureau de la Protection Animale de la Direction Générale de l'Alimentation) - Ministère de la Recherche (avec notamment l'Agence Nationale de la Recherche) - Union Européenne - Conseil de l'Europe
Associations de protection des animaux	- PMAF (La Protection Mondiale des Animaux de Ferme) - OABA (Œuvre d'Assistance aux Bêtes d'Abattoir) - SNDA (Société Nationale pour la Défense des Animaux) - LFDA (Fondation Ligue Française des Droits de l'Animal) - World Society for the Protection of Animals (WSPA) - Confédération Nationale des SPA (Sociétés de Protection des Animaux) de France - Pro Anima - la Fondation Brigitte Bardot - PETA France (Pour un Traitement Éthique des Animaux)
Grand public	- INC (Institut national de la consommation) - UFC-Que Choisir (Union Fédérale des Consommateurs)

Évaluer le bien-être nécessite donc la mise en place d'un programme de recherche impliquant des interactions entre différentes disciplines. Il existe plusieurs niveaux d'interactions entre chercheurs de disciplines différentes travaillant sur un projet commun (Rosenfield 1992 cité par Lund et al. 2006 ; Lavoie et Fougeyrollas 1998) :

- **Multidisciplinaire** (ou pluridisciplinaire) qualifie un travail mené en parallèle ou de manière séquentielle par des disciplines différentes. Ainsi, les membres d'une équipe multidisciplinaire sont peu en interaction entre eux. La compréhension de la situation se situe davantage sur des visions analytiques et parallèles.
- **Interdisciplinaire** qualifie un travail qui nécessite l'établissement de relations entre plusieurs disciplines. Ainsi, les membres d'une équipe interdisciplinaire interagissent beaucoup, ce qui permet une compréhension intégrée des problèmes et des besoins. Les échanges d'informations résultant des nombreuses interactions permettent un enrichissement mutuel des connaissances.
- **Transdisciplinaire** qualifie un travail dépassant les cloisonnements entre les disciplines, en se basant sur des concepts, théories et approches partagés. Les membres d'une équipe transdisciplinaire, provenant d'horizons différents, mettent au point ensemble une méthodologie commune.

Évaluer le bien-être semble donc être une thématique dépassant la multidisciplinarité. Il s'agirait plutôt d'une thématique interdisciplinaire, au sens où son traitement nécessite, à certaines étapes clefs, des interactions fortes entre chercheurs de disciplines différentes : sciences animales, sciences sociales, mais aussi en recherche opérationnelle (comme nous le verrons par la suite). On pourrait même voir l'évaluation du bien-être comme un travail transdisciplinaire, au sens où elle nécessite la mise en place d'un projet reposant sur une base commune construite ensemble par des chercheurs de différentes disciplines, ce qui est le cas dans le projet européen Welfare Quality® (cf. section suivante). Selon Rosenfield (1992), une approche transdisciplinaire permettrait de fournir un cadre théorique systémique et global pour la définition et l'analyse des facteurs sociaux, économiques, politiques, environnementaux et humains de certaines thématiques, telles que la santé et le bien-être des hommes (thématique qui la préoccupe).

Des interactions entre les chercheurs (des différentes disciplines) et d'autres acteurs (représentants de consommateurs, distributeurs, éleveurs, institutions...) semblent nécessaires du fait que l'évaluation du bien-être animal implique, en plus de choix scientifiques, des choix politiques, économiques et éthiques.

1.2 Le projet Welfare Quality®

Welfare Quality® (titre complet : *Integration of animal welfare in the food quality chain : from public concern to improved welfare and transparent quality*) est un projet intégré européen cofinancé par la Commission de l'Union Européenne. Il fait partie de la priorité "Qualité et sécurité alimentaire" du 6e PCRD[16]. Il a débuté le 1er mai 2004, pour une durée de 5 ans (jusqu'au 30 avril 2009). Son coût total est de 17 millions d'euros, dont 14,6 millions d'euros sont couverts par l'Union Européenne.

1.2.1 Objectifs du projet

Welfare Quality® a pour but de proposer des outils pour la mise en place de systèmes d'information auprès des consommateurs en matière de bien-être animal. Pour ce faire, le projet a pour ambition de :

1. Développer des mesures permettant d'apprécier le bien-être animal ;
2. Développer un système standardisé européen pour l'évaluation globale du bien-être animal (en ferme ou à l'abattoir) et pour l'information destinée aux acheteurs et consommateurs ;
3. Mettre au point des stratégies pour remédier à des problèmes de bien-être ;
4. Intégrer en réseau, à l'échelle européenne, l'expertise de spécialistes du domaine pluridisciplinaire qu'est le bien-être animal.

Welfare Quality® rassemble de nombreux groupes d'experts (scientifiques ou non) à travers l'Europe afin de répondre aux objectifs de recherche sous-jacents :

- analyser les attentes des consommateurs en matière d'information sur le bien-être des animaux, identifier les stratégies de communication efficaces ;
- étudier le marché actuel et potentiel pour des produits respectueux du bien-être animal, les systèmes d'information et les modes d'inspection ;
- développer des systèmes fiables et harmonisés d'évaluation du bien-être des animaux en ferme ;
- proposer un standard d'information auprès des consommateurs ;

[16] Le 6e Programme Cadre de Recherche et Développement (PCRD) de l'Union Européenne prévoit le financement de projets de recherche pour un montant de 17,5 milliards d'euros sur la période 2002-2006. Les priorités thématiques sont les suivantes : Sciences de la vie, génomique, biotechnologies ; Technologies de la société de l'information ; Nanotechnologies ; Aéronautique et espace ; Qualité et sécurité alimentaire ; Développement durable, changement climatique, écosystèmes ; Citoyens et gouvernance.

- proposer des solutions pratiques pour améliorer le bien-être des animaux en ferme, et par là-même aider les éleveurs à rejoindre d'éventuels programmes de certification en bien-être animal ;
- concevoir et évaluer le transfert à la pratique du système d'évaluation et du standard d'information, et des solutions pratiques identifiées dans le projet.

Le projet est centré sur trois espèces animales : les bovins (vaches laitières, veaux de boucherie et taurillons à l'engrais), les porcs (porcs à l'engrais, truies et porcelets) et les volailles (poulets de chair et poules pondeuses).

Les attentes sociétales et les marchés sont étudiés dans sept pays européens : la France, la Hongrie, l'Italie, la Norvège, les Pays-Bas, le Royaume-Uni, la Suède.

Le projet est coordonné par trois organismes de recherche : Wageningen University and Research Centre (Pays-Bas)[17], l'Université de Cardiff (Royaume-Uni), et l'INRA[18] (France). Welfare Quality® réunit 44 instituts et universités qui apportent ainsi l'expertise nécessaire, à la fois en sciences sociales (ou humaines) et en sciences animales. Les chercheurs sont ainsi basés dans 13 pays européens (Allemagne, Autriche, Belgique, Danemark, Espagne, France, Irlande, Italie, Norvège, Pays-Bas, République Tchèque, Royaume-Uni, Suède). Quatre pays d'Amérique latine (Brésil, Chili, Mexique, Uruguay) ont rejoint le projet en 2007 (cf. Figure 1.6).

[17]suite à la mobilité du porteur du projet (le Pr. Harry Blokhuis), initialement rattaché à l'Université de Wageningen, la Sveriges lantbruksuniversitet (Suède) est devenue un des organismes coordinateurs du projet

[18]Institut National de la Recherche Agronomique

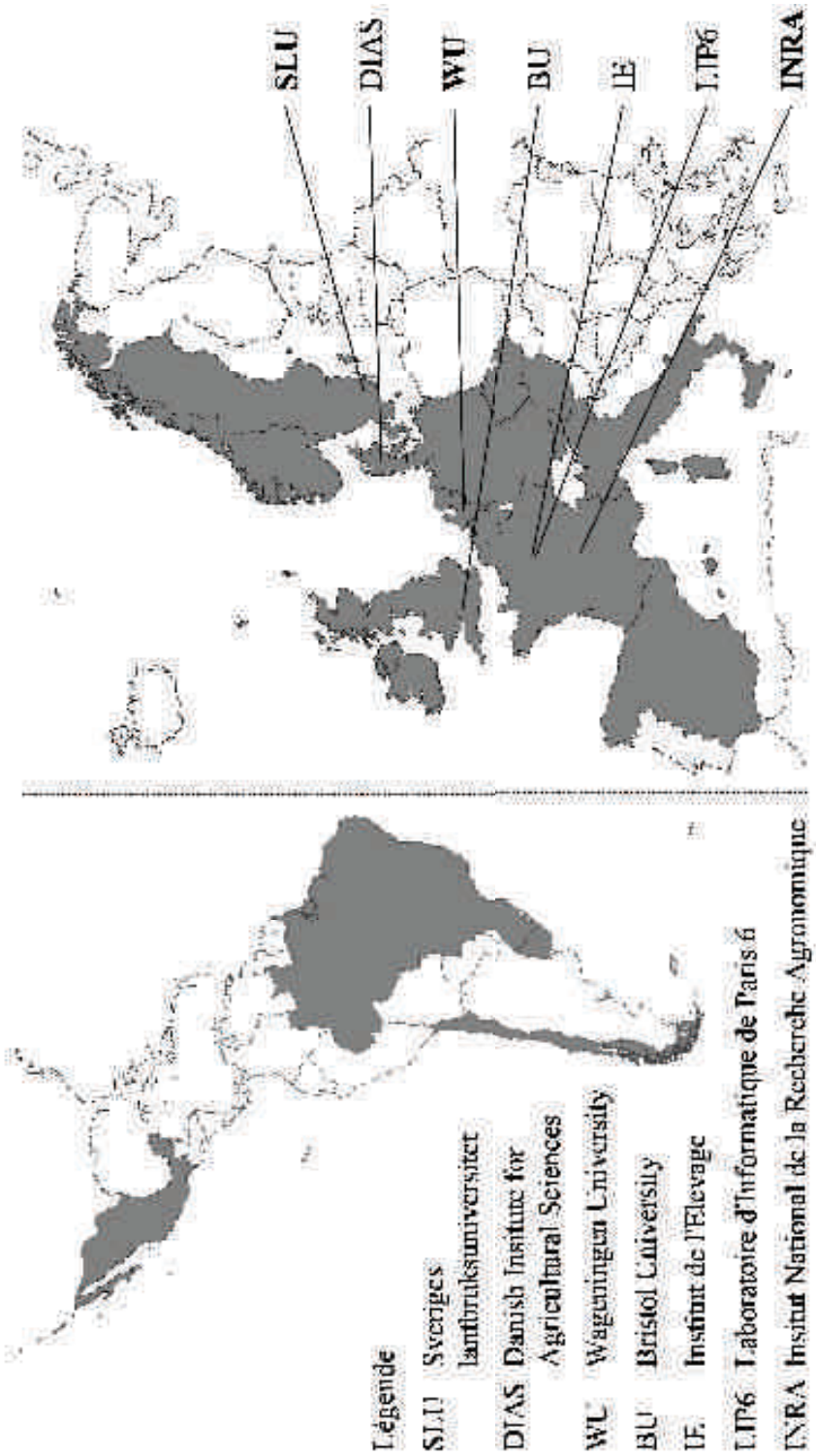


FIG. 1.6 – Liste des pays partenaires du projet Welfare Quality® ainsi que des organismes de recherche impliqués dans le WP2.3 (les organismes coordinateurs du projet apparaissent en gras)

1.2.2 Organisation du projet

Welfare Quality® est organisé en quatre sous-projets de recherche[19] (Veissier et al. 2005) :

- le **sous-projet 1** étudie les attentes sociétales et les pratiques des consommateurs, des distributeurs et des producteurs en relation avec le bien-être animal, il évalue la possibilité de mettre en place des systèmes de certification basés sur le respect du bien-être des animaux ;
- le **sous-projet 2** développe une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir ;
- le **sous-projet 3** met au point et évalue des solutions pratiques permettant de remédier à des problèmes de bien-être clairement identifiés ;
- le **sous-projet 4** générera les standards d'évaluation et d'information, transférera les connaissances aux utilisateurs, évaluera l'acceptabilité du système par les utilisateurs.

Cette organisation permet de faciliter la mise en relation des différents domaines d'expertise tout en optimisant le travail et le développement des compétences multidisciplinaires nécessaires à la réalisation des différents objectifs scientifiques et techniques définis dans le projet.

Par ailleurs, des activités transversales (dialogue science-société, formation et mobilité des jeunes chercheurs, transfert des connaissances et qualité scientifique) renforcent les travaux des quatre sous-projets de recherche. Un bureau scientifique a également été mis en place afin d'évaluer les propositions de recherche (pour les activités qui n'ont pas été attribuées à un partenaire donné en début de projet) et la qualité de certains livrables clés du projet. Il est constitué de plusieurs scientifiques reconnus sur le plan international et non-partenaires du projet. Un comité conseil, lui aussi extérieur au projet, est constitué de représentants d'associations de consommateurs, distributeurs, producteurs ou protecteurs des animaux, ainsi que de l'Union Européenne et de l'Office International des Épizooties[20]. Il donne son avis sur les choix stratégiques en cours de projet. Ce comité conseil permet de s'assurer que le travail réalisé dans le projet est en accord avec les attentes des principaux utilisateurs potentiels du système d'évaluation/conseil/information visé par Welfare Quality®.

[19] les quatre sous-projets de recherche sont présentés dans cette section, mais de plus amples détails sont fournis dans l'Annexe A du manuscrit

[20] La composition de ce Comité Conseil est détaillée en Annexe B.

1.2.2.1 Sous-projet 1 : Attentes des consommateurs, des distributeurs et des éleveurs vis-à-vis du bien-être animal et de produits respectueux du bien-être des animaux dont ils sont issus

L'objectif de ce sous-projet est de mettre en évidence les attitudes et les pratiques des consommateurs, des distributeurs et des éleveurs vis-à-vis du bien-être animal, et ainsi de pouvoir évaluer dans quelle mesure de nouvelles stratégies d'évaluation, d'amélioration et d'information sur le bien-être pourraient être mises en application avec succès.

Ce sous-projet 1 s'organise en trois groupes de tâches (WP[21]) concernant respectivement les consommateurs, les distributeurs et les éleveurs. Chacun de ces WP cherche à identifier comment les consommateurs, les distributeurs et les éleveurs perçoivent le bien-être animal, et à déterminer comment il serait possible de les persuader d'adopter des standards de bien-être plus rigoureux.

1.2.2.2 Sous-projet 2 : Développement d'une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir

L'objectif global du sous-projet 2 et de développer une méthodologie standardisée d'évaluation en routine du bien-être chez les bovins, les porcs et les volailles, de la ferme à l'abattoir. Ce système d'évaluation est basé essentiellement sur des mesures prises directement sur les animaux, mais aussi sur l'environnement (ressources et gestion), notamment pour mettre en évidence d'éventuelles causes de mal-être. Le développement de ce système d'évaluation inclut le fait de déterminer quelle procédure serait la plus appropriée pour synthétiser les différentes mesures effectuées sur le terrain et pouvoir ainsi produire une évaluation globale du bien-être.

Ce sous-projet est subdivisé en quatre groupes de tâches étroitement liés visant tout d'abord à définir le schéma général pour l'évaluation du bien-être. Puis, pour chaque espèce animale concernée par le projet, une liste de mesures valides, répétables, faisables et permettant d'évaluer tous les aspects du bien-être sera établie. L'information produite par ces mesures sera ensuite synthétisée au niveau de la ferme selon une méthodologie précise d'évaluation. Enfin, les protocoles de visites en ferme seront testés sur des élevages européens afin de s'assurer de la faisabilité de la procédure d'évaluation.

[21] Workpackage

1.2.2.3 Sous-projet 3 : Développement de stratégies permettant d'améliorer le bien-être des animaux d'élevage

L'objectif principal du sous-projet 3 est de développer, puis tester, des stratégies pratiques afin d'améliorer le bien-être des animaux en ferme. Ces stratégies potentielles peuvent inclure à la fois des actions sur l'environnement des animaux et des approches génétiques, dans le but de réduire l'occurrence de certains comportements ou états physiologiques dangereux pour les animaux, de fournir un environnement sain mais stimulant, et d'améliorer les relations homme-animal en apportant des conseils aux éleveurs. Ces stratégies correctives sont appliquées aux situations identifiées comme étant sources d'intérêt pour les consommateurs ou causes de mal-être pour les animaux. À terme, ces stratégies devraient permettre d'aider les éleveurs à atteindre un bon niveau de bien-être sur leur ferme. Pour ce faire, les solutions proposées devront donc être faciles à appliquer et efficaces, tout en satisfaisant des exigences économiques afin d'être viables.

Le sous-projet 3 est divisé en six groupes de tâches, chacun abordant un problème de bien-être particulier : stress lié à la manipulation des animaux, traits phénotypiques dangereux, comportements sources de blessures, boiteries, mortalité néo-natale, et stress social. Ces différents points couvrent les principaux problèmes perçus comme étant importants par les consommateurs européens, et offrant des marges de progrès importantes.

1.2.2.4 Sous-projet 4 : Vers la mise en application d'une procédure standardisée d'évaluation et d'information du bien-être ainsi que vers celle de stratégies d'amélioration du bien-être

Les principaux objectifs de ce sous-projet sont de générer des standards pour l'évaluation du bien-être et la transmission de l'information, de communiquer sur les connaissances acquises durant le projet, de mettre en place des projets de recherche sur des thématiques similaires, et de produire des recommandations sur la (ou les) meilleure(s) stratégie(s) pour la mise en application du système d'évaluation et d'information et sur des stratégies correctrices proposées par le projet.

Ces différents objectifs seront traités par un ensemble de sept groupes de tâches.

Alors que tous les autres sous-projets se déroulent en parallèle et sont presque terminés, le sous-projet 4 a démarré il y a juste quelques mois (cf. Figure 1.7). Il permettra de réunir les connaissances et productions des trois autres sous-projets, et ainsi de s'assurer que le système d'évaluation du bien-être, le système d'information lié au produit,

et les stratégies d'améliorations en élevage seront mis en oeuvre selon les mécanismes les plus appropriés.

L'organisation temporelle des quatre sous-projets au cours des cinq années de Welfare Quality® est présentée sur la Figure 1.7.

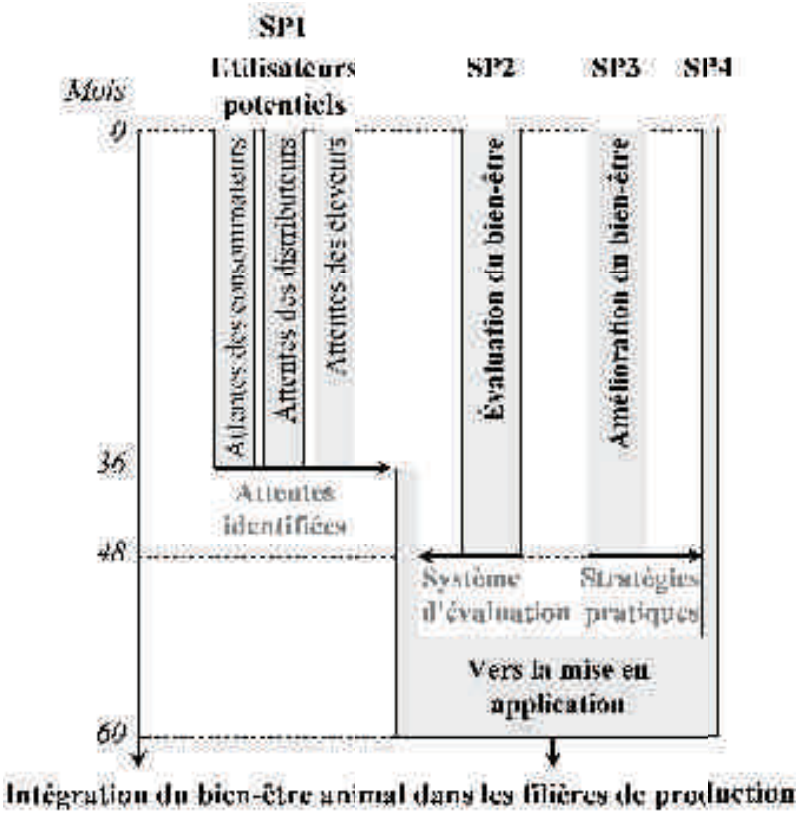


FIG. 1.7 – Organisation de Welfare Quality® : les quatre sous-projets de recherche

1.2.3 Schéma général du système visé par Welfare Quality®

Le schéma général du système visé par Welfare Quality® (cf. Figure 1.8) permet de répondre aux trois grandes finalités du projet : l'évaluation, le conseil et l'information. En effet, la procédure d'évaluation (en gris clair sur la figure) fournit les bases nécessaires à la fois au conseil et à l'information (en gris foncé). Welfare Quality® a pour but de fournir un outil dont les sorties pourront être utilisées par les utilisateurs futurs du système, soit pour proposer un diagnostic et des solutions aux éleveurs, transporteurs et abatteurs (*conseil*), ou pour informer les consommateurs ou distributeurs sur le niveau

de bien-être assuré aux animaux durant leur vie (*information*). Ces deux finalités du système dépendent donc de ses utilisateurs futurs.

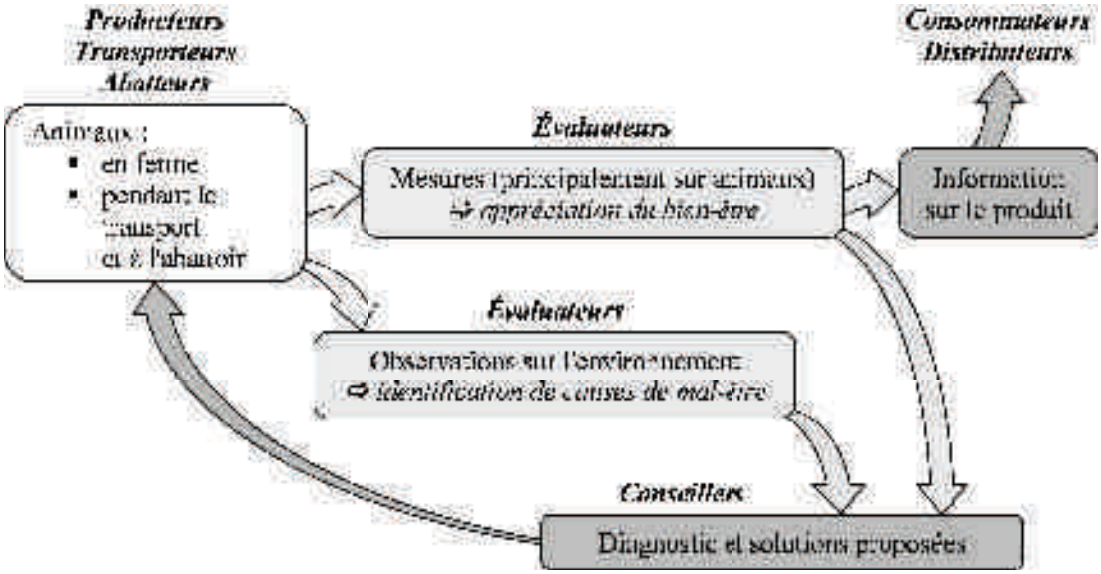


FIG. 1.8 – Schéma général du système visé par le projet Welfare Quality® - modifié d'après Veissier et al. (2005)

Le rôle du projet est donc de fournir toutes les informations nécessaires aux utilisateurs. Pour ce faire, le système d'évaluation repose sur la réalisation de nombreuses mesures, prises directement sur les animaux mais aussi prises sur les ressources disponibles et sur la manière dont les hommes (éleveurs, transporteurs et abatteurs) gèrent les animaux. La question de la nature des mesures à prendre en compte pour évaluer le bien-être animal porte souvent à discussion (voir par exemple Johnsen et al. 2001 ; Winckler et al. 2003). Trois écoles s'affrontent alors, entre ceux qui ne se basent quasiment que sur des mesures prises sur l'environnement et la gestion des animaux (ex. Bartussek 1999 ; Main et al. 2001), ceux qui, au contraire, pensent qu'il faut mettre l'accent sur les mesures prises directement sur les animaux (ex. Capdeville et Veissier 2001 ; Winckler et al. 2003), et enfin ceux qui recommandent d'associer les deux types de mesures (ex. Johnsen et al. 2001).

Les mesures basées sur les animaux consistent à analyser les comportements des animaux, leur santé, ou certains paramètres de production, afin d'évaluer comment les animaux réagissent à ce qui les entoure, alors que les mesures basées sur l'environnement consistent simplement à décrire les systèmes de logement et la gestion du troupeau par l'éleveur (Johnsen et al. 2001 ; Bennedsgaard et Thamsborg 2000).

Les mesures basées *sur l'environnement* peuvent être réalisées assez facilement et demandent peu de temps. De plus, elles sont généralement faciles à standardiser et peuvent être répétées sans difficultés. Enfin, ces mesures constituent souvent une excellente base pour la résolution de problèmes de bien-être.

Les mesures observées directement *sur les animaux* sont généralement considérées comme étant des mesures plus directes du bien-être au sens où elles cherchent à évaluer directement l'état de l'animal. En effet, puisque le bien-être se réfère aux caractéristiques de l'animal, plutôt qu'à ce qu'on lui donne (Broom 1996), il semble logique de les privilégier. Elles seraient donc préférables aux mesures sur l'environnement pour évaluer le bien-être des animaux. Malheureusement, elles sont souvent beaucoup plus difficiles à réaliser et les observations demandent plus de temps (notamment pour tout ce qui nécessite des observations comportementales). De plus, les résultats collectés peuvent se révéler difficiles à interpréter en termes de bien-être (Johnsen et al. 2001). Il est donc primordial pour ce type de mesures de les tester pour leur validité, leur répétabilité et leur faisabilité (Winckler et al. 2003).

Au sein de Welfare Quality® des mesures sont réalisées à la fois sur les animaux et sur leur environnement, mais pour des utilisations différentes. En effet, afin d'évaluer le bien-être des animaux, les mesures sont essentiellement observées directement sur les animaux (Keeling et Bock 2007). Par contre, pour tout ce qui relève du diagnostic et de la recherche de solutions, les mesures basées sur les ressources disponibles et la gestion des animaux seront prises en compte (cf. Figure 1.8).

1.2.4 Quelle est la place de mon travail de thèse au sein du projet ?

Mon projet de thèse s'inscrit directement dans le Sous-projet 2 (SP2) de Welfare Quality® : *Développement d'une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir*. Plus particulièrement, mon travail s'insère dans le groupe de tâches 2.3 (noté par la suite WP2.3) visant à construire un modèle d'évaluation globale du bien-être[22].

Le SP2 se compose de quatre groupes de tâches distincts (cf. Annexe A) mais s'enchaînant, chaque WP utilisant les sorties des autres (cf. Figure 1.9). Ainsi, pour la mise au point du modèle d'évaluation globale, le WP2.3 a suivi les principes énoncés par le WP2.1, et s'est basé sur la liste des mesures et la façon de les interpréter produites par le WP2.2. Enfin, les données recueillies sur le terrain par le WP2.4 ont servi à affiner la construction du modèle.

[22]Les organismes de recherche partenaires du WP2.3 sont listés sur la Figure 1.6

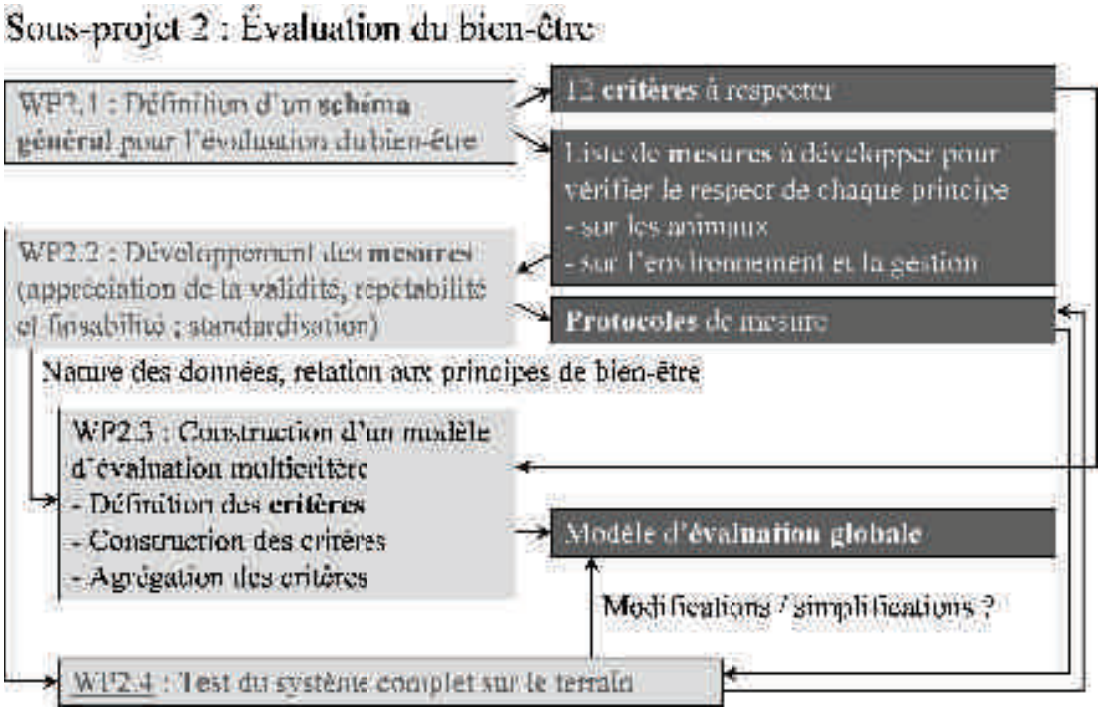


FIG. 1.9 – Organisation du Sous-projet 2 : Évaluation du bien-être

Au sein du WP2.3, mon rôle a été de conduire sur les vaches laitières en ferme une étude pilote aboutissant à la construction d'un modèle d'évaluation servant de guide aux modèles devant être réalisés sur les autres types d'animaux (pour les bovins : les veaux et les taurillons à l'engrais, mais aussi pour les porcs et les volailles). En effet, par souci de cohérence au sein de Welfare Quality® l'ensemble des modèles suit une même méthodologie pour sa réalisation, que ce soit au niveau de la construction (ex. *quelle structure adopter pour le modèle ? quelle méthode de calcul utiliser ?*) ou au niveau de son paramétrage (ex. *faut-il se baser sur des avis d'experts ? si oui, lesquels ? et comment le faire ?*).

1.3 La problématique générale de la thèse

1.3.1 Le bien-être animal : un concept multidimensionnel

Il existe de nombreuses définitions du bien-être animal. Alors que certains mettent en avant le fonctionnement biologique des animaux (par ex. Barnett et Hemsworth 1990) d'autres mettent l'accent sur leur ressenti émotionnel (par ex. Dawkins 1980). Les définitions du bien-être données dans la littérature suggèrent des approches différentes

du bien-être et ont des influences variées sur la recherche, via notamment le choix des mesures à prendre en compte pour évaluer le bien-être animal (Mason et Mendl 1993).

Stafleu et al. (1996) distinguent trois types de définitions du bien-être : les définitions lexicales, explicatives et opérationnelles.

Les définitions *lexicales* permettent d'établir le sens général commun donné au concept de bien-être. En ce sens, les définitions lexicales se doivent de présenter les aspects du bien-être sur lesquels tout le monde s'accorde. De telles définitions sont donc générales et souvent courtes. Stafleu et al. (1996) en retiennent trois principales :

- le bien-être animal est un état dans lequel l'animal se sent bien ;
- le bien-être animal est un état dans lequel l'animal ne souffre pas et ne ressent pas de douleur ;
- le bien-être animal est un état dans lequel l'animal est physiquement et psychologiquement en harmonie avec lui-même et avec son environnement[23].

Des définitions aussi larges et générales permettent à la science et à la société de se mettre d'accord sur la notion de bien-être et permettent de définir le champ de recherche dans lequel les scientifiques devront travailler.

Les définitions *explicatives* ne cherchent pas à élucider la signification générale du concept, mais présentent comment certains aspects du concept de bien-être peuvent être reliés à d'autres concepts scientifiques. En ce sens les aspects moraux ou émotionnels sont exclus de ce type de définition. Ces définitions permettent de produire un cadre théorique d'étude aux chercheurs s'intéressant au bien-être animal.

Parmi les nombreuses définitions explicatives existant dans la littérature, on peut en citer quelques unes :

- Selon Broom (1986, 1987), le bien-être animal est l'état dans lequel se trouve un animal qui essaie de s'adapter efficacement à son milieu, c'est-à-dire avec un "coût" acceptable, sans souffrance ni épuisement ; le bien-être correspond alors à l'état physique de l'animal ainsi qu'à ses capacités physiques d'adaptation.
- Certains auteurs associent les aspects physique et psychologique, en définissant le bien-être comme l'état résultant de l'absence de souffrance et de la présence d'expériences positives (Duncan 2005) ou comme la synthèse d'une bonne santé mentale et physique (Webster 2005 ; Dawkins 2006).
- Cette idée d'associer les aspects physique et psychologique est aussi présente chez Fraser. Selon lui le bien-être physique se traduit par un bon état de santé de l'animal et le bien-être psychologique par un bien-être comportemental, c.-à-d.

[23] ce qui correspond à la définition du bien-être donnée par Hurnik et al. 1985 (citée par Hurnik 1990)

l'existence de comportements normaux et l'absence de comportements anormaux (Fraser 1989). Fraser complète ensuite sa définition en dégageant trois principes communs aux définitions existantes du bien-être animal (Fraser 1993 ; Fraser 1995) :
- l'absence de souffrance (douleur, peur, faim, soif...) ;
- le fonctionnement normal de l'organisme (absence de maladie, de blessure, de malnutrition...) ;
- l'existence d'expériences positives (confort, expression des comportements propres à l'espèce, jeu...).

Enfin, les définitions *opérationnelles* décrivent les paramètres concrets qui devraient être utilisés, ainsi que la façon de les mesurer, de sorte à obtenir une évaluation du bien-être. Ces définitions listent les principes devant être respectés en élevage pour assurer le bien-être des animaux, et formant donc autant de points de contrôle. C'est par exemple le cas de la définition du bien-être fournie en 1965 dans le *rapport Brambell* (Brambell 1965), elle-même reprise est complétée dans les *cinq libertés* du Farm Animal Welfare Council (1992) :

1. **Absence de faim et de soif**, par accès libre à de l'eau fraîche et à une alimentation équilibrée ;
2. **Absence d'inconfort**, en fournissant un environnement adapté incluant un abri et une zone de repos confortable ;
3. **Absence de souffrance**, de blessure et de maladie, par la prévention, un diagnostic précoce et des soins rapides ;
4. **Liberté d'exprimer des comportements normaux**, grâce à un espace suffisant, des équipements adaptés et la présence de congénères ;
5. **Absence de peur et d'anxiété**, en assurant des conditions et un traitement des animaux qui évite la souffrance mentale.

Ces *cinq libertés* proposent donc une liste des besoins physiologiques, comportementaux, sanitaires, environnementaux, et même psychologiques des animaux.

Une caractéristique majeure du concept de bien-être se détache clairement des définitions pouvant être proposées : la nature **multidimensionnelle** du bien-être, couvrant des aspects aussi différents que la bonne santé de l'animal ou sa capacité à ressentir des émotions positives.

Stafleu et al. (1996) soulignent alors qu'en passant des définitions lexicales à explicatives, puis opérationnelles, le concept de bien-être s'éloigne de plus en plus des

considérations morales qui étaient à l'origine du questionnement pour devenir plus objectivement mesurable.

Welfare Quality® cherche justement à répondre à ce besoin d'objectiver l'évaluation du bien-être, et reposera donc plutôt sur une approche opérationnelle. En effet, l'un des objectifs du projet est de proposer un système d'évaluation utilisable sur le terrain et fournissant des résultats cohérents. Le bien-être doit ainsi être évalué par des mesures concrètes pouvant être réalisées sur le terrain. La définition du bien-être utilisée dans Welfare Quality® repose donc sur une liste de critères couvrant tous les aspects du bien-être, et se rapproche ainsi de la définition donnée par les *cinq libertés*. Le choix de cette liste est détaillée dans le Chapitre 4.

1.3.2 L'évaluation du bien-être animal repose sur la réalisation de nombreuses mesures

Comme nous venons de le voir, le bien-être animal est un concept multidimensionnel. Si on reprend l'exemple de la définition donnée par Fraser (1993), le bien-être couvre l'absence de souffrance, le fonctionnement normal de l'organisme et l'existence d'expériences positives ; chacun de ces principes comprenant plusieurs aspects, par exemple l'absence de souffrance comprend l'absence de faim prolongée, de soif, de peur, d'inconfort et de détresse.

Certains auteurs ont proposé de n'utiliser qu'un seul indicateur pour évaluer le bien-être d'un animal. Ces indicateurs universels sont supposés fournir une vue d'ensemble synthétique du niveau de bien-être des animaux. On peut ainsi citer l'utilisation du dosage des corticostéroïdes (proposé par Barnett et Hemsworth 1990), le dosage des protéines de phase aiguë (proposé par Geers et al. 2003), ou la longévité (proposé par Hurnik 1990). Malheureusement, aucune de ces mesures ne permet de couvrir l'ensemble des dimensions du bien-être. Par exemple, les protéines de phase aiguë semblent être spécifiques des dommages causés aux tissus (suite à des maladies ou des agressions entre les individus) et peuvent ne pas refléter l'existence d'un stress psychologique (Piñeiro et al. 2005). Or, le nombre de cas de maladies est indépendant de la réalisation de comportements normaux propres à l'espèce considérée. Ces derniers sont pourtant mis en avant dans de nombreuses définitions du bien-être (comme par exemple les *cinq libertés*). De même, le taux de cortisol ne semble pas pouvoir à lui seul couvrir l'ensemble des dimensions du bien-être animal. Un taux de cortisol élevé s'explique par un état de stress de l'animal et/ou par le fait que ses besoins métaboliques ne sont pas couverts, mais il ne permet pas de savoir si l'animal est malade. La longévité ne semble pas non plus pouvoir correspondre à un indicateur universel du bien-être. En effet, ce n'est pas

parce qu'un animal meurt relativement jeune d'un accident (ou du fait de la pratique de renouvellement du troupeau suivie par l'éleveur) qu'il n'a pas été élevé dans de bonnes conditions de bien-être avant.

Par conséquent, plusieurs mesures semblent nécessaires pour pouvoir produire une évaluation couvrant tous les aspects du bien-être d'un animal (voir par exemple Friend 1980 ; Dawkins 1980 ; Webster 1997 ; Rutter 1998). On distingue quatre grandes catégories d'indicateurs servant à évaluer le bien-être animal : sanitaires, zootechniques, physiologiques et comportementaux.

1.3.2.1 Les indicateurs sanitaires

Un bon état sanitaire d'un animal se traduit par l'absence de maladies et de blessures (Dantzer 1983). L'apparition de pathologies, pouvant être cause de malaise ou de douleurs pour l'animal (Veissier et al. 2007), peut être le révélateur d'un stress de l'animal. En effet, les défenses immunitaires de l'individu sont amoindries en cas de stress, ce qui favorise le développement d'agents pathogènes opportunistes (Veissier et al. 1999).

1.3.2.2 Les indicateurs zootechniques

Il est possible de tenir compte du gain de poids, de la production de lait ou d'oeufs pour évaluer le bien-être des animaux du fait que ces indicateurs peuvent entre autres être affectés en cas de stress chronique (Veissier et al. 2007).

Beaucoup d'éleveurs se basent sur de bons résultats zootechniques pour dire que leurs animaux sont en situation de bien-être, mais les faits peuvent parfois les contredire (Dantzer 1983). Par exemple, les poulets de chair, sélectionnés pour avoir une croissance plus rapide et un poids plus important, présentent de plus en plus de difficultés locomotrices. Sanotra et al. (2003) ont clairement montré que plus les poulets de chair sont lourds, plus ils ont de risque de boiter et de souffrir de pododermatites ou de déformations du squelette.

1.3.2.3 Les indicateurs physiologiques

Il existe deux grandes voies d'activation neuroendocrinienne répondant à une situation de stress de l'animal :

- activation de la branche sympathique du système nerveux autonome → libération de catécholamines (noradrénaline et adrénaline), avec pour conséquence directe une augmentation de l'activité cardiaque,
- activation de l'axe corticotrope → libération de corticoïdes (notamment cortisol).

La mesure des taux d'hormones (noradrénaline et adrénaline), de l'activité cardiaque et du taux de cortisol (plasmatique, urinaire ou salivaire) permet de détecter des stress ponctuels (Veissier et al. 1999).

1.3.2.4 Les indicateurs comportementaux

Les modifications comportementales en élevage intensif sont de trois types, suivant le degré de sévérité et l'incidence économique (Ewbank 1973) :

- anomalies comportementales aisément reconnaissables avec incidences pathologiques et économiques, ex : succions ;
- anomalies comportementales aisément reconnaissables sans incidences pathologiques ni économiques, ex : mâchonnement des barreaux, activités à vide... ;
- modifications décelables uniquement lors d'observations systématiques et orientées.

Cependant ces indicateurs peuvent se révéler parfois difficiles à interpréter, du fait notamment du rôle important des facteurs environnementaux, pas toujours faciles à identifier et à limiter (Dantzer 1983). De plus, la mesure sur le terrain de ces indicateurs comportementaux peut parfois nécessiter beaucoup de temps.

Winckler et al. (2003) proposent une liste réduite de mesures pour l'évaluation en ferme du bien-être des bovins : les boiteries, les blessures, l'état d'engraissement, la propreté, le comportement de coucher et de lever, les comportements sociaux agonistiques, les comportements oraux anormaux, la relation homme-animal, les soins apportés par l'éleveur, des indicateurs positifs de bien-être, des caractéristiques du mode de logement des animaux, l'incidence des maladies et la mortalité. On peut remarquer que les mesures physiologiques ne sont par recommandées pour une évaluation du bien-être dans des élevages commerciaux. En effet, les mesures physiologiques, pourtant très efficaces, ont cependant le gros inconvénient de demander des protocoles très lourds à mettre en place. Ainsi, pour avoir des mesures pertinentes il faut pouvoir réaliser de nombreux prélèvements (notamment prises de sang). Dans le cas des prises de sang, il est souvent nécessaire de poser un cathéter sur chaque animal, ce qui est impossible à réaliser ailleurs qu'en conditions contrôlées en station expérimentale. De plus, l'action même de la prise de sang s'avère stressante pour l'animal et est donc susceptible de faire varier le taux de cortisol plasmatique.

1.3.3 Dans quels buts évaluer le bien-être ?

L'évaluation du bien-être peut servir différents objectifs et donc différentes approches du bien-être animal (Main et al. 2003 ; Veissier et al. 2007).

Dans le cadre d'une *approche analytique*, l'objectif de l'évaluation du bien-être est d'identifier et d'étudier les facteurs de variations du bien-être des animaux. Cette approche est utilisée à des fins de recherche, comme par exemple pour décrire l'impact des conditions de vie des animaux sur leur état (à la fois physique et mental). On peut par exemple citer les travaux de Veissier et al. (1998) étudiant l'impact d'une alimentation essentiellement liquide sur les activités orales des veaux, ou encore ceux de Lensink et al. (2001) étudiant l'influence du comportement de l'éleveur et des conditions de logement des veaux sur leurs réactions comportementales face aux hommes, ainsi que sur leurs réactions comportementales et physiologiques suite à leur manipulation et à leur transport. L'objectif premier de ce type d'approche analytique est de mieux comprendre les besoins des animaux, leurs préférences et leurs aversions (Rushen 1986 ; Dawkins 1990). Ces travaux peuvent alors aboutir à une augmentation des connaissances, puis à des solutions pratiques permettant de résoudre certains problèmes de bien-être.

Dans une logique d'*intervention et de conseil*, des observations exhaustives sont conduites dans un élevage, dont le propriétaire s'est généralement porté volontaire pour cette intervention. L'objectif est alors d'avoir une vue d'ensemble de l'élevage avec des observations pouvant porter à la fois sur des aspects sanitaires, zootechniques et comportementaux. Ces observations permettent d'identifier d'éventuels problèmes, puis leurs causes, pour pouvoir proposer des mesures correctrices.

L'évaluation du bien-être peut aussi être réalisée dans une *logique de recommandation*, qui se distingue du conseil de par sa portée générale (au lieu de l'application au cas par cas du conseil). Les connaissances acquises sont utilisées pour mettre en place des recommandations ou des réglementations sur la façon dont les animaux doivent être traités et logés. On peut citer par exemple le rôle important des groupes d'experts de l'Autorité Européenne de Sécurité des Aliments auprès de la Commission Européenne. C'est en effet sur la base de leurs rapports et recommandations que la Commission se fonde pour proposer des réglementations à l'Union Européenne. Ces rapports font une revue des situations d'élevage en les évaluant de manière exhaustive au moyen de différents critères tenant compte des risques potentiels sur les animaux.

La mise en application de ces réglementations rentre par contre dans le cadre d'une *logique de certification* visant à vérifier la conformité des élevages avec les lois en vigueur. De façon similaire, les systèmes de certification (et de labellisation) faisant référence au

respect du bien-être animal font appel à ce type de logique. L'objectif étant de vérifier que l'éleveur est bien en conformité avec l'ensemble des points listés dans le cahier des charges du système de certification. Dans ce cas, les vérifications sur le terrain sont généralement conduites par des organismes tiers, et ce dans un souci d'objectivité de l'évaluation, contrairement à la logique de conseil qui peut se traduire par des procédures d'auto-évaluations destinées aux éleveurs.

Enfin, l'évaluation du bien-être peut aussi être réalisée dans le cadre de la *conception de systèmes*, afin par exemple de choisir le système innovant maximisant un certain nombre d'indicateurs de bien-être sur lesquels différents systèmes "candidats" auront été testés *ex ante*.

Les outils utilisés en pratique pour évaluer le bien-être sont différents selon les objectifs de l'évaluation (cf. Tableau 1.3). En effet, l'exhaustivité de l'évaluation n'est pas forcément nécessaire dans une approche analytique, alors que c'est une des conditions requises dans le cadre d'un programme de certification. Par ailleurs, la construction d'une procédure formalisée et standardisée d'évaluation globale du bien-être est nécessaire pour une certification, alors que dans le cadre d'un diagnostic-conseil, le conseiller n'a pas besoin d'établir une évaluation globale du bien-être sur la ferme considérée.

Les objectifs de Welfare Quality® (cf. section précédente) s'inscrivent clairement dans une logique de certification, puis d'intervention-conseil. De l'identification de la logique à suivre dépend le type de modèle à prendre en compte ainsi que la problématique qui se pose alors.

TAB. 1.3 – Caractéristiques de l'évaluation du bien-être en fonction des objectifs visés

Objectif	**Liste d'indicateurs**	**Objets vérifiés**	**Construction d'une évaluation globale**
Approche analytique	Centrée sur les éléments étudiés	Environnement et/ou animaux	Non
Intervention /conseil	Exhaustive	Environnement et animaux	Utile, mais plutôt à des niveaux de synthèse intermédiaires
Recommandation	Centrée sur les éléments étudiés	Plutôt environnement	Non
Certification	Exhaustive	Environnement et/ou animaux	Nécessaire, à formaliser
Conception de système	Exhaustive	Environnement et animaux	Nécessaire, pas forcément formelle

1.3.4 Construire une évaluation multicritère du bien-être animal

Le bien-être animal étant une notion multidimensionnelle, son évaluation peut naturellement reposer sur la construction d'un modèle d'évaluation multicritère. Il faut alors formuler la problématique à laquelle nous faisons face dans Welfare Quality®, puis établir le type de modèle le plus approprié pour la résoudre.

1.3.4.1 Quel type de problématique ?

La première étape de la construction d'un modèle consiste à préciser le problème qu'il doit servir à résoudre. L'identification de la problématique à laquelle nous devons répondre doit reposer sur les objectifs du projet Welfare Quality®.

De façon générale, en aide à la décision se pose le problème de définir quel type de décision on souhaite prendre sur un ensemble d'actions possibles. Les *actions* (pouvant être aussi appelées *alternatives*) correspondent aux différents objets entre lesquels l'évaluation (ou la décision) doit s'opérer. Dans le cas de Welfare Quatlity®, ce sont sur des fermes que l'évaluation du bien-être animal porte.

Nous devons tout d'abord identifier l'ensemble des actions concernées par le modèle d'évaluation. La décision ne vise-t-elle qu'un groupe de fermes identifiées (ex. un ensemble de n fermes appartenant à une même coopérative) ou cherche-t-elle à pouvoir

être automatisée sur des groupes de fermes différents et évolutifs (ex. sur l'ensemble des fermes européennes désirant s'inscrire dans une démarche qualité utilisant le système d'évaluation produit par Welfare Quality®) ? Dans le cas du projet, c'est la seconde proposition qui, dès le départ, a été retenue.

Une fois clarifié l'ensemble des actions sur lequel s'appliquera le modèle d'évaluation du bien-être, il est nécessaire de clarifier le résultat que l'on souhaite obtenir à la fin du processus d'évaluation, et ainsi de formuler la problématique rencontrée dans Welfare Quality®. Pour ce faire, le problème auquel nous faisons face peut être situé par rapport à quatre problématiques de référence (Roy 1985) : la description, le choix, le tri et le classement (cf. Tableau 1.4). Cherche-t-on seulement à fournir une description organisée des différentes fermes (description) ? Souhaite-t-on pouvoir affecter chaque ferme à un ensemble de catégories de bien-être pré-définies (tri) ? Veut-on pouvoir classer les fermes les unes par rapport aux autres (classement) ? Ou enfin, cherche-t-on à établir quelle est la meilleure ferme (choix) ?

TAB. 1.4 – Les quatre problématiques de référence en évaluation multicritère - d'après Roy (1985, p. 74)

Problématique	**Objectif**	**Résultat**
Description	Éclairer la décision par une description, dans un langage approprié, des actions et de leurs conséquences	Une description ou une procédure cognitive
Choix	Éclairer la décision par le choix d'un sous-ensemble aussi restreint que possible en vue d'un choix final d'une seule action, ce sous-ensemble contenant les "meilleures" actions (optimums) ou, à défaut, des actions "satisfaisantes" (satisfecums).	Un choix ou une procédure de sélection
Tri	Éclairer la décision par un tri résultant d'une affectation de chaque action à une catégorie ; les catégories étant définies *a priori* en fonction de normes ayant trait à la suite à donner aux actions qu'elles sont destinées à recevoir.	Un tri ou une procédure d'affectation
Classement	Éclairer la décision par un rangement obtenu en regroupant tout ou partie (les "plus satisfaisantes") des actions en classes d'équivalence, ces classes étant ordonnées, de façon complète ou partielle, conformément aux préférences.	Un rangement ou une procédure de classement

Parmi ces quatre types de problématiques, la *description* est en fait un préalable aux trois autres. En effet, elle ne va pas jusqu'à formuler une recommandation (contrairement aux trois autres problématiques), mais formalise la description des différentes actions et ainsi sert de base indispensable pour la définition de recommandations. L'objectif de Welfare Quality® concernant le système d'évaluation du bien-être consiste à fournir une évaluation globale du bien-être animal sur les différentes fermes concernées, et ce de façon standardisée et automatisée. Il s'agit donc de produire, via le modèle d'évaluation, une recommandation (évaluation globale du bien-être). La problématique de description n'est alors pas suffisante pour remplir cet objectif.

Par ailleurs, le projet a pour objectif de produire un système d'évaluation fournissant une information sur les fermes appropriée pour une utilisation dans le cadre de procédures de certification "bien-être". Dans tout système de certification, la question est de savoir si la ferme observée peut être labellisée ou non. On peut donc y voir la définition de deux catégories de bien-être : le niveau de bien-être sur la ferme est soit 0- insuffisant, soit 1- suffisant, pour pouvoir être labellisée. En fonction de la nature de la certification (par ex. marché de niche assurant un excellent niveau de bien-être ou, au contraire, label européen assurant le respect de la législation en vigueur), les catégories devant être définies seront différentes. Par contre, Welfare Quality® ne cherche pas à produire un système qui permettrait de classer les fermes les unes par rapport aux autres ou à pointer la ou les meilleures fermes européennes. On cherche à connaître le niveau de bien-être sur une exploitation donnée, et non à savoir si elle est meilleure ou moins bonne que les fermes voisines. Par contre, il sera important de produire un système permettant de suivre les évolutions du niveau de bien-être dans les fermes pour vérifier, si un système de certification est mis en place, qu'il permet effectivement d'améliorer les conditions de vie des animaux. De plus, cela permettrait à un éleveur de voir si les efforts qu'il a accomplis ont bien permis d'améliorer le bien-être de ses animaux.

L'objectif de Welfare Quality® est de répondre aux attentes sociétales ayant été identifiées (cf. la première section de ce chapitre introductif) :

- définition d'un label européen garantissant le respect des réglementations en vigueur dans l'Union Européenne touchant au respect du bien-être animal ;
- garantie d'un bon niveau de bien-être (supérieur à celui requis par la législation) afin que le bien-être animal puisse être inclus dans des démarches qualité généralistes (bien-être associé à une bonne qualité gustative des produits...) ;
- identification des élevages assurant un excellent niveau de bien-être et pouvant ainsi se démarquer des autres élevages au sein d'un marché de niche.

Chacune de ces attentes impliquent la définition de règles binaires d'affectation à des catégories reflétant différents niveaux de bien-être. Par définition, le choix et la signification des catégories de bien-être nécessaires dépendra principalement des utilisations potentielles pouvant être faites du système d'évaluation du bien-être proposé par le projet Welfare Quality® (cf. ci-dessus). Il s'agira alors de caractériser les conditions d'appartenance à chacune des différentes catégories, de sorte à ne classer une ferme que dans une seule des catégories.

Le système d'évaluation produit par Welfare Quality® reposera donc sur une procédure d'affectation permettant de trier les fermes d'après des normes pré-établies et définissant des catégories de bien-être, sans qu'il soit pour autant nécessaire d'attribuer une note globale à chacune des fermes. Dans la classification établie par Roy (1985), le problème que Welfare Quality® cherche à résoudre se rapproche donc de la **problématique de tri**.

1.3.4.2 Quel type de modèle ?

Une fois la problématique clarifiée, nous devons préciser quel type de modèle serait le plus approprié pour y répondre. Il existe différents types de modèles d'évaluation : les modèles descriptifs, normatifs et prescriptifs. Ils jouent des rôles différents dans une procédure de décision (Bell et al. 1988 ; Roy 1993).

Les modèles *descriptifs* servent à décrire une situation pré-existante stable et indépendante de l'observation. Par exemple, les statistiques multivariées sont souvent utilisées pour décrire une population, et des variables synthétiques sont calculées pour résumer la variabilité entre les observations.

Les modèles *normatifs* nous disent comment les choses devraient être ou comment les personnes devraient agir. Ils visent à fournir des procédures d'évaluation permettant de vérifier l'adéquation des éléments observés à des normes pré-établies. C'est par exemple le cas de nombreuses procédures de labellisation.

Enfin, les modèles *prescriptifs* sont utilisés afin d'aider les personnes à prendre de meilleures décisions et ainsi d'améliorer leur activité. L'approche prescriptive ne repose sur aucune situation pré-existante qui devrait être décrite, mais vise plutôt à collecter et organiser des informations pertinentes afin de faciliter la construction de recommandations pour atteindre des objectifs fixés.

Ces distinctions entre les différents types de modèles sont essentielles puisqu'en fonction de chacun, son mode de construction, d'évaluation et de validation variera. Comme Bell et al. (1988) le soulignent :

- les modèles descriptifs sont évalués en fonction de leur validité empirique, c'est-à-dire s'ils correspondent ou non aux choix observés ;
- les modèles normatifs sont évalués en fonction de leur adéquation théorique, c'est-à-dire s'ils permettent ou non de fournir des idéalisations acceptables ou des choix rationnels ;
- les modèles prescriptifs sont évalués en fonction de leur valeur pragmatique, c'est-à-dire s'ils permettent ou non d'aider les personnes à prendre de meilleures décisions.

Comme nous avons pu le voir dans la section précédente, nous ne considérons pas que l'évaluation du bien-être animal devrait être vue comme un simple problème descriptif. Par contre, le bien-être animal correspond plutôt à un problème de définition de normes, d'évaluation des pratiques actuelles et de création de recommandations. En effet, au vu de la volonté de l'Union Européenne d'établir des standards européens de bien-être et de celle d'autres utilisateurs potentiels de mettre en place des procédures de certification, le modèle d'évaluation du bien-être aura donc un caractère normatif. De plus, au vu de l'objectif de Welfare Quality® de produire un système d'évaluation pouvant prescrire des recommandations en termes de diagnostic et de conseil auprès des éleveurs, le modèle d'évaluation revêt aussi un caractère prescritptif. La procédure formelle d'évaluation du bien-être animal peut donc être vue comme un modèle à la fois normatif et prescriptif.

Par ailleurs, différentes approches peuvent être suivies pour mettre en place les trois types de modèles : soit il existe des préférences pré-existantes au modèle et déjà formalisées, soit, ce qui est généralement le cas, les préférences doivent être construites ou révélées au cours de la mise en place du modèle. Il s'agit dans ce deuxième cas d'adopter une *approche constructive* lors de la création du modèle. C'est au cours de ce processus de construction que les utilisateurs potentiels du système d'évaluation (les décideurs) vont forger leurs préférences, en pouvant les réviser au fur et à mesure des questions qui leur seront posées par l'analyste (chargé de mettre en place le modèle d'évaluation) (voir par exemple Tsoukiàs 2007). Avec une approche constructive, la construction du modèle doit se faire en collaboration avec les décideurs.

Du fait de l'absence de préférences globales universelles pré-existant à la construction de notre modèle, l'évaluation du bien-être requiert donc la mise en application d'une approche constructive dans laquelle une procédure interactive sera utilisée pour aider les décideurs à produire une évaluation globale du bien-être qui leur permettra soit

d'établir des normes, soit d'évaluer les fermes (ou abattoirs) et de conseiller les éleveurs (ou abatteurs).

L'approche constructive qui sera adoptée au niveau de l'évaluation globale pourra localement reposer sur des méthodes descriptives, comme par exemple l'utilisation de méthodes d'optimisation afin de trouver les fonctions mathématiques rendant le mieux compte des évaluations d'experts. Pour Tsoukiàs (2007), aucune objection ne s'oppose à la mise en place de ce type de modèles, utilisant parfois des outils qui ne sont *a priori* pas adaptés au modèle global.

Nous proposons donc de mettre en place *un modèle à la fois normatif et prescriptif*, reposant sur *une approche constructive* impliquant les experts du domaine et les utilisateurs potentiels du futur système d'évaluation.

1.4 L'objectif de mon travail

Mon travail de thèse s'inscrit directement au sein du Sous-projet 2 (SP2) de Welfare Quality® : *Développement d'une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir*. Plus précisément, mon travail s'insère dans le WP2.3, groupe de tâches qui vise à construire un modèle d'évaluation globale du bien-être, et ce pour les huit types d'animaux concernés par le projet (vaches laitières, veaux de boucherie, taurillons à l'engrais, porcs à l'engrais, truies, porcelets, poulets de chair et poules pondeuses) et, pour la plupart, au cours des deux périodes identifiées dans le projet : en ferme et durant la phase de transport-abattage.

J'ai été chargée de **construire le modèle d'évaluation globale du bien-être des vaches laitières en ferme**. Ce modèle devra être suffisamment générique et stable pour pouvoir être, par la suite, transposé aux autres couples (*type d'animal*, *période*) concernés par Welfare Quality®. Le modèle "vaches laitières en ferme" servant alors d'exemple pour les autres modèles d'évaluation, il devra reposer sur l'établissement d'un langage et d'une méthodologie communs à l'ensemble des espèces animales.

En effet, le projet Welfare Quality® vise à développer des modèles d'évaluation cohérents entre eux, utilisant les mêmes principes de construction, des méthodes de synthèse de l'information similaires, et des méthodes de consultation des experts harmonisées entre les différents modèles. Ce sont tous ces principes et méthodes qui sont devenus la base de travail du WP2.3 et qu'il nous a donc fallu identifier et mettre au point.

Ce modèle permettra de répondre à une **problématique de tri ordonné** (permettant de définir plusieurs classes de bien-être d'inacceptable à excellent). Il devra être à la fois *normatif* (de par son application dans le cadre d'un label européen et/ou de procédures de certification "bien-être") et *prescriptif* (de par son utilisation pour servir de base à un outil de diagnostic-conseil auprès des éleveurs, transporteurs et abatteurs), en adoptant *une approche constructive*.

De plus, comme nous l'avons vu dans l'introduction, le bien-être animal est un concept multidimensionnel complexe et dont l'évaluation globale repose nécessairement sur un ensemble de contraintes qu'il nous a aussi fallu identifier. Nous avons donc travaillé en constante interaction avec les partenaires de Welfare Quality® qui mettaient au point les mesures, afin de pouvoir développer un modèle d'évaluation qui prenne en compte tous les aspects du bien-être.

Chapitre 2

Revue des méthodes actuellement proposées pour une évaluation globale du bien-être animal

Résumé de la revue des méthodes actuellement proposées pour une évaluation globale du bien-être animal :

Plusieurs systèmes ont d'ores et déjà été proposés pour produire une évaluation globale du bien-être des animaux d'élevage. Ces systèmes d'évaluation ont été proposés dans le but de conseiller les éleveurs, d'aider à la prise de décision publique ou de réaliser des études dans le cadre de programmes de recherche. Ils se basent généralement sur les informations fournies par de nombreuses mesures qui sont ensuite synthétisées en une seule évaluation globale au niveau de l'élevage. Pour ce faire, différentes méthodes peuvent être utilisées pour agréger l'information.

Dans les systèmes d'évaluation actuellement proposés, les données recueillies sur une ferme peuvent être :
- analysées par un ou plusieurs experts qui en tirent une conclusion générale, mais sans que le processus de décision n'ait été formalisé ;
- comparées à des exigences minimales définies pour chacune des mesures prises en compte ;
- converties en des données correspondant chacune au rang de la ferme sur la mesure considérée, ces rangs étant ensuite sommés ;
- converties en scores évalués sur une échelle de valeur, ces scores étant ensuite agrégés via une moyenne pondérée (ex. TGI35L) ou via des règles *ad hoc*.

Cependant, les systèmes d'évaluation existant à l'heure actuelle, au moins quand ils sont utilisés de manière exclusive, peuvent se révéler être insuffisamment sensibles (par ex. en ne permettant de distinguer que deux catégories de fermes) ou non applicables en routine. De plus, ils peuvent ne pas refléter correctement la nature multidimensionnelle du bien-être animal ou ne par tenir compte de l'importance relative des différentes mesures. Il s'agit pourtant de conditions requises dans le contexte du projet Welfare Quality®.

Ces méthodes peuvent cependant se révéler être très utiles dans un système d'évaluation où elles seraient utilisées de manière complémentaire. Elles pourraient en effet être utilisées à différentes étapes de la construction du modèle d'évaluation globale, en fonction des contraintes identifiées à chacune des étapes de la procédure d'agrégation.

Chapitre 2

Revue des méthodes actuellement proposées pour une évaluation globale du bien-être animal

Il n'existe pas de mesure universelle et unique du bien-être. Au contraire, le bien-être est un concept multidimensionnel et il convient d'évaluer chacune de ses dimensions. De plus, l'appréciation globale du bien-être ne peut être dérivée de la simple description de l'existant (par exemple, l'état corporel moyen dans un troupeau, l'incidence des mammites...). Les données obtenues par de telles mesures doivent être interprétées en termes de bien-être.

Plusieurs méthodes ont d'ores et déjà été proposées pour produire une évaluation globale à partir d'un ensemble de mesures évaluant le bien-être. Cela va de l'agrégation non formalisée par des experts à l'utilisation de calculs précis comme des sommes pondérées de scores obtenus sur chacune des mesures.

L'objet de cette partie est de présenter les différentes méthodes proposées jusqu'à présent pour produire une évaluation globale du bien-être, en soulignant à la fois leurs forces et leurs faiblesses. L'objectif n'est pas de juger les différentes méthodes, mais d'analyser et de mettre en évidence les limites des calculs utilisés pour synthétiser les informations, notamment dans le contexte d'un système de labellisation "bien-être". Cette analyse repose sur l'identification de la construction théorique propre à chaque méthode. En fonction de la construction suivie, il est possible d'identifier dans quel(s) contexte(s) chaque méthode sera la plus appropriée. Nous évaluerons plus particulièrement l'adéquation de chacune des méthodes à une utilisation dans le cadre d'un système de certification du bien-être animal. Pour ce type d'utilisation, nous avons besoin d'une

méthode :

- facilement explicable aux différents utilisateurs potentiels (éleveurs, consommateurs...) ;
- utilisable en routine sur un grand nombre de fermes (ou d'abattoirs), l'ensemble des fermes (ou abattoirs) concernées pouvant fortement varier entre deux évaluations ;
- permettant d'encourager les producteurs à améliorer le niveau de bien-être de leurs animaux, en mettant en évidence les problèmes de bien-être les plus sérieux afin qu'ils puissent être étudiés et corrigés en priorité ;
- répétable, afin que les éleveurs et les consommateurs puissent avoir confiance dans l'évaluation proposée ;
- à la fois précise et sensible, afin de mettre en évidence des améliorations (ou au contraire des dégradations) des conditions de vie des animaux.

Cette synthèse des méthodes proposées jusqu'à présent pour produire une évaluation globale du bien-être animal a fait l'objet d'une revue plus complète (*"Aggregation of measures to produce an overall assessment of animal welfare - Part 1 - A review of existing methods"* Botreau et al. 2007) dont le texte figure à la fin de cette section.

2.1 Agrégation non formalisée de plusieurs mesures de bien-être

Quand plusieurs mesures sont utilisées pour produire une évaluation globale du bien-être, la procédure d'agrégation peut ne pas avoir été formalisée. C'est par exemple le cas quand il est demandé à un expert (ou un groupe d'experts) de donner son opinion sur une ferme en se basant sur un rapport où toutes les mesures effectuées sur le terrain sont réunies.

Ce type d'évaluation, dont le résultat est explicite mais pas la procédure de synthèse de l'information, est actuellement utilisé pour apporter des conseils aux éleveurs. Par exemple Sørensen et al. (2001) et Hegelund et al. (2003) se basent sur un ensemble de mesures prises à la fois sur les animaux et sur leur environnement pour mettre en évidence des problèmes de bien-être, identifier leurs causes et proposer une stratégie qui permettra aux éleveurs d'améliorer le bien-être de leurs animaux.

Cette méthode a l'avantage d'être simple puisqu'aucun outil mathématique n'est nécessaire à sa réalisation.

Cependant, la procédure de synthèse de l'information manque généralement de transparence. En fonction du (ou des) expert(s) interrogé(s), l'évaluation globale pourra

être différente. Par exemple, on peut penser qu'un éthologiste prendra plus particulièrement en compte le comportement des animaux, alors qu'un vétérinaire accordera plus d'importance à l'état sanitaire du troupeau.

De plus, l'évaluation demandée à l'expert repose souvent sur de nombreuses mesures. Or, Lacroix et al. (1997) ont démontré que lorsque le nombre d'informations à prendre en compte devient trop important, l'homme ne peut pas les traiter toutes en même temps et certaines d'entre elles risquent de ne pas être prises en compte.

Enfin, ce type d'évaluation non formalisée ne semble pas adaptée à un usage en routine puisqu'il ne peut pas, par définition, être automatisé.

En conclusion, il semble que cette méthode soit très appropriée dans le cas d'un conseil apporté au cas par cas à des éleveurs, ainsi qu'à des instances réglementaires (cf. par exemple l'utilisation de ce type de méthode en Suisse pour autoriser ou non l'utilisation de nouveaux systèmes de logement ou de nouveaux équipements pour les animaux d'élevage, Wechsler 2001, 2003, 2005).

Quels sont les avantages d'une procédure d'évaluation formalisée ?

Les procédures de synthèse de l'information formalisées ont plusieurs avantages par rapport aux procédures venant d'être décrites :

- Le processus de formalisation nécessite que les décideurs (ici, les experts chargés d'évaluer le bien-être) identifient précisément les éléments à prendre en compte et ceux qui peuvent être relégués au second rang. Parmi l'ensemble des mesures qui sont réunies dans les rapports qui leur sont soumis, il est probable que seule une partie de cette information soit utile à la prise de décision. Au sein d'une procédure d'évaluation formalisée, cette distinction entre données utiles à la décision et données purement informatives doit être clairement spécifiée et être le résultat d'un consensus entre les différents décideurs.
- En plus d'aider à préciser quelles sont les informations à utiliser, le processus de formalisation permet d'expliciter comment utiliser ces informations. Avec une procédure non formalisée il peut se révéler parfois extrêmement délicat d'identifier le raisonnement suivi par le (ou les) décideur(s). L'évaluation est alors le résultat d'une "boîte noire" sur laquelle il semble donc difficile de pouvoir obtenir un consensus. De plus, la procédure d'évaluation perd alors son potentiel explicatif.
- Du fait que les éléments à prendre en compte ainsi que la manière de les utiliser pour l'évaluation sont clairement identifiés dans un procédure formalisée, ce type d'évaluation pourrait sans grande difficulté être automatisée. Cet atout est notamment utile dans le cadre d'évaluations devant être répétées plusieurs fois, comme

c'est le cas dans le cadre d'une utilisation dans une procédure de certification. Toutes ces caractéristiques ne sont pas toujours nécessaires (ce qui est notamment le cas dans le cadre des utilisations présentées dans cette première section), mais peuvent se révéler être particulièrement utiles, voire indispensables, dans bien des utilisations.

2.2 Définition d'exigences minimales pour chacune des mesures effectuées

La plupart des systèmes de certification reposent sur un ensemble de mesures dont le résultat obtenu sur la ferme est comparé à une norme pré-établie. C'est par exemple le cas du *Freedom Food Scheme* mis en place au Royaume-Uni en 1994, avec pour objectif de fournir aux distributeurs et aux consommateurs l'assurance du respect du bien-être animal dans les fermes participant au programme. Les mesures prises en compte sont pour la très grande majorité prises sur l'environnement ou la gestion du troupeau par l'éleveur (cf. Main et al. 2001). Pour chacune des mesures, le résultat est comparé à une valeur de référence. Voici quelques exemples d'exigences minimales pour des troupeaux de vaches laitières (Royal Society for the Prevention of Cruelty to Animals 2004) :

- pour des vaches laitières de race Holstein nourries *ad libitum*, il doit y avoir au moins 25 cm d'auge par vache ;
- pour un troupeau de 50 vaches la longueur totale des abreuvoirs doit être d'au moins 2.25 m ;
- en stabulation libre avec aire paillée, pour des vaches pesant entre 651 et 700 kg, l'aire de couchage et l'aire d'exercice doivent avoir des superficies respectives d'au moins 6.25 m^2 et 7.50 m^2 par vache ;
- les sols en béton lisse doivent être rainurés ou alors traités avec une matière anti-dérapante ;
- ...

Il s'agit de méthodes claires et simples, facilement explicables au grand public. De plus, contrairement aux méthodes reposant sur une agrégation non formalisée, ces méthodes sont très faciles à standardiser. Leur mise en œuvre est donc tout à fait recommandée dans le cadre de systèmes de certification basés sur un ensemble de conditions à respecter, et conduisant à une réponse sous la forme "conformité" *vs* "rejet".

Cependant, certaines caractéristiques de ces méthodes peuvent se révéler être de véritables inconvénients.

En effet, quand elles sont appliquées de façon stricte, elles conduisent à une réponse en "tout ou rien". Il suffit que sur une seule mesure la ferme obtienne un résultat lé-

gèrement en deçà du minimum requis pour que cette ferme soit rejetée. Ces méthodes peuvent donc se révéler "sur-réactives". Il existe toutefois des manières d'introduire de la flexibilité. Au cas où une anomalie serait relevée sur la ferme, cette dernière pourrait se voir accorder un délai afin de prendre des mesures correctrices. Il serait aussi possible d'envisager d'accorder un certain pourcentage de mesures sur lesquelles la ferme pourrait ne pas être en conformité.

De plus, dans ces méthodes toutes les mesures prises en compte ont la même importance, et la non-conformité sur une seule de ces mesures constitue un *veto*. Il serait cependant plus approprié d'introduire des nuances en accordant plus d'importance aux éléments causant le plus de souffrance aux animaux.

En conclusion, de telles méthodes sont tout à fait adaptées à une utilisation dans le cadre de systèmes de certification, mais il pourrait être souhaitable d'y incorporer plus de flexibilité, et aussi de tenir compte de l'importance relative des différentes mesures.

2.3 Calcul de la somme (ou moyenne) des rangs obtenus par chaque ferme sur les différentes mesures

Whay et al. (2003) ont comparé le bien-être de veaux sur 45 fermes en se basant sur 19 mesures (cf. Tableau 2.1). Sur chaque mesure les 45 fermes ont été classées de la meilleure (rang 1) à la moins bonne (rang 45). Chaque ferme était donc décrite par 19 rangs partiels, et un rang global était alors calculé comme étant la moyenne des 19 rangs partiels. Les 45 fermes étaient alors classées sur la base de leur rang global. D'un point de vue mathématique, ceci correspond à la méthode de Borda (Bouyssou et al. 2000, p. 129).

De telles méthodes sont claires et faciles à comprendre et à standardiser. Elles permettent à chaque éleveur de se positionner au sein du groupe étudié et ainsi de mieux se rendre compte de la situation dans laquelle ses animaux sont élevés.

Cependant, dans ces méthodes, tous les rangs partiels (c.-à-d. toutes les mesures) ont la même importance pour le calcul du rang global. Il est malgré tout possible d'affecter des pondérations aux différentes mesures, par exemple, en comptant dans le calcul de la moyenne deux fois le rang partiel correspondant à une mesure considérée comme étant deux fois plus importante que les autres.

La principale limite de ces méthodes est liée au fait que le résultat obtenu par une ferme dépend de la population observée. En effet, le rang relatif qu'obtient une ferme par rapport à une autre peut dépendre de la présence d'une ferme tiers (cf. exemple

TAB. 2.1 – Liste des 19 mesures prises en compte dans Whay et al. (2003) pour évaluer le bien-être *(NB : Obs = Observé ; Est = Estimé ; Reg = D'après les registres tenus par l'éleveur)*

Aspect du bien-être	**Mesure**
Santé respiratoire	Fréquence respiratoire élevée *(% Obs)*
	Respiration anormale *(% Obs)*
	Toux *(% Obs)*
	Cas de pneumonie *(pour 100 veaux par an, Est)*
	Cas de pneumonie *(pour 100 veaux par an, Reg)*
Nutrition	Veaux maigres (Score d'état < 2) *(% Obs)*
	Rumen météorisé *(% Obs)*
	Rumen vide *(% Obs)*
	Rumen plein ('hay belly') *(% Obs)*
	Périnée et queue sales *(% Obs)*
	Diarrhée *(% Obs)*
	Cas de diarrhées *(pour 100 veaux par an, Est)*
Apparence générale	Blessures, lésions corporelles *(% Obs)*
	Poil terne *(% Obs)*
	Poil piqué *(% Obs)*
	Perte de poils *(% Obs)*
	Membres postérieurs sales *(% Obs)*
	Flancs sales *(% Obs)*
	Apathie *(% Obs)*

fournit dans l'article présenté à la fin de cette section), ce qui peut poser des problèmes de justification des résultats auprès des personnes concernées. De telles méthodes ne peuvent donc être utilisées que dans des situations très spécifiques où le groupe de fermes observées reste le même d'une année sur l'autre, ce qui, dans la grande majorité des cas, ne correspond pas aux systèmes de labellisation sur la base du volontariat.

2.4 Calcul de la somme (ou moyenne) des scores obtenus sur chaque mesure

Alors que les sommes de rangs ne permettent des comparaisons qu'au sein d'une population donnée, les sommes de scores produisent des évaluations dans l'absolu (c.-à-d. indépendantes de la population observée). De telles méthodes ont été largement développées dans le cadre de l'évaluation du bien-être animal (Bartussek 1999 ; Keeling et Svedberg 1999 ; Horning 2001 ; Scott et al. 2001 ; Bracke et al. 2002).

Ces méthodes se basent sur une procédure commune. D'abord, les données brutes recueillies sur le terrain, et exprimées avec des unités très différentes (ex. en kg pour le poids de l'animal, en % de vaches présentant des mammites, en note/score pour l'état d'engraissement...), sont transformées en scores partiels sur une échelle commensurable, c.-à-d. une échelle s'interprétant en termes de bien-être de la même façon quelle que soit la mesure considérée. Ensuite, des poids sont attribués, de manière explicite ou implicite, aux différentes mesures afin de tenir compte de leurs importances relatives. Enfin, un score global est calculé, correspondant à la somme (ou à la moyenne) pondérée des scores partiels obtenus sur chacune des mesures.

Les sommes pondérées de scores sont très couramment utilisées, notamment du fait qu'elles sont faciles à expliquer, au moins dans les grandes lignes, et qu'elles sont couramment utilisées dans d'autres contextes (ex. la notation des élèves). Le score global obtenu au niveau des fermes permet de comparer ces dernières tout en fournissant un jugement dans l'absolu de chaque ferme, c.-à-d. indépendant des autres élevages qui ont été évalués.

Cependant, les sommes de scores souffrent de plusieurs limites.

Il n'est possible de calculer des sommes ou des moyennes que sur des données quantitatives (Bouyssou et al. 2000, p. 47-48), et non sur des données ordinales (cf. exemple détaillé dans le Tableau 2 de l'article présenté à la fin de cette section). Or, les scores attribués aux mesures sont-ils réellement quantitatifs ? (voir discussion dans le chapitre suivant, dans l'article Botreau et al. 2007, p. 1193).

Par ailleurs, les sommes de scores autorisent pleinement les compensations entre

les différentes mesures. Ainsi, un problème majeur de bien-être peut être masqué par plusieurs petits avantages sur d'autres aspects du bien-être. Ceci n'est pas forcément souhaitable, notamment quand il s'agit de synthétiser l'information obtenue sur des aspects très différents du bien-être (voir discussion dans le chapitre suivant, dans l'article Botreau et al. 2007, p. 1190-1191).

Enfin, les sommes de scores ne permettent pas de favoriser des situations de compromis, caractérisées par des résultats moyens sur l'ensemble des critères plutôt que contrastés (Bouyssou et al. 2000, p. 43-45). En effet, une ferme présentant des résultats moyens sur l'ensemble des mesures peut obtenir le même score global, voire un résultat inférieur, qu'une ferme ayant obtenu de très bons résultats sur certaines mesures et de très mauvais sur d'autres. Cependant, du point de vue de l'animal, il serait peut-être préférable pour lui de vivre dans des conditions moyennes plutôt qu'être soumis à des aspects très négatifs, même si par ailleurs ses conditions de vie sont bonnes. Un exemple illustrant ce point est détaillé dans l'article présenté à la fin de cette section (Botreau et al. 2007, p. 1184).

En conclusion, les sommes de scores présentent l'avantage majeur d'être intuitives et simples à expliquer. Elles présentent cependant un certain nombre de conditions d'utilisation et de limites qui peuvent les rendre parfois inadaptées. Par exemple, il semble hasardeux de calculer des sommes de scores sur des aspects très différents qui, en théorie, ne devraient pas se compenser l'un l'autre, comme cela semble être le cas en bien-être animal.

2.5 Utilisation de règles définies au cas par cas afin de limiter les compensations entre les différents éléments

Capdeville et Veissier (2001) se basent eux aussi sur des scores calculés pour chacune des mesures réalisées sur le terrain, mais au lieu de les sommer pour obtenir un score global, ils se basent sur des règles établies au cas par cas, dans le but de limiter les compensations entre les scores. Pour chaque combinaison de scores partiels (exprimés en 4 niveaux, de A 'excellent niveau de bien-être' à D 'mauvais niveau de bien-être'), un score global est défini, le plus souvent inférieur à la moyenne des scores partiels. Ainsi un A et un D donneront un C.

Cependant, cette méthode ne permet d'agréger qu'un petit nombre de scores partiels à la fois (au maximum 4), ce qui impose le regroupement des mesures en de nombreux sous-ensembles, puis la réalisation de plusieurs agrégations successives. Les regroupements ainsi opérés n'ont alors pas forcément de signification biologique, et la méthode

reste assez opaque dans son ensemble et difficile à expliquer.

2.6 Conclusion

Les méthodes qui ont jusqu'à présent été proposées pour produire une évaluation globale du bien-être peuvent se révéler, quand elles sont utilisées de façon exclusives, insuffisamment sensibles (en produisant une réponse de type "tout ou rien", ce qui est par exemple le cas du label "Freedom Food", Main et al. 2001) ou non applicables en routine (ce qui est par exemple le cas des évaluations reposant sur l'agrégation non formalisée de mesure, ex. Sørensen et al. 2001 ; Hegelund et al. 2003), elles peuvent aussi parfois ne pas refléter correctement l'aspect multidimensionnel du bien-être ou ne pas prendre en compte le fait que certaines mesures peuvent être plus importantes que d'autres (ce qui est par exemple le cas du label "Freedom Food" ou des évaluations reposant sur le calcul de la somme des rangs, Whay et al. 2003). Le Tableau 2.2 présente brièvement les principaux systèmes proposés jusqu'à présent ainsi que leurs avantages et inconvénients. Le choix d'une méthode de synthèse de l'information repose en partie sur l'objectif de l'utilisation du système d'évaluation du bien-être que l'on souhaite mettre en place. La dernière colonne du Tableau 2.2 met en évidence pour quel(s) type(s) d'utilisation les systèmes d'évaluation passés en revue dans cette section sont les plus appropriés.

Utilisées séparément les unes des autres, ces différentes méthodes ne semblent donc pas permettre un résultat pleinement satisfaisant dans un contexte de labellisation assurant le respect d'un bon (voire d'un excellent) niveau de bien-être. Toutefois, en les combinant et en les utilisant à différentes étapes de l'agrégation des mesures de bien-être, il serait vraisemblablement possible d'améliorer le résultat final. Dans tous les cas, il faut bien garder à l'esprit les différentes contraintes liées à la procédure d'agrégation afin de produire un système répondant le mieux possible aux objectifs fixés dans le projet Welfare Quality® pour un tel système d'évaluation globale du bien-être. Ce point sera abordé dans le chapitre suivant.

Tab. 2.2 – Principales méthodes actuellement proposées pour calculer un score global de bien-être animal

Description	Avantages	Inconvénients	Utilisations recommandées
Agrégation non formalisée par un expert qui analyse des données recueillies en ferme	Se base uniquement sur les données brutes recueillies en ferme	Manque de transparence Impossible à standardiser Le nombre important d'informations rend difficile la tâche de l'expert	Évaluation au cas par cas d'un nombre limité d'élevages, de systèmes d'élevage ou d'équipements
Respect d'exigences minimales Pour chaque mesure la ferme est comparée au seuil exigé, puis les anomalies sont additionnées	Claire et simple Facile à standardiser Permet de vérifier la conformité à une norme	Toutes les mesures ont la même importance Ne permet pas de comparer des fermes	Vérification en routine que des exigences sont respectées
Somme de rangs Les fermes sont classées de la meilleure à la pire pour chaque mesure, puis les rangs sont additionnés	Claire et simple Facile à standardiser Permet de classer les fermes entre elles	Les mesures ont la même importance Ne permet des comparaisons qu'au sein d'un ensemble de fermes donné	Comparaisons entre élevages à l'intérieur d'une population fixe
Somme de scores Les données sont transformées en scores de satisfaction, puis les scores sont additionnés	Relativement intuitive Permet d'obtenir un score absolu pour une ferme donnée	Difficile de définir une échelle de bien-être Compensations entre mesures pleinement autorisées Ne favorise pas les compromis	Comparaisons entre élevages ou certification d'élevages, à condition d'autoriser les compensations

Animal (2007), **1:8**, pp 1179–1187 © The Animal Consortium 2007
doi: 10.1017/S1751731107000535

animal

Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods

R. Botreau[1,2†], M. Bonde[3], A. Butterworth[4], P. Perny[5], M. B. M. Bracke[6], J. Capdeville[2] and I. Veissier[1]

[1]INRA, UR1213 Herbivores, Site de Theix, F-63122 Saint-Genès-Champanelle, France; [2]Institut de l'Elevage, BP18, Castanet Tolosan F-31321, France; [3]Department of Animal Health, Welfare and Nutrition, Danish Institute of Agricultural Sciences, POB 50, Tjele DK-8830, Denmark; [4]University of Bristol Clinical Veterinary Science, Langford UK-BS40 5DU, UK; [5]Laboratoire Informatique de Paris 6, Université Pierre et Marie Curie, 8 rue du Capitaine Scott, Paris F-75015, France; [6]Animal Sciences Group, Wageningen University and Research Centre, POB 65, Lelystad NL-8200 AB, The Netherlands

(Received 10 January 2007; Accepted 6 June 2007)

Several systems have been proposed for the overall assessment of animal welfare at the farm level for the purpose of advising farmers or assisting public decision-making. They are generally based on several measures compounded into a single evaluation, using different rules to assemble the information. Here we discuss the different methods used to aggregate welfare measures and their applicability to certification schemes involving welfare. Data obtained on a farm can be (i) analysed by an expert who draws an overall conclusion; (ii) compared with minimal requirements set for each measure; (iii) converted into ranks, which are then summed; or (iv) converted into values or scores compounded in a weighted sum (e.g. TGI35L) or using ad hoc *rules. Existing methods used at present (at least when used exclusively) may be insufficiently sensitive or not routinely applicable, or may not reflect the multidimensional nature of welfare and the relative importance of various welfare measures. It is concluded that different methods may be used at different stages of the construction of an overall assessment of animal welfare, depending on the constraints imposed on the aggregation process.*

Keywords: animal welfare, assessment, livestock, methodology

Introduction

During the last 30 years, extensive experimental work has been carried out to collect information and to inform the societal debate on animal welfare. The aim of much of this study has been to describe the impact of living conditions on animals and to gain a better understanding of their needs, preferences or aversions (Rushen, 1986; Dawkins, 1990). These studies have resulted in recommendations to producers or policy makers on how animals should be kept for farming or other purposes (see for instance the reports from the Scientific Committee on Animal Health and Animal Welfare of the European Commission, 2007). More recently, researchers have attempted to assess the actual welfare of farm animals on farms or at slaughter to provide information on the influences on animal welfare of their living (or slaughter) conditions. Such information is considered by many to be essential in the context of certification schemes based, at least in part, on animal welfare (e.g. Bartussek, 1999; Main *et al.*, 2001). These schemes can help improve animal welfare through market-driven strategies (Bock and Van Leeuwen, 2005) and may require a formal model for the overall evaluation of animal welfare (Fraser, 2003).

Welfare is a multidimensional concept: it embraces absence of suffering, high levels of biological functioning – including absence of diseases – and the potential for animals to have 'positive experience' (Fraser, 1993). Each of these principles subsumes several criteria, e.g. absence of suffering comprises absence of prolonged pain, hunger, thirst, fear, discomfort and distress. Some single measures have been proposed to provide a broad assessment of animal welfare including corticosteroids (Barnett and Hemsworth, 1990) (acute phase proteins) (Hurnik, 1990) and longevity (Geers *et al.*, 2003). However, none of these single measures attempts to cover all the dimensions of welfare. For instance, an animal can be diseased with no impairment of its corticotropic axis activity and vice versa.

† E-mail: rbotrea@clermont.inra.fr

Acute phase proteins seem specific to tissue damage (due to disease or aggression between animals) and may not be related to psychological stress (Piñeiro *et al*., 2005). Longevity depends at least in part on the occurrence of diseases and diseases are rather independent of the performance of natural behaviour, which is considered essential to animal welfare (Farm Animal Welfare Council (FAWC), 1992). Hence, it appears likely that several measures are necessary to obtain a comprehensive view of any particular animal's welfare (Dawkins, 1980; Friend, 1980; Webster, 1997; Rutter, 1998).

Sets of measures, made on either the farm or housing environment, or on the animals themselves, have been defined for the purpose of (i) advising farmers on how to improve the welfare of their animals (Sørensen *et al*., 2001); (ii) checking compliance with legislative requirements – for example, as carried out in Sweden in association with the banning of battery cages (Keeling and Svedberg, 1999); (iii) implementing welfare certification schemes (e.g. Freedom Food Scheme: Main *et al*., 2001); or (iv) comparing systems to refine legislation (Bracke *et al*., 2002; for a review see Main *et al*., 2003). Most of these goals require combining measures to form an overall assessment, either to fully describe a situation (e.g. comparing systems) or to make an absolute assessment on it (e.g. checking compliance with standards).

Several types of evaluation models – descriptive, normative and prescriptive – can be distinguished and they play different roles in a decision process based on several measures (Bell *et al*., 1988; Roy, 1993). Descriptive models are use to describe a pre-existing situation that is stable and independent of any observation. For example, multivariate statistics are often used to describe populations and compound variables (i.e. principal components) are calculated to summarise the variability between observations. This approach provides the ability to describe and compare observed situations (as in Veissier *et al*., 2004).

Normative models tell us how things should be or how people should act, and aim at providing evaluation procedures to check the adequacy of observed behaviours relative to pre-defined norms. This is, for example, the case for many formal labelling processes (i.e. certification, hotel/restaurant rating, etc.) and could also be the case for animal welfare evaluation. As observed by French (1984), '… normative analyses tell us how we should behave in particular circumstances, whereas descriptive theories conjecture how things are behaving'.

Finally, prescriptive models are used to help people make better decisions and to improve their activity. The prescriptive approach does not assume any pre-existing situation that should be described but rather aims to gather and organise relevant information so as to facilitate the construction of recommendations to reach a goal. These distinctions in the nature of the model are essential, because they might change the perspective in the construction of the model and the way it should be evaluated or validated. As noted by Bell *et al*. (1988):

> descriptive models are evaluated by their empirical validity, that is the extent to which they correspond to observed choices. Normative models are evaluated by their theoretical adequacy, that is, the degree to which they provide acceptable idealizations or rational choices. Prescriptive models are evaluated by their pragmatic value, that is their ability to help people make better decision.

We do not consider that the evaluation of animal welfare should be seen as a simple descriptive problem because there is no pre-existing universal view of animal welfare that can be measured. More likely, animal welfare assessment is a matter of the definition of norms, evaluation of current practice with respect to these norms and the creation of recommendations (e.g. Fraser, 1995 and 2003). As such, the formal evaluation procedure for assessing animal welfare could be seen as both a normative and a prescriptive model.

Several methods have been proposed to compound the results from welfare measures to make an overall assessment. They range from informal aggregation by experts to the use of precise calculations such as weighted sums of scores obtained for each measure. Here we review the methods that have been proposed for the overall assessment of animal welfare, highlighting their strengths and weaknesses. The relevance of the measures *per se* will not be addressed here as this issue has been largely dealt with elsewhere (e.g. Winckler *et al*., 2003). Our objective is neither to form an overall judgement of the various methods used to produce a welfare assessment nor to rank them. By contrast, we intend to analyse the content and limits of the calculation methods proposed to aggregate welfare measures. This analysis relies on the identification of the theoretical construction of each method. According to the characteristics of each calculation method, we will then analyse in which context(s) a given method could be best used. This should in turn help in choosing the best methods for a specific use, e.g. the comparison of animal units in benchmarking or to provide policy makers with decision tools. More specifically, we will evaluate the potential of each aggregation method for use in the context of certification schemes for animal welfare. For this application, we need an aggregation method that (i) can be easily explained to stakeholders (producers, consumers, etc.); (ii) can be used routinely on large numbers of animal units (farms, slaughter-plants, zoos, etc.) where the composition of these sets varies in time; and (iii) encourages producers to improve animal welfare. The method must therefore be repeatable – so that both farmers and end-users trust the results – and precise – in order to be able to monitor improvements or on the contrary decrements in welfare – while lending greatest importance to the more serious welfare problems – so that major problems can be given first priority and be cured first.

Non-formal aggregation of several welfare measures

When several measures are used to make an overall assessment, the aggregation may not be explicit. This is the case when one expert (or a group of experts) is asked to form an opinion on an animal unit (e.g. a farm or a slaughterhouse) based on a number of observations related to animal welfare. Such non-explicit aggregation of measures is currently used to advise farmers. For instance, Sørensen *et al.* (2001) developed a system for an 'ethical account' of dairy and pig farms, and Hegelund *et al.* (2003) developed a similar system for egg production in organic farms. They take into account animal-based measures (behaviour, health) and environment-based measures (system, management). Welfare problems and their potential causes are identified and a strategy to improve the welfare status of the animals is proposed to the farmer.

Informal appraisal of measures can also be used in certification or authorisation processes. This is the case for systems developed by the Swedish Board of Agriculture to approve or reject new housing systems or equipment for laying hens (Algers *et al.*, 1995; Ekstrand *et al.*, 1997; quoted by Johnsen *et al.*, 2001) or by the Federal Veterinary Office from Switzerland to authorise mass-produced housing systems and equipment designed for farm animals (Wechsler, 2001, 2003 and 2005). Data on production, health, mortality and behaviour of the animals are collected in experimental investigations and on-farm inspections. These are then brought together in a detailed report from which experts are asked to form an opinion on new equipment.

This method seems simple since no mathematical tools are necessary.

However, the rationale behind the process of aggregation is generally not transparent. The opinion is the result of the reasoning of one expert (or several experts), and experts may have different views on the interpretation of measures. For instance, because of their different backgrounds veterinarians may lend more importance to health, whereas ethologists may emphasise behaviour, and so these methods are prone to variability between people.

In addition, animal welfare assessment is often based on a large set of data. As mentioned by Lacroix *et al.* (1997), humans may become saturated by large amounts of detailed information, resulting in some information being discounted when non-formal aggregation of information is used.

Finally, for each evaluation to be performed on a given animal unit, these methods require the opinion and input of 'experts'. Consequently, informal aggregation is not suitable for routine assessments on large numbers of animal units (farms, slaughterhouses) by inexpert persons.

Even so, informal appraisal by experts seems highly appropriate for advising policy makers. In Switzerland, such evaluation has proved efficient for licensing housing systems and equipment. Similarly, the reports from the European Food Safety Agency can be viewed as a summary of expert opinion on farming systems on the basis of several indicators, generally listed at the beginning of the report (e.g. the report on the welfare of calves; Algers *et al.*, 2006).

Definition of minimal requirements for measures

Benchmarking systems are based on minimum requirements that a farm, a housing system or equipment have to meet. A series of measures are taken and a threshold is set for each measure. This system is the basis of the Freedom Food Scheme, which was set up as a farm assurance and food-labelling scheme by the Royal Society for the Prevention of Cruelty to Animals (RSPCA) in 1994. The measures considered are essentially based on resources, management and health records (health plan, diseases, etc.; Main *et al.*, 2001). The Freedom Food scheme has been adapted to North American farms through the certification scheme *'Certified Humane Raised and Handled'* developed by the association for Humane Farm Animal Care (HFAC, 2003). Similarly, Von Borell *et al.* (2001) proposed a method based on hazard analysis and critical control points (HACCP) for pig housing. These authors consider the impact of different housing designs on welfare, health, management and the environment. Key threats to welfare (critical control points) are identified, and adequate controls are then determined for these points. This method may be regarded as a checklist with the same number of points as the critical control points determined in the HACCP procedure.

The methods based on comparison with a list of minimal requirements are straightforward and can be easily explained to laypersons and, unlike a non-formal aggregation, they can be standardised and suit the aim of assuring compliance with pre-set standards. As a consequence, they are perfectly adapted to be implemented in certification schemes based on the fulfilment of a list of specifications, leading to an *'accept/reject'* answer. However, they present some characteristics specific to this aim that may prove to be drawbacks if such methods are to be applied to achieving other objectives.

First, when applied strictly, they are 'all-or-nothing' processes. If the value obtained by a farm for one measure is below the pre-set threshold, then that farm will be considered below standard – and in a certification process, this farm may be rejected. Because of this limitation, these methods may be 'over-reactive': on a farm classified as 'good' a deterioration in only one aspect can result in exclusion from the certification scheme (e.g. a disease outbreak, even if all other parameters remain the same); no distinction is made between a farm that fails on only one aspect and a farm that fails on many aspects. However, some flexibility can be introduced. For instance, in the Freedom Food scheme, when a requirement is not met, the farmer is given 8 weeks to make corrective measures (RSPCA, undated). Hence, problems on a good farm that

fails on one aspect may be corrected within an agreed time, whereas this would be difficult to achieve on a farm that failed on too many (or deep seated) aspects. Another way to make the method more flexible is to allow a certain proportion of non-compliances. A more mathematically elaborate method would be to define reference profiles (of sets of welfare measures or dimensions) and compare observations with these profiles according to majority rules (Bouyssou *et al*., 2000, p. 226). In this case, the results could take the form of the assignment of a farm to a class corresponding to a certain level of welfare, such as low, moderate or high. This method has not yet been applied in the context of animal welfare.

Second, in these methods, all measures have the same impact, and non-compliance for one measure constitutes a 'veto'. However, some elements, such as painful disease states (e.g. severe lameness), may induce more suffering than others (e.g. lack of positive experiences). Some schemes, such as the EUREPGAP certification standards developed by retailers and their global suppliers (EUREPGAP, nd), include welfare requirements classified as major requirements, minor requirements or recommendations. It is likely that a less-strict compliance is requested for recommendations than for minor requirements and in turn for minor requirements than for major ones.

In conclusion, comparison with a list of minimal requirements (inspection) could be used as part of certification schemes. More-subtle results are obtained when a certain degree of flexibility is introduced and the relative contribution of the different measures for animal welfare is taken into account.

Sum (or mean) of ranks

Whay *et al*. (2003) compared dairy calves' welfare on 45 farms on the basis of 19 measures corresponding to respiratory health, nutrition and general appearance. For each measure the farms were ranked from best (rank 1) to worst (rank 45). Each farm was thus described by 19 partial ranks, and an overall rank was then calculated as the mean of the 19 partial ranks. The farms were sorted according to their overall ranks. Mathematically, this corresponds to the Borda method, i.e. a sum of partial ranks (Bouyssou *et al*., 2000, p. 129).

In a similar vein, the Bristol Welfare Assurance Programme proposed to rank farms according to the quintile in which the farm is situated among a given population: the farm scores 0 when in the best quintile for a given measure and 4 when in the worst quintile (Webster, 2005). Then the sum of the partial scores obtained by the farm on the various measures is calculated. This method can be assimilated to a sum of ranks since the score obtained by a farm depends on how this farm performs in comparison with other farms, i.e. in this method the partial scores are simplified expressions of partial ranks. An even greater simplification of ranks is produced when farms are simply compared with the average value obtained over the population of the inspected farms, as in Huxley *et al*. (2004) for organic dairy herds.

Such sorting methods are clear and easy to understand and standardise, and can help farmers to position their farm among others and to make them more aware of the conditions in which their animals are living.

In these methods, each partial rank (i.e. each measure) has the same importance in the final overall rank. For instance, in the method proposed by Whay *et al*. (2003), hind limb dirtiness was as important as pneumonia. However, it is possible to assign a weighting to measures, for instance by counting the rank obtained for a particular measure twice in the calculation of the final score if this measure is considered to be twice as important as each of the others.

The main limit of this method is that the overall ranking of a farm depends on the population observed: a better ranking obtained by one farm over another results not only from their relative positions in the rankings for the different measures but also from their relative positions with respect to the other farms examined (Bouyssou *et al*., 2000, p. 16; see Table 1 for a numerical example). The Borda method should be used only in very specific situations in which the set of farms is perfectly well-defined and stable over time. This will not generally be the case in a voluntary labelling system.

Sum (or mean) of scores

While a sum of ranks allows comparisons only within a given sample of observations, a sum of scores produces absolute values (i.e. independent of the sample observed). Such methods are widely developed in the context of animal welfare (Bartussek, 1999; Keeling and Svedberg, 1999; Horning, 2001; Scott *et al*., 2001; Bracke *et al*., 2002).

A common framework can be seen for all these methods. First, raw data expressed in very different units (e.g. kg for body weight, % of animals affected by mastitis and *score* for body condition) are converted into a 'partial welfare score' on a numerical value scale with a meaning that is common to all the measures (i.e. a commensurable scale). Second, weightings are assigned to the values obtained for the different measures to account for the impact of each measure on animal welfare. These weightings are explicit in Scott *et al*. (2001) and Bracke *et al*. (2002). They are implicit in the Tier Gerechtheits Index (TGI) defined by Bartussek (1999) or Sundrum and Rubelowski (2001) where they correspond to the size of the value scale: e.g. in TGI35L, a 0 to 3 scale is used for space allowance and a -0.5 to 0.5 scale is used for cleanliness. Third, an overall score is calculated as the weighted sum (or weighted mean) of the partial scores obtained for each measure.

Some authors, however, do not use weightings. For example, El Balaa and Marie (2004) proposed a method to assess the welfare of small ruminants where a total of

Table 1 *Illustration of how a sum of ranks depends on the sample of units compared with each other*

	Scores		Partial ranks			
Farms	Measure 1	Measure 2	Measure 1	Measure 2	Sum of ranks	Comparison between A and B
First set of farms						
A	10	15	3	1	**4**	A is preferred to B
B	16	8	1	4	**5**	
C1	6	9	4	3	7	
D1	12	13	2	2	4	
Second set of farms						
A	10	15	3	1	**4**	B is preferred to A
B	16	8	1	2	**3**	
C2	9	5	4	4	8	
D2	13	7	2	3	5	

Two farms A and B are compared on the basis of two distinct measures. At the top of the table, where farms A and B are compared with farms C1 and D1, farm A ranks higher than B. At the bottom of the table, where farms A and B are compared with farms C2 and D2, farm B ranks higher than A

Table 2 *(a and b) Illustration of how differences in scaling can affect welfare outcomes when a sum of scores is calculated*

(a) 'Lying-down movements' is expressed on a three-level scale: normal movement after only one intention (0); several intentions before a normal lying movement (1); interrupted movement (one foreleg is bent then the animal gets up) (2). 'Reactions to human' is expressed on a five-level scale: the animal approaches the experimenter (0); it does not move (0.5); it stops eating (1); it makes one step back (1.5); it makes several steps back (2). The welfare of A and B is concluded to be equal

	Scale	Animal A	Animal B	Comparison between A and B
Lying-down movements	0/1/2	2	1	B is better than A
Response to human approach	0/0.5/1/1.5/2	0	1	A is better than B
Sum of scores		2	2	The welfare of A and B are similar

(b) 'Lying down movements' is expressed on a three-level scale (the same as for Table 2a), and 'reactions to human' is expressed on a three-level scale: the animal approaches (0); it does not move (1); it stops eating or moves back (2). The welfare of A is considered to be higher than that of B

	Scale	Animal A	Animal B	Comparison between A and B
Lying-down movements	0/1/2	2	1	B is better than A
Response to human approach	0/1/2	0	2	A is better than B
Sum of scores		2	3	The welfare of A is higher than the welfare of B

Two animals A and B are compared on the basis of lying movements (continuous lying movement is considered to be better for animal welfare) and responses to humans (not being afraid of humans being considered better). Animal A displays an interrupted lying down movement and approaches the experimenter while Animal B had no interruption of lying and stops eating when the experimenter comes near. The preference between A and B is assigned according to the sum of partial scores obtained for these two measures

43 measures is used to check compliance with the five freedoms defined by FAWC (1992). For each freedom, they calculated the mean of the scores obtained for the measures related to that freedom, giving the same importance to all measures.

Weighted sums (or means) of scores are very popular and are easily understood by non-scientists, at least in their general principles. Partial scores can be used to point out strong *v.* weak points of each farm assessed and so can help farmers to choose ways to improve the welfare of their animals. The overall score allows comparisons between animal units while an absolute judgement of a farm, independently of the others, can still be made. However, summing of scores suffers from several limitations. These are presented below.

First, welfare scores must be cardinal values, i.e. data assessed on interval or ratio scales (Bouyssou *et al.*, 2000, pp. 47–48), to be appropriate for the calculation of a weighted sum. However, it is often assumed, rather than shown, that the intervals between levels of the scales used are equivalent. Uncertainty about intervals between ordinal scales can lead to confusing results. This point is illustrated with a numerical example in Table 2. Ideally, the scales for different welfare measures should be converted into a unified scale (e.g. a scale between 0, worst for welfare, and 1, best for welfare) before (weighted) sums (or means) of scores are calculated.

Second, sums of scores allow full compensation. Spoolder *et al.* (2003) pointed out that when a sum of scores is calculated, a serious welfare disadvantage can be

compensated for by a number of minor advantages, even though this may not always be legitimate. Full compensation between welfare aspects may not be desirable. This point will be further discussed in part 2 of the present dissertation on overall assessment of animal welfare (Botreau *et al.*, 2007). One way to limit compensation could be to specify minimal requirements below which the animal unit is discarded (Perny, 1998). Bracke *et al.* (2002) introduced very high weighting factors for some levels of partial scores ($\pm 10\,000$ compared with ± 1, ± 2, ± 3, ± 5 used in most cases in the model). These very high weightings act as minimum requirements, limiting the possibility of compensating for a serious welfare problem with a number of small benefits.

In addition, when compensation between measures is legitimate, we can accept that a small increase in an average score for one measure (e.g. going from 15% to 20% of mastitis in a dairy herd) may be compensated for by a small decrease in an average score for another measure (e.g. going from 20% to 15% in lameness). However, compensation may not be acceptable at other points on the scale. For instance, an increase from 0% to 5% of mastitis may not be compensated for by a decrease from 20% to 15% of lameness because 0 mastitis means that there is no risk to spread mastitis from one animal to another, whereas at 5% there is some risk. Hence, when using sums of scores, attempts should be made to ensure that trade-offs are constant.

Third, sums (or means) of scores do not favour situations of compromise (Bouyssou *et al.*, 2000, pp. 43–45). A farm that obtains average scores for all measures (i.e. a homogeneous profile) could obtain the same overall score as a farm with very low scores on some measures and very high ones on others (i.e. a heterogeneous profile). However, from an animal's point of view, it may be preferable to live on a farm with moderate results on all measures than to be subjected to very poor environmental conditions irrespective of the other aspects. This point can be illustrated looking at three animals compared on the basis of their responses to human approach and their lying-down movement (the scales are the same as those used in Table 2b). In this example, Animal *X* and Animal *Y* have contrasting partial scores:

- *X* is not at all afraid of humans – i.e. scores 0 on measure 1 – but displays serious difficulties lying down – i.e. scores 2 on measure 2;
- *Y* is very afraid of humans – i.e. scores 2 on measure 1 – but has no problem at all lying down – i.e. scores 0 on measure 2;
- *Z* is moderately good for both measures – slightly disturbed by the presence of humans – i.e. scores 1 on measure 1 – and hesitates before lying down – i.e. scores 1 on measure 2.

The sum of scores equals 2 for *X*, *Y* and *Z*. However, a best compromise might be to have a mild response to human approach and to have only one intention before lying, rather than to approach humans but to have interrupted lying down movements. But whatever the weightings used to sum the two scores, *Z* will never be strictly preferred to *X* and *Y*: let w_1 and w_2 be the weights associated with measure 1 and measure 2;

- to consider *Z* in a better state than *X*, w_1 and w_2 shall be chosen so that $w_1 + w_2 < 2w_2$, i.e. $w_1 < w_2$;
- to consider *Z* in a better state than *Y*, w_1 and w_2 shall be chosen so that $w_1 + w_2 < 2w_1$, i.e. $w_2 < w_1$.

These two conditions '$w_1 < w_2$' and '$w_2 < w_1$' are obviously not consistent!

In conclusion, the sum of scores is very intuitive. It requires a number of conditions (cardinal data and constant trade-offs) that are difficult to check in practice. In addition, it may allow compensation where compensation should be restricted, and cannot favour compromises. Hence, it is hazardous to use such a method to produce an overall assessment based on non-compensatory aspects. However, it may be an appropriate method for compounding subsets of welfare measures related to the same welfare aspect, where compensation between measures can be considered legitimate (e.g. summing the occurrences of major diseases to obtain an overall view of the health status of a herd, as in Mounier *et al.*, 2006). In the same vein, multivariate analyses could be used to construct welfare indices. The first components of principal component analyses (PCA) are linear combinations of variables, i.e. weighted sums with the weight assigned to a variable depending on its distribution in relation to other variables and thus on its potential to describe variations in the population studied (Lebart and Fenelon, 1975). PCA have been successfully used on subsets of variables related to the same welfare issue. For instance, Veissier and Capdeville (2004) considered three main criteria describing the comfort of cows cubicles: difficulties in lying down and getting up movements, injuries, cleanliness. For each criterion, several measures were defined, a PCA was run on the data obtained from a survey on 70 farms, and the first component of PCA was considered as a summary variable to compare several cubicle designs. Such a use of PCA appears relevant in that context because within a criterion the variables can be considered as different ways of looking at the same problem (e.g. difficulties in lying down are generally accompanied by difficulties in getting up). However, setting the weights for separate criteria on the basis of a PCA seems to us inappropriate because there is no functional links between criteria. Furthermore, the limitations of weighted sums highlighted above would fully apply to PCA used to combine several welfare criteria.

Use of *ad hoc* rules limiting compensation

Capdeville and Veissier (2001) proposed using *ad hoc* rules to make an overall assessment of animal welfare for loose-housed dairy cows. Their rules were based on expert

Table 3 *Currently proposed methods for the overall assessment of animal welfare at farm level, with their advantages and limitations, and their potential use*

Method	Advantages	Limitations	Potential uses
Non-formal aggregation, an expert's opinion is produced on the examination of a report gathering all the data collected	• Based only on the raw data collected on farm, i.e. with no calculation	• Lack of transparency • Impossible to standardise for routine use by non-experts • A large amount of data is difficult to compound by the expert	• Perfect for the assessment of a few animal units or housing systems/equipments. This method should be limited to an analysis on a case-by-case basis.
'Checklist', each measure is compared with a minimal requirement	• Clear and simple • Easy to standardise • Allows correspondence to a standard/norm to be checked	• Yields an 'all-or-nothing' response • All measures have the same importance • Does not allow comparisons between farms	• Routine use, to check that all requirements are fulfilled, leading to a yes/no answer. This method can be used to check the respect for the current laws or within certification schemes based on the strict fulfilment of a series of specifications.
Sum of ranks, ranks obtained on several measures by a farm within a given set are summed	• Clear and simple • Easy to standardise • Allows ranking of farms	• All measures have the same importance • Allows comparisons between farms only within a given set of farms	• Routine use, to rank animal units belonging to a pre-set group of farms.
Weighted sum of scores, scores obtained on several measures by a farm are summed	• Relatively intuitive • Allows an absolute score for any farm to be produced	• Compensations fully authorised between measures • Does not favour situations of compromise	• Routine use, leads to an answer more sensitive than a mere yes/no answer. This method can be used to compare farms and/or to implement certification or labelling schemes where interactions between welfare items are permitted

opinion and were designed to limit compensation between measures. A set of 16 animal needs was defined, derived from the five freedoms (FAWC, 1992). Forty-nine measures were recorded on-farm to assess how these needs were met (essentially animal-based measures such as lying movements, lameness, injuries and agonistic social behaviour). As in the methods based on sum of scores (see above), a value scale was used: 'A' corresponded to a very high level of welfare; 'B', a moderately high level of welfare; 'C', a moderately low level of welfare; and 'D', a very low level of welfare. The difference from the previous methods lies in the rules used to aggregate the information. For each possible combination of partial scores, an overall value was decided on and this is generally lower than the mean of partial scores. For instance, a score 'A' on one measure with a score 'B' on another measure results in a 'B' for the overall value for the two measures taken together. Capdeville and Veissier (2001) defined three types of rules to aggregate information according to the importance of measures for the welfare of animals. The rules were stricter, i.e. allowed less compensation, or even forbade it, for measures they considered especially important (e.g. severe injuries).

In this method, and also in some methods using weighted sums, the aggregation process is performed in different stages: first a few measures are grouped together, then groups of measures are aggregated, etc. For instance, 'lying-down and standing-up movements' and 'resting posture' are considered together to assess resting behaviour before being included in the need for physical comfort. This hierarchical process limits problems of overweighting items that are assessed in numerous measures. Finally, a score is given to the farm for each basic need and then for the five freedoms together.

The *ad hoc* rules defined by Capdeville and Veissier (2001), however, could group only a few items together (maximum = 4). This required making many small sets of measures and then making many sequential aggregations. Consequently, the grouping may not always have a biological meaning and the entire method is rather opaque and difficult to explain.

In conclusion, this method limits compensation between measures and can favour compromise situations. However, the opacity of the method makes it difficult to appraise its legitimacy and to transpose it to other animal types.

Discussion

The advantages and limits of each category of methods for aggregating welfare measures analysed in this paper are summarised in Table 3, together with specifications on the context in which they are most appropriate. The choice of a method may in part depend on the objective for the use of a welfare assessment. Comparisons with thresholds are better suited to check compliance with requirements, whereas weighted sums can help to compare farms or farming systems. However, we feel that none of the existing methods proposed to produce an overall assessment of animal welfare is yet fully satisfactory to be used within a certification scheme, which includes several welfare classes (i.e. more sensitive than a mere yes/no classification). The most common shortcomings that limits application in a certification scheme are:

(i) allowing too much compensation between welfare aspects (e.g. in sum of ranks or sum of scores), and this may result in overshadowing a serious welfare problem by many small welfare improvements;
(ii) assuming all measures have the same importance (e.g. in definition of minimal requirements for measures or sum of ranks), and this does not help focussed improvements to be targeted on the most serious problems;
(iii) difficulty in creating a formal structure for routine use (e.g. in non-formal aggregation or use of *ad hoc* rules), which reduces the potential for using the method on a large scale and may result in difficulties in communicating the results;
(iv) providing assessments that are meaningful only for a given data set (sum of ranks), which may lead to competition between farmers rather than true welfare improvements.

Some authors propose systems that include different methods for the aggregation of measures. For example, Keeling and Svedberg (1999) proposed the use of 36 welfare measures on poultry units. For 12 measures, they defined thresholds corresponding to minimum legislative requirements. When these requirements were satisfied, 24 other measures were assessed on the farm and were further compounded using a weighted sum to produce an overall estimation of the welfare of animals.

Such a mixture of aggregation methods have already been used in other contexts, for instance in the European New Car Assessment Programme (EURONCAP, 2004). New cars are assessed for the level of protection displayed for adult occupants, children and pedestrians. The child protection assessment leads to a rating from 0 to 5 stars, according to a sum of partial scores obtained for each issue of child protection. Several methods are used to attribute partial scores: for some issues a checklist with minimum requirements is used, for some other issues decision trees are used.

This use of 'mixed methods' is probably to be encouraged so that a range of specific features linked to welfare assessment can best be taken into account. For instance, compensations may be allowed between measures that represent different viewpoints on the same problem (e.g. jointly considering difficulties both in lying down, and in getting up, to assess comfort around resting). However, it seems hazardous to consider that, for example, good health can fully compensate for the reduced ability to express natural behaviours. When several welfare dimensions are to be combined, non-compensatory methods such as the definition of minimal requirements seem more appropriate. Hence, different methods may be used at different stages in the construction of the overall assessment depending on the constraints imposed on the aggregation process. Therefore, the constraints imposed on the aggregation of welfare measures need to be clearly identified at a number of points in the aggregation process to generate meaningful results. These constraints are further described and analysed in Part 2 of the present dissertation on the overall assessment of animal welfare (Botreau *et al.*, 2007).

Acknowledgements

The present study is part of the Welfare Quality® research project, which has been co-financed by the European Commission, within the sixth Framework Programme, contract no. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

References

Algers B, Ekstrand C, Geismar J, Gunnarsson S, Odén K, Onila M and Svedberg J 1995. Utvärden av OLI voletage inhysningssystem för värphöns. I. enlighet med SJV:s program för förprövning av ny teknik. Swedish University of Agricultural Sciences, Skara, Sweden.

Algers B, Broom D, Canali E, Hartung J, Smulders F, Van Rennen CG and Veissier I 2006. Scientific opinion on the risk of poor welfare in intensive calf farming systems: an update of the Scientific Veterinary Committee Report on the Welfare of calves. The EFSA Journal 366, 1–144.

Barnett JL and Hemsworth PH 1990. The validity of physiological and behavioural measures of animal welfare. Applied Animal Behaviour Science 25, 177–187.

Bartussek H 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. Livestock Production Science 61, 179–192.

Bell DE, Raiffa H and Tversky A 1988. Decision making: descriptive, normative and prescriptive interactions. Cambridge University Press, Cambridge.

Bock B and van Leeuwen F 2005 Socio-political and market developments of animal welfare schemes. In Farm animal welfare concerns, consumers, retailers and producers – WelfareQuality® reports no. 1 (ed. J Roex and M Miele), pp. 115–167. Cardiff University.

Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, van Reenen CG and Veissier I 2007. Aggregation of measures to produce an overall assessment of animal welfare. Part 2: analysis of constraints. Animal 1.

Bouyssou D, Marchant T, Pirlot M, Perny P, Tsoukias A and Vincke P 2000. Evaluation and decision models – a critical perspective. Kluwer Academic Publishers, Dordrecht.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002. Decision support system for overall welfare assessment in pregnant sows. A model structure and weighting procedure. Journal of Animal Science 80, 1819–1834.

Capdeville J and Veissier I 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30, 62–68.

Dawkins MS 1980. Animal suffering: the science of animal welfare. Chapman & Hall Ltd, London.

Dawkins MS 1990. From an animal's point of view: motivation, fitness, and animal welfare. Psychological Science 13, 1–61.

El Balaa R, Marie M 2004. Evaluation du bien-être animal dans les élevages de petits ruminants. Animal welfare evaluation in small ruminant husbandry. Proceedings of the 11th Rencontres Recherches Ruminants, p. 210. Institut de l'Elevage – INRA, Paris.

Ekstrand C, Odén K, Gunnarsson S, Onila M, Algers B and Svedberg J 1997. Utvärden av OLI voletage inhysningssystem för värphöns. I. enlighet med SJV:s program för förprövning av ny teknik. Swedish University of Agricultural Sciences, Skara, Sweden.

EUREPGAP nd. EUREPGAP – the global partnership for safe and sustainable agriculture. Retrieved June 4, 2007, from http://eurep.org/.

European New Car Assessment Programme, 2004. Child protection assessment protocol. Retrieved June 4, 2007, from http://www.euroncap.com/Downloads/31f3c2e9-9206-4d74-9e2f-cc9908430621/Child-Assessment-Protocol-ver-1-0c.pdf.aspx.

Farm Animal Welfare Council 1992. FAWC updates the five freedoms. The Veterinary Record 17, 357.

Fraser D 1993. Assessing animal well-being: common sense, uncommon science. Proceedings of the food animal well-being conference proceedings and deliberations (ed. USDA and Purdue University Office of Agricultural Research Programs), pp. 37–54. Purdue University Office of Agricultural Research Programs, West Lafayette, Indiana.

Fraser D 1995. Science, values and animal welfare: exploring the 'inextricable connection'. Animal Welfare 4, 103–117.

Fraser D 2003. Assessing animal welfare at the farm and group level: the interplay of science and values. Animal Welfare 12, 433–443.

French S 1984. Fuzzy decision analysis: some criticism. TIMS/Studies in the Management Sciences 20, 29–44.

Friend TH 1980. Stress: what is it and how can it be quantified? International Journal for the Study of Animal Problems 1, 366–374.

Geers R, Petersen B, Huysmans K, Knura-Deszczka S, De Becker M, Gymnich S, Henot D, Hiss S and Sauerwein H 2003. On-farm monitoring of pig welfare by assessment of housing, management, health records and plasma haptoglobin. Animal Welfare 12, 643–647.

Hegelund L, Sørensen JT and Johansen NF 2003. Developing a welfare assessment system for use in commercial organic egg production. Animal Welfare 12, 649–653.

Horning B 2001. The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 42–47.

Humane Farm Animal Care, 2003. Certified humane raised and handled. Retrieved June 4, 2007, from http://certifiedhumane.com.

Hurnik JF 1990. World's poultry science association invited lecture: animal welfare: ethical aspects and practical considerations. Poultry Science 69, 1827–1834.

Huxley JN, Burke J, Roderick S, Main DCJ and Whay HR 2004. Animal welfare assessment benchmarking as a tool for health and welfare planning in organic dairy herds. Veterinary Record (The) 155, 237–239.

Johnsen PF, Johannesson T and Sandoe P 2001. Assessment of farm animal welfare at herd level: many goals, many methods. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 26–33.

Keeling L, Svedberg J 1999. Legislation banning conventional battery cages in Sweden and a subsequent phase-out programme. Proceedings of the Congress 'Regulation of Animal Production in Europe' (ed. M Kunisch and H Eckel), pp. 73–78.

Lacroix R, Huijbers J, Tiemessen R, Lefebvre D, Marchand D and Wade KM 1997. Fuzzy set-based analytical tools for dairy herd improvement. Applied Engineering in Agriculture 14, 79–85.

Lebart L and Fenelon JP 1975. Statistique et informatique appliqués. Editions Dunod, Paris.

Main DCJ, Webster F and Green LE 2001. Animal welfare assessment in farm assurance schemes. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 108–113.

Main DCJ, Kent JP, Wemelsfelder F, Ofner E and Tuyttens FAM 2003. Applications for methods of on-farm welfare assessment. Animal Welfare 12, 523–528.

Mounier L, Colson S, Roux M, Dubroeucq H, Boissy A, Ingrand S and Veissier I 2006. Links between specialization in the finishing of bulls, mixing, farmers' attitudes towards animals and the production of finishing bulls: a survey on French farms. Animal Science 82, 561–568.

Perny P 1998. Multicriteria filtering methods based on concordance and non-discordance principles. Annals of Operations Research 80, 137–165.

Piñeiro C, Morales M, Piñeiro M, Ruiz de la Torre JL, Mateos GG, Manteca X. 2005. Psychological stress caused by changes in feeding management did not affect the concentration of acute phase proteins in pigs. Proceedings of the Fifth International Colloquium on Animal Acute Phase Proteins.

Roy B 1993. Decision science or decision-aid science? European Journal of Operation Research 66, 184–204.

Royal Society for the Prevention of Cruelty to Animals undated. Flow chart for the Freedom Food scheme. Retrieved June 4, 2007, from http://www.rspca.org.uk/servlet/Satellite?pagename=RSPCA/RSPCARedirect&pg=Producer section&marker=1&articleId=1121758148522.

Rushen J 1986. The validity of behavioural measures of aversion: a review. Applied Animal Behaviour Science 16, 309–323.

Rutter SM, 1998. Assessing the welfare of intensive and extensive livestock. Proceedings of the Workshop Pasture Ecology and Animal Intake, pp. 1–9.

Scientific Committee on Animal Health and Animal Welfare of the European Commission undated. Reports from the Scientific Committee on Animal Health and Animal Welfare of the European Commission. Retrieved June 4, 2007, from http://ec.europa.eu/food/fs/sc/scah/outcome_en.html.

Scott EM, Nolan AM and Fitzpatrick JL 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 5–10.

Sørensen JT, Sandoe P and Halberg N 2001. Animal welfare as one among several values to be considered at farm level: the idea of an ethical account for livestock farming. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 11–16.

Spoolder H, De Rosa G, Horning B, Waiblinger S and Wemelsfelder F 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529–534.

Sundrum A and Rubelowski I 2001. The meaningfulness of design criteria in relation to the mortality of fattening bulls. Acta Agriculturæ Scandinavica, Section A, Animal Science Supplementum 30, 48–52.

Veissier I, Capdeville J and Delval E 2004. Cubicle housing systems for cattle: comfort of dairy cows depends on cubicle adjustment. Journal of Animal Science 82, 3321–3337.

Von Borell E, Bockisch FJ, Buscher W, Hoy S, Krieter J, Muller C, Parvizi N, Richter T, Rudovsky A, Sundrum A and Van Den Weghe H 2001. Critical control points for on-farm assessment of pig housing. Livestock Production Science 72, 177–184.

Webster J 1997. Applied ethology: what use is it to animal welfare? Advances in ethology 32Supplements to Ethology, 10.

Webster J 2005. The assessment and implementation of animal welfare: theory into practice. Revue Scientifique Et Technique-Office International Des Epizooties 24, 723–734.

Wechsler B 2001. Pretesting of mass-produced farm animal housing systems in Switzerland 20 years of experience. Proceedings of the international symposium of the second technical section of CIGR on animal welfare considerations in livestock housing systems (ed. Polish Committee of Agricultural Engineering), pp. 55–67. Poligmar, Zielona Góra, Poland.

Wechsler B 2003. Testing of mass-produced farm animal housing systems with regard to animal welfare. Proceedings of the 54th annual meeting of the European Association for Animal Production (ed. Y van der Honing), p. 125. Wageningen Pers, Wageningen.

Wechsler B 2005. An authorisation procedure for mass-produced farm animal housing systems with regard to animal welfare. Livestock Production Science 94, 71–74.

Whay HR, Main DCJ, Green LE and Webster AJF 2003. An animal-based welfare assessment of group-housed calves on UK dairy farms. Animal Welfare 12, 611–617.

Winckler C, Capdeville J, Gebresenbet G, Horning B, Roiha U, Tosi M and Waiblinger S 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Animal Welfare 12, 619–624.

Chapitre 3

Identification des contraintes pour une évaluation globale du bien-être animal

Résumé de l'identification des contraintes pour une évaluation globale du bien-être animal :

L'évaluation du bien-être animal correspond à un problème multicritère pour lequel la mise en place d'une procédure constructive est nécessaire afin de pouvoir synthétiser les nombreuses mesures en une évaluation globale. Cette construction dépend cependant des caractéristiques propres à l'évaluation du bien-être animal. Il convient donc, avant toute chose, d'identifier ces caractéristiques qui ont des répercussions directes sur le choix et le développement de la procédure d'évaluation globale.

Ces caractéristiques peuvent être liées :

- *au concept même de bien-être*, notamment le fait que le bien-être est une notion multidimensionnelle dont les différentes dimensions ne peuvent pas totalement se compenser (par ex. la bonne santé d'un animal ne peut pas entièrement compenser le fait que cet animal soit soumis à de fortes restrictions comportementales) ;
- *à l'interprétation des mesures en termes de bien-être*, ces mesures pouvant être plus ou moins pertinentes et apporter des contributions différentes à l'évaluation globale (certaines mesures peuvent être jugées plus importantes que d'autres) ;
- *à la façon dont les données sont recueillies sur le terrain*, les données brutes pouvant notamment être exprimées sur des échelles de natures très différentes (allant de données quantitatives au niveau individuel à des données ordinales directement mesurées au niveau du troupeau).

Une fois les caractéristiques propres à l'évaluation globale du bien-être animal identifiées, le modèle d'évaluation qui sera proposé dans cette thèse devra en tenir compte, notamment en choisissant les méthodes de synthèse de l'information les mieux adaptées.

Chapitre 3

Identification des contraintes pour une évaluation globale du bien-être animal

Avant de construire un outil d'évaluation globale du bien-être animal en ferme, les caractéristiques de l'évaluation globale du bien-être doivent être considérées. En effet, elles ont des répercussions directes sur le choix et le développement d'une procédure d'évaluation globale.

Ces caractéristiques peuvent être liées au concept même de bien-être, à l'interprétation des mesures en termes de bien-être, ou à la façon dont les données sont recueillies sur le terrain. Les contraintes méthodologiques qui pèsent sur la construction d'un modèle d'évaluation globale du bien-être ont été présentées et discutées dans l'article *"Aggregation of measures to produce an overall assessment of animal welfare - Part 2 : analysis of constraints"* (Botreau et al. 2007), figurant à la fin de ce chapitre, et dont nous résumons ici les principaux messages.

3.1 Caractéristiques liées au concept de bien-être animal

Le bien-être animal est défini au niveau de l'individu, alors que l'évaluation globale en ferme est produite au niveau du troupeau. Faut-il autoriser ou limiter les compensations entre individus ? Alors que certains prônent "le plus grand bonheur du plus grand nombre" (approche *utilitariste*, cf. par exemple Singer 1990), d'autres mettent en avant le fait que chaque individu a des droits essentiels qui doivent être respectés (approche *déontologique*, cf. par exemple Feinberg 1980, Regan 1992). En jugeant acceptable une

situation dans laquelle la majorité des animaux vivent dans de bonnes conditions alors qu'une minorité d'entre eux souffrent, les premiers autorisent pleinement les compensations entre individus. A l'inverse, les seconds partent du principe qu'on ne peut pas justifier de bons résultats si les moyens utilisés ont violé les droits d'au moins un individu. En d'autres termes, ils ne jugent pas les compensations entre individus comme éthiquement acceptables. Une approche plus pragmatique pourrait consister à préférer une situation dans laquelle tous les animaux vivent dans des conditions moyennes à une situation dans laquelle la moitié des animaux vivent dans d'excellentes conditions alors que l'autre moitié se retrouve dans des conditions très dégradées, sources de mal-être pour les animaux.

Le bien-être est une notion multidimensionnelle. Avant toute évaluation globale, il convient donc d'en identifier toutes les dimensions. De plus, ces dimensions ne peuvent probablement pas totalement se compenser. Par exemple, le bon état de santé ne compense pas les restrictions comportementales qui leur sont imposées (telles que l'isolement social). C'est ainsi que le premier rapport sur le bien-être des veaux établi par le Comité scientifique vétérinaire de la Commission Européenne (Broom et al. 1995, p. 97) conclut que le bien-être des veaux est médiocre dans de petites cases individuelles, malgré une incidence plus faible des maladies. Même si des débats existent encore autour de la compensation entre les dimensions du bien-être, il semble prudent d'utiliser des méthodes de synthèse de l'information qui permettent de limiter les compensations entre les dimensions du bien-être.

L'évaluation du bien-être animal doit suivre une procédure scientifiquement valide, tout en étant acceptée par ses utilisateurs potentiels (tels que des groupements de producteurs, des distributeurs, des associations de protection des animaux, des associations de consommateurs, ou encore des institutions officielles telles que l'Union Européenne ou l'OIE[1]). Comme cela a été dit dans le chapitre introductif, l'évaluation du bien-être animal implique forcément la définition de normes (au sens large) : qu'appelle-t-on un excellent niveau de bien-être ? qu'est-ce qui n'est pas acceptable ?

De ce fait les scientifiques ont pour rôle de produire un système d'évaluation reposant sur des mesures dont la validité aura été préalablement vérifiée. Toutefois, le choix des critères (sur lesquels on se base pour évaluer le niveau de bien-être) doit être partagé par les utilisateurs potentiels d'une telle évaluation globale du bien-être et les limites entre les différentes classes de bien-être (pouvant aller d'*"inacceptable"* à *"excellent"*) doivent être définies en concertation avec eux, car la science ne peut statuer sur ce qui

[1]Organisation Internationale des Épizooties

est acceptable ou non pour les animaux. Il ne peut y avoir d'évaluation purement scientifique du bien-être, des jugements de valeurs intervenant forcément dans sa réalisation (Fraser 1995 ; 2003), comme le précise Fraser (1995) *"our conception of animal welfare inherently involves value notion about what is better or worse, more important or less important, for the quality of life of animals"* . Selon Tannenbaum (1991, cité par Fraser 1995) les scientifiques ont tort de traiter l'évaluation du bien-être animal comme un sujet purement technique où les aspects scientifiques peuvent être examinés de façon totalement indépendante de considérations éthiques. Ceci est lié à la nature interdisciplinaire du bien-être animal (cf. Introduction) qui inclut des dimensions scientifiques, éthiques, économiques et politiques (Lund et al. 2006).

En effet, le scientifique peut donner un ordre de grandeur, il peut dire si une variation va dans le sens d'une amélioration ou d'une dégradation du bien-être, mais il ne peut pas dire si, dans l'absolu, quelque chose est bien ou non. Le scientifique a un avis éclairé mais ne peut pas donner de jugement de valeur (en tout cas pas uniquement en tant que scientifique, mais aussi en tant que citoyen).

Malheureusement, il se peut que les différents utilisateurs potentiels aient des opinions et des objectifs différents. Dans ce cas, il sera alors nécessaire d'identifier ces différences d'opinion, puis de trouver un terrain d'entente, commun à tous les utilisateurs. Pour ce faire, il faudrait produire un système d'évaluation facilement compréhensible par les utilisateurs et dont la conception aura impliqué des représentants de ces mêmes utilisateurs.

3.2 Caractéristiques liées à l'interprétation des mesures en termes de bien-être

La pertinence des mesures par rapport au bien-être animal doit être vérifiée. En effet, les mesures effectuées en ferme pour évaluer le bien-être des animaux ne permettent pas de mesurer directement l'état mental de bien-être des animaux. Elles n'en constituent que des indices qui ont besoin d'être interprétés en termes de bien-être. Par exemple, l'état d'engraissement d'une vache laitière ne peut servir de mesure pour évaluer la faim ressentie par l'animal que s'il est interprété en relation avec le stade de lactation de la vache. Lors du choix et du développement des mesures il est donc primordial de tester leur pertinence pour apprécier un ou plusieurs aspects du bien-être. Ceci pourra être réalisé en comparant ces mesures avec d'autres déjà considérées comme valides, en étudiant l'effet de différents traitements appliqués aux animaux (par ex. on pourrait valider une mesure de la relation homme-animal en comparant des animaux ayant reçus

des contacts doux *vs* rudes), ou à défaut en obtenant un consensus auprès d'experts du domaine.

Par ailleurs, certaines mesures peuvent être liées à plusieurs dimensions du bien-être. Par exemple, le faible état d'engraissement d'un animal peut être dû à un manque de nourriture ou à une maladie chronique. Le problème est alors d'évaluer les différentes dimensions du bien-être (liées à l'alimentation, à la santé...) en évitant d'utiliser plusieurs fois la même mesure interprétée de la même manière, ce qui risquerait de lui donner, par construction, plus d'importance dans l'évaluation globale. Il est cependant possible de tenir compte d'une même mesure pour évaluer différentes dimensions du bien-être à condition de l'interpréter de manière différente. Ainsi, il serait possible de focaliser sur les animaux maigres pour évaluer l'insuffisance de l'approvisionnement en nourriture des animaux, et de focaliser sur les animaux ayant un état d'engraissement trop élevé pour mettre en évidence d'éventuels risques pour la santé des animaux. Dans ce cas là, l'état d'engraissement est clairement interprété de manière différente dans les deux situations.

De plus, les mesures n'ont pas toutes la même importance quant à l'étendue du problème de bien-être qu'elles recouvrent. Par exemple, une blessure légère (type perte de poils) cause sans doute moins de douleur à l'animal qu'une boiterie avérée qui le gêne à chaque déplacement. Dans le cadre d'une évaluation globale du bien-être il est donc essentiel d'identifier l'importance relative de chacune des mesures à prendre en compte. Pour ce faire, il est souvent fait appel à des experts. Il est alors nécessaire de choisir les experts consultés en équilibrant les champs disciplinaires (vétérinaire, éthologistes, zootechniciens...) et les points de vue.

Enfin, certaines mesures peuvent être liées. Il existe par exemple chez les veaux une relation de cause à effet entre parasitisme digestif (inclus dans l'évaluation de la santé des veaux) et gain de poids (pouvant servir à évaluer si les animaux sont correctement nourris). En effet, Bangoura et Daugschies (2007) ont clairement montré que des veaux infectés par des spores de *Eimeria zuernii*, et donc atteints de coccidiose, présentent un gain de poids nettement inférieur à celui de veaux sains, ce qui a par conséquent des répercussions sur le poids des veaux. Au moment de la construction de l'évaluation globale du bien-être il est alors important d'identifier les relations pouvant exister entre les mesures afin d'éviter les erreurs d'interprétation. Dans l'exemple précédent, on pourrait s'abstenir d'évaluer si les veaux sont insuffisamment nourris (veaux trop maigres) chez les individus atteints de coccidiose (ce qui se traduit notamment par des diarrhées, c.-à-d. quelque chose de facilement repérable chez les veaux).

3.3 Caractéristiques liées à la collecte des données

En fonction des mesures, les données peuvent être exprimées sur différents types d'échelles. Certaines données sont quantitatives (par ex., la fréquence des comportements agonistiques, les distances de fuite...) alors que d'autres sont exprimées sur des échelles ordinales (par ex., la réaction face à l'homme pourra être considérée comme forte, modérée ou faible). Avec les échelles ordinales il est impossible de calculer des moyennes car de telles données ne fournissent des informations que sur le rang des objets les uns par rapport aux autres et les écarts entre les différents degrés de l'échelle ne sont pas forcément identifiables. Ainsi, une réaction forte est plus importante qu'une réaction modérée, laquelle est plus importante qu'une réaction faible, mais on ne sait pas si l'écart entre les catégories est le même. Par ailleurs, même avec des mesures quantitatives, l'interprétation en termes de bien-être ne suit pas forcément une fonction linéaire. Par exemple, deux vaches qui fuient l'homme respectivement à 6 et 8 m seront toutes deux considérées comme ayant très peur de l'homme, alors qu'une vache qui fuit l'homme à 2 m sera considérée comme étant nettement plus peureuse qu'une vache qui accepte de se faire toucher (ce qui correspond à une distance de fuite de 0 m). Pourtant, dans les deux cas, la différence de distance de fuite entre les animaux comparés est égale à 2 m.

Les mesures peuvent différer en précision. En effet, la répétabilité entre observateurs et entre moments d'observation varie d'une mesure à l'autre. Au cas où des mesures se révéleraient peu précises, il serait nécessaire d'améliorer l'évaluation en considérant simultanément plusieurs mesures relatives au même aspect du bien-être. C'est par exemple le cas de l'évaluation du confort autour du repos des vaches laitières, plusieurs mesures permettent d'évaluer cet aspect du bien-être mais elles sont généralement peu répétables, ou alors observables seulement sur une petit nombre d'animaux (et donc pas forcément représentatives de l'ensemble du troupeau). Cependant, en les considérant simultanément, elles sont susceptibles de fournir une bonne image de l'état du troupeau.

Il est parfois difficile d'évaluer la plage de variation d'une mesure au sein d'une population. Cependant, cette information est nécessaire pour pouvoir interpréter les mesures en fonction de la réalité de l'élevage, et ainsi faire en sorte que le système d'évaluation ne classe pas toutes le fermes dans la même catégorie, mais au contraire permette de les distinguer les unes par rapport aux autres. Par exemple, si on souhaite offrir la possibilité d'une marge de progression, le système ne devra pas considérer d'emblée que toutes les fermes sont excellentes.

Par ailleurs, si une catégorie *"excellente"* est définie, il conviendra de vérifier qu'il

existe d'ores et déjà des élevages qui en remplissent les conditions afin qu'appartenir à cette catégorie de bien-être apparaisse réalisable.

Un modèle d'évaluation globale du bien-être en ferme devrait donc tenir compte des caractéristiques propres à l'évaluation du bien-être, à l'interprétation et à la nature des données recueillies, et ce en choisissant les méthodes de synthèse de l'information les mieux adaptées.

Animal (2007), **1:8**, pp 1188–1197 © The Animal Consortium 2007
doi: 10.1017/S1751731107000547

animal

Aggregation of measures to produce an overall assessment of animal welfare. Part 2: analysis of constraints

R. Botreau[1,2†], M. B. M. Bracke[3], P. Perny[4], A. Butterworth[5], J. Capdeville[2]
C. G. Van Reenen[3] and I. Veissier[1]

[1]*INRA, UR1213 Herbivores, Site de Theix, F-63122 Saint-Genès-Champanelle, France;* [2]*Institut de l'Elevage, BP18, Castanet Tolosan F-31321, France;* [3]*Animal Sciences Group, Wageningen University and Research Centre, POB 65, Lelystad NL-8200 AB, The Netherlands;* [4]*Laboratoire Informatique de Paris 6, Université Pierre et Marie Curie, 8 rue du Capitaine Scott, Paris F-75015, France;* [5]*University of Bristol Clinical Veterinary Science, Langford, Bristol BS40 5DU, UK*

(Received 10 January 2007; Accepted 6 June 2007)

The overall assessment of animal welfare is a multicriterion evaluation problem that needs a constructive strategy to compound information produced by many measures. The construction depends on specific features such as the concept of welfare, the measures used and the way data are collected. Welfare is multidimensional and one dimension probably cannot fully compensate for another one (e.g. good health cannot fully compensate for behavioural deprivation). Welfare measures may vary in precision, relevance and their relative contribution to an overall welfare assessment. The data collected are often expressed on ordinal scales, which limits the use of weighted sums to aggregate them. A sequential aggregation is proposed in the Welfare Quality® project, first from measures to welfare criteria (corresponding to dimensions with pre-set objectives) and then to an overall welfare assessment, using rules determined at each level depending on the nature and number of variables to be considered and the level of compensation to be permitted. Scientific evidence and expert opinion are used to refine the model, and stakeholders' approval of general principles is sought. This approach could potentially be extended to other problems in agriculture such as the overall assessment of the sustainability of production systems.

Keywords: animal welfare, assessment, livestock, methodology

Introduction

During the last decade, numerous trade groups (producers, processors, retailers and restaurant chains) have developed certification systems with their suppliers. Many of these inspection schemes include a section on animal welfare (e.g. *IKB* (*Integrale KetenBeheersing*) by the Dutch meat industry, *Swedish Broiler Control*, *Filières Qualité Carrefour* in France and McDonald's Europe). However, there is no common standard for assessing animal welfare and for providing consumers with the relevant information.

One objective of the European Welfare Quality® project is to design an overall assessment of animal welfare that will help provide consumers with information on the products they buy (Blokhuis *et al.*, 2003). For this purpose, we need a system that (i) can be used routinely and on many different farms (throughout Europe), (ii) is sensitive to fluctuations in the welfare status of animals on these farms, (iii) reflects the welfare status of the herd as a whole, (iv) remains transparent for stakeholders (producers, retailers, consumers and citizens) and (v) corresponds to the current state of the art in animal welfare science.

Welfare is a multidimensional concept. Several requirements that have to be met to ensure welfare have been identified (e.g. the classical five freedoms (Farm Animal Welfare Council (FAWC), 1992)). This multidimensionality implies that welfare is most adequately assessed through a number of measures, each linked to a specific welfare dimension (or to several welfare dimensions). In turn, measures must be integrated to support an overall judgement (e.g. Bartussek, 1999; Capdeville and Veissier, 2001; Bracke *et al.*, 2002a). Although science can help assign a relative importance to welfare dimensions, the decision is inherently value based (Fraser, 1995). Nevertheless, science can help define welfare indices, and formalise the judgement made on their relative importance by societal groups.

In Part 1 of the present dissertation (Botreau *et al.*, 2007a), a review of the methods currently proposed to construct an overall welfare assessment from several measures is presented. To date, since each single method

† E-mail: rbotrea@clermont.inra.fr

presents advantages and disadvantages, none seems fully adequate to form the basis of a European standard for evaluating welfare in animal units (farms and slaughter plants) and to supply information on animal welfare to end-users. This review highlighted that (i) non-formal aggregation by experts has to be performed by the same expert assessors to generate comparable results, (ii) sum of ranks generates comparisons only within definite populations and (iii) weighted sums may be problematic because they allow full compensation between welfare aspects, which might conflict with the multidimensional nature of animal welfare.

Before constructing a tool for an overall assessment of animal welfare, the specific features linked to animal welfare assessment that constrain the aggregation of welfare measures should be considered. These features may be linked to the concept of welfare *per se*, to the interpretation of measures in terms of welfare, or to data collection. Here, we identify these specific features and suggest solutions to overcome current difficulties. These solutions are currently being investigated in the Welfare Quality® project, which will be briefly described.

Specific features linked to the concept of welfare

Welfare is a multidimensional concept

Welfare is a multidimensional concept. For example, the FAWC (1992) defined five basic requirements for animal welfare: freedom from hunger and thirst, freedom from discomfort, freedom from pain, injury or disease, freedom to express normal behaviour and freedom from fear and distress. Other arrangements of the basic requirements of animals have been proposed, e.g. the classification proposed by Fraser (1993), which consists of freedom from suffering, high levels of biological functioning and the existence of positive experiences. These dimensions are sometimes considered to be more or less independent, i.e. their fulfilment depends on different aspects of the environment that are likely to be experienced differently by animals. For instance, feeling sick is likely to be unrelated to feeling frustrated (for example, due to some behavioural activity being thwarted). The dimensions listed above may be, in turn, subdivided into separate independent items. For instance, feeling sick (nausea) and being injured have distinct causes and consequences for the animal despite being both categorised as elements of 'health'. Hence, many authors have advocated the consideration of more numerous aspects of welfare.

In a literature review, Bracke *et al.* (1999b) formulated 13 basic welfare needs: ingestion (including the need for food and water), rest, social contact, reproduction-related needs (including sexual interaction, nest building, maternal needs), kinesis/movement, exploration (including exploring novelty and foraging), play, body care, evacuation (defecation/urination), thermoregulation, respiration, health (including illness and injury-related needs) and safety (including the perception of danger and social conflict). These needs are considered to correspond to different biological control systems in which welfare-relevant emotions/feelings such as hunger, feeling ill, feeling pain, fear, etc. play a role. Capdeville and Veissier (2001) defined a similar list of 16 basic needs (no hunger, no thirst, no malnutrition, no physical stress, no climatic stress, no disease, no injury, feeding behaviour, locomotion, resting, social behaviour, sexual behaviour, maternal behaviour, no frightening events, opportunity for avoidance, good contacts with humans). Scientific reports on the welfare of certain animals produced by the European Food Safety Authority generally start with a list of the needs of the animals under consideration. For instance, in calves, 14 needs are identified ranging from needs essential for life (need to breath, need to sleep, etc.) to needs to perform certain behaviours like exploration or mastication of food (Algers *et al.*, 2006).

As with any multidimensional evaluation model, it is important to define a clear set of criteria, i.e. dimensions with set objectives (values to be achieved, or optimised), which can be used to check the welfare of an animal. The set of criteria should be exhaustive (i.e. all important aspects should be considered) and minimal (i.e. only aspects necessary to take a decision are included), and it should be possible to interpret any aspect separately from the other aspects. Additionally, the methods should have the approval of the stakeholders (Bouyssou, 1990). As regards farm animal welfare, these stakeholders may not only be scientists but may also be other people with an interest in animal welfare (consumers or citizens in general, producers, etc.). Based on these considerations, a set of 12 principles was proposed to develop systems for welfare monitoring in the Welfare Quality® project: absence of prolonged hunger, absence of prolonged thirst, comfort around resting, thermal comfort, ease of movement, absence of injuries, absence of disease, absence of pain induced by management procedures, expression of social behaviour, expression of other behaviour, good human–animal relationship, absence of general fear. Each item is assumed to be perceived by the animal differently from the others and (or) to be linked to different causal factors (for discussion see Botreau *et al.*, 2007c).

Animal welfare is defined at the individual level, whereas an overall assessment is usually produced at farm level

Animal welfare is generally defined at individual level. For instance, Broom (1986) defined the welfare of an animal in regards to the attempts this individual makes to cope with its environment. Dawkins (1980) stresses that production parameters, e.g. growth, morbidity or mortality, are not relevant to animal welfare, partly because they are taken at farm level: the production of a farm can be satisfactory even if some animals are in poor conditions.

The question of compensation between individuals has been addressed in animal ethics, with the two major theories – *utilitarianism* v. *deontology* (or *rights theory*) – likely to result in opposite interpretations.

Utilitarianism seeks 'the greatest good for the greatest number' (from Joseph Priestley, read by Bentham, 'father' of utilitarianism, in 1768). Hence, solutions that offer the best balance between the sum of satisfactions and the sum of frustrations are to be preferred. Some philosophers like Singer (1990), extended utilitarianism to animals. Utilitarianism does not imply good welfare for everyone, for while some animals may be living in good welfare conditions, a minority of others may suffer from bad welfare conditions. In this sense, utilitarianism allows compensation between animals.

By contrast, *deontology* stresses that any individual has basic rights (e.g. Feinberg, 1980) and means that, in principle, it is not possible to justify good results obtained using means which violate the rights of any single individual (Regan, 1992). In a deontological approach, compensation between individuals is not seen as ethical.

A pragmatic approach would consider animal suffering to be of prime importance while accepting that a certain percentage of animals will suffer. In an attempt to assess how this can work in practice, we consulted scientists working on animal welfare and asked them to score farms where various proportions of animals in more or less severe conditions related to lameness, injuries and body condition score were presented. In all cases, their answers showed that their judgement was more positive when all animals were in medium conditions than when some animals were in very poor conditions and some others in excellent conditions (e.g. an increase in animals which were not lame never outbalanced an increase of the same extent in animals which were severely lame) (Botreau *et al*., 2007b).

To make an evaluation at farm level, when welfare can be considered to have its effect on individuals it is proposed not to recommend the use of average values taken at farm level alone, but rather to try to describe also the variation within the farm by, for example, considering the standard variation of a continuous variable or by splitting measures into classes differing in severity.

Welfare dimensions may not fully compensate for each other

As already mentioned in Part 1 of the present dissertation (Botreau *et al*., 2007a), the various dimensions of animal welfare may not compensate each other.

Utilitarianism and deontology are both aimed at whole populations and can be used to debate whether compensation between individuals should be, from an ethical point of view, permitted or not. However, they can be both translated at the individual level to determine whether compensation between welfare dimensions should or should not be allowed. Transposed at individual level, utilitarianism leads to the conclusion that individuals try to maximise their state of wellbeing, expressed by the surplus of pleasure over distress. Hence, animals should be able to compensate bad situations on some welfare aspects by good situations on other aspects. Aerts *et al*. (2006), who shared this view, explained this by the capacity of animals to adapt to different situations. In the deontology perspective, each individual must have all his basic needs fulfilled, and the fact that this individual is far above basic requirements for some needs cannot compensate for basic requirements corresponding to other needs being not satisfied.

The most common approach to compensation between welfare dimensions lies between utilitarianism and deontology, allowing some compensation but never full compensation. For instance, Heleski *et al*. (2005) conducted a survey among veterinarians from 27 US veterinary colleges. They showed that 71% of respondents described their attitude toward farm animal welfare as 'we can use animals for the greater human good but have an obligation to provide for the majority of the animals' physiologic and behavioral needs'. Such a statement is clearly somewhere between *utilitarianism* (justifying the use of animals by the fact that this benefits to humans) and *deontology* (recognising that animals have rights). This intermediate view can result in a consideration that the level of animal welfare on a farm is good if a great majority of the animals present on the farm are living in good welfare conditions. At individual level, it could be possible to consider that compensations between welfare dimensions should be limited as they correspond to basic needs. This is illustrated in the first report on calf welfare by the Scientific Veterinary Committee of the European Commission (Broom *et al*., 1995, p. 97). This report concluded that the welfare of calves was very poor in small individual crates despite the lower incidence of disease. The update of this report stresses the difficulties in balancing the provision of social contacts against the increase in health problems (Algers *et al*., 2006, p. 61).

To check whether compensation could be allowed between the 12 welfare principles defined in the Welfare Quality® project (see above), we consulted 20 researchers involved in the project, 14 animal scientists and six social scientists. They were considered expert in animal welfare according to their previous researches and their contribution to the project. We asked them to produce welfare scores from various combinations of the fulfilment of the 12 principles. All except one considered that compensations should be highly limited between welfare dimensions (unpublished data).

Of course, the final judgement on whether we can allow compensation or not between welfare dimensions should come from animals themselves. Studies designed to produce an overall assessment of animal welfare have only recently started to appear in the literature. When several factors are likely to affect animals in a similar way (i.e. likely to influence the same variables measured on animals), interactions between factors can be analysed and if there is no interaction, then this suggests that items cannot compensate each other. For instance, in the work by Raussi *et al*. (2003), calves were exposed to different amounts of social and/or human contacts, with the hypothesis that human contacts would be more effective for

calves deprived of social contacts. However, no interactions between these two types of contact were found, and the authors concluded that human contacts cannot compensate for the lack of social contacts with animals of the same kind, and vice versa. To complete this approach, it would be very useful to design experiments where animals are offered several alternatives to check to what extent one alternative can compensate for the lack of another. Operant conditioning might help in this regard: two reinforcements each consisting of various proportions of two rewards could be presented in a concurrent schedule.

Taken together, ethical, societal and animal studies suggest that dimensions of welfare may not compensate for each other (or at least not fully). Therefore, a cautious approach is to use methods that can (but not necessarily) limit compensation between welfare dimensions.

To avoid having to fix whether compensations between welfare dimensions have to be limited or not, the process of aggregation can be stopped at criterion level. For instance, Beyer (1998) considered three different dimensions (housing system, animal care and management of the exercise yard) and stopped the aggregation process at this level. Thus, instead of producing one overall evaluation, this system yields three different scores, one per dimension. To limit compensation, while still providing an overall assessment, Capdeville and Veissier (2001) introduced specific rules to aggregate information so that a good score could not fully balance a bad one. Similarly, Mellor and Reid (1994) defined five component scores (one for each freedom) and suggested setting the overall score at the lowest component score. However, in this case, an improvement in any component other than the worst one will not increase the overall score, which may deter farmers from making improvements. For instance, a pig on slatted flooring that is seriously ill has poor welfare, but if this pig is provided with a comfortable straw-bed, its welfare will improve, at least to some extent, even though the main problem is the illness, not the floor. Another way to limit compensation is to add constraints by defining thresholds below which a value cannot be compensated for (minimum requirements) (Bracke *et al.*, 2002a; Spoolder *et al.*, 2003) or to use a more sophisticated algorithm that assigns higher weightings to lower component scores (e.g. Yager, 1988).

The welfare of an animal is interpreted by humans

Animal welfare should be assessed from the point of view of the animal (Dawkins, 1990). When a single aspect of welfare is considered, the animal's point of view may perhaps be obtained using measures of preferences (e.g. demand curves). However, it is much more difficult, although maybe not totally unrealistic, to determine how an animal would rank very different aspects of welfare occurring at different time scales, for instance, being afraid of something and being sick. The assessment of an animal's welfare as a whole will of course always be to some extent an assessment from a human point of view (see discussion in Fraser, 2003). The list of principles chosen in Welfare Quality® (see above) received preliminary agreement from focus groups of consumers (49 in all from seven European countries) and from the Advisory Committee of Welfare Quality® composed of representatives of main stakeholder groups (consumers, retailers, producers, animal welfare advocates and policy makers). The Advisory Committee will also be asked to give its views on the aggregation of criteria.

For an overall assessment model to be understood and accepted by all stakeholders (producers, consumers, citizens, scientists, etc.), the model should ideally lead to a consensus about what matters from the animal's point of view. Thus, a major challenge is identifying and reconciling points of difference between a scientific assessment of animal welfare and the opinion of the different stakeholders, in order to be able to communicate a welfare standard. This requires the principles of the methods used to aggregate information to be understandable, and understood, by lay people so that they can give their informed opinion and may impact on the fine-tuning of the assessment model.

Specific features linked to the welfare interpretation of the measures

Welfare is a prolonged mental state, resulting from how the animal experiences its environment over time (Dawkins, 1980; Duncan, 1996; Bracke *et al.*, 1999a). Measures used to assess animal welfare are not direct measures of mental state but only indices that need to be interpreted in terms of welfare. This introduces specific constraints.

The validity of measures in relation to animal welfare needs appraisal

Two types of measures are used to assess animal welfare; those measuring aspects of the animals' environment and those measuring aspects of the animals themselves (also known as design criteria and performance criteria) (Anonymous, 2001). The relationships between environment-based measures and their possible effects on animals are not always straightforward. For instance, a farmer may respond to suboptimal air conditions in a barn with better surveillance and health care. Animal-based measures such as behavioural and health parameters are generally considered to be more closely linked to the welfare of animals (Capdeville and Veissier, 2001; Winckler *et al.*, 2003; Whay *et al.*, 2003c). Ideally, the relation between animal-based measures and actual mental states of animals should be checked. However, the identification of mental states in animals is an emerging area of research (Desire *et al.*, 2002) and these relations are still unknown today. In practice, welfare measures to be used on farms are validated by comparing with more sophisticated methods (concurrent validity), by studying the effects of specific treatments (predictive validity), or alternatively by checking experts

agree on considering these measures as valid (validity based on scientific consensus).

In Welfare Quality®, concurrent validity will be given priority. When no reference method exists, predictive validity will be looked. However, in many cases we will be able to obtain only consensus validity for the time being of the project.

Some measures may be linked to several welfare dimensions

Different measures are needed to assess the different dimensions of welfare, e.g. body condition score may be used as an index of hunger, and the level of abnormal behaviour as an indicator of the animal's inability to express natural behaviour. Some measures may provide information about more than one welfare dimension. For example, a low body condition score can be due to prolonged lack of food (hunger) or a chronic disease (health). Similarly, elevated cortisol levels may be indicative of a wide range of welfare problems. The same problem (of one measure being indicative of multiple welfare dimensions) also arises with environment-based measures. For instance, space affects numerous aspects of welfare, including mobility and social contact. Consequently, this measure can have a greater impact on the overall assessment than a measure linked to only one dimension (Horning, 2001). Unwarranted double counting, however, has to be avoided. This can be limited when the interpretation of the measure in terms of welfare is different from one dimension to another, as in Bracke *et al*. (2002a), where space for exercise and space for social interactions were distinguished.

Measures do not all have the same importance for animal welfare

The assignment of relative weightings to determine the contribution of a measure to overall welfare is always a critical point. There is a general risk of assigning less importance to dimensions that are described by fewer measures. For instance, it is common to assess the comfort of the resting area through several indices (e.g. for cows: difficulties in lying down and in getting up, dog sitting positions, position of the cows in a stall or cubicle, animals lying outside the lying area), while lameness may be assessed with only one measure (e.g. % of lame animals) as in Capdeville and Veissier (2001). Physical comfort should not be given a greater importance than lameness in the overall assessment solely because it requires more measures. Lameness probably affects animals more than discomfort around resting since it is a painful condition. To overcome this problem, the welfare dimensions that need to be covered (i.e. welfare criteria and possibly subcriteria) must be clearly identified and assigned a relative importance. Only then, for each dimension covered by several measures, should the relative contribution of the different measures to that dimension be identified.

Again, the point of view of animal should be taken into account when this is available. However, most experiments designed to assess the preferences of animals offer several alternatives of the same nature: several foods or several types of housing, etc. (e.g. Klopper *et al*., 1981; Barber *et al*., 2004). It would be of value to consider which aspect of welfare the animal prioritises as more important, e.g. health, comfort around resting or the possibility to express behaviour, but experiments providing such information have not yet been carried out.

Expert opinion can be used (i) when no study has been yet run to address a specific point (but related studies can help form an opinion of what is most probable to be) and/or (ii) when scientific evidence alone cannot solve a problem (Roqueplo, 1997). Point (i) is highly relevant for animal welfare due to lack of information on the importance that animals may attribute to the various welfare aspects, but the consequences of the non-fulfilment of each aspect can be assessed thanks to indirect parameters such as mortality, morbidity, expression of behaviour (Algers *et al*., 2006). Point (ii) is also very relevant for animal welfare. As described by Fraser (1995), animal welfare cannot be addressed in an entirely objective way and the importance attributed to the various dimensions of welfare is inevitably value based. As a consequence, the weighting of various welfare aspects can be based on expert opinion.

Butterworth *et al*. (2004) and Haslam and Kestin (2003) used conjoint analyses to define weightings: a limited number of measures were selected with pre-set possible levels, and permutations between possible values of measures were then presented to experts who were asked to give their opinion. Weightings were then extracted assuming a linear combination of measures. This method is only possible with a limited number of measures and with pre-set values for these measures (e.g. Yes/No or few ordinal levels). When these conditions are not met, experts can be consulted using the Delphi method as in Anonymous (2001) or Whay *et al*. (2003b), where the arguments and opinions of experts are taken into account in several proposals following an iterative procedure. Alternatively, weighting coefficients can be derived from a classification of scientific evidence (Bracke *et al*., 2002a) and further validated by comparison with expert opinion (Bracke *et al*., 2002b). In general, it is not recommended to ask experts directly to assign weightings to measures because weightings depend on the method used to aggregate the measures. Experts can be asked to give their opinion on situations or data sets, and weightings can then be calculated to match their answers. This is the strategy followed in Welfare Quality®.

The choice of experts can have a large impact on the results. For instance, veterinarians might attribute more importance to health factors, where ethologists may attribute more importance to the expression of normal behaviour. Hence, each time experts are consulted, one has to ensure that experts are chosen according to their knowledge of the field and that various points of views are balanced within the group of experts (Roqueplo, 1997).

Relations may exist between measures

Besides the characteristics of the data, any aggregation process should take into account the links between measures (Bouyssou, 1990).

Welfare measures may be affected by external factors which do not directly affect welfare *per se*, and such measures may need to be corrected for a meaningful welfare interpretation to be made. For instance, in dairy cows, body condition scores may need to be corrected for the stage of lactation. Similarly, when assessing pain, lameness scores may need to be corrected for the walking requirements imposed by the housing system. For example, lameness may have more severe consequences for cows at pasture than when they are in tie-stalls, because of the different walking requirements, e.g. to get food and water.

Measures used to assess welfare may also affect one another. For example, cows that suffer from lameness have shorter avoidance distances, a measure that is often taken to assess cows' fear of humans (Špinka *et al*., 2005). A solution to avoid misinterpretation could be to exclude lame animals when assessing fear responses.

When constructing a multicriterion evaluation, such relationships between measures must be identified in order to avoid double counting (Bouyssou, 1990). When links between measures are unavoidable then those measures should be considered jointly (i.e. in the same criterion).

Specific features linked to the collection of data

Data can be collected on different types of scale depending on the measures

Measures to assess animal welfare are generally expressed with units on several types of scale (Scott *et al*., 2001). Some measures may be expressed on cardinal (i.e. quantitative) scales, as for example the frequency of aggressive behaviour during a defined period, or the flight distance of an animal when a human approaches. Other measures, however, are expressed on ordinal scales, i.e. observations are assigned to ordered categories. For instance, we may consider that an animal has mild, moderate or strong reactions to handling. 'Strong' is greater than 'moderate' and 'moderate' is greater than 'mild'. For ordinal scales, average scores may be misleading. In the above example, if all the animals respond moderately the same 'average' is obtained as if one half of the group does not respond and the other half responds strongly. But, since the difference between 'mild' and 'moderate' may be smaller (or larger) than the difference between 'moderate' and 'strong', the first group may actually be on average less (or more) responsive than the second.

Furthermore, the relation between a measure and its interpretation in terms of animal welfare is often not proportional. For instance, two animals that flee when the experimenter is 8 and 6 m away will be considered similar in terms of fear of humans. By contrast, an animal that flees when the experimenter is 2 m away will be considered more frightened by humans than an animal that accepts being touched (i.e. flight distance of 0 m). Yet, in both cases the difference in flight distance is actually the same (2 m).

In some cases, optimal values may be neither minimum nor maximum values, but somewhere in between. For instance, for gregarious species, both social isolation and being in a very large group may be detrimental to animal welfare: being in a small group, which corresponds to what is observed in natural conditions may be most appropriate. In poultry, feather pecking is much more frequent in large groups (>100 hens) than in small groups of 15 to 60 birds (Bilcik and Keeling, 2000).

Hence, when raw data are converted onto a value scale, from 'poor welfare' to 'good welfare', the conversion may not necessarily be a linear function. Again this function may be derived from expert opinion.

When a welfare assessment is performed, raw data are often converted into scores on a value scale composed of discrete values. The scores are then often processed as cardinal data (interval or ratio), whereas in some cases it would be more appropriate to consider them as ordinal data and avoid calculating means or sums.

Measures differ in precision

Measures used to assess welfare may range in precision. For example, there can be variations between observers and between days of observation. Environment-based measures often tend to be more reliable than animal-based measures. For instance, it is easy to measure the length of a stall, and it is unlikely that this will vary greatly between observers and from one day to another. By contrast, measures taken on animals tend to be subject to variation. For instance, when estimating the reaction of an animal facing a human on a predefined ordinal scale (e.g. no/mild/moderate/strong/very strong reaction), the observer may hesitate between two successive levels of the scale. In addition, the behaviour of an animal is never exactly the same between repetitions of the test and so it is essential to determine the reliability of each welfare measure. An aggregation model should take into account variation in reliability between measures. This can be done, for example, by setting discrimination thresholds, below which differences between values are not considered significant (Perny, 1998 or Bouyssou *et al*., 2000, p. 180), or by the use of fuzzy logic (Lacroix *et al*., 1998).

In Welfare Quality®, the method used to construct criteria varies depending on the precision of measures, and, in the case of measures with low precision, it is the intention to make as many measures as is practically possible to derive a criterion score.

Data are sometimes missing

A problem inherent to measures taken under practical conditions, as opposed those taken in experimental conditions, is the difficulty in recording them, and this can result in missing data. For instance, health records may not be kept accurately, so that disease prevalences cannot be

assessed properly. Missing values may be handled by substituting other related data, however, in some cases, it may prove impossible to carry out the assessment because critical data are missing.

It is sometimes difficult to assess the range of variation of a measure within a population

Another problem for overall welfare assessment arises when the range of variation of a measure is not known. Observations in experimental conditions are usually not sufficient to obtain such information and it may be necessary to run observations on a large scale on farms (or during transport, or at slaughter). These observations are very demanding and are rarely carried out to identify the distribution of all measures. In Welfare Quality®, surveys in a range of European countries are planned and it is hoped this will result in the required information on minimum, maximum, mean and standard error for cardinal measures, and information on percentiles for qualitative and ordinal measures. This information is needed to adjust the aggregation model to the characteristics of the measures, so that the overall assessment remains sensitive (i.e. farms that apparently differ do not obtain the same final assessment).

Discussion

An ideal model for an overall assessment of animal welfare should deal adequately with all the requirements linked to welfare assessment. The most challenging requirement is that welfare is a multidimensional concept and dimensions may not fully compensate for one another. Methods used to synthesise information can be more or less compensatory, and the choice of a method shall depend on the level of compensation one wants to allow. In addition, welfare measures vary in precision, relevance and importance, and the type of data collected have prompted us to search for calculations other than simple (and intuitively appealing) weighted sums.

Assessing welfare involves describing how well the animals experience their world based on the best possible judgement of their situation. This judgement requires detailed knowledge of the available scientific information. Such information is necessary to avoid errors in interpreting a given measure (e.g. wallowing in pigs could be interpreted solely as a sign of contentment, whereas this behaviour is often displayed when the animal is overheated (Baldwin and Ingram, 1967)). But this judgement cannot solely be based on science and on the data collected from experiments. As mentioned by Fraser (1995), the assignment of a relative importance to welfare dimensions is at least partly subjective. Researchers need to be confident that these dimensions and their relative importance match the expectations of societal groups involved in the keeping, selling or protection of farm animals. In addition, science cannot tell us what is socially acceptable or not – and so threshold values are usually set according to expert opinion (e.g. disease prevalence values above which welfare is considered to be poor and where remedial measures are required at herd level can be set from veterinary advice) (Whay *et al.*, 2003a). Finally, ethical, economical and political issues may also come into play (Commission for the European Communities, 2002).

The great variability of evaluation problems encountered in practice has led scientists with different backgrounds (management sciences, mathematical psychology, economics, operations research and computer sciences) to develop a variety of formal models and methodologies in decision theory to support evaluation tasks and decision making activities, e.g. the outranking approach based on ordinal aggregation methods, multiattribute utility theory (MAUT; e.g. Roy, 1996) based on cardinal aggregation methods (additive or non-additive, e.g. Choquet integrals (Grabisch, 1996) and generalised additive independence utility (Gonzales and Perny, 2005), etc.). These models and methodologies concern different important issues relevant to multidimensional evaluation such as measurement problems (preferences, perceptions and performance by numerical information), aggregation problems (overall evaluations from multidimensional and possibly conflicting viewpoints), uncertainty and imprecision modelling (Vincke, 1992; Roy, 1996; Bouyssou *et al.*, 2000).

The objectives of Welfare Quality® are (i) to construct a standardised overall welfare assessment for cattle, pigs and poultry for use at a European scale and (ii) to establish a standardised Europe-wide communication system for information on products (Blokhuis *et al.*, 2003). The overall assessment produced by Welfare Quality® could be included in an information system (assigning animal units to a few predefined categories, from low welfare to high welfare), on a voluntary basis. To have a complete evaluation of the welfare level of animals throughout their lives, farms, hauliers and slaughterhouses will have to be assessed.

To make an overall assessment of animal welfare, it is planned to follow a hierarchical aggregation process, going first from the measures (performed in the field, e.g. on farms) to the 12 welfare principles identified in Welfare Quality® (Botreau *et al.*, 2007b). These principles are absence of prolonged hunger, absence of prolonged thirst, comfort around resting, thermal comfort, ease of movement, absence of injuries, absence of disease, absence of pain induced by management procedures, expression of social behaviour, expression of other behaviour, good human–animal relationship and absence of general fear. To facilitate the communication with consumers, these 12 elements have been grouped into four main criteria: feeding, housing, health and behaviour (and called the 12 welfare basic elements 'subcriteria'). These four criteria will subsequently be aggregated to form an overall assessment (Table 1).

The subcriteria will be constructed using mathematical methods (e.g. weighted sums, comparison with minimal requirements) chosen according to the number of measures contained in each subcriterion, their nature and the precision with which it is assumed they can be made.

Table 1 *Hierarchical structure designed in Welfare Quality® to integrate numerous measures (~30) into an overall welfare assessment*

Measures (examples)	Principles	Criteria	Overall assessment
~30	12	4	1
+	Compensations		-
Body condition score	Absence of prolonged hunger	Good feeding	Overall assessment
Quantity of water provision	Absence of prolonged thirst		
Cleanliness, abnormal rising	Comfort around resting	Good housing	
Panting	Thermal comfort		
Tethering, slipping	Ease of movement		
Injuries, lameness	Absence of injuries	Good health	
Mastitis, diarrhoea	Absence of disease		
Dehorning, tail docking	Absence of pain induced by management procedures		
Agonistic behaviours	Expression of social behaviours	Appropriate behaviour	
Stereotypical behaviours	Expression of other behaviours		
Avoidance distance	Good human-animal relationship		
Reaction facing a novel situation	Absence of general fear		

First, 12 principles will be checked thanks to a combination of relevant measures. Second, the information will be compounded into four criteria and finally aggregated to form one overall assessment. Different mathematical methods will be used to process the information allowing decreasing the level of compensation along the hierarchical structure.

The appropriate subcriteria will be combined to evaluate each criterion. The method chosen to aggregate the subcriteria into criteria will limit compensations when this appears to better match evaluation given by experts, by assigning more importance to the lowest subcriterion-scores, thus hopefully encouraging producers to correct the more severe problems first. The aggregation of criteria, to create an overall assessment, will be performed using comparisons with pre-set profiles in order to be able to limit further compensations. Hence, the higher the aggregation is in the hierarchical structure the more limited may be the compensation between components.

Stakeholders are involved during the construction of the assessment method to enhance its potential for further implementation in practice. The information resulting from this evaluation will need to be expressed in a compounded way to inform consumers. Finally, the method will have to remain flexible enough to follow the evolution of farming, transport and slaughter conditions, and that of societal concerns, and, last but not least, to be updated according to the state of the art in animal welfare science and societal expectations.

An ideal method for the overall assessment of animal welfare should satisfy all the specific features and recommendations presented here. Even though this may not prove totally feasible, we intend to produce a 'best possible' method for the overall assessment of animal welfare based on scientific evidence, expert opinion, and stakeholders' views. This work is in progress and the final model of overall welfare assessment will be described in forthcoming papers.

In conclusion, the routine overall assessment of animal welfare needs a formal model of multicriterion evaluation. The construction of such a model requires bridging animal sciences, social sciences and methodologies developed in decision theory. The design of the strategy outlined in Welfare Quality® may also be applicable to a wide range of similar problems found in animal production such as defining an overall model for the assessment of sustainability of farming systems.

Acknowledgements

The present study is part of the Welfare Quality® research project which has been co-financed by the European Commission, within the sixth Framework Programme, contract no. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

References

Aerts S, Lips D, Spencer S, Decuypere E and Tavernier J 2006. A new framework for the assessment of animal welfare: integrating existing knowledge from a practical ethics perspective. Journal of Agricultural and Environmental Ethics 19, 67–76.

Algers B, Broom D, Canali E, Hartung J, Smulders F, van Rennen CG and Veissier I 2006. Scientific opinion on the risk of poor welfare in intensive calf farming systems: an update of the Scientific Veterinary Committee Report on the Welfare of Calves. The EFSA Journal 366, 1–144.

Anonymous 2001. Scientists' assessment of the impact of housing and management on animal welfare. Journal of Applied Animal Welfare Science 4, 3–52.

Baldwin BA and Ingram DL 1967. Behavioural thermoregulation in pigs. Physiology and Behavior 2, 15–16.

Barber CL, Prescott NB, Wathes CM, Le Sueur C and Perry G 2004. Preferences of growing ducklings and turkey poults for illuminance. Animal Welfare 13, 211–224.

Bartussek H 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. Livestock Production Science 61, 179–192.

Beyer S 1998. Konstruktion und Überprüfung eines Bewerungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit. Wissenschftlicher Fachverlag, Giessen.

Bilcik B and Keeling LJ 2000. Relationship between feather pecking and ground pecking in laying hens and the effect of group size. Applied Animal Behaviour Science 68, 55–66.

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. Animal Welfare 12, 445–455.

Botreau R, Bonde M, Butterworth A, Perny P, Bracke MBM, Capdeville J and Veissier I 2007a. Aggregation of measures to produce an overall assessment of animal welfare. Part 1: a review of existing methods. Animal 1, 1179–1187.

Botreau R, Perny P, Capdeville J and Veissier I 2007b. Construction of product information from animal welfare assessment. Proceedings of the Second Welfare Quality® Stakeholder Conference, pp. 33–36.

Botreau R, Veissier I, Butterworth A, Bracke MBM and Keeling LJ 2007c. Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, 225–228.

Bouyssou D 1990. Building criteria: a prerequisite for MCDA. In Readings in multiple criteria decision-aid (ed. CA Bana e Costa), pp. 58–80. Springer Verlag, Heidelberg.

Bouyssou D, Marchant T, Pirlot M, Perny P, Tsoukias A and Vincke P 2000. Evaluation and decision models – a critical perspective. Kluwer Academic Publishers, Dordrecht.

Bracke MBM, Spruijt BM and Metz JHM 1999a. Overall animal welfare assessment reviewed. Part 1: is it possible? Netherlands Journal of Agricultural Science 47, 273–291.

Bracke MBM, Spruijt BM and Metz JHM 1999b. Overall animal welfare reviewed. Part 3: welfare assessment based on needs and supported by expert opinion. Netherlands Journal of Agricultural Science 47, 307–322.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a. Decision support system for overall welfare assessment in pregnant sows A: model structure and weighting procedure. Journal of Animal Science 80, 1819–1834.

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002b. Decision support system for overall welfare assessment in pregnant sows B: validation by expert opinion. Journal of Animal Science 80, 1835–1845.

Broom DM 1986. Indicators of poor welfare. The British Veterinary Journal 142, 524–526.

Broom DM, Blokhuis WJ, Canali E, Dijkhuizen AA, Fallom R, Le Neindre P, Saloniemi H and Webster AJF 1995. Report of the scientific veterinary committee, Animal welfare section on the welfare of calves. Commission of the European Communities.

Butterworth A, Sadler L, Knowles TG and Kestin SC 2004. Evaluating possible indicators of insensibility and death in Cetacea. Animal Welfare 13, 13–17.

Capdeville J and Veissier I 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 62–68.

Commission for the European Communities 2002. Communication from the commission to the Council and the European Parliament on animal welfare legislation on farmed animals in third countries and the implications for EU.

Dawkins MS 1980. Animal suffering: the science of animal welfare. Chapman and Hall Ltd, London.

Dawkins MS 1990. From an animal's point of view: motivation, fitness, and animal welfare. Psychological Science 13, 1–61.

Desire L, Boissy A and Veissier I 2002. Emotions in farm animals: a new approach to animal welfare in applied ethology. Behavioural Processes 602, 165–180.

Duncan IJH 1996. Animal welfare defined in terms of feelings. Acta Agriculturae Scandinavica, Section A, Animal Science, Supplementum 27, 29–35.

Farm Animal Welfare Council 1992. FAWC updates the five freedoms. The Veterinary Record 17, 357.

Feinberg J 1980. Human duties and animal rights. In Rights, justice and the bounds of liberty (ed. J Feinberg), pp. 185–206. Princeton University Press, Princeton.

Fraser D 1993. Assessing animal well-being: common sense, uncommon science. Proceedings of the food animal well-being-conference proceedings and deliberations (ed. USDA and Purdue University Office of Agricultural Research Programs), pp. 37–54. Purdue University Office of Agricultural Research Programs, West Lafayette, Indiana.

Fraser D 1995. Science, values and animal welfare: exploring the 'inextricable connection'. Animal Welfare 4, 103–117.

Fraser D 2003. Assessing animal welfare at the farm and group level: the interplay of science and values. Animal Welfare 12, 433–443.

Gonzales C, Perny P 2005. GAI networks for decision making under certainty. Proceedings of the 19th international joint conference on artificial intelligence – multidisciplinary workshop on advances in preference handling (ed. R Brafman and U Junker), pp. 100–105.

Grabisch M 1996. The application of fuzzy integrals in multicriteria decision making. European Journal of Operational Research 89, 445–456.

Haslam SM and Kestin SC 2003. use of conjoint analysis to weight welfare assessment measures for broiler chickens in UK husbandry systems. Animal Welfare 12, 669–675.

Heleski CR, Mertig AG and Zanella AJ 2005. Results of a National Survey of US Veterinary College Faculty regarding attitudes toward farm animal welfare. Journal of the American Veterinary Medical Association 226, 1538–1546.

Horning B 2001. The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 42–47.

Klopper FD, Kilgour R and Matthews LR 1981. Paired comparison analysis of palatabilities of twenty foods to dairy cows. Proceedings of the New Zealand Society of Animal Production 41, 242–247.

Lacroix R, Strasser M, Kok R and Wade KM 1998. Performance analysis of a fuzzy decision-support system for culling of dairy cows. Canadian Agricultural Engineering 40, 139–152.

Mellor DJ and Reid CSW 1994. Concepts of animal well-being and predicting the impact of procedures on experimental animals. In Proceedings of the improving the well-being of animals in the research environment (ed. RM Baker, G Jenkin and DJ Mellor), pp. 3–18. ANZCCART, Sydney.

Perny P 1998. Multicriteria filtering methods based on concordance and non-discordance principles. Annals of Operations Research 80, 137–165.

Raussi S, Lensink BJ, Boissy A, Pyykkönen M and Veissier I 2003. The impact of social and human contacts on calves' behaviour and stress responses. Animal Welfare 12, 191–203.

Regan T 1992. Pour les droits des animaux. Les Cahiers Antispécistes 5.

Roqueplo P 1997. Entre savoir et décision, l'expertise scientifique. INRA editions, Paris, France.

Roy B 1996. Multicriteria methodology for decision aiding. Kluwer Academic, Dordrecht.

Scott EM, Nolan AM and Fitzpatrick JL 2001. Conceptual and methodological issues related to welfare assessment: a framework for measurement. Acta Agriculturæ Scandinavica, Section A, Animal Science, Supplementum 30, 5–10.

Singer P 1990. The significance of animal suffering. Psychological Science 13, 9–12.

Špinka M, Dembele I, Panamá J and Stìhulová I 2005. Lame dairy cows have shorter avoidance distances. Proceedings of the 39th international congress of the International Society for Applied Ethology, Sagamihara, Japan.

Spoolder H, De Rosa G, Horning B, Waiblinger S and Wemelsfelder F 2003. Integrating parameters to assess on-farm welfare. Animal Welfare 12, 529–534.

Vincke P 1992. Multicriteria decision aid. Wiley, New York, USA.

Whay HR, Main DCJ, Green LE and Webster AJF 2003a. assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records. The Veterinary Record 153, 197–202.

Whay HR, Main DCJ, Green LE and Webster AJF 2003b. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: consensus of expert opinion. Animal Welfare 12, 205–217.

Whay HR, Main DCJ, Green LE and Webster AJF 2003c. An animal-based welfare assessment of group-housed calves on UK dairy farms. Animal Welfare 12, 611–617.

Winckler C, Capdeville J, Gebresenbet G, Horning B, Roiha U, Tosi M and Waiblinger S 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Animal Welfare 12, 619–624.

Yager RR 1988. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190.

Chapitre 4

Structure générale du modèle d'évaluation développé dans la thèse

Résumé de la structure générale du modèle d'évaluation développé dans la thèse :

Le bien-être étant un concept multidimensionnel, son évaluation globale en routine nécessite un modèle d'évaluation multicritère. La première étape de la construction d'un tel modèle repose sur la définition de la famille de critères sur lesquels les fermes seront évaluées en termes de bien-être animal. Cette famille de critères doit être exhaustive (n'oubliant aucune dimension du bien-être) tout en restant minimale (afin d'éviter les redondances). Les critères doivent être isolables, c.-à-d. pouvoir s'interpréter indépendamment les uns des autres. Il est important que ces critères soient opérationnels, c.-à-d. qu'ils puissent constituer une base solide et commune à l'ensemble des acteurs de la décision (dans notre cas, les scientifiques mais aussi les utilisateurs potentiels du système d'évaluation) pour la poursuite de la procédure d'aide à la décision. Enfin, la famille de critères doit rester lisible (en ne contenant qu'un petit nombre de critères).

Nous proposons une famille composée de quatre critères (l'alimentation adaptée, le logement correct, la bonne santé et le comportement approprié), subdivisés en 12 sous-critères (par ex. l'alimentation adaptée se décompose en deux sous-critères : l'absence de faim et l'absence de soif).

En conséquence, nous proposons un modèle d'évaluation basé sur une agrégation séquentielle en évaluant d'abord les 12 sous-critères à partir des mesures de bien-être observées en ferme, puis en les agrégeant afin d'évaluer les quatre critères, et enfin en synthétisant l'information fournie par les critères en une évaluation globale.

Chapitre 4

Structure générale du modèle d'évaluation développé dans la thèse

Nous avons décidé d'explorer les méthodes développées en recherche opérationnelle pour l'évaluation multicritère, et reposant sur l'identification et l'évaluation des différentes dimensions (critères) composant l'objet à évaluer (ici : le bien-être animal). La première étape de cette approche constructive est donc de définir la famille de critères à prendre en compte, puis de dresser la stratégie à suivre pour les construire et les agréger en une évaluation globale à l'échelle de l'élevage.

4.1 Structure générale d'un processus de décision

Tsoukiàs (2007) considère le modèle d'évaluation $\mathcal{M}$ comme le n-uplet suivant :

$$\mathcal{M} = \langle \mathcal{A}, \{\mathcal{D}, \mathcal{E}\}, \mathcal{H}, \mathcal{U}, \mathcal{R} \rangle$$

où :

- $\mathcal{A}$ correspond à l'ensemble des alternatives sur lesquelles le modèle s'applique. Dans notre cas, il s'agit de l'ensemble des fermes pouvant être concernées par le système d'évaluation du bien-être animal.
- $\mathcal{D}$ correspond à l'ensemble des dimensions (ou attributs) sur lesquelles les éléments de $\mathcal{A}$ sont évaluées. Il s'agit pour nous de l'ensemble des mesures permettant d'évaluer le bien-être en ferme et dont la liste a été établie par le WP2.2.

- $\mathcal{E}$ correspond à l'ensemble des échelles sur lesquelles les attributs sont évalués. De manière formelle, pour toute alternative i et tout attribut d, cela revient à écrire : $\forall i \ \forall d \in \mathcal{D},\ d_i : \mathcal{A} \longrightarrow E_i \in \mathcal{E}$. Comme nous le verrons par la suite, dans notre cas, les échelles $E \in \mathcal{E}$ peuvent être de natures très différentes.
- $\mathcal{H}$ correspond à l'ensemble des critères sur lesquels les éléments de $\mathcal{A}$ sont évalués. Un critère étant une relation de préférence formalisant une relation binaire entre deux alternatives ou une fonction représentant la valeur, dans notre cas en termes de bien-être animal, de chaque alternative sur une échelle d'utilité.
- $\mathcal{U}$ correspond à l'ensemble des structures d'incertitude et/ou états épistémiques s'appliquant à $\mathcal{D}$ et $\mathcal{H}$, et qui sont exprimés par les décideurs (dans notre cas les scientifiques et les autres utilisateurs potentiels du système d'évaluation).
- $\mathcal{R}$ correspond à l'ensemble des opérateurs permettant à l'information disponible sur $\mathcal{A}$ d'être synthétisée en une évaluation globale, en passant pour cela par $\mathcal{D}$ et $\mathcal{H}$. Ces opérateurs doivent avoir un sens à la fois d'un point de vue théorique (comme par exemple s'assurer de l'adéquation des traitements à la nature des données) et d'un point de vue opérationnel (les méthodes utilisées devant être facilement explicables aux utilisateurs futurs de sorte à ce que ces derniers puissent comprendre et utiliser les résultats issus du modèle).

Le modèle d'évaluation globale du bien-être des vaches laitières en ferme que nous, analystes, devons proposer aux décideurs repose donc sur la définition de l'ensemble des éléments décrits ci-dessus. Pour l'instant, $\mathcal{A}$ correspond potentiellement à l'ensemble des fermes laitières européennes, mais en fonction de l'utilisation qui sera faite du système d'évaluation proposé $\mathcal{A}$ ne pourra correspondre qu'à une petite partie de ces fermes. $\mathcal{D}$ est quant à lui presque établi (des modifications sont encore à prévoir dans la liste des mesures à effectuer en ferme). Il résulte du travail du WP2.2, et sera affiné dans les mois à venir en fonction de la faisabilité sur le terrain du protocole de visite regroupant l'ensemble des mesures. Dans la suite de ce manuscrit nous nous proposons de définir $\mathcal{H}$ et $\mathcal{R}$. Le modèle établi par Tsoukiàs (2007) est général, toutes ses composantes ne sont pas forcément utiles dans notre cas. Nous avons choisi de mettre en place un modèle permettant de répondre à un problème d'évaluation dans le certain. Par conséquent, nous n'avons pas eu à représenter les incertitudes. La composante $\mathcal{U}$ n'est donc pas prise en compte dans notre modèle.

4.2 Choix de la famille de critères

Notre premier objectif ici est donc de proposer une famille de critères pouvant servir de base à l'évaluation multicritère du bien-être animal. Ce travail a fait l'objet de l'article *"Definition of criteria for overall assessment of animal welfare"* (Botreau et al. 2007), présenté à la fin de ce chapitre. Étant donné que la notion de critère n'est pas définie ni discutée dans cet article, nous nous permettons de le faire dans les pages suivantes.

La famille de critères doit permettre de décrire les différents aspects du bien-être animal. Elle doit pour cela respecter des propriétés théoriques et pratiques (Bouyssou 1990 ; Roy 1996).

4.2.1 Les propriétés d'une famille de critères

Dans une approche multicritère, le but est d'établir des comparaisons sur la base de l'évaluation d'alternatives (dans notre cas : des fermes) effectuée sur plusieurs critères (ici, permettant de décrire le bien-être des animaux sur la ferme). Le succès de la procédure d'aide à la décision dépend donc largement du choix de ces critères et de leur construction (Bouyssou 1990).

D'après le Petit Larousse, un critère est un *"principe qui permet de distinguer une chose d'une autre, d'émettre un jugement, une estimation"*. C'est avec cette signification qu'il est employé en Recherche Opérationnelle. Plus particulièrement, dans le contexte de l'Aide à la Décision, les critères doivent permettre d'établir des **jugements de préférence** en relation avec la décision (Roy 1985), que ce jugement soit relatif (telle ferme est meilleure ou pire que telle autre sur le critère considéré) ou absolu (sur le critère considéré la ferme est bonne ou mauvaise). Chaque critère permet donc de comparer des alternatives (ici des fermes) sur un "point de vue particulier" (Roy 1985).

Soit X l'ensemble des alternatives (ici, l'ensemble des fermes considérées), le critère (ou fonction critère) $f : X \longrightarrow \mathbb{R}$ permet de comparer deux alternatives x et y sur l'ensemble des conséquences couvertes par le critère f. Le décideur accepte alors de modéliser cette comparaison conformément à :

$$f(x) \geq f(y) \Rightarrow x \succsim_f y \tag{4.1}$$

où $\succsim_f$ représente une relation de préférence sur l'ensemble des conséquences couvertes par le critère f. En d'autres termes : si la valeur de x sur la fonction critère f est supérieure à celle de y, alors, sur ce critère, x est préférée à y.

Dans une approche multicritère, l'analyste (c.-à-d. celui qui construit la procédure

d'évaluation) cherche à construire plusieurs critères se référant chacun à un point de vue spécifique. Ces points de vue représentent les différents axes sur lesquels les décideurs se basent pour justifier, transformer ou discuter leurs préférences (Bouyssou 1990). Les comparaisons qui sont ainsi déduites des différents critères doivent alors être interprétées en termes de préférences partielles, chacune étant réduite aux aspects couverts par le point de vue utilisé pour définir le critère. Ainsi, il est possible de comparer deux alternatives sur l'ensemble des conséquences couverte par un critère, toutes choses étant égales par ailleurs (c.-à-d. sur les autres aspects, non couverts par le critère).

La première étape de la conception d'un modèle d'évaluation consiste donc à définir la famille de critères sur laquelle la construction va reposer. Cette famille doit être conçue de sorte à être intelligible par tous ceux qui auront à la connaître et doit réaliser un consensus aussi vaste que possible (Roy 1985).

Dans le cas où la famille de critères se compose de n critères $(f_1, f_2, \ldots, f_n)$, la ferme x sera donc caractérisée par le vecteur $(f_1(x), f_2(x), \ldots, f_n(x))$. Il est important que ce vecteur, dans son ensemble, représente correctement la ferme x.

Ainsi, une famille de critères doit posséder certaines caractéristiques (Bouyssou 1990 ; Roy 1985) pour pouvoir être appliquée dans des procédures d'aide à la décision.

Comme le signale Roy (1985), le choix de la famille de critères doit obéir à trois exigences logiques permettant de définir le concept de **"famille cohérente de critères"**. Une famille de critères est dite cohérente si elle vérifie les trois principes suivants :

- **exhaustivité** : tous les points de vue importants sont pris en compte. Ceci implique que si $\forall i \ f_i(x) = f_i(y)$, chaque acteur doit être d'accord pour considérer que x et y sont indifférents. En d'autres termes, aucun critère décisif ne doit être oublié.
- **cohésion** : les préférences partielles modélisées par les critères sont cohérentes avec les préférences globales (c.-à-d. intégrant la totalité des conséquences) exprimées sur les alternatives (fermes). Cette condition implique que si la ferme x est considérée meilleure que la ferme y sur l'ensemble des points de vue ($x \succsim y$), le même jugement (x 'meilleure que y') doit être porté sur la ferme z qui est jugée comme étant au moins aussi bonne que x sur chacun des critères ($\forall i \ f_i(z) \geq f_i(x) \Rightarrow z \succsim y$).
- **non-redondance** : la famille de critères ne contient pas de critères non nécessaires (dont la suppression conduirait à une famille de critères satisfaisant encore les deux premières conditions). Ainsi, la famille ne doit pas contenir de critères redondants

ou non pertinents. En cas de critères redondants (c.-à-d. porteurs d'une même information), outre la lourdeur inutile du modèle, le risque serait d'attribuer trop d'importance à un critère qui apparaîtrait sous deux formes similaires dans la famille de critères.

Les exigences de monotonie et de non-redondance impliquent une certaine indépendance des critères. En effet, le résultat de la comparaison de deux fermes x et y, dont les performances ne diffèrent que sur un seul critère, ne doit pas être influencé par les performances (identiques pour x et y) obtenues sur les autres critères de la famille. Roy (1985) parle alors d'**isolabilité** de chaque critère vis-à-vis de l'ensemble des n-1 autres critères de la famille.

Pomerol et Romero (2000, p. 312) soulignent que la principale de ces exigences concerne l'exhaustivité de la famille de critères, la qualité de la décision finale en dépendant. Ils font aussi remarquer que si les conditions de cohérence sont généralement satisfaites avec un décideur rationnel, elle n'est pas pour autant facile à tester. En effet, tester la cohérence de la famille de critères suppose une bonne connaissance des préférences globales du décideur ce qui se révèle souvent difficile à obtenir. En effet, si la procédure d'évaluation est nécessaire c'est souvent parce que le décideur n'arrive pas à exprimer clairement ses préférences globales.

Bouyssou (1990) signale l'existence d'autres propriétés, plus pratiques, que devrait avoir une famille de critères. Cette famille devrait être :

- **opérationnelle** : la famille devrait être considérée par tous les acteurs de la décision comme une base solide et commune pour la poursuite de la procédure d'aide à la décision. Les critères devraient être transparents et faciles à comprendre, en évitant de créer des "boîtes noires" au cours de la procédure d'agrégation de l'information.
- **lisible** : la famille de critères ne devrait être composée que d'un nombre limité de critères. En effet, au cours de la procédure d'agrégation, l'analyste a besoin d'identifier et d'évaluer les correspondances entre les valeurs obtenues sur les différents critères. Afin de traiter cette tâche, il est généralement considéré qu'une douzaine de critères est un maximum. Le fait de n'avoir qu'un nombre réduit de critères à prendre en compte peut aussi se justifier par les limites de la capacité du cerveau humain à traiter simultanément un nombre trop important de données. Comme le soulignent Lacroix et al. (1998), face à trop d'informations les hommes peuvent saturer et ne plus pouvoir intégrer d'informations supplémentaires. Pomerol et Romero (2000, p. 313) réduisent même le nombre souhaitable de critères à sept.

4.2.2 Présentation de notre famille de critères

La littérature propose plusieurs listes de principes devant être respectés pour que le bien-être de l'animal soit assuré. On peut citer par exemple les *cinq libertés* proposées initialement dans le "rapport Brambell" (1965), puis repris par le Farm Animal Welfare Council (1992) :

1. **Absence de faim et de soif**, par accès libre à de l'eau fraîche et à une alimentation équilibrée ;
2. **Absence d'inconfort**, en fournissant un environnement adapté incluant un abri et une zone de repos confortable ;
3. **Absence de douleur**, de blessure et de maladie, par la prévention, un diagnostic précoce et des soins rapides ;
4. **Liberté d'exprimer des comportements normaux**, grâce à un espace suffisant, des équipements adaptés et la présence de congénères ;
5. **Absence de peur et d'anxiété**, en assurant des conditions et un traitement des animaux qui évite la souffrance mentale.

Ces *cinq libertés* proposent donc une liste des besoins physiologiques, comportementaux, sanitaires, environnementaux, et même psychologiques des animaux. Cette liste a été largement utilisée dans la conception de systèmes de certification (ex. le *Freedom Food scheme* mis en place au Royaume-Uni, Main et al. 2001) et dans la législation européenne concernant le bien-être des animaux. Cependant, les *cinq libertés* présentent certaines limites. Plusieurs libertés se recouvrent, par exemple, une zone de couchage inconfortable (Liberté 2) est souvent associée à des blessures (Liberté 3). Les libertés 2 et 5 sont très générales et peuvent couvrir à elles deux la plupart des aspects du bien-être.

D'autres auteurs ont proposé des listes beaucoup plus détaillées des besoins des animaux. Par exemple, Bracke (1999), après avoir fait une revue des différentes listes disponibles dans la littérature, a listé 13 besoins basiques pour assurer le bien-être : ingestion (incluant le besoin de nourriture et d'eau), repos, contact social, besoins liés à la reproduction (incluant l'activité sexuelle, la construction du nid, les soins maternels), déplacement/mouvement, exploration (incluant l'exploration, la nouveauté et la recherche de nourriture), jeu, toilettage, évacuation des déchets (défécation et miction), thermorégulation, respiration, santé (incluant les besoins liés à la maladie et aux blessures), et sécurité (incluant la perception du danger et des conflits sociaux). Ces besoins sont supposés représenter les différents systèmes de contrôle dans lesquels les émotions/sentiments liés au bien-être (tels que la faim, se sentir malade, la douleur, la peur...) jouent un rôle.

Capdeville et Veissier (2001) ont défini une liste similaire de 16 besoins : absence de faim, de soif, de malnutrition, de stress physique, de stress climatique, de maladie, de blessure ; locomotion ; repos ; comportement alimentaire, social, sexuel, maternel ; absence d'événement anxiogène ; possibilité de fuite ; bons contacts avec l'homme.

Nous nous sommes basés sur la littérature existante (notamment les trois articles cités ci-dessus) pour établir la famille de critères d'évaluation du bien-être. Pour ce faire, nous avons suivi quelques principes généraux :

- Les critères doivent pouvoir être applicables à tous les animaux d'élevage (pour des espèces, mais aussi pour des âges et des sexes différents).
- Le bien-être étant directement lié aux états mentaux des animaux (Duncan 2002), son évaluation devrait reposer autant que possible sur des mesures prises directement sur les animaux et groupées en fonction de ce que les animaux perçoivent (cf. la section 2.3 de l'Introduction). Par exemple, une zone de couchage mal adaptée aux animaux peut conduire à la fois à des difficultés comportementales observées chez les animaux (ex. difficultés quand les animaux se couchent ou se lèvent) et à des blessures. Cependant, nous considérons que dans le premier cas les animaux ressentent de l'inconfort alors qu'avec les blessures les animaux ressentent de la douleur. Nous avons donc décidé de considérer ces deux aspects séparément, bien que la cause de la gêne soit la même dans les deux cas. Par ailleurs, les blessures, quelles que soient leurs causes, ont été considérées conjointement puisque dans tous les cas elles sont source de douleur pour les animaux. Néanmoins, nous avons décidé d'isoler la douleur causée par les interventions pratiquées sur les animaux par les éleveurs (ex. castration, écornage...) du fait de leur cause clairement identifiée et du fait que ces interventions ne sont généralement pratiquées qu'une seule fois sur les animaux. Ces pratiques causent donc des blessures ponctuelles entraînant généralement une douleur aiguë chez les animaux, contrairement aux autres blessures qui peuvent être récurrentes et causer une douleur plus modérée.
- Alors que les compensations entre les différents éléments d'un même critère sont autorisées, les compensations entre les critères devraient être limitées. Par exemple, le fait qu'il y ait de bonnes relations homme-animal ne compense pas le manque de contacts sociaux entre les animaux, ce que par exemple Raussi et al. (2003) ont montré chez les veaux[1].
- Les dangers ayant une très faible probabilité de survenir ne sont pas pris en compte. C'est par exemple le cas du feu pouvant ravager une étable ou d'une attaque de prédateur sur un troupeau au pâturage.

[1] Le thème de la compensation a été abordé avec plus de détail dans le Chapitre 3.

TAB. 4.1 – Quelques listes de principes et de besoins qui nous ont servi a définir les sous-critères et critères

Cinq libertés	**Bracke** (1999)	**Capdeville et Veissier** (2001)	***Welfare Quality®*** *Sous-critères*	*Critères*
Absence de faim et de soif	Ingestion (incluant le besoin de nourriture et d'eau)	Absence de faim Absence de malnutrition	*Absence de faim prolongée*	*Alimentation adaptée*
		Absence de soif	*Absence de soif prolongée*	
Absence d'inconfort	Repos	Repos	*Confort autour du repos*	*Logement correct*
	Thermorégulation	Absence de stress climatique	*Confort thermique*	
	Déplacement/mouvement	Locomotion	*Facilité de déplacement*	
Absence de souffrance, de blessure, de maladie	Santé (incluant les besoins liés à la maladie et aux blessures)	Absence de blessure	*Absence de blessures*	*Bonne santé*
	Évacuation des déchets (défécation/miction)	Absence de maladie	*Absence de maladies*	
	Respiration			
		Absence de stress physique	*Absence de douleur causée par les pratiques d'élevage*	
Liberté d'exprimer des comportements normaux	Contact social	Comportement social	*Expression des comportements sociaux*	*Comportement approprié*
	Besoins liés à la reproduction (incluant le sexe, la construction du nid, les besoins maternels)	Comportement sexuel		
		Comportement maternel	*Expression des autres comportements*	
	Exploration (incluant l'exploration, la nouveauté et la recherche de nourriture)	Comportement alimentaire		
	Jeu			
	Toilettage			
Absence de peur et d'anxiété		Bons contacts avec les hommes	*Bonne relation Homme-Animal*	
	Sécurité (incluant la perception du danger et des conflits sociaux)		*Absence de peur (en général)*	
		Absence d'événement anxiogène		
		Possibilité de fuite		

En confrontant les cinq libertés et des listes de besoins présentées dans la littérature, tout en suivant les principes présentés ci-dessus, nous avons pu définir une liste de quatre critères (cf. Tableau 4.1, colonne 5) représentant les grandes dimensions du bien-être animal (valables pour tous les types d'animaux et toutes les périodes de vie) :

- **Alimentation adaptée** où l'apport en nourriture et en eau sera évalué,
- **Logement correct** où les conséquences des conditions de logement sur les animaux seront évaluées,
- **Bonne santé** où l'état sanitaire des animaux ainsi que les sources de douleur pour les animaux seront évalués,
- **Comportement approprié** où l'expression des comportements des animaux, aussi bien entre eux qu'envers l'homme ou leur environnement, sera évaluée.

Ces critères correspondent chacun à une aire fonctionnelle précise (l'alimentation, le logement, la santé et le comportement) et permettent de répondre à quatre grandes questions pouvant constituer un support efficace de communication auprès des utilisateurs potentiels du système d'évaluation :

- *Les animaux sont-ils correctement nourris et l'approvisionnement en eau est-il suffisant ?*
- *Les animaux sont-ils correctement logés ?*
- *Les animaux sont-ils en bonne santé ?*
- *Les animaux expriment-ils des comportements reflétant des états émotionnels optimisés ?*

Ces quatre critères se subdivisent de façon logique en 12 dimensions de préférences sous-jacentes, que nous appelleront des *sous-critères* (cf. Tableau 4.1, colonne 4). Cette étape de construction intermédiaire permet de distinguer des dimensions différentes du bien-être animal, bien que liées à la même aire fonctionnelle. Ainsi, le Critère *"Alimentation adaptée"* se décompose logiquement en deux sous-critères permettant d'évaluer respectivement la faim et la soif des animaux.

Nous nous retrouvons alors avec une liste de quatre critères et de douze sous-critères, chaque critère étant composé de deux à quatre sous-critères (cf. les dernières colonnes du Tableau 4.1).

TAB. 4.2 – Famille de critères utilisée dans Welfare Quality® pour construire une évaluation globale du bien-être.

Critères	Sous-critères		Signification	Spécifications
Alimentation adaptée	SC1	Absence de faim prolongée	Les animaux devraient être nourris de sorte à ce qu'ils ne souffrent pas de faim prolongée	
	SC2	Absence de soif prolongée	L'approvisionnement en eau doit être suffisant, de sorte que les animaux ne souffrent pas de soif prolongée	
Logement correct	SC3	Confort autour du repos	Les animaux ne devraient pas souffrir d'inconfort, notamment au niveau de leur zone de couchage	Évalué à travers le comportement (incluant les mouvements de coucher et de lever) mais pas les blessures (inclues dans le SC5)
	SC4	Confort thermique	Les animaux devraient évoluer dans un environnement thermique adéquat	
	SC5	Facilité de déplacement	Les animaux devraient pouvoir se déplacer librement	En ne prenant pas en compte les problèmes de santé (inclus dans les SC6, 7 et 8), ni les mouvements autour du repos (inclus dans le SC3)
Bonne santé	SC6	Absence de blessures	Les animaux ne devraient pas être blessés physiquement	Excepté celles produites par des maladies ou des interventions volontaires sur les animaux (ex. mutilations)[a]
	SC7	Absence de maladies	Les animaux devraient être exempts de toute maladie	Absence de problèmes cliniques autres que les blessures[b]
	SC8	Absence de douleur causée par les pratiques d'élevage	Les pratiques d'élevage ne devraient pas être cause de douleur pour les animaux	Ex. mutilations et étourdissement/abattage
Comportement approprié[c]	SC9	Expression des comportements sociaux	Les animaux devraient avoir la possibilité d'exprimer des comportements sociaux naturels, n'étant pas source de danger ou cause de blessures	Équilibre entre des aspects négatifs (ex. agressions) et positifs (ex. léchages sociaux)
	SC10	Expression des autres comportements	Les animaux devraient avoir la possibilité d'exprimer d'autres comportements naturels qui leur sont profitables	Équilibre entre des aspects négatifs (ex. stéréotypies) et positifs (ex. exploration)
	SC11	Bonne relation Homme-Animal	De bonnes relations entre l'homme et l'animal participent au bien-être des animaux	
	SC12	Absence de peur (en général)	Les animaux ne devraient pas ressentir d'émotions négatives telles que la peur, la frustration ou l'apathie	Ne tient pas compte de la peur des hommes (inclue dans le SC11)

[a] Pour les porcelets non sevrés la 'mortalité' est prise en compte avec les blessures parce que la mort est généralement due à l'écrasement par la mère.

[b] Ce sous-critère inclut la mortalité des jeunes animaux (à l'exception des porcelets non sevrés), durant le transport, et celle causée par des accidents ou des maladies en ferme. La mortalité liée à la stratégie de renouvellement n'est pas prise en compte.

[c] les 'comportements sociaux' (9) et la 'peur de l'homme' (11) correspondent à deux composantes importantes du bien-être animal (Hemsworth et Coleman 1998 ; Boe et Faerevik 2003). Nous les avons isolés respectivement des 'autres comportements' (10) et de la 'peur en général' (12) afin d'éviter de masquer les effets de ces deux dernières composantes sur le bien-être des animaux.

Avant de pouvoir confirmer que cette liste de quatre critères correspond bien à la *famille de critères* sur laquelle reposera notre modèle d'évaluation, nous devons préalablement vérifier qu'elle possède les propriétés présentées dans la section précédente.

Cette liste semble être *exhaustive* puisque tous les aspects du bien-être animal cités dans la littérature peuvent y trouver leur place (cf. Tableau 4.1 pour une illustration avec les *5 libertés*, les besoins définis par Bracke 1999 et ceux définis par Capdeville et Veissier 2001).

Cette liste semble aussi être *non-redondante*, chaque critère correspondant à un aspect important du bien-être animal et les redondances ayant été évitées grâce à des spécifications claires sur les frontières entre les différents sous-critères (cf. Tableau 4.2, colonnes 2 à 5), ce qui, du fait, permet d'éviter les redondances au niveau des critères.

Alors que deux des trois propriétés théoriques d'une famille de critères sont respectées, nous n'avons cependant pas testé la *cohésion* de cette liste de critères. Nous nous retrouvons en effet dans le cas, décrit par Pomerol et Romero (2000, p. 312), où il ne semble pas possible de tester la cohérence de la famille de critères auprès de l'ensemble des décideurs (incluant à la fois les scientifiques et les utilisateurs potentiels du futur système d'évaluation du bien-être), d'autant plus que personne ne semble pouvoir exprimer clairement ses préférences globales, dans l'absolu, sur un jeu de fermes. Il n'existe en effet pas de "golden standard" permettant de juger des fermes sur le niveau de bien-être qu'elles assurent à leurs animaux.

Chacun des éléments composant cette liste semble être *isolable*, c.-à-d. pouvoir être interprété indépendamment des autres. Il peut cependant exister quelques dépendances fonctionnelles entre certains critères. Par exemple, chez les vaches laitières, la peur de l'homme (intervenant dans l'évaluation du Critère *"Comportement approprié"*) s'évalue généralement par la mesure de la distance de fuite de l'animal face à l'approche d'un homme qu'il ne connaît pas (plus cette distance est grande, plus la vache est considérée comme ayant peur de l'homme), alors que l'évaluation de la santé des animaux inclut l'observation des boiteries. Cependant, Špinka et al. (2005) ont démontré que les vaches souffrant de boiterie ont des distances de fuite plus courtes que celles ne boitant pas. L'évaluation de la relation homme-animal et celle de la santé semblent donc être liées. Dans la mesure du possible, afin de minimiser ce type de dépendances fonctionnelles, il serait souhaitable d'affiner l'évaluation de chaque élément de la liste en jouant sur les

mesures et la manière dont elles sont observées. Dans le cas de l'exemple, il serait alors recommandé d'évaluer les distances de fuite uniquement chez les animaux ne boitant pas, afin d'éviter de sur-évaluer l'absence de peur de l'homme. En appliquant ce type de solutions quand cela se révélera nécessaire, nous pourrons construire des critères isolables.

Avant de pouvoir dire que cette liste de quatre critères est *opérationnelle* (au sens donné par Bouyssou 1990), il est nécessaire de vérifier que l'ensemble des parties prenantes (ex. les utilisateurs potentiels du futur système d'évaluation du bien-être) considèrent cette liste comme une base solide et commune sur laquelle l'évaluation globale du bien-être peut reposer. Les différentes parties prenantes du système d'évaluation sont à la fois les éleveurs, les distributeurs, les consommateurs, mais aussi les associations de protection animale, les législateurs et les scientifiques. Le projet Welfare Quality® a pour atout d'être ouvert aux points de vue de ces différents groupes de personnes, à la fois via le travail réalisé dans le sous-projet 1 (cf. Introduction) et par l'implication de représentants de ces groupes dans le Comité Conseil du projet[2]. Ces deux voies de consultation nous ont permis de tester l'acceptabilité de la liste des quatre critères que nous proposons. En effet, la liste a été présentée aux membres du Comité Conseil, qui l'ont validée. Par ailleurs, les partenaires du Sous-projet 1 ont organisé 55 réunions en petits groupes composés de consommateurs/citoyens à travers sept pays européens (la France, la Hongrie, l'Italie, les Pays-Bas, la Norvège, la Suède et le Royaume-Uni). Cette étude, rapportée par Miele et Evans (2005), a permis de montrer qu'il y avait un appui sociétal fort concernant la liste proposée des critères sur lesquels baser l'évaluation du bien-être. Veissier et Evans (2007) rapportent même le commentaire d'un consommateur italien pour qui les critères sélectionnés représentent ce qui avait été dit tout au long de la réunion, mais d'une manière plus approfondie. Veissier et Evans (2007) soulignent cependant quelques différences entre les attentes des consommateurs/citoyens et le point de vue des scientifiques de Welfare Quality® :

1. L'évaluation du bien-être proposée dans Welfare Quality® se base, dans la mesure du possible, sur des mesures observées directement sur les animaux, afin de mieux appréhender ce qu'ils ressentent. En conséquence, le système d'évaluation ne préjuge pas des atouts ou défauts des différents systèmes d'élevage. Au contraire, les consommateurs sont persuadés que des systèmes extensifs permettent un niveau de bien-être nettement supérieur à celui permis par des systèmes d'élevage intensifs, dits alors "industriels". Les consommateurs mettent clairement en avant le fait

[2]La composition de ce Comité Conseil est détaillée en Annexe B

que les animaux d'élevage doivent évoluer dans un environnement le plus naturel possible.

2. Les consommateurs tendent à avoir une approche holistique du bien-être animal. Ils ne conçoivent alors le bien-être que comme un tout, qui ne peut être divisé en composantes indépendantes, qu'ils considèrent alors comme artificielles.

De manière générale, tout le monde (scientifiques et consommateurs) s'accorde donc à dire que les animaux ne devraient pas souffrir. Au sein du projet, l'évaluation se base donc en priorité sur des mesures faites sur les animaux. Toutefois, des mesures effectuées sur l'environnement seront aussi prises en considération, que ce soit pour évaluer le bien-être ou pour servir de base à un diagnostic en élevage, puis à un conseil auprès de l'éleveur. Ainsi, le système d'élevage (ex. stabulation libre à logette, étable entravée...) n'entrera pas en ligne de compte pour l'évaluation du bien-être, ce qui permettra effectivement de comparer les systèmes entre eux.

Par ailleurs, la vision holistique du bien-être animal n'est pas forcément en contradiction avec la présentation multidimensionnelle qui en est faite au sein du projet. En fait, l'approche holistique a pour conséquence le fait de ne considérer une ferme comme fournissant un bon niveau de bien-être à ses animaux que si elle est bonne sur l'ensemble des quatre critères listés. Ceci nous conforte dans l'idée qu'il est nécessaire de limiter les compensations entre les différentes dimensions du bien-être. Au cours de la construction de l'évaluation, nous devrons donc utiliser une méthode d'agrégation permettant de limiter ces compensations.

Au final, il semble donc approprié de dire que la liste des quatre critères, subdivisés en 12 sous-critères, peut servir de base commune, partagée par l'ensemble des parties prenantes, pour évaluer le bien-être des animaux.

La liste de critères proposée se compose de seulement quatre critères. Elle est donc *lisible* et facilement compréhensible. De plus, elle représente une base de communication simple vers le grand public (et tout spécialement les consommateurs) en présentant les quatre grands points devant être vérifiés en ferme.

La liste des quatre critères semble donc présenter toutes les propriétés, théoriques et pratiques, requises pour une famille de critères. Nous pouvons donc considérer qu'elle constitue notre famille de critère, sur laquelle nous baserons notre modèle d'évaluation du bien-être animal.

4.3 Une construction progressive pour aller des mesures à l'évaluation globale

Nous proposons de suivre une procédure d'évaluation progressive, partant des mesures effectuées sur le terrain à l'évaluation globale en passant par des étapes intermédiaires (les 12 sous-critères et les 4 critères décrits ci-dessus).

L'information qui sera produite aux différents niveaux de la construction restera disponible afin de pouvoir expliquer les résultats obtenus. Ceci permettra de produire un système le plus transparent possible en évitant au maximum les "boîtes noires". De plus, cette information pourra avoir différents destinataires : au niveau des mesures et des sous-critères l'information pourra servir de base à un diagnostic en élevage puis à un conseil apporté à l'éleveur, alors qu'aux niveaux plus synthétiques (critères et évaluation globale) les informations disponibles pourront être intégrées dans un système d'information à destination des distributeurs et du grand public, notamment des consommateurs (cf. Figure 4.1).

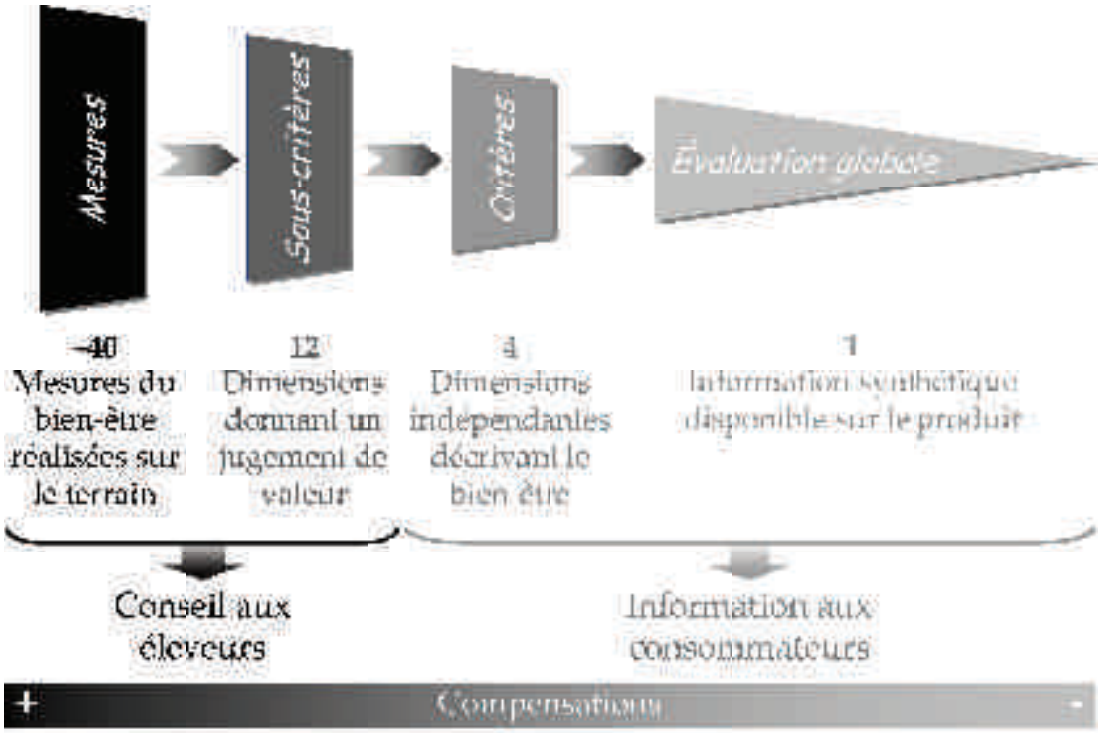


FIG. 4.1 – Système d'agrégation progressive de l'information collectée en ferme

A chaque niveau de l'évaluation, la construction devra tenir compte du nombre d'éléments à considérer, de leur nature, de leur précision, et de leur importance relative.

Par ailleurs, les méthodes de calcul devront permettre d'autoriser ou de limiter les compensations entre les éléments à prendre en compte, et ce en fonction de ce que les experts considéreront comme étant approprié. De manière générale, il semble que plus on se rapproche de l'évaluation globale, plus les compensations doivent être limitées. En effet, au niveau de la construction des sous-critères les différentes mesures à agréger permettent d'évaluer des aspects similaires du bien-être, et de fait il ne semble pas inapproprié d'autoriser les compensations entre les mesures, au moins dans une certaine limite. Au niveau des critères, les sous-critères à agréger correspondent déjà à des aspects différents du bien-être, comme par exemple la faim et la soif, qui correspondent à deux facettes bien distinctes de l'alimentation des animaux. Il semble assez évident qu'on ne peut pas dire que les animaux sont dans des conditions satisfaisantes parce qu'ils n'ont pas faim alors que leur approvisionnement en eau est insuffisant, ce qui pourrait être le cas si on autorisait pleinement les compensations. Ceci est d'autant plus vrai au niveau de l'agrégation des critères pour former l'évaluation globale au niveau de la ferme. Par définition, les quatre critères correspondent aux quatre grandes dimensions du bien-être, indépendantes les unes des autres. A ce niveau, dire que le bon état sanitaire des animaux compense totalement le fait qu'ils souffrent de restrictions comportementales (telles que l'isolement) semble difficile à justifier.

Le système proposé utilisera donc des méthodes de calcul et de synthèse de l'information qui seront différentes d'un niveau à l'autre :

- au niveau des **sous-critères** : utilisation de méthodes de calculs appropriées (sommes pondérées, comparaison à des normes, définition d'arbre de décision...), choisies en fonction des mesures qui sont à prendre en compte, autorisant généralement les compensations entre les mesures, tout en les limitant pour une même mesure entre les individus ;
- au niveau des **critères** : utilisation d'une méthode de calcul permettant de tenir compte du fait que certains sous-critères peuvent être plus importants que d'autres, ainsi que du fait que les compensations entre les sous-critères doivent être plus limitées ;
- au niveau de l'**évaluation globale** : utilisation d'une méthode de comparaison à des profils pré-établis délimitant plusieurs catégories de bien-être, afin de limiter fortement les compensations entre les critères tout en fournissant une information adaptée aux différentes utilisations potentielles du système d'évaluation globale du bien-être.

Toutefois, la question du niveau de compensation à autoriser ou non devra être tranchée à partir des avis d'experts. Pour ce faire, nous n'avons pas directement posé la question des compensations aux experts. Nous leur avons soumis des jeux de données virtuelles (c.-à-d.composés de fermes virtuelles caractérisées par les valeurs des mesures, leurs scores au niveau des sous-critères ou des critères). Nous leur avons demandé de porter un jugement sur chacune des fermes composant ces jeux de données. C'est à partir de leurs réponses que nous avons déduit le niveau de compensation qu'ils souhaitent autoriser. Les experts consultés sont différents en fonction du niveau de l'agrégation. Ainsi, au niveau des sous-critères nous avons consulté des chercheurs en sciences animales impliqués dans le choix et le développement des mesures sur lesquelles reposent l'évaluation du bien-être. Au niveau de la construction des critères nous avons consulté à la fois des chercheurs en sciences animales et en sciences sociales, ces derniers pouvant alors restituer l'avis des consommateurs, distributeurs et éleveurs questionnés au sein du Sous-projet 1 (cf. Annexe A). Enfin, pour l'agrégation des critères en une évaluation globale, nous avons consulté des scientifiques mais aussi des représentants d'utilisateurs potentiels du système d'évaluation, membres du Comité Conseil du projet (cf. Annexe B).

Ainsi nous avons pu impliquer, directement ou indirectement, des utilisateurs potentiels du système d'évaluation, en plus de chercheurs en sciences sociales et animales. Leur implication au cours de la construction permettra probablement de faciliter l'acceptabilité du système qui sera produit au sein du projet Welfare Quality® et ainsi de maximiser ses chances d'être mis en application dans le futur au niveau européen.

© 2007 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK

Animal Welfare 2007, 16: 225-228
ISSN 0962-7286

Definition of criteria for overall assessment of animal welfare

R Botreau*[‡], I Veissier[†], A Butterworth[§], MBM Bracke[#] and LJ Keeling[¶]

[†] INRA, UR1213 Herbivores, Site de Theix, Saint-Genès-Champanelle F-63122, France
[‡] Institut de l'Elevage, BP18, Castanet Tolosan F-31321, France
[§] University of Bristol Clinical Veterinary Science, Langford, BS40 5DU, UK
[#] Animal Sciences Group, Wageningen University and Research Centre, PO Box 65, Lelystad NL-8200 AB, The Netherlands
[¶] Department of Animal Environment and Health, University of Agricultural Sciences, PO Box 7038, Uppsala SE-750 07, Sweden
* Contact for correspondence and requests for reprints: rbotrea@clermont.inra.fr

Abstract

Welfare is multidimensional, comprising good health, comfort, expression of behaviour, and so on. Its overall assessment therefore requires a multicriteria evaluation. The set of criteria shall be exhaustive (no missing item), minimal (only necessary items), agreed by stakeholders, and legible (a limited number of criteria). Furthermore, the interpretation from one criterion shall not depend on that from another. We propose a set of 12 subcriteria grouped into four criteria: feeding, housing, health and optimised emotional states. This work will assist in developing measures to be used on-farm to form a European standard for overall assessment of animal welfare.

Keywords: *animal welfare, integration, multicriteria evaluation, overall assessment*

Introduction

Consumers are concerned about the living conditions of farm animals. As a consequence of this, numerous trade groups (producers, distributors, retailers, chain restaurants) have created certification schemes which include animal welfare (eg Integrale Keten Beheersing in the Netherlands, Swedish Broiler Control, Filières Qualité Carrefour in France, McDonald's Europe, RSPCA Freedom Food in the UK). These initiatives aim at allowing consumers to choose 'high welfare standard' products, but the assessment of welfare varies from one scheme to another. There is no European standard to assess animal welfare while bringing the information to consumers. The EU project Welfare Quality® seeks to fill this gap for cattle, pigs and poultry, by proposing an overall assessment system which is scientifically valid and widely accepted by stakeholders (Blokhuis *et al* 2003).

Welfare is multidimensional (Mason & Mendl 1993; Fraser 1995) and so its assessment corresponds to a multicriteria evaluation problem. The first step towards solving this problem is to define a set of criteria relevant for animal welfare, while following theoretical and practical requirements for a multicriteria evaluation (Bouyssou 1990; Roy 1996). Our objective here is to propose such a set of criteria suitable for an overall evaluation of animal welfare.

Properties of a set of criteria

The set of criteria, which makes an overall assessment possible, should fulfil the following requirements (Bouyssou 1990):

• It must be *exhaustive*, ie containing every important viewpoint.

• It must be *minimal*, ie containing only necessary criteria (banning redundant or irrelevant criteria).

• Criteria must be *independent* of each other. The interpretation from one criterion shall not depend on that from another criterion. Moreover, to avoid double counting there should be, as far as possible, no functional links between criteria.

• The set of criteria should be *agreed* by all stakeholders and considered as a sound basis for operating a practical assessment. The criteria and their application should be transparent and easy to understand, avoiding 'black boxes' in the aggregation procedure.

• To be '*legible*' the set of criteria should be composed of a limited number of criteria. In fact, to implement an aggregation procedure, it is necessary to show the values obtained for the different criteria correspond to one another. To handle this task, it is generally considered that twelve criteria is a maximum.

Table 1 Set of criteria and subcriteria used in WelfareQuality® to develop an overall welfare assessment.

Criteria	Subcriteria	Specifications
Good feeding	1. Absence of prolonged hunger.	
	2. Absence of prolonged thirst.	
Good housing	3. Comfort around resting.	Assessed through behaviour (including rising up and lying down movements) but not injuries (included in 5).
	4. Thermal comfort.	
	5. Ease of movement.	Not considering health problems (included in 6, 7, 8) and movements around resting (included in 3).
Good health	6. Absence of injuries.	Except those produced by a disease or voluntary interventions (eg mutilations)[1].
	7. Absence of disease.	Absence of clinical problems other than injuries[2].
	8. Absence of pain induced by management procedures.	Eg mutilations and stunning.
Appropriate behaviour[3]	9. Expression of social behaviours.	Balance between negative (eg aggression) and positive (eg social licking) aspects.
	10. Expression of other behaviours.	Balance between negative (eg stereotypies) and positive (eg exploration) aspects.
	11. Good human-animal relationship.	No fear of humans.
	12. Absence of general fear.	Except fear of humans.

[1] For suckling piglets 'mortality' is considered with injuries because death is mostly caused by crushing by the sow.
[2] This includes mortality for young animals (except suckling piglets) and during transport. Mortality at other times is not considered because it largely depends on management and culling strategies.
[3] 'Social behaviours' (9) and 'fear of humans' (11) are very important components of farm animals welfare (Hemsworth & Coleman 1998; Boe & Faerevik 2003). They have been isolated respectively from 'other behaviours' (10) and 'general fear' (12), to avoid masking the effects of these latter elements.

Proposal for a list of welfare items fulfilling these properties

The literature offers several lists of principles that need to be fulfilled to achieve animal welfare, eg the *five freedoms*: 1) freedom from hunger and thirst, 2) freedom from discomfort, 3) freedom from pain, injury and disease, 4) freedom to express normal behaviour, and 5) freedom from fear and distress (Farm Animal Welfare Council 1992). These *five freedoms* have been widely used in certification schemes (eg Freedom Food scheme; Main *et al* 2001) and in EU regulations. However, they present some limitations. Several freedoms overlap, for example, an uncomfortable resting place (Freedom 2) is often associated with injuries (Freedom 3). Freedoms 2 and 5 are very general and can cover most of welfare aspects. Other authors have provided more detailed lists of animal needs (reviewed by Bracke *et al* 1999) while others propose lists of measures (eg Winckler *et al* 2003). Taking the existing literature into consideration, we applied the following principles to choose a set of welfare items:

• Welfare items shall be applicable to any farm animal species.

• Since welfare relates to mental states (Duncan 2002), the assessment should be based on measures taken on animals and grouped according to what the animal perceives. For instance, poor resting areas may lead to behavioural difficulties (eg in lying down) and to injuries, but we considered that the former results in discomfort and the latter in pain; hence they were considered separately. By contrast, injuries, whatever their cause, were considered together because they are all likely to result in pain. We nevertheless kept pain induced by management procedures (castration, slaughter, etc) separate because of the clearly identified cause.

• Compensation within a given item may be allowed while less compensation is allowed between items (eg good human-animal relationships cannot fully compensate for lack of social contacts [Raussi *et al* 2003]).

• Hazards with a low probability (like fire or attacks by predators) will not be included.

We thus defined a list of 12 items (Table 1, column 2) which fulfils most of the requirements for a set of criteria.

This list appears to be *exhaustive* since the aspects of animal welfare quoted in the literature (eg Bracke *et al* 1999; Capdeville & Veissier 2001) can find their place.

This list seems *minimal* as each item has been shown to be important for welfare and redundancies between items have been avoided thanks to clear specifications on the limits of each item (Table 1, column 3).

Each item appears to be capable of *independent interpretation*. Nevertheless, some functional dependencies between

© 2007 Universities Federation for Animal Welfare

items were noticed. For example, if fear of humans is measured by flight distance, health and observed fear are dependent because lame animals flee less than healthy ones (Špinka *et al* 2005). Refining the assessment of each item (eg by assessing fear of humans only in non-lame animals) will help minimise such dependencies.

We consulted the advisory committee of WelfareQuality ® composed of representatives of consumers, distributors, producers, animal advocates and policy makers. They were in *agreement* on this list. In addition, discussions were organised in 49 focus groups of consumers (in seven European countries) and showed that consumers consider our list as a relevant basis to assess welfare.

As presented previously, twelve items is at the upper limit agreed for *legibility*. Even if all criteria are assessed on a two-level scale,

$$2^2 \times \sum_{\alpha=1}^{10} (12 - \alpha) = 264$$

possible combinations should be explored to perform the final aggregation into the overall assessment.

Reorganisation of the 12 welfare items into four criteria

To be fully supported by users, the criteria should be easy to understand. To ease communication with stakeholders, particularly consumers, and because it would make the set more legible (as discussed in the previous section), it was decided to use a smaller number of items.

We grouped the 12 items into criteria defined so that compensation was minimal between criteria (eg good health is unlikely to compensate for the lack of appropriate behaviour) while keeping the 12 previous items as subcriteria (Table 1, column 1). The resulting four criteria correspond to the functional areas: feeding, housing, health, behaviour, and to the general questions:

- Are the animals properly fed and supplied with water?
- Are the animals properly housed?
- Are the animals healthy?
- Does the behaviour of the animals reflect optimised emotional states?

Conclusion and animal welfare implications

Reliable science-based assessment of animal welfare requires carefully constructed assessment tools. In this paper, we propose an exhaustive and minimal set of 12 subcriteria, independent in their interpretation, regrouped into four criteria to ease communication, especially with consumers (Table 1). This leads to a hierarchical structure where the information at each level will remain available, particularly at the subcriterion level. This makes the whole process transparent, so helping farmers identify those improvements on which they should focus their efforts.

This set of criteria will be used within WelfareQuality® to develop welfare assessment systems. Subsequent steps include the choice of appropriate measures to evaluate each subcriterion, deciding on a mathematical expression for each subcriterion, synthesis of the subcriteria into the criteria with clear formal aggregation (avoiding black boxes), and aggregation of the criteria into the overall assessment.

Such an assessment scheme to properly capture the welfare state of the animals will be of great potential benefit and should help lead to improvement in the animals' quality of life.

Acknowledgements

The present study is part of the WelfareQuality® research project which has been co-financed by the European Commission, within the 6th Framework Programme, contract no. FOOD-CT-2004-506508. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

References

Blokhuis HJ, Jones RB, Geers R, Miele M and Veissier I 2003 Measuring and Monitoring Animal Welfare: Transparency in the Food Product Quality Chain. *Animal Welfare 12:* 445-455

Boe KE and Faerevik G 2003 Grouping and Social Preferences in Calves, Heifers and Cows. *Applied Animal Behaviour Science 80:* 175-190

Bouyssou D 1990 Building criteria: a prerequisite for MCDA. In: Bana e Costa CA (ed) *Readings in Multiple Criteria Decision-Aid.* Springer Verlag: Heidelberg, Germany

Bracke MBM, Spruijt BM and Metz JHM 1999 Overall animal welfare assessment reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science 47:* 307-322

Capdeville J and Veissier I 2001 A Method of Assessing Welfare in Loose Housed Dairy Cows at Farm Level, Focusing on Animal Observations. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplement 30:* 62-68

Duncan IJH 2002 Poultry welfare: Science or subjectivity? *British Poultry Science 43:* 643-652

Farm Animal Welfare Council 1992 FAWC updates the five freedoms. *Veterinary Record 17:* 357

Fraser D 1995 Science, values and animal welfare: exploring the 'inextricable connection'. *Animal Welfare 4:* 103-117

Hemsworth PH and Coleman GJ 1998 *Human-Livestock Interactions: The Stockperson and the Productivity and Welfare of Intensively Farmed Animals*. CAB International: Oxon/New York

Main DCJ, Webster F and Green LE 2001 Animal Welfare Assessment in Farm Assurance Schemes. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplement 30:* 108-113

Mason G and Mendl M 1993 Why is there no simple way of measuring animal welfare? *Animal Welfare 2:* 301-319

Raussi S, Lensink BJ, Boissy A, Pyykkonen M and Veissier I 2003 The Effect of Contact With Conspecifics and Humans on Calves' Behaviour and Stress Responses. *Animal Welfare 12:* 191-203

Roy B 1996 *Multicriteria Methodology for Decision Aiding*. Kluwer Academic: Dordrecht, The Netherlands

Špinka M, Dembele I, Panamá J and Stìhulová I 2005 Lame dairy cows have shorter avoidance distances. In: *Proceedings of the 39th International Congress of the International Society for Applied Ethology*. 20-24 August 2005. Sagamihara, Japan

Winckler C, Capdeville J, Gebresenbet G, Horning B, Roiha U, Tosi M and Waiblinger S 2003 Selection of Parameters for On-Farm Welfare-Assessment Protocols in Cattle and Buffalo. *Animal Welfare 12:* 619-624

© *2007 Universities Federation for Animal Welfare*

Chapitre 5

La construction des sous-critères

Résumé de la construction des sous-critères :

Tous les sous-critères sont exprimés sur une échelle d'utilité commune $[0, 100]$ où 0 et 100 représentent respectivement la pire et la meilleure situation envisageable sur une ferme.

Pour construire les sous-critères nous nous servons de deux méthodes principales :
- la définition d'un arbre de décision permettant de déterminer un nombre fini de situations possibles en ferme ;
- le calcul d'un index issu d'une combinaison linéaire des mesures recodées au niveau du troupeau sur une échelle ordinale à trois ou quatre niveaux. Une fonction d'utilité est ensuite définie à l'aide de fonctions splines pour recoder l'index sur l'échelle d'utilité.

Dans tous les cas, le paramétrage de la construction repose sur la réaction d'experts sur des jeux de données virtuelles. À ce stade de la construction, les experts consultés sont des chercheurs en sciences animales ayant participé, au sein du projet Welfare Quality®, au choix et au développement des mesures utilisées pour évaluer les sous-critères.

Chapitre 5

La construction des sous-critères

5.1 Quelles sont les mesures servant à évaluer le bien-être dans Welfare Quality® ?

L'évaluation du bien-être repose sur de nombreuses mesures effectuées sur le terrain. Au sein du projet Welfare Quality®, la tâche consistant à établir la liste des mesures à prendre en compte revient au WP2.2 (cf. Figure 1.9 du chapitre introductif), et ce pour l'ensemble des espèces et des périodes (élevage ou transport/abattage) considérées dans le projet.

5.1.1 Identification des mesures servant à évaluer chacun des 12 sous-critères

Le WP2.2 avait pour rôle de proposer une liste de mesures valides, répétables et faisables en ferme pour évaluer le bien-être des animaux. Pour l'évaluation du bien-être des vaches laitières en ferme, nos partenaires du WP2.2 sont partis d'une première liste de 50 mesures à observer directement sur les animaux. Ils ont ensuite testé ces mesures pour les trois caractéristiques souhaitées : validité, répétabilité et faisabilité. À l'issue du WP2.2, seules 26 mesures ont été retenues.

Pour construire le modèle d'évaluation du bien-être des vaches laitières en ferme, nous nous sommes basés sur 25 de ces mesures. En effet, deux des mesures retenues par le WP2.2 consistent à évaluer la distance de fuite des vaches face à l'approche de la personne réalisant les observations, l'une avec les vaches libres dans l'étable et l'autre effectuée sur les vaches bloquées au cornadis[1]. Ces deux mesures permettent

[1]Barrière permettant le passage d'une tête d'animal, utilisée pour distribuer efficacement l'alimentation ou immobiliser les animaux (notamment nécessaire pour vacciner ou inséminer les vaches)

d'évaluer la peur des vaches face à un homme inconnu (c.-à-d. autre que l'éleveur). Il existe cependant une différence importante au niveau du temps nécessaire pour réaliser ces deux mesures (pour une trentaine de vaches, environ une heure pour la distance mesurée dans l'étable contre seulement une vingtaine de minutes pour celle mesurée au cornadis). De plus, seule la mesure effectuée sur des animaux attachés est réalisable pour des vaches en étable entravée[2]. Ayant à l'esprit que les utilisateurs du système d'évaluation risquent de demander de réduire la durée de la visite en ferme et que le système devrait être utilisable quel que soit le type de logement, nous avons décidé de n'utiliser dans le modèle d'évaluation qu'une seule de ces deux mesures : celle effectuée sur des animaux bloqués.

En parallèle des mesures observables sur les animaux, de nombreuses mesures sur les ressources et la conduite du troupeau par l'éleveur ont aussi été définies par le WP2.2. Parmi cet ensemble de mesures, nous en avons utilisé 19 pour compléter le modèle d'évaluation du bien-être des vaches laitières en ferme lorsque certains sous-critères ne pouvaient pas être évalués par des mesures sur animaux. Ainsi, pour l'absence de soif, un état de déshydratation visible sur les animaux n'aurait permis de distinguer que des cas extrêmes. Nous avons alors préféré baser l'évaluation du Sous-critère 2 sur la qualité de l'approvisionnement en eau des animaux.

Au total, **44 mesures** (25 effectuées sur les animaux et 19 sur leur environnement) sont prises en compte pour évaluer le bien-être des vaches laitières en ferme. Ces 45 mesures servent à construire neuf des douze sous-critères définis dans le Chapitre 4. Chacun de ces neuf sous-critères se compose de 1 à 9 mesures (cf. Tableau 5.1). Il existe donc trois sous-critères pour lesquels aucune mesure n'est pour l'instant disponible. Il s'agit des sous-critères 4 (*Confort thermique*), 10 (*Expression des autres comportements*) et 12 (*Absence de peur*). Nous avons cependant décidé de conserver l'ensemble des 12 sous-critères dans la construction de notre modèle puisque certaines mesures sont encore en développement dans Welfare Quality®. Ces trois sous-critères couvrent des aspects jugés importants pour les vaches laitières, et nous espérons donc que des mesures appropriées seront disponibles d'ici la fin du projet. Par exemple, il semble possible de pouvoir à terme inclure la possibilité pour l'ensemble des vaches (qu'elles soient en stabulation libre ou en étable entravée) d'aller à la pâture dans le Sous-critère 10 correspondant à l'expression de comportements normaux autres que sociaux. La liste

[2]Contrairement à des étables où les animaux sont libres de leurs mouvements (étable avec logettes ou aire paillée), les vaches en étable entravée sont attachées. Ce système de logement est encore assez répandu en France (environ 20% des vaches françaises sont logées dans des étables entravées), notamment dans les zones de montagne où ce système présente l'avantage de nécessiter moins de surface au sol

définitive des mesures recommandées pour évaluer le bien-être des vaches laitières ne sera disponible qu'en juillet 2008.

Le choix des mesures à utiliser pour l'évaluation de chaque sous-critère est le résultat de discussions au sein du WP2.3 (composé de chercheurs en sciences animales). Malgré les spécifications permettant de clarifier les frontières entre les sous-critères (cf. Tableau 4.2 du Chapitre 4), l'affectation de certaines mesures s'est révélée problématique. En effet, certaines mesures peuvent entrer dans la définition de plusieurs sous-critères. Il convient alors de définir à quel sous-critère on décide de rattacher la mesure en question ou de préciser la façon de l'interpréter en fonction du sous-critère, et ce afin d'éviter de prendre en compte plusieurs fois la même information dans l'évaluation.

C'est par exemple le cas des boiteries, qui peuvent être soit dues à une blessure (par exemple liée à une tarsite[3] causée par le frottement répété du tarse de la vache sur l'arête bétonnée de sa logette) soit liées à une maladie (il existe par exemple un lien de cause à effet entre acidose[4] et fourbure[5], cause de boiterie). Pour les vaches laitières, nous avons décidé de tenir compte des boiteries uniquement dans l'évaluation du Sous-critère 6 *"Absence de blessures"* considérant que leurs boiteries sont plus souvent dues à des blessures qu'à des maladies.

L'état d'engraissement pourrait lui aussi entrer dans la construction de deux sous-critères : l'*absence de faim prolongée* (SC1) et l'*absence de maladies* (SC7). En effet, une vache peut être maigre car elle n'a pas accès à suffisamment de nourriture ou parce qu'elle est atteinte d'une maladie chronique. Toutefois, au niveau du troupeau, un pourcentage élevé d'animaux maigres semble plus susceptible d'être lié à un problème d'alimentation qu'à une pathologie. Nous avons donc décidé de ne tenir compte de cette mesure que dans la construction du SC1.

5.1.2 Caractéristiques des mesures choisies pour évaluer le bien-être animal dans Welfare Quality®

Les 44 mesures recueillies en ferme pour évaluer le bien-être des vaches laitières sont réalisées à des échelles différentes. Alors que certaines fournissent directement une information au niveau du troupeau (ex. le nombre d'abreuvoirs mis à la disposition des animaux), d'autres sont observées au niveau individuel sur un nombre de vaches

[3]Inflammation des tissus mous périarticulaires entourant les petites articulations tarsiennes

[4]Le pH ruminal d'une vache se situe normalement près de la neutralité à environ 6,5 mais peut varier selon la composition de la ration ingérée et la gestion de l'alimentation. Des rations riches en concentré peuvent significativement réduire le pH du rumen. Certaines bactéries présentes dans le rumen vont alors rapidement dégrader l'amidon de la ration en acide lactique ce qui a pour effet d'acidifier le rumen.

[5]Congestion et inflammation du pied de l'animal

TAB. 5.1 – Liste des mesures utilisées pour construire les sous-critères

Critères	Sous-critères		Mesures
Alimentation adaptée	SC1	Absence de faim prolongée	État d'engraissement (pourcentage d'animaux maigres)
	SC2	Absence de soif prolongée	Approvisionnement en eau des animaux : - nombre d'auges - longueur totale des auges - nombre de bols (avec ou sans réservoir d'eau) - débit de l'eau - propreté et nombre d'animaux
Logement correct	SC3	Confort autour du repos	Temps mis par les vaches pour se coucher Pourcentage d'animaux entrant en collision avec une partie du matériel de logement (ex. barres délimitant les logettes) au moment où ils se couchent Pourcentage de vaches couchées partiellement en dehors de la zone de couchage Propreté : - de la mamelle - du flanc et de la cuisse - des pattes arrières
	SC4	Confort thermique	-
	SC5	Facilité de déplacement	Les vaches sont-elles attachées ? Ont-elles accès à une zone d'exercice extérieure ? (nombre de jours par an et nombre d'heures par jour) Vont-elles à la pâture ? (nombre de jours par an et nombre d'heures par jour)
Bonne santé	SC6	Absence de blessures	Boiterie Altérations du tégument : - Nombre de blessures légères (pertes de poils) par vache - Nombre de blessures sérieuses (plaies ou gonflements) par vache
	SC7	Absence de maladies	Problèmes respiratoires (toux, éternuements, écoulements nasal et oculaire, fréquence respiratoire élevée) Problèmes digestifs (diarrhée) Problèmes reproductifs (mammite et perte vaginale) Mortalité
	SC8	Absence de douleur causée par les pratiques d'élevage	Mutilations réalisées en élevage : - écornage (% de vaches écornées, méthode utilisée, utilisation de produits anesthésiques et analgésiques) - queue coupée (% de vaches ayant la queue coupée, nombre de vertébres restantes, utilisation de produits anesthésiques et analgésiques)
Comportement approprié	SC9	Expression des comportements sociaux	Fréquences des comportements agonistiques (coup de tête entraînant ou non le déplacement de la vache cible, poursuite après un coup, coup entraînant le lever d'un animal couché, combat)
	SC10	Expression des autres comportements	-
	SC11	Bonne relation Homme-Animal	Distance de fuite au cornadis
	SC12	Absence de peur (en général)	-

représentatif de l'ensemble du troupeau (ex. l'état d'engraissement).

Les 44 mesures recueillies en ferme pour évaluer le bien-être des vaches laitières produisent des données brutes exprimées sur des échelles très différentes (cf. colonne 2 du Tableau 5.2) :

- Certaines mesures sont exprimées sur des échelles *cardinales*. C'est par exemple le cas de la fréquence des comportements sociaux agonistiques, correspondant au nombre moyen de comportements observés par vache et par heure. La distance de fuite de la vache face à un homme qui s'approche d'elle (il s'agit ici de la distance entre le mufle de la vache et la main de la personne qui s'approche de l'animal au moment où l'animal commence à reculer).
- D'autres mesures sont exprimées sur des échelles *ordinales*. C'est par exemple le cas de l'évaluation des boiteries chez les vaches. Trois catégories d'animaux ont été définies : les vaches ne boitant pas (démarche équilibrée et régulière avec répartition homogène du poids de l'animal sur ses quatre pattes), les vaches boitant légèrement (démarche avec des foulées irrégulières due à une répartition non homogène du poids de l'animal sur ses quatre pattes) et les vaches boitant sérieusement (démarche très irrégulière causée par une forte réticence de la vache à s'appuyer sur au moins l'une de ses pattes).
- Enfin, certaines mesures sont *nominales*. C'est entre autres le cas de la méthode d'écornage des animaux. Trois méthodes différentes sont possibles pour écorner des vaches : 1. retrait des bourgeons de cornes chez le veau par la chaleur, 2. retrait des bourgeons par l'utilisation d'une solution basique (soude), et 3. écornage mécanique chez la vache adulte.

Pour pouvoir être utilisées pour évaluer les différents sous-critères, ces données brutes ont besoin d'être interprétées en termes de bien-être animal. Il est important de rappeler que la relation existant entre la mesure et son interprétation en termes de bien-être n'est pas forcément linéaire, et ce, même quand la donnée brute est cardinale (cf. Chapitre 3 avec l'exemple de la distance de fuite[6]). Ainsi, si les données brutes doivent être converties sur une échelle de valeur allant de "niveau médiocre de bien-être" à "excellent niveau de bien-être", il est peu vraisemblable que cette conversion suive une fonction linéaire.

Par ailleurs, les mesures composant un même sous-critères peuvent avoir des contributions plus ou moins importantes dans la construction du sous-critère concerné. C'est

[6]Par exemple, des vaches ayant des distances de fuite de 6 et 8 m seront toutes deux considérées comme ayant très peur de l'homme, alors qu'une vache qui fuit l'homme à 2 m sera considérée comme étant nettement plus peureuse qu'une vache qui accepte de se faire toucher (distance de fuite de 0 m).

par exemple le cas du SC3 *"Confort autour du repos"* composé de six mesures se répartissant en deux groupes : d'une part les mesures comportementales (temps mis par les vaches pour se coucher, % de vaches entrant en collision lorsqu'elles se couchent et % de vaches couchées partiellement en dehors de la zone de couchage) et d'autre part les mesures de propreté (propreté de la mamelle, propreté du flanc et de la cuisse et propreté des pattes arrières). Or, la contribution des mesures comportementales à l'évaluation du SC3 est jugée par les chercheurs en sciences animales beaucoup plus importante que celle des mesures de propreté. En effet, seules des vaches très sales auront un état de bien-être diminué. Il sera donc aussi nécessaire de tenir compte dans la construction des sous-critères de l'importance relative des différentes mesures qui les composent.

Enfin, certaines mesures offrent une meilleure précision que d'autres. Par exemple, l'évaluation des blessures (SC6) se fait par l'intermédiaire de deux catégories de mesures, toutes deux effectuées au niveau individuel : l'évaluation des boiteries et celle des altérations du tégument (avec différenciation des blessures légères et sérieuses). Alors que la mesure des boiteries ne fournit qu'une évaluation en trois classes (ne boite pas / boite légèrement / boite sévèrement), et donc peu précise, l'évaluation des altérations du tégument fournit au contraire des informations très détaillées nous donnant pour chaque vache le nombre de blessures légères et le nombre de blessures sérieuses.

Existence de relations fonctionnelles entre certaines mesures

Il existe des relations entre certaines mesures servant à évaluer le bien-être animal. En effet, nous savons que chez les vaches laitières des animaux qui boitent présentent des distances de fuite face à l'homme plus réduites (Špinka et al. 2005). Pour pouvoir éviter ce lien fonctionnel entre ces deux mesures, il serait possible de ne considérer que les distances de fuite de vaches ne boitant pas, ou tout au moins ne présentant pas de boiterie sévère. Il suffirait pour cela que l'identifiant de chaque vache soit précisé en regard de son évaluation des boiteries et de la mesure de sa distance de fuite. Il subsisterait cependant un problème, lié au nombre d'animaux devant être observés. En effet, cela aurait pour conséquence de réduire le nombre d'animaux observés pour évaluer la qualité de la relation homme-animal. Ce nombre doit en effet rester suffisamment grand pour que l'échantillon soit représentatif de l'ensemble du troupeau. Sur une ferme ne présentant que peu de boiteries, ce ne serait pas perturbant, mais il existe des fermes où un très fort pourcentage de vaches boitent. C'est par exemple le cas d'une des fermes visitées dans le cadre du WP2.4 sur laquelle 73% des vaches boitaient sérieusement, pour un total de plus de 86% de vaches présentant une boiterie. Sur un tel élevage il serait alors impossible d'évaluer la distance de fuite. Il serait donc en fait peut-être

préférable d'accepter de sous-estimer légèrement les distances de fuite.

Il faut cependant souligner que l'étude menée par Špinka et al. (2005) traitait de la distance de fuite d'animaux libres de leurs mouvements, or la mesure que nous utilisons est évaluée chez des animaux attachés ou bloqués au cornadis. Il est alors possible que la relation entre boiterie et distance de fuite soit beaucoup plus réduite que dans l'étude de Špinka et al. (2005). Si tel s'avère être le cas, en prenant en compte toutes les vaches observées (boiteuses ou non) nous ne commettrions qu'une très légère sous-estimation de la peur de l'homme. Il serait aussi possible d'introduire un facteur de correction établi en fonction de résultats expérimentaux. Dans l'état actuel des connaissances sur la relation de cause à effet entre les boiteries et les distances de fuite au cornadis, nous proposons de ne pas tenir compte de l'existence de cette relation et acceptons donc de légèrement sous-évaluer la peur de l'homme.

Toutefois, il existe pour d'autres types d'animaux des relations entre mesures différentes. Il existe par exemple chez les veaux une relation de cause à effet entre parasitisme digestif et gain de poids. En effet, Bangoura et Daugschies (2007) ont clairement montré que des veaux infectés par des spores de *Eimeria zuernii*, et donc atteints de coccidiose, présentent un gain de poids nettement inférieur à celui de veaux sains. Dans leur étude, tous les veaux infectés ont présenté des diarrhées, contrairement aux animaux sains (non infectés). Le risque serait alors de pénaliser doublement l'éleveur, une fois du fait des diarrhées (dans le Sous-critère 7) et une fois du fait de la maigreur des animaux (dans le Sous-critère 1). Là encore, il serait plus sage de ne tenir compte du poids des veaux que sur des veaux n'étant pas atteints de diarrhées (facilement observable lors d'une visite en ferme), mais nous risquerions alors de nous retrouver face au problème de la taille de l'échantillon considéré. Dans le cadre d'une évaluation annuelle des élevages, une solution pourrait alors être de reprendre l'évaluation du poids des animaux observée l'année précédente ($n-1$) sur le même élevage, en partant du principe que si l'éleveur n'a pas changé de pratique d'alimentation entre les deux années, il n'y a pas de raison pour que les faibles poids d'animaux observés l'année n s'expliquent par l'alimentation, mais plutôt par le problème sanitaire.

5.2 Principes généraux de la construction

Comme nous venons de le voir, le nombre de mesures à agréger varie d'un sous-critère à l'autre (de 1 à 9 mesures). De plus, les mesures servant à construire un sous-critère peuvent être observées au niveau individuel ou directement au niveau du troupeau, peuvent s'exprimer sur des échelles différentes et n'ont pas forcément la même importance. Enfin, les informations fournies par les mesures peuvent être plus ou moins précises.

Par conséquent, différentes méthodes seront utilisées pour synthétiser l'information au niveau des différents sous-critères de sorte à adapter chaque fois la méthode au nombre de mesures à agréger, à leur nature, à leur précision et à leur importance relative. Ainsi, dans certains cas, nous avons décidé de définir des arbres de décision, pour d'autres sous-critères nous avons eu recours à des sommes pondérées ou encore à la définition de seuils délimitant des catégories (cf. Tableau 5.2). Cependant, dans tous les cas, nous avons utilisé une échelle d'évaluation commune à l'ensemble des sous-critères, allant de 0 à 100, avec les conventions suivantes :

- **0** correspond à la pire situation envisageable sur une ferme (c.-à-d. la situation en-deçà de laquelle aucune détérioration supplémentaire du niveau de bien-être n'est possible),
- **20** correspond à une situation mauvaise en-deçà de laquelle le niveau de bien-être est jugé inacceptable,
- **50** correspond à une situation neutre, le niveau de bien-être n'est ni bon ni mauvais,
- **100** correspond à la meilleure situation envisageable sur une ferme (c.-à-d. la situation au-delà de laquelle aucune amélioration supplémentaire du niveau de bien-être n'est possible).

Ainsi, les 12 sous-critères listés dans le Tableau 5.1 seront formellement définis comme des fonctions $g_j : X \longrightarrow [0, 100],\ j = 1, \ldots, 12$ où X représente l'ensemble des fermes concernées par l'évaluation du bien-être de leurs animaux.

Afin d'interpréter correctement les données brutes en termes de bien-être, de définir les seuils de référence et l'importance relative des différentes mesures (quand cela était nécessaire), et finalement de construire les fonctions d'utilité appropriées, nous avons consulté des chercheurs en sciences animales impliqués, au sein de Welfare Quality®, dans le développement et le choix des mesures servant à évaluer le bien-être (WP2.2). En effet, à cette étape de la construction nous devons faire appel à des personnes ayant

une très bonne connaissance des mesures, de la façon dont elles sont recueillies sur le terrain et de la manière de les interpréter en termes de bien-être pour les animaux.

Au cours de cette première étape de calibrage du modèle, nous n'avons pu consulter que cinq ou six chercheurs (différents en fonction du sous-critère) impliqués dans le WP2.2. La liste des mesures n'étant toujours pas définitive, il nous est impossible de demander à plus de personnes de donner leur expertise sur le sujet. En effet, nous demandons aux experts consultés de nous consacrer du temps. Du fait, de l'évolution constante de l'ensemble des mesures à prendre en compte, nous avons déjà dû recontacter plusieurs fois les experts. Dans ces conditions, il semble nécessaire d'attendre la liste définitive des mesures (été 2008) pour pouvoir contacter plus de personnes et ainsi vérifier et, au besoin, ajuster le paramétrage de la construction.

TAB. 5.2 – Types de construction des sous-critères évaluant le bien-être des vaches laitières en ferme

SC	*Nb mes.*	*Type de données*	Construction
g_1	*1*	*Mesure binaire au niveau individuel (très maigre vs normal)*	% de vaches très maigres transformé en score en utilisant une interpolation par des splines minimisant le carré des écarts
g_2	*6*	*Mesures cardinales ou ordinales au niveau de la ferme*	Méthode lexicographique définissant 4 situations possibles, scores directement attribués à chacune des 4 situations
g_3	*6*	*Mesures cardinales ou ordinales, au niveau individuel ou au niveau de la ferme*	Chaque mesure est transformée sur une échelle ordinale au niveau de la ferme : pas de problème / problème léger / problème sérieux. Calcul d'un index basé sur une combinaison linéaire des proportions de problèmes légers et sérieux, en considérant que les comportements autour du repos sont 3 fois plus importants que les mesures de propreté. Index transformé en utilisant une interpolation par des splines minimisant le carré des écarts
g_5	*5*	*Mesures cardinales ou ordinales au niveau de la ferme*	Construction d'un arbre de décision définissant 5 situations possibles, scores directement attribués à chacune des 5 situations
g_6	*3*	*Mesures cardinales ou ordinales au niveau individuel*	Calcul de 2 index (1 pour les altérations du tégument et 1 pour les boiteries), chacun se basant sur une combinaison linéaire des % de vaches présentant un problème léger *vs* un problème sérieux, puis transformé en score en utilisant une interpolation par des splines minimisant le carré des écarts. Les 2 scores partiels sont synthétisés en calculant une intégrale de Choquet
g_7	*9*	*Mesures cardinales ou ordinales, au niveau individuel ou au niveau de la ferme, à des échelles de temps différentes (visite vs année)*	Chaque mesure est transformée sur une échelle ordinale au niveau de la ferme : pas de problème / problème léger / problème sérieux. Calcul d'un index basé sur une combinaison linéaire des proportions de problèmes légers et sérieux. Index transformé en utilisant une interpolation par des splines minimisant le carré des écarts
g_8	*8*	*Mesures binaires ou nominales au niveau de la ferme*	Pour chacune des deux mutilations (écornage et queue coupée), construction d'un arbre de décision définissant respectivement 13 et 9 situations possibles. Scores partiels directement attribués à chacune des situations. Le score final correspond au plus mauvais des deux scores partiels obtenu sur au moins 10% des animaux
g_9	*5*	*Mesures cardinales au niveau de la ferme*	Calcul d'un index basé sur une combinaison linéaire du nombre de coups de tête et du nombre d'autres interactions agonistiques, transformé en utilisant une interpolation par des splines minimisant le carré des écarts
g_{11}	*1*	*Mesure cardinale au niveau individuel*	Pour chaque animal observé, sa distance de fuite est transformée sur une échelle ordinale : pas de problème / problème léger / problème modéré / problème sérieux. Calcul d'un index basé sur une combinaison linéaire des % d'animaux observés dans chacun des 4 niveaux de l'échelle. Index transformé en utilisant une interpolation par des splines minimisant le carré des écart

NB : Les sous-critères 4, 10 et 12 ne sont pas présentés dans ce tableau puisqu'il n'y a pas pour le moment de mesures pour les évaluer

5.3 Exemples de construction

Nous n'allons pas détailler dans ce chapitre les constructions de l'ensemble des sous-critères. Nous allons juste présenter quelques exemples illustrant les différentes techniques utilisées pour construire les sous-critères. La construction de l'ensemble des sous-critères est détaillée dans le livrable D2.8b, présenté en Annexe C du manuscrit.

5.3.1 Exemple 1 : Sous-critère g_2 *"Absence de soif prolongée"*

Le premier exemple concerne le Sous-critère g_2 *"Absence de soif prolongée"*. Son évaluation repose sur six mesures effectuées sur la qualité de l'approvisionnement en eau des animaux telles que le nombre de points d'eau, leur propreté ou le débit d'eau disponible aux points d'eau (cf. Tableau 5.1). Ces six mesures effectuées sur la ferme permettent de répondre à trois questions décrivant les différents aspects de l'approvisionnement en eau des animaux :

1. Le nombre d'abreuvoirs est-il suffisant ?
2. Les abreuvoirs sont-t-ils propres ?
3. Y-a-t-il au moins deux abreuvoirs accessibles par chaque animal ?

Ces trois aspects de l'approvisionnement en eau peuvent être considérés hiérarchiquement du fait de leurs importances respectives : par exemple, si le nombre d'abreuvoirs n'est pas suffisant, cela ne sert à rien de regarder si ils sont propres puisque de toute façon l'approvisionnement en eau sera jugé comme étant insuffisant. Ceci nous a conduit à définir l'arbre de décision présenté sur la Figure 5.1. La définition d'un arbre de décision pour définir l'échelle de notation sur laquelle évaluer une dimension (ici le Sous-critère 2) a aussi été utilisée dans d'autres applications, comme par exemple dans le cas de la machine à trier les paquets à la Poste (Renard 1986).

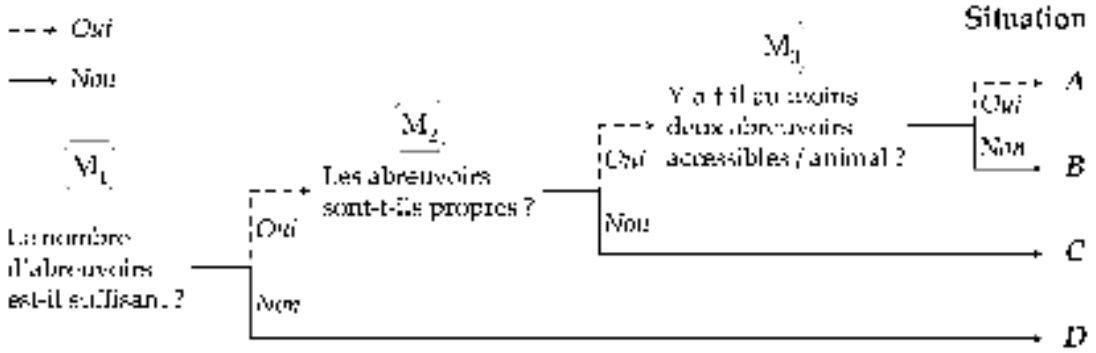


FIG. 5.1 – Arbre de décision utilisé pour le Sous-critère *"Absence de soif prolongée"*

Le premier point à vérifier concerne le nombre d'abreuvoirs en état de fonctionnement (c.-à-d. ayant un débit d'eau suffisant) mis à disposition des animaux. Ces abreuvoirs peuvent être de deux types différents : les *bacs d'abreuvement* (notés par la suite *auges*), dans lesquels plusieurs vaches peuvent boire en même temps, et les *bols*, dans lesquels une seule vache peut boire. Le nombre d'abreuvoirs en état de fonctionnement est comparé à une recommandation établie en fonction du nombre de vaches dans le troupeau : il doit y avoir au moins 6 cm d'auge par vache et/ou au moins un bol pour 10 vaches. Le second point concerne la propreté des abreuvoirs. Enfin, si et seulement si les deux conditions précédentes sont vérifiées, il est préférable que les animaux aient accès à au moins deux abreuvoirs : au cas où un abreuvoir tombe en panne, les vaches peuvent toujours avoir accès au second le temps que l'éleveur détecte le problème et le répare.

L'arbre de décision ainsi établi définit quatre situations possibles pour une ferme, de A (la meilleure) à D (la moins bonne) (cf. Figure 5.1).

Pour répondre aux trois questions sur lesquelles repose l'arbre de décision, nous nous sommes basés sur les 6 mesures relatives à l'approvisionnement, auxquelles se rajoute l'information concernant le type de logement (stabulation libre *vs* étable entravée). Ces mesures vont donc renseigner sept variables différentes. À ces variables issues des mesures, se rajoute une autre variable calculée : le nombre maximum de vaches pouvant s'abreuver correctement en fonction des abreuvoirs disponibles et du débit d'eau. Ces huit variables sont succinctement présentées dans le Tableau 5.3.

TAB. 5.3 – Liste des variables, issues des mesures ou calculées, utilisées pour construire g_2

Variable	Domaine	Signification
n_a	$\mathbb{R}$	nombre d'auges
l_a	$\mathbb{R}$	longueur totale des auges
n_b	$\mathbb{R}$	nombre de bols (avec ou sans réservoir d'eau)
d	$\{0, 1, 2\}$	débit de l'eau *(0- insuffisant, 1- suffisant, 2- partiellement suffisant)*
p	$\{0, 1\}$	propreté des abreuvoirs *(0- sale, 1- propre)*
T	$\{L, E\}$	type d'étable : stabulation libre *(L) vs* étable entravée *(E)*
n_v	$\mathbb{R}$	nombre de vaches
n_{aut}	$\mathbb{R}$	nombre maximum de vaches recommandé en fonction du nombre d'abreuvoirs

Chaque ferme observée est alors affectée à l'une des quatre situations possibles grâce à une fonction $h : X \longrightarrow \{A, B, C, D\}$; où X correspond à l'ensemble des fermes devant être évaluées. La fonction h est définie via l'Algorithme 1. Cet algorithme permet de

dire à quelle situation (A, B, C ou D) correspond une ferme donnée, en fonction des valeurs obtenues par cette ferme sur les variables définies dans le Tableau 5.3.

Algorithme 1 : Procédure d'affectation d'une ferme aux quatre situations possibles, de A à D

```
h(x) ← D
n_aut(x) ← 0
if T(x) = L then
    if d(x) = 0 then h(x) ← D
    else if d(x) = 1 then
        n_aut(x) ← (l_a(x)/6) + (n_b(x) × 10)
        if n_aut(x) < n_v(x) then h(x) ← D; le nombre d'abreuvoirs est insuffisant
        else
            if p(x) = 0 then h(x) ← C; les abreuvoirs sont sales
            else
                if (n_a(x) + n_b(x)) < 2 then h(x) ← B; il n'y a qu'un seul abreuvoir
                else h(x) ← A
    else if d(x) = 2 then
        n_aut(x) ← n_aut(x)/2
        if n_aut(x) < n_v(x) then h(x) ← D; le nombre d'abreuvoirs est insuffisant
        else
            if p(x) = 0 then h(x) ← C; les abreuvoirs sont sales
            else
                if (n_a(x) + n_b(x)) < 2 then h(x) ← B; il n'y a qu'un seul abreuvoir
                else h(x) ← A
else
    if n_b(x)/n_v(x) < 0,5 then h(x) ← D; le nombre d'abreuvoirs est insuffisant
    else
        if p(x) = 0 then h(x) ← C; les abreuvoirs sont sales
        else
            if n_b(x) < 1 then h(x) ← B; il n'y a qu'un abreuvoir pour 2 vaches
            else h(x) ← A
end
```

Une fonction d'utilité u est alors utilisée pour transformer la valeur $h(x)$ obtenue par la ferme x en score de sous-critère $g_2(x)$, défini comme suit :

$$g_2(x) = u(h(x)) \tag{5.1}$$

Cette fonction d'utilité u est définie à dire d'experts. Nous avons demandé à cinq experts d'attribuer un score à chacune des quatre situations possibles définies par l'arbre

de décision. Pour chaque situation z, $u(z)$ correspond à la moyenne des évaluations données par les experts à la situation z (cf. Tableau 5.4).

TAB. 5.4 – Évaluations données par les experts pour le Sous-critère 2

Situation			**Score**			**Moyenne**
z	E1	E2	E3	E4	E5	$u(z)$
A	100	100	100	100	100	**100**
B	70	50	60	70	50	**60**
C	30	40	50	20	20	**32**
D	0	0	15	0	0	**3**

5.3.2 Exemple 2 : Sous-critère g_6 *"Absence de blessures"*

Ce second exemple concerne le Sous-critère 6 *"Absence de blessures"*. Ce sous-critère est évalué sur deux types de blessures différents : d'une part les *altérations du tégument* et d'autre part les *boiteries*. Dans les deux cas, les évaluations sont réalisées sur un échantillon d'animaux représentatif de l'ensemble du troupeau. Cependant, les animaux observés pour les boiteries ne sont pas forcément les mêmes que ceux dont les altérations du tégument ont été évaluées. Il n'est dont pas possible de considérer en même temps ces deux types de blessures. En effet, cela reviendrait à ne prendre en compte que les animaux évalués à la fois pour les boiteries et les altérations du tégument, et donc à ne plus avoir un échantillon d'animaux représentatif du troupeau. Nous avons donc dû considérer séparément les deux types de blessures, et ainsi attribuer à une ferme x donnée deux notes distinctes, une pour les altérations du tégument ($g_6^1(x)$) et une pour les boiteries ($g_6^2(x)$). Ces deux notes seront ensuite combinées afin d'obtenir le score de sous-critère $g_6(x)$.

Les constructions suivies pour les deux notes $g_6^1(x)$ et $g_6^2(x)$ sont très similaires. Nous allons ici plus détailler la construction pour les *"altérations du tégument"*.

5.3.2.1 Évaluation des altérations du tégument : note $g_6^1(x)$

L'évaluation des altérations du tégument repose sur deux mesures cardinales observées sur les mêmes vaches :

- nombre de blessures légères (abrasions, pertes de poils) ;
- nombre de blessures sérieuses (plaies ouvertes, gonflements).

Au niveau individuel, ces deux mesures sont combinées en une seule évaluation exprimée sur une échelle ordinale de trois niveaux. Ces trois catégories, présentées dans le

Tableau 5.5, ont été définies en collaboration avec le chercheur responsable de l'évaluation des altérations du tégument au sein du WP2.2. Elles correspondent à trois niveaux de problème différents vis-à-vis du bien-être de la vache : absence de problème, problème modéré, problème sérieux.

Cette simplification de l'information recueillie en ferme (on part de deux mesures cardinales pour arriver à une seule mesure ordinale) repose sur l'hypothèse selon laquelle l'animal perçoit essentiellement la douleur causée par sa blessure la plus grave. Si cette simplification de l'information se révèle être insuffisamment discriminante (après des tests sur des jeux de données réelles), on pourra alors imaginer la définition de plus de catégories afin de distinguer par exemple les vaches ne présentant qu'une seule blessure sérieuse de celles qui en présentent plusieurs.

TAB. 5.5 – Évaluation des altérations du tégument au niveau individuel

3 catégories sont définies au niveau individuel :		Problème vis-à-vis du bien-être :
0	La vache ne présente aucune blessure	Pas de problème
1	La vache présente au moins une blessure légère, mais pas de blessure sérieuse	Modéré
2	La vache présente au moins une blessure sérieuse	Sérieux

Au niveau de la ferme, nous prenons alors en compte les pourcentages d'animaux dans chacune des catégories 0, 1 et 2, notés respectivement $p_0(x)$, $p_1(x)$ et $p_2(x)$ pour la ferme x.

Il nous a alors fallu établir la procédure à suivre pour agréger ces trois pourcentages, et ainsi construire correctement la note g_6^1. Pour ce faire, nous avons décidé de consulter cinq experts en les faisant réagir sur un jeu de données virtuelles. Ce jeu de données était composé de 11 fermes virtuelles caractérisées par leurs pourcentages de vaches dans chacune des trois catégories (cf. Tableau 5.6). Ces 11 fermes couvrent l'étendue des cas possibles, du pire (*Ferme 3*, toutes les vaches observées présentent au moins une blessure sérieuse) au meilleur (*Ferme 1*, toutes les vaches observées ne présentent aucune blessure).

Nous avons demandé aux experts, dans un premier temps, de classer ces 11 fermes de la pire à la meilleure, puis, dans une deuxième temps, d'attribuer un score (sur l'échelle de 0 à 100 définie précédemment) à chacune d'elle.

TAB. 5.6 – Jeu de données virtuelles permettant d'établir la construction

Ferme	$p_0(x)$	$p_1(x)$	$p_2(x)$	**Rang**	**Score**
Ferme 1	100	0	0		
Ferme 2	0	100	0		
Ferme 3	0	0	100		
Ferme 4	50	50	0		
Ferme 5	0	50	50		
Ferme 6	50	0	50		
Ferme 7	50	25	25		
Ferme 8	20	50	30		
Ferme 9	0	25	75		
Ferme 10	75	25	0		
Ferme 11	75	0	25		

Les experts ont généralement classé de la même façon les 11 fermes, seules quelques inversions mineures (moins de 10 points de différence au niveau des scores) se retrouvent chez les Experts 2 et 5 (cf. les dix premières colonnes du Tableau 5.7 présentant les réponses des experts). Nous avons alors recherché si une combinaison linéaire des trois pourcentages p_0, p_1 et p_2 pouvait refléter l'ordre de préférence des fermes, ordre de préférence recueillant la quasi-unanimité des experts. Pour ce faire, nous avons décidé de calculer un index $h_1(x)$ pour toute ferme $x \in X$, défini comme suit :

$$h_1(x) = 100 - \frac{\sum_{k=0}^{2} w_k \, p_k(x)}{w_2} \tag{5.2}$$

où w_k correspond au poids attribué à la catégorie k. Cette définition formelle repose sur le fait que $p_0(x) + p_1(x) + p_2(x) = 1$ pour tout $x \in X$. De plus, cet index est défini de sorte à fournir un résultat $\in [0, 100]$, où 0 correspond à la pire situation envisageable et 100 à la meilleure (cf. la section 4.2). Ainsi, en notant x_* la moins bonne ferme possible (caractérisée par $p_2(x_*) = 100$) et x^* la meilleure ferme possible (caractérisée par $p_0(x^*) = 100$), l'index $h_1(x)$ est construit de telle sorte que $h_1(x^*) = 100$ et $h_1(x_*) = 0$ (ce qui explique pourquoi nous divisons par w_2 la somme pondérée des $p_k(x)$).

Pour définir les poids $w_k,\ k = 0, \ldots, 2$, nous nous sommes basés sur les réponses des experts sur le jeu de données virtuelles. Les poids ont été obtenus par régression ordinale en utilisant un solveur (avec pour contrainte de trouver une combinaison de poids entiers et dont la somme est minimale produisant un index respectant le classement des fermes). Nous avons ainsi obtenu les poids suivants : $w_0 = 0$, $w_1 = 1$ et $w_2 = 5$.

En plus de classer les fermes, nous avons demandé aux experts de donner leur propre évaluation de $g_6^1(x)$ pour chaque ferme de référence x, en utilisant l'échelle standardisée $[0, 100]$ utilisée pour exprimer les performances sur l'ensemble des sous-critères. Les résultats obtenus sont présentés dans le Tableau 5.7.

TAB. 5.7 – Évaluations données par les experts aux *altérations du tégument*

	Situation			**Scores donnés par les experts Ei**						**Index**	**Score**
	$p_0(x)$	$p_1(x)$	$p_2(x)$	E1	E2	E3	E4	E5	*Moyenne*	$h_1(x)$	$g_6^1(x)$
weights	*0*	*1*	*5*								
Ferme 3	**0**	**0**	**100**	0	0	0	0	0	*0*	**0**	**0**
Ferme 9	**0**	**25**	**75**	3	1	5	5	5	*4*	**20**	**4**
Ferme 5	**0**	**50**	**50**	7	2	10	10	10	*8*	**40**	**8**
Ferme 6	**50**	**0**	**50**	15	3	13	15	15	*12*	**50**	**11**
Ferme 8	**20**	**50**	**30**	20	10	18	20	15	*17*	**60**	**16**
Ferme 7	**50**	**25**	**25**	25	4	22	25	25	*20*	**70**	**22**
Ferme 11	**75**	**0**	**25**	30	5	25	30	25	*23*	**75**	**26**
Ferme 2	**0**	**100**	**0**	35	15	60	50	20	*36*	**80**	**31**
Ferme 4	**50**	**50**	**0**	50	20	70	70	30	*48*	**90**	**51**
Ferme 10	**75**	**25**	**0**	75	75	85	80	50	*73*	**95**	**71**
Ferme 1	**100**	**0**	**0**	100	100	100	100	100	*100*	**100**	**100**

Bien que l'index $h_1(x)$, défini comme une agrégation linéaire des pourcentages p_0, p_1, p_2 explique parfaitement les rangs affectés par les experts aux fermes, la comparaison des colonnes "Moyenne" (moyenne des scores donnés par les experts à chacune des fermes de référence) et "Index" du Tableau 5.7 montre clairement que l'index h_1 n'est pas suffisant pour produire une valeur de performance absolue correspondant aux valeurs prescrites par les experts sur les fermes de référence. Il est donc nécessaire de recourir à une fonction d'utilité non-linéaire u_1 afin de transformer l'index $h_1(x)$ en score de sous-critère partiel $g_6^1(x)$, en utilisant l'équation suivante :

$$g_6^1(x) = u_1\left(h_1(x)\right) \tag{5.3}$$

La fonction $u_1(x)$ doit satisfaire des contraintes au niveau des bornes : $u_1(100) = 100$ et $u_1(0) = 0$. De plus, le score ainsi calculé doit être en accord avec le classement induit par $h_1(x)$ sur les fermes de référence. Enfin, u_1 devrait permettre d'approximer assez correctement les évaluations prescrites par les experts (cf. Tableau 5.7). Du fait de la non-linéarité de la préférence observée pour le critère g_6^1 par rapport à h_1, u_1 a été définie comme une fonction spline, ce qui permet une grande souplesse dans l'approximation du jugement observé. La construction de u_1 se base sur l'ajustement aux opinions des

experts par des fonctions B-splines, de sorte à minimiser le carré des écarts entre les valeurs fournies par les experts et les scores calculés (Bartels et al. 1987 ; Plass et Stone 1983).

Une fonction spline cubique se compose de plusieurs morceaux de courbes, chaque morceau correspondant à un polynôme de degré 3. Afin de conserver un bon contrôle de la forme de la courbe sans pour autant multiplier artificiellement le nombre de morceaux, nous avons décidé de définir dans un premier temps une fonction spline composée de cinq morceaux, c'est-à-dire avec quatre noeuds intérieurs que nous avons choisi équidistants. Le fait de choisir d'utiliser quatre noeuds équidistants (aux valeurs 20, 40, 60 et 80 de l'index) permet de limiter des distorsions dans la tension des portions de courbe (entre deux noeuds très éloignés la courbe sera beaucoup plus "tendue" qu'entre deux noeuds proches) et permet aussi de s'adapter à toutes les situations. En effet, avec quatre noeuds intérieurs peu importe que la courbe souhaitée présente un ou plusieurs points d'inflexion, et peu importe la position de ces points d'inflexion puisque qu'il y aura de toute façon un noeud proche du (ou des) point(s) d'inflexion.

Les fonctions B-splines cubiques permettent d'obtenir une fonction de continuité C^2 (cf. Annexe D). Avec quatre noeuds intérieurs, nous allons obtenir une courbe composée de cinq morceaux de courbes polynomiales qui ont la particularité de se rejoindre au niveau des noeuds. Nous obtenons ainsi une seule courbe continue[7]. De plus, au niveau de chaque noeud, les deux fonctions qui s'y rejoignent ont des dérivées d'ordre 1 égales[8] (à leur point de rencontre les fonctions ont la même tangente), ce qui évite d'avoir des ruptures de pente au niveau de chaque noeud dans la courbe finale. Enfin, au niveau de chaque noeud les dérivées d'ordre 2 des fonctions qui s'y rejoignent sont elles aussi égales[9] (à leur point de rencontre les fonctions ont la même courbure), ce qui permet d'avoir au final une courbe encore plus lisse.

La définition de la fonction spline revient donc à chercher un polygone de contrôle composé de points équidistants, dont le lissage génère une fonction spline qui minimise δ, la somme des carrés des écarts entre le modèle et les scores prescrits par les experts, qui s'écrit formellement :

$$\delta = \sum_{p=1}^{n_e} w_p \left(\sum_{k=1}^{n_f} (u_1(h_1(x_k)) - y_p(x_k))^2 \right) \tag{5.4}$$

[7]ce qui correspond à la définition d'une continuité d'ordre 0 C^0
[8]ce qui correspond à la définition d'une continuité d'ordre 1 C^1
[9]ce qui correspond à la définition d'une continuité d'ordre 2 C^2

où n_e correspond au nombre d'experts (pour le moment $n_e = 5$) et n_f au nombre de fermes (ici $n_f = 11$). Les fermes de référence sont notées $x_1, \ldots, x_{11}$ et $y_p(x_k)$ correspond à la valeur attribuée par l'expert p à la ferme x_k pour les "altérations du tégument". La quantité $u_1(h_1(x_k))$ correspond au score calculé pour la ferme x_k par le modèle et w_p correspond au poids attribué à l'expert p (pour le moment $w_p = 1$ pour tout p). En attribuant des poids aux experts (w_p à l'expert p), il est possible d'attacher plus ou moins d'importance à l'opinion des différents experts en fonction du degré de confiance qu'ils s'attribuent et/ou du niveau d'expertise qu'ils ont dans le cadre de la procédure d'évaluation. Pour le moment, nous avons décidé d'attribuer le même poids à tous les experts.

Afin de lisser le polygone de contrôle nous avons utilisé une spline cubique (cf. équations (5.5) et (5.6)) générée par des combinaisons de quatre fonctions B-splines représentée sur la Figure 5.2 et définies comme suit :

$$\begin{aligned} b_{-1}(\lambda) &= \frac{1}{6}(1 - 3\lambda + 3\lambda^2 - \lambda^3) \\ b_0(\lambda) &= \frac{1}{6}(4 - 6\lambda^2 + 3\lambda^3) \\ b_1(\lambda) &= \frac{1}{6}(1 + 3\lambda + 3\lambda^2 - 3\lambda^3) \\ b_2(\lambda) &= \frac{1}{6}(\lambda^3) \end{aligned} \tag{5.5}$$

Ces fonctions sont construites de sorte que :

$$\forall \lambda \in [0, 1], \quad \sum_{i=-1}^{2} b_i(\lambda) = 1 \tag{5.6}$$

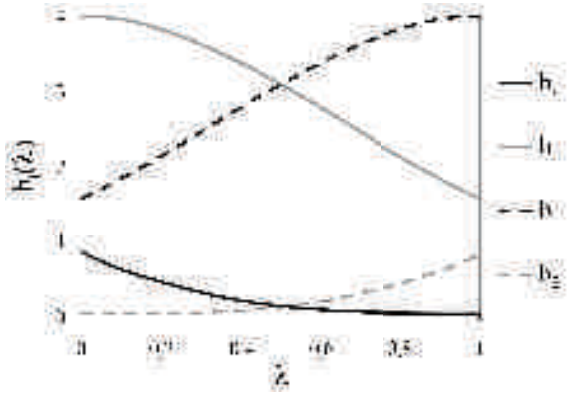


FIG. 5.2 – Tracé des quatre fonctions B-splines basiques

À partir des réponses fournies par les experts, nous avons appris huit points de contrôle (cf. Table 5.8) en utilisant un solveur de sorte à minimiser l'erreur globale δ tout en obligeant la courbe à passer par les points (0,0) et (100,100). Nous avons ainsi utilisé les points de contrôle suivants :

TAB. 5.8 – Points du polygone de contrôle représentés graphiquement sur la Figure 5.3

x (index)	-20	0	20	40	60	80	100	120
y (utilité)	0.5	-1.6	6.0	2.3	24.3	11.7	103.0	176.3

Parmi ces huit points de contrôles on retrouve les 4 noeuds intérieurs (à 20, 40, 60 et 80), les limites inférieure (0) et supérieure (100) de l'index, ainsi qu'un point en-deçà de 0 (-20) et un au-delà de 100 (120). Ces deux derniers points interviennent dans le calcul de l'utilité pour des index appartenant respectivement à $[0, 20]$ ou à $[80, 100]$.

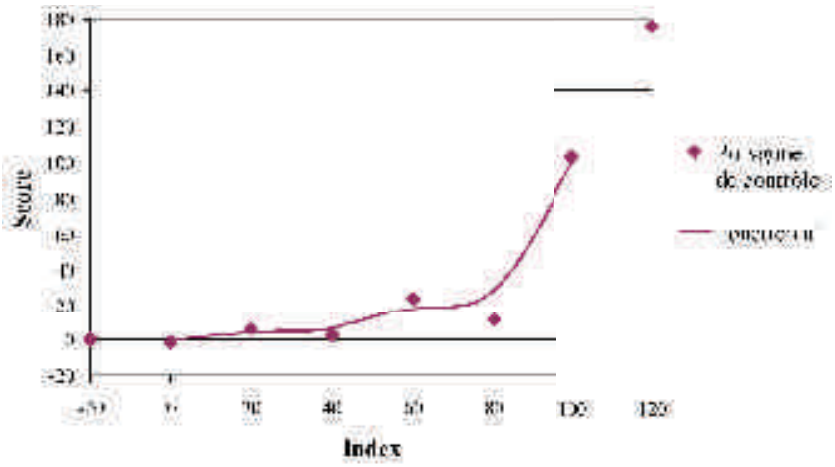


FIG. 5.3 – Polygone de contrôle et fonction spline u_1 en résultant

Nous obtenons ainsi une courbe composée de cinq portions de courbes (polynômes de degré 3), chaque portion de courbe étant définie par une combinaison convexe de quatre points de contrôle successifs, formant ainsi une enveloppe convexe, en utilisant pour le calcul des B-splines les coefficients suivants :

$$u_1(z) = \alpha_z\, b_{-1}(\bar{z}) + \beta_z\, b_0(\bar{z}) + \gamma_z\, b_1(\bar{z}) + \delta_z\, b_2(\bar{z}) \tag{5.7}$$

où $\bar{z} = \frac{z-a}{b-a}$ pour tout $z \in [a, b]$. Les coefficients α_z, β_z, γ_z et δ_z forment l'enveloppe de contrôle considérée en fonction de la valeur de z (cf. exemple pour $z \in [40, 60]$ sur la Figure 5.4) ; les valeurs utilisées pour générer la courbe u_1 sont présentées dans le Tableau 5.9.

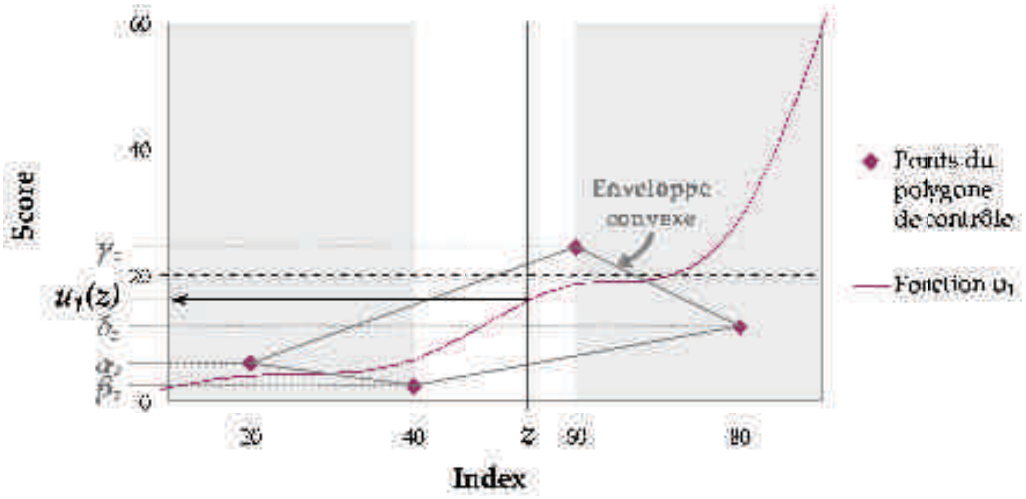


FIG. 5.4 – Illustration des points de contrôle intervenant dans le calcul de l'utilité pour un index $\in [40, 60]$

TAB. 5.9 – Polygone de contrôle pour la fonction spline u_1

z	α_z	β_z	γ_z	δ_z
[0, 20]	0.5	-1.6	6.0	2.3
[20, 40]	-1.6	6.0	2.3	24.3
[40, 60]	6.0	2.3	24.3	11.7
[60, 80]	2.3	24.3	11.7	103.0
[80,100]	24.3	11.7	103.0	176.3

Nous obtenons ainsi la fonction non-linéaire u_1 (cf. Figure 5.5) passant obligatoirement par les points (0,0) et (100,100), tout en minimisant le carré des écarts entre les scores donnés par les experts et le score g_6^1 calculé.

Dans le cas des altérations du tégument, l'utilisation de fonctions B-splines, comme décrite ci-dessus, fournit un résultat tout à fait correct. Cependant nous pouvons faire deux remarques liées à notre construction, la première portant sur le choix d'utiliser des B-splines (il existe bien d'autre types de fonctions splines, cf. Bartels et al. 1987) et la seconde sur le choix du nombre et de la position des noeuds intérieurs.

1. La fonction d'utilité obtenue peut ne pas être monotone

La fonction d'utilité obtenue est une fonction polynomiale par morceaux avec des contraintes de continuité, mais en utilisant des B-splines nous ne pouvons imposer aucune contrainte de monotonie à la fonction. Or, avec une construction reposant d'abord sur une fonction linéaire afin de calculer un index restituant l'ordre de préférence des experts sur les fermes, la fonction d'utilité définie dans un second temps doit absolu-

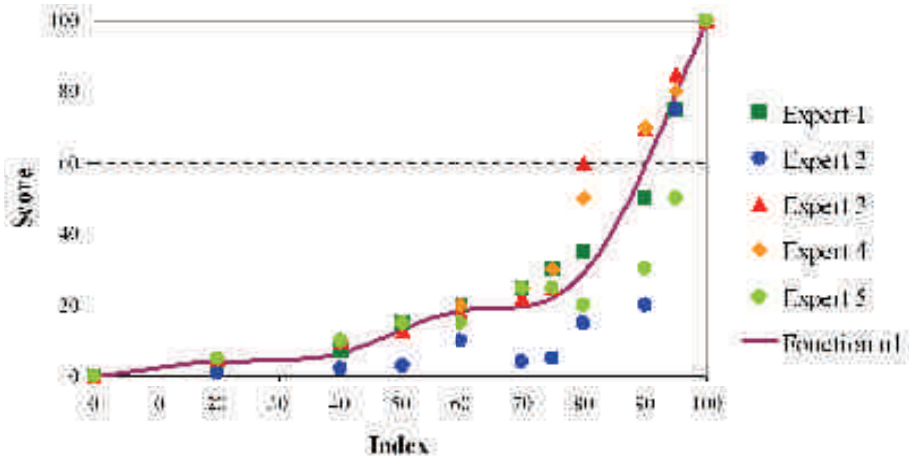


FIG. 5.5 – SC6 - *Altérations du tégument* : approximation en utilisant la fonction d'utilité u_1 (déterminée à l'aide de B-splines avec 4 noeuds intérieurs)

ment être croissante de sorte à maintenir la cohérence avec l'ordre de préférence établi par les experts. Dans le cas des *altérations du tégument*, la fonction u_1 ainsi obtenue est strictement croissante. Malheureusement, dans d'autres situations où la définition d'une fonction d'utilité (reposant toujours sur un index représentant le classement des fermes donné par les experts) a été nécessaire, nous avons parfois obtenu des fonctions non monotones. Ce fut par exemple le cas pour la construction du score partiel du Sous-critère 6 lié à l'évaluation des boiteries (la construction détaillée de la note g_6^2 sera détaillée un peu plus loin). En effet, en utilisant pour définir la fonction d'utilité des fonctions B-splines avec quatre noeuds intérieurs équidistants, on obtient une courbe présentant des "vagues" (cf. dernière case de la Figure 5.6).

Nous avons alors remis en question l'utilisation de fonctions B-splines, et nous avons opté pour l'utilisation de I-splines qui, elles, fournissent des courbes qui sont par nature monotones (Ramsay 1988). Pour définir les fonctions I-splines permettant de minimiser le carré des écarts, nous avons utilisé le logiciel *GenStat®*[10], en utilisant la bibliothèque de procédures *Biometris*[11].

L'utilisation de fonctions I-splines permet d'obtenir des fonctions croissantes. Cependant, cela revient à ajouter des contraintes et on pourrait s'attendre à observer une dégradation de l'erreur résiduelle (correspondant à la somme des carrés des écarts entre le modèle et les notes données par les experts). Sur l'exemple des *boiteries* (présenté sur

[10] *GenStat® for Windows™ 10th Edition* - version 10.1.0.72 (©2007) - VSN International Ltd (Hempstead, Royaume-Uni)

[11] *Biometris Procedure Library for GenStat 10th Edition*, éditée par P.W. Goedhart et J.T.N.M. Thissen

la Figure 5.6), on constate que cette dégradation est minime (en passant d'une erreur de 6283 pour les B-splines à 6288 en utilisant les I-splines). Il apparaît donc que les I-splines permettent de respecter nos contraintes sans pour autant dégrader l'ajustement aux avis des experts.

2. Il est possible d'obtenir une courbe plus lisse en diminuant le nombre de noeuds

En plus de la monotonie, il pourrait être souhaitable d'obtenir des fonctions un peu plus lisses. Dans la quasi-totalité des cas, plus l'index d'une ferme est élevé, plus une petite dégradation de cet index sera pénalisée par les experts. Il semble donc que la sévérité des experts augmente avec l'index, et c'est là le point principal que notre construction doit refléter. Dans la plupart des cas, il ne semble pas que les points d'inflexion locaux pouvant apparaître en utilisant des B-splines avec quatre noeuds intérieurs aient réellement un sens. Nous avons alors cherché une solution permettant d'augmenter la tension de la courbe (c.-à-d. de diminuer les "vagues" sur les courbes) tout en conservant une bonne qualité d'ajustement aux données.

Une solution pourrait alors être de diminuer le nombre de noeuds, et d'ainsi augmenter la tension des portions de courbes polynomiales constituant la fonction d'utilité. Nous avons initialement choisi de définir quatre noeuds intérieurs équidistants, ce qui nous permet, quelle que soit la forme de la courbe souhaitée (cf. plus haut) d'obtenir une courbe satisfaisante. Il est cependant possible, en se basant sur une représentation graphique du nuage de points correspondant aux évaluations données par les experts aux fermes du jeu de données virtuelles, d'identifier la (ou les) zone(s) d'inflexion de la courbe souhaitée et ainsi de choisir le nombre de points d'inflexion de la courbe, ainsi que leurs positions. Il pourrait alors être souhaitable de limiter le nombre de noeuds, ce qui par ailleurs simplifierait l'écriture de la fonction dans le modèle. Il faut cependant que le nombre de noeuds soit suffisant pour fournir une approximation satisfaisante de l'opinion des experts. En utilisant le logiciel GenStat® nous avons pu facilement faire varier le nombre de noeuds. Il s'est avéré qu'avec un seul noeud il était possible d'obtenir des résultats tout à fait satisfaisants.

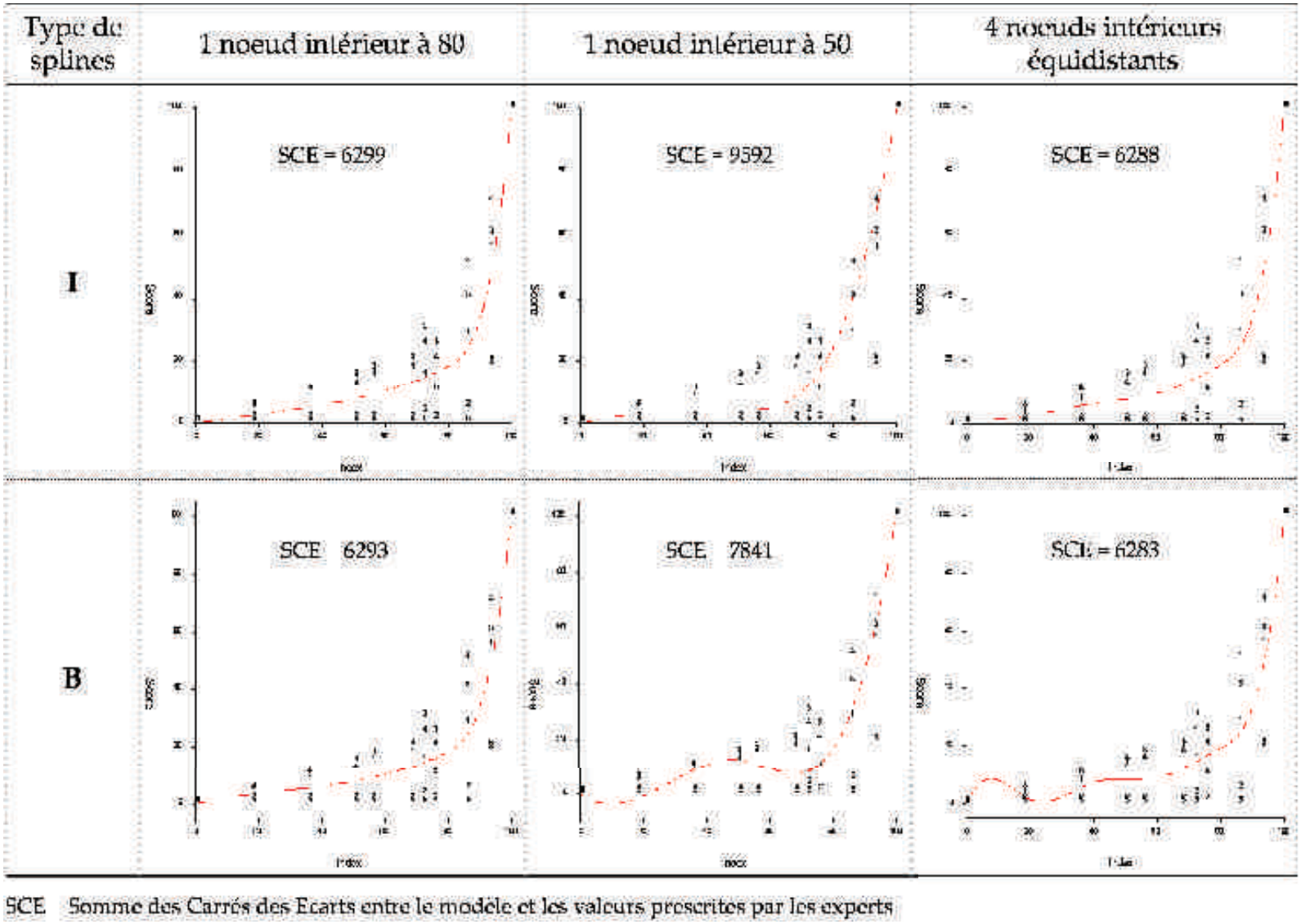


FIG. 5.6 – Comparaison des fonctions d'utilité obtenues avec des B- ou des I-splines, avec 1 noeud (placé à 80 ou 50) ou à 4 noeuds intérieurs, sur l'exemple du score partiel g_6^2 *"Boiteries"*

Il faut cependant souligner qu'en cas de diminution du nombre de noeuds, la position du (ou des) noeud(s) joue un rôle très important. Comme nous pouvons le voir dans le cas de l'évaluation des boiteries (cf. Figure 5.6), le positionnement du noeud à 80 (c.-à-d. dans la zone d'inflexion du nuage de points) donne un résultat excellent, et ce même avec des B-splines (avec une erreur résiduelle de seulement 6293, contre 6283 avec quatre noeuds), alors qu'en positionnant le noeud à 50 nous obtenons de nouveau un phénomène de "vagues" au niveau de la courbe qui s'accompagne d'une dégradation de l'erreur (7841 contre 6293 pour un noeud à 80). En utilisant des I-splines, le positionnement du noeud semble être encore plus important : nous constatons en effet une augmentation de la somme des carrés des écarts de plus de 50% entre l'utilisation d'un noeud positionné à 80 (erreur = 6299) et celle d'un noeud à 50 (erreur = 9592).

Le passage des B-splines à des I-splines, s'accompagnant d'une réduction du nombre de noeuds, apparaît donc comme une méthode mieux adaptée que celle initialement choisie. Il faut cependant apporter un soin particulier au choix du nombre de noeuds et à leurs positions.

Par souci de standardisation de la méthode, nous avons donc décidé de ne plus utiliser pour définir les fonctions d'utilité que des fonctions I-splines, avec généralement un seul noeud dont l'emplacement est déterminé graphiquement en fonction du point d'inflexion du nuage de points correspondant aux avis des experts[12].

Pour les altérations du tégument, nous obtenons ainsi la fonction non-linéaire u_1 représentée sur la Figure 5.7. Nous avons graphiquement positionné le noeud à 75. Le programme écrit dans GenStat® nous permet d'obtenir les équations des deux polynômes de degré 3 générant les portions de courbe situées respectivement avant et après le noeud. La fonction u_1, définie comme suit :

$$u_1(z) = a_z + b_z\, z + c_z\, z^2 + d_z\, z^3 \tag{5.8}$$

où les coefficients a_z, b_z, c_z et d_z varient en fonction de z (cf. Tableau 5.10).

TAB. 5.10 – Coefficients de la fonction polynomiale de degré 3 u_1

Coefficients	$z \in [0, 75]$	$z \in [75, 100]$
a_z	0	-1222,703766
b_z	0,217244452	49,12371424
c_z	-0,002958965	-0,654982845
d_z	6,28369E-05	0,002960161

[12]NB : si plusieurs points d'inflexion apparaissent graphiquement de façon claire, alors deux noeuds seront définis.

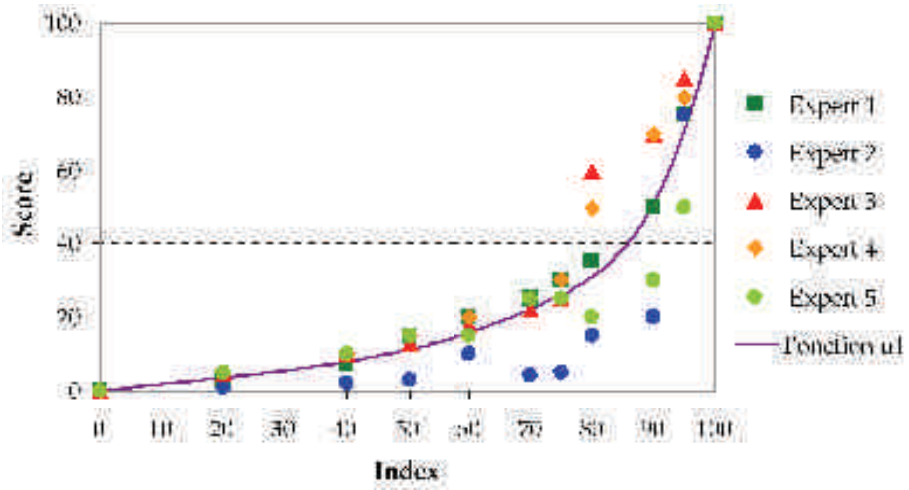


FIG. 5.7 – SC6 - *Altérations du tégument* : approximation en utilisant la fonction d'utilité u_1 (déterminée à l'aide de I-splines avec 1 noeud intérieur à 75)

Une fois la fonction non-linéaire u_1 déterminée, il est désormais possible de calculer la note obtenue pour les *"altérations du tégument"* par n'importe quelle nouvelle ferme x, à partir des données recueillies sur cette ferme. Il suffit en effet de calculer l'index $h_1(x)$ à partir des données $p_0(x)$, $p_1(x)$, $p_2(x)$, puis d'en déduire la note $g_6^1(x)$ en utilisant l'Équation (5.3).

Le calcul de $g_6^1(x)$ en utilisant l'Équation (5.3) peut aussi être effectué sur les fermes de références $x_1, \ldots, x_{11}$. Les résultats obtenus sont présentés dans la dernière colonne du Tableau 5.7 et représentés graphiquement sur la Figure 5.7. Comme nous pouvons le voir, le modèle rend bien compte des valeurs de référence fournies par les experts dans leur globalité, sans toutefois tenir compte des différences de sévérité pouvant exister entre certains experts. Par exemple, on voit clairement que l'Expert 2 est plus sévère que l'Expert 3, par conséquent pour une ferme x donnée la valeur $g_6^1(x)$ se retrouve au milieu, mais parfois éloignée, des valeurs $y_2(x)$ et $y_3(x)$.

5.3.2.2 Évaluation des boiteries : note $g_6^2(x)$

L'évaluation des boiteries au niveau de la ferme repose sur une construction similaire à celle décrite ci-dessus pour les altérations du tégument. Cependant, l'étape de simplification de l'information pour passer de mesures cardinales à une mesure ordinale au niveau individuel n'est pas nécessaire. En effet, l'évaluation de la boiterie d'un animal est directement évaluée sur une échelle ordinale à trois niveaux (cf. Tableau 5.11), correspondant là aussi à trois niveaux de sévérité du problème vis-à-vis du bien-être

(absence de problème, problème modéré, problème sérieux).

TAB. 5.11 – Évaluation des boiteries au niveau individuel

3 catégories sont définies au niveau individuel :		Problème vis-à-vis du bien-être :
0	La vache ne boite pas (démarche équilibrée et régulière avec répartition homogène du poids de l'animal sur ses quatre pattes)	Pas de problème
1	la vache boite légèrement (démarche avec des foulées irrégulières due à une répartition non homogène du poids de l'animal sur ses quatre pattes)	Modéré
2	la vache boite sérieusement (démarche très irrégulière causée par une forte réticence de la vache à s'appuyer sur au moins l'une de ses pattes)	Sérieux

Au niveau de la ferme, une ferme x est donc caractérisée par ses pourcentages de vaches dans chacune des trois catégories décrites ci-dessus, notés respectivement $p_0(x)$, $p_1(x)$ et $p_2(x)$. En suivant la même procédure de construction que celle décrite pour les *altérations du tégument*, nous sommes capables de calculer, pour n'importe quelle nouvelle ferme x, sa note $g_6^2(x)$ correspondant au score de sous-critère partiel obtenu par la ferme sur les boiteries (cf. Équation 5.9).

$$g_6^2(x) = u_2\left(h_2(x)\right) = u_2\left(100 - \frac{\sum_{k=0}^{2} w_k\, p_k(x)}{w_2}\right) \tag{5.9}$$

où $w_0 = 0$, $w_1 = 2$ et $w_2 = 7$.

Nous avons obtenu la fonction u_2 (cf. Equation 5.10) en utilisant Genstat® pour trouver la fonction I-spline avec un noeud intérieur à 80 permettant de minimiser la somme des carrés des écarts. La fonction u_2 ainsi obtenue est représentée sur la Figure 5.8.

$$u_2(z) = a_z + b_z\, x + c_z\, x^2 + d_z\, x^3 \tag{5.10}$$

où les coefficients a_z, b_z, c_z et d_z varient en fonction de z (cf. Tableau 5.12).

TAB. 5.12 – Coefficients de la fonction polynomiale de degré 3 u_1

Coefficients	$z \in [0, 80]$	$z \in [80, 100]$
a_z	0	-4325,222235
b_z	0,174639325	162,370479
c_z	-0,002182992	-2,029631053
d_z	3,67844E-05	0,008484485

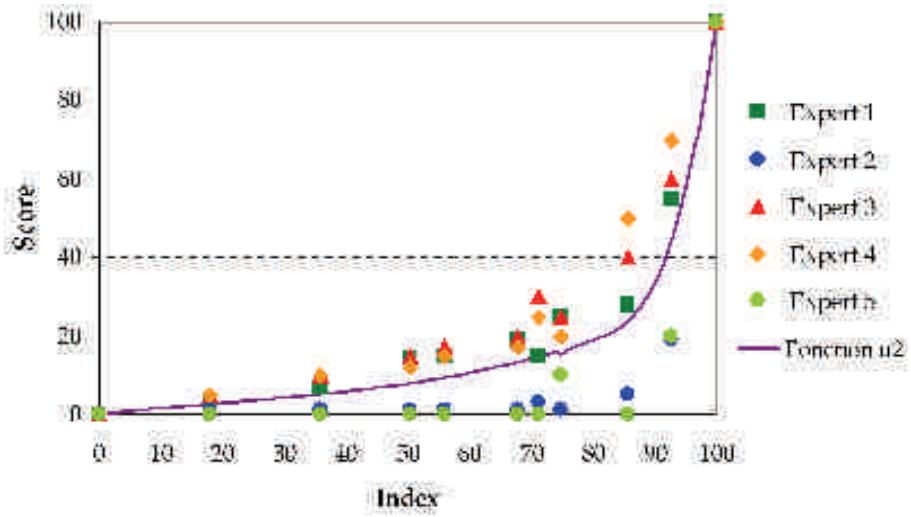


FIG. 5.8 – SC6 - *Boiteries* : approximation en utilisant la fonction d'utilité u_2 (déterminée à l'aide de I-splines avec 1 noeud intérieur à 80)

5.3.2.3 Agrégation des notes $g_6^1(x)$ et $g_6^2(x)$ pour obtenir le score de sous-critère $g_6(x)$

Afin de définir le score de sous-critère $g_6(x)$ permettant de noter la ferme x vis-à-vis de l'*Absence de blessures*, nous devons agréger les scores partiels obtenus sur chacun des deux types de blessures (*altérations du tégument* et *boiteries*), notés respectivement $g_6^1(x)$ et $g_6^2(x)$. Pour ce faire, nous avons défini une fonction d'agrégation $\varphi : [0, 100] \longrightarrow [0, 100]$ (cf. Équation 5.11).

$$g_6(x) = \varphi\left(g_6^1(x), g_6^2(x)\right) \tag{5.11}$$

Nous avons demandé à des experts de réagir sur un petit jeu de données composé de trois fermes virtuelles caractérisées par leurs notes sur les altérations du tégument et les boiteries. Afin d'identifier le raisonnement suivi par les experts, nous leur avons demandé de donner à chaque ferme un score pour le Sous-critère *"Absence de blessures"*. Les trois fermes choisies pour le jeu de données (cf. Tableau 5.13) permettent d'identifier deux aspects importants du raisonnement pouvant être suivis par les experts :

1. Les experts attribuent-ils plus d'importance aux boiteries qu'aux altérations du tégument (ou inversement) ?
2. Les experts attribuent-ils plus d'importance à la plus mauvaise des deux notes (et ainsi limitent les compensations), ou à l'inverse plus d'importance à la meilleure note (et dans ce cas-là favorisent les compensations) ?

L'analyse des réponses fournies par les experts (présentées dans le Tableau 5.13) montre que ceux-ci attribuent plus d'importance aux boiteries, tout en donnant un peu plus de poids à la plus mauvaise note. Pour tenir compte de ces deux types de raisonnement concomitants, nous avons utilisé l'intégrale de Choquet pour φ dans l'Équation (5.11) pour calculer le score de sous-critère g_6. La dernière colonne du Tableau 5.13 présente le score de sous-critère calculé, en le comparant avec les scores donnés par les experts (et pour chaque ferme la moyenne des scores des experts), on peut facilement constater que l'intégrale de Choquet fournit des résultats en parfaite adéquation avec les réponses des experts.

TAB. 5.13 – Réponses des experts pour l'agrégation de notes g_6^1 et g_6^2 pour évaluer le Sous-critère 6 *"Absence of injuries"* (g_6)

Farm x	$g_6^1(x)$	$g_6^2(x)$	E1	E2	E3	E4	*Moyenne* $g_6(x)$	**Score**
x	**40**	**60**	50	50	50	55	*51*	**51**
y	**50**	**50**	50	50	50	50	*50*	**50**
z	**60**	**40**	40	50	50	45	*46*	**46**

La justification du choix de l'intégrale de Choquet pour φ est beaucoup plus détaillée dans le Chapitre 6, dédié à la construction des critères à partir des sous-critères.

5.3.3 Exemple 3 : Sous-critère g_3 *"Confort autour du repos"*

Le Sous-critères *"Confort autour du repos"* est évalué sur la base de six mesures réalisées en ferme (cf. Tableau 5.14). Ces mesures sont prises soit au niveau individuel, soit directement au niveau du troupeau. Elles sont aussi exprimées sur des échelles différentes (cardinales ou ordinales) et n'ont pas forcément la même importance vis-à-vis du bien-être animal. Ces six mesures permettent de couvrir deux aspects différents du Sous-critère 3 : le comportement des vaches autour du repos (les vaches indiquant par leurs comportements si leur logement est bien conçu) et la propreté des vaches (le fait que des vaches soient sales pouvant principalement s'expliquer par un défaut de nettoyage de l'étable, et notamment de la zone de couchage).

Ces six mesures sont de natures très différentes. En ce qui concerne les mesures prises au niveau individuel, elles sont observées sur des animaux pouvant être différents. Par ailleurs, l'information recueillie en ferme et permettant d'évaluer le Sous-critère 3 est très riche. Pour ces raisons, il est possible de simplifier l'information de base pour obtenir des données plus facilement traitables au niveau du troupeau tout en permettant d'évaluer de manière suffisamment précise le *"Confort autour du repos"*. Nous avons donc décidé de

TAB. 5.14 – Liste des mesures utilisées pour construire le Sous-critère *"Confort autour du repos"*

Aspect du confort	**Mesures**	**Niveau d'observation**	**Nature de l'échelle**
Comportement autour du repos	Temps mis par les vaches pour se coucher	Individu	Cardinale
	Pourcentage d'animaux entrant en collision avec une partie du matériel de logement (ex. barres délimitant les logettes) au moment où ils se couchent	Ferme	Cardinale
	Pourcentage de vaches couchées partiellement en dehors de la zone de couchage	Ferme	Cardinale
Propreté	Propreté de la mamelle	Individu	Ordinale
	Propreté du flanc et de la cuisse	Individu	Ordinale
	Propreté des pattes arrières	Individu	Ordinale

transformer chaque mesure en une donnée ordinale au niveau du troupeau. Nous avons ainsi défini trois catégories pouvant s'appliquer à chacune des mesures. Ces catégories sont les mêmes que celles utilisées pour le Sous-critère 6 *"Absence de blessures"* : absence de problème, problème modéré, problème sérieux.

Pour transformer les mesures M_i sur cette échelle en trois niveaux, commune à l'ensemble des mesures, nous avons dû définir pour chaque mesure des seuils permettant de définir les limites entre les trois catégories de problème. Pour la mesure i, ces seuils sont notés T_i^1 et T_i^2, respectivement pour la limite supérieure de la catégorie "pas de problème" et pour la limite inférieure de la catégorie "problème sérieux". Ainsi, chaque mesure est recodée de la manière décrite dans l'Algorithme 2.

Pour définir ces seuils, nous avons consulté pour chaque mesure les chercheurs ayant participé à son choix et à son développement dans Welfare Quality®. Les valeurs ainsi retenues sont présentées dans le Tableau 5.15.

La première étape est alors d'obtenir pour chaque mesure une évaluation au niveau du troupeau. Pour deux des mesures, la donnée brute est déjà au niveau du troupeau, mais pour les quatre autres ce n'est pas le cas.

En ce qui concerne le temps mis par les vaches pour se coucher, les chercheurs ayant développé cette mesure considèrent qu'au niveau du troupeau la donnée à prendre en

Algorithme 2 : Pour chacune des mesures, procédure d'affectation d'une ferme x à l'une des trois catégories de problème

$M_i(x) \leftarrow$ donnée obtenue au niveau du troupeau par la ferme x sur la mesure i
$T_i^1 \leftarrow$ seuil fixé par les experts pour la mesure i pour marquer la limite supérieure de la catégorie "pas de problème"
$T_i^2 \leftarrow$ seuil fixé par les experts pour la mesure i pour marquer la limite inférieure de la catégorie "problème sérieux"

for $i = 1$ **to** 6 **do**
  if $M_i(x) \leq T_i^1$ **then** $M_i(x) \leftarrow 0$ *"pas de problème"*
  else if $M_i(x) \leq T_i^2$ **then** $M_i(x) \leftarrow 1$ *"problème modéré"*
  else $M_i(x) \leftarrow 2$ *"problème sérieux"*
end

TAB. 5.15 – Seuils définis à l'aide des experts pour chacune des six mesures

Mesure		T_i^1	T_i^2
M_1	Temps moyen mis par les vaches pour se coucher	5,2s	6,3s
M_2	Pourcentage d'animaux entrant en collision avec une partie du matériel de logement (ex. barres délimitant les logettes) au moment où ils se couchent	3%	5%
M_3	Pourcentage de vaches couchées partiellement en dehors de la zone de couchage	20%	30%
M_4	Pourcentage de vaches sales au niveau de la mamelle	20%	50%
M_5	Pourcentage de vaches sales au niveau du flanc et de la cuisse	10%	19%
M_6	Pourcentage de vaches sales au niveau des pattes arrières	10%	19%

compte est le temps moyen mis par les vaches observées pour se coucher.

Les trois mesures de propreté sont prises au niveau individuel sur une échelle allant de 1 (propre) à 4 (très sale). Les chercheurs ayant proposé ces mesures considèrent que, pour une zone donnée, seules les vaches sales ou très sales sont à prendre en compte, et proposent ainsi une simplification de l'échelle en une donnée binaire "propre" (note de propreté ≤ 2) *vs* "sale" (note de propreté ≥ 3). Au niveau du troupeau, nous avons décidé de considérer, pour chacune des zones, le pourcentage de vaches sales.

Dans un deuxième temps, nous devons calculer à partir des mesures M_i le score de sous-critère g_3. Cette étape se déroule en trois temps :

1. Pour chaque mesure $i \in \{1, ..., 6\}$, nous comparons les données M_i obtenues sur la ferme $x \in X$ (soit directement, soit après la première étape consistant à passer d'une mesure individuelle à une mesure au niveau du troupeau) aux seuils T_i^1 et T_i^2. Ainsi, nous recodons l'information fournie par les mesures en une donnée ordinale (cf. Algorithme 2).

2. Ensuite, nous comptons le nombre de points considérés comme n'étant pas problématiques, le nombre de ceux jugés modérément problématiques, et enfin le nombre de ceux jugés sérieusement problématiques.
   Les six mesures permettent d'évaluer deux aspects différents du Sous-critère 3 : le *comportement autour du repos* et la *propreté*. Ces deux aspects n'ont pas forcément la même importance vis-à-vis de l'évaluation du *"Confort autour du repos"*. Il pourrait donc être nécessaire d'introduire une pondération des mesures afin d'aboutir à un équilibre au niveau de l'importance des deux aspects. Nous nous retrouvons alors avec un nombre d'*alertes* pour chaque catégorie de problème : n_0 pour la catégorie *"pas de problème"*, n_1 pour la catégorie *"problème modéré"* et n_2 pour la catégorie *"problème sérieux"*. Ce nombre d'alertes tient compte de l'importance relative accordée à chacun des deux aspects "comportement autour du repos" et "propreté", se traduisant par des poids à affecter aux différentes mesures et notés respectivement ω_c et ω_p (cf. Algorithme 3). Pour déterminer les poids ω_c et ω_p nous avons demandé aux chercheurs impliqués dans le choix et le développement de ces mesures de nous indiquer l'importance relative qu'ils souhaitent voir attribuer à chacun des aspects. Nous sommes rapidement arrivés à un consensus, avec les valeurs suivantes : $\omega_c = 3$ et $\omega_p = 1$.

Algorithme 3 : Procédure de comptage du nombre d'alertes obtenues par une ferme x dans chacune des trois catégories de problèmes

$M_i(x) \leftarrow$ donnée ordinale (exprimée sur l'échelle en trois niveaux 0, 1 ou 2) au niveau du troupeau obtenue par la ferme x sur la mesure i
$\omega_c \leftarrow 3$ (poids attribué par les experts à l'aspect *"comportement autour du repos"*)
$\omega_p \leftarrow 1$ (poids attribué par les experts à l'aspect *"propreté"*)
$n_0(x) \leftarrow 0$
$n_1(x) \leftarrow 0$
$n_2(x) \leftarrow 0$

for $i = 1$ **to** 3 **do**
  if $M_i(x) = 0$ **then** $n_0(x) \leftarrow n_0(x) + \omega_c$
  else if $M_i(x) = 1$ **then** $n_1(x) \leftarrow n_1(x) + \omega_c$
  else $n_2(x) \leftarrow n_2(x) + \omega_c$
end
for $i = 4$ **to** 6 **do**
  if $M_i(x) = 0$ **then** $n_0(x) \leftarrow n_0(x) + \omega_p$
  else if $M_i(x) = 1$ **then** $n_1(x) \leftarrow n_1(x) + \omega_p$
  else $n_2(x) \leftarrow n_2(x) + \omega_p$
end

3. Chaque ferme x est alors caractérisée par son nombre d'alertes dans chacune des trois catégories de problème, notés respectivement $n_0(x)$, $n_1(x)$ et $n_2(x)$. En suivant la même procédure de construction que celle décrite pour le Sous-critère 6, nous sommes capables de calculer, pour n'importe quelle nouvelle ferme x, sa note $g_3(x)$ correspondant au score de sous-critère obtenu par la ferme sur le Sous-critère 3 *"Confort autour du repos"* : nous calculons d'abord un index $h(x)$ permettant de refléter l'ordre de préférence établi par les experts sur un jeu de données virtuelles, puis nous définissons une fonction non-linéaire u qui permet d'obtenir le score $g_3(x)$ à partir de $h(x)$ (cf. Équation 5.12).

$$g_3(x) = u\left(h(x)\right) = u\left(\frac{100}{n_{tot}} - \frac{\sum_{k=0}^{2} w_k\, n_k(x)}{w_2}\right) \tag{5.12}$$

où $w_0 = 0$, $w_1 = 4$ et $w_2 = 9$, avec $n_{tot} = 3\,\omega_c + 3\,\omega_p = 12$ le nombre total d'alertes.

Nous avons obtenu la fonction u (cf. Equation 5.13) en utilisant Genstat® pour trouver la fonction I-spline avec un noeud intérieur à 65 permettant de minimiser les carré des écarts. La fonction u_2 ainsi obtenue est représentée sur la Figure 5.9.

$$u(z) = a_z + b_z\, z + c_z\, z^2 + d_z\, z^3 \tag{5.13}$$

où les coefficients a_z, b_z, c_z et d_z varient en fonction de z (cf. Tableau 5.16).

TAB. 5.16 – Coefficients de la fonction polynomiale de degré 3 u

Coefficients	$z \in [0, 65]$	$z \in [65, 100]$
a_z	0	-35,20034002
b_z	0,78646981	2,411100892
c_z	-0,009321937	-0,034316262
d_z	0,000109077	0,000237253

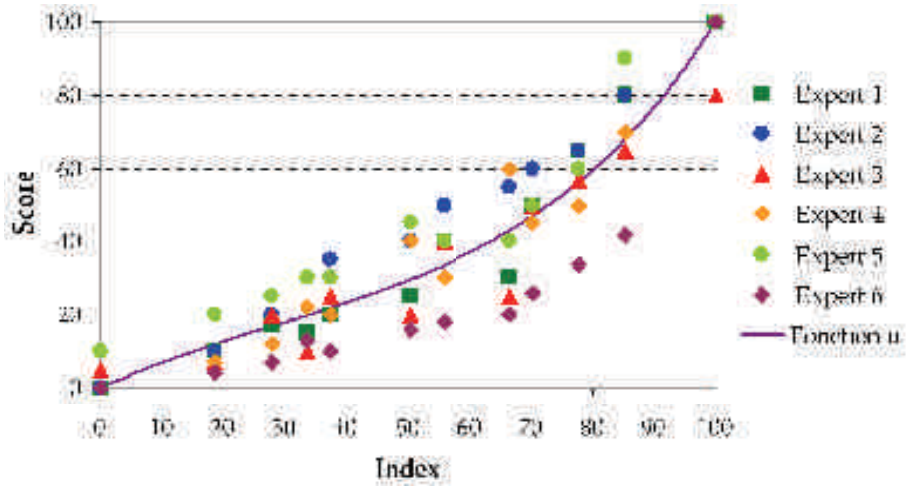


FIG. 5.9 – SC3 : approximation en utilisant la fonction d'utilité u (déterminée à l'aide de I-splines avec 1 noeud intérieur à 65)

5.4 Expérimentations sur jeux de données réelles et affinage de la construction des sous-critères

Dans le cadre du Groupe de tâches WP2.4, des observations ont été menées dans 4 pays européens (l'Autriche, l'Allemagne, l'Italie et la République Tchèque) sur un total de 91 fermes, dont 71 en stabulation libre et 20 en étable entravée (cf. Tableau 5.17 pour le détail par pays). Le calcul des scores de sous-critères sur ces 91 fermes a mis en évidence pour certains sous-critères une probable trop grande sévérité des experts consultés.

TAB. 5.17 – Répartition géographique et par type d'étable des fermes observées dans le WP2.4

Pays	**Stabulation libre**	**Étable entravée**	***Total***
Autriche	15	10	*25*
Allemagne	16	10	*26*
Italie	25	0	*25*
République Tchèque	15	0	*15*
Total	*71*	*20*	***91***

Les experts consultés n'avaient pas toujours d'idée précise des valeurs pouvant être obtenues dans des fermes commerciales (certaines des mesures n'ayant été testées jusque là que dans des conditions expérimentales). Nous avons alors décidé de revenir auprès d'eux en leur présentant de nouveau les mêmes jeux de données virtuelles que la première fois mais, cette fois-ci, accompagnés d'informations supplémentaires quant à la distribution des 91 fermes enquêtées lors du WP2.4 sur les différentes mesures. Les experts ont alors légèrement revu leurs évaluations initiales, sans toutefois être réellement moins sévères.

En effet, comme il est facile de le voir sur les histogrammes représentant la distribution des fermes après correction des paramètres de calcul des scores de sous-critères (Figures 5.10, 5.11, 5.12 et 5.13), les évaluations obtenues restent parfois très sévères. Par exemple, aucune ferme ne dépasse le score de 60 pour le Sous-critère 3 *"Confort autour du repos"* (voir Figure 5.11).

Deux explications sont alors possibles. Soit les fermes enquêtées ne sont pas bonnes sur la plupart des sous-critères. Si on admet que l'échantillon de 91 fermes est représentatif de l'élevage en Europe, il faudra alors admettre que le bien-être des vaches laitières en Europe est loin d'être bon sur plusieurs aspects. Soit les experts consultés ont été trop sévères sur certains sous-critères.

Pour la construction du modèle final d'évaluation du bien-être des vaches laitières en ferme (c.-à-d. une fois la liste finale des mesures établie), il sera primordial de consulter beaucoup plus d'experts (une vingtaine d'experts semblerait être une bonne base[13]), en incluant, en plus de chercheurs, des professionnels de l'élevage et des vétérinaires ayant une bonne connaissance à la fois de la réalité des élevages et des mesures. Il faudrait également disposer d'un plus grand nombre de données issues de fermes européennes. Si les évaluations restent aussi sévères, il faudra alors en conclure que le niveau de bien-être assuré dans la grande majorité des 91 fermes enquêtées dans le WP2.4 n'est pas bon.

En conclusion, nous proposons d'exprimer tous les sous-critères sur une échelle d'utilité commune $[0, 100]$, en utilisant pour les construire deux méthodes principales :

- la définition d'un arbre de décision permettant de déterminer un nombre fini de situations possibles en ferme ;
- le calcul d'un index issu d'une combinaison linéaire des mesures recodées au niveau du troupeau sur une échelle ordinale à trois ou quatre niveaux. Une fonction d'utilité est ensuite définie à l'aide de fonctions splines pour recoder l'index sur l'échelle d'utilité.

Dans tous les cas, le paramétrage de la construction repose sur la réaction d'experts (spécialistes des mesures impliquées dans la construction des sous-critères) sur des jeux de données virtuelles. Le test de la construction des sous-critères sur un jeu de fermes réelles a cependant mis en évidence la nécessité de revoir le paramétrage des constructions une fois la liste définitive des mesures établie.

[13] ce qui correspond par exemple au nombre d'experts consultés par Bracke et al. 1999 pour construire leur modèle d'évaluation du bien-être des porcs

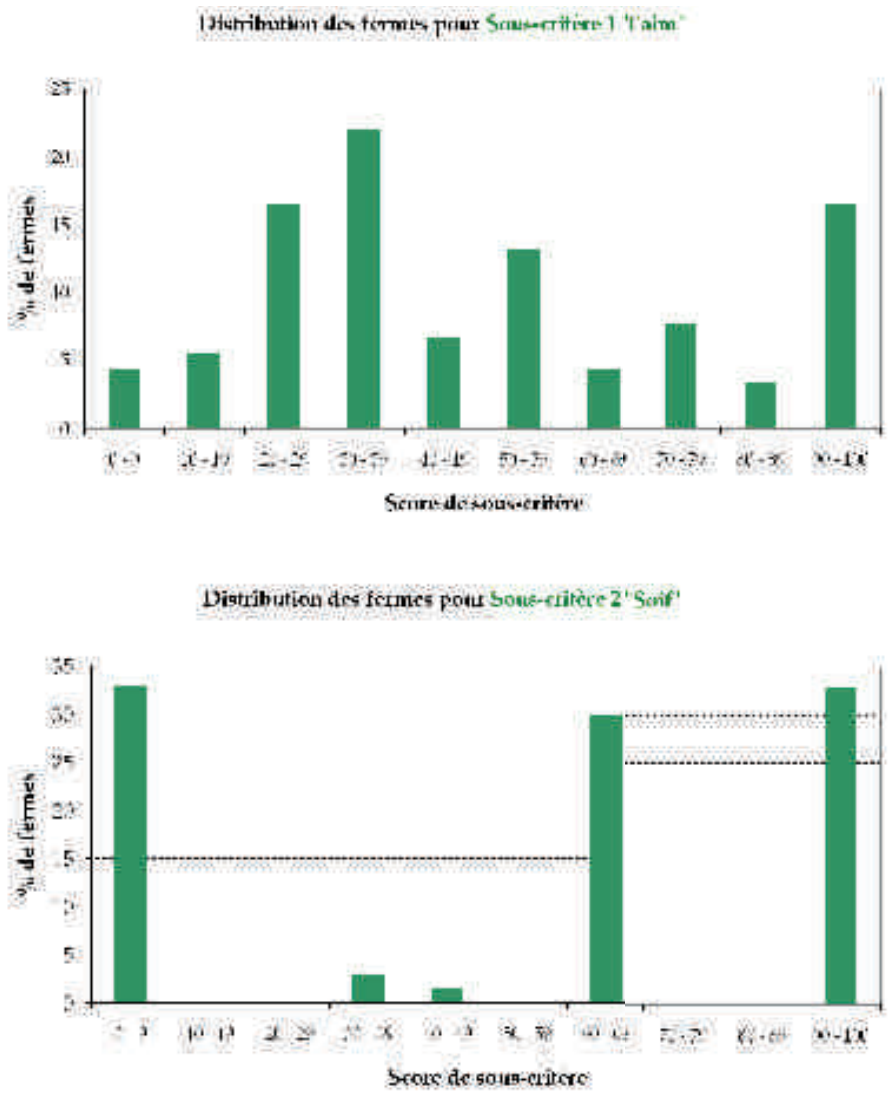


FIG. 5.10 – Distribution des fermes sur les sous-critères composant le Critère *"Alimentation adaptée"*

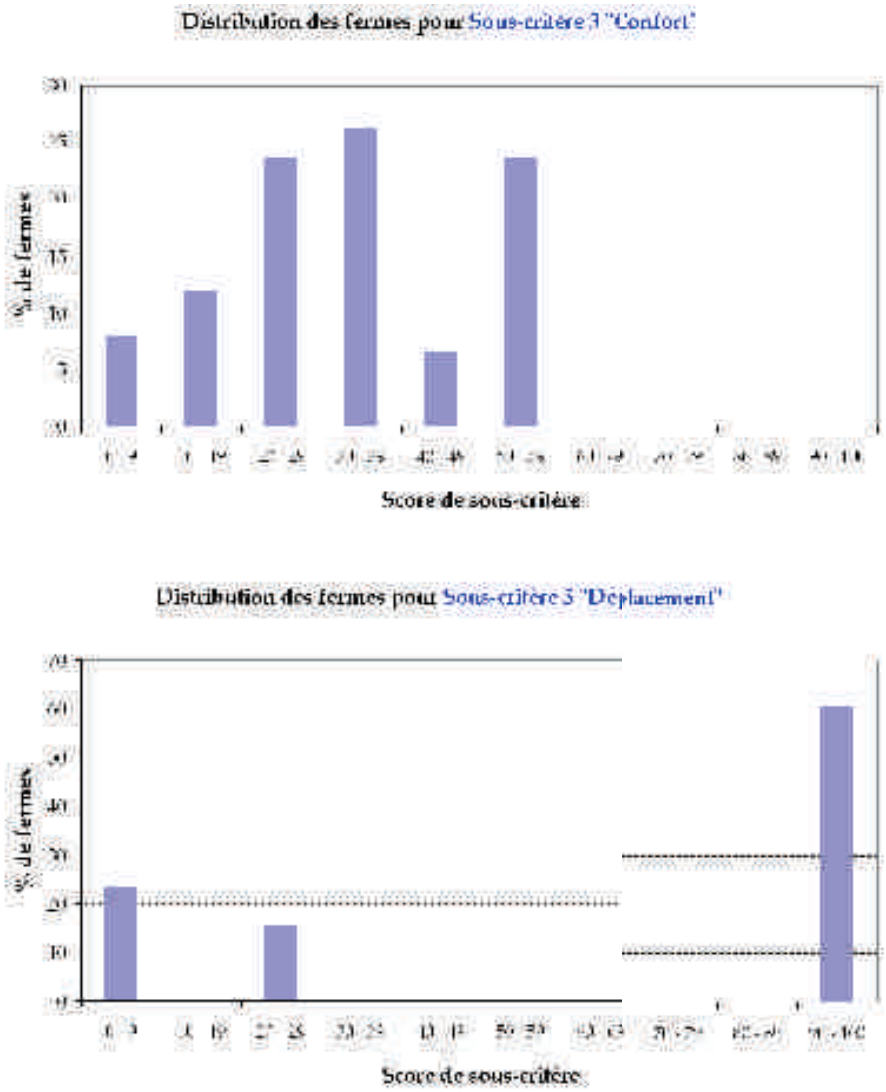


FIG. 5.11 – Distribution des fermes sur les sous-critères composant le Critère *"Logement correct"*

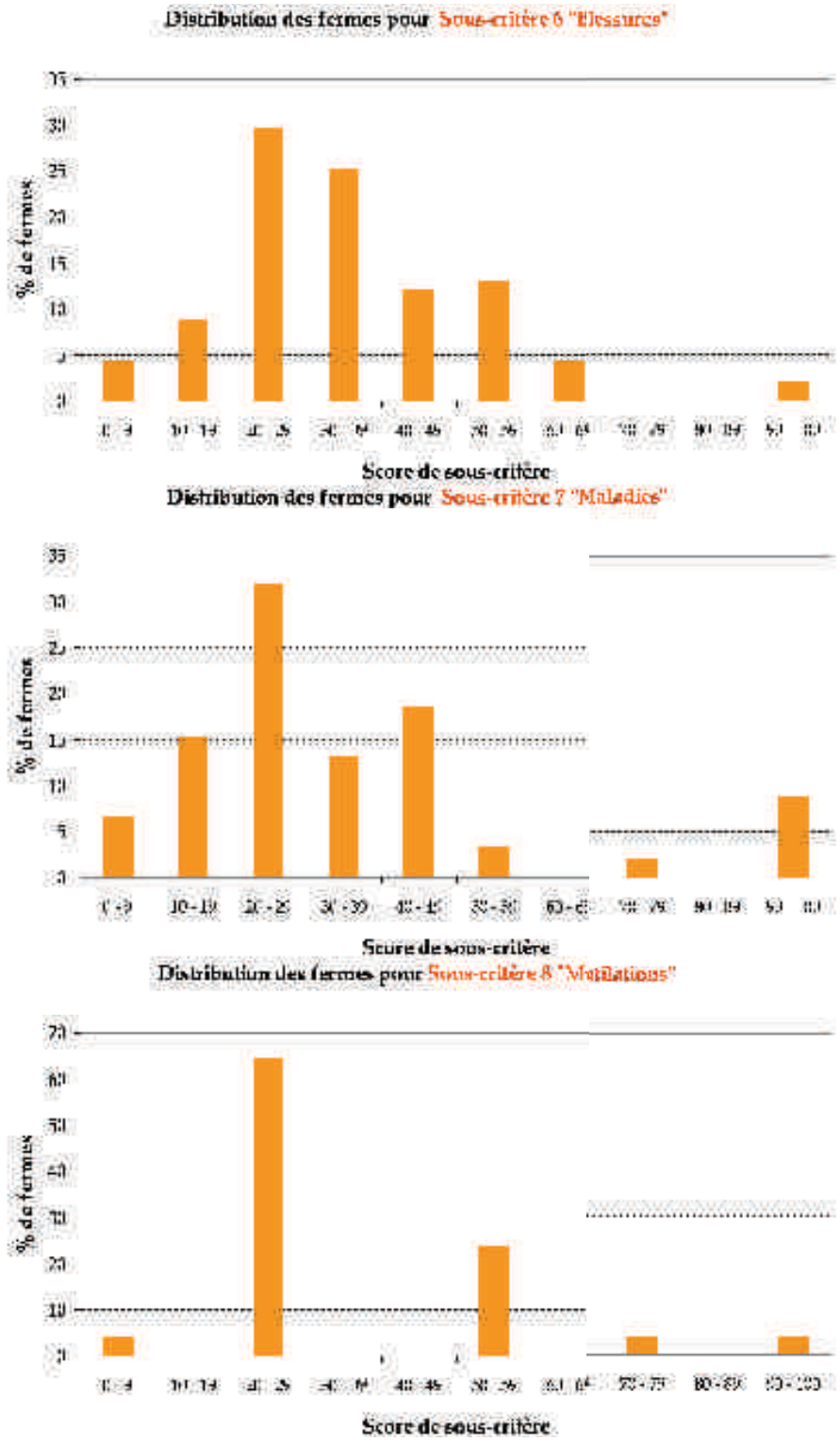


FIG. 5.12 – Distribution des fermes sur les sous-critères composant le Critère *"Bonne santé"*

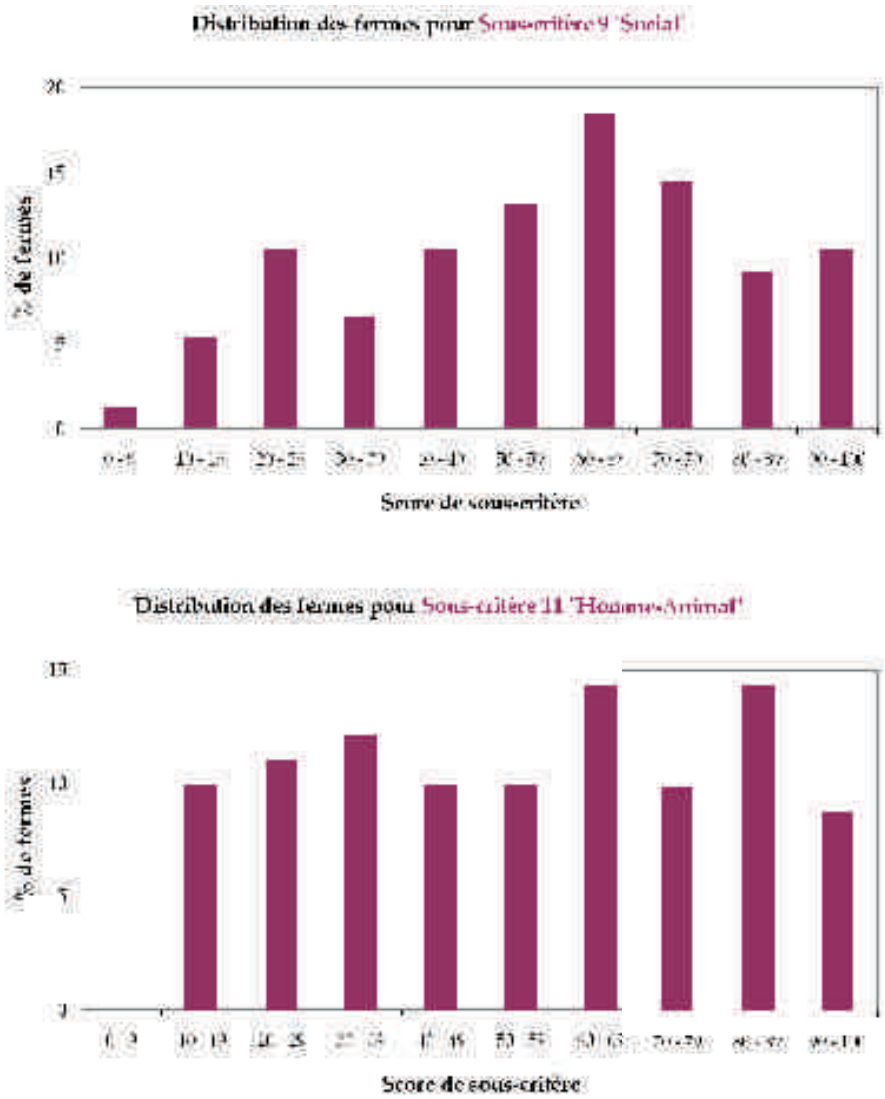


FIG. 5.13 – Distribution des fermes sur les sous-critères composant le Critère *"Comportement approprié"*

Chapitre 6

La construction des critères

Résumé de la construction des critères :

Les scores de sous-critères sont agrégés pour évaluer les quatre critères : *Alimentation adaptée*, *Logement correct*, *Bonne santé* et *Comportement approprié*.

Pour ce faire, nous utilisons l'intégrale de Choquet. Cet agrégateur nous permet de refléter les deux comportements suivis par les experts lors de l'agrégation des scores de sous-critères :

- attribuer plus de poids à certains sous-critères jugés plus importants que les autres dans la construction des critères ;
- pénaliser une ferme sur ses mauvais scores en attribuant plus d'importance aux scores de sous-critères les plus bas (en d'autres termes, limiter les compensations entre les sous-critères).

Dans tous les cas, le paramétrage de la construction repose sur la réaction d'experts sur des jeux de données virtuelles. À ce stade de la construction, les experts consultés sont des chercheurs en sciences animales et en sciences sociales impliqués dans les décisions du projet Welfare Quality®.

Chapitre 6

La construction des critères

Dans le chapitre précédent, nous avons défini les 12 sous-critères comme des fonctions $g_j : X \longrightarrow [0, 100],\ j = 1, \ldots, 12$ où X représente l'ensemble des fermes devant être évaluées. Pour la construction de l'ensemble des sous-critères, nous avons utilisé la même échelle d'évaluation $[0, 100]$. Nous devons maintenant définir formellement les critères permettant d'évaluer si l'alimentation est adaptée aux animaux, si le logement est correct, si les animaux sont en bonne santé et si leur comportement est approprié. Chaque critère est le résultat de l'agrégation des sous-critères qui le composent (de 2 à 4 sous-critères en fonction du critère). De manière plus précise, au regard du Tableau 4.2 présentant la famille de critères utilisée, les fonctions de critère $f_i : X \longrightarrow [0, 100]$ ont besoin d'être définies :

- Critère f_1 (Alimentation adaptée) est défini par l'agrégation de g_1, g_2, et s'écrit formellement $f_1(x) = \chi_1(g_1(x), g_2(x))$
- Critère f_2 (Logement correct) est défini par l'agrégation de g_3, g_4, g_5, et s'écrit formellement $f_2(x) = \chi_2(g_3(x), g_4(x), g_5(x))$
- Critère f_3 (Bonne santé) est défini par l'agrégation de g_6, g_7, g_8, et s'écrit formellement $f_3(x) = \chi_3(g_6(x), g_7(x), g_8(x))$
- Critère f_4 (Comportement approprié) est défini par l'agrégation de $g_9, g_{10}, g_{11}, g_{12}$, et s'écrit formellement $f_4(x) = \chi_4(g_9(x), g_{10}(x), g_{11}(x), g_{12}(x))$

6.1 Consultation d'experts sur des jeux de données virtuelles

Afin de nous aider à définir pour chaque critère $f_i,\ i = 1, \ldots, 4$ la fonction d'agrégation χ_i la plus adaptée, puis d'en préciser les paramètres, nous avons consulté des chercheurs à la fois en sciences animales et sociales, impliqués dans le projet Welfare Quality®. À ce stade de la construction, nous n'avons pas besoin de poser uniquement des questions à des spécialistes des mesures (les données brutes étant déjà interprétées). Par contre, les points de vue des utilisateurs potentiels peuvent être pris en compte. C'est pour cela qu'en plus de chercheurs en sciences animales, connaissant bien les mesures servant à évaluer le bien-être, nous avons aussi consulté des chercheurs en sciences sociales ayant au sein de Welfare Quality® mené des études approfondies sur les attentes des consommateurs, des distributeurs et des éleveurs. Ainsi, ces scientifiques pourront, par leurs réponses, inclure les points de vues des principaux utilisateurs potentiels du système d'évaluation. Quatorze experts ont ainsi répondu à nos questions sur l'évaluation du bien-être des vaches laitières en ferme. Nous leur avons demandé de réagir sur des jeux de données virtuelles (un par critère), composés de fermes caractérisées par les scores qu'elles obtiennent sur les q sous-critères devant être considérés pour un critère donné. Nous leur avons demandé d'attribuer un score de critère (sur l'échelle $[0, 100]$ précédemment utilisée pour construire les sous-critères) à chacune des fermes composant le jeu de données considéré.

Un exemple de jeu de donné ayant été présenté aux experts est fourni sur la Figure 6.1, il s'agit du jeu de données correspondant au Critère 2 *"Logement correct"*, résultant de l'agrégation de trois sous-critères : le SC3 *"Confort autour du repos"*, le SC4 *"Confort thermique"* et le SC5 *"Facilité de déplacement"*. Nous avons demandé aux experts de réagir sur des données issues de la totalité des 12 sous-critères, même si certains ne peuvent pour l'instant pas être évalués faute de mesures faisables et répétables. En effet, dans la liste définitive des mesures nous espérons que figureront des éléments permettant d'évaluer les sous-critères qui sont pour l'instant "orphelins". À terme, l'ensemble des sous-critères pourront donc être évalués en ferme, c'est pourquoi nous les avons d'ores et déjà tous inclus dans la construction des critères. Pour l'évaluation du Critère 2, c'est par exemple le cas du *confort thermique* pour lequel aucune mesure n'est pour le moment disponible, mais pour lequel des mesures sont à l'étude.

D'après les réponses fournies par les experts, nous serons capables d'identifier leur(s) mode(s) de raisonnement. Une fois identifiés, nous pourrons alors modéliser la procédure d'agrégation des sous-critères en utilisant l'agrégateur le mieux adapté.

Criterion 2: Good housing

SC3: Comfort around resting	SC4: Thermal comfort	SC5: Ease of movement	Criterion score
25	60	75	
25	75	50	
40	50	80	
40	60	60	
50	25	75	
50	40	60	
50	60	50	
50	80	40	
50	75	25	
60	40	50	
60	50	40	
75	25	50	
75	50	25	

Legend
25
40
50
60
75

You may explain below your reasoning

FIG. 6.1 – Jeu de données présenté aux experts pour construire les fonctions f_i : exemple du Critère 2 *"Logement correct"*

6.2 Synthèse de l'information fournies par les sous-critères

Il s'agit de définir $f_i(x)$ de sorte qu'elle mesure la performance de la ferme $x \in X$ vis-à-vis du Critère f_i, ce critère étant le résultat de l'agrégation de q sous-critères ($q = 2$, 3 ou 4, en fonction du critère considéré). Nous devons donc maintenant déterminer les fonctions d'agrégations $\chi_i : [0, 100] \longrightarrow [0, 100]$, $i = 1, \ldots, 4$ permettant au mieux de refléter le raisonnement suivi par les experts pour agréger les sous-critères. De telles fonctions d'agrégations sont utilisées pour, en quelque sorte, moyenner les performances obtenues sur les sous-critères correspondants. En tant qu'opérateurs de compromis, ces fonctions sont sensées respecter les axiomes suivants :

$$\begin{array}{ll} Compensation: & \min(z_1, \ldots, z_q) \leq \chi_i(z_1, \ldots, z_q) \leq \max(z_1, \ldots, z_q) \\ Monotonie: & (\forall j = 1, \ldots, q,\ z_j \leq y_j) \Rightarrow \chi_i(z_1, \ldots, z_q) \leq \chi_i(y_1, \ldots, y_q) \end{array}$$

Nous avons focalisé notre analyse sur trois opérateurs de compromis différents :

- la somme pondérée (notée ***WS*** pour *Weighted Sum* dans les tableaux et les formules) ;
- la moyenne pondérée ordonnée (notée ***OWA*** pour *Ordered Weighted Average*) ;
- l'intégrale de Choquet (notée ***Choquet***).

Le choix de ces trois agrégateurs est argumenté ci-dessous.

6.2.1 Présentation des agrégateurs testés : *WS*, *OWA* et *Choquet*

Deux caractéristiques souhaitées pour la construction des critères sont rapidement apparues au cours de discussions avec des experts.

D'une part, il semble important de conserver la possibilité de tenir compte de l'importance relative des différents sous-critères impliqués dans la construction d'un même critère. Par exemple, en ce qui concerne le Critère *"Alimentation adaptée"*, certains experts considèrent que le Sous-critère 2 *"Absence de soif prolongée"* est plus important que le Sous-critère 1 *"Absence de faim prolongée"*. Dans ce cas, pour refléter par le modèle l'opinion de ces experts, la méthode choisie pour agréger g_1 et g_2 devrait permettre d'attribuer plus de poids à g_2. Pour ce faire, il est possible de recourir à une somme pondérée quasi-arithmétique S de la forme :

$$S(z_1, \ldots, z_q) = \phi^{-1}\left(\sum_{j=1}^{q} w_j\, \phi(z_j)\right) \tag{6.1}$$

où $z_j = g_j(x)$ est le score obtenu par la ferme x sur le Sous-critère j ; où w_j, $i = 1, \ldots, q$,

sont des coefficients positifs sommant à 1 ($\sum_{j=1}^{q} w_j = 1$) ; et où ϕ est une fonction continue et strictement monotone.

Au sein de cette famille d'agrégateurs, on retrouve les sommes pondérées *géométrique* (avec $\phi(x) = \ln(x)$), *harmonique* (avec $\phi(x) = 1/x$), *quadratique* (avec $\phi(x) = x^2$) et bien sûr, la plus simple et la plus commune : la ***somme pondérée*** *arithmétique* (avec $\phi(x) = x$), définie par :

$$WS(z_1, \ldots, z_q) = \sum_{j=1}^{q} w_j \, z_j \tag{6.2}$$

où $z_j = g_j(x)$ est le score obtenu par la ferme x sur le Sous-critère j ; et où w_j, $i = 1, \ldots, q$, sont des coefficients positifs sommant à 1 ($\sum_{j=1}^{q} w_j = 1$).

D'autre part, il est aussi apparu que les experts donnaient plus d'importance aux plus mauvais scores. Ainsi, ils proposent de limiter la possibilité de compenser de mauvais scores sur certains sous-critères par de bons scores sur d'autres. Afin de refléter correctement ce principe lors de l'agrégation, une solution naturelle consiste à focaliser uniquement sur le composant z_k le plus mauvais au sein du vecteur $(z_1, \ldots, z_q)$, puis de considérer les autres composants comme des éléments secondaires. Ceci conduit à utiliser l'opérateur *leximin* (voir par exemple Dubois et al. 2001). L'opérateur leximin repose sur un ordonnancement des fermes comparées (ici les fermes x et y) définit comme suit :

Soient deux fermes $x = (g_1(x), \ldots, g_q(x))$ et $y = (g_1(y), \ldots, g_q(y))$

Soient les vecteurs $\tilde{x}$ et $\tilde{y}$ correspondant respectivement aux vecteurs x et y dont les coordonnées ont été réordonnées par ordre croissant

$$\begin{aligned} x >^{lex} y \quad \text{ssi} \quad & \exists\, k \in 1, \ldots, q \ \text{ tq} \\ & \forall\, l < k, \quad \tilde{x}_l = \tilde{y}_l \ \text{ et } \ \tilde{x}_k > \tilde{y}_k \end{aligned} \tag{6.3}$$

Un des problèmes connus de l'opérateur leximin est sont excès de pessimisme dans l'évaluation. En ne comparant des fermes que sur leur(s) plus mauvaise(s) composante(s), cela masque complètement des profils pouvant être très différents sur les autres composantes. Par exemple si on compare les deux fermes x et y, définies comme suit : $x = (35, 50, 50, 50)$ et $y = (35, 45, 100, 100)$, avec leximin alors $x >^{lex} y$. Les excellents résultats de y sur deux des quatre sous-critères ne sont pas pris en compte puisque sur le deuxième plus mauvais sous-critère y a obtenu un score légèrement inférieur à x (45 *vs* 50).

Une solution alternative permettant d'obtenir un meilleur contrôle de l'impact relatif de bonnes et mauvaises évaluation est l'***OWA*** introduite par Yager (1988) et définie

comme suit :

$$OWA(z_1, ..., z_q) = \sum_{j=1}^{q} \omega_j \, z_{(j)} \tag{6.4}$$

où $\{(1), \ldots, (q)\}$ est une permutation de $\{1, \ldots, q\}$ telle que $z_{(j-1)} \leq z_{(j)}$ pour tout $j = 2, ..., q$. Pour obtenir le comportement souhaité lors de l'agrégation, il suffit d'utiliser des poids définis de telle sorte que plus faible est le score, plus important est le poids qui lui est attribué. De manière formelle, cela donne : $\omega_j > \omega_{j+1},\ j = 1, \ldots, q-1$.

Il est important de signaler que, contrairement à la somme pondérée, l'OWA ne permet pas d'attribuer des poids différents aux sous-critères, ceci étant dû à la propriété de symétrie suivante :

Symétrie : $\forall \sigma$ permutation de $\{1, \ldots, q\},\ OWA(z_1, \ldots, z_q) = OWA(z_{\sigma(1)}, \ldots, z_{\sigma(q)})$

Afin de pouvoir tenir compte des différences d'importance pouvant exister entre les sous-critères, tout en permettant au modèle d'attribuer plus de poids aux moins bons scores, nous avons tout d'abord pensé nous servir d'un opérateur dérivant de l'OWA, définit par Torra (1997) : la WOWA (pour *Weighted OWA*). Cet opérateur repose sur la définition de deux jeux de poids différents (chacun étant représenté par un vecteur de dimensions q) : l'un correspondant aux poids attribués aux q sous-critères ($w = \{w_1, \ldots, w_q\}$) et l'autre correspondant aux poids attribués aux rangs obtenus par les q sous-critères une fois réordonnés de manière croissante ($\omega = \{\omega_1, \ldots, \omega_q\}$). Cet agrégateur se définit formellement comme suit[1] :

$$WOWA(z_1, ..., z_q) = \sum_{j=1}^{q} \left[\varphi\left(\sum_{i=j}^{q} w_{(i)} \right) - \varphi\left(\sum_{i=j+1}^{q} w_{(i)} \right) \right] z_{(j)} \tag{6.5}$$

où $\{(1), \ldots, (q)\}$ est une permutation de $\{1, \ldots, q\}$ telle que $z_{(j-1)} \leq z_{(j)}$ pour tout $j = 2, ..., q$;

et où φ est une fonction monotone croissante telle que : $\begin{cases} \varphi(0) = 0 \\ \varphi\left(\frac{j}{q}\right) = \sum_{i=q-j+1}^{q} \omega_{(i)} \\ \varphi(1) = 1 \end{cases}$

Pour obtenir lors de l'agrégation le comportement que nous souhaitons, il suffit d'utiliser des poids définis de telle sorte que plus faible est le score, plus important est le poids qui lui est attribué. De manière formelle, cela donne : $\omega_j > \omega_{j+1},\ j = 1, \ldots, q-1$.

Malheureusement, le problème de cet agrégateur réside dans la détermination des

[1] adapté d'après Torra (1997) de sorte à utiliser la même permutation que celle définit ci-dessus pour l'OWA

paramètres de pondération. En effet, la révélation conjointe des deux jeux de poids w et ω à partir des réponses des experts sur les jeux de données virtuelles semble délicate.

Une généralisation intéressante ici serait l'***intégrale de Choquet*** (Choquet 1954 ; Grabisch 1996 ; Grabisch et al. 2000). Il s'agit en fait d'un agrégateur plus général qui inclut à la fois la somme pondérée, l'OWA et l'OWA pondérée comme cas particuliers. L'intégrale de Choquet permet de contrôler de manière précise l'importance accordée à chacune des composantes dans l'agrégation tout en permettant de contrôler l'impact sur le score final des bonnes et des mauvaises composantes. Formellement, l'intégrale de Choquet s'écrit :

$$C(z_1, ..., z_q) = \sum_{j=1}^{q} \left[z_{(j)} - z_{(j-1)}\right] v(A_{(j)}) \tag{6.6}$$

$$= \sum_{j=1}^{q} \left[v(A_{(j)}) - v(A_{(j+1)})\right] z_{(j)} \tag{6.7}$$

avec pour convention $z_{(0)} = 0 \leq z_{(1)} \leq \ldots \leq z_{(q)}$, $A_{(j)} = \{(j), ..., (q)\}$ pour $j = 1, \ldots, q$, $A_{(q+1)} = \emptyset$, et v est une fonction de capacité définie pour n'importe quel sous-ensemble de sous-critères entrant dans la composition du critère considéré. Cette capacité est une fonction dont la définition est soumise aux contraintes suivantes :

$$\begin{cases} v(\emptyset) = 0 \\ v(\{1, \ldots, q\}) = 1 \\ A \subseteq B \Rightarrow v(A) \leq v(B) \end{cases} \tag{6.8}$$

À partir de l'Équation (6.7) définissant l'intégrale de Choquet, on peut facilement montrer que la somme pondérée, l'OWA et l'OWA pondérée en sont des cas particuliers :

- si $v(A) = \varphi\left(\sum_{i \in A} w_{(i)}\right)$, l'Équation (6.7) est alors identique à l'Équation (6.5) ;
- si $v(A) = \sum_{i \in A} w_{(i)}$,
  $$\begin{aligned} C(z_1, ..., z_q) &= \textstyle\sum_{j=1}^{q} \left[\left(\sum_{i=j}^{q} w_{(i)}\right) - \left(\sum_{i=j+1}^{q} w_{(i)}\right)\right] z_{(j)} \\ &= \textstyle\sum_{j=1}^{q} w_{(j)}\, z_{(j)} \\ &= \textstyle\sum_{j=1}^{q} w_j\, z_j \\ &= WS(z_1, ..., z_q) \end{aligned}$$
- si v est définie comme étant une capacité symétrique, c.-à-d. ne dépendant que du cardinal de l'ensemble auquel elle s'applique, alors $v(A) = f(|A|)$ et
  $$\begin{aligned} C(z_1, ..., z_q) &= \textstyle\sum_{j=1}^{q} \left[f(|A_{(j)}|) - f(|A_{(j+1)}|)\right] z_{(j)} \\ &= \textstyle\sum_{j=1}^{q} \left[f(q-j+1) - f(q-j)\right] z_{(j)} \end{aligned}$$
  Le terme $[f(q-j+1) - f(q-j)]$ ne dépend que de j, il peut donc être noté plus

simplement ω_j, et alors :

$$\begin{aligned} C(z_1, ..., z_q) &= \textstyle\sum_{j=1}^{q} \omega_j \, z_{(j)} \\ &= OWA(z_1, ..., z_q) \end{aligned}$$

Dans l'Équation (6.6), l'ensemble $A_{(j)}$ correspond à l'ensemble des sous-critères obtenant une note supérieure ou égale à $z_{(j)}$; la valeur $v(A_{(j)})$ représente le poids attaché à la coalition $A_{(j)}$ dans l'agrégation. Le poids attaché à un sous-critère j peut être défini comme la contribution moyenne de j à n'importe quelle coalition A. Ceci est formalisé par la *valeur de Shapley* définie comme suit (Grabisch et Roubens 2000) :

$$\phi_j = \sum_{A \subseteq \{1,...,q\} \setminus \{j\}} \frac{(q - |A| - 1)!|A|!}{q!} \; [v(A \cup \{j\}) - v(A)] \tag{6.9}$$

où q est le cardinal de l'ensemble des sous-critères devant être agrégés et A représente n'importe quel sous-ensemble de sous-critères au sein de cet ensemble.

La capacité v n'est pas forcément additive, c.-à-d. que $v(A \cup B)$ peut être différent de $v(A) + v(B)$ et ce, même quand $A \cap B = \emptyset$. Cette propriété de non-additivité est utile pour refléter les synergies positives ou négatives entre les sous-critères. Le niveau d'interaction entre les sous-critères, tout comme le signe (positif ou négatif) de l'interaction, peut être mesuré grâce à l'*indice d'interaction*, défini ci-dessous pour l'ensemble des paires de sous-critères (Grabisch et Roubens 2000) :

$$I_{i,j} = \sum_{A \subseteq \{1,...,q\} \setminus \{i,j\}} \frac{(q - |A| - 2)!|A|!}{(q-1)!} \; [v(A \cup \{ij\}) - v(A \cup \{i\}) - v(A \cup \{j\}) + v(A)] \tag{6.10}$$

ou plus généralement pour un nombre quelconque de critères par la formule (Marichal 2000) :

$$I_B = \sum_{A \subseteq \{1,...,q\} \setminus B} \frac{(q - |A| - |B|)!|A|!}{(q - |B| + 1)!} \left[\sum_{C \subseteq B} (-1)^{|B|-|C|} v(A \cup C) \right] \qquad \forall B \subseteq \{1, \ldots, q\} \tag{6.11}$$

Le signe de $I_{i,j}$ reflète le type d'interaction entre les sous-critères. Par exemple, quand $I_{i,j} > 0$ il y a une plus-value à être bien noté à la fois sur le Sous-critère i et sur le Sous-critère j. Au contraire, quand $I_{i,j} < 0$ le fait d'être bien noté sur les deux sous-critères est vu comme une redondance (il suffit d'être bien noté sur l'un des deux sous-critères pour être considéré comme étant bon sur l'ensemble des deux sous-critères). Le cas où $I_{i,j} = 0$ correspond à des critères n'interagissant pas. Nous ne détaillons pas ici les caractéristiques de l'intégrale de Choquet, mais pour des informations plus précises

sur l'utilisation de l'intégrale de Choquet dans le cadre de l'agrégation multicritère, il est notamment possible de se référer à Grabisch (1996), Grabisch et Roubens (2000) et Grabisch et al. (2002).

Nous nous sommes basés sur les trois agrégateurs venant d'être décrits (*somme pondérée*, *OWA* et *intégrale de Choquet*) pour construire les critères. Nous avons cependant dû choisir l'agrégateur le mieux adapté à notre situation. C'est ce que la partie suivante met en évidence.

6.2.2 Comparaison des agrégateurs pour la construction des critères et choix de l'agrégateur le mieux adapté à notre situation

Afin d'illustrer la procédure que nous avons choisie pour agréger les scores de sous-critères, nous allons maintenant détailler la construction de quelques critères nous servant d'exemples.

6.2.2.1 Analyse des réponses d'un seul experts sur le cas le plus simple : le Critère f_1 *"Alimentation adaptée"*, composé de deux sous-critères

A l'aide d'un exemple, partons tout d'abord du cas le plus simple : l'analyse des réponses d'un seul expert pour le Critère 1 *"Alimentation adaptée"*. Le jeu de données virtuelles présenté à l'expert ainsi que ses réponses sont présentés dans le Tableau 6.1. À l'aide de ses réponses, notre but ici est d'identifier la fonction critère f_1 permettant d'agréger les deux sous-critères (SC1 *"Absence de faim prolongée"* et SC2 *'Absence de soif prolongée"*) qui composent le Critère 1, en respectant le raisonnement suivi par l'expert lors de ses réponses.

TAB. 6.1 – Exemple de réponses données par un expert pour le Critère 1

Ferme	Scores de sous-critère		**Expert**[a]
x_k	$g_1(x_k)$	$g_2(x_k)$	$y(x_k)$
x_1	25	75	**40**
x_2	40	60	**48**
x_3	50	50	**50**
x_4	60	40	**43**
x_5	75	25	**30**

[a]Note donnée par l'expert pour le Critère 1 à chacune des fermes virtuelles composant le jeu de données

Le jeu de données présenté à l'expert est composé de cinq fermes fictives (x_k, $k = 1, \ldots, 5$) évaluées sur les sous-critères g_1 et g_2. À chaque ferme x_k, l'expert a alors attribué une note globale $y(x_k)$ sensée représentée la valeur de la ferme vis-à-vis du Critère 1.

Afin d'illustrer le raisonnement suivi par l'expert, nous pouvons d'abord regarder les notes qu'il a données aux fermes x_2, x_3 et x_4. En fait, pour un critère composé uniquement de deux sous-critères, un jeu de données contenant trois fermes est suffisant pour en déduire les fonctions de capacité v. Le jeu de données que nous avons utilisé contient, de façon délibérée, des fermes supplémentaires permettant d'introduire des redondances de sorte à pouvoir tester la cohérence des réponses de l'expert, et aussi leur robustesse en cas de variations des écarts d'utilité entre les deux sous-critères (en comparant par exemple les valeurs attribuées aux fermes x_1 et x_2 caractérisées respectivement par les vecteur $(25, 75)$ et $(40, 60)$).

La ferme x_2 se caractérise par un mauvais score sur le Sous-critère g_1 (la faim) mais aussi par un bon score sur le Sous-critère g_2 (la soif). À l'inverse, la ferme x_4 se caractérise par un bon score sur g_1 et un mauvais sur g_2. En proposant un score de critère plus faible à la ferme x_4 (43) qu'à la ferme x_2 (48), l'expert recommande une sanction plus forte quand le score le plus mauvais est obtenu sur g_2. Ceci révèle que pour cet expert le Sous-critère g_2 est plus important que le Sous-critère g_1. Par ailleurs, en attribuant un score de critère de seulement 48 (c.-à-d. inférieur à la moyenne des scores de sous-critères, 50) à la ferme x_2 alors que cette dernière se caractérise par un bon score sur la soif ($g_2(x_2) = 60$), l'expert adopte une attitude sévère qu'il explique par sa volonté d'accorder plus de poids aux scores les moins bons. Un tel expert suit donc simultanément deux principes complémentaires d'agrégation pour évaluer les fermes sur le Critère 1 :

1. il attribue plus d'importance au Sous-citère g_2 qu'au Sous-critère g_1 ;
2. il attribue plus d'importance au moins bon des deux scores obtenus par la ferme.

Par conséquent, en utilisant une somme pondérée ou un OWA nous ne pourrons capter qu'une partie de son raisonnement. Pour obtenir une meilleure approximation des valeurs prescrites par l'expert pour le Critère 1, et ainsi mieux représenter son opinion, nous pouvons nous attendre à ce que l'intégrale de Choquet soit plus appropriée. Cette hypothèse sera d'ailleurs clairement confirmée par les calculs effectués sur les réponses de cet expert.

De manière générale, en se basant sur un échantillon de n_f fermes x_k, $k = 1, \ldots, n_f$, et sur les évaluations $g_1(x_k)$ et $g_2(x_k)$ obtenues par la ferme x_k sur les deux sous-

critères, nous voulons trouver la fonction d'agrégation χ_1 qui minimise δ, la somme des carrés des écarts entre les réponses des experts $y(x_k)$ et les prédictions du modèle $f_1(x_k) = \chi_1(g_1(x_k), g_2(x_k))$:

$$\delta = \sum_{k=1}^{n_f} (f_1(x_k) - y(x_k))^2 \tag{6.12}$$

Nous avons ici comparé les trois options présentées précédemment pour χ_1 :

- WS : $f_1(x_k) = w_1 g_1(x_k) + w_2 g_2(x_k)$
- OWA : $f_1(x_k) = \omega_1 \min\{g_1(x_k), g_2(x_k)\} + \omega_2 \max\{g_1(x_k), g_2(x_k)\}$
- Choquet : $f_1(x_k) = \begin{cases} g_1(x_k) + (g_2(x_k) - g_1(x_k))v_2 & \text{si } g_2(x_k) > g_1(x_k) \\ g_2(x_k) + (g_1(x_k) - g_2(x_k))v_1 & \text{sinon} \end{cases}$

Pour WS, les variables de décision sont w_1 et w_2. Elles sont soumises aux contraintes suivantes : $w_1 \geq 0,\ w_2 \geq 0$ et $w_1 + w_2 = 1$. De même pour OWA, les variables de décision sont ω_1 et ω_2 et elles sont soumises aux contraintes suivantes : $\omega_1 \geq 0,\ \omega_2 \geq 0$ et $\omega_1 + \omega_2 = 1$. Pour l'intégrale de Choquet, nous admettons que $v(\emptyset) = 0$ et $v(\{1,2\}) = 1$ (conditions standard de normalisation pour un opérateur de compromis). Ainsi, seules deux variables restent à déterminer : $v(\{1\})$ et $v(\{2\})$, notées respectivement v_1 et v_2 pour plus de simplicité dans l'écriture. Ces deux variables sont soumises aux conditions suivantes : $0 \leq v_i \leq 1,\ i = 1, 2$, dues aux conditions de monotonicité (cf. Équation (6.8). Nous avons tenu compte de toutes ces contraintes dans le solveur nous ayant permis d'optimiser δ.

En se basant sur l'exemple donné dans le Tableau 6.1, la minimisation de l'erreur quadratique conduit aux paramètres de pondération donnés dans le Tableau 6.3 et fournissant les résultats d'approximation présentés dans le tableau 6.2.

TAB. 6.2 – Comparaison de WS, OWA et Choquet pour le Critère 1

Ferme	Scores de sous-critère		**Expert**[a]	Score de critère f_1		
x_k	g_1	g_2	y	WS	OWA	Choquet
x_1	25	75	**40**	55	36	40
x_2	40	60	**48**	52	44	46
x_3	50	50	**50**	50	50	50
x_4	60	40	**43**	48	44	42
x_5	75	25	**30**	45	36	31
			δ	*491*	*66*	*5*

[a]Note donnée par l'expert pour le Critère 1 à chacune des fermes virtuelles composant le jeu de données

TAB. 6.3 – Paramètres définis pour l'agrégation des deux sous-critères composant le Critère 1 avec un expert

WS		OWA		Choquet			
w_1	0.38	ω_1	0.72	v_1	0.11	ϕ_1	0.40
w_2	0.62	ω_2	0.28	v_2	0.31	ϕ_2	0.60
						$I_{1,2}$	0.58

En comparant les valeurs de δ obtenues respectivement avec la somme pondérée, l'OWA et l'intégrale de Choquet (dernière ligne du Tableau 6.2), nous constatons que l'*intégrale de Choquet* fournit le meilleur résultat, avec une erreur très faible ($\delta = 5$), et que la *somme pondérée* donne clairement le plus mauvais résultat ($\delta = 491$). L'*OWA*, quant à elle, fournit un résultat correct mais restant moins bon que celui obtenu avec Choquet.

En regardant de plus près le modèle impliquant l'intégrale de Choquet, les valeurs de Shapley obtenues (0.6 pour la soif *vs* 0.4 pour la faim, cf. Table 6.3) montrent que nous attribuons plus de poids à la soif (g_2) qu'à la faim (g_1). Les valeurs ϕ_1 et ϕ_2 (respectivement 0.40 et 0.60) sont d'ailleurs très proches des poids w_1 et w_2 obtenus avec la somme pondérée (respectivement 0.38 et 0.62). Par ailleurs, nous obtenons un indice d'interaction positif et élevé ($I_{12} = 0.58$, cf. Table 6.3), ce qui indique que dans le modèle pour qu'une ferme soit considérée comme bonne elle doit l'être sur les deux sous-critères, ce qui revient à attribuer plus d'importance au moins bon des deux scores. Cet indice d'interaction élevé est en accord avec les résultats du modèle utilisant l'OWA, pour lequel nous obtenons des poids ω_1 et ω_2 de respectivement 0.72 et 0.28. L'intégrale de Choquet permet donc de rendre très bien compte du raisonnement suivi par l'expert (cf. plus haut). De plus, les conclusions que nous avons tirées des réponses de l'expert sont aussi en accord avec les explications qu'il nous a laissées pour justifier ses évaluations.

6.2.2.2 Analyse des réponses d'un seul experts sur le Critère f_2 *"Logement correct"*, composé de trois sous-critères

Prenons maintenant l'exemple du Critère 2 (f_2), composé de trois sous-critères (g_3, g_4 et g_5). Nous avons procédé de la même manière que pour le Critère 2, en commençant par faire réagir un expert sur un jeu de données virtuelles qui, cette fois-ci, est composé de 13 fermes virtuelles (cf. Tableau 6.4). Ce sont les réponses de cet expert qui nous ont permis de déterminer les paramètres de pondération par minimisation de la somme des

carrés des écarts δ définie comme suit :

$$\delta = \sum_{k=1}^{n_f} (f_2(x_k) - y(x_k))^2 \tag{6.13}$$

avec $n_f = 13$ fermes. L'optimisation de cette fonction nous a permis de déterminer les trois paramètres nécessaires quand f_2 correspond à une somme pondérée ou à une OWA, et les six valeurs de capacité de l'intégrale de Choquet (cf. Tableau 6.5).

TAB. 6.4 – Jeu de données virtuelles pour le Critère 2 *"Logement correct"* et exemple de réponses données par un expert

Ferme	Scores de sous-critère			**Expert**[a]	Score de critère f_2		
x_k	g_3	g_4	g_5	y	WS	OWA	Choquet
x_1	25	50	75	**41**	52	40	41
x_2	40	50	60	**47**	51	46	46
x_3	50	50	50	**50**	50	50	50
x_4	25	75	50	**40**	48	40	40
x_5	40	60	50	**46**	49	46	47
x_6	50	25	75	**43**	54	40	44
x_7	50	40	60	**52**	52	46	51
x_8	50	60	40	**44**	48	40	44
x_9	50	75	25	**35**	46	46	34
x_{10}	60	40	50	**51**	51	46	49
x_{11}	75	25	50	**42**	52	40	42
x_{12}	60	50	40	**45**	49	46	47
x_{13}	75	50	25	**38**	48	40	38
				δ	*688*	*109*	*14*

[a]Note donnée par l'expert pour le Critère 2 à chacune des fermes virtuelles composant le jeu de données

Nous pouvons constater que l'optimisation de δ en utilisant un solveur (sous les contraintes explicitées précédemment) fournit de très bonnes estimations des valeurs prescrites par l'expert quand on utilise l'intégrale de Choquet. En effet, la somme des carrés des écarts obtenue avec Choquet ($\delta = 14$) est inférieure à celle obtenue avec l'OWA ($\delta = 109$), cette dernière étant elle-même très nettement inférieure à celle obtenue avec WS ($\delta = 688$). Le fait que l'intégrale de Choquet fournisse de meilleurs résultats n'est pas surprenant, par contre, il est intéressant de noter que dans le cas de cet expert pour le Critère 2 l'intégrale de Choquet apparaît nettement comme l'agrégateur le plus

approprié (la différence de δ entre Choquet et OWA étant significative).

Les paramètres de pondération ainsi obtenus sont présentés dans le Taleau 6.5.

TAB. 6.5 – Paramètres définis pour l'agrégation des trois sous-critères composant le Critère 2 avec un expert

WS		OWA		Choquet			
w_3	0.33	ω_1	0.59	v_3	0.07	ϕ_3	0.29
w_4	0.25	ω_2	0.20	v_4	0.02	ϕ_4	0.29
w_5	0.42	ω_3	0.20	v_5	0.10	ϕ_5	0.42
				$v_{3,4}$	0.31	$I_{3,4}$	0.15
				$v_{3,5}$	0.48	$I_{3,5}$	0.24
				$v_{4,5}$	0.54	$I_{4,5}$	0.35
						$I_{3,4,5}$	-0.15

En utilisant l'intégrale de Choquet pour modéliser le raisonnement suivi par l'expert, le calcul des valeurs de Shapley ϕ_i (cf. Équation 6.9) et des indices d'interaction $I_{i,j}$ (cf. Équation 6.10) nous permet d'interpréter la logique d'agrégation de l'expert (cf. Tableau 6.5). Nous pouvons tout d'abord noter que $\phi_4 \simeq \phi_3 < \phi_5$ ce qui n'est pas très éloigné de ce que l'expert nous a expliqué vouloir faire : *"le SC5 est le plus important, suivi du SC3, le SC4 est presque accessoire pour les vaches laitières"*. De plus, nous constatons que les indices d'interaction sont positifs et non négligeables (avec des valeurs allant de 0.15 à 0.35 pour les indices d'interaction entre deux sous-critères), sauf pour l'indice d'interaction entre les trois sous-critères qui est négatif (il n'y a pas de moins-value supplémentaire à être mauvais sur les trois sous-critères au lieu de l'être sur seulement deux d'entre eux). Ceci est aussi en accord avec le souhait de l'expert de limiter les compensations entre les mauvais et les bons scores de sous-critère, les situations à favoriser étant celles qui réalisent un compromis équilibré entre les bons et les mauvais scores. Par conséquent nous pouvons dire que l'intégrale de Choquet nous permet de refléter correctement les attentes de l'expert, et nous fournit une approximation nettement meilleure que celle permise par la somme pondérée ou l'OWA.

6.2.2.3 Généralisation à l'analyse des réponses des 14 experts sur le Critère f_2 *"Logement correct"*

Afin d'obtenir une construction des critères plus fiable, nous avons consulté plusieurs experts et non un seul. En fait, 14 experts ont répondu à nos questions (via les jeux de données virtuelles) pour les vaches laitières en ferme. Leurs réponses sont présentées dans le Tableau 6.6.

En généralisant l'approche que nous avons adoptée dans le cas où un seul expert était pris en compte, nous avons pu définir les paramètres de l'agrégateur (WS, OWA ou Choquet) utilisé dans le modèle considéré, là encore par minimisation de l'erreur globale δ définie comme suit :

$$\delta = \sum_{p=1}^{n_e} w_p \left(\sum_{k=1}^{n_f} (f_2(x_k) - y_p(x_k))^2 \right) \tag{6.14}$$

où n_e est le nombre d'experts, n_f est le nombre de fermes de référence présentées dans le jeu de données, $y_p(x_k)$ correspond à l'évaluation donnée par l'expert p à la ferme x_k pour le Critère 2, $f_2(x_k)$ est le score calculé sur la ferme x_k par le modèle et w_p est le poids attribué à l'expert p. Pour le moment, nous avons décidé de donner la même importance à tous les experts ($\forall\ p = 1, \ldots, n_e, w_p = 1$) mais l'introduction de poids différents en fonction des experts pourrait s'avérer utile pour tenir compte dans la procédure d'évaluation de la pertinence et du niveau d'expertise des personnes consultées, ou encore du niveau de confiance que les experts ont dans leurs réponses.

Comme nous pouvons le voir dans le Tableau 6.6 en comparant les évaluations moyennes données par les experts (colonne $\bar{y}$) aux scores f_2 calculés en utilisant l'intégrale de Choquet (colonne C), ce modèle permet de très bien approximer les évaluations moyennes des experts. Ceci est confirmé par la faible valeur de δ obtenue avec Choquet. Nous pouvons noter que l'approximation réalisée avec Choquet est légèrement meilleure que celle réalisée en utilisant une OWA et apporte une amélioration très nette par rapport à l'utilisation d'une somme pondérée (avec respectivement des valeurs de $\delta/nombre\ d'experts$ égales à 590, 598 et 2862, cf. la dernière ligne du Tableau 6.6).

TAB. 6.6 – Évaluations données par les experts et calcul du score pour le Critère 2 *"Logement correct"*

k	g_3	g_4	g_5	***Réponses des experts***[a]															**Score de critère f_2**		
				y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	$\bar{y}$	WS	OWA	**C**
1	**25**	**50**	**75**	20	35	35	25	25	25	50	54	35	25	25	25	25	41	**32**	50	32	**31**
2	**40**	**50**	**60**	30	45	45	40	40	50	50	52	50	40	40	40	40	47	**44**	50	43	**42**
3	**50**	**50**	**50**	40	50	50	50	50	50	50	50	50	50	50	50	50	50	**49**	50	50	**50**
4	**25**	**75**	**50**	20	35	35	25	25	25	45	58	35	25	25	25	25	40	**32**	51	32	**33**
5	**40**	**60**	**50**	30	45	45	40	40	50	50	53	50	40	40	40	40	46	**44**	50	43	**44**
6	**50**	**25**	**75**	20	35	35	25	25	27	40	46	35	25	25	25	25	43	**31**	49	32	**32**
7	**50**	**40**	**60**	40	45	45	40	40	50	50	48	50	40	40	40	40	52	**44**	50	43	**44**
8	**50**	**60**	**40**	50	45	45	40	40	50	50	52	50	50	50	50	50	44	**45**	50	43	**44**
9	**50**	**75**	**25**	40	35	35	25	25	25	45	54	35	25	25	25	25	35	**32**	51	32	**32**
10	**60**	**40**	**50**	40	45	45	40	40	50	40	47	50	40	40	40	40	51	**43**	50	43	**43**
11	**75**	**25**	**50**	40	35	35	25	25	27	30	42	35	25	25	25	25	42	**31**	49	32	**31**
12	**60**	**50**	**40**	40	45	45	40	40	50	50	48	50	40	40	40	40	45	**44**	50	43	**44**
13	**75**	**50**	**25**	40	35	35	25	25	25	30	46	35	25	25	25	25	38	**31**	50	32	**32**
															δ / nombre d'experts				*2862*	*598*	*590*

[a]Notes données par les experts pour le Critère 2 aux fermes virtuelles composant le jeu de données

Les valeurs des paramètres de pondération déterminées à l'aide du solveur sont présentées dans le Tableau 6.7. Sur l'ensemble de 14 experts consultés, aucun des trois sous-critères sur lesquels se base la construction du Critère 2 ne semble être plus important qu'un autre (les valeurs de Shapley sont du même ordre de grandeur pour les trois sous-critères, entre 0.32 et 0.35). Par contre, les indices d'interaction sont très élevés (entre 0.36 et 0.56 en fonction de la coalition de sous-critères considérée). Considérés ensemble, les experts semblent donc vouloir limiter fortement les compensations entre les sous-critères en mettant beaucoup plus de poids sur les plus mauvais scores. Si on regarde les résultats obtenus avec l'OWA, nous constatons que cet agrégateur donne des résultats presque aussi bons que ceux obtenus avec Choquet en attribuant énormément d'importance au plus mauvais score ($\omega_1 = 0.72$) et un poids plus modéré au score intermédiaire ($\omega_2 = 0.28$), le meilleur score, quant à lui, n'est même pas pris en compte dans le calcul ($\omega_3 = 0$).

TAB. 6.7 – Paramètres définis pour l'agrégation des trois sous-critères composant le Critère 2 avec les 14 experts

WS		OWA		Choquet			
w_3	0.33	ω_1	0.72	v_3	0.03	ϕ_3	0.32
w_4	0.35	ω_2	0.28	v_4	0.03	ϕ_4	0.35
w_5	0.33	ω_3	0.00	v_5	0.04	ϕ_5	0.33
				$v_{3,4}$	0.19	$I_{3,4}$	0.41
				$v_{3,5}$	0.14	$I_{3,5}$	0.36
				$v_{4,5}$	0.20	$I_{4,5}$	0.41
						$I_{3,4,5}$	0.56

6.2.2.4 Généralisation à l'ensemble des quatre critères : l'intégrale de Choquet apparaît comme étant l'agrégateur le mieux adapté

Généralisation et résultats obtenus pour les quatre critères

Nous avons traité de la même façon tous les critères, qu'ils soient composés de deux sous-critères (f_1), de trois sous-critères (f_2 et f_3) ou de quatre sous-critères (f_4). Pour chaque critère nous avons comparé les trois agrégateurs listés préalablement : la somme pondérée, l'OWA et l'intégrale de Choquet, en se basant à chaque fois sur les réponses des 14 experts. Le Tableau 6.8 présente la moyenne par expert de la somme des carrés des écarts ($= \delta/nombre\ d'experts$) obtenue pour chacun des quatre critères. Nous pouvons ainsi constater que dans tous les cas la somme pondérée apparaît clairement comme étant la solution la moins efficace. Au contraire, l'intégrale de Choquet fournit

systématiquement la meilleure approximation de l'avis des experts, même si l'OWA semble être aussi un bon choix d'agrégateur pour construire les critères. Il existe toujours en effet une différence en faveur de Choquet entre les valeurs du δ par expert, mais cette différence est faible (de 8 à 35 en fonction du critère). Le fait que l'OWA et l'intégrale de Choquet soient presque aussi efficaces l'une que l'autre s'explique par le nombre important d'experts ayant décidé d'interdire les compensations en donnant un score de critère égal au plus mauvais des scores de sous-critères (c.-à-d. en choisissant comme agrégateur la fonction min). Ce comportement correspond à un cas particulier d'OWA dans laquelle les paramètres sont les suivants : $\omega_1 = 1$ et pour $i \neq 1$ $\omega_i = 0$. Pour tous ces experts, l'OWA donne le même résultat que l'intégrale de Choquet, avec une somme des carrés des écarts nulle. Les experts ayant adopté un raisonnement moins drastique ont cependant presque tous voulu limiter les compensations. Ainsi, sur l'ensemble des experts il semble normal que l'OWA donne des résultats satisfaisants.

TAB. 6.8 – Qualité de l'approximation obtenue avec le somme pondérée, l'OWA et l'intégrale de Choquet

	$\sum$(Erreur^2) / nombre d'experts		
	WS	OWA	**Choquet**
Critère 1	1256	324	**291**
Critère 2	2862	598	**590**
Critère 3	2933	566	**531**
Critère 4	2075	736	**722**

Les paramètres de pondération établis pour les quatre critères et les trois agrégateurs différents sont présentés dans le Tableau 6.9.

Malgré les faibles écarts d'importance accordés aux différents sous-critères, on peut tout de même noter la bonne adéquation entre les valeurs de Shapley obtenues avec l'intégrale de Choquet et les poids obtenus avec la somme pondérée. Par exemple, pour le Critère 4, nous obtenons avec WS les poids suivants : $w_9 = 0.24$, $w_{10} = 0.22$, $w_{11} = 0.25$ et $w_{12} = 0.30$; alors qu'avec l'intégrale de Choquet, les valeurs de Shapley associées aux quatre sous-critères sont les suivantes : $\phi_9 = 0.24$, $\phi_{10} = 0.21$, $\phi_{11} = 0.25$ et $\phi_{12} = 0.30$.

TAB. 6.9 – Paramètres définis pour l'agrégation des trois sous-critères composant le Critère 2 avec les 14 experts

Critère	WS		OWA		Choquet			
C1	w_1	0.43	ω_1	0.91	v_1	0.02	ϕ_1	0.42
	w_2	0.57	ω_2	0.09	v_2	0.17	ϕ_2	0.58
							$I_{1,2}$	0.82
C2	w_3	0.33	ω_1	0.72	v_3	0.03	ϕ_3	0.32
	w_4	0.35	ω_2	0.28	v_4	0.03	ϕ_4	0.35
	w_5	0.33	ω_3	0.00	v_5	0.04	ϕ_5	0.33
					$v_{3,4}$	0.19	$I_{3,4}$	0.41
					$v_{3,5}$	0.14	$I_{3,5}$	0.36
					$v_{4,5}$	0.20	$I_{4,5}$	0.41
							$I_{3,4,5}$	0.56
C3	w_6	0.33	ω_1	0.83	v_6	0.05	ϕ_6	0.33
	w_7	0.34	ω_2	0.09	v_7	0.05	ϕ_7	0.35
	w_8	0.34	ω_3	0.09	v_8	0.06	ϕ_8	0.33
					$v_{6,7}$	0.11	$I_{6,7}$	0.46
					$v_{6,8}$	0.06	$I_{6,8}$	0.39
					$v_{7,8}$	0.10	$I_{7,8}$	0.43
							$I_{6,7,8}$	0.88
C4	w_9	0.24	ω_1	0.58	v_9	0.11	ϕ_9	0.24
	w_{10}	0.22	ω_2	0.27	v_{10}	0.07	ϕ_{10}	0.21
	w_{11}	0.25	ω_3	0.07	v_{11}	0.11	ϕ_{11}	0.25
	w_{12}	0.30	ω_4	0.07	v_{12}	0.14	ϕ_{12}	0.30
					$v_{9,10}$	0.11	$I_{9,10}$	0.10
					$v_{9,11}$	0.12	$I_{9,11}$	0.09
					$v_{9,12}$	0.17	$I_{9,12}$	0.11
					$v_{10,11}$	0.12	$I_{10,11}$	0.10
					$v_{10,12}$	0.17	$I_{10,12}$	0.13
					$v_{11,12}$	0.20	$I_{11,12}$	0.10
					$v_{9,10,11}$	0.33	$I_{9,10,11}$	0.22
					$v_{9,10,12}$	0.39	$I_{9,10,12}$	0.21
					$v_{9,11,12}$	0.44	$I_{9,11,12}$	0.27
					$v_{10,11,12}$	0.40	$I_{10,11,12}$	0.18
							$I_{9,10,11,12}$	-0.10

Optimisation du nombre de paramètres dans l'intégrale de Choquet

Il apparaît donc que l'intégrale de Choquet est l'agrégateur le mieux adapté à la construction de nos critères. Cependant, le nombre de paramètres à déterminer croît de façon exponentielle avec le nombre de sous-critères, avec un total de $2^q - 2$ paramètres pour l'agrégation de q sous-critères. Par exemple, nous avons besoin d'apprendre (grâce à un solveur) six coefficients pour définir une fonction critère reposant sur trois sous-critères (ce qui est le cas des critères f_2 et f_3), tandis que 14 coefficients sont nécessaires pour f_4 puisqu'il se base sur quatre sous-critères. Ainsi, en utilisant l'intégrale de Choquet nous obtenons une meilleure approximation qu'avec l'OWA ou la somme pondérée, mais pour ce faire nous devons définir un nombre de paramètres plus important (cf. Tableau 6.10). La différence est notamment visible pour le Critère 4 *"Comportement approprié"* pour lequel nous devons déterminer 14 paramètres avec Choquet contre seulement 4 avec l'OWA.

TAB. 6.10 – Nombre de paramètres devant être déterminés en fonction de l'agrégateur utilisé

Critère	**Nombre de paramètres**			
	WS	OWA	Choquet	*Choquet 2-additive*
C1	2	2	2	*2*
C2	3	3	6	*6*
C3	3	3	6	*6*
C4	4	4	14	*10*

De manière générale, pour réduire le nombre de paramètres à déterminer pour l'intégrale de Choquet, Grabisch et al. (Accepté) proposent d'utiliser une *capacité k-additive* ($k < q$) qui ne nécessite que $\sum_{l=1}^{k} C_q^l = \sum_{l=1}^{k} \frac{q!}{l!(q-l)!}$ coefficients (une valeur de capacité pour chaque ensemble de k ou moins de k sous-critères).

Afin d'illustrer le potentiel des capacités k-additives, prenons le cas où nous souhaitons définir un critère à partir d'un ensemble $N = \{1, \ldots, q\}$ de sous-critères à agréger avec l'intégrale de Choquet (Équation (6.6)), en utilisant une capacité v. Nous pouvons alors considérer la *représentation de Möbius* de v définie par :

$$m_v(B) = \sum_{A \subseteq B} (-1)^{b-a} v(A), \qquad \forall\, B \subseteq N \tag{6.15}$$

où $a = |A|$ et $b = |B|$. Ainsi, la capacité v peut s'exprimer en utilisant la *représentation de Möbius* par :

$$v(C) = \sum_{B \subseteq C} m_v(B), \qquad \forall\, C \subseteq N \tag{6.16}$$

La capacité v est dite *k-additive* si sa représentation de Möbius satisfait $m_v(B) = 0$ pour tout $B \subseteq N$ tel que $b > k$ et s'il existe au moins un sous-ensemble B de cardinalité k tel que $m_v(B) \neq 0$. Pour différencier par l'écriture la capacité générale de la capacité k-additive, cette dernière est notée v_k et, d'aprés l'Équation (6.16) s'écrit :

$$v_k(C) = \sum_{\substack{B \subseteq C \\ b \leq k}} m_v(B), \qquad \forall\, C \subseteq N \tag{6.17}$$

Une **capacité 1-additive** correspond à une capacité purement additive. En effet, d'après la définition venant d'être donnée une capacité 1-additive se caractérise par $m_v(B) = 0$ pour tout $B \subseteq N$ tel que $b > 1$ ce qui, d'après les Équations (6.15) et (6.17), donne :

$$v_1(B) = \sum_{\substack{A \subseteq B \\ a = 1}} v_1(A) \tag{6.18}$$

Les indices d'interactions sont par conséquent tous nuls dans le cas d'une capacité 1-additive et l'intégrale de Choquet correspond alors à une simple somme pondérée. Plus généralement, une capacité k-additive ne peut modéliser des interactions qu'entre k composantes au maximum (Grabisch et al. Accepté).

Comme nous avons pu le voir précédemment, l'utilisation d'une somme pondérée (ce qui revient à utiliser l'intégrale de Choquet avec une capacité 1-additive) ne donne pas des résultats satisfaisants, mais qu'en est-il de l'utilisation de l'intégrale de Choquet avec une **capacité 2-additive** ? L'utilisation d'un capacité 2-additive au lieu de la capacité générale permet de réduire significativement le nombre de paramètres à déterminés, passant de $2^q - 2$ à seulement $q(q+1)/2$ (une valeur de capacité pour chaque paire de sous-critères et chaque singleton). Dans le cas du Critère 4, cela revient à déterminer 10 coefficients au lieu des 14 nécessaires avec une capacité générale. Il est cependant nécessaire de vérifier que le modèle reste suffisamment souple pour pouvoir approximer correctement les avis des experts.

En conséquence de l'Équation (6.17), si v est 2-additive alors les coefficients correspondant aux coalitions de trois sous-critères $\{i, j, k\}$ s'écrivent comme suit :

$$v(ijk) = m_v(ij) + m_v(jk) + m_v(ik) + m_v(i) + m_v(j) + m_v(k)$$

Puisque $m_v(ij) = v(ij) - v(i) - v(j)$ et $m_v(i) = v(i)$ du fait de l'Équation (6.15), $v(ijk)$ peut facilement être dérivée des capacités des sous-ensembles contenant un ou deux

sous-critères :

$$v(ijk) = v(ij) + v(jk) + v(ik) - v(i) - v(j) - v(k) \tag{6.19}$$

Ainsi, en utilisant une capacité 2-additive pour l'intégrale de Choquet qui sert à la construction de f_4 à partir des sous-critères g_9, g_{10}, g_{11} et g_{12}, nous avons la possibilité de réduire le nombre de coefficients devant être appris de 14 à seulement 10. Cette idée est particulièrement intéressante dans le cas où le postulat de 2-additivité ne dégrade pas significativement la qualité de l'approximation des réponses des experts par l'intégrale de Choquet. Afin de tester ce point sur nos données, nous avons comparer les δ obtenus en utilisant l'intégrale de Choquet, avec ou sans la contrainte de 2-additivité, à ceux obtenus avec l'OWA et la somme pondérée. Sans surprise, l'erreur minimale est obtenue pour l'intégrale de Choquet avec une capacité générale ($\delta = 722$). Il est par contre intéressant de constater que l'ajout de la contrainte de 2-additivité est très pertinent puisque qu'on réduit de façon notable le nombre de paramètres tout en conservant un résultat proche de l'optimalité ($\delta = 726$, soit une erreur de seulement 4 points supérieure à celle obtenue avec la capacité générale). L'erreur ainsi obtenue reste inférieure à celle obtenue avec l'OWA ($\delta = 736$) et est bien sûr nettement inférieure à celle obtenue avec la somme pondérée ($\delta = 2075$), qui correspond à une capacité 1-additive.

En conclusion, l'utilisation d'une capacité 2-additive pour le calcul de l'intégrale de Choquet apparaît donc comme un excellent compromis entre la souplesse de notre modèle et sa complexité, ce qui est en parfait accord avec Grabisch (2006). Grabisch (2006) justifie de plus le choix d'une capacité 2-additive par le fait qu'il est de toute façon difficile pour un décideur d'appréhender la signification d'interactions à plus de deux critères.

Nous avons donc décidé d'utiliser l'intégrale de Choquet avec une capacité 2-additive pour la construction du Critère 4 résultant de l'agrégation de quatre sous-critères. Les paramètres déterminés dans le cas d'une capacité générale[2] et d'une capacité 2-additive sont présentés dans le Tableau 6.11.

Les valeurs de Shapley obtenues avec la capacité générale et avec la capacité 2-additive sont presque identiques. Il est par contre impossible de comparer les indices d'interactions obtenus pour chaque paire de sous-critères avec la capacité générale et la capacité 2-additive. En effet, dans la capacité générale il y a en plus des interactions obtenues sur les paires de sous-critères, des interactions au niveau de chaque triplet de sous-critères et une interaction entre les quatre sous-critères.

[2]résultats déjà présentés dans le Tableau 6.9, mais de nouveau rappelés ici pour faciliter la comparaison

TAB. 6.11 – Comparaison des valeurs des capacités générale et 2-additive de l'intégrale de Choquet, des valeurs de Shapley et des indices d'interaction obtenus lors de la construction du Critère 4 avec les 14 experts

	Capacité générale	Capacité 2-additive		Capacité générale	Capacité 2-additive
v_9	0.11	0.00	ϕ_9	0.24	0.24
v_{10}	0.07	0.00	ϕ_{10}	0.21	0.22
v_{11}	0.11	0.00	ϕ_{11}	0.25	0.25
v_{12}	0.14	0.00	ϕ_{12}	0.30	0.29
$v_{9,10}$	0.11	0.14	$I_{9,10}$	0.10	0.11
$v_{9,11}$	0.12	0.15	$I_{9,11}$	0.09	0.13
$v_{9,12}$	0.17	0.19	$I_{9,12}$	0.11	0.16
$v_{10,11}$	0.12	0.14	$I_{10,11}$	0.10	0.11
$v_{10,12}$	0.17	0.17	$I_{10,12}$	0.13	0.17
$v_{11,12}$	0.20	0.21	$I_{11,12}$	0.10	0.18
$v_{9,10,11}$	0.33	(0.41)	$I_{9,10,11}$	0,22	0
$v_{9,10,12}$	0.39	(0.48)	$I_{9,10,12}$	0,21	0
$v_{9,11,12}$	0.44	(0.54)	$I_{9,11,12}$	0,27	0
$v_{10,11,12}$	0.40	(0.50)	$I_{10,11,12}$	0,18	0
			$I_{9,10,11,12}$	-0,10	0

Le passage d'une capacité générale à une capacité 2-additive sur le Critère 4 ne dégrade que très peu la qualité de l'approximation tout en permettant de réduire le nombre de paramètres et en s'interprétant de la même façon en termes d'importance relative des sous-critères.

Distribution des 91 fermes observées lors du WP2.4

En conclusion, l'intégrale de Choquet (avec pour f_4 une capacité 2-additive) semble dans tous les cas être l'agrégateur le mieux adapté pour construire les critères à partir des différents sous-critères qui les composent.

En utilisant cet agrégateur pour traiter les données recueillies dans le WP2.4 sur 91 fermes, nous obtenons, critère par critère, les distributions présentées dans les Figures 6.2 et 6.3.

On peut cependant se demander s'il est vraiment nécessaire d'avoir recours à un agrégateur aussi compliqué alors que l'OWA fournit des résultats qui restent tout à fait satisfaisants.

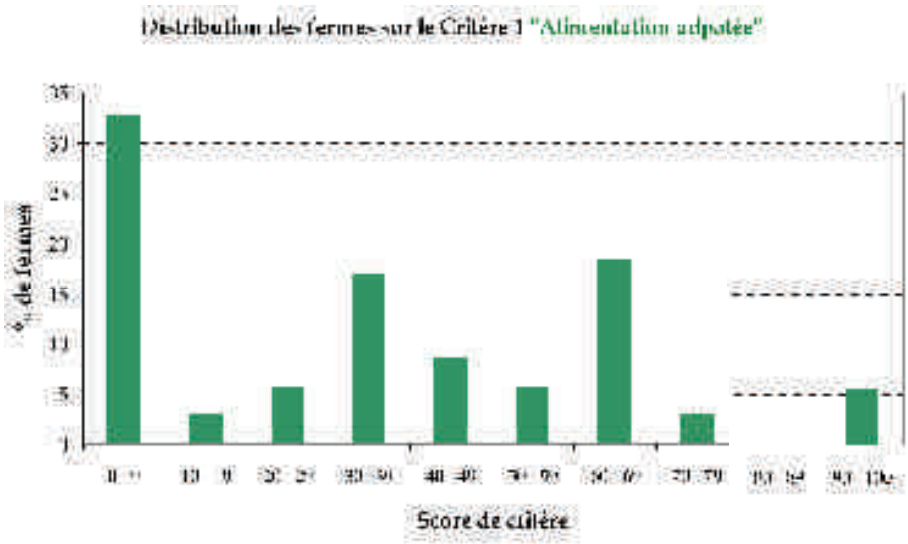
Comme nous l'avons dit précédemment, les bons résultats obtenus avec l'OWA s'expliquent par la grande sévérité des experts qui, à la quasi-unanimité, ont souhaité limi-

ter très fortement les compensations entre les sous-critères composant un même critère (voire même pour certains les interdire). Nous avions déjà constaté la sévérité des personnes consultées lors de la construction des sous-critères. Le fait d'être autant sévère au niveau des critères n'améliore pas la distribution des 91 fermes visitées par le WP2.4. En effet, nous constatons par exemple qu'aucune ferme ne dépasse le score de 50 sur le critère 3 (cf. Figure 6.3) et que près de 33% des fermes obtiennent un score inférieur à 10 sur le Critère 1 (cf. Figure 6.2). Nous avons déjà prévu de recontacter les experts pour la construction des sous-critères, mais il semble aussi qu'il faille revenir auprès des 14 personnes impliquées dans la construction des critères en leur demandant s'il est nécessaire pour eux de suivre un raisonnement aussi drastique. Dans ce cas, il est possible que les différences d'importance entre les sous-critères puissent davantage s'exprimer. Avec les paramètres actuels, c'est principalement les poids liés aux scores qui influent sur les scores de critères (d'où la bonne adéquation entre les résultats de l'OWA et ceux de l'intégrale de Choquet).

Par ailleurs, nous souhaitons, par souci d'homogénéité, utiliser le même agrégateur pour la construction des critères des différents couple (*types d'animaux*, *période de vie*[3]). Pour l'instant nous n'avons traité que les réponses des experts pour les vaches laitières en ferme. Il est possible que dans les autres cas l'OWA fournisse des résultats moins performants que ceux obtenus avec l'intégrale de Choquet. Par exemple, si pour les porcs à l'engraissement la soif apparaît comme étant beaucoup plus importante que la faim au sein du Critère 1, l'écart entre l'OWA et l'intégrale de Choquet se creusera et il sera alors nettement préférable d'utiliser le second agrégateur.

Pour ces deux raisons il semble donc beaucoup plus prudent d'utiliser l'agrégateur le plus généraliste (l'intégrale de Choquet) qui permet de refléter toutes les situations possibles. De plus, en utilisant des capacités 2-additives on a pu montrer que les résultats obtenus pour le Critère 4 (le seul à nécessiter cette modification) restent excellents tout en permettant de réduire significativement le nombre de paramètres à définir. Nous avons décidé d'utiliser l'intégrale de Choquet pour la construction de l'ensemble des critères, qu'ils concernent les vaches laitières en ferme ou n'importe quel autre type d'animaux.

[3]en ferme ou durant la phase de transport/abattage



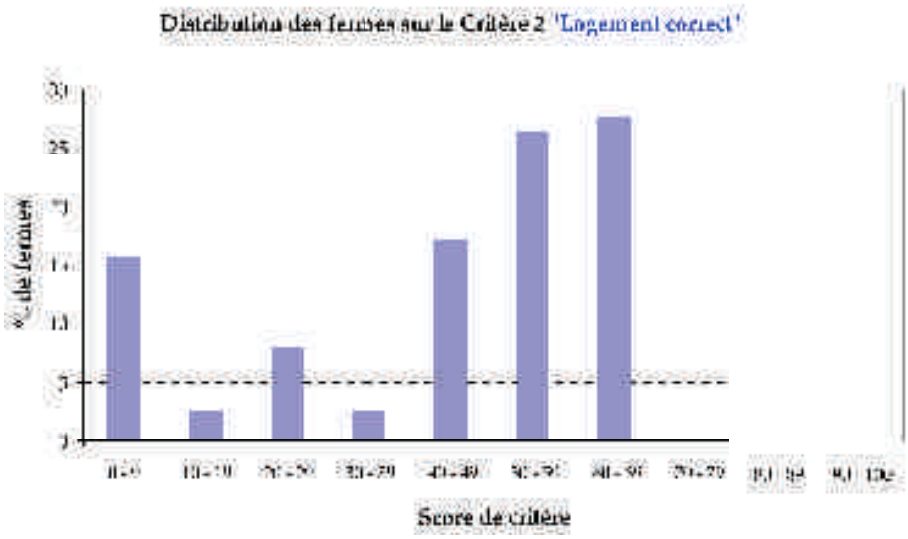


FIG. 6.2 – Distribution des fermes sur les Critères 1 *"Alimentation adaptée"* et 2 *"Logement correct"*

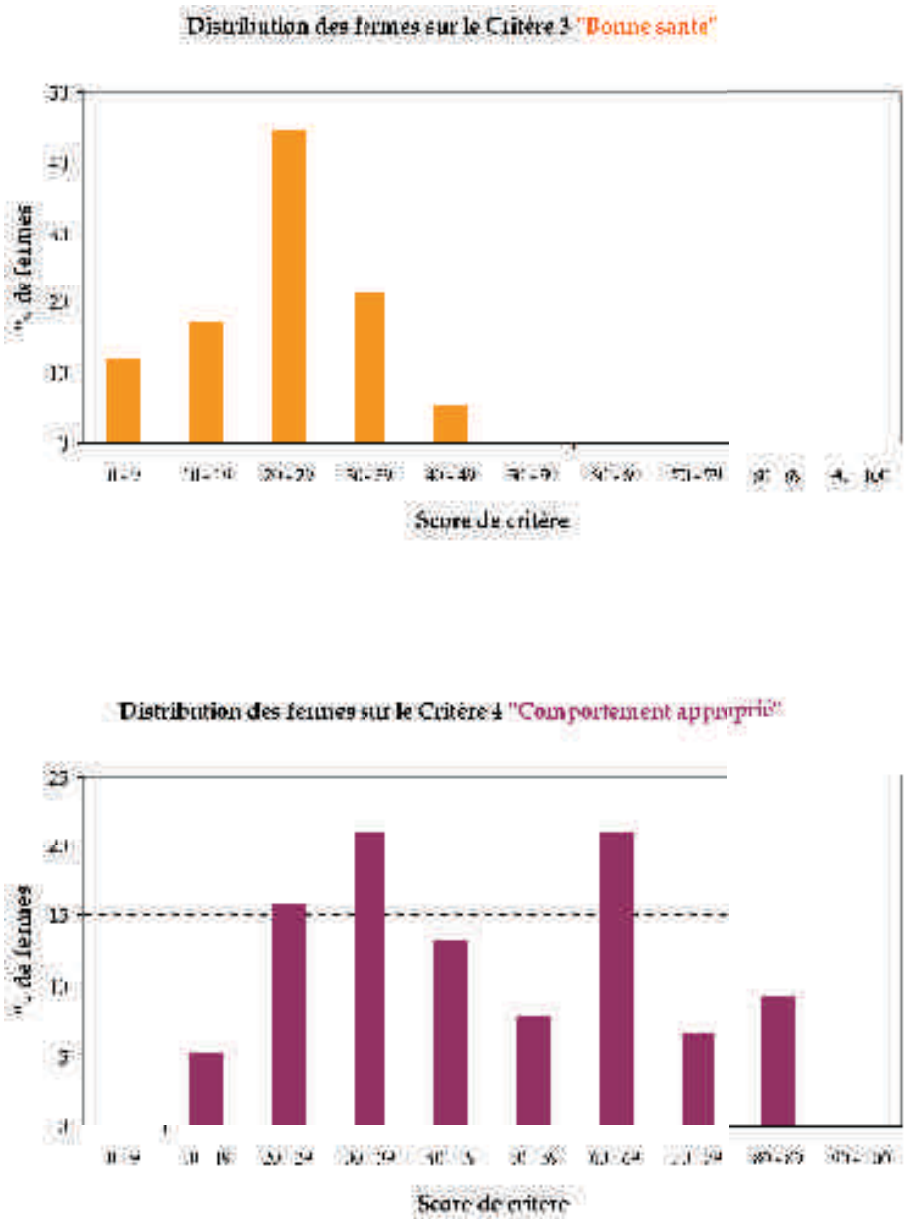


FIG. 6.3 – Distribution des fermes sur les Critères 3 *"Bonne santé"* et 4 *"Comportement approprié"*

En conclusion, l'intégrale de Choquet, agrégateur généralisant à la fois la somme pondérée et l'OWA, nous permet de tenir compte simultanément des deux types de raisonnements suivis par les experts : attribuer plus de poids à certains sous-critères et/ou pénaliser les fermes sur leurs mauvais résultats. Par ailleurs, nous avons montré qu'il était possible d'optimiser le nombre de paramètres devant être déterminés tout en conservant une bonne flexibilité du modèle. Enfin, nous avons pu voir que les paramètres de l'intégrale de Choquet pouvaient s'interprêter grâce aux calculs des valeurs de Shapley et des indices d'interaction. Ces valeurs et indices nous ont en effet permis de révéler des traits de comportement qui ont été validés par les experts eux-mêmes.

Chapitre 7

L'agrégation des critères

Résumé de l'agrégation des critères :

L'objectif étant d'affecter chaque ferme observée à une catégorie de bien-être pré-définie, l'agrégation des scores de critère repose sur une problématique de TRI ordonné.

Pour ce faire, nous utilisons une méthode de type ELECTRE TRI dans laquelle les performances des fermes sont comparées de manière descendante aux profils de référence permettant de délimiter les différentes catégories de bien-être (c.-à-d. en regardant d'abord si la ferme peut être affectée à la meilleure catégorie, puis, si ce n'est pas le cas, en regardant si elle peut l'être à la seconde meilleure catégorie, et ainsi de suite...).

Quatre catégories correspondant à des niveaux de bien-être croissants ont été définies (*Non classé*, *Normal*, *Bon* et *Excellent*) de sorte à répondre aux attentes des principaux utilisateurs potentiels (éleveurs, distributeurs, vétérinaires, associations de protection animale, institutions).

Les profils de référence délimitant ces catégories et les règles d'appartenance aux différentes catégories doivent être déterminées en accord avec ces utilisateurs potentiels du système d'évaluation. Nous leur proposons une base de discussion devant les aider dans leur choix :
- profils de référence $z_1 = 75$, $z_2 = 50$ et $z_3 = 20$;
- règles d'appartenance : quatre procédures différentes sont proposées, reposant sur des règles de concordance et de discordance plus ou moins strictes (unanimité, majorité qualifiée, ainsi que deux jeux de règles adaptées aux différents profils).

L'analyse est complétée par des simulations à partir de fermes virtuelles ainsi que par l'étude de la répartition des fermes observées dans le WP2.4 dans les différentes catégories de bien-être.

Chapitre 7

L'agrégation des critères

Comme nous avons pu le voir dans le chapitre introductif, la majorité des consommateurs associent le bien-être à la qualité (aussi bien gustative que sanitaire) des produits, au respect de l'environnement, et aux vertus des produits pour la santé humaine (Miele et Evans 2007). Pour aider les utilisateurs potentiels du système d'évaluation à prendre des décisions (notamment aider les consommateurs dans leurs actes d'achat), le système d'évaluation du bien-être développé Welfare Quality® devra pouvoir servir de support à un système d'information délivrant une information synthétique, claire et facilement compréhensible. Cette information sera issue du résultat de la dernière étape de construction du modèle d'évaluation : l'agrégation des quatre critères (*Alimentation adaptée*, *Logement correct*, *Bonne santé* et *Comportement approprié*) en une évaluation globale. Cette dernière requiert tout d'abord la définition de catégories de bien-être ordonnées correspondant chacune à un niveau de bien-être défini et permettant de répondre aux attentes des différents utilisateurs en termes d'information. Il faudra ensuite mettre en place une procédure d'affectation de chaque ferme pouvant être observée à la catégorie appropriée, ce qui correspond à un problématique de tri ordonné (selon la terminologie de Roy 1985).

7.1 Définition du nombre et et de la signification des catégories de bien-être

Afin de produire un système d'évaluation sur lequel pourra reposer un système d'information pertinent (c.-à-d. répondant aux attentes des utilisateurs), il nous semble primordial d'inclure des représentants des principaux utilisateurs dans la mise en place de cette étape finale de la construction de l'évaluation globale du bien-être. Leurs opi-

nions semblent en effet être particulièrement utiles et pertinentes pour la construction de la procédure d'agrégation des quatre critères précédemment définis afin de produire l'évaluation globale.

7.1.1 Identification des utilisations potentielles pouvant être faites de l'outil d'évaluation

Les attentes en matière d'information sur le bien-être animal varient en fonction des utilisateurs. Nous cherchons donc à mettre en en place un système d'évaluation pouvant être utilisé de différentes manières, mais pour ce faire nous avons d'abord dû identifier les attentes en termes d'utilisation(s) des différents utilisateurs potentiels. Par conséquent, nous avons consulté le Comité Conseil du projet (composé de représentants des principaux utilisateurs, cf. Annexe B). Nous avons demandé à ses membres de réagir sur un certain nombre de scénarios potentiels d'utilisation en soulignant les utilisations que leurs organisations respectives considèrent comme utiles pour elles-mêmes.

En collaboration avec des chercheurs en sciences sociales impliqués dans Welfare Quality® nous avons établi quatre scénarios d'utilisation du système d'évaluation et nous les avons soumis aux membres du Comité Conseil :

1. Le système d'évaluation de Welfare Quality® pourrait aider à définir le niveau minimum de bien-être considéré comme acceptable. Il pourrait alors se substituer à la réglementation en vigueur dans l'Union Européenne quant à la définition des standards minimum de bien-être, ou encore aider à la mise en place d'un **étiquetage obligatoire**[1]. Un tel outil de labellisation nécessite la réalisation des inspections sur le terrain par des organismes tiers ou des représentants de l'autorité légale.
2. Le système d'évaluation pourrait aussi aider à distinguer des situations dans lesquelles le niveau de bien-être des animaux non seulement dépasse le niveau minimum acceptable (défini ci-dessus) mais en plus atteint un niveau élevé, voire très élevé. Un tel outil pourrait alors être utilisé dans le cadre d'une **labellisation volontaire**, nécessitant que les inspections soient réalisées par des organismes tiers certificateurs. Plusieurs types de labellisation sont alors envisageables :
   - soit dans le cadre d'un label spécifique dédié au respect de hautes exigences en matière de bien-être animal,

[1] comme par exemple le label Européen proposé par (Kyprianou 2007) afin de différencier les produits issus d'animaux élevés dans des conditions permettant un niveau de bien-être animal compatible avec les réglementations de l'Union Européenne des produits importés dont on ne connaît pas les conditions exactes d'élevage (en particulier quand dans un pays il n'existe pas de réglementations en matière de bien-être animal)

- soit dans le cadre d'un label qualité déjà existant, ce label pouvant être plutôt axé sur la qualité du produit (comme par exemple le *Label Rouge* français) ou se disant déjà respectueux du bien-être animal (comme par exemple le label anglais *Freedom Food*).

3. Le système d'évaluation pourrait aussi se révéler être très utile dans le cadre d'un **outil d'auto-évaluation** pour aider les éleveurs à identifier des problèmes de bien-être ou des facteurs de risque sur leurs fermes, et suivre les progrès qu'ils réalisent en mettant en place des mesures correctrices (en comparant les scores obtenus avant et après que des mesures aient été prises). Dans ce cas, l'outil d'évaluation devra nécessairement être accompagné d'instructions précises quant à la réalisation de l'évaluation. La mise en place d'un système d'information standardisée n'est cependant pas une obligation pour cet outil, les informations fournies au niveau des quatre critères et des 12 sous-critères offrant déjà une information synthétique et utile aux éleveurs.

4. Le système d'évaluation pourrait servir d'**outil de recherche** afin d'évaluer le niveau de bien-être des animaux vivant dans différentes conditions, ou encore pour évaluer celui de nouvelles races dans les conditions actuelles d'élevage. Un tel outil pourrait alors aussi être utilisé dans un **cadre réglementaire** pour aider les instances législatives à prendre des décisions quant à l'autorisation ou non de nouveaux équipements, systèmes de logements, ou races. Dans ce cas, il n'est pas nécessaire d'avoir un système d'information standardisé. Un avis synthétique d'expert peut être suffisant.

Nous avons alors recueilli les réactions des membres du Comité Conseil sur ces différents scénarios. En voici les principaux commentaires :

- Le représentant de l'*Eurogroup for Animals*[2] a réagi positivement sur la plupart des scénarios. Il met notamment en avant la mise en place d'un étiquetage obligatoire (Scénario 1), mais dans une application différente de celle reposant sur un logo européen (que je vais désormais considéré comme le Scénario 1a). Il propose en effet la mise en place d'un étiquetage obligatoire en plusieurs niveaux de bien-être (se rapprochant ainsi du type d'étiquetage indiquant la consommation en énergie des appareils électro-ménagers) (noté désormais Scénario 1b). Il ne voit cependant pas comment il serait possible d'utiliser le système d'évaluation pour vérifier le respect de la législation en vigueur (Scénario 1a), basée sur des obligations de moyens, alors que le système d'évaluation de Welfare Quality® est basé

[2] Association regroupant différentes organisations européennes de protection animale

au contraire sur les résultats (c.-à-d. sur les animaux eux-mêmes). Il considère toutefois que le système produit sera utile pour la recherche et pour fournir aux instances réglementaires des informations qui leur permettront de développer un jour des réglementations basées sur les résultats (Scénario 4). Ensuite, il considère le Scénario 2 (labellisation sur la base du volontariat) comme intéressante, en émettant toutefois des réserves quant à l'inclusion du respect du bien-être dans un label qualité généraliste. Il voit aussi un intérêt à la création d'un outil permettant aux éleveurs, aux transporteurs et aux abattoirs d'avoir un retour sur le niveau de bien-être qu'ils assurent aux animaux dont ils ont la charge (Scénario 3).

- Le représentant de l'*EurSafe*[3] met en avant l'intérêt du système d'évaluation dans le cadre de labels sur la base du volontariat, en soulignant la possibilité de considérer au moins deux niveaux de référence différents : un premier situé au-dessus du niveau minimum requis par la législation et un second permettant de distinguer des fermes assurant un excellent niveau de bien-être. Ensuite, dans une moindre mesure, il cite l'intérêt du Scénario 1 (étiquetage obligatoire) puis du 3 (outil d'auto-évaluation). Par contre, il ne considère pas que le système d'évaluation mis en place dans Welfare Quality® soit réellement pertinent dans un but de recherche.
- Le représentant d'*Aviagen*[4] considère comme nécessaire la mise en place d'un outil d'évaluation standardisé au niveau européen, mais ne voit la mise en place d'un label et/ou d'un étiquetage "bien-être" que comme un élément secondaire. En ce sens, ses attentes se rapproche du Scénario 3 correspondant à une utilisation en tant qu'outil d'(auto-)évaluation (nécessitant toutefois une évaluation synthétique). En cas de mise en place d'un étiquetage, c'est celui reposant sur une labellisation volontaire (Scénario 2) qui lui semble être le plus favorable. Il émet cependant des réserves quant au coût engendré par les inspections qui seraient alors nécessaires.
- Le représentant d'*Ahold*[5] met en avant les Scénarios 1 (étiquetage obligatoire) et 2 (labellisation volontaire). Il qualifie l'utilisation du système en tant qu'outil d'auto-évaluation d'intéressante mais ne pouvant apporter de réel changement. Par contre, il considère que l'utilisation en tant qu'outil de recherche n'est pas réaliste.

[3] *European Society for Agricultural and Food Ehics*, organisation indépendante qui fournit un lieu de discussion reconnu au niveau européen aux personnes ayant un intérêt professionnel pour les questions éthiques liées à l'agriculture et aux filières agro-alimentaires

[4] Leader mondial des sélectionneurs de volailles

[5] Entreprise Néerlandaise dont l'activité est centrée sur la grande distribution et le commerce alimentaire

- Le représentant de la *Federation of Veterinarians of Europe* (FVE) considère que les scénarios les plus pertinents sont ceux qui concernent l'utilisation du système d'évaluation dans le cadre d'un système de labellisation sur la base du volontariat (Scénario 2) ainsi qu'en tant qu'outil d'auto-évaluation et de suivi des progrès effectués (Scénario 3).
- Le représentant de la DG-SANCO[6] considère que dans le futur le système d'évaluation proposé par Welfare Quality® pourra être utilisé pour mettre en place une labellisation européenne obligatoire basée sur des obligations de résultat. Il considère aussi d'un oeil favorable l'utilisation dans le cadre de labellisation(s) volontaire(s). Enfin, il souligne l'intérêt du système d'évaluation en tant qu'outil de recherche.

Il en ressort que les quatre scénarios proposés semblent tous intéresser au moins un membre du Comité Conseil (cf. Tableau 7.1), chaque membre représentant les attentes de son organisation. Un seul nouveau scénario a été proposé par les personnes consultées, il s'agit du Scénario 1b proposant un étiquetage obligatoire en plusieurs niveaux. Cependant, le représentant d'Eurogroup signale qu'il pourrait y avoir dans le futur d'autres utilisations auxquelles on ne pense pas à l'heure actuelle. Le système d'évaluation devrait donc être suffisamment flexible pour pouvoir s'adapter à des situations non envisagées pour le moment.

7.1.2 Conséquences en termes de nombre et de signification des catégories de bien-être

Chacune des utilisations potentielles identifiées à partir des scénarios nécessite la définition de plusieurs catégories, d'un *mauvais* à un *excellent* niveau de bien-être (cf. Tableau 7.2). C'est en nous basant sur ces utilisations potentielles que nous avons pu définir le nombre de catégories pré-définies de bien-être devant être prises en compte pour que l'outil produit puisse être utilisé dans tous les scénarios envisagés.

Nous pouvons constater que les trois catégories nécessaires pour la mise en place d'une labellisation sur la base du volontariat ne recouvrent pas les deux catégories nécessaires pour un étiquetage obligatoire de type logo européen (logo UE). En effet, pour pouvoir prétendre à un label qualité, une ferme doit au minimum assurer à ses animaux un niveau de bien-être significativement supérieur à ce que la réglementation en vigueur demande, alors que les deux catégories requises par l'étiquetage obligatoire de type logo UE ne permettent que de distinguer les fermes remplissant les conditions imposées par la loi de celles qui ne les ont pas encore atteintes. Ainsi, ne serait-ce que

[6]Direction Générale "Santé et la protection des Consommateurs" de la Commission Européenne

TAB. 7.1 – Perception des différents scénarios par les principaux membres du Comité Conseil de Welfare Quality®

Organisations	**Scénario 1** Étiquetage obligatoire		**Scénario 2** Label volontaire	**Scénario 3** Outil d'évaluation	**Scénario 4** Outil de recherche
	1a[a]	1b[b]			
Eurogroup	−	+	+	+	+
EurSafe	+		+	+	−
Aviagen	−		+	+	+
Ahold	+		+	+	−
FVE			+	+	
DG-SANCO	+		+		+

[a]information binaire type 'logo UE'
[b]information en plusieurs niveaux

pour ces deux utilisations potentielles, au moins quatre catégories de bien-être semblent nécessaires. Ces quatre catégories correspondent d'ailleurs aux attentes dans le cadre d'une utilisation pour un étiquetage obligatoire fournissant une information en plusieurs niveaux de bien-être. Dans le cas des deux dernières utilisations (outil d'évaluation sans étiquetage et outil de recherche), aboutir à une évaluation globale formalisée n'apparaît pas forcément comme une nécessité. Cependant, quatre catégories distinctes semblent permettre d'obtenir une évaluation générale suffisante pour un premier classement ou comparaison. Pour ces deux derniers scénarios, cette première impression devra de toute façon être complétée par les données précises au niveau des critères et des sous-critères.

Au vu des scénarios retenus par les utilisateurs potentiels consultés, nous avons décidé de définir les quatre catégories de bien-être suivantes :

- **Excellent** : le niveau de bien-être des animaux est particulièrement bon. La ferme (ou l'abattoir) peut correspondre aux exigences d'un marché de niche via un label assurant aux consommateurs des produits de très haute qualité, ce label pouvant être dédié au bien-être animal.
- **Bon** : le niveau de bien-être des animaux est bon (mais pas excellent). Les bonnes pratiques d'élevages appliquées sont suffisantes pour garantir un bon niveau de bien-être au sein d'un label qualité plus généraliste pouvant inclure des garanties

TAB. 7.2 – Utilisations potentielles du système d'évaluation globale du bien-être et nombre de catégories devant alors être considérées

Utilisations potentielles de l'outil d'évaluation	**Catégories de bien-être**	
	Nombre	Signification
Étiquetage obligatoire *de type logo UE*	*2*	En-dessous *vs* au-dessus du minimum légal exigé (ou d'un équivalent si aucune réglementation n'existe ou si celle-ci ne porte pas sur les mêmes éléments)
Étiquetage obligatoire *fournissant une évaluation du bien-être en plusieurs niveaux*	*4*	Quatre catégories différenciant les niveaux de bien-être *Mauvais*, *Normal*, *Bon* et *Excellent*
Label	*3*	Au moins trois catégories différenciant des niveaux de bien-être *Insuffisant* pour faire partie d'un label, *Bon* (suffisant pour être inclus dans un label qualité) ou *Excellent* (pouvant être inclus dans un label dédié au bien-être)
Outil d'évaluation et de suivi des améliorations et/ou détériorations	*Plusieurs*	*Mauvais* ——————— *Excellent* niveau de bien-être
Outil d'évaluation dans un objectif de recherche ou de prise de décision réglementaire	*Plusieurs*	*Mauvais* ——————— *Excellent* niveau de bien-être

sur d'autres caractéristiques du produit (comme par exemple la qualité gustative). Il s'agit dans ce cas d'un type de label volontaire différent de celui se basant sur la catégorie "Excellent", en étant en effet un peu moins exigeant.

- **Normal** : le niveau de bien-être des animaux est acceptable, mais insuffisant pour permettre à la ferme (ou à l'abattoir) d'entrer dans un système de labellisation volontaire. Il correspond par contre à un niveau de bien-être suffisant pour que la ferme soit considérée comme acceptable (c.-à-d. respectant au moins les exigences réglementaires ou équivalentes).
- **Non classé** : le niveau de bien-être des animaux est faible et considéré comme insuffisant pour que la ferme puisse être classée (et ainsi pouvoir intégrer un système de certification).

Ces quatre catégories de bien-être pourront être utilisées différemment en fonction de l'utilisation potentielle qui en sera faite (cf. Figure 7.1).

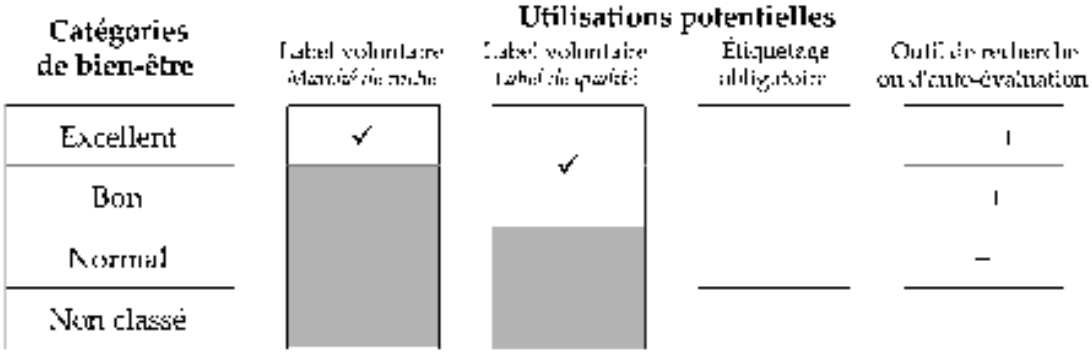


FIG. 7.1 – Utilisation des catégories de bien-être pour chacun des usages pouvant être fait du système d'évaluation. En fonction de l'usage qui en sera fait, la réponse pourra être binaire (ex. *S'agit-il d'une ferme excellente ou non ?*) ou graduelle (*La ferme est-elle −, +, ++ ou +++ ?*).

7.2 Procédures d'affectation aux quatre catégories de bien-être

Comme nous l'avons établi à la fin du chapitre introductif (p. 48-51), nous sommes face à une problématique de tri ordonné afin d'affecter chaque ferme à l'une des quatre catégories ordonnées de bien-être venant d'être définies (*"Excellent"*, *"Bon"*, *"Normal"* et *"Non classé"*) à partir des scores obtenus par cette ferme sur les quatre critères de bien-être f_1, f_2, f_3 et f_4. Nous devons donc désormais établir une méthode permettant d'atteindre cet objectif.

Par principe, nous ne considérons pas dans Welfare Quality® qu'il soit pertinent de produire une évaluation sous la forme d'une note globale de bien-être. En effet, les quatre critères construits au Chapitre 6 représentent des dimensions très différentes du bien-être animal. Synthétiser les scores de critères en une note globale exprimée sur une échelle $[0, 100]$ reviendrait donc, à notre avis, à "additionner des choux et des carottes" (comme par exemple l'*alimentation* et le *logement* des animaux). Ce sentiment est renforcé par le fait que les critères de bien-être ne peuvent pas se compenser (cf. Chapitre 3, p. 75-76,et Chapitre 4, p. 105). Résumer l'évaluation multidimensionnelle du bien-être par une note globale nous semble donc inopportun, même si avec une telle méthode il serait tout à fait possible de définir des catégories, en fixant par exemple des seuils sur la note globale[7]. Nous nous proposons plutôt d'utiliser une véritable méthode d'analyse multicritère sur l'espace des quatre grandes dimensions du bien-être animal formé par les critères.

Nous avons donc choisi de suivre une procédure de tri multicritère. Parmi les procédures d'affectation établies dans la littérature, nous avons retenu les procédures d'affectation de type ELECTRE TRI (voir par exemple Moscarola et Roy 1977 ; Mousseau et Slowinski 1998) ou une extension de cette méthode permettant d'introduire de la logique floue (voir par exemple Perny 1998 ; Mousseau et al. 1999 ; Mousseau et al. 2000).

[7]Il serait en effet possible d'utiliser une intégrale de Choquet pour agréger les scores de critères et ainsi obtenir au niveau de la ferme un score global exprimé sur l'échelle $[0, 100]$. En utilisant des valeurs de capacité telles que la procédure d'agrégation se caractériserait par une tendance non-compensatoire. Il faudrait ensuite placer trois seuils sur l'échelle $[0, 100]$ de sorte à délimiter les quatre catégories de bien-être.

7.2.1 Présentation de la méthode ELECTRE TRI

La méthode ELECTRE TRI[8] permet d'affecter une alternative (une ferme dans notre cas) $x_k = (x_{k1}, \ldots, x_{km})$ (avec $x_{kj} = f_j(x_k), j = 1, \ldots, 4$) à une des n catégories C_i $(i = 1, \ldots, n)$ ordonnées pré-définies. Pour ce faire, cette méthode se base sur les performances x_{kj} $(j = 1, \ldots, m)$ intrinsèques de l'alternative sur l'ensemble des critères et les compare à des profils de référence $z_i = (z_{i1}, \ldots, z_{im})$ $(i = 1, \ldots, n)$, z_i représentant la limite supérieure de la catégorie C_{i-1} et la limite inférieure de la catégorie C_i.

Dans notre cas, $n = 4$ (quatre catégories C_1=*"Excellent"*, C_2=*"Bon"*, C_3=*"Normal"* et C_4=*"Non classé"*) et $m = 4$ (quatre critères f_1, f_2, f_3 et f_4).

ELECTRE TRI construit une relation de préférence $\succ$ qui valide ou invalide l'affirmation $x_k \succ z_i$ *"la ferme x_k est au moins aussi bonne que z_i"* (*"x_k bat le profil z_i"*).

Pour valider cette affirmation, deux conditions doivent être vérifiées :

- **Concordance** : pour que $x_k \succ z_i$ soit acceptée, une majorité "suffisante" de critères doit être en faveur de cette affirmation,
- **Non-discordance** : quand les conditions de concordance sont vérifiées, aucun des critères de la minorité ne doit s'opposer "trop fortement" à l'affirmation $x_k \succ z_i$.

Deux types de paramètres de préférence entre les critères interviennent dans la construction de $\succ$:

- l'ensemble des poids $(w_1, w_2, \ldots, w_m)$ est utilisé dans le test de concordance quand l'importance relative des coalitions de critères en faveur de l'affirmation $x_k \succ z_i$ est calculée,
- l'ensemble des seuils de veto $(v_{i1}, v_{i2}, \ldots, v_{im})$ $\forall i \in \{1, \ldots, n\}$ est utilisé dans le test de discordance. v_{ij} représente la plus petite différence $z_{ij} - x_{kj}$ incompatible avec l'affirmation $x_k \succ z_i$.

Afin de tenir compte de l'imprécision et de l'incertitude sur les données, il est possible d'introduire des seuils d'indifférence q_{ij} et de préférence p_{ij} qui constituent alors une information de préférence intra-critère.

Dans la méthode ELECTRE TRI, la relation de préférence $\succ$ est construite de sorte à permettre la comparaison entre l'alternative x_k et le profil z_i. $\succ$ est construite en suivant les étapes suivantes :

1. Calculer les *indices de concordance partielle* $c_j(x_k, z_i)$ et $c_j(z_i, x_k)$, exprimant respectivement dans quelle mesure l'affirmation "x_k bat z_i" est vraie et dans quelle mesure l'affirmation "z_i bat x_k par rapport au critère j" est vraie. L'indice de concordance partielle $c_j(x_k, z_i)$ se calcule comme suit (cf. Figure 7.2) :

[8]Nous présentons dans cette section la méthode ELECTRE TRI 2.0 (Mousseau et al. 1999, 2000)

$$\begin{cases} c_j(x_k, z_i) = 0 & \text{si} \quad x_{kj} \leq (z_{ij} - p_{ij}) \\ c_j(x_k, z_i) = \frac{x_{kj} - z_{ij} + p_{ij}}{p_{ij} - q_{ij}} & \text{si} \quad (z_{ij} - p_{ij}) < x_{kj} \leq (z_{ij} - q_{ij}) \\ c_j(x_k, z_i) = 1 & \text{si} \quad (z_{ij} - q_{ij}) < x_{kj} \end{cases} ;$$

2. Calculer les *indices de concordance globale* $c(x_k, z_i)$ et $c(z_i, x_k)$, exprimant dans quelle mesure les évaluations de x_k et de z_i sur tous les critères sont concordantes avec l'affirmation "x_k bat z_i" (respectivement "z_i bat x_k"). L'indice de concordance $c(x_k, z_i)$ se calcule comme suit :

$$c(x_k, z_i) = \frac{\sum_{j=1\ldots m} w_j \, c_j(x_k, z_i)}{\sum_{j=1\ldots m} w_j} \quad ;$$

3. Calculer les *indices de discordance partielle* $d_j(x_k, z_i)$ et $d_j(z_i, x_k)$, exprimant dans quelle mesure le critère f_j est opposé à l'affirmation "x_k bat z_i" (respectivement "z_i bat x_k"). Un critère f_j est dit discordant avec l'affirmation "x_k bat z_i" si sur ce critère z_i est préféré à x_k (c.-à-d. $c_j(z_i, x_k) = 1$ et $c_j(x_k, z_i) = 0$). Le critère f_j oppose un veto quand la différence $z_{ij} - x_{kj}$ dépasse le seuil de veto v_{ij}. L'indice de discordance partielle $d_j(x_k, z_i)$ se calcule comme suit (cf. Figure 7.2) :

$$\begin{cases} d_j(x_k, z_i) = 0 & \text{si} \quad x_{kj} > (z_{ij} - p_{ij}) \\ d_j(x_k, z_i) = \frac{z_{ij} - x_{kj} - p_{ij}}{v_{ij} - p_{ij}} & \text{si} \quad (z_{ij} - v_{ij}) < x_{kj} \leq (z_{ij} - p_{ij}) \\ d_j(x_k, z_i) = 1 & \text{si} \quad (z_{ij} - v_{ij}) \geq x_{kj} \end{cases} ;$$

4. Calculer le degré de crédibilité de la relation de préférence floue (sur la base du calcul des *indices de crédibilité* $\sigma(x_k, z_i)$) ;

5. Déterminer la relation de préférence $\succ$. En cas d'introduction de logique floue, il faut traduire la relation de préférence floue en relation de préférence précise ($\succ$). Il faut alors déterminer λ la plus petite valeur de l'indice de crédibilité compatible avec l'affirmation "x_k bat z_i", $\sigma(x_k, z_i) \geq \lambda \Rightarrow x_k \succ z_i$.

La procédure d'affectation revient à définir la manière dont une alternative x_k doit être comparée aux profils de référence afin de déterminer la catégorie à laquelle x_k devrait être affectée. ELECTRE TRI propose deux procédures d'affectation possibles :

- la procédure d'affectation *pessimiste* (ou *descendante*) qui consiste à comparer x_k successivement aux profils z_i avec $i = 1, 2, \ldots, n$, si z_i est le premier profil tel que $x_k \succ z_i$, alors il faut affecter x_k à la catégorie C_i ;
- la procédure d'affectation *optimiste* (ou *ascendante*) qui consiste à comparer x_k successivement aux profils z_i avec $i = n, n-1, \ldots, 1$, si z_i est le premier profil tel que $z_i \succ x_k$, alors il faut affecter x_k à la catégorie C_{i-1}.

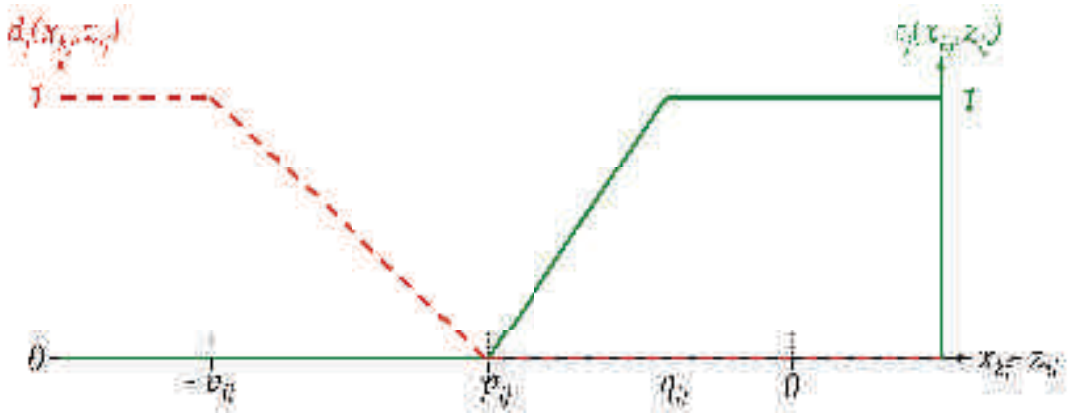


FIG. 7.2 – Définition des indices de concordance et de discordance partielles

Pour que les catégories soient définies de manière cohérente, les profils doivent respecter les deux conditions suivantes :

- Condition 1 : $\forall j \in 1, \ldots, m$, $\forall i \in 1, \ldots, n-1 \quad z_{ij} \geq z_{(i+1)j}$;
- Condition 2 : $\forall j \in 1, \ldots, m$, $\forall i \in 1, \ldots, n-1 \quad (z_{ij} - p_{ij}) \geq (z_{(i+1)j} + p_{(i+1)j})$.

La Condition 1 permet de s'assurer que les catégories sont ordonnées (C_1 meilleure que C_2, C_2 meilleure que C_3...), alors que la Condition 2 permet de s'assurer que les catégories peuvent être distinguées.

7.2.2 Présentation de la méthode que nous avons adoptée

Nous proposons d'utiliser une méthode qui se rapproche de la méthode ELECTRE TRI sans pour autant être exactement similaire. En effet, au lieu d'utiliser le principe de concordance et le principe de non-discordance, nous proposons de définir pour l'appartenance à chaque catégorie C_i deux niveaux d'exigence différents (soit deux test de concordance différents) : un *niveau d'exigence minimale* et un *niveau d'aspiration*[9] à partir duquel la coalition de critères concordants est jugée suffisante pour que la ferme soit affectée à la catégorie C_i.

Nous proposons de définir le niveau d'exigence minimale sur la catégorie C_{i+1} (en exigeant l'unanimité sur z_{i+1}, c.-à-d. $c(x_k, z_{i+1}) = 1$) et le niveau d'aspiration sur la catégorie C_i (en comparant donc x_k à z_i, l'affirmation "x_k bat z_i" si et seulement si $c(x_k, z_i) \geq \lambda_i$, $\lambda_i \in [0, 1]$ à définir). Ainsi, pour pouvoir définir le niveau d'exigence minimale pour la catégorie C_3 "Normal" nous avons besoin de faire appel à un profil

[9] terminologie issue de la théorie de la rationalité limité de H. Simon et correspondant à des méthodes de décision visant à considérer comme bonnes ou satisfaisantes des actions, même si elles ne sont pas optimales (cf. Parthenay 2005)

z_4. S'il était défini comme la limite inférieure de C_4[10], il correspondrait au vecteur $z_4 = (0, 0, 0, 0)$, or nous désirons que z_4 prenne des valeurs supérieures à 0, ce qui permet de nous assurer qu'une ferme obtenant un très faible score ($< z_{4j}$ sur un critère f_j) n'a aucune chance d'être affectée à la catégorie C_3.

Nous devons donc définir quatre profils de référence z_i $i \in \{1, 2, 3, 4\}$ (cf. Figure 7.3).

L'utilisation du niveau d'exigence minimale, issu de la comparaison avec le profil de la catégorie inférieure en termes de bien-être, se rapproche de la notion de discordance. Cependant, alors que les seuils de veto dans le principe de discordance sont fixés de manière relative par rapport au profil z_i, la comparaison par rapport à z_{i+1} correspondrait à un seuil de veto fixé par rapport à z_{i+1} et donc indépendamment de z_i.

Nous ne proposons pas d'introduire de la logique floue dans la méthode, même si, comme nous venons de le voir avec la méthode ELECTRE TRI ce serait parfaitement possible, en définissant pour cela les seuils d'indifférence et de préférence.

Nous définissons alors une relation de préférence $\succsim$ au lieu de la relation $\succ$, basée sur les indices de concordance partielle définis comme suit :

$$\begin{cases} c_j(x_k, z_i) = 0 & \text{si} \quad x_{kj} \leq z_{ij} \\ c_j(x_k, z_i) = 1 & \text{si} \quad x_{kj} \geq z_{ij} \end{cases}$$

de sorte à obtenir $c(x_k, z_i) = 1$ même si $\forall j \; x_{kj} = z_{ij}$.

Nous proposons de suivre une *procédure de tri descendante* : pour une ferme donnée, nous regardons d'abord si elle peut être affectée à la meilleure catégorie (*"Excellent"*), si ce n'est pas le cas, nous regardons si elle peut être affectée à la seconde meilleure catégorie (*"Bon"*), et ainsi de suite... Ainsi, dans l'exemple présenté sur la Figure 7.3, le vecteur z_1 représente la limite inférieure de la catégorie *"Excellent"*. Pour n'importe quelle ferme x_k (correspondant à la ligne en pointillé sur la Figure 7.3), nous commençons par comparer x_k à z_1. Si x_k est au moins aussi bonne que z_1 (noté $x_k \succsim z_1$) alors x_k est affectée à la catégorie C_1 : *"Excellent"*. Si ce n'est pas le cas (non($x_k \succsim z_1$)), x_k est comparée à un profil moins exigeant z_2 qui représente la limite inférieure de la seconde catégorie (C_2 : *"Bon"*). Si $x_k \succsim z_2$ alors x_k est affectée à C_2, sinon nous la comparons au troisième profil de référence z_3 et ainsi de suite... Si $x_k \succsim z_3$ alors elle est affectée à la catégorie C_3 : *"Normal"*, sinon, la ferme x_k est affectée à la dernière catégorie C_4. De manière plus formelle, la procédure d'affectation s'écrit comme suit :

Remarque : une ferme ne peut être affectée qu'à une et une seule catégorie. En effet,

[10] c.-à-d. en appliquant la même règle que pour les trois autres catégories

Algorithme 4 : Procédure d'affectation des fermes aux différentes catégories

```
for k = 1 to n_f do
    i ← 1
    while not (x_k ≳ z_i) and i < 4 do
        i ← i + 1
    end
    x_k → C_i
end
```

même si par définition une ferme $x_k \succsim z_i$ alors $x_k \succsim z_{i+1}$[11], l'Algorithme 4 stoppe la procédure de calcul dès qu'il trouve un i tel que $x_k \succsim z_i$ et affecte alors la ferme x_k à la catégorie C_i, et ce même si les performances de la ferme lui auraient aussi permis d'être affectée à la catégorie C_{i+1}.

7.2.3 Choix des profils délimitant les catégories de bien-être

Nous proposons d'utiliser les profils de référence suivants (cf. Figure 7.3) :

- **Profil 1** (limite inférieure de la catégorie C_1 "Excellent") fixé à 75 pour les quatre critères, soit $z_1 = (75, 75, 75, 75)$,
- **Profil 2** (limite inférieure de la catégorie C_2 "Bon") fixé à 50 pour les quatre critères, soit $z_2 = (50, 50, 50, 50)$,
- **Profil 3** (limite inférieure de la catégorie C_3 "Normal") fixé à 20 pour les quatre critères, soit $z_3 = (20, 20, 20, 20)$, ce qui correspond à la valeur présentée au experts au cours de la construction des sous-critères et des critères comme la valeur en-deçà de laquelle le niveau de bien-être est jugé inacceptable (cf. Chapitre 5, p. 118),
- **Profil 4** fixé à 10 (soit la moitié de z_3) pour les quatre critères, soit $z_4 = (10, 10, 10, 10)$.

7.2.4 Définition des procédures d'affectation aux différentes catégories de bien-être

L'objectif de cette section est de proposer plusieurs procédures d'affectation des fermes aux différentes catégories de bien-être, ces procédures étant plus ou moins strictes. Le choix final de la procédure d'affectation se fera à la fin du projet Welfare Quality®, après avoir testé les différentes options sur un jeu de données important. L'utilisation

[11] $x_k \succsim z_i \Rightarrow c(x_k, z_{i+1}) = 1 \geq \lambda_{i+1}$; par ailleurs les profils étant définis de sorte à ce que $z_{ij} \geq z_{(i+1)j} \Rightarrow z_i \succsim z_{i+1}$, alors $c(x_k, z_{i+1}) = 1 \Rightarrow c(x_k, z_{i+2}) = 1$; et on a donc $(c(x_k, z_{i+2}) = 1$ et $c(x_k, z_{i+1}) \geq \lambda_{i+1}) \Rightarrow x_k \succsim z_{i+1}$

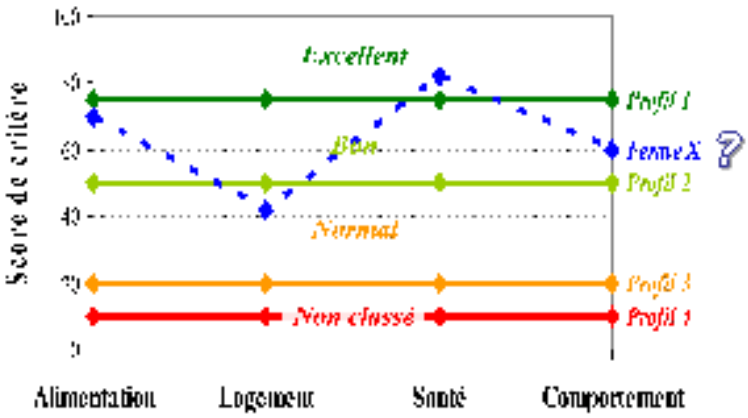


FIG. 7.3 – Agrégation des critères : comparaison aux profils de référence

d'un jeu de données représentatif de l'élevage en Europe nous permettra alors de nous assurer :

- que chaque catégorie inclut un pourcentage significatif de fermes,
- et que l'évaluation en termes de bien-être animal pouvant être faite par des experts sur les différentes fermes du jeu de données est bien en accord avec la signification attribuée à la catégorie (*Non classé*, *Normal*, *Bon* ou *Excellent*) à laquelle le modèle affecte la ferme considérée.

Nous définissons ici quatre procédures d'affectation différentes, mais la liste n'est bien évidemment pas exhaustive. Comme nous l'avons vu lors de la présentation de la méthode générale que nous avons choisi d'utiliser, pour savoir si une ferme x_k peut être affectée à la catégorie C_i, il faut dans un premier temps vérifier que $c(x_k, z_{i+1}) = 1$, et si cette exigence minimale est vérifiée, alors il faut vérifier que :

$$c(x_k, z_i) = \frac{\sum_{j=1\ldots4} w_j\, c_j(x_k, z_i)}{\sum_{j=1\ldots4} w_j} \geq \lambda_i \tag{7.1}$$

Avec l'Équation 7.1 nous prévoyons de pouvoir attribuer plus ou moins d'importance aux différents critères. Lors des réunions menées jusqu'à présent avec des scientifiques et de représentants d'utilisateurs potentiels du système d'évaluation du bien-être, la possibilité d'attribuer des poids w_j différents aux critères f_j n'a pas fait l'objet d'attentes particulières, il semble plutôt que les critères doivent être équipondérés ($\forall j \in \{1, 2, 3, 4\}\ w_j = 1$).

Les quatre procédures diffèrent donc sur les valeurs attribuées au λ_i, $i = 1, \ldots, 4$:

1. règle d'unanimité, appliquée à l'ensemble des profils, soit $\forall i\ \lambda_i = 1$;

2. règle se basant sur une majorité qualifiée (3 contre 1), appliquée à l'ensemble des profils, soit $\forall i\ \lambda_i < 1$ (en cas d'équipondération des critères $\forall i\ \lambda_i = 0.75$) ;
3. premier jeu de règles différentes en fonction du profil z_i considéré, dont la sévérité augmente avec i, avec notamment une exigence d'unanimité par rapport à z_3, soit $\lambda_1 < \lambda_2 < \lambda_3 = 1$;
4. second jeu fait sur le même modèle que la Procédure 3, mais avec des λ_i plus faibles.

Ces quatre procédures utilisent l'Algorithme 5 pour définir la relation $\succsim$, seuls les paramètres de pondération et les valeurs des λ_i changent d'une procédure à l'autre.

Algorithme 5 : Définition de $\succsim$ dans les Procédures 1 à 4

for $i = 1$ *to* 3 **do**
  $c(x_k, z_i) \leftarrow 0$
  $c(x_k, z_{i+1}) \leftarrow 0$
  for $j = 1$ *to* 4 **do**
    if $x_{kj} \geq z_{ij}$ **then** $c_j(x_k, z_i) \leftarrow 1$ **and** $c_j(x_k, z_{i+1}) \leftarrow 1$
    else
      $c_j(x_k, z_i) \leftarrow 0$
      if $x_{kj} \geq z_{(i+1)j}$ **then** $c_j(x_k, z_{i+1}) \leftarrow 1$
      else $c_j(x_k, z_{i+1}) \leftarrow 0$
    end
  end
  $c(x_k, z_i) \leftarrow \frac{\sum_{j=1\ldots4} w_j\, c_j(x_k, z_i)}{\sum_{j=1\ldots4} w_j}$
  $c(x_k, z_{i+1}) \leftarrow \frac{\sum_{j=1\ldots4} w_j\, c_j(x_k, z_{i+1})}{\sum_{j=1\ldots4} w_j}$
  if $c(x_k, z_{i+1}) = 1$ **and** $c(x_k, z_i) \geq \lambda_i$ **then** $x_k \succsim z_i$
  else non$(x_k \succsim z_i)$
end

Dans cet algorithme nous prévoyons la possibilité de pondérer les critères les uns par rapport aux autres via la définition du vecteur $w = (w_1, w_2, w_3, w_4)$. Toutefois, nous allons par la suite considérer que les *critères sont équipondérés*.

7.2.4.1 Procédure 1 : règle d'unanimité, appliquée à l'ensemble des profils

Dans cette première procédure, une ferme x_k ne peut appartenir à la catégorie C_i que si $\forall\, j = 1, \ldots, 4\ c_j(x_k, z_i) = 1$, ce qui revient à définir $\forall i\ \lambda_i = 1$. Avec cette procédure très stricte, si le niveau d'aspiration est respecté, alors le niveau d'exigence minimale (par rapport à z_{i+1}) l'est forcément du fait de la définition de catégories ordonnées.

Cette procédure est particulièrement stricte puisqu'elle interdit toute compensation entre les scores de critères. En effet, il suffit par exemple que la ferme x_k ne batte pas le profil z_1 sur un seul critère j pour que la ferme ne puisse atteindre la catégorie C_1 *"Excellent"*, et ce même si sur le critère j elle obtient un résultat tout à fait correct. Ainsi, même si $x_k = (100, 69, 100, 100)$, cette ferme ne pourra pas être classée dans la catégorie C_1 à cause de son résultat de 69 sur f_2. Cette ferme atteint le niveau d'exigence minimum sur C_1, mais à cause de son score sur f_2 elle n'atteint pas le niveau d'aspiration sur C_1. Pourtant, avec des scores atteignant le maximum sur les autres critères et ce résultat tout à fait correct sur f_2, il pourrait être assez logique de considérer x_k dans la catégorie *"Excellent"*.

7.2.4.2 Procédure 2 : règle se basant sur une majorité qualifiée, appliquée à l'ensemble des profils

Dans le Procédure 2, la règle d'affectation est un peu plus souple puisque pour être affectée à la catégorie C_i la ferme x_k n'a qu'à battre le profil z_i sur trois des quatre critères, ce qui correspond à une règle de majorité qualifiée. Cependant, en ne considérant que le respect du niveau d'aspiration, cette règle pourrait permettre à la ferme $x_k = (80, 80, 5, 80)$ d'être affectée à la classe C_1 *"Excellent"*, et ce malgré le résultat particulièrement mauvais qu'elle a obtenu sur f_3. Ainsi, il serait possible de totalement compenser ce résultat inacceptable par d'excellents résultats sur les trois autres critères. Un tel phénomène de compensation à ce stade de la construction ne peut être toléré (cf. discussion sur les compensations dans le Chapitre 3). Cet exemple nous permet d'illustrer l'intérêt de prendre aussi en compte le niveau d'exigence minimale : pour pouvoir être affectée à C_1, x_k doit au minimum battre z_2 à l'unanimité, et en plus battre z_1.

Avec une telle règle d'affectation, la ferme $x_1 = (80, 80, 5, 80)$ serait affectée à la catégorie C_4 *"Non classé"* du fait de son très mauvais résultat sur f_3 ($x_{k3} = 5 < z_{43} = 10$). Cet exemple permet de bien illustrer le pouvoir non-compensatoire de l'utilisation du niveau d'exigence minimale. En effet, en ce contentant du niveau d'aspiration, cette ferme aurait été affectée à la catégorie *"Excellent"* (la meilleure catégorie) alors qu'avec le niveau d'exigence minimale elle se retrouve dans la plus mauvaise catégorie *"Non classé"*.

La niveau d'aspiration défini dans la Procédure 2 se base sur une règle de majorité qualifiée à 3 contre 1, ce qui revient à définir $\forall i \ \lambda_i = 0.75$.

7.2.4.3 Procédure 3 : premier jeu de règles différentes en fonction du profil z_i considéré, dont la sévérité augmente avec i

Les deux procédures qui viennent d'être présentées ont pour principale caractéristique d'utiliser, quel que soit le profil considéré, la même règle pour vérifier le respect des niveaux d'aspiration. Cependant, compte tenu de la difficulté à atteindre le niveau d'excellence pour un critère donné (cf. dernière section du Chapitre 6), on pourrait très bien établir des règles plus ou moins strictes en fonction du profil considéré. En effet, la catégorie *"Normal"* correspondant à des fermes qui doivent au moins respecter la législation en vigueur (ou son équivalent), il ne semble pas aberrant de considérer que pour être affectée à la catégorie *"Normal"* une ferme doive battre z_3 sur l'ensemble des critères (règle d'unanimité sur z_3). Au contraire, il pourrait sembler suffisant pour affecter une ferme à la catégorie *"Excellent"* que cette ferme offre un bon niveau de bien-être partout (c.-à-d. respect du niveau d'exigence minimale) et qu'en plus elle offre un niveau de bien-être excellent sur une partie des critères (c.-à-d. respect du niveau d'aspiration).

Ainsi, alors que pour la comparaison au profil z_3 on pourrait être très strict en imposant l'unanimité, la règle pour qu'une ferme soit affecté à la catégorie "Excellent" (comparaison au profil z_1) pourrait être beaucoup plus souple en qualifiant une ferme d'excellente si elle fournit un niveau de bien-être "bon" sur deux des critères et que sur les deux autres elle fournit un "excellent" niveau de bien-être (sur deux critères la ferme bat z_1 alors que sur les deux autres elle ne bat que z_2). Cette différenciation des règles en fonction des profils de référence semble en accord avec le souhait de plusieurs experts, exprimé soit au cours de réunions soit lors de la consultation pour la construction de critères, de considérer qu'une ferme qui obtient des résultats médiocres sur trois critères et mauvais sur le dernier devrait recevoir une pénalité plus forte qu'une ferme excellente sur trois critères et "seulement" bonne sur le dernier. Ainsi, plus une ferme obtient de mauvais résultats, moins on tolère ses points faibles. Ainsi, la Procédure 3 propose les règles d'affectation suivantes :

- **comparaison au profil** z_1 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_2) = 1$ et $c(x_k, z_1) \geq \lambda_1$, avec $\lambda_1 = 0.50$
  $\hookrightarrow$ *x_k doit battre z_2 à l'unanimité et battre z_1 sur deux critères* ;
- **comparaison au profil** z_2 **:** la ferme x_k bat z_2 ($x_k \succsim z_2$) si et seulement si $c(x_k, z_3) = 1$ et $c(x_k, z_2) \geq \lambda_2$, avec $\lambda_2 = 0.75$
  $\hookrightarrow$ *x_k doit battre z_3 à l'unanimité et battre z_2 sur trois critères* ;
- **comparaison au profil** z_3 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_3) = 1$, c.-à-d. avec $\lambda_3 = 1$
  $\hookrightarrow$ *x_k doit battre z_3 à l'unanimité.*

7.2.4.4 Procédure 4 : second jeu de règles différentes en fonction du profil z_i considéré, dont la sévérité augmente avec i

Alors que dans la Procédure 3 nous proposions une règle assez stricte, notamment pour battre le Profil 3 (règle d'unanimité sur z_3 pour respecter le niveau d'aspiration), nous proposons ici des règles plus souples, mais en gardant l'idée principale : adapter la règle d'affectation en fonction de la catégorie C_i par rapport à laquelle on veut tester la ferme (en comparant les performances de la ferme considérée x_k à z_{i+1} pour le niveau d'exigence minimale et par rapport à z_i pour le niveau d'aspiration), en étant de plus en plus strict quand i augmente. La Procédure 4 repose sur les règles suivantes :

- **comparaison au profil** z_1 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_2) = 1$ et $c(x_k, z_1) \geq \lambda_1$, avec $\lambda_1 = 0.25$
  $\hookrightarrow$ *x_k doit battre z_2 à l'unanimité et battre z_1 sur un critère* ;
- **comparaison au profil** z_2 **:** la ferme x_k bat z_2 ($x_k \succsim z_2$) si et seulement si $c(x_k, z_3) = 1$ et $c(x_k, z_2) \geq \lambda_2$, avec $\lambda_2 = 0.50$
  $\hookrightarrow$ *x_k doit battre z_3 à l'unanimité et battre z_2 sur deux critères* ;
- **comparaison au profil** z_3 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_4) = 1$ et $c(x_k, z_3) \geq \lambda_3$, avec $\lambda_3 = 0.75$
  $\hookrightarrow$ *x_k doit battre z_4 à l'unanimité et battre z_3 sur trois critères.*

Dans cette procédure il suffit d'être excellent sur un critère et bon sur les autres pour appartenir à la catégorie "Excellent". Si ce n'est pas le cas, il suffit de fournir un niveau de bien-être bon sur deux critères et normal sur les deux autres pour appartenir à la catégorie "Bon". Nous durcissons encore un peu plus la règle pour la comparaison au profil z_3 puisque nous n'acceptons qu'une seule performance inférieure à z_3 pour affecter une ferme à la catégorie *"Normal"*. Ainsi, si pour un seul des critères la valeur obtenue par la ferme est inférieure à z_4, alors la ferme sera affectée à la catégorie *"Non classé"*.

7.3 Comparaison de l'impact des quatre procédures sur l'évaluation globale

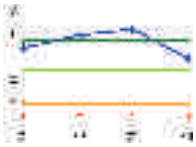
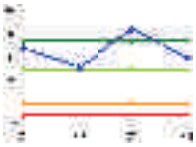
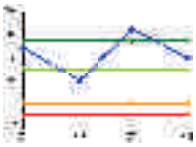
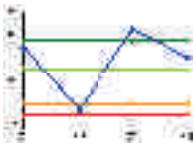
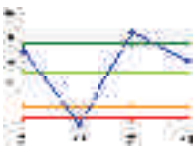
7.3.1 Comparaison sur un jeu de fermes virtuelles

Le choix de la procédure d'affectation joue un rôle direct sur l'évaluation finale donnée aux fermes à évaluer. Pour en voir mieux les effets, nous avons tester ces quatre procédures sur un jeu de cinq fermes virtuelles qui ont les mêmes évaluations sur les critères f_1, f_3 et f_4, mais qui diffèrent sur le critère f_2 (cf. les quatre premières colonnes du Tableau 7.3). En fonction du score obtenu par les fermes sur f_2 et de la procédure choisie pour définir les règles d'affectation aux différentes catégories, les fermes peuvent être affectées aussi bien à la catégorie *"Excellent"* qu'à la catégorie *"Non classé"* (cf. Tableau 7.3).

7.3.2 Comparaison sur le jeu de fermes réelles évaluées lors du WP2.4

Sur les 91 fermes visitées lors du WP2.4, seules 70 sont exploitables au niveau final de l'agrégation (du fait de données manquantes, certains sous-critères ne peuvent pas être évalués pour certaines fermes et en conséquence certains critères ne le peuvent pas non plus). Appliquées à ces 70 élevages, les quatre procédures conduisent à des distributions très différentes (cf. Figures 7.4, 7.5, 7.6 et 7.7). Cependant, aucune de ces quatre procédures ne permet d'obtenir des fermes dans la catégorie *"Excellent"* et au minimum 56% des fermes se retrouvent dans la catégorie *"Non classé"*. Du fait des scores faibles obtenus par les fermes sur certains critères, ces distributions ne sont pas surprenantes. Ces résultats, qui méritent d'être confortés sur un plus large échantillon de fermes, alertent toutefois sur les risques de règles d'agrégation très strictes, comme la règle d'unanimité.

TAB. 7.3 – Conséquences des différentes procédures sur l'affectation des catégories

	Fermes			**Procédures**			
Noms	Scores de critères			**1**	**2**	**3**	**4**
X_1	f_1	69		Bon	Bon	Excellent	Excellent
	f_2	**80**					
	f_3	85					
	f_4	60					
X_2	f_1	69		Bon	Bon	Bon	Excellent
	f_2	**52**					
	f_3	85					
	f_4	60					
X_3	f_1	69		Normal	Bon	Bon	Bon
	f_2	**40**					
	f_3	85					
	f_4	60					
X_4	f_1	69		Non classé	Normal	Non classé	Normal
	f_2	**15**					
	f_3	85					
	f_4	60					
X_5	f_1	69		Non classé	Non classé	Non classé	Non classé
	f_2	**5**					
	f_3	85					
	f_4	60					

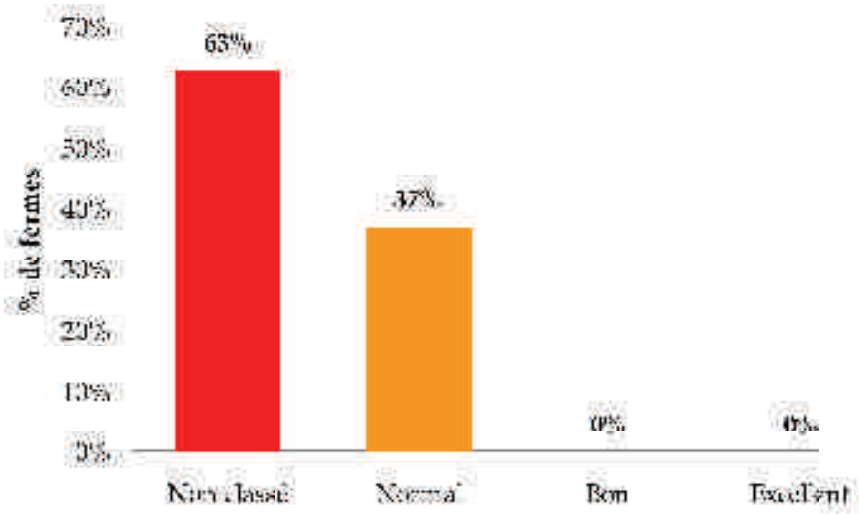


FIG. 7.4 – Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 1**

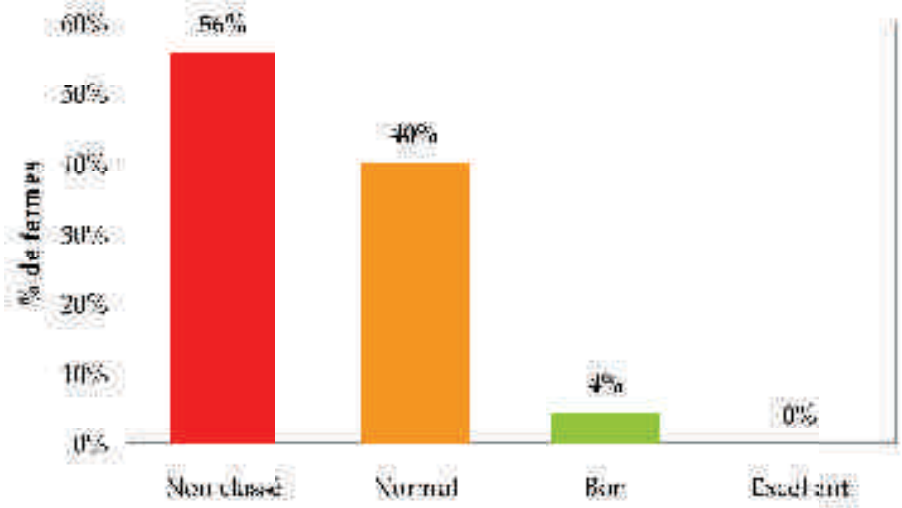


FIG. 7.5 – Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 2**

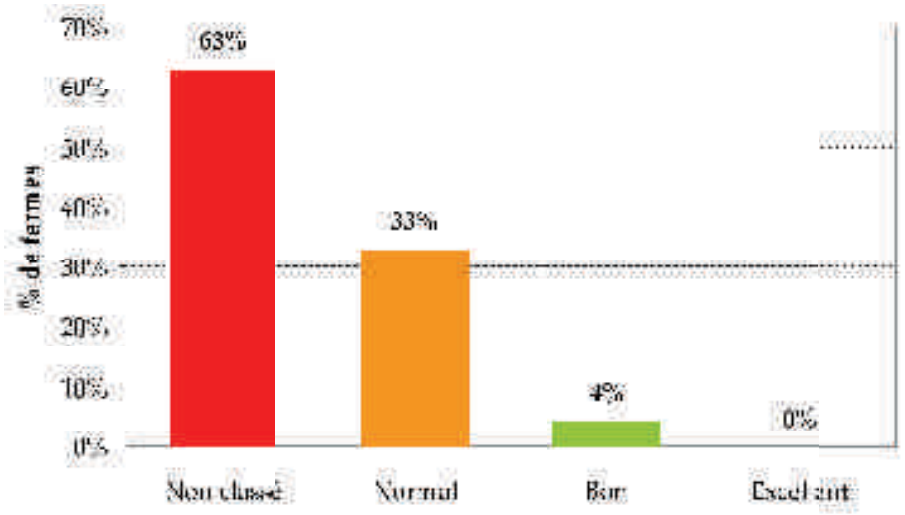


FIG. 7.6 – Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 3**

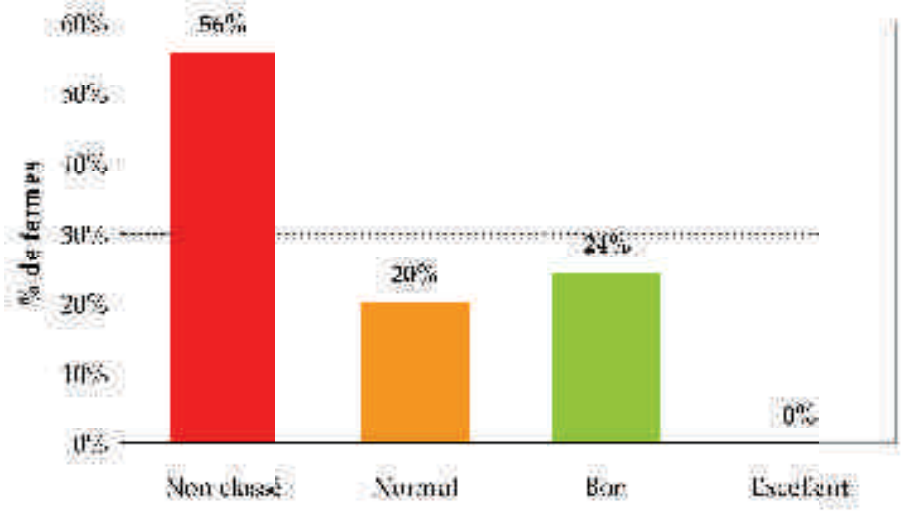


FIG. 7.7 – Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 4**

Il est toutefois possible d'obtenir une distribution des fermes convenable, dans laquelle chaque catégorie a des fermes qui lui sont affectées, en obtenant relativement peu de fermes dans les catégories *"Non classé"* et *"Excellent"*, et avec une majorité de fermes dans la catégorie *"Normal"*.

En effet, il est possible d'obtenir une telle distribution des fermes (cf. Figure 7.8) en suivant la Procédure 4 d'affectation aux catégories de bien-être, et en se servant des profils de référence présentés dans le Tableau 7.4.

TAB. 7.4 – Valeurs des profils de référence z_i

	f_1	f_2	f_3	f_4
z_1	65	65	55	65
z_2	40	40	30	35
z_3	6	10	10	10
z_4	3	5	5	5

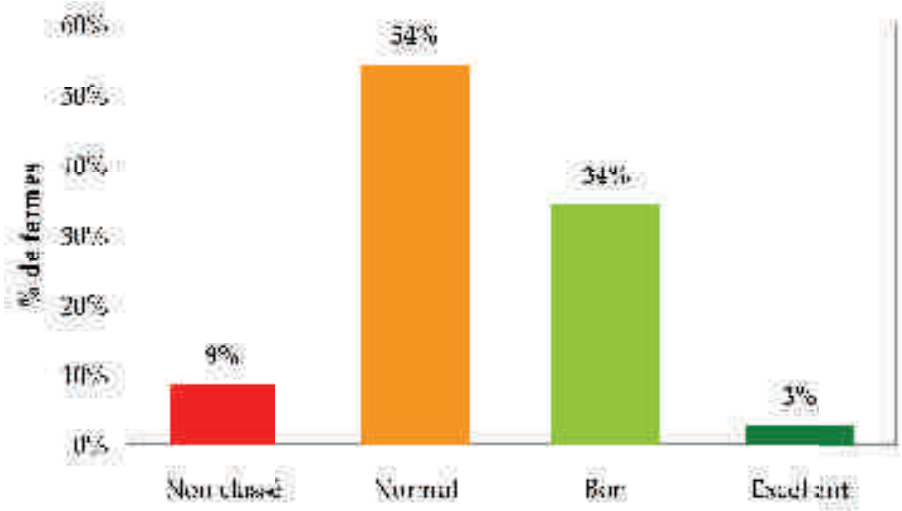


FIG. 7.8 – Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 4** et des profils de références modifiés

Au vu de ces résultats, nous conseillons l'utilisation de la Procédure 4 qui, bien qu'étant la plus souple, permet d'introduire un gradient de sévérité croissant avec i, caractéristique qui semble être demandée par les experts et les utilisateurs potentiels du système d'évaluation.

Toutefois, lors du paramétrage du modèle final, ces quatre procédures seront présentées aux utilisateurs, et pourront être accompagnées de procédures additionnelles.

Pour faciliter la présentation de ces différentes procédures auprès des experts consultés, nous proposons d'utiliser un fichier interactif[12] permettant de visualiser en direct les modifications apportées à la répartition des fermes visitées (dans le cadre du WP2.4, et peut-être aussi du WP4.6) en fonction des procédures et du paramétrage (valeurs des profils, seuil d'indifférence...). Il pourrait être fructueux de visualiser cette distribution non seulement pour les fermes laitières mais aussi pour les autres productions couvertes par le projet Welfare Quality® afin de décider de règles valables quel que soit le type de fermes. Cette visualisation immédiate de la distribution des fermes permet aux experts de voir si la répartition obtenue correspond à leurs attentes. Par exemple, on peut supposer que se retrouver avec 50% des fermes dans la catégorie *"Non classé"* ne sera pas jugé acceptable par les professionnels de l'élevage et les institutions européennes car ils estiment que la plupart des fermes européennes assurent déjà un niveau de bien-être acceptable, en particulier du fait des règlements européens de protection animale. Nous avons toutefois vu qu'il est possible en jouant sur les paramètres de la construction de modifier la répartition des fermes. Il serait donc possible de revoir les paramètres pour obtenir une distribution des élevages correspondant aux attentes des utilisateurs potentiels.

[12]un tel fichier a d'ores et déjà été créé sous *Excel* et utilisé lors d'une réunion avec des scientifiques impliqués dans Welfare Quality®

MULTICRITERIA EVALUATION OF ANIMAL WELFARE AT FARM LEVEL: AN APPLICATION OF MCDA METHODOLOGIES

Raphaëlle BOTREAU *, Jacques CAPDEVILLE †, Patrice PERNY ‡, Isabelle VEISSIER *

Abstract. In this paper we present a contribution of Multicriteria Decision Analysis to the evaluation of Animal Welfare in European farms. We present the main features of the overall evaluation process designed to assess animal welfare on a farm from measures taken on the animals or their environment. We review the main steps of the construction of subcriteria and criteria. Then we discuss the aggregation problem and the design of a multicriteria sorting procedure to assess overall welfare from those criteria. This study is illustrated on cattle but is designed to be transposed to other animal species (e.g. pigs, poultry) so as to provide a general evaluation system for animal welfare on farms.

Keywords: Multicriteria Analysis, Choquet Integral, Multicriteria Sorting methods, Animal Welfare.

1 Introduction

In occidental societies, consumers are more and more concerned about the living conditions of farm animals from whom they consume the products [7]. As a consequence, trade groups (producers, processors, retailers, and chain restaurants) have included animal welfare in their commercial strategies, with the development of certification schemes (e.g. Integrale Keten Beheersing in the Netherlands, Swedish Broiler Control, Filières Qualité Carrefour in France, McDonald's Europe, RSPCA Freedom Food in the UK). These initiatives aim at providing consumers with relevant information to facilitate the choice of "high welfare standard" products.

Actually, the assessment of welfare varies from one scheme to another, due to the lack of standard for animal welfare assessment at farm level, and for product

*INRA, UR1213 Herbivores, Site de Theix, Saint-Genès-Champanelle F-63122, France
†Institut de lElevage, BP18, Castanet Tolosan F-31321, France
‡LIP6, University of Paris 6, 104 avenue du Président Kennedy, Paris F-75016, France

information on animal welfare to consumers. This statemement was at the origin of the EU project *Welfare Quality*®, the aim of which is to propose an overall assessment system of the welfare of cattle, pigs and poultry, that is scientifically grounded. This system could provide EU with a clear view on the current status of animal welfare in European countries, help farmers to improve their current practices with a monitoring tool, and provides consumers with new insights on product quality linked to animal welfare.

The elaboration of an overall assessment system requires the definition of a formal evaluation model that will capture the main aspects of the problem. As argued in [4], formal models make explicit and, to a large extent, unambiguous representations of a given problem. The modeling process requires to make a substantial effort to formulate the problem, the overall objective, to decide which type of information should be included in the process, what is important and what is not. This facilitates the communication among the various actors (scientists experts in animal welfare, practitioners, stakeholders) of the decision process. This effort can only be beneficial as it forces people to think harder and deeper about the problem, about the current state of affairs, about ways of improvement. Once a formal model has been established, a battery of formal techniques (often implemented on a computer) become available for drawing any kind of conclusion and provide explanations. Moreover, formal techniques, as explicit decision procedures, lend themselves easily to criticism and might be discussed, modified, improved. In this respect, this paper present the main features of the multicriteria evaluation procedure we have proposed, within the Welfare Quality® project, to assess welfare status of animals on farms, with the example of dairy cows' welfare assessment. The paper is organized as follows: Section 2 presents the structure of the evaluation process from the measures taken on farms to the overall evaluation. Section 3 discusses the definition of subcriteria from initial measures. Section 4 presents the construction of criteria from subcriteria. Finally, Section 5 gives the main lines of the multicriteria sorting procedure used to assess welfare on farms by comparison to reference profiles.

2 Animal Welfare: a multidimensional concept

It is widely acknowledged that animal welfare is a multidimensional concept [16, 9]. It includes multiple aspects like health, comfort, expression of natural behaviour, and so on. These aspects are unlikely to compensate for each other. This point has been more extensively discussed in [2]. Hence the overall assessment of the welfare of animals corresponds to a multicriteria evaluation problem that needs a constructive strategy to compound information collected on farms.

In its advice to the UK government on welfare requirements for farm animals, the Farm Animal Welfare Council [8] identified five freedoms: freedom from hunger and thirst; freedom from discomfort; freedom from pain, injury and disease; freedom to express normal behaviour; and freedom from fear and distress. The five freedoms are widely used either in certification schemes (e.g. Freedom Food scheme [15]), in codes of practices or regulations, e.g. EU legislation to protect animals. Several freedoms

overlap each other, e.g. when natural lying behaviour cannot be performed properly the animals may feel discomfort and even be injured. We thus proposed a list of welfare criteria that differs slightly from the five freedoms [3]. These criteria are "good feeding", "good housing", "good health", and "appropriate behaviour". With these criteria, we want to check respectively if the animals do not suffer from being improperly fed, if they do not suffer from being inappropriately housed, if they are maintained in good health and then free from pain, and if their behaviour reflects optimised emotional states. Each of the four criteria is independent from the others and corresponds to a specific management factor on a farm, having clear implications for farmers' practices and strategic choices.

We then defined the various aspects that should enter in the definition of these welfare criteria. This led us to consider several subcriteria for each criterion:

- **Good feeding** is based on *absence of prolonged hunger* (Subcriterion 1) and *thirst* (Subcriterion 2);
- **Good housing** is based on *comfort around resting* (Subcriterion 3), *thermal comfort* (Subcriterion 4), and *ease of movement* (Subcriterion 5);
- **Good health** is based on *absence of injuries* (Subcriterion 6), the *absence of diseases* (Subcriterion 7), and the *absence of pain due to management procedures* (Subcriterion 8);
- **Appropriate behaviour** is based on *expression of social behaviour* (Subcriterion 9), the *expression of behaviours other than social behaviour* (Subcriterion 10), a *good human-animal relationship* (Subcriterion 11), and the *absence of general fear* (Subcriterion 12).

Each subcriterion will be assessed from specific measures taken on the farm. In the Welfare Quality® project, it was decided to take measures directly on animals rather than on their environment, so as to produce evaluations closed to what animals may feel. The measures consist essentially of clinical observations (body condition, symptoms of diseases, injuries) and behavioral observations (lying postures and movements, social interactions, responses to humans...). When no animal-based measure is available to elaborate a subcriterion, measures are taken on the resources or the management of the farm (e.g. provision of water, ambiance in a barn, access to an exercise area, procedures used for mutilations if any). After discussion between experts involved in the project, 32 measures have been developed and validated by animal scientists to assess the welfare of dairy cows. As a consequence the evaluatin procedure will follow a progressive aggregation process going from the 32 measures to the 12 subcriteria, then from these subcriteria to the 4 criteria, and finally to one overall assessment (cf. Figure 1).

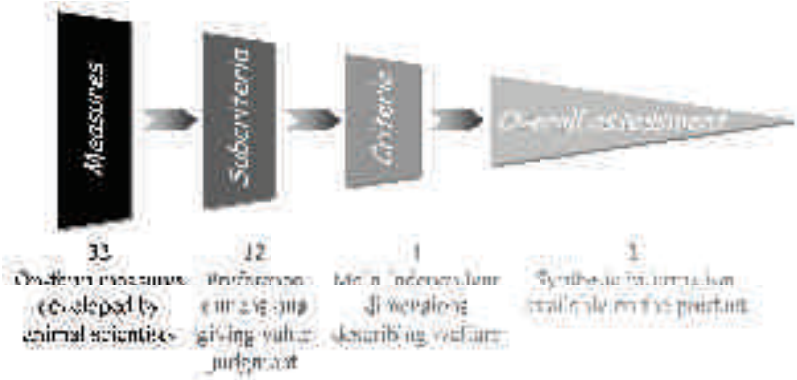


Figure 1: Evaluation structure for the assessment of animal welfare

3 Subcriteria construction

3.1 General methodology

Thirty-two measures are collected on farms to assess cow welfare, leading to raw data expressed on very different scales. Some measures are expressed on cardinal scales, as for example the frequency of aggressive behaviour between animals during a given period, or the flight distance of an animal when a human approaches (i.e. the distance between the animal and the approaching observer at the time the animal starts moving away). Other measures are expressed on ordinal scales. For instance, an animal can have *no* injury, *mild* injuries (at least one mild injury, like an hair patch, but no serious injury) or *severe* injuries (at least one serious injury, like an open lesion). These raw data obtained on farms need to be interpreted in terms of animal welfare, in order to be used in the construction of the subcriteria. It must be clear that the relation between a measure and its interpretation in terms of animal welfare is often not proportional. For instance, two animals that flee when the experimenter is 8 m and 6 m away will be considered similar in terms of fear of humans. By contrast, an animal that flees when the experimenter is 2 m away will be considered more frightened by humans than an animal that accepts being touched (i.e. flight distance of 0 m). Yet in both cases, the difference in flight distance is actually the same (2 m). Hence if raw data are to be converted onto a value scale, from 'poor welfare' to 'good welfare', the conversion may not necessarily follow a linear function.

Each subcriterion is composed of 1 to 9 measures (cf. Table 1). As it is shown in Table 1, there is no measure yet to inform the subcriteria 4 (*Thermal comfort*), 10 (*Expression of other behaviour*), and 12 (*Absence of general fear*). We nevertheless kept the 12 subcriteria in our construction because some measures are still under development in the project Welfare Quality® and we expect that appropriate measures will be available in the near future for all these Subcriteria.

Table 1: Set of criteria, subcriteria, and measures

Criteria	Subcriteria	Measures
Good feeding	1. Absence of prolonged hunger	Body condition score (percentage of very lean animals)
	2. Absence of prolonged thirst	Water supply (number of functioning water bowls, cleanliness, number of water bowls accessible by an animal)
Good housing	3. Comfort around resting	Time needed to lie down Percentage of animals colliding with housing equipment during lying down Percentage of animals lying with hindquarter on edge Cleanliness scores (udder, flank and upper legs, lower legs)
	4. Thermal comfort	-
	5. Ease of movement	Presence of tethering Access to outdoor loafing Access to pasture
Good health	6. Absence of injuries	Lameness score (lameness prevalence) Integument alterations (hairless patches and lesions/swellings)
	7. Absence of diseases	Respiratory disorders (coughing, sneezing, nasal discharge, ocular discharge, increased respiratory rate) Enteric disorders (diarrhoea) Reproductive disorders (mastitis, vulvar discharge) Mortality
	8. Absence of pain due to management procedure	Routine mutilations (for dehorning and tail docking: procedure, age, use of anaesthetics / analgesics)
Appropriate behaviour	9. Expression of social behaviours	Incidence of agonistic behaviours (head butts and other behaviours)
	10. Expression of other behaviours	-
	11. Good human-animal relationship	Avoidance distance at the feeding place
	12. Absence of general fear	-

The use of measures involved in the definition of each subcriterion have been discussed and decided after discussion with animal scientists. Although this is not the case in general, some measures might enter in the definition of several subcriteria. For example, body condition score may be relevant to assess *hunger* (when many animals are lean on a farm, this is probably due to underfeeding) but also to assess *diseases* (a diseased animal may loose weight). In this example, the way body condition score is interpreted is the same for hunger (Subcriterion 1) and diseases (Subcriterion 7). In order to avoid double-counting body condition score, we decided to allocate this measure only to Subcriterion 1 (hunger), and not to Subcriterion 7 (diseases), because the most probable cause of high percentage of lean animals in a herd is the lack of appropriate food and because Subcriterion 7 is informed by many other measures.

In the construction of subcriteria, the raw data may be expressed on different scales and may have different contributions. As a consequence the construction process is different from one subcriterion to another, adapting the aggregation method according to the number of measures, their nature, precision and relative importance. Several methods are thus used such as lexicographic valuation trees, weighted sums, or comparisons to thresholds (cf. Table 2). For each subcriterion, we use the same absolute evaluation scale, ranging from 0 to 100, with the following conventions:

- **0** corresponds to the worst situation one can find on a farm (i.e. the situation below which it is considered there cannot be further decrements in welfare),
- **50** corresponds to a neutral situation, the level of welfare is not too bad but not good,
- **100** corresponds to the best situation one can find on a farm (i.e. the situation above which it is considered there cannot be further improvements in welfare).

Hence, the 12 subcriteria listed in Table 1 will be formally defined as functions $g_j : X \longrightarrow [0, 100],\ j = 1, \ldots, 12$ where X is the set of farms to be evaluated.

In order to interpret correctly the raw data in terms of welfare, to define reference thresholds and the relative importance of the different measures (when necessary), and finally to construct appropriate utility functions, we consulted animal scientists. At that stage of the construction, we indeed need to ask experts who perfectly know the measures, how they are collected on farms and how they can be interpreted in terms of welfare. Actually, during this first calibration stage, we have consulted five or six animal scientists (depending on the subcriterion) that were involved in the choice and development of measures in Welfare Quality®. In the future, when these measures will be used in routine observations by practitioners, it will be possible to involve more experts and recalibrate the construction.

Table 2: Types of construction of the welfare subcriteria

SC	*# meas.*	*Type of data*	Construction
g_1	*1*	*Binary measure at cow level (very lean vs. normal)*	% of very lean cows transformed into score using least-squares B-splines curve fitting
g_2	*3*	*Binary measures at farm level*	Lexicographic method leading to 4 possible cases, scores directly attributed to each of these 4 cases
g_3	*6*	*Cardinal or ordinal measures on resting and cleanliness, at cow level*	Each measure is transformed onto an ordinal scale: no/moderate/severe problem Calculation of an index based on a linear combination of the proportions of moderate & severe problems, considering resting behaviour 3 times as important as cleanliness Index transformed using least-squares B-splines curve fitting
g_5	*3*	*Binary measures at farm level*	Construction of a decision tree leading to 5 possible cases, scores directly attributed to each of these 5 cases
g_6	*2*	*Ordinal measures at cow level*	Calculation of 2 indices (1 for tegument alterations & 1 for lameness) each based on a linear combination of the % of cows presented mild *vs.* severe problems, transformed using least-squares B-splines curve fitting. These two sub-scores are then synthesized thanks to Choquet integral
g_7	*9*	*Cardinal measures at farm level, at different time scales*	Each measure is transformed onto an ordinal scale: no/moderate/severe problem Calculation of an index based on a linear combination of the proportions of moderate & severe problems Index transformed using least-squares B-splines curve fitting
g_8	*6*	*Binary or qualitative measures at farm level*	For each of the two types of mutilation (dehorning and tail docking), construction of a decision tree leading respectively to 13 and 9 possible cases. Scores are directly attributed to each of these possible cases Overall score = the worst score obtained on at least 10% of the animals
g_9	*2*	*Cardinal measures at farm level*	Calculation of an index based on a linear combination of the no. of head butts and the no. of other agonistic behaviours, transformed using least-squares B-splines curve fitting
g_{11}	*1*	*Cardinal measure at cow level*	For each animal observed on farm, its avoidance distance is transformed onto an ordinal scale: no/moderate/important/serious problem Calculation of an index based on a linear combination of the % of animals observed in each of the 4 levels of the scale Index transformed using least-squares B-splines curve fitting

3.2 Examples

To give an idea of techniques used for the construction of the subcriteria, we will consider two typical examples: Subcriterion 2 (Absence of prolonged thirst) and Subcriterion 6 (Absence of injuries).

3.2.1 Example 1: Subcriterion "Absence of prolonged thirst"

To assess the absence of prolonged thirst on a farm, three questions based on resources and management are used, respectively linked to the number of functioning drinkers, the drinker cleanness, and the availability of 2 drinkers per animal. These aspects can be considered hierarchically due to their respective dependence and importance. This led us to propose the evaluation tree pictured on Figure 2.

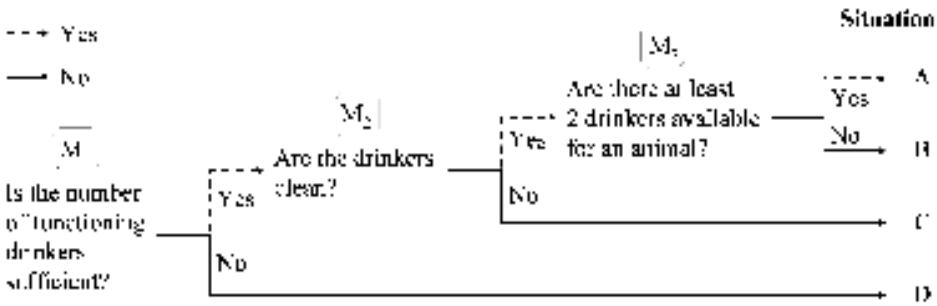


Figure 2: The tree used for Subcriterion *Absence of prolonged thirst*

The first point to be checked on a farm is the number of functioning drinkers and this number is compared to a recommendation for the size of a group of cows. The second point is the cleanliness of the water. Finally, if the two precedent conditions are satisfied, it is better for the animals to have access to at least two drinkers: in case one drinker breaks down, the cows still have access to the second one. This decision tree leads to four possible situations, from A (the best one) to D (the worst one)(cf. Figure 2). We asked the experts to assign a score to each of the four possible situations. Each observed farm is assigned to one of the four possible situations thanks to a function $h : X \longrightarrow \{A, B, C, D\}$; where X is the set of farms to be evaluated. A utility function u is then used to transform the value $h(x)$ obtained by farm x into the subcriterion-score $g_2(x)$ defined as follows:

$$g_2(x) = u(h(x)) \tag{1}$$

For each situation z, $u(z)$ corresponds to the mean of evaluations given by experts to situation z (cf. Table 3).

Table 3: Experts' evaluations for subcriterion 2

Situation			**Score**			**Mean**
z	E1	E2	E3	E4	E5	$u(z)$
A	100	100	100	100	100	**100**
B	70	50	60	70	50	**60**
C	30	40	50	20	20	**32**
D	0	0	15	0	0	**3**

3.2.2 Example 2: Subcriterion "Absence of injuries"

Subcriterion 6 (*Absence of injuries*) is assessed through two measures observed at individual level: *alterations of the tegument* and *lameness*. For each measure, a sample of cows, representative of the herd, is inspected. The animals inspected for lameness may not be the same as the ones inspected for alterations of their tegument. Thus, we calculate two independent grades at herd level, one for *alterations of the tegument* and one for *lameness* (respectively $g_6^1(x)$ and $g_6^2(x)$). These grades are then combined so as to get the subcriterion-score $g_6(x)$.

The constructions followed for the two grades are very similar. Let us present briefly the one followed for "*alterations of the tegument*". At cow level, three categories are defined (cf. Table 4), corresponding to different levels of problems according to welfare (i.e. absence of problem, moderate problem, severe problem). Hence, the data we had to deal with are ordinal at cow level.

Table 4: Evaluation of alterations of the tegument at cow level

3 categories are defined at cow level:		Problem in respect to welfare:
0	The cow does not have any injury	No problem
1	The cow presents at least one mild injury (e.g. an hair patch), *but no serious injury*	Moderate
2	The cow presents at least on serious injury (e.g. open lesion or swelling)	Severe

At farm level we take into account the percentages of cows in each of the categories 0, 1, and 2, denoted respectively $p_0(x)$, $p_1(x)$, and $p_2(x)$ for farm x. On the basis of expert's opinions, it has been decided that a linear combination of these percentages was able to rank farms by decreasing order of preference. Hence, we decided to calculate an index $h_1(x)$ for all farm $x \in X$, defined as follows:

$$h_1(x) = 100 - \frac{\sum_{k=0}^{2} w_k \, p_k(x)}{w_2} \tag{2}$$

This formal definition relies on the fact that $p_0(x) + p_1(x) + p_2(x) = 1$ for all $x \in X$.

Moreover, this index is defined so as to yield a result in [0,100], where 0 corresponds to the worst situation and 100 to the best one. Thus, denoting x_* the worst possible farm (i.e. characterized by $p_2(x_*) = 100$) and x^* the best possible farm (i.e. characterized by $p_0(x^*) = 100$), the index $h_1(x)$ is constructed so that $h_1(x^*) = 100$ and $h_1(x_*) = 0$ (which explains why we divide by w_2 the weighted sum of $p_k(x)$).

To set the weights $w_k,\ k = 0, \ldots, 2$, we asked the experts to react on a virtual dataset composed by 11 reference farms. They were requested to rank these farms from the worst to the best. At this stage, experts usually provide very similar ranks even when some minor differences might be observed. The weights are then derived by ordinal regression using a solver. Here the obtained weights are w_0=0, w_1=1, and w_2=5.

Then experts where asked to provide an absolute score representing their own evaluation of $g_6^1(x)$ for each reference farm x, using the standardized scale $[0, 100]$ used to express performances on all subcriteria. The results obtained are given in the following table:

Table 5: Experts' evaluations for *alterations of the tegument*

	Situation			**Scores given by experts Ei**						**Index**	**Score**
	$p_0(x)$	$p_1(x)$	$p_2(x)$	E1	E2	E3	E4	E5	*Mean*	$h_1(x)$	$g_6^1(x)$
weights	*0*	*1*	*5*								
Farm 3	**0**	**0**	**100**	0	0	0	0	0	*0*	**0**	**0**
Farm 9	**0**	**25**	**75**	3	1	5	5	5	*4*	**20**	**4**
Farm 5	**0**	**50**	**50**	7	2	10	10	10	*8*	**40**	**7**
Farm 6	**50**	**0**	**50**	15	3	13	15	15	*12*	**50**	**13**
Farm 8	**20**	**50**	**30**	20	10	18	20	15	*17*	**60**	**19**
Farm 7	**50**	**25**	**25**	25	4	22	25	25	*20*	**70**	**19**
Farm 11	**75**	**0**	**25**	30	5	25	30	25	*23*	**75**	**22**
Farm 2	**0**	**100**	**0**	35	15	60	50	20	*36*	**80**	**29**
Farm 4	**50**	**50**	**0**	50	20	70	70	30	*48*	**90**	**59**
Farm 10	**75**	**25**	**0**	75	75	85	80	50	*73*	**95**	**79**
Farm 1	**100**	**0**	**0**	100	100	100	100	100	*100*	**100**	**100**

Although index $h_1(x)$ defined as a linear aggregation of percentages p_0, p_1, p_2 fully explains the ranks attached to farms by experts, Table 5 shows that the h_1 index is not sufficient to produce an absolute performance value matching prescribed values on reference farms. It is therefore necessary to resort to a non-linear utility function u_1 in order to transform index $h_1(x)$ into subcriterion-score $g_6^1(x)$ using the following equation:

$$g_6^1(x) = u_1\left(h_1(x)\right) \tag{3}$$

Function $u_1(x)$ has to satisfy boundary constraints $u_1(100) = 100$ and $u_1(0) = 0$. Moreover, it must be compatible with the ranking induced by $h_1(x)$ on reference farms. Finally u_1 should reasonably approximate the prescribed evaluations provided by experts (see Table 5). Due to the necessity of getting a non linear utility function, u_1 has been defined as a cubic B-spline function. The construction of u_1 is based on

least square fitting of experts' opinion using B-spline functions [1, 20]. We search a polygon of control composed of equidistant points, the smoothing of which generates of spline function that best minimizes δ, the sum of square deviations between the model and the prescribed scores given by experts, which formally writes:

$$\delta = \sum_{p=1}^{n_e} w_p \left(\sum_{k=1}^{n_f} (u_1(h_1(x_k)) - y_p(x_k))^2 \right) \tag{4}$$

where n_e is the number of experts (here $n_e = 5$) and n_f is the number of farms (here $n_f = 11$). Reference farms are denoted $x_1, \ldots, x_{11}$ and $y_p(x_k)$ is the value of farm x_k seen by expert p for "alterations of the tegument". The quantity $u_1(h_1(x_k))$ is the score calculated on farm x_k by the model for "alterations of the tegument" and w_p is the weight assigned to expert p (for the moment $w_p = 1$ for all p). By assigning weights to experts (w_p to expert p), it is possible to attach more or less importance to the opinion of the different experts depending of their confidence or level of expertise in the evaluation process. For the moment, we assigned the same importance to all the experts.

To smooth the polygon of control we used a cubic spline (cf. equations (5) and (6)) generated by combinations of four B-splines functions defined as follows:

$$\begin{aligned} b_{-1}(\lambda) &= \frac{1}{6}(1 - 3\lambda + 3\lambda^2 - \lambda^3) \\ b_0(\lambda) &= \frac{1}{6}(4 - 6\lambda^2 + 3\lambda^3) \\ b_1(\lambda) &= \frac{1}{6}(1 + 3\lambda + 3\lambda^2 - 3\lambda^3) \\ b_2(\lambda) &= \frac{1}{6}(\lambda^3) \end{aligned} \tag{5}$$

They are constructed so that:

$$\forall \lambda \in [0, 1], \quad \sum_{i=-1}^{2} b_i(\lambda) = 1 \tag{6}$$

We have learned here eight control points (cf. Table 6) using a solver to minimize the overall error δ while constraining the curve to go through points (0,0) and (100,100). This yields the following control points:

Table 6: Points of the polygon of control plotted on Figure 3

x (index)	-20	0	20	40	60	80	100	120
y (utility)	0.5	-1.6	6.0	2.3	24.3	11.7	103.0	176.3

We thus obtained a curve composed of five pieces (polynomials of degree 3), each piece being defined as a convex combination of control vertices using the following

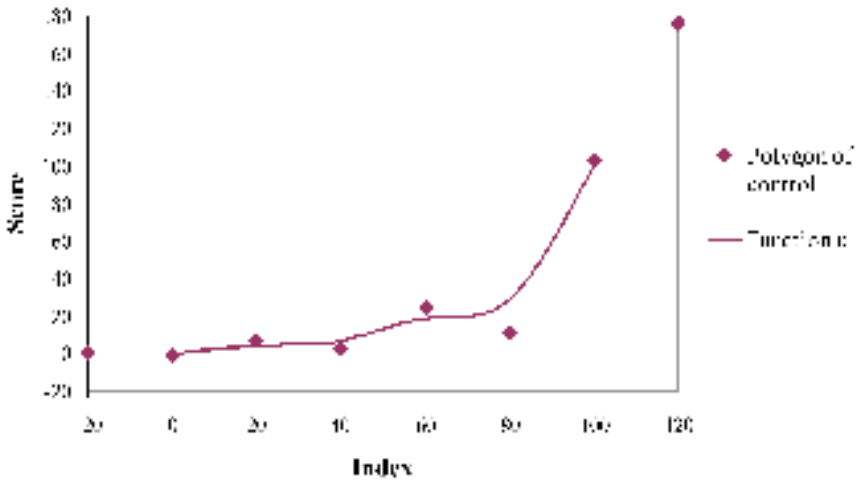


Figure 3: Polygon of control and the resulting spline function u_1

B-splines coefficients:

$$u_1(z) = \alpha_z\, b_{-1}(\bar{z}) + \beta_z\, b_0(\bar{z}) + \gamma_z\, b_1(\bar{z}) + \delta_z\, b_2(\bar{z}) \tag{7}$$

where $\bar{z} = \frac{z-a}{b-a}$ for all $z \in [a, b]$. The coefficients α_z, β_z, γ_z, and δ_z acting as control vertices vary with z; the values used to generate curve u_1 are presented in Table 7.

Table 7: Control vertices for spline function u_1

z	α_z	β_z	γ_z	δ_z
[0, 20]	0.5	-1.6	6.0	2.3
[20, 40]	-1.6	6.0	2.3	24.3
[40, 60]	6.0	2.3	24.3	11.7
[60, 80]	2.3	24.3	11.7	103.0
[80,100]	24.3	11.7	103.0	176.3

With the determination of this non-linear function u_1 (cf. Figure 4), it is now possible to calculate the "alterations of the tegument" subscore obtained by any new farm, from data collected on that farm. Indeed, for any new farm x, it is sufficient to compute the index $h_1(x)$ from data $p_0(x), p_1(x), p_2(x)$, and then to derive $g_6^1(x)$ using Equation (3).

The calculation of $g_6^1(x)$ using Equation (3) can also be performed on reference farms $x_1, \ldots, x_{11}$. The results are provided in the right column of Table 5 and pictured on Figure 4. As can be seen, the model approximate very well the reference values provided by expert. Further experimentations on other samples have confirmed the adequacy of the construction.

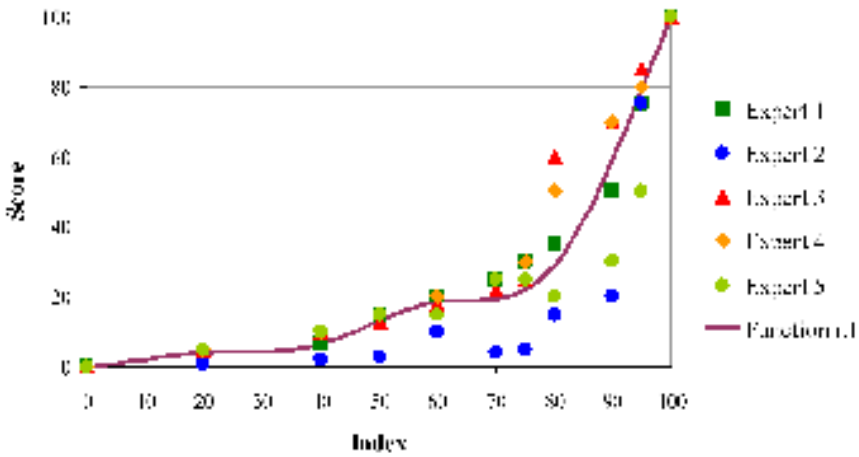


Figure 4: SC6: approximation using spline utility function u_1

The evaluation of "lameness" at farm level is essentially similar to the one described for "alterations of the tegument". Lameness is assessed on a three levels ordinal scale (cf. Table 8), corresponding respectively to three categories of problem concerning the respect of welfare (absence of problem, moderate problem, severe problem).

Table 8: Evaluation of lameness at cow level

3 categories are defined at cow level:		Problem in respect to welfare:
0	The cow is **not lame**	No problem
1	The cow is **lame**: imperfect temporal rhythm in stride creating a limp (irregular foot fall - not equal timing between hoof-beats, weight not borne for equal time on each of the four feet)	Moderate
3	The cow is **severly lame**: extreme asymmetry in gait, or more than one limb affected	Severe

At farm level, a farm x is characterized by its percentages of animals within each of the three categories (absence of problem, moderate problem, severe problem), respectively denoted $p_0(x)$, $p_1(x)$, and $p_2(x)$. Following the same process of construction as for *alterations of the tegument*, we are able to calculate, for any new farm x, a sub-score $g_6^2(x)$ for *lameness* (cf. equation 8).

$$g_6^2(x) = u_2\left(h_2(x)\right) = u_2\left(100 - \frac{\sum_{k=0}^{2} w_k\, p_k(x)}{w_2}\right) \tag{8}$$

To define subcriterion-score $g_6(x)$ giving a grade to farm x with respect to the *Absence of injuries*, we have to aggregate the sub-scores obtained for each of the two measures observed on farm x (*alterations of the tegument* and *lameness*, respectively $g_6^1(x)$ and $g_6^2(x)$), defining a function of aggregation $\varphi : [0,100] \longrightarrow [0,100]$ (cf. equation 9).

$$g_6(x) = \varphi\left(g_6^1(x), g_6^2(x)\right) \tag{9}$$

We asked experts to react on a dataset composed of virtual farms characterized by their sub-scores on lameness and alterations of the tegument. In order to identify their reasoning, we asked them to give to each farm a score for subcriterion *Absence of injuries*. A precise analysis of the results led us to use a Choquet integral for φ in Equation (9) so as to calculate the subcriterion score g_6. Test on data samples have shown that the model proposed performs well in practice. As an illustration, Table 9 compares the values predicted by the model to experts' evaluations on a small sample of typical examples. More details justifying the choice of the Choquet integral for φ will be presented in the next section dedicated to the construction of criteria.

Table 9: A sample of experts' answers for subcriterion 6 *Absence of injuries*

Farm x	$g_6^1(x)$	$g_6^1(x)$	E1	E2	E3	E4	**Score** $g_6(x)$
x	**40**	**60**	50	50	50	55	**51**
y	**50**	**50**	50	50	50	50	**50**
z	**60**	**40**	40	50	50	45	**46**

4 Criteria construction

In the previous section, we have defined 12 subcriteria as functions $g_j : X \longrightarrow [0,100],\ j = 1,\ldots,12$ where X is the set of farms to be evaluated. Note that we used the same evaluation scale $[0,100]$ for all subcriteria. We now have to define formally the criteria measuring good feeding, good housing, good health and appropriate behaviour by aggregation of some of the 12 subcriteria. More precisely, considering Table 1, the following criterion functions $f_i : X \longrightarrow [0,100]$ need to be defined:

- Criterion f_1 (Good feeding) is defined by aggregation of g_1, g_2, which formally writes $f_1(x) = \chi_1(g_1(x), g_2(x))$
- Criterion f_2 (Good housing) is defined by aggregation of g_3, g_4, g_5, which formally writes $f_2(x) = \chi_2(g_3(x), g_4(x), g_5(x))$
- Criterion f_3 (Good health) is defined by aggregation of g_6, g_7, g_8, which formally writes $f_3(x) = \chi_3(g_6(x), g_7(x), g_8(x))$
- Criterion f_4 (Appropriate behaviour) is defined by aggregation of $g_9, g_{10}, g_{11}, g_{12}$, which formally writes $f_4(x) = \chi_4(g_9(x), g_{10}(x), g_{11}(x), g_{12}(x))$

4.1 General principles

Here $f_i(x)$ measures the performance of farm x with respect to criterion f_i, the latter being defined from q subcriteria ($q = 2$, 3 or 4, depending on the criterion considered). We have now to determine the aggregation functions $\chi_i : [0, 100] \longrightarrow [0, 100]$, $i = 1, \ldots, 4$. Such aggregation functions are used to average, in some sense, the performances obtained on the contributing subcriteria. As compromise aggregation operators, they are expected to fulfill the following axioms:

$$\begin{array}{ll} Compensation: & \min(z_1, \ldots, z_q) \leq \chi_i(z_1, \ldots, z_q) \leq \max(z_1, \ldots, z_q) \\ Monotony: & (\forall j = 1, \ldots, q,\ z_j \leq y_j) \Rightarrow \chi_i(z_1, \ldots, z_q) \leq \chi_i(y_1, \ldots, y_q) \end{array}$$

Two desired characteristics of the construction rapidly appeared from the discussions with the experts. On the one hand, it seems

important to keep the possibility to take into account the relative importance attached to the different subcriteria. For example, looking at Criterion "Good feeding", some experts consider that Subcriterion 2 (thirst) is more important than Subcriterion 1 (hunger).

In that case, to reflect this expert opinion, the method chosen to aggregate g_1 and g_2 should allow to put more weight on g_2. Hence we might think to resort to a weighted quasi-arithmetic mean. Within this family, the most common aggregator is the *weighted sum* defined by:

$$WS(z_1, \ldots, z_q) = \sum_{j=1}^{q} w_j z_j \tag{10}$$

where $z_j = g_j(x)$ is the score obtained by farm x on Subcriterion j, and w_j are positive coefficients adding up to 1.

On the other hand, it appears that the experts may also want to focus on worst scores, which have a particular importance. Thus, they have proposed to limit the possibility of compensating bad grades on some subcriteria by good grades on some others so as to encourage farmers to improve first the most problematic points observed on their farms. To reflect this principle while keeping good discrimination properties, a natural solution is to focus on the worst component z_k in the vector $(z_1, \ldots, z_q)$ and then to use the other components as secondary criteria. This leads to the leximin criterion, see e.g. [6]. One known problem with the leximin is overpessimism in the evaluation (the worst component can hide very different profiles on the other components). An alternative choice that gives a better control of the relative impact of good and bad evaluations is the *ordered weighted average* introduced by [23] and defined as follows:

$$OWA(z_1, ..., z_q) = \sum_{i=1}^{q} w_j z_{(j)} \tag{11}$$

where $\{(1), \ldots, (q)\}$ is a permutation of $\{1, \ldots, q\}$ such that $z_{(j-1)} \leq z_{(j)}$ for all $j = 2, ..., q$. To get the desired behaviour in the aggregation, it is sufficient to use weights defined so that lower is the score, higher is the weight. Formally it gives:

$w_j > w_{j+1},\ j = 1, \ldots, q-1$. Note that, contrary to the weighted sum, OWA does not allow to attach different weights to subcriteria due to the following symmetry property:

$$Symmetry: \ \forall \sigma \text{ permutation of } \{1, \ldots, q\},\ OWA(z_1, \ldots, z_q) = OWA(z_{\sigma(1)}, \ldots, z_{\sigma(q)})$$

An interesting improvement here is the Choquet integral [5, 10, 13]. This is indeed a more general aggregator that both includes weighted sums and OWA as special cases. It permits a fine control of the importance attached to criteria in the aggregation but also of the impact of good and bad grades on the final score. Formally, the Choquet integral writes:

$$C(z_1, ..., z_q) = \sum_{j=1}^{q} \left[z_{(j)} - z_{(j-1)}\right] v(A_{(j)}) \tag{12}$$

$$= \sum_{j=1}^{q} \left[v(A_{(j)}) - v(A_{(j+1)})\right] z_{(j)} \tag{13}$$

with the convention $z_{(0)} = 0 \leq z_{(1)} \leq \ldots \leq z_{(q)}$, $A_{(j)} = \{(j), ..., (q)\}$ for $j = 1, \ldots, q$, $A_{(q+1)} = \emptyset$, and v is a capacity function defined for any subset of subcriteria entering in the composition of the criterion. This capacity is a set function subject to the following constraints:

$$\begin{cases} v(\emptyset) = 0 \\ v(\{1, \ldots, q\}) = 1 \\ A \subseteq B \Rightarrow v(A) \leq v(B) \end{cases} \tag{14}$$

In Equation (12), the set $A_{(j)}$ is the set of subcriteria giving a grade greater or equal than $z_{(j)}$; the value $v(A_{(j)})$ represents the weight attached to coalition $A_{(j)}$ in the aggregation. The weight attached to a subcriterion j can be defined as the average contribution of j to any coalition A. This is formalized by the Shapley value defined as follows [14]:

$$\phi_j = \sum_{A \subseteq \{1,\ldots,q\} \setminus \{j\}} \frac{(q - |A| - 1)!|A|!}{q!} [v(A \cup \{j\}) - v(A)] \tag{15}$$

where q is the cardinal of the set of subcriteria to be aggregated and A is any subset of subcriteria within this set.

Capacity v is not necessarily additive, i.e. $v(A \cup B)$ might be different from $v(A) + v(B)$ even when $A \cap B = \emptyset$. This non-additivity is useful to reflect positive or negative synergies between criteria. The level of interaction between subcriteria as well as the sign (positive or negative) of the interaction can be measured by the following indices defined for all pair of subcriteria [14]:

$$I_{ij} = \sum_{A \subseteq \{1,\ldots,q\} \setminus \{i,j\}} \frac{(q - |A| - 2)!|A|!}{(q-1)!} \times [v(A \cup \{ij\}) - v(A \cup \{i\}) - v(A \cup \{j\}) + v(A)] \tag{16}$$

The sign of I_{ij} reflects the type of interaction among subcriteria. For instance, when $I_{ij} > 0$ there is a prime to be well graded both on criterion i and j. On the contrary, when $I_{ij} < 0$ being well graded on both subcriteria is seen as a redundancy that should not be counted twice. The case $I_{ij} = 0$ corresponds to non-interacting criteria. The reader is referred to [10, 14, 11] to get more detailed information about the use of Choquet integral in multicriteria aggregation.

To help us define the appropriate aggregation function χ_i for every criterion $f_i,\ i = 1, \ldots, 4$, and then its parameters (weights or capacity values), we consulted both animal and social scientists involved in Welfare Quality®. At that stage of the construction process, it was not necessary to ask questions to specialists of the measures (raw data were yet interpreted), and stakeholders' points of view (through social scientists' expertise) could be taken into account. Fourteen experts answered for dairy cows. We asked them to react on virtual datasets, one per criterion, composed of several virtual farms characterised by their scores on the q subcriteria to be considered. They were requested to assign a criterion value (on the [0,100] scale) to each farm. For the sake of illustration, we discuss now the consctruction of some criteria.

4.2 Criterion f_1 (Good feeding)

Let us consider Table 10 as an example of expert's answers obtained for criterion 1 (f_1).

Table 10: Example of answer obtained for criterion 1

Farm	Subcriterion scores		**Expert**
x_k	$g_1(x_k)$	$g_2(x_k)$	$y(x_k)$
x_1	25	75	**40**
x_2	40	60	**48**
x_3	50	50	**50**
x_4	60	40	**43**
x_5	75	25	**30**

The dataset proposed to the expert presents 5 fictitious farms (cf. Table 10) evaluated on two subcriteria g_1 and g_2. The expert has completed these lines by giving an overall grade $y(x_k)$ supposed to reflect the value of farm x_k with respect to criterion 1. To illustrate the reasoning followed by this expert, we can first have a look at the scores assigned by the expert to farms x_2, x_3, and x_4. Indeed, in theory with only two criteria, three farms are sufficient to infer the capacity function v. The table deliberately includes additional farms to get some redundancies so as to check the internal consistency of the answers as well as their robustness to variations of the range of utility gaps from a subcriterion to another.

Farm x_2 has a bad score on Subcriterion g_1 (absence of prolonged hunger) but a good one on Subcriterion g_2 (absence of prolonged thirst). Conversely, the farm

x_4 has a good score on g_1 but a bad score on g_2. Finally, farm x_3 is fairly good both on g_1 and g_2. By prescribing a lower criterion value to farm x_4 (43) than to farm x_2 (48), the expert gives higher penalty when the worst score is obtained with g_2. This reveals that criterion g_2 (thirst) is more important than subcriterion g_1 (hunger) for the expert. Moreover, by prescribing a criterion value of 48 only (i.e. less than the average of subcriterion-scores, 50) to farm x_2 with $g_2(x_2) = 60$, the expert adopts a conjunctive attitude that he could explain as follows: he wish to put more importance on lower scores. Such an expert followed simultaneously two complementary principles in the assessment of farms: 1) assign more importance to subcriterion g_2, 2) focus on worst scores. As a consequence, when using WS or OWA we would capture only a part of this behaviour. To obtain a better approximation of prescribed criterion values, in accordance with the expert's opinion, we might expect that the Choquet integral will be more appropriate. This intuition will be clearly confirmed by our numerical experimentations as shown below.

Using a sample of n_f farms $x_k,\ k = 1, \ldots, n_f$, and the evaluations $g_1(x_k)$ and $g_2(x_k)$, we wish to find the aggregation function χ_1 that minimises δ, the sum of square deviations between expert's answers $y(x_k)$ and the model prediction $f_1(x_k) = \chi_1(g_1(x_k), g_2(x_k))$:

$$\delta = \sum_{k=1}^{n_f} (f_1(x_k) - y(x_k))^2 \tag{17}$$

Here, we have compared three possible options for χ_1 :

- Weigthed Sum : $f_1(x_k) = w_1 g_1(x_k) + w_2 g_2(x_k)$
- OWA : $f_1(x_k) = w_1 \min\{g_1(x_k), g_2(x_k)\} + w_2 \max\{g_1(x_k), g_2(x_k)\}$
- Choquet : $f_1(x_k) = \begin{cases} g_1(x_k) + (g_2(x_k) - g_1(x_k))v_2 & \text{if } \ g_2(x_k) > g_1(x_k) \\ g_2(x_k) + (g_1(x_k) - g_2(x_k))v_1 & \text{otherwise} \end{cases}$

For the WS and OWA aggregators, the decision variables are w_1 and w_2. They are subject to the following constraints: $w_1 \geq 0,\ w_2 \geq 0$ and $w_1 + w_2 = 1$. For the Choquet integral, we assume that $v(\emptyset) = 0$ and $v(\{1, 2\}) = 1$ (standard normalization conditions for a compromise operator). Hence only two variables remain $v(\{1\})$ and $v(\{2\})$ denoted v_1 and v_2 respectively for simplicity. These variables are submitted to the following constraints: $0 \leq v_i \leq 1,\ i = 1, 2$ due to monotonicity conditions. Using the examples given in Table 10 the minimization of the quadratic error leads to weighting parameters given in Table 12 with the approximation results given in Table 11.

By comparing the values of δ obtained with WS, OWA and Choquet respectively (last line of Table 11), we can see that Choquet gives the best result, with a very small error δ, and that WS gives clearly the worst one. Considering the Choquet model , we have to assign more weight to thirst (g_2) than to hunger (g_1) (Shapley values obtained with Choquet reflect perfectly that with 0.6 *vs.* 0.4, cf. Table 12), while assigning much more importance to the lowest score than to the best one (that is fully in accordance with the fact that the Index of interaction calculated with Choquet is high: $I_{12} = 0.58$, cf. Table 12). This perfectly matches with experts explanations given to justify his evaluations.

Table 11: Comparison of WS, OWA and Choquet for Criterion 1

Farm	Subcriterion scores		**Expert**	Criterion score f_1		
x_k	g_1	g_2	y	WS	OWA	Choquet
x_1	25	75	**40**	55	36	40
x_2	40	60	**48**	52	44	46
x_3	50	50	**50**	50	50	50
x_4	60	40	**43**	48	44	42
x_5	75	25	**30**	45	36	31
			δ	*491*	*66*	*5*

Table 12: Parameters defined for sub-aggregation (Criterion 1)

WS		OWA		Choquet			
w_1	0.38	w_1	0.72	v_1	0.11	ϕ_1	0.40
w_2	0.62	w_2	0.28	v_2	0.31	ϕ_2	0.60
						I_{12}	0.58

4.3 Criterion f_2 (Good housing)

If we now consider Criterion 2 (f_2), composed of three subcriteria (g_3, g_4, and g_5), we proceed as for Criterion 1. The dataset includes now 13 virtual farms (cf. Table 13) allowing the determination of weighting parameters by minimization of the sum of square deviations between values $f_2(x_k)$ given by the model and experts' answers $y(x_k)$, for all reference farms x_k:

$$\delta = \sum_{k=1}^{n_f} (f_2(x_k) - y(x_k))^2 \tag{18}$$

with $n_f = 13$ farms. The optimization of this function yields the 3 weighting parameters when f_2 is a weighted sum or an OWA, and the 6 capacity values for the Choquet integral. For example, we present in Table 13 the answers given by an expert for Criterion 2. We see that optimization provides a very good estimate of expert's prescribed values. Note that for Choquet, the sum of square deviations equals 14, that is less than the δ obtained with OWA (109), and much less than the δ obtained with WS (688).

The weighting parameters obtained are given in Table 14. By calculating the Shapley values (ϕ_i) (cf. Equation 15) and the interaction indices (I_{ij}) (cf. Equation 16) we can interpret the reasoning followed by the expert seen through the Choquet model (cf. Table 14). Note first that $\phi_3 < \phi_4 < \phi_5$ which is in accordance with what the expert explained to us (Subcriterion 5 is the most important, then Subcriterion 4, and finally 3). In addition, we can observe that the interaction indices are positive and non-negligible (values range from 0.15 to 0.35). This is also in accordance with the wish of the expert to limit compensations between bad and good subcriterion-scores,

Table 13: Virtual dataset for Criterion 2 "Good housing" and example of answer

Farm	Subcriterion scores			**Expert**	Criterion score f_2		
x_k	g_3	g_4	g_5	y	WS	OWA	Choquet
x_1	25	50	75	**41**	52	40	41
x_2	40	50	60	**47**	51	46	46
x_3	50	50	50	**50**	50	50	50
x_4	25	75	50	**40**	48	40	40
x_5	40	60	50	**46**	49	46	47
x_6	50	25	75	**43**	54	40	44
x_7	50	40	60	**52**	52	46	51
x_8	50	60	40	**44**	48	40	44
x_9	50	75	25	**35**	46	46	34
x_{10}	60	40	50	**51**	51	46	49
x_{11}	75	25	50	**42**	52	40	42
x_{12}	60	50	40	**45**	49	46	47
x_{13}	75	50	25	**38**	48	40	38
				δ	*688*	*109*	*14*

Table 14: Parameters defined for sub-aggregation (Criterion 2) with one expert

WS		OWA		Choquet			
w_3	0.33	w_1	0.59	v_3	0.07	ϕ_3	0.29
w_4	0.25	w_2	0.20	v_4	0.02	ϕ_4	0.29
w_5	0.42	w_3	0.20	v_5	0.10	ϕ_5	0.42
				v_{34}	0.31	I_{34}	0.15
				v_{35}	0.48	I_{35}	0.24
				v_{45}	0.54	I_{45}	0.35

favouring situations achieving a well-balanced compromise. As a consequence we can say that the use of Choquet integral allows us to reflect properly the expectations of the expert, and provides clearly a better approximate than Weighted Sum or Ordered Weighted Average.

To get a safer construction of criteria, we consulted more than one expert. Actually, 14 experts answered to our questions for dairy cows. Their answers are presented in Table 15. Extending the approach adopted for a single expert, we defined the parameters of the aggregator (WS, OWA or Choquet) by minimization of the overall error δ defined by:

$$\delta = \sum_{p=1}^{n_e} w_p \left(\sum_{k=1}^{n_f} (f_2(x_k) - y_p(x_k))^2 \right) \tag{19}$$

where n_e is the number of experts, n_f is the number of examples (farms), $y_p(x_k)$ is the evaluation given by expert p to farm x_k for Criterion 2, $f_2(x_k)$ is the score calculated on farm x_k by the model and w_p is the weight assigned to expert p. Actually, we decided to give the same importance to all the experts ($\forall\, p = 1, \ldots, n_e, w_p = 1$) but weights might be useful to take into account relevance, expertise or self-confidence of experts in the evaluation process.

As can be seen in Table 15, by comparing the averages of experts' evaluations (column $\bar{y}$) to the scores f_2 calculated by the Choquet integral (column C), we can see that the model very well approximates the average of experts' evaluations. This is confirmed by the low value of δ obtained with Choquet. Note that the approximation with Choquet is slightly better than with OWA and clearly improves what we could obtain with WS (cf. δ values in the last line of Table 15).

4.4 Criteria f_3 and f_4

Criteria f_3 and f_4 are treated similarly as f_1 and f_2. Actually, for all criteria, we compare Weighted Sum, Ordered Weighted Average and Choquet integral. In Table 17 the average sums of square deviations per expert are presented for the four criteria. In all cases, the use of WS appears to be clearly the least efficient, and the use of Choquet integral appears to be the best choice, even if OWA appears to be also a good choice. The fact that Choquet integral and OWA have a similar efficiency in our case is due to the fact that a majority of experts decided to a put criterion values close to the minimum of the subcriterion scores, a behaviour that can be simply approximated by an OWA (e.g. $w_1 = 1$ and for $i \neq 1$ $w_i = 0$).

Choosing the Choquet integral as the aggregator to construct criteria from the contributing subcriteria appears to be systematically the best choice (giving the smallest δ). However, the number of parameters to be determined expentionally grows with the number of subcriteria ($2^q - 2$ for q subcriteria). For example we need to learn 6 coefficients to defined a criterion from three subcriteria (that is the case for f_2 and f_3), while 14 coefficients are needed for f_4. As proposed in [12], to reduce the number of parameters in Choquet Integrals, we can use a *2-additive capacity* that only requires $q(q+1)/2$ coefficients (a capacity value for each pair of criteria and each singleton).

Table 15: Experts' evaluations calculation of the score for Criterion f_2 "Good housing"

k	g_3	g_4	g_5	***Experts' answers***															**Criterion score f_2**		
				y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	$\bar{y}$	WS	OWA	**C**
1	**25**	**50**	**75**	20	35	35	25	25	25	50	54	35	25	25	25	25	41	**32**	50	32	**31**
2	**40**	**50**	**60**	30	45	45	40	40	50	50	52	50	40	40	40	40	47	**44**	50	43	**42**
3	**50**	**50**	**50**	40	50	50	50	50	50	50	50	50	50	50	50	50	50	**49**	50	50	**50**
4	**25**	**75**	**50**	20	35	35	25	25	25	45	58	35	25	25	25	25	40	**32**	51	32	**33**
5	**40**	**60**	**50**	30	45	45	40	40	50	50	53	50	40	40	40	40	46	**44**	50	43	**44**
6	**50**	**25**	**75**	20	35	35	25	25	27	40	46	35	25	25	25	25	43	**31**	49	32	**32**
7	**50**	**40**	**60**	40	45	45	40	40	50	50	48	50	40	40	40	40	52	**44**	50	43	**44**
8	**50**	**60**	**40**	50	45	45	40	40	50	50	52	50	50	50	50	50	44	**45**	50	43	**44**
9	**50**	**75**	**25**	40	35	35	25	25	25	45	54	35	25	25	25	25	35	**32**	51	32	**32**
10	**60**	**40**	**50**	40	45	45	40	40	50	40	47	50	40	40	40	40	51	**43**	50	43	**43**
11	**75**	**25**	**50**	40	35	35	25	25	27	30	42	35	25	25	25	25	42	**31**	49	32	**31**
12	**60**	**50**	**40**	40	45	45	40	40	50	50	48	50	40	40	40	40	45	**44**	50	43	**44**
13	**75**	**50**	**25**	40	35	35	25	25	25	30	46	35	25	25	25	25	38	**31**	50	32	**32**
																		δ / # of experts	*2862*	*598*	*590*

Table 16: Parameters defined for sub-aggregation (Criterion 2) with the 14 experts

WS		OWA		Choquet			
w_3	0.33	w_1	0.72	v_3	0.03	ϕ_3	0.32
w_4	0.35	w_2	0.28	v_4	0.03	ϕ_4	0.35
w_5	0.33	w_3	0.00	v_5	0.04	ϕ_5	0.33
				v_{34}	0.19	I_{34}	0.41
				v_{35}	0.14	I_{35}	0.36
				v_{45}	0.20	I_{45}	0.41

Table 17: Approximation with Weighted Sum, OWA and Choquet

	$\sum$(Error^2) / no. of experts		
	Weighted sum	OWA	**Choquet**
Criterion 1	1256	324	**291**
Criterion 2	2862	598	**590**
Criterion 3	2933	566	**531**
Criterion 4	2075	736	**722**

In order to illustrate the potential of 2-additive capacities in our context, suppose we want to define a criterion from a set $N = \{1, \ldots, q\}$ of subcriteria to be aggregated with a Choquet integral (Equation (12)), using a capacity v. We can consider the *Möbius representation* of v defined by:

$$m_v(B) = \sum_{A \subseteq B} (-1)^{b-a} v(A), \qquad \forall\, B \subseteq N \tag{20}$$

where $a = |A|$ and $b = |B|$. Hence, the capacity v can be expressed using the *Möbius representation* by:

$$v(C) = \sum_{B \subseteq C} m_v(B), \qquad \forall\, C \subseteq N \tag{21}$$

Capacity v is said to be *2-additive* if its Möbius representation satisfies $m_v(B) = 0$ for all $B \subseteq N$ such that $b > 2$ and there exists at least one subset C of cardinality 2 such that $m_v(C) \neq 0$. As a consequence of Equation (21), if v is 2-additive then the coefficients corresponding to coalitions of 3 subcriteria $\{ijk\}$ write as follows:

$$v(ijk) = m_v(ij) + m_v(jk) + m_v(ik) + m_v(1) + m_v(2) + m_v(3)$$

Since $m_v(ij) = v(ij) - v(i) - v(j)$ and $m_v(i) = v(i)$ due to Equation (20), $v(ijk)$ can easily be derived from the capacities of subsets containing 1 or 2 subcriteria as follows:

$$v(ijk) = v(ij) + v(jk) + v(ik) - v(i) - v(j) - v(k) \tag{22}$$

Hence, by using a 2-additive capacity for the Choquet integral that defines f_4 from subcriteria $g_9, g_{10}, g_{11}, g_{12}$, we have the possibility to reduce the number of coefficients to be learned from 14 to 6, as for the calculation of f_2 and f_3. This idea is particularly

interesting when the 2-additivity assumption does not downgrade significantly the quality of the approximation of experts' answers by a Choquet integral. To test this on our data, we have compared the δ obtained using a Choquet integral, with and without the 2-additivity constraint, to results obtained with OWA and weighted sum. The minimal error is obtained for Choquet with a general capacity ($\delta = 722$). Note that it is worth adding the 2-additivity constraints here since we remain near to optimality ($\delta = 726$). As usual OWA and WS give less interesting results (736 and 2075 respectively). We can see that, in this case, using a 2-additive capacity reduces the number of parameters of the model (and therefore the elicitation burden) while keeping good performances in the approximation of expert's opinion.

5 Aggregation of criteria

The project Welfare Quality® aims at developing a European standard for the assessment of animal welfare, and to implement it into an information standard for all the stakeholders (producers, processors, retailers, consumers, institutions, and researchers). To produce an appropriate evaluation and informative system, we think it is important and useful to include stakeholders' views in the construction of the overall welfare assessment and information system.

Their opinion seems to be especially useful and relevant for the last stage of the construction process, to go from the four criterion values to the overall assessment. First of all, we consulted some stakeholders' representatives (from European farmers, animals protectors, consumers, and vets associations, institutions like OIE[1], and scientists), members of the Advisory Committee of the project. We asked them individually to express what purposes their organisation foresees for a welfare assessment tool. Four main uses of the welfare assessment system were thus identified:

- it could help to define the minimum level of animal welfare that is acceptable (it could either substitute the current EU regulation on minimum animal welfare standard or help implementing a regulation on *Compulsory Labelling*);
- it could help to distinguish instances in which the welfare of the animals not only exceeds the minimum acceptable level but attains high or very high ones (it could help implementing *Voluntary Labelling*, using independent third party inspections);
- it could be fruitfully used as a *'self-assessment' management tool* to help farmers identify welfare problems or risks and to monitor improvements;
- it could be used as a *research tool* to evaluate the welfare status of animals housed in various systems as well as that of emerging breeds, this could also serve for *legislative purposes* to take a decision on the authorisation of new systems or equipments.

[1] Organisation Internationale des Epizooties

Each of the uses implies the definition of several welfare categories from "unacceptable" to "excellent" (cf. Table 18). To implement such uses in the future, we have to produce an evaluation system which is able to allocate each observed farm to one of the pre-defined categories of welfare, in other words we are facing an *ordered sorting problem*, according to the terminology of [22].

Table 18: Potential uses of an overall welfare assessment and consequences on the number of welfare categories to be considered

Use of the assessment tool	**Welfare categories**	
	#	Meaning
Compulsory labelling	*2*	Below *vs.* above minimal legal requirements or equivalent (if no legislation exists)
Voluntary labelling	*3 or 4*	at least Excellent / Enhanced / Acceptable levels of welfare
Self-assessment tool for farmers	*Several*	Very low ——— Excellent welfare
Systems' evaluation tool (for research or legislative purposes)	*Several*	Very low ——— Excellent welfare

First of all, we have to make a decision on the number of welfare categories to be considered and on their definition. According to the information gathered from stakeholders on the potential uses of the welfare assessment system, at least four welfare categories seem to be necessary to be able to answer all the stakeholders' expectations:

- **Excellent**: the welfare of the animals is very high. The animal unit may correspond to a niche market, within a very high quality voluntary label, maybe dedicated to animal welfare.
- **Enhanced**: the welfare of animals is good (but not excellent). Good farming practices are applied, and this may serve to guarantee a good level of welfare within a more general quality label (that is another voluntary labelling system, less demanding).
- **Acceptable**: the welfare of animals is acceptable, but insufficient to be used into a voluntary labelling system.
- **Not classified**: the welfare of animals is low and considered unacceptable. Within a compulsory labelling scheme, such animal units would be rejected.

These four welfare categories can be used differently depending on the potential use considered (cf. Figure 5).

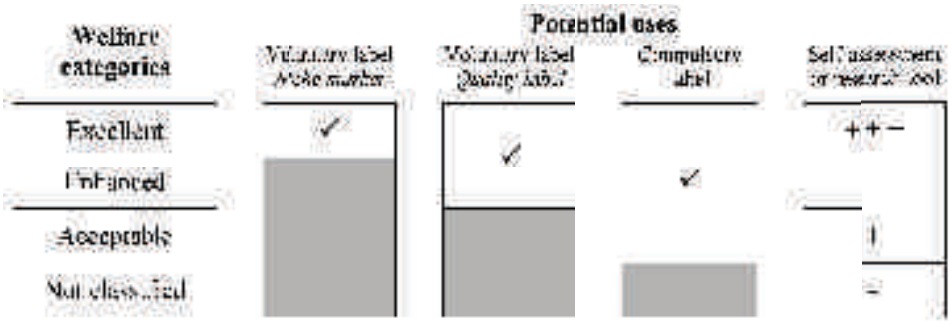


Figure 5: Utilization of the welfare categories within the different potential uses. Depending on the potential use considered, the response maybe binary (e.g. *Is it an excellent farm or not?*) or gradual (*Is the farm −, +, ++ or +++?*).

We have now to define the welfare categories "Excellent", "Enhanced", "Acceptable", and "Not classified" from the four welfare criteria f_1, f_2, f_3, f_4 as well as a multicriteria sorting procedure assigning farms to categories. These four criteria use the same evaluation scale but should not be aggregated directly into a single comprehensive score because they represent four distinct dimensions that refer to different viewpoints on welfare and that do not compensate each other. Instead, it is more appropriate to perform a real mulcriteria analysis in this four-dimensional space so as to make assignement of farms into categories. In such a context, we propose the use of an assignment procedure of type Electre TRI (see [17, 18]) or a fuzzy extension of it (see [19]). In such procedures, the assignement of a given farm to a category will be based on the intrinsic performances of the farm and their comparison to the reference performance profiles representing the lower bound of categories. We give an illustration of a simple procedure of this type on Figure 6.

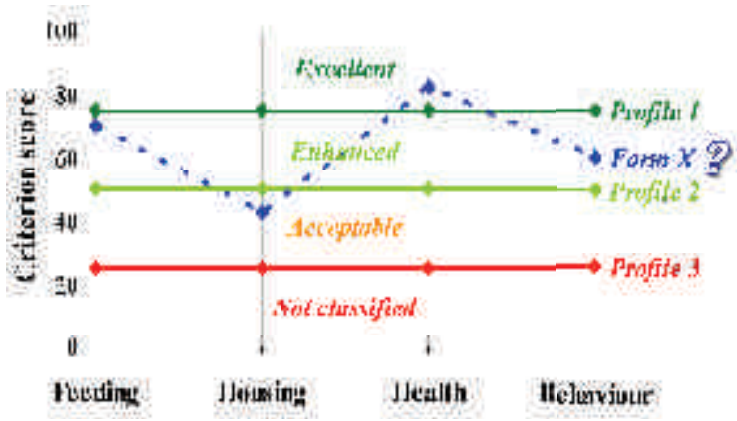


Figure 6: Aggregation of criteria: Comparison to pre-defined profiles

In this example, the vector $z_1 = (75, 75, 75, 75)$ represents the lower bound of the category "Excellent". For any farm x_k characterized by the performance vector $(x_{k1}, \ldots, x_{k4})$ with $x_{ki} = f_i(x_k), i = 1, \ldots, 4$ (dotted line on Figure 6), we first compare x_k *to* z_1. If x_k is at least as good as z_1 (denoted $x_k \succsim z_1$) then x_k is assigned to category C_1: "Excellent". If not, x_k is compared to a less demanding profile $z_2 = (50, 50, 50, 50)$ that represents the lower bound of the second category (C_2: "Enhanced"). If $x_k \succsim z_2$ then x_k is assigned to C_2, otherwise it is compared to a the third reference profile $z_3 = (25, 25, 25, 25)$ and so on... If $x_k \succsim z_3$ then it is assigned into category C_3: "Acceptable" otherwise, it will fall in the last category C_4. More formally, the assignment procedure simply writes as follows:

Algorithm 1: Assignment procedure

```
for k = 1 to n_f do
    i ← 1
    while not(x_k ≳ z_i) and i < 4 do
        i ← i + 1
    end
    x_k → C_i
end
```

As usual in outranking methods of ELECTRE type, deciding whether $x_k \succsim z_i$ or not at a given step of the algorithm is based on the concordance and non-discordance principles. The *concordance principle* requires that a sufficient majority of criteria are in accordance with the assertion $x_k \succsim z_i$ whereas the *discordance principle* checks that no criterion strongly contradicts this assertion (this requires to define veto thresholds, for more details see [21, 19]). The exact calibration of reference profiles as well as the conditions required to satisfy concordance and non-discordance tests are actually under discussion. They will be made after testing the procedure on a large sample of reference farms so as to make sure that 1) the categories cover the whole range of possibilities, 2) every category includes a significant percentage of farms in the sample, and 3) expert's assessment of animal welfare in farms assigned to each category are consistent with the label attached to the category (Excellent, Enhanced, Acceptable, Not classified). The final choice will be made in close cooperation with the various stakeholders and potential users (farmers, consumers, and policy makers), after tuning parameters (profiles, veto thresholds, weights) so as to satisfy the 3 above conditions. We will then check that the results of the assessment model match their expectations. Focus groups discussions in selected EU countries (France, United-Kingdom, the Netherlands, and a Scandinavian country) are already planned.

6 Discussion

The application we describe in this paper allows producing synthetic information about the welfare status of animals on a farm from data collected on animals or their environment. Thanks to the formalisation of the aggregation process, the evaluation

can be performed in a standardised way and bring a sorting of farms into four welfare categories: excellent welfare, enhanced welfare, acceptable welfare, not classified. The method will be extended to the other animal types under study in WelfareQuality®: beef cattle, veal calves, fattening pigs, sows, laying hens, and broilers. As a matter of fact, the criteria and subcriteria have been decided whatever the animal species. Although their importance may vary from one species to another, the general construction will remain the same. This will make it possible to have a general evaluation system to assess animal welfare in European farms. Such a system presents several advantages:

First, Welfare Quality® is developing training packages and methods to ensure that the observers that collect data on farms produce reliable data. Thanks to this training and the formalisation of the aggregation process, the evaluation can be performed routinely in an objective way.

Second, the sensitivity of the assessment will allow comparisons between farms, farming systems, countries, etc. Such comparisons are essential to promote good farming systems and practices. Data collected in commercial farms could thus be more easily used to guide policy makers. In addition, it will allow detecting farms where the welfare of animals is very high, so that these farms can be selected for a certification scheme promoting animal welfare. It should also allow detecting farms where the welfare of animals is unacceptable and thus help the EU Commission in its decision to attribute subsidies, in the absence of a specific directive for a given animal type. Finally, the intermediate grades (acceptable and enhanced welfare) together with the two extreme grades (not classified and excellent) shall help farmers figure out the welfare of their animals and monitor progresses they may make thanks to the adoption of remedial measures on the farm (e.g. health plan, improvement of housing...).

Third, the assessment remains transparent. Not only the overall assessment but all intermediate scores will remain available and can be use to produce automatic explanation of an evaluation. These intermediate scores, at subcriterion or criterion level, as well as the possibility of getting explanations are essential for farmers. They can help producers understand their final result, orientate the identification of causes of poor welfare, while encouraging them to take efficient remedial measures (i.e. likely to affect the result).

Finally, the assessment remains flexible. As a matter of fact, the architecture of the overall model has been set up in order to fit all farm animal productions and various objectives such as the selection of best farms, the advice to farmers, or the comparisons of farming systems. Nevertheless, all parameters have been explicitly described: thresholds to distinguish between a moderate *vs.* a severe problem, weights attributed to these problems, parameters of the Choquet integrals to synthesise the information at criterion-level, reference profiles and membership rules during the final aggregation. These parameters may be changed according to new scientific knowledge (e.g. a severe problem may be further considered as only moderate because proved not to affect animals very much) or due to changes in societal expectations (e.g. if the welfare of animals improve significantly in all farms, end-users may want to be more selective when considering a farm as excellent).

The formal model proposed here makes explicit the information used in the evaluation process as well as the evaluation rules adopted. It can be implemented through an automatic multicriteria evaluation procedure, including the different dimensions of animal welfare. Such a procedure can be accessible on the web so as to provide online evaluations with various additional possibilities (explanations, recommendations). These are the main advantages of resorting to formal multicriteria methods in the evaluation of animal welfare.

References

[1] Bartels R., Beatty J., Barsky B., *An introduction to splines for use in computer graphics and geometric modelling*, Morgan Kaufmmann Publishers, inc., 1987.

[2] Botreau R., Bracke M., Perny P., Butterworth A., Capdeville J., van Reenen D., Veissier I., Aggregation of measures to produce an overall assessment of animal welfare: Part 2 - Analysis of constraints, *Animal*, **1**, 8, (2007), 1188–1197.

[3] Botreau R., Veissier I., Butterworth A., Bracke M., Keeling L., Definition of criteria for overall assessment of animal welfare, *Animal Welfare*, **16**, (2007), 225–228.

[4] Bouyssou D., Marchant T., Perny P., Pirlot M., Tsoukis A., Vincke P., *Evaluation and decision models: a critical perspective*, Kluwer Academic Publishers, 2000.

[5] Choquet G., Theory of capacities, *Annales de l'Institut Fourier*, **5**, (1953), 131–295.

[6] Dubois D., Fortemps P., Pirlot M., Prade H., Leximin Optimality and Fuzzy Set Theoretic Operations, *European Journal of Operational Research*, **130**, (2001), 20–28.

[7] European Commission, *Attitudes of EU citizens towards Animal Welfare - Special Eurobarometer 270/Wave 66.1 - TNS Opinion & Social*, 2007.

[8] Farm Animal Welfare Council, FAWC updates the five freedoms, *The Veterinary Record*, **17**, (1992), 357.

[9] Fraser D., Science, values and animal welfare: exploring the 'inextricable connection', *Animal Welfare*, **4**, (1995), 103–117.

[10] Grabisch M., The application of fuzzy integrals in multicriteria decision making, *European Journal of Operational Research*, **89**, (1996), 445–456.

[11] Grabisch M., Duchne J., Lino F., Perny P., Subjective evaluation of discomfort in sitting position, *Fuzzy Optimization and Decision Making*, **1**, 3, (2002), 287–312.

[12] Grabisch M., Kojadinovic I., Meyer M., A review of capacity identification methods for Choquet integral based multi-attribute utility theory, Applications of the Kappalab R package, to appear.

[13] Grabisch M., Murofushi T., Sugeno M., Kacprzyk J., *Fuzzy Measures and Integrals. Theory and Applications*, Physica Verlag, Berlin, 2000.

[14] Grabisch M., Roubens M., *Application of the Choquet integral in multicriteria decision making*, Physica Verlag, 2000, 348–374.

[15] Main D., Webster F., Green L., Animal Welfare Assessment in Farm Assurance Schemes, *Acta Agriculturae Scandinavica, Section A, Animal Science*, **Supplementum 30**, (2001), 108–113.

[16] Mason G., Mendl M., Why is there no simple way of measuring animal welfare?, *Animal Welfare*, **2**, (1993), 301–319.

[17] Moscarola J., Roy B., Procdure automatique d'examen de dossiers fonde sur une segmentation trichotomique en prsence de critres multiples, *RAIRO Recherche Oprationnelle*, **11**, (1977), 145–173.

[18] Mousseau V., Slowinski R., Inferring an ELECTRE TRI Model from Assignment Examples, *Journal of Global Optimization*, **12**, (1998), 157–174.

[19] Perny P., Multicriteria Filtering Methods based on concordance and non-discordance principle, *Annals of Operations Research*, , 80, (1998), 137–165.

[20] Plass M., Stone M., Curve-Fitting with piecewise paramatric cubics, *Computer Graphics-SIGGRAPH'83 proceedings*, **17**, 3, (1983), 229–239.

[21] Roy B., The outranking approach and the foundations of ELECTRE methods, *Theory and Decision*, , 31, (1991), 49–73.

[22] Roy B., *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers, 1993.

[23] Yager R., On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Trans. on Sys., Man and Cyb.*, **18**, (1988), 183–190.

Chapitre 8

Vers un outil de diagnostic et de conseil efficace

Résumé sur l'aptitude du modèle à servir de base à un diagnostic-conseil :

En maintenant l'information disponible à chaque étape de construction de l'évaluation globale, le modèle permet d'expliquer les évaluations obtenues par les fermes. Dans ce chapitre, nous nous basons sur des exemples de fermes visitées lors du WP2.4 afin de montrer la capacité du modèle à mettre en évidence les points faibles et les points forts des fermes. Pour ce faire, nous partons de l'évaluation globale qu'elles ont obtenue et nous redescendons, étape par étape, jusqu'aux données brutes recueillies sur le terrain. Cette procédure d'explication du résultat pourra être automatisée afin de faciliter le diagnostic.

Nous constatons que le modèle d'évaluation proposé a un pouvoir incitatif. Il permet en effet d'encourager les éleveurs à améliorer le niveau de bien-être de leurs animaux, et notamment sur les aspects les plus problématiques (dans la plupart des cas).

Pour être pleinement efficace, la phase de diagnostic consistant à mettre en évidence les mesures sur lesquelles l'éleveur doit porter son attention, devra être complétée par une phase de conseil à l'éleveur. Les causes potentielles des problèmes observés pourront alors être identifiées et des solutions correctrices pourront être proposées à l'éleveur.

Chapitre 8

Vers un outil de diagnostic et de conseil efficace

Dans ce chapitre nous nous basons sur quelques exemples de fermes visitées lors du WP2.4 pour illustrer la capacité du modèle à expliquer les résultats (phase de diagnostic) en montrant les points forts et les points faibles de l'élevage. Cette phase de diagnostic peut se poursuivre par un conseil efficace sur le terrain (auprès des éleveurs dans le cas du modèle "vaches laitières en ferme").

Pour illustrer ce point nous avons décidé d'utiliser la Procédure 4, présentant les règles d'affectation aux différentes catégories les plus souples. Le Scénario 4 repose sur les règles suivantes (rappel du Chapitre 7) :

- **comparaison au profil** z_1 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_2) = 1$ et $c(x_k, z_1) \geq \lambda_1$, avec $\lambda_1 = 0.25$
  $\hookrightarrow$ *x_k doit battre z_2 à l'unanimité et battre z_1 sur un critère* ;
- **comparaison au profil** z_2 **:** la ferme x_k bat z_2 ($x_k \succsim z_2$) si et seulement si $c(x_k, z_3) = 1$ et $c(x_k, z_2) \geq \lambda_2$, avec $\lambda_2 = 0.50$
  $\hookrightarrow$ *x_k doit battre z_3 à l'unanimité et battre z_2 sur deux critères* ;
- **comparaison au profil** z_3 **:** la ferme x_k bat z_1 ($x_k \succsim z_1$) si et seulement si $c(x_k, z_4) = 1$ et $c(x_k, z_3) \geq \lambda_3$, avec $\lambda_3 = 0.75$
  $\hookrightarrow$ *x_k doit battre z_4 à l'unanimité et battre z_3 sur trois critères.*

Cette procédure d'affectation, en relation avec la catégorie C_i, repose sur un double niveau d'exigence : un niveau d'exigence minimale lié à la comparaison des performances de la ferme x_k au profil z_{i+1} marquant la limite basse de la catégorie C_{i+1}, associé à un niveau d'aspiration lié à la comparaison des performances de la ferme x_k au profil z_i.

Rappel de la structure du modèle :

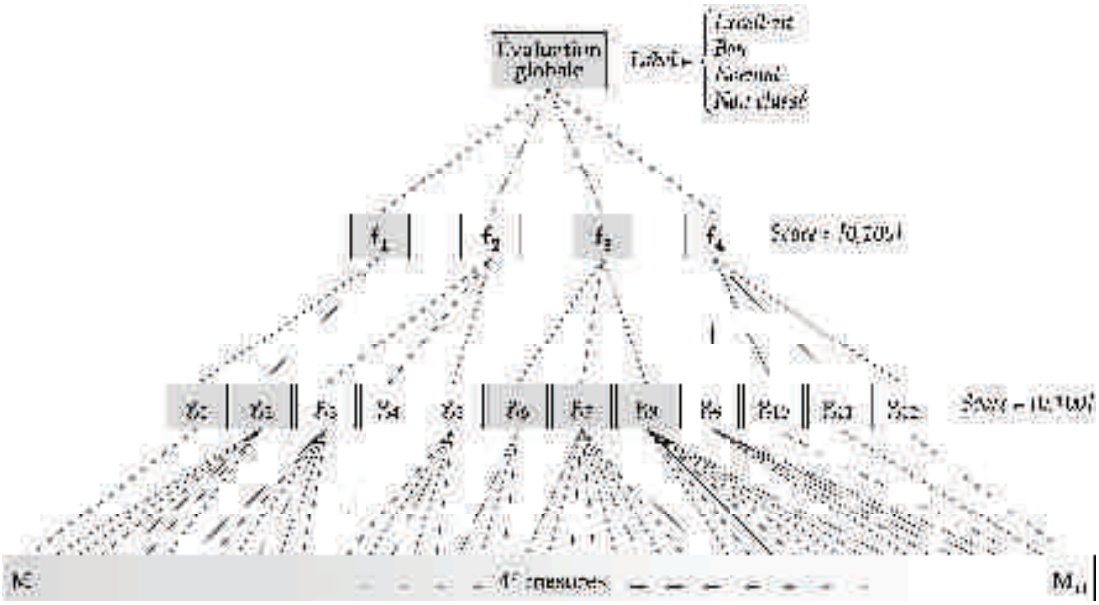


FIG. 8.1 – Structure séquentielle du modèle d'évaluation

La première étape de la phase de diagnostic consiste à présenter à l'éleveur ses performances à chaque étape de la construction du modèle (évaluation globale, critères, sous-critères et mesures). Il serait possible de présenter ces valeurs sous la forme d'un diagramme similaire à celui présenté sur la Figure 8.1. C'est avec cette information détaillée, allant de la catégorie à laquelle la ferme a été affectée aux données brutes recueillies sur l'élevage, qu'il sera possible de dresser un diagnostic pertinent, permettant de pointer les mesures se révélant problématiques sur l'élevage. La phase de conseil pourra alors débuter, en identifiant les causes potentielles de ces mauvais résultats au niveau de certaines mesures afin de pouvoir, par la suite, suggérer à l'éleveur des actions correctrices adaptées à l'élevage, et ce, dans la perspective d'aider l'éleveur à passer dans la catégorie supérieure.

Nous venons de voir dans le chapitre précédent que le paramétrage du modèle ne permet à aucune des fermes observées dans le WP2.4 d'être affectée à la catégorie *"Excellent"*. Il permet cependant de couvrir les trois autres catégories de bien-être. Nous présentons ici trois exemples de fermes par catégorie, en partant de leur évaluation globale puis en redescendant aux critères, puis aux sous-critères et enfin aux mesures.

Les évaluations obtenues par chacune de ces fermes au niveau des critères et des

sous-critères sont présentées sous la forme d'un schéma utilisant les symboles présentés dans la Figure 8.2.

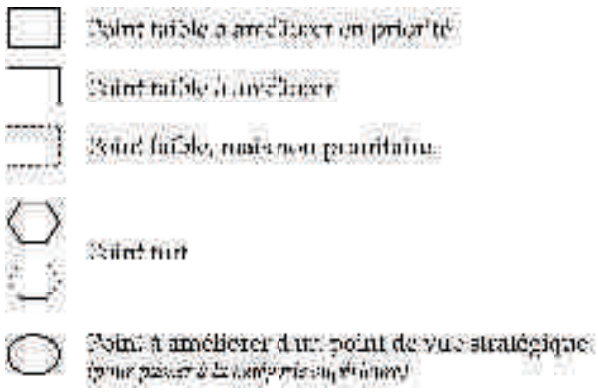


FIG. 8.2 – Symboles des points forts et faibles des fermes, utilisés dans les Figures 8.3 à 8.11

8.1 Exemples de fermes affectées à la catégorie *"Non classé"*

Dans cette première section, nous présentons des exemples de fermes ayant été affectées à la **catégorie *"Non classé"***, en montrant pour chacune d'elles les points faibles à améliorer en priorité.

Afin d'être affectée à la catégorie *"Normal"*, une ferme doit battre z_3 sur trois critères et obtenir un score supérieur ou égal à z_4 sur le quatrième critère. Cette ferme doit donc obtenir au moins 20 points sur trois critères et au moins 10 sur le quatrième.

8.1.1 Exemple 1

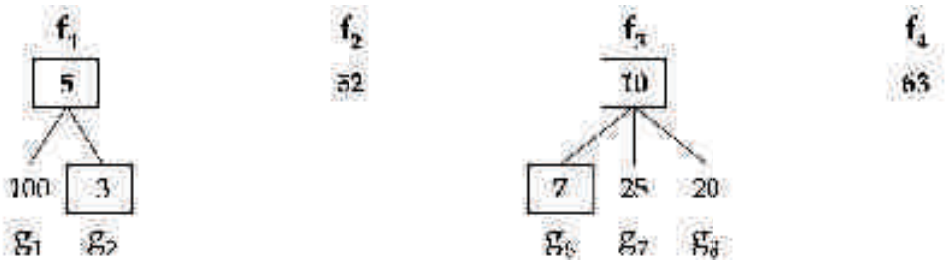


FIG. 8.3 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 1

La Ferme 1 présente deux points faibles importants au niveau des critères : sur le Critère 1 *"Alimentation adaptée"* et sur le Critère 3 *"Bonne santé"*, avec respectivement des scores de 5 et 10.

Le critère à examiner en priorité est le Critère 1, sur lequel la ferme obtient un score inférieur à 10 (score correspondant au niveau d'exigence minimale pour la catégorie "Normal"). Une brève analyse au niveau des sous-critères montre que le problème est lié au Sous-critère 2 *"Absence de soif prolongée"* et non au Sous-critère 1 *"Absence de faim prolongée"* sur lequel la ferme reçoit le score maximal de 100. Pour apporter une aide utile à l'éleveur, il est alors possible de descendre au niveau des mesures servant à évaluer g_2, sous-critère pour lequel la ferme a reçu le plus mauvais score possible. Un score de 3 sur g_2 correspond à un nombre insuffisant d'abreuvoirs. En effet, en regardant de plus près les données brutes on constate que sur cette ferme il n'y a que deux bacs d'abreuvement d'une longueur totale de 75 cm pour un groupe de 30 animaux. Or, avec ces deux abreuvoirs seules 12 vaches devraient être dans le groupe. Pour remédier au problème, l'éleveur doit installer des abreuvoirs supplémentaires (par exemple deux bols). Étant donné que par ailleurs les abreuvoirs sont propres et que les animaux ont accès à deux abreuvoirs, en ajoutant simplement deux bols la ferme pourrait recevoir la note maximale.

En effectuant ces modifications au niveau de l'approvisionnement en eau, la Ferme 1 pourrait alors être affectée à la catégorie *"Normal"* (mais de justesse au vu du score obtenu sur le Critère 3).

La Ferme 1 pourrait ensuite être affectée à la catégorie *"Bon"* sous réserve d'amélioration du Critère 3 (en passant d'un score de 10 à au moins 20). Ce critère se compose de trois sous-critères dont un seul est vraiment problématique : le Sous-critère 6 *"Absence de blessures"*. Comme nous avons pu le voir dans le Chapitre 5, l'évaluation de ce sous-critère repose sur deux éléments distincts : les altérations du tégument et les boiteries. Sur ces deux aspects des blessures, la Ferme 1 obtient respectivement des scores de 13 (53% des vaches présentent au moins une blessure modérée et 37% au moins une blessure sérieuse) et 4 (13% des vaches boitent légèrement et 73% boitent sérieusement). Le rôle du conseiller sera alors d'aider l'éleveur à identifier ce qui cause les blessures et les boiteries chez les animaux, puis, dans un deuxième temps, de lui proposer des solutions correctrices.

8.1.2 Exemple 2

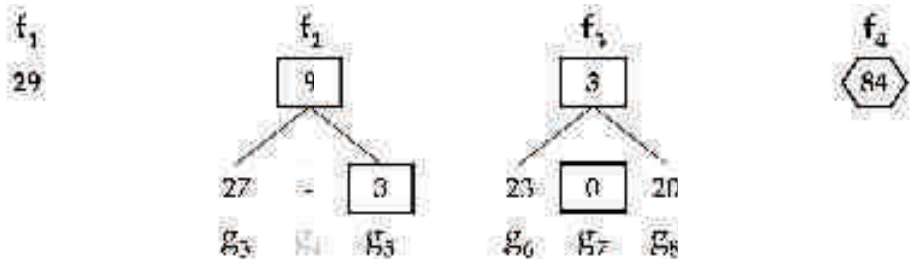


FIG. 8.4 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 2

Ce second exemple concerne une ferme présentant un profil très hétérogène, avec un point fort sur le Critère 4 *"Comportement approprié"* (avec un score de 84) et deux points particulièrement faibles sur les Critères 2 *"Logement correct"* et 3 *"Bonne santé"* (avec respectivement des scores de 9 et 3). Chacun des scores obtenus sur les Critères 2 et 3 suffit à expliquer l'affectation de la Ferme 2 à la catégorie *"Non classé"* puisque ces scores sont inférieurs au niveau d'exigence minimale. Cette ferme doit impérativement améliorer les Critères 2 et 3, au moins pour obtenir des scores supérieurs ou égaux à 10. Dans un deuxième temps, plusieurs possibilités s'offrent à l'éleveur pour faire en sorte que son exploitation soit affectée à la catégorie *"Normal"* puisque le passage d'un seul des deux critères problématiques au-dessus de z_3 suffit. L'éleveur pourra alors choisir la solution qui est la plus facile à mettre en oeuvre.

Sur le Critère 3, contrairement à la Ferme 1, le problème ici est lié au Sous-critère 7 *"Absence de maladies"* avec un score de 0. Un tel score est obtenu par cette ferme car elle présente des problèmes sanitaires à la fois au niveau respiratoire (par ex. 18% des vaches sur cette ferme présentent un écoulement nasal et 25% un écoulement oculaire), digestif (7% des vaches présentent des diarrhées) et reproductif (par ex. 11% des vaches ont des pertes vaginales). Il semble donc y avoir un problème sanitaire général et très important sur cet élevage, qui nécessite la mise en place d'un plan d'action sanitaire.

Au niveau du Critère 2 le problème se situe au niveau du Sous-critère 5 *"Facilité de déplacement"* sur lequel la Ferme 2 obtient là aussi la plus mauvaise note possible (3). En effet, les vaches sont maintenues attachées durant toute l'année, elles n'ont pas accès à une zone d'exercice extérieure et elles ne sont jamais mises à la pâture. Une solution pour améliorer ce point pourrait être d'installer une aire d'exercice extérieure à côté de l'étable (si il y a suffisamment de place à proximité du bâtiment) afin que l'éleveur puisse y sortir ses vaches régulièrement.

8.1.3 Exemple 3

FIG. 8.5 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 3

La Ferme 3 présente un point faible important sur le Critère 1. Le problème se situe à la fois au niveau de l'approvisionnement en eau (g_2, avec un score de 3) et au niveau de l'approvisionnement en nourriture (g_1, avec un score de 12). Un score de 3 sur le Critère 1 est suffisant pour expliquer l'affectation de la ferme à la catégorie "Non classé". En passant au-dessus de 10 sur ce critère, la ferme sera alors affectée à la catégorie *"Normal"*, et en passant au-dessus de z_3 (c.-à-d. si elle obtenait au minimum 20) elle pourrait même être affectée à la catégorie *"Bon"* du fait de ses deux bons résultats sur les Critères 2 et 4.

Au niveau du SC2, comme pour la Ferme 1, le problème est dû à un nombre d'abreuvoirs insuffisant. En effet, les points d'eau existants ne permettent qu'à 20 vaches en lactation d'avoir un abreuvement correct, alors que 80 vaches sont présentes. L'éleveur doit impérativement installer des points d'eau supplémentaires. Une fois ces nouveaux abreuvoirs installés, la ferme pourra obtenir le score maximal sur g_2 (comme la Ferme 1).

Au niveau du g_1, le score s'explique par le fait que 50% des vaches présentes dans le troupeau sont maigres. Le conseiller pourra alors proposer à l'éleveur de revoir son plan d'alimentation, en optimisant la quantité et la nature des aliments à fournir à chaque vache en fonction de son stade de lactation et de sa production laitière. Ainsi, l'éleveur devrait pouvoir corriger le mauvais état d'engraissement de ses animaux, tout en minimisant les coûts alimentaires.

8.2 Exemples de fermes affectées à la catégorie *"Normal"*

Nous présentons dans cette deuxième section des exemples de fermes ayant été affectées à la **catégorie *"Normal"***, en montrant pour chacune d'elles les points à améliorer pour lui permettre de passer dans la catégorie supérieure.

Pour pouvoir être affectée à la catégorie *"Bon"*, une ferme doit battre z_2 sur deux critères et obtenir un score supérieur à z_3 sur les deux autres critères. Cette ferme doit donc obtenir au moins 50 points sur deux critères et au moins 20 sur les deux autres.

8.2.1 Exemple 4

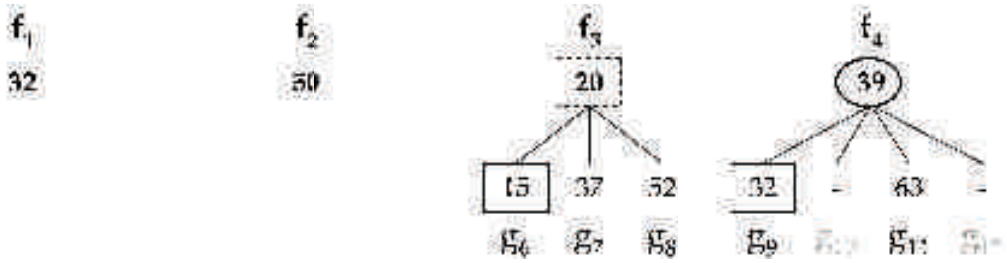


FIG. 8.6 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 4

Cette ferme est typiquement une ferme de la catégorie *"Normal"* puisque pour l'ensemble des critères elle obtient des scores compris entre z_3 et z_2. Pour passer à la catégorie *"Bon"* l'éleveur pourra améliorer le Critère 1 et/ou le Critère 4 (en les faisant passer respectivement de 32 et 39 à au moins 50), l'amélioration d'un seul des deux suffisant. Du point de vue des animaux il pourra cependant être souhaitable qu'il améliore aussi le Critère 3 qui n'obtient qu'un score de 20, tout juste équivalent à z_3. Dans tous les cas, il reste des améliorations substantielles à apporter pour passer à la catégorie supérieure.

Sur le Critère 3 le problème principal est dû à g_6 qui n'obtient qu'un score de 15. En regardant de plus près l'évaluation des altérations du tégument et des boiteries, on se rend compte qu'elle obtient un score partiel de seulement 4 pour les altérations du tégument contre 21 pour les boiteries. L'éleveur devrait donc regarder en priorité ce qui blesse ses animaux, et notamment ce qui occasionne des plaies et/ou des gonflements (75% des vaches présentant au moins une blessure sérieuse). En fonction du type de blessures et de leur positionnement, il serait intéressant de regarder avec l'éleveur ce qui, dans le bâtiment, peut blesser les animaux.

8.2.2 Exemple 5

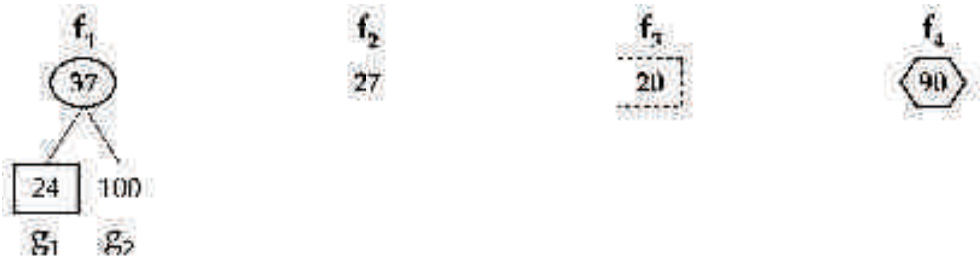


FIG. 8.7 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 5

La Ferme 5 présente un atout majeur sur le Critère 4 (avec un score de 90). Malheureusement pour elle, cet atout est masqué par des résultats médiocres à moyens sur les trois autres critères. Cependant, l'éleveur pourrait assez facilement faire passer son exploitation à la catégorie *"Bon"* en portant ses efforts sur le Critère 1. En effet, en passant de 37 à 50 sur ce critère, la ferme serait affectée à la catégorie supérieure. L'approvisionnement en eau ne pose aucun problème (score de 100), ce qui n'est pas le cas de l'approvisionnement en nourriture (score de 24, soit 28% des vaches qui sont maigres). Comme pour la Ferme 3, un plan d'alimentation mieux conçu pourrait règler le problème. Toutefois, de même que pour la Ferme 4, il serait bénéfique aux animaux d'améliorer le Critère 3.

8.2.3 Exemple 6

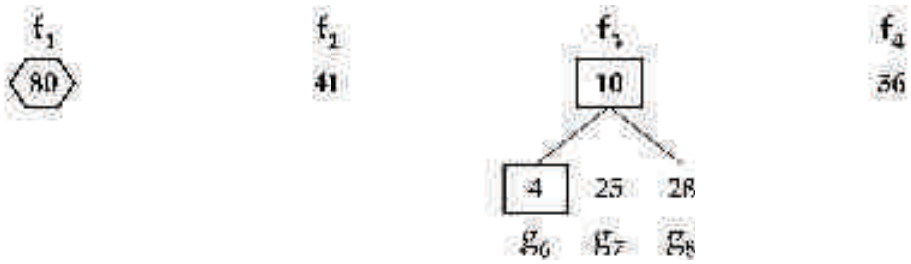


FIG. 8.8 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 6

La Ferme 6 présente un profil hétérogène, avec un point fort sur le Critère 1 (score de 80), mais un point particulièrement faible sur le Critère 3 (avec un score de seulement 10). Pour passer à la catégorie supérieure, l'éleveur doit impérativement améliorer le Critère 3, en faisant en sorte qu'il atteigne au moins un score de 20.

Au niveau du Critère 3, le problème est lié au SC6 sur lequel la ferme n'obtient qu'un

score de 4. La Ferme 6 obtient ce score du fait que la totalité des vaches observées présentent des blessures sérieuses, et que 83% des vaches boîtent. Compte tenu du nombre important à la fois de boiteries et de blesssures, il pourrait être utile de vérifier en priorité l'aire de couchage à logettes de l'étable.

8.3 Exemples de fermes affectées à la catégorie *"Bon"*

Nous présentons dans cette dernière section des exemples de fermes ayant été affectées à la **catégorie *"Bon"***, en montrant pour chacune d'elles les points à améliorer pour lui permettre de passer dans la catégorie supérieure.

Pour pouvoir être affectée à la catégorie *"Excellent"*, une ferme doit battre z_1 sur un critère et obtenir un score supérieur ou égal à z_2 sur les trois autres critères. Cette ferme doit donc obtenir au moins 75 points sur un critère et au moins 50 sur les trois autres.

8.3.1 Exemple 7

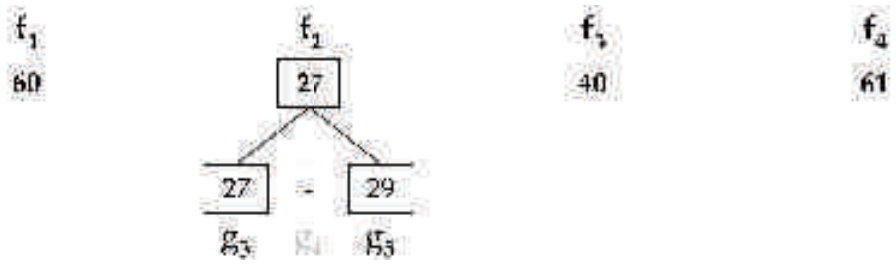


FIG. 8.9 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 7

La Ferme 7 a beaucoup de progrès à accomplir pour passer à la catégorie supérieure. Elle devra s'améliorer sur l'ensemble des critères, et tout particulièrement sur le Critère 2 pour lequel elle devra passer d'un score de 27 à au moins 50, ainsi que sur le Critère 3 sur lequel elle devra passer de 40 à 50.

Au vu des scores obtenus par la Ferme 7 sur les Sous-critères 3 et 5 composant le Critère 2, l'éleveur devra porter ses efforts à la fois sur le confort autour du repos de ses vaches (g_3, score de 27) et sur la facilité de déplacement (g_5, score de 29).

Le principal problème à solutionner est le manque flagrant de propreté des animaux. En effet, sur cette ferme 81% des vaches sont sales (voire très sales) au niveau de la mamelle, 92% le sont au niveau du flanc et de la cuisse et 77% le sont au niveau des

pattes. Il semble donc qu'il y ait un manque d'entretien et de nettoyage des stalles. L'éleveur pourrait y remédier en les nettoyant plus souvent.

De plus, quand elles se couchent les vaches mettent un peu trop de temps et heurtent trop souvent l'équipement des stalles. Les stalles ne sont peut-être pas correctement adaptées aux animaux (par ex. barre au garrot trop basse pour des animaux dont la taille a significativement augmenté au cours des vingt dernières années). Solutionner ce problème semble beaucoup plus délicat et coûteux puisqu'il est lié directement au bâtiment.

L'éleveur a donc tout intérêt à mettre l'effort sur la propreté, ce qui devrait déjà permettre une augmentation significative du score sur le Critère 2.

8.3.2 Exemple 8

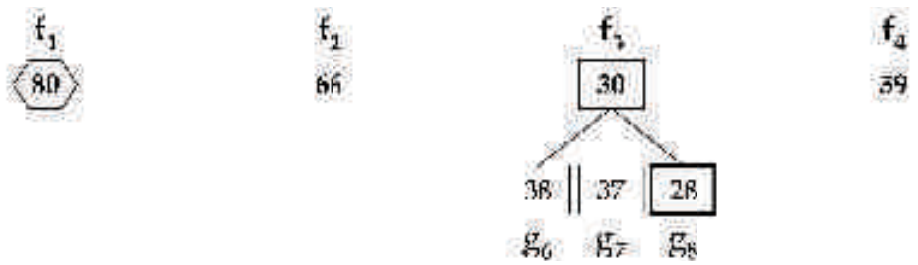


FIG. 8.10 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 8

La Ferme 8 présente un point fort au niveau du Critère 1 avec un score de 80 ($> z_1$) et un point faible sur le Critère 3 avec un score de seulement 30. En améliorant le Critère 3 (en faisant passer le score de 30 à au moins 50) cette ferme pourrait être affectée à la catégorie *"Excellent"*.

Pour ce faire, l'éleveur devra faire des efforts sur l'ensemble des sous-critères composant le Critère 3, et notamment sur le Sous-critère 8 *"Absence de douleur causée par les pratiques d'élevage"*. Le problème sur cet élevage vient de la pratique d'écornage utilisée. En effet, les bourgeons de cornes chez les veaux sont brûlés par cautérisation sans anesthésique ni analgésique, ce qui correspond à un score de 28. En employant un anti-inflammatoire au moment de la cautérisation, le score passerait à 51. L'éleveur a donc tout intérêt à modifier sa pratique d'écornage pour ainsi moins faire souffrir les animaux.

8.3.3 Exemple 9

FIG. 8.11 – Scores obtenus au niveau des critères et des sous-critères par la Ferme 9

La Ferme 9 appartient de justesse à la catégorie *"Bon"*, en effet, elle bat de justesse z_3 sur le Critère 4 *"Comportement approprié"* avec un score de seulement 20. Pour passer à la catégorie supérieure, l'éleveur devra donc apporter des améliorations sur la plupart des critères, mais devra surtout faire passer le score du Critère 4 de 20 à au moins 50 (les autres critères étant d'ores et déjà au-dessus de z_2, seuil d'exigence minimale pour la catégorie *"Excellent"*).

En examinant de plus près le Sous-critère 9 (*"Expression des comportements sociaux"*) et le Sous-critère 11 (*"Bonne relation homme-animal"*), on peut constater que ces deux sous-critères sont à améliorer.

Le score sur g_9 est le plus problématique avec une valeur de seulement 19, s'expliquant par un nombre élevé d'interactions agonistiques entre les animaux du troupeau (avec en moyenne près de 3.5 comportements agressifs par vache et par heure). Le conseiller devra alors voir avec l'éleveur pour quelles raisons le nombre de comportements agressifs est si élevé.

Le score obtenu sur g_{11} n'est pas non plus très élevé puisqu'il n'obtient qu'un score de 26 qui s'explique par le fort pourcentage d'animaux présentant des distances de fuite longues (en effet, 35% des vaches du troupeau reculent quand l'homme s'approche à plus de 1 m d'elles). Le conseiller pourra alors étudier l'attitude de l'éleveur vis-à-vis de ses animaux et tenter de lui donner des solutions pour améliorer la relation entre ses vaches et lui.

8.4 Vers une automatisation de la procédure de diagnostic-conseil

Il est possible de distinguer deux procédures différentes utiles à l'éleveur et pouvant être en grande partie automatisées :

1. **procédure de diagnostic**, permettant d'expliquer pourquoi la ferme x_k a été affectée à la catégorie C_i,
2. **procédure de conseil**, permettant d'aider l'éleveur à porter ses efforts sur les éléments qui en ont le plus besoin, ou sur les éléments dont l'amélioration lui permettrait de passer à la catégorie supérieure.

8.4.1 Procédure de diagnostic

Pour être affectée à la catégorie C_i, la ferme x_k doit avoir des performances telles que $x_k \succsim z_i$ et non$(x_k \succsim z_{i-1})$. L'affirmation $x_k \succsim z_i$ est vraie si et seulement si la coalition concordante $C(x_k, z_i)$ est suffisamment forte et que $C(x_k, z_{i+1}) = \{1, 2, 3, 4\}$. L'affirmation $x_k \succsim z_{i-1}$ est fausse si et seulement si $C(x_k, z_{i+1})$ n'est pas suffisamment forte ou si $\exists\ j\ tq\ j\ \notin\ C(x_k, z_i)$.

TAB. 8.1 – Tableau des performances et des indices de concordance partielle de x_6 (les données de couleur grise sont indépendantes des performances de x_6)

	f_1	f_2	f_3	f_4	
w_j	1	1	1		1
z_{1j}	75	75	75		75
z_{2j}	50	50	50		50
z_{3j}	20	20	20		20
z_{4j}	10	10	10		10
x_{kj}	**80**	**41**	**10**	**56**	
$c_j(x_6, z_1)$	1	0	0	0	
$c_j(x_6, z_2)$	1	0	0	1	
$c_j(x_6, z_3)$	1	1	0	1	
$c_j(x_6, z_4)$	1	1	1	1	

TAB. 8.2 – Tableau des coalitions de critères concordants et des indices de concordance de la Ferme 6 (les données de couleur grise sont indépendantes des performances de x_6)

	z_1	z_2	z_3	z_4	
$C(x_6, z_i)$	$\{1\}$	$\{1,4\}$	$\{1,2,4\}$	$\{1,2,3,4\}$	
$c(x_6, z_i)$	0.25	0.5	0.75	1	
λ_i	0.25	0.5	0.75		1

À partir des indices de concordance il est possible d'expliquer pourquoi la ferme est affectée à la catégorie "Normal" : sa coalition de critères concordants avec $x_6 \succsim z_3$ est insuffisante pour permettre à la Ferme 6 d'être affectée à la catégorie "Bon". En effet cette coalition ne comprend que trois critères alors qu'il faudrait les quatre :

$$(1) \quad \left.\begin{array}{l} c(x_6, z_4) = 1 \\ c(x_6, z_3) = 0.75 = \lambda_3 \end{array}\right\} \quad \Rightarrow \quad x_6 \succsim z_3$$

$$(2) \quad c(x_6, z_3) = 0.75 < 1 \quad \Rightarrow \quad \text{non}(x_6 \succsim z_2)$$

$$\text{d'où} \quad \left.\begin{array}{l} x_6 \succsim z_3 \\ \text{non}(x_6 \succsim z_2) \end{array}\right\} \quad \Rightarrow \quad x_6 \in C_3 \text{ "Normal"}$$

Il est alors possible de pointer automatiquement le critère n'ayant pas obtenu une performance suffisante pour battre z_3, il s'agit du seul critère f_j pour lequel $c_j(x_6, z_3) = 0$: le critère f_3.

En parallèle de ce diagnostic au niveau des critères, un message d'alerte devrait parvenir à l'éleveur concernant f_3 puisque $z_{43} \leq f_3(x_6) \leq (z_{43} + \varepsilon)$. En cas de rehaussement ou de légère dégradation de la performance de x_6 sur f_3, x_6 serait retrogradée à la catégorie C_4 "Non classé".

La procédure permettant de pointer le(s) critère(s) empêchant le passage dans la catégorie supérieure et la procédure permettant d'alerter l'éleveur sur une performance trop proche d'un profil indiquent, dans le cas de la Ferme 6, toutes les deux le même critère : f_3.

On peut alors imaginer l'ouverture par le logiciel d'une boîte de dialogue demandant à l'éleveur sur quel critère (dans le cas de la Ferme 6 il n'y en a qu'un) il souhaite voir s'afficher les scores qu'il a obtenu au niveau des sous-critères composant le critère f_j en question (pour la Ferme 6 il s'agit de f_3).

Le logiciel pourrait alors afficher un diagramme du type de ceux présentés dans les exemples détaillés dans les sections précédentes, permettant de visualiser les sous-

critères ayant obtenu des scores inférieurs ou égaux à l'objectif à atteindre sur f_j, correspondant à $z_{(i+1)j}$ pour une affectation à C_i[1]. Ainsi, si $g_l(x_k) \leq z_{(i+1)j}$ alors colorer la case correspondant au Sous-critère l sur le diagramme. Dans le cas de la Ferme 6, ce serait alors la case correspondant au Sous-critère 6 qui serait mise en évidence puisque c'est le seul critère entrant dans la construction de f_3 qui ait obtenu un score $g_6(x_6) \leq z_{43}$.

Une fois de plus, le logiciel pourrait ouvrir une boîte de dialogue demandant à l'éleveur sur quels sous-critères il souhaiterait voir le détail des mesures. Dans le cas de notre exemple, seul le critère g_6 serait proposé. Le logiciel pourrait alors faire apparaître un tableau présentant toutes les données brutes recueillies sur le terrain et servant à évaluer le sous-critère considéré, avec un bref rappel de la manière dont les données brutes ont été traitées pour pouvoir produire le score de sous-critère. Avec ce tableau ne présentant que quelques mesures, l'éleveur pourra facilement identifier les problèmes.

Nous pouvons noter que les boîtes de dialogues proposées ici sont facultatives, il est tout à fait possible de faire en sorte que le logiciel réponde automatiquement aux questions, en détaillant tous les critères identifiés dans la procédure comme cause de non-affectation à la catégorie supérieure ou comme source de risque potentiel (performance sur ce critère trop proche de la valeur d'un profil).

8.4.2 Procédure de conseil

La procédure de conseil se baserait sur celle de diagnostic en reprenant le tableau des indices de concordance, les conclusions concernant les critères empêchant l'affectation à la catégorie supérieure, et les conclusions sur les critères présentant une performance trop proche d'un profil. Elle présenterait à l'éleveur une synthése des *objectifs à atteindre au niveau des critères* pour passer à la catégorie supérieure :

- Quels sont les critères sur lesquels l'éleveur a intérêt à porter ses efforts ?
- Sur combien de critères, au minimum, l'éleveur doit-il porter ses efforts pour passer à la catégorie supérieure ?
- Quel score l'éleveur doit-il obtenir au minimum sur ce(s) critère(s) ?

Dans le cas de critères équipondérés, pour une ferme x_k affectée à la catégorie C_i, $[4 \times (1 - c(x_k, z_i))]$ représente le nombre de critères sur lesquels la ferme doit obligatoirement

[1] pour agréger les scores de sous-critères en un score critère nous utilisons l'intégrale de Choquet. Cet agrégateur est un opérateur de compromis qui fournit donc une score compris entre le score de sous-critère minimal et le score maximal. Ainsi, si on veut être sûr que l'objectif sera atteint sur le critère, on peut s'assurer qu'aucun score de sous-critère n'est inférieur à la valeur à atteindre sur le critère.

s'améliorer pour respecter le niveau d'exigence minimale de la catégorie supérieure C_{i-1}, et $\{f_1, f_2, f_3, f_4\} \backslash C(x_k, z_i)$ représente la coalition de critères devant impérativement être améliorés. Le détail de l'évaluation de ces critères pourrait alors être présenté à l'éleveur, selon la même procédure que celle décrite dans la section précédente pour identifier les sous-critères pénalisant, puis les mesures responsables de ces mauvais résultats.

Une boîte de dialogue pourrait s'ouvrir pour demander à l'éleveur si il désire simuler les améliorations requises sur cet ensemble de critères. Dans ce cas, la procédure pourrait alors s'intéresser au niveau d'aspiration de la catégorie C_{i-1}. $4 \times (\lambda_{i-1} - c(x_k, z_{i-1}))$ représenterait alors le nombre de critères devant être améliorés sur la ferme x_k pour que celle-ci puisse être affectée à la catégorie C_{i-1}, et $\{1, 2, 3, 4\} \setminus C(x_k, z_{i-1})$ représente l'ensemble des critères susceptibles d'être améliorés. Si $|\{1, 2, 3, 4\} \setminus C(x_k, z_{i-1})| > [4 \times (\lambda_{i-1} - c(x_k, z_{i-1}))]$ alors plusieurs scénarios d'amélioration se présentent à l'éleveur. Puisque le nombre de critères pouvant être améliorés est supérieur au nombre de critères devant être améliorés, l'éleveur devra choisir le (ou les) critère(s) sur le(s)quel(s) il portera ses efforts. Ce choix pourra être déterminé en fonction de la facilité (ou au contraire de la difficulté) d'amélioration des différents critères "candidats".

La procédure de conseil pourrait alors reprendre la partie de l'outil de diagnostic permettant de questionner l'éleveur via des boîtes de dialogues interactives afin de permettre à l'éleveur de connaître les scores de sous-critères, puis les données brutes, correspondant aux critères sur lesquels il a décidé de porter ses efforts.

La procédure automatisée permettra d'aider l'éleveur, mais c'est l'éleveur qui au final décidera de sa stratégie.

En conclusion, le modèle peut encourager les éleveurs à améliorer le niveau de bien-être de leurs animaux en leur montrant comment passer d'une catégorie donnée à la catégorie supérieure. Le système de conseil pourrait cependant induire des effets pervers dans certaines situations dans lesquelles l'éleveur n'a pas intérêt à porter en priorité ses efforts sur le point le plus faible, mais plutôt sur le critère qui a obtenu un score assez proche du profil supérieur (cf. Exemples 4 et 5). Ces risques seraient cependant limités puisque :

- ces risques ne concernent que les procédures ne demandant pas l'unanimité,
- conserver une performance faible est risqué pour l'éleveur car en cas de rehaussement des profils de référence (où de légère dégradation de ses performances), le fait d'avoir garder cette valeur faible pourrait conduire à un déclassement de la ferme, alors qu'en cherchant à améliorer un autre critère, plus proche du profil de référence supérieur, l'éleveur espérait au contraire passer à la catégorie supérieure.

Nous pouvons aussi voir à l'aide de ces neuf exemples que le modèle d'évaluation, construit sur la base d'une procédure d'agrégation séquentielle, peut fournir une base à un outil de diagnostic (pouvant être automatisé), qui, pour être pleinement efficace, devra être complété par un outil de conseil aux éleveurs. Il permet en effet de visualiser facilement les points forts et les points faibles d'un élevage, et permet de cibler les mesures qui posent problème. C'est alors le travail du conseiller de mettre en évidence les causes de ces mauvaises mesures et de proposer des actions correctrices.

Les solutions correctrices ne sont pas toutes difficiles à mettre en place. Il se peut par exemple que des blessures soient simplement causées par une arête en béton trop franche dans le couloir que les vaches empruntent pour aller à la traite et à laquelle les vaches se blessent quotidiennement. D'autres problèmes peuvent cependant être un peu plus délicats à identifier, puis à régler. Par exemple, dans le cas de la Ferme 9 où de fréquents comportements agonistiques sont observés, le conseiller devra rechercher la cause de ce surcroît d'agressivité au sein du troupeau. Il devra alors réaliser une visite approfondie de la ferme, une observation détaillée des animaux ainsi qu'un entretien avec l'éleveur sur ses pratiques. Il pourra alors mettre en évidence des causes très variées, telles que l'introduction régulière de nouveaux individus dans le troupeau (ce qui fait qu'une nouvelle hiérarchie doit s'installer au sein du troupeau, d'où une augmentation des comportements agressifs) ou encore le positionnement inadéquat des points d'alimentation, d'eau ou d'autres éléments tels que des brosses (auxquelles les vaches adorent se frotter).

Discussion

Résumé de la discussion :

Le modèle d'évaluation du bien-être des vaches laitières en ferme proposé dans cette thèse présente plusieurs atouts par rapport aux modèles d'évaluation existants. En premier lieu, il respecte l'aspect multidimensionel du bien-être. En se basant sur une construction formalisée de l'évaluation du bien-être, il permet de produire des résultats transparents via une procédure d'évaluation facile à standardiser et à automatiser. Il présente de surcroît un bon pouvoir explicatif permettant au modèle de servir de base à un diagnostic-conseil efficace en élevage. De plus, le modèle d'évaluation proposé est flexible, pouvant s'adapter à des modifications de l'élevage et des attentes sociétales en Europe, et intégrer des connaissances scientifiques nouvelles.

Dans la suite de Welfare Quality®, la liste des mesures de bien-être pourra être affinée (d'ici l'été 2008) et des observations complémentaires seront réalisées dans des fermes européennes. Il sera alors nécessaire de revoir la construction des sous-critères et des critères. Les paramètres de ces constructions pourront alors être définis en faisant appel à un nombre plus important d'experts. Le choix final des règles d'agrégation des critères en une évaluation globale du bien-être devra alors être effectué, en étroite collaboration avec des utilisateurs potentiels du système d'évaluation. Une fois le modèle final d'évaluation établi, il serait alors souhaitable d'en vérifier la sensibilité, la robustesse et la légitimité, puis de le valider (notamment au niveau expérimental et opérationnel).
Le modèle d'évaluation développé pour les vaches laitières en ferme pourra alors être transposé aux autres types d'animaux, en ferme et durant la phase de transport/abattage, concernés par le projet Welfare Quality®.

Enfin, ce type de modèle d'évaluation multicritère peut s'appliquer à d'autres concepts multidimensionnels dans le domaine agricole, et notamment afin d'évaluer la durabilité des élevages.

Discussion

Comme nous avons pu le voir dans le chapitre introductif, le respect du bien-être des animaux de rente fait l'objet d'attentes sociétales de plus en plus fortes. C'est dans ce contexte que le projet européen Welfare Quality® a été mis en place. Ce projet vise à construire un standard en matière d'évaluation et d'information sur le bien-être des bovins, des porcs et des volailles, aussi bien en ferme que dans la phase de transport et d'abattage. Le bien-être animal se compose de plusieurs dimensions, telles que la santé, le comportement ou l'absence de stress. Son évaluation repose donc sur un ensemble de mesures, complémentaires les unes des autres. L'objet de la thèse était d'élaborer un modèle d'évaluation globale du bien-être des vaches laitières reflétant l'aspect multidimensionnel du bien-être. Nous avons proposé dans ce travail une approche, utilisant des méthodes développées en évaluation multicritère, basée sur une structure séquentielle dans laquelle plusieurs critères complémentaires, non réductibles à un seul, sont construits puis agrégés afin d'obtenir une évaluation globale.

Dans un premier temps nous avons effectué une revue des méthodes actuellement proposées pour produire une évaluation globale du bien-être animal. Ces différentes méthodes présentent chacune des avantages et des inconvénients par rapport à une utilisation répondant aux objectifs de Welfare Quality®.

Nous avons ensuite identifié les contraintes liées à l'évaluation globale du bien-être animal. Ces caractéristiques sont liées au concept même de bien-être, à l'interprétation des mesures en terme de bien-être et à la façon dont les données sont recueillies sur le terrain.

Une fois identifiés les problèmes liés à la construction d'un modèle d'évaluation globale du bien-être animal, nous avons défini un ensemble de quatre critères (*Alimentation adaptée*, *Logement correct*, *Bonne santé* et *Comportement approprié*) et 12 sous-critères, valables quelle que soit l'espèce animale considérée.

La première étape de l'élaboration du modèle d'évaluation a donc consisté à construire les sous-critères à partir de 44 mesures effectuées sur le terrain. L'évaluation de chaque

sous-critère est présentée sous la forme d'un score allant de 0 (la pire situation envisageable) à 100 (la situation idéale), où le 50 correspond à une situation neutre, ni bonne ni mauvaise. Cette échelle d'utilité est commune à l'ensemble des sous-critères. Le calcul des scores de sous-critères à partir des données brutes varie en fonction du nombre de mesures à agréger, de leur nature et de leur précision (sommes pondérées associées à une fonction d'utilité, méthode lexicographique...).

Ensuite, les sous-critères composant un même critère ont été agrégés. Pour ce faire, nous avons utilisé une méthode qui permet de limiter les compensations entre les sous-critères en attribuant plus d'importance aux sous-critères ayant obtenus les moins bons scores.

Enfin, pour agréger les critères en une évaluation globale nous avons proposé d'utiliser une méthode de comparaison à des profils prédéfinis délimitant des catégories ordonnées de bien-être.

A chaque étape, le modèle d'évaluation a été paramétré sur la base d'avis d'experts. Les experts interrogés étaient des chercheurs en sciences animales (comportement, santé ou bien-être animal) pour la construction des sous-critères, des chercheurs en sciences animales et en sciences sociales pour l'agrégation des sous-critères en critères, et ces chercheurs et des utilisateurs futurs du standard d'évaluation pour l'étape d'agrégation finale.

Un modèle "vaches laitières" présentant plusieurs atouts

Grâce à l'approche séquentielle, reposant sur une famille de quatre critères et douze sous-critères, et à la limitation des compensations entre les grandes dimensions du bien-être animal, notre modèle d'évaluation permet de tenir compte de la *multidimensionalité du bien-être animal.*

Contrairement à certaines méthodes actuellement proposées pour produire une évaluation globale du bien-être animal, le modèle d'évaluation proposé ici permet de *tenir compte de l'importance relative des différentes mesures*, mais aussi de différences éventuelles d'importance entre les sous-critères constituant un même critère et de la manière dont ces sous-critères interviennent dans l'évaluation des critères. De par la construction des sous-critères, il permet aussi d'utiliser des méthodes de calcul adaptées au nombre de mesures à prendre en compte et à leur nature, ce qui n'aurait pas pu être le cas si une seule et même méthode avait été utilisée pour construire l'ensemble de sous-critères.

Un soin particulier a aussi été apporté à la répartition des mesures pour la construction des sous-critères, et ce afin d'éviter de prendre en compte *plusieurs fois la même*

information dans des sous-critères différents.

En mettant en place ce modèle d'évaluation du bien-être, nous pouvons *formaliser le raisonnement* suivi par des scientifiques (en sciences sociales et animales), tout en incluant les attentes et le mode de raisonnement souhaité par des utilisateurs potentiels du système d'évaluation globale du bien-être animal qui résultera du projet Welfare Quality®. Même s'il reste du travail au sein du projet pour arriver à un modèle recevant l'assentiment de toutes les parties intéressées par la mise en place d'un tel système, les méthodes de consultation ont d'ores et déjà été en grande partie établies dans le cadre de la thèse. Les discussions devront être poursuivies avec les futurs utilisateurs du système d'évaluation afin d'établir les choix finaux pour l'agrégation des critères (notamment le choix des règles d'appartenance aux différentes catégories). Les scénarios et simulations établis dans la thèse serviront de base à ces discussions.

Le modèle proposé repose sur une structure d'agrégation séquentielle utilisant des méthodes pouvant être expliquées, au moins dans les grandes lignes, aux utilisateurs, ce qui permet de maintenir l'information disponible à chaque étape de l'agrégation. Il en résulte un modèle d'évaluation qui évite l'écueil des "boîtes noires" et rend la procédure d'évaluation *transparente*. Il en découle aussi que la procédure d'évaluation est *facile à standardiser*. Ce point peut se révéler être une condition indispensable dans le cadre d'une évaluation devant être réalisée sur un grand nombre de fermes, comme cela peut être le cas dans une procédure de certification.

Les mesures à prendre en compte ainsi que la manière de les utiliser pour l'évaluation sont clairement identifiées dans la procédure formalisée que nous proposons. Associé à cette formalisation de la procédure, le fait qu'elle soit aussi facilement standardisable permet à la procédure d'évaluation que nous proposons de pouvoir, sans grande difficulté, être *automatisée*. Cet atout est particulièrement utile dans le cadre d'évaluations devant être répétées plusieurs fois, comme c'est le cas dans une procédure de certification, utilisation principale visée par Welfare Quality®.

Par ailleurs, l'information générée à chaque étape de la procédure d'évaluation (des mesures à l'évaluation globale) permet de justifier les résultats obtenus et de servir de base à des recommandations (comme nous avons pu le montrer sur quelques exemples de fermes dans le Chapitre 8). Ainsi, le modèle d'évaluation peut aisément servir à un diagnostic suivi d'un conseil auprès des éleveurs.

Le modèle, tel qu'il est paramétré dans la thèse, *limite fortement les compensations* entre les sous-critères composant un même critère, et permet aussi, en fonction

de la procédure considérée, de les limiter entre les critères. Cette caractéristique non-compensatoire du modèle d'évaluation a pour effet d'encourager les éleveurs à améliorer en priorité les points les plus problématiques observés sur leurs fermes. En effet, dans le cadre d'une visite à intervalle régulier (par exemple tous les deux ans), l'éleveur qui lors de la première visite a été mal évalué, en raison d'un aspect particulier qui était défaillant, peut en y remédiant améliorer de manière très significative son évaluation lors de la visite suivante.

Par ailleurs, le modèle permet facilement de mettre en évidence les points pénalisants, mais aussi les points forts de l'élevage, et ce, d'une façon pouvant être assez facilement automatisée. Ainsi, l'éleveur n'aura pas de difficultés à identifier les points les plus problématiques présents sur son exploitation (phase de diagnostic). Avec la mise en place d'une procédure de conseil auprès des éleveurs, ces derniers pourront identifier les causes potentielles des problèmes de bien-être observés et y remédier. De plus, le WP4.3 de Welfare Quality® va constituer une base documentaire permettant de recommander des stratégies correctrices pour chaque sous-critère posant problème. Le modèle permet donc à la fois d'encourager les éleveurs à améliorer les conditions de vie de leurs animaux et de les aider dans leur démarche en leur proposant un outil de diagnostic selon une procédure pouvant être en partie automatisée ainsi que des solutions correctrices aux principaux problèmes pouvant être rencontrés. Ce diagnostic devrait fortement orienter le conseil auprès des éleveurs afin qu'ils mettent en place les actions correctrices les plus efficaces pour leur élevage.

Ainsi, la capacité du modèle à pouvoir expliquer les évaluations obtenues, associée au fait qu'il incite les éleveurs à améliorer en priorité les problèmes les plus sérieux, devrait permettre de *favoriser l'instauration d'un cercle vertueux*, bénéfique à l'amélioration des conditions d'élevage en Europe (voir par exemple Webster 2005). En effet, en permettant aux éleveurs de cibler les points à améliorer en priorité et en les incitant à les résoudre, les actions correctrices auront alors toutes les chances d'être efficaces.

Les améliorations apportées par l'éleveur peuvent être visibles même au niveau de l'évaluation globale grâce au *pouvoir de discrimination* du modèle. Il permet en effet d'affecter les fermes à quatre catégories ordonnées de bien-être, ce qui permet un suivi des améliorations plus fin que dans le cadre de systèmes ne produisant qu'une évaluation binaire du type "bon" *vs* "insuffisant" (ce qui est par exemple le cas des méthodes actuelles de type "checklist", cf. Chapitre 2).

Nous avons choisi de comparer les fermes à des normes prédéfinies de bien-être, et non de comparer directement les fermes entre elles. Nous pourrions toutefois le faire indirectement, par comparaison de l'appartenance respective des fermes aux différentes

catégories de bien-être. Cette comparaison indirecte des fermes ne permet qu'une relativement faible discrimination des fermes (toutes les fermes affectées à une même catégorie ne pouvant être différenciées). Cependant, si la possibilité de classer les fermes les unes par rapport aux autres devenait une exigence des utilisateurs du système d'évaluation, il serait possible de le faire en modifiant la règle d'agrégation des scores de critères. En effet, en utilisant une méthode, telle que l'intégrale de Choquet, produisant un score global à partir des critères, il serait possible de produire un classement des fermes, de celle offrant aux animaux le moins bon niveau de bien-être à la meilleure. Nous mettons cependant en garde les futurs utilisateurs contre une utilisation de ce classement pour établir le niveau de bien-être offert aux animaux par chacune d'elles. Il serait en effet possible de considérer que les 10% des fermes les mieux classées devraient être affectées à la catégorie *"Excellent"*, et ainsi de suite... Tout d'abord, si l'ensemble des fermes offre un niveau de bien-être très faible aux animaux, certains élevages, un peu meilleurs que les autres, seraient alors considérés comme excellents, ce qui ne traduirait pas une réalité. De même, dans le cas contraire où toutes les fermes respectent le bien-être des animaux, certaines d'entre elles seraient considérées comme mauvaises. De plus, les fermes concernées constituent un ensemble susceptible de varier entre deux évaluations successives (par ex. les fermes désirant être labellisées au sein d'une procédure de certification peuvent varier d'une année sur l'autre) et le classement dépend des fermes qui constituent cet ensemble. Enfin, même si l'ensemble des fermes considérées reste stable entre deux évaluations, dans le cas où toutes les fermes améliorent le niveau de bien-être de leurs animaux entre les deux visites, les éleveurs ne verraient pas leurs efforts récompensés puisque le classement final des fermes risquerait d'être le même. Nous nous sommes au contraire placés dans une problématique de tri et notre modèle permet à une ferme d'être affectée une année donnée à la catégorie *"Normal"* de voir ses efforts récompensés par un passage à la catégorie *"Bon"* l'année suivante. Pour cela, nous proposons une procédure permettant de *produire une évaluation intrinsèque des fermes*, eu égard à *des normes définies* dans le paramétrage du modèle (ces normes pouvant évoluer si nécessaire, cf. paragraphe suivant).

Le système d'évaluation qui sera produit pourra en plus être *flexible*, via des réajustements du paramétrage du modèle, voire, si nécessaire, une révision de la construction des sous-critères. Ainsi, le système pourra s'adapter aux évolutions :

- de l'élevage en Europe,
- des attentes sociétales,
- des connaissances scientifiques dans le domaine du bien-être animal.

Dans le cas d'une utilisation dans une procédure de certification, il ne semble cepen-

dant pas bénéfique de procéder à des réajustements fréquents afin de ne pas déstabiliser les utilisateurs du système d'évaluation, et en particulier les éleveurs. Par contre, le système dans son ensemble pourra être revu au bout de quelques années si le besoin s'en fait sentir. On peut ici faire le parallèle avec les normes de classement des hôtels en France qui sont revues régulièrement. Ainsi l'arrêté du 14/04/1953 fixant la procédure et les normes de classement des hôtels de tourisme a été abrogé par l'arrêté du 16/12/1964, lui-même modifié par la suite. Les normes et la procédure de classement des hôtels actuellement en vigueur en France ont été mises en place par l'arrêté du 14/02/1986, six catégories d'hôtels sont alors définies : 0, 1, 2, 3, 4 étoiles et 4 étoiles luxe[2].

Enfin, le système d'évaluation permet de répondre aux attentes en termes d'utilisation des différents utilisateurs potentiels (listées dans le Chapitre 7).

Vers le modèle final d'évaluation du bien-être des vaches laitières en ferme

Nous proposons ici quelques voies de réflexion et d'amélioration du modèle qui pourront être explorées dans la suite du projet Welfare Quality®.

Gestion des données manquantes

Comme la grande majorité des méthodes servant à évaluer le bien-être animal, notre modèle suppose la réalisation de l'ensemble des mesures sur l'ensemble des fermes (ou abattoirs) évaluées. Dans le cas des vaches laitières en ferme, cela impose la réalisation des 44 mesures nécessaires à l'évaluation du bien-être des animaux. Cependant, en cas de données ponctuellement manquantes, il serait alors possible d'introduire une composante d'incertitude dans notre modèle d'évaluation, ce qui correspondrait alors à tenir compte de la composante $\mathcal{U}$ du modèle général de Tsoukiàs (2007) (cf. première section du Chapitre 4). Il serait alors possible d'envisager différents scénarios sur les données manquantes. Afin de limiter l'incertitude sur ces données, il pourrait être envisagé de se baser sur les valeurs obtenues par la ferme lors de la (ou les) précédente(s) visite(s), ce qui permettrait de réduire le nombre de scénarios.

Toutefois, cette solution ne pourrait concerner que les données ponctuellement manquantes. En effet, si l'éleveur ne tient pas correctement son registre sanitaire (où les

[2]Source : *www.legifrance.fr*

maladies de ses animaux ainsi que les traitement prodigués doivent être notés), ce manquement répété aux consignes serait alors considéré comme un motif de rejet de la ferme (par exemple, la ferme ne pourra pas participer au système de certification "bien-être"). Les éleveurs désirant être évalués devraient donc être tenus de mettre à jour correctement leurs registres.

Remaniement de la liste des mesures et réalisation d'observations supplémentaires en fermes

La liste définitive des mesures n'est pas encore disponible. Comme nous avons pu le voir dans le Chapitre 5, certains sous-critères ne peuvent être évalués pour le moment faute de mesures, alors qu'il semble que ces sous-critères soient pertinents pour les vaches laitières. Des mesures sont donc actuellement à l'étude au sein de Welfare Quality®. En parallèle, le WP2.5 sera chargé de proposer des scénarios avec des protocoles de visite plus ou moins longs. En cas de simplification du protocole de visite, la liste des mesures devra être remaniée, avec probablement suppression de certaines mesures et/ou simplification de la manière de les observer.

Il sera alors possible de compléter le modèle pour les sous-critères jusqu'à présent "orphelins", et de simplifier et/ou revoir la construction de certains sous-critères s'il s'avère que les mesures qui les composent ont changé.

En parallèle, les observations réalisées dans le WP2.4 seront complétées et il sera possible de travailler sur le jeu de données complet, contenant 91 fermes laitières. De plus, d'autres observations seront réalisées en fermes dans le cadre du WP4.6, ces données supplémentaires viendront s'ajouter aux données du WP2.4. Le modèle pourra alors être testé sur un nombre de fermes européennes plus important, ce qui permettra d'augmenter la représentativité de l'échantillon utilisé pour visualiser la répartition des fermes à laquelle on pourra s'attendre en Europe.

Affinage de la construction du modèle d'évaluation

Comme nous avons pu le voir dans les Chapitres 5 et 6, le paramétrage du modèle proposé dans cette thèse conduit à des règles sévères (au vu de la répartition des fermes obtenue sur les différents sous-critères et critères). En se basant sur la liste révisée des mesures (et donc sur des constructions des sous-critères modifiées) et sur des données recueillies sur un nombre de fermes plus important, il sera possible de revoir le paramétrage des deux premières étapes de construction de l'évaluation : le paramétrage de

la construction des sous-critères, et celui des intégrales de Choquet permettant de les agréger en critères. Il sera donc nécessaire de consulter de nouveau les experts déjà sollicités (pour les sous-critères et les critères).

Au niveau des sous-critères, il sera particulièrement intéressant de consulter plus de personnes. En effet, nous n'avons consulté dans la thèse que cinq à six experts pour les sous-critères, tous ces experts étant des scientifiques impliqués dans Welfare Quality®. Ce travail nous a permis de mettre en place la méthodologie de consultation des experts, mais il sera bien évidemment nécessaire de compléter ces premières expertises par la consultation de plusieurs autres experts. À ce niveau de la construction, les experts consultés doivent avoir une bonne connaissance des mesures et de leur signification (puisque c'est à cette étape que les données brutes sont interprétées en termes de bien-être). Il semble alors logique de consulter en priorité les chercheurs en sciences animales ayant développé et testé en fermes les mesures. Cependant, il existe d'autres personnes ayant une très bonne connaissance de l'élevage et de certaines des mesures que nous utilisons (comme par exemple l'incidence des pathologies ou l'état d'engraissement des vaches). Il peut s'agir de professionnels de l'élevage (tels que des conseillers des instituts techniques ou des chambres d'agriculture) et de vétérinaires. À terme, l'objectif sera d'avoir consulté au moins une vingtaine d'experts, apportant à la construction leurs connaissances à la fois pratiques et théoriques de l'élevage en Europe, pour affiner le modèle d'évaluation.

Au niveau des critères, nous avons dans la thèse consulté des chercheurs en sciences animales et sociales impliqués dans Welfare Quality®. Des réunions sont d'ores et déjà prévues pour revoir avec eux leurs premières évaluations (en leur montrant notamment la sévérité du raisonnement consistant à attribuer à un critère le score minimal observé sur les sous-critères). En complément, il sera là aussi possible d'élargir le spectre des experts consultés.

Au cours de ce processus de consultation, il sera primordial de revenir sur la **répartition des fermes laitières** pour montrer les évolutions qui seront apparues du fait des modifications du paramétrage de la construction. Il sera alors possible de s'engager dans un processus de consultation itératif, dont le résultat correspondrait à un paramétrage reflétant le raisonnement souhaité par les experts tout en permettant d'obtenir une répartition des fermes acceptable et cohérente avec l'idée qu'ils se font du niveau de bien-être assuré dans les élevages européens.

Du fait de la diversité des experts et de leurs domaines d'expertise, il sera peut-être utile de pondérer les opinions des experts. Le modèle proposé dans la thèse permet

d'inclure cette spécification[3], même si elle n'est pas exprimée pour l'instant. Pour aller dans une telle direction, la difficulté sera alors d'identifier les pondérations à affecter aux différents experts. Lors de la consultation des experts, nous avons demandé à chacun d'entre eux le niveau de confiance qu'ils accordaient à leurs réponses (en se servant pour cela d'une échelle ordinale en cinq niveaux : *Not at all confident* / *Quite not confident* / *Moderately confident* / *Quite confident* / *Very confident*). Mais le problème réside dans l'interprétation de ce niveau de confiance, certaines personnes ayant toujours tendance à déprécier leur avis, alors que d'autres sont toujours sûres d'elles, même si elles peuvent parfois avoir des connaissances moins poussées. Ces appréciations pourront tenir compte de la confiance moyenne d'un expert, qui pourra refléter plus son caractère que ses connaissances propres. Ainsi, il pourra être judicieux de considérer la différence entre un indice donné (pour une évaluation par un expert) et la moyenne des indices de confiance de cet expert.

Par ailleurs, nous avons pu constater des divergences d'opinion, parfois fortes entre experts. Ce fut notamment le cas pour la construction du Sous-critère 8 évaluant les mutilations, en particulier l'écornage (cf. Annexe C). Ces disparités devraient s'atténuer en présentant aux experts les données recueillies dans les élevages visités dans le cadre de Welfare Quality®. Par exemple, on peut se demander s'il est légitime d'attribuer une note de 5 à la pratique d'écornage la plus courante dans les élevages européens (cautérisation des bourgeons de corne sans anesthésie ni analgésie). En effet, même si cette pratique est loin d'être la meilleure en termes de bien-être, une telle notation se révélera en fait inacceptable pour les professionnels de l'élevage, d'autant plus que dans la plupart des pays les éleveurs ne sont pas autorisés à pratiquer eux-mêmes des anesthésies, mêmes locales. Le risque serait alors de voir les professionnels de l'élevage s'opposer au modèle et alors refuser d'utiliser le système d'évaluation proposé. Au lieu de cela, il semble beaucoup plus efficace d'attribuer une note moyenne, acceptable pour les éleveurs tout en les pénalisant de manière significative pour de telles pratiques. Ainsi le système d'évaluation pourrait être utilisé, et de ce fait, aurait pour conséquences d'inciter les éleveurs à modifier leurs pratiques. Par ailleurs, nous espérons qu'en demandant à plus d'experts venant d'horizons différents, l'écart-type entre les réponses des experts diminuera (il restera toujours quelques personnes aux extrêmes, mais la majorité pourrait se retrouver dans une position de consensus).

Le choix final des profils et des règles d'affectation devra se faire en étroite coopération avec des représentants des principaux utilisateurs potentiels du système d'éva-

[3]cf. par exemple définition de δ dans le Chapitre 5, p. 128

luation. Une fois les profils déterminés, il faudra donc définir, avec des scientifiques de Welfare Quality® et les membres du Comité Conseil du projet, la valeur du seuil d'indifférence (si son utilisation est demandée), les valeurs des seuils de veto à utiliser ainsi que choisir la procédure d'agrégation des critères à suivre, sur la base des quatre procédures proposées dans la thèse (par ex. majorité *vs* unanimité).

La première étape de cette discussion sur la procédure d'affectation des fermes aux catégories de bien-être pourra reposer sur la consultation des membres du Comité Conseil de Welfare Quality® en leur demandant combien, selon leurs estimations, il y a de fermes excellentes en Europe, et à l'inverse combien de fermes offrent un niveau de bien-être inacceptable aux animaux. Il sera alors possible de voir si de grandes disparités existent entre ces représentants des principaux utilisateurs potentiels ou si un consensus semble déjà se constituer pour ces deux catégories extrêmes. De plus, la détermination des objectifs en termes de répartition des fermes parmi les quatre catégories de bien-être faciliterait le choix de la procédure à suivre pour passer des scores de critères à l'évaluation globale.

Réalisation des analyses de sensibilité et de robustesse et vérification de la légitimité du modèle final

Tsoukiàs (2007) souligne que trois précautions doivent être prises avant de formuler une recommandation finale. Il recommande d'effectuer :

- une *analyse de sensibilité*, afin de voir comment réagit le modèle face à des modifications de ses paramètres. Le mot paramètre correspond ici à l'ensemble des données incluses dans le modèle ainsi qu'aux paramètres définis en cours de construction du modèle. Ces deux types d'entrée du modèle peuvent être imprécises et sujettes à variation. L'analyse de sensibilité est donc réalisée sur un modèle utilisant des procédures définies et paramétrées, et consiste à faire varier les paramètres au voisinage des valeurs qui leurs sont attribuées dans le modèle, correspondant aux valeurs optimales.

  Dans notre cas, les données recueillies sur le terrain sont évaluées sur un échantillon d'animaux jugé représentatif de l'ensemble du troupeau. Il est cependant probable qu'en fonction des animaux observés il y ait de légères modifications des données recueillies.

  Le paramétrage de notre modèle, quant à lui, est basé sur des opinions d'experts. Il est probable qu'en fonction des experts consultés les valeurs des paramètres soient légèrement modifiées (par ex. que se passerait-il si nous consultions un ex-

pert supplémentaire ?).

Il sera donc important de vérifier que de légères modifications des paramètres n'ont pas d'impact significatif sur le résultat final. L'analyse de sensibilité consistera à faire varier chaque paramètre l'un après l'autre afin de voir dans quelle plage de valeurs des paramètres le résultat reste inchangé (Ben Mena 2001a). Étant donné le nombre de paramètres à faire varier, il pourrait être pratique d'utiliser une solution informatisée afin de réaliser cette analyse de sensibilité[4].

- une *analyse de robustesse*. Contrairement à l'analyse de sensibilité, l'analyse de robustesse vise à tester la variabilité du résultat obtenu avec le modèle en cas d'utilisation de procédures différentes, paramétrées différemment et sur des données pouvant elles aussi être différentes. En effet, au cours de la construction du modèle d'évaluation nous avons fait des hypothèses et des choix sur les méthodes et leur paramétrage. De plus, les données elles-mêmes sont susceptibles de varier en fonction des conditions d'inspection des fermes. Nous avons ainsi introduit une part d'arbitraire dans la construction et il importe donc de vérifier que la variabilité des résultats en fonction de différents scénarios (établis sur des procédures, des paramètres et des données différentes) n'est pas trop importante (Roy 2002). Il serat ainsi possible de tester des scénarios faisant intervenir des modifications des conditions d'observation en ferme (par ex. observations en été ou en hiver), des choix de procédures d'affectation des fermes aux catégories de bien-être ou des fonctions d'agrégation des sous-critères en critères (par ex. intégrale de Choquet *vs* OWA), voire même des modifications, plausibles à moyen terme, de la réglementation européenne (par ex. l'interdiction des étables entravées pour les vaches laitières). Dans ce type d'analyse, la principale difficulté revient à identifier la liste des scénarios plausibles, sans pour autant que cette liste soit trop longue (Roy 2002).
- une *vérification de la légitimité*, afin de s'assurer qu'à la fois le modèle et les résultats qu'il fournit sont en accord avec les attentes des utilisateurs potentiels, aussi bien en termes de compréhension du problème qu'en termes de résultats.

  Pour être légitime, un modèle doit aussi être accepté par l'ensemble de ses utilisateurs. Pour faciliter l'approbation des utilisateurs Bouyssou et al. (2000, p. 226-227) propose même de montrer aux utilisateurs qu'il est possible d'implémenter le modèle dans un tableur (par ex. Excel), et que par conséquent les méthodes d'agrégation, même si elles ne sont pas connues des utilisateurs, n'en sont pas pour autant complexes. Il est aussi prévu dans le WP4.2 de développer un logiciel

[4]du type de celle proposée par Ben Mena (2001b) pour une méthode de type ELECTRE III

accessible sur Internet et pouvant servir à des démonstrations de l'utilisation du modèle.
Par ailleurs, des réunions en groupes seront organisées par le WP4.4 avec des consommateurs et des éleveurs afin de tester leur acceptation du système

Validation du modèle final d'évaluation globale du bien-être

Le modèle d'évaluation qui résultera du processus de construction (résultant de la définition de l'ensemble des éléments décrits dans la première section du Chapitre 4) devra ensuite être validé. Pour ce faire, Tsoukiàs (2007) et Pala et al. (1999) citent plusieurs procédures de validation complémentaires :

- la validation conceptuelle (visant à vérifier la pertinence des concepts utilisés lors du processus de construction),
- la validation logique (visant à vérifier la cohérence logique du modèle d'évaluation),
- la validation expérimentale (visant à vérifier l'adéquation entre les résultats fournis par le modèle et des données expérimentales),
- la validation opérationnelle (visant à vérifier la faisabilité de la mise en application et de l'utilisation du modèle dans le cadre d'une utilisation en routine).

La *validation conceptuelle* a pour but de vérifier que les théories et hypothèses sur lesquelles repose notre modèle d'évaluation sont correctes et pertinentes. Tout au long de la procédure de construction du modèle, des chercheurs en sciences animales, spécialistes du bien-être animal, ont été consultés et impliqués. La structure du modèle leur a été présentée plusieurs fois, et a à chaque fois reçu leur approbation. De même, l'hypothèse principale, relative à la non-compensation entre dimensions du bien-être, a été largement vérifiée au cours des consultations d'experts. Enfin, la famille de critères sur laquelle se base notre modèle a reçu une très large approbation, aussi bien au niveau scientifique qu'au niveau des utilisateurs potentiels du système d'évaluation.

La *validation logique* vise à vérifier si le modèle permet de décrire correctement la problématique, s'il permet bien d'y répondre, sans oublier d'éléments ou de relations importantes. Lors de la construction du modèle d'évaluation que nous proposons dans cette thèse (et qui servira de modèle pour les modèles finaux d'évaluation qui seront produits à la fin de Welfare Quality®), nous avons appliqué les principes de construction ainsi que des méthodes de calcul déjà établis dans la littérature et reconnus par les spécialistes de l'évaluation multicritère. Nous pouvons donc supposer que les concepts

sur lesquels se base le modèle ont été validés par des experts du domaine (voir par ex. Roy 1985 ; Bouyssou et al. 2000 ; Vincke 1992 ; Perny 1998).

Selon Pala et al. (1999), la validation logique du modèle repose aussi sur la vérification du programme informatique ou des feuilles de calcul permettant d'implémenter le modèle. Le modèle que nous proposons dans la thèse ne fait pas encore l'objet d'un logiciel, mais ce travail est prévu dans le WP4.2. Le code du programme qui sera alors établi devra être vérifié.

La *validation expérimentale* traite de la qualité du modèle proposé, les experts du domaine devant y adhérer. Cette validation dépend des résultats obtenus sur les analyses de sensibilité et de robustesse, mais aussi sur une vérification de l'adéquation entre les résultats fournis par le modèle et des données expérimentales recueillies sur le jeu de fermes laitières européennes visitées dans le cadre de Welfare Quality®. Comme nous l'avons montré dès le chapitre introductif, il n'existe pas d'étalon pour l'évaluation globale du bien-être en ferme. Il est cependant possible de vérifier l'adéquation entre les résultats du modèles et l'impression générale ressentie par les personnes ayant réalisé les observations sur le terrain. Pour ce faire, nous leur avons demandé de donner leur évaluation globale du niveau de bien-être sur chacune des fermes en utilisant pour cela une *Visual Analogue Scale*[5]. Ce type d'échelle visuelle est notamment utilisée pour évaluer la douleur ressentie par un patient. On demande alors à ce dernier de marquer sur une ligne horizontale de 100 mm allant de "aucune douleur" à "douleur extrême" où se situe son niveau de douleur (Gould et al. 2001). Nous avons demandé aux personnes ayant réalisé les observations en ferme d'utiliser le même type d'échelle pour indiquer leur évaluation du bien-être des animaux sur chacune des fermes visitées (cf. Figure 8.12).

FIG. 8.12 – Visual Analogue Scale utilisée par les personnes ayant réalisé les observation dans le WP2.4 pour évaluer le niveau de bien-être qu'elles avaient perçu sur les différentes fermes

En plus de cette évaluation globale exprimée sur la *Visual Analogue Scale*, nous leur avons demandé de faire un commentaire sur chacun des quatre critères.

Les évaluations ainsi réalisées par le WP2.4 nous permettrons de vérifier qu'il n'y a pas de désaccord profond entre les évaluations globales fournies par notre modèle et

[5]échelle visuelle permettant de mesurer une caractéristique ou une attitude qui est supposée prendre des valeurs s'exprimant sur un continuum de valeurs qui ne peut être aisément mesuré directement (Gould et al. 2001)

celles données par les observateurs. Les évaluations recueillies auprès de WP2.4 n'ont qu'une valeur qu'indicative puisqu'elles ne correspondent qu'au ressenti de l'observateur au cours de la visite. Cependant, en cas de désaccord profond entre ce ressenti et les résultats du modèle (par exemple si notre modèle affecte une ferme à la catégorie *"Excellent"* alors que lors de la visite l'observateur l'a considérée comme n'assurant qu'un niveau de bien-être médiocre), il nous faudra comprendre les divergences, par exemple en discutant avec les évaluateurs et en leur faisant expliciter leur évaluation. Au besoin, la construction du modèle et notamment sa dernière étape - l'agrégation des critères - pourra alors être revue.

La *validation opérationnelle* vise à vérifier l'applicabilité du modèle d'évaluation ainsi que son utilité pour répondre aux attentes de ses utilisateurs.

Afin de favoriser la mise en pratique du système d'évaluation proposé dans Welfare Quality®, le WP4.2 est chargé de mettre au point un logiciel permettant de saisir aisément les données sur le terrain, associé à une base de données permettant d'archiver au niveau européen l'ensemble des données recueillies en fermes et leurs évaluations (plate-forme Internet), elle-même associée à un module de calcul des scores à partir des données recueillies sur le terrain. Ceci permettra de fournir un prototype aux utilisateurs potentiellement intéressés. En parallèle, le WP4.4 sera chargé de réaliser des enquêtes auprès de distributeurs pour voir comment ils mettent en place actuellement des procédures de certification et dans quelles mesures il serait possible de mettre en place un système de certification "bien-être" sur la base de l'outil proposé dans Welfare Quality®.

En ce qui concerne l'évaluation de l'utilité du système, le travail réalisé par le WP4.6 sera particulièrement intéressant. En assurant un suivi des élevages, il permettra en effet d'apprécier l'efficacité du modèle pour évaluer des progrès réalisés et servir de support au conseil auprès des éleveurs.

Généralisation du modèle d'évaluation et transfert vers les autres couples (*type d'animal*, *période*)

Le WP2.3 a pour rôle de mettre en place les modèles d'évaluation pour l'ensemble des types d'animaux et des périodes étudiés dans Welfare Quality®. Le modèle proposé dans la thèse ne concerne que les vaches laitières en ferme. Le modèle que nous proposons se base sur une famille de quatre critères valable quel que soit le couple (*type d'animal*, *période*) concerné. Cependant de légères variations pourront avoir lieu au niveau des

sous-critères, mais le principe d'une agrégation séquentielle en sous-critères puis en quatre critères communs prévaut quel que soit le type d'animal ou la période considérés. De plus, le modèle "vaches laitières en ferme" que nous proposons présente les méthodes d'agrégation des informations et de consultation des experts à suivre pour les autres modèles.

La construction de ces modèles est déjà bien avancée pour la plupart des animaux, mais là encore il faudra attendre les listes définitives des mesures pour pouvoir finaliser les modèles.

Perspectives de recherche : vers une application du modèle pour l'évaluation d'autres notions multidimensionnelles liées à l'élevage et à l'agriculture

Le bien-être animal n'est pas le seul concept multidimensionnel qui s'applique aux élevages et il est vraisemblable que la démarche d'évaluation multicritère suivie pour cette thèse puisse s'appliquer à d'autres concepts, tels que la durabilité des activités d'élevage, thématique prenant de plus en plus d'importance. Le concept de durabilité est apparu pour la première fois dans un texte officiel en 1987 et s'est vu consacré lors de la conférence de Rio organisée par les Nations Unies en 1992 (Landais 1998). Landais (1998) définit quatre critères de performances à travers lesquels il est possible d'évaluer la durabilité des exploitations agricoles (cf. Figure 8.13) :

- la *viabilité* : l'exploitation doit permettre à l'agriculteur de dégager des revenus suffisants pour vivre correctement ;
- la *vivabilité* : l'exploitation doit assurer à l'agriculteur et à sa famille d'avoir une bonne qualité de vie, dépendant de facteurs liés à sa charge de travail et à l'organisation de ce dernier, mais aussi à des facteurs extérieurs à l'exploitation comme son insertion dans les réseaux professionnels locaux ou l'image que peuvent avoir les autres personnes de son travail[6] ;
- la *transmissibilité* : l'agriculteur devrait pouvoir transmettre son exploitation à la génération suivante, mais cela pose assez souvent des problèmes (liés aux relations socio-économiques intervenant dans la vivabilité, à des problèmes pratiques liés à la succession ou encore à l'agrandissement quasi-inéluctable des exploitations) ;
- la *reproductibilité* : les effets négatifs de l'activité agricole sur l'environnement

[6] le respect du bien-être animal peut alors faire partie de l'évaluation de la vivabilité

devraient être limités au maximum de sorte à ne pas remettre en cause la poursuite de l'activité.

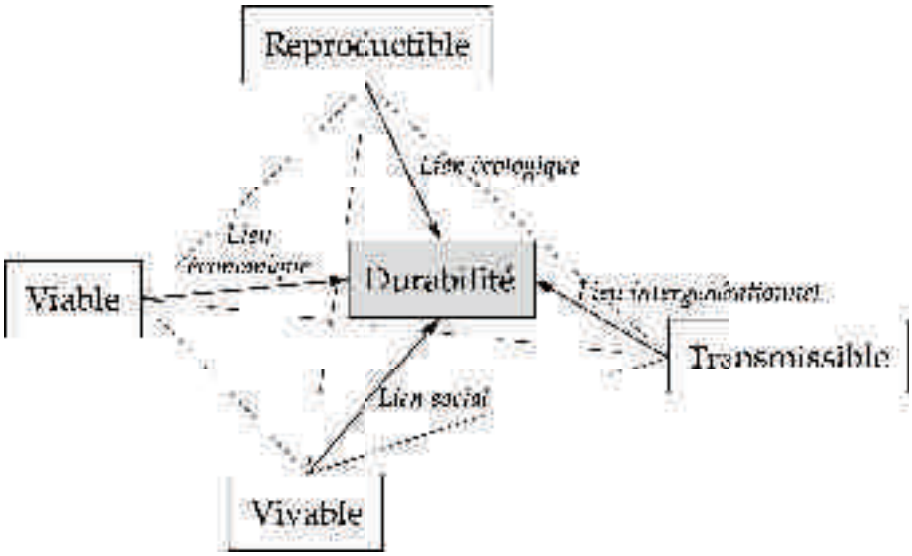


FIG. 8.13 – Les quatre piliers de la durabilité des exploitations agricoles (d'après Landais 1998)

D'autres auteurs proposent d'évaluer la durabilité des élevages sur la base de trois critères, correspondant aux aspects économiques, sociaux et écologiques des exploitations (voir par exemple van Calker et al. 2006).

Dans tous les cas ces dimensions sont plus ou moins indépendantes et ne se compensent pas l'une l'autre. Par exemple, en se basant sur les trois dimensions de la durabilité identifiées par van Calker et al. (2006), il semble impossible de dire que la durabilité d'une exploitation est correcte parce qu'elle obtient de très bons résultats sur la dimension écologique alors que par ailleurs ses résultats économiques sont catastrophiques. Les techniques d'aide à la décision multicritère sont donc parfaitement adaptées à l'évaluation de la durabilité. C'est notamment ce type de modèle que van Calker et al. (2006) ont utilisé pour développer une application permettant d'évaluer la durabilité de fermes laitières néerlandaises.

Le travail de formalisation d'une procédure d'évaluation globale du bien-être animal, basé sur le cas des vaches laitières en ferme, m'a permis de prendre conscience des risques et difficultés à surmonter lors de la définition d'un concept multidimensionnel et de la conception d'une procédure d'évaluation. Ce travail de modélisation multicritère, réalisé sur le bien-être animal, peut se transposer à d'autres notions multidimensionnelles concernant le monde agricole. Au niveau de l'exploitation, le respect du bien-être animal est un élément important, notamment pour la perception que l'éleveur peut avoir de son travail mais aussi pour l'image véhiculée auprès de ses voisins et des citoyens en général. Cependant, ce n'est pas suffisant pour les éleveurs. Ils doivent aussi pouvoir vivre de leur exploitation, avec des conditions de travail leur assurant à eux aussi un bon niveau de bien-être. Nous arrivons alors au concept de durabilité des élevages, pour lequel des recherches sont désormais engagées afin de produire une évaluation globale et de concevoir de nouveaux systèmes de production.

Glossaire

Glossaire

ACTA	Association de Coordination Technique Agricole
CESE	Comité Économique et Social Européen. C'est un organe consultatif de la Commission Européenne formé de représentants des principaux acteurs socio-économiques européens
DG-SANCO	Direction Générale "Santé et la protection des Consommateurs". Au sein de la Commission Européenne, la DG-SANCO a pour mission de contribuer à améliorer la santé, la sécurité et la confiance des citoyens européens
FVE	Federation of Veterinarians of Europe. Il s'agit d'une oragnisation regroupant des organisations vétérinaires de 35 pays européens
GATT	General Trade Agreement on Tariffs and Trade (accord général sur les tarifs douaniers et le commerce). Ce traité a pour but d'harmoniser les politiques douanières des parties signataires. C'est un accord multilatéral de libre-échange qui globalement est censé faire baisser les prix pour les consommateurs, mieux utiliser les facteurs de production, favoriser l'emploi dans les secteurs où chaque pays détient un avantage comparatif
ICTA	Instituts et Centres Techniques Agricoles
INRA	Institut National de la Recherche Agronomique
OIE	Office Internationale des Épizooties (Organisation Mondiale de la Santé Animale mais a gardé son acronyme historique OIE). L'OIE est l'organisation intergouvernementale chargée d'améliorer la santé animale dans le monde.
OMC	Organisation Mondiale du Commerce. C'est la seule organisation internationale qui s'occupe des règles régissant le commerce entre les pays

MDD	Marques De Distributeurs. Est considéré comme produit vendu sous MDD un produit dont les caractéristiques ont été définies par l'entreprise ou le groupe d'entreprises qui en assure la vente au détail et qui est le propriétaire de la marque sous laquelle il est vendu. Elles sont à opposer aux marques nationales (ou internationales). Par exemple, on peut trouver au rayon traiteur d'un hypermarché Auchan des plats préparés de marque *Auchan* (MDD) ou de marque *Fleury Michon* (marque nationale).
OWA	Moyenne pondérée ordonnée (*Ordered Weighted Average*)
SP	Sous-projet du projet Welfare Quality®
WP	Groupe de tâches, au sein d'un sous-projet de Welfare Quality®
WS	Somme pondérée (*Weighted Sum*)

Annexes

Annexe A

Organisation du projet Welfare Quality®

Le projet Welfare Quality®, présenté dans Veissier *et al.* (Veissier, Blokhuis, Geers, Jones, and Miele 2005), s'organise en quatre sous-projets de recherche complémentaires (cf. Figure A.1) :

- le **sous-projet 1** étudie les attitudes sociales et les pratiques des consommateurs, des distributeurs et des producteurs en relation avec le bien-être animal et évalue la possibilité de mettre en place des systèmes respectueux du bien-être animal ;
- le **sous-projet 2** développe une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir ;
- le **sous-projet 3** développe et évalue des solutions pratiques à des problèmes de bien-être ;
- le **sous-projet 4** générera les standards d'évaluation et d'information, transférera les connaissances aux utilisateurs, évaluera l'acceptabilité du système par les utilisateurs potentiels.

Cette organisation permet de faciliter la mise en relation des différents domaines d'expertise tout en optimisant le travail et le développement des compétences multidisciplinaires nécessaire à la réalisation des différents objectifs scientifiques et techniques définis dans le projet.

Par ailleurs, des activités horizontales (dialogue science-société, formation et mobilité des jeunes chercheurs, transfert des connaissances et évaluation scientifique) renforcent les activités des quatre sous-projets de recherche.

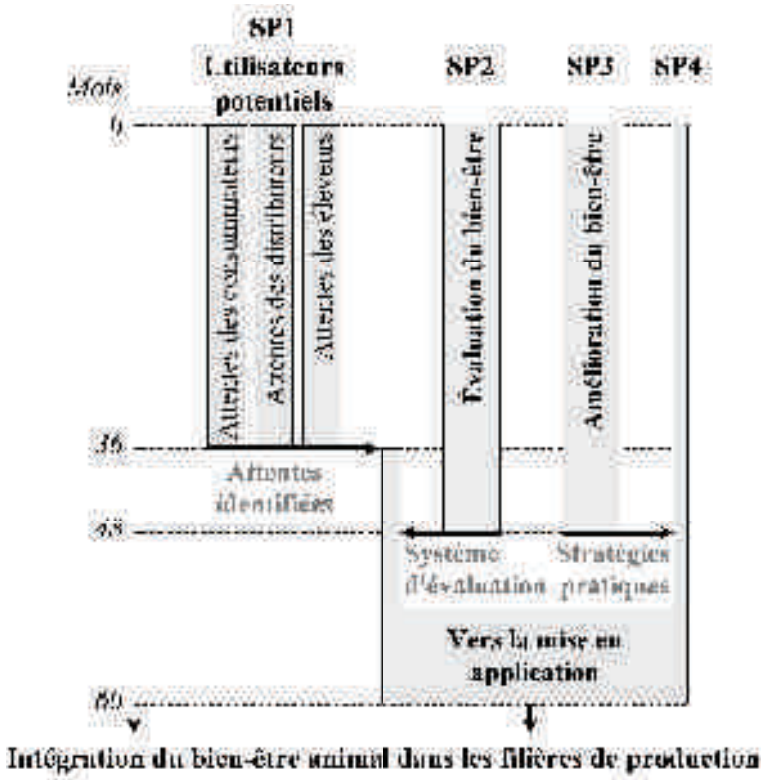


FIG. A.1 – Organisation des quatre sous-projets de recherche au sein de Welfare Quality®

A.1 Sous-projet 1 : Attentes des consommateurs, des distributeurs et des éleveurs vis-à-vis du bien-être animal et de produits respectueux du bien-être des animaux dont ils sont issus

L'objectif de ce sous-projet est de mettre en évidences les attitudes et les pratiques des consommateurs, des distributeurs et des éleveurs vis-à-vis du bien-être animal, et ainsi de pouvoir évaluer dans quelle mesure de nouvelles stratégies d'évaluation, d'amélioration et d'information sur le bien-être pourraient être mise en application avec succès.

Ce sous-projet 1 s'organise en trois groupes de tâches concernant respectivement les consommateurs, les distributeurs et les éleveurs. Chacun de ces groupes de tâches cherche à identifier comment ces groupes perçoivent le bien-être animal, et à déterminer comment il serait possible de les persuader d'adopter des standards de bien-être plus rigoureux :

Groupe de tâches 1.1
Analyser les attentes des consommateurs vis-à-vis de produits respectueux du bien-être, du type d'information souhaitée, et de la stratégie de communication et d'information qui serait la plus efficace.

Groupe de tâches 1.2
Évaluer l'état à la fois actuel et potentiel du marché vis-à-vis de produits respectueux du bien-être animal, les caractéristiques des labels "bien-être" existants, et les systèmes d'inspections utilisés actuellement.

Groupe de tâches 1.3
Identifier les barrières potentielles au développement de produits respectueux du bien-être auxquelles font face les éleveurs.

A.2 Sous-projet 2 : Développement d'une méthodologie standardisée pour l'évaluation du bien-être des animaux de la ferme à l'abattoir

L'objectif global du sous-projet 2 et de développer une méthodologie standardisée d'évaluation en routine du bien-être chez les bovins, les porcs et les volailles, de la ferme à l'abattoir. Ce système d'évaluation est basé essentiellement sur des mesures prises directement sur les animaux, mais aussi sur l'environnement (ressources et gestion), notamment pour mettre en évidence d'éventuelles causes de mal-être. Le développement de ce système d'évaluation inclut le fait de déterminer quelle procédure serait la plus appropriée pour synthétiser les différentes mesures effectuées sur le terrain et pouvoir ainsi produire une évaluation globale du bien-être.

Ce sous-projet est subdivisé en quatre groupes de tâches étroitement liés :

Groupe de tâches 2.1
Définir une liste de mesures (d'après les connaissances existantes) qui pourraient être utiliser sur le terrain (en ferme et à l'abattoir) et qui permettraient de couvrir les différentes dimensions du bien-être.

Groupe de tâches 2.2
Développer des procédures de contrôle du bien-être sur le terrain, en se basant uniquement sur les mesures dont la validité, la répétabilité et la faisabilité auront été testées dans cet atelier.

Groupe de tâches 2.3
Définir une méthode d'évaluation multicritère du bien-être animal, fournissant une évaluation globale du niveau de bien-être sur l'élevage à partir des différentes mesures sélectionnées par le groupe de tâches 2.2.

Groupe de tâches 2.4
Définition des protocoles finaux d'évaluation du bien-être, se basant sur des mesures qui permettent de différencier les fermes les unes des autres, tout en restant relativement simples et facilement réalisables.

A.3 Sous-projet 3 : Développement de stratégies permettant d'améliorer le bien-être des animaux d'élevage

L'objectif principal du sous-projet 3 est de développer puis tester des stratégies pratiques afin d'améliorer les bien-être des animaux en ferme. Ces stratégies potentielles peuvent inclure à la fois des actions sur l'environnement des animaux et des approches génétiques, dans le but de réduire l'occurrence de certains comportements ou états physiologiques dangereux pour les animaux, de fournir un environnement sain mais stimulant, et d'améliorer les relations homme-animal en apportant des conseils aux éleveurs. Ces stratégies correctives sont appliquées aux situations identifiées comme étant sources d'intérêt pour les consommateurs ou causes de mal-être pour les animaux. À terme, ces stratégies devraient permettre d'aider les éleveurs à atteindre un bon niveau de bien-être sur leur ferme.

Des solutions viables doivent satisfaire à la fois des exigences économiques et en termes de bien-être. Elles doivent aussi être applicables sur le terrain (c.-à-d. accessibles et faciles à mettre en application par l'éleveur. De telles solutions pratiques n'impliquent pas nécessairement l'adoption de systèmes d'élevage extensifs (tels que l'élevage biologique). Des systèmes intensifs peuvent aussi assurer le bien-être des animaux, en leur permettant de couvrir leurs besoins les plus importants.

Le sous-projet 3 est divisé en six ateliers, chacun abordant un problème de bien-être particulier : stress lié à la manipulation des animaux, traits phénotypiques dangereux, comportements sources de blessures, les boiteries, la mortalité néo-natale, et le stress social. Ces différents points couvrent les principaux problèmes perçus comme étant importants par les consommateurs européens, et offrant des marges de progrès importantes. Chacun de ces groupes de tâches cherche donc à déterminer d'éventuelles solutions pour :

Groupe de tâches 3.1
Minimiser les stress liés à la manipulation des animaux.

Groupe de tâches 3.2
Réduire grâce à la sélection génétique certains problèmes de bien-être.

Groupe de tâches 3.3
Éliminer certains comportements sources de blessures.

Groupe de tâches 3.4
Réduire les boiteries chez les bovins et les poulets de chair.

Groupe de tâches 3.5

Minimiser la mortalité néonatale des porcelets.

Groupe de tâches 3.6

Diminuer les stress social.

A.4 Sous-projet 4 : Vers la mise en application d'une procédure standardisée d'évaluation et d'information du bien-être ainsi que vers celle de stratégies d'amélioration du bien-être

Les principaux objectifs de ce sous-projet sont de générer des standards pour l'évaluation du bien-être et la transmission de l'information, de communiquer sur les connaissances acquises durant le projet, de mettre en place des projets de recherche sur des thématiques similaires, et produire des recommandations sur la (ou les) meilleure(s) stratégie(s) pour la mise en application du système d'évaluation et d'information et des stratégies correctrices proposés par le projet.

Alors que tous les autres sous-projets sont presque terminés, le sous-projet 4 a démarré il y a juste quelques mois. Il permettra de réunir les connaissances et productions des trois autres sous-projets, et ainsi de s'assurer que le système d'évaluation du bien-être, le système d'information lié au produit, et les stratégies d'améliorations en élevage seront mis en oeuvre selon les mécanismes les plus appropriés.

Le sous-projet 4 est composé de sept groupes de tâches permettant de couvrir les différents objectifs fixés pour ce sous-projet.

La publication dans un format déterminé des standards d'évaluation du bien-être et d'information constitue l'une des issues clef de ce sous-projet. Ces standards devront pouvoir permettre d'appliquer la procédure d'évaluation quel que soient le pays et le système d'élevage concernés. Afin d'assurer aux consommateurs un certain niveau de bien-être, ces standards pourront être utilisés dans le cadre de procédures de certification ou de réglementations. L'information liée au niveau de bien-être assuré aux animaux devrait permettre de servir de base à un système d'information destiné aux consommateurs, tout en respectant les contraintes identifiées dans le sous-projet 1.

La transmission des connaissances et des résultats aux utilisateurs potentiels du système d'évaluation et d'information développé dans le sous-projet 2, ainsi que la communication sur les stratégies d'amélioration identifiées dans le sous-projet 3, seront rendues possibles par la production d'une base de données disponibles sur Internet et par la mise à disposition des futurs évaluateurs d'outils pédagogiques et de session d'entraînement.

Les principaux points de recherche consistent donc à évaluer : la réponses des consommateurs face au systèùe d'information lié au produit, l'impact socio-économique de ce système d'information, l'impact sur le bien-être animal de l'évaluation qui en est

faite sur le terrain, et ce que pensent les principaux utilisateurs potentiels du système d'évaluation produit par le sous-projet 2.

Groupe de tâches 4.1
Finaliser et publier les standards concernant les systèmes d'évaluation du bien-être et d'information liée au produit.

Groupe de tâches 4.2
Fournir les informations techniques nécessaires aux utilisateurs potentiels du système d'évaluation du bien-être.

Groupe de tâches 4.3
Former les utilisateurs potentiels en les informant sur les stratégies pouvant être mises en place pour améliorer le bien-être.

Groupe de tâches 4.4
Étudier les réponses des consommateurs face aux systèmes proposés pour fournir de l'information sur les produits.

Groupe de tâches 4.5
Évaluer l'impact socio-économique des systèmes permettant de fournir de l'information sur les produits.

Groupe de tâches 4.6
Identifier et évaluer les mécanismes optimaux pour une utilisation en ferme du système d'évaluation du bien-être afin d'évaluer les stratégies visant à améliorer le bien-être.

Groupe de tâches 4.7
Formuler des recommandations pour les instances réglementaires sur la mise en application potentielle du système d'évaluation du bien-être animal, du système d'information liée au produit, et des stratégies d'améliorations.

Annexe B

Composition du Comité Conseil du projet Welfare Quality®

Le Comité Conseil du projet Welfare Quality® est composé de représentants des principaux utilisateurs potentiels du système d'évaluation et d'information qui sera proposé. Il est donc composé de représentants de consommateurs, d'associations de défense des animaux, de distributeurs et de producteurs, chacun d'entre eux étant fort intéressé par les résultats du projet.

Le Comité Conseil a trois principaux rôles. Il a tout d'abord pour rôle de conseiller le Comité de Direction sur les principaux objectifs et approches du projet et sur la prise en compte de questions et de stratégies spécifiques. Il sert aussi à conseiller et à assiter les coordinateurs des différents sous-projets. Enfin, ses membres doivent se baser sur leur expertise collective pour s'assurer que les rapports de recherche et les propositions préparés par les partenaires de Welfare Quality® ou des scientifiques externes au projet sont clairs, innovants, pertinents et opportuns. Ces objectifs sont atteints grâce à des contacts réguliers avec le Comité de Direction, des évaluations du travail réalisé dans le projet et des réunions ou ateliers sur des sujets précis organisé avec des partenaires du projet.

Liste des membres actuels du Comité Conseil, avec leurs noms et l'organisation qu'ils représentent :

- Mme Nancy De Bryine et M. Jan Vaarten, *Federation of Veterinarians of Europe*
- Dr Sarah Kahn et Dr Leopoldo Stuardo, *International Organisation for Animal Health (OIE)*
- Dr Sonja van Tichelen, *Eurogroup for Animal Welfare (EUROGROUP)*
- Professeur Peter Sandoe, *European Society for Agricultural Ethics*

- Professeur Jan Daelemans, *European Pig Producers Group*
- M. Per-Ake Sahlberg, *European Dairy Farmers Group*
- Dr Ken Laughlin, *European Forum of Farm Animal Breeders (anciennement FAIP)*
- Dr Nigel Garbutt, *Global Partnership for Safe and Sustainable Agriculture (EU-REPGAP)*
- Mmes Christina Nygaard et Mmes Roxanne Feller, *COPA-COGECA*
- M. Aldin Hilbrands, *Royal Ahold*
- M. Patrik Holm-Thisner, *McDonald's Europe*
- M. Ian Burton, *PAI Group, Product Authentication Inspectorate Ltd*

Notes :

- Le Dr Andrea Gavinelli (*DG-SANCO*) participe en tant qu'observateur aux différentes réunions du Comité Conseil.
- La liste des membres du Comité Conseil peut évoluer en fonction des circonstances.

Annexe C

Description de la construction de l'ensemble des sous-critères

Dans cette annexe sont présentées les fiches décrivant la construction des sous-critères, avec présentation très succincte de la méthode (les méthodes étant détaillées dans le manuscrit) et suivi de la définition et des évolutions des paramètres. Ces fiches seront incluses dans le livrable D2.8b du projet Welfare Quality®.

Subcriterion 1: Absence of prolonged hunger

Definition of the subcriterion

In this subcriterion, we aim at assessing whether the animals suffer or not from prolonged hunger. On farms, the focus is on chronic hunger and thus on the detection of lean animals. For transport and slaughter, the focus is more on short term hunger, depending on the duration of food removal before slaughter.

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11101	Body condition score	Binary	Individual

To assess chronic hunger on-farm, only **one** measure foreseen in the project seems to be useful, and that measure is **body condition score** (**BCS**).

Body condition score of the animal *k* is assessed on an ordinal scale (with 3 levels: 0. too thin animal, 1. normal, and 2. too fat)➔ M_k

Since we are interested in assessing the level of hunger of the animals, we do not consider as problematic the animals which are too fat, and we consider very lean animals as suffering from chronic hunger.
Hence, we consider only two kinds of animals, those which are very lean (corresponding to the **level 0** of the ordinal scale used on farm to collect data) and the others (corresponding to both **levels 1 and 2** of the scale used on farm).

Analysis of the constraints

Since there is only one measure to produce a score for this subcriterion, we want to preserve the precision of the information given by the raw data.

In addition we have to deal with raw data observed at the individual level.

The subcriterion score qualifies the ***prevalence of extreme leanness in the herd****.*

Type of mathematical construction proposed

Normally, body score has to be corrected for by the breed and the stage of lactation of the animal in order to be correctly interpreted in terms of “level of hunger”. However, we are only considering extreme cases, for which, whatever the stage of lactation of the animals, we will consider that animals are clearly too lean. So we decided not to take into account the stage of lactation of the animals.

At herd level, the information we will have to deal with is the percentage p_0 of animals which are considered as very lean (animals assigned to category 0 of the ordinal scale used on farm).

The greater p_0, the lower the subcriterion score. As a consequence we have first to calculate an intermediate value called 'index': **Index =** $h(x) = 100 - p_0(x)$

We will have to transform the **index** into a value score (∈ [0,100]). This transformation will be defined thanks to experts' opinion on a given dataset.

Subcriterion score = $g_1(x)$ = $u(h(x))$ = $u(100 - p_0(x))$ with p_0 = the % of very lean animals and u the function used to transform the index h into a score

Parameters to be defined

Information is needed to define the most appropriate function *u*.

Adjustments of the mathematical construction

- **Questions asked to experts**

*To choose the best function **u**, a set of data is proposed to experts who have to give an absolute evaluation on the 0 – 100 scale for each data presented in the dataset.*

We asked 5 experts to score the 11 virtual farms presented in Table 1.
The experts asked were U. Knierim, I. Veissier, H.R. Whay, C. Winckler, J. Capdeville.

Table 1: Dataset submitted to experts. They had to attribute a subcriterion score to each farm

	p_0	Score
Farm 1	0%	
Farm 2	10%	
Farm 3	20%	
Farm 4	30%	
Farm 5	40%	
Farm 6	50%	
Farm 7	60%	
Farm 8	70%	
Farm 9	80%	
Farm 10	90%	
Farm 11	100%	

- **Adjustment of the construction according to answers from experts**

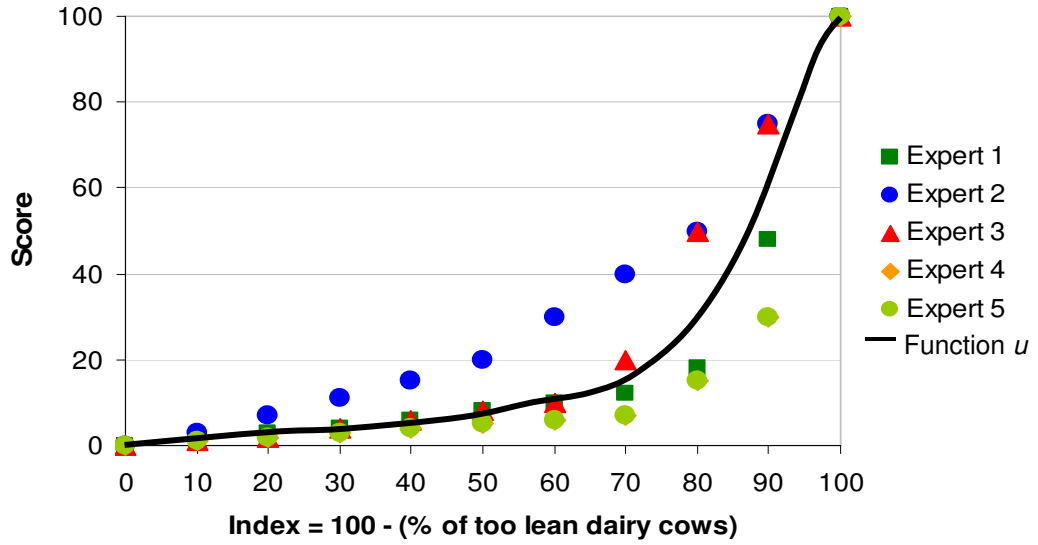
The experts' evaluations of farms with various percentages of very lean dairy cows are shown in Table 2.

B-spline functions were defined so as to best match expert scores (i.e. to minimise the sum of square errors with experts' scores) (Figure 2).

Table 2: Answers from experts

	p_0	Score					Index
		Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	
Farm 1	0%	*100*	*100*	*100*	*100*	*100*	100
Farm 2	10%	*48*	*75*	*75*	*30*	*30*	90
Farm 3	20%	*18*	*50*	*50*	*15*	*15*	80
Farm 4	30%	*12*	*40*	*20*	*7*	*7*	70
Farm 5	40%	*10*	*30*	*10*	*6*	*6*	60
Farm 6	50%	*8*	*20*	*8*	*5*	*5*	50
Farm 7	60%	*6*	*15*	*6*	*4*	*4*	40
Farm 8	70%	*4*	*11*	*4*	*3*	*3*	30
Farm 9	80%	*3*	*7*	*2*	*2*	*2*	20
Farm 10	90%	*1*	*3*	*1*	*1*	*1*	10
Farm 11	100%	*0*	*0*	*0*	*0*	*0*	0

Figure 2: B-spline function to attribute a subcriterion-score to 'Absence of prolonged hunger'



- **Test on dataset(s)**

The distribution of the % of lean cows in the 64 dairy farms visited during WP2.4 is shown in Table 3 and on Figure 3.

able 3: Summary information on the distribution of the % of very lean cows among the 64 dairy farms visited in WP2.4

	% of too lean cows
Mean	**16**
Standard deviation	16,5
Min	0,0
Max	63,6
1st quartile	3,3
2nd quartile	12,5
3rd quartile	21,8
4th quartile	63,6

We contacted again the experts in order to ask them to refine (if they consider it was necessary) their first opinion, taking into account the information on the distribution of farms on SC1, provided in Table 3, Figure 3 and also in the third column of Table 4.

Figure 3: Distribution of the % of very lean cows among the 64 dairy farms visited in WP2.4

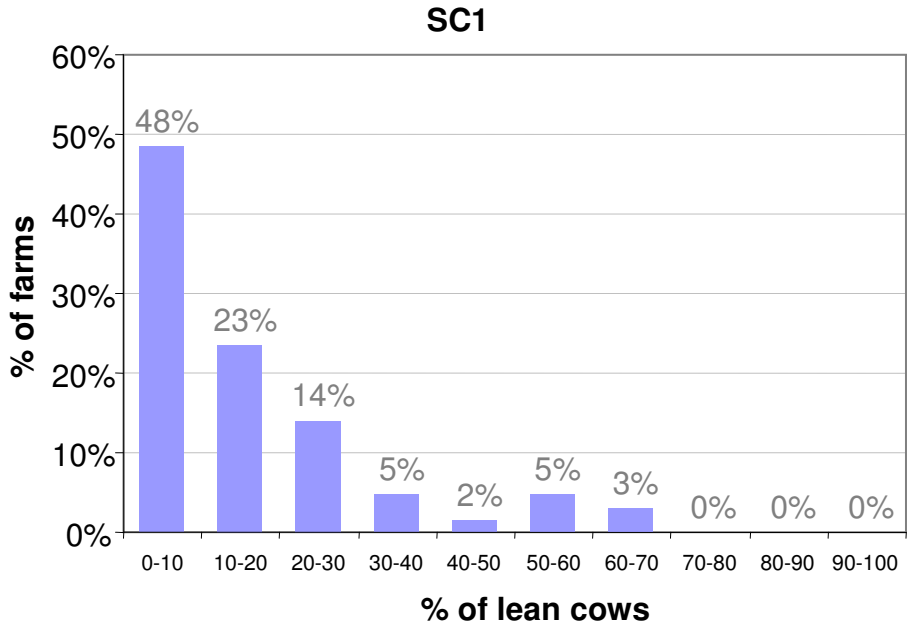


Table 4: Second dataset submitted to experts. They had to attribute a subcriterion score to each farm, taking into account the information provided on the distribution of the % of lean cows among the 64 dairy farms visited in WP2.4

	% of too thin animals (body score < 2)	% of farms with a % of too thin cows **inferior or equal** to the farm considered	***Score***
Farm 1	0	20%	
Farm 2	10	48%	
Farm 3	20	72%	
Farm 4	30	86%	
Farm 5	40	91%	
Farm 6	50	92%	
Farm 7	60	97%	
Farm 8	70	100%	
Farm 9	80	100%	
Farm 10	90	100%	
Farm 11	100	100%	

- **Final adjustment according to test on dataset**

Table 5: Results obtained from the experts asked

	p_0	Index	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*	*Caluclated score*
Farm 1	0%	100	*100*	*100*	*100*	*100*	*100*	100	100
Farm 2	10%	90	*60*	*75*	*65*	*20*	*20*	48	46
Farm 3	20%	80	*30*	*55*	*40*	*5*	*15*	29	30
Farm 4	30%	70	*22*	*40*	*30*	*5*	*10*	21	22
Farm 5	40%	60	*15*	*30*	*25*	*5*	*5*	16	16
Farm 6	50%	50	*10*	*25*	*20*	*5*	*0*	12	12
Farm 7	60%	40	*8*	*20*	*17*	*0*	*0*	9	8
Farm 8	70%	30	*5*	*10*	*15*	*0*	*0*	6	6
Farm 9	80%	20	*3*	*5*	*10*	*0*	*0*	4	4
Farm 10	90%	10	*1*	*2*	*5*	*0*	*0*	2	2
Farm 11	100%	0	*0*	*0*	*0*	*0*	*0*	0	0

To determine the most appropriate function ***u***, we applied another technique using I-splines instead of B-splines (because we knew that the scores varied in a monotonic way with the index), and with only one interior knot, here placed at 80. The function ***u*** is composed of two pieces of cubic curves, and writes as follows:

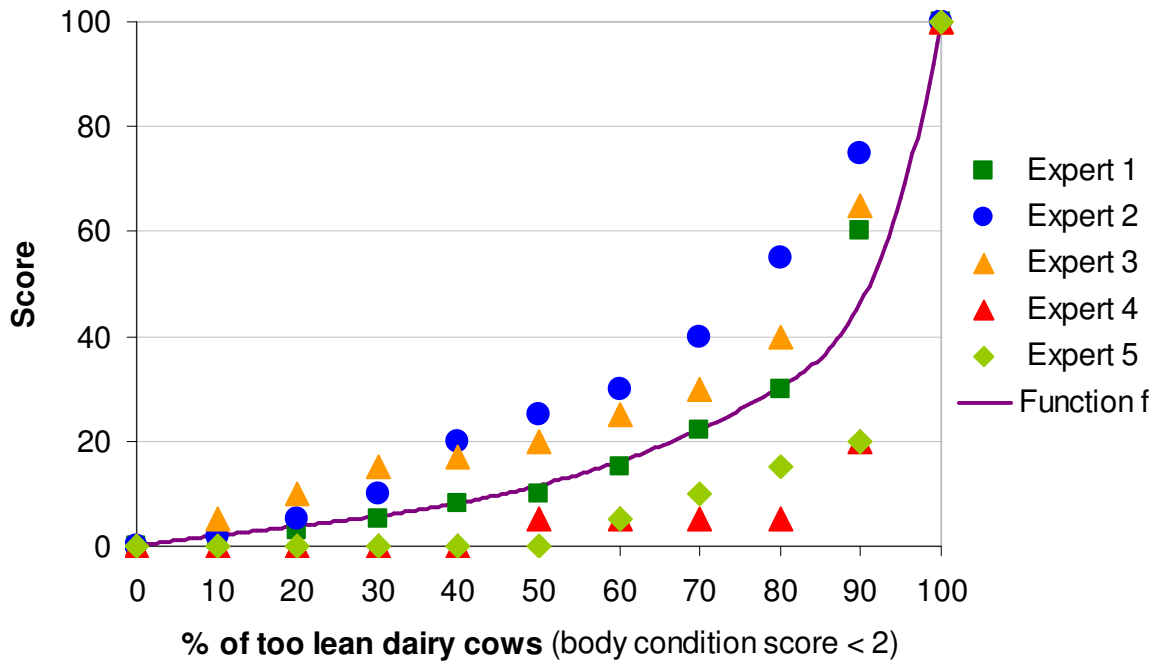
$$u(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the position of x in relation to the interior knot. These coefficients are presented in Table 6.

Table 6: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,80]$	$x \in [80,100]$
a_x	0	-2961.31541
b_x	0.22165963	111.270985
c_x	-0.00277075	-1.39088729
d_x	5.9271E-05	0.00584309

Figure 4: *I*-spline function (with one interior knot at 80) to attribute a subcriterion-score to 'Absence of prolonged hunger'



Calculation

Let *x* be a farm

Variables from raw data:
Let *N* be the number of animals observed for their body scores in the herd
Let M_k be the body condition score of the kth animal observed on farm x

Calculated variables:
Let n_0 be the number of animals considered as too lean
Let n_1 be the number of animals considered as normal
Let n_2 be the number of animals considered as too fat
Let p_0 be the percentage of too lean animals
Let p_1 be the percentage of normal animals
Let p_2 be the percentage of too fat animals

Variables with predefined values:
Let a_1, b_1, c_1, d_1, a_2, b_2, c_2 and d_2 be the parameters of the two pieces of the utility function
Let K be the position of the knot (i.e. limit between the two pieces of the utility function)

1. From raw data to data at farm level

$$n_0(x) \leftarrow 0$$
$$n_1(x) \leftarrow 0$$
$$n_2(x) \leftarrow 0$$

$p_0(x) \leftarrow 0$

$p_1(x) \leftarrow 0$

$p_2(x) \leftarrow 0$

For k = 1, …, *N(x)*

If $M_k(x) = 0$ then $n_0(x) \leftarrow n_0(x) + 1$

Else, if $M_k(x) = 1$ then $n_1(x) \leftarrow n_1(x) + 1$

Else $n_2(x) \leftarrow n_2(x) + 1$

End

$$p_0(x) \leftarrow \frac{n_0(x)}{N(x)}$$

$$p_1(x) \leftarrow \frac{n_1(x)}{N(x)}$$

$$p_2(x) \leftarrow \frac{n_2(x)}{N(x)}$$

2. From data at farm level to subcriterion score

$a_1 = 0$
$b_1 = 0.22165963$
$c_1 = -0.00277075$
$d_1 = 5.9271E-05$

$a_2 = -2961.31541$
$b_2 = 111.270985$
$c_2 = -1.39088729$
$d_2 = 0.00584309$

K = 80

$h(x) \leftarrow 0$
$g_1(x) \leftarrow 0$

$h(x) = 100 - p_0(x)$
If $h(x) < K$ then $g_1(x) \leftarrow a_1 + b_1\, h(x) + c_1\, h(x)^2 + d_1\, h(x)^3$
Else $g_1(x) \leftarrow a_2 + b_2\, h(x) + c_2\, h(x)^2 + d_2\, h(x)^3$

Subcriterion 2: Absence of prolonged thirst

Definition of the subcriterion

In this subcriterion, we aim at assessing whether the animals suffer or not from thirst. On farms, it is likely that thirst, when not enough water is available, does not lead to dehydration; hence the subcriterion will be assessed through water availability and not through signs of dehydration. By contrast, during transport and before slaughter, water can be unavailable for a prolonged time and signs of dehydration may be detected.

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11102	Number of water troughs	Cardinal	Herd or Group
11103	Total length of water troughs	Cardinal	Herd or Group
11104	Number of water bowls	Cardinal	Herd or Group
11105	Water flow	Ordinal	Herd or Group
11106	Cleanliness of drinkers	Ordinal	Herd or Group
11107	Number of animals	Cardinal	Herd or Group

Because signs of dehydration on animals are observable only in extreme cases, it has been decided to assess absence of prolonged thirst through five resources- and management-based measures (with an additional one considering the number of animals in the herd or group). For each group of animals the observer shall answer the questions:

- *Is the number of functioning drinkers sufficient?* (There exist some recommendations for these measures. For example, there should be at least one water bowl for 10 lactating cows[1]. If the ratio of animals per drinker is lower than the recommendation, then it is considered that the animals have sufficient access to water to cover their needs. If this ratio increases, the animals may queue and get thirsty, with a progressive increase of potential thirst with the increase of the animals/drinker ratio)
- *Are the drinkers clean?*
- *Are there at least 2 drinkers available for an animal?*

[1] If we look at the recommendations produced for the Council of Europe concerning water provision to dairy cows:

*"There should be enough water available for **at least 10%** of housed cattle to drink at any one time. Water bowls and troughs shall be kept clean and checked daily to make sure they are not blocked or damaged, and the water is flowing freely."*
[Extract of the draft revised recommendation concerning cattle, made for the Council of Europe by the Committee of the European Convention for the protection of animals kept for farming purposes, this text is not yet public]

If we consider this recommendation as a basis for our calculation, this leads to consider that there should be at least one "drinker place" for 10 cows.

Analysis of the constraints

Each group of dairy cows is described according to the 3 item listed above, which are evaluated separately. The three measures must be combined in order to bring a finite number of possible combinations. Then, a judgement on the quality of water provision on farm from unacceptable to very good must be attributed to each of the combinations.

If the cows are separated into several groups, this will require some integration of the information from groups to farm.

The subcriterion score will thus represent the ***intensity of the negative deviation to a recommendation for supply in water (with 100 when the recommendation is fulfilled).***

Type of mathematical construction proposed

To combine the three measures, we propose to use a lexicographic method, with a hierarchical organisation of the measures.

All the measures are considered as binary (Yes/No answer).

The fact that each animal can have access to at least two drinkers is considered as less important (positioned at the third place of the lexicographic tree), since we have yet considered the functioning state of the drinkers and their cleanliness. These two elements are the most important causes of problem, making an animal need access to another drinker.

This organisation leads to the lexicographic tree presented in Figure 1.

➔ *This method* ***forbids*** *compensations between measures. This is justified because the measures don't assess similar aspects and they* ***need to be considered together*** *to assess properly water provision.*

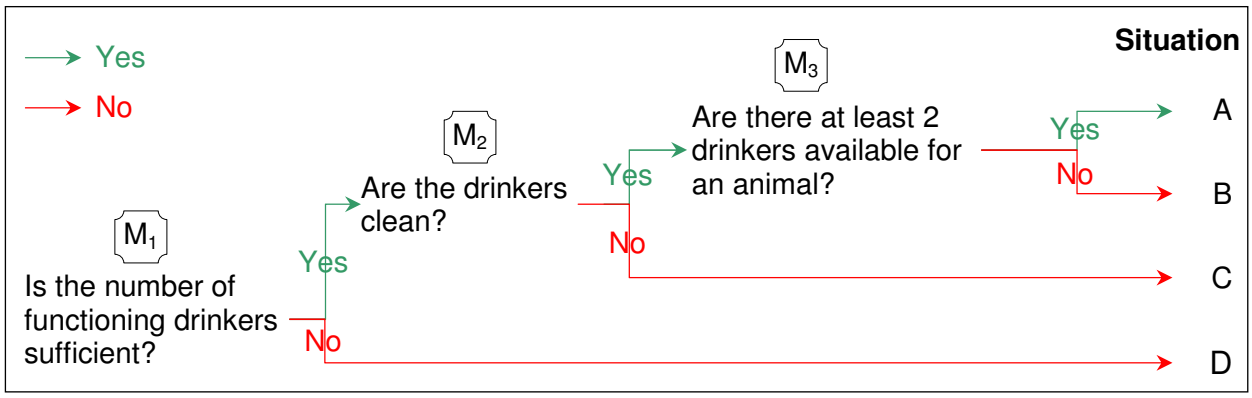
For each measure, we have to define the 'threshold' above which the answer will be 'yes'. In other words, we have to define clearly what we intend by 'yes' for each measure.
Example:

- *For M_1, if there are at least 1 drinker / 10 cows, the answer is 'yes'.*

Then, **the subcriterion-score to be attributed to a group of cows submitted to a given combination is calculated as the mean score attributed by experts to this combination**.

To aggregate, if necessary, the information provided on several groups, we consider the worst situation observed for at least 15% of the dairy cows.

Figure 1: Lexicographic tree defining 4 possible combinations of the measures for water provision



Parameters to be defined

First, we need to define how to answer the questions M1, M2 and M3, that implies to set some thresholds.

Second, we need to define the scores attributed to each of the four possible combinations (A, B, C and D).

Adjustments of the mathematical construction

- **Questions asked to experts**

Definition of the thresholds needed to be able to answer the questions:
We asked direct questions to the partners who developed the measures.

Scores attributed to the four possible situations defined in the decision tree (Figure 1):
Figure 1 was shown to experts. They were asked to estimate scores y_1 to y_4.
The experts asked were I. Veissier, E. Canali, B. Algers, J. Capdeville, S. Waiblinger.

- **Adjustment of the construction according to answers from experts**

Definition of the thresholds needed to be able to answer the questions:

The way we answer the questions M1, M2 and M3 depends on the type of barn: loose-housing system (noted LOOSE) or tied-stalls (noted TIED).

M1- Is the number of functioning drinkers sufficient?

LOOSE:

1- For each group and each type of drinkers, we look at the measure 'total length':

C. Winckler told us to calculate the total length of the drinkers available to the group of animals considered; then, to divide the total length by the number of animals in the group; and finally to compare this length per cow to the ratio 8 cm / cow:

If length per cow **< 8 cm**, then the answer is **NO**
Else, length per cow ≥ **8 cm**, then the answer is YES

2- If at the question 1- the answer is YES, then we look at the measure 'water flow sufficient':

No	/	Yes	/	Partly
0	/	1	/	2
NO		**YES**		**We divide by 2 the length per cow** (in each group) **then we go back to the first question:**

If the length per cow divided by 2 is **< 8cm** → **NO**
Else → **YES**

Then, if there are several groups of cows, we compare the results obtained in each group and we keep the worst situation observed on <u>at least 10% of the animals</u>.

TIED:

1- Look at the measure 'no. drinkers per cow':
If this number is ≥ **0.5**, then the answer is **YES**
Else, the answer is **NO**

2- If answer to the first question is YES, then we look at the measure 'water flow sufficient':

No	/	Yes	/	Partly
0	/	1	/	2
NO		**YES**		**We divide by 2 the number of drinkers per cow then we go back to the first question:**

If the number divided by 2 is ≥ **0.5** → **YES**
Else → **NO**

M2- Are the drinkers clean?

The measure on cleanliness of the drinkers is expressed as:

No	/	Yes	/	Partly
0	/	1	/	2
NO		**YES**		

M3- Are there at least 2 drinkers available for an animal?

LOOSE:
For each group, we look at the total number of drinkers (*troughs + bowls*) given to animals:
If this number is ≥ **2**, then the answer is **YES**
Else, the answer is **NO**

Then, if there are several groups of cows, we compare the results obtained in each group and we keep the worst situation observed on <u>at least 15% of the animals</u>.

TIED:
Measure 'no. Drinkers per cow':
If this number is ≥ **1**, then the answer is **YES**
*(**1** means that there is always 1 drinker between two cows)*
Else, the answer is **NO**

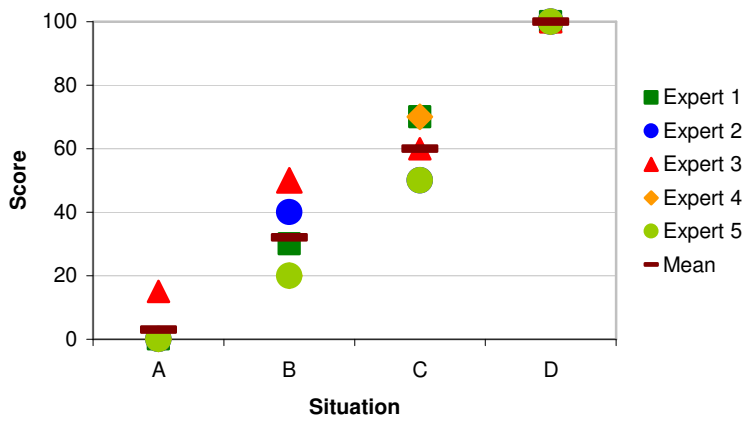
Scores attributed to the four possible situations defined in the decision tree (Figure 1):

The result of experts' evaluation of the four possible combinations is presented in Table 1.

Table 1: Evaluation of the four combinations by experts

	Situation	***Score***					
		Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	***Mean***
Farm 1	A	*0*	*0*	*15*	*0*	*0*	***3***
Farm 2	B	*30*	*40*	*50*	*20*	*20*	***32***
Farm 3	C	*70*	*50*	*60*	*70*	*50*	***60***
Farm 4	D	*100*	*100*	*100*	*100*	*100*	***100***

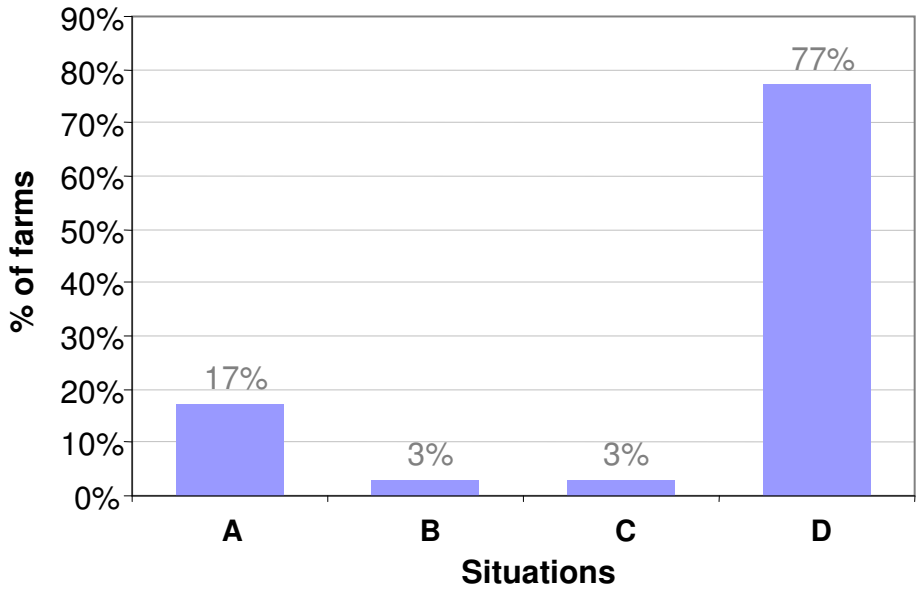
Figure 2: Graphic representation of the evaluation of the four combinations by experts



- **Test on dataset(s)**

We looked at the first 35 farms visited during WP2.4 to see the distribution of farms on SC1. We obtained the following results, both presented on Table 3 and on Figure 3.

Figure 3: Distribution of farms on SC2 (on 35 farms)



- **Final adjustment according to test on dataset**

Facing these results, we decided to contact another time the experts to define the thresholds and the way to be followed to answer the questions M1, M2 and M3. We proposed to them the following suggestions to revise the construction:

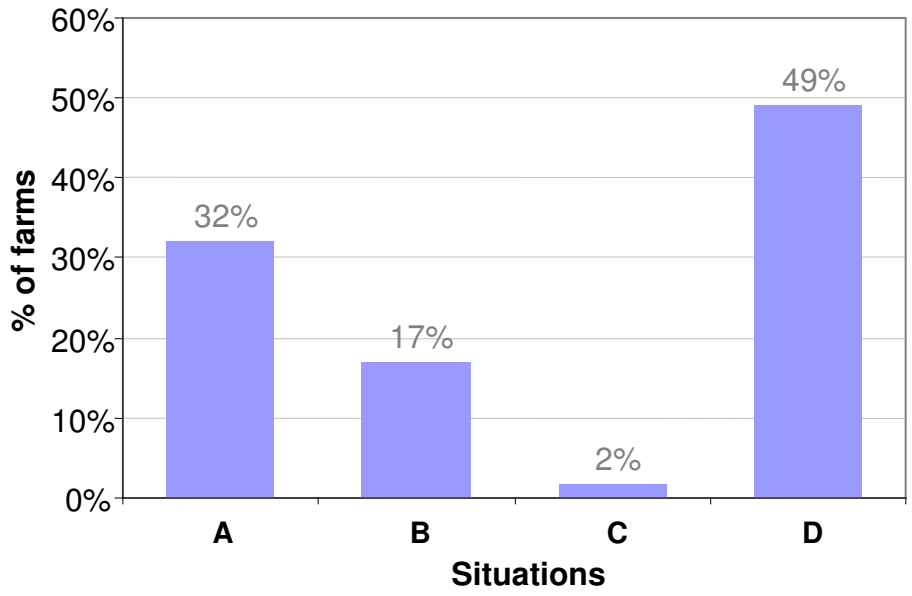
- consider separately water troughs and water bowls:
  - consider that a water bowl is sufficient to provide water to 10 cows (Recommendation from the Council of Europe)
  - take 4 cm as the minimum trough length per cow (in other words, a 40 cm trough for 1 cow to drink alone, a 80 cm trough for two cows to drink together, etc.)

- not to take into account water flow in tied barns or, alternatively, decide of a higher limit for water flow in tied barns. Please, provide us with a new limit in that case.

The five experts consulted reacted on these suggestions. These were accepted (without taking into account water flow for tied-stalls), except for the minimum length of water trough per cow that is set at 6 cm.

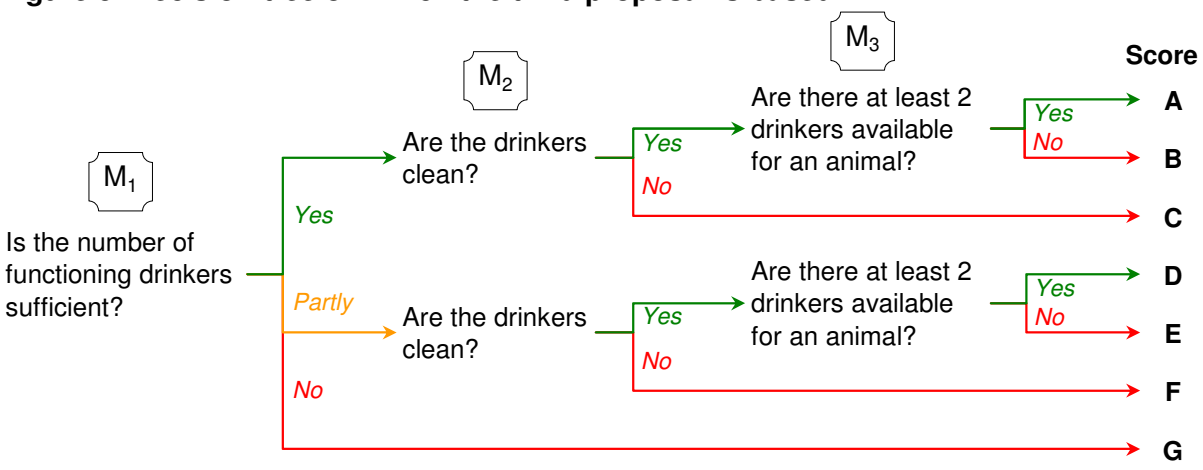
We now had the possibility to test the new proposal on 70 farms visited during WP2.4. The new distribution of farms is presented on Figure 4.

Figure 4: Second distribution of farms on SC2 (on 70 farms)



With this second proposal, there are still 49% of farms in the worst situation. This was surprising, especially for lactating cows since water is essential to produce milk. As a consequence we decided to elaborate a third proposal adding three new possible situations to the initial decision-tree (Figure 5).

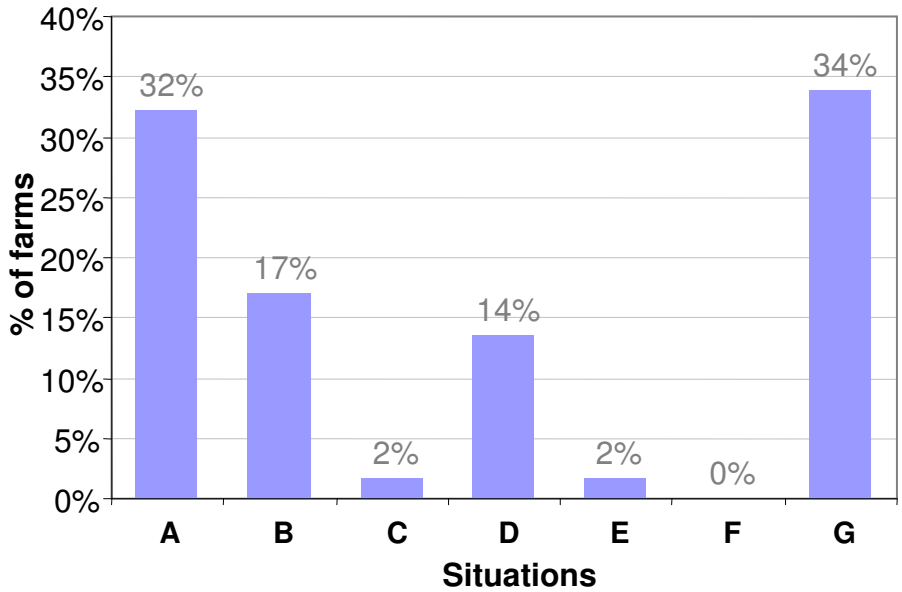
Figure 5: Decision-tree on which the third proposal is based



To be considered as sufficient, there must be at least 1 water bowl for 10 cows and/or 6 cm of trough per cow, taking into account the water flow (cf. the detailed description given for the first proposal). To be considered as ***partly sufficient***, there must be at least **1 water bowl for 15 cows** and/or **4 cm of trough per cow**.

With this proposal we obtained the distribution of farms presented on Figure 6.

Figure 6: Third distribution of farms on SC2 (on 70 farms)



This proposal will need to be confirmed once the final list of measures will have been produced. For the moment, we didn't ask experts to give their opinion on this new proposal. The scores attributed for Situations A, B, C and G are the same as the ones previously attributed. For the three new situations (D, E and F), we decided to set temporary values. The scores attributed to the seven situations are presented on Table 2.

Table 2: Scores attributed to the seven possible situations defined in the decision-tree (cf. Figure 5 for explanations on A, B… G)

Situation	Score
A	100
B	60
C	32
D	60
E	40
F	20
G	3

Calculation

Let x be a farm, and $x_1, \ldots, x_k$ be the name of the k groups in farm x

Variables from raw data:
Let n_t be the number of water troughs in the group of animals considered
Let l_t be the total length of water troughs in the group
Let n_b be the number of water bowls in the group
Let f be the evaluation of water flow (0. insufficient / 1. sufficient / 2. partly sufficient)
Let c be the evaluation of drinkers' cleanliness (0. dirty / 1. clean / 2. partly clean)
Let t be the type of housing system (L: loose / T. tied)
Let n be the number of animals in the group

Calculated variables:
Let n_{fa} be the maximum number of animals in the group to fit with full recommendations
Let n_{pa} be the maximum number of animals in the group to fit with partial recommendations
Let M_i be the answer to the ith question considered for farm x

Variables with predefined values:
Let r_{tf} and r_{tp} be respectively the recommendation concerning the minimum length of water trough per cow to have a fully sufficient and a partially sufficient water provision
Let r_{bf} and r_{bp} be respectively the recommendation concerning the maximum number of cows per water bowl to have a fully sufficient and a partially sufficient water provision
Let y_A to y_G be the scores attributed to the 7 possible values of $h(x)$

1. From raw data to data at farm level

$r_{tf} = 6$
$r_{tp} = 4$
$r_{bf} = 10$
$r_{bp} = 15$

For i=1 to k do

$n_a(x_i) \leftarrow 0$
$M_1(x_i) \leftarrow 0$
$M_2(x_i) \leftarrow 0$
$M_3(x_i) \leftarrow 0$

```
If t(x_i)=L        then
        If f(x_i)=0        then    M_1(x_i) ← 0
        Else, if f(x_i)=1  then
                n_fa(x_i) ← (l_t(x_i)/r_tf) + (n_b(x_i)×r_bf)
                n_pa(x_i) ← (l_t(x_i)/r_tp) + (n_b(x_i)×r_bp)
                If n_pa(x_i) < n(x_i)              then    M_1(x_i) ← 0
                        Else, if n_fa(x_i) < n(x_i)  then    M_1(x_i) ← 1
                                                     Else    M_1(x_i) ← 2
                        Else, if f(x_i)=1  then
        Else, if f(x_i)=2  then
                n_fa(x_i) ← n_fa(x_i)/2
                n_pa(x_i) ← n_pa(x_i)/2
                If n_pa(x_i) < n(x_i)              then    M_1(x_i) ← 0
                        Else, if n_fa(x_i) < n(x_i)  then    M_1(x_i) ← 1
                                                     Else    M_1(x_i) ← 2
```

Else ($t(x_i)$=T)
If $n_b(x_i)/n(x_i)<0.5$ then $M_1(x_i) \leftarrow 0$
Else $M_1(x_i) \leftarrow 1$

If $c(x_i)=0$ then $M_2(x_i) \leftarrow 0$
Else $M_2(x_i) \leftarrow 1$

If $t(x_i)$=L then
If $(n_t(x_i)+n_b(x_i))<2$ then $M_3(x_i) \leftarrow 0$
Else $M_3(x_i) \leftarrow 1$
Else ($t(x_i)$=T)
If $n_b(x_i)<1$ then $M_3(x_i) \leftarrow 0$
Else $M_3(x_i) \leftarrow 2$

End

2. From data at farm level to subcriterion score

$y_A = 100$
$y_B = 60$
$y_C = 32$
$y_D = 60$
$y_E = 40$
$y_F = 20$
$y_G = 3$

$g_2(x) \leftarrow 100$

For i=1 to k do

$h(x_i) \leftarrow 0$
$g_2(x_i) \leftarrow 0$

If $M_1(x_i)=0$ then $h(x_i) \leftarrow G$ and $g_2(x_i) \leftarrow y_G$
Else, if $M_1(x_i)=1$
if $M_2(x_i)=0$ then $h(x_i) \leftarrow F$ and $g_2(x_i) \leftarrow y_F$
Else, if $M_3(x_i)=0$ then $h(x_i) \leftarrow E$ and $g_2(x_i) \leftarrow y_E$
Else $h(x_i) \leftarrow D$ and $g_2(x_i) \leftarrow y_D$
Else ($M_1(x_i)=2$)
if $M_2(x_i)=0$ then $h(x_i) \leftarrow C$ and $g_2(x_i) \leftarrow y_C$
Else, if $M_3(x_i)=0$ then $h(x_i) \leftarrow B$ and $g_2(x_i) \leftarrow y_B$
Else $h(x_i) \leftarrow A$ and $g_2(x_i) \leftarrow y_A$

If $\dfrac{n(x_i)}{\sum_{i=1}^{k} n(x_i)} < 0.15$ then $g_2(x) \leftarrow g_2(x)$
Else, if $g_2(x_i) < g_2(x)$ then $g_2(x) \leftarrow g_2(x_i)$
Else $g_2(x) \leftarrow g_2(x)$

End

Subcriterion 3: Comfort around resting

Definition of the subcriterion

This subcriterion assesses the comfort of the animals around resting, through essentially behavioural observations, but not through the injuries that are included in Subcriterion 6 "Absence of injuries".

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11108	Time needed to lie down	Cardinal	Individual
11109	Percentage of animals colliding with housing equipment during lying down	Cardinal	Herd
11110	Percentage of animals lying partly or completely outside the supposed lying area	Cardinal	Herd
11111	Cleanliness: Udders	Ordinal	Individual
11112	Cleanliness: Flank & upper legs	Ordinal	Individual
11113	Cleanliness: Legs	Ordinal	Individual

In order to assess 'Comfort around resting' for **Dairy Cows**, 6 measures are to be considered, 3 for behaviour around resting and 3 for cleanliness.

'**Cleanliness**' will be assessed on three areas of the body ('lower legs', 'udder' and 'flank & upper legs') on a 4 level scale (ordinal scale with 4 levels: from 1-clean to 4-very dirty with plaques of manure). **For each area**, the information we will have to deal with at the herd level is **the percentage of animals observed in the classes 3 & 4** (dirty animals).

For **'behaviour around resting'**:
The 'Time needed to lie down' is evaluated at herd level and represents the **mean time** (in seconds) from a minimum of 6 lying down movements.
For 'Frequencies of lying partly or completely outside the supposed lying area', we consider the % of animals showing these behaviours.
For 'Frequencies of collisions* with housing equipment during lying down', we consider the number of collisions seen and/or heard divided by the number of lying down or rising events. The frequency is expressed as a %.

Analysis of the constraints

Since six measures are included in this subcriterion, it is not essential to use an aggregation process which keeps all the precision given by raw data for each measure.
We propose to consider at the herd level only 3 categories for each measure.

*The subcriterion score qualifies the **frequency of problems encountered by animals when resting**.*

Type of mathematical construction proposed

For each measure, we consider 3 levels from a welfare point of view:

- Normal (no problem)
- Moderate problem
- Serious problem

To do that, we need to establish for each measure two thresholds to which we can compare the measures observed.

Table 1: Organisation and names of the thresholds associated to the measures.

	Data measured	Normal	Moderate problem	Serious problem
Time needed to lie down	M_1	$\leq \mathbf{T_{1,1}}$	$T_{1,1} < \leq T_{1,2}$	$> \mathbf{T_{1,2}}$
Frequency of animals lying partly or completely outside the supposed lying area	M_2	$\leq \mathbf{T_{2,1}}$	$T_{2,1} < \leq T_{2,2}$	$> \mathbf{T_{2,2}}$
Frequency of collisions with housing equipment during lying down	M_3	$\leq \mathbf{T_{3,1}}$	$T_{3,1} < \leq T_{3,2}$	$> \mathbf{T_{3,2}}$
Cleanliness: lower legs	M_4	$\leq \mathbf{T_{4,1}}$	$T_{4,1} < \leq T_{4,2}$	$> \mathbf{T_{4,2}}$
Cleanliness: udder	M_5	$\leq \mathbf{T_{5,1}}$	$T_{5,1} < \leq T_{5,2}$	$> \mathbf{T_{5,2}}$
Cleanliness: flank and upper legs	M_6	$\leq \mathbf{T_{6,1}}$	$T_{6,1} < \leq T_{6,2}$	$> \mathbf{T_{6,2}}$

For some measures taken at individual level (M_4, M_5 and M_6), we need another threshold to which we can compare the measure obtained on each animal: $\mathbf{t_4}$, $\mathbf{t_5}$ and $\mathbf{t_6}$.

The subcriterion-score's calculation will be performed in **two steps**, the first one concerning the measures observed at the individual level (M_1, M_4, M_5 and M_6), and the second one concerning all the measures.

First step of calculation, for the measures observed at the individual level (M_1, M_4, M_5 and M_6):

For the <u>time needed to lie down</u>, since only few animals can be observed during an on-farm visit, the best way to aggregate the information collected into a data at herd level is to consider the mean time needed to lie down by the observed animals.
Let $M_{1k}(x)$ be the time measured for the animal k on farm x, and $n_1(x)$ be the number of animals observed for "time needed to lie down" on farm x.

$$M_1(x) = \frac{\sum_{k=1}^{n_1(x)} M_{1k}(x)}{n_1(x)}$$

For the <u>measures of cleanliness</u> (M_4, M_5 and M_6), we will calculate the percentage of animals with a cleanliness score **≥ 3** (i.e. **dirty animals**).
$t_4 = t_5 = t_6 = 3$

Second step of calculation, for all the measures:

1. For each measure i ($\in$ [1,6]), we will compare the data $M_i(x)$ measured on farm x (or obtained through the first step of the calculation) to the thresholds $T_{i,1}$ and $T_{i,2}$. Actually, we recode the cardinal information obtained on each measure into an ordinal one.

2. Then, we count the number of measures considered as **moderately** problematic and the number of measures considered as **seriously** problematic.
   As there are 3 measures for *cleanliness* and 3 measures for *behaviour around resting*, we may introduce weights so to obtain a balance between the two types of measures. For example, we could attribute an overall weight of ¼ for cleanliness and ¾ for resting behaviour. To do that we could count **three times** the measures about *behaviour around resting*, and only **once** the *cleanliness* measures.

   So, the information we will have to deal with will be:
   - Number of measures considered as normal $N_0(x)$
   - Number of measures considered as moderately problematic $N_1(x)$
   - Number of measures considered as seriously problematic $N_2(x)$

3. We will then calculate the subcriterion score g_3 as follows:

$$g_3(x) = u\left(100\left(1 - \frac{\sum_{i=0}^{2} w_i \left(\frac{N_i(x)}{N_{tot}} \times 100\right)}{w_2}\right)\right) = u(h(x)) = u(Index) \qquad \text{with } N_{tot} = \sum_{i=0}^{2} N_i$$

   We may need to use a function ***u*** to produce a subcriterion score corresponding to the score that could have been given by experts.

Parameters to be defined

We need to define:
- the thresholds at herd level ($\mathbf{T_{1,1}}$ to $\mathbf{T_{6,2}}$),
- the relative weight of cleanliness in comparison to behavioural measures,
- the weights $\mathbf{w_i}$ attributed to the three categories in respect to the severity of the problem for welfare around resting
- the choice of the most appropriate **function *u*** to calculate the subcriterion score $\boldsymbol{g_3}$.

Adjustments of the mathematical construction

- **Questions asked to experts**

To define the thresholds ($\mathbf{T_{1,1}}$ to $\mathbf{T_{6,2}}$)

We asked two experts (WP2.2 partners who worked on these measures) to set such thresholds.
The experts were U. Knierim and B. Whay.

To set the thresholds $\mathbf{T_{1,1}}$ to $\mathbf{T_{6,2}}$ we asked experts to fill in Table 2.

Table 2: Table presented to experts in order to define the thresholds attributed to the measures.

	Normal	Moderate problem	Serious problem
Time needed to lie down	≤ s	s < ≤ s	> s
Frequency of animals lying partly or completely outside the supposed lying area	≤ %	% < ≤ %	> %
Frequency of collisions with housing equipment during lying down	≤ %	% < ≤ %	> %
Cleanliness: lower legs	≤ %	% < ≤ %	> %
Cleanliness: udder	≤ %	% < ≤ %	> %
Cleanliness: flank and upper legs	≤ %	% < ≤ %	> %

To set the relative weight of cleanliness in comparison to behaviour measures

We asked the two precedent experts what relative weight they attribute to behaviour around resting and cleanliness.

To set the weights w_i and choose the most appropriate function f

We made experts react on virtual dataset (Table 3), asking them first to rank the virtual farms from the best one to the worst one, and then to give a score to each of the farms presented in the dataset.
The asked experts were I. Veissier, J. Capdeville, H.R. Whay, U. Knierim, E. Canali, C. Winckler.

The dataset was composed of virtual farms described by the number of measures within each category 'normal', 'moderate problem', and 'serious problem', respectively n_0, n_1 and n_2.

Table 3: Dataset submitted to experts. They had to sort the farms from best to worst, then to attribute a subcriterion score to each farm.

	n_0	n_1	n_2	*Rank*	*Score*
Farm 1	12	0	0		
Farm 2	8	4	0		
Farm 3	8	0	4		
Farm 4	6	6	0		
Farm 5	6	0	6		
Farm 6	4	8	0		
Farm 7	4	0	8		
Farm 8	0	12	0		
Farm 9	0	8	4		
Farm 10	0	6	6		
Farm 11	0	4	8		
Farm 12	0	0	12		

Data presented in Table 3 correspond to a situation where experts decided to attributed three times more importance to measures about resting behaviour than to measures about cleanliness.

The ranks given by experts will help us define the most appropriate weights $\mathbf{w_i}$.
The scores attributed by experts will help us find the most appropriate function *u*.

- **Adjustment of the construction according to answers from experts**

To define the thresholds (t_1, then $T_{1,1}$ to $T_{6,2}$)

The thresholds $\mathbf{T_{1,1}}$ to $\mathbf{T_{6,2}}$ have been defined with U. Knierim and B. Whay:

Table 4: Thresholds defined by experts for the different measures.

	Normal	Moderate problem	Serious problem
Time needed to lie down	≤ **5.20 s**	5.20 s < ≤ 6.30 s	> **6.30 s**
Frequency of animals lying partly or completely outside the supposed lying area	≤ **3**%	3% < ≤ 5%	> **5**%
Frequency of collisions with housing equipment during lying down	≤ **20**%	20% < ≤ 30%	> **30**%
Cleanliness: lower legs	≤ **20%**	20% < ≤ 50%	> **50%**
Cleanliness: udder	≤ **10%**	10% < ≤ 19%	> **19%**
Cleanliness: flank and upper legs	≤ **10%**	10% < ≤ 19%	> **19%**

To set the relative weight of cleanliness in comparison to behaviour measures

Experts consider that we should attribute an overall weight of ¾ for resting behaviour and ¼ for cleanliness, this leads to consider three times the values obtained for resting behaviour and only once the cleanliness values.

As a consequence $\boldsymbol{N_{tot} = 12}$ *(= 3×3 + 1×3).*

To set the weights w_i and choose the most appropriate function f

The results of experts' evaluation of farms with various numbers of measures within each category (Normal / Moderate problem / Serious problem) are shown in Table 5 (Columns 5-10). We calculated the mean of subcriterion-scores attributed by the 6 experts; then the farms were sorted according to the mean subcriterion-score they obtained (Table 5, Columns 11-12).

The weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ were determined by linear optimisation so as to match the ranking of farms according to mean score attributed to experts:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 4$
$\mathbf{w_2} = 9$

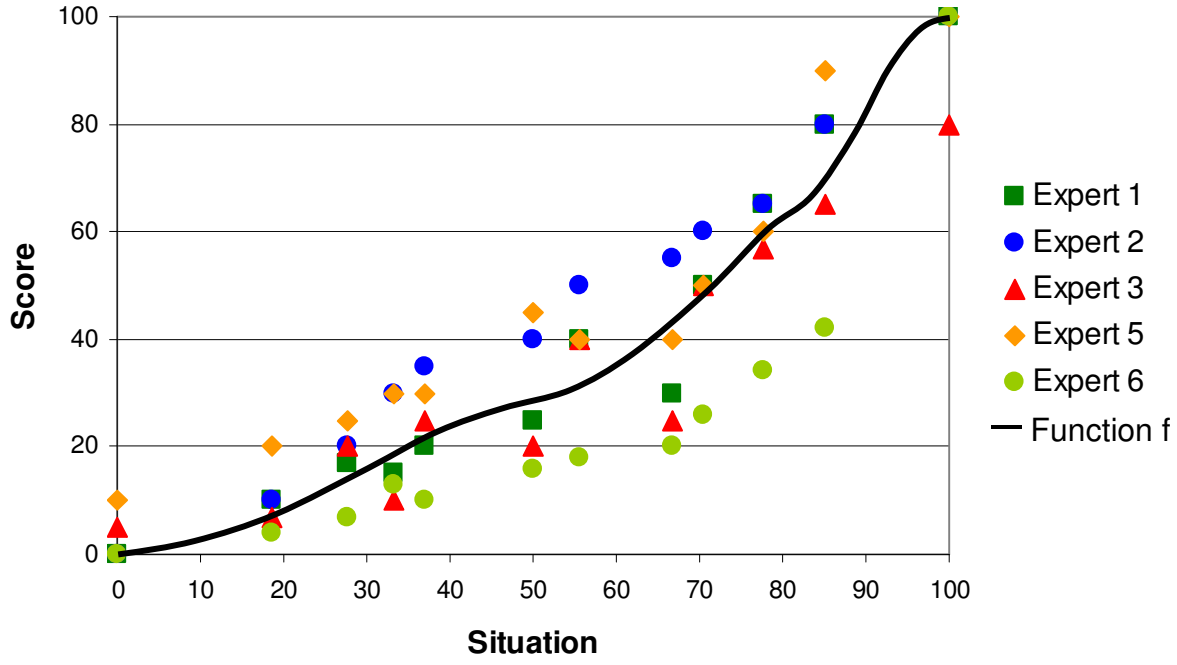
The result of the index resulting from the weighted sum obtained by each farm is shown in Table 5, column 13.

Finally B-spline functions were defined so as to match best experts socres (i.e. minimise the sum of square errors with experts' scores) (Figure 1).

Table 5: Answers from experts

	Situation			Score							Rank	Index	Calculated score
	p0	p1	p2	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Mean			
Farm 12	0	0	12	0	0	5	0	10	0	3	12	0	**0**
Farm 11	0	4	8	10	10	7	7	20	4	10	11	19	**7**
Farm 10	0	6	6	17	20	20	12	25	7	17	10	28	**14**
Farm 7	4	0	8	15	30	10	22	30	13	20	9	33	**19**
Farm 9	0	8	4	20	35	25	20	30	10	23	8	37	**22**
Farm 5	6	0	6	25	40	20	40	45	16	31	7	50	**29**
Farm 8	0	12	0	40	50	40	30	40	18	36	6	56	**32**
Farm 3	8	0	4	30	55	25	60	40	20	38	5	67	**44**
Farm 6	4	8	0	50	60	50	45	50	26	47	4	70	**49**
Farm 4	6	6	0	65	65	57	50	60	34	55	3	78	**60**
Farm 2	8	4	0	80	80	65	70	90	42	71	2	85	**71**
Farm 1	12	0	0	100	100	80	100	100	100	97	1	100	**100**

Figure 1: B-spline function to attribute a subcriterion-score to 'Comfort around resting'



- **Test on dataset(s)**

We looked at 51 farms visited during WP2.4 to see the distribution of these farms on SC3. We obtained the results presented, in a cumulative way, in the fourth column of Table 6.

We then asked another time experts to react on the reorganised dataset (cf. Table 6), including the information on the distribution of farms on SC3.

Table 6: Second dataset submitted to experts. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	Normal	Moderate problem	Serious problem	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 12	0	0	12	4%		
Farm 11	0	4	8	8%		
Farm 10	0	6	6	18%		
Farm 7	4	0	8	31%		
Farm 9	0	8	4	33%		
Farm 5	6	0	6	71%		
Farm 8	0	12	0	75%		
Farm 3	8	0	4	78%		
Farm 6	4	8	0	78%		
Farm 4 *	6	6	0	98%		
Farm 2	8	4	0	100%		
Farm 1	12	0	0	100%		

- **Final adjustment according to test on dataset**

The new answers of experts are presented in Table 7.

Table 7: Results obtained from the experts asked

	Situation			Score							*Rank*	Index	**Calculated score**
	p0	p1	p2	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Expert 6*	*Mean*			
Farm 12	0	0	12	*0*	*0*	*5*	*0*	*10*	*0*	3	*12*	0	**0**
Farm 11	0	4	8	*10*	*10*	*10*	*10*	*20*	*4*	11	*11*	19	**12**
Farm 10	0	6	6	*20*	*20*	*15*	*15*	*25*	*7*	17	*10*	28	**17**
Farm 7	4	0	8	*22*	*30*	*21*	*25*	*30*	*13*	24	*9*	33	**20**
Farm 9	0	8	4	*30*	*40*	*45*	*20*	*30*	*10*	29	*8*	37	**22**
Farm 5	6	0	6	*40*	*50*	*16*	*40*	*35*	*16*	33	*7*	50	**30**
Farm 8	0	12	0	*55*	*55*	*35*	*30*	*40*	*18*	39	*6*	56	**34**
Farm 3	8	0	4	*60*	*60*	*40*	*50*	*40*	*20*	45	*5*	67	**43**
Farm 6	4	8	0	*70*	*60*	*70*	*55*	*40*	*26*	54	*4*	70	**47**
Farm 4	6	6	0	*80*	*65*	*50*	*60*	*45*	*34*	56	*3*	78	**56**
Farm 2	8	4	0	*90*	*70*	*75*	*70*	*50*	*42*	66	*2*	85	**68**
Farm 1	12	0	0	*100*	*100*	*80*	*100*	*100*	*100*	97	*1*	100	**100**

The weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ remain the same as previously:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 4$
$\mathbf{w_2} = 9$

To determine the most appropriate function ***u***, we applied another technique using I-splines instead of B-splines, and with only one interior knot, here placed at 65. The function ***u*** is composed of two pieces of cubic curves, and writes as follows:

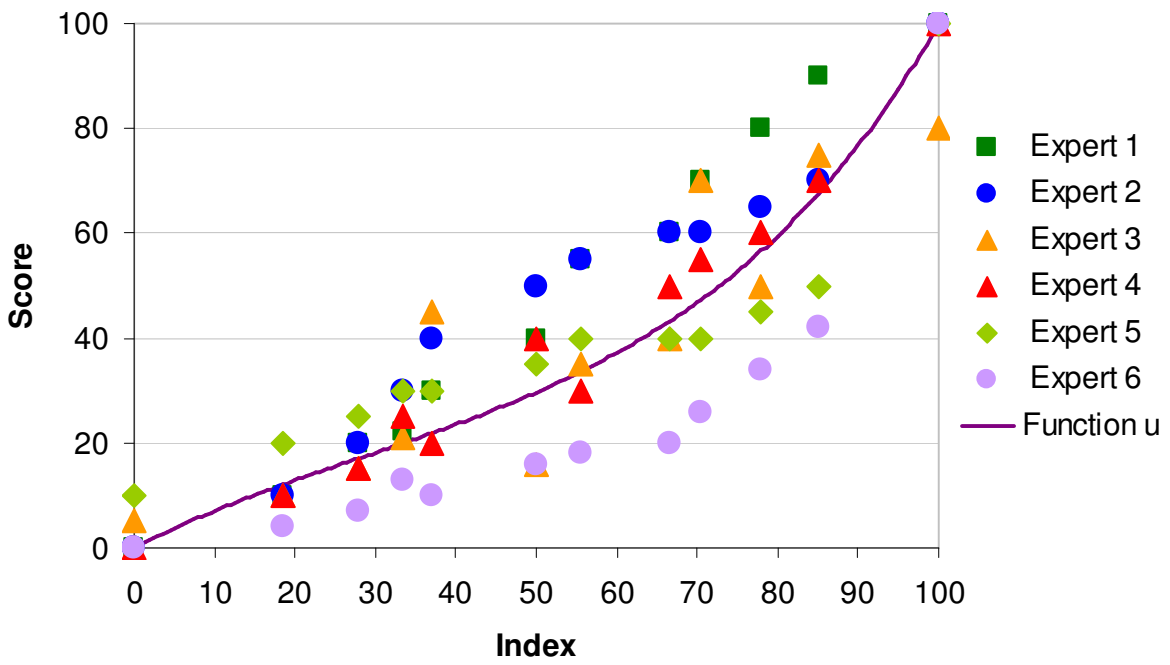
$$u(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 8.

Table 8: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,65]$	$x \in [65,100]$
a_x	0	-35.20034
b_x	0.78646981	2.41110089
c_x	-0.00932194	-0.03431626
d_x	0.00010908	0.00023725

Figure 2: I-spline function (with one interior knot at 65) to attribute a subcriterion-score to 'Comfort around resting'



Calculation

Let x be a farm

Variables from raw data:
Let n_i be the number of animals observed for the measure i ($i \in \{1,2,4,5,6\}$)
Let M_{ik} be the data obtained by the kth animal observed on the farm x for measure i
Let M_3 be the % of lying cows which lie partly outside lying area

Calculated variables:
Let d_2 be the number of animals colliding with housing equipment during lying down
Let d_4 be the number of animals with dirty legs
Let d_5 be the number of animals with dirty udder
Let d_6 be the number of animals with dirty flanks and upper legs
Let M_1 be the mean duration of lying down across animals
Let M_2 be the % of lying down movements with collision with housing equipment
Let M_4 be the % of cows with dirty legs
Let M_5 be the % of cows with dirty udder
Let M_6 be the % of cows with dirty flanks and upper legs
Let c_i be the category ($i \in \{1, \ldots,6\}$) to which the measure i is assigned ($c_i \in \{0,1,2\}$)
Let N_j be the number of measures in the jth category ($j \in \{0,1,2\}$)

Variables with predefined values:
Let t_4, t_5 and t_6 be the thresholds above whose animals are considered as dirty
Let $T_{1,1}$ to $T_{6,2}$ be the thresholds used to delimit the three categories of problems in respect to welfare
Let w_r and w_c be the weights associated to measures linked respectively to *behaviour around resting* and *cleanliness*
Let w_j be the weights associated to the category j
Let N_{tot} be total number of evaluations to be taken into account
Let a_1, b_1, c_1, d_1, a_2, b_2, c_2 and d_2 be the parameters of the two pieces of the utility function
Let K be the position of the knot (i.e. limit between the two pieces of the utility function)

1. From raw data to data at farm level

$t_4=3$
$t_5=3$
$t_6=3$

$d_2(x) \leftarrow 0$
$d_4(x) \leftarrow 0$
$d_5(x) \leftarrow 0$
$d_6(x) \leftarrow 0$
$M_1(x) \leftarrow 0$
$M_2(x) \leftarrow 0$
$M_4(x) \leftarrow 0$
$M_5(x) \leftarrow 0$
$M_6(x) \leftarrow 0$

$$M_1(x) = \frac{\sum_{k=1}^{n_1(x)} M_{1k}(x)}{n_1(x)}$$

For i = 2, 4, 5 and 6
    $d_i(x) \leftarrow 0$
    For k = 1, …, $n_i(x)$
        If $M_{ik}(x) \geq t_i$ then $d_i(x) \leftarrow d_i(x) + 1$
        Else $d_i(x) \leftarrow d_i(x)$
    End
    $M_i(x) \leftarrow \frac{d_i(x)}{n_i(x)} \times 100$
End

2. From data at farm level to subcriterion score

$T_{1,1} = 5.2$ $T_{1,2} = 6.3$
$T_{2,1} = 3$ $T_{2,2} = 5$
$T_{3,1} = 20$ $T_{3,2} = 30$
$T_{4,1} = 20$ $T_{4,2} = 50$
$T_{5,1} = 10$ $T_{5,2} = 19$
$T_{6,1} = 10$ $T_{6,2} = 19$

$w_r = 3$
$w_c = 1$

$w_0 = 0$
$w_1 = 4$
$w_2 = 9$

$a_1 = 0$
$b_1 = 0.78646981$
$c_1 = -0.00932194$
$d_1 = 0.00010908$

$a_2 = -35.20034$
$b_2 = 2.41110089$
$c_2 = -0.03431626$
$d_2 = 0.00023725$

K = 65

$c_0(x) \leftarrow 0$
$c_1(x) \leftarrow 0$
$c_2(x) \leftarrow 0$

$N_0(x) \leftarrow 0$
$N_1(x) \leftarrow 0$
$N_2(x) \leftarrow 0$

$h(x) \leftarrow 0$
$g_3(x) \leftarrow 0$

$N_{tot} \leftarrow 3\ w_r + 3\ w_c$

For i = 1, ..., 6
    If $M_i(x) \geq T_{i,2}$ then $c_i(x) \leftarrow 2$

Else if $M_i(x) < T_{i,1}$ then $c_i(x) \leftarrow 0$
Else $c_i(x) \leftarrow 1$

End

For i = 1 to 3
If $c_i(x) = 0$ then $N_0(x) \leftarrow N_0(x) + w_r$
Else, if $c_i(x) = 1$ then $N_1(x) \leftarrow N_1(x) + w_r$
Else $N_2(x) \leftarrow N_2(x) + w_r$

End

For i = 4 to 6
If $c_i(x) = 0$ then $N_0(x) \leftarrow N_0(x) + w_c$
Else, if $c_i(x) = 1$ then $N_1(x) \leftarrow N_1(x) + w_c$
Else $N_2(x) \leftarrow N_2(x) + w_c$

End

$$h(x) \leftarrow \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{j=0}^{2} w_j \; N_j(x)}{w_2} \right) \right)$$

If $h(x) < K$ then $g_3(x) \leftarrow a_1 + b_1\, h(x) + c_1\, h(x)^2 + d_1\, h(x)^3$
Else $g_3(x) \leftarrow a_2 + b_2\, h(x) + c_2\, h(x)^2 + d_2\, h(x)^3$

Subcriterion 5: Ease of movement

Definition of the subcriterion

This subcriterion aims at assessing the locomotion of the animals, i.e. the opportunity to move and the ease of locomotion, linked closely to the quality of the environment and not due to the health of the animal (health is considered elsewhere, in Subcriteria 6, 7 and 8)

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11114	Are cows tethered?	Binary	Herd
11115	Number of days per year of outdoor loafing area (OLA) access	Cardinal	Herd
11116	Number of hours per day of OLA	Cardinal	Herd
11117	Number of days per year at pasture	Cardinal	Herd
11118	Number of hours per day at pasture	Cardinal	Herd

Animals can be kept indoor all around the year or can have access to pasture during summer. In addition, when animals are tethered in the barn they can be released regularly (from twice a week to daily outdoor runs, with a great variability of the duration of these releases from one farm to another). Despite the fact that it may be better to release the cows three times a week for several hours than daily short releases, it seems impossible to take into account this point. We propose to simplify the assessment of the tethering system defining only 4 categories:

- Tethered all year around
- Tethered all year around with regular exercise
- Tethered during winter and at pasture during summer
- Tethered during winter, but with regular exercise, and at pasture during summer

Since it seems easier to ask questions on the number of days with access to pasture or outdoor loafing area (OLA) and the average number of hours per day, we will have to derive the information we need from the information collected on farms.

Analysis of the constraints

There is no constraint on this assessment since only few different situations are taken into account.

*The subcriterion score should qualify the **possibility of walking movements**.*

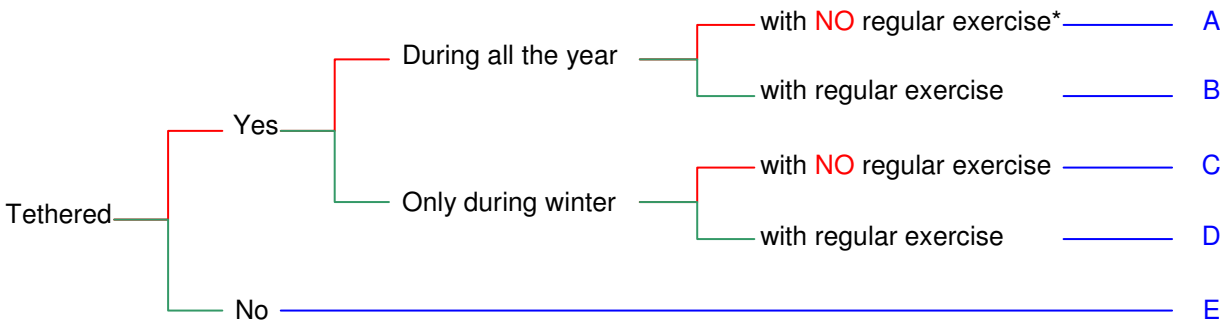
Type of mathematical construction proposed

We plan to construct an ordinal scale composed of 5 categories (from A to E) (Figure 1):

We propose to attribute a score to each category of the ordinal scale used:

- A → y_A (probably 0)
- B → y_B
- C → y_C
- D → y_D
- E → y_E (probably 100)

Figure 1: Combination of conditions for tethering.



* 'regular exercise' = from twice a week to daily outdoor (or possibly indoor) runs (which last at least 1h)

To answer the questions asked within the decision tree, we have to define three new variables:

- *T*: Number of days the animals are ***tethered***
- *E*: Number of days the animals have ***exercise*** (i.e. number of days during which animals have access to an OLA or pasture for at least 1 hour but less then 6 hours)
- *O*: Number of days the animals are ***outside*** (i.e. number of days during which animals have access to an OLA or pasture for at least 6 hours)

with $T + O = 365$

Then, if the animals spend **at least 350 days** per year 'oustide', we consider they are not tethered. Else, if they spend **at least 100 days** per year 'oustide', we consider they are tethered only during winter. Else (i.e. $O < 100$ days), we consider they are tethered during all the year.

To know if the animals, when they are tethered, have regular exercise, we compare the **ratio *E* / *T*** to a threshold. If this ratio is **above 0.5** (i.e. when animals are not outside, at least 1 every 2 days they have exercise), then we consider they have regular exercise.

Parameters to be defined

For X={A,B…E}, we need to define with experts the score y_X corresponding to the category X of the ordinal scale used.

Adjustments of the mathematical construction

▪ Questions asked to experts

We made experts react on the decision tree (Figure 1), asking them to give a score to each of the situations (y_A to y_E).
The asked experts were I. Veissier, J. Capdeville, E. Canali, C. Winckler, U. Knierim.

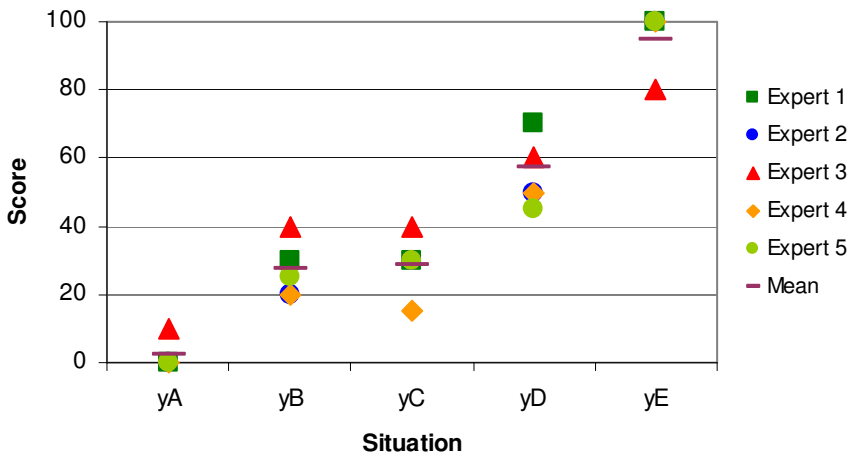
▪ Adjustment of the construction according to answers from experts

The mean of scores attributed by the experts to a given combination is considered the subcriterion-score to be attributed to a farm corresponding to that combination.

Table 1: Evaluation of the 5 combinations by experts.

Situation	Score					
	Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*
y_A	*0*	*0*	*10*	*0*	*0*	2,5
y_B	*30*	*20*	*40*	*20*	*25*	27,5
y_C	*30*	*30*	*40*	*15*	*30*	28,75
y_D	*70*	*50*	*60*	*50*	*45*	57,5
y_E	*100*	*100*	*80*	*100*	*100*	95

Figure 2: Graph representing experts' answers.

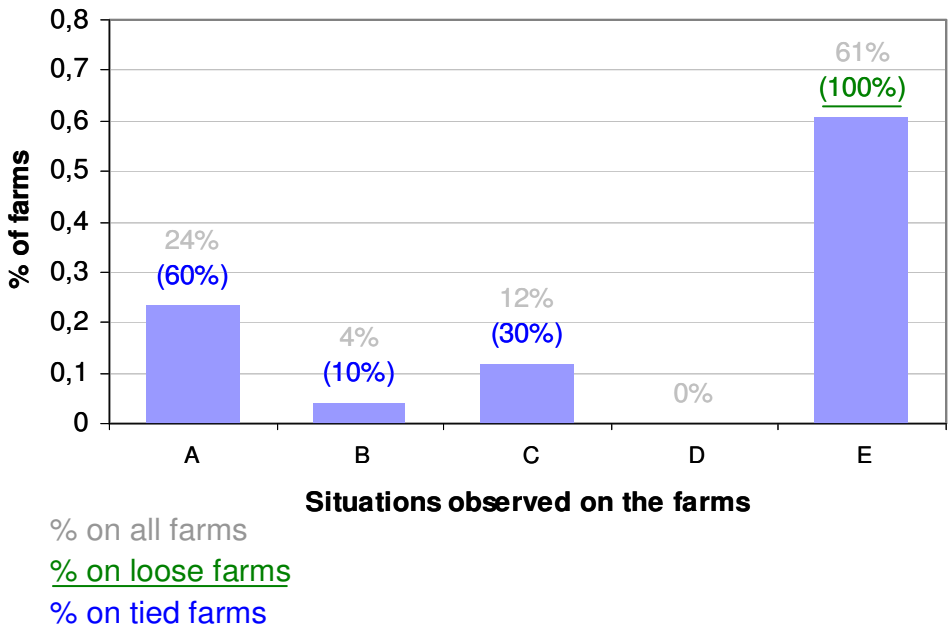


- **Test on dataset(s)**

We looked at the first 51 farms (31 in loose-housing systems and 20 in tied-barns) visited during WP2.4 to see the distribution of farms on SC5. We obtained the distribution of farms presented on Figure 3.

We then recontacted the experts asking them to refine their first assessments (if they considered it was necessary).

Figure 3: Distribution of 51 dairy farms visited in WP2.4



- **Final adjustment according to test on dataset**

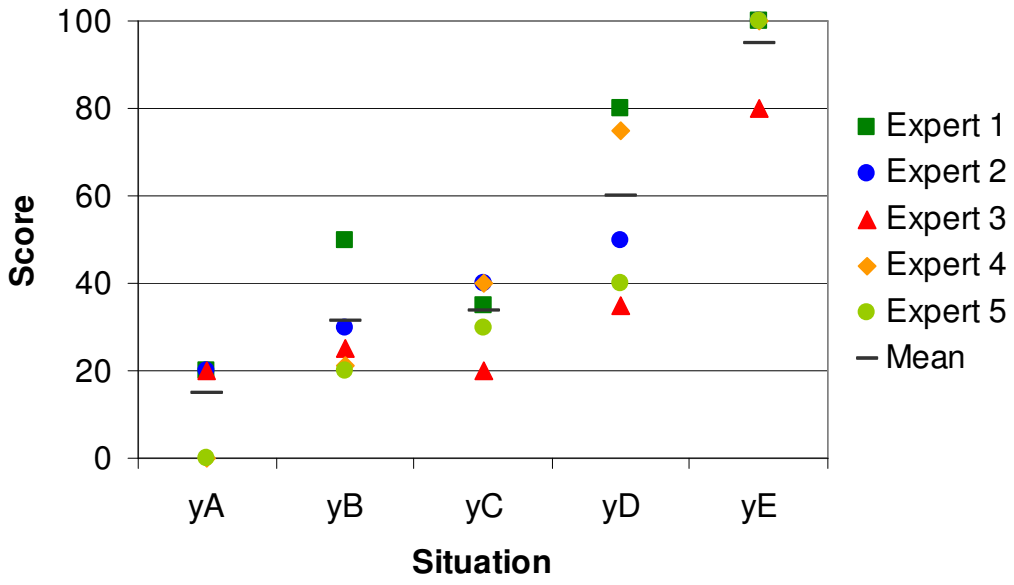
We showed the resulting distribution of farms to experts and we asked them to refine their first assessments (if they considered it was necessary). Their new answers are presented in Table 2 and on Figure 3.

Their second evaluations are slightly less severe.

Table 2: Answers from experts

Situation	Score					
	Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*
y_A	*20*	*20*	*20*	*0*	*0*	15
y_B	*50*	*30*	*25*	*21*	*20*	32
y_C	*35*	*40*	*20*	*40*	*30*	34
y_D	*80*	*50*	*35*	*75*	*40*	60
y_E	*100*	*100*	*80*	*100*	*100*	95

Figure 3: Graph representing experts' answers.



Calculation

Let x be a farm

Variables from raw data:
Let t be the type of housing system (L: loose / T. tied)
Let n_{OLA} be the number of days animals have access to an outdoor loafing area
Let d_{OLA} be the average number of hours per day animals have access to OLA
Let n_P be the number of days animals have access to pasture
Let d_P be the average number of hours per day animals have access to pasture

Calculated variables:
Let n_t be the number of days animals are kept *tethered*
Let n_e be the number of days tethered animals have *exercise*
Let n_o be the number of days tethered animals are kept *outside*
Let M_i be the answer to the ith question considered for farm x:
Question 1: Are animals tethered? 0. Yes, 1. No
Question 2: In case they are tethered, are they tethered during all the year (0) or only during winter (1)?
Question 3: In case they are tethered, do animals have regular exercise? 0. No, 1. Yes

Variables with predefined values:
Let d_e and d_o be the minimum duration of exercise (outdoor or indoor) for the cows to be considered respectively as *"tethered with exercise"* and *"outside"* (i.e. not tethered)
Let n_1 be the minimum number of "outside" days for the animals to be considered as not tethered
Let n_2 be the minimum number of "outside" days for the animals to be considered as tethered only during winter
Let p be the minimum ratio $\frac{n_e(x)}{n_t(x)}$ for the tethered animals be considered as having regular exercise
Let y_A to y_E be the scores attributed to the 5 possible values of $h(x)$

1. From raw data to data at farm level

$d_e = 1$
$d_o = 6$

$n_t(x) \leftarrow 0$
$n_e(x) \leftarrow 0$
$n_o(x) \leftarrow 0$

If ($d_e \leq d_{OLA}(x) < d_o$ AND $d_e \leq d_P(x) < d_o$) then $n_e(x) \leftarrow n_{OLA}(x) + n_P(x)$ and $n_o(x) \leftarrow 0$

Else, if ($d_e \leq d_{OLA}(x) < d_o$ AND $d_P(x) \geq d_o$) then $n_e(x) \leftarrow n_{OLA}(x)$ and $n_o(x) \leftarrow n_P(x)$

Else, if ($d_{OLA}(x) \geq d_o$ AND $d_e \leq d_P(x) < d_o$) then $n_e(x) \leftarrow n_P(x)$ and $n_o(x) \leftarrow n_{OLA}(x)$

Else, if ($d_{OLA}(x) \geq d_o$ AND $d_P(x) \geq d_o$) then $n_e(x) \leftarrow 0$ and $n_o(x) \leftarrow n_{OLA}(x) + n_P(x)$

$n_t(x) \leftarrow 365 - n_o(x)$

2. From data at farm level to subcriterion score

$y_A = 15$
$y_B = 32$
$y_C = 34$
$y_D = 60$
$y_E = 95$

$n_1 = 350$
$n_2 = 100$
$p = 0.5$

$M_1(x) \leftarrow 0$
$M_2(x) \leftarrow 0$
$M_3(x) \leftarrow 0$

$h(x) \leftarrow 0$
$g_5(x) \leftarrow 0$

If $t(x)$=L then $M_1(x)$=1
Else ($t(x)$=T)
    If $n_o(x) \geq n_1$ then $M_1(x)$=1
    Else, if $n_o(x) \geq n_2$ then $M_2(x)$=1
    Else $M_2(x)$=0
    If $\frac{n_e(x)}{n_t(x)} \geq p$ then $M_3(x)$=1
    Else $M_3(x)$=0

If $M_1(x)$=1 then $h(x) \leftarrow$ E and $g_5(x) \leftarrow y_E$
Else ($M_1(x)$=0)
    If $M_2(x)$=1
        If $M_3(x)$=1 then $h(x) \leftarrow$ D and $g_5(x) \leftarrow y_D$
        Else ($M_3(x)$=0) then $h(x) \leftarrow$ C and $g_5(x) \leftarrow y_C$
    Else ($M_2(x)$=0)
        If $M_3(x)$=1 then $h(x) \leftarrow$ B and $g_5(x) \leftarrow y_B$
        Else ($M_3(x)$=0) then $h(x) \leftarrow$ A and $g_5(x) \leftarrow y_A$

Subcriterion 6: Absence of injuries

Definition of the subcriterion

In this subcriterion, we aim at assessing external injuries due to physical causes and not to pathogens (which are included in Subcriterion 7). Voluntary interventions on animals (e.g. mutilations) are excluded and considered in a separate subcriterion (8). Lameness, whatever the cause, is also included in Subcriterion 6 as it is difficult to identify its causes and more likely due to a mechanical cause.

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11119	Number of mild alterations of the tegument due to injuries	Cardinal	Individual
11120	Number of severe alterations of the tegument due to injuries	Cardinal	Individual
11121	Lameness	Ordinal	Individual

Skin alterations: For each animal observed the number and severity of alterations of the tegument (present on the whole body of the animal and caused by injuries) are recorded.

Lameness: Lameness is assessed on a three level scale:
- 0. Not lame
- 1. Lame: imperfect temporal rhythm in stride creating a limp (irregular foot fall – not equal timing between all hoof-beats, weight not borne for equal time on each of the four feet)
- 2. Severely lame: extreme asymmetry in gait, or more than one limb affected

Analysis of the constraints

Skin alterations and lameness are observed at the individual level, but maybe not on the same animals, so it is impossible to consider lameness as an injury in order to produce scores at the individual level which could be aggregated at the farm level into the subcriterion score. So we will have first to calculate scores at the farm level for skin alterations and lameness and, in a second time, we will have to integrate these two scores into the subcriterion score.

In order to integrate the partial scores obtained for skin alterations and lameness, we may want to attribute different importance to lameness and skin alterations, while limiting compensations between these two elements.

The subcriterion score should qualify both ***the incidence and the severity of injuries****.*

Type of mathematical construction proposed

First stage of the aggregation:

The measures are observed at the individual level, so a first stage will be to aggregate the individual measures into an assessment at the herd level.

In table 1 and Table 2 the categories defined for lameness and skin alterations are shown.

Since we defined three levels at individual level for skin alterations and lameness, the first stage of aggregation follows the same calculation for the two measures.
The data we have to aggregate are the percentages of animals within each level of the scale used to assess lameness and skin alterations (3 levels for each measure):

Level	0	1	2
% of animals	$p_0^i(x)$	$p_1^i(x)$	$p_2^i(x)$

with i = 1 for lameness and 2 for skin alterations

$$\text{Lameness score} = g_6^1(x) = u_1\left(100 - \frac{\sum_{k=0}^{2} w_k^1 \; p_k^1(x)}{w_2^1}\right)$$

$$\text{Skin alterations score} = g_6^2(x) = u_2\left(100 - \frac{\sum_{k=0}^{2} w_k^2 \; p_k^2(x)}{w_2^2}\right)$$

Non-linear transformations (by the use of the functions u_1 and u_2) will probably be necessary.

Table 1: Categories defined for lameness.

Three categories are defined at the individual level:		Problem in respect to lameness
0	The cow is **not lame**	No problem
1	The cow is **lame:** imperfect temporal rhythm in stride creating a limp (irregular foot fall - not equal timing between hoof-beats, weight not borne for equal time on each of the four feet)	Moderate
2	The cow is **severly lame**: extreme asymmetry in gait, or more than one limb affected	Severe

Table 2: Categories defined for skin alterations.

Three categories are defined at the individual level:		Problem in respect to skin alterations
0	The cow does not have any injury (neither hair patch*, nor lesion/swelling**)	No problem
1	The cow has at least one mild alteration of the tegument (i.e. hair patch(es)), but no severe injury (i.e. no lesion or swelling)	Moderate
2	The cow has at least one severe injury (i.e. lesion(s) and/or swelling(s))	Severe

** Hair patch: area with air loss or extensive thinning of the coat as a response to parasites, skin not damaged, hyperkeratosis possible*

*** Lesion/Swelling: area with damaged skin either in form of a scab or a wound, dermatitis due to ectoparasites, (partly) missing teats and overt swelling*

Second stage of the aggregation:

In order to aggregate $g_6^1(x)$ (skin alterations score) and $g_6^2(x)$ (lameness score), we decided to use the Choquet integral:

$$\textit{Subcr. score} = g_6(x) = C\left(g_6^1(x), g_6^2(x)\right) = \begin{cases} g_6^1(x) + \left(g_6^2(x) - g_6^1(x)\right)v_2 & \textit{if} \quad g_6^2(x) \geq g_6^1(x) \\ g_6^2(x) + \left(g_6^1(x) - g_6^2(x)\right)v_1 & \textit{else} \end{cases}$$

Parameters to be defined

We have to define the weights w_k^1 and w_k^2, to find the most appropriate non-linear functions u_1 and u_2, and the weights v_i.

We also have to define the parameters for the Choquet integral that best fit the reasoning followed by the experts to aggregate Lameness score and Skin alterations score.

Adjustments of the mathematical construction

- **Questions asked to experts**

To parameter the construction for Lameness and Skin alterations:

We made experts react on virtual dataset (Table 3), asking them first to rank the virtual farms from the best one to the worst one, and then to give a score to each of the farms presented in the dataset.
The asked experts were U. Knierim, H.R. Whay, I. Veissier, J. Capdeville, C. Winckler.

The dataset was composed of virtual farms described by the number of measures within each category 'normal', 'moderate problem', and 'serious problem', respectively n_0, n_1 and n_2.

Table 3: Dataset submitted to experts. They had to sort the farms from best to worst, then to attribute a score to each farm for lameness or skin alterations.

	n_0	n_1	n_2	*Rank*	*Score*
Farm 1	12	0	0		
Farm 2	8	4	0		
Farm 3	8	0	4		
Farm 4	6	6	0		
Farm 5	6	0	6		
Farm 6	4	8	0		
Farm 7	4	0	8		
Farm 8	0	12	0		
Farm 9	0	8	4		
Farm 10	0	6	6		
Farm 11	0	4	8		
Farm 12	0	0	12		

The ranks given by experts will help us define the most appropriate weights $\mathbf{w_i}$.
The scores attributed by experts will help us find the most appropriate function *u*.

To parameter the aggregation of Lameness score and Skin alterations score:

We asked experts to react on a small dataset to determine how to merge the two sub-scores into the subcriterion-score.

Table 4: Dataset submitted to experts to aggregate Lameness score and Skin alterations score.

	Lamness score	Skin alterations score	*Score*
Farm 1	60	40	
Farm 2	50	50	
Farm 3	40	60	

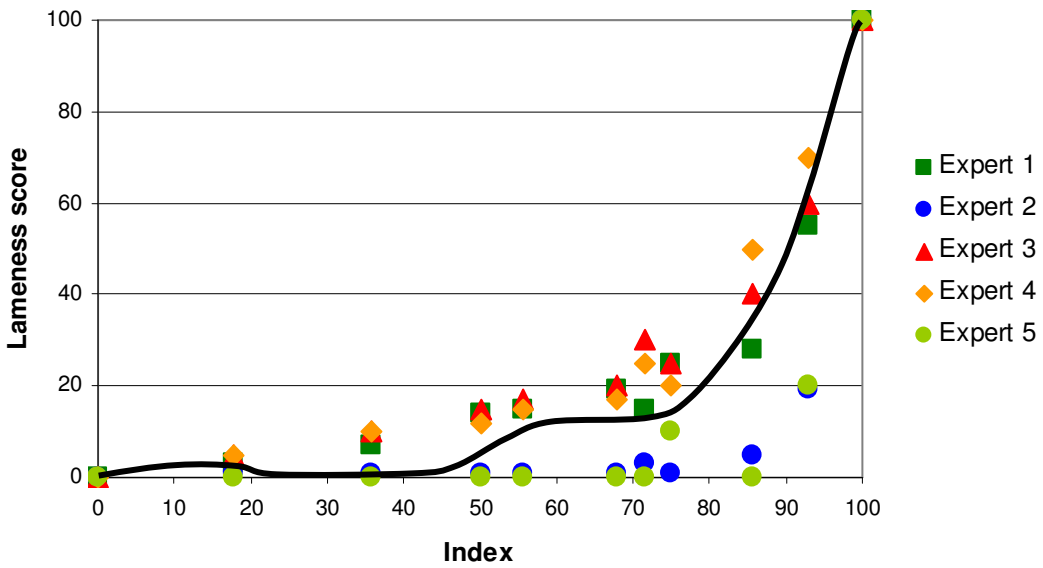
- **Adjustment of the construction according to answers from experts**

Lameness:

Table 5a: Answers from experts for lameness.

	Situation				*Score*				
	p0	p1	p2	Σ					
Weights =	*0*	*2*	*7*		*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*
Farm 3	0	0	100	0	*0*	*0*	*0*	*0*	*0*
Farm 9	0	25	75	18	*3*	*1*	*5*	*5*	*0*
Farm 5	0	50	50	36	*7*	*1*	*10*	*10*	*0*
Farm 6	50	0	50	50	*14*	*1*	*15*	*12*	*0*
Farm 8	20	50	30	56	*15*	*1*	*17*	*15*	*0*
Farm 7	50	25	25	68	*19*	*1*	*20*	*17*	*0*
Farm 2	0	100	0	71	*15*	*3*	*30*	*25*	*0*
Farm 11	75	0	25	75	*25*	*1*	*25*	*20*	*10*
Farm 4	50	50	0	86	*28*	*5*	*40*	*50*	*0*
Farm 10	75	25	0	93	*55*	*19*	*60*	*70*	*20*
Farm 1	100	0	0	100	*100*	*100*	*100*	*100*	*100*

Figure 1a: B-spline functions allowing attributing a subcriterion-score to lameness.

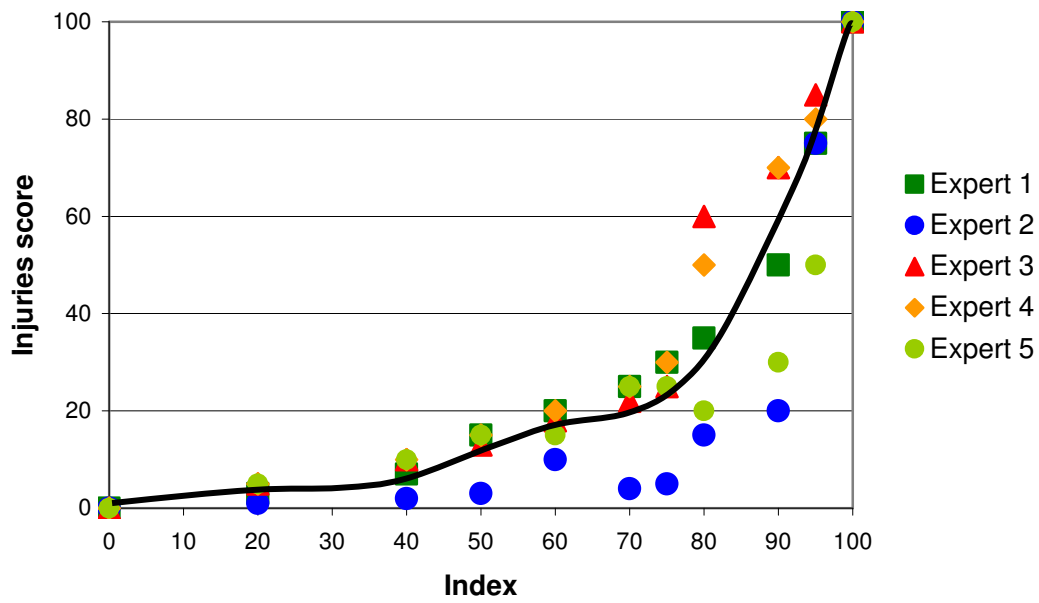


Skin alterations:

Table 5b: Answers from experts for skin alterations.

	Situation				Score				
	p0	**p1**	**p2**	**Index**	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*
Weights =	*0*	*1*	*5*						
Farm 3	0	0	100	0	*0*	*0*	*0*	*0*	*0*
Farm 9	0	25	75	20	*3*	*1*	*5*	*5*	*5*
Farm 5	0	50	50	40	*7*	*2*	*10*	*10*	*10*
Farm 6	50	0	50	50	*15*	*3*	*13*	*15*	*15*
Farm 8	20	50	30	60	*20*	*10*	*18*	*20*	*15*
Farm 7	50	25	25	70	*25*	*4*	*22*	*25*	*25*
Farm 11	75	0	25	75	*30*	*5*	*25*	*30*	*25*
Farm 2	0	100	0	80	*35*	*15*	*60*	*50*	*20*
Farm 4	50	50	0	90	*50*	*20*	*70*	*70*	*30*
Farm 10	75	25	0	95	*75*	*75*	*85*	*80*	*50*
Farm 1	100	0	0	100	*100*	*100*	*100*	*100*	*100*

Figure 1b: B-spline functions allowing attributing a subcriterion-score to skin alterations.



Aggregation of Lameness and Skin alterations scores:

We asked four experts to react on the following small dataset (Table 6). We then used their answers to define the two capacity values v_1 and v_2 that minimises the square deviation between experts' answers and the calculated score. We obtained $\mathbf{v_1=0.11}$ and $\mathbf{v_2=0.31}$.

Table 6: Aggregation of the lameness and skin alterations subscores

	Lamness score	Skin alterations score	*Score*				*Mean*	Calculated score
			Expert 1	*Expert 2*	*Expert 3*	*Expert 4*		
Farm 1	60	40	*50*	*50*	*50*	*55*	*51*	**51**
Farm 2	50	50	*50*	*50*	*50*	*50*	*50*	**50**
Farm 3	40	60	*40*	*50*	*50*	*45*	*46*	**46**

- **Test on dataset(s)**

We looked at 64 farms visited during WP2.4 to see the distribution of these farms on SC3. We obtained the results presented, in a cumulative way, in the fourth column of the Tables 6a (for *lameness*) and 6b (for *skin alterations*).

We then asked another time experts to react on the reorganised dataset (cf. Tables 6a and 6b), including the information on the distribution of farms on SC6.

Lameness:

Table 7a: Second dataset submitted to experts, for *lameness*. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	p0	p1	p2	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 3	0	0	100	0%		
Farm 9	0	25	75	0%		
Farm 5	0	50	50	3%		
Farm 6	50	0	50	6%		
Farm 8	20	50	30	9%		
Farm 7	50	25	25	17%		
Farm 2	0	100	0	23%		
Farm 11	75	0	25	28%		
Farm 4	50	50	0	53%		
Farm 10	75	25	0	83%		
Farm 1	100	0	0	100%		

Skin alterations:

Table 7b: Second dataset submitted to experts, for *skin alterations*. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	p0	p1	p2	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 3	0	0	100	6%		
Farm 9	0	25	75	28%		
Farm 5	0	50	50	41%		
Farm 6	50	0	50	50%		
Farm 8	20	50	30	58%		
Farm 7	50	25	25	67%		
Farm 11	75	0	25	73%		
Farm 2	0	100	0	84%		
Farm 4	50	50	0	98%		
Farm 10	75	25	0	100%		
Farm 1	100	0	0	100%		

- **Final adjustment according to test on dataset**

The new answers of experts are presented in Tables 7a and 7b.

Lameness:

Table 8a: Answers from experts for *lameness*

	Situation			Score						*Rank*	Index	**Calculated score**
	p0	p1	p2	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*			
Farm 3	0	0	100	*0*	*0*	*0*	*0*	*0*	0	*11*	0	**0**
Farm 9	0	25	75	*2*	*0*	*0*	*5*	*2*	2	*10*	18	**2**
Farm 5	0	50	50	*5*	*0*	*5*	*10*	*4*	5	*9*	36	**5**
Farm 6	50	0	50	*13*	*0*	*10*	*15*	*10*	10	*8*	50	**9**
Farm 8	20	50	30	*15*	*0*	*15*	*17*	*10*	11	*7*	56	**12**
Farm 7	50	25	25	*17*	*10*	*20*	*22*	*20*	18	*6*	68	**19**
Farm 2	0	100	0	*16*	*15*	*35*	*25*	*20*	22	*5*	71	**22**
Farm 11	75	0	25	*21*	*10*	*40*	*30*	*25*	25	*4*	75	**25**
Farm 4	50	50	0	*22*	*17*	*60*	*50*	*40*	38	*3*	86	**37**
Farm 10	75	25	0	*65*	*20*	*80*	*70*	*50*	57	*2*	93	**58**
Farm 1	100	0	0	*100*	*100*	*100*	*100*	*100*	100	*1*	100	**100**

The weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ remain the same as previously:

$$\mathbf{w}_0^1 = 0$$

$$\mathbf{w}_1^1 = 2$$

$$\mathbf{w}_2^1 = 7$$

To determine the most appropriate function $\boldsymbol{u_1}$, we applied another technique using I-splines instead of B-splines, and with only one interior knot, here placed at 78. The function $\boldsymbol{u_1}$ is composed of two pieces of cubic curves, and writes as follows:

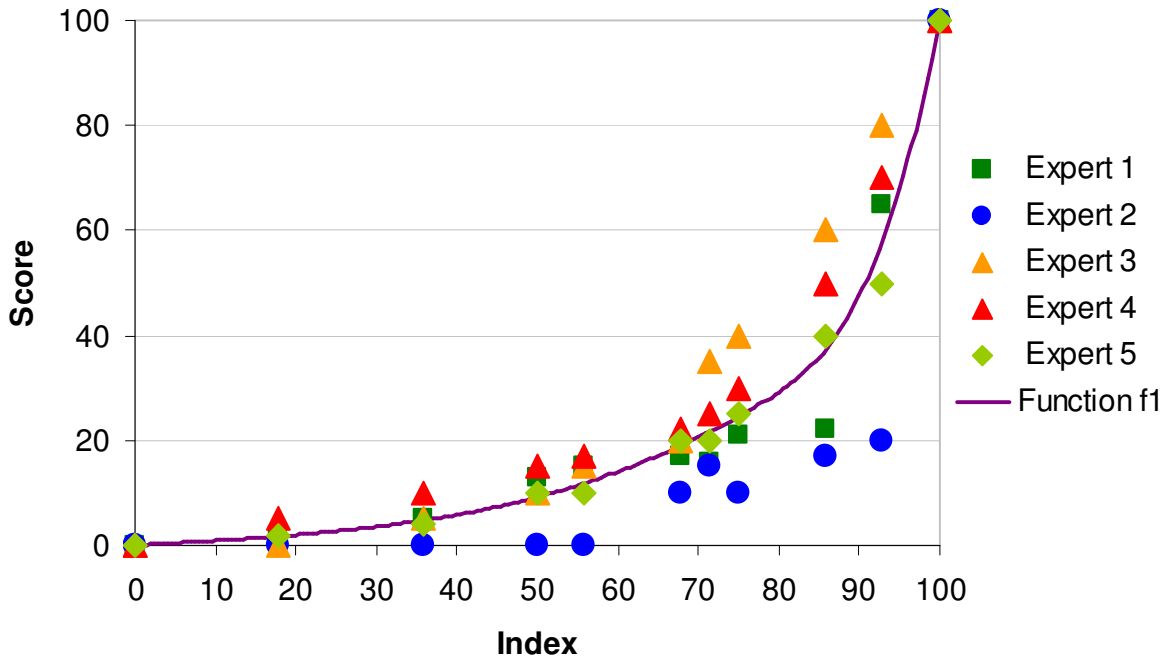
$$u_1(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 6.

Table 9a: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,78]$	$x \in [78,100]$
a_x	0	-2060.598372
b_x	0.098799325	79.35251952
c_x	-0.000954972	-1.017025859
d_x	5.34439E-05	0.004395605

Figure 2a: I-spline functions (with one interior knot at 78) allowing attributing a subcriterion-score to 'Comfort around resting', for *lameness*



Skin alterations:

Table 8b: Answers from experts for *skin alterations*

	Situation			Score						Rank	Index	Calculated score
	p0	p1	p2	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*			
Farm 3	0	0	100	*0*	*0*	*0*	*0*	*0*	0	*11*	0	**0**
Farm 9	0	25	75	*5*	*0*	*10*	*15*	*5*	7	*10*	20	**7**
Farm 5	0	50	50	*10*	*20*	*15*	*22*	*10*	15	*9*	40	**15**
Farm 6	50	0	50	*15*	*20*	*20*	*30*	*20*	21	*8*	50	**21**
Farm 8	20	50	30	*20*	*25*	*40*	*40*	*20*	29	*7*	60	**30**
Farm 7	50	25	25	*30*	*50*	*45*	*50*	*30*	41	*6*	70	**42**
Farm 11	75	0	25	*40*	*50*	*50*	*60*	*35*	47	*5*	75	**49**
Farm 2	0	100	0	*45*	*70*	*75*	*70*	*50*	62	*4*	80	**57**
Farm 4	50	50	0	*70*	*85*	*80*	*80*	*60*	75	*3*	90	**77**
Farm 10	75	25	0	*80*	*90*	*90*	*90*	*80*	86	*2*	95	**88**
Farm 1	100	0	0	*100*	*100*	*100*	*100*	*100*	100	*1*	100	**100**

The weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ remain the same as previously:

$$\mathbf{w}_0^2 = 0$$

$$\mathbf{w}_1^2 = 1$$

$$\mathbf{w}_2^2 = 5$$

To determine the most appropriate function $\boldsymbol{u_2}$, we applied another technique using I-splines instead of B-splines, and with only one interior knot, here placed at 65. The function $\boldsymbol{u_2}$ is composed of two pieces of cubic curves, and writes as follows:

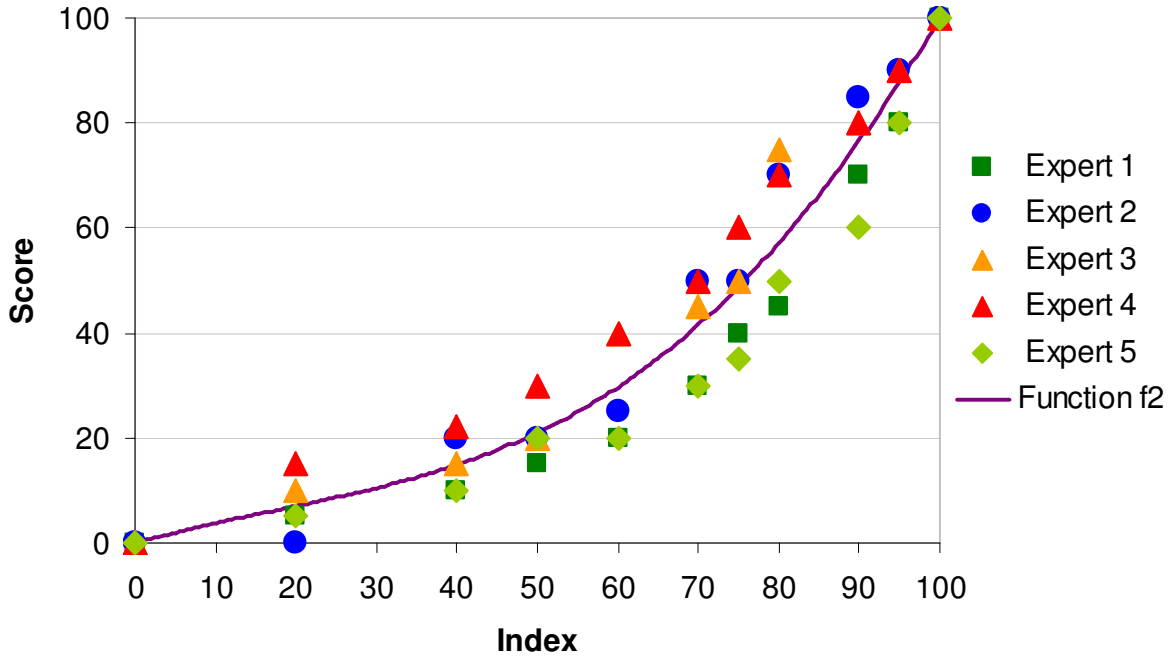
$$u_2(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 6.

Table 9b: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,65]$	$x \in [65,100]$
a_x	0	29.8965855
b_x	0.43167869	-0.94444993
c_x	-0.00650436	0.01453
d_x	0.00012589	1.9248E-05

Figure 2b: I-spline functions (with one interior knot at 65) allowing attributing a subcriterion-score to 'Comfort around resting', for *skin alterations*



Calculation

Let *x* be a farm

Variables from raw data:
Let N_1 be the number of animals observed for *lameness*
Let N_2 be the number of animals observed for skin *alterations*
Let M_{ik} be the data obtained by the kth animal observed on the farm *x* for measure *i* (1. severity of lameness, 2. number of hair patches, 3. number of lesions/swellings)

Calculated variables:
Let n_0^i, n_1^i and n_2^i be the number of animals assigned respectively to the categories 0, 1 and 2 for *lameness* if *i*=1 or for *skin alterations* if *i*=2
Let p_0^i, p_1^i and p_2^i be the % of animals assigned respectively to the categories 0, 1 and 2 for *lameness* if *i*=1 or for *skin alterations* if *i*=2
Let h^i be the index obtained for *lameness* if *i*=1 or for *skin alterations* if *i*=2
Let g_6^i be the subscore obtained for *lameness* if *i*=1 or for *skin alterations* if *i*=2
Let g_6 be the subcriterion-score for *"Absence of injuries"*

Variables with predefined values:
Let α and β be the thresholds respectively for hair patches and lesions/swellings used to delimit the three categories of problems defined for *alterations of the tegument*
Let w_0^i, w_1^i and w_2^i be the weights assigned to the three categories of problems, with (*i*=1 for *lameness* and *i*=2 for *skin alterations*)
Let a_1^i, b_1^i, c_1^i, d_1^i, a_2^i, b_2^i, c_2^i and d_2^i be the parameters of the two pieces of the utility function u_i
Let K^i be the position of the knot, i.e. the limit between the two pieces of the utility function u_i
Let v_i be the capacity value associated to the subscore $g_6^i(x)$

1. From raw data to data at farm level

$\alpha = 1$
$\beta = 1$

$n_0^1(x) \leftarrow 0$
$n_1^1(x) \leftarrow 0$
$n_2^1(x) \leftarrow 0$

$n_0^2(x) \leftarrow 0$
$n_1^2(x) \leftarrow 0$
$n_2^2(x) \leftarrow 0$

$p_0^1(x) \leftarrow 0$
$p_1^1(x) \leftarrow 0$
$p_2^1(x) \leftarrow 0$

$p_0^2(x) \leftarrow 0$

$p_1^2(x) \leftarrow 0$

$p_2^2(x) \leftarrow 0$

For k = 1, …, $N_1(x)$

If (M_{1k}(x)=0) then $n_0^1(x) \leftarrow n_0^1(x)+1$

Else, if (M_{1k}(x)=1) then $n_1^1(x) \leftarrow n_1^1(x)+1$

Else $n_2^1(x) \leftarrow n_2^1(x)+1$

End

$$p_0^1(x) \leftarrow \frac{n_0^1(x)}{N_1(\mathrm{x})}$$

$$p_1^1(x) \leftarrow \frac{n_1^1(x)}{N_1(\mathrm{x})}$$

$$p_2^1(x) \leftarrow \frac{n_2^1(x)}{N_1(\mathrm{x})}$$

For k = 1, …, $N_2(x)$

If (M_{2k}(x) < α and M_{3k}(x) < β) then $n_0^2(x) \leftarrow n_0^2(x)+1$

Else, if (M_{2k}(x) ≥ α and M_{3k}(x) < β) then $n_1^2(x) \leftarrow n_1^2(x)+1$

Else $n_2^2(x) \leftarrow n_2^2(x)+1$

End

$$p_0^2(x) \leftarrow \frac{n_0^2(x)}{N_2(\mathrm{x})}$$

$$p_1^2(x) \leftarrow \frac{n_1^2(x)}{N_2(\mathrm{x})}$$

$$p_2^2(x) \leftarrow \frac{n_2^2(x)}{N_2(\mathrm{x})}$$

2. From data at farm level to subcriterion score

$w_0^1 = 0$

$w_1^1 = 2$

$w_2^1 = 7$

$a_1^1 = 0$	$a_2^1 = -2060.598372$
$b_1^1 = 0.098799325$	$b_2^1 = 79.35251952$
$c_1^1 = -0.000954972$	$c_2^1 = -1.017025859$
$d_1^1 = 5.34439E-05$	$d_2^1 = 0.004395605$

$K^1 = 65$

$w_0^2 = 0$
$w_1^2 = 1$
$w_2^2 = 5$

$a_1^2 = 0$	$a_2^2 = 29.8965855$
$b_1^2 = 0.43167869$	$b_2^2 = -0.94444993$
$c_1^2 = -0.00650436$	$c_2^2 = 0.01453$
$d_1^2 = 0.00012589$	$d_2^2 = 1.9248E-05$

$K^2 = 65$

$v_1 =$
$v_2 =$

$h^1(x) \leftarrow 0$
$h^2(x) \leftarrow 0$
$g_6^1(x) \leftarrow 0$
$g_6^2(x) \leftarrow 0$
$g_6(x) \leftarrow 0$

For i = 1 to 2 do

$$h^i(x) = 100 - \frac{\sum_{k=0}^{2} w_k^i \; p_k^i(x)}{w_2^i}$$

If $h^i(x) < K^i$ then $g_6^i(x) \leftarrow a_1^i + b_1^i \; h^i(x) + c_1^i \left(h^i(x)\right)^2 + d_1^i \left(h^i(x)\right)^3$

Else $g_6^i(x) \leftarrow a_2^i + b_2^i \; h^i(x) + c_2^i \left(h^i(x)\right)^2 + d_2^i \left(h^i(x)\right)^3$

End

If $g_6^1(x) \le g_6^2(x)$ then $g_6(x) = g_6^1(x) + v_2\left(g_6^2(x) - g_6^1(x)\right)$

Else $g_6(x) = g_6^2(x) + v_1\left(g_6^1(x) - g_6^2(x)\right)$

Subcriterion 7: Absence of diseases

Definition of the subcriterion

This subcriterion aims at assessing the health of the animals through the occurrence of diseases (either metabolic diseases or diseases produced by pathogens (virus, microbes, parasites…).

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11122	Coughing	Cardinal	Herd
11123	Nasal discharge	Binary	Individual
11124	Ocular discharge	Binary	Individual
11125	Increased respiratory rate	Binary	Individual
11126	Diarrhoea	Binary	Individual
11127	Vulva discharge	Binary	Individual
11128	Mastitis	Cardinal	Herd
11129	Dystocia incidence	Cardinal	Herd
11130	Downer cows	Cardinal	Herd
11131	Mortality: animals culled due to diseases or accidents	Cardinal	Herd

As for injuries, the measure “Health status” is composed of several evaluations (as many evaluations as categories of diseases). Incidence of diseases is assessed in two different ways, using very different time scales: evaluation on the animals during the visit (e.g. for nasal discharge), and evaluation through farm records (at the year scale) (e.g. for mortality).

Analysis of the constraints

Some diseases may be more painful for animals than others; however, it is extremely difficult to compare the degree of pain of two diseases that lead to different symptoms, on a different time scale, etc.

We have to deal with different time scales.

Since we have to aggregate quite an important number of measures (10), it is possible not to use all the precision given by raw data.

*The subcriterion score should qualify **the incidence of diseases**.*

Type of mathematical construction proposed

We propose to use a method similar to the one proposed for the Subcriterion 3 "Comfort around resting", i.e. to compare each measure to two different warning thresholds in order to qualify three categories (cf. Table 1): **normal** situation, **serious** problem *(i.e. a remedial plan at farm level MUST be taken above this threshold)*, and an intermediate situation corresponding to a **moderate** problem.

In addition we decided to consider together coughing and increased respiratory rate since they are both syndromes of the respiratory sphere. For these two measures, the worst one will determine the warning level of the herd.
Similarly, we also decided to group nasal and ocular discharges because they are often due to similar causes.

So, in total we have to deal with **8 evaluations (N = 8)**, each of them being categorised in one of the three categories 'normal', 'moderate problem', and 'serious problem'. In order to delimit these three categories, we will need set two thresholds for each measures: $T_{i,1}$ and $T_{i,2}$ for Measure *i* (cf. Table 1).

Table 1: Organisation and names of the thresholds associated to the measures.

Measure	Data measured	Normal	Moderate problem	Serious problem
Nasal discharge	M_1	$< T_{1,1}$	$T_{1,1} \leq \ldots < T_{1,2}$	$\geq T_{1,2}$
Ocular discharge	M_2	$< T_{2,1}$	$T_{2,1} \leq \ldots < T_{2,2}$	$\geq T_{2,2}$
Coughing	M_3	$< T_{3,1}$	$T_{3,1} \leq \ldots < T_{3,2}$	$\geq T_{3,2}$
Increased Respiratory Rate	M_4	$< T_{4,1}$	$T_{4,1} \leq \ldots < T_{4,2}$	$\geq T_{4,2}$
Diarrhoea	M_5	$< T_{5,1}$	$T_{5,1} \leq \ldots < T_{5,2}$	$\geq T_{5,2}$
Mastitis	M_6	$< T_{6,1}$	$T_{6,1} \leq \ldots < T_{6,2}$	$\geq T_{6,2}$
Vulva discharge	M_7	$< T_{7,1}$	$T_{7,1} \leq \ldots < T_{7,2}$	$\geq T_{7,2}$
Dystocia incidence	M_8	$< T_{8,1}$	$T_{8,1} \leq \ldots < T_{8,2}$	$\geq T_{8,2}$
Downer cows	M_9	$< T_{9,1}$	$T_{9,1} \leq \ldots < T_{9,2}$	$\geq T_{9,2}$
Mortality: animals culled due to diseases or accidents	M_{10}	$< T_{10,1}$	$T_{10,1} \leq \ldots < T_{10,2}$	$\geq T_{10,2}$

NB: Differences of importance between measures are taken into account implicitly through the definition of the thresholds (lower thresholds for more important/severe syndromes).

Once each measure categorised, we can calculate the number of measures in each of the three categories:

- N_0 = number of measures in the category "Normal"
- N_1 = number of measures in the category "Moderate problem"
- N_2 = number of measures in the category "Serious problem"

Then, we will calculate the subcriterion score with a calculation based on the weighted sum of the number of measures in each category, attributing more weight to more serious problem (and no weight for the category normal): $w_0=0 < w_1 < w_2$.

The use of a non-linear function *u* to transform the linear combination in order to be in accordance with experts' opinion will surely be necessary.

$$g_7(x) = u\left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^{2} w_k N_k(x)}{w_2}\right)\right)$$

Parameters to be defined

We will have to define two elements:

- The thresholds $T_{i,1}$ and $T_{j,2}$
- the weights w_j,
- the utility function *u*.

Adjustments of the mathematical construction

▪ Questions asked to experts

Thresholds: direct questions to experts.

Weights and utility function *u*:

We made experts react on virtual dataset (Table 2), asking them first to rank the virtual farms from the best one to the worst one, and then to give a score to each of the farms presented in the dataset.
The asked experts were I. Veissier, J. Capdeville, H.R. Whay, U. Knierim, E. Canali, C. Winckler.

The ranks will allow us to determine the most appropriate weights, and the scores will allow us to define the utility function.

Table 2: Dataset submitted to experts. They had to sort the farm from best to worst, then to attribute subcriterion scores.

	Normal	Moderate problem	Serious problem	*Rank*	*Score*
Farm 1	8	0	0		
Farm 2	6	2	0		
Farm 3	6	0	2		
Farm 4	4	2	2		
Farm 5	4	4	0		
Farm 6	4	0	4		
Farm 7	2	6	0		
Farm 8	2	3	3		
Farm 9	2	0	6		
Farm 10	0	4	4		
Farm 11	0	6	2		
Farm 12	0	2	6		
Farm 13	0	8	0		
Farm 14	0	0	8		

▪ Adjustment of the construction according to answers from experts

Definition of the thresholds:
We asked three vets and researchers to define the thresholds. They were able to define the threshold corresponding to the lower bound of a serious problem ($T_{i,2}$). They agreed on these thresholds and we proposed to set each threshold $T_{i,1}$ as a function of $T_{i,2}$:

$$T_{i,1} = \frac{T_{i,2}}{2}$$

This proposal has been accepted by the consulted persons.
The results of this consultation process are presented in Table 3.

Table 3: Thresholds defined by experts for the different measures.

Measure		Way used to collect the information*	Description of the measure	$T_{i,1}$	$T_{i,2}$
Nasal discharge	M	Visit	% of animals with nasal discharge	5%	10%
Ocular discharge	M_2	Visit	% of animals with ocular discharge	3%	6%
Coughing	M_3	Visit	nb of coughs (during 15 min) / nb of animals in the herd	4%	8%
Increased respiratory rate	M_4	Visit	% of animals with an increased respiratory rate	5%	10%
Diarrhoea	M_5	Records	% of animals with diarrhoea	3%	6%
Mastitis	M_6	Records	nb of cases of mastitis during a year / nb of animals in the herd	10%	20%
Vulvar discharge	M_7	Visit	% of animals with vulvar discharge	3%	6%
Dystocia	M_8	Records	nb of cases of dystocia during a year / nb of animals in the herd	4%	8%
Downer cows	M_9	Records	nb of downer cows during a year / nb of animals in the herd	3%	6%
Mortality	M_{10}	Records	nb of dead animals * during a year / nb of animals in the herd	2%	4%

Blue: direct observation
Green: farm records

Weights and utility function:
The results of experts' evaluation of farms with various numbers of measures within each category (Normal / Moderate problem / Serious problem) are shown in Table 4 (Columns 5-9). We calculated the mean of subcriterion-scores attributed by the 6 experts; then the farms were sorted according to the mean subcriterion-score they obtained (Table 5, Columns 10-11).

It was not possible to find the weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ by linear optimisation so as to match the ranking of farms according to mean score attributed to experts. This could be due to inconsistencies *between* and/or *within* experts. Looking individually at experts' answers, it was possible to find weights for only three of them (among the five consulted), as presented in Table 4. Hence we can conclude that some inconsistencies also exist within experts for two of them, in addition to inconsistencies between experts, as visible in Table 5.

Table 4: The most appropriate weights obtained by linear optimisation on each expert.

Expert	w_0	w_1	w_2
I. Veissier	0	1	5
J. Capdeville	-	-	-
B. Whay	0	1	5
U. Knierim	-	-	-
E. Canali	0	4	7

However, it has been possible to find weights that minimise the quadratic error due to inversions in the ranking of farms:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 1$
$\mathbf{w_2} = 3$

The remaining quadratic error is nil.
The result of the index resulting from the weighted sum obtained by each farm is shown in Table 5, column 12.

To define the most appropriate B-spline function we used a software package, developed by a student at University Paris 6, that also included the definition of weights. Since the first stage of the calculation resulted in a error (impossible to find weights matching the ranking of farms given by experts), it was not possible to run the second stage: the definition of a B-spline function that minimise the sum of square errors between experts' scores and calculated ones).
As a consequence, we tried to find a polynomial function which minimised the sum of quadratic errors with experts' scores. By linear optimisation we found the following polynomial function of degree 3:

$$u(x) = a\,x^3 + b\,x^2 + c\,x + d$$

with a = 0.0001912
b = 0.0178798
c = 0.8761845
d = 0

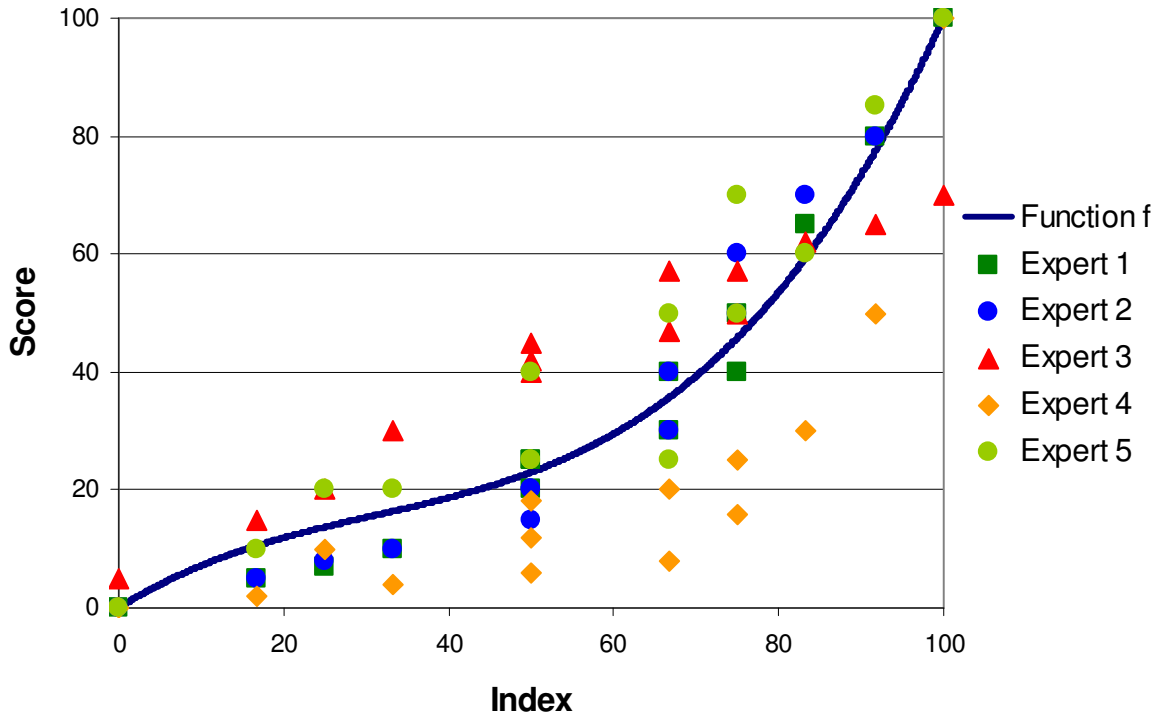
We tested polynomial function of degree 2, 3, 4,5 and 6. We choose a polynomial function of degree 3 since it leads to the best R^2.

This simple polynomial function *u* leads to a quadratic error of 9989 for 5 experts.

Table 5: Results obtained from the experts asked.

	Situation			Score						Rank	Index
	p0	p1	p2	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Mean		
Farm 14	0	0	8	0	0	5	0	0	1	14	0
Farm 12	0	2	6	5	5	15	2	10	7	13	17
Farm 9	2	0	6	7	8	20	10	20	13	12	25
Farm 10	0	4	4	10	10	30	4	20	15	11	33
Farm 8	2	3	3	25	15	42	12	25	24	10	50
Farm 11	0	6	2	25	20	45	6	25	24	9	50
Farm 6	4	0	4	20	20	40	18	40	28	8	50
Farm 13	0	8	0	40	30	57	8	25	32	7	67
Farm 4	4	2	2	30	40	47	20	50	37	6	67
Farm 7	2	6	0	50	60	57	16	50	47	5	75
Farm 3	6	0	2	40	50	50	25	70	47	4	75
Farm 5	4	4	0	65	70	62	30	60	57	3	83
Farm 2	6	2	0	80	80	65	50	85	72	2	92
Farm 1	8	0	0	100	100	70	100	100	94	1	100

Figure 1: Polynomial function of degree 5 to attribute a subcriterion-score to 'Absence of diseases'



- **Test on dataset(s)**

We looked at 64 farms visited during WP2.4 to see the distribution of these farms on SC7. We obtained the results presented, in a cumulative way, in the fifth column of Table 6.

However, we only could calculate estimations of the scores since all the measures taken from farm records (*mastitis*, *dystocia*, *downer cows*, and *mortality*) were not recorded during WP2.4 visits.

Table 6: Second dataset submitted to experts. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	Normal	Moderate problem	Serious problem	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 14	0	0	8	5%		
Farm 12	0	2	6	9%		
Farm 9	2	0	6	14%		
Farm 10	0	4	4	14%		
Farm 6	4	0	4	56%		
Farm 8	2	3	3	56%		
Farm 11	0	6	2	56%		
Farm 4	4	2	2	70%		
Farm 3	6	0	2	91%		
Farm 13	0	8	0	91%		
Farm 7	2	6	0	91%		
Farm 5	4	4	0	95%		
Farm 2	6	2	0	95%		
Farm 1	8	0	0	100%		

- **Final adjustment according to test on dataset**

We didn't judge appropriate to consult again experts since the information we had from WP2.4 data revealed that some information may be impossible to obtain (that concerns all the measures to be obtained from farm records). In addition, a new measure seems to be useful and feasible: mean number of sneezes per animal. As a consequence, some discussion is still needed on the measures to be used to assess the absence of diseases. This will surely be done within WP2.5. Since for the moment we are sure the list of measures will be revised, we preferred wait Subtask 2.3.1.4 to consult again the experts.

However, we decided to revise the definition of the function *u*, as presented below.

The weights $\mathbf{w_0}$, $\mathbf{w_1}$ and $\mathbf{w_2}$ remain the same as previously:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 1$
$\mathbf{w_2} = 3$

To determine the most appropriate function ***u***, we applied another technique using I-splines instead of B-splines, and with only one interior knot, here placed at 65. The function ***u*** is composed of two pieces of cubic curves, and writes as follows:

$$u(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

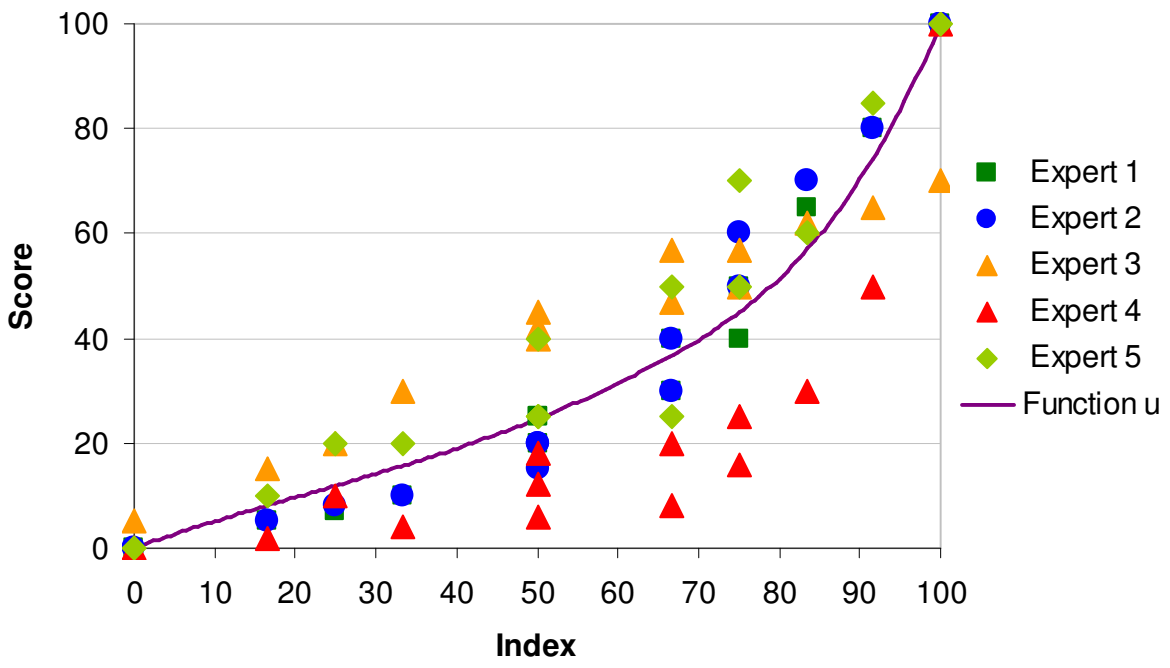
with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 8.

Table 8: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,65]$	$x \in [65,100]$
a_x	0	-150.976977
b_x	0.55017	7.49967224
c_x	-0.00478157	-0.11100951
d_x	7.2509E-05	0.0006111

Using I-spline function instead of the simple polynomial function previously defined, we reduced the quadratic error from 9989 to 9368 for 5 experts.

Figure 2: I-spline function (with one interior knot at 65) to attribute a subcriterion-score to 'Absence of diseases'



Calculation

Let x be a farm

Variables from raw data:
Let n_i be the number of animals observed for the measure i ($i \in \{1,2,4,5,7\}$)
Let M_{ik} be the data obtained by the kth animal observed on the farm x for measure i ($i \in \{1,2,4,5,7\}$)
Let M_3 be the $\frac{\text{nb of coughs (during 15 min)}}{\text{nb of animals in the herd}} \times 100$
Let M_6 be the $\frac{\text{nb of cases of mastitis during a year}}{\text{nb of animals in the herd}} \times 100$
Let M_8 be the $\frac{\text{nb of cases of dystocia during a year}}{\text{nb of animals in the herd}} \times 100$
Let M_9 be the $\frac{\text{nb of downer cows during a year}}{\text{nb of animals in the herd}} \times 100$
Let M_{10} be the $\frac{\text{nb of dead animals during a year}}{\text{nb of animals in the herd}} \times 100$

Calculated variables:
Let d_1 be the number of cows with nasal discharge
Let d_2 be the number of cows with ocular discharge
Let d_4 be the number of cows with increased respiratory rate
Let d_5 be the number of cows with diarrhoea
Let d_7 be the number of cows with vulvar discharge
Let M_1 be the % of cows with nasal discharge
Let M_2 be the % of cows with ocular discharge
Let M_4 be the % of cows with increased respiratory rate
Let M_5 be the % of cows with diarrhoea
Let M_7 be the % of cows with vulvar discharge
Let c_i be the category ($i \in \{1, \ldots,10\}$) to which the measure i is assigned ($c_i \in \{0,1,2\}$)
Let N_j be the number of measures in the jth category ($j \in \{0,1,2\}$)

Variables with predefined values:
Let $T_{1,1}$ to $T_{10,2}$ be the thresholds used to delimit the three categories of problems in respect to welfare
Let N_{tot} be total number of evaluations to be taken into account
Let w_j be the weights associated to the category j
Let a_1, b_1, c_1, d_1, a_2, b_2, c_2 and d_2 be the parameters of the two pieces of the utility function
Let K be the position of the knot (i.e. limit between the two pieces of the utility function)

1. From raw data to data at farm level

$d_1(x) \leftarrow 0$
$d_2(x) \leftarrow 0$
$d_4(x) \leftarrow 0$
$d_5(x) \leftarrow 0$
$d_7(x) \leftarrow 0$
$M_1(x) \leftarrow 0$
$M_2(x) \leftarrow 0$
$M_4(x) \leftarrow 0$
$M_5(x) \leftarrow 0$
$M_7(x) \leftarrow 0$

For i = 1, 2, 4, 5 and 7
    $d_i(x) \leftarrow 0$
    For k = 1, …, $n_i(x)$
        If $M_{ik}(x) = 1$ then $d_i(x) \leftarrow d_i(x) + 1$
        Else $d_i(x) \leftarrow d_i(x)$
    End
    $M_i(x) \leftarrow \frac{d_i(x)}{n_i(x)} \times 100$
End

2. From data at farm level to subcriterion score

$T_{1,1} = 5$ $T_{1,2} = 10$
$T_{2,1} = 3$ $T_{2,2} = 6$
$T_{3,1} = 4$ $T_{3,2} = 8$
$T_{4,1} = 5$ $T_{4,2} = 10$
$T_{5,1} = 3$ $T_{5,2} = 6$
$T_{6,1} = 10$ $T_{6,2} = 20$
$T_{7,1} = 3$ $T_{7,2} = 6$
$T_{8,1} = 4$ $T_{8,2} = 8$
$T_{9,1} = 3$ $T_{9,2} = 6$
$T_{10,1} = 2$ $T_{10,2} = 4$

$N_{tot} = 8$

$w_0 = 0$
$w_1 = 1$
$w_2 = 3$

$a_1 = 0$
$b_1 = 0.55017$
$c_1 = -0.00478157$
$d_1 = 7.2509E-05$

$a_2 = -150.976977$
$b_2 = 7.49967224$
$c_2 = -0.11100951$
$d_2 = 0.0006111$

K = 65

$c_0(x) \leftarrow 0$
$c_1(x) \leftarrow 0$
$c_2(x) \leftarrow 0$

$N_0(x) \leftarrow 0$
$N_1(x) \leftarrow 0$
$N_2(x) \leftarrow 0$

$h(x) \leftarrow 0$
$g_7(x) \leftarrow 0$

If ($M_1(x) \geq T_{1,2}$ or $M_2(x) \geq T_{2,2}$) then $c_1(x) \leftarrow 2$
Else if ($M_1(x) \geq T_{1,1}$ or $M_2(x) \geq T_{2,1}$) then $c_1(x) \leftarrow 1$
Else $c_1(x) \leftarrow 0$

If ($M_3(x) \geq T_{3,2}$ or $M_4(x) \geq T_{4,2}$) then $c_2(x) \leftarrow 2$
Else if ($M_3(x) \geq T_{3,1}$ or $M_4(x) \geq T_{4,1}$) then $c_2(x) \leftarrow 1$
Else $c_2(x) \leftarrow 0$

For i = 5 to 10
If $M_i(x) \geq T_{i,2}$ then $c_{i-2}(x) \leftarrow 2$
Else if $M_i(x) < T_{i,1}$ then $c_{i-2}(x) \leftarrow 0$
Else $c_{i-2}(x) \leftarrow 1$
End

For i = 1 to 8
If $c_i(x) = 0$ then $N_0(x) \leftarrow N_0(x) + 1$
Else, if $c_i(x) = 1$ then $N_1(x) \leftarrow N_1(x) + 1$
Else $N_2(x) \leftarrow N_2(x) + 1$
End

$$h(x) \leftarrow \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{j=0}^{2} w_j \; N_j(x)}{w_2} \right) \right)$$

If $h(x) < K$ then $g_7(x) \leftarrow a_1 + b_1\, h(x) + c_1\, h(x)^2 + d_1\, h(x)^3$
Else $g_7(x) \leftarrow a_2 + b_2\, h(x) + c_2\, h(x)^2 + d_2\, h(x)^3$

Subcriterion 8: Absence of pain induced by management procedures

Definition of the subcriterion

This subcriterion aims at assessing the pain due to management procedures like mutilations or stunning. The pain which is here assessed is not due to an injury or a disease, but to human intervention.

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11132	Dehorning - % of dehorned animals	Cardinal	Herd
11133	Dehorning - method	Qualitative	Herd
11134	Dehorning - use of anaesthtics	Binary	Herd
11135	Dehorning - use of analgesics	Binary	Herd
11136	Tail docking - % of tail-docked animals	Cardinal	Herd
11137	Tail docking - amount of tail cut	Ordinal	Herd
11138	Tail docking - use of anaesthtics	Binary	Herd
11139	Tail docking - use of analgesics	Binary	Herd

All theses measures are assessed at **herd level**, considering the current practices used by the farmer.
If the farmer changed his practices recently (maybe less than 2 or 3 years ago), then it would be fair not to penalise him for his old practices (e.g. mutilations without the use of anaesthetics), even if some animals in the herd were mutilated following these old practices.
At herd level, the worst current practice is considered (if it concerns at least 15% of the animals).

<u>Dehorning</u>
What is important is:
- the age of the animals at the time of the mutilation
  ➔ *no dehorning / on calves / on adults*
- the method used to dehorn
  ➔ *thermal / chemical*
- whether anaesthetics and/or analgesics have been used
  ➔ *nothing /* anaesthetics / analgesics / anaesthetics and analgesics

Dehorning on *adults* should be considered unacceptable even if it concerns only one animal in the herd. If the farmer dehorned an adult cow because the animal came from another farm, he should have chosen a dehorned cow instead of non- dehorned one. The only case dehorning on adults is acceptable is when dehorning is motivated by medical reasons (e.g. a cow which brake itself one horn).

<u>Tail docking</u>
What is important is the way tail docking is performed (use of analgesics…) and the amount of tail cut:
- whether anaesthetics and/or analgesics have been used
  ➔ *nothing /* anaesthetics / analgesics / anaesthetics and analgesics
- the age of the animals at the time of the mutilation
  ➔ *no tail docking / some hair left / no hair left*

Analysis of the constraints

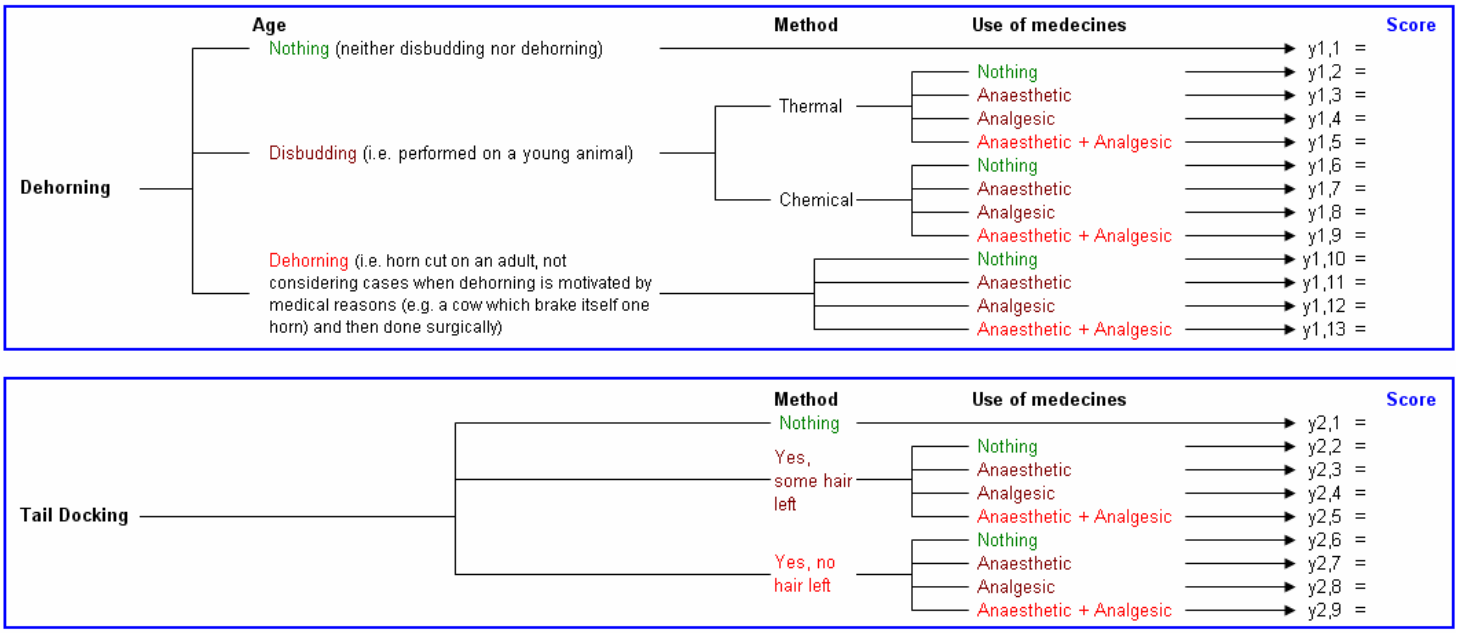
Crossing the two measures would lead to too numerous possibilities. We decided to consider the two mutilations separately, and then to consider at farm level the worst practice (dehorning or tail docking).

The subcriterion score should qualify the ***intensity of the negative deviation to a recommendation for the conditions of realisation of routine mutilations****.*

Type of mathematical construction proposed

We propose to attribute a score to each of the combinations according to the severity and pain they pose on animals.

Figure 1: Combinations of conditions for dehorning and tail docknig.



Parameters to be defined

The only elements we will have to define are the scores $y_{i,j}$ for i=1 to 2 and j=1to13 (for dehorning) or 9 (for tail docking).

Adjustments of the mathematical construction

- **Questions asked to experts**

Figure 1 was shown to experts. They were asked to estimate scores $y_{1,1}$ to $y_{1,13}$ and scores $y_{2,1}$ to $y_{2,9}$.
The experts asked were I. Veissier, J. Capdeville, H.R. Whay, U. Knierim, E. Canali.

- **Adjustment of the construction according to answers from experts**

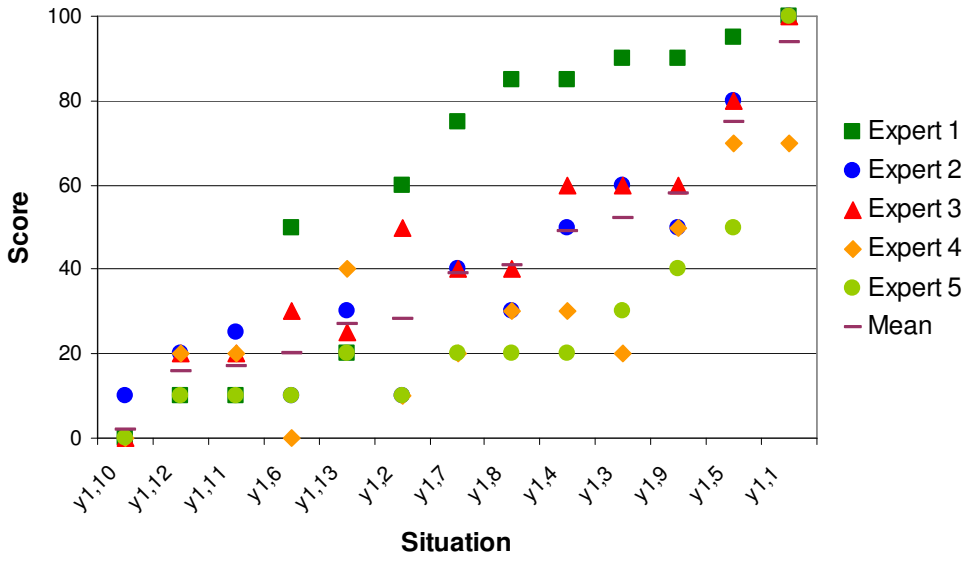
Dehorning:

The mean of scores attributed by the experts (Table 1a) to a given combination is considered the subcriterion-score to be attributed to this combination.

Table 1a: Evaluation of the 13 combinations by experts.

Situation	Score					
	Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*
y1,10	*0*	*10*	*0*	*0*	*0*	2
y1,12	*10*	*20*	*20*	*20*	*10*	16
y1,11	*10*	*25*	*20*	*20*	*10*	17
y1,6	*50*	*10*	*30*	*0*	*10*	20
y1,13	*20*	*30*	*25*	*40*	*20*	27
y1,2	*60*	*10*	*50*	*10*	*10*	28
y1,7	*75*	*40*	*40*	*20*	*20*	39
y1,8	*85*	*30*	*40*	*30*	*20*	41
y1,4	*85*	*50*	*60*	*30*	*20*	49
y1,3	*90*	*60*	*60*	*20*	*30*	52
y1,9	*90*	*50*	*60*	*50*	*40*	58
y1,5	*95*	*80*	*80*	*70*	*50*	75
y1,1	*100*	*100*	*100*	*70*	*100*	94

Figure 2a: Graph representing experts' answers for dehorning.



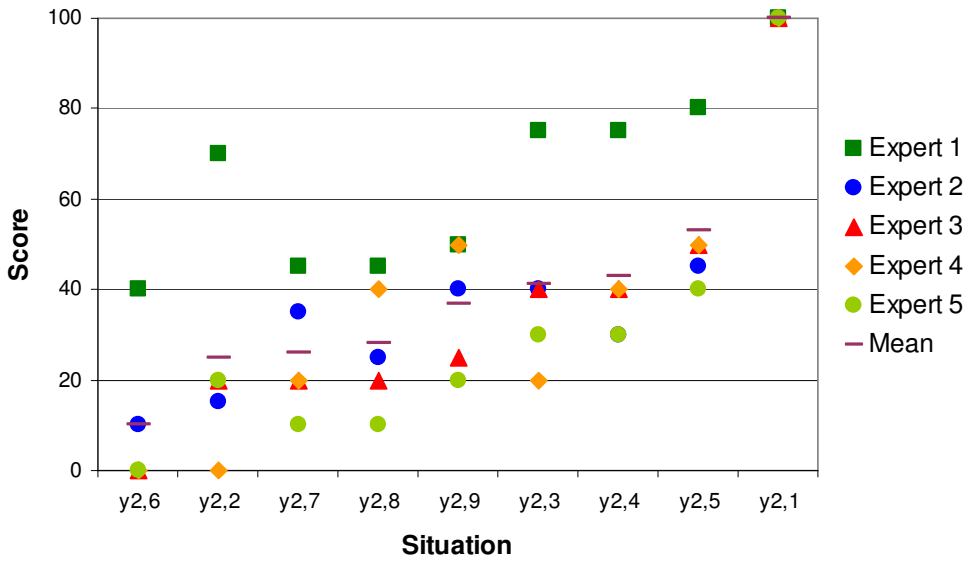
Tail docking:

The mean of scores attributed by the experts (Table 1b) to a given combination is considered the subcriterion-score to be attributed to this combination.

Table 1b: Evaluation of the 9 combinations by experts.

Situation	Score					
	Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*
y2,6	*40*	*10*	*0*	*0*	*0*	10
y2,2	*70*	*15*	*20*	*0*	*20*	25
y2,7	*45*	*35*	*20*	*20*	*10*	26
y2,8	*45*	*25*	*20*	*40*	*10*	28
y2,9	*50*	*40*	*25*	*50*	*20*	37
y2,3	*75*	*40*	*40*	*20*	*30*	41
y2,4	*75*	*30*	*40*	*40*	*30*	43
y2,5	*80*	*45*	*50*	*50*	*40*	53
y2,1	*100*	*100*	*100*	*100*	*100*	100

Figure 2b: Graph representing experts' answers for tail docking.



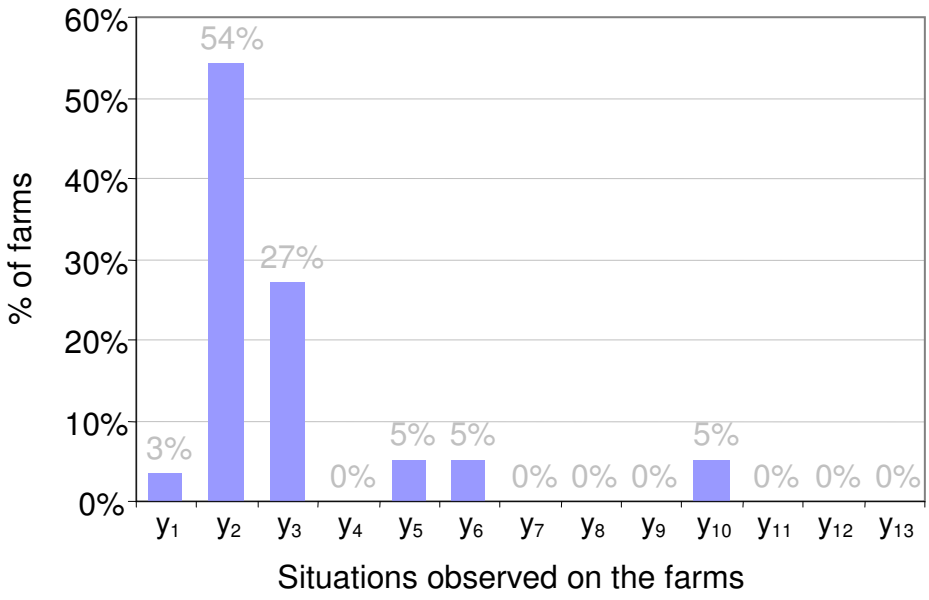
- **Test on dataset(s)**

We looked at the first 50 farms visited during WP2.4 to see the distribution of farms on dehorning.

None of the visited farms carried out tail docking.

We obtained the results presented on Figure 3.

Figure 3: Distribution of farms on *dehorning* (on 50 farms)



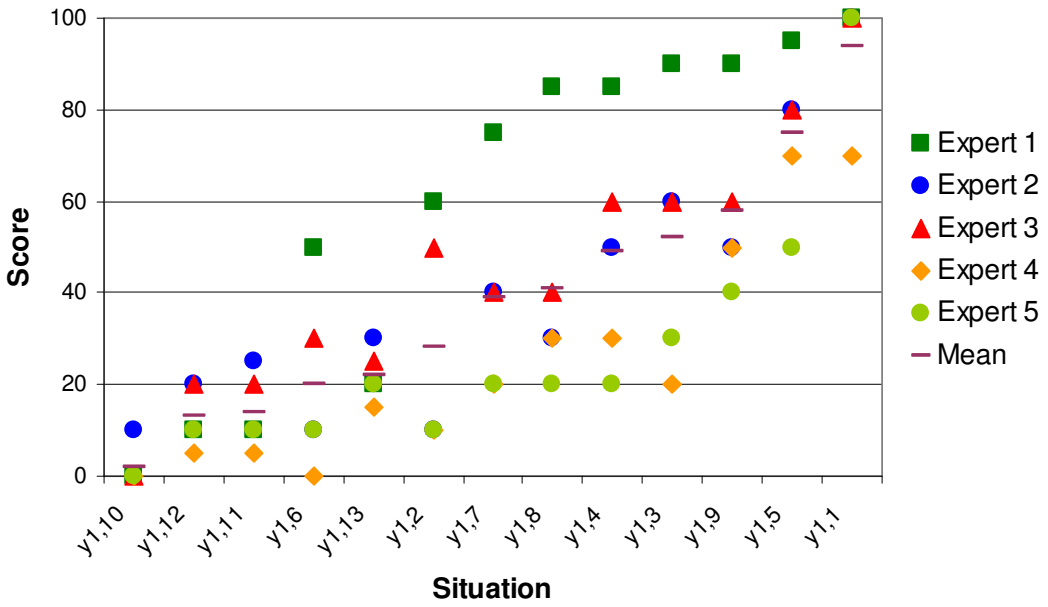
- **Final adjustment according to test on dataset**

We contacted again the experts in order to ask them to refine (if they consider it was necessary) their first opinion, taking into account the information on the distribution of farms on SC8, provided in Figure 3. Some of them did not change their first opinion.

Table 3: Answers from experts on the 13 possible combinations for dehorning.

Situation	Score					
	Expert 1	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*
y1,10	*0*	*10*	*0*	*0*	*0*	2
y1,12	*10*	*20*	*20*	*5*	*10*	13
y1,11	*10*	*25*	*20*	*5*	*10*	14
y1,6	*50*	*10*	*30*	*0*	*10*	20
y1,13	*20*	*30*	*25*	*15*	*20*	22
y1,2	*60*	*10*	*50*	*10*	*10*	28
y1,7	*75*	*40*	*40*	*20*	*20*	39
y1,8	*85*	*30*	*40*	*30*	*20*	41
y1,4	*85*	*50*	*60*	*30*	*20*	49
y1,3	*90*	*60*	*60*	*20*	*30*	52
y1,9	*90*	*50*	*60*	*50*	*40*	58
y1,5	*95*	*80*	*80*	*70*	*50*	75
y1,1	*100*	*100*	*100*	*70*	*100*	94

Figure 4: Graph representing experts' answers for dehorning.



Calculation

Let *x* be a farm

Variables from raw data:
Let M_i be the answer to the ith question considered for farm *x:*

- M_1: % of dehorned animals
- M_2: method used to dehorn animals (0. no dehorning / 1. chemical / 2. thermocauter / 3. surgery)
- M_3: use of anaesthetics while dehorning? (0. No / 1. Yes)
- M_4: use of analgesics while dehorning? (0. No / 1. Yes)
- M_5: % of tail docked animals
- M_6: method used to tail dock animals (0. no tail docking / 1. rubber / 2. surgery)
- M_7: use of anaesthetics while tail docking? (0. No / 1. Yes)
- M_8: use of analgesics while tail docking? (0. No / 1. Yes)

Calculated variables:

Let h^i be the index obtained for *dehorning* if i=1 or for *tail docking* if i=2

Let g_6^i be the subscore obtained for *dehorning* if i=1 or for *tail docking* if i=2

Let g_6 be the subcriterion-score for *"Absence of pain due to management procedures"*

Variables with predefined values:
Let $y_{1,1}$ to $y_{1,13}$ and $y_{2,1}$ to $y_{2,9}$ be the scores attributed respectively to the 13 possible values of $h^1(x)$ and to the 9 possible values of $h^2(x)$

1. From raw data to data at farm level

Raw data are yet at farm level, they don't need any calculation to be stored in the database

2. From data at farm level to subcriterion score

$y_{1,1} = 94$
$y_{1,2} = 28$
$y_{1,3} = 52$
$y_{1,4} = 49$
$y_{1,5} = 75$
$y_{1,6} = 20$
$y_{1,7} = 39$
$y_{1,8} = 41$
$y_{1,9} = 58$
$y_{1,10} = 2$
$y_{1,11} = 14$
$y_{1,12} = 13$
$y_{1,13} = 22$

$y_{2,1} = 100$
$y_{2,2} = 25$
$y_{2,3} = 41$
$y_{2,4} = 43$
$y_{2,5} = 53$
$y_{2,6} = 10$
$y_{2,7} = 26$
$y_{2,8} = 28$
$y_{2,9} = 37$

$h^1(x) \leftarrow 0$
$h^2(x) \leftarrow 0$
$g_8^1(x) \leftarrow 0$
$g_8^2(x) \leftarrow 0$
$g_8(x) \leftarrow 0$

If $M_2(x) = 0$ then $h^1(x) \leftarrow 1$ and $g_8^1(x) \leftarrow y_{1,1}$
Else, if $M_1(x) < 0.15$ then $h^1(x) \leftarrow 1$ and $g_8^1(x) \leftarrow y_{1,1}$
Else, if $M_2(x) = 2$ then
If ($M_3(x) = 0$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 2$ and $g_8^1(x) \leftarrow y_{1,2}$
Else, if ($M_3(x) = 1$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 3$ and $g_8^1(x) \leftarrow y_{1,3}$
Else, if ($M_3(x) = 0$ and $M_4(x) = 1$) then $h^1(x) \leftarrow 4$ and $g_8^1(x) \leftarrow y_{1,4}$
Else $h^1(x) \leftarrow 5$ and $g_8^1(x) \leftarrow y_{1,5}$
Else, if $M_2(x) = 1$ then
If ($M_3(x) = 0$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 6$ and $g_8^1(x) \leftarrow y_{1,6}$
Else, if ($M_3(x) = 1$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 7$ and $g_8^1(x) \leftarrow y_{1,7}$
Else, if ($M_3(x) = 0$ and $M_4(x) = 1$) then $h^1(x) \leftarrow 8$ and $g_8^1(x) \leftarrow y_{1,8}$
Else $h^1(x) \leftarrow 9$ and $g_8^1(x) \leftarrow y_{1,9}$
Else ($M_2(x) = 3$)
If ($M_3(x) = 0$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 10$ and $g_8^1(x) \leftarrow y_{1,10}$
Else, if ($M_3(x) = 1$ and $M_4(x) = 0$) then $h^1(x) \leftarrow 11$ and $g_8^1(x) \leftarrow y_{1,11}$
Else, if ($M_3(x) = 0$ and $M_4(x) = 1$) then $h^1(x) \leftarrow 12$ and $g_8^1(x) \leftarrow y_{1,12}$
Else $h^1(x) \leftarrow 13$ and $g_8^1(x) \leftarrow y_{1,13}$

If $M_6(x) = 0$ then $h^2(x) \leftarrow 1$ and $g_8^2(x) \leftarrow y_{2,1}$

Else, if $M_5(x) < 0.15$ then $h^2(x) \leftarrow 1$ and $g_8^2(x) \leftarrow y_{2,1}$

Else, if $M_6(x) = 1$ then

If ($M_7(x) = 0$ and $M_8(x) = 0$) then $h^2(x) \leftarrow 2$ and $g_8^2(x) \leftarrow y_{2,2}$

Else, if ($M_7(x) = 1$ and $M_8(x) = 0$) then $h^2(x) \leftarrow 3$ and $g_8^2(x) \leftarrow y_{2,3}$

Else, if ($M_7(x) = 0$ and $M_8(x) = 1$) then $h^2(x) \leftarrow 4$ and $g_8^2(x) \leftarrow y_{2,4}$

Else $h^2(x) \leftarrow 5$ and $g_8^2(x) \leftarrow y_{2,5}$

Else, if $M_6(x) = 2$ then

If ($M_7(x) = 0$ and $M_8(x) = 0$) then $h^2(x) \leftarrow 6$ and $g_8^2(x) \leftarrow y_{2,6}$

Else, if ($M_7(x) = 1$ and $M_8(x) = 0$) then $h^2(x) \leftarrow 7$ and $g_8^2(x) \leftarrow y_{2,7}$

Else, if ($M_7(x) = 0$ and $M_8(x) = 1$) then $h^2(x) \leftarrow 8$ and $g_8^2(x) \leftarrow y_{2,8}$

Else $h^2(x) \leftarrow 9$ and $g_8^2(x) \leftarrow y_{2,9}$

If $g_6^1(x) \leq g_6^2(x)$ then $g_6(x) = g_6^1(x)$

Else $g_6(x) = g_6^2(x)$

Subcriterion 9: Expression of social behaviours

Definition of the subcriterion

In this subcriterion, we aim at assessing the balance between the expression of negative (e.g. aggression) and positive (e.g. social licking) social behaviours.

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11140	Frequencies of agonistic social behaviour: Head butts	Cardinal	Herd or Group
11141	Frequencies of agonistic social behaviour: Displacements	Cardinal	Herd or Group
11142	Frequencies of agonistic social behaviour: Chasing	Cardinal	Herd or Group
11143	Frequencies of agonistic social behaviour: Fighting	Cardinal	Herd or Group
11144	Frequencies of agonistic social behaviour: Chasing up	Cardinal	Herd or Group

These measures are realised at the herd (or group) level and expressed as the average frequency of events per animal and per hour.

Analysis of the constraints

Two degrees of severity are distinguished among these behavioural measures. On the one hand we have to consider agnostic behaviours which did not cause a displacement of the target animal (Head butts, noted HB). On the other hand we have all the other agonistic behaviours causing displacement (noted D).

Behaviours that cause the displacement of an animal are judged more severe than simple head butts. The method of calculation should reflect this difference of importance of the two types of behaviours.

The subcriterion-score will thus represent ***the negative impact of agonistic behaviours on the animals.***

Type of mathematical construction proposed

Let i be the index of the measure ($i \in \{1,\ldots,5\}$).
Let M_i be the frequency of the behaviour i.

Let y_1 be the frequency of head butts and y_2 be the frequency of displacements:

$$\begin{cases} y_1(x) = M_1(x) \\ y_2(x) = \sum_{i=2}^{5} M_i(x) \end{cases}$$

We propose to calculate the subcriterion score on the basis of a linear combination of y_1 and y_2:

$$g_9(x) = u\left(100 \times \frac{(160\,w_1 + 340\,w_2) - \left(y_1(x)\ w_1 + y_2(x)\ w_2\right)}{160\,w_1 + 340\,w_2}\right)$$

Since there is no absolute worst situation (contrary to the other subcriteria, e.g. SC1 where the worst situation is characterised by 100% of cows with a body condition score < 2), we need define a reference farm that will correspond to what we consider as the worst situation or the situation below which all farms will obtain a score of 0. We propose for the reference farm a combination of 160 head butts per hour and 340 other agonistic behaviours per hour. In case a farm obtains an even worse combination of frequencies, a score of 0 will be assigned to that farm. These values (160 head butts and 340 other agonistic behaviours) were defined according to data we observed on farms in a previous study (not in Welfare Quality®) and also to data obtained by WP2.2 where a maximum of about 500 agonistic interactions were observed.

Probably, a non-linear utility function u will be necessary.

Parameters to be defined

We will have to define the weights w_i, and the most appropriate function u.

Adjustments of the mathematical construction

- **Questions asked to experts**

We made experts react on virtual dataset (Table 1), asking them first to rank the virtual farms from the best one to the worst one, and then to give a score to each of the farms presented in the dataset.
The asked experts were I. Veissier, U. Knierim, C. Winckler and S. Laister.

Table 1: Dataset submitted to experts. They had to sort the farms from best to worst, then to attribute a subcriterion score to each farm.

	Attacks		Displacements	*Rank*	*Score*
Farm 1	0	among which	0		
Farm 2	25	among which	0		
Farm 3	25	among which	12		
Farm 4	25	among which	25		
Farm 5	50	among which	0		
Farm 6	50	among which	25		
Farm 7	50	among which	50		
Farm 8	100	among which	0		
Farm 9	100	among which	50		
Farm 10	100	among which	100		
Farm 11	150	among which	50		
Farm 12	150	among which	100		
Farm 13	200	among which	70		
Farm 14	200	among which	140		
Farm 15	300	among which	100		
Farm 16	300	among which	200		
Farm 17	400	among which	130		
Farm 18	400	among which	260		
Farm 19	500	among which	170		
Farm 20	500	among which	340		

- **Adjustment of the construction according to answers from experts**

The experts' evaluations of farms with various levels of agonistic interactions between animals are shown in Table 2.

Table 2: Answers from experts

	Attacks		Displacements	Score					Rank	Index
				Expert 1	Expert 2	Expert 3	Expert 4	Mean		
Farm 20	500	among which	340	0	19	0	0	5	20	**0**
Farm 18	400	among which	260	5	20	7	7	10	19	**22**
Farm 19	500	among which	170	5	21	8	8	11	18	**28**
Farm 16	300	among which	200	10	23	14	14	15	17	**41**
Farm 17	400	among which	130	15	25	13	13	17	16	**43**
Farm 15	300	among which	100	20	32	20	20	23	15	**57**
Farm 14	200	among which	140	22	30	23	23	25	14	**59**
Farm 12	150	among which	100	30	36	35	35	34	13	**70**
Farm 13	200	among which	70	35	43	34	34	37	12	**71**
Farm 10	100	among which	100	37	42	41	41	40	11	**75**
Farm 11	150	among which	50	43	45	44	44	44	10	**78**
Farm 9	100	among which	50	45	52	50	50	49	9	**83**
Farm 7	50	among which	50	55	68	57	57	59	8	**87**
Farm 8	100	among which	0	70	60	63	63	64	6	**91**
Farm 6	50	among which	25	65	67	64	64	65	7	**91**
Farm 4	25	among which	25	80	82	68	68	75	5	**94**
Farm 1	0	among which	0	100	100	70	70	85	1	
Farm 5	50	among which	0	85	80	90	90	86	4	**96**
Farm 3	25	among which	12	90	90	95	95	93	3	**96**
Farm 2	25	among which	0	95	95	100	100	98	2	**98**

Farm 1 corresponds to a particular case on which experts disagree. According to the answers experts provided to us, 2 experts considered that Farm 1 was the best one, while the 2 others considered that Farm 1 was slightly better than Farm 4.
To determine the weights w_1 and w_2 we decided not to take this farm into account.

The weights $\mathbf{w_1}$ and $\mathbf{w_2}$ were determined by linear optimisation so as to match the ranking of farms according to mean score attributed by experts:

$\mathbf{w_1} = 5$
$\mathbf{w_2} = 14$

The index resulting from the weighted sum obtained by each farm is shown in Table 2, column 11.

As for Subcriterion 7, since the determination of weights taking into account all farms (i.e. including Farm 1) lead to an error, we couldn't define B-splines function to transform the linear index into the subcriterion score.

- **Test on dataset(s)**

We looked at 51 farms visited during WP2.4 to see the distribution of these farms on SC9. We obtained the results presented, in a cumulative way, in the fourth column of Table 3.

We then asked another time experts to react on the reorganised dataset (Table 3), including the information on the distribution of farms on SC9.

Table 3: Second dataset submitted to experts. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	Attacks		Displacements	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 20	500	among which	340	*0%*		
Farm 18	400	among which	260	*2%*		
Farm 19	500	among which	170	*2%*		
Farm 16	300	among which	200	*6%*		
Farm 17	400	among which	130	*6%*		
Farm 15	300	among which	100	*18%*		
Farm 14	200	among which	140	*24%*		
Farm 12	150	among which	100	*29%*		
Farm 13	200	among which	70	*29%*		
Farm 10	100	among which	100	*37%*		
Farm 11	150	among which	50	*43%*		
Farm 9	100	among which	50	*63%*		
Farm 7	50	among which	50	*71%*		
Farm 8	100	among which	0	*76%*		
Farm 6	50	among which	25	*78%*		
Farm 4	25	among which	25	*82%*		
Farm 1	0	among which	0			
Farm 5	50	among which	0	*86%*		
Farm 3	25	among which	12	*86%*		
Farm 2	25	among which	0	*94%*		

- **Final adjustment according to test on dataset**

The new answers of experts are presented in Table 4.

Table 4: Answers from experts

	Situation			Score						Rank	Index	Calculated score
	y_1+y_2		y_2	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Mean			
Farm 20	500	among which	340	0	5	0	0	0	1	20	0	**0,0**
Farm 18	400	among which	260	5	10	10	10	10	9	19	22	**7,3**
Farm 19	500	among which	170	10	5	15	14	14	12	18	27	**9,1**
Farm 16	300	among which	200	15	21	17	16	16	17	17	41	**15,1**
Farm 17	400	among which	130	20	10	20	18	18	17	16	43	**16,3**
Farm 15	300	among which	100	30	21	30	20	20	24	15	57	**26,8**
Farm 14	200	among which	140	32	35	35	25	25	30	14	59	**29,5**
Farm 12	150	among which	100	40	45	40	30	30	37	13	70	**43,0**
Farm 13	200	among which	70	42	35	45	40	40	40	12	71	**43,3**
Farm 10	100	among which	100	44	60	50	50	50	51	11	75	**49,9**
Farm 11	150	among which	50	50	45	55	60	60	54	10	78	**55,7**
Farm 9	100	among which	50	60	60	65	70	70	65	9	83	**63,9**
Farm 7	50	among which	50	70	80	75	75	75	75	8	87	**72,8**
Farm 8	100	among which	0	80	60	77	85	85	77	7	91	**79,9**
Farm 6	50	among which	25	82	80	80	85	85	82	6	91	**81,1**
Farm 4	25	among which	25	95	80	92	90	90	89	4	95	**89,7**
Farm 1	0	among which	0	100	100	90	80	80	90	1,5	100	**100,0**
Farm 5	50	among which	0	90	95	85	90	90	90	5	94	**86,0**
Farm 3	25	among which	12	95	95	95	90	90	93	3	96	**90,5**
Farm 2	25	among which	0	100	95	100	100	100	99	1,5	98	**94,8**

The experts agree to consider Farm 20 (characterised by 160 head butts per hour per 100 cows and 340 other agonistic behaviours) as the worst possible situations. All farms with a combination of frequencies of head butts and other agonistic behaviours even worse than Farm 20 should then receive a score of 0.

The weights $\mathbf{w_1}$ and $\mathbf{w_2}$ remain slightly changed:

$\mathbf{w_1} = 4$
$\mathbf{w_2} = 11$

To determine the most appropriate function ***u***, we applied another technique using I-splines instead of B-splines, and with only one interior knot, here placed at 70. The function ***u*** is composed of two pieces of cubic curves, and writes as follows:

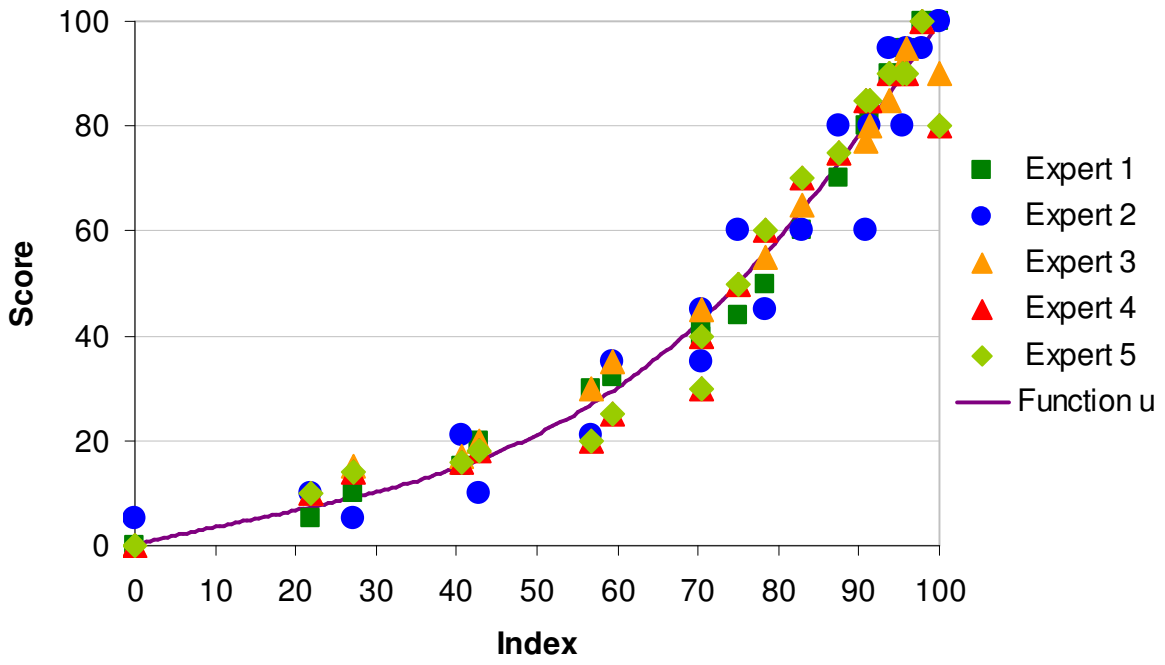
$$u(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 5.

Table 5: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,70]$	$x \in [70,100]$
a_x	0	86.8086955
b_x	0.3964767	-3.32175776
c_x	-0.00558523	0.04745369
d_x	0.00012269	-0.00012917

Figure 1: I-spline function (with one interior knot at 70) to attribute a subcriterion-score to 'Comfort around resting'



Calculation

Let *x* be a farm

Variables from raw data:
Let M_i be the frequency of the ith behaviour (per hour and per 100 animals) on farm *x:*

- M_1: Head butts
- M_2: Displacements
- M_3: Chasing
- M_4: Fighting
- M_5: Chasing up

Calculated variables:
Let y_1 and y_2 be the frequency (in number of behaviours per hour and per 100 animals) of respectively *head butts* and other *agonistic behaviours*
Let *h* be the index obtained for *"Expression of social behaviours"*
Let g_9 be the subcriterion-score for *"Expression of social behaviours"*

Variables with predefined values:
Let n_1 and n_2 be the highest frequencies of respectively *head butts (without displacement)* and *other agonistic behaviours (causing displacement)* to be observed on a farm. If the combination of the frequencies observed on a given farm leads to an even worse situation, a score of 0 will be assigned to that farm
Let w_1 and w_2 be the weights associated respectively to *head butts (without displacement)* and *other agonistic behaviours (causing displacement)*
Let a_1, b_1, c_1, d_1, a_2, b_2, c_2 and d_2 be the parameters of the two pieces of the utility function
Let K be the position of the knot (i.e. limit between the two pieces of the utility function)

1. From raw data to data at farm level

Raw data are yet at farm level, they don't need any calculation to be stored in the database

2. From data at farm level to subcriterion score

$n_1 = 160$
$n_2 = 340$

$w_1 = 4$
$w_2 = 11$

$a_1 = 0$
$b_1 = 0.3964767$
$c_1 = -0.00558523$
$d_1 = 0.00012269$

$a_2 = 86.8086955$
$b_2 = -3.32175776$
$c_2 = 0.04745369$
$d_2 = -0.00012917$

$K = 70$

$y_1(x) \leftarrow 0$
$y_2(x) \leftarrow 0$

$h(x) \leftarrow 0$
$g_9(x) \leftarrow 0$

$$y_1(x) \leftarrow M_1(x)$$

$$y_2(x) \leftarrow \sum_{i=2}^{5} M_i(x)$$

$$h(x) \leftarrow 100 \times \frac{(w_1\, n_1 + w_2\, n_2) - (w_1\, y_1(x) + w_2\, y_2(x))}{w_1\, n_1 + w_2\, n_2}$$

If $h(x) < 0$ then $h(x) \leftarrow 0$
Else $h(x) \leftarrow h(x)$

If $h(x) < K$ then $g_9(x) \leftarrow a_1 + b_1\, h(x) + c_1\, h(x)^2 + d_1\, h(x)^3$
Else $g_9(x) \leftarrow a_2 + b_2\, h(x) + c_2\, h(x)^2 + d_2\, h(x)^3$

Subcriterion 11: Good human-animal relationship

Definition of the subcriterion

In this subcriterion, we aim at evaluating the level of animals' fear of humans or, on the contrary, the positive reactions of animals towards humans (search for the contact with humans).

Measures to be used to assess the subcriterion

Id_Meas.	Measure	Nature	Level of observation
11145	Reactions of the animals to a moving person: avoidance distance at the feeding place	Ordinal	Individual

Analysis of the constraints

It is unlikely that the measures, although continuous like flight distances, are to be interpreted continuously. For instance, there is probably more difference between an animal that has a flight distance at 0 (i.e. accept the contact) and another at 1 m, than between this animal and one that flies at 2 m. Hence it is probably more appropriate to use ordinal scales with few levels.

Type of mathematical construction proposed

Avoidance distance at the feeding place

We defined, in collaboration with WP2.2 partners who developed the measure, 4 levels:

- 0: the animal can be touched
- 1: avoidance distance ≤ 50cm
- 2: 50cm < avoidance distance < 1m
- 3: avoidance distance ≥ 1m

So, at the herd level, we will have to deal with the percentage of animals in each of these classes (from 0 to 3):

Classes	0	1	2	3
% of animals	p_0	p_1	p_2	p_3

We propose to calculate a weighted sum with the percentages of animals in each category weighted by the severity of the avoidance distance in respect to animal welfare.
The following formula allows the score to range between 0 and 100 with the worst situation set at 0 (we call this intermediate score *h(x)* the *index*):

$$h(x) = 100 - \frac{\sum_{i=0}^{3} w_i \; p_i(x)}{w_3}$$

Some transformation of this index may be necessary to better match to experts' opinion:

$$g_{11}(x) = u(h(x)) = u\left(100 - \frac{\sum_{i=0}^{3} w_i \; p_i(x)}{w_3}\right)$$

Parameters to be defined

We need:

- to set the weights (w_0, w_1, w_2, and w_3) to be used in the weighted sum,
- to define the most appropriate function *u* to transform the index into the subcriterion-score.

Adjustments of the mathematical construction

▪ Questions asked to experts

We made experts react on a virtual dataset (Table 1), asking them first to rank the virtual farms from the best one to the worst one, and then to give a score to each of the farms presented in the dataset.
The asked experts were I. Veissier, X. Boivin, J. Capdeville, C. Winckler, S. Waiblinger.

Table 1: Dataset submitted to experts. They had to sort the farm from best to worst, then to attribute subcriterion scores.

	p_0	p_1	p_2	p_3	*Rank*	*Score*
Farm 1	100	0	0	0		
Farm 2	75	25	0	0		
Farm 4	50	50	0	0		
Farm 3	75	0	25	0		
Farm 6	25	75	0	0		
Farm 8	0	100	0	0		
Farm 7	25	50	25	0		
Farm 9	0	75	25	0		
Farm 5	50	0	25	25		
Farm 10	0	50	50	0		
Farm 11	0	50	25	25		
Farm 12	0	25	75	0		
Farm 13	0	0	100	0		
Farm 14	0	0	75	25		
Farm 15	0	0	50	50		
Farm 16	0	0	25	75		
Farm 17	0	0	0	100		

▪ Adjustment of the construction according to answers from experts

The results of experts' evaluation of farms with various numbers of measures within each category are shown in Table 2 (Columns 6-10). We calculated the mean of subcriterion-scores attributed by the 6 experts; then the farms were sorted according to the mean subcriterion-score they obtained (Table 5, Columns 11-12).

The weights $\mathbf{w_0}$, $\mathbf{w_1}$, $\mathbf{w_2}$ and $\mathbf{w_3}$ were determined by linear optimisation so as to match the ranking of farms according to mean score attributed to experts:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 8$
$\mathbf{w_2} = 28$
$\mathbf{w_3} = 72$

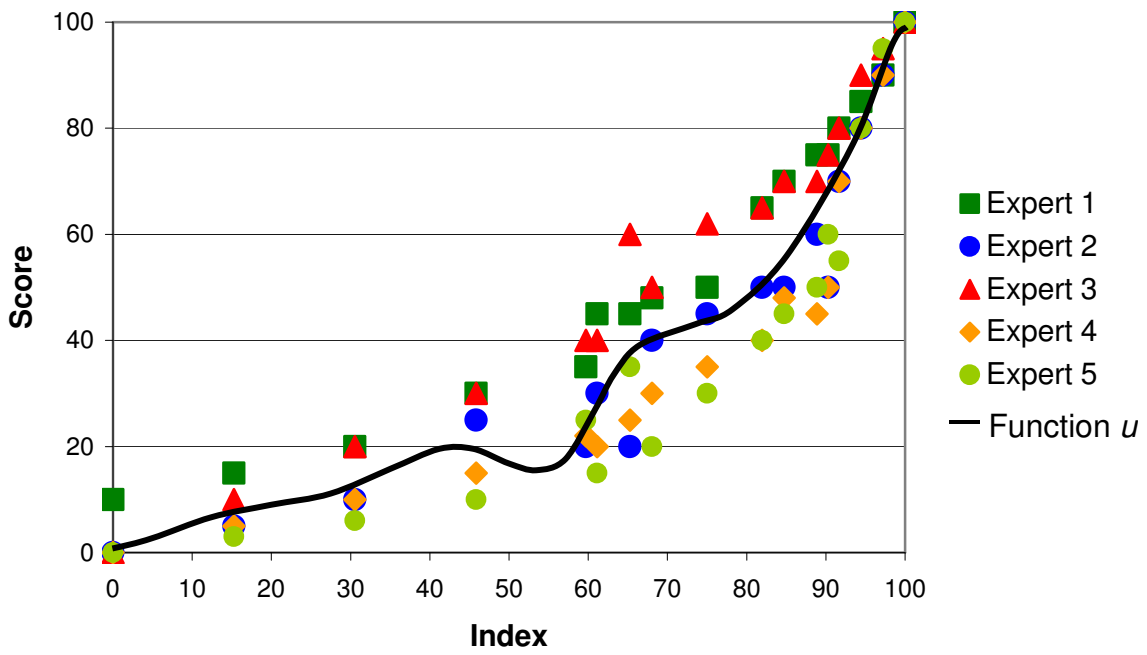
The index resulting from the weighted sum obtained by each farm is shown in Table 2, column 13.

Finally B-spline functions were defined so as to match best experts socres (i.e. minimise the sum of square errors with experts' scores) (Figure 1).

Table 2: Results obtained from the experts asked.

	Situation				Score						Rank	Index	Calculated score
	p_0	p_1	p_2	p_3	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*			
Farm 17	0	0	0	100	*10*	*0*	*0*	*0*	*0*	*2*	*17*	0	**0**
Farm 16	0	0	25	75	*15*	*5*	*10*	*5*	*3*	*8*	*16*	15	**7**
Farm 15	0	0	50	50	*20*	*10*	*20*	*10*	*6*	*13*	*15*	31	**12**
Farm 14	0	0	75	25	*30*	*25*	*30*	*15*	*10*	*22*	*14*	46	**19**
Farm 11	0	50	25	25	*35*	*20*	*40*	*22*	*25*	*28*	*13*	60	**23**
Farm 13	0	0	100	0	*45*	*30*	*40*	*20*	*15*	*30*	*12*	61	**28**
Farm 5	50	0	25	25	*45*	*20*	*60*	*25*	*35*	*37*	*11*	65	**37**
Farm 12	0	25	75	0	*48*	*40*	*50*	*30*	*20*	*38*	*10*	68	**40**
Farm 10	0	50	50	0	*50*	*45*	*62*	*35*	*30*	*44*	*9*	75	**43**
Farm 9	0	75	25	0	*65*	*50*	*65*	*40*	*40*	*52*	*8*	82	**51**
Farm 7	25	50	25	0	*70*	*50*	*70*	*48*	*45*	*57*	*7*	85	**56**
Farm 8	0	100	0	0	*75*	*60*	*70*	*45*	*50*	*60*	*6*	89	**65**
Farm 3	75	0	25	0	*75*	*50*	*75*	*50*	*60*	*62*	*5*	90	**69**
Farm 6	25	75	0	0	*80*	*70*	*80*	*70*	*55*	*71*	*4*	92	**73**
Farm 4	50	50	0	0	*85*	*80*	*90*	*80*	*80*	*83*	*3*	94	**81**
Farm 2	75	25	0	0	*90*	*90*	*95*	*90*	*95*	*92*	*2*	97	**92**
Farm 1	100	0	0	0	*100*	*100*	*100*	*100*	*100*	*100*	*1*	100	**100**

Figure 1: B-spline function to attribute a subcriterion-score to 'Comfort around resting'.



- **Test on dataset(s)**

We looked at 64 farms visited during WP2.4 to see the distribution of these farms on SC11. We obtained the results presented, in a cumulative way, in the fifth column of Table 3.

We then asked another time experts to react on the reorganised dataset (cf. Table 3), including the information on the distribution of farms on SC11.

Table 3: Second dataset submitted to experts. They had to attribute a subcriterion-score to each farm, taking into account the information provided on the distribution of farms

	p0	p1	p2	p3	% of farms **as good as or worse** than the farm considered	*Rank*	*Score*
Farm 17	0	0	0	100	0%		
Farm 16	0	0	25	75	0%		
Farm 15	0	0	50	50	0%		
Farm 14	0	0	75	25	3%		
Farm 11	0	50	25	25	8%		
Farm 13	0	0	100	0	8%		
Farm 5	50	0	25	25	8%		
Farm 12	0	25	75	0	8%		
Farm 10	0	50	50	0	14%		
Farm 9	0	75	25	0	16%		
Farm 7	25	50	25	0	23%		
Farm 8	0	100	0	0	41%		
Farm 3	75	0	25	0	48%		
Farm 6	25	75	0	0	52%		
Farm 4	50	50	0	0	69%		
Farm 2	75	25	0	0	86%		
Farm 1	100	0	0	0	100%		

- **Final adjustment according to test on dataset**

The new answers of experts are presented in Table 4.

The weights $\mathbf{w_0}$, $\mathbf{w_1}$, $\mathbf{w_2}$ and $\mathbf{w_3}$ slightly changed:

$\mathbf{w_0} = 0$
$\mathbf{w_1} = 3$
$\mathbf{w_2} = 11$
$\mathbf{w_3} = 26$

Table 4: Results obtained from the experts asked.

	Situation				Score						Rank	Index	Calculated score
	p_0	p_1	p_2	p_3	*Expert 1*	*Expert 2*	*Expert 3*	*Expert 4*	*Expert 5*	*Mean*			
Farm 17	0	0	0	100	*10*	*0*	*0*	*0*	*0*	*2*	*17*	0	**0**
Farm 16	0	0	25	75	*20*	*5*	*5*	*5*	*3*	*8*	*16*	14	**9**
Farm 15	0	0	50	50	*40*	*10*	*10*	*10*	*6*	*15*	*15*	29	**15**
Farm 14	0	0	75	25	*50*	*12*	*20*	*15*	*10*	*21*	*14*	43	**21**
Farm 13	0	0	100	0	*65*	*21*	*25*	*20*	*15*	*29*	*13*	58	**29**
Farm 11	0	50	25	25	*64*	*15*	*22*	*22*	*25*	*30*	*12*	59	**29**
Farm 5	50	0	25	25	*66*	*20*	*25*	*25*	*35*	*34*	*11*	64	**33**
Farm 12	0	25	75	0	*68*	*25*	*30*	*30*	*20*	*35*	*10*	65	**34**
Farm 10	0	50	50	0	*70*	*30*	*40*	*35*	*30*	*41*	*9*	73	**41**
Farm 9	0	75	25	0	*72*	*40*	*45*	*40*	*40*	*47*	*8*	81	**50**
Farm 7	25	50	25	0	*75*	*49*	*50*	*48*	*45*	*53*	*7*	84	**55**
Farm 8	0	100	0	0	*80*	*75*	*60*	*45*	*50*	*62*	*6*	88	**64**
Farm 3	75	0	25	0	*85*	*50*	*70*	*50*	*60*	*63*	*5*	89	**67**
Farm 6	25	75	0	0	*88*	*80*	*80*	*70*	*55*	*75*	*4*	91	**71**
Farm 4	50	50	0	0	*92*	*90*	*90*	*80*	*80*	*86*	*3*	94	**80**
Farm 2	75	25	0	0	*95*	*95*	*95*	*90*	*95*	*94*	*2*	97	**89**
Farm 1	100	0	0	0	*100*	*100*	*100*	*100*	*100*	*100*	*1*	100	**100**

To determine the most appropriate function ***u***, we applied I-spline with only one interior knot, here placed at 70. The function ***u*** is composed of two pieces of cubic curves, and writes as follows:

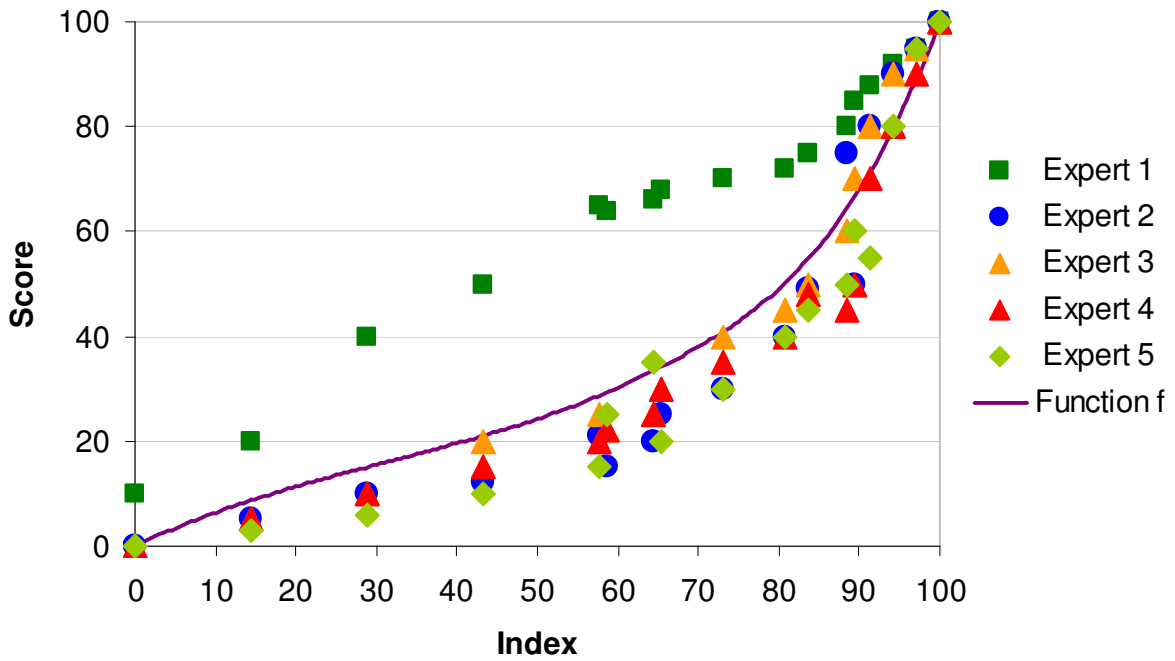
$$u(x) = a_x + b_x\, x + c_x\, x^2 + d_x\, x^3$$

with coefficients a_x, b_x, c_x and d_x depending on the value of x. These coefficients are presented in Table 5.

Table 5: Values of the coefficients a_x, b_x, c_x and d_x

Coefficients	$x \in [0,70]$	$x \in [70,100]$
a_x	0	-261,745092
b_x	0,73841231	11,9499006
c_x	-0,01077701	-0,17071286
d_x	0,00011433	0,00087388

Figure 2: I-spline function (with one interior knot at 70) to attribute a subcriterion-score to 'Comfort around resting'



Calculation

Let *x* be a farm

Variables from raw data:
Let *N* be the number of animals observed for *avoidance distance*
Let M_k be the data obtained by the kth animal observed on the farm *x* for *avoidance distance*

Calculated variables:
Let n_0, n_1, n_2 and n_3 be the number of animals assigned respectively to the categories 0, 1, 2 and 3 for *avoidance distance*
Let p_0, p_1, p_2 and p_3 be the % of animals assigned respectively to the categories 0, 1, 2 and 3 for *avoidance distance*
Let *h* be the index obtained for
Let g_6 be the subcriterion-score for *"Absence of injuries"*

Variables with predefined values:
Let *α*, *β* and *γ* be the thresholds used to delimit the four categories of problems defined for *avoidance distance*
Let w_0, w_1, w_2 and w_3 be the weights assigned to the four categories of problems
Let a_1, b_1, c_1, d_1, a_2, b_2, c_2 and d_2 be the parameters of the two pieces of the utility function
Let K be the position of the knot (i.e. limit between the two pieces of the utility function)

1. From raw data to data at farm level

$\alpha = 0$
$\beta = 50$
$\gamma = 100$

$n_0(x) \leftarrow 0$
$n_1(x) \leftarrow 0$
$n_2(x) \leftarrow 0$
$n_3(x) \leftarrow 0$

$p_0(x) \leftarrow 0$
$p_1(x) \leftarrow 0$
$p_2(x) \leftarrow 0$
$p_3(x) \leftarrow 0$

For k = 1, …, *N(x)*
If $M_k(x) \leq \alpha$ then $n_0(x) \leftarrow n_0(x) + 1$
Else, if $M_k(x) \leq \beta$ then $n_1(x) \leftarrow n_1(x) + 1$
Else, if $M_k(x) < \gamma$ then $n_2(x) \leftarrow n_2(x) + 1$
Else $n_3(x) \leftarrow n_3(x) + 1$
End

$$p_0(x) \leftarrow \frac{n_0(x)}{N(x)}$$
$$p_1(x) \leftarrow \frac{n_1(x)}{N(x)}$$
$$p_2(x) \leftarrow \frac{n_2(x)}{N(x)}$$
$$p_3(x) \leftarrow \frac{n_3(x)}{N(x)}$$

2. From data at farm level to subcriterion score

$w_0 = 0$
$w_1 = 3$
$w_2 = 11$
$w_3 = 26$

$a_1 = 0$
$b_1 = 0.73841231$
$c_1 = -0.01077701$
$d_1 = 0.00011433$

$a_2 = -261.745092$
$b_2 = 11.9499006$
$c_2 = -0.17071286$
$d_2 = 0.00087388$

$K = 70$

$$h(x) \leftarrow 0$$
$$g_{11}(x) \leftarrow 0$$

$$h(x) = 100 - \frac{\sum_{k=0}^{3} w_k \; p_k(x)}{w_3}$$

If $h(x) < K$ then $g_{11}(x) \leftarrow a_1 + b_1 \; h(x) + c_1 \; (h(x))^2 + d_1 \; (h(x))^3$

Else $g_{11}(x) \leftarrow a_2 + b_2 \; h(x) + c_2 \; (h(x))^2 + d_2 \; (h(x))^3$

Annexe D

Démonstration de la continuité d'ordre 2 des fonctions B-splines cubiques

B-splines basiques :

$$\begin{aligned} b_{-1}(\lambda) &= \frac{1}{6}(1 - 3\lambda + 3\lambda^2 - \lambda^3) \\ b_0(\lambda) &= \frac{1}{6}(4 - 6\lambda^2 + 3\lambda^3) \\ b_1(\lambda) &= \frac{1}{6}(1 + 3\lambda + 3\lambda^2 - 3\lambda^3) \\ b_2(\lambda) &= \frac{1}{6}(\lambda^3) \end{aligned} \tag{D.1}$$

Dérivées d'ordre 1 des B-splines basiques :

$$\begin{aligned} b'_{-1}(\lambda) &= \frac{1}{6}(-3 + 6\lambda - 3\lambda^2) \\ b'_0(\lambda) &= \frac{1}{6}(-12\lambda + 9\lambda^2) \\ b'_1(\lambda) &= \frac{1}{6}(3 + 6\lambda - 9\lambda^2) \\ b'_2(\lambda) &= \frac{1}{6}(3\lambda^2) \end{aligned} \tag{D.2}$$

Dérivées d'ordre 2 des B-splines basiques :

$$\begin{aligned} b''_{-1}(\lambda) &= \frac{1}{6}(6\lambda - 6\lambda) \qquad \text{(D.3)} \\ b''_{0}(\lambda) &= \frac{1}{6}(-12 + 18\lambda) \\ b''_{1}(\lambda) &= \frac{1}{6}(6\lambda - 18\lambda) \\ b''_{2}(\lambda) &= \frac{1}{6}(6\lambda) \end{aligned} \qquad \text{(D.4)}$$

Au niveau d'un noeud (placé en b), les deux portions de courbes consécutives se rejoignent. Chacune de ces portions de courbes (l'une exprimée sur l'intervalle $[a, b]$ et l'autre sur l'intervalle $[b, c]$, avec $a < b < c$) est issue du calcul d'une combinaison linéaire des quatre splines basiques. Pour le calcul des deux fonctions qui se rejoignent à un noeud (peu importe lequel), cinq coefficients différents sont nécessaires (cf. Chapitre 5), nous les noterons α_i avec $i \in \{1, 2, 3, 4, 5\}$.

La portion de courbe avant le noeud (sur $[a, b]$) est définie par la fonction u_1 définie comme suit :

$$u_1(z) = \alpha_1\, b_{-1}(\bar{z}) + \alpha_2\, b_0(\bar{z}) + \alpha_3\, b_1(\bar{z}) + \alpha_4\, b_2(\bar{z}) \qquad \text{(D.5)}$$

où $\bar{z} = \frac{z-a}{b-a}$ pour tout $z \in [a, b]$.

En b, $\bar{z} = 1$.

La portion de courbe après le noeud (sur $[b, c]$) est définie par la fonction u_1 définie comme suit :

$$u_2(z) = \alpha_2\, b_{-1}(\bar{z}) + \alpha_3\, b_0(\bar{z}) + \alpha_4\, b_1(\bar{z}) + \alpha_5\, b_2(\bar{z}) \qquad \text{(D.6)}$$

où $\bar{z} = \frac{z-b}{c-b}$ pour tout $z \in [b, c]$.

En b, $\bar{z} = 0$.

Continuité d'ordre 0 :

$$\begin{cases} u_1(1) = \alpha_1\, b_{-1}(1) + \alpha_2\, b_0(1) + \alpha_3\, b_1(1) + \alpha_4\, b_2(1) \\ u_2(0) = \alpha_2\, b_{-1}(0) + \alpha_3\, b_0(0) + \alpha_4\, b_1(0) + \alpha_5\, b_2(0) \end{cases}$$

$$\Longrightarrow u_1(1) = u_2(0) = \alpha_2 \left(\frac{1}{6}\right) + \alpha_3 \left(\frac{4}{6}\right) + \alpha_4 \left(\frac{1}{6}\right) \qquad \text{(D.7)}$$

Continuité d'ordre 1 :

$$\begin{cases} u_1'(1) = \alpha_1\, b_{-1}'(1) + \alpha_2\, b_0'(1) + \alpha_3\, b_1'(1) + \alpha_4\, b_2'(1) \\ u_2'(0) = \alpha_2\, b_{-1}'(0) + \alpha_3\, b_0'(0) + \alpha_4\, b_1'(0) + \alpha_5\, b_2'(0) \end{cases}$$
$$\Longrightarrow u_1'(1) = u_2'(0) = \alpha_2 \left(-\frac{3}{6}\right) + \alpha_4 \left(\frac{3}{6}\right) \tag{D.8}$$

Continuité d'ordre 2 :

$$\begin{cases} u_1''(1) = \alpha_1\, b_{-1}''(1) + \alpha_2\, b_0''(1) + \alpha_3\, b_1''(1) + \alpha_4\, b_2''(1) \\ u_2''(0) = \alpha_2\, b_{-1}''(0) + \alpha_3\, b_0''(0) + \alpha_4\, b_1''(0) + \alpha_5\, b_2''(0) \end{cases}$$
$$\Longrightarrow u_1''(1) = u_2''(0) = \alpha_2 - 2\,\alpha_3 + \alpha_4 \tag{D.9}$$

Références bibliographiques

Bibliographie

Autrement (1997). Le consommateur partagé. In M. Paillat (Ed.), *Le mangeur et l'animal - Mutations de l'élevage et de la consommation*, pp. 135–148. Paris, France : Autrement.

Bangoura, B. and A. Daugschies (2007). Parasitological and clinical parameters of experimental *Eimeria Zuernii* infection in calves and influence on weight gain and haemogram. *Parasitology Research 100*, 1331–1340.

Barnett, J. and P. Hemsworth (1990). The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science 25*(1-2), 177–187.

Bartels, R., J. Beatty, and B. Barsky (1987). *An introduction to splines for use in computer graphics and geometric modeling.* USA : Morgan Kaufmann publishers, INC.

Bartussek, H. (1999). A review of the animal needs index (ani) for the assessment of animals' well-being in the housing systems for austrian proprietary products and legislation. *Livestock Production Science 61*, 179–192.

Bell, D., H. Raiffa, and A. Tversky (1988). *Decision making : Descriptive, normative and prescriptive interactions.* Cambridge, UK : Cambridge University Press.

Ben Mena, S. (2001a). Méthodes de surclassement et analyse de robustesse. *Biotechnologie Agronomie Société Environnement 5*, 37–41.

Ben Mena, S. (2001b). Une solution informatisée à lŠanalyse de sensibilité dŠelectre iii. *Biotechnologie Agronomie Société Environnement 5*, 31–35.

Bennedsgaard, T. and S. Thamsborg (2000). Comparison of welfare assessment in organic dairy herds by the tgi200-protocol and a factor model based on clinical examinations and production parameters. In M. Hovi and R. G. Trujillo (Eds.), *Second*

NAHWOA Workshop (Diversity of livestock systems and definition of animal welfare), Cordoba, Spain, pp. 143–150.

Bitoun, O. (2007). La suisse championne du monde des MDD (sur www.lsa.fr).

Bock, B. and F. van Leeuwen (2005). Socio-political and market developments of animal welfare schemes. In J. Roex and M. Miele (Eds.), *Farm Animal Welfare Concerns, consumers, retailers and producers - Welfare Quality® Reports No. 1*, pp. 115–167. Cardiff, United Kingdom : Cardiff University.

Boe, K. E. and G. Faerevik (2003). Grouping and social preferences in calves, heifers and cows. *Applied Animal Behaviour Science 80*(3), 175–190.

Botreau, R., M. Bonde, A. Butterworth, P. Perny, M. Bracke, J. Capdeville, and I. Veissier (2007). Aggregation of measures to produce an overall assessment of animal welfare : Part 1 - a review of existing methods. *Animal 1*(8), 1179–1187.

Botreau, R., M. Bracke, P. Perny, A. Butterworth, J. Capdeville, C. van Reenen, and I. Veissier (2007). Aggregation of measures to produce an overall assessment of animal welfare : Part 2 - analysis of constraints. *Animal 1*(8), 1188–1197.

Botreau, R., I. Veissier, A. Butterworth, M. Bracke, and L. Keeling (2007). Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16*(2), 225–228.

Bourdon, J. (2003). Recherche agronomique et bien-être des animaux d'élevage - histoire d'une demande sociale. *Histoire et Sociétés Rurales 19*, 221–239.

Bouyssou, D. (1990). Building criteria : a prerequisite for mcda. In C. Bana e Costa (Ed.), *Readings in Multiple Criteria Decision-Aid*, pp. 58–80. Heidelberg, Germany : Springer Verlag.

Bouyssou, D., T. Marchant, M. Pirlot, P. Perny, A. Tsoukiàs, and P. Vincke (2000). *Evaluation and Decision Models - A Critical Perspective*. Dordrecht, The Netherlands : Kluwer Academic Publishers.

Bracke, M., B. Spruijt, and J. Metz (1999). Overall animal welfare assessment reviewed. part 3 : Welfare assessment based on needs and supported by expert opinion. *Netherlands Journal of Agricultural Science 47*, 307–322.

Bracke, M. B. M., B. M. Spruijt, J. H. M. Metz, and W. Schouten (2002). Decision support system for overall welfare assessment in pregnant sows a : Model structure and weighting procedure. *Journal of Animal Science 80*(7), 1819–1834.

Brambell, R. (1965). *Report of the Technical Committee to enquire into the welfare of animals kept under intensive livestock husbandry systems.* London, United Kingdom : Her MajestyŠs Stationery Office.

Broom, D. (1986). Indicators of poor welfare. *British Veterinary Journal 142*, 524–526.

Broom, D. (1987). Applications of neurobiological studies to farm animal welfare. In P. Wiepkema and P. van Andrichem (Eds.), *Biology of stress in farm animals*, pp. 101–110. The Haguen, The Netherlands : M. Nijhof Publ.

Broom, D. (1996). Animal welfare defined in terms of attempts to cope with the environment. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 27*, 22–28.

Broom, D., W. Blokhuis, E. Canali, A. Dijkhuizen, R. Fallom, P. Le Neindre, H. Saloniemi, and A. Webster (1995). Report of the scientific veterinary committee, animal welfare section on the welfare of calves. Technical report, European Commission.

Burgat, F. and R. Dantzer (1997). Une nouvelle préoccupation : le bien-être animal. In M. Paillat (Ed.), *Le mangeur et l'animal - Mutations de l'élevage et de la consommation*, pp. 69–86. Paris, France : Autrement.

Capdeville, J. and I. Veissier (2001). A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 62–68.

Choquet, G. (1954). Theory of capacities. *Annales de l'Institut Fourier 5*, 131–295.

Commission of the European Communities (2002). Communication from the commission to the council and the european parliament on animal welfare legislation on farmes animals in third countries and the implications for the eu - com(2002)626.

Dantzer, R. (1983). Protection animale en élevage intensif. *Journées Recherche Porcine 15*, 25–36.

Dantzer, R. (1995). Confort et bien-être des animaux en élevage intensif. *Le Point Vétérinaire 26*, 1027–1034.

Dawkins, M. (1980). *Animal suffering : the science of animal welfare.* London : Chapman and Hall Ltd.

Dawkins, M. (1990). From an animal's point of view : Motivation,fitness, and animal welfare. *Psychological science 13*(1), 1–61.

Dawkins, M. (2006). A user's guide to animal welfare science. *Trends in Ecology and Evolution 21*, 77–82.

Dockès, A. (2000). Les représentations de l'animal et du bien-être animal par les éleveurs et les intervenants en élevage (sur http ://wcentre.tours.inra.fr/BienEtre/rapport.htm).

Dockès, A. and F. Kling-Eveillard (2006). Farmers' and advisers' representations of animals and animal welfare : Ethics in animal agriculture. *Livestock Science 103*(3), 243–249.

Dockès, A. and F. Kling-Eveillard (2007). Livestock farmers' perception of animals and animal welfare. *INRA Productions Animales 20*(1), 23–28.

Dubois, D., P. Fortemps, M. Pirlot, and H. Prade (2001). Leximin optimality and fuzzy set theoretic operations. *European Journal of Operational Research 130*, 20–28.

Duncan, I. (2002). Poultry welfare : Science or subjectivity ? *British Poultry Science 43*, 643–652.

Duncan, I. (2005). Science-based assessment of animal welfare : Farm animals. *Revue Scientifique Et Technique-Office International Des Epizooties 24*(2), 483–492.

Dupas, F. (2005). Le statut juridique de l'animal en france et dans les états membres de l'union européenne : historique, bases juridiques actuelles et conséquences pratiques. Master's thesis, École Nationale Vétérinaire de Toulouse.

European Commission (2005). *Attitudes of consumers towards the welfare of farmed animals - Special Eurobarometer 229/Wave 63.2 - TNS Opinion & Social.*

European Commission (2007a). *Attitudes of consumers towards the welfare of farmed animals - Special Eurobarometer 229 (2) / Wave 64.4 - TNS Opinion & Social.*

European Commission (2007b). *Attitudes of EU citizens towards Animal Welfare - Special Eurobarometer 270/Wave 66.1 - TNS Opinion & Social.*

European Economic and Social Committee (2007). *Opinion of the European Economic and Social Committee on Animal Welfare - Labelling (Exploratory opinion).* Brussels, Belgium : European Economic and Social Committee.

Ewbank, R. (1973). The trouble with being a farm animal. *New Scientist 60*, 172–173.

Fabre, A. (1995). Bien-être des animaux dans les élevages : enjeux et perspectives d'une réglementation nationale. *Le Point Vétérinaire 27*, 283–292.

Farm Animal Welfare Council (1992). Fawc updates the five freedoms. *Veterinary Record (The) 17*, 357.

Feinberg, J. (1980). Human duties and animal rights. In J. Feinberg (Ed.), *Rights, Justice and the Bounds of Liberty*, pp. 185–206. Princeton, USA : Princeton University Press.

Fraser, D. (1989). Welfare and well-being. *Veterinary Record 125*, 332–333.

Fraser, D. (1993). Assessing animal well-being : common sense, uncommon science. In U. Programs and P. U. O. of Agricultural Research (Eds.), *Food Animal Well-Being - Conference Proceedings and Deliberations*, Indianapolis, Indiana, pp. 37–54. Purdue University Office of Agricultural Research Programs.

Fraser, D. (1995). Science, values and animal welfare : exploring the 'inextricable connection'. *Animal Welfare 4*, 103–117.

Fraser, D. (2003). Assessing animal welfare at the farm and group level : the interplay of science and values. *Animal Welfare 12*(4), 433–443.

Friend, T. (1980). Stress : what is it and how can it be quantified ? *International Journal For The Study Of Animal Problems 1*, 366–374.

Gavinelli, A. (2007). Animal welfare labellin : competitiveness, consumer information and better regulation for the eu. In *second Welfare Quality® stakeholder conference*,

Berlin, Germany, pp. 23–24.

Geers, R., B. Petersen, K. Huysmans, S. Knura-Deszczka, M. De Becker, S. Gymnich, D. Henot, S. Hiss, and H. Sauerwein (2003). On-farm monitoring of pig welfare by assessment of housing, management, health records and plasma haptoglobin. *Animal Welfare 12*(4), 643–647.

Gould, D., D. Kelly, L. Goldstone, and J. Gammon (2001). Examining the validity of pressure ulcer risk assessment scales : developing and using illustrated patient simulations to collect the data - information point : Visual analogue scale. *Journal of Clinical Nursing 10*, 697–706.

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research 89*(3), 445–456.

Grabisch, M. (2006). L'utilisation de l'intégrale de choquet en aide multicritère à la décision. *Newsletter of the European WorkingGroup "Multicriteria Aid for Decisions" 3*(14), 5–10.

Grabisch, M., J. Duchêne, F. Lino, and P. Perny (2002). Subjective evaluation of discomfort in sitting position. *Fuzzy Optimization and Decision Making 1*(3), 287–312.

Grabisch, M., I. Kojadinovic, and P. Meyer (Accepted). A review of capacity identification methods for Choquet integral based multi-attribute utility theory - applications of the kappalab r package. *European Journal of Operational Research*.

Grabisch, M., T. Murofushi, and M. Sugeno (2000). *Fuzzy Measures and Integrals - Theory and Applications*. Berlin, Germany : Physica Verlag.

Grabisch, M. and M. Roubens (2000). Application of the Choquet integral in multicriteria decision making. In G. M., T. Murofushi, and M. Sugeno (Eds.), *Fuzzy Measures and Integrals - Theory and Applications*, pp. 348–374. Physica Verlag.

Guéméné, D. and J. Faure (2004). Productions avicoles, bien-être et législation européenne. *INRA Productions Animales 17*, 59–68.

Hasson, A. (1994). Le point de vue d'une protectionniste. In M. Picard, R. Porter, and J. Signoret (Eds.), *Un point sur... Comportement et bien-être animal*, pp. 11–15.

Paris, France : INRA Éditions.

Hegelund, L., J. T. Sørensen, and N. F. Johansen (2003). Developing a welfare assessment system for use in commercial organic egg production. *Animal Welfare 12*(4), 649–653.

Hemsworth, P. and G. Coleman (1998). *Human- Livestock Interactions : The Stockperson and the Productivity and Welfare of Intensively Farmed Animals.* Oxon/New-York : CAB International.

Horning, B. (2001). The assessment of housing conditions of dairy cows in littered loose housing systems using three scoring methods. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 42–47.

Hurnik, J. (1990). World's poultry science association invited lecture : Animal welfare : ethical aspects and practical considerations. *Poultry Science 69*, 1827–1834.

Hurnik, J., A. Webster, and P. Siegel (1985). *Dictionary of farm animal behaviour* (1st ed.). Guelph, Canada : Univesrity of Guelph.

Ingenbleek, P. (2007). Retailers dealing with welfare schemes. In *second Welfare Quality® stakeholder conference*, Berlin, Germany, pp. 47–51.

Johnsen, P. F., T. Johannesson, and P. Sandoe (2001). Assessment of farm animal welfare at herd level : many goals, many methods. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 26–33.

Kant, E. (1986). Métaphysique des moeurs (1797). In *Œuvres philosophiques III*, pp. 733. Paris, France : Gallimard.

Keeling, L. and B. Bock (2007). Turning welfare principles into practice : approach followed in welfare quality®. In *second Welfare Quality® stakeholder conference*, Berlin, Germany, pp. 25–28.

Keeling, L. and J. Svedberg (1999). Legislation banning conventional battery cages in sweden and a subsequent phase-out programme. In M. Kunisch and H. Eckel (Eds.), *Congress 'Regulation of Animal Production in Europe'*, Wiesbaden, Germany, pp. 73–78.

Kjærnes, U. (2007). Ethics and action : e relational perspective on food trends and consumer concerns. In W. Zollitsch, C. Winckler, S. Waiblinger, and A. Haslberger (Eds.), *7th Congress of the European Society for Agricultural and Food Ethics*, Vienna, Austria.

Kjaernes, U., M. Miele, and J. Roex (2007). *Attitudes of Consumers, Retailers and Producers to Farm Animal Welfare*. Welfare Quality.

Kjærnes, U., E. Roe, and B. Bock (2007). Societal concerns on farm animal welfare. In *second Welfare Quality® stakeholder conference*, Berlin, Germany, pp. 13–18.

Kyprianou, M. (2007). Prospects for an eu animal welfare labelling scheme. *AWIC Bulletin 13*, 35–36.

Lacroix, R., J. Huijbers, R. Tiemessen, D. Lefebvre, D. Marchand, and K. Wade (1997). Fuzzy set-based analytical tools for dairy herd improvement. *Applied engineering in agriculture 14*, 79–85.

Lacroix, R., M. Strasser, R. Kok, and K. Wade (1998). Performance analysis of a fuzzy decision-support system for culling of dairy cows. *Canadian agricultural engineering 40*, 139–152.

Landais, E. (1998). Agriculture durable : les fondements d'un nouveau contrat social ? *Courrier de l'environnement de l'INRA 33*, 5–22.

Latouche, K. (1999). French people's willingness to pay for farm animal welfare. In M. Kunisch and H. Eckel (Eds.), *International congress - Regulation of animal production in Europe*, Wiesbaden, Germany, pp. 108–112. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.v.

Lavoie, J. and P. Fougeyrollas (1998). De la pluri, à la multi vers l'interdisciplinarité de l'approche-programme (sur http ://www.med.univ-rennes1.fr/sisrai/art/approche-programme.html).

Le Denmat, M. (1994). Le point de vue de l'éleveur sur le bien-être des porcs. In M. Picard, R. Porter, and J. Signoret (Eds.), *Un point sur... Comportement et bien-être animal*, pp. 17–19. Paris, France : INRA Éditions.

Le Neindre, P. (2003). Le bien-être des animaux de rente. In C. Baudoin (Ed.),

L'éthologie appliquée aujourd'hui - Vol. 1 : Bien-être, élevages et expérimentation, pp. 25–31. Paris, France : Éditions ED.

Lensink, B. J., S. Raussi, X. Boivin, M. Pyykkonen, and I. Veissier (2001). Reactions of calves to handling depend on housing condition and previous experience with humans. *Applied Animal Behaviour Science 70*(3), 187–199.

Lund, V., G. Coleman, S. Gunnarsson, M. Appleby, and K. Karkinen (2006). Animan welfare science - working at the interface between the natural and social sciences. *Applied Animal Behaviour Science 97*, 37–49.

Main, D. C. J., J. P. Kent, F. Wemelsfelder, E. Ofner, and F. A. M. Tuyttens (2003). Applications for methods of on-farm welfare assessment. *Animal Welfare 12*(4), 523–528.

Main, D. C. J., F. Webster, and L. E. Green (2001). Animal welfare assessment in farm assurance schemes. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 108–113.

Marichal, J. (2000). An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems 8*(6), 800–807.

Mason, G. and M. Mendl (1993). Why is there no simple way of measuring animal welfare ? *Animal Welfare 2*(4), 301–319.

Miele, M. and A. Evans (2005). European consumers' views about farm animal welfare. In *1st Welfare Quality® stakeholder conference*, Brussels, Belgium.

Miele, M. and A. Evans (2007). Rationale behind the welfare quality® assessment of animal welfare. In *2nd Welfare Quality® stakeholder conference*, Berlin, Germany, pp. 19–22.

Moscarola, J. and B. Roy (1977). Procédure automatique d'examen de dossiers fondée sur une segmentation trichotomique en présence de critères multiples. *RAIRO Recherche Opérationnelle 11*, 145–173.

Mousseau, V. and R. Slowinski (1998). Inferring an electre tri model from assignment examples. *Journal of Global Optimization 12*, 157–174.

Mousseau, V., R. Slowinski, and P. Zielniewicz (1999). Electre tri 2.0a : Methodological guide and user's documentation, lamsade technical report no 111. Technical report, Université de Paris-Dauphine.

Mousseau, V., R. Slowinski, and P. Zielniewicz (2000). A user-oriented implementation of the electre tri method integrating preference elicitation support. *Computers & Operations Research 27*(7), 757–777.

Pala, ., J. Vennix, and J. Kleijnen (1999). Validation in soft OR, hard OR and system dynamics : a critical comprison and contribution to the debate. In J. Vennix and B. Cavana (Eds.), *Proceedings of the 17th International Conference of The System Dynamic Society and the 5th Australian & New Zealand Systems Conference*, New York, United States of America. The System Dynamics Society.

Parthenay, C. (2005). Herbert simon : rationalité limitée, théorie des organisations et sciences artificielles (sur http ://www.grjm.net/documents/claude_parthenay/Parthenay_Simon.pdf).

Perny, P. (1998). Multicriteria filtering methods based on concordance and non-discordance principle. *Annals of Operations Research 80*, 137–165.

Piñeiro, C., M. Morales, M. Piñeiro, J. Ruiz de la Torre, G. Mateos, and X. Manteca (2005). Psychological stress caused by changes in feeding management did not affect the concentration of acute phase proteins in pigs. In *5th International Colloquium on Animal Acute Phase Proteins*, Dublin, Ireland.

Plass, M. and M. Stone (1983). Curve-fitting with piecewise parametric cubics. *ACM SIGGRAPH Computer Graphics 17*(3), 229–239.

Pomerol, J.-C. and S. Barba-Romero (2000). *Multicriterion decision in management : principles and pratice*. Boston, United States of America : Kluwer Academic Publishers.

Poulain, J. (1997). Mutations et modes alimentaires. In M. Paillat (Ed.), *Le mangeur et l'animal - Mutations de l'élevage et de la consommation*, pp. 103–121. Paris, France : Autrement.

Ramsay, J. (1988). Monotone regression splines in action. *Statistical Science 3*, 425–461.

Raussi, S., B. J. Lensink, A. Boissy, M. Pyykkonen, and I. Veissier (2003). The effect of contact with conspecifics and humans on calves' behaviour and stress responses. *Animal Welfare 12*(2), 191–203.

Regan, T. (1992). Pour les droits des animaux. *Les Cahiers antispécistes 5*.

Renard, F. (1986). Utilisation d'electre tri dans l'analyse des réponses à un appel d'offres : le cas de la machine de tri paquets à la direction générale des postes. *Cahier du LAMSADE 73*.

Rosenfield, P. (1992). The potential of transdisciplinarity research for sustaining and extending linkages between the health and social sciences. *Social Science & Medicine 35*, 1343–1357.

Roy, B. (1985). *Méthodologie Multicritère d'Aide a la Décision*. Paris, France : Economica.

Roy, B. (1993). Decision science or decision-aid science ? *European Journal of Operation Research 66*, 184–204.

Roy, B. (1996). *Multicriteria Methodology for Decision Aiding*. Dordrechtn, The Netherlands : Kluwer Academic.

Roy, B. (2002). Robustesse de quoi et vis-à-vis de quoi mais aussi robustesse pourquoi en aide à la décision ? *Newsletter of the European WorkingGroup "Multicriteria Aid for Decisions" 3*(6), 1–6.

Royal Society for the Prevention of Cruelty to Animals (2004). *RSPCA Welfare Standards for dairy cattle*. Horsham, United Kingdom : RSPCA.

Rushen, J. (1986). The validity of behavioural measures of aversion : a review. *Applied Animal Behaviour Science 16*, 309–323.

Rutter, S. (September 24-25, 1996 1998). Assessing the welfare of intensive and extensive livestock. In *Workshop Pasture Ecology and Animal Intake*, Dublin, pp. 1–9.

Sanotra, G., C. Berg, and J. Lund (2003). A comparison between leg problems in danish and swedish broiler production. *Animal Welfare 12*, 677–683.

Scott, E. M., A. M. Nolan, and J. L. Fitzpatrick (2001). Conceptual and methodological issues related to welfare assessment : a framework for measurement. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 5–10.

Singer, P. (1990). The significance of animal suffering. *Psychological Science 13*(1), 9–12.

Sørensen, J. T., P. Sandoe, and N. Halberg (2001). Animal welfare as one among several values to be considered at farm level : the idea of an ethical account for livestock farming. *Acta Agriculturae Scandinavica, Section A, Animal Science Supplementum 30*, 11–16.

Stafleu, F., F. Grommers, and J. Vorstenbosch (1996). Animal welfare : Evolution and erosion of a moral concept. *Animal Welfare 5*(3), 225–234.

Tannenbaum, J. (1991). Ethics and animal welfare : the inextricable connection. *Journal of the American Veterinary Medical Association 198*, 1360–1376.

Thiermann, A. and S. Babcock (2005). Animal welfare and international trade. *Revue Scientifique Et Technique-Office International Des Epizooties 24*, 747–755.

Torra, V. (1997). The weighted owa operator. *International Journal of Intelligent Systems 12*(2), 153–166.

Tsoukiàs, A. (2007). On the concept of decision aiding process. *Annals of Operations Research 154*, 3–27.

van Calker, K., P. Berentsen, C. Romero, G. Giesen, and R. Huirne (2006). Development and application of a multi-attribute sustainability function for dutch dairy farming systems. *Ecological Economics 57*(4), 640–658.

Veissier, I., C. Beaumont, and F. Levy (2007). Research into animal welfare : Aims, methods and purpose. *Productions Animales 20*(1), 3–9.

Veissier, I., H. Blokhuis, R. Geers, B. Jones, and M. Miele (2005). The welfare quality project : From consumer expectations to welfare certifications in animal husbandry. *Bulletin De L'academie Veterinaire De France 158*(3), 263–267.

Veissier, I., R. Botreau, J. Capdeville, and P. Perny (2007). L'évaluation en ferme du bien-être des animaux : objectifs, outils disponibles, utilisations, exemple du projet

welfare quality®. In *14èmes Rencontres Recherches Ruminants*, Paris, France.

Veissier, I., A. Butterworth, B. Bock, and E. Roe (Accepted). European approaches to ensure good animal welfare. *Applied Animal Behaviour Science Special Issue 40 years after the Brambell report.*

Veissier, I., A. Ramirez de la Fe, and P. Pradel (1998). Nonnutritive oral activities and stress responses of veal calves in relation to feeding and housing conditions. *Applied Animal Behaviour Science 57*, 35–49.

Veissier, I., C. Sarignac, and J. Capdeville (1999). Les méthodes d'appréciation du bien-être des animaux d'élevage. *INRA Productions Animales 12*, 113–121.

Verley, P. (1997). *L'Échelle du monde, essai sur l'industrialisation de l'Occident.* Paris, France : Gallimard.

Vincke, P. (1992). *Multicriteria Decision Aid.* New-York, USA : Wiley.

Špinka, M., I. Dembele, J. Panamá, and I. Stehulová (2005). Lame dairy cows have shorter avoidance distances. In R. Kusunose and S. Sato (Eds.), *Proceedings of the 39th International Congress of the International Society for Applied Ethology*, pp. 83. ISAE2005.

Webster, J. (1997). Applied ethology : what use is it to animal welfare ? *Advances in ethology 32 Supplements to Ethology*, 10.

Webster, J. (2005). *Animal welfare limping towards Eden* (2nd ed.). Oxford, United Kingdom : Blackwell Publishing.

Wechsler, B. (2001). Pretesting of mass-produced farm animal housing systems in switzerland 20 years of experience. In P. C. o. A. Engineering (Ed.), *International Symposium of the 2nd Technical Section of C.I.G.R. on Animal Welfare Considerations in Livestock Housing Systems*, Szklarska Poreba, Poland, pp. 55–67. Poligmar.

Wechsler, B. (2003). Testing of mass-produced farm animal housing systems with regard to animal welfare. In Y. van der Honing (Ed.), *54th Annual Meeting of the European Association for Animal Production*, Rome, Italy, pp. 125. Wageningen Pers.

Wechsler, B. (2005). An authorisation procedure for mass-produced farm animal housing systems with regard to animal welfare. *Livestock Production Science 94*(1-2),

71–74.

Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster (2003). Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens : Consensus of expert opinion. *Animal Welfare 12*(2), 205–217.

Winckler, C., J. Capdeville, G. Gebresenbet, B. Horning, U. Roiha, M. Tosi, and S. Waiblinger (2003). Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. *Animal Welfare 12*(4), 619–624.

Yager, R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics 18*(1), 183–190.

Table des figures

Table des figures

1.1 Évolution de la population urbaine et de la population non agricole en France, de 1965 à 2005 (Source : banque de données de la FAO[1]) 9

1.2 Réponses des consommateurs à la question *"Comment jugez-vous le niveau de bien-être animal en élevage ?"*, appliquée au cas des poules pondeuses et des vaches laitières, en fonction du nombre de fois où ils ont visité des élevages 11

1.3 Mode de fonctionnement et interactions des instances européennes en charge de la protection animale (adapté d'après Guéméné et Faure 2004) 17

1.4 Évolution de la quantité de viande importée et exportée par l'Union Européenne, de 1995 à 2005 (Source : banque de données de la FAO[2]) . 24

1.5 Le bien-être animal : au centre de nombreux questionnements 28

1.6 Liste des pays partenaires du projet Welfare Quality® ainsi que des organismes de recherche impliqués dans le WP2.3 (les organismes coordinateurs du projet apparaissent en gras) 33

1.7 Organisation de Welfare Quality® : les quatre sous-projets de recherche 37

1.8 Schéma général du système visé par le projet Welfare Quality® - modifié d'après Veissier et al. (2005) 38

1.9 Organisation du Sous-projet 2 : Évaluation du bien-être 40

4.1 Système d'agrégation progressive de l'information collectée en ferme . . 104

5.1 Arbre de décision utilisé pour le Sous-critère *"Absence de soif prolongée"* 121

5.2 Tracé des quatre fonctions B-splines basiques 129

5.3 Polygone de contrôle et fonction spline u_1 en résultant 130

5.4 Illustration des points de contrôle intervenant dans le calcul de l'utilité pour un index $\in [40, 60]$ 131

5.5 SC6 - *Altérations du tégument* : approximation en utilisant la fonction d'utilité u_1 (déterminée à l'aide de B-splines avec 4 noeuds intérieurs) 132

5.6 Comparaison des fonctions d'utilité obtenues avec des B- ou des I-splines, avec 1 noeud (placé à 80 ou 50) ou à 4 noeuds intérieurs, sur l'exemple du score partiel g_6^2 *"Boiteries"* 134
5.7 SC6 - *Altérations du tégument* : approximation en utilisant la fonction d'utilité u_1 (déterminée à l'aide de I-splines avec 1 noeud intérieur à 75) 136
5.8 SC6 - *Boiteries* : approximation en utilisant la fonction d'utilité u_2 (déterminée à l'aide de I-splines avec 1 noeud intérieur à 80) 138
5.9 SC3 : approximation en utilisant la fonction d'utilité u (déterminée à l'aide de I-splines avec 1 noeud intérieur à 65) 144
5.10 Distribution des fermes sur les sous-critères composant le Critère *"Alimentation adaptée"* 147
5.11 Distribution des fermes sur les sous-critères composant le Critère *"Logement correct"* 148
5.12 Distribution des fermes sur les sous-critères composant le Critère *"Bonne santé"* 149
5.13 Distribution des fermes sur les sous-critères composant le Critère *"Comportement approprié"* 150

6.1 Jeu de données présenté aux experts pour construire les fonctions f_i : exemple du Critère 2 *"Logement correct"* 153
6.2 Distribution des fermes sur les Critères 1 *"Alimentation adaptée"* et 2 *"Logement correct"* 175
6.3 Distribution des fermes sur les Critères 3 *"Bonne santé"* et 4 *"Comportement approprié"* 176

7.1 Utilisation des catégories de bien-être pour chacun des usages pouvant être fait du système d'évaluation. En fonction de l'usage qui en sera fait, la réponse pourra être binaire (ex. *S'agit-il d'une ferme excellente ou non ?*) ou graduelle (*La ferme est-elle −, +, ++ ou +++ ?*). 186
7.2 Définition des indices de concordance et de discordance partielles 190
7.3 Agrégation des critères : comparaison aux profils de référence 193
7.4 Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 1** 200
7.5 Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 2** 200
7.6 Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 3** 201

7.7 Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 4** 201
7.8 Répartition dans les quatre catégories de bien-être des 70 fermes visitées lors du WP2.4 avec la **Procédure 4** et des profils de références modifiés 202

8.1 Structure séquentielle du modèle d'évaluation 236
8.2 Symboles des points forts et faibles des fermes, utilisés dans les Figures 8.3 à 8.11 237
8.3 Scores obtenus au niveau des critères et des sous-critères par la Ferme 1 237
8.4 Scores obtenus au niveau des critères et des sous-critères par la Ferme 2 239
8.5 Scores obtenus au niveau des critères et des sous-critères par la Ferme 3 240
8.6 Scores obtenus au niveau des critères et des sous-critères par la Ferme 4 241
8.7 Scores obtenus au niveau des critères et des sous-critères par la Ferme 5 242
8.8 Scores obtenus au niveau des critères et des sous-critères par la Ferme 6 242
8.9 Scores obtenus au niveau des critères et des sous-critères par la Ferme 7 243
8.10 Scores obtenus au niveau des critères et des sous-critères par la Ferme 8 244
8.11 Scores obtenus au niveau des critères et des sous-critères par la Ferme 9 245
8.12 Visual Analogue Scale utilisée par les personnes ayant réalisé les observation dans le WP2.4 pour évaluer le niveau de bien-être qu'elles avaient perçu sur les différentes fermes 263
8.13 Les quatre piliers de la durabilité des exploitations agricoles (d'après Landais 1998) 266

A.1 Organisation des quatre sous-projets de recherche au sein de Welfare Quality® 272

Liste de mes publications

Liste de mes publications

Articles

Article 1 Botreau, R., M. Bonde, A. Butterworth, P. Perny, M. Bracke, J. Capdeville, and I. Veissier (2007). Aggregation of measures to produce an overall assessment of animal welfare : Part 1 - a review of existing methods. *Animal 1*(8), 1179-1187.

Article 2 Botreau, R., M. Bracke, P. Perny, A. Butterworth, J. Capdeville, C. van Reenen, and I. Veissier (2007). Aggregation of measures to produce an overall assessment of animal welfare : Part 2 - analysis of constraints. *Animal 1*(8), 1188-1197.

Article 3 Botreau, R., I. Veissier, A. Butterworth, M. Bracke, and L. Keeling (2007). Definition of criteria for overall assessment of animal welfare. *Animal Welfare 16*(2), 225-228.

Article 4 Botreau, R., J. Capdeville, P. Perny, and I. Veissier (soumis). Multicriteria evaluation of animal welfare at farm level : an application of MCDA methodologies. *Foundations of Computing and Decision Sciences.*

Communications orales

Oral 1 Botreau, R., M. B. M. Bracke, M. Bonde, P. Perny, A. Butterworth, L. J.Keeling, J. Capdeville, K. van Reenen, B. Engel, and I. Veissier (2005). Aggregation of parameters to produce an overall assessment of animal welfare : a review of methods and constraints. In Proceedings of the 3rd International Workshop on the Assessment of Animal Welfare at Farm and Group Level (WAFL- 05), Vienne (Autriche), 22-24 septembre 2005, p. 30.

Oral 2 Botreau ,R. (2006). Modélisation et évaluation multicritère du bien-être des animaux de ferme - Étude centrée sur les bovins. Journées de l'École Doctorale ABIES, Paris (France), 8-9 mars 2006.

Oral 3 Botreau, R., J. Capdeville, P. Perny, and I. Veissier (2006). Évaluation globale du bien-être animal en vue de son intégration dans des démarches qualité. 11èmes journées des Sciences du Muscle et des Technologies de la Viande, Clermont-Ferrand (France), 4-5 octobre 2006. *Viandes & Produits Carnés Hors série*, 49-50.

Oral 4 Botreau, R., J. Capdeville, P. Perny, and I. Veissier (2007). Multicriteria evaluation of animal welfare at farm level : an application of MCDA methodologies. In Proceedings of the 65th meeting of the European Working Group Multiple Criteria Decision Aiding (MCDA'65), Poznan (Pologne), 12-14 avril 2007, pp. 73-74.

Oral 5 Botreau, R., P. Perny, J. Capdeville, and I. Veissier (2007). Construction of product information from animal welfare assessment. In Proceedings of the Second Welfare Quality® stakeholder conference, Berlin (Allemagne), 3-4 mai 2007, pp. 33-36.

Oral 6 Botreau, R. (2007).Méthodologie d'évaluation globale du bien-être animal en vue de son intégration dans des démarches qualité. Séminaire AGRI Bien-Être Animal *"Réponses de la recherche aux interrogations sur le bien-être animal. Les enjeux de l'expertise pour la production de connaissances robustes"*, Paris (France), 4 juillet 2007.

Oral 7 Botreau, R., and I. Veissier (2007). Construction of an overall assessment of animal welfare - the case of dairy herds in Welfare Quality®. In Proceedings of the Cattle Consultancy Days 2007 (annual conference for bovine veterinary practitioners and production consultants), Nyborg (Danemark), 29-30 août 2007, pp. 33-40.

Oral 8 Botreau, R., P. Perny, J. Capdeville, and I. Veissier (2007). Évaluation multicritère du bien-être. In Proceedings des 2èmes Journées d'Animation Scientifique du Département de Physiologie Animal et Systèmes d'Élevage (INRA), Tours (France), 22-24 octobre 2007, p. 190.

Oral 9 Veissier, I., R. Botreau, J. Capdeville, and P. Perny (2007). L'évaluation en ferme du bien-être des animaux : objectifs, outils disponibles, utilisations, exemple du projet Welfare Quality®. 14èmes Rencontres Recherches Ruminants, Paris (France), 5-6 décembre 2007.

Communications affichées

Poster 1 Botreau, R., I. Veissier, J. Capdeville, M. Bracke, A. Butterworth, L. Keeling, M. Bonde, T. Rousing, G. de Rosa, K. van Reenen, and P. Perny (2006). Multicriteria assessment of farm animals' quality of life. In Proceedings of the UFAW/BVA Ethics Committee Internation Symposium "Quality of Life : the heart of the matter", Londres (Royaume-Uni), 13-14 septembre 2006. *Animal Welfare 16*(Supplement), 166.

Poster 2 Botreau, R., J. Capdeville, P. Perny, and I. Veissier (2006). Les difficultés d'une évaluation globale du bien-être animal. 13èmes Rencontres Recherches Ruminants, Paris (France), 6-7 décembre 2007. *Rencontres Recherches Ruminants 13*, 41.

dessin de Coulib

Résumé *Évaluation multicritère du bien-être animal (ex. des vaches laitières en ferme)*

Le respect du bien-être des animaux d'élevage fait l'objet d'attentes sociétales de plus en plus fortes. Dans ce contexte, le projet européen Welfare Quality® vise à construire un standard en matière d'évaluation et d'information sur le bien-être animal. Le bien-être est composé de plusieurs dimensions (santé, comportement, absence de stress...) et son évaluation doit donc reposer sur un ensemble de mesures complémentaires. L'objet de la thèse est d'élaborer un modèle d'évaluation globale du bien-être des vaches laitières reflétant l'aspect multidimensionnel du bien-être. Notre modèle suit une approche multicritère. Il repose sur une structure séquentielle dans laquelle plusieurs critères sont construits puis agrégés afin d'obtenir une évaluation globale.

Une fois identifiés les problèmes liés à la construction d'un modèle d'évaluation globale du bien-être animal, nous définissons un ensemble de 4 critères (*Alimentation adaptée*, *Logement correct*, *Bonne santé* et *Comportement approprié*) et 12 sous-critères. Pour construire les sous-critères à partir des 44 mesures effectuées en ferme, nous proposons de les évaluer sur la même échelle d'utilité [0,100]. Les données brutes sont alors agrégées en utilisant des méthodes adaptées au nombre de mesures, à leur nature et à leur importance relative (sommes pondérées associées à une fonction d'utilité, méthode lexicographique...). Les sous-critères composant un même critère sont ensuite agrégés en utilisant un méthode permettant de tenir compte du fait que certains sous-critères sont plus importants que d'autres, tout en limitant les compensations entre les sous-critères (intégrale de Choquet). Enfin, l'agrégation des critères pour former une évaluation globale utilise une méthode de comparaison à des profils prédéfinis délimitant quatre catégories de bien-être (correspondant à une problématique de tri ordonné). A chaque étape, le modèle d'évaluation est paramétré sur la base d'avis d'experts, ces experts étant des scientifiques ou des utilisateurs futurs du système d'évaluation.

Le système d'évaluation du bien-être qui résultera du projet Welfare Quality® (en 2009) pourra être utilisé à plusieurs fins : étiquetage obligatoire européen, label "bien-être", conseil aux éleveurs recherche (la capacité du modèle à pouvoir expliquer les résultats étant très utile dans ces deux derniers cas).

Mots-clés : *Bien-être animal, Évaluation multicritère, Aide à la décision, Bovins, Élevage*

Abstract *Multicriterion evaluation of animal welfare (example of on-farm dairy cows)*

There is an increasing societal concern for the welfare of farm animals. The European project Welfare Quality® aims at developing a standard for the assessment of and information about animal welfare. Welfare is composed of several dimensions (health, behaviour, absence of stress...). Hence, its overall evaluation is based on a set of complementary measures. The objective of this PhD thesis is to develop a model for the overall assessment of the welfare of dairy cows, which respects the multidimensionality of the welfare concept. Our model follows a multicriterion approach. It follows a sequential evaluation process where several criteria are constructed and aggregated into an overall assessment.

After identifying the constraints imposed on the construction of an evaluation model of overall welfare, we define a set of 4 criteria (*Good feeding*, *Good housing*, *Good health* and *Appropriate behaviour*) and 12 subcriteria. To construct the subcriteria from the 44 measures observed on farms, we propose to assess them on a common utility scale [0,100]. Raw data are thus aggregated using methods adapted to the number of measures, their nature and their relative importance (weighted sums associated to a utility function, decision trees...). Subcriteria involved in the same criterion are the aggregated using a method which allows to consider that subcriteria are more or less important, while limiting compensations between subcriteria (Choquet integral). Finally, the aggregation of criteria into an overall assessment is based on a comparison to predefined profiles delimiting four categories of welfare (corresponding to an ordered sorting procedure). At each stage, the evaluation model is parameterized on the basis of experts' opinions. The experts are scientists or potential users of the future evaluation system).

The evaluation system that will be produced by Welfare Quality® (in 2009) can be used for several purposes : a compulsory European labelling system, a "welfare-friendly" labelling system, advice to farmers or research (with the capacity of the model to explain the results being adequate for these two later purposes).

Keywords : *Animal welfare, Multicriterion evaluation, Decision-aid, Cattle, Farming*