Modelisation and estimation of heterogeneous variances in nonlinear mixed models

Mylene Duval

UFR Génétique,
Elevage et Reproduction

Département
Génétique Animale

N°/__/__/__/__/__/__/__/__/__/__/

Thèse

Pour obtenir le grade de

DOCTEUR D'AGROPARISTECH

Discipline : Mathématiques appliquées à la génétique animale

présentée et soutenue publiquement par

Mylène DUVAL

Le 8 décembre 2008

MODELISATION ET ESTIMATION DE VARIANCES HETEROGENES DANS LES MODELES NON LINEAIRES MIXTES

Directeur de thèse : Jean-Louis FOULLEY
Co-directrice de thèse : Christèle ROBERT-GRANIE

Jury

Etienne VERRIER	Professeur, AgroParisTech	Président du Jury
Didier CONCORDET	Professeur, Ecole Nat[le] Vétérinaire, Toulouse	Rapporteur
Eric PARENT	Ingénieur, ENGREF	Rapporteur
Adeline SAMSON	Maître de Conférences, Paris-V	Examinateur
Christèle ROBERT-GRANIE	Directrice de Recherches, INRA, Toulouse	Examinateur
Jean-Louis FOULLEY	Directeur de Recherches, INRA, Jouy-en-Josas	Examinateur

Avant-Propos

Cette thèse a été réalisée au sein de la Station d'Amélioration Génétique des Animaux de l'INRA de Castanet-Tolosan. Je remercie Mr Didier BOICHARD, chef du Département de Génétique Animale, Mrs Eduardo MANFREDI et Alain DUCOS, ancien et actuel directeurs de la Station, Mr Jean-Pierre BIDANEL, directeur de la Station de Génétique Quantitative et Appliquée, pour les moyens qu'ils ont mis à ma disposition afin de réaliser à bien ce travail.

Je remercie les membres du jury, en particulier Etienne VERRIER qui m'a fait l'honneur d'être le Président du Jury, Didier CONCORDET et Eric PARENT, pour avoir accepté de rapporter ce travail et pour l'intérêt qu'ils y ont porté, Adeline SAMSON pour sa participation à ce jury.

Je tiens tout particulièrement à remercier Jean-Louis FOULLEY et Christèle ROBERT-GRANIE qui ont accepté de m'encadrer tout au long de ma thèse. Leur patience, leurs conseils et leurs compétences scientifiques m'ont aidée à progresser et m'ont guidée durant ces trois années. Encore merci.

Merci à l'ensemble de mes collègues enseignants du département de Génie Mathématiques de l'INSA de Toulouse, qui m'ont aidée à faire mes premiers pas en tant que professeur. Je remercie sincèrement Béatrice LAURENT-BONNEAU pour son soutien et ses conseils dans mon choix de carrière professionnelle.

Je souhaite maintenant remercier tous les membres de la Station d'Amélioration Génétique des Animaux de l'INRA de Castanet-Tolosan. Ils m'ont accueillie chaleureusement et chaque jour, ce fut avec plaisir que je retrouvais toute l'équipe. Merci aux "grands animateurs" de la pause café : Isabelle, Roger et tous les autres que je n'oublie pas mais que je ne citerai pas pour n'omettre personne.

Mes remerciements vont également aux électroniciens et informaticiens, qui ont toujours été disponibles pour m'aider à régler les tracasseries informatiques, et ont fortement contribué à la bonne humeur de notre couloir, merci beaucoup ! !

Je suis infiniment reconnaissante à Virginie et Isabelle, qui ont partagé avec moi les bons comme les mauvais moments de la thèse. Elles ont su m'écouter, m'épauler et me communiquer toute l'énergie nécessaire pour avancer dans les moments difficiles. Ce fut un réel plaisir de passer ces trois années à leurs côtés.

Merci aussi aux trois autres mousquetaires : Anne, Bérengère et Françoise, pour les discussions “entre filles”, au cours de nos petits repas du midi. Je suis reconnaissante aussi à ma copine thèsarde, Kim-Anh, avec laquelle j’ai partagé mes problèmes de doctorante et d’enseignante, elle était un réel soutien pour moi.

C’est maintenant à mes parents et mon frère que j’adresse toute mon affection et ma reconnaissance pour m’avoir encouragée et aidée à en arriver là.
Et enfin, à celui qui partage ma vie depuis quelques années, c’est bien plus qu’un merci que j’adresse...

Cette thèse a bénéficié d'une bourse du Ministère de la Recherche attribuée par l'Ecole Doctorale ABIES.
Elle a été réalisée au sein de la Station d'Amélioration Génétique des Animaux (SAGA) de l'INRA de Castanet-Tolosan.

INRA SAGA
Chemin de Borde Rouge
BP 52627
31 326 Castanet-Tolosan Cedex

Résumé

Les modèles non linéaires occupent une place à part dans la méthodologie des modèles mixtes. Contrairement aux modèles linéaire et linéaire généralisé qui s'apparentent souvent à des boites noires, la fonction d'ajustement des données dans le cas non linéaire provient en général de l'intégration d'une équation différentielle ce qui confère à ces modèles une dimension "explicative" beaucoup plus riche et souvent plus parcimonieuse. D'autre part, l'estimation des paramètres y est difficile du fait de l'impossibilité d'une intégration analytique des effets aléatoires. Comme dans tous les modèles mixtes notamment ceux appliqués aux données longitudinales, ils permettent bien de prendre en compte la variabilité entre et intra unités expérimentales. Mais, là comme ailleurs, le statut des résidus supposés habituellement indépendants et identiquement distribués suivant une loi normale de variance homogène reste problématique car fréquemment irréaliste.
L'objet de ce travail était de présenter quelques possibilités de modélisation de ces variances résiduelles qui prennent en compte la grande hétérogénéité potentielle de celles-ci, mais dans un souci délibéré d'économie vis-à-vis du nombre de nouveaux paramètres impliqués dans ces fonctions. C'est pourquoi, en sus de la relation classique moyenne-variance, nous avons opté pour une approche paramétrique de type "modèle mixte" sur les logvariances. Nous avons choisi une méthode d'inférence classique basée sur la théorie du maximum de vraisemblance et, dans ce cadre complexe, nous avons considéré un algorithme de type EM stochastique plus précisément l'algorithme dit SAEM-MCMC. La structure de modèle mixte à la fois sur les paramètres de position et de dispersion se prête particulièrement bien à la mise en oeuvre de ces algorithmes EM. La phase MCMC, a nécessité la mise au point et le calibrage de distributions instrumentales adaptées à cette situation ainsi que la définition de critères permettant de contrôler la convergence de l'algorithme. Le tout a été validé numériquement dans le cadre linéaire et non linéaire par comparaison à des algorithmes EM analytiques quand ils existaient (cas linéaire) ou à d'autres algorithmes numériques tels ceux basés sur la quadrature de Gauss.
Ces techniques ont été illustrées par l'analyse de profils de comptage de cellules somatiques de vaches laitières. Plusieurs modèles linéaire et non linéaires sont comparés et montrent clairement l'intérêt d'une modélisation mixte des variances résiduelles.

Mots clés : modèles mixtes non linéaires, hétéroscédasticité, vraisemblance, algorithme SAEM.

Abstract
Modelisation and estimation of heterogeneous variances in nonlinear mixed models

Nonlinear mixed models stand apart in mixed models methodology. Contrary to linear and generalized linear models, often used as black boxes, the trajectory function fit in nonlinear models generally comes from the integration of differential equations. This provides a biological interpretation for the parameters, whereas models are often more parsimonious. However, the estimation of the parameters in nonlinear mixed models is complex because random effects cannot be integrated out of the likelihood in closed form. As in all mixed models, especially those used to analyze longitudinal data, nonlinear models are well adapted to take into account between and within-cluster variation. However, one of the common assumptions of such models is that of independent, identically distributed residuals with a common variance; this assumption is unrealistic in many fields of applications.

The objective of this study was to propose some models for the residual variance to take into account the potential heterogeneity of variance of the residuals, while limiting the number of parameters in these models. In this sense, we used a parametric approach based on a linear mixed model on the logvariance, as well as the classical power "mean-variance" function.

A classical inference method based on maximum likelihood theory was selected and we considered a stochastic EM algorithm, the SAEM-MCMC algorithm. The mixed model structure applied to the position and dispersion parameters is well adapted to the implementation of EM algorithms. Some instrumental distributions adapted to the analysis of these models, as well as some convergence criteria, were proposed in the MCMC step. The overall algorithm was numerically validated in both linear and nonlinear models, by comparing its results with those of an analytical EM algorithm (in the linear case) or other algorithms like those based on Gaussian quadrature.

Finally, an application to the analysis of somatic cell scores in dairy cattle was presented. Several linear and nonlinear models were compared showing a clear gain obtained taking into account the heterogeneity of variances.

Key-words: nonlinear mixed models, heteroskedasticity, likelihood, SAEM algorithm.

Table des matières

Liste des tableaux

Table des figures

Introduction générale

Pour mieux comprendre les phénomènes naturels comme ceux intervenant par exemple en chimie, physique et biologie, on essaie d'analyser les résultats en construisant des modèles, élaborés à partir de nos connaissances du phénomène ou bien à partir de l'approche théorique que l'on en a.
En biologie comme dans le domaine médical, l'étude de ces phénomènes passe généralement par des observations (par exemple la taille d'une portée en biologie animale, des prélèvements sanguins dans le domaine médical) répétées sur chacun des individus d'une même population, ordonnées dans le temps ou l'espace.

Deux sources de variabilité apparaissent sur les jeux de données repétées : la variabilité entre les observations mesurées sur un même individu et la variabilité entre les individus eux-mêmes. Le modèle mixte est un outil statistique permettant de mettre en évidence une relation entre la variable (la réponse observée) et des covariables explicatives, en prenant en compte ces deux types de variations (Laird et Ware, 1982).
On distingue généralement deux types de modèles mixtes : les modèles dont la fonction espérance est linéaire en les paramètres et les modèles dans lesquels la fonction de régression est non linéaire.

Les modèles linéaires mixtes peuvent s'écrire de la manière générale suivante :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

où $\mathbf{y}$ correspond au vecteur des observations, $\mathbf{X}$ et $\mathbf{Z}$ des matrices d'incidence ou de design, formées de 1 et de 0, $\boldsymbol{\beta}$ le vecteur des effets fixes du modèle, $\mathbf{u}$ un vecteur aléatoire Gaussien. L'erreur résiduelle est modélisée par $\mathbf{e}$, vecteur aléatoire supposé Gaussien.
Ce modèle est dit linéaire car la fonction moyenne modélisée par la partie $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ est linéaire en $\boldsymbol{\beta}$ et $\mathbf{u}$.
Ces modèles ont eu beaucoup de champs d'application (Robinson, 1991), en particulier dans le domaine de la génétique quantitative animale. Henderson a eu un rôle pionnier dans l'étude des modèles linéaires mixtes en proposant à la fois des méthodes opérationnelles d'estimation des composantes de la variance soit quadratiques (Henderson, 1953) soit de maximum de vraisemblance (Henderson, 1973) et en développant la théorie des meilleurs prédicteurs linéaires tel que le BLUP (Henderson et al., 1959).
Cependant, de nombreux caractères comme la taille, observée au cours du temps, ne

peuvent être modélisés de manière linéaire avec satisfaction. Dans ce cas, il est plus judicieux d'utiliser un modèle non linéaire mixte : le nombre de paramètres à estimer dans le modèle sera moins important et l'interprétation biologique des paramètres sera plus réaliste.
Les modèles non linéaires mixtes (Sheiner et al., 1972) ne diffèrent des modèles linéaires mixtes que par la non-linéarité de la fonction moyenne par rapport à $\boldsymbol{\beta}$ et $\mathbf{u}$. Ils sont utilisés dans maints champs d'application : épidémiologie, pharmacocinétique, courbes de croissance, etc..

Lorsqu'une population se divise en sous-groupes (classes sociales d'individus, races d'animaux), il peut être intéressant d'étudier les variations de comportements entre ces groupes. Lorsque les comportements seront identiques, on parlera de groupes homogènes, et dans le cas contraire de groupes hétérogènes. Les facteurs responsables de l'hétérogénéité d'un caractère peuvent être multiples : le sexe, l'âge, la saison, le troupeau...
La littérature recense l'existence de variances hétérogènes dans différents champs de statistiques appliquées : en biostatistiques (Pinheiro et Bates, 2000), en économie (Judge et al, 1985 ; Engle, 1982).
Lors de la prise en compte des différents facteurs d'hétérogénéité, le nombre de cellules de combinaison des facteurs peut devenir considérable et par conséquent les variances sont inestimables. Il devient donc impératif de les modéliser. Dans ce sens, plusieurs modélisations des variances hétérogènes résiduelles ont été proposées : une dépendance exponentielle de la moyenne par la variance (Box et Hill, 1974) et un modèle linéaire structural sur la log-variance comme on le fait généralement sur la moyenne (Davidian et Carroll, 1987).

Dans ce travail, nous nous sommes intéressée à la modélisation et à l'estimation des variances hétérogènes résiduelles dans les modèles non-linéaires mixtes (NLMM).

Diverses méthodes ont été proposées pour l'estimation des paramètres dans les modèles non-linéaires mixtes. Les premières méthodes apparues dans les années 80 étaient basées sur la linéarisation au premier ordre de la vraisemblance des observations, au voisinage soit d'une valeur nulle des effets aléatoires (méthode dite FO, Sheiner et Beal, 1980), soit de l'espérance conditionnelle de ceux-ci (méthode dite FOCE, Lindstrom et Bates, 1990). Comme les erreurs d'approximation peuvent être grandes (Davidian et Giltinian, 1995 ; Pinheiro et Bates, 1995 ; Lindstrom et Bates, 1990), des méthodes exactes basées sur la théorie du maximum de vraisemblance (ML) ont été étudiées. Ces méthodes nécessitent l'intégration des effets aléatoires pour obtenir la densité des observations. Deux voies principales ont été adoptées à cet égard : soit des méthodes de quadrature de Gauss, soit des méthodes passant par une intégration indirecte (algorithme EM) avec l'appui de techniques stochastiques (Celeux et Diebolt, 1985 ; Wei et Tanner, 1990 ; Delyon et al., 1999).

L'objectif de cette thèse était de modéliser des fonctions de variances résiduelles dans les modèles non linéaires mixtes, et d'adapter l'algorithme SAEM (Delyon et al., 1999) à

l'estimation des paramètres de ces fonctions.

Le chapitre 1 est consacré à une étude bibliographique. Quelques applications des modèles non linéaires dans différents champs d'application ainsi que des rappels sur les principales méthodes d'estimation dans ces modèles seront présentés. Puis nous commenterons les différentes modélisations et prises en compte des variances hétérogènes résiduelles présentes dans la littérature depuis les années 80.

Les chapitres 2 et 3 concernent la phase de développement et de validation de notre méthode d'estimation. Dans le chapitre 2, nous présenterons un article actuellement soumis à "Computational Statistics". Des critères y sont proposés pour fixer les paramètres de l'algorithme SAEM-MCMC, afin d'estimer de manière précise l'estimateur du maximum de vraisemblance.
L'article présenté dans le chapitre 3, soumis à "Statistical Modelling", propose une modélisation des variances hétérogènes dans les modèles non linéaires mixtes ainsi qu'une adaptation de l'algorithme SAEM-MCMC à l'estimation de ces variances. Notre méthode d'estimation dans ce cadre est validée et comparée à des méthodes généralement utilisées dans ce domaine.

Le chapitre 4 de ce manuscrit s'intéresse à modéliser l'évolution des scores de cellules somatiques chez les bovins laitiers, en fonction du stade de lactation. Plusieurs fonctions linéaires et non linéaires sont proposées pour modéliser la moyenne puis des variances hétérogènes sont introduites dans les différents modèles. Nous montrons entre autres que les variances hétérogènes améliorent considérablement le modèle, en termes de critères AIC et BIC.

Une discussion générale fait l'objet du dernier chapitre.

Chapitre 1

Modèles non linéaires mixtes et variances hétérogènes : éléments bibliographiques

1.1 Les modèles non linéaires mixtes

1.1.1 Les applications

Depuis plusieurs années, on préfère modéliser l'espérance du modèle en tenant compte des connaissances que l'on a du phénomène et des objectifs de l'analyse. Par exemple dans les études de cinétique enzymatique, de nombreuses relations chimiques connues et validées peuvent conduire de façon naturelle à l'élaboration d'un modèle. De même, les schémas à compartiments reconstituant les différents échanges dans l'organisme, peuvent être traduits par des systèmes d'équations différentielles.
En biologie, les mécanismes mis en jeu sont souvent mal connus et trop nombreux. On choisit donc, à partir de l'allure du phénomène étudié, une famille de courbes régies par des équations les plus simples possible et qui retracent bien le phénomène.
Les paramètres d'un modèle non-linéaire sont biologiquement ou physiquement interprétables. De plus, pour la modélisation d'un même phénomène, un modèle non linéaire comptera généralement moins de paramètres qu'un modèle linéaire. Il n'est pas conseillé de choisir une fonction avec beaucoup de paramètres dans le souci d'expliquer finement le phénomène. En effet, le nombre de paramètres à estimer risque d'être trop grand et la précision des estimateurs amoindrie.

Dans ce paragraphe, nous distinguerons différentes situations dans lesquelles nous sommes amenés à modéliser le phénomène étudié à l'aide d'une fonction non linéaire. Quelques exemples appliqués (Huet et al., 1992), dont certains seront repris dans les chapitres suivants, étayerons ces situations.

Choix du modèle guidé par la théorie

Dans certains domaines, une étude précise et localisée des phénomènes peut aboutir à l'élaboration d'un système d'équations différentielles dont les conditions initiales sont fournies par l'expérience. C'est en particulier le cas des modèles à compartiments, très utilisés en pharmacocinétique, science étudiant le devenir d'un médicament dans l'organisme au cours du temps, en terme de processus d'absorption, de distribution, de métabolisme et d'élimination. Les compartiments du modèle correspondent aux différents organes étudiés et au plasma, le contenu des compartiments étant décrit en fonction des échanges possibles à l'aide d'équations différentielles.

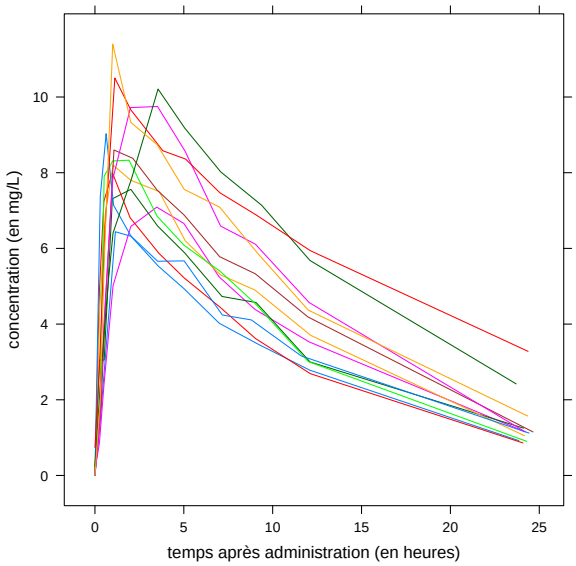


FIGURE 1.1 – Concentration sanguine en théophylline chez 12 patients après administration orale d'une même dose

La figure 1.1 représente un jeu de données typique d'une étude en pharmacocinétique (Pinheiro et Bates, 2000 ; Davidian et Giltinan, 2003). Il s'agit d'une expérience dans laquelle une même dose de théophylline (antiasmathique) a été administrée à 12 patients. Des prélèvements sanguins ont été effectués à différents temps afin de mesurer la concentration de théophylline dans l'organisme. Les profils observés pour chaque individu ont la même tendance : la concentration sanguine en théophylline augmente très vite pour atteindre un pic avant 5 heures puis diminue progressivement. Pour modéliser ce type de données, Pinheiro et Bates (2000) proposent la fonction non linéaire suivante, basée sur un modèle à un compartiment et solution du système différentiel correspondant :

$$C(t) = \frac{Dk_e k_a}{Cl(k_a - k_e)}[\exp(-k_e t) - \exp(-k_a t)]$$

où $C(t)$ correspond à la concentration en théophylline au temps t pour un individu auquel on a administré la dose D au temps $t = 0$. Le paramètre k_a correspond au taux d'absorption de la théophylline, k_e au temps d'élimination et Cl la clairance, définie comme le volume sanguin totalement débarrassé de la théophylline par unité de temps.
Le vecteur des paramètres $\boldsymbol{\phi} = (k_e, k_a, Cl)$ traduit tout le phénomène observé pour chaque individu.
Une telle analyse a pour but d'obtenir une valeur moyenne de $\boldsymbol{\phi}$ et de déterminer les facteurs de variation des paramètres entre les individus. Si les paramètres varient en fonction de l'âge, du poids, ou tout autre facteur, il devient possible d'adapter un régime commun aux patients dont les caractéristiques sont similaires.

Les équations des courbes de croissances, utilisées pour décrire l'évolution de la taille ou du poids par exemple, peuvent elles aussi être obtenues comme solution de systèmes différentiels. Elles sont utiles pour décrire des phénomènes de croissance dans de nombreux domaines : la physiologie (ex : évolution de la taille au cours du temps), la nutrition, la génétique (prise en compte de l'hérédité de certains caractères de croissance) et la bactériologie (ex : l'évolution d'une famille de bactéries au cours du temps). Plaçons-nous, par exemple, dans le domaine de l'aménagement de la forêt.
La figure 1.2 représente une étude effectuée sur cinq orangers (Pinheiro et Bates, 2000),

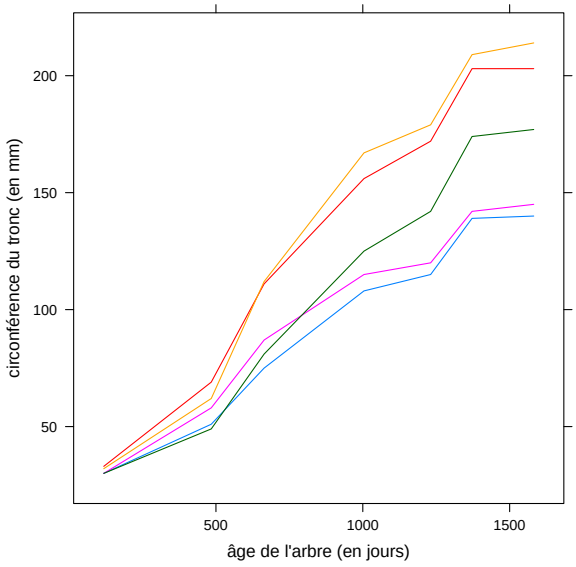


FIGURE 1.2 – Circonférence du tronc de 5 orangers

pour lesquels on a mesuré la circonférence du tronc à sept reprises au cours des cinq premières années. Il s'agit d'un cas typique d'étude de courbes de croissance. Plusieurs types de fonctions non linéaires ont été proposés pour modéliser ce genre de profils, nous en présentons deux :

– Le modèle de croissance dit monomoléculaire ou de Mitscherlich :

$$C(t) = \phi_1 - \phi_2 \exp(-\phi_3 t)$$

où $C(t)$ correspond à la circonférence du tronc à l'âge t. On peut interpréter biologiquement les paramètres du modèle de la manière suivante : ϕ_1 est la circonférence asymptotique de l'arbre, ϕ_2 la différence entre la circonférence ϕ_1 et la circonférence mesurée au début de l'étude, i.e. $\phi_2 = \phi_1 - C(0)$, et ϕ_3 est un paramètre caractérisant la vitesse de croissance $C'(t)$ rapportée à l'écart entre la circonférence actuelle et la circonférence adulte de l'arbre, i.e. $\phi_3 = C'(t)/\big(\phi_1 - C(t)\big)$.
Cette fonction de croissance particulièrement simple découle directement de l'équation différentielle :

$$C'(t) \propto \alpha - C(t)$$

où α est la circonférence asymptotique, et cette équation se fonde sur l'hypothèse que la vitesse de croissance instantanée est proportionnelle non pas à la circonférence actuelle $C(t)$ (ce qui conduirait à une croissance infinie) mais à la croissance $\alpha - C(t)$ qui reste à accomplir.

– Le modèle de Gompertz (Pinheiro et Bates, 2000) :

$$C(t) = \phi_1 \exp\big[-\phi_2 \exp(-\phi_3 t)\big]$$

où ϕ_1 correspond à la circonférence asymptotique du tronc, $\phi_2 = \log\big(\phi_1/C(0)\big)$ est un paramètre lié aux conditions initiales ; quant à ϕ_3 qualifié de degré de maturité, il représente la vitesse de croissance relative $C'(t)/C(t)$ au point d'inflexion t^* de la courbe.

Bien modéliser la croissance des arbres en fonction de certains facteurs de variation comme la composition du sol ou la région de culture est un atout essentiel à l'aménagement des forêts.

Dans certains cas, les systèmes différentiels retraçant les phénomènes physiques n'ont pas de solution analytique (Donnet et Samson, 2007). On peut alors approximer la fonction de régression à l'aide d'une solution numérique approchée du système, de forme non linéaire.

Choix du modèle guidé par l'allure du phénomène

En biologie, les mécanismes mis en jeu au cours de l'élaboration des phénomènes sont très souvent peu connus, on choisit donc une fonction retraçant le mieux possible ces phénomènes. Etudions un exemple appartenant au domaine de la pathologie animale : l'étude des inflammations de la mamelle chez les ruminants laitiers. Les mammites sont des inflammations de la glande mammaire, essentiellement d'origine infectieuse, qui se

traduisent par un afflux massif de leucocytes (globules blancs) du sang vers la mamelle. Ces infections ont des conséquences économiques importantes pour l'éleveur causées par la perte de production laitière et les traitements vétérinaires coûteux.
Les moyens directs utilisés pour déceler les mammites étant coûteux, on s'intéresse plutôt à l'étude d'un critère corrélé au diagnostic des mammites : le comptage de cellules somatiques (CCS). La concentration en cellules somatiques dans le lait traduit généralement l'état général de la mammelle et donc indirectement son état infectieux (Rupp, 2000).
La figure 1.3 présente le profil moyen des logarithmes des CCS au cours d'une lactation, obtenu lors d'une analyse effectuée sur 159 génisses du domaine expérimental INRA-Le pin au Haras (Robert-Granié et al., 2004). Une étude plus approfondie de ce jeu de données fait l'objet du chapitre 4.
Rodriguez-Zas et al. (2000) ont présenté quelques fonctions non linéaires adaptées à la

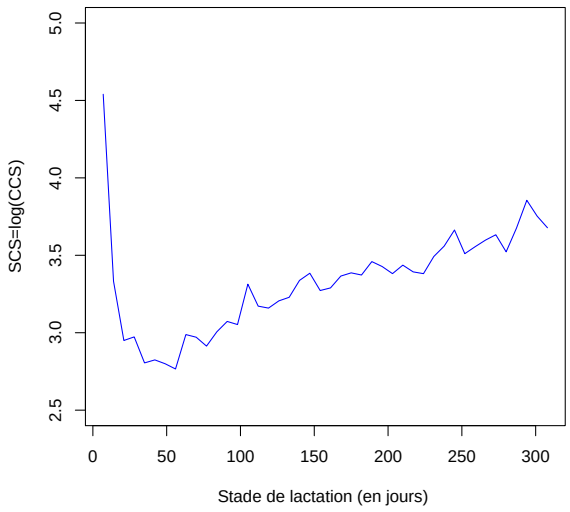


FIGURE 1.3 – Profil moyen des logarithmes des comptages de cellules somatiques au cours d'une lactation, lors d'une étude sur des bovins laitiers.

modélisation des profils des logarithmes des comptages de cellules somatiques au cours d'une lactation. Une bonne connaissance de la variation des CCS entre les animaux, suivant la saison, la lactation, l'âge de la vache et autres facteurs explicatifs sera utile à la prise en compte du caractère "résistance aux mammites" dans le programme de sélection des ruminants laitiers.

En résumé, nous avons présenté dans ce paragraphe des situations dans lesquelles pour chaque individu, les données observées (par exemple : concentration en théophylline, circonférence du tronc) étaient répétées et continues en fonction du temps, l'allure du phénomène étudié étant non linéaire.

1.1.2 Les modèles

Spécification du modèle

Dans les exemples précédents, plusieurs points communs aux jeux de données peuvent être identifiés :
i) La réponse $\mathbf{y}$ est constituée de données répétées (par exemple : concentration en théophylline, circonférence du tronc) mesurées sur chaque individu d'une population (par exemple : humain, arbre).
ii) Pour chaque individu, les observations dépendent non linéairement d'un ensemble de paramètres $\boldsymbol{\phi}$.
iii) Les profils individuels ont la même forme de courbe mais la valeur du vecteur $\boldsymbol{\phi}$ peut différer d'un individu à l'autre.

La structure des jeux de données nous conduit naturellement à construire un modèle hiérarchique à deux niveaux : un premier niveau dans lequel on explicite les structures de moyenne et variance résiduelle, et un deuxième niveau dans lequel on caractérise la variation entre les individus. Ces modèles ont été proposés tout d'abord par Sheiner et al. (1972) puis Sheiner et Beal (1980) dans le but de prendre en compte la variation entre individus.
Dans cette partie, nous rappellerons la typologie de ces modèles, en présentant tout d'abord les modèles à effets fixes homogènes, dans lesquels la variation entre les individus n'est pas prise en compte.

Modèles non linéaires à effets fixes homogènes

On note y_{ij} la $j$ème observation (concentration en théophylline, circonférence du tronc de l'arbre...) mesurée sur l'individu i, $i = 1, ..., N$, sous la condition t_{ij}, $j = 1, ..., n_i$. Dans beaucoup d'applications t_{ij} correspond au temps auquel a été effectuée la mesure.

On peut considérer un modèle basique pour étudier les situations présentées dans le paragraphe précédent.

$$y_{ij} = f(\mathbf{z}_{ij}, \boldsymbol{\beta}) + \sigma\varepsilon^*_{ij}, \tag{1.1}$$

$\forall i \in \{1, ..., N\}\ \forall j \in \{1, .., n_i\}$.

f est ici une fonction non linéaire représentant la fonction moyenne, comme $C(t)$ dans les exemples précédents, dépendante du vecteur de paramètres $\boldsymbol{\beta}(m \times 1)$ et des variables de régression $\mathbf{z}_{ij}$. Dans les exemples d'application présentés précédemment, $\mathbf{z}_{ij}$ était le temps t_{ij} et $\boldsymbol{\beta}$ correspondait au vecteur des paramètres (k_a, k_e, Cl) pour l'application en pharmacocinétique et (ϕ_1, ϕ_2, ϕ_3) pour l'application aux orangers.
La fonction $f(\mathbf{z}_{ij}, \boldsymbol{\beta})$ représente ce qui se passe "en moyenne" pour l'individu i. En réalité, de nombreux processus biologiques et physiques ne peuvent être pris en compte dans la

modélisation de la moyenne (en particulier lorsqu'on explique un phénomène à l'aide d'un modèle à compartiments), les observations vont donc avoir tendance à fluctuer autour de cette moyenne. Cette variation est prise en compte dans les erreurs résiduelles du modèle, notées $\sigma\varepsilon_{ij}^*$, supposées indépendantes identiquement distribuées (i.i.d.) Gaussiennes centrées et de variance σ^2. Dans ce modèle, la variance résiduelle est supposée constante et donc ne dépend pas des individus.
Le choix de la loi Gaussienne dans la modélisation des erreurs résiduelles peut être expliqué par le fait que la somme d'aléas faiblement dépendants les uns des autres, issus de plusieurs sources d'écart à la fonction moyenne f, peut être considérée Gaussienne grâce au théorème central limite.

Le premier inconvénient lié à ce modèle est l'absence de variation inter-individuelle, due par exemple à des caractéristiques différentes de sous-population (par exemple l'âge, le sexe) ou à des phénomènes non expliqués (des processus biologiques et physiques). On peut prendre en compte cette variation en ajoutant un deuxième niveau au modèle (1.1), consacré à la spécification du vecteur $\boldsymbol{\beta}$.

Modèles non linéaires à effets mixtes homogènes

Le premier modèle non linéaire à effets mixtes a été défini par Lindstrom et Bates (1990), puis a été repris largement par de nombreux auteurs. Il généralise le modèle linéaire mixte de Laird et Ware (1982) et le modèle (1.1). Il peut s'écrire de la manière suivante :

$$y_{ij} = f(\mathbf{z}_{ij}, \boldsymbol{\phi}_i) + \sigma\varepsilon_{ij}^*, \tag{1.2}$$

$\forall i \in \{1, ..., N\},\ \forall j \in \{1, .., n_i\}$.

Les erreurs intra-individuelle, modélisées par $\sigma\varepsilon_{ij}^*$, sont i.i.d. Gaussiennes centrées de variance σ^2. La fonction moyenne f dépend d'un paramètre individuel aléatoire $\boldsymbol{\phi}_i(m \times 1)$, de covariables $\mathbf{z}_{ij}$. On modélise $\boldsymbol{\phi}_i$ par l'équation suivante :

$$\boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i \tag{1.3}$$

avec $\mathbf{b}_i$ un vecteur Gaussien centré, de matrice de covariance $\mathbf{D}$ et $\boldsymbol{\beta}(S \times 1)$ un vecteur "moyenne" (ou vecteur d'effets fixes) inconnu.
La matrice $\mathbf{A}_i(m \times S)$ est une matrice de design supposée connue. Par exemple $\mathbf{A}_i$ peut-être une matrice composée de 0 ou de 1, prenant en compte les caractéristiques individuelles de l'individu i, et $\boldsymbol{\beta}$ les différents niveaux possibles d'un caractère.
Dans l'exemple de pharmacocinétique, la masse corporelle w_i de chaque individu est renseignée dans la base de données. Il est alors possible de modéliser la relation entre les paramètres $\boldsymbol{\phi}_i = (k_e, k_a, Cl)_i$ et la masse w_i par un modèle linéaire, en utilisant une matrice de design de taille 3×6 :

$$\mathbf{A}_i = \begin{pmatrix} 1 & w_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & w_i & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & w_i \end{pmatrix}$$

et un vecteur de paramètres de taille 6×1 :

$$\boldsymbol{\beta} = (\mu_{0ke}, \mu_{1ke}, \mu_{0ka}, \mu_{1ka}, \mu_{0Cl}, \mu_{1Cl})'$$

où $\mu_{0ke}, \mu_{0ka}, \mu_{0Cl}$ sont les intercepts correspondants à chaque paramètre k_e, k_a, et Cl, et $\mu_{1ke}, \mu_{1ka}, \mu_{1Cl}$ sont les pentes correspondantes.

La matrice $\boldsymbol{B}_i$ est une matrice de design, qui permet par exemple de différencier les effets fixes des effets aléatoires dans le vecteur $\boldsymbol{\phi}_i$. Dans l'exemple précédent, supposons que k_e soit un effet fixe et les paramètres k_a et k_l soient aléatoires, alors on peut écrire :

$$\mathbf{B}_i = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

et $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})'$.

Avec les hypothèses précédentes, on obtient le système suivant :

$$\begin{cases} k_{ei} = \mu_{0ke} + \mu_{1ke} w_i \\ k_{ai} = \mu_{0ka} + \mu_{1ka} w_i + b_{i1} \\ Cl_i = \mu_{0Cl} + \mu_{1Cl} w_i + b_{i2} \end{cases}$$

Dans cet exemple, on a supposé que le paramètre k_e n'était pas aléatoire et variait seulement en fonction du poids, c'est-à-dire que deux personnes de même poids obtenaient une valeur identique pour k_e. Or dans la réalité, il n'est en général pas plausible de penser que certains paramètres physiologiques restent fixes et ne varient pas entre des individus de mêmes caractéristiques, car tous les mécanismes physiologiques existants sont rarement connus.
Néanmoins, nous sommes obligés de laisser fixes certains paramètres, comme par exemple des paramètres de puissance dans les équations, afin que les méthodes d'estimation soient numériquement stables.

En dépit de sa simplicité, l'écriture des effets aléatoires sous la forme $\mathbf{B}_i\mathbf{b}_i$ avec une matrice d'indice $\mathbf{B}_i$ dans laquelle certaines lignes peuvent prendre des valeurs nulles peut conduire à des problèmes d'estimation. De plus, les coefficients sans facteur individuel aléatoire nécessitent un traitement inférentiel spécifique.
Dans notre étude, nous avons utilisé une autre approche. Contrairement au modèle (1.2),

nous séparons directement les effets fixes des effets aléatoires dans la fonction moyenne. Voici la modélisation générale des modèles non linéaires mixtes homogènes qui sera utilisée dans la suite de ce manuscrit :

$$y_{ij} = f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i) + \sigma\varepsilon^*_{ij} \tag{1.4}$$

où y_{ij} est la $j$ème observation, $j \in \{1, .., n_i\}$, mesurée sur l'individu i, $i \in \{1, ..., N\}$. $f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)$ est paramétrisée par un vecteur de covariables $\mathbf{z}_{ij}$ (ex : temps), un vecteur aléatoire $\boldsymbol{\phi}_i$ de taille $(m \times 1)$ propre à l'individu i. On suppose que les vecteurs $\boldsymbol{\phi}_i$ sont modélisés de la façon suivante :

$$\boldsymbol{\phi}_i = \mathbf{A}_i\boldsymbol{\mu} + \mathbf{u}_i$$

où les $\mathbf{u}_i$ sont i.i.d. $\mathcal{N}(0, \boldsymbol{\Gamma})$, c'est-à-dire Gaussiens centrés et de matrice de variance-covariance $\boldsymbol{\Gamma}(m \times m)$. $\mathbf{A}_i\boldsymbol{\mu}$ représente la moyenne des $\boldsymbol{\phi}_i$ et spécifie le lien entre $\boldsymbol{\phi}_i$ et p covariables $\mathbf{A}_i$ $(m \times p)$ (exemple : sexe, traitement) de coefficients $\boldsymbol{\mu}$. La matrice de variance-covariance $\boldsymbol{\Gamma}$ quantifie la variation inter-individuelle.

Le vecteur $\boldsymbol{\beta}$ correspond au vecteur des effets fixes. Son rôle est similaire à celui de $\boldsymbol{\phi}_i$ excepté qu'il ne contient pas de partie aléatoire.

1.2 Variances hétérogènes : existence et modélisation

1.2.1 Existence des variances hétérogènes

La variance, qui mesure la dispersion d'une variable aléatoire autour de sa moyenne, peut dépendre de certains facteurs et en particulier peut varier entre les individus et entre les observations d'un même individu. La figure 1.4 présente les profils de circonférence du tronc de cinq orangers, déjà présentés dans la section 1.1. La variabilité des mesures augmente au cours de l'expérience. Il serait donc judicieux de prendre en compte ce phénomène dans l'écriture du modèle. Dans ce paragraphe, nous donnerons quelques exemples tirés de différents champs d'application, dans lesquels la présence de variances hétérogènes a été mise en évidence.

Exemple en génétique animale

Un des buts principaux de la génétique animale est l'amélioration des animaux domestiques par le biais de la sélection sur certains caractères (Minvielle, 1998). On évalue le potentiel génétique de chacun des animaux candidats à la sélection à l'aide d'un modèle linéaire mixte dans lequel les effets aléatoires non résiduels correspondent aux valeurs

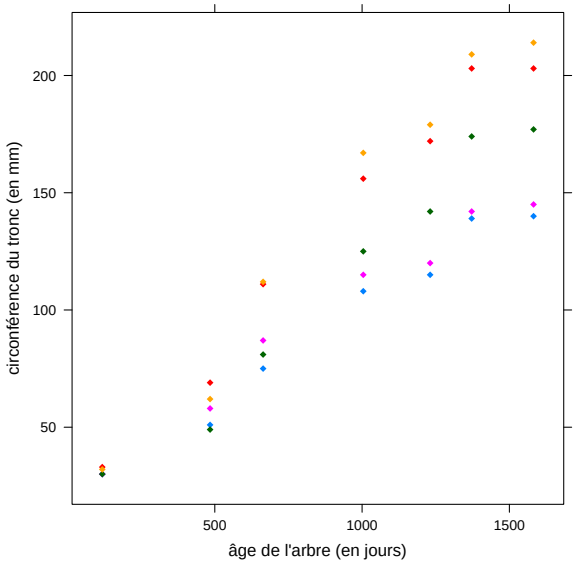


FIGURE 1.4 – Mesure de la circonférence du tronc de cinq orangers en fonction de leur âge.

génétiques des animaux. Prenons un modèle simple :

$$y_{ij} = \mathbf{a}'_{ij}\boldsymbol{\beta} + s_i + e_{ij}$$

où y_{ij} représente la performance du j-ème descendant du père i, $\mathbf{a}'_{ij}$ le vecteur ligne des covariables de milieu relatives à ce descendant, l'effet aléatoire s_i, supposé normal centré, correspond à la valeur génétique du père i et e_{ij} est l'effet résiduel propre au j-ème descendant, supposé normal centré.
L'efficacité de la sélection dépend du rapport $\mathrm{Var}(s_i)/\big[\mathrm{Var}(s_i)+\mathrm{Var}(e_{ij})\big]$. En modèle homogène, ce rapport, noté ρ, est le coefficient de corrélation intra-classe $\rho = \sigma_s^2/(\sigma_s^2 + \sigma_e^2)$, qui est proportionnel à ce que les généticiens appellent l'héritabilité h^2 ($h^2 = 4\rho$). On conçoit très bien que la variance intra-famille $\mathrm{Var}(e_{ij})$ n'est pas nécessairement homogène et peut dépendre de facteurs de milieu. De même la contribution d'un père s_i peut elle aussi varier d'une condition de milieu (j) à l'autre (j'), et on écrira selon Foulley et Quaas (1995) : $s_i = \sigma_{s_j} s_i^*$, où $s_i^* \sim \mathcal{N}(0, 1)$.

Exemple en microéconomie

Battese et Bonyhady (1981) s'intéressent à l'exemple des fonctions de dépenses de nourriture des ménages. Les dépenses de nourriture observées peuvent dépendre du nombre de personnes composant le ménage et des revenus annuels du ménage. En général, on suppose que les dépenses de nourriture observées sont mieux expliquées pour des ménages

à faibles revenus que pour des ménages à forts revenus. En ce sens, la variabilité des dépenses observées dépend des revenus et donc n'est pas constante d'un ménage à l'autre.

Exemple en éducation

Browne et al. (2002) relatent une expérience, analysée à l'origine par Goldstein et al. (1993), dans laquelle on relève les résultats obtenus lors d'un examen par 4059 élèves issus de 65 écoles de la ville de Londres. Les résultats obtenus par tous les élèves à l'âge de 11 ans lors d'un examen de lecture (appelé LRT) sont la principale variable explicative. Lorsqu'on partitionne les résultats LRT en 7 groupes de même taille, en fonction du score, nous remarquons que la variabilité des résultats des élèves dépend du groupe mais aussi du sexe. Dans ce cas, la prise en compte de variances hétérogènes dans le modèle d'étude semble justifiée.

Dans les modèles non linéaires présentés dans la section 1.1., l'hétérogénéité de variances peut être de deux types : l'hétérogénéité de la variance résiduelle σ_{ij}^2 et l'hétérogénéité de la structure de covariance des effets aléatoires $\mathbf{\Gamma}$. Dans la suite de ce document, nous nous sommes intéressés plus particulièrement à la modélisation de la variance résiduelle, dont le choix dépend fortement de la nature des données et de l'application.

1.2.2 Modélisation des variances résiduelles

La prise en compte des variances hétérogènes résiduelles dans les modèles linéaires et non linéaires mixtes fait l'objet de nombreuses publications dans divers domaines. Finalement quel que soit le type de modèles mixtes dans lequel on se place, la modélisation des variances hétérogènes peut être menée de manière identique. Plusieurs directions peuvent être suivies : on peut modéliser la variance résiduelle de manière discrète, paramétrique, semi-paramétrique ou encore non-paramétrique. Nous avons étudié principalement le cas paramétrique.

Pour commencer, l'idée de base serait de prendre une variance résiduelle σ_{ij}^2 par observation, mais il serait optimiste de penser que l'estimateur de chaque variance serait précis puisqu'il serait calculé à partir d'une seule observation. De plus, l'estimateur de chaque variance ne pourrait être consistant, c'est-à-dire converger vers le vrai paramètre lorsque le nombre d'observations tend vers l'infini, puisqu'on augmenterait le nombre de paramètres à estimer avec le nombre d'observations.
Dans le cadre des modèles mixtes, les auteurs se sont particulièrement intéressés à réduire le nombre de paramètres dans la modélisation des variances hétéogènes. Leur démarche a été de :
i) rassembler les observations dans plusieurs groupes en fonction de certaines caractéristiques communes puis supposer que dans chaque groupe la variance est constante mais varie entre

les groupes,
ii) modéliser la variance de manière structurale comme on le ferait sur la fonction moyenne, ou bien prendre en compte une relation entre la moyenne et la variance.

Une variance constante par groupe d'observations

Nous allons tout d'abord nous intéresser à la première possibilité : regrouper les observations suivant certaines caractéristiques. Ces caractéristiques communes peuvent être naturellement indiquées par les covariables utilisées sur la fonction moyenne, ou bien les observations peuvent être regroupées par région géographique, par taille, etc...
Ensuite, dans chaque groupe g (g=1,...,G), la variance résiduelle σ_g^2 est constante mais varie entre les groupes (ex : Hedeker et Mermelstein, 2007).

Cette modélisation des variances a l'inconvénient de ne pas pouvoir varier suivant des variables explicatives. De plus, il peut être difficile de déterminer les différents groupes de sorte que la variance soit constante intra groupe. Enfin et surtout, elle a l'inconvénient de générer un nombre considérable de paramètres si les groupes considérés sont formés par combinaison de plusieurs facteurs. En effet, avec k_l niveaux pour le facteur l, on a potentiellement $\Pi_l k_l$ variances différentes à estimer.

Modéliser la variance par un modèle structural

Considérons tout d'abord le cas où la variance résiduelle ne dépend pas de la fonction moyenne f du modèle. Plusieurs auteurs ont proposé de modéliser la variance par une régression linéaire de la forme générale (Cook et Weisberg, 1983) :

$$\sigma_{ij}^2 = h(w'_{ij}\delta) \tag{1.5}$$

où $h : \mathbb{R} \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$, δ le vecteur de paramètres de dispersion et w'_{ij} le vecteur d'incidence correspondant. Dans cette modélisation, les variances résiduelles se comportent toutes suivant une régression linéaire identique.
Certaines fonctions de lien h ont été plus utilisées par les auteurs, c'est le cas notamment des fonctions polynomiales ou de la fonction exponentielle (Cook et Weisberg, 1983 ; Judge et al., 1985 ; Aitkin, 1987 ; Foulley et al., 1992).
Dans ce sens, Harvey (1976) propose de modéliser le logarithme de la variance par un modèle linéaire de la forme $\sigma_{ij}^2 = \exp\left[(w'_{ij}\delta)^2\right]$ ou encore $\log(\sigma_{ij}^2) = (w'_{ij}\delta)^2$. La forme log-linéaire de la variance assure la positivité de l'estimateur de la variance et s'applique lorsque les effets expliquant la variance sont multiplicatifs. Cette modélisation est largement utilisée dans le cadre des modèles linéaires mixtes en génétique animale (Robert-Granié et al., 1999 ; San Cristobal et al., 2002).
Afin de rajouter un aléa tout en gardant les avantages de la forme logarithmique, Foulley et al. (1992) introduisent un modèle mixte sur la log-variance, avec des variables explicatives $\mathbf{w}_{ij}$ et $\mathbf{q}_{ij}$, des effets fixes $\boldsymbol{\delta}$ et des effets aléatoires $\mathbf{v}$:

$$\log(\sigma_{ij}^2) = \mathbf{w'_{ij}}\boldsymbol{\delta} + \mathbf{q'_{ij}}\mathbf{v} \tag{1.6}$$

où $\mathbf{v} \sim \mathcal{N}(0, \mathbf{\Lambda})$.
Dans la même idée, Lee et Nelder (2006) proposent de modéliser $\log(\sigma_{ij}^2)$ avec une partie aléatoire : $\log(\sigma_{ij}^2) = \gamma + b_i$, où γ correspond à l'effet moyenne et b_i suit une distribution centrée à queues épaisses. Cette modélisation correspond mieux aux variances dont la distribution est plus étalée qu'une Gaussienne.

Modéliser la variance à l'aide d'une relation moyenne-variance

Dans de nombreux domaines comme l'économie (Judge et al., 1985) et la pharmacocinétique (Beal et Sheiner, 1988), l'écart-type résiduel σ_{ij} semble lié à la fonction moyenne f par une relation linéaire. Dans ce sens, on choisit souvent de modéliser la variance résiduelle par la fonction puissance (Box et Hill, 1974) :

$$\sigma_{ij}^2 = \delta_1 f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)^{\delta_2} \tag{1.7}$$

où δ_1 et δ_2 sont deux réels à estimer.
Lorsque $\delta_2 = 2$, le modèle (1.7) revient à établir un modèle homogène sur le logarithme des observations. En effet grâce à un développement limité de $\log(y_{ij})$, on peut montrer que $\sigma_{ij}^2 / f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)^2$ est un bon estimateur de $Var(\log y_{ij})$. On appelle ce modèle un modèle à coefficient de variation constant.
D'autres transformations de la famille Box-Cox des données (Box et Cox, 1964 ; Box et Hill, 1974) peuvent être utilisées de manière similaire pour rendre un modèle homogène lorsqu'initialement la fonction de variance résiduelle était définie par un modèle de puissance. L'importance de se ramener à un modèle homogène est liée à la difficulté des méthodes d'estimation concernant la prise en compte des variances hétérogènes.

En pharmacocinétique, il semble que le modèle à coefficient de variation constant soit souvent retenu (Beal et Sheiner, 1988). Néanmoins dans la plupart des études, il est difficile de fixer δ_2 a priori, il est donc plus raisonnable de l'estimer avec l'ensemble des paramètres du modèle à partir des données.
Dans une étude sur les performances de poids de lapins, Blasco et al. (2003) propose de faire varier la variance résiduelle en fonction du temps en la modélisant avec la même fonction de Gompertz que la fonction moyenne, mais avec des paramètres constants. Contrairement au modèle (1.7), la variance résiduelle ne dépend de la fonction moyenne que par sa forme et non par ses paramètres.
Certains auteurs associent les deux modélisations : Foulley (2004) pose $\sigma_{ij}^2 = \eta_{ij}^{\alpha} \exp(\mathbf{w}_{ij}' \boldsymbol{\delta})$, où η_{ij} est associé à $E(y_{ij})$ et α est un réel à estimer. De même, Lu et al. (2006) associe un modèle structural et une relation moyenne-variance de la manière suivante : $\sigma_{ij}^2 = v\big(f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)\big) w_j$, où v est une fonction régulière, et w_j est une variable aléatoire log-normale.

1.3 Méthodes d'estimation dans les modèles non linéaires à effets mixtes

Explicitons pour commencer le vecteur $\boldsymbol{\theta}$ des paramètres à estimer. Dans le modèle général homogène (1.4), il s'écrit $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2)$. Lorsque le modèle prend en compte des variances hétérogènes, il faut y ajouter les paramètres de modélisation de la variance résiduelle. Rappelons ici que la modélisation choisie pour les modèles non linéaires mixtes sépare directement les effets fixes des effets aléatoires dans la fonction moyenne.
Nous noterons dans la suite $\mathbf{y}$ le vecteur des observations et $\boldsymbol{\phi}$ le vecteur composé des $\boldsymbol{\phi}_i$.

L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour estimer $\boldsymbol{\theta}$. Il s'agit avant tout d'évaluer la vraisemblance des observations notée $p(\boldsymbol{\theta}; \mathbf{y})$, puis d'estimer $\boldsymbol{\theta}$ par la valeur de $\boldsymbol{\theta}$ qui maximise cette vraisemblance. L'estimateur du maximum de vraisemblance est "consistant" et asymptotiquement normal, ainsi on peut construire des régions de confiance pour l'estimateur.

Ici, la vraisemblance des observations s'écrit :

$$p(\boldsymbol{\theta}; \mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\phi}; \boldsymbol{\theta}) d\boldsymbol{\phi} \tag{1.8}$$

où $p(\mathbf{y}, \boldsymbol{\phi}; \boldsymbol{\theta})$ est la vraisemblance du couple $(\mathbf{y}, \boldsymbol{\phi})$.

Au lieu d'évaluer la vraisemblance des observations (1.8) via la loi du couple $(\mathbf{y}, \boldsymbol{\phi})$, nous aurions pu étudier la loi du couple $(\mathbf{y}, \mathbf{u})$, comme l'ont fait beaucoup d'auteurs, en particulier pour les méthodes de linéarisation (cf section 1.3.1). Ce choix aura principalement une incidence sur les méthodes d'estimation basées sur l'algorithme Expectation-Maximisation (Dempster et al., 1977) noté EM, présentées dans la section 1.3.2.

L'équation (1.8) peut aussi s'écrire :

$$p(\boldsymbol{\theta}; \mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\phi}; \boldsymbol{\theta}) p(\boldsymbol{\phi}; \boldsymbol{\theta}) d\boldsymbol{\phi} \tag{1.9}$$

où $p(\mathbf{y}|\boldsymbol{\phi}; \boldsymbol{\theta})$ est la vraisemblance des observations sachant $\boldsymbol{\phi}$ et $\boldsymbol{\theta}$ et $p(\boldsymbol{\phi}; \boldsymbol{\theta})$ est la distribution de $\boldsymbol{\phi}$.

Dans le cas d'un modèle non linéaire homogène, c'est-à-dire avec une variance résiduelle constante σ^2, les spécifications du modèle (1.4) nous amènent à écrire les vraisemblances de la manière suivante :

$$p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\theta}) = \frac{1}{(2\Pi\sigma^2)^{Ntot/2}} \exp\Big[-\frac{1}{2\sigma^2} \sum_{i,j} \big(y_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)\big)^2 \Big]$$

où $Ntot = \sum_i n_i$.

$$p(\boldsymbol{\phi};\boldsymbol{\theta}) = \frac{1}{(2\Pi)^{N/2}|\boldsymbol{\Gamma}|^{N/2}} \exp\big[-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{\phi}_i - \mathbf{A}_i\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{\phi}_i - \mathbf{A}_i\boldsymbol{\mu})\big]$$

Pour les fonctions moyennes f linéaires en $\boldsymbol{\phi}$, la théorie des modèles linéaires mixtes peut s'appliquer et l'intégrale (1.8) conduit à des estimateurs de maximum de vraisemblance via l'algorithme EM (Dempster et al., 1977) par exemple.

Dans le cas général, la non linéarité de la fonction f ne permet pas d'évaluer cette intégrale, et donc il n'est pas possible de dériver directement la vraisemblance des observations afin d'estimer le vecteur des paramètres $\boldsymbol{\theta}$ par l'estimateur du maximum de vraisemblance. Des techniques numériques peuvent être utilisées pour approcher la valeur de l'intégrale, mais elles peuvent être coûteuses en temps de calcul. Dans ce sens, de nombreuses études basées sur des approximations de $p(\boldsymbol{\theta};\mathbf{y})$ ont été conduites dans le but de se ramener à un modèle linéaire. Nous rapporterons et commenterons dans cette partie les principales méthodes d'estimations utilisées dans les modèles non linéaires mixtes.
Ces méthodes ont été développées dans le cadre des modèles non linéaires mixtes homogènes ; certaines ont été adaptées à l'estimation de quelques fonctions de variances hétérogènes.

1.3.1 Méthodes de linéarisation de la vraisemblance

Le principal objectif des méthodes de linéarisation est de transformer la vraisemblance des observations dans le but d'obtenir un modèle approximé linéaire, et ensuite pouvoir utiliser les méthodes classiques. Ces méthodes sont très simples à implémenter et sont peu coûteuses en temps de calculs grâce aux travaux effectués dans le domaine des modèles linéaires mixtes.

La linéarisation "First-Order (FO)" de Beal et Sheiner (1982)

Beal et Sheiner (1982) ont choisi d'intégrer la vraisemblance conjointe du couple $(\mathbf{y},\mathbf{u})$ par rapport à $\mathbf{u}$ pour obtenir la vraisemblance des observations, contrairement à nous qui nous sommes intéressés à la distribution de $(\mathbf{y},\boldsymbol{\phi})$. Nous resterons fidèles à leur choix.

Un développement de Taylor à l'ordre 1 sur la fonction f est mené autour de $\mathbf{u}_i = 0$, en supposant que le terme en $\mathbf{u}_i\varepsilon^*_{ij}$ est négligeable.
Le modèle (1.4) devient :

$$y_{ij} \approx f(\mathbf{z}_{ij},\boldsymbol{\beta},\mathbf{A}_i\boldsymbol{\mu}) + \mathbf{Z}_i(\mathbf{z}_{ij},\boldsymbol{\beta},\mathbf{A}_i\boldsymbol{\mu})\mathbf{u}_i + \sigma\varepsilon^*_{ij} \tag{1.10}$$

où $Z_i(\mathbf{z}_{ij},\boldsymbol{\beta},\mathbf{A}_i\boldsymbol{\mu}+\mathbf{u}^*_i) = \frac{\partial f(\mathbf{z}_{ij},\boldsymbol{\beta},\mathbf{A}_i\boldsymbol{\mu}+\mathbf{u}_i)}{\partial \mathbf{u}_i}\big|_{\mathbf{u}_i=\mathbf{u}^*_i}$.

Dans l'équation (1.10), on obtient les deux premiers moments suivants :
$\mathbb{E}(y_{ij}) = f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu})$ et $\mathbf{V}_{ij} = \mathbb{V}ar(y_{ij}) = \mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu})\boldsymbol{\Gamma}\mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu})' + \sigma^2$.

Utilisant l'approximation (1.10) du modèle, la vraisemblance des observations du modèle approximé est la densité d'un vecteur gaussien de moyenne et variance linéaires en $\mathbf{u}_i$.

Le modèle (1.10) ressemble à un modèle hiérarchique linéaire à deux niveaux, mais avec des différences notables : la matrice $\mathbf{Z}_i$ n'est pas une matrice de design et dépend de $\boldsymbol{\beta}$ et $\boldsymbol{\mu}$, de plus la fonction $f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu})$ est non-linéaire en $\boldsymbol{\beta}$.
Beal et Sheiner (1982) proposent ensuite d'estimer le vecteur des paramètres en minimisant la fonction objective suivante :

$$L_{FO}(\theta) = \sum_{ij}\left[\log(\mathbf{V}_{ij}) + \frac{1}{V_{ij}}\big(y_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A_i}\boldsymbol{\mu})\big)^2\right] \tag{1.11}$$

Pour minimiser la vraisemblance L_{FO}, des techniques numériques comme l'algorithme de Newton-Raphson peuvent être utilisées. Cette méthode, implémentée dans la fonction Nlmixed de SAS (Littell et al., 2006), a tendance à fournir des estimateurs biaisés de $\boldsymbol{\theta}$; en particulier lorsque les éléments diagonaux de la matrice de covariance $\boldsymbol{\Gamma}$ sont élévés (Vonesh, 1992).

La linéarisation "First-Order Conditional" (FOCE)

La linéarisation FOCE fut initialement introduite par Lindstrom et Bates (1990) dans le but d'améliorer la linéarisation FO. Elle est disponible via le package nlme du logiciel R, et dans SAS via Proc Nlmixed. Dans la méthode précédente, un développement de Taylor était réalisé autour de $\mathbf{u} = 0$, alors que dans la linéarisation FOCE, il est réalisé autour du mode a posteriori de $\mathbf{u}$, correspondant au meilleur prédicteur linéaire non biaisé (BLUP, Harville, 1976) des modèles linéaires mixtes.
Soit $\hat{\mathbf{u}}_i$ l'estimation courante de $\mathbb{E}[\mathbf{u}_i|\mathbf{y}; \boldsymbol{\theta}]$ égale à l'expression

$$\mathbb{E}[\mathbf{u}_i|\mathbf{y}; \boldsymbol{\theta}] = \sigma^2\boldsymbol{\Gamma}\mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i)'\big(\mathbf{y}_i - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i)\big)$$

L'approximation FOCE du modèle (1.4) est la suivante :

$$y_{ij} \approx f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i) + \mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i)(\mathbf{u}_i - \hat{\mathbf{u}}_i) + \sigma\varepsilon^*_{ij} \tag{1.12}$$

où :
$\mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \boldsymbol{u}^*_i) = \frac{\partial f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \boldsymbol{u}_i)}{\partial \mathbf{u}_i}\big|_{\mathbf{u}_i = \mathbf{u}^*_i}$.

Dans l'équation (1.12), on obtient les deux premiers moments suivants :
$\mathbb{E}(y_{ij}) = f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i) - \mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\boldsymbol{u}}_i)\hat{\boldsymbol{u}}_i$ et
$\mathbf{V}_{ij} = \mathbb{V}ar(y_{ij}) = \mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i)\boldsymbol{\Gamma}\mathbf{Z}_i(\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{A}_i\boldsymbol{\mu} + \hat{\mathbf{u}}_i)' + \sigma^2$.

Comme pour la méthode FO, la vraisemblance des observations du modèle approximé est la densité d'un vecteur gaussien de moyenne et variance linéaires en $\mathbf{u}_i$. Elle s'écrit donc facilement en fonction de $\mathbb{E}(y_{ij})$ et $\mathbf{V}_{ij}$. Lindstrom et Bates (1990) proposent ensuite d'utiliser un algorithme de Newton Raphson pour obtenir un estimateur de $\boldsymbol{\theta}$.
Al-Zaid et Yang (2001) présentent une approximation similaire dans laquelle un algorithme EM est utilisé à partir du modèle approximé (1.12). Comme on peut s'y attendre, ces méthodes ont tendance à mal estimer le vecteur des paramètres lorsque la distribution des observations s'éloigne d'une distribution normale.
D'après Vonesh (1992), la méthode d'approximation FOCE fournit de meilleurs estimateurs que la méthode FO en particulier pour les composantes de variances. Un inconvénient important des méthodes FO et FOCE est lié au peu de propriétés asymptotiques connues (Ramos et Pantula, 1995).
Wolfinger (1993) démontre comment la version REML (Restricted Maximum Likelihood) de l'approximation FOCE peut-être obtenue à l'aide d'une approximation de Laplace.

1.3.2 Méthodes basées sur la théorie du maximum de vraisemblance

Face aux difficultés d'estimations lorsque le nombre d'observations par individu est faible ou lorsque les variances individuelles sont grandes, des méthodes exactes basées sur la théorie du maximum de vraisemblance ont été proposées (par exemple, Wei et Tanner, 1990). Le principal obstacle à la maximisation de la vraisemblance des observations est l'évaluation de l'intégrale. Les méthodes exactes que nous proposons ici sont toutes basées sur des méthodes d'intégration par Monte Carlo (MC). Vonesh et al. (2002) rapporte que ces méthodes fournissent de bons résultats lorsqu'il y a peu de paramètres aléatoires mais deviennent vite très coûteuses en temps de calculs lorsque ceux-ci augmentent.
L'échantillonnage d'importance (importance sampling) est très souvent utilisé dans ce genre de situations, car la méthode est simple à implémenter et efficace. Il s'agit d'estimer l'intégrale par une moyenne empirique de l'intégrande évaluée en des points simulés sous une loi instrumentale. Le succès de cette méthode dépend fortement de la densité instrumentale utilisée pour simuler les effets aléatoires. La densité instrumentale minimisant la variance de l'estimateur de la vraisemblance des observations étant $p(\boldsymbol{\phi}|\mathbf{y};\boldsymbol{\theta})$ (Robert et Casella, 2004), non explicite, certains auteurs ont proposé d'autres densités instrumentales (Pinheiro et Bates, 1995).

Demidenko (1997) et Vonesh et al. (2002) ont montré que la méthode FOCE est asymptotiquement identique aux méthodes exactes lorsque le nombre d'individus N et le nombre d'observations par individu n_i tendent vers l'infini.

Quadrature de Gauss

L'application des méthodes de quadratures de Gauss dans le calcul de la vraisemblance des observations des modèles non linéaires mixtes a été proposée par Davidian et Gallant (1992).
La quadrature de Gauss est généralement utilisée pour estimer une intégrale en la remplaçant par une somme pondérée évaluée en un certain nombre de points du domaine d'intégration. Les poids et abscisses utilisés avec la plupart des noyaux d'intégration sont détaillés dans Abramowitz et Stegun (1964).
Ici pour obtenir la vraisemblance des observations, l'intégrale à calculer est de dimension multiple (de la taille du vecteur aléatoire) et dans ce cas, les méthodes de quadrature de Gauss peuvent s'avérer numériquement complexes (Davis et Rabinowitz, 1984). Néanmoins la structure quasi-Gaussienne de l'intégrande dans le modèle présenté simplifie les calculs en transformant l'intégrale multiple en la somme d'intégrales simples.
Soient x_l, w_l, $l = 1, ..., QG$ les abscisses et les poids utilisés dans la méthode de quadrature de Gauss, le noyau que nous avons choisi est la distribution Gaussienne centrée réduite. Comme les $\mathbf{y}_i$ sont indépendantes, la vraisemblance de $\mathbf{y}$ peut-être écrite comme le produit des vraisemblances des $\mathbf{y}_i (i = 1, ..., N)$. On estime la vraisemblance des observations de la manière suivante :

$$
\begin{aligned}
p(\boldsymbol{\theta}; \mathbf{y_i}) = \; & cste \times \int \tfrac{1}{\sigma^{n_i}} \tfrac{1}{|\boldsymbol{\Gamma}|^{1/2}} \exp\left[-\tfrac{1}{2\sigma^2} \textstyle\sum_j \left(\mathbf{y}_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)\right)^2 \right] \\
& \times \exp\left[-\tfrac{1}{2}\left(\boldsymbol{\phi}_i - \mathbf{A}_i\boldsymbol{\mu}\right)' \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{\phi}_i - \mathbf{A}_i\boldsymbol{\mu}\right) \right] d\boldsymbol{\phi}_i \\
= \; & cste \times \int \tfrac{1}{\sigma^{n_i}} \exp\left[-\tfrac{1}{2\sigma^2} \times \textstyle\sum_j \left(\mathbf{y}_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\Gamma}^{\frac{1}{2}} \mathbf{x}_{l_1,...,l_m} + \mathbf{A}_i\boldsymbol{\mu})\right)^2 \right] \\
& \times \exp\left[-\tfrac{1}{2}\|\mathbf{x}\|^2 \right] d\mathbf{x} \\
\simeq \; & cste \times \textstyle\sum_{l_1=1}^{N_{QG}} ... \sum_{l_m=1}^{N_{QG}} \tfrac{1}{\sigma^{n_i}} \\
& \times \exp\left[-\tfrac{1}{2\sigma^2} \textstyle\sum_j \left(\mathbf{y}_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\Gamma}^{\frac{1}{2}} \mathbf{x}_{l_1,...,l_m} + \mathbf{A}_i\boldsymbol{\mu}_i)\right)^2 \right] \textstyle\prod_{k=1}^{m} w_{l_k}
\end{aligned}
$$

où $\mathbf{x}_{l_1,...,l_m} = (x_{l_1}, ..., x_{l_m})'$.

La méthode de quadrature de Gauss classique a été améliorée en centrant les abscisses autour de $\mathbb{E}(\boldsymbol{\phi}_i|\mathbf{y}; \boldsymbol{\theta})$ au lieu de $\boldsymbol{\mu}$ (Pinheiro et Bates, 1995). Les résultats obtenus avec cette méthode d'estimation sont proches de ceux obtenus par l'échantillonnage d'importance avec un noyau similaire.

Lorsque le vecteur des effets aléatoires est de grande dimension, la quadrature de Gauss devient numériquement difficile et le temps de calculs très long.

Algorithmes de Monte-Carlo EM (MCEM) et Approximation Stochastique d'EM (SAEM)

Les méthodes que nous allons présenter dans ce paragraphe sont basées sur la théorie des algorithmes Expectation-Maximization (Dempster et al., 1977), plus communément appelés EM. Cette classe d'algorithmes fournit un estimateur du maximum de vraisemblance dans les modèles où certaines données sont dites "manquantes" ou "latentes". Dans cette étude, le vecteur $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_N)$ correspondant aux paramètres aléatoires du modèle n'est pas observable et est considéré comme une donnée manquante.
L'algorithme EM général est itératif et composé de deux étapes : une étape Espérance et une étape Maximisation.

Soit $\boldsymbol{\theta}^{(k-1)}$ la valeur courante du paramètre obtenue à l'itération $k-1$ de l'algorithme EM. A l'itération k, EM est composé des deux étapes suivantes :

- Etape E : on évalue l'expression $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) = \mathbb{E}\big[\log p(\mathbf{y}, \boldsymbol{\phi}; \boldsymbol{\theta})|\mathbf{y}; \boldsymbol{\theta}^{(\mathbf{k-1})}\big]$, c'est-à-dire l'espérance conditionnelle de la vraisemblance du couple $(\mathbf{y}, \boldsymbol{\phi})$ sous la distribution des données manquantes $\boldsymbol{\phi}$ sachant les observations $\mathbf{y}$ et la valeur courante des paramètres $\boldsymbol{\theta}^{(k-1)}$.

- Etape M : on réactualise la valeur de $\boldsymbol{\theta}$ en affectant à $\boldsymbol{\theta}^{(k)}$ la valeur de $\boldsymbol{\theta}$ qui maximise $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$.

Cet algorithme est en pratique très efficace et fréquemment utilisé pour l'étude des modèles linéaires mixtes (Searle et al., 1992). Les propriétés de convergence de l'algorithme vers un point stationnaire sous des conditions de régularité de la vraisemblance ont été démontrées dans Dempster et al. (1977) et Wu (1983).

L'expression $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ peut de manière égale être écrite de la façon suivante en faisant apparaître le signe intégrale de l'espérance conditionnelle :

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) = \int \log p(\mathbf{y}, \boldsymbol{\phi}; \boldsymbol{\theta}) p(\boldsymbol{\phi}|\mathbf{y}; \boldsymbol{\theta}^{(k-1)}) d\boldsymbol{\phi} \tag{1.13}$$

Dans les modèles non linéaires mixtes, l'intégrale (1.13) n'est pas sous une forme explicite. Dans ce sens, tout l'intérêt des algorithmes développés dans ce paragraphe est de bien estimer l'intégrale (1.13) grâce à des méthodes stochastiques, c'est-à-dire en générant des variables aléatoires. Contrairement à l'algorithme EM déterministe, les algorithmes EM stochastiques n'assurent pas nécessairement la convergence vers le même point stationnaire de la vraisemblance quelles que soient les valeurs initiales des paramètres.

Dans un premier temps, Wei et Tanner (1990) proposent de manière naturelle d'estimer l'intégrale (1.13) par une moyenne de Monte Carlo :

$$\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) = \frac{1}{S_k}\sum_{s=1}^{S_k} \log p(\mathbf{y}, \boldsymbol{\phi}_s; \boldsymbol{\theta})$$

où $\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_{S_k}$ sont simulés de manière indépendante sous la distribution $p(\boldsymbol{\phi}|\mathbf{y}; \boldsymbol{\theta}^{(k-1)})$. La loi forte des grands nombre assure que $\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ est une bonne estimation de $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ pourvu que S_k soit assez grand.

En pratique, cette méthode ne peut pas être appliquée directement puisque la loi des données manquantes $\boldsymbol{\phi}$ sachant les observations $\mathbf{y}$ et la valeur courante des paramètres $\boldsymbol{\theta}^{(k-1)}$, notée $p(\boldsymbol{\phi}|\mathbf{y}; \boldsymbol{\theta}^{(k-1)})$, n'est pas connue.
Des algorithmes basés sur les méthodes de Chaînes de Markov et Monte Carlo (MCMC) peuvent être utilisés pour obtenir les variables simulées $(\boldsymbol{\phi_s})_{s=1,...,S_k}$: l'algorithme d'acceptation-rejet, l'algorithme de Metropolis-Hastings, ou encore l'algorithme d'échantillonnage de Gibbs (Robert et Casella, 2003).
La précision de l'estimateur $\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ dépend fortement de la valeur de S_k. Faut-il simuler un nombre constant $S_k = S$ de variables à chaque itération EM ou au contraire augmenter le nombre de simulations au fur à mesure des itérations EM ?
McCulloch (1997) et Booth et Hobert (1999) rapportent quelques résultats de leurs études dans le cadre des modèles linéaires généralisés. Même en choisissant d'augmenter progressivement le nombre de simulations, ces auteurs montrent qu'il fallait simuler plus de $S = 60000$ variables à l'itération finale pour obtenir un estimateur $\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ précis.

Walker (1996) propose de simuler les $\boldsymbol{\phi_s}$ de $\hat{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$ sous la distribution a priori $\mathcal{N}(\mathbf{A}_i\boldsymbol{\mu}, \boldsymbol{\Gamma})$ de $\boldsymbol{\phi}_i$ en passant par une méthode d'échantillonnage d'importance. D'après les résultats obtenus, le choix de la loi instrumentale a un impact sur la valeur de S_k.

Dans le même ordre d'idées, Wang (2007) propose d'utiliser un mélange de distributions simple et efficace basé sur les distributions a priori et a posteriori de $\boldsymbol{\phi}$ et obtient de meilleurs résultats que l'algorithme MCEM classique.

Afin de réduire le nombre de variables simulées dans l'algorithme MCEM, Delyon et al. (1999) proposent de recycler les variables simulées aux itérations précédentes par une procédure d'approximation de Robbins et Monroe (1951), et remplacent l'étape E de l'algorithme EM par deux étapes : une étape de simulation des variables $(\boldsymbol{\phi}_s)_{s=1,...,S}$ et une étape d'approximation stochastique. Leur algorithme, noté SAEM pour "Stochastic Approximation of EM", est donc composé de 3 étapes :

Soit $\boldsymbol{\theta}^{(k-1)}$ la valeur de $\boldsymbol{\theta}$ obtenue à l'étape $k-1$,

– Etape de simulation : Simuler $\boldsymbol{\phi}^{(k)}$ sous la distribution $p(\boldsymbol{\phi}|\mathbf{y};\boldsymbol{\theta}^{(k-1)})$

– Etape d'approximation stochastique : actualiser l'estimateur de $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)})$, noté $Q_k(\boldsymbol{\theta})$, par

$$Q_k(\boldsymbol{\theta}) = Q_{k-1}(\boldsymbol{\theta}) + \gamma_k\big[\log\ p(\boldsymbol{\phi}^{(k)},\mathbf{y};\boldsymbol{\theta}^{(k-1)}) - Q_{k-1}(\boldsymbol{\theta})\big]$$

où $(\gamma_k)_k$ est une suite décroissante de réels positifs,

– L'étape de maximisation est la même que celle de EM : $\boldsymbol{\theta}^{(k)} = arg\max_{\boldsymbol{\theta}} Q_k(\boldsymbol{\theta})$.

Sous des propriétés de régularité de la vraisemblance, Delyon et al. (1999) ont démontré la convergence de l'algorithme SAEM vers l'estimateur du maximum de vraisemblance.

Dans le cadre général des modèles non linéaires mixtes, nous avons déjà vu que la distribution $p(\boldsymbol{\phi}|\mathbf{y};\boldsymbol{\theta}^{(k-1)})$ n'est pas toujours connue. Dans ce sens, Kuhn et Lavielle (2004) présentent l'algorithme SAEM-MCMC dans lequel l'étape de simulation de SAEM est accomplie par une méthode de type MCMC, comme l'algorithme de Metropolis-Hastings (Robert et Casella, 2004). Ce dernier algorithme construit une chaîne de Markov dont la distribution stationnaire correspond à $p(\boldsymbol{\phi}|\mathbf{y};\boldsymbol{\theta}^{(k-1)})$.
Lorsque la vraisemblance des données complètes appartient à la famille exponentielle, Kuhn et Lavielle (2005) simplifient l'écriture de l'algorithme en faisant apparaître des statistiques exhaustives, et montrent la convergence de l'algorithme couplé SAEM-MCMC.

Théoriquement, la convergence de l'algorithme SAEM est assurée en ne simulant qu'une seule variable $\boldsymbol{\phi}$ à l'étape S. Néanmoins en pratique, plusieurs (de dix à quelques centaines selon les jeux de données) sont nécessaires pour obtenir un estimateur précis de $\boldsymbol{\theta}$. Kuhn et Lavielle (2005) montrent que l'algorithme SAEM-MCMC est moins coûteux en terme de temps de calcul que l'algorithme MCEM. Dans le même sens, Wang (2007) présente une étude pour laquelle il obtient un rapport de temps égal à 6 entre ces deux méthodes.

L'algorithme SAEM-MCMC a été implémenté dans le logiciel Monolix
(http ://group.Monolix.org/), dans le cadre de modèles homogènes et prise en compte de variances hétérogènes par une relation moyenne-variance. Cet algorithme très efficace comporte néanmoins un inconvénient majeur : de nombreux paramètres utilisés lors des simulations et lors de l'étape d'approximation doivent être calibrés avant de lancer l'algorithme.

Comme de nombreux algorithmes itératifs basés sur l'algorithme EM, la vitesse de convergence de l'algorithme SAEM-MCMC peut s'avérer lente dans certains cas. Lavielle et Meza (2007) proposent une adaptation de l'algorithme PX-EM, développé par Liu et al. (1998) dans le cadre des modèle linéaires mixtes. Celle-ci permet d'accélérer l'algorithme SAEM et permet aussi d'éviter les minima locaux de la vraisemblance.

1.3.3 Algorithme basé sur une pseudo vraisemblance

Concordet et Nuñez (2002) proposent une méthode basée sur une pseudo-vraisemblance, placée à la limite des méthodes de linéarisation et des méthodes de maximum de vraisemblance. L'idée générale de la méthode est la suivante : puisque la vraisemblance des observations est difficile à obtenir, on suppose que les observations suivent une loi normale dont la moyenne et la variance correspondent aux deux premiers moments de la vraie distribution des observations. Ensuite on estime le vecteur des paramètres $\boldsymbol{\theta}$ en minimisant un critère des moindres carrés. Sous certaines conditions de régularité de la vraisemblance, l'estimateur obtenu est consistant et asymptotiquement Gaussien. Concordet et Nuñez (2002) rapportent une application dans laquelle cette méthode fournit de meilleurs estimateurs que FOCE et que les méthodes de quadrature de Gauss.

1.3.4 Méthodes bayésiennes

La théorie bayésienne est basée sur la version continue du théorème de Bayes s'énonçant de la manière suivante. Pour une loi (dite a priori) $\Pi(\boldsymbol{\theta})$ sur le vecteur des paramètres $\boldsymbol{\theta}$ et un vecteur des observations $\mathbf{y}$ de densité $p(\mathbf{y}|\boldsymbol{\theta})$, la distribution de $\boldsymbol{\theta}$ conditionnellement à $\mathbf{y}$ (loi dite a posteriori) a pour densité :

$$\Pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

L'approche bayésienne s'attache à émettre des inférences sur le vecteur $\boldsymbol{\theta}$, à partir des variables $\mathbf{y}$ observées, contrairement à l'approche fréquentiste, qui caractérise le comportement des observations conditionnellement à $\boldsymbol{\theta}$ (pour plus de détails sur la théorie bayésienne, voir Robert, 1992).
S'appuyant sur le théorème de Bayes, la structure hiérarchique du modèle (1.4) rend l'approche bayésienne candidate aux méthodes d'estimation possibles. De plus, l'avancée dans les calculs d'intégrales grâce aux méthodes de Monte Carlo et de Chaînes de Markov (MCMC) rendent les méthodes bayésiennes numériquement accessibles (Smith et Roberts, 1993).

Une loi a priori est choisie pour $\boldsymbol{\theta}$, en particulier pour le vecteur d'effets fixes $\boldsymbol{\beta}$ de la fonction moyenne du modèle (1.4), les paramètres de la variance résiduelle, et la moyenne et variance des paramètres aléatoires : $\boldsymbol{\mu}$ et $\boldsymbol{\Gamma}$.
On va se placer dans le cas où la variance résiduelle est homogène ($\sigma_{ij}^2 = \sigma^2 \ \forall i, j$) pour simplifier les notations. La densité a posteriori de $(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2)$ est donnée par l'expression suivante (Davidian et Giltinan, 1995, chap. 8) :

$$p(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2|\mathbf{y}) = \frac{\prod_{i=1}^{N} p(\mathbf{y}_i|\boldsymbol{\phi}_i, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2)p(\boldsymbol{\phi}_i|\boldsymbol{\mu}, \boldsymbol{\Gamma})p(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2)}{p(\mathbf{y})} \tag{1.14}$$

La densité marginale de chacun des paramètres est ensuite obtenue par intégration de (1.14) et on estime les paramètres en prenant la moyenne ou le mode de la densité marginale.

Le passage de la densité conditionnelle à la densité marginale pour chacun des paramètres peut-être réalisé à l'aide d'algorithmes MCMC comme les algorithmes de Gibbs (Wakefield et al., 1994).
Bien que la distribution a posteriori de $\boldsymbol{\theta}$ soit importante, il peut être nécessaire de résumer la distribution a posteriori sous une valeur de $\boldsymbol{\theta}$ "optimale", comme dans l'approche fréquentiste. Cette valeur "optimale" n'est pas unique : elle est liée à une fonction de perte due au choix du meilleur estimateur de $\boldsymbol{\theta}$, noté $\hat{\boldsymbol{\theta}}$. Les fonctions de perte les plus fréquemment utilisées sont les suivantes :

- La perte quadratique si
$$L(\boldsymbol{\theta}, \delta(y)) = (\boldsymbol{\theta} - \delta(y))^2$$
- La perte absolue si
$$L(\boldsymbol{\theta}, \delta(y)) = |\boldsymbol{\theta} - \delta(y)|$$

où $\delta(y)$ est un estimateur de $\boldsymbol{\theta}$.
La valeur du paramètre optimal est celle qui minimise la fonction de risque espérée a posteriori par rapport à $\boldsymbol{\theta}$ (Berger, 1985) :

$$R(\boldsymbol{\theta}, \delta(y)) = \int_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \delta(y)) \Pi(\boldsymbol{\theta}|y) d\boldsymbol{\theta}$$

C'est l'espérance de la distribution a posteriori pour la fonction de perte quadratique et la médiane de cette même loi pour la fonction de perte absolue.

Dans la recherche directe d'estimateurs ponctuels de $\boldsymbol{\theta}$ sans expliciter la densité a posteriori, des méthodes basées sur l'algorithme EM et l'approximation de Laplace ont été proposées (Racine-Poon, 1985 ; Tierney et Kadane, 1986). Les méthodes type EM semblent fournir de mauvais estimateurs lorsque le nombre d'observations par individu est petit ou moyen, et l'approximation de Laplace est très sensible à la paramétrisation utilisée. Le passage indirect via la densité a posteriori en utilisant les méthodes MCMC sont jugées elles, plus flexibles et plus performantes (Wakefield et al., 1994).
Dans notre étude, nous avons plusieurs fois comparé nos résultats à ceux obtenus par le logiciel WinBUGS (http ://www.mrc-bsu.cam.ac.uk/bugs/), qui fournit un estimateur bayésien des paramètres dans les modèles non linéaires mixtes.
De telles méthodes permettent avantageusement de prendre en compte des contraintes biologiques ou physiques sur le vecteur des paramètres grâce au choix de la loi a priori (Gelman en al., 1996). Néanmoins si peu d'informations sur les paramètres sont disponibles, les méthodes fréquentistes peuvent être plus adaptées.

La plupart de ces méthodes ont été adaptées à l'étude de certains modèles à variances hétérogènes :

– Dans SAS via la procédure Nlmixed et dans R via la procédure nlme (Pinheiro et Bates, 1995), les modèles linéaires fixes de la forme (1.5) et la relation moyenne-variance (1.7) sont étudiés pour les méthodes de linéarisation et de quadrature.

– Dans Monolix, on peut estimer des variances hétérogènes modélisées par une relation moyenne variance en utilisant l'algorithme SAEM-MCMC.

– Le logiciel WinBUGS permet de prendre en compte les modèles de variances hétérogènes de la forme (1.6) et (1.7).

Finalement, exceptées les méthodes bayésiennes implémentées dans le logiciel WinBUGS, aucune des méthodes d'estimation ne permet de prendre en compte à la fois un modèle linéaire mixte sur la log-variance ou une relation moyenne-variance.

L'objectif de la thèse est d'adapter l'algorithme SAEM-MCMC à la prise en compte d'hétérogénéité de variance via les modèles (1.6) ou (1.7), puis de comparer la méthode obtenue aux méthodes existantes (chapitre 3).
Pour ce faire, des critères ont tout d'abord été proposés pour calibrer les paramètres de l'algorithme SAEM-MCMC (chapitre 2).

Chapitre 2

Quelques critères pour calibrer les paramètres de l'algorithme SAEM-MCMC

L'article suivant a été soumis à "Computational Statistics" le 18 mars 2008, nous n'avons pas eu de retour à ce jour.

Criteria to calibrate the parameters of the SAEM-MCMC algorithm in maximum likelihood estimation for nonlinear mixed effects models

Mylène Duval, Christèle Robert-Granié

INRA UR 631 SAGA,F-31 326 Castanet-Tolosan, France
email: Mylene.Duval@toulouse.inra.fr

Tel: 05-61-28-51-94, Fax: 05-61-28-53-53

Abstract

The SAEM-MCMC algorithm is a powerful tool used for computing maximum likelihood estimators in the wide class of nonlinear mixed effects models. In this article, we propose some criteria to calibrate the parameters of the SAEM-MCMC algorithm. We compare on a real data set and by simulations the outputs of several procedures: the SAEM-MCMC algorithm with parameters chosen using our criteria, the SAEM-MCMC algorithm with five other sets of parameters, and the exact NLMIXED procedure of SAS. We show that the value of the estimates depends on the choice of the set of parameters. In particular we need to run long markov chains in the Metropolis-Hastings algorithm to obtain an accurate estimator, what may be time consuming.

Key words: nonlinear mixed models, maximum likelihood, stochastic algorithm, Metropolis-Hastings, SAEM-MCMC.

2.1 Introduction

Nonlinear mixed effects models (NLMM) are more and more frequently used for analysis of longitudinal data and repeated measurements in pharmacokinetics, growth and other studies. Comparing to linear mixed models, parameters of such models provide a better biological interpretation of the mechanisms involved and the corresponding models are also more parcimonious. The main interest of this paper is to obtain good parameter estimates using maximum likelihood estimation in nonlinear mixed effects models.
Several procedures have already been proposed to estimate parameters of NLMM. The first ones were based on linearization of the log-likelihood such as first-order (F0) and first-order conditional expectation (FOCE) approximations (Sheiner and Beal, 1980; Lindstrom and Bates, 1990). Since errors can be large in the approximation of the observed log-likelihood (Davidian and Giltinian, 1995), some methods based on exact maximum likelihood (ML) were proposed such as Gaussian quadrature and methods based on Monte Carlo methods. However, integration via Gaussian quadrature can be difficult and inaccurate in cases with high dimensionality, in this way stochastic tools may be a powerful alternative. Wei and Tanner (1990) proposed the MCEM algorithm, in which the E-step of the EM algorithm is approximated using a large sample of simulated data and so it is highly time consuming. For instance, Booth and Hobert (1999) reported some results from a study on a real data set: they simulated around 60,000 samples for the final iteration. Delyon et al (1999) proposed a method which promises convergence with fewer simulations: the SAEM algorithm. In this method, the E-step of EM algorithm is replaced by a Simulation step and a Stochastic Approximation step. When the conditional distribution of the missing effects given the observations is unknown, Kuhn and Lavielle (2004, 2005) combined the SAEM algorithm with a MCMC procedure, such as the Metropolis-Hastings algorithm, and called SAEM-MCMC algorithm. In practice, the main problem of this method is to adequately calibrate its parameters to obtain good parameter estimates.
The aim of this study was to present some criteria which determine the parameters of the SAEM-MCMC algorithm. We discuss if this choice of parameters is relevant or not. The paper is organized as follows. In section 2, we introduce the model and the SAEM-MCMC algorithm. In section 3, we propose some criteria to determine the parameters of the algorithm. Section 4 is devoted to the comparison of the outputs obtained with the SAEM-MCMC algorithm using several sets of parameters on real and simulated data.

2.2 The nonlinear mixed effects model and the SAEM-MCMC algorithm

2.2.1 The model

We consider the following model:

$$y_{ij} = f(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i) + g(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i, \alpha)\, \varepsilon_{ij}$$

where y_{ij} is the jth observation $j \in \{1, .., n_i\}$ of the ith individual $i \in \{1, ..., N\}$.
$f(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)$ describes the mean response and depends on a vector of explanatory variables $\boldsymbol{z}_{ij}$ (e.g. time) and a vector of random effects $\boldsymbol{\phi}_i$ $(m \times 1)$. We assume that $\boldsymbol{\phi}_i$ are i.i.d. Gaussian random variables $\mathcal{N}(\boldsymbol{A}_i\boldsymbol{\mu}, \boldsymbol{\Gamma})$. $\boldsymbol{A}_i\boldsymbol{\mu}$ specifies the link between the expectation of $\boldsymbol{\phi}_i$ and p covariates $\boldsymbol{A}_i$ $(m \times p)$ (e.g. gender, treatment) with coefficients $\boldsymbol{\mu}$. The variance covariance matrix $\boldsymbol{\Gamma}$ quantifies the inter-individual variation. Vector $\boldsymbol{\beta}$ corresponds to additional fixed effects with no random counterpart.
The ε_{ij}'s are supposed i.i.d. $\mathcal{N}(0, \sigma^2)$ independent of the $\boldsymbol{\phi}_i$. The variance function g is dependent on f and a parameter vector α. In general $g = f^\alpha$, reflecting the possible character of intra individual variability.

2.2.2 The SAEM-MCMC algorithm

The aim of this study was to compute the maximum likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2, \alpha)$ by maximizing the likelihood of the observations $p(\boldsymbol{y}, \boldsymbol{\theta})$.

The EM algorithm (Dempster et al., 1977) is a popular and efficient iterative tool used to compute the maximum likelihood of $\boldsymbol{\theta}$. It consists on taking the random vector of the model $\boldsymbol{\phi}$ as missing data and monitoring two steps. At iteration k of the EM algorithm, the E-step evaluates $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k-1)}) = \mathbb{E}[p(\boldsymbol{y}, \boldsymbol{\phi}; \boldsymbol{\theta})|\boldsymbol{y}, \boldsymbol{\theta}^{(k-1)}]$, i.e. the conditional expectation of the log-likelihood of the complete data vector $(\boldsymbol{y}, \boldsymbol{\phi})$ given the observed data $\boldsymbol{y}$ and the current value of the vector of parameters $\boldsymbol{\theta}$. The M-step consists in computing the value $\boldsymbol{\theta}^{(k)}$ that maximizes $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k-1)})$.

The complete log-likelihood can be written as:

$$\begin{aligned}\log p(\boldsymbol{y}, \boldsymbol{\phi}; \boldsymbol{\theta}) = \quad & const - \textstyle\sum_{ij} \log g(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i, \alpha) - \frac{1}{2} \sum_{ij} \frac{(y_{ij} - f(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i))^2}{g(\boldsymbol{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i, \alpha)^2} \\ & - \tfrac{N}{2} \log(|\boldsymbol{\Gamma}|) - \tfrac{1}{2} \textstyle\sum_{i=1}^{N} (\boldsymbol{\phi}_i - \boldsymbol{A}_i\boldsymbol{\mu})' \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - \boldsymbol{A}_i\boldsymbol{\mu})\end{aligned}$$

In nonlinear mixed models, function f does not depend linearly on the random effects, so the E-step is generally not in closed form. Delyon et al. (1999) propose to replace the E-step of EM by a Simulation step and a Stochastic Approximation step. The stochastic approximation step is a way to reduce the amount of simulations we need in a classical Monte Carlo EM, by "recycling" the simulated vectors. In the simulation step, missing values $\boldsymbol{\phi}$ are simulated under the distribution of $\boldsymbol{\phi}$ given the observed data and $\boldsymbol{\theta}$, noted

$p(\boldsymbol{\phi}|\boldsymbol{y},\boldsymbol{\theta})$. When this distribution is unknown, Kuhn and Lavielle (2004, 2005) propose to combine the SAEM algorithm with a MCMC method such as the Metropolis-Hastings algorithm.
In the case where the M-step may not be directly performed, we use a Newton-Raphson method, as proposed in Kuhn and Lavielle (2005) and Wang (2007).

Given $\boldsymbol{\theta}^{(k-1)}$, estimated parameter value of $\boldsymbol{\theta}$ at iteration $k-1$, the SAEM-MCMC algorithm is composed of the following steps at iteration k:

- Simulation step: Draw L random vectors $(\boldsymbol{\phi}^{(k,l)}, l = 1, ..., L)$ under the conditional distribution $p(\boldsymbol{\phi}|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)})$ by simulating L independent chains in the Metropolis-Hastings algorithm presented in section 2.3.

- Stochastic approximation step:
  We define $\boldsymbol{S}_{1,i}^{(k)} = \mathbb{E}[\boldsymbol{\phi}_i|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}]$ and $\boldsymbol{S}_{2,i}^{(k)} = \mathbb{E}[\boldsymbol{\phi}_i\boldsymbol{\phi}_i'|\boldsymbol{y},\boldsymbol{\theta}^{(k-1)}]$, where the expectations are taken under the conditional distribution of the missing data $\boldsymbol{\phi}$ given the observations and the current value of $\boldsymbol{\theta}$.
  We update the following quantities used to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$:

$$\boldsymbol{S}_{1,i}^{(k)} = \boldsymbol{S}_{1,i}^{(k-1)} + \gamma_k\big[\sum_{l=1}^{L}\boldsymbol{\phi}_i^{(k,l)}/L - \boldsymbol{S}_{1,i}^{(k-1)}\big]$$

$$\boldsymbol{S}_{2,i}^{(k)} = \boldsymbol{S}_{2,i}^{(k-1)} + \gamma_k\big[\sum_{l=1}^{L}\boldsymbol{\phi}_i^{(k,l)}\boldsymbol{\phi}_i^{(k,l)\,\prime}/L - \boldsymbol{S}_{2,i}^{(k-1)}\big]$$

  where $(\gamma_k)_k$ is a decreasing sequence of positive scalars such that $\gamma_k = 1$ if $k < K$ and $\gamma_k = (k-K)^{-1}$ otherwise.

  Regarding $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma^2, \alpha)$, we have the following recursion:

$$\boldsymbol{\eta}^{(k)} = \boldsymbol{\eta}^{(k-1)} - \gamma_k(\boldsymbol{B}_k^{-1}\boldsymbol{M}_k)$$

  where $\boldsymbol{B}_k = \sum_{l=1}^{L}\partial^2_{\boldsymbol{\eta}} log\; p(\boldsymbol{y},\boldsymbol{\phi}^{(k,l)};\boldsymbol{\theta})/L$ and $\boldsymbol{M}_k = \sum_{l=1}^{L}\partial_{\boldsymbol{\eta}} log\; p(\boldsymbol{y},\boldsymbol{\phi}^{(k,l)};\boldsymbol{\theta})/L$.

- Maximization step:
  After maximizing the Q function with respect to the parameters, we obtain:

$$\boldsymbol{\mu}^{(k)} = \Big(\sum_{i=1}^{N}\boldsymbol{A}_i'\big(\boldsymbol{\Gamma}^{(k-1)}\big)^{-1}\boldsymbol{A}_i\Big)^{-1}\sum_{i=1}^{N}\boldsymbol{A}_i'\big(\boldsymbol{\Gamma}^{(k-1)}\big)^{-1}\boldsymbol{S}_{1,i}^{(k)}$$

$$\boldsymbol{\Gamma}^{(k)} = \tfrac{1}{N}\sum_{i=1}^{N}\Big(\boldsymbol{S}_{2,i}^{(k)} - \boldsymbol{A}_i\boldsymbol{\mu}^{(k)}\boldsymbol{S}_{1,i}^{(k)\prime} - \boldsymbol{S}_{1,i}^{(k)}\boldsymbol{\mu}^{(k)\prime}\boldsymbol{A}_i' + \boldsymbol{A}_i\boldsymbol{\mu}^{(k)}\boldsymbol{\mu}^{(k)\,\prime}\boldsymbol{A}_i'\Big)$$

and $\boldsymbol{\beta}^{(k)}$, $\sigma^{2\,(k)}$, $\alpha^{(k)}$ are estimated as the components of $\boldsymbol{\eta}^{(k)}$

Remark In order to have the convergence of the sequence $(\boldsymbol{\theta}^{(k)})_k$, the sequence $(\gamma_k)_k$ has to be chosen such that each γ_k must belong to [0,1] and the series $\sum \gamma_k$ must diverge, $\sum \gamma_k^2$ must converge. Kuhn and Lavielle (2004) propose to take the sequence $(\gamma_k)_k$ such that $\gamma_k = 1$ for $1 \leq k \leq K$ and $\gamma_k = (k-K)^{-1}$ else, where K is an integer that can be fixed between 50 and 100. In practice since the sequence $(\gamma_k)_k$ is decreasing quickly, the choice of K is important.

2.2.3 The Metropolis-Hastings algorithm

Under mild conditions (Kuhn and Lavielle, 2004), the Metropolis-Hastings algorithm produces L independent ergodic markov chains with stationnary distribution $p(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta})$. These chains are used in the simulation step of the SAEM-MCMC algorithm.

We define $m_1 \in \mathbb{N}$ the iteration at which we change of instrumental distribution, $\rho_1 \in \mathbb{R}^+$ and $\rho_2 \in \mathbb{R}^+$.

At the kth step of the SAEM-MCMC algorithm, we note $\boldsymbol{\phi}_i^{(k,l,t-1)} = \boldsymbol{x}^{(t)}$ the element obtained at the $(t-1)$th step of the lth chain simulated by the Metropolis-Hastings algorithm. At the tth step, the Metropolis-Hastings algorithm is implemented as follows.

- If $t < m_1$, then generate $\boldsymbol{W}_t \sim \mathcal{N}(\boldsymbol{A}_i \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Gamma}^{(k-1)}) \rightarrow$ the value obtained is noted $\boldsymbol{w}_t$.
  Let the acceptance rate: $\rho(\boldsymbol{x}^{(t)}, \boldsymbol{w}_t) = \min\left(\frac{p((y_{ij})_j|\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{w}_t}}{p((y_{ij})_j|\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{x}^{(t)}}}, 1\right)$
- If $t \geq m_1$, then generate $\rho \sim \mathcal{U}_{[\rho_1,\rho_2]}$ and $\boldsymbol{W}_t \sim \mathcal{N}(\boldsymbol{x}^{(t)}, \rho\boldsymbol{\Gamma}^{(k-1)})$
  This instrumental distribution is slightly different of the one used in Monolix software. the value obtained is noted w_t.
  The acceptance rate is: $\rho(\boldsymbol{x}^{(t)}, \boldsymbol{w}_t) = \min\left(\frac{\left[p((y_{ij})_j|\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{w}_t}\right]\left[p(\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{w}_t}\right]}{\left[p((y_{ij})_j|\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{x}^{(t)}}\right]\left[p(\boldsymbol{\phi}_i;\boldsymbol{\theta}^{(k-1)})_{|\boldsymbol{\phi}_i=\boldsymbol{x}^{(t)}}\right]}, 1\right)$
- $\boldsymbol{\phi}_i^{(k,l,t)} = \begin{cases} \boldsymbol{w}_t & \text{with probability } \rho(\boldsymbol{x}^{(t)}, \boldsymbol{w}_t) \\ \boldsymbol{x}^{(t)} & \text{with probability } 1-\rho(\boldsymbol{x}^{(t)}, \boldsymbol{w}_t) \end{cases}$

For the sake of simplicity, we note $\boldsymbol{\phi}_i^{(k,l)} = \boldsymbol{\phi}_i^{(k,l,T)}$, where T is the length of the chain.

2.2.4 Estimations of the log-likelihood and standard errors

The computation of the likelihood of the observations is needed to perform the likelihood ratio test. It can be approximated using an importance sampling integration method (Robert and Casella, 2004; Walker, 1996):

$$\hat{p}(\boldsymbol{\theta}; \boldsymbol{y}) = \frac{1}{S}\sum_{j=1}^{S} p(\boldsymbol{y}|\boldsymbol{\phi}_j, \boldsymbol{\theta})\frac{p(\boldsymbol{\phi}_j, \boldsymbol{\theta})}{h(\boldsymbol{\phi}_j, \boldsymbol{\theta})}$$

where $p(\boldsymbol{\phi}_j, \boldsymbol{\theta})$ is the distribution of the missing data and $(\boldsymbol{\phi}_j)_{j=1,\ldots,S}$ are drawn from the importance distribution $h(\boldsymbol{\phi}_j, \boldsymbol{\theta})$. The choice of the function h that minimizes the variance of the estimator $\hat{p}(\boldsymbol{\theta}; \boldsymbol{y})$ is $p(\boldsymbol{\phi}_j|\boldsymbol{y}, \hat{\boldsymbol{\theta}})$ (Robert and Casella, 2004). Since this density is unknown, we choose for $h(\boldsymbol{\phi}_j, \boldsymbol{\theta})$ a Gaussian distribution with mean $\mathbb{E}(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta})$ and variance $Var(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta})$ calculated using the simulated $\boldsymbol{\phi}_j$'s when convergence is reached for $\boldsymbol{\theta}$.

The computation of the standard errors (SE) of the estimated parameters is needed to perform the Wald test. The SE can be evaluated as the diagonal elements of the inverse of the Fisher information matrix estimate, of which evaluation is complex because it has no analytic form.
Using the Louis' formula (Louis, 1982) and following Delyon et al (1999), we approximate the observed information matrix of $\hat{\boldsymbol{\theta}}$ as follows:
At the stochastic approximation step, we compute

$$\boldsymbol{R}_1 = \sum_{l=1}^{L} \partial_{\boldsymbol{\theta}} \log p(\boldsymbol{y}, \boldsymbol{\phi}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})/L$$

$$\boldsymbol{R}_2 = \sum_{l=1}^{L} \partial^2_{\boldsymbol{\theta}} \log p(\boldsymbol{y}, \boldsymbol{\phi}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})/L$$

$$\boldsymbol{R}_3 = \sum_{l=1}^{L} \left(\partial_{\boldsymbol{\theta}} \log p(\boldsymbol{y}, \boldsymbol{\phi}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})\right) \left(\partial_{\boldsymbol{\theta}} \log p(\mathbf{y}, \boldsymbol{\phi}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})\right)'/L$$

Then

$$\boldsymbol{\Delta}_k = \boldsymbol{\Delta}_{k-1} + \gamma_k[\boldsymbol{R}_1 - \boldsymbol{\Delta}_{k-1}]$$
$$\boldsymbol{G}_k = \boldsymbol{G}_{k-1} + \gamma_k[\boldsymbol{R}_2 + \boldsymbol{R}_3 - \boldsymbol{G}_{k-1}]$$
$$\boldsymbol{H}_k = \boldsymbol{G}_k - \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k'$$

and we approximate the Fisher information matrix by the inverse of $-\mathbf{H}_k$ at convergence of $\hat{\boldsymbol{\theta}}$.

2.3 The criteria

In practice, the following parameters need to be calibrated:
- the value of m_1 (iteration at which we change of instrumental distribution),
- the parameters of the second instrumental distribution ρ_1 and ρ_2,
- the value of T (the length of the chain),
- the value of L (the number of independent chains),
- the value of K, iteration at which the stochastic approximation step is effective,
- the number of iterations in the SAEM algorithm, noted I.

These parameters are defined as follows: we run a markov chain with the Metropolis-Hastings algorithm for one representative individual of the population studied, taking the initial value $\boldsymbol{\theta}_0$ for the vector of parameter $\boldsymbol{\theta}$. Then we calibrate the parameters with the following methods and we use them for all individuals.

• **Parameters of the Metropolis-Hastings algorithm**:

The simulated markov chain does not have the same behavior in the first and in the second part of the chain. Figure 2.1 shows a simulated chain using the Metropolis-Hastings algorithm. Parameter m_1 was fixed at 50,000 and the length of the chain at $100,000$. In the first 50,000 iterations with the first instrumental distribution, the markov chain stays during many iterations at the same values on the contrary to the second part of the markov chain (from iteration 50,000 to 100,000) where it moves around a certain mean value. This fact can be explained by the following: the first instrumental distribution simulates larger variables than the second instrumental distribution and so these variables are more rejected in the first case than in the second case.
Nevertheless the first instrumental distribution is useful when initial values $\boldsymbol{\theta}_0$ are far from the maximum likelihood estimate of $\boldsymbol{\theta}$; the simulations are more scattered. In this sense, we calibrated $m_1 = T/10$, where T is the length of the markov chain.

In the first part of the chain, we simulate $\boldsymbol{W}_t$ with the prior distribution of the random effects. In the second part of the chain, $\boldsymbol{W}_t$ is simulated with a Gaussian random walk centered on the precedent iteration of the chain. We fixed ρ_1 and ρ_2 such that the acceptance rate was between 25% and 40 %, as proposed in Robert and Casella (2004).

Is it necessary that the chain reaches the stationary distribution to better estimate the conditional expectations? Indeed in the MCEM algorithm (Wei and Tanner, 1990), the longer the chains are, the more precise are the estimations of the conditional expectations. Nevertheless in the SAEM-MCMC algorithm the stochastic approximation step may be a way to exempt the convergence of the chain towards the stationnary distribution. In this study, we prefer to run enough iterations and test the convergence of the chains towards the stationnary distribution thanks to the Gelman and Rubin (1992) criterion. This criterion provides a diagnosis of the markov chain convergence by comparing within-chain and between-chain variances. With these considerations, we can determine the length of the markov chain (T).

As proposed by Kuhn and Lavielle, convergence of the algorithm can be improved by running L markov chains instead of only one in the simulation step of the SAEM algorithm. In practice we can run several chains (L is between 1 and 10) and compare the behavior of each of them. If their properties (mean, variance) are very different, we may choose $L = 5$ or $L = 10$ chains, else, $L = 1$ chain can be enough.

• **Stopping rule and "smoothing criterion" for the SAEM algorithm** :
We note

$$\boldsymbol{e}^{(k)} = \left(\frac{|\boldsymbol{\theta}_j{}^{(k)} - \boldsymbol{\theta}_j{}^{(k-1)}|}{|\boldsymbol{\theta}_j{}^{(k)}|} \right)_j \tag{2.1}$$

a vector of same length as parameter vector $\boldsymbol{\theta}$ and we note $e_j^{(k)}$ the jth component of $\boldsymbol{e}^{(k)}$.

In order to determine parameter K, the iteration at which the sequence $(\gamma_k)_k$ is decreasing in the SAEM-MCMC algorithm, we studied a "smoothing" criterion that is based on the variation of the $e(i)$'s.
At each iteration k (for $k > 9$) of the SAEM algorithm, we fit a linear regression on the ten last iterations of each component of $\boldsymbol{e}$. When the slope of the linear regression is not increasing, we perform 15 additional iterations to ensure that we are really in the neighborhood of the maximum likelihood estimator. So we calculate the slope

$$\zeta_j^{(k)} = \frac{\sum_{m=1}^{10} \left(m - \frac{1}{10} \sum_{s=1}^{10} s \right) \left(e_j^{(k-m+1)} - \frac{1}{10} \sum_{s=0}^{9} e_j^{(k-s)} \right)}{\sum_{m=1}^{10} (m - \frac{1}{10} \sum_{s=1}^{10} s)^2}$$

And we define K by

$$K = 15 + \min \left\{ k;\, \forall j,\, \zeta_j^{(k)} - \zeta_j^{(k-1)} < 0 \right\}$$

Figure 2.2 represents an illustration of the evolution for one component e_1 of $\boldsymbol{e}$ (for example a fixed parameter of the model) during the iterations of the SAEM algorithm. With the proposed procedure, K was fixed at 43 iterations. We can note that the slope of the curve is high at the beginning and then after iteration around 25, e_1 varies slightly around a small positive value. At this moment, the curve can be smoothed.

We use the stopping rule suggested by Booth and Hobert (1999) and used by Wang (2007) in similar problems to diagnose the convergence. We propose to stop the SAEM algorithm when $\max_j e_j^{(k)} < \delta_2$ with $\delta_2 = 0.0001$. We note I the number of iterations of the SAEM-MCMC algorithm.

2.4 Application and simulation

In this section, we use real data and simulated data to compare our SAEM-MCMC algorithm with parameters calibrated using our criteria, called "SAEM-MCMC with criteria", with several versions of the SAEM-MCMC algorithm using six different sets of parameters presented in Table 2.1 and with the SAS NLMIXED procedure based on the approximation of the likelihood of the observations by adaptative Gaussien quadrature.

SAEM-MCMC with criteria is implemented via a fortran program and all computations are carried out in Fortran 95. We choose the well known orange tree data, presented by Pinheiro and Bates (1995a, 1995b) and available for example on Splus, to illustrate our algorithm. This data set consists of seven measurements of the trunk circumference of each of five orange trees.
Following Pinheiro and Bates (1995a), a logistic curve is used for modelling the trunk circumference y_{ij} of tree i at age t_{ij}

$$\boldsymbol{y}_{ij} = \frac{\phi_{1i}}{1 + \exp\big(-\frac{t_{ij}-\phi_{2i}}{\beta}\big)} + \varepsilon_{ij},\ \forall i \in \{1, ..., N\}\ \forall j \in \{1, .., n_i\} \tag{2.2}$$

The errors $(\varepsilon_{ij})_{ij}$'s are assumed i.i.d. $\mathcal{N}(0, \sigma^2)$, the $\boldsymbol{\phi}_i = (\phi_{1i}, \phi_{2i})$'s are i.i.d. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\Gamma} = \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix}$ is the covariance matrix of the $\boldsymbol{\phi}_i$'s.
Let $\boldsymbol{\theta} = (\mu_1, \mu_2, \beta, \boldsymbol{\Gamma}, \sigma^2)$ the vector of parameters.

The parameters $\phi_{1i}, \phi_{2i}, \beta$ have a physical interpretation: ϕ_{1i} corresponds to the asymptotic trunk circumference, ϕ_{2i} represents the age at which the tree attains half of its asymptotic trunk circumference, and β is the growth scale. Here we assume that parameter β is a fixed effect.

Several versions of the SAEM-MCMC algorithm using six different sets of parameters (Table 2.1) are considered: five sets of parameters are chosen such as the importance of each parameter is revealed, and the last set of parameters is calibrated using the criteria presented in section 3.
For example, set 1 is the set of parameters given in Kuhn and Lavielle (2005): 10 iterations with the first instrumental distribution, 0 with the second one, i.e. $T = 10$ and $m_1 = 11$, $L = 5$ independent chains, and K is fixed at 100 iterations.
In sets 1 and 2, only the first instrumental distribution is used, with longer markov chains in set 2 than in set 1. In sets 3 and 4, we use the two instrumental distributions, with longer chains in set 4 than in set 3. Set 5 is composed of the same parameters of set 4 but with the value of K (iteration at which the sequence $(\gamma_k)_k$ is decreasing) fixed at 50 iterations.
The exact version of SAS NLMIXED is used to validate the best set of parameters.

Then in order to check the gain obtained with the criteria in the estimation of the maximum likelihood estimator of θ, we simulate 100 data sets on the same data structure.
Data are simulating using the following parameters:
$N = 100$, $n_i = 15$, $\mu_1 = 20$, $\mu_2 = 70$, $\beta = 30$, the 15 points of observations t_{ij} are following: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, $\boldsymbol{\Gamma} = \begin{pmatrix} 10 & -1 \\ -1 & 40 \end{pmatrix}$ and $\sigma^2 = 0.5$.

For the real data set, the following initial values are used: $\mu_1^{(0)}$=150, $\mu_2^{(0)}$=600, $\beta^{(0)}$=200,

$\tau_1^{2(0)}$=500, $\tau_2^{2(0)}$=200, $\tau_{12}^{(0)}$=0 and $\sigma^{2(0)}$=10.
And for the simulating data sets, the following initial values are used: $\mu_1^{(0)}$=10, $\mu_2^{(0)}$=35, $\beta^{(0)}$=15, $\tau_1^{2(0)}$=5, $\tau_2^{2(0)}$=30, $\tau_{12}(0)$=0 and $\sigma^{2(0)}$=0.1.

Our criteria calibrate the parameters of the SAEM-MCMC algorithm for the real data: $T = 300$, $m_1 = 30$, with a Gelman and Rubin criterion at 1.06, $L = 5$, $\rho_1 = 0$, $\rho_2 = 0.4$ with an average acceptation rate at 25%. The "smoothing" criterion is fixed at $K = 37$ iterations, and the stopping rule stopped the algorithm at $I = 172$ iterations. The computing time is equal to 28 seconds.
Results are presented in Table 4.1. Except for set 1, the values of $\hat{\theta}$ obtained with the SAEM-MCMC algorithm and the different values of parameters considered are similar and close to the ones obtained with our algorithm with criteria and the NLMIXED procedure. However, we can note that adding a second instrumental distribution in the Metropolis-Hastings algorithm improve estimates, in particular the variance-covariance parameters. In the same way, longer are the markov chains in the Metropolis-Hastings algorithm, more precise are the estimators. In this study, the choice of parameter K does not seem to be important: the results obtained with sets 4 and 5 are similar.
Finally, results obtained with the SAEM-MCMC algorithm with criteria indicate that it is not necessary to run many iterations with the first instumental distribution ($m_1 = 30$). The "smoothing" criterion K starts at 38 iterations, indicating that the algorithm converges rapidly towards the maximum likelihood estimates. The results obtained are close to the NLMIXED estimates.

The criteria are used to calibrate the parameters of the SAEM-MCMC algorithm for the first simulated data set. We obtain: $T = 800$, $m1 = 80$, with Gelman and Rubin criterion at 1.04, $L = 1$, $\rho_1 = 0, \rho_2 = 0.4$ with an average rate of acceptation at 30%. Then the same parameters are used for the other data sets. We obtain a mean parameter K equal to 38.
Estimations of mean estimates (fixed effects and variance-covariance components) on the simulated data set are presented in Table 2.3. Mean time of running of the SAEM-MCMC algorithm is noted *time* (in seconds) and let $\bar{I}$ the mean number of iterations in the SAEM-MCMC algorithm. Means are computed on the 100 simulated data sets.
Table 2.4 presents the bias, the Mean Squared Error (MSE) of the estimator of $\boldsymbol{\theta}$ obtained with the six parameters sets of the SAEM-MCMC algorithm considered and the NLMIXED results. Standard errors of parameter estimates and MQE are also presented in Table 2.5 for our method (the same results are also obtained with sets 4 and 5) and NLMIXED procedure.

Concerning the estimates of μ_1, μ_2 and β, we obtained same results for all the procedures. For the estimates of the variance-covariance parameters τ_1^2, τ_2^2, τ_{12} and σ^2, sets 1 to 4 provided heterogeneous results. Estimates from sets 1 and 3 are far from estimates obtained with sets 2 and 4, and the comparison of the log-likelihood value at the estimates implies that using long markov chains provide better estimates of the parameters

of variances and covariances.
We obtained better estimates and smaller MSE for set 3 than for set 1. Then the comparison between outputs from sets 2 and 4 implies we obtain better estimates when we use two instrumental distributions instead of one.
We obtained similar results for sets 4 and 5, so parameter K is not necessary to calibrate at a first time for these data.
We compared computational time obtained with sets 4, 5 and SAEM-MCMC with criteria procedure: our method provided the best time.
With sets 4 and 5, SAEM-MCMC with criteria and NLMIXED procedures provided same estimates, biases, MSE and MQE.

2.5 Discussion - Conclusion

At a first time, we can see that our criteria provided different parameters for the simulated data set and the real data set, in particular for the parameters of the Metropolis-Hastings algorithm. This is due to the different variability within and between the markov chains in the two cases. For the two data sets, the outputs obtained with the different methods are not similar, essentially for the variance-covariance parameters. We need to run long chains in the Metropolis-Hastings algorithm, using two instrumental distributions, to obtain precise and close estimates to the ones obtained with the exact SAS-NLMIXED procedure.
Gelman and Rubin criterion implies that the chains reach the stationary distribution in the Metropolis-Hastings algorithm. Here, applications illustrated that in this case, convergence of the SAEM-MCMC is ensured. Nevertheless, we obtained close results with simulating shorter chains in the simulated data set. So this criterion may be stringent in many studies.
Comparing the computational time of all the SAEM-MCMC versions with estimates close to the true value of the vector of parameters, our method was not computationally high time.

In summary, the SAEM algorithm combines the statistical properties of an exact maximum likelihood method with computational efficiency. This algorithm allows to use classical model comparison criteria such as AIC (the Information Criterion of Akaike) and BIC (the Bayesian Information Criterion of Schwartz) for longitudinal data analysis using nonlinear mixed effects models. In this article, we propose some criteria to calibrate the different parameters of the SAEM-MCMC algorithm, and slight differences on the instrumental distributions are proposed compared to the ones implemented in Monolix software. We show on the Orange tree data and on a simulated data set that these criteria are relevant and we obtain close parameter estimates, standard errors, biases and MSE to the SAS-NLMIXED ones, with a good performance of computational time.

2.6 Acknowledgment

The authors are grateful to Jean-Louis Foulley, Jean-Michel Marin, Didier Concordet, Djalil Chafaï, Julie Antic and Béatrice Laurent for interesting discussions and useful comments.

2.7 References

Booth GJ, Hobert PJ (1999) Maximizing generalized linear mixed models likelihoods with an automated Monte Carlo EM algorithm. J. Roy. Statist. Soc. B, 61,265-285.
Davidian M, Giltinan DM (1995) Nonlinear Models for repeated Measures Data. Chapman & Hall, New York.
Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the EM algorithm. Ann. Statist., 27(1), 94-128.
Dempster AP, Laird NM, Rubin, DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B, 39, 1-38.
Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-511.
Kuhn E, Lavielle M (2004) Coupling a stochastic approximation version of EM with a MCMC procedure. ESAIM Probab. Stat., 8, 115-131.
Kuhn E, Lavielle M (2005) Maximum Likelihood estimation in non linear mixed effects models. Comput. Statist. Data Anal., 49, 1020-1038.
Lindstrom M, Bates D (1990) Nonlinear mixed effects models for repeated measures data. Biometrics, 44, 673-687.
Louis TA (1982) Finding the observed information matrix when using the EM algorithm. J. Roy. Stat. Soc. Ser. B, 44 (2), 226-233.
Pinheiro JC, Bates DM (1995a) Mixed-effects Models in S and S-PLUS. Springer, New York.
Pinheiro JC, Bates DM (1995b) Approximations to the log-likelihood function in the nonlinear mixed effects model. J. Comput. Graph. Statist., 4, 12-35.
Robert CP, Casella G (2004) Monte Carlo Statistical methods. Springer, New York.
Sheiner LB, Beal SL (1980) Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. J. Pharm. Biopharm., 8, 553-571.
Walker S (1996) An EM algorithm for nonlinear random effects models. Biometrics, 52, 934-944.
Wang J (2007) EM algorithms for nonlinear mixed effects models. Comput. Statist. Data Anal., 51, 3244-3256.
Wei GCG, Tanner MA (1990) A Monte Carlo implementation of the EM algorithm and the Poor's Man's data augmentation algorithms. J. Amer. Statist. Assoc., 85 (411), 699-704.

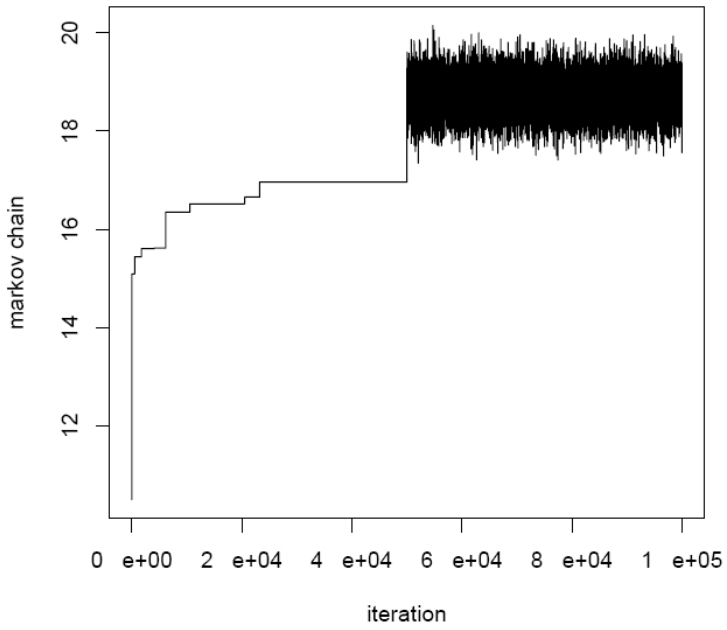


Figure 2.1: Simulation of a markov chain of the Metropolis-Hastings algorithm.

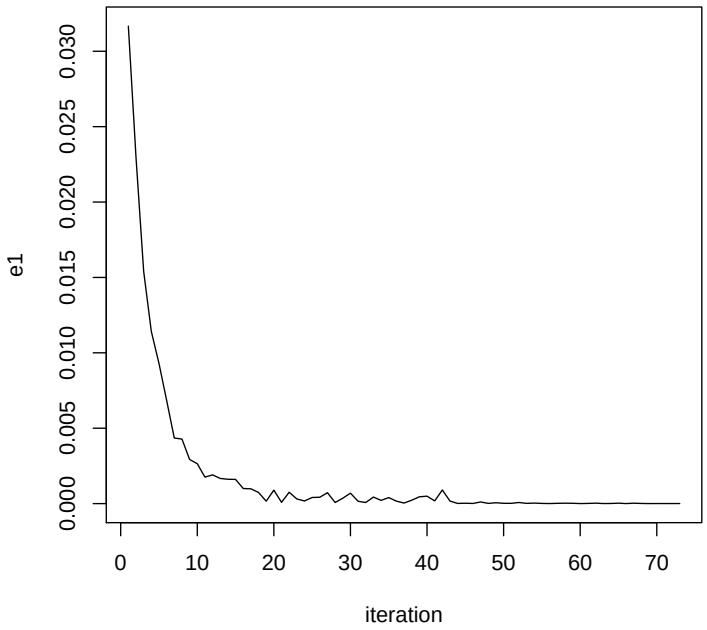


Figure 2.2: Illustration of the evolution of e_1 during the iterations of the SAEM-MCMC algorithm.

Set number	T	m_1	L	K
set 1	10	11	5	100
set 2	100	101	5	100
set 3	20	11	5	100
set 4	200	101	5	100
set 5	200	101	5	50

T: number of iterations in the SAEM algorithm

m_1: iteration at which we change of instrumental distribution in the Metropolis-Hastings algorithm

L: number of independent chains simulated in the Metropolis-Hastings algorithm

K: iteration at which the stochastic approximation step is effective in the SAEM algorithm

Table 2.1: Presentation of the five sets of parameters used in the SAEM-MCMC algorithm.

$\boldsymbol{\theta}$	μ_1	μ_2	β	τ_1^2	τ_2^2	τ_{12}	σ^2	$LogL^{(a)}$	$I^{(b)}$
Initial value	150	600	200	500	200	0	10	-	-
set 1	173.80	611.32	288.43	634.07	0.14	-7.16	192.08	-139.40	363
set 2	188.37	700.54	336.80	1150.03	660.80	871.12	59.50	-131.06	173
set 3	177.74	631.45	301.72	891.28	509.79	674.03	130.59	-136.42	335
set 4	193.14	727.83	353.15	1224.71	849.54	975.64	58.06	-130.96	264
set 5	192.1	722.3	349.3	1185.00	730.74	843.81	58.10	-130.87	132
SAEM-MCMC with criteria	190.60	713.72	344.00	1169.23	983.73	876.79	57.04	-130.87	172
NLMIXED	191.34	717.03	346.50	1198.87	1468.33	1017.603	55.99	-130.75	-

$^{(a)} LogL = \log \hat{p}(\boldsymbol{\theta}; \boldsymbol{y})$

$^{(b)} I$: number of iterations in the SAEM algorithm

Table 2.2: Values of $\hat{\boldsymbol{\theta}}$ for the SAEM-MCMC algorithm with the six sets of parameters, and NLMIXED procedure on the Orange trees data.

$\boldsymbol{\theta}$	μ_1	μ_2	β	τ_1^2	τ_2^2	τ_{12}	σ^2	$LogL^{(a)}$	$time^{(b)}$	$\bar{I}^{(c)}$
True value	20	70	30	10	40	-1	0.5	-	-	-
Initial value	10	35	15	5	30	0	0.1	-	-	-
set 1	19.12	69.90	27.96	4.18	2.44	-1.79	2.55	-2630.41	1.20	268
set 2	19.38	71.50	28.61	8.99	27.90	-3.13	0.72	-2100.18	7.46	270
set 3	19.40	71.73	28.81	6.62	8.94	-4.20	1.70	-2346.14	2.27	249
set 4	19.41	71.52	28.67	9.74	39.04	-2.21	0.60	-2071.35	11.24	200
set 5	19.41	71.52	28.68	9.74	39.04	-2.22	0.60	-2071.25	8.13	142
SAEM-MCMC with criteria	19.41	71.51	28.67	9.77	40.17	-2.09	0.59	-2071.25	7.97	143
NLMIXED	19.41	71.49	28.68	9.81	40.15	-2.12	0.59	-2071.00	-	-

The mean estimates are based on 100 simulations.

$^{(a)}LogL = \log \hat{p}(\boldsymbol{\theta}; \boldsymbol{y})$

$^{(b)}time$: computational mean time in seconds

$^{(c)}\bar{I}$: mean number of iterations in the SAEM algorithm

Table 2.3: Estimation of the mean estimates on the simulated data set.

$\boldsymbol{\theta}$	μ_1	μ_2	β	τ_1^2	τ_2^2	τ_{12}	σ^2
Bias							
set 1	-0.88	-0.10	-2.04	-5.82	-37.56	-0.79	2.05
set 2	-0.62	1.50	-1.39	-1.00	-12.10	-2.13	0.22
set 3	-0.60	1.73	-1.19	-3.38	-31.06	-3.20	1.20
set 4	-0.59	1.52	-1.33	-0.26	-0.96	-1.21	0.10
set 5	-0.59	1.52	-1.32	-0.26	-0.96	-1.22	0.10
SAEM-MCMC with criteria	-0.59	1.51	-1.33	-0.23	0.17	-1.09	0.09
NLMIXED	-0.59	1.49	-1.32	-0.19	0.15	-1.12	0.09
MSE							
set 1	0.87	1.01	4.45	34.64	1413.27	2.65	4.27
set 2	0.46	2.65	2.00	3.25	172.70	8.63	0.05
set 3	0.45	3.49	1.54	12.80	975.30	13.00	1.47
set 4	0.43	2.69	1.81	2.49	36.66	5.30	0.01
set 5	0.43	2.70	1.80	2.49	36.60	5.34	0.01
SAEM-MCMC with criteria	0.43	2.66	1.81	2.48	36.94	5.06	0.01
NLMIXED	0.43	2.59	1.80	2.36	35.77	5.20	0.01

The biases and MSE of the estimates are based on 100 simulations.

Table 2.4: Estimation of the biases and MSE of the estimates on the simulated data set.

$\boldsymbol{\theta}$	μ_1	μ_2	β	τ_1^2	τ_2^2	τ_{12}	σ^2
SAEM-MCMC with criteria							
$\hat{\boldsymbol{\theta}}$	19.41	71.51	28.67	9.77	40.17	-2.09	0.59
$\hat{\sigma}(\hat{\boldsymbol{\theta}})$	0.32	0.69	0.23	1.41	6.52	2.16	0.02
MQE	0.53	3.15	1.86	4.50	80.29	9.80	0.01
NLMIXED							
$\hat{\boldsymbol{\theta}}$	19.41	71.49	28.68	9.81	40.15	-2.12	0.59
$\hat{\sigma}(\hat{\boldsymbol{\theta}})$	0.32	0.69	0.23	1.41	6.52	2.16	0.02
MQE	0.53	3.08	1.86	4.40	78.97	9.94	0.01

The estimates, standard errors and MQE of the estimates are based on 100 simulations.

Table 2.5: Value of estimates, standard errors and MQE of $\hat{\boldsymbol{\theta}}$ for our SAEM-MCMC procedure and the NLMIXED procedure on the simulated data.

Chapitre 3

Modélisation et estimation des variances hétérogènes dans les modèles non linéaires mixtes

L'article suivant a été soumis à "Statistical Modelling" le 9 avril 2008. Nous n'avons pas eu de retour à ce jour.

Une annexe à ce chapitre est présentée à la fin du manuscrit. Le jeu de données concernant une expérience de sélection sur des poulets, menée par F. Ricard (INRA, Station de Nouzilly), y est présenté, ainsi que les codes et algorithmes utilisés dans ce chapitre, dans le cas particulier d'un modèle de variance hétérogène.

Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm

Mylène Duval[(1)*], Christèle Robert-Granié[(1)], Jean-Louis Foulley[(2)]

[(1)] *INRA UR 0631 SAGA, Chemin de Borde Rouge - BP 52627, F-31 326 Castanet-Tolosan, France*
[*] *responsible author*
Mylene.Duval@toulouse.inra.fr
Christele.Robert-Granie@toulouse.inra.fr
[(2)] *INRA UR 0337 SGQA, domaine de Vilvert, F-78350 Jouy-en-Josas, France*
jean-louis.foulley@jouy.inra.fr

Abstract

The SAEM-MCMC algorithm is a powerful tool for computing maximum likelihood estimators in the wide class of nonlinear mixed effects models. We propose in this article an adaptation of this algorithm to the estimation of heterogeneous variances in such models. Two residual variance models are considered: a linear mixed model on the logvariance, with fixed and random effects, and a mean-variance relationship. As compared to other procedures implemented in R, SAS and Monolix, our algorithm provides more flexibility in modelling variance functions and reliability in the outputs.
This algorithm was numerically validated in the case of a heteroskedastic linear mixed model by comparing its outputs with those of a standard EM algorithm applied to Pothoff and Roy's data. Finally, an application to real data involving a selection experiment on growth in chicken is presented in which that algorithm was compared to outputs of SAS-Nlmixed, nlme, Monolix and WinBUGS softwares.

Key-words: nonlinear mixed models, SAEM-MCMC, heteroskedasticity, maximum likelihood estimation.

3.1 Introduction

Nonlinear mixed models (NLMM) are the tools of choice for analyzing unbalanced repeated data arising from complex biological mechanisms such as growth curves or pharmacokinetic and pharmacodynamics (PKPD) trials. Moreover, as compared to linear models, parameters of such models provide a better biological interpretation of the mechanisms involved, and the corresponding models are also more parsimonious.

Most studies based on these models assume that variances are homogenous across strata within the sampled population. However, this assumption turns out to be unrealistic in many practical situations encountered in various areas of applied statistics, e.g. biostatistics (Pinheiro and Bates, 2000) and economics (Judge et al., 1985; Engle, 1982). In particular, heteroskedasticity arises naturally in within-group or residual variances. There are different ways to model heterogeneous variances including, among others, mean-variance relationships (Box and Hill, 1974) and linear models on log-variances: see Davidian and Carroll (1987) for a general discussion. This study will deal with estimation methods of such variance functions in the context of nonlinear models with random effects.

To that effect, several procedures have already been proposed and investigated to estimate parameters of NLMM. The first were based on linearization of the data model such as first-order (FO) approximations (Sheiner and Beal, 1980), or better, first-order conditional expectation (FOCE) approximations (Lindstrom and Bates, 1990) or equivalently, Laplace's approximations (Wolfinger, 1993). Since these methods can be severely biased, alternatives based on exact maximum likelihood (ML) have been implemented involving either Gaussian quadratures or stochastic tools. However, integration via Gaussian quadratures can be extremely difficult or inaccurate, in particular in cases with high dimensionality. Stochastic EM is a powerful alternative to the latter, especially the so-called Stochastic Approximate EM procedure (SAEM) introduced by Delyon et al. (1999) and applied initially to NLMM by Kuhn and Lavielle (2004, 2005). Most softwares for NLMM already include heterogeneous variances in their syntax, but, either they rely on approximate inference procedures (e.g. nlme) or, if not, they often fail to converge (e.g. SAS-Nlmixed) or offer a limited choice of variance functions (e.g. Monolix).

Thus, the objective of this paper is to propose a general algorithm involving both a more diversified assortment of variance functions and more reliable estimating procedures for computing ML estimation of parameters in heteroskedastic NLMM via the SAEM-MCMC algorithm.

The paper is organized as follows. In section 2, we introduce a general class of NLMM with different residual variance functions. Then in section 3, the SAEM-MCMC algorithm is presented. In section 4, we deal with two numerical applications. The first is a validation of our algorithm on a linear mixed model. The second is an application to a real data set of growth curves in chicken, the outputs of which are compared to those of

other methods and softwares.

3.2 The heteroskedastic nonlinear mixed Model

Let us consider the general class of the following nonlinear mixed models (NLMM):

$$y_{ij} = f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i) + g(\boldsymbol{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij})\varepsilon^*_{ij} \tag{3.1}$$

where y_{ij} is the j-th observation $j \in \{1, .., n_i\}$ of the i-th individual $i \in \{1, ..., I\}$. $f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)$ describes the mean response as a function f of $\mathbf{z}_{ij}$, a vector of explanatory variables (e.g. time) and $\boldsymbol{\phi}_i$ a $(m \times 1)$ vector of random effects pertaining to individual i. It is assumed that $\boldsymbol{\phi}_i$ are i.i.d. Gaussian random variables $\mathcal{N}(\mathbf{A}_i\boldsymbol{\mu}, \boldsymbol{\Gamma})$ with mean $\mathbf{A}_i\boldsymbol{\mu}$ and variance $\boldsymbol{\Gamma}$. $\mathbf{A}_i\boldsymbol{\mu}$ specifies the link between the expectation of $\boldsymbol{\phi}_i$ and p covariates $\mathbf{A}_i$ $(m \times p)$ (e.g. gender, treatment) with coefficients $\boldsymbol{\mu}$. The variance covariance matrix $\boldsymbol{\Gamma}$ quantifies the inter-individual variation. Vector $\boldsymbol{\beta}$ corresponds to additional fixed effects involved in some elements of individual coefficients (similar to vector $\boldsymbol{\phi}_i$) but having no random counterpart.
In this model, the intra-individual variation is accounted for by the product $g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij})\varepsilon^*_{ij}$ where ε^*_{ij} are i.i.d. $\mathcal{N}(0, 1)$ random errors and $\sigma_{ij} = g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij})$ stands for the variance function described with arguments in the same way as in function f.
More specifically, we are considering the two following variance functions to describe heteroskedasticity in residual variances. The first approach consists of modelling residual variances as a structural linear mixed model involving explanatory covariates $\mathbf{w}_{ij}$, $\mathbf{q}_{ij}$ and the corresponding fixed and random effects $\boldsymbol{\delta}$, $\mathbf{v}$ via a loglink function:

$$\log(\sigma^2_{ij}) = \mathbf{w}'_{\mathbf{ij}}\boldsymbol{\delta} + \mathbf{q}'_{\mathbf{ij}}\mathbf{v} \tag{3.2}$$

where $\mathbf{v} \sim \mathcal{N}(0, \boldsymbol{\Lambda})$.
This model was proposed initially by Foulley et al. (1992) as a mixed model extension of the classical linear model for logvariances. It can be viewed as part of multilevel mixed linear models developed by Goldstein (1987) and was also investigated later on by Lin et al. (1997), Brown et al. (2002), Rigby and Stasinopoulos (2005), Lu et al. (2006) and Lee and Nelder (2006) in the context of linear and generalized linear mixed models. It provides great flexibility in modelling potential sources of variation in the residual variances with the key idea of parsimony thanks to the introduction of random effects. The second approach assumes that the variance is proportional to a power of the (conditional) mean (Box and Hill, 1974; Davidian and Carroll, 1987):

$$\sigma^2_{ij} = \delta_1 f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)^{\delta_2} \tag{3.3}$$

This is a very popular model for taking into account in an easy way a scale effect linking the variance to the mean as it arises e.g., with growth curve data (Blasco et al., 2003), household expenditures (Battese and Bonyhady, 1981) and in PKPD analysis (Davidian and Giltinan, 1995, chapter 9). Notice that for $\delta_2 = 2$, it reduces to the case of a constant

coefficient of variation which in turn is usually handled via a logtransformation of the data. The power transformation may be a very reasonable assumption but in most cases, the coefficient δ_2 remains unknown and has to be estimated.

3.3 A monitored SAEM-MCMC algorithm

3.3.1 Computation of the ML estimations

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\delta}, \boldsymbol{\Lambda})$ denote the vector of parameters of model (3.1). We have to compute the maximum likelihood (ML) estimates of these parameters by maximizing the loglikelihood of the data vector considered as a function of $\boldsymbol{\theta}$.
This can be done conveniently in mixed models via the EM algorithm (Dempster et al., 1977) using the random effects as missing data $\mathbf{u}$ and proceeding according to the following two steps.
The E-step evaluates $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ defined as the conditional expectation
$\mathbb{E}[\log p(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u})|\mathbf{y}, \boldsymbol{\theta} = \boldsymbol{\theta}^*]$ of the loglikelihood of the complete data vector $(\mathbf{y}, \mathbf{u})$ given the observed data $\mathbf{y}$ and the current value $\boldsymbol{\theta}^*$ of the parameter vector whereas the M-step maximises it with respect to the parameters so as to update the values of $\boldsymbol{\theta}^*$.

More precisely, as far as the variance function (3.2) is concerned, using $\mathbf{u} = (\boldsymbol{\phi}, \mathbf{v})$,

$$\begin{aligned}\log p(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}) = \quad & const - \textstyle\sum_{ij} \log g(\mathbf{w_{ij}}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij}) - \frac{1}{2} \sum_{ij} \frac{(y_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i))^2}{g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij})^2} \\ & -\tfrac{I}{2} \log(|\boldsymbol{\Gamma}|) - \tfrac{1}{2} \textstyle\sum_{i=1}^{I} (\boldsymbol{\phi}_i - \mathbf{A}_i \boldsymbol{\mu})' \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - \mathbf{A}_i \boldsymbol{\mu}) \\ & -\tfrac{1}{2} \log(|\boldsymbol{\Lambda}|) - \tfrac{1}{2} \mathbf{v}' \boldsymbol{\Lambda}^{-1} \mathbf{v}\end{aligned}$$

and similarly for variance function (3.3), using just $\mathbf{u} = \boldsymbol{\phi}$

$$\begin{aligned}\log p(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}) = \quad & const - \textstyle\sum_{ij} \log g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij}) - \frac{1}{2} \sum_{ij} \frac{(y_{ij} - f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i))^2}{g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_{ij})^2} \\ & -\tfrac{I}{2} \log(|\boldsymbol{\Gamma}|) - \tfrac{1}{2} \textstyle\sum_{i=1}^{I} (\boldsymbol{\phi}_i - \mathbf{A}_i \boldsymbol{\mu})' \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - \mathbf{A}_i \boldsymbol{\mu})\end{aligned}$$

Unfortunatly in NLMM's, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is not in closed form and we have to rely upon a stochastic version of it. Namely, this refers to the SAEM algorithm, defined by Delyon et al. (1999) which proved to be more efficient than a classical Monte Carlo EM due to a recycling of simulations from one iteration to the next. When simulation cannot

be performed directly, it is replaced by a MCMC procedure (e.g. Metropolis-Hastings algorithm) as proposed by Kuhn and Lavielle (2004). In practice the main problem of this SAEM-MCMC algorithm is to adequately calibrate its parameters. Here, we proceed along the same lines as defined by Duval and Robert-Granié (2007) for homoskedastic NLMM's (see details in appendix).
In addition, we achieve the M step via a Levenberg-Marquardt algorithm instead of a classical Newton-Raphson algorithm by introducing an adaptive sequence $(\lambda_k)_k$ of real numbers.

Given $\boldsymbol{\theta}^{(k-1)}$, the estimated value of $\boldsymbol{\theta}$ at iteration $k-1$, the SAEM algorithm is implemented as follows at iteration k.

- Simulation step: draw L random vectors $\left(\mathbf{u}^{(k,l)} = (\boldsymbol{\phi}^{(k,l)}, \mathbf{v}^{(k,l)}); l = 1, ..., L\right)$ from the conditional distribution $p(\mathbf{u}|\mathbf{y}, \boldsymbol{\theta}^{(k-1)})$ using L independent Metropolis-Hastings chains (see details in appendix).

- Stochastic approximation step:
  let us define $\boldsymbol{S}_{1,i}^{(k)} = \mathbb{E}[\boldsymbol{\phi}_i|\mathbf{y}, \boldsymbol{\theta}^{(k-1)}]$, $\boldsymbol{S}_{2,i}^{(k)} = \mathbb{E}[\boldsymbol{\phi}_i\boldsymbol{\phi}_i'|\mathbf{y}, \boldsymbol{\theta}^{(k-1)}]$ and $\boldsymbol{S}_3^{(k)} = \mathbb{E}[\mathbf{v}\mathbf{v}'|\mathbf{y}, \boldsymbol{\theta}^{(k-1)}]$, these quantities are updated as follows

$$\boldsymbol{S}_{1,i}^{(k)} = \boldsymbol{S}_{1,i}^{(k-1)} + \gamma_k\big[\sum_{l=1}^{L} \boldsymbol{\phi}_i^{(k,l)}/L - \boldsymbol{S}_{1,i}^{(k-1)}\big] \tag{3.4}$$

$$\boldsymbol{S}_{2,i}^{(k)} = \boldsymbol{S}_{2,i}^{(k-1)} + \gamma_k\big[\sum_{l=1}^{L} \boldsymbol{\phi}_i^{(k,l)}\boldsymbol{\phi}_i^{(k,l)\,\prime}/L - \boldsymbol{S}_{2,i}^{(k-1)}\big] \tag{3.5}$$

$$\boldsymbol{S}_3^{(k)} = \boldsymbol{S}_3^{(k-1)} + \gamma_k\big[\sum_{l=1}^{L} \boldsymbol{v}^{(k,l)}\boldsymbol{v}^{(k,l)'}/L - \boldsymbol{S}_3^{(k-1)}\big] \tag{3.6}$$

  where $(\gamma_k)_k$ is a decreasing sequence of positive real scalars such that $\gamma_k = 1$ if $k < K$ and $\gamma_k = (k-K)^{-1}$ otherwise.
  The three previous formulae pertain to parameters $\boldsymbol{(\mu, \Gamma, \Lambda)}$. Regarding $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\delta})$, we have the following recursion:

$$\boldsymbol{\eta}^{(k)} = \boldsymbol{\eta}^{(k-1)} - \gamma_k(\boldsymbol{M}_k^{-1}\boldsymbol{D}_k) \tag{3.7}$$

  where $\mathbf{D}_k = \sum_{l=1}^{L} \partial_{\boldsymbol{\eta}} \log p(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}^{(k,l)})/L$, $\mathbf{M}_k = \mathbf{B}_k + \lambda_k diag(\mathbf{B}_k)$ with
  $\mathbf{B}_k = \sum_{l=1}^{L} \partial^2_{\boldsymbol{\eta}} \log p(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}^{(k,l)})/L$ and $(\lambda_k)_k$ is an adaptative sequence of real numbers which is increasing if $Q(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k-1)})$ is itself increasing from one iteration to the next, and decreasing otherwise.

– Estimation step:
Maximisation of the Q function with respect to the parameters gives

$$\boldsymbol{\mu}^{(k)} = \left(\sum_{i=1}^{I} \boldsymbol{A}_i' \big(\boldsymbol{\Gamma}^{(k-1)}\big)^{-1} \boldsymbol{A}_i \right)^{-1} \sum_{i=1}^{I} \boldsymbol{A}_i' \big(\boldsymbol{\Gamma}^{(k-1)}\big)^{-1} \boldsymbol{S}_{1,i}^{(k)} \tag{3.8}$$

$$\boldsymbol{\Gamma}^{(k)} = \frac{1}{I} \sum_{i=1}^{I} \left(\boldsymbol{S}_{2,i}^{(k)} - \boldsymbol{A}_i \boldsymbol{\mu}^{(k)} \boldsymbol{S}_{1,i}^{(k)'} - \boldsymbol{S}_{1,i}^{(k)} \boldsymbol{\mu}^{(k)'} \boldsymbol{A}_i' + \boldsymbol{A}_i \boldsymbol{\mu}^{(k)} \boldsymbol{\mu}^{(k)\,\prime} \boldsymbol{A}_i' \right) \tag{3.9}$$

$$\boldsymbol{\Lambda}^{(k)} = \boldsymbol{S}_3^{(k)} \tag{3.10}$$

and $\boldsymbol{\beta}^{(k)}$ and $\boldsymbol{\delta}^{(k)}$ as the components of $\boldsymbol{\eta}^{(k)}$ defined previously.

These are formulae for the variance function defined in (3.2). As far as the second variance function is concerned (3.3), the same formulae apply ignoring equations (3.6) and (3.10) pertaining to $\boldsymbol{S}_3^{(k)}$ and $\boldsymbol{\Lambda}^{(k)}$.

3.3.2 Estimations of the log-likelihood and standard errors

The integration with respect to the random effects of the model involved in the likelihood of the observed data $p(\boldsymbol{\theta}; \mathbf{y})$ can be estimated conveniently via an importance sampling scheme

$$\hat{p}(\boldsymbol{\theta}; \mathbf{y}) = \frac{1}{S} \sum_{s=1}^{S} p(\mathbf{y}|\mathbf{u}_s, \boldsymbol{\theta}) \frac{p(\mathbf{u}_s, \boldsymbol{\theta})}{\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})}$$

where $p(\mathbf{u}, \boldsymbol{\theta})$ is the distribution of the missing data and $\mathbf{u}_1$,...,$\mathbf{u}_s$,...,$\mathbf{u}_S$ are drawn from the importance distribution $\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})$.
Here we choose for $\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})$, a Gaussian distribution with mean $\mathbb{E}(\mathbf{u}_s|\mathbf{y}, \boldsymbol{\theta})$ and variance $\mathbb{V}ar(\mathbf{u}_s|\mathbf{y}, \boldsymbol{\theta})$ calculated from draws of $\mathbf{u}_s$ obtained at the simulation step of the SAEM algorithm when convergence is reached for $\boldsymbol{\theta}$.
This choice should be efficient as this distribution is expected to be close to the true conditional distribution $p(\mathbf{u}_s|\mathbf{y}, \boldsymbol{\theta})$ which is known to be the one providing a minimum variance estimator of $p(\boldsymbol{\theta}; \mathbf{y})$ (Robert and Casella, 2004, theorem 3.12 page 95).

We can also obtain an estimation of the Fisher information matrix using Louis' (1982) formula. Let us define the following quantities obtained by simulation:

$$\boldsymbol{R}_1 = \sum_{l=1}^{L} \partial_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})/L$$

$$\boldsymbol{R}_2 = \sum_{l=1}^{L} \partial_{\boldsymbol{\theta}}^2 \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k-1)})/L$$

$$\boldsymbol{R}_3 = \sum_{l=1}^{L} \left(\partial_{\boldsymbol{\theta}} \log p(\boldsymbol{y},\mathbf{u}^{(k,l)};\boldsymbol{\theta}^{(k-1)})\right) \left(\partial_{\boldsymbol{\theta}} \log p(\mathbf{y},\mathbf{u}^{(k,l)};\boldsymbol{\theta}^{(k-1)})\right)'/L$$

Then we calculate at the stochastic approximation step

$$\boldsymbol{\Delta}_k = \boldsymbol{\Delta}_{k-1} + \gamma_k[\boldsymbol{R}_1 - \boldsymbol{\Delta}_{k-1}]$$

$$\boldsymbol{G}_k = \boldsymbol{G}_{k-1} + \gamma_k[\boldsymbol{R}_2 + \boldsymbol{R}_3 - \boldsymbol{G}_{k-1}]$$

$$\boldsymbol{H}_k = \boldsymbol{G}_k - \boldsymbol{\Delta}_k\boldsymbol{\Delta}_k'$$

and we approximate the Fisher information matrix by the inverse of $-\mathbf{H}^{(k)}$ at convergence of $\hat{\boldsymbol{\theta}}$.

3.4 Numerical applications

The SAEM-MCMC algorithm presented here was applied to two data sets. The first one described by a linear model allows us to validate the algorithm in comparison with an analytical EM. The second one deals with growth data in chicken modelled via a Gompertz function and illustrates the potential of the algorithm for analyzing real data sets via NLMM's.

3.4.1 Validation on a linear model: Pothoff and Roy's data

A validation of our algorithm was first carried out on the dental growth data presented originally by Pothoff and Roy (1964) and analyzed later on in more details by Jenrich and Schulter (1986) and Verbeke and Molenberghs (1999).
Here, the data set analyzed is the incomplete data set conceived by Little and Rubin (1987) after deleting nine observations at age 10. Among all the possible models envisioned with respect to mean and covariance structures, we restricted our attention to the random intercept model (or compound symmetry) which turned out to be the simplest one consistent with the data according to Verbeke and Molenberghs (1999). This model can be written as:

$$y_{hij} = A_{hi} + B_{hi}t_j + \sigma_{hij}\varepsilon^*_{hij}$$

where y_{hij} is the j-th measurement (distance from the center of the pituary to the pteryomaxillary fissure in 1/10 mm) made on the i-th child nested within the h-th gender (h=1,2 for boys and girls respectively) at age t_j (8, 10, 12 and 14 years); A_{hi} represents the individual intercept assumed i.i.d. $\mathcal{N}(\boldsymbol{\mu}_h, \tau^2)$ and B_{hi} stands for the slope. In a compound symmetry model, B_{hi} reduces to its fixed components term $B_{hi} = \beta_h$, the regression coefficient of response on age for gender h. σ_{hij} designates the residual standard deviation for observation hij and the corresponding error term ε^*_{hij} are supposed i.i.d. $\mathcal{N}(0,1)$.

Attention here was restricted to variance functions comprising purely fixed effects since only such functions are anatically tractable in EM theory applied to LMM's (Foulley et

al, 1992). We consider a general variance function involving variation with respect to age and gender as follows:

$$\log \sigma^2_{hij} = \delta_h + \delta^*_h t_j$$

This allows to describe several sub-models of interest e.g.:
- M0: "Homogeneity": $\delta_1 = \delta_2$, $\delta^*_1 = \delta^*_2 = 0$;
- M1: "Age effect": $\delta_1 = \delta_2$, $\delta^*_1 = \delta^*_2$;
- M2: "Gender effect": $\delta_1 \neq \delta_2$, $\delta^*_1 = \delta^*_2 = 0$;
- M3: "Age + Gender effects": $\delta_1 \neq \delta_2$, $\delta^*_1 = \delta^*_2$;
- M4: "Age + Gender+ Age*Gender effects": $\delta_1 \neq \delta_2$, $\delta^*_1 \neq \delta^*_2$;

The SAEM-MCMC algorithm corresponding to this model was implemented according to Duval and Robert-Granié (2007): see details in the appendix about the criteria values used to calibrate the algorithm.
Results for the different submodels are shown in table 3.1. Likelihood ratio tests clearly highlight the significant effect of gender on the residual variance. Outputs of the SAEM-MCMC algorithm for Model M4 were compared with those of an analytical EM algorithm derived along the same lines as in Foulley et al. (1992) (table 3.2). As can be expected, the two algorithms gave very close results in terms of estimations and their SE. In addition, we checked that these values also agreed with the outputs obtained via SAS-Proc Mixed which is based not on EM but on a second order algorithm (Newton-Raphson or Fisher scoring).

3.4.2 Application to a non linear mixed model : growth curves in poultry

Data come from a selection experiment on growth in broiler chickens carried out at INRA, Nouzilly station, France, by F Ricard. A divergent selection was applied to weights at 8 (W8) and 36 (W36) weeks of age and resulted into the following 5 lines: +- (high at W8, low at W36); -+ (low at W8, high at W36); ++ (high at both W8 and W36), – (low at both W8 and W36) and a control (C) unselected. The data set analyzed concerns a subsample of 10 females per line born at the last generation and recorded at 12 different times (0,4,6,8,12,16,20,24,28,32 and 40 weeks of age).

Following Mignon-Grasteau et al. (2000) and Meza et al. (2007), these data were analyzed using a Gompertz function corresponding to the following model:

$$y_{hij} = A_{hi} \exp\big(- B_{hi} \exp(-C_{hi} t_j/100)\big) + \sigma_{hij}\varepsilon^*_{hij} \tag{3.11}$$

where y_{hij} is the weight performance of the i-th (i=1,...,10) hen within the h-th (h=1,...,5) line at age t_j. A_{hi} represents the asymptotic mature weight of animal hi; B_{hi} is a parameter linking the adult weight to the birth weight and C_{hi} is the so called "maturation rate". We assumed that the A_{hi}'s are i.i.d. $A_{hi} \sim \mathcal{N}(a_h, \tau^2_a)$, the C_{hi}'s are i.i.d. $C_{hi} \sim \mathcal{N}(c_h, \tau^2_c)$, and correlated to the A_{hi}'s according to $\tau_{ac} = Cov(A_{hi}, C_{hi})$. Here, due to convergence

problems encountered with some softwares, B_{hi} was restricted to its fixed line effect β_h. σ_{hij} is the residual standard deviation for observation hij and the error terms ε^*_{hij} are assumed i.i.d. with $\varepsilon^*_{hij} \sim \mathcal{N}(0,1)$ independent of the A_{hi}'s and the C_{hi}'s.

Among all possible variance functions involving the effects of age, line and individual, we considered the five ones defined as follows:

V0: $\log(\sigma^2_{hij}) = \delta$ for the homogeneous case;
V1: $\log(\sigma^2_{hij}) = \delta_1 + \delta_2 t^*_j + \delta_3 {t^*_j}^2$, i.e. a quadratic adjustement of logvariances according to age defined here for computing convenience as $t^*_j = (t_j - 20)/100$;
V2: $\log(\sigma^2_{hij}) = \delta_{1h} + \delta_{2h} t^*_j + \delta_{3h} {t^*_j}^2$, the same as V1 but with line specific adjustements;
V3: $log(\sigma^2_{hij}) = \delta_{1hi} + \delta_{2h} t^*_j + \delta_{3h} {t^*_j}^2$, the same as V2 but with a random intercept component $\delta_{1hi} \sim \mathcal{N}(\delta_{1h}, \tau^2_d)$;
V4: $\sigma^2_{hij} = r f^p_{hij}$, with $f_{hij} = A_{hi} \exp\big(- B_{hi} \exp(-C_{hi} t_j/100)\big)$.

Functions V0, V1, V2 and V3 are special cases of linear models on logvariances defined in (3.2) while V4 corresponds to the power of the mean model defined in (3.3).

Global performance of the five variance functions are reported on table 3.3. As shown by all the criteria displayed (deviance, AIC and BIC) modelling residual variances using any of the four functions mentioned considerably improves the efficiency of the analysis. The addition of the line effect over that of age is especially critical making models more effective as indicated by the very small P-value of the log-likelihood ratio test, and the large change in the AIC (-134.7) and BIC (-111.8) values when contrasting models V1 and V2 respectively.
An additional improvement of the AIC (-10.5) and BIC (-8.6) can also be obtained by considering the intercept δ_{1h} of this function as random (V4) which clearly illustrates the potential interest of using random effects in the variance functions. On the other hand, the power of the mean function, which is the most common function used to take heteroskedasticity into account does not perform better than the Age $\times$ Line fixed function V2.

An important aspect of this study consisted in contrasting the use of different softwares on this real data application. Softwares listed for the output comparison in addition to our SAEM-MCMC algorithm were the Nlmixed procedure of SAS (Littell et al., 2006), Monolix running on Matlab, nlme for Splus and R (Pinheiro and Bates, 2000) and finally WinBUGS. SAS-Nlmixed and Monolix are based on "exact" ML procedures via Gaussian quadratures and Stochastic EM algorithms respectively. Nlme is a FOCE approximation method while WinBUGS performs Bayesian Posterior Inference via Gibbs sampling.

All these softwares do not provide the same flexibility in variance modelling, so it was not possible to make a complete comparison in their outputs for all variance functions. V4 can be used in all these softwares, but V1 and V2 cannot be set up in Monolix while

V3 -which in this example was the best model- is only available in our SAEM-MCMC algorithm and WinBUGS.
The estimates of parameters for this model are shown in table 3.4 under these two procedures. ML and Bayes estimates are in good agreement, especially for fixed effects of both location and dispersion parameters. Regarding variance components, as expected Bayes estimates are somewhat higher than ML estimates due to the well known bias of the latter (Meza et al., 2007).

Table 3.5 provides a comparison of outputs for the V2 variance function using SAEM-MCMC, Nlmixed, nlme and WinBUGS. We observe that the SAEM-MCMC and Nlmixed estimates are very consistent with one another, and this provides a validation of our algorithm in the nonlinear case. Nlme outputs are also very close to the previous estimates which indicates a good behaviour of this FOCE algorithm in such an example. The largest comparison between softwares is shown on table 3.6 for the power variance function. Again, the SAEM-MCMC and Nlmixed estimates coincide perfectly whereas some differences occur with nlme and Monolix results. Regarding standard errors problems arise with Nlmixed, Monolix and nlme due to either non positive definite Hessian matrices (Nlmixed, nlme) or the way some parameters are handled (β_h treated as random with a small variance in Monolix). As in the former table, WinBUGS produces higher variance components estimations for that model V4 than the other algorithms.
Finally, SAEM-MCMC and WinBUGS turn out to be the most flexible and secure softwares among those compared to estimate parameters using either ML or Bayes inference.

3.5 Discussion - Conclusion

This paper represents a further step in the development of nonlinear mixed models by enriching them with a large class of variance functions. The key idea is that heterogeneity of residual variance is not an exception but should be part of model construction as it is for subclass means. Doing so turns out to be very effective for improving the efficiency of the statistical analysis (as well illustrated in the chicken example shown here) as in other practical instances.
In addition to the classical "power of the mean" function, we consider a mixed model extension of the classical linear model on logvariances which provides much flexibility in modelling potential sources of variation in residual variances.
This approach is especially interesting as it takes into account major sources of variation in situations with little information per subclass or experimental unit in a very parsimonious way. Treating the source of variation as random is equivalent to constructing a shrunk estimator of variances combining a population and subject (or subclass) specific estimators: see e.g., Jaffrézic et al. (2007) for variance estimation in differentially expressed gene studies.
Nevertheless, other variance functions could have been envisioned especially in the case of repeated data structures such as those based on stochastic processes (ARCH models of

Engle, 1982) or on semi-parametric techniques (e.g. B-splines reported by Torres, 2001). An important issue that was not covered here consists of incorporating heterogeneity in components of variance and covariance of random effects. This could be done along the same lines as described in linear mixed models by Foulley and Quaas (1995).

Regarding inferential aspects, here the choice was made to rely on maximum likelihood in contrast to other approximated techniques. In fact, we strongly believe that exact statistical procedures are always to be preferred to approximations even though some of them, such as FOCE (via here nlme) can work quite well in some examples. Unfortunately, this is not always the case and it is difficult to know a priori in which conditions and with which models!
Here, our attention was focused on ML estimation of all the parameters of the model including dispersion parameters. Correction of bias induced by ML on such parameters could be achieved using REML estimations by integrating fixed effects out the $f(.)$ function. Again, there is conceptually no difficulty to do that via the EM algorithm as shown by Meza et al. (2007) by including those fixed effects as part of the missing data vector.

Finally, ML estimation was obtained with the SAEM-MCMC algorithm. This is a very simple algorithm which works quite well provided its parameters are well calibrated. It is also clearly more reliable than Gaussian quadratures in this kind of model as illustrated here on table 3.6. This does not mean that the story is over especially concerning the stochastic procedure chosen for the integration step. To that goal, procedures based either on Population Monte Carlo (Cappé et al., 2004) on quasi Monte-Carlo integration (Pan and Thompson, 2007) might merit future testing.

3.6 References

Battese, G.E., Bonyhady, B.P. (1981) Estimation of household expenditure functions: an application of a class of heteroscedastic regression models. The Economic Record, 57, 80-85.

Blasco, A., Piles, M., Varona, L. (2003) A bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. Genetics Selection Evolution, 35, 21-41.

Box, G.E.P., Hill, W.J. (1974) Correcting inhomogeneity of variances with power transformation wighting. Technometrics, 16, 385-389.

Brown, W.J., Draper, D., Golstein, H., Rasbash, J. (2002) Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. Computational Statistics and Data Analysis, 39, 203-225.

Cappé, O., Guillin, A., Marin, J.M., Robert, C.P. (2004) Population Monte Carlo. Journal of Computational and Graphical Statistics, 13, 907-929.

Davidian, M., Carroll, J. (1987) Variance Function Estimation. Journal of the American Statistical Association, 82, 1079-1091.

Davidian, M., Giltinan, D.M. (1995) Nonlinear mixed models for repeated Measures Data. New York: Chapman and Hall.

Delyon, B., Lavielle, M., Moulines, E. (1999) Convergence of a stochastic approximation version of the EM algorithm. Annals of Statistics, 27, 94-128.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B, 39, 1-38.

Duval, M., Robert-Granié, C. (2007) SAEM-MCMC: some criteria.
http://hal.archives-ouvertes.fr/hal-00189580/fr/

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of variance of united Kingdom inflations. Econometrica 50, 987-1008.

Foulley, J.L., San Cristobal, M., Gianola, D., Im, S. (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. Computational Statistics and Data Analysis, 13, 291-305.

Foulley, J.L., Quaas, R.L. (1995) Heterogeneous variances in Gaussian linear mixed models. Genetic Selection Evolution, 27, 211-228.

Gelman, A., Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-511.

Goldstein, H. (1987) Multilevel models in educational and social research. New York: Oxford University Press.

Jaffrézic, F., Marot, G., Degrelle, S., Hue, I., et al. (2007) A structural mixed model for variances in differential gene expression studies. Genetical Research, 89, 19-25.

Jenrich, R.I., Schulter, M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices. Biometrics, 42, 805-820.

Judge G.G., Griffiths W.E., Carter Hill R., Lutkepohl H., et al. (1985) The theory and practice of econometrics. New-York: J Wiley and Sons.

Kuhn, E., Lavielle, M. (2004) Coupling a stochastic approximation version of EM with a MCMC procedure. ESAIM Probability and Statistics, 8, 115-131.

Kuhn, E., Lavielle, M. (2005) Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics and Data Analysis, 49, 1020-1038.
Lee, Y, Nelder, J.A. (2006) Double hierarchical generalized linear models (with discussion). Journal of the Royal Statistical Society: Series C, 55, 139-180.
Lin, X., Ray, J., Harlow, S.D. (1997) Linear mixed models with heterogeneous within-cluster variances. Biometrics, 53, 910-923.
Lindstrom, M., Bates, D. (1990) Nonlinear mixed effects models for repeated measures data. Biometrics, 44, 673-687.
Littell, R., Milliken, G., Stroup, W., Wolfinger, R., et al. (2006) SAS for Mixed Models. New York: SAS Institute.
Little, R.J.A., Rubin, D.B. (1977) Statistical analysis with missing data. New-York: Wiley and Sons.
Louis, T.A. (1982) Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society: Series B, 44, 226-233.
Lu, J-C, Chen, D., Zhou, W. (2006) Quasi-likelihood estimation for GLM with random scales. Journal of Statistical Planning and Inference, 136, 401-429.
Meza, C., Jaffrézic, F., Foulley, J.L. (2007) REML estimation of variance parameters in nonlinear mixed effects models using SAEM algorithm. The Biometrical Journal, 49, 876-888.
Mignon-Grasteau, S., Piles, M., Varona, L., Poivey, J.P., et al. (2000) Genetic analysis of growth curve parameters for male and female chickens resulting from selection on shape of growth curve. Journal of Animal Science, 78, 2532-2531.
Pan, J., Thompson, R. (2006) Quasi Monte Carlo estimation in generalized linear mixed models. Computational Statistics and Data analysis, 51, 5765-5775.
Pinheiro, J.C., Bates, D.M. (2000) Mixed-effects Models in S and S-PLUS. New York: Springer.
Pothoff, R.F., Roy, S.N. (1964) A generalized multivariate analysis of variance function useful especially for growth curve problems. Biometrika, 51, 313-326.
Rigby, R.A., Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape (with discussion). Applied Statistics, 54, 507-554.
Robert, C.P., Casella, G. (2004) Monte Carlo Statistical methods. New York: Springer.
Sheiner, L.B., Beal, S.L. (1980) Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. Journal of pharmacokinetics and Biopharmaceutics, 8, 553-571.
Torres, R.A. (2001) Markov chain Monte Carlo methods for estimating the covariance structure of longitudinal data - an application to dairy cattle. PhD thesis, Cornell University.
Verbeke G., Molenberghs G. (1999) Linear models for longitudinal data. New York: Springer Verlag.
Wolfinger, R.D. (1993) Laplacian's approximation for nonlinear mixed models. Biometrika, 80, 791-795.

3.7 Appendix

3.7.1 The Metropolis-Hastings algorithm

The simulation step of the SAEM-MCMC algorithm is implemented via a Metropolis-Hastings algorithm (Robert and Casella, 2004). This MCMC procedure produces Markov chains with stationary distribution being the conditional distribution of the missing data $\mathbf{u}$ given the observations $\mathbf{y}$ and the vector of parameters $\boldsymbol{\theta}$.

Let us define $p^{(k-1)}(\mathbf{u})$ the density of $\mathbf{u}$, $\mathbf{V}^{(k-1)}$ the variance of $\mathbf{u}$ at iteration $k-1$, and $p(\mathbf{y}|\boldsymbol{\theta},\mathbf{u})$ the density of the observations given the vector of the parameters $\boldsymbol{\theta}$ and the missing values $\mathbf{u}$.

Let $m_1 \in \mathbb{N}$, $\rho_1 \in \mathbb{R}^+$, $\rho_2 \in \mathbb{R}^+$, $\rho_2 > \rho_1$ and $\boldsymbol{\theta}^{(k-1)}$ the estimated parameter value of $\boldsymbol{\theta}$ at iteration $k-1$ of the SAEM-MCMC algorithm.

Let $\mathbf{x}^{(t)} = \mathbf{u}^{(k,l,t-1)}$ obtained at the $(t-1)$th step of the lth chain of the Metropolis-Hastings algorithm performed at the kth step of the SAEM-MCMC algorithm. The next step is implemented as follows.

- If $t < m_1$, then generate $\mathbf{W}_t$ from $p^{(k-1)}(\mathbf{u})$ with acceptance rate $\rho(\mathbf{x}^{(t)}, \mathbf{w}_t) = \min\left(\frac{p(\mathbf{y}|\mathbf{w}_t,\boldsymbol{\theta}^{(k-1)})}{p(\mathbf{y}|\mathbf{x}^{(t)},\boldsymbol{\theta}^{(k-1)})}, 1\right)$
- If $t \geq m_1$, then generate $\rho \sim \mathcal{U}_{[\rho_1,\rho_2]}$ and $\mathbf{W}_t$ from $\mathcal{N}(\mathbf{x}^{(t)}, \rho\mathbf{V}^{(k-1)})$ with acceptance rate $\rho(\mathbf{x}^{(t)}, \mathbf{w}_t) = \min\left(\frac{p(\mathbf{y}|\mathbf{w}_t,\boldsymbol{\theta}^{(k-1)})\, p^{(k-1)}(\mathbf{w}_t)}{p(\mathbf{y}|\mathbf{x}^{(t)},\boldsymbol{\theta}^{(k-1)})\, p^{(k-1)}(\mathbf{x}^{(t)})}, 1\right)$
- Finally $\mathbf{u}^{(k,l,t)} = \begin{cases} \mathbf{w}_t & \text{with probability } \rho(\mathbf{x}^{(t)}, \mathbf{w}_t) \\ \mathbf{x}^{(t)} & \text{with probability } 1-\rho(\mathbf{x}^{(t)}, \mathbf{w}_t) \end{cases}$

Then for the sake of simplicity, we note $\mathbf{u}^{(k,l)} = \mathbf{u}^{(k,l,T)}$, where T is the length of the chain.

3.7.2 Some criteria to calibrate the parameters of the SAEM-MCMC algorithm

Duval and Robert-Granié (2007) presented criteria to choose the parameter values of the SAEM-MCMC algorithm. These parameters were calibrated as follows: we run a Markov chain using the Metropolis-Hastings algorithm for one representative individual taking the initial value $\boldsymbol{\theta}_0$ of the parameters and then the following criteria are used. Finally these parameters are used to run the Metropolis-Hastings algorithm for all the individuals.

• The iteration at which we switch to another instrumental distribution in the Metropolis-Hastings algorithm, noted m_1:
The different behaviors of the chain under the two instrumental distributions lead us to

choose $m_1 = T/10$, according to Duval and Robert-Granié (2007).

• **The second instrumental distribution of the Metropolis-Hastings algorithm**: ρ_1 and ρ_2 are chosen such that the acceptance rate is between 30% and 40 %, according to Robert and Casella (2004).

• **The length of each chain in the Metropolis-Hastings algorithm, noted** T: We run long Markov chains in the Metropolis-Hasting, such as they are under the stationary distribution. Convergence was tested according to the Gelman and Rubin criteria (1992), by comparing within-chain and between-chain variances.

• **The number of independent chains simulated in the Metropolis-Hastings algorithm, noted** L: In practice we run several chains and compare the behavior of each of them. If their properties (mean, variance) are significantly different, we may choose $L = 5$ or $L = 10$ chains, else, $L = 1$ chain can be enough (Kuhn and Lavielle, 2005).

• **Choice of the parameter** K **for the sequence** $(\gamma_k)_k$: Let

$$e^{(k)} = \max_j \left(\frac{|\theta_j^{(k)} - \theta_j^{(k-1)}|}{|\theta_j^{(k)}|} \right) \tag{3.12}$$

We choose K as the first iteration k such that $e^{(k)} < 10^{-2}$.

• **Stopping rule**: The SAEM-MCMC algorithm is stopped when $e^{(k)} < 10^{-4}$.

3.7.3 Parameters of the SAEM-MCMC algorithm used in Pothoff and Roy's data analysis

Criteria provided the following parameters: $T = 300$, $m_1 = 30$, $L = 5$, $\rho_1 = 0.5$, $\rho_2 = 1.5$ and $K = 222$. An acceptance rate around 30% and a Gelman and Rubin criterion at 1.07 were obtained. Our algorithm stopped at 362 iterations and the EM algorithm at 53 iterations.

$\boldsymbol{\theta}$	μ_1	μ_2	β_1	β_2	τ^2	δ_1	δ_2	δ_1^*	δ_2^*	$-2L^{(a)}$	$\Delta(-2L)^{(b)}$
initial value $\boldsymbol{\theta_0}$	200	200	5	12	300	2	7	-2	-2	-	-
M0: $\delta_1 = \delta_2$ and $\delta_1^* = \delta_2^* = 0$											
$\hat{\boldsymbol{\theta}}$	211.29	225.97	9.78	15.74	311.38	5.30				857.23	0.00
SE	6.41	5.51	1.87	1.67	99.92	0.33					
M1: $\delta_1 = \delta_2$ and $\delta_1^* = \delta_2^*$											
$\hat{\boldsymbol{\theta}}$	211.47	225.77	9.74	15.91	313.50	5.51		-0.14		856.73	0.50
SE	6.70	5.62	1.94	1.67	101.66	0.33		0.20			
M2: $\delta_1^* = \delta_2^* = 0$											
$\hat{\boldsymbol{\theta}}$	211.05	226.04	9.85	15.72	328.00	4.08	5.68			838.87	18.36
SE	5.84	5.88	1.06	1.93	101.64	0.26	0.21				
M3: $\delta_1^* = \delta_2^*$											
$\hat{\boldsymbol{\theta}}$	210.97	225.65	9.91	16.00	334.89	4.34	5.96	-0.19		838.13	19.10
SE	5.96	6.33	1.06	2.05	103.25	0.39	0.38	0.21			

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$, $^{(b)}\Delta(-2L)$: variation of the value of $-2L$ between Homogeneous Model and the corresponding Model

M0: Homogeneous, M1: Age, M2: Gender, M3: Age+Gender

Table 3.1: SAEM-MCMC estimates ($\hat{\boldsymbol{\theta}}$) and standard errors (SE) for the analysis of Pothoff and Roy's data for variance functions M0 to M3.

θ	μ_1	μ_2	β_1	β_2	τ^2	δ_1	δ_2	δ_1^*	δ_2^*	$-2L^{(a)}$
initial value	200	200	5	12	300	2	7	-2	-2	-
$\hat{\boldsymbol{\theta}}_{EM}$	210.89	225.71	9.94	15.97	336.37	4.37	5.94	-0.21	-0.18	838.13
SE	5.99	6.18	1.07	1.98	104.70	0.63	0.42	0.42	0.25	-
$\hat{\boldsymbol{\theta}}_{SAEM-MCMC}$	210.90	225.66	9.94	15.99	336.24	4.36	5.96	-0.20	-0.18	838.13
SE	6.01	6.22	1.08	1.99	104.74	0.63	0.42	0.43	0.25	-

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$

M4: Age+Gender+Age*Gender

Table 3.2: Comparison of EM and SAEM-MCMC estimates ($\hat{\boldsymbol{\theta}}$) and standard errors (SE) for the analysis of Pothoff and Roy's set with the variance function M4.

Model	Parameters	$-2L^{(a)}$	Comparison	$\Delta(-2L)$	P-value	AIC	BIC
V0	19	7525.6	-	-	-	7563.6	7599.9
V1	21	7161.6	V0 vs V1	364.0	0.0000	7203.6	7243.8
V2	33	7002.9	V1 vs V2	159.6	0.0000	7068.9	7132.0
V3	34	6990.4	V2 vs V3	12.5	0.0008	7058.4	7123.4
V4	20	7053.1	V0 vs V4	472.5	0.0000	7093.1	7131.4

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$

V0: "Homogeneous", V1: "Age", V2: "Age×Line", V3: "Age × Line+Ind", V4: "Power"

Table 3.3: Comparison of models for residual variance functions based on ML estimation via the SAEM-MCMC algorithm.

parameters	initial value	WinBUGS		SAEM-MCMC	
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE
a_0	2500	3011.00	93.02	3025.50	93.61
β_0	4	4.38	0.04	4.37	0.04
c_0	15	11.96	0.52	11.90	0.48
τ_a^2	50000	59960.00	13860.00	51972.00	11357.00
τ_c^2	5	2.43	0.56	1.98	0.44
τ_{ac}	-100	-160.40	67.05	-160.77	55.33
τ_d^2	10	0.14	0.01	0.10	0.10
δ_{10}	5	10.53	0.14	10.96	0.60
δ_{20}	10	16.36	1.31	16.53	1.35
δ_{30}	-100	-114.60	20.65	-117.81	32.63
$-2L^{(a)}$		6558.00$^{(b)}$		6990.40	
AIC		6814.00$^{(c)}$		7058.40	

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$

$^{(b)}$ "Dhat" in WinBUGS terminology i.e. posterior expectation of deviance

$^{(c)}$ DIC criterion

Table 3.4: Parameter estimates ($\hat{\boldsymbol{\theta}}$) with their standard errors (SE) for the chicken data with residual variance functions V3 using WinBUGS and SAEM-MCMC.

Parameter	initial value	SAEM-MCMC		Nlmixed		nlme		Winbugs	
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE
a_0	2500	3025.20	85.66	3027.06	86.07	3026.10	81.53	3018.00	93.47
β_0	4	4.37	0.03	4.37	0.03	4.37	0.03	4.34	0.04
c_0	15	11.88	0.48	11.88	0.48	11.88	0.47	11.93	0.53
τ_a^2	50000	52230.00	11247.00	52167.00	11265.00	52168.00		60310.00	13870.00
τ_c^2	5	1.96	0.43	1.96	0.43	1.95		2.42	0.56
τ_{ac}	-100	-159.21	54.69	-158.59	54.85	-158.31		-157.80	66.65
δ_{10}	5	11.02	0.33	11.01	0.33	11.00		10.95	0.34
δ_{20}	10	16.48	1.24	16.49	1.24	16.47		16.34	1.26
δ_{30}	-100	-119.67	19.45	-118.85	16.68	-118.20		-113.10	20.39
$-2L^{(a)}$		7002.88		7003.00		7002.88		6613.70$^{(b)}$	
AIC		7068.90		7069.00		7068.90		6834.95$^{(c)}$	
BIC		7132.00		7132.10		7132.00			

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}};\boldsymbol{y})$

$^{(b)}$ "Dhat" in WinBUGS terminology i.e. posterior expectation of deviance

$^{(c)}$ DIC criterion

Table 3.5: Parameter estimates ($\hat{\boldsymbol{\theta}}$) with their standard errors (SE) for residual variance function V2 for SAEM-MCMC, Nlmixed, nlme procedures and WinBUGS software on chicken data.

Parameter	initial value	SAEM-MCMC		Nlmixed		Monolix		Winbugs		nlme	
$\boldsymbol{\theta}$	$\boldsymbol{\theta_0}$	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE
a_0	2500	3164.70	83.98	3164.60	83.66	3160.00	80.90	3139.00	88.11	3153.10	84.36
β_0	5	4.43	0.03	4.43	0.03	4.41	0.03	4.43	0.03	4.42	0.03
c_0	15	11.76	0.43	11.76	0.43	11.70	0.42	11.83	0.49	11.75	0.43
τ_a^2	50000	54694.00	12941.00	54674.00	0.11[(d)]	48400.00	11100.00	64200.00	15740.00	54251.00	
τ_c^2	5	1.58	0.40	1.58	0.39	1.55	0.38	2.16	0.54	1.53	
τ_{ac}	-100	-137.15	53.30	-137.17	41.56	-120.51		-150.10	67.43	-130.83	
r	exp(-1)	0.04	0.06	0.03	0.01	0.01		0.04	0.02	0.03	
p	1	1.75	0.12	1.75	0.06	1.90		1.72	0.06	1.74	
$-2L^{(a)}$		7053.10		7054.10		7061.04		6719.20[(b)]		7055.20	
AIC		7093.10		7094.40		7101.00		6911.95[(c)]		7095.20	
BIC		7131.34		7132.70		7139.30				7133.40	

[(a)] $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$

[(b)] "Dhat" in WinBUGS terminology i.e. posterior expectation of deviance

[(c)] DIC criterion

[(d)] Warning of SAS: hessian not positive definite

Table 3.6: Parameter estimates ($\hat{\boldsymbol{\theta}}$) with their standard errors (SE) for residual variance function V4 for SAEM-MCMC and Nlmixed procedures, Monolix and WinBUGS softwares, and nlme procedure on chicken data.

Chapitre 4

Application : modéliser les cinétiques de scores de cellules somatiques chez les bovins laitiers

4.1 Introduction

Les mammites sont des inflammations de la glande mammaire, essentiellement d'origine infectieuse, qui se traduisent par un affux massif des leucocytes (globules blancs) du sang vers la mamelle. Ces infections sont fréquentes et représentent la maladie la plus coûteuse chez les bovins laitiers (Seegers et al., 1997). Elles ont des conséquences économiques importantes pour l'éleveur : perte de production laitière (Lescourret et Coulon, 1994), baisse du prix de vente du lait issu des vaches infectées et coûts des traitements vétérinaires.
Le diagnostic direct des mammites étant coûteux et difficile à effectuer intra exploitation agricole, les généticiens quantitatifs se sont intéressés à un critère génétiquement corrélé et prélevé lors des contrôles laitiers : le comptage de cellules somatiques, noté CCS (Mrode et Swanson, 1996 ; Rupp et Boichard, 1999). La concentration de cellules somatiques dans le lait traduit généralement l'état général de la mammelle et donc indirectement son état infectieux.

Les analyses statistiques de ce caractère sont le plus souvent basées sur les scores des cellules somatiques (SCS), correspondant à une fonction du logarithme des CCS, de distribution normale et de variance théoriquement homogène (Ali et Shook, 1980).
Mrode et Swanson (1996) ont montré que l'héritabilité (part de la variance génétique sur la variance totale) des SCS est de l'ordre de 0.15 et se trouve bien supérieure à celle des mammites, proche de 0.02. Dans ce sens, la sélection pour la diminution des SCS, plus facile à mettre en place, pourra en théorie diminuer la susceptibilité aux mammites (Mrode et Swanson, 1996 ; Colleau et Le Bihan-Duval, 1995). Rupp et Boichard (1997) ont mis en évidence une corrélation génétique défavorable entre la production laitière et les scores cellulaires, de sorte que les CCS et donc la susceptibilité aux mammites auront

tendance à augmenter sous l'effet de la sélection laitière. Les éleveurs ont donc intérêt à sélectionner à la fois sur la production et sur les scores cellulaires, ce qui implique une bonne connaissance de l'évolution des scores cellulaires au cours de la lactation.
Généralement, les SCS évoluent de manière inversée par rapport à la quantité de lait produite au cours de la lactation : après le début de la lactation, les SCS décroissent jusqu'à atteindre un minimum vers 60 jours puis réaugmentent (Wiggans et Shook, 1987). De plus il a été montré que les effets environnementaux comme le troupeau, le stade de lactation, la saison, l'âge de la vache au vêlage sont d'importantes sources de variation des SCS et donc une augmentation des SCS n'est pas toujours la conséquence d'une infection mammaire.

Les fonctions mathématiques non linéaires sont naturellement candidates pour décrire l'évolution des SCS au cours de la lactation : Rodriguez-Zas et al. (2000) en proposent plusieurs dont certaines obtenues par le biais des courbes de lactation. Robert-Granié et al. (2004) proposent d'utiliser un modèle linéaire, dont la fonction moyenne est décrite par un pôlynome fractionnaire (Royston et Altman, 1994) et pour lequel plusieurs fonctions de variance résiduelle sont étudiées. Les modèles non linéaires ont un avantage sur les modèles linéaires : les paramètres ont généralement une interprétation physique, qui permet ensuite d'assurer un suivi biologique à l'étude.
Robert-Granié et al. (2004) montrent que le modèle linéaire mixte utilisé est amélioré en termes de critères AIC et BIC, lorsque des variances hétérogènes sont introduites. Dans le cadre des modèles linéaires mixtes, Foulley et al. (1992) proposent de modéliser les variances hétérogènes à l'aide d'un modèle linéaire mixte, prenant en compte différentes sources de variation. Une autre approche initialement présentée en pharmacocinétique dans le cadre des modèles non linéaires mixtes par Davidian et Carroll (1987), concerne la prise en compte d'une relation moyenne-variance.

L'objectif de cette étude est de caractériser les profils moyens et individuels des SCS dans cette population bovine, en prenant en compte les variations inter-individus et intra-individu, à l'aide d'un modèle mixte à variances hétérogènes. Plusieurs modèles linéaires et non linéaires sont proposés et comparés. Les paramètres des modèles sont estimés via une méthode de maximum de vraisemblance stochastique basée sur l'algorithme EM, proposée par Kuhn et Lavielle (2004, 2005) et appelée SAEM-MCMC.
Cette méthode fut reprise par Duval et Robert-Granié (2007) en y intégrant des critères de convergence dans le cadre homogène, et par Duval et al. (2008) en adaptant la méthode à l'estimation des variances hétérogènes dans les modèles non linéaires mixtes.

Notre étude est basée sur des données relatives à 6448 contrôles hebdomadaires de cellules somatiques de lait effectués sur 159 génisses Holstein et croisées Holstein $\times$ Normandes, entretenues au domaine expérimental INRA - Le pin au Haras.
Dans la partie "Matériel et méthodes", nous décrirons le jeu de données et les modèles envisagés pour modéliser la cinétique des SCS au cours de la lactation. Différents modèles d'hétérogénéité de variances seront étudiés pour prendre en compte la variation intra-

individu. Enfin, dans la dernière partie, les résultats seront présentés et discutés.

4.2 Matériel et méthodes

4.2.1 Les données

Cette étude est basée sur une expérience conduite entre 1998 et 1999 dans un domaine expérimental de l'INRA : Le Pin au Haras, Normandie, France. Cette expérience avait pour objectif d'étudier la relation entre les scores de cellules somatiques et les mammites chez les bovins laitiers. Le fichier de données est constitué de 6448 mesures de SCS effectuées chaque semaine sur 159 génisses de race Holstein et croisées Holstein × Normande, du 5-ème jour au 305-ème jour après vêlage. Le nombre d'observations par animal n'est pas constant, en moyenne 40 observations par individu sont disponibles au cours de la lactation. Plusieurs caractéristiques ont été relevées au cours de l'expérience : la race de l'animal, l'année de vêlage (1998 ou 1999) et la saison de vêlage (3 niveaux). Elles sont décrites dans les tableaux 4.1 et 4.2.

4.2.2 Le modèle statistique

Le modèle général

Pour analyser ces données longitudinales, on a choisi de modéliser les profils individuels par un modèle mixte de la forme :

$$y_{ij} = f(t_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i) + \sigma_{ij}\varepsilon^*_{ij} \tag{4.1}$$

$$\boldsymbol{\phi_i} \sim_{i.i.d.} \mathcal{N}(\mathbf{A}_i\boldsymbol{\mu}, \boldsymbol{c}) \tag{4.2}$$

où y_{ij} est la j-ème observation, $j \in \{1, ..., n_i\}$, mesurée sur la vache i, $i \in \{1, ..., N\}$, au temps t_{ij}.
$f(t_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)$ représente la fonction moyenne et f est paramétrée par le temps t_{ij}, un vecteur $\boldsymbol{\beta}$ représentant les effets fixes, et un vecteur aléatoire $\boldsymbol{\phi}_i$. Les $\boldsymbol{\phi}_i$ sont supposés i.i.d. Gaussiens de moyenne $\mathbf{A}_i\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Gamma}$, la relation $\mathbf{A}_i\boldsymbol{\mu}$ exprime comment $\boldsymbol{\phi}_i$ varie en fonction des covariables $\boldsymbol{\mu}$ via la matrice de design $\mathbf{A}_i$.
Les erreurs résiduelles sont modélisées par $\sigma_{ij}\varepsilon^*_{ij}$, où les ε^*_{ij} sont i.i.d. Gaussiennes centrées réduites.

Le choix du modèle linéaire ou non linéaire réside dans la forme de la fonction f linéaire ou non linéaire en fonction de $\boldsymbol{\beta}$ et $\boldsymbol{\phi}_i$.

Modèles pour la moyenne

Différentes fonctions moyennes proposées en particulier par Rodriguez-Zas et al. (2000) et Robert-Granié et al. (2004) ont été étudiées.

Concernant les fonctions de type linéaire, Robert-Granié et al. (2004) ont comparé 3 modèles de fonction moyenne différentes pour modéliser la cinétique des scores de cellules somatiques de lait chez les bovins laitiers :

[M1] Un polynome fractionnaire d'ordre 2 de la forme

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3)\big) = \phi_1 + \phi_2 t_{ij}^{-1/3} + \phi_3 t_{ij}^{-1/3}\log(t_{ij})$$

[M2] Un polynome conventionnel d'ordre 3 de la forme

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3,\phi_4)\big) = \phi_1 + \phi_2 t_{ij} + \phi_3 t_{ij}^2 + \phi_4 t_{ij}^3$$

[M3] La fonction proposée par Ali et Schaeffer (1987) :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3,\phi_4,\phi_5)\big) = \phi_1 + \phi_2 t_{ij}^* + \phi_3 {t_{ij}^*}^2 + \phi_4\log(1/t_{ij}^*) + \phi_5[\log(1/t_{ij}^*)]^2$$

où $t_{ij}^* = t_{ij}/305$.

Pour ces trois modèles, les paramètres $\boldsymbol{\phi}$ n'ont pas vraiment d'interprétation biologique.

Les modèles non-linéaires suivants ont été proposés entre autres dans Rodriguez-Zas et al. (2000) :

[M4] La fonction non linéaire définie par Wood (1967) :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3)\big) = \phi_1 t_{ij}^{\phi_2}\exp(-\phi_3 t_{ij})$$

où ϕ_1 est un paramètre associé à la valeur initiale des SCS en début de lactation, ϕ_2 représente la pente de la partie décroissante en début de lactation, et ϕ_3 la pente de la partie croissante en fin de lactation.

[M5] La fonction non linéaire de Mitscherlich (Rook et al., 1993) de la forme :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3)\big) = [1-\phi_1\exp(\phi_2 t_{ij})]\exp(-\phi_3 t_{ij})$$

où ϕ_1 est associé à l'échelle de la courbe, ϕ_2 au taux de décroissance de la courbe en début de lactation, ϕ_3 au taux de croissance des SCS après avoir atteint le minimum de la courbe.

[M6] La fonction non linéaire proposée par Morant et Gnanasakthy (1989) :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3,\phi_4)\big) = \phi_1 \exp[\phi_2 t^*_{ij} + \phi_3 t^{*2}_{ij} + \phi_4/t_{ij}]$$

où ϕ_1 est associé à la valeur initiale des SCS au cours de la lactation, ϕ_2 à la pente de la courbe en fin de lactation, ϕ_3 au taux auquel change la pente au cours de la lactation, ϕ_4 au taux de décroissance des SCS en début de lactation.

[M7] La fonction non linéaire présentée dans Davidian et Giltinan (1995) :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3,\phi_4)\big) = \phi_1 \exp(\phi_2 t^*_{ij}) + \phi_3 \exp(\phi_4 t^*_{ij})$$

où ϕ_1 est associé à la valeur initiale des SCS au cours de la lactation, ϕ_2 à la pente de la courbe en début de lactation, ϕ_3 à un facteur d'échelle, ϕ_4 au taux de croissance des SCS en fin de lactation.

[M8] La fonction proposée par Robert-Granié et al. (2004) en supposant que la puissance est un paramètre aléatoire, noté RGNL :

$$f\big(t_{ij},(\phi_1,\phi_2,\phi_3)\big) = \phi_1 + \phi_2 t^{\phi_4}_{ij} + \phi_3 t^{\phi_4}_{ij} \log(t_{ij})$$

Comme pour les modèles linéaires présentés précédemment, les paramètres de cette fonction n'ont pas d'interprétation biologique.

Les modèles non linéaires ont plusieurs avantages par rapport aux modèles linéaires : généralement les paramètres sont biologiquement interprétables, et les modèles sont parcimonieux. Néanmoins les modèles linéaires sont plus souvent utilisés car les méthodes d'estimations sont beaucoup moins complexes et très souvent implémentées dans les logiciels usuels.
Dans les fonctions moyennes [M1] à [M8], les paramètres peuvent être aléatoires (composantes de $\boldsymbol{\phi}$) ou fixes (composantes de $\boldsymbol{\beta}$).

Modèles sur la variance résiduelle

Les modèles de variance résiduelle suivants ont été étudiés :

[V0] Le modèle homogène : $\sigma^2_{ij} = \sigma^2 \ \forall i,j$,

[V1] Une relation moyenne-variance : $\sigma^2_{ij} = \delta_1 f^{\delta_2}_{ij}$, où f_{ij} correspond à la fonction moyenne, δ_1 et δ_2 sont des constantes à estimer,

puis quatre modèles linéaires mixtes basés sur la log-variance résiduelle (Foulley et al., 1992) :

[V2] $\log(\sigma^2_{ij}) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t^*_{ij}$, où $\boldsymbol{\delta}_0$ et $\boldsymbol{\delta}_1$ sont des vecteurs d'effets fixes à déterminer,

[V3] $\log(\sigma_{ij}^2) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t_{ij}^* + \boldsymbol{\delta}_2 t_{ij}^{*2}$, où $\boldsymbol{\delta}_0$, $\boldsymbol{\delta}_1$ et $\boldsymbol{\delta}_2$ sont des vecteurs d'effets fixes,

[V4] $\log(\sigma_{ij}^2) = \boldsymbol{\delta}_{0i} + \boldsymbol{\delta}_{1i} t_{ij}^*$, où $(\boldsymbol{\delta}_{0i}, \boldsymbol{\delta}_{1i}) \sim \mathcal{N}((\boldsymbol{\delta}_0, \boldsymbol{\delta}_1), \Delta)$,

[V5] $\log(\sigma_{ij}^2) = \boldsymbol{\delta}_{0i} + \boldsymbol{\delta}_{1i} t_{ij}^* + \boldsymbol{\delta}_{2i} t_{ij}^{*2}$, $(\boldsymbol{\delta}_{0i}, \boldsymbol{\delta}_{1i}, \boldsymbol{\delta}_{2i}) \sim \mathcal{N}((\boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2), \Delta)$.

Ce choix n'est pas exhaustif mais permet de prendre en compte plusieurs structures différentes pour la variance résiduelle. Concernant les modèles [V2] à [V5], des covariables peuvent agir sur les paramètres du modèle.

4.2.3 Méthode d'estimation

Lorsque le modèle (4.1) est homogène, le vecteur des paramètres à estimer s'écrit $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}, \sigma^2)$. Il faut y ajouter les paramètres de modélisation de la variance si le modèle est hétérogène.

Lorsque le modèle est linéaire mixte, l'algorithme Expectation-Maximisation (Dempster et al., 1977), basé sur la théorie du maximum de vraisemblance, fournit des estimateurs convergents et précis, il est fréquemment utilisé pour estimer le vecteur des paramètres $\boldsymbol{\theta}$. Dans le cadre des modèles non linéaires mixtes, cet algorithme ne peut plus être directement utilisé car l'intégrale intervenant dans une des étapes itératives n'est plus sous forme explicite. Dans ce sens, plusieurs méthodes exactes basées sur une estimation de Monte-Carlo de cette intégrale ont été proposées (Wei et Tanner, 1990 ; Walker, 1996 ; Wang, 2007).
La méthode d'estimation utilisée dans cette étude, notée SAEM-MCMC, fut proposée par Kuhn et Lavielle (2004, 2005). Elle est décrite dans le chapitre 2 de ce manuscrit. Comme les méthodes précédentes, elle est basée sur un algorithme EM stochastique, son efficacité et son faible coût de calcul tient au recyclage des simulations effectué à chaque itération de l'algorithme.

L'algorithme SAEM-MCMC fournit un estimateur exact contrairement aux méthodes de linéarisation, implémentées dans les logiciels classiques, comme certaines procédures Nlmixed de SAS et nlme de R. L'algorithme SAEM-MCMC connaît néanmoins un inconvénient majeur : plusieurs constantes et paramètres de l'algorithme doivent être calibrés par l'utilisateur lui-même. Duval et Robert-Granié (2007) ont proposé des critères afin de fixer ces paramètres (cf chapitre 2). L'adaptation de l'algorithme SAEM-MCMC à l'estimation des variances hétérogènes dans les modèles non linéaires mixtes a été proposée par Duval et al. (2008), et permet de prendre en compte des modèles de variance résiduelle plus variés que ne le font les autres logiciels (par exemple : Monolix, Nlmixed de SAS, nlme de R).
L'algorithme SAEM-MCMC sera utilisé dans cette étude pour estimer le vecteur des paramètres $\boldsymbol{\theta}$ des différents modèles étudiés. La plupart des résultats obtenus ont été validés

en les comparant à ceux obtenus par le logiciel d'estimation bayésienne WinBUGS.

4.2.4 Sélection de modèles

La sélection d'un modèle est complexe car le choix de la fonction moyenne dépend du choix de la fonction de variance résiduelle et vice versa (Verbeke et Molenberghs, 2000). Beaucoup de combinaisons de fonctions de moyenne et variance peuvent être comparées. Nous avons choisi d'étudier huit fonctions moyenne (modèles [M1] à [M8]), pour lesquelles les paramètres peuvent être supposés aléatoires ou fixes, corrélés ou pas et dépendant de covariables à déterminer. Pour chacun des modèles de moyenne, les cinq fonctions de variance résiduelle peuvent être considérées, et pour ces fonctions, les paramètres sont aléatoires ou fixes, corrélés ou pas et dépendant de covariables à choisir.
Face au grand nombre de modèles à tester, nous avons adopté la stratégie de sélection de modèle suivante.

Etape 1 : choix des meilleurs modèles linéaire et non linéaire homogènes.
Chaque fonction moyenne a été intégrée au modèle (4.1), à variances homogènes (i.e. $\sigma_{ij} = \sigma^2 \, \forall i, j$), dans lequel tous les paramètres $\boldsymbol{\phi}$ étaient supposés aléatoires et corrélés. De plus, nous avons supposé que le modèle était saturé au niveau des effets fixes jusqu'aux interactions du premier ordre. Dans ce sens, pour chacun des paramètres, les covariables suivantes étaient prises en compte dans le modèle : la campagne (2 niveaux), la race (2 niveaux), la saison de vêlage (3 niveaux) et les interactions de premier ordre entre ces effets simples, i.e. campagne × race, campagne × saison et race × saison.

Les paramètres des différents modèles ont été estimés via l'algorithme SAEM-MCMC, et les modèles ont été comparés à l'aide de critères de sélection de modèles, définis dans le paragraphe suivant.
La première étape de l'étude consistait à choisir la meilleure fonction linéaire et la meilleure fonction non-linéaire qui s'ajustaient le mieux aux profils moyens des CCS.

Etape 2 : déterminer les effets aléatoires et les effets fixes de la fonction moyenne.
Pour les modèles linéaire et non linéaire choisis à l'étape 1, la matrice de corrélation associée, notée $\boldsymbol{\rho}_{\boldsymbol{\Gamma}}$, a été construite. Pour des corrélations faibles, nous pouvions supposer que les paramètres correspondants n'étaient pas corrélés entre eux. Dans ce cas, un nouveau modèle a été construit sous cette hypothèse de non corrélation, puis les deux modèles ont été comparés à l'aide d'un test de rapport de vraisemblance, défini dans le paragraphe suivant. Enfin, les composantes de $\boldsymbol{\phi}$, pour lesquelles la variance semblait faible, ont été testées en tant que paramètres fixes.

Etape 3 : Sélection des covariables sur les paramètres de la fonction moyenne.
Dans cette étape, la significativité des covariables sur chaque paramètre de la fonction moyenne était testée en utilisant un test de Wald (Verbeke et Molenberghs, 2000).

Etape 4 : choix du modèle de variance résiduelle.
Les fonctions de variance (modèles [V0] à [V5]) ont été intégrées aux modèles obtenus à l'étape 3, en supposant tout d'abord que les paramètres des fonctions de variance ne dépendaient pas de covariables.
Les modèles ont ensuite été comparés à l'aide des critères de sélection de modèles AIC et BIC. Au final, nous avons retenu les deux meilleurs modèles de variance pour les meilleurs modèles linéaire et non-linéaire.

Etape 5 : choix des covariables pour chaque paramètre des fonctions de variance.
De même que pour la fonction moyenne, la nature aléatoire des paramètres des fonctions de variance obtenues à l'étape 4 a été étudiée à l'aide du test de rapport de vraisemblance. Ensuite les covariables pouvant agir sur chacun des paramètres des fonctions de variance ont été testées à l'aide d'un test de Wald. Cette dernière étape est similaire aux étapes 2 et 3 effectuées au niveau de la fonction moyenne.

Critères de sélection de modèles et procédures de test

Sélection des covariables sur les effets fixes
Comme la deuxième étape détermine les paramètres aléatoires et fixes de la moyenne, la sélection des covariables peut s'effectuer sur la moyenne des effets aléatoire $\boldsymbol{\mu}$ ou sur le vecteur des effets fixes $\boldsymbol{\beta}$, et de la même manière sur les paramètres de la variance. Dans tous les cas, on procède à un test de Wald en utilisant la matrice estimée des erreurs standards des paramètres.
Supposons qu'on veuille tester les covariables constituant le vecteur $\boldsymbol{\mu}$. Soit $\hat{\boldsymbol{\mu}}$ l'estimateur de $\boldsymbol{\mu}$ obtenu par la méthode d'estimation SAEM-MCMC et $\hat{\boldsymbol{V}}$ l'estimation de la matrice des variances d'échantillonnage, notée $\boldsymbol{V}$. On souhaite tester l'hypothèse suivante :

$$H_0 : \boldsymbol{k}'\boldsymbol{\mu} = \mathbf{0}_r$$

contre son alternative

$$H_1 : \boldsymbol{k}'\boldsymbol{\mu} \neq \mathbf{0}_r$$

où $\boldsymbol{k}$ est une matrice de taille $m \times r$ et 0_r est le vecteur nul de taille r.
Sous H_0, la statistique du test de Wald

$$(\boldsymbol{k}'\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{k}'\boldsymbol{\mu})$$

suit asymptotiquement une loi de khi-deux à $rang(\boldsymbol{k})$ degré de liberté.

Test de rapport de vraisemblance

Supposons qu'on veuille comparer deux modèles emboités M_0 et M_1, de dimension respective P_0 et P_1. On note l'hypothèse nulle

$$H_0 : M_0 = M_1$$

contre son alternative

$$H_1 : M_0 \neq M_1$$

Sous H_0, la statistique

$$LRT = 2\big[L_{M_1}(\boldsymbol{\theta}; y) - L_{M_0}(\boldsymbol{\theta}; y)\big]$$

suit asymptotiquement une loi du $\chi^2_{P_1-P_0}$, où $L_{M_0}(\boldsymbol{\theta}; y)$ et $L_{M_1}(\boldsymbol{\theta}; y)$ correspondent respectivement à la log-vraisemblance des observations sous les modèles M_0 et M_1.
L'hypothèse H_0 du test est rejetée au niveau asymptotique α si la valeur du test LRT est supérieure à une valeur critique LRT_c telle que $P[\chi^2_{P_1-P_0} \geq LRT_c] = \alpha$.

Dans notre étude, nous avons testé à plusieurs reprises la nullité de la variance de certains effets aléatoires. Dans ce cas, le modèle M_0 correspond au modèle dans lequel la variance est nulle (le paramètre n'est donc pas aléatoire), et M_1 est le modèle dans lequel la variance testée est positive. La procédure de test définie dans ce cas pose quelques problèmes. En effet, elle consiste à tester un paramètre en bordure de l'espace paramétrique ($H_0 : \sigma^2 = 0$ contre $H_1 : \sigma^2 > 0$, la valeur 0 étant à la frontière de l'espace des paramètres puisque σ^2 est définie sur $\mathbb{R}^+$).
Dans ce cas, la statistique de test LRT suit asymptotiquement la loi suivante : $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_0$, où χ^2_0 est la masse de Dirac en 0 (Gourieroux et Montfort, 1989 ; Self et Liang, 1987).

Critères de sélection de modèles

Les différents modèles étudiés ont été comparés en utilisant les tests classiques de sélection de modèles : le test de rapport de vraisemblance lorsque les modèles étaient emboités, les critères d'Akaike (AIC, Akaike, 1973) et de Schwarz (BIC, Schwarz, 1978) sinon.

Critère AIC : $AIC(\mathcal{M}) = -2L(\hat{\boldsymbol{\theta}}; \mathbf{y}) + 2 * nb.par.$, où $nb.par.$ correspond au nombre de paramètres à estimer dans ce modèle.

Critère BIC : $BIC(\mathcal{M}) = -2L(\hat{\boldsymbol{\theta}}; \mathbf{y}) + log(N) * nb.par.$, où N est le nombre d'individus dans l'étude et $nb.par.$ le nombre de paramètres à estimer dans ce modèle.

Plus la valeur du critère AIC(ou BIC) est petite, meilleur est le modèle d'un point de vue statistique.

4.3 Résultats

4.3.1 Etape 1 : Choix des meilleurs modèles linéaire et non linéaire homogènes

D'après Robert-Granié et al. (2004), le polynôme conventionnel [M2] décrit mal la courbe des SCS en début de lactation, et la fonction correspondante au modèle fixe d'Ali et Schaeffer (1987), noté [M3], tend à moins bien décrire l'évolution des SCS en fin de lactation que la courbe définie par le pôlynome fractionnaire [M1]. Bien que les modèles [M1] et [M3] soient statistiquement équivalents, la fonction d'Ali et Schaeffer est plus complexe que le pôlynome fractionnaire, en terme de nombre de paramètres à estimer.
Dans ce sens et d'après les résultats obtenus par Robert-Granié et al. (2004), nous avons choisi le pôlynome fractionnaire comme modèle linéaire de référence, à comparer avec le meilleur modèle non-linéaire étudié.

Le tableau 4.3 présente les estimations de la log-vraisemblance et la valeur des critères AIC et BIC pour chacune des fonctions moyennes étudiées [M1] et [M4] à [M8]. A ce stade de l'analyse, nous avons supposé que les modèles étaient à variances homogènes, saturés sur les effets aléatoires et saturés sur les effets fixes jusqu'aux interactions de premier ordre.
D'après le tableau 4.3, les valeurs de la log-vraisemblance des observations des modèles [M1] et [M4] à [M8] sont situées entre 17000 et 17900. Parmi les modèles dont la fonction moyenne est non linéaire, les modèles [M4] (Wood), [M6] (Morant et Gnanasakthy) et [M8] (notée "Robert-Granié et al. non linéaire") ont les valeurs des critères AIC et BIC les plus faibles. Le modèle ayant la valeur la plus petite est le modèle défini par Morant et Gnanasakthy (1989), il s'agit statistiquement du meilleur modèle.

Dans un second temps, il est important de comparer les courbes obtenues pour les différents modèles puisque l'objectif principal de cette étude est de modéliser au mieux l'évolution des SCS au cours de la lactation.
La figure 4.1 représente les courbes des profils moyens bruts et des profils moyens définis par les courbes de Wood, "Robert-Granié et al. non linéaire" et Morant et Gnanasakthy. La fonction de Wood ne décrit pas bien la courbe des profils moyens tout au long de la lactation : la pente en début de lactation n'est pas assez faible et la courbe augmente trop vite en fin de lactation. L'estimation de l'intercept de la fonction "Robert-Granié et al. non linéaire" est un peu trop grande, la courbe correspondante est donc décalée vers le haut et décrit moins bien la courbe des profils moyens que la courbe de Morant et Gnanasakthy.
Contrairement à la fonction de Wood, les fonctions de Morant et Gnanasakthy [M6] et "Robert-Granié et al. non linéaire" [M8] présentent un point d'inflexion en milieu de lactation, et par conséquent correspondent mieux à la forme de la courbe des profils moyens des SCS en fin de lactation.
De cette étape, le modèle de Morant et Gnanasakthy est, parmi les modèles testés, le

modèle non linéaire le mieux adapté à l'étude de ces données.

La figure 4.2 représente la courbe des profils moyens bruts et des profils moyens définis par les courbes de Morant et Gnanasakthy ([M6]) et Robert-Granié et al. ([M1]). Les courbes sont quasiment confondues : celle de Robert-Granié et al. atteint une valeur minimum plus petite que celle de Morant et Gnanasakthy. Le minimum semble être atteint au même stade de la lactation pour les deux fonctions.

Dans la suite de l'étude, nous avons mené les étapes 2 à 5 à partir de ces deux modèles, notés respectivement [MG] pour celui de Morant et Gnanasakthy et [R] pour celui de Robert-Granié et al.

4.3.2 Etape 2 : Déterminer les effets fixes et aléatoires de la fonction moyenne

A ce stade de l'étude, les modèles obtenus à l'étape 1 sont à variances homogènes, saturés en effets aléatoires et saturés sur les effets fixes jusqu'aux interactions de premier ordre.

Modèle non linéaire [MG]

Rappelons ici la fonction moyenne non linéaire, proposée par Morant et Gnanasakthy (1989), choisie comme la meilleure des fonctions non linéaires testées à l'étape 1 :

$$f\big(t_{ij}, (\phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i})\big) = \phi_{1i} \exp[\phi_{2i} t^*_{ij} + \phi_{3i} t^{*\,2}_{ij} + \phi_{4i}/t_{ij}]$$

où $(\phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i}) \sim \mathcal{N}(\mathbf{A}_i \boldsymbol{\mu}, \boldsymbol{\Gamma})$.

Après estimation de la matrice de covariance $\boldsymbol{\Gamma}$, nous avons obtenu les estimations suivantes concernant les variances (et leur erreurs standards) : $Var(\phi_{1i}) = 2.24(0.32)$, $Var(\phi_{2i}) = 2.66(0.43)$, $Var(\phi_{3i}) = 1.77(0.30)$, $Var(\phi_{4i}) = 10.09(2.38)$, et la matrice de corrélation correspondant à $\boldsymbol{\Gamma}$:

$$\boldsymbol{\rho}_{\boldsymbol{\Gamma}} = \begin{pmatrix} 1 & -0.59 & 0.42 & -0.77 \\ -0.59 & 1 & -0.94 & 0.54 \\ 0.42 & -0.94 & 1 & -0.42 \\ -0.77 & 0.54 & -0.42 & 1 \end{pmatrix}$$

Au vu des estimations des corrélations, nous avons testé trois sous-modèles :
[MG1] dans lequel $\boldsymbol{\phi_3}$ n'est pas corrélé avec $(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_4})$,
[MG2] dans lequel $\boldsymbol{\phi_4}$ n'est pas corrélé avec $(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \boldsymbol{\phi_3})$,

[MG3] dans lequel $(\boldsymbol{\phi_3}, \boldsymbol{\phi_4})$ sont de covariance non-nulle, non corrélés avec $(\boldsymbol{\phi_1}, \boldsymbol{\phi_2})$.

Dans le tableau 4.4, les modèles [MG], [MG1], [MG2] et [MG3] sont comparés à l'aide du test de rapport de vraisemblance, et des critères AIC et BIC.
D'après ce tableau, le modèle [MG] avec une matrice de covariance $\boldsymbol{\Gamma}$ pleine est statistiquement meilleur que [MG1] ($\boldsymbol{\phi}_4$ non corrélé avec les autres paramètres), [MG2] ($\boldsymbol{\phi}_3$ non corrélé avec les autres paramètres) et [MG3] (($\boldsymbol{\phi}_3$,$\boldsymbol{\phi}_4$) non corrélés avec ($\boldsymbol{\phi}_1$,$\boldsymbol{\phi}_2$)).

En conclusion, le meilleur modèle concernant le modèle de Morant et Gnanasakthy est le modèle saturé en effets aléatoires.

Modèle linéaire [R]

Le polynôme fractionnaire proposé par Robert-Granié et al. (2004), et choisi à l'étape 1 comme modèle linéaire de référence s'écrit de la manière suivante :

$$f\big(t_{ij}, (\phi_{1i}, \phi_{2i}, \phi_{3i})\big) = \phi_{1i} + \phi_{2i} t_{ij}^{-1/3} + \phi_{3i} t_{ij}^{-1/3} \log(t_{ij})$$

où $(\phi_{1i}, \phi_{2i}, \phi_{3i}) \sim \mathcal{N}(\mathbf{A}_i \boldsymbol{\mu}, \boldsymbol{\Gamma})$.
Nous avons obtenu les estimations de variances (et erreurs standard suivantes) : $Var(\phi_{1i}) = 30.02(4.16)$, $Var(\phi_{2i}) = 17.99(3.55)$, $Var(\phi_{3i}) = 35.08(5.20)$ et la matrice de corrélation correspondant à la matrice de covariance $\boldsymbol{\Gamma}$:

$$\boldsymbol{\rho}_\Gamma = \begin{pmatrix} 1.00 & 0.09 & 0.94 \\ 0.09 & 1.00 & 0.36 \\ 0.94 & 0.36 & 1.00 \end{pmatrix}$$

Au vu de ces estimations, nous avons testé le sous-modèle suivant :
[R1] dans lequel $\boldsymbol{\phi_2}$ n'est pas corrélé avec $(\boldsymbol{\phi_1}, \boldsymbol{\phi_3})$.

Le tableau 4.5 présente la statistique de test du rapport de vraisemblance et les critères de sélection de modèles pour comparer les modèles [R] et [R1]. Les résultats obtenus montrent que le paramètre $\boldsymbol{\phi_2}$ est corrélé aux paramètres $\boldsymbol{\phi_1}$ et $\boldsymbol{\phi_3}$.

4.3.3 Etape 3 : Sélection des covariables sur le vecteur moyenne $\boldsymbol{\mu}$

Les tableaux 4.6 et 4.7 présentent les résultats des tests de Wald pour déterminer les covariables significatives du vecteur moyenne $\boldsymbol{\mu}$, pour chacun des paramètres $\boldsymbol{\phi}_*$ dans les modèles [R] et [MG].
Dans ces tableaux, sont présentées les valeurs du test de Wald, et les P-valeurs correspondantes, pour les modèles [MG] et [R].

A partir des résultats issus de ces tableaux, les covariables suivantes ont été sélectionnées pour la fonction moyenne [MG] :

- Aucune covariable n'a d'effet significatif sur les paramètres $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_4$,
- La saison de vêlage a un effet significatif sur les paramètres $\boldsymbol{\phi}_2$ et $\boldsymbol{\phi}_3$.

Concernant le modèle [R], aucune covariable n'a d'effet significatif sur les paramètres $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ et $\boldsymbol{\phi}_3$.

4.3.4 Etape 4 : Choix du modèle de variance résiduelle

Les tableaux 4.8 et 4.9 présentent pour les modèles non linéaire [MG] et linéaire [R] les valeurs de la log-vraisemblance, des critères AIC et BIC, du test du rapport de vraisemblance pour les modèles de variance résiduelle [V0] à [V5].

D'après ces tableaux, quelle que soit la structure de variance résiduelle choisie, on améliore les modèles [MG] et [R] en introduisant des variances hétérogènes. En terme de critère AIC, on gagne 100 points en passant du modèle à variance homogène [V0] au modèle [V1] (relation moyenne-variance) et 2000 points des modèles de variance [V0] à [V5] (modèle linéaire mixte sur la log-variance).
On peut remarquer que pour les deux modèles [MG] et [R], les fonctions de variances résiduelles basées sur un modèle structural (modèles [V2] à [V5]) sont statistiquement meilleures que la relation moyenne-variance (modèle [V1]).
D'après les tableaux 4.8 et 4.9, plus la structure de la fonction de variance est complexe en terme de paramètres (c'est-à-dire de [V2] à [V5]), meilleur est le modèle en terme de critères AIC, BIC et de test du rapport de vraisemblance. Le passage d'un modèle structural composé d'effets fixes à un modèle structural mixte (modèles [V2]-[V3] à [V4]-[V5]) améliore considérablement le modèle (une différence de $-2L(\hat{\boldsymbol{\theta}}; y)$ environ égale à 1900), que ce soit pour les modèles [R] ou [MG].

D'après ces tableaux, la fonction de variance [V5] fournit les meilleurs valeurs des critères AIC et BIC pour les modèles [R] et [MG].
La figure 4.3 présente une comparaison graphique des fonctions de variances [V0], [V1] et [V5] pour le modèle [MG]. Nous avons aussi représenté en noir la variance empirique obtenue à partir des résidus du modèle à variances homogènes. La courbe associée à la fonction de variance [V5] décrit mieux que les autres fonctions de variance, la décroissance de la courbe de variance empirique en fin de lactation. Néanmoins la croissance de la courbe de variance empirique en début de lactation n'est prise en compte par aucune des fonctions de variance.

4.3.5 Etapes 5 : Choix des covariables pour les paramètres des fonctions de variances [V4] et [V5]

Nous avons d'abord vérifié que les coefficients des matrices de corrélation associées aux matrices de covariance $\boldsymbol{\Delta}$ des fonctions [V4] et [V5] étaient supérieurs à 0.50 pour chacun des modèles [R] et [MG]. Nous avons donc choisi les modèles de variance [V4] et [V5] saturés en effets aléatoires, puis nous avons déterminé les covariables ayant un effet significatif sur chaque paramètre de ces mêmes fonctions de variance.

Le tableau 4.10 (respectivement 4.11) présente les valeurs du test de Wald pour le modèle [MG] (respectivement [R]) et les fonctions de variance [V4] et [V5].
Concernant le modèle [MG] et la fonction de variance [V4], seules les covariables "campagne" et "race" ont une effet significatif sur les paramètres $\boldsymbol{\delta_0}$ et $\boldsymbol{\delta_1}$. Concernant la fonction de variance [V5], aucune covariable n'a d'effet significatif sur $\boldsymbol{\delta_0}$, $\boldsymbol{\delta_1}$ et $\boldsymbol{\delta_2}$.

Concernant le modèle [R] et la fonction de variance [V4], les covariables ayant un effet significatif sont différentes de celles obtenus pour le modèle [MG] : il s'agit des interactions "saison de vêlage*campagne", "saison de vêlage*race", et des effets simples "saison de vêlage", "campagne", "race", sur le paramètre $\boldsymbol{\delta_0}$, et uniquement des effets simples "campagne" et "race" sur le paramètre $\boldsymbol{\delta_1}$.
Concernant la fonction de variance [V5], comme pour le modèle [MG], aucune covariable n'a d'effet significatif sur les paramètres $\boldsymbol{\delta_0}$, $\boldsymbol{\delta_1}$ et $\boldsymbol{\delta_2}$.

Le tableau 4.12 présente une comparaison des modèles linéaires et non linéaires obtenus après les principales étapes de l'étude, avec les valeurs des critères AIC, BIC et du test du rapport de vraisemblance.
D'après ce tableau, on a théoriquement amélioré le modèle à chaque étape, avec une nette amélioration par l'introduction des variances hétérogènes. Les meilleurs modèles non linéaire et linéaire obtenus sont finalement :

- Le modèle non linéaire proposé par Morant et Gnanasakthy (1989) pour lequel :
  - les paramètres $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, $\boldsymbol{\phi}_3$ et $\boldsymbol{\phi}_4$ sont supposés aléatoires et corrélés,
  - $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_4$ ne dépendent pas de covariable, $\boldsymbol{\phi}_2$ et $\boldsymbol{\phi}_3$ dépendent de la saison de vêlage de la vache,
  - le modèle de variance résiduelle est le suivant : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i}t^*_{ij} + \delta_{2i}t^{2*}_{ij}$, où $(\delta_{0i}, \delta_{1i}, \delta_{2i}) \sim \mathcal{N}((\delta_0, \delta_1, \delta_2), \boldsymbol{\Delta})$, $(\delta_0, \delta_1, \delta_2)$ sont des intercepts et la matrice $\boldsymbol{\Delta}$ est une matrice de covariance pleine.

- Le modèle linéaire proposé par Robert-Granié et al. (2004) où :
  - les paramètres $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, et $\boldsymbol{\phi}_3$ sont supposés aléatoires et corrélés,
  - $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, et $\boldsymbol{\phi}_3$ ne dépendent d'aucune covariable,
  - le modèle de variance résiduelle est le suivant : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i}t^*_{ij} + \delta_{2i}t^{2*}_{ij}$, où $(\delta_{0i}, \delta_{1i}, \delta_{2i}) \sim \mathcal{N}((\delta_0, \delta_1, \delta_2), \boldsymbol{\Delta})$, $(\delta_0, \delta_1, \delta_2)$ sont des intercepts et la matrice $\boldsymbol{\Delta}$ est

une matrice de covariance pleine.

Enfin, dans le tableau 4.13, nous présentons les estimations et erreurs standards des paramètres obtenus pour les deux modèles finaux linéaire et non linéaire. En terme de critères AIC et BIC, le modèle non linéaire de Morant et Gnanasakthy (1989) (modèle [MG]) est meilleur que le modèle linéaire proposé par Robert-Granié et al. (modèle [R]).

Le graphique 4.4 présente les profils moyens bruts ainsi que les profils moyens associés aux modèles finaux de Robert-Granié et al. et Morant et Gnanasakthy. Les courbes correspondantes aux modèles [R] et [MG] sont très proches l'une de l'autre et décrivent assez bien l'allure des profils moyens bruts. Or nous avons obtenu une nette différence entre les deux modèles en termes de critères AIC et BIC. Dans ce sens, nous avons tracé quelques profils individuels pour savoir si la différence entre ces deux modèles ne se situait pas à ce niveau. La figure 4.5 présente une comparaison des profils brut et des modèles [R] et [MG] pour l'individu 39. La courbe correspondante au modèle linéaire tend à garder la même forme que la courbe des profils moyens, tandis que celle correspondante au modèle non linéaire semble mieux s'adapter à ce profil individuel d'allure très différente de celle des profils moyens. Seul un exemple de profil individuel est présenté ici, mais d'autres profils ont été étudiés et des conclusions similaires ont été obtenues.

4.4 Discussion et conclusion

La stratégie de sélection de modèles adoptée ici est la suivante : nous avons d'abord déterminé les meilleures fonctions "moyenne" dans un modèle à variances homogènes, avec choix des paramètres fixes et aléatoires et détermination des covariables significatives pour chaque paramètre. En second lieu, nous avons introduit des fonctions de variance résiduelle dans ces modèles.
D'autres stratégies auraient pu être menées : par exemple comparer dès le début de l'étude des modèles hétérogènes, dans lesquels tous les paramètres étaient fixes, puis ensuite tester leur nature aléatoire. Dans le cadre des modèles linéaires, Verbeke et Molenberghs (2000) proposent d'étudier d'abord un modèle saturé en effets fixes et aléatoires au niveau de la moyenne, dans lequel des fonctions de variance résiduelle sont comparées. Ensuite la nature aléatoire des paramètres et les covariables agissant sur les effets fixes sont testées. Aucun élément ne permet de conclure quant à la meilleure stratégie à adopter.
Dans notre étude, nous sélectionnons les modèles de manière séquentielle à l'aide de tests et de critères de sélection, sans pour autant s'inquiéter du niveau global obtenu. En fait puisque les modèles finaux sont comparés seulement à l'aide des critères AIC et BIC, le niveau global des procédures de tests n'influe pas sur le choix final des modèles.

Concernant les covariables agissant sur les paramètres de la fonction "moyenne", nous obtenons des résultats différents pour les modèles de Morant et Gnanasakthy (1989) et de Robert-Granié et al. (2004). La covariable "saison de vêlage" a un effet significatif sur les

paramètres $\boldsymbol{\phi}_2$ et $\boldsymbol{\phi}_3$ de la fonction de Morant et Gnanasakthy (1989), ce qui implique que la saison de vêlage de la vache a un impact sur le taux de SCS observé dans l'organisme au milieu et en fin de lactation. Les paramètres de la fonction de Robert-Granié et al. (2004) ne dépendent d'aucune covariable. Etant donné que les paramètres du modèle proposé par Robert-Granié et al. (2004) n'ont pas d'interprétation biologique comparable, nous ne pouvons pas déterminer les mécanismes expliquant cette différence.
Le même phénomène se produit au niveau du choix des covariables de la structure de variance [V4], à ceci près que dans ce cas, nous avons obtenu des covariables significatives différentes pour les paramètres $\boldsymbol{\delta}_0$ et $\boldsymbol{\delta}_1$ d'une même fonction de variance. Ici, cette différence peut s'expliquer par le fait que des effets différents ont été pris en compte sur la fonction moyenne pour les deux modèles et donc que le comportement de la résiduelle face aux covariables testées est différent.

Concernant le choix de la meilleure fonction de variance pour chaque modèle, le modèle mixte sur la log-variance a fourni les valeurs des critères AIC et BIC les plus faibles. Nous aurions pu augmenter la complexité du modèle en augmentant encore le degré du polynôme. En effet, le graphique 4.3 montre que la fonction de variance [V5], décrit mieux la décroissance de la courbe empirique des résidus au cours de la lactation que la relation moyenne-variance, mais en début de lactation, le pic de croissance ne semble pas être pris correctement en compte au niveau de la courbe moyenne, ceci pour chacune des fonctions de variance étudiées. Rappelons tout de même que la fonction de variance [V5], avec ses paramètres de nature aléatoire, prend en compte la variabilité individuelle.
Augmenter le degré du pôlynome du modèle structural de variance résiduelle pourrait être plus performant, mais impose aussi d'estimer plus de paramètres, et dans ce cas nous pourrions être confrontée à des problèmes de convergence de l'algorithme d'estimation. De plus, les pôlynomes de degré élevé ont tendance à moins bien s'ajuster aux extremités des courbes de données, avec des effets de vagues. Une autre solution serait de rechercher une modélisation de la variance à l'aide de fonctions semi ou non paramétriques telles que les B-splines ou méthodes à noyaux (De Boor, 1978 ; Ruppert et al., 2003).

Lorsqu'on étudie de plus près le modèle de Morant et Gnanasakthy (1989), une forme plus naturelle aurait été d'introduire le paramètre $\boldsymbol{\phi}'_1 = \log(\boldsymbol{\phi}_1)$ dans la fonction exponentielle, comme ceci :

$$y_{ij} = \exp(\boldsymbol{\phi}'_1 + \boldsymbol{\phi}_2 t^*_{ij} + \boldsymbol{\phi}_3 t^{2\,*}_{ij} + \boldsymbol{\phi}_4 / t_{ij}) + \sigma^2_{ij}\varepsilon^*_{ij} \tag{4.3}$$

Contrairement à notre étude, ici le paramètre $\boldsymbol{\phi}'_1 = \log(\boldsymbol{\phi}_1)$ est supposé de distribution normale et non log-normale. Nous avons estimé les paramètres de ce modèle, et les valeurs des critères AIC et BIC étaient un peu plus élevées que celles obtenues au cours de l'étude. Dans le même esprit, la structure du modèle non linéaire de Morant et Gnanasakthy (1989) nous permet aussi d'interpréter ce modèle comme un modèle linéaire appliqué au logarithme des données, avec un reparamétrage sur le paramètre $\boldsymbol{\phi}_1$. En effet, il peut s'écrire :

$$w_{ij} = \log(y_{ij}) = \boldsymbol{\phi}'_1 + \boldsymbol{\phi}_2 t^*_{ij} + \boldsymbol{\phi}_3 t^{2\,*}_{ij} + \boldsymbol{\phi}_4 / t_{ij} + \sigma^2_{ij}\varepsilon^*_{ij} \tag{4.4}$$

Ce modèle correspond en fait au modèle de Morant et Gnanasakthy (1989) dans lequel la fonction de variance résiduelle est la relation moyenne-variance (Box et Cox, 1964).
Il serait intéressant de comparer le modèle linéaire (4.4) obtenu sur le logarithme des données et le modèle final de Morant et Gnanasakthy (1989). Comme les deux modèles ne sont pas appliqués au même jeu de données, on ne peut pas comparer directement les valeurs des critères AIC et BIC obtenues pour chaque modèle. Néanmoins le passage de la log-vraisemblance des observations y_{ij} à la log-vraisemblance des nouvelles observations $w_{ij} = \log(y_{ij})$ du modèle (4.4) peut s'obtenir de la manière suivante :

$$\begin{aligned} -2L(\hat{\boldsymbol{\theta}}_1; \mathbf{y}) &= -2L(\hat{\boldsymbol{\theta}}; \mathbf{w}) - 2\log(|\mathbf{J}|) \\ -2L(\hat{\boldsymbol{\theta}}_2; \mathbf{y}) &= -2L(\hat{\boldsymbol{\theta}}; \mathbf{w}) + 2\sum_{i,j} \log(y_{ij}) \end{aligned} \tag{4.5}$$

où $L(\hat{\boldsymbol{\theta}}_1; \mathbf{y})$ est la log-vraisemblance des observations du modèle non linéaire de Morant et Gnanasakthy (1989), $L(\hat{\boldsymbol{\theta}}_2; \mathbf{w})$ est la log-vraisemblance des observations w_{ij} du modèle (4.4) et $|\mathbf{J}|$ est la valeur absolue du déterminant de la matrice jacobienne correspondante au passage des données $\mathbf{w}$ aux données $\mathbf{y}$. $\hat{\boldsymbol{\theta}}_1$ et $\hat{\boldsymbol{\theta}}_2$ correspondent aux estimateurs de $\boldsymbol{\theta}$.
En tenant compte de l'expression (4.5) et en passant des valeurs de la log-vraisemblance des observations $\mathbf{w}$ à la log-vraisemblance des observations $\mathbf{y}$ (seuil minimum des observations y_{ij} à 0.14), on obtient la valeur du critère AIC (respectivement BIC) égale à 18058 (18187), plus élevée que celle obtenue par le modèle final de Morant et Gnanasakthy (1989) que nous avons étudié.
D'un point de vue graphique, nous avons comparé la distribution des résidus standardisés obtenus avec le modèle final de Morant et Gnanasakthy et le modèle linéaire (4.4).
La figure 4.6 présente les histogrammes des résidus standardisés obtenus avec les deux modèles, ainsi que la densité de la loi Gaussienne centrée réduite tracée en rouge. Plus l'histogramme se rapproche de la densité gaussienne, meilleur est le modèle, dans le sens où il décrit mieux les données.
D'après ce graphe, les résidus obtenus via le modèle final de Morant et Gnanasakthy semblent mieux se rapprocher d'un échantillon indépendant distribué sous la densité Gaussienne centrée réduite.
En complément de la figure 4.6, nous avons tracé sur la figure 4.7 les QQplots des résidus standardisés obtenus par ces deux modèles. Le QQPlot (ou Quantile to Quantile Plot) est un graphique dont l'objectif est de tester la conformité entre la distribution empirique d'une variable et une distribution théorique donnée. Nous l'appliquons ici au test de conformité à la distribution normale centrée réduite. La qualité du modèle que nous avons étudié passe par une bonne adéquation entre la droite $y = x$ et le nuage de points des résidus.
La figure 4.7 appuie nos remarques au sujet des histogrammes et des valeurs comparées des critères AIC et BIC : les résidus standardisés obtenus par le modèle linéaire sur le logarithme des données ne forment pas un échantillon distribué sous la loi normale centrée réduite. Le nuage de points correspondant à ce modèle se rapproche beaucoup moins bien de la diagonale que ne le fait celui correspondant au modèle de Morant et Gnanasakthy.

Puisque le modèle linéaire sur le logarithme des données correspond au modèle de Morant et Gnanasakthy dans lequel la fonction de variance résiduelle est la relation moyenne-variance, les remarques précédentes valident les résultats du tableau 4.8, dans lequel nous avons obtenu une nette diminution du critère AIC pour la fonction de variance [V5] par rapport à [V1].

Pour conclure cette étude d'un point de vue statistique, le modèle non linéaire de Morant et Gnanasakthy est meilleur que le modèle linéaire de Robert-Granié et al. Néanmoins, si l'on s'intéresse à la représentation graphique des deux courbes moyennes obtenues (figure 4.4), les deux modèles décrivent de manière similaire et assez bien la courbe des profils moyens. Par contre au niveau du profil individuel présenté sur la figure 4.5, le modèle de Morant et Gnanasakthy décrit beaucoup mieux l'allure de la courbe que le modèle de Robert-Granié et al. Ce graphique montre que d'une part l'aspect mixte et d'autre part l'aspect non linéaire du modèle [MG] permettent de bien prendre en compte la variabilité individuelle et particulièrement pour des individus ayant un profil individuel éloigné du profil moyen des CCS.

Année de vêlage	Nombre d'animaux (nombre d'observations)		Total
	Holstein	Holstein × Normande	
1998	36 (1438)	38 (1492)	74 (2930)
1999	19 (779)	66 (2739)	85 (3518)
Total	55 (2217)	104 (4231)	159 (6448)

TABLE 4.1 – Nombre d'animaux et d'observations par race et par année de vêlage.

Saison de vêlage	Nombre d'animaux (nombre d'observations)		Total
	Holstein	Holstein × Normande	
1 : août - sept.	26 (1055)	70 (2847)	96 (3902)
2 : oct. - nov.	15 (622)	23 (958)	38 (1580)
3 : déc. - mai	14 (540)	11 (426)	25 (966)
Total	55 (2217)	104 (4231)	159 (6448)

TABLE 4.2 – Nombre d'animaux et d'observations par race et par saison de vêlage.

Modèles [a]	par. [b]	$-2L^{(c)}$	$AIC^{(d)}$	$BIC^{(e)}$
[M1] Robert-Granié et al.	37	17481	17555	17668
[M4] Wood	37	17448	17522	17636
[M5] Mitscherlich	37	17871	17945	18059
[M6] Morant et Gnanasakthy	51	17069	17171	17327
[M7] Davidian et Giltinan	51	17642	17744	17901
[M8] Robert-Granié et al. Non Linéaire	51	17282	17384	17541

[a] Ces modèles sont à variances homogènes, saturés sur les effets aléatoires et saturés sur les effets fixes jusqu'aux interactions de premier ordre

[b] par. correspond aux nombre de paramètres à estimer dans le modèle

[c] $-2L$: L correspond à l'estimation de la log-vraisemblance des observations

[d] $AIC = -2L + 2 * par.$

[e] $BIC = -2L + \log(N) * par.$

TABLE 4.3 – Critères de sélection de modèles pour différentes fonctions sur la moyenne.

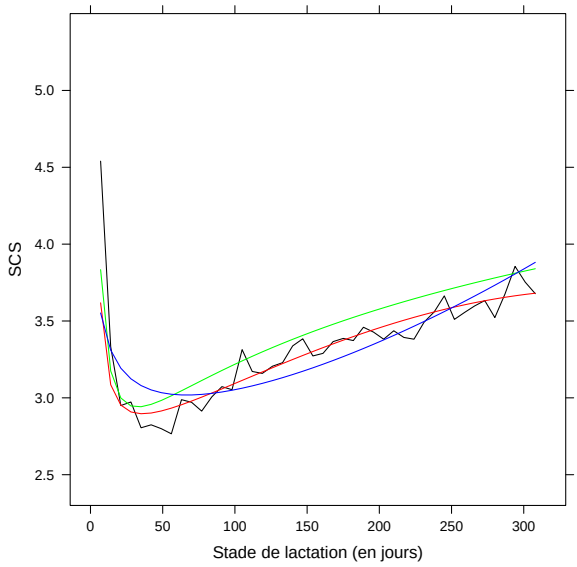


FIGURE 4.1 – Comparaison des profils moyens de SCS (en noir) avec les courbes correspondant au modèle de Wood (bleu), au modèle de Morant et Gnanasakthy (en rouge) et au modèle Robert-Granié et al. Non Linéaire (en vert), en fonction du stade de lactation.

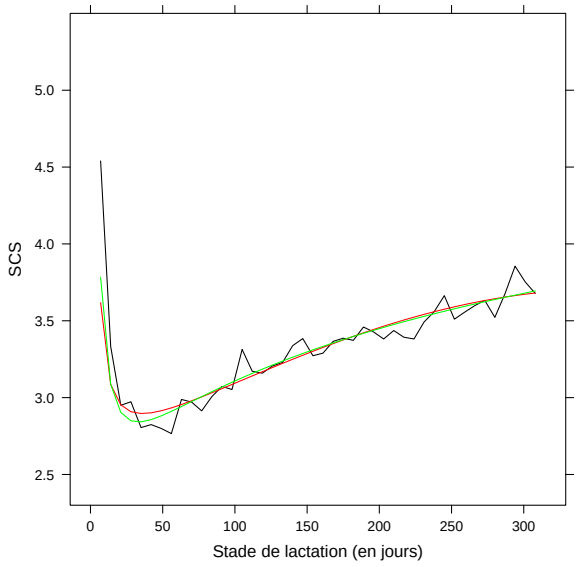


FIGURE 4.2 – Comparaison des profils moyens des SCS (en noir) avec les courbes correspondant au modèle de Morant et Gnanasakthy (en rouge) et au modèle linéaire de Robert-Granié et al. (en vert), en fonction du stade de lactation.

Modèles	par.$^{(a)}$	$-2L^{(b)}$	Comparaison	$diff(-2L)^{(c)}$	P-valeur	AIC	BIC
[MG]	51	17068.62	-	-	-	17170.62	17327.13
[MG1]	48	17409.54	[MG] vs [MG1]	340.92	<0.001	17505.57	17652.88
[MG2]	48	17193.52	[MG] vs [MG2]	124.90	<0.001	17289.52	17436.83
[MG3]	46	17461.78	[MG] vs [MG3]	393.16	<0.001	17553.79	17694.96
			[MG1] vs [MG3]	52.24	<0.001		
			[MG2] vs [MG3]	268.26	<0.001		

$^{(a)}$ par. : nombre de paramètres du modèle

$^{(b)}$ $-2L$: -2*log(vraisemblance des observations)

$^{(c)}$ $diff(-2L)$: différence des valeurs de $-2L$ pour les deux modèles considérés

[MG] : $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, $\boldsymbol{\phi}_3$ et $\boldsymbol{\phi}_4$ sont aléatoires et corrélés

[MG1] : $\boldsymbol{\phi}_3$ non corrélé avec ($\boldsymbol{\phi}_1$,$\boldsymbol{\phi}_2$,$\boldsymbol{\phi}_4$)

[MG2] : $\boldsymbol{\phi}_4$ non corrélé avec ($\boldsymbol{\phi}_1$,$\boldsymbol{\phi}_2$,$\boldsymbol{\phi}_3$)

[MG3] : ($\boldsymbol{\phi}_3, \boldsymbol{\phi}_4$) de covariance non-nulle, non corrélé avec ($\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$)

TABLE 4.4 – Etude du modèle homogène [MG] avec des structures de covariance $\boldsymbol{\Gamma}$ différentes.

Modèles	par.$^{(a)}$	$-2L$ $^{(b)}$	Comparaison	$diff(-2L)^{(c)}$	P-valeur	AIC	BIC
[R]	37	17480.90	-	-	-	17554.90	17668.45
[R1]	35	17589.10	[R] vs [R1]	108.20	<0.001	17659.10	17766.51

$^{(a)}$ par. : nombre de paramètres du modèle

$^{(b)}$ $-2L$: -2*Log(vraisemblance des observations)

$^{(c)}$ $diff(-2L)$: différence des valeurs de $-2L$ pour les deux modèles considérés

[R] : $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ et $\boldsymbol{\phi}_3$ sont aléatoires et corrélés

[R1] : $\boldsymbol{\phi}_2$ non corrélé avec ($\boldsymbol{\phi}_1$,$\boldsymbol{\phi}_3$)

TABLE 4.5 – Etude du modèle homogène [R] avec des structures de covariance $\boldsymbol{\Gamma}$ différentes.

Covariables	Valeur du test	P-valeur
Sur le paramètre ϕ_1		
saison de vêlage (3 niveaux)	2.97	0.23
campagne (2 niveaux)	0.42	0.52
race (2 niveaux)	0.17	0.68
saison de vêlage*campagne	0.25	0.88
saison de vêlage*race	1.37	0.50
campagne*race	0.74	0.39
Sur le paramètre ϕ_2		
saison de vêlage (3 niveaux)	11.87	0.003*
campagne (2 niveaux)	0.13	0.72
race (2 niveaux)	0.14	0.71
saison de vêlage*campagne	0.74	0.69
saison de vêlage*race	0.51	0.77
campagne*race	0.009	0.92
Sur le paramètre ϕ_3		
saison de vêlage (3 niveaux)	14.11	0.0009*
campagne (2 niveaux)	0.84	0.36
race (2 niveaux)	0.03	0.86
saison de vêlage*campagne	0.42	0.81
saison de vêlage*race	1.16	0.56
campagne*race	0.0002	0.99
Sur le paramètre ϕ_4		
saison de vêlage (3 niveaux)	2.51	0.29
campagne (2 niveaux)	0.13	0.72
race (2 niveaux)	0.77	0.38
saison de vêlage*campagne	0.99	0.61
saison de vêlage*race	1.39	0.50
campagne*race	0.05	0.82

* Significatif au niveau $\alpha = 0.05$

TABLE 4.6 – Sélection des covariables sur les paramètres de la fonction moyenne [MG] à l'aide du test de Wald.

Covariables	Valeur du test	P-valeur
Sur le paramètre ϕ_1		
saison de vêlage (3 niveaux)	3.58	0.17
campagne (2 niveaux)	1.35	0.25
race (2 niveaux)	3.64	0.06
saison de vêlage*campagne	1.06	0.59
saison de vêlage*race	0.35	0.84
campagne*race	0.14	0.71
Sur le paramètre ϕ_2		
saison de vêlage (3 niveaux)	1.22	0.54
campagne (2 niveaux)	0.99	0.32
race (2 niveaux)	0.17	0.68
saison de vêlage*campagne	5.32	0.07
saison de vêlage*race	1.13	0.57
campagne*race	0.11	0.74
Sur le paramètre ϕ_3		
saison de vêlage (3 niveaux)	1.92	0.38
campagne (2 niveaux)	2.81	0.09
race (2 niveaux)	2.58	0.11
saison de vêlage*campagne	0.15	0.93
saison de vêlage*race	0.12	0.94
campagne*race	0.10	0.92

* Significatif au niveau $\alpha = 0.05$

TABLE 4.7 – Sélection des covariables sur les paramètres de la fonction moyenne [R] à l'aide du test de Wald.

Modèles	par.[(a)]	$-2L$[(b)]	Comparaison	$diff(-2L)$[(c)]	P-valeur[(d)]	AIC	BIC
[V0]	19	17111.44	-	-	-	17149.44	17207.75
[V1]	20	17005.87	[V1] vs [V0]	105.57	<0.001	17045.87	17107.24
[V2]	20	16852.04	[V2] vs [V0]	259.40	<0.001	16892.04	16953.42
[V3]	21	16790.92	[V3] vs [V0]	320.52	<0.001	16832.92	16897.36
			[V3] vs [V2]	61.12	<0.001		
[V4]	23	15216.69	[V4] vs [V0]	1894.75	<0.001	15262.69	15333.28
			[V4] vs [V2]	1635.35	<0.001		
[V5]	27	14948.98	[V5] vs [V0]	2162.46	<0.001	15002.98	15085.84
			[V5] vs [V3]	1841.94	<0.001		
			[V5] vs [V4]	267.71	<0.001		

[(a)] par. : nombre de paramètres du modèle

[(b)] $-2L$: $-2*$log(vraisemblance des observations)

[(c)] $diff(-2L)$: différence des valeurs de $-2L$ entre les deux modèles considérés

[(d)] P-valeur correspond à la P-valeur obtenue dans le test du rapport de vraisemblance entre les deux modèles comparés

[V0] : modèle à variances homogènes, i.e. $\sigma^2_{ij} = \sigma^2 \, \forall i, j$
[V1] : $\sigma^2_{ij} = \delta_1 f_{ij}^{\delta_2}$
[V2] : $\log(\sigma^2_{ij}) = \delta_0 + \delta_1 t^*_{ij}$
[V3] : $\log(\sigma^2_{ij}) = \delta_0 + \delta_1 t^*_{ij} + \delta_2 {t^*_{ij}}^2$
[V4] : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i} t^*_{ij}$
[V5] : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i} t^*_{ij} + \delta_{2i} {t^*_{ij}}^2$

TABLE 4.8 – Etude de différentes fonctions de variance résiduelle sur le modèle [MG].

Modèles	par.[(a)]	$-2L$[(b)]	Comparaison	$diff(-2L)$[(c)]	P-valeur[(d)]	AIC	BIC
[V0]	10	17521.71	-	-	-	17541.71	17572.40
[V1]	11	17401.29	[V1] vs [V0]	120.42	<0.001	17423.29	17457.05
[V2]	11	17338.30	[V2] vs [V0]	183.41	<0.001	17360.29	17394.05
[V3]	12	17326.98	[V3] vs [V0]	194.73	<0.001	17350.98	17387.81
			[V3] vs [V2]	11.32	0.0008		
[V4]	14	15517.62	[V4] vs [V0]	2004.09	<0.001	15545.62	15588.58
			[V4] vs [V2]	1820.68	<0.001		
[V5]	18	15201.10	[V5] vs [V0]	2320.61	<0.001	15237.10	15292.34
			[V5] vs [V3]	2125.88	<0.001		
			[V5] vs [V4]	316.52	<0.001		

(a) par. : nombre de paramètres du modèle

(b) $-2L$: -2*log(vraisemblance des observations)

(c) $diff(-2L)$: différence des valeurs de $-2L$ entre les deux modèles considérés

(d) P-valeur correspond à la P-valeur obtenue dans le test du rapport de vraisemblance entre les deux modèles comparés

[V0] : modèle à variances homogènes, i.e. $\sigma^2_{ij} = \sigma^2 \, \forall i, j$
[V1] : $\sigma^2_{ij} = \delta_1 f_{ij}^{\delta_2}$
[V2] : $\log(\sigma^2_{ij}) = \delta_0 + \delta_1 t^*_{ij}$
[V3] : $\log(\sigma^2_{ij}) = \delta_0 + \delta_1 t^*_{ij} + \delta_2 {t^*_{ij}}^2$
[V4] : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i} t^*_{ij}$
[V5] : $\log(\sigma^2_{ij}) = \delta_{0i} + \delta_{1i} t^*_{ij} + \delta_{2i} {t^*_{ij}}^2$

TABLE 4.9 – Etude de différentes fonctions de variance résiduelle sur le modèle [R].

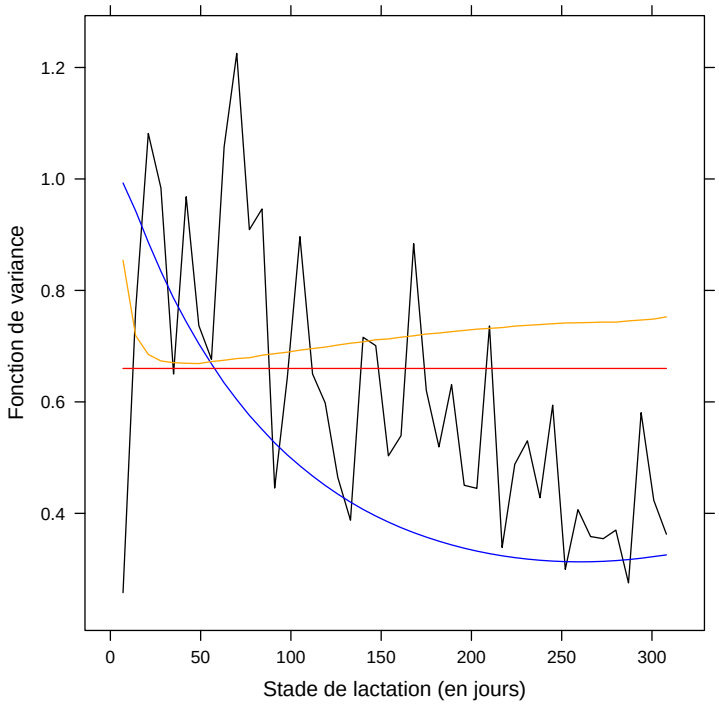


FIGURE 4.3 – Comparaison des fonctions de variance obtenues pour le modèle non linéaire de Morant et Gnanasakthy (1989) : la variance empirique en fonction du temps (en noir), la variance constante [V0] du modèle homogène (en rouge), la relation moyenne-variance [V1] (en orange) et la fonction de variance [V5] (en bleu).

	Modèles de variance résiduelle			
	[V4]		[V5]	
Covariables	Valeur du test	P-valeur	Valeur du test	P-valeur
Sur le paramètre $\boldsymbol{\delta}_0$				
saison de vêlage (3 niveaux)	2.74	0.25	0.46	0.79
campagne (2 niveaux)	9.39	0.002*	0.23	0.63
race (2 niveaux)	12.32	<0.001*	2.92	0.08
saison de vêlage*campagne	4.60	0.10	1.56	0.46
saison de vêlage*race	6.65	0.04	2.61	0.27
campagne*race	0.31	0.58	0.35	0.55
Sur le paramètre $\boldsymbol{\delta}_1$				
saison de vêlage (3 niveaux)	1.75	0.42	2.40	0.30
campagne (2 niveaux)	10.45	0.001*	1.90	0.17
race (2 niveaux)	10.36	0.001*	0.03	0.86
saison de vêlage*campagne	2.23	0.33	0.08	0.96
saison de vêlage*race	1.48	0.48	0.22	0.90
campagne*race	0.21	0.65	0.11	0.74
Sur le paramètre $\boldsymbol{\delta}_2$				
saison de vêlage (3 niveaux)			2.46	0.29
campagne (2 niveaux)			1.75	0.19
race (2 niveaux)			0.24	0.62
saison de vêlage*campagne			0.10	0.95
saison de vêlage*race			0.31	0.86
campagne*race			0.04	0.84

* Significatif au niveau $\alpha = 0.05$

TABLE 4.10 – Sélection des covariables pour les fonctions de variance [V4] et [V5] dans le modèle [MG], à l'aide du test de Wald (statistique de test et P-valeur associée).

	Modèles de variance résiduelle			
	[V4]		[V5]	
Covariables	Valeur du test	P-valeur	Valeur du test	P-valeur
Sur le paramètre $\boldsymbol{\delta}_0$				
saison de vêlage (3 niveaux)	0.29	0.87	0.29	0.87
campagne (2 niveaux)	0.07	0.79	0.33	0.57
race (2 niveaux)	23.14	$<0.001^*$	3.20	0.07
saison de vêlage*campagne	7.33	0.02^*	1.78	0.41
saison de vêlage*race	9.02	0.01^*	2.25	0.32
campagne*race	0.18	0.67	0.11	0.74
Sur le paramètre $\boldsymbol{\delta}_1$				
saison de vêlage (3 niveaux)	2.10	0.35	1.49	0.47
campagne (2 niveaux)	12.07	$< 0.001^*$	2.00	0.16
race (2 niveaux)	8.11	0.004^*	0.01	0.92
saison de vêlage*campagne	1.99	0.37	0.18	0.91
saison de vêlage*race	1.56	0.46	0.56	0.76
campagne*race	0.06	0.81	0.003	0.96
Sur le paramètre $\boldsymbol{\delta}_2$				
saison de vêlage (3 niveaux)			1.42	0.49
campagne (2 niveaux)			2.82	0.09
race (2 niveaux)			0.29	0.59
saison de vêlage*campagne			0.21	0.90
saison de vêlage*race			0.82	0.66
campagne*race			0.0008	0.98

* Significatif au niveau $\alpha = 0.05$

TABLE 4.11 – Sélection des covariables pour les fonctions de variance [V4] et [V5] dans le modèle [R], à l'aide du test de Wald (statistique de test et P-valeur associée).

Modèles	par.[(a)]	$-2L$[(b)]	Comparaison	$diff(-2L)$[(c)]	P-valeur[(d)]	AIC	BIC	$diff(BIC)$[(e)]
[MG.A]	51	17068.62	-	-	-	17170.62	17327.13	0.00
[MG.B]	19	17111.44	[MGB] vs [MGA]	42.82	0.10	17149.44	17207.75	119.38
[MG.C]	27	15202.09	[MGC] vs [MGB]	1909.35	<0.001	15256.09	15338.95	1868.80
[MG.D]	27	14948.98	[MGD] vs [MGB]	2162.46	<0.001	15002.98	15085.84	253.11
[R.A]	37	17480.90	-	-	-	17554.90	17668.45	0.00
[R.B]	10	17521.71	[RB] vs [RA]	40.81	0.04	17541.71	17572.40	96.05
[R.C]	24	15476.73	[RC] vs [RA]	2044.98	<0.001	15524.73	15598.39	1974.01
[R.D]	18	15201.10	[RD] vs [RA]	2320.61	<0.001	15237.10	15292.34	306.05

[(a)] par. : nombre de paramètres du modèle

[(b)] $-2L$: -2*log(vraisemblance des observations)

[(c)] $diff(-2L)$: différence des valeurs de $-2L$ entre les deux modèles considérés

[(d)] P-valeur obtenue dans le test de maximum de vraisemblance entre les deux modèles comparés

[(e)] $diff(BIC)$: différence des critères BIC entre les deux modèles considérés

[.A] : modèle à variances homogènes, saturé en effets aléatoires et en effets fixes

[.B] : modèle à variances homogènes obtenu après sélection des covariables sur la moyenne

[.C] : modèle [V4] après sélection des covariables sur la structure de variance résiduelle

[.D] : modèle [V5] après sélection des covariables sur la structure de variance résiduelle

TABLE 4.12 – Comparaison des modèles étendus [MG] et [R] obtenus lors des différentes étapes.

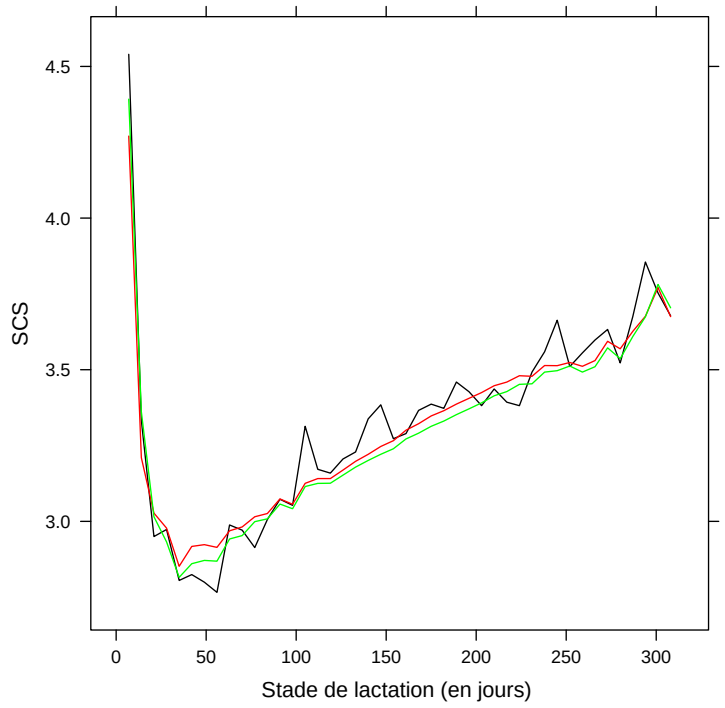


FIGURE 4.4 – Comparaison des profils de courbes obtenus avec le modèle [MG] hétérogène de fonction de variance [V5] (en rouge) et le modèle [R] hétérogène de fonction de variance [V5] (en vert).

Paramètres	Modèle [R] final		Modèle [MG] final	
μ_1	7.59	(0.48)	2.66	(0.14)
μ_2 intercept	2.91	(0.52)	1.58	(0.27)
μ_2 effet "saison de vêlage" 1			-1.19	(0.30)
μ_2 effet "saison de vêlage" 2			-0.79	(0.35)
μ_3 intercept	-5.18	(0.55)	-1.21	(0.24)
μ_3 effet "saison de vêlage" 1			1.22	(0.27)
μ_3 effet "saison de vêlage" 2			0.68	(0.31)
μ_4			2.59	(0.83)
δ_1	-0.05	(0.25)	0.04	(0.18)
δ_2	-2.45	(1.06)	-2.85	(0.79)
δ_3	1.50	(1.00)	1.69	(0.75)
$\Gamma(1,1)$	26.83	(4.50)	2.01	(0.32)
$\Gamma(1,2)$	-0.48	(3.18)	-1.02	(0.36)
$\Gamma(1,3)$	-26.32	(4.98)	0.47	(0.26)
$\Gamma(1,4)$			-3.15	(0.93)
$\Gamma(2,2)$	15.11	(5.47)	1.78	(0.60)
$\Gamma(2,3)$	-5.86	(4.21)	-1.23	(0.46)
$\Gamma(2,4)$			1.78	(1.34)
$\Gamma(3,3)$	29.57	(5.85)	0.98	(0.36)
$\Gamma(3,4)$			-0.83	(1.02)
$\Gamma(4,4)$			8.81	(3.30)
$\Delta(1,1)$	1.67	(0.46)	1.75	(0.33)
$\Delta(1,2)$	-5.84	(1.88)	-5.95	(1.40)
$\Delta(1,3)$	4.52	(1.65)	4.48	(1.32)
$\Delta(2,2)$	44.37	(9.00)	42.34	(7.50)
$\Delta(2,3)$	-44.19	(8.54)	-41.54	(7.48)
$\Delta(3,3)$	47.36	(8.51)	44.21	(7.73)
$-2L$	15201.10		14948.97	
AIC	15237.10		15002.97	
BIC	15292.34		15085.84	

TABLE 4.13 – Estimations et erreurs standards (entre parenthèses) des paramètres obtenus pour les deux meilleurs modèles finaux [R] et [MG].

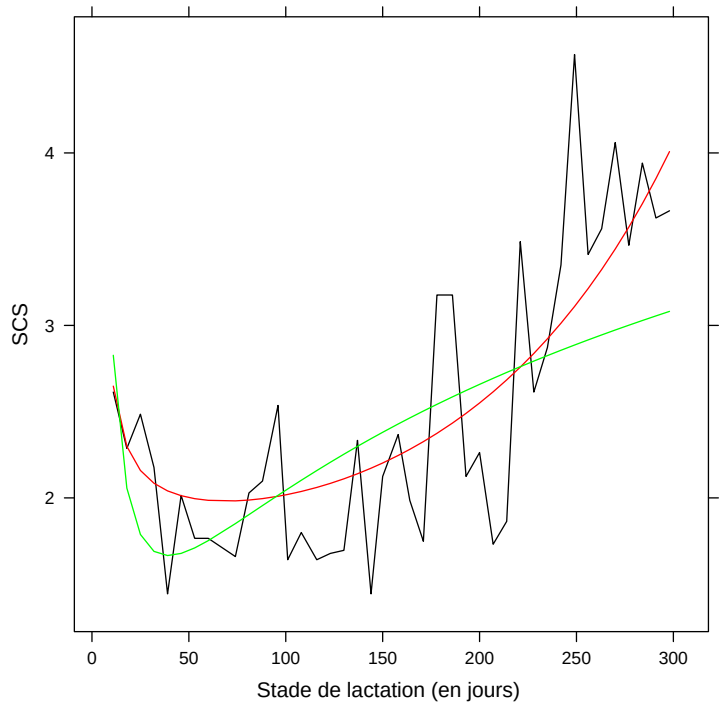


FIGURE 4.5 – Comparaison des profils de courbes individuelles (individu 39) obtenus avec le modèle [MG] hétérogène de fonction de variance [V5] (en rouge) et le modèle [R] hétérogène de fonction de variance [V5] (en vert).

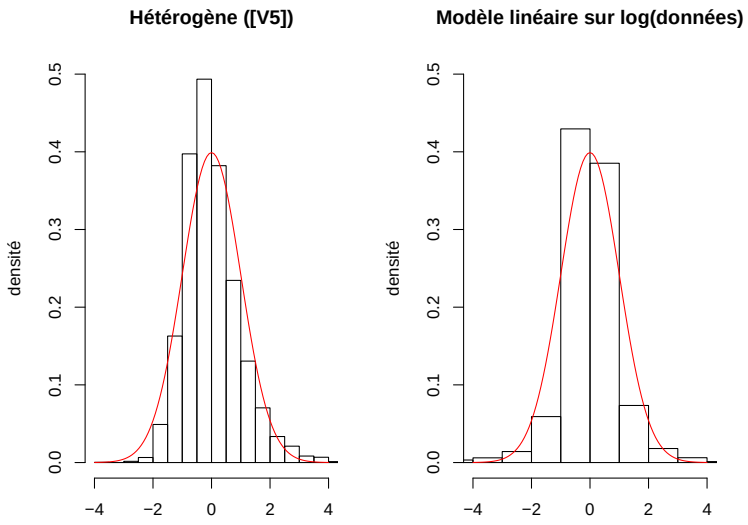


FIGURE 4.6 – Comparaison des histogrammes des résidus obtenus avec le modèle [MG] hétérogène de fonction de variance [V5] (à gauche) et le modèle linéaire sur le logarithme des données (à droite).

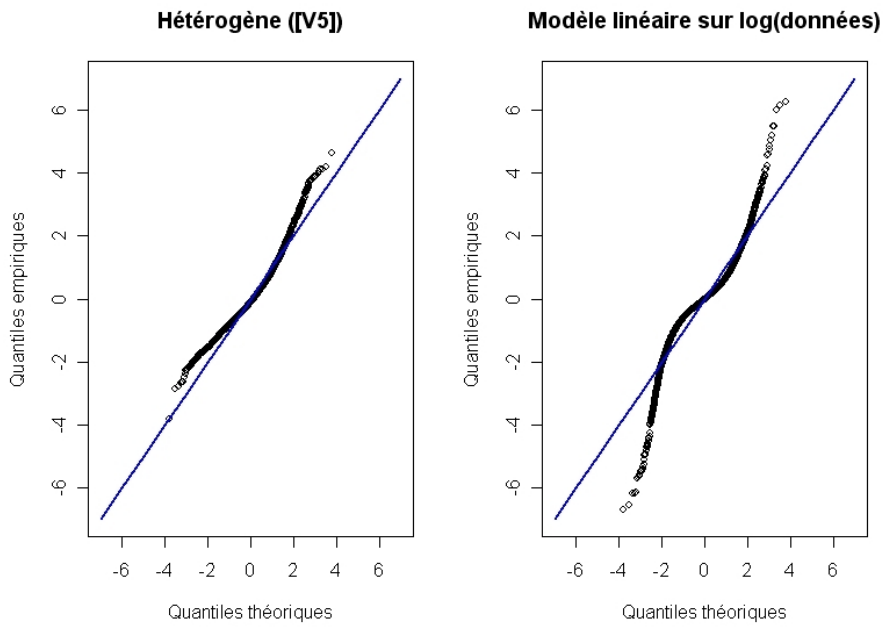


FIGURE 4.7 – Comparaison des QQplots des résidus standardisés obtenus avec le modèle [MG] hétérogène de fonction de variance [V5] (à gauche) et le modèle linéaire associé basé sur le logarithme des données (à droite).

Discussion générale

Dans l'écriture d'un modèle mixte non linéaire, les structures de la fonction "moyenne" et de la variance résiduelle doivent être spécifiées. Ces deux parties ne sont pas indépendantes : la modélisation d'une des parties influe sur l'estimation des paramètres de l'autre partie et vice versa. Dans notre étude, nous nous sommes particulièrement intéressée à la modélisation des variances résiduelles du modèle, en supposant que la fonction "moyenne" décrivait correctement les données. Néanmoins, en pratique, la modélisation des deux parties est nécessaire. Nous avons présenté dans le chapitre 4 de ce manuscrit une étude dans laquelle des fonctions "moyennes" ont été comparées à l'aide d'un modèle homogène, puis des fonctions de variance ont été introduites dans le modèle. D'autres stratégies auraient pu être menées, il serait intéressant de comparer les résultats obtenus. Comme dans beaucoup de problèmes de ce genre (cf par exemple la régression multivariée), il n'y a malheureusement pas de procédé optimum de sélection.

Concernant la structure de la fonction "moyenne", on doit choisir entre un modèle linéaire et non linéaire, puis entre un modele à effets fixes et un modèle comportant des effets aléatoires. Contrairement aux modèles à effets fixes, les modèles mixtes permettent de prendre en compte la variabilité entre unités expérimentales, et sont matérialisés par la nature aléatoire de certains paramètres. Dans certaines situations, il n'est pas facile de statuer entre les deux types d'effets, notamment si l'on s'en réfère à la façon dont ces effets ont été échantillonnés. L'exemple classique des troupeaux et des taureaux traités respectivement comme fixes et aléatoires en sélection animale illustre bien la gêne du modélisateur qui pourrait très bien soutenir l'argumentation inverse. En fait, dans beaucoup de cas, certains paramètres sont traités comme aléatoires de façon à obtenir des estimateurs "shrinkés" à la James Stein dont on sait qu'ils ont de meilleures propriétés que ceux des moindres carrés.

Dans notre étude, nous avons choisi de modéliser les variances résiduelles de façon paramétrique. Deux modèles flexibles ont été proposés : une relation "puissance" entre la moyenne et la variance et un modèle linéaire mixte sur la log-variance. La relation moyenne-variance sert essentiellement à prendre en compte les effets d'échelle, une puissance au carré correspondant à un coefficient de variation constant. Beaucoup pensent qu'on peut y remédier par une simple transformation des données de type Box-Cox. En fait, de telles transformations sont beaucoup moins anodines qu'il n'y paraît. Elles sont souvent très sensibles aux "outliers" et peuvent également détruire complètement des

relations remarquables (linéarité par exemple) entre la réponse et les covariables (cf un exemple dans Wolfinger, 1996). Dans les applications aux jeux de données réelles que nous avons traitées, la prise en compte des variances résiduelles hétérogènes a amélioré le modèle dans tous les cas. D'autres modèles, issus d'un cadre non-paramétrique ou semi-paramétrique, pourraient être considérés notamment si les covariables explicatives de l'hétérogénéité de variances sont continues (De Boor, 1978 ; Torres, 2001 ; Ruppert et al., 2003).

Intéressons nous maintenant plus particulièrement à l'algorithme SAEM-MCMC, choisi comme méthode d'estimation tout au long de notre travail. Comparée à la plupart des méthodes classiques, notre méthode est plus flexible et fournit de très bons résultats dans l'estimation des deux structures de variance étudiées, y compris celle faisant intervenir un modèle mixte. De nombreux auteurs ont montré que les méthodes d'approximation comme SAS-Nlmixed ou R-nlme, plus simples et plus rapides que les méthodes "exactes" basées sur l'algorithme EM stochastique, pouvaient dans certains cas fournir des erreurs d'approximation importantes (Davidian et Giltinan, 1995 ; Pinheiro et Bates, 1995 ; Lindstrom et Bates, 1990). Dans les applications que nous avons traitées, nous avons obtenu des résultats assez proches entre l'algorithme SAEM-MCMC et ces méthodes approchées. Néanmoins en pratique, et par sécurité, il vaut certainement mieux utiliser une méthode "exacte" (aux erreurs près du processus stochastique, qui interviennent à la phase E de l'algorithme) comme l'algorithme SAEM-MCMC. De plus, les fonctions de variance que nous avons étudiées, et qui sont susceptibles d'améliorer considérablement l'efficacité des modèles, ne sont pas toutes disponibles dans les logiciels basés sur ces méthodes approchées.

Parmi les méthodes exactes basées sur la théorie du maximum de vraisemblance (ML), plusieurs méthodes ont été proposées : les méthodes de quadratures de Gauss, difficiles à utiliser en pratique quand le nombre de paramètres de nature aléatoire du modèle augmente, ainsi que les procédures stochastiques basées sur l'algorithme EM. Dans notre étude, nous avons choisi d'estimer le vecteur des paramètres via l'algorithme SAEM-MCMC, dont l'avantage par rapport à l'algorithme MCEM est le gain considérable de temps de calcul, dû au recyclage des simulations à chaque itération.
Dans cet algorithme, de nombreux paramètres sont calibrés par l'utilisateur : entres autres la taille des chaînes dans l'algorithme de Metropolis-Hastings et le nombre d'itérations de l'algorithme SAEM. Nous avons montré sur des exemples simples que ces paramètres pouvaient compromettre la convergence de l'algorithme vers l'estimateur du maximum de vraisemblance, c'est pourquoi nous avons proposé dans le chapitre 2, des critères pour fixer ces paramètres et établir la convergence.

Lorsqu'on étudie les propriétés d'un algorithme, dans le but de l'appliquer à des données ou de l'adapter à d'autres situations, il faut s'assurer au préalable qu'il fournit les résultats attendus pour des modèles simples : il s'agit de la validation de l'algorithme. Elle permet de vérifier que les programmes informatiques ne comportent pas d'erreurs et valide les

techniques numériques à la base de la méthode d'estimation.
Tout au long de notre travail de thèse, nous avons toujours essayé de valider les résultats obtenus par des méthodes de maximum de vraisemblance standard en confrontant nos résultats à ceux obtenus par d'autres procédés numériques : soit par l'algorithme EM lui-même mais dans ses versions analytiques (cas du modèle linéaire), soit par les méthodes classiques utilisées dans le cadre des modèles non linéaires mixtes par des logiciels produisant des estimations ML "exactes" tels que SAS-Nlmixed et Monolix. Lorsque nous avons adapté l'algorithme SAEM-MCMC à l'estimation des variances hétérogènes, la validation de l'algorithme fut plus délicate car la procédure Nlmixed de SAS et le logiciel Monolix ne traitent pas l'ensemble de ces structures de variances hétérogènes. Nos résultats ont alors été comparés à ceux obtenus par le logiciel d'estimation bayésienne WinBUGS dans un cadre non informatif pour se rapprocher le plus possible des estimations ML.

Bien que l'algorithme SAEM-MCMC soit performant pour l'estimation des paramètres dans les modèles hétérogènes étudiés, cet algorithme a tendance à fournir des estimateurs biaisés surtout pour les composantes de variance, comme toutes les méthodes de maximum de vraisemblance. Meza et al. (2007) ont adapté l'algorithme SAEM à l'estimation du maximum de vraisemblance résiduel (REML) dans les modèles non linéaires mixtes homogènes, et assurent que l'estimation REML est moins biaisée que l'estimation ML, y compris pour des jeux de données incomplètes de type MAR. Il serait donc intéressant d'implémenter REML-SAEM dans le cadre de l'estimation des variances hétérogènes. Dans la version homogène de l'algorithme REML-SAEM de Meza et al. (2007), le vecteur des effets fixes est considéré comme appartenant à l'ensemble des données manquantes et est donc traité comme une variable aléatoire gaussienne avec une variance tendant vers l'infini. Ce faisant, grâce à l'algorithme EM, les effets fixes sont automatiquement "éliminés" par intégration (principe de la vraisemblance marginalisée ou intégrée vis-à-vis des paramètres de nuisance, Berger, Liseo et Wolpert, 1999). Dans la version hétérogène de cet algorithme, il faudrait procéder de même (Foulley, 2004).
Pour pallier aux difficultés rencontrées dans la mise en oeuvre de SAEM-REML, l'estimation bayésienne avec des a priori non informatifs choisis correctement est une alternative possible du fait de l'interprétation bayésienne de REML (Harville, 1974).

Vis-à-vis de la phase de simulation proprement dite, Donnet et Samson (2008) proposent de ne pas simuler toutes les données manquantes en même temps comme dans l'algorithme de Metropolis-Hastings classique mais d'utiliser plutôt un algorithme de Gibbs, dans lequel on simule séparément les données manquantes issues de la moyenne et celles issues de la fonction de variance. Cette procédure fournit une chaîne de markov uniformement ergodique sous certaines conditions, mais elle peut demander beaucoup de temps de calcul car, à chaque étape de l'algorithme de Gibbs, on a recours à deux algorithmes de Metropolis-Hastings sous des lois invariantes complexes.

Au lieu de s'efforcer d'augmenter la vitesse de convergence de la méthode MCMC utilisée pour simuler les variables aléatoires au niveau de l'étape S de l'algorithme SAEM, il pour-

rait être intéressant d'appliquer l'algorithme PX-EM, soit dans une version basique comme celle adoptée par Lavielle et Meza (2007) soit dans une version plus élaborée (Foulley et Van Dyk, 2004) pour accélérer l'algorithme général du moins dans ses premières itérations. Cette procédure tend à augmenter la vitesse de convergence de l'algorithme et à éviter les maxima locaux de la vraisemblance. Rappelons que l'algorithme PX-EM (Liu et al., 1998) élargit l'espace paramétrique d'origine en introduisant un ou plusieurs paramètres auxiliaire au vecteur des paramètres d'origine à estimer. Dans sa version stochastique, la convergence de l'algorithme vers un maximum de vraisemblance local est assurée sous les mêmes conditions que l'algorithme SAEM, pourvu que la procédure d'accélération ne soit utilisée que lors des premières itérations (une dizaine suffit en pratique).

Bibliographie

Abramowitz, M., Stegun, I. A., 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables. New-York: Dover.

Aitkin, M., 1987. Modelling variance heterogeneity in normal regression using GLIM. Applied statistics, 36, 332-339.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. B. N. PETROV and F. CSAKI, eds. Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 267-281.

Ali, A. K. A., Schaeffer, L. R., 1987. Accounting for covariances among test day milk yields in dairy cows. Canadian Journal of Animal Science, 67, 637-644.

Ali, A. K. A., Shook, G. E., 1990. An optimum transformation for somatic cell concentration in milk. Journal of Dairy Science, 63, 487-490.

Al-Zaid, M., Yang, S. S., 2001. An approximate EM algorithm for nonlinear mixed effects models. Biometrical Journal, 7, 881-893.

Battese, G. E., Bonyhady, B. P., 1981. Estimation of household expenditure functions: an application of a class of heteroscedastic regression models. The Economic Record, 57, 80-85.

Beal, S. L., Sheiner, L. B., 1982. Estimating population kinetics. CRC Critical Reviews in Biomedical Engineering, 8, 195-222.

Beal, S. L., Sheiner, L. B., 1988. Heteroscedastic nonlinear regression. Technometrics, 30, 327-338.

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis. New York: Springer-Verlag.

Berger, J, Liseo, B, Wolpert, R. L, 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science, 14, 1-28.

Blasco, A., Piles, M., Varona, L., 2003. A bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. Genetics Selection Evolution, 35, 21-41.

Booth, G. J., Hobert P. J., 1999. Maximizing generalized linear mixed models likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society, Ser. B, 61,265-285.

Box, G. E. P., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society, Ser. B, 26, 211-252.

Box, G. E. P., Hill, W. J., 1974. Correcting inhomogeneity of variances with power transformation weighting. Technometrics, 16, 385-389.

Browne, W. J., Draper, D., Goldstein, H., Rasbash, J., 2002. Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. Computational Statistics and Data Analysis, 39, 203-225.

Celeux, G., Diebolt, J., 1985. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. Computational Statistics Quaterly, 2, 73-82.

Colleau, J. J., Le Bihan-Duval, E., 1995. A simulation study of selection methods to improve mastitis resistance of dairy cows. Journal of Dairy Science, 78, 659-671.

Concordet, D., Nunez, O. G., 2002. A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. Computational Statistics and Data Analysis, 39, 187-201.

Cook, R. D., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. Biometrika, 70, 1-10.

Davidian, M., Carroll, J., 1987. Variance Function Estimation. Journal of the American Statistical Association, 82, 1079-1091.

Davidian, M., Gallant, A. R., 1992. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. Journal of Pharmacokinetics and Biopharmaceutics, 20, 529-556.

Davidian, M., Giltinan, D. M., 1995. Non linear models for repeated measurement data, London: Chapman & Hall.

Davidian, M., Giltinan, D. M., 2003. Nonlinear models for repeated measurements: An overview and update. Editor's Invited paper, Journal of Agricultural, Biological, and

Environmental Statstics, 8, 387-419.

Davis, P. J., Rabinowitz, P., 1984. Methods of numerical integration. Second edition. New York: Academic Press.

De Boor, C., 1978. A practical guide to splines. Berlin: Springer.

Delyon, B., Lavielle, M., Moulines, E., 1999. Convergence of a stochastic approximation version of the EM algorithm. Annals of Statistics, 27, 94-128.

Demidenko, E., 1997. Asymptotic properties of nonlinear mixed-effects models. In Modelling longitudinal and spatially correlated data : methods, Applications, and future directions. Eds. T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. D. Wolfinger. New-York: Springer.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Ser. B, 39, 1-38.

Donnet, S., Samson, A., 2007. Estimation of parameters in incomplete data models defined by dynamical systems. Journal of Statistical Planning and Inference, 137, 2815-2831.

Donnet, S., Samson, A., 2008. Parametric inference for mixed models defined by stochastic differential equations. ESAIM: Probability and Statistics, 12, 196-218.

Duval, M., Robert-Granié, C., 2007. Criteria to calibrate the parameters of the SAEM-MCMC algorithm in maximum likelihood estimation for nonlinear mixed effects models. Soumis.

Duval, M., Robert-Granié, C., Foulley, J. L., 2008. Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm. Soumis.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of variance of united Kingdom inflations. Econometrica, 50, 987-1008.

Foulley, J. L., San Cristobal, M., Gianola, D., Im, S., 1992. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. Computational Statistics and Data Analysis, 13, 291-305.

Foulley, J. L., Quaas, R. L., 1995. Heterogeneous variances in Gaussian linear mixed models. Genetic Selection Evolution, 27, 211-228.

Foulley, J. L., 2004. Including mean-variance relationships in heteroskedastic mixed linear models: theory and applications. Interstat.

Foulley, J. L., Van Dyk , D. A., 2000. The PX EM algorithm for fast stable fitting of Henderson's mixed model. Genetic Selection Evolution, 32, 143-163.

Gelman, A., Bois, F., Jiang, L. M., 1996. Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distribustions. Journal of the American Statistical Association, 91, 1400-1412.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., 1993. A multilevel analysis of school examination results. Oxford Review. Education 19, 425-433.

Gourieroux, C., Montfort, A., 1989. Statistique et modèles économétriques. Collection "Economie et statistiques avancées" ENSAE. Vol. 2. France: Economica.

Harvey, A. C., 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica, 44, 461-465.

Harville, D. A., 1974. Bayesian inference for variance components using only error contrats. Biometrika, 61, 383-385.

Harville, D. A., 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. Annals of Statistics, 4, 384-395.

Hedeker, D., Mermelstein, R. J., 2007. Mixed-effects regression models with heterogeneous variance: analyzing ecological momentary assessment data of smoking. In T.D. Little, J.A. Bovaird, N.A. Card (Eds.), Modeling Contextual Effects in Longitudinal Studies. Erlbaum: Mahwah, NJ.

Henderson, C. R., 1953. Estimation of variance and covariance components. Biometrics, 9, 226-252.

Henderson, C. R. O., Kempthorne, O., Searle, S. R., Von Krosigk, C. N., 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics, 15, 192-218.

Henderson, C. R., 1973. Sire evaluation and genetic trends. In Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. J. L. Lush. American Society of Animal Science, 10-41.

Huet, S., Jolivet, E., Messéan, A., 1992. La régression non-linéaire (méthodes et applications en biologie). Editions INRA, France.

Judge, G. G., Griffiths, W. E., Carter, Hill R., Lutkepohl, H., Lee, T. C., 1985. The

theory and practice of econometrics. New-York: J Wiley & Sons.

Kuhn, E., Lavielle, M., 2004. Coupling a stochastic approximation version of EM with a MCMC procedure. ESAIM Probability and Statistics, 8, 115-131.

Kuhn, E., Lavielle, M., 2005. Maximum Likelihood estimation in nonlinear mixed effects models. Computational Statistics and Data Analysis, 49, 1020-1038.

Laird, N. M., Ware, J. H., 1982. Random-effects models for longitudinal data. Biometrics, 38, 963-974.

Lavielle, M., Meza, C., 2007. A Parameter Expansion version of the SAEM algorithm. Statistics and Computing, 17, 121-130.

Lee, Y., Nelder, J.A., 2006. Double hierarchical generalized linear models (with discussion). Journal of the Royal Statistical Society, Ser. C, 55, 139-180.

Lescourret, F., Coulon, J. B., 1994. Modeling the impact of mastitis on milk production by dairy cows. Journal of Dairy Science, 77, 2289-2301.

Lin, X., Ray, J., Harlow, S. D., 1997. Linear mixed models with heterogeneous withincluster variances. Biometrics, 53, 910-923.

Lindstrom, M. J., Bates, D. M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics, 46, 673-687.

Littell, R., Milliken, G., Stroup, W., Wolfinger, R., Schabenberger, O., 2006. SAS for Mixed Models. New York: SAS Institute.

Liu, C., Rubin, D., Wu, Y., 1998. Parameter expansion to accelerate EM: the PX-EM. Biometrika, 85, 755-770.

Lu, J-C, Chen, D., Zhou, W., 2006. Quasi-likelihood estimation for GLM with random scales. Journal of Statistical Planning and Inference, 136, 401-429.

McCulloch, C. E., 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association, 92, 162-169.

Meza, C., Jaffrézic, F., Foulley, J. L., 2007. REML Estimation of Variance Parameters in Nonlinear Mixed Effects Models Using the SAEM Algorithm. Biometrical Journal, 49, 876-888.

Minvielle, F., 1998. La sélection animale. Collection "Que sais-je ?", France.

Morant, S. V., Gnanasakthy, A., 1989. A new approach to the mathematical formulation of lactation curves. Animal Production, 49, 151-162.

Mrode, R. A., Swanson, G. J. T., 1996. Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle. Animal Breeding Abstract, 64, 847-857.

Pinheiro, J. C., Bates, D. M., 1995. Approximations to the log-likelihood function in the nonlinear mixed effects model. Journal of Computational and Graphical Statistics, 4, 12-35.

Pinheiro, J. C., Bates, D. M., 2000. Mixed-effects in S and S-Plus. New-York: Springer.

Racine-Poon, A., 1985. A bayesian approach to non-linear random effects models. Biometrics, 41, 1015-1024.

Ramos, R. Q., Pantula, 1995. Estimation of nonlinear random coefficient models. Statistics and Probability Letters, 24, 49-56.

Robbins, H., Monroe, S., 1951. A stochastic Approximation Method. Annals of Mathematical Statistics, 22, 400-407.

Robert, C., 1992. L'analyse statistique bayésienne. Paris : Economica.

Robert, C. P., Casella, G. 2004. Monte Carlo Statistical methods. New York: Springer.

Robert-Granié, C., Bonaïti, B., Boichard, D., and Barbat, A., 1999. Accounting for variance heterogeneity in french dairy cattle genetic evaluation. Livestock Production Science, 60, 343-357.

Robert-Granié, C., Foulley, J. L., Meza, E., Rupp, R., 2004. Statistical analysis of somatic cell scores via mixed model methodology for longitudinal data. Animal Research, 53, 259-273.

Robinson, G. K., 1991. That BLUP Is a Good Thing: The Estimation of Random Effects. Statistical Science, 6, 15-51.

Rodriguez-Zas, S. L., Gianola, D., Shook, G. E., 2000. Evaluation of models for somatic cell score lactation patterns in Holsteins. Livestock Production Science, 67, 19-30.

Rook, A. J., France, J., Dhanoa, M. S., 1993. On the mathematical description of lactation curves. The journal of Agricultural Science, 121, 97-102.

Royston, P., Altman, D. G., 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Applied Statistics, 43, 429-467.

Rupp, R., Boichard, D., 1997. Evaluation génétique des bovins laitiers sur les comptages de cellules somatiques pour l'amélioration de la résistance aux mammites. Rencontres Recherches Ruminants, 4, 211-214.

Rupp, R., Boichard, D., 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holstein. Journal of Dairy Science, 82, 2198-2204.

Rupp, R., 2000. Analyse génétique de la résistance aux mammites chez les ruminants laitiers. Thèse de doctorat. Institut National Agronomique Paris-Grignon, France.

Ruppert, D., Wand, M. P., Carroll, R. J., 2003. Semiparametric Regression. New York: Cambridge University Press.

San Cristobal, M., Robert-Granié, C., Foulley, J. L., 2002. Hétéroscédasticité et modèles linéaires mixtes : théorie et applications en génétique quantitative. Journal de la Société Française de Statistique, 143, 1-2.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics, 6, 461-646.

Searle, S. R., Casella, G., McCulloch, C. E., 1992. Variance components. New York: Wiley and Sons.

Seegers, J., Menard, J. L., Dejean, O., Weber, M., 1997. Cell count evolution and clinical mastitis frequency in milk recording herds of the OPTILAIT area (South west). Rencontres Recherches Ruminants, 4, 279.

Self, S. G., Liang, K. Y., 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non standard conditions. Journal of American Statistical Association, 82, 605-610.

Sheiner, L. B., Rosenberg, B., Melmon, K. L., 1972. Modeling of individual pharmacokinetics for computer-aided drug dosing. Computational Biomedical Research, 5, 441-459.

Sheiner, L. B., Beal, S. L., 1980. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. Journal of pharmacokinetics and biopharmaceutics, 553-570.

Smith, A. F. M., Roberts, G. O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society, Ser. B, 55, 3-23.

Tierney, L., Kadane, J., 1986. Accurate approximations for posterior moments and marginal densities. Journal of American Statistical Association, 81, 82-86.

Torres, R. A. A, 2001. MCMC methods for estimating the covariance structure of longitudinal data - An application to dairy cattle data. PhD thesis, Cornell University.

Verbeke, G., Molenberghs, G., 2000. Linear mixed models for longitudinal data. New-York: Springer.

Vonesh, E. F., 1992. Non-linear models for the analysis of longitudinal data. Statistics in Medecine, 11, 1929-1954.

Vonesh, E. G., Wang, H., Nie, L., Majumdar, D., 2002. Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. Journal of the American Statistical Association, 97, 271-283.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., Gelfand, A. E., 1994. Bayesian analysis of linear and non-linear population models by using the Gibbs Sampler. Applied Statistics, 43, 201-221.

Walker, S., 1996. An EM algorithm for nonlinear random effects models. Biometrics, 52, 934-944.

Wang, J., 2007. EM algorithms for nonlinear mixed effects models. Computational Statistics and Data Analysis, 51, 3244-3256.

Wei, G. C. G., Tanner, M.A., 1990. A Monte Carlo implementation of the EM algorithm and the Poor's Man's data augmentation algorithms. Journal of the American Statistical Association, 85, 699-704.

Wiggans, G. R., Shook, G. E., 1987. A lactation measure of somatic cell count. Journal of Dairy Science, 70, 2666-2675.

Wolfinger, R., 1993. Laplace's approximation for nonlinear mixed models. Biometrika, 80, 791-795.

Wolfinger, R. D., 1996. Heterogeneous variance-covariance structure for repeated measures. Journal of Agricultural, Biological, and Environmental Statistics, 1, 205-230.

Wood, P. D. P., 1967. Algebraic model of the lactation curve in cattle. Nature, 216, 164-165.

Wu, C. F. J., 1983. On the convergence properties of the EM algorithm. Annals of Statistics, 11, 95-103.

Annexes

Dans cette annexe, nous proposons quelques compléments à la section 3.4.2 du chapitre 3, afin de permettre à tout chercheur de reproduire les résultats obtenus. Nous présentons le jeu de données concernant une expérience sur la sélection de poulets, menée par F. Ricard (INRA, Station de Nouzilly), ainsi que les codes des méthodes utilisées dans l'étude du modèle de variance hétérogène V2. Cet exemple a été choisi pour permettre une présentation la plus large possible des logiciels. En particulier, nous n'avons pas pu considérer des modèles potentiellement mieux adaptés car ils ne pouvaient pas être traités par certains logiciels (cas d'un modèle de variance avec effets aléatoires) ou ne convergeaient pas (cas du modèle où tous les paramètres individuels de croissance sont aléatoires).
Un tableau complètera les résultats obtenus dans le tableau 3.5. Les temps de calculs entre les méthodes n'ont pas pu être comparés car les logiciels ont été utilisés sur des machines différentes, et certains logiciels ne comportent pas de critères d'arrêts standards (par exemple : Winbugs).

Voici le plan de l'annexe :
1) Jeu de données chez le poulet,
2) Descriptif de notre algorithme SAEM-MCMC dans le cas général d'un modèle de variance à effets fixes,
3) Calcul des dérivées premières, secondes et croisées pour l'étude de la fonction de variance V2,
4) Code de l'algorithme,
5) Code et sortie de la procédure nlme de R,
6) Code et sortie de la procédure Nlmixed de SAS,
7) Code et sortie du logiciel WinBUGS,
8) Tableau complet des résultats.

Pour une analyse comparative des résultats, on se référera à l'article de Duval et al. (2008).

1) Jeu de données chez le poulet (INRA, F. Ricard)

Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids
1	1	0	44	5	1	16	1460	9	1	36	2320
1	1	4	205	5	1	20	1950	9	1	40	2750
1	1	6	355	5	1	24	2210	10	1	0	36
1	1	8	640	5	1	28	2530	10	1	4	240
1	1	12	1020	5	1	32	2760	10	1	6	390
1	1	16	1600	5	1	36	2880	10	1	8	560
1	1	20	2310	5	1	40	2980	10	1	12	1060
1	1	24	2880	6	1	0	33	10	1	16	1740
1	1	28	3650	6	1	4	175	10	1	20	2240
1	1	32	3800	6	1	6	330	10	1	24	2680
1	1	36	3520	6	1	8	570	10	1	28	3150
1	1	40	3240	6	1	12	1000	10	1	32	2920
2	1	0	46	6	1	16	1600	10	1	36	2880
2	1	4	180	6	1	20	2050	10	1	40	2830
2	1	6	300	6	1	24	2380	11	2	0	41
2	1	8	520	6	1	28	2830	11	2	4	375
2	1	12	980	6	1	32	2530	11	2	6	660
2	1	16	1580	6	1	36	2520	11	2	8	980
2	1	20	2190	6	1	40	2500	11	2	12	1380
2	1	24	2620	7	1	0	38	11	2	16	1520
2	1	28	3240	7	1	4	225	11	2	20	1730
2	1	32	3180	7	1	6	380	11	2	24	1980
2	1	36	3090	7	1	8	620	11	2	28	2100
2	1	40	2970	7	1	12	1210	11	2	32	1900
3	1	0	34	7	1	16	1800	11	2	36	1850
3	1	4	230	7	1	20	2070	11	2	40	1790
3	1	6	370	7	1	24	2430	12	2	0	39
3	1	8	600	7	1	28	3000	12	2	4	390
3	1	12	1040	7	1	32	2640	12	2	6	650
3	1	16	1640	7	1	36	2700	12	2	8	960
3	1	20	2150	7	1	40	2750	12	2	12	1420
3	1	24	2550	8	1	0	38	12	2	16	1590
3	1	28	3100	8	1	4	170	12	2	20	1750
3	1	32	3150	8	1	6	270	12	2	24	2080
3	1	36	3040	8	1	8	430	12	2	28	2060
3	1	40	2920	8	1	12	935	12	2	32	2120
4	1	0	38	8	1	16	1440	12	2	36	2050
4	1	4	200	8	1	20	2050	12	2	40	1990
4	1	6	330	8	1	24	2360	13	2	0	33
4	1	8	530	8	1	28	2820	13	2	4	370
4	1	12	830	8	1	32	2820	13	2	6	620
4	1	16	1240	8	1	36	2820	13	2	8	960
4	1	20	1760	8	1	40	2810	13	2	12	1370
4	1	24	1990	9	1	0	35	13	2	16	1600
4	1	28	2360	9	1	4	180	13	2	20	1740
4	1	32	2680	9	1	6	280	13	2	24	1890
4	1	36	2470	9	1	8	500	13	2	28	2000
4	1	40	2700	9	1	12	970	13	2	32	1900
5	1	0	38	9	1	16	1600	13	2	36	1950
5	1	4	265	9	1	20	2090	13	2	40	1880
5	1	6	430	9	1	24	2470	14	2	0	32
5	1	8	600	9	1	28	2740	14	2	4	330
5	1	12	1050	9	1	32	2420	14	2	6	550

Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids
14	2	8	850	18	2	36	2120	23	3	16	2460
14	2	12	1290	18	2	40	2130	23	3	20	2790
14	2	16	1490	19	2	0	37	23	3	24	3130
14	2	20	1640	19	2	4	390	23	3	28	2850
14	2	24	1880	19	2	6	655	23	3	32	2900
14	2	28	2040	19	2	8	920	23	3	36	2900
14	2	32	1960	19	2	12	1310	23	3	40	2950
14	2	36	1930	19	2	16	1450	24	3	0	38
14	2	40	1900	19	2	20	1590	24	3	4	370
15	2	0	30	19	2	24	1790	24	3	6	675
15	2	4	235	19	2	28	1930	24	3	8	1060
15	2	6	420	19	2	32	1600	24	3	12	2280
15	2	8	660	19	2	36	1700	24	3	16	2650
15	2	12	1160	19	2	40	1800	24	3	20	3100
15	2	16	1460	20	2	0	36	24	3	24	3470
15	2	20	1630	20	2	4	310	24	3	28	3360
15	2	24	2000	20	2	6	530	24	3	32	3470
15	2	28	2040	20	2	8	840	24	3	36	3470
15	2	32	1810	20	2	12	1300	24	3	40	3380
15	2	36	1790	20	2	16	1490	25	3	0	39
15	2	40	1770	20	2	20	1680	25	3	4	385
16	2	0	35	20	2	24	1940	25	3	6	670
16	2	4	275	20	2	28	1930	25	3	8	1000
16	2	6	510	20	2	32	1950	25	3	12	2200
16	2	8	800	20	2	36	1990	25	3	16	2680
16	2	12	1130	20	2	40	1950	25	3	20	3120
16	2	16	1270	21	3	0	43	25	3	24	3520
16	2	20	1450	21	3	4	380	25	3	28	3390
16	2	24	1730	21	3	6	685	25	3	32	3600
16	2	28	1750	21	3	8	1050	25	3	36	3600
16	2	32	1700	21	3	12	2060	25	3	40	3780
16	2	36	1690	21	3	16	2450	26	3	0	42
16	2	40	1740	21	3	20	2890	26	3	4	350
17	2	0	35	21	3	24	3130	26	3	6	685
17	2	4	370	21	3	28	2980	26	3	8	1100
17	2	6	635	21	3	32	2840	26	3	12	2160
17	2	8	920	21	3	36	2840	26	3	16	2460
17	2	12	1390	21	3	40	2690	26	3	20	2850
17	2	16	1520	22	3	0	42	26	3	24	3350
17	2	20	1650	22	3	4	440	26	3	28	3290
17	2	24	1760	22	3	6	675	26	3	32	3230
17	2	28	1940	22	3	8	920	26	3	36	3230
17	2	32	1790	22	3	12	1410	26	3	40	3130
17	2	36	1810	22	3	16	1760	27	3	0	42
17	2	40	1820	22	3	20	2020	27	3	4	385
18	2	0	37	22	3	24	2250	27	3	6	635
18	2	4	305	22	3	28	2450	27	3	8	940
18	2	6	630	22	3	32	2520	27	3	12	1470
18	2	8	880	22	3	36	2460	27	3	16	1850
18	2	12	1360	22	3	40	2790	27	3	20	2150
18	2	16	1540	23	3	0	34	27	3	24	2580
18	2	20	1840	23	3	4	415	27	3	28	3050
18	2	24	2040	23	3	6	690	27	3	32	2880
18	2	28	2290	23	3	8	1060	27	3	36	2850
18	2	32	2120	23	3	12	2140	27	3	40	2690

Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids	Individu	Lignée	Age	Poids
28	3	0	42	32	4	24	1810	37	4	6	345
28	3	4	385	32	4	28	1910	37	4	8	510
28	3	6	635	32	4	32	1830	37	4	12	920
28	3	8	940	32	4	36	1800	37	4	16	1380
28	3	12	1470	32	4	40	1910	37	4	20	1650
28	3	16	1850	33	4	0	36	37	4	24	1820
28	3	20	2150	33	4	4	230	37	4	28	1840
28	3	24	2580	33	4	6	375	37	4	32	1750
28	3	28	3050	33	4	8	570	37	4	36	1700
28	3	32	2880	33	4	12	950	37	4	40	1800
28	3	36	2850	33	4	16	1230	38	4	0	32
28	3	40	2690	33	4	20	1450	38	4	4	230
29	3	0	34	33	4	24	1670	38	4	6	400
29	3	4	290	33	4	28	1760	38	4	8	560
29	3	6	520	33	4	32	1960	38	4	12	920
29	3	8	820	33	4	36	1680	38	4	16	1220
29	3	12	1380	33	4	40	1690	38	4	20	1460
29	3	16	1820	34	4	0	36	38	4	24	1510
29	3	20	2330	34	4	4	215	38	4	28	1520
29	3	24	2700	34	4	6	340	38	4	32	1550
29	3	28	2990	34	4	8	540	38	4	36	1540
29	3	32	2880	34	4	12	860	38	4	40	1580
29	3	36	3030	34	4	16	1190	39	4	0	33
29	3	40	2960	34	4	20	1380	39	4	4	180
30	3	0	39	34	4	24	1610	39	4	6	265
30	3	4	290	34	4	28	1610	39	4	8	380
30	3	6	545	34	4	32	1660	39	4	12	690
30	3	8	830	34	4	36	1580	39	4	16	1030
30	3	12	1420	34	4	40	1630	39	4	20	1310
30	3	16	2010	35	4	0	35	39	4	24	1570
30	3	20	2330	35	4	4	195	39	4	28	1750
30	3	24	2710	35	4	6	320	39	4	32	1670
30	3	28	3020	35	4	8	450	39	4	36	1620
30	3	32	2850	35	4	12	730	39	4	40	1560
30	3	36	3000	35	4	16	1170	40	4	0	33
30	3	40	3100	35	4	20	1470	40	4	4	225
31	4	0	33	35	4	24	1780	40	4	6	335
31	4	4	210	35	4	28	1660	40	4	8	510
31	4	6	305	35	4	32	1630	40	4	12	850
31	4	8	460	35	4	36	1550	40	4	16	1180
31	4	12	850	35	4	40	1460	40	4	20	1470
31	4	16	1190	36	4	0	31	40	4	24	1650
31	4	20	1550	36	4	4	200	40	4	28	1900
31	4	24	1780	36	4	6	325	40	4	32	1700
31	4	28	1800	36	4	8	500	40	4	36	1840
31	4	32	1880	36	4	12	870	40	4	40	1900
31	4	36	1910	36	4	16	1170	41	5	0	40
31	4	40	1980	36	4	20	1520	41	5	4	355
32	4	0	37	36	4	24	1680	41	5	6	575
32	4	4	175	36	4	28	1770	41	5	8	860
32	4	6	325	36	4	32	1630	41	5	12	1340
32	4	8	480	36	4	36	1650	41	5	16	1700
32	4	12	800	36	4	40	1770	41	5	20	2010
32	4	16	1230	37	4	0	31	41	5	24	2140
32	4	20	1540	37	4	4	220	41	5	28	2350

Individu	Lignée	Age	Poids
41	5	32	2150
41	5	36	2120
41	5	40	2200
42	5	0	39
42	5	4	330
42	5	6	530
42	5	8	800
42	5	12	1230
42	5	16	1610
42	5	20	1840
42	5	24	2130
42	5	28	2210
42	5	32	2160
42	5	36	2280
42	5	40	2310
43	5	0	33
43	5	4	300
43	5	6	515
43	5	8	820
43	5	12	1330
43	5	16	1670
43	5	20	1860
43	5	24	2060
43	5	28	2200
43	5	32	2260
43	5	36	2320
43	5	40	2310
44	5	0	31
44	5	4	250
44	5	6	480
44	5	8	730
44	5	12	1140
44	5	16	1580
44	5	20	1820
44	5	24	2100
44	5	28	2160
44	5	32	2180
44	5	36	2120
44	5	40	2100
45	5	0	30
45	5	4	305
45	5	6	510
45	5	8	740
45	5	12	1100
45	5	16	1480
45	5	20	1750
45	5	24	2050
45	5	28	2090
45	5	32	1870
45	5	36	1920
45	5	40	2020
46	5	0	34
46	5	4	260
46	5	6	445
46	5	8	680

Individu	Lignée	Age	Poids
46	5	12	1110
46	5	16	1520
46	5	20	1740
46	5	24	2050
46	5	28	2180
46	5	32	2020
46	5	36	2100
46	5	40	2210
47	5	0	40
47	5	4	300
47	5	6	415
47	5	8	660
47	5	12	1110
47	5	16	1490
47	5	20	1690
47	5	24	1900
47	5	28	2280
47	5	32	2240
47	5	36	2340
47	5	40	2230
48	5	0	40
48	5	4	315
48	5	6	600
48	5	8	780
48	5	12	1520
48	5	16	2030
48	5	20	2380
48	5	24	2680
48	5	28	2930
48	5	32	2640
48	5	36	2500
48	5	40	2580
49	5	0	38
49	5	4	285
49	5	6	520
49	5	8	780
49	5	12	1380
49	5	16	1770
49	5	20	2070
49	5	24	2380
49	5	28	2640
49	5	32	2750
49	5	36	2970
49	5	40	2900
50	5	0	36
50	5	4	280
50	5	6	535
50	5	8	800
50	5	12	1120
50	5	16	1760
50	5	20	2080
50	5	24	2420
50	5	28	2580
50	5	32	2460
50	5	36	2490

Individu	Lignée	Age	Poids
50	5	40	2520

2) Algorithme SAEM-MCMC : descriptif du programme

Pour les notations utilisées dans cette note, nous vous renvoyons à la section 3.3 et à l'appendice 3.7.
Dans ce descriptif, on suppose que le modèle de variance ne comporte pas d'effet aléatoire.

Procédure MCMC :
Dans cette procédure, il s'agit de construire des chaines de Markov via l'algorithme de Metropolis-Hastings.
On construit *L* chaînes de longueur *T***.** Pour chaque chaîne, jusqu'à l'itération t =m_1, on génère des variables aléatoires $\mathbf{W}_t$ sous la densité de **u**, puis à partir de l'itération m_1, les variables $\mathbf{W}_t$ sont générées sous une loi Gaussienne centrée sur la valeur de l'itération précédente.
A chaque itération, la chaîne se déplace vers la nouvelle valeur $\mathbf{W}_t$ avec probabilité ρ.

subroutine MCMC utilisée à l'étape (k) du PROGRAM Modele
Cette sous-routine renvoie les $\mathbf{u}^{(k,l)}$, l=1,…,L

Tant que $l \leq L$ **faire**

Tant que $t \leq T$ **faire**

On simule $\mathbf{W}_t$ sous la densité instrumentale suivante :
Si $t < m_1$, la distribution instrumentale utilisée est la loi a priori de **u**,
Sinon, la distribution instrumentale est une loi Gaussienne centrée sur l'itération précédente $\mathbf{u}^{(k,l,t-1)}$ et de variance $\rho V^{(k-1)}$, où ρ est issu d'une loi uniforme sur l'intervalle $[\rho_1;\rho_2]$ et $\mathbf{V^{(k-1)}}$ est l'estimation de la variance de **u** à l'étape (k-1).

On calcule ensuite la probabilité d'acceptation $\rho\left(u^{(k,l,t-1)},W_t\right)$, puis on génère une loi uniforme X sur [0 ;1].
Si $\rho\left(u^{(k,l,t-1)},W_t\right) > X$, alors la chaîne fait un saut sur la nouvelle valeur : $\mathbf{u}^{(k,l,t)} = \mathbf{W}_t$
Sinon la chaîne ne bouge pas : $\mathbf{u}^{(k,l,t)} = \mathbf{u}^{(k,l,t-1)}$

Stop

La valeur de la dernière itération $\mathbf{u}^{(k,l,T)}$ de la chaîne est stockée dans $\mathbf{u}^{(k,l)}$
Stop

Remarque : dans le code, on ne garde pas seulement la dernière itération de chaque chaîne mais les (T/2) dernières itérations.

PROGRAM modele
Définition des variables
Initialisation des paramètres
Lecture du fichier de données

Tant que k ne vérifie pas le critère d'arrêt, **faire**

Calculer γ_k (paramètre de lissage)
Générer les $\mathbf{u}^{(k,l)}$, l=1,…,L, sous la distribution conditionnelle de **u** sachant les observations et la valeur courante du paramètre, via la procédure MCMC.
Approximer $\mathbf{S}_{1i}^{(k)}$ comme l'espérance conditionnelle de **u** sachant **y** et $\boldsymbol{\theta}^{(k)}$, et $\mathbf{S}_{2i}^{(k)}$ comme l'espérance conditionnelle de **uu**' sachant **y** et $\boldsymbol{\theta}^{(k)}$, en utilisant les approximations de la section 3.3.
On calcule ensuite les dérivées première et seconde de la log-vraisemblance complète par rapport au vecteur η, que l'on stocke dans les matrices D_k et B_k.
Puis l'approximation de $\eta^{(k)}$ devient

$$\eta^{(k)} = \eta^{(k-1)} - \gamma_k (M_k{}^{-1} D_k)$$

Ce qui nous permet d'estimer les effets fixes du modèle et de la variance.

Le vecteur moyenne des effets aléatoires $\boldsymbol{\mu}^{(k)}$ est estimé par :

$$\mu^{(k)} = \left(\sum_{i=1}^{I} A_i' \left(\Gamma^{(k-1)} \right)^{-1} A_i \right)^{-1} \sum_{i=1}^{I} A_i' \left(\Gamma^{(k-1)} \right)^{-1} S_{1,i}{}^{(k)}$$

La matrice de covariance **Γ**, est estimée par :

$$\Gamma^{(k)} = \frac{1}{I} \sum_{i}^{I} S_{2,i}{}^{(k)} - A_i \mu^{(k)} S_{1,i}{}^{(k)\prime} - S_{1,i}{}^{(k)} \mu^{(k)\prime} A_i' + A_i \mu^{(k)} \mu^{(k)\prime} A_i'$$

Les dérivées premières, secondes, et croisées de la log-vraisemblance par rapport à chacun des paramètres sont calculées pour estimer la matrice des erreurs standards. Dans chacune de ces dérivées, le vecteur **u** est remplacé par $\mathbf{u}^{(k,l)}$, l=1,…,L, puis une valeur moyenne (sur L) est donnée. L'étape d'approximation est ensuite effectuée sur toutes les dérivées.

La valeur de la vraisemblance est calculée en la nouvelle valeur des paramètres.

Le critère d'arrêt est testé.

Stop
La matrice des erreurs standard est donnée.
La valeur finale des paramètres et la valeur de la vraisemblance des observations sont données.

3) Calcul des dérivées premières, secondes et croisées dans ce cas particulier

$$y_{hij} = A_{hi} \exp\big(- B_{hi} \exp(-C_{hi} T_j)\big) + \sigma_{hij} \varepsilon^*_{hij},$$

$\forall i \in \{1, ..., I\}$ $\forall j \in \{1, .., n_i\}$, avec $I = 50$, $n_i = 12$,
h dans $\{lignée1, lignée2, lignée3, lignée4, lignée5\}$.

- $T_j = \frac{t_j}{100}$
- $\varepsilon^*_{ij} \sim_{iid} \mathcal{N}(0, 1)$
- $A_{hi} \sim_{iid} \mathcal{N}(a_h, \tau_a^2)$
- $B_{hi} = \beta_h$ non aléatoire
- $C_{hi} \sim_{iid} \mathcal{N}(c_h, \tau_c^2)$
- $\tau_{ac} = cov(A_{hi}, C_{hi})$
- $\log(\sigma^2_{hij}) = \delta_{1h} + \delta_{2h} t^\star_{ij} + \delta_{3h} {t^\star_{ij}}^2$, où $t^\star_{ij} = (t_{ij} - 20)/100$

Paramètres à estimer :

$\boldsymbol{\theta} = ((a_h, \beta_h, c_h, \delta_{1h}, \delta_{2h}, \delta_{3h})_{h=lignée\ 1,..,5}, \tau_a^2, \tau_c^2, \tau_{ac})$

Les données manquantes utilisées sont : $\boldsymbol{\phi} = (A, C)$.
On note $Ntot = I \times n_i = 600$.

Vraisemblance complète

On a $L = \log p(\mathbf{y}, \boldsymbol{\phi}; \boldsymbol{\theta}) = cste - \sum_{h,i,j} \frac{1}{2} \log(\sigma^2_{hij}) - \frac{I}{2} \log(\tau_a^2 \tau_c^2 - \tau_{ac}^2) - \frac{1}{2} \sum_{h,i,j} \frac{1}{\sigma^2_{hij}} (y_{hij} - f_{hij})^2 - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \Gamma^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$
où $f_{hij} = A_{hi} \exp\big(- B_{hi} \exp(-C_{hi} T_j)\big)$ et $\mathbf{\Gamma} = \begin{pmatrix} \tau_a^2 & \tau_{ac} \\ \tau_{ac} & \tau_c^2 \end{pmatrix}$

Dérivées premières

On note $det = \tau_a^2 \tau_c^2 - \tau_{ac}^2$, $mata = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $matc = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $matac = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$\frac{\partial L}{\partial \beta_l} = \sum_{l,i,j} \frac{1}{\sigma^2_{lij}} \big(y_{lij} - f_{lij}\big) \frac{\partial f_{lij}}{\partial \beta_l}$ avec $\frac{\partial f_{lij}}{\partial \beta_l} = -f_{lij} \exp(-C_{li} T_{ij})$

$\frac{\partial L}{\partial \tau_a^2} = -\frac{I}{2} \frac{\tau_c^2}{\tau_a^2 \tau_c^2 - \tau_{ac}^2} + \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial L}{\partial \tau_c^2} = -\frac{I}{2} \frac{\tau_a^2}{\tau_a^2 \tau_c^2 - \tau_{ac}^2} + \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial L}{\partial \tau_{ac}} = -\frac{I}{2} \frac{-2\tau_{ac}}{\tau_a^2 \tau_c^2 - \tau_{ac}^2} + \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial L}{\partial (a_l, c_l)} = \sum_{i=1, i \in l}^{I} \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - (a_l, c_l))$

$\frac{\partial L}{\partial \sigma^2_{lij}} = -\frac{1}{2\sigma^2_{lij}} + \frac{1}{2(\sigma^2_{lij})^2} \big(y_{lij} - f_{lij}\big)^2$

$\frac{\partial L}{\partial \delta_{1l}} = \sum_{ij, i \in l} \frac{\partial L}{\partial \sigma^2_{lij}} \frac{\partial \sigma^2_{lij}}{\partial \log(\sigma^2_{lij})} \frac{\partial \log(\sigma^2_{lij})}{\partial \delta_{1l}}$

$\frac{\partial L}{\partial \delta_{1l}} = \sum_{ij, i \in l} \frac{\partial L}{\partial \sigma^2_{lij}} \sigma^2_{lij}$

$\frac{\partial L}{\partial \delta_{2l}} = \sum_{ij, i \in l} \frac{\partial L}{\partial \sigma^2_{lij}} \sigma^2_{lij} t^{\star}_{ij}$

$\frac{\partial L}{\partial \delta_{3l}} = \sum_{ij, i \in l} \frac{\partial L}{\partial \sigma^2_{lij}} \sigma^2_{lij} t^{2\star}_{ij}$

Dérivées secondes et croisées

$\frac{\partial^2 L}{\partial (a_l, c_l)^2} = -\sum_{i=1, i \in l}^{I} \boldsymbol{\Gamma}^{-1}$

$\frac{\partial^2 L}{\partial (a_l, c_l) \partial \tau_a^2} = -\sum_{i=1, i \in l}^{I} \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - (a_l, c_l))$

$\frac{\partial^2 L}{\partial (a_l, c_l) \partial \tau_c^2} = -\sum_{i=1, i \in l}^{I} \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - (a_l, c_l))$

$\frac{\partial^2 L}{\partial (a_l, c_l) \partial \tau_{ac}} = -\sum_{i=1, i \in l}^{I} \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} (\boldsymbol{\phi}_i - (a_l, c_l))$

$\frac{\partial^2 L}{\partial \tau_a^2 \partial \tau_a^2} = \frac{I}{2} \frac{(\tau_c^2)^2}{(\tau_a^2 \tau_c^2 - \tau_{ac}^2)^2} - \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial^2 L}{\partial \tau_a^2 \partial \tau_c^2} = -\frac{I}{2} \frac{\tau_a \tau_c - \tau_{ac}^2 - \tau_a^2 \tau_c^2}{(\tau_a^2 \tau_c^2 - \tau_{ac}^2)^2} - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big) - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial^2 L}{\partial \tau_a^2 \partial \tau_{ac}} = \frac{I}{2} \frac{-2\tau_c^2 \tau_{ac}}{(\tau_a^2 \tau_c^2 - \tau_{ac}^2)^2} - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big) - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} \, mata \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial^2 L}{\partial \tau_c^2 \partial \tau_c^2} = \frac{I}{2} \frac{(\tau_a^2)^2}{(\tau_a^2 \tau_c^2 - \tau_{ac}^2)^2} - \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$\frac{\partial^2 L}{\partial \tau_c^2 \partial \tau_{ac}} = \frac{I}{2} \frac{-2\tau_a^2 \tau_{ac}}{(\tau_a^2 \tau_c^2 - \tau_{ac}^2)^2} - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big) - \frac{1}{2} \sum_{i=1}^{I} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)' \boldsymbol{\Gamma}^{-1} \, matac \, \boldsymbol{\Gamma}^{-1} \, matc \, \boldsymbol{\Gamma}^{-1} \big(\boldsymbol{\phi}_i - (a_i, c_i)\big)$

$$\frac{\partial^2 L}{\partial \tau_{ac}\partial \tau_{ac}} = \frac{I}{2}\frac{2\tau_a\tau_c+2\tau_{ac}^2}{(\tau_a^2\tau_c^2-\tau_{ac}^2)^2} - \sum_{i=1}^{I}\big(\boldsymbol{\phi}_i-(a_i,c_i)\big)'\boldsymbol{\Gamma}^{-1}\,matac\,\boldsymbol{\Gamma}^{-1}\,matac\,\boldsymbol{\Gamma}^{-1}\big(\boldsymbol{\phi}_i-(a_i,c_i)\big)$$

$$\frac{\partial^2 L}{\partial \beta_l \partial \beta_l} = \sum_{ij,i\in l}\frac{1}{\sigma_{lij}^2}(y_{lij}\frac{\partial^2 f_{lij}}{\partial \beta_l} - \big(\frac{\partial f_{lij}}{\partial \beta_l}\big)^2 - f_{lij}\frac{\partial^2 f_{lij}}{\partial \beta_l^2})$$

$$\frac{\partial^2 L}{\partial \beta_l \partial \delta_{1l}} = -\sum_{ij,i\in l}\frac{1}{\sigma_{hij}^2}\big(y_{lij}-f_{lij}\big)\frac{\partial f_{lij}}{\partial \beta_l}$$

$$\frac{\partial^2 L}{\partial \beta_l \partial \delta_{2l}} = -\sum_{ij,i\in l}\frac{t_{ij}^*}{\sigma_{hij}^2}\big(y_{lij}-f_{lij}\big)\frac{\partial f_{lij}}{\partial \beta_l}$$

$$\frac{\partial^2 L}{\partial \beta_l \partial \delta_{3l}} = -\sum_{ij,i\in l}\frac{t_{ij}^{*\,2}}{\sigma_{hij}^2}\big(y_{lij}-f_{lij}\big)\frac{\partial f_{lij}}{\partial \beta_l}$$

$$\frac{\partial^2 L}{\partial \delta_{1l} \partial \delta_{1l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{1}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

$$\frac{\partial^2 L}{\partial \delta_{2l} \partial \delta_{2l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{t_{ij}^{*\,2}}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

$$\frac{\partial^2 L}{\partial \delta_{3l} \partial \delta_{3l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{t_{ij}^{*\,4}}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

$$\frac{\partial^2 L}{\partial \delta_{1l} \partial \delta_{2l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{t_{ij}^{*}}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

$$\frac{\partial^2 L}{\partial \delta_{1l} \partial \delta_{3l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{t_{ij}^{*;2}}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

$$\frac{\partial^2 L}{\partial \delta_{2l} \partial \delta_{3l}} = -\frac{1}{2}\sum_{ij,i\in l}\frac{t_{ij}^{*;3}}{\sigma_{lij}^2}\big(y_{lij}-f_{lij}\big)^2$$

Les autres dérivées sont nulles.

4) Algorithme SAEM-MCMC : code de l'algorithme en fortran

Programme Fortran : Modèle de variance $V2 : \log(\sigma_{hij}{}^2) = \delta_{1h} + \delta_{2h} t_j * + \delta_{3h} t_j *^2$

subroutine MCMC(seed11,seed22,seed33,seed44,seed55,seed66,rho1,rho2,&
tau2a,tau2c,covac,sigma2ij,L1,L2,L3,lignee,iterMCMC,m1,&
burn,nbChaines,age,N,Ni,y,Ntot,Nind2,phic1,phic3,seed1,seed2,seed3,seed4,seed5,seed6)

USE normale !module qui génère des variables gaussiennes, par la méthode de Marsaglia et Bray(1964)

```
IMPLICIT  NONE
!declaration des variables appelees
integer, intent(in) :: m1,iterMCMC, burn, nbChaines, N, Ntot, Ni
double precision, intent(in) :: seed11, seed22, seed33, seed44, seed55, seed66
double precision, intent(in) ::  tau2a, tau2c, covac
double precision, dimension(Ntot), intent(in) :: y, age, sigma2ij
double precision, dimension(5), intent(in) :: L1, L2, L3
integer,dimension(Ntot), intent(in) ::lignee
double precision, intent(in) :: rho1,rho2
double precision,intent(out) :: seed1, seed2, seed3, seed4, seed5, seed6
double precision, dimension((iterMCMC-burn)*nbChaines,N), intent(out) :: phic1, phic3
integer,dimension(N+1), intent(in) :: Nind2
!
!declaration des variables locales
integer :: i, j, l
double precision ::emsim, emsimul, pr
double precision, dimension(1,1) ::emsimul2, emsim2
double precision, dimension(2,N) :: sim, simul
double precision :: X, vr1, vr3
double precision, dimension(12) :: usim, usimul
integer, parameter :: NMAX=4
double precision, dimension(NMAX,NMAX) :: A, h
double precision, dimension(2,2) :: chol, matinv
integer :: info=0
integer, parameter :: lwork=64*NMAX
integer :: info2
integer, dimension (NMAX) :: IPIV2
double precision, dimension(lwork) :: work
double precision, dimension(2,1) :: tt
double precision, dimension(2,1) :: LL
double precision :: rho
external f07fdf, f07adf



!execution
100 FORMAT(5F8.1,5F5.2,5F6.2,F8.1,F8.2,F8.2,F9.1)

!générer des données normales
seed1=seed11
seed2=seed22
seed3=seed33
seed4=seed44
seed5=seed55
seed6=seed66

do i=1,2
```

```
do j=1,2
 chol(i,j)=0
 matinv(i,j)=0
end do
end do


!calcul de la décomposition de Choleski de la matrice de covariance
 A(1,1)=tau2a
 A(1,2)=covac
 A(2,2)=tau2c
 call f07fdf('U',2,A,NMAX,info)

 if (info==0) then
 chol(1,1:2)=A(1,1:2)
 chol(2,2)=A(2,2)
 else
 write(*,*) 'la matrice de cov n''est pas définie positive'
 end if

!calcul de l'inverse de la matrice de covariance
 h(1,1)=tau2a
 h(1,2)=covac
 h(2,1)=covac
 h(2,2)=tau2c

 call F07ADF(2,2,h,NMAX,IPIV2,info2)

if (info2.eq.0) then
 call F07AJF(2,h,NMAX,IPIV2,work,lwork,info2)
  else
 write(*,*) 'probleme inversion matrice hessienne'
end if

 if (info2.ne.0) then
 write(*,*) 'info2', info2
 end if

 matinv(1:2,1:2)=h(1:2,1:2)

!la boucle sur les chaînes correspond à la construction de nbChaines chaînes indépendantes
 chaines : do l=1,nbChaines

do i=1,N
  ! appel des lois normales
  call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr1,seed1,seed2,seed3,seed4,seed5,seed6)
  call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr3,seed1,seed2,seed3,seed4,seed5,seed6)
  sim(1:2,i)=matmul(transpose(chol),(/vr1,vr3/))+(/L1(lignee(Nind2(i)+1)),L3(lignee(Nind2(i)+1))/)
end do

!la boucle sur j correspond à la construction de chaque iteration de la chaîne l
itMCMC : do j=1,iterMCMC

emsim=0
emsimul=0
emsimul2=0
emsim2=0

indiv : do i=1,N
```

```
 LL(1:2,1)=(/L1(lignee(Nind2(i)+1)),L3(lignee(Nind2(i)+1))/)
call unif(seed1,seed2,seed3,seed4,seed5,seed6,X,seed1,seed2,seed3,seed4,seed5,seed6)

if (j<m1) then
 call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr1,seed1,seed2,seed3,seed4,seed5,seed6)
 call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr3,seed1,seed2,seed3,seed4,seed5,seed6)
 simul(1:2,i)=matmul(transpose(chol),(/vr1,vr3/))+(/L1(lignee(Nind2(i)+1)),L3(lignee(Nind2(i)+1))/)

else
call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr1,seed1,seed2,seed3,seed4,seed5,seed6)
call norm(seed1,seed2,seed3,seed4,seed5,seed6,0.0d0, 1.0d0, vr3,seed1,seed2,seed3,seed4,seed5,seed6)

call unif(seed1,seed2,seed3,seed4,seed5,seed6,rho,seed1,seed2,seed3,seed4,seed5,seed6)
rho=rho2-(rho2-rho1)*rho


 simul(1:2,i)=matmul(rho*transpose(chol),(/vr1,vr3/))+sim(1:2,i)

end if
usim=sim(1,i)*exp(-L2(lignee(Nind2(i)+1))*exp(-sim(2,i)*age(((i-1)*Ni+1):(i*Ni))))
usimul=simul(1,i)*exp(-L2(lignee(Nind2(i)+1))*exp(-simul(2,i)*age(((i-1)*Ni+1):(i*Ni))))

emsim=-sum(1./(2.*sigma2ij(((i-1)*Ni+1):(i*Ni)))*(y(((i-1)*Ni+1):(i*Ni))-usim)**2)
emsimul=-sum(1./(2.*sigma2ij(((i-1)*Ni+1):(i*Ni)))*(y(((i-1)*Ni+1):(i*Ni))-usimul)**2)
tt(1:2,1)=sim(1:2,i)-LL(1:2,1)
emsim2=-1./2.*matmul(matmul(transpose(tt),matinv),tt)
tt(1:2,1)=simul(1:2,i)-LL(1:2,1)
emsimul2=-1./2.*matmul(matmul(transpose(tt),matinv),tt)

 !probabilité d'acceptation
if (j<m1) then
pr=exp(min((emsimul-emsim),0.0d0))
  else
pr=exp(min((emsimul+emsimul2(1,1)-emsim-emsim2(1,1)),0.0d0))
end if

call unif(seed1,seed2,seed3,seed4,seed5,seed6,X,seed1,seed2,seed3,seed4,seed5,seed6)

if (pr>X) then
sim(1:2,i)=simul(1:2,i)
else
sim(1:2,i)=sim(1:2,i)
end if

end do indiv

if  (j>burn) then
phic1(((iterMCMC-burn)*(l-1)+j-burn),1:N)=sim(1,1:N)
phic3(((iterMCMC-burn)*(l-1)+j-burn),1:N)=sim(2,1:N)
end if

end do itMCMC
end do chaines
RETURN
end subroutine
```

```
PROGRAM modele
USE normale

implicit none
!y : les observations : les poids
!L1,L2,L3 : vecteur correspondant aux moyennes des paramètres phi1,phi2,phi3 de Gompertz
!length(L1)=nombre de lignees (1 parametre par lignee)
!matcov : matrice de covariance pour les 3 effets aléatoires
!Ntot : nombre total d'observations
!N : nombre d'individus
!kk boucle de lecture de fichier
!aa,bb,cc paramètre du log de variance résiduelle

!iterMCMC : le nombre d'iterations dans chaque appel de Metropolis Hastings
!iterMCEM=nb_chaines : le nombre de chaines indépendantes dans Metropolis Hastings
!m1 : étape à laquelle on change de noyau dans Metropolis
!burn : les itérations de Metropolis sur lesquelles on ne calcule pas les approximations

!déclaration des constantes

double precision, parameter ::  Pi=3.141593

!les racines d'entrées
double precision, parameter :: x10=16807,x11=282475249,x12=1622650073
double precision, parameter :: x20=984943658,x21=1144108930,x22=470211272

!paramètres à contrôler avant de lancer le programme
integer, parameter :: nbtheta=33
integer, parameter :: Ntot=600,N=50,iterMCMC=500,m1=50,burn=250,Ni=12
integer, parameter :: nbchaines=5
double precision :: rho1=0.25, rho2=1.0,lam=2000.0d0

!entier de boucles
integer :: kk,i,ierror,K=0, sst, llt,as,u, us

!initialisation des paramètres
double precision, dimension(5) :: L1
double precision, dimension(5) :: L2
double precision, dimension(5) :: L3
double precision, dimension(5) :: aah,bbh,cch
double precision :: tau2c, tau2a, covac

!vecteur de lecture du fichier d'entrée
double precision,dimension(Ntot) :: age,agestar
integer,dimension(Ntot) :: lignee
double precision,dimension(Ntot) ::  y
integer,dimension(Ntot) ::  indiv

!racines pour générer les vecteurs aléatoires
double precision :: seed1, seed2, seed3, seed4, seed5, seed6

double precision, dimension (Ntot) :: sigma2ij
integer, dimension(N) :: Nind
integer, dimension(N+1) :: Nind2
double precision, dimension((iterMCMC-burn)*nbChaines,N) :: phic1, phic3
double precision, dimension(N) :: espphi1,esp2phi1,espphi3,esp2phi3, espphi1phi3
double precision :: gam
integer, dimension(5) :: comblignee
double precision, dimension(5) :: LL1, LL3
```

```
!pour inversion newt-raph
integer, parameter :: nbraph=20,NMAX=nbraph*2,lwork=64*NMAX
integer :: info2
integer, dimension (NMAX) :: IPIV2
double precision, dimension(nbraph,nbraph) :: one
double precision, dimension(lwork) :: work
double precision, dimension(NMAX,NMAX) ::  hess2
double precision, dimension(nbraph,nbraph) ::  hess
double precision, dimension(nbraph) :: theta, dtheta

!critere d'arrêt
double precision, dimension(nbtheta) :: sumarret,erreur,num,pente,ok
double precision, dimension(10) :: sss
double precision :: den
double precision, dimension(nbtheta,10) :: tet2

!matrice de covariance
double precision, dimension(10) :: dalga, dalgadt2a, dalgadt2c, dalgadtac
double precision, dimension(10,10) :: d2alga
double precision, dimension(5) :: db, dbdaah,dbdbbh,dbdcch
double precision, dimension(5,5) :: d2b
double precision ::  dt2a,dt2c,dtac,dt2adt2c,dt2adtac,dt2cdtac, d2t2a, d2t2c, d2tac
double precision, dimension(5) :: daah,dbbh,dcch,daahbbh,daahcch,dbbhcch
double precision, dimension(5,5) :: d2aah,d2bbh,d2cch
double precision :: detcov
double precision, dimension(2,2) :: mata,matac,matc
double precision, dimension(1,1) :: sq
double precision, dimension(nbtheta,1) ::  derpretheta,derpretheta2,derpretheta3

double precision, dimension(nbtheta,nbtheta) :: dersectheta,dersectheta2, dersectheta3, estcov, dersectheta4
double precision :: phii1, phii3
double precision, dimension(2,1) :: phii
double precision::  alpij2,dalpijdL22,d2alpijdL22
double precision, dimension(2,1) :: mmm, mmma, mmmc, mmmac
double precision, dimension(2,2) ::  invmat

!inversion matrice de covariance des estimateurs
integer, parameter :: NMAX3=nbtheta*2,lwork3=64*NMAX3
double precision, dimension(lwork3) :: work3
integer :: info3
integer, dimension (NMAX3) :: IPIV3
double precision, dimension(NMAX3,NMAX3) ::  cov3
double precision ::  dt

!estimation de la vraisemblance
double precision ::  vrai, vraicroi
double precision:: vrai2
double precision, dimension(Ni) :: mm
double precision, dimension(2) :: alu
integer, parameter :: NMAXc=4
double precision :: maxi
integer,parameter :: nbChainesv=100, burnv=iterMCMC-10
double precision, dimension((iterMCMC-burnv)*nbChainesv) :: vrai3
double precision, dimension((iterMCMC-burn)*nbChaines) :: vrai4
double precision, dimension((iterMCMC-burnv)*nbChainesv,N) :: phic1v, phic3v
 double precision, dimension(2,1) :: tt, ttv
double precision, dimension(N) :: detcovv
double precision, dimension(2,2) :: invmatv
```

```
double precision, dimension(2,N) :: meanv, varv
double precision, dimension(N) :: covv
double precision, dimension(1,1) :: ar, ar2
integer :: es, es2

!*****************fin des declarations
!initialisation des paramètres

L1=2500.0+(/0.0,0.0,0.0,0.0,0.0/)
L2=4.0+(/0.0,0.0,0.0,0.0,0.0/)
L3=15.0+(/0.0,0.0,0.0,0.0,0.0/)
tau2a=50000.0
tau2c=5.0
 covac=-100.0
aah=5.0d0+(/0.0,0.0,0.0,0.0,0.0/)
bbh=10.0d0+(/0.0,0.0,0.0,0.0,0.0/)
cch=-100.0d0+(/0.0,0.0,0.0,0.0,0.0/)

!L1=2396.0+(/652.0d0,-466.0d0,701.0d0,-608.0d0,0.0d0/)
!L2=3.82+(/1.41,-0.37,0.45,0.59,0.0/)
!L3=15.0+(/-1.47,2.86,1.60,0.62,0.0/)
!tau2a=57349.0
!tau2c=1.57
! covac=-83.0
!aa=10.52
!bb=10.68
!cc=-106.82

mata(1,1:2)=(/1,0/)
mata(2,1:2)=(/0,0/)
matc(1,1:2)=(/0,0/)
matc(2,1:2)=(/0,1/)
matac(1,1:2)=(/0,1/)
matac(2,1:2)=(/1,0/)

vraicroi=0

do kk=1,nbraph
 do us=1,nbraph
  one(kk,us)=0
 enddo
end do

100 FORMAT(5F9.2,5F9.2,5F9.2,F10.1,F10.2,F10.2,15F10.3)

write(*,100) (/L1,L2,L3, tau2a, tau2c, aah,bbh,cch/)

do kk=1,N
espphi1(kk)=0
esp2phi1(kk)=0
espphi3(kk)=0
esp2phi3(kk)=0
espphi1phi3(kk)=0
end do

sumarret=(/L1,L2,L3,tau2a, tau2c,covac,aah,bbh,cch/)

open(unit=8,file='chickmod.txt',status='old',action='read',iostat=ierror)

! lecture du fichier
```

```
openif : if(ierror==0) then   !open was ok

 do kk=1,Ntot
 read(8,*,iostat=ierror) indiv(kk), lignee(kk), age(kk), y(kk)
 end do

end if openif
 close(8)
agestar=(age-20.)/100.
age=age/100.

Nind=(/(0,kk=1,N)/)
do kk=1,Ntot
Nind(indiv(kk))=Nind(indiv(kk))+1
end do
do kk=2,N
Nind(kk)=Nind(kk)+Nind(kk-1)
end do
Nind2=(/0,Nind/)

 comblignee=(/0,0,0,0,0/)
do kk=1,N
 comblignee(lignee(Nind2(kk)+1))=comblignee(lignee(Nind2(kk)+1))+1
end do

!ouverture et écriture du fichier d'ecriture

!open(unit=101,file='reponsechick1val.txt',status='replace',action='write')

seed1=x10
seed2=x11
seed3=x12
seed4=x20
seed5=x21
seed6=x22

theta=(/L2,aah,bbh,cch/)

i=1

SAEM :do
do kk=1,Ntot
as=indiv(kk)

sigma2ij(kk)=exp(aah(lignee(Nind2(as)+1))+bbh(lignee(Nind2(as)+1))*agestar(kk)+cch(lignee(Nind2(as)+1))*a
gestar(kk)**2)
end do


detcov=tau2a*tau2c-covac**2
invmat(1,1:2)=1./detcov*(/tau2c,-covac/)
invmat(2,1:2)=1./detcov*(/-covac,tau2a/)

!def de gam
if ((i>K).and.(K>0)) then
  gam=1./(i-K)
  else
  gam=1
end if
```

```
 call MCMC(seed1,seed2,seed3,seed4,seed5,seed6,rho1,rho2,&
tau2a,tau2c,covac,sigma2ij,L1,L2,L3,lignee,iterMCMC,m1,&
burn,nbChaines,age,N,Ni,y,Ntot,Nind2,phic1,phic3,seed1,seed2,seed3,seed4,seed5,seed6)

LL1=(/0,0,0,0,0/)
LL3=(/0,0,0,0,0/)
do kk=1,N
espphi1(kk)=espphi1(kk)+gam*(sum(phic1(1:((iterMCMC-burn)*nbChaines),kk))/((iterMCMC-
burn)*nbChaines)-&
espphi1(kk))
esp2phi1(kk)=esp2phi1(kk)+gam*(sum(phic1(1:((iterMCMC-burn)*nbChaines),kk)**2)/((iterMCMC-
burn)*nbChaines)-&
esp2phi1(kk))
espphi3(kk)=espphi3(kk)+gam*(sum(phic3(1:((iterMCMC-burn)*nbChaines),kk))/((iterMCMC-
burn)*nbChaines)-&
espphi3(kk))
esp2phi3(kk)=esp2phi3(kk)+gam*(sum((phic3(1:((iterMCMC-burn)*nbChaines),kk))**2)/((iterMCMC-
burn)*nbChaines)-&
esp2phi3(kk))
LL1(lignee(Nind2(kk)+1))=LL1(lignee(Nind2(kk)+1))+sum(phic1(1:((iterMCMC-burn)*nbChaines),kk))/&
((iterMCMC-burn)*nbChaines)
LL3(lignee(Nind2(kk)+1))=LL3(lignee(Nind2(kk)+1))+sum(phic3(1:((iterMCMC-burn)*nbChaines),kk))/&
((iterMCMC-burn)*nbChaines)
espphi1phi3(kk)=espphi1phi3(kk)+gam*(sum(phic1(1:((iterMCMC-
burn)*nbChaines),kk)*phic3(1:((iterMCMC-burn)*nbChaines),kk))/&
((iterMCMC-burn)*nbChaines)-espphi1phi3(kk))

end do

!calcul de l'esp cond de la log vraisemblance complete
!determinant et inversion de la matrice de covariance des effets al
detcov=tau2a*tau2c-covac**2
invmat(1,1:2)=1./detcov*(/tau2c,-covac/)
invmat(2,1:2)=1./detcov*(/-covac,tau2a/)

do es=1,((iterMCMC-burn)*nbChaines)
vrai2=0
do es2=1,N
 tt(1,1)=phic1(es,es2)-L1(lignee((es2-1)*Ni+1))
 tt(2,1)=phic3(es,es2)-L3(lignee((es2-1)*Ni+1))
 ar=-1./2*matmul(matmul(transpose(tt),invmat),tt)
alu=(/phic1(es,es2),phic3(es,es2)/)
mm=alu(1)*exp(-L2(lignee((es2-1)*Ni+1))*exp(-alu(2)*age(((es2-1)*Ni+1):(es2*Ni))))
vrai2=vrai2-sum(1./(2.*sigma2ij(((es2-1)*Ni+1):(es2*Ni)))*(y(((es2-1)*Ni+1):(es2*Ni))-mm)**2)+ar(1,1)
end do
vrai4(es)=vrai2
end do
maxi=maxval(vrai4)
vrai4=vrai4-maxi

vrai=-1./2.*sum(log(2.*Pi*sigma2ij))-N/2.*log(detcov)+log(sum(exp(vrai4)/((iterMCMC-
burn)*nbChaines)))+maxi
write(*,*) 'espcondvraicomp',vrai

if (vrai>vraicroi) then
lam=lam/10.
else
lam=lam*10.
end if
```

```
vraicroi=vrai
write(*,*)'lam',lam
write(*,*)'vrai',vrai

!matrice de covariance

do sst=1,nbtheta
derpretheta2(sst,1)=0
 do kk=1,nbtheta
  dersectheta2(sst,kk)=0
  dersectheta4(sst,kk)=0
 end do
end do
dt=0
do sst=1,((iterMCMC-burn)*nbChaines)
dalga=(/0,0,0,0,0,0,0,0,0,0/)!derivee de log L par rapport à (a,c)
dalgadt2a=(/0,0,0,0,0,0,0,0,0,0/)!derivee croisee de log L par rapport à (a,c) et tau_a^2
dalgadt2c=(/0,0,0,0,0,0,0,0,0,0/)!derivee croisee de log L par rapport à (a,c) et tau_c^2
dalgadtac=(/0,0,0,0,0,0,0,0,0,0/)!derivee croisee de log L par rapport à (a,c) et tau_ac
dt2a=0!derivee premiere par rapport à tau_a^2
dt2c=0!derivee premiere par rapport à tau_c^2
dtac=0!derivee premiere par rapport à tau_ac
dt2adt2c=0!derivee croisee par rapport à tau_a^2 et tau_c^2
dt2adtac=0!derivee croisee par rapport à tau_a^2 et tau_ac
dt2cdtac=0!derivee croisee par rapport à tau_c^2 et tau_ac
d2t2a=0!derivee seconde par rapport à tau_a^2
d2t2c=0!derivee seconde par rapport à tau_c^2
d2tac=0!derivee seconde par rapport à tau_ac
do llt=1,10
 do kk=1,10
  d2alga(llt,kk)=0!hessienne par rapport à (a,c)
 end do
end do
do llt=1,5
 do kk=1,5
  d2b(llt,kk)=0!hessienne pâr rapport à b
  d2aah(llt,kk)=0!hessienne pâr rapport à aah
  d2bbh(llt,kk)=0!hessienne pâr rapport à bbh
  d2cch(llt,kk)=0!hessienne pâr rapport à cch
 end do
end do

db=(/0,0,0,0,0/)!derivee premiere par rapport à b
daah=(/0,0,0,0,0/)!derivee premiere par rapport à aah
dbbh=(/0,0,0,0,0/)!derivee premiere par rapport à bbh
dcch=(/0,0,0,0,0/)!derivee premiere par rapport à cch
daahbbh=(/0,0,0,0,0/)!derivee croisee par rapport à aah et bbh
daahcch=(/0,0,0,0,0/)!derivee croisee par rapport à aah et cch
dbbhcch=(/0,0,0,0,0/)!derivee croisee par rapport à bbh et cch
dbdaah=(/0,0,0,0,0/)!derivee croisee par rapport à b et aah
dbdbbh=(/0,0,0,0,0/)!derivee croisee par rapport à b et bbh
dbdcch=(/0,0,0,0,0/)!derivee croisee par rapport à b et cch

 do kk=1,N
 !derivees premieres
 phii1=phic1(sst,kk)-L1(lignee(Nind2(kk)+1))
 phii3=phic3(sst,kk)-L3(lignee(Nind2(kk)+1))
 phii(1:2,1)=(/phii1,phii3/)
 mmm=matmul(invmat,phii)
 dalga(lignee(Nind2(kk)+1))=dalga(lignee(Nind2(kk)+1))+mmm(1,1)
```

```
 dalga(5+lignee(Nind2(kk)+1))=dalga(5+lignee(Nind2(kk)+1))+mmm(2,1)

sq=-1./2.*tau2c/detcov+1./2.*matmul(matmul(matmul(matmul(transpose(phii),invmat),mata),invmat),phii)
dt2a=dt2a+sq(1,1)

sq=-1./2.*tau2a/detcov+1./2.*matmul(matmul(matmul(matmul(transpose(phii),invmat),matc),invmat),phii)
dt2c=dt2c+sq(1,1)

sq=covac/detcov+1./2.*matmul(matmul(matmul(matmul(transpose(phii),invmat),matac),invmat),phii)
dtac=dtac+sq(1,1)

sq=-1./2.*matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matc),invmat),mata),&
invmat),phii)-
1./2.*matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),mata),invmat),matc),invmat),phi
i)
dt2adt2c=dt2adt2c-1./2.*(detcov-tau2a*tau2c)/(detcov**2)+sq(1,1)

sq=&
-
1./2.*matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matac),invmat),mata),invmat),p
hii)-&

1./2.*matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),mata),invmat),matac),invmat),p
hii)
 dt2adtac=dt2adtac-(tau2c*covac)/(detcov**2)+sq(1,1)

sq=-
1./2.*matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matac),invmat),matc),invmat),p
hii)-&

1./2.*matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matc),invmat),matac),invmat),p
hii)
 dt2cdtac=dt2cdtac-(tau2a*covac)/(detcov**2)+sq(1,1)

sq=matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),mata),invmat),mata),invmat),phii)
d2t2a=d2t2a+1./2.*(tau2c*tau2c)/(detcov**2)-sq(1,1)

sq=matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matc),invmat),matc),invmat),phii)
d2t2c=d2t2c+1./2.*(tau2a*tau2a)/(detcov**2)-sq(1,1)

sq=matmul(matmul(matmul(matmul(matmul(matmul(transpose(phii),invmat),matac),invmat),matac),invmat),phi
i)
d2tac=d2tac+(detcov+2.*(covac*covac))/(detcov**2)-sq(1,1)


 mmma=matmul(matmul(matmul(invmat,mata),invmat),phii)
 mmmc=matmul(matmul(matmul(invmat,matc),invmat),phii)
 mmmac=matmul(matmul(matmul(invmat,matac),invmat),phii)
 dalgadt2a(lignee(Nind2(kk)+1))=dalgadt2a(lignee(Nind2(kk)+1))-mmma(1,1)
 dalgadt2a(5+(lignee(Nind2(kk)+1)))=dalgadt2a(5+(lignee(Nind2(kk)+1)))-mmma(2,1)
 dalgadt2c(lignee(Nind2(kk)+1))=dalgadt2c(lignee(Nind2(kk)+1))-mmmc(1,1)
 dalgadt2c(5+(lignee(Nind2(kk)+1)))=dalgadt2c(5+(lignee(Nind2(kk)+1)))-mmmc(2,1)
 dalgadtac(lignee(Nind2(kk)+1))=dalgadtac(lignee(Nind2(kk)+1))-mmmac(1,1)
 dalgadtac(5+(lignee(Nind2(kk)+1)))=dalgadtac(5+(lignee(Nind2(kk)+1)))-mmmac(2,1)
 d2alga(lignee(Nind2(kk)+1),lignee(Nind2(kk)+1))=d2alga(lignee(Nind2(kk)+1),lignee(Nind2(kk)+1))-
invmat(1,1)
 d2alga(lignee(Nind2(kk)+1),5+lignee(Nind2(kk)+1))=d2alga(lignee(Nind2(kk)+1),5+lignee(Nind2(kk)+1))-
invmat(1,2)
 d2alga(5+lignee(Nind2(kk)+1),lignee(Nind2(kk)+1))=d2alga(lignee(Nind2(kk)+1),5+lignee(Nind2(kk)+1))
```

```
d2alga(5+lignee(Nind2(kk)+1),5+lignee(Nind2(kk)+1))=d2alga(5+lignee(Nind2(kk)+1),5+lignee(Nind2(kk)+1))
-invmat(2,2)

 end do

 do kk=1,Ntot
as=indiv(kk)
alpij2=phic1(sst,as)*exp(-L2(lignee(Nind2(as)+1))*exp(-phic3(sst,as)*age(kk)))
daah(lignee(Nind2(as)+1))=daah(lignee(Nind2(as)+1))-1./2.+1./(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
dbbh(lignee(Nind2(as)+1))=dbbh(lignee(Nind2(as)+1))-agestar(kk)/2.+agestar(kk)/(2.*sigma2ij(kk))*(y(kk)-
alpij2)**2
dcch(lignee(Nind2(as)+1))=dcch(lignee(Nind2(as)+1))-
agestar(kk)**2/2.+agestar(kk)**2/(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
daahbbh(lignee(Nind2(as)+1))=daahbbh(lignee(Nind2(as)+1))-agestar(kk)/(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
daahcch(lignee(Nind2(as)+1))=daahcch(lignee(Nind2(as)+1))-agestar(kk)**2/(2.*sigma2ij(kk))*(y(kk)-
alpij2)**2
dbbhcch(lignee(Nind2(as)+1))=dbbhcch(lignee(Nind2(as)+1))-agestar(kk)**3/(2.*sigma2ij(kk))*(y(kk)-
alpij2)**2
d2aah(lignee(Nind2(as)+1),lignee(Nind2(as)+1))=d2aah(lignee(Nind2(as)+1),lignee(Nind2(as)+1))-&
1./(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
d2bbh(lignee(Nind2(as)+1),lignee(Nind2(as)+1))=d2bbh(lignee(Nind2(as)+1),lignee(Nind2(as)+1))-&
agestar(kk)**2/(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
d2cch(lignee(Nind2(as)+1),lignee(Nind2(as)+1))=d2cch(lignee(Nind2(as)+1),lignee(Nind2(as)+1))-&
agestar(kk)**4/(2.*sigma2ij(kk))*(y(kk)-alpij2)**2
dalpijdL22=-alpij2*exp(-phic3(sst,as)*age(kk))
db(lignee(Nind2(as)+1))=db(lignee(Nind2(as)+1))+1./sigma2ij(kk)*(y(kk)-alpij2)*dalpijdL22
d2alpijdL22=-dalpijdL22*exp(-phic3(sst,as)*age(kk))
d2b(lignee(Nind2(as)+1),lignee(Nind2(as)+1))=d2b(lignee(Nind2(as)+1),lignee(Nind2(as)+1))+1./sigma2ij(kk)*
(&
(y(kk)*d2alpijdL22)-dalpijdL22**2-alpij2*d2alpijdL22)

dbdaah(lignee(Nind2(as)+1))=dbdaah(lignee(Nind2(as)+1))-1./sigma2ij(kk)*(y(kk)-alpij2)*dalpijdL22
dbdbbh(lignee(Nind2(as)+1))=dbdbbh(lignee(Nind2(as)+1))-agestar(kk)/sigma2ij(kk)*(y(kk)-alpij2)*dalpijdL22
dbdcch(lignee(Nind2(as)+1))=dbdcch(lignee(Nind2(as)+1))-agestar(kk)**2/sigma2ij(kk)*(y(kk)-
alpij2)*dalpijdL22
 end do


 !ordre L2,L1,L3,tau2a,tau2c,cov(a,c),aa,bb,cc
derpretheta(1:nbtheta,1)=(/db,dalga,dt2a,dt2c,dtac,daah,dbbh,dcch/)
derpretheta2(1:nbtheta,1)=derpretheta2(1:nbtheta,1)+(/db,dalga,dt2a,dt2c,dtac,daah,dbbh,dcch/)
dersectheta(1:5,1:5)=d2b
dersectheta(1,19)=dbdaah(1)
dersectheta(2,20)=dbdaah(2)
dersectheta(3,21)=dbdaah(3)
dersectheta(4,22)=dbdaah(4)
dersectheta(5,23)=dbdaah(5)
dersectheta(19:23,1:5)=transpose(dersectheta(1:5,19:23))

dersectheta(1,24)=dbdbbh(1)
dersectheta(2,25)=dbdbbh(2)
dersectheta(3,26)=dbdbbh(3)
dersectheta(4,27)=dbdbbh(4)
dersectheta(5,28)=dbdbbh(5)
dersectheta(24:28,1:5)=transpose(dersectheta(1:5,24:28))

dersectheta(1,29)=dbdcch(1)
dersectheta(2,30)=dbdcch(2)
dersectheta(3,31)=dbdcch(3)
```

```
dersectheta(4,32)=dbdcch(4)
dersectheta(5,33)=dbdcch(5)
dersectheta(29:33,1:5)=transpose(dersectheta(1:5,29:33))

dersectheta(6:15,6:15)=d2alga
dersectheta(6:15,16)=dalgadt2a
dersectheta(6:15,17)=dalgadt2c
dersectheta(6:15,18)=dalgadtac
dersectheta(16,6:15)=dersectheta(6:15,16)
dersectheta(17,6:15)=dersectheta(6:15,17)
dersectheta(18,6:15)=dersectheta(6:15,18)
dersectheta(16,16:18)=(/d2t2a,dt2adt2c,dt2adtac/)
dersectheta(16:18,16)=dersectheta(16,16:18)
dersectheta(17,17:18)=(/d2t2c,dt2cdtac/)
dersectheta(17:18,17)=dersectheta(17,17:18)
dersectheta(18,18)=d2tac

do kk=0,4
dersectheta(19+kk,24+kk)=daahbbh(kk+1)
dersectheta(24+kk,19+kk)=daahbbh(kk+1)
dersectheta(19+kk,29+kk)=daahcch(kk+1)
dersectheta(29+kk,19+kk)=dersectheta(19+kk,29+kk)
dersectheta(24+kk,29+kk)=dbbhcch(kk+1)
dersectheta(29+kk,24+kk)=dersectheta(24+kk,29+kk)
dersectheta(19:23,19:23)=d2aah
dersectheta(24:28,24:28)=d2bbh
dersectheta(29:33,29:33)=d2cch
enddo
dersectheta4=dersectheta4+dersectheta
dersectheta2=dersectheta2+dersectheta+matmul(derpretheta,transpose(derpretheta))

end do

derpretheta3=derpretheta3+gam*(derpretheta2/((iterMCMC-burn)*nbChaines)-derpretheta3)
dersectheta3=dersectheta3+gam*(dersectheta2/((iterMCMC-burn)*nbChaines)-dersectheta3)

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!estimation des paramètres des effets aleatoires
L1=L1+gam*(LL1/comblignee-L1)
L3=L3+gam*(LL3/comblignee-L3)

 tau2a=sum(esp2phi1)/N-sum(L1**2*comblignee/N)
 covac=sum(espphi1phi3)/N-sum(L1*L3*comblignee/N)
 tau2c=sum(esp2phi3)/N-sum(L3**2*comblignee/N)

!estimation de aa,bb,cc et L2 par Newton Raphson

do kk=1,nbraph
do sst=1,nbraph
hess2(kk,sst)=0
end do
end do
hess2(1:5,1:5)=dersectheta4(1:5,1:5)/((iterMCMC-burn)*nbChaines)
hess2(6:20,6:20)=dersectheta4(19:33,19:33)/((iterMCMC-burn)*nbChaines)
hess2(6:20,1:5)=dersectheta4(19:33,1:5)/((iterMCMC-burn)*nbChaines)
hess2(1:5,6:20)=transpose(hess2(6:20,1:5))


do kk=1,nbraph
```

```
one(kk,kk)=hess2(kk,kk)
end do
hess2(1:nbraph,1:nbraph)=hess2(1:nbraph,1:nbraph)+lam*one

 call F07ADF(nbraph,nbraph,hess2,NMAX,IPIV2,info2)

if (info2.eq.0) then

 call F07AJF(nbraph,hess2,NMAX,IPIV2,work,lwork, info2)
  else
write(*,*) 'probleme inversion matrice hessienne'
end if
if (info2.ne.0) then
write(*,*) 'probleme inversion matrice hessienne'
end if

hess=hess2(1:nbraph,1:nbraph)
dtheta=(/derpretheta2(1:5,1)/((iterMCMC-burn)*nbChaines),derpretheta2(19:33,1)/((iterMCMC-
burn)*nbChaines)/)

theta=theta-gam*matmul(hess,dtheta)

L2=theta(1:5)
aah=theta(6:10)
bbh=theta(11:15)
cch=theta(16:20)


!fin estimation des parametres
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!temps d'arret
if (i>1) then
erreur=abs((sumarret-(/L1,L2,L3,tau2a, tau2c,covac,aah,bbh,cch/))/&
(/L1,L2,L3,tau2a,tau2c,covac,aah,bbh,cch/))
if(maxval(erreur)<0.0001) then
exit
end if

end if

if (i<=10) then
tet2(1:nbtheta,i)=erreur
end if


if ((i>10).AND.(K==0)) then
do u=1,9
tet2(1:nbtheta,u)=tet2(1:nbtheta,u+1)
end do
tet2(1:nbtheta,10)=erreur
sss=(/(u,u=1,10)/)-sum((/(u,u=1,10)/))/10
den=sum(sss**2)
do u=1,nbtheta
num(u)=sum(sss*(tet2(u,1:10)-sum(tet2(u,1:10))))
if (((num(u)/den-pente(u))<0).and.(i>11)) then
ok(u)=1
end if
end do
if(maxval(erreur)<0.01) then
```

```
K=i+15
end if
pente=num/den
write(120,*) pente
end if

sumarret=(/L1,L2,L3,tau2a,tau2c,covac,aah,bbh,cch/)
write(*,*) 'i=',i

i=i+1
!if (i/100.0-i/100==0.0) then
write(*,*) 'i=',i
write(*,100) (/L1,L2,L3, tau2a, tau2c,covac,aah,bbh,cch/)
!end if

end do SAEM
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!calcul de l'estimation de la matrice de covariance des estimateurs

 cov3(1:nbtheta,1:nbtheta)=dersectheta3-matmul(derpretheta3,transpose(derpretheta3))
 cov3=-cov3

 call F07ADF(nbtheta,nbtheta,cov3,NMAX3,IPIV3,info3)

if (info2.eq.0) then

 call F07AJF(nbtheta,cov3,NMAX3,IPIV3,work3,lwork3, info3)
  else
write(*,*) 'probleme inversion matrice covariance estimateur'
end if
if (info3.ne.0) then
write(*,*) 'probleme inversion matrice covariance estimateur'
end if

estcov=cov3(1:nbtheta,1:nbtheta)

open(unit=102,file='estvar4.txt',status='replace',action='write')
write(102,*) 'L1',L1
write(102,*)'L2',L2
write(102,*)'L3',L3
write(102,*)'tau2a',tau2a
write(102,*)'tau2c',tau2c
write(102,*)'tauac',covac
write(102,*)'aah',aah
write(102,*)'bbh',bbh
write(102,*)'cch',cch

write(*,*)
write(*,*) 'i=',i
write(*,*)
write(*,*) 'K=',K
write(*,*)
!write(*,*)'var estimateurs',estcov(1,1),estcov(2,2),estcov(3,3),estcov(4,4),&
!estcov(5,5),estcov(6,6),&
!estcov(7,7),estcov(8,8),estcov(9,9),estcov(10,10),estcov(11,11),estcov(12,12),&
!estcov(13,13),estcov(14,14),estcov(15,15),estcov(16,16),estcov(17,17),&
!estcov(18,18),estcov(19,19)

write(102,*)'ecart-type estimateurs'
```

```
do kk=1,nbtheta
write(102,*) sqrt(estcov(kk,kk))
end do
close(102)

!estimation de la vraisemblance

!determinant et inversion de la matrice de covaraince des effets al
detcov=tau2a*tau2c-covac**2
invmat(1,1:2)=1./detcov*(/tau2c,-covac/)
invmat(2,1:2)=1./detcov*(/-covac,tau2a/)

 call MCMC(seed1,seed2,seed3,seed4,seed5,seed6,rho1,rho2,&
tau2a,tau2c,covac,sigma2ij,L1,L2,L3,lignee,iterMCMC,m1,&
burnv,nbChainesv,age,N,Ni,y,Ntot,Nind2,phic1v,phic3v,seed1,seed2,seed3,seed4,seed5,seed6)

!estimation des paramètres de la loi de \phi_i|y_i,\theta + calcul du det de la mat de cov
do i=1,N
meanv(1,i)=1./((iterMCMC-burnv)*nbChainesv)*sum(phic1v(1:((iterMCMC-burnv)*nbChainesv),i))
meanv(2,i)=1./((iterMCMC-burnv)*nbChainesv)*sum(phic3v(1:((iterMCMC-burnv)*nbChainesv),i))
varv(1,i)=1./((iterMCMC-burnv)*nbChainesv)*sum(phic1v(1:((iterMCMC-burnv)*nbChainesv),i)**2)-
meanv(1,i)**2
varv(2,i)=1./((iterMCMC-burnv)*nbChainesv)*sum(phic3v(1:((iterMCMC-burnv)*nbChainesv),i)**2)-
meanv(2,i)**2
 covv(i)=1./((iterMCMC-burnv)*nbChainesv)*sum(phic1v(1:((iterMCMC-burnv)*nbChainesv),i)*&
phic3v(1:((iterMCMC-burnv)*nbChainesv),i))-meanv(1,i)*meanv(2,i)
detcovv(i)=varv(1,i)*varv(2,i)-covv(i)**2
end do

do es=1,((iterMCMC-burnv)*nbChainesv)
vrai2=0
do es2=1,N

!inversion de la matrice de correlation
invmatv(1,1:2)=1./detcovv(es2)*(/varv(2,es2),-covv(es2)/)
invmatv(2,1:2)=1./detcovv(es2)*(/-covv(es2),varv(1,es2)/)

 tt(1,1)=phic1v(es,es2)-L1(lignee((es2-1)*Ni+1))
 tt(2,1)=phic3v(es,es2)-L3(lignee((es2-1)*Ni+1))
 ttv(1,1)=phic1v(es,es2)-meanv(1,es2)
 ttv(2,1)=phic3v(es,es2)-meanv(2,es2)
 ar=-1./2.*log(detcov)-1./2*matmul(matmul(transpose(tt),invmat),tt)
 ar2=-1./2*log(detcovv(es2))-1./2.*matmul(matmul(transpose(ttv),invmatv),ttv)

alu=(/phic1v(es,es2),phic3v(es,es2)/)
mm=alu(1)*exp(-L2(lignee((es2-1)*Ni+1))*exp(-alu(2)*age(((es2-1)*Ni+1):(es2*Ni))))
vrai2=vrai2-Ni/2.*log(2*Pi)-sum(1./2*log(sigma2ij(((es2-1)*Ni+1):(es2*Ni)))+1./(2.*sigma2ij(((es2-
1)*Ni+1):(es2*Ni)))*&
(y(((es2-1)*Ni+1):(es2*Ni))-mm)**2)+ar(1,1)-ar2(1,1)

end do
vrai3(es)=vrai2

end do
maxi=maxval(vrai3)
vrai3=vrai3-maxi
vrai=log(sum(exp(vrai3)/((iterMCMC-burnv)*nbChainesv)))+maxi
write(*,*) 'nn,vrai',vrai

END PROGRAM
```

5) Procédure Nlme de R – Fonction de variance V2

Programme R :

```
tutu<-read.table("chickmod.txt")
names(tutu)<-c("animal","strain","age","poids")
tutu$strain<-as.factor(tutu$strain)
tutu$animal<-as.factor(tutu$animal)
tutu$poids[tutu$animal==8&tutu$age==12]<-935
tutu$poids[tutu$animal==26&tutu$age==28]<-3290

library(nlme)
attach(tutu)
data0.fr<-groupedData(poids ~ age | animal,data=data.frame(animal,poids,age,strain))

gompInit<-
function (mCall, data, LHS)
{
   xy <- sortedXyData(mCall[["x"]], LHS, data)
   if (nrow(xy) < 4) {
      stop("too few distinct input values to fit the Gomp model")
   }
   xyL <- xy
   xyL$y <- log(abs(xyL$y))
   pars <- NLSstAsymptotic(xyL)
   pars <- coef( nls(y ~ exp(-B * exp(-C*x)), data = xy, alg = "plinear",
      start=c(B=pars[["b1"]],C=exp(pars[["lrc"]]))))
   val <- pars[c(3, 1, 2)]
   names(val) <- mCall[c("Asym", "B", "C")]
   val
}

gomp<-selfStart(gomp,initial=gompInit)
fm.lis<-nlsList(poids ~ gomp(age/100,Asym,B,C), data=data0.fr)

fm1.nlme<-nlme(fm.lis, random=Asym+C ~ 1)

# pour obtenir les valeurs initiales des paramètres
tutuFix<-fixef(fm1.nlme)

fm2.nlme<-update(fm1.nlme, fixed=Asym+B+C ~ strain,
start=c(tutuFix[1],0,0,0,0,tutuFix[2],0,0,0,0,tutuFix[3],0,0,0,0))

fm3.nlme=update(fm2.nlme,weights=varComb(varIdent(11,~1|strain),varExp(2.5,~((age-
20)/100)|strain),varExp(25,~(((age-20)/100)^2)|strain)))

fm3.nlme

#pour obtenir le résumé de nlme
al=summary(fm3.nlme)

#pour extraire effets fixes, dans notre paramétrisation
parfixes=as.real(al$coefficients$fixed)
a=c(parfixes[1],c(parfixes[1]+ parfixes[2 :5]))
```

```
beta=c(parfixes[6],c(parfixes[6]+ parfixes[7 :10]))
gamma=c(parfixes[11],c(parfixes[11]+ parfixes[11 :15]))

a[1]
beta[1]
gamma[1]

#pour extraire la matrice de corrélation
mat=al$corFixed

#pour extraire les standard errors  des effets fixes
varf=as.real(al$tTable[1:15,2])

#fonction pour calculer les écarts-type des effets fixes
ecart=function(sX,sY,corXY){
ecart=sqrt(sX**2+sY**2+2*corXY*sX*sY)}

#st.error des effets fixes
#Pour a :
sterror_a=c(varf[1],ecart(varf[1],varf[2],mat[1,2]),ecart(varf[1],varf[3],mat[1,3]),
ecart(varf[1],varf[4],mat[1,4]),ecart(varf[1],varf[5],mat[1,5]))
#Pour beta :
sterror_beta=c(varf[6],ecart(varf[6],varf[7],mat[6,7]),ecart(varf[6],varf[8],mat[6,8]),
ecart(varf[6],varf[9],mat[6,9]),ecart(varf[6],varf[10],mat[6,10]))
#Pour gamma :
sterror_gamma=c(varf[11],ecart(varf[11],varf[12],mat[11,12]),ecart(varf[11],varf[13],mat[11,13]),
ecart(varf[11],varf[14],mat[11,14]),ecart(varf[11],varf[15],mat[11,15]))

sterror_a[1]
sterror_beta[1]
sterror_gamma[1]

#-2L
-2*al$logLik

#AIC et BIC
al$AIC
```

Sortie R :

```
> fm3.nlme
Nonlinear mixed-effects model fit by maximum likelihood
  Model: poids ~ gomp(age/100, Asym, B, C)
  Data: data0.fr
  Log-likelihood: -3501.44
  Fixed: Asym + B + C ~ strain
Asym.(Intercept)     Asym.strain2     Asym.strain3     Asym.strain4
    3026.0977046    -1142.9785833       18.6940243    -1230.0455003
    Asym.strain5    B.(Intercept)        B.strain2        B.strain3
    -669.9149557       4.3705873       -0.4303941       -0.0919132
       B.strain4        B.strain5    C.(Intercept)        C.strain2
      -0.4597886       -0.3828771       11.8826229        8.5346429
       C.strain3        C.strain4        C.strain5
       5.1530869        2.3112739        4.0076694
```

```
Random effects:
 Formula: list(Asym ~ 1, C ~ 1)
 Level: animal
 Structure: General positive-definite, Log-Cholesky parametrization
                StdDev     Corr
Asym.(Intercept) 228.404934 As.(I)
C.(Intercept)      1.397428 -0.496
Residual          94.512406

Combination of variance functions:
 Structure: Different standard deviations per stratum
 Formula: ~1 | strain
 Parameter estimates:
       4        2        5        1        3
1.000000 1.494123 1.425506 2.588603 2.994911
 Structure: Exponential of variance covariate, different strata
 Formula: ~((age - 20)/100) | strain
 Parameter estimates:
       4        2        5        1        3
6.679758 4.316743 3.669669 8.236348 3.926180
 Structure: Exponential of variance covariate, different strata
 Formula: ~(((age - 20)/100)^2) | strain
 Parameter estimates:
        4        2        5        1        3
-38.79024 -65.43141 -48.04176 -59.10317 -58.47274
Number of Observations: 600
Number of Groups: 50

> a[1]
[1] 3026.098
> beta[1]
[1] 4.370587
> gamma[1]
[1] 11.88262
>  sterror_a[1]
[1] 81.52629
> sterror_beta[1]
[1] 0.03153598
> sterror_gamma[1]
[1] 0.4769568
> al$AIC
7068.88
> -2* al$logLik
7002.88
attr(,"df")
[1] 14
attr(,"class")
[1] "logLik"
```

Pour notre paramétrisation des variances résiduelles :
On obtient les coefficients delta1_h, delta2_h et delta3_h comme ceci :
delta1_h=log(residual**2*power0**2), delta2_h =2*power1, delta3_h =2*power2
où :
- power0 correspond aux parameters de Formula: ~1 | strain
- power1 aux parameters de Formula: ~((age - 20)/100) | strain
- power2 aux parameters de Formula: ~(((age - 20)/100)^2) | strain

delta1_0=log((2.589463*94.477876)**2)
> delta1_0
[1] 10.99963

delta2_0=2*8.236347
> delta2_0
[1] 16.47269

delta3_0=2*-59.10317
> delta3_0
[1] -118.2063

6) Procédure nlmixed de SAS – Fonction de variance V2

*Programme SAS;

```
libname lib '/ugen/ugencro/Mylene';   filename inA1 'chickmod.txt'; run;

data A1; infile inA1; input animal strain age weight; run;

data a2; set a1;
strain1=0;strain2=0; strain3=0; strain4=0; strain5=0;
if strain=1 then strain1=1;
if strain=2 then strain2=1;
if strain=3 then strain3=1;
if strain=4 then strain4=1;
if strain=5 then strain5=1;
run;

proc nlmixed data=a2 tech=trureg;
parms a0=2500 a1=0 a2=0 a3=0 a4=0 b0=4 b1=0 b2=0 b3=0 b4=0 c0=15 c1=0 c2=0 c3=0 c4=0
f0=10 f1=0 f2=0 f3=0 f4=0 d0=10 d1=0 d2=0 d3=0 d4=0 e0=-100 e1=0 e2=0 e3=0 e4=0 s2au=50000
s2cu=2 sacu=0;
aa=a0+a1*strain1+a2*strain2+a3*strain3+a4*strain4;
a=aa+au;
b=b0+b1*strain1+b2*strain2+b3*strain3+b4*strain4;
cc=c0+c1*strain1+c2*strain2+c3*strain3+c4*strain4;
c=cc+cu;
pred=a*exp(-b*exp(-c*(age/100)));
f=f0+f1*strain1+f2*strain2+f3*strain3+f4*strain4;
d=d0+d1*strain1+d2*strain2+d3*strain3+d4*strain4;
e=e0+e1*strain1+e2*strain2+e3*strain3+e4*strain4;
s2=exp(f+d*(age-20)/100+e*((age-20)/100)**2);
model weight ~ normal(pred,s2);
random au cu ~ normal([0,0],[s2au,sacu,s2cu]) subject=animal;
estimate 'param a0' a0+a1;
estimate 'param a1' a0+a2;
estimate 'param a2' a0+a3;
estimate 'param a3' a0+a4;
estimate 'param a4' a0;
estimate 'param b0' b0+b1;
estimate 'param b1' b0+b2;
estimate 'param b2' b0+b3;
estimate 'param b3' b0+b4;
estimate 'param b4' b0;
estimate 'param c0' c0+c1;
estimate 'param c1' c0+c2;
estimate 'param c2' c0+c3;
estimate 'param c3' c0+c4;
estimate 'param c4' c0;
estimate 'param f0' f0+f1;
estimate 'param f1' f0+f2;
estimate 'param f2' f0+f3;
estimate 'param f3' f0+f4;
estimate 'param f4' f0;
estimate 'param d0' d0+d1;
estimate 'param d1' d0+d2;
estimate 'param d2' d0+d3;
```

```
estimate 'param d3' d0+d4;
estimate 'param d4' d0;
estimate 'param e0' e0+e1;
estimate 'param e1' e0+e2;
estimate 'param e2' e0+e3;
estimate 'param e3' e0+e4;
estimate 'param e4' e0;
run;
```

*** Sortie SAS;**

Initial Parameters

a0	a1	a2	a3	a4	b0	b1	b2	b3	b4	c0	c1	c2	c3	c4	f0	f1	f2
2500	0	0	0	0	4	0	0	0	0	15	0	0	0	0	10	0	0

f3	f4	d0	d1	d2	d3	d4	e0	e1	e2	e3	e4	s2au	s2cu	sacu	NegLogLike
0	0	10	0	0	0	0	-100	0	0	0	0	50000	2	0	3970.26868

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Radius
1*	70	3785.45804	184.8106	291.321	-174.673
2*	105	3699.0215	86.43654	41.91704	-81.2786
3*	140	3666.60418	32.41733	30.81899	-44.239
4*	175	3652.38826	14.21591	37.42978	-14.7881
5*	212	3620.63439	31.75388	90.28021	-68.7966
6*	247	3542.96243	77.67195	29.80093	-87.9863
7*	282	3503.46456	39.49787	9.848434	-39.5112
8*	317	3501.68889	1.775677	1.741462	-1.802
9	352	3501.53753	0.151353	0.254304	-0.14003
10	387	3501.51105	0.026484	0.007769	-0.02563
11	422	3501.51015	0.000896	0.000249	-0.00089
12	457	3501.51015	2.51E-7	6.989E-8	-2.51E-7

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	7003.0
AIC (smaller is better)	7069.0
AICC (smaller is better)	7073.0
BIC (smaller is better)	7132.1

Parameter Estimates

Label	Estimate	Standard Error	DF	t Value	Pr>|t|	Alpha	Lower	Upper
param a0	3027.06	86.0680	48	35.17	<.0001	0.05	2854.01	3200.11
param a1	1882.99	72.5382	48	25.96	<.0001	0.05	1737.14	2028.83
param a2	3044.73	74.2661	48	41.00	<.0001	0.05	2895.41	3194.05
param a3	1795.17	79.0128	48	22.72	<.0001	0.05	1636.31	1954.04
param a4	2356.18	73.1586	48	32.21	<.0001	0.05	2209.09	2503.28
param b0	4.3711	0.03231	48	135.30	<.0001	0.05	4.3062	4.4361
param b1	3.9407	0.03402	48	115.84	<.0001	0.05	3.8723	4.0091
param b2	4.2807	0.06996	48	61.19	<.0001	0.05	4.1400	4.4213
param b3	3.9114	0.03446	48	113.51	<.0001	0.05	3.8421	3.9807
param b4	3.9889	0.06251	48	63.81	<.0001	0.05	3.8632	4.1146
param c0	11.8833	0.4754	48	25.00	<.0001	0.05	10.9274	12.8392
param c1	20.4219	0.4834	48	42.24	<.0001	0.05	19.4499	21.3939
param c2	17.0434	0.5410	48	31.51	<.0001	0.05	15.9557	18.1310
param c3	14.2088	0.5217	48	27.24	<.0001	0.05	13.1600	15.2577
param c4	15.8969	0.5354	48	29.69	<.0001	0.05	14.8204	16.9733
param f0	11.0099	0.3285	48	33.51	<.0001	0.05	10.3494	11.6704
param f1	9.8995	0.2342	48	42.27	<.0001	0.05	9.4286	10.3703
param f2	11.2923	0.2921	48	38.67	<.0001	0.05	10.7051	11.8795
param f3	9.1135	0.3481	48	26.18	<.0001	0.05	8.4136	9.8135
param f4	9.8056	0.3060	48	32.05	<.0001	0.05	9.1904	10.4208
param d0	16.4942	1.2376	48	13.33	<.0001	0.05	14.0058	18.9826
param d1	8.6447	1.3209	48	6.54	<.0001	0.05	5.9888	11.3006
param d2	7.8570	1.2861	48	6.11	<.0001	0.05	5.2711	10.4428
param d3	13.3276	1.5777	48	8.45	<.0001	0.05	10.1553	16.4998
param d4	7.3570	1.3545	48	5.43	<.0001	0.05	4.6336	10.0803
param e0	-118.85	19.6824	48	-6.04	<.0001	0.05	-158.42	-79.2714
param e1	-130.80	13.6875	48	-9.56	<.0001	0.05	-158.32	-103.28
param e2	-117.04	16.7892	48	-6.97	<.0001	0.05	-150.79	-83.2806
param e3	-78.6453	21.3134	48	-3.69	0.0006	0.05	-121.50	-35.7919
param e4	-96.0204	18.7088	48	-5.13	<.0001	0.05	-133.64	-58.4039
s2au	52167	11265	48	4.63	<.0001	0.05	29517	-176E-13
s2cu	1.9585	0.4320	48	4.53	<.0001	0.05	2.8271	1.323E-9
sacu	-158.59	54.8510	48	-2.89	0.0057	0.05	-48.3010	1.35E-10

7) Logiciel WinBUGS : fonction de variance V2

Concernant les résultats obtenus par ce logiciel, nous présentons dans les pages suivantes :
- le code de la procédure,
- les estimations des paramètres obtenues par ce logiciel,
- quelques exemples de graphiques de densités a posteriori,
- quelques exemples de graphiques de convergences,
- quelques exemples de graphiques d'autocorrélation.

JLF 10/2007 chicken_A1R-B1F-C1R-ABC-EXPAGE50-A1F-B1F-C1F

$$y_{hij} = A_{hi} \exp\left[-B_{hi} \exp\left(-C_{hi} T_j\right)\right] + e_{hij}$$

avec $T_j = t_j / 100$ (multiplication du coeft C par 100)

$$A_{hi} = \underbrace{\alpha_h}_{\substack{\text{strain} \\ \text{(fixed)}}} + \underbrace{a_{hi}}_{\substack{\text{animal within strain} \\ \text{(aléatoire)}}} \qquad a_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_a^2\right) \Leftrightarrow A_{hi} \sim_{iid} \mathcal{N}\left(\alpha_h, \sigma_a^2\right)$$

$$B_{hi} = \beta_h + b_{hi} \quad \text{here} \quad B_{hi} = \beta_h$$

$$C_{hi} = \gamma_h + c_{hi} \qquad c_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_c^2\right) \Leftrightarrow C_{hi} \sim_{iid} \mathcal{N}\left(\gamma_h, \sigma_c^2\right)$$

$$\sigma_{ac} = Cov\left(a_{hi}, c_{hi}\right); \; e_{hij} \sim_{iid} \mathcal{N}\left(0, \sigma^2_{e_{t(j)}}\right)$$

$$h = 1(MP), 2(PM), 3(PP), 4(MM), 5(Témoin\ C)$$

Modèle de variance résiduelle par souche

$$\boxed{\sigma_{e_{ht}} = \exp\left(\underline{a}_h + \underline{b}_h t^* + \underline{c}_h t^{*2}\right)} \quad t^* = (t - 20)/100 \text{ où } t = age(sem)$$

Ici $\underline{a} = asigma, \underline{b} = bsigma, \underline{c} = csigma$

#warm up suggested: 10 000 runs

```
model {
    for (i in 1:K) {
    for (j in 1:n) {
    Y[i, j] ~ dnorm(eta[i, j], tauR[i,j])

# model for location/Gompertz function
    eta[i, j] <- phi[i,1]*exp(-muB[i]*exp(-phi[i,2]*x[j]/100))
# model for dispersion
    tauR[i,j]<-1/exp(asig[i]+bsig[i]*t[j]+csig[i]*t[j]*t[j])
    varR[i,j]<-1/tauR[i,j]
      }
phi[i, 1:2] ~ dmnorm(mu[i,1:2], tau[1:2,1:2])

# model for location/continued
mu[i,1]<-alpha[1]*x1[i]+alpha[2]*x2[i]+alpha[3]*x3[i]+alpha[4]*x4[i]+alpha[5]*x5[i]
muB[i]<-beta[1]*x1[i]+beta[2]*x2[i]+beta[3]*x3[i]+beta[4]*x4[i]+beta[5]*x5[i]
mu[i,2]<-gamma[1]*x1[i]+gamma[2]*x2[i]+gamma[3]*x3[i]+gamma[4]*x4[i]+gamma[5]*x5[i]
# model for dispersion/continued
asig[i]<-asigma[1]*x1[i]+asigma[2]*x2[i]+asigma[3]*x3[i]+asigma[4]*x4[i]+asigma[5]*x5[i]
bsig[i]<-bsigma[1]*x1[i]+bsigma[2]*x2[i]+bsigma[3]*x3[i]+bsigma[4]*x4[i]+bsigma[5]*x5[i]
csig[i]<-csigma[1]*x1[i]+csigma[2]*x2[i]+csigma[3]*x3[i]+csigma[4]*x4[i]+csigma[5]*x5[i]
    }

# residual variance by age (j) and breed (k)
for (j in 1:n){
```

```
t[j]<-(x[j]-20)/100
for (k in 1:5){
 varE[j,k]<-exp(asigma[k]+bsigma[k]*t[j]+csigma[k]*t[j]*t[j])
      }
      }

#calculation of contrasts between breeds and control
for (k in 1:4){
alphaC[k]<-alpha[k]-alpha[5]
betaC[k]<-beta[k]-beta[5]
gammaC[k]<-gamma[k]-gamma[5]
   }

#Priors
#on location parameters

for (k in 1:5){
alpha[k]  ~ dnorm(0.0, 1.0E-6)
beta[k]  ~ dnorm(0.0, 1.0E-6)
gamma[k]  ~ dnorm(0.0, 1.0E-6)

asigma[k]  ~  dnorm(0.0, 1.0E-6)
bsigma[k]  ~  dnorm(0.0, 1.0E-6)
csigma[k]  ~  dnorm(0.0, 1.0E-6)
}

#on between individual var covar  parameters
tau[1:2,1:2]  ~ dwish(R[1:2,1:2],4)
var[1:2,1:2]<-inverse(tau[1:2,1:2])

}

Data
   list(n = 12, K = 50, x = c(0,4,6,8,12,16,20,24,28,32,36,40),
x1=c(1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
x2=c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
x3=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
x4=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0),
x5=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1),
      Y = structure(
         .Data = c(44,205,355,640,1020,1600,2310,2880,3650,3800,3520,3240,
46,180,300,520,980,1580,2190,2620,3240,3180,3090,2970,
34,230,370,600,1040,1640,2150,2550,3100,3150,3040,2920,
38,200,330,530,830,1240,1760,1990,2360,2680,2470,2700,
38,265,430,600,1050,1460,1950,2210,2530,2760,2880,2980,
33,175,330,570,1000,1600,2050,2380,2830,2530,2520,2500,
38,225,380,620,1210,1800,2070,2430,3000,2640,2700,2750,
38,170,270,430,935,1440,2050,2360,2820,2820,2820,2810,
35,180,280,500,970,1600,2090,2470,2740,2420,2320,2750,
36,240,390,560,1060,1740,2240,2680,3150,2920,2880,2830,
41,375,660,980,1380,1520,1730,1980,2100,1900,1850,1790,
39,390,650,960,1420,1590,1750,2080,2060,2120,2050,1990,
33,370,620,960,1370,1600,1740,1890,2000,1900,1950,1880,
32,330,550,850,1290,1490,1640,1880,2040,1960,1930,1900,
30,235,420,660,1160,1460,1630,2000,2040,1810,1790,1770,
35,275,510,800,1130,1270,1450,1730,1750,1700,1690,1740,
35,370,635,920,1390,1520,1650,1760,1940,1790,1810,1820,
```

37,305,630,880,1360,1540,1840,2040,2290,2120,2120,2130,
37,390,655,920,1310,1450,1590,1790,1930,1600,1700,1800,
36,310,530,840,1300,1490,1680,1940,1930,1950,1990,1950,
43,380,685,1050,2060,2450,2890,3130,2980,2840,2840,2690,
42,440,675,920,1410,1760,2020,2250,2450,2520,2460,2790,
34,415,690,1060,2140,2460,2790,3130,2850,2900,2900,2950,
38,370,675,1060,2280,2650,3100,3470,3360,3470,3470,3380,
39,385,670,1000,2200,2680,3120,3520,3390,3600,3600,3780,
42,385,635,940,1470,1850,2150,2580,3050,2880,2850,2690,
42,350,685,1100,2160,2460,2850,3350,**3290**,3230,3230,3130,
34,255,465,830,1480,1950,2500,2820,3170,2760,2940,2780,
34,290,520,820,1380,1820,2330,2700,2990,2880,3030,2960,
39,290,545,830,1420,2010,2330,2710,3020,2850,3000,3100,
33,210,305,460,850,1190,1550,1780,1800,1880,1910,1980,
37,175,325,480,800,1230,1540,1810,1910,1830,1800,1910,
36,230,375,570,950,1230,1450,1670,1760,1960,1680,1690,
36,215,340,540,860,1190,1380,1610,1610,1660,1580,1630,
35,195,320,450,730,1170,1470,1780,1660,1630,1550,1460,
31,200,325,500,870,1170,1520,1680,1770,1630,1650,1770,
31,220,345,510,920,1380,1650,1820,1840,1750,1700,1800,
32,230,400,560,920,1220,1460,1510,1520,1550,1540,1580,
33,180,265,380,690,1030,1310,1570,1750,1670,1620,1560,
33,225,335,510,850,1180,1470,1650,1900,1700,1840,1900,
40,355,575,860,1340,1700,2010,2140,2350,2150,2120,2200,
39,330,530,800,1230,1610,1840,2130,2210,2160,2280,2310,
33,300,515,820,1330,1670,1860,2060,2200,2260,2320,2310,
31,250,480,730,1140,1580,1820,2100,2160,2180,2120,2100,
30,305,510,740,1100,1480,1750,2050,2090,1870,1920,2020,
34,260,445,680,1110,1520,1740,2050,2180,2020,2100,2210,
40,300,415,660,1110,1490,1690,1900,2280,2240,2340,2230,
40,315,600,780,1520,2030,2380,2680,2930,2640,2500,2580,
38,285,520,780,1380,1770,2070,2380,2640,2750,2970,2900,
36,280,535,800,1120,1760,2080,2420,2580,2460,2490,2520),
.Dim = c(50, 12)),
R=structure(.Data=c(20.0E4,0,0,10),.Dim=c(2,2)))⇦

Inits **list**(alpha = c(3000,3000,3000,3000,3000),
beta=c(4,4,4,4,4),
gamma=c(12,12,12,12,12),
asigma=c(10,10,10,10,10), bsigma=c(0.1,0.1,0.1,0.1,0.1),
csigma=c(0,0,0,0,0),
tau = structure(.Data=c(2.0E-5,0,0,1),.Dim=c(2,2)))⇦

JLF 10/2007 chicken_A1R-B1F-C1R-ABC-EXPAGE50-A1F-B1F-C1F

JLF 09/2008 chicken_A1R-B1F-C1R-ABC-EXPAGE50-A1F-B1F-C1F

$$y_{hij} = A_{hi} \exp\left[-B_{hi} \exp\left(-C_{hi} T_j\right)\right] + e_{hij}$$

avec $T_j = t_j / 100$ (multiplication du coeft C par 100)

$$A_{hi} = \underbrace{\alpha_h}_{\substack{\text{strain} \\ \text{(fixed)}}} + \underbrace{a_{hi}}_{\substack{\text{animal within strain} \\ \text{(aléatoire)}}} \qquad a_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_a^2\right) \Leftrightarrow A_{hi} \sim_{iid} \mathcal{N}\left(\alpha_h, \sigma_a^2\right)$$

$$B_{hi} = \beta_h + b_{hi} \quad \text{here} \quad B_{hi} = \beta_h$$

$$C_{hi} = \gamma_h + c_{hi} \qquad c_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_c^2\right) \Leftrightarrow C_{hi} \sim_{iid} \mathcal{N}\left(\gamma_h, \sigma_c^2\right)$$

$$\sigma_{ac} = Cov\left(a_{hi}, c_{hi}\right); \; e_{hij} \sim_{iid} \mathcal{N}\left(0, \sigma^2_{e_{t(j)}}\right)$$

$$h = 1(MP), 2(PM), 3(PP), 4(MM), 5(Témoin\ C)$$

Modèle de variance résiduelle par souche

$$\boxed{\sigma_{e_{ht}} = \exp\left(\underline{a}_h + \underline{b}_h t^* + \underline{c}_h t^{*2}\right)} \quad t^* = (t - 20)/100 \text{ où } t = age(sem)$$

Ici $\underline{a} = asigma, \underline{b} = bsigma, \underline{c} = csigma$

#warm up suggested: 10 000 runs

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha[1]	3017.0	92.82	1.196	2838.0	3015.0	3205.0	10001	70000
alpha[2]	1871.0	77.67	0.3953	1717.0	1872.0	2024.0	10001	70000
alpha[3]	3040.0	80.02	0.4754	2882.0	3040.0	3197.0	10001	70000
alpha[4]	1795.0	85.2	1.297	1629.0	1795.0	1965.0	10001	70000
alpha[5]	2343.0	78.57	0.4181	2189.0	2343.0	2497.0	10001	70000
alphaC[1]	673.7	121.9	1.296	436.5	672.9	918.1	10001	70000
alphaC[2]	-471.9	110.6	0.5984	-689.4	-471.8	-255.5	10001	70000
alphaC[3]	697.0	111.9	0.6576	475.3	697.0	916.1	10001	70000
alphaC[4]	-547.9	115.7	1.369	-774.2	-548.2	-319.5	10001	70000
asigma[1]	10.97	0.3451	0.009211	10.27	10.98	11.63	10001	70000
asigma[2]	9.883	0.232	0.007022	9.432	9.879	10.34	10001	70000
asigma[3]	11.15	0.3119	0.007803	10.54	11.15	11.76	10001	70000
asigma[4]	9.054	0.358	0.01259	8.314	9.071	9.727	10001	70000
asigma[5]	9.767	0.3088	0.009816	9.164	9.766	10.38	10001	70000
beta[1]	4.377	0.03613	9.288E-4	4.313	4.375	4.455	10001	70000
beta[2]	3.938	0.03626	8.022E-4	3.863	3.939	4.006	10001	70000
beta[3]	4.351	0.08242	0.002393	4.198	4.347	4.525	10001	70000
beta[4]	3.912	0.03688	0.001009	3.842	3.911	3.986	10001	70000
beta[5]	3.979	0.06433	0.002221	3.846	3.982	4.099	10001	70000
betaC[1]	0.3975	0.07393	0.002441	0.2614	0.3954	0.5477	10001	70000
betaC[2]	-0.04177	0.07419	0.002382	-0.1828	-0.04343	0.1109	10001	70000
betaC[3]	0.3716	0.1059	0.003328	0.1696	0.3677	0.5904	10001	70000
betaC[4]	-0.0676	0.07425	0.00241	-0.2068	-0.06972	0.08271	10001	70000
bsigma[1]	16.45	1.255	0.03375	14.0	16.47	18.91	10001	70000
bsigma[2]	8.774	1.381	0.05086	6.147	8.757	11.5	10001	70000
bsigma[3]	7.896	1.357	0.03488	5.158	7.914	10.53	10001	70000
bsigma[4]	13.43	1.563	0.05007	10.33	13.46	16.41	10001	70000

bsigma[5]	7.505	1.427	0.03918	4.81	7.46	10.38	10001	70000
csigma[1]	-113.4	20.58	0.5426	-151.1	-114.4	-70.88	10001	70000
csigma[2]	-127.3	14.13	0.406	-154.2	-127.5	-99.19	10001	70000
csigma[3]	-106.9	18.66	0.4843	-142.0	-107.4	-69.06	10001	70000
csigma[4]	-72.97	21.55	0.7904	-111.6	-73.98	-28.35	10001	70000
csigma[5]	-90.91	19.03	0.6289	-127.0	-91.45	-52.09	10001	70000
gamma[1]	11.93	0.525	0.004371	10.89	11.93	12.96	10001	70000
gamma[2]	20.44	0.5304	0.004856	19.39	20.44	21.48	10001	70000
gamma[3]	17.06	0.593	0.009351	15.89	17.06	18.23	10001	70000
gamma[4]	14.18	0.5707	0.01022	13.04	14.18	15.29	10001	70000
gamma[5]	15.89	0.5802	0.01073	14.74	15.89	17.03	10001	70000
gammaC[1]	-3.959	0.7836	0.01168	-5.502	-3.956	-2.417	10001	70000
gammaC[2]	4.55	0.7889	0.01194	3.005	4.547	6.112	10001	70000
gammaC[3]	1.175	0.8321	0.01458	-0.4618	1.175	2.819	10001	70000
gammaC[4]	-1.712	0.8139	0.01501	-3.313	-1.711	-0.1208	10001	70000
var[1,1]	60330.0	13940.0	90.97	38750.0	58410.0	92750.0	10001	70000
var[1,2]	-158.6	66.59	0.4741	-309.3	-151.5	-46.19	10001	70000
var[2,1]	-158.6	66.59	0.4741	-309.3	-151.5	-46.19	10001	70000
var[2,2]	2.42	0.5592	0.003855	1.545	2.346	3.719	10001	70000

	Dbar	**Dhat**	**DIC**	**pD**
Y	6725.0	6614.0	6835.0	110.5
total	6725.0	6614.0	6835.0	110.5

Minimum deviance
6661.0

LF 10/2007 chicken_A1R-B1F-C1R-ABC-EXPAGE50-A1F-B1F-C1F

$$y_{hij} = A_{hi} \exp\left[-B_{hi} \exp\left(-C_{hi} T_j\right)\right] + e_{hij}$$

avec $T_j = t_j / 100$ (multiplication du coeft C par 100)

$$A_{hi} = \underbrace{\alpha_h}_{\substack{\text{strain} \\ \text{(fixed)}}} + \underbrace{a_{hi}}_{\substack{\text{animal within strain} \\ \text{(aléatoire)}}} \qquad a_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_a^2\right) \Leftrightarrow A_{hi} \sim_{iid} \mathcal{N}\left(\alpha_h, \sigma_a^2\right)$$

$$B_{hi} = \beta_h + b_{hi} \quad \text{here} \quad B_{hi} = \beta_h$$

$$C_{hi} = \gamma_h + c_{hi} \qquad c_{hi} \sim_{iid} \mathcal{N}\left(0, \sigma_c^2\right) \Leftrightarrow C_{hi} \sim_{iid} \mathcal{N}\left(\gamma_h, \sigma_c^2\right)$$

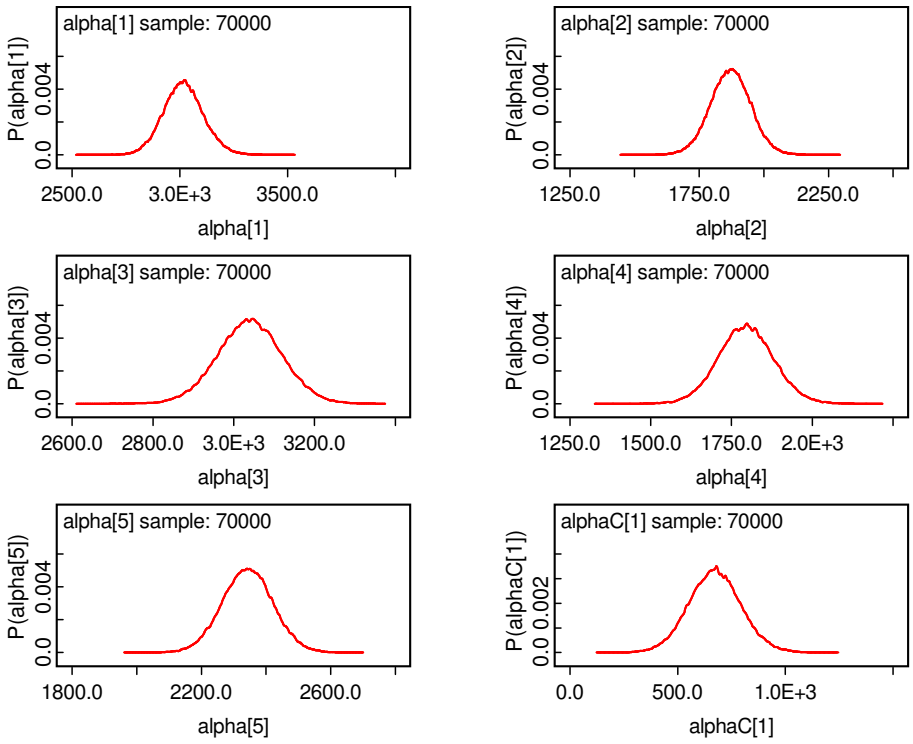
$$\sigma_{ac} = Cov\left(a_{hi}, c_{hi}\right); \; e_{hij} \sim_{iid} \mathcal{N}\left(0, \sigma^2_{e_{t(j)}}\right)$$

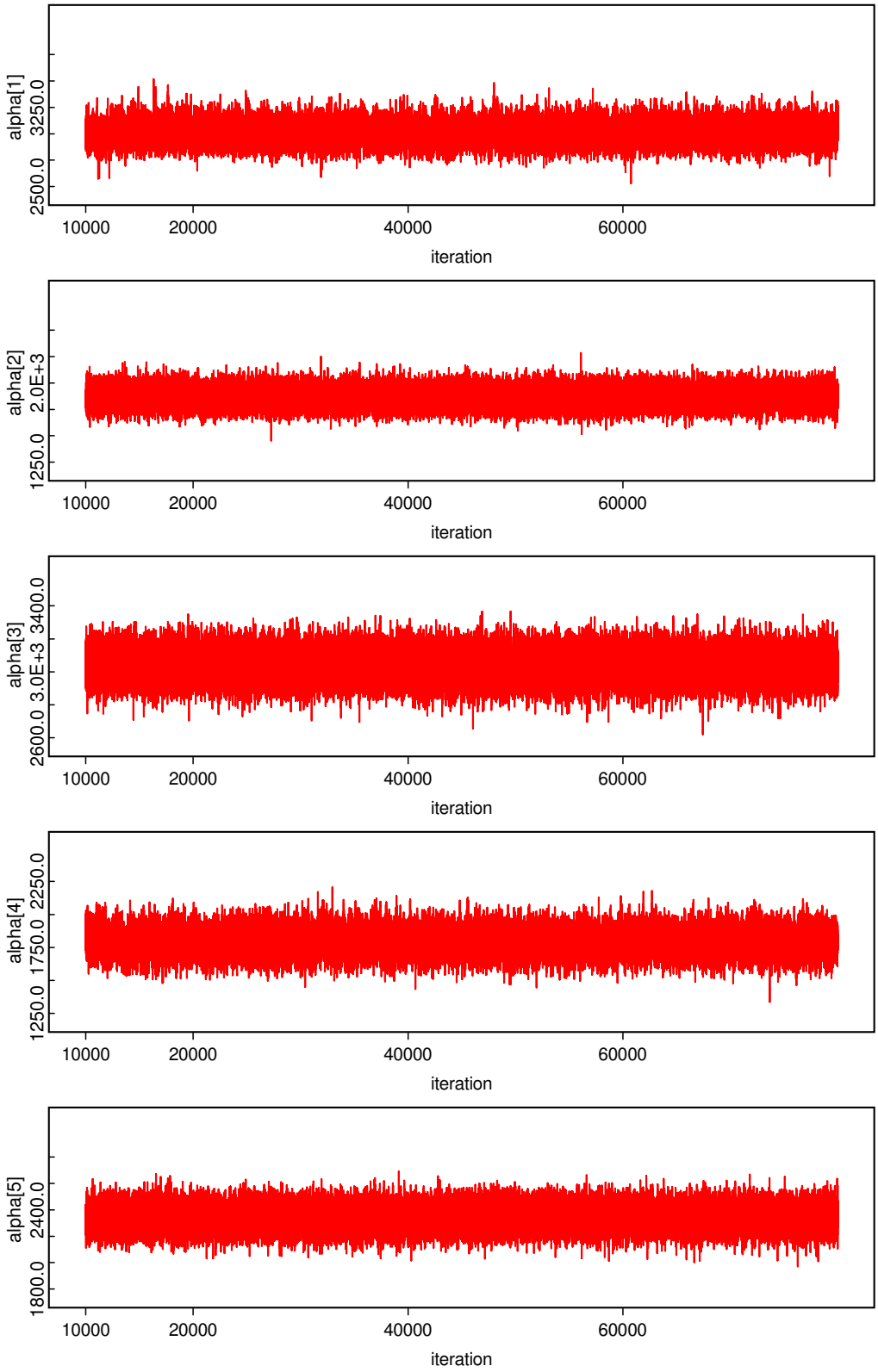
$$h = 1(MP), 2(PM), 3(PP), 4(MM), 5(Témoin\ C)$$

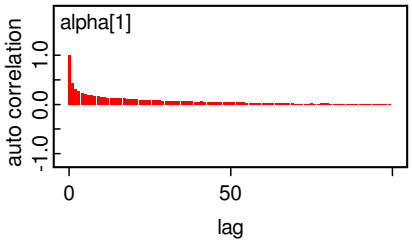
Modèle de variance résiduelle par souche

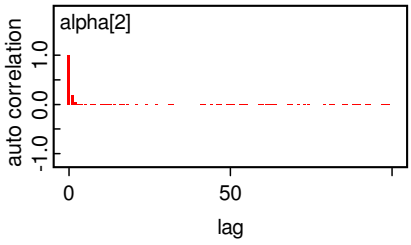
$$\boxed{\sigma_{e_{ht}} = \exp\left(\underline{a}_h + \underline{b}_h t^* + \underline{c}_h t^{*2}\right)} \quad t^* = (t - 20)/100 \text{ où } t = age(sem)$$

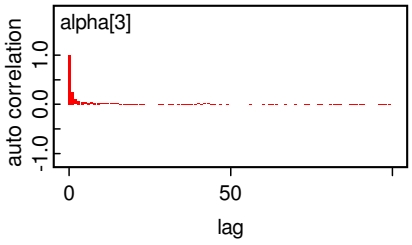
Ici $\underline{a} = asigma, \underline{b} = bsigma, \underline{c} = csigma$

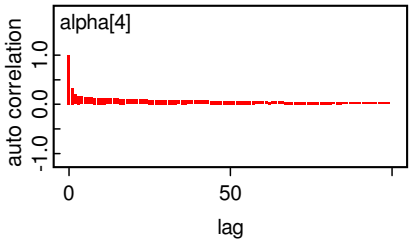


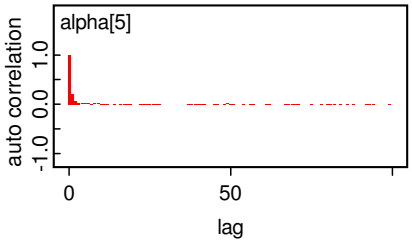

alpha[1]
2500.0
3250.0
10000
20000
40000
60000
iteration
alpha[2]
1250.0
2.0E+3
10000
20000
40000
60000
iteration
alpha[3]
2600.0
3.0E+3
3400.0
10000
20000
40000
60000
iteration
alpha[4]
1250.0
1750.0
2250.0
10000
20000
40000
60000
iteration
alpha[5]
1800.0
2400.0
10000
20000
40000
60000
iteration

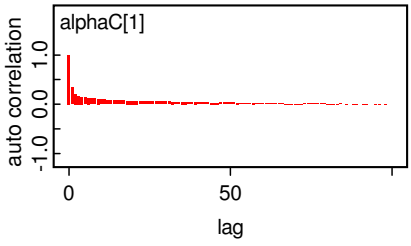

alpha[1]
auto correlation
-1.0 0.0 1.0
0
50
lag

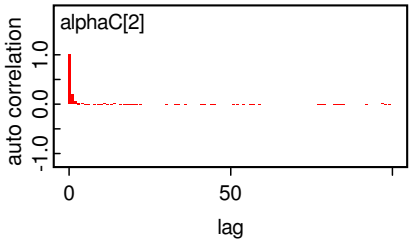

alpha[2]
auto correlation
-1.0 0.0 1.0
0
50
lag

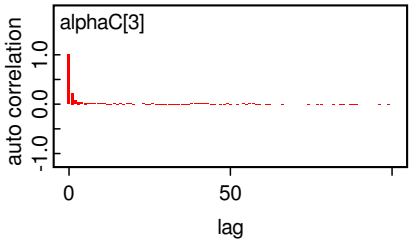

alpha[3]
auto correlation
-1.0 0.0 1.0
0
50
lag

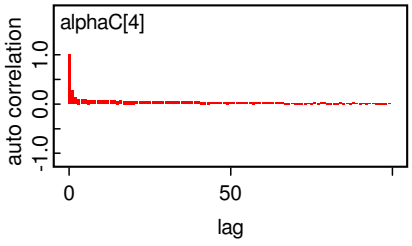

alpha[4]
auto correlation
-1.0 0.0 1.0
0
50
lag

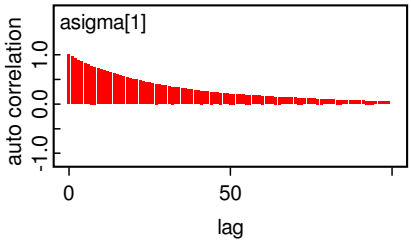

alpha[5]
auto correlation
-1.0 0.0 1.0
0
50
lag

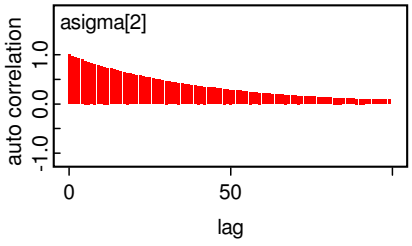

alphaC[1]
auto correlation
-1.0 0.0 1.0
0
50
lag

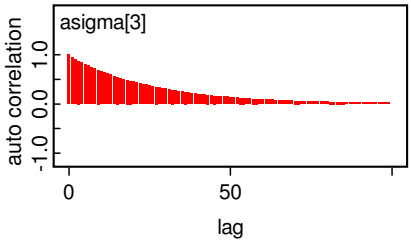

alphaC[2]
auto correlation
-1.0 0.0 1.0
0
50
lag


alphaC[3]
auto correlation
-1.0 0.0 1.0
0
50
lag


alphaC[4]
auto correlation
-1.0 0.0 1.0
0
50
lag


asigma[1]
auto correlation
-1.0 0.0 1.0
0
50
lag


asigma[2]
auto correlation
-1.0 0.0 1.0
0
50
lag


asigma[3]
auto correlation
-1.0 0.0 1.0
0
50
lag

Paramètre	valeur initiale	SAEM-MCMC		Nlmixed	
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE
a_0	2500	3025.20	85.66	3027.06	86.07
a_1	2500	1883.10	72.55	1882.99	72.54
a_2	2500	3044.40	74.24	3044.73	74.26
a_3	2500	1793.30	77.73	1795.17	79.01
a_4	2500	2356.20	73.17	2356.20	73.15
β_0	4	4.37	0.03	4.37	0.03
β_1	4	3.94	0.03	3.94	0.03
β_2	4	4.28	0.07	4.28	0.07
β_3	4	3.91	0.03	3.91	0.03
β_4	4	3.99	0.06	3.99	0.06
c_0	15	11.88	0.48	11.88	0.48
c_1	15	20.41	0.48	20.42	0.48
c_2	15	17.05	0.55	17.04	0.54
c_3	15	14.22	0.51	14.21	0.52
c_4	15	15.90	0.53	15.90	0.54
τ_a^2	50000	52230.00	11247.00	52167.00	11265.00
τ_c^2	5	1.96	0.43	1.96	0.43
τ_{ac}	-100	-159.21	54.69	-158.59	54.85
δ_{10}	5	11.02	0.33	11.01	0.33
δ_{11}	5	9.90	0.23	9.90	0.23
δ_{12}	5	11.30	0.29	11.29	0.29
δ_{13}	5	9.13	0.32	9.11	0.35
δ_{14}	5	9.80	0.30	9.81	0.30
δ_{20}	10	16.48	1.24	16.49	1.24
δ_{21}	10	8.62	1.31	8.64	1.32
δ_{22}	10	7.85	1.29	7.86	1.29
δ_{23}	10	13.28	1.50	13.33	1.59
δ_{24}	10	7.37	1.35	7.36	1.36
δ_{30}	-100	-119.67	19.45	-118.85	16.68
δ_{31}	-100	-130.76	13.68	-130.80	13.68
δ_{32}	-100	-117.72	16.63	-117.04	16.79
δ_{33}	-100	-79.73	19.56	-78.65	21.31
δ_{34}	-100	-96.00	18.05	-96.02	18.71
$-2L^{(a)}$		7002.88		7003.00	
AIC		7068.90		7069.00	
BIC		7132.00		7132.10	

$^{(a)}$ $-2L = -2\log\hat{p}(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$

TAB. 4.14 – Valeurs de $\hat{\theta}$ et des erreurs standards SE obtenues par les méthodes SAEM-MCMC et NlMixed pour la fonction de variance V2.

Paramètre	valeur initiale	nlme		WinBUGS	
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	$\hat{\boldsymbol{\theta}}$	SE	$\hat{\boldsymbol{\theta}}$	SE
a_0	2500	3026.10	81.53	3018.00	93.47
a_1	2500	1883.10	73.46	1871.00	77.24
a_2	2500	3044.80	75.07	3039.00	79.97
a_3	2500	1796.10	75.11	1796.00	83.88
a_4	2500	2356.20	74.00	2343.00	78.34
β_0	4	4.37	0.03	4.38	0.04
β_1	4	3.94	0.03	3.94	0.04
β_2	4	4.28	0.07	4.36	0.09
β_3	4	3.91	0.04	3.91	0.04
β_4	4	3.99	0.05	3.98	0.06
c_0	15	11.88	0.48	11.93	0.53
c_1	15	20.42	0.49	20.45	0.53
c_2	15	17.04	0.54	17.07	0.59
c_3	15	14.19	0.49	14.17	0.57
c_4	15	15.89	0.50	15.89	0.57
τ_a^2	50000	52168.00		60310.00	13870.00
τ_c^2	5	1.95		2.42	0.56
τ_{ac}	-100	-158.31		-157.80	66.65
δ_{1_0}	5	11.00		10.95	0.34
δ_{11}	5	9.90		9.91	0.24
δ_{12}	5	11.29		11.13	0.30
δ_{13}	5	9.10		9.04	0.33
δ_{14}	5	9.81		9.77	0.30
δ_{20}	10	16.47		16.34	1.26
δ_{21}	10	8.63		8.65	1.36
δ_{22}	10	7.85		7.90	1.36
δ_{23}	10	13.36		13.45	1.52
δ_{24}	10	7.34		7.46	1.40
δ_{30}	-100	-118.20		-113.10	20.39
δ_{31}	-100	-130.86		-128.50	13.88
δ_{32}	-100	-116.96		-106.40	17.93
δ_{33}	-100	-77.53		-72.04	19.96
δ_{34}	-100	-96.09		-90.74	18.77
$-2LogL^{(a)}$		7002.88		$6613.70^{(b)}$	
AIC		7068.90		$6834.95^{(c)}$	
BIC		7132.00			

$^{(b)}$ "Dhat" dans la terminologie WinBUGS : espérance a posteriori de la deviance

$^{(c)}$ Critère DIC

TAB. 4.15 – Valeurs de $\hat{\theta}$ et des erreurs standards SE obtenues par Winbugs et nlme pour la fonction de variance V2.

Publications scientifiques

Articles scientifiques :
A1- Jaffrézic, F., de Koning, D-J., Boettcher, P. J., Bonnet, A., Buitenhuis, B., Closset, R., Déjean, S., Delmas, C., Detilleux, J. C., Dovc, P., Duval, M., Foulley, J. L., Hedegaard, J., Hornshøj, H., Hulsegge, I.B., Janss, L., Jensen, K., Jiang, L., Lavric, M., Lê Cao, K.-A., Lund, M.S., Malinverni, Marot, G., Nie, H., Petzl, W., Pool, W.H., Robert-Granié, C., SanCristobal, M., van Schothorst, E. M., Schuberth, H. J., Sørensen, P., Stella, A., Tosser-Klopp, G., Waddington, D., Watson, M., Yang, W., Zerbe, H., Seyfert, H.-M., 2007. Analysis of the real EADGENE data set : Comparison of methods and guidelines for data normalization and selection of differentially expressed genes. Genetics Selection Evolution, 39, 6, 633-650.

A2- Sorensen, P., Bonnet, A., Buitenhuis, B., Closset, R., Déjean, S., Delmas, C., Duval, M., Glass, L., Hedegaard, H., Hornshoj, J., Hulsegge, I.B., Jaffrézic, F., Jensen, K., Jiang, L., de Koning, D-J, Lê Cao, K.-A., Niei, H., Petzl, W., Pool, M. H., Robert-Granié, C., San Cristobal, M., Sando Lund, M., van Schothorst, E. M., Schuberth, H-J., Seyfert, H.-M., Tosser-Klopp, G., Waddington, D., Watson, M., Yang, W., Zerbe, H., 2007. Analysis of the real EADGENE data set : Multivariate approaches and post analysis. Genetics Selection Evolution, 39, 6, 651-668.

A3- Watson, M., Perez-Alegre, M., Denis Baron, M., Delmas, C., Dovc, P., Duval, M., Foulley, J. L., Garrido-Pavon, J.J., Hulsegge, B., Jaffrézic, F., Jiménez-Marin, A., Lavric, M., Lê Cao, K.-A., Marot, G., Mouzaki, D., Pool, M.H., Robert-Granié, C., San Cristobal, M., Tosser-Klop, G., Waddington, D., Koning, D.J. de, 2007. Analysis of a simulated microarray dataset : Comparison of methods for data normalisation and detection of differential expression. Genetics Selection Evolution, 39, 6, 669-683.

A4- Duval, M., Delmas, C., Laurent, B., Robert-Granié, C., 2007. A procedure based on partial sums of order statistics to detect differentially expressed genes. http ://hal.archives-ouvertes.fr/hal-00302355/fr/

A5- Duval, M., Robert-Granié, C., 2007. Criteria to calibrate the parameters of the SAEM-MCMC algorithm in maximum likelihood estimation for nonlinear mixed effects models. Soumis à Computational Statistics.

A6- Duval, M., Robert-Granié, C., Foulley, J. L., 2008. Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm. Soumis à Statistical Modelling.

Communications à des congrès :
B1- Nozière, P., Duval, M., Brossart, L., Courtial, C., Hoch, T. Modélisation dynamique

de l'absorption ruminale des acides gras volatils : influence du pH. 10ème rencontre recherche ruminants (3R), Paris, France, 3-4 Décembre, 2003.

B2- Duval, M., Degrelle, S., Delmas, C., Hue, I., Laurent, B., Robert-Granié, C. A novel procedure to determine differentially expressed genes between several conditions. 4th Workshop on statistical analysis of post genomic data, Toulouse, France, March 30-31, 2006.

B3- Duval, M., Degrelle, S., Delmas, C., Hue, I., Laurent, B., Robert-Granié, C. A novel procedure to determine differentially expressed genes between two conditions. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil, August 13-18, 2006.

B4- Duval, M., Degrelle, S., Delmas, C., Hue, I., Laurent, B., Robert-Granié, C. A simple procedure to determine differentially expressed genes between several conditions. Workshop "Statistique Mathématique et Applications", Luminy, France, 13-17 Novembre, 2006.

B5- Lê Cao, K.-A., San Cristobal, M., Tosser-Klopp, G., Delmas, C., Duval, M., Robert-Granié, C. Gene profile clustering, selection of predictive genes using random forests and a stochastic algorithm, regulatory networks in a transcriptomic kinetics on bovine mastitis. EADGENE "Data Analysis Workshop",Tune, Denmark, November 1-3, 2006.

B6- Delmas, C., Robert-Granié, C., Duval, M., Lê Cao, K.-A., San Cristobal, M., Tosser-Klop, G. Two new procedures to detect differentially expressed genes. EADGENE "Data Analysis Workshop",Tune, Denmark, November 1-3, 2006.

B7- Duval, M., Robert-Granié, C. Modèles non linéaires mixtes : SAEM-MCMC en pratique. 39èmes journées de la Société Française de Statistique, Angers, France, 11-15 Juin, 2007.

B8- Duval, M., Robert-Granié, C., Foulley, J. L. Heterogeneous variances in nonlinear mixed effects models via the SAEM-MCMC algorithm. 24th International Biometric Conference, Dublin, Ireland, July 13-18, 2008.

Rapports diplomants :
C1- Duval M., 2003. Modélisation de la cinétique d'absorption des AGV chez les ruminants. Rapport de stage de Maîtrise, INRA Clermont-Ferrand, Juillet.

C2- Duval M., 2005. Sélection de modèles et applications aux données d'expression des gènes. Rapport de stage de Master2, INRA Toulouse, Juin.