



**HAL**  
open science

# Identifier des légumineuses à graines productives en Europe par synthèses quantitatives de données à large échelle

Charles Cernay

► **To cite this version:**

Charles Cernay. Identifier des légumineuses à graines productives en Europe par synthèses quantitatives de données à large échelle. Sciences agricoles. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLA014 . tel-02888685

**HAL Id: tel-02888685**

**<https://pastel.hal.science/tel-02888685>**

Submitted on 3 Jul 2020

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

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

N°2016SACLA014

THESE DE DOCTORAT  
DE  
L'UNIVERSITE PARIS-SACLAY  
PREPAREE A  
L'INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT  
(AGROPARISTECH)

ECOLE DOCTORALE N°581

Agriculture, alimentation, biologie, environnement et santé

Spécialité de doctorat : Sciences agronomiques

par

**Charles Cernay**

**Identifier des légumineuses à graines productives en Europe  
par synthèses quantitatives de données à large échelle**

*Identifying productive grain legumes in Europe  
by large-scale quantitative data syntheses*

**Thèse de doctorat soutenue à Paris, le 10 novembre 2016, devant le Jury :**

Philippe Lescoat	Professeur, INRA, AgroParisTech, Université Paris-Saclay	Président
Chantal Le Mouél	Directrice de Recherche, INRA	Rapporteur
Eric Malézieux	Directeur de Recherche, CIRAD	Rapporteur
David Makowski	Directeur de Recherche, INRA, AgroParisTech, Université Paris-Saclay	Directeur
Elise Pelzer	Chargée de Recherche, INRA, AgroParisTech, Université Paris-Saclay	Invitée



« Rien ne va de soi. Rien n'est donné. Tout est construit. »

Gaston Bachelard, *La formation de l'esprit scientifique*, 1938





## Avant-propos

Edimbourg, septembre 2016. C'est une belle fin de journée. J'ai pris mon sac à dos, et je me promène dans les rues de la ville. On entend ici le bruit des mouettes – et au loin, le murmure d'une cornemuse. Je me suis assis sur un banc, et j'ai songé à cette histoire. Celle de chiche le pois et du haricot écossais. Une aventure de légumineuses qui a commencé il y a trois années.

Je remercie très chaleureusement mes encadrants, David et Elise. Merci à toi, David, pour m'avoir suivi tout au long de ce chemin. J'ai beaucoup apprécié de travailler à tes côtés. Je tiens à te remercier pour la formidable qualité de ton encadrement, ta bonne humeur, ta grande disponibilité, et ton dynamisme. Merci à toi, Elise, de m'avoir guidé dans le monde des légumineuses. Je tiens à te remercier pour ta bienveillance, ta confiance, ton énergie, et ton aide à déchiffrer certains hiéroglyphes issus de différentes expérimentations agronomiques aux quatre coins de la planète. Je remercie très vivement Tamara qui a participé à la construction de la thèse. Au-delà de ton aide précieuse, j'ai beaucoup apprécié ton enthousiasme, et ton optimisme. Je remercie chaleureusement Jean-Marc Meynard pour nos échanges constructifs au début de la thèse. Je tiens à remercier Marie-Hélène Jeuffroy pour m'avoir donné l'opportunité de découvrir la communauté scientifique des légumineuses. Merci à vous deux, pour m'avoir permis d'apprécier la richesse des sciences agronomiques. Je remercie Philippe Lescoat d'avoir accepté d'être président du jury et examinateur. Je tiens à remercier Chantal Le Mouël et Eric Malézieux d'avoir acceptés d'être rapporteurs. Merci à Sofia pour son remarquable stage. Un grand merci à Guénaëlle Hellou, Olivier Réchauchère, Anne Schneider, et Anne-Sophie Voisin, pour avoir participé aux comités de thèse ; vos remarques ont mis en lumière des idées centrales. Je remercie également toutes les personnes qui ont participé au bon déroulement de la thèse. Merci à Laure Hossard qui a construit le modèle relationnel de la base de données. Merci à Jean-Marc Teulé qui m'a appris le langage SQL. Merci à Alain Bône qui m'a aidé dans la recherche d'articles scientifiques, parfois émanant de contrées éloignées, et de temps anciens. Je remercie Diana pour sa gentillesse et ses conseils. Je tiens à remercier Anabelle, Marion, Rémy, Luce, et Lucie, pour leur bonne humeur. Je tiens à remercier Amandine, Cynthia, Jean-Baptiste, Valentin, Maud, Céline, Maude, Aurya, et Anaïs, pour les apéros et les parties de tarot. Un grand merci à toute l'UMR Agronomie pour son accueil chaleureux. Merci à Michèle pour la gestion efficace des aspects administratifs de la thèse. Une mention spéciale à Damien. Merci à toi de m'avoir fait relever la tête des codes et des scripts. Barcelone fût riche de péripéties ensolleillées. Je remercie également les collocataires grignonnais avec qui j'ai partagé d'agréables moments. Une dédicace à Mathilde et Louis. Enfin, merci à mes parents, et à mes deux frères, pour leur soutien et leurs encouragements, chemin faisant.



## Articles scientifiques dans des revues à comité de lecture

---

- 1.Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M. & Makowski, D. Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* **5**, 11171 (2015). [Publié]
- 2.Cernay, C., Pelzer, E., & Makowski, D. A global experimental dataset for assessing grain legume production. *Sci. Data* **3**, 160084 (2016). [Publié]
- 3.Cernay, C., Makowski, D. & Pelzer, E. Ranking grain legume productivity in Europe, North America and Oceania. (2016). [Soumis pour publication]
- 4.Cernay, C., Pelzer, E. & Makowski, D. Preceding grain legume cultivation increases cereal yields under low nitrogen input conditions. (2016). [Soumis pour publication]

## Bases de données scientifiques dans des revues à comité de lecture

---

- 1.Cernay, C., Pelzer, E. & Makowski, D. A global experimental dataset for assessing grain legume production. *Dryad Digital Repository* (2016). [Publié]

## Communications scientifiques orales

---

- 1.Cernay, C., Pelzer, E., Meynard, J.-M. & Makowski, D. Comparer les performances agronomiques et environnementales d'une diversité de légumineuses à graines cultivables en Europe. LabEx BASC: Gif-sur-Yvette, France, March 2015.
- 2.Cernay, C., Makowski, D. & Pelzer, E. Face au pois, qui fait le poids ? Hiérarchiser les rendements et les quantités en protéines de légumineuses à graines en Europe, Amérique du Nord et Océanie. 1<sup>st</sup> Francophone Conference on Legumes: Dijon, France, May-June 2016.
- 3.Cernay, C., Makowski, D. & Pelzer, E. To pea or not to pea? Comparing grain legume species from field experiments in Europe, North America and Oceania. 14<sup>th</sup> Congress of the European Society for Agronomy: Edinburgh, Scotland, September 2016.

## Communications scientifiques écrites

---

- 1.Cernay, C., Pelzer, E., Ben-Ari, T., Meynard, J.-M. & Makowski, D. Comparison of yields, protein contents and environmental impacts of grain legumes for designing legume-based cropping systems in Europe. Proceedings of the 5<sup>th</sup> International Symposium for Farming Systems Design: Montpellier, France, September 2015.
- 2.Cernay, C., Makowski, D. & Pelzer, E. Face au pois, qui fait le poids ? Hiérarchiser les rendements et les quantités en protéines de légumineuses à graines en Europe, Amérique du Nord et Océanie. Proceedings of the 1<sup>st</sup> Francophone Conference on Legumes: Dijon, France, May-June 2016.
- 3.Cernay, C., Makowski, D. & Pelzer, E. To pea or not to pea? Comparative insights on legume grain and grain protein yields from Europe, North America and Oceania. Proceedings of the 14<sup>th</sup> Congress of the European Society for Agronomy: Edinburgh, Scotland, September 2016.

## Posters

---

- 1.Cernay, C., Pelzer, E., Meynard, J.-M. & Makowski, D. Evaluation agronomique et environnementale des systèmes de cultures incluant des légumineuses à l'échelle mondiale. LabEx BASC: Gif-sur-Yvette, France, January 2014.
- 2.Cernay, C., Pelzer, E., Meynard, J.-M. Ben-Ari, T. & Makowski, D. Global comparison of agronomic and environmental performances of grain legumes to grow in Europe. LabEx BASC: Gif-sur-Yvette, France, March 2015.
- 3.Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M. & Makowski, D. Measuring inter-annual yield variability for a diversity of grain legumes across Europe and the Americas. 5<sup>th</sup> International Symposium for Farming Systems Design: Montpellier, France, September 2015.

## Communications de vulgarisation

---

- 1.Finale Université Paris-Saclay, « Ma Thèse en 180 Secondes », 1<sup>er</sup> Prix du Jury, Gif-sur-Yvette, France, Avril 2014.
- 2.Finale Ile-de-France, « Ma Thèse en 180 Secondes », Paris, France, Mai 2014.
- 3.Journées Ecole Doctorale ABIES, 1<sup>er</sup> Prix du Jury, Paris, France, Avril 2015.
- 4.Café-science, « Les légumineuses, aliments du futur ? », Maule, France, Mai 2016.



La thèse a été soutenue par l'Agence Nationale de la Recherche (ANR) dans le cadre du programme « Investissements d'Avenir » (ANR-10-IDEX-0003-02) et le LabEx BASC (ANR-11-LABX-0034).



## Sommaire

Introduction	13
Chapitre 1	37
Chapitre 2	53
Chapitre 3	77
Chapitre 4	109
Discussion	128
Conclusion	173
Références	179
Annexes	195





# Introduction



# Introduction

## 1. Les légumineuses à graines, une diversité d'espèces cultivées à intérêts alimentaire et agroenvironnemental

### 1.1. Une diversité d'espèces sous-exploitée

Les légumineuses sont des espèces végétales appartenant à la famille des Fabaceae (Leguminosae). Cette famille de plantes comprend près de 800 genres et de 20000 espèces, inféodées à une large diversité de conditions climatiques et de types de sols (Smýkal et al., 2015 ;Yahara et al., 2013). Ces espèces peuvent être cultivées selon des modalités variées (Graham et Vance, 2003 ; Sinclair et Vadez, 2012), en peuplement monospécifique ou plurispécifique, et récoltées pour la production de graines ou de fourrages. Certaines espèces de légumineuses peuvent être cultivées sans être récoltées, en tant que couverts ou cultures intermédiaires. Bien qu'il existe plusieurs modalités de culture des légumineuses à l'échelle mondiale, la thèse ne concerne que des espèces à cycle de croissance et de développement annuel, en peuplement monospécifique, et récoltées pour l'utilisation de leurs graines à maturité physiologique. La production mondiale des légumineuses à graines est limitée par rapport à celles des céréales. En 2014, leur production mondiale ne représentait que 17% de celle du blé (*Triticum aestivum*), du maïs (*Zea mays*), et du riz (*Oryza sativa*), contre 14% en 1961 (FAOSTAT, 2016). La production mondiale des légumineuses à graines représente un faible nombre d'espèces d'intérêt économique significatif (Akibode et Maredia, 2011 ; Foyer et al., 2016). En 2014, la production mondiale de soja (*Glycine max*) représentait 73% de la production totale de légumineuses à graines, contre 31% en 1961 (FAOSTAT, 2016). Entre 1961 et 2014, le soja (*Glycine max*) représentait, en moyenne annuelle, 75% et 90% des surfaces de légumineuses à graines en Amérique du Sud et du Nord, respectivement (FAOSTAT, 2016). Durant la même période, 70%, 76%, 76% et 78% des surfaces des légumineuses à graines étaient récoltées, en moyenne annuelle, pour trois ou quatre espèces en Europe, Océanie, Asie, et Afrique, respectivement (FAOSTAT, 2016).

### 1.2. Des cultures à intérêt alimentaire

Les graines des légumineuses sont une source d'acides aminés indispensables (e.g., la lysine) et d'énergie en alimentation animale (Duranti, 2006 ; Vaz Patto et al., 2015). Chez les monogastriques (e.g., porcs et volailles), les graines de céréales constituent la principale source d'énergie de la ration alimentaire (Jezierny et al., 2010 ; Martens et al., 2012). Mais, du fait de leur faible teneur protéique, elles doivent être complétées par une source riche en protéines telles que celles issues des graines de légumineuses (Jezierny et al., 2010 ; Martens et al., 2012). Chez les ruminants, les fourrages constituent la base de la ration alimentaire. Cependant, la complémentation par une source de protéines – par exemple, issues des graines de légumineuses – est nécessaire, particulièrement dans les rations à base d'ensilage de maïs (Doyle et al., 2005). Chez les poissons, les légumineuses à graines peuvent également représenter des sources de protéines intéressantes d'origine végétale, en tant que substitut partiel aux farines d'origine animale (Gatlin et al., 2007).

La contribution des légumineuses à graines à la sécurité alimentaire des pays du Sud est significative (Akibode et Maredia, 2011 ; Foyer et al., 2016 ; Nedumaran et al., 2013). En effet, les graines des légumineuses contiennent – pour la majorité des espèces – des protéines de qualité pour l'alimentation humaine. Elles présentent une faible teneur en matière grasses, une richesse en fibres, et un faible indice glycémique (Duranti, 2006 ; Tharanathan et Mahadevamma, 2003 ; Vaz Patto et al., 2015). Parce qu'elles sont pauvres en acides aminés soufrés et en tryptophane, les graines des légumineuses sont souvent utilisées en complément de celles des céréales (Temba et al., 2016). Les graines des légumineuses ont des propriétés nutritionnelles qui varient en fonction de l'espèce (Duranti, 2006 ; Koivunen et al., 2016 ; Tharanathan et Mahadevamma, 2003). Elles peuvent partiellement se substituer aux protéines d'origine animale dans les régimes alimentaires (Bassett et al., 2010 ; Davis et al., 2010 ; Multari et al., 2015). De récentes études cliniques soulignent les effets favorables d'une consommation régulière de légumineuses à graines sur la santé (Bassett et al., 2010 ; Mudryj et al., 2014). L'industrie agroalimentaire peut utiliser les graines des légumineuses comme des ingrédients technofonctionnels (Asif et al., 2013). Bien que certains facteurs nutritionnels indésirables dans les graines diminuent la digestibilité des protéines des légumineuses, des traitements technologiques peuvent les inactiver (Gilani et al., 2005). Plusieurs produits dérivés de graines de légumineuses sont actuellement utilisés sous forme de farines incorporées dans des pâtes alimentaires ou de concentras et d'isolats dans des boissons (Asif et al., 2013 ; Laleg et al., 2016), alternatives aux boissons lactées ou produits carnés.

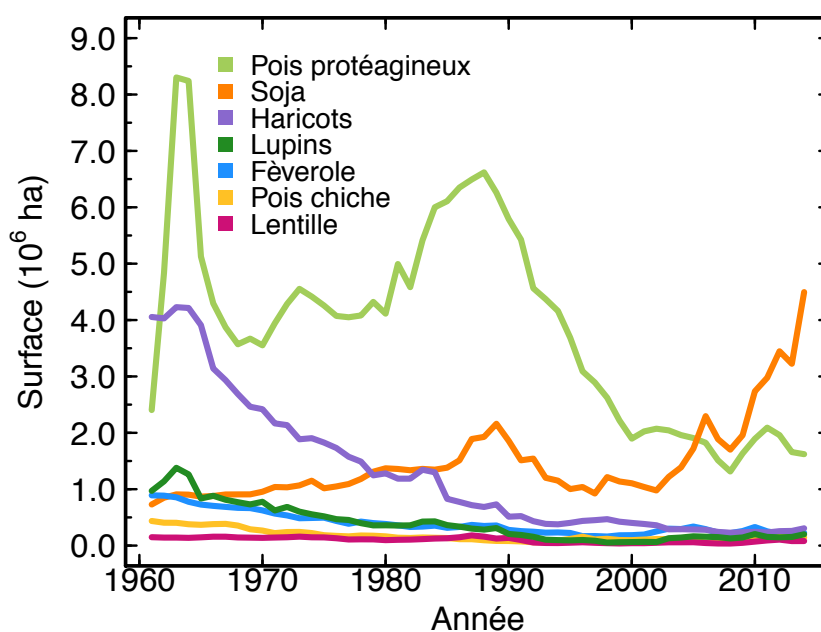
### 1.3. Des cultures à intérêt agroenvironnemental

A l'exception de quelques sous-familles botaniques proches, les légumineuses sont les seules espèces capables de fixer symbiotiquement l'azote atmosphérique ( $N_2$ ) (Graham et Vance, 2003 ; Sinclair et Vadez, 2012). La fixation de  $N_2$  est réalisée par une symbiose des bactéries *Rhizobium* ou *Bradyrhizobium* dans des nodosités différenciées sur le système racinaire de la plante. L'efficacité de la fixation symbiotique varie en fonction de l'espèce de légumineuses (Hardarson et Atkins, 2003). L'azote fixé peut bénéficier non seulement à la légumineuse, mais également à la culture qui la suit dans la séquence culturale (Chalk, 1998). En effet, une culture de légumineuses requiert peu ou pas de fertilisation azotée (Peoples et al., 2009). Plusieurs expérimentations agronomiques ont montré que les rendements des céréales et des oléagineux cultivés après des légumineuses étaient, en général, supérieurs à ceux des céréales cultivées après des céréales ou à ceux des oléagineux cultivés après des oléagineux (Armstrong et al., 1997 ; Badaruddin et Meyer, 1994 ; Jensen et al., 2004 ; Williams et al., 2014). Il existe plusieurs facteurs expliquant cet effet des légumineuses sur les rendements des cultures suivantes. Ces facteurs peuvent être endogènes ou exogènes au cycle de l'azote. Certains auteurs suggèrent que les légumineuses augmentent les concentrations en azote minéral du sol disponibles pour les cultures suivantes, en raison de la minéralisation de leurs résidus de culture ; aussi, les légumineuses absorberaient de moindres quantités d'azote du sol comparées à celles absorbées par des céréales et des oléagineux (Angus et al., 2015 ; Kirkegaard et al., 2008 ; Preissel et al., 2015 ; Seymour et al., 2012). L'effet des légumineuses sur les rendements des cultures suivantes peut également s'expliquer par des facteurs autres que ceux endogènes au cycle de l'azote : l'introduction des cultures de légumineuses permettrait la rupture des cycles des bioagresseurs spécifiques des céréales et des oléagineux ; l'insertion d'une famille botanique différente, le changement des caractéristiques de leurs rhizosphères spécifiques, le décalage des dates de semis, résultant des cultures de légumineuses, faciliteraient la gestion des adventices les plus fréquentes, et limiterait ainsi le recours aux produits phytosanitaires à l'échelle de la séquence culturale (Bennett et al., 2012 ; Crotty et al., 2015 ; Schneider et al., 2015). Cependant, l'effet a priori positif des légumineuses en tant que cultures précédentes semblerait dépendre de la dose de fertilisation d'azote minéral appliquée sur les cultures suivantes (Badaruddin et Meyer, 1994 ; Cheruiyot et al., 2003 ; Dakora et al., 1987 ; Ghosh et Singh, 1996 ; Jensen et al., 2004). Par ailleurs, cet effet n'a pas que des aspects positifs ; la minéralisation de la matière organique et d'autres processus libèrent de l'azote sous des formes labiles lors de la dégradation des résidus de culture de légumineuses. Ces phénomènes peuvent favoriser un risque de fuite d'azote vers l'environnement, en particulier la lixiviation de nitrate ( $NO_3^-$ ) (Peoples et al., 2009). En termes de fuites gazeuses, les cultures de légumineuses émettent moins de protoxyde d'azote ( $N_2O$ ) que les cultures de non-légumineuses fertilisées, et peuvent avoir un effet positif sur les bilans des émissions de gaz à effet de serre (Jeuffroy et al., 2013 ; Schmeer et al., 2014).

## 2. Regards croisés sur les légumineuses à graines en Europe, Amérique du Nord et Océanie

### 2.1. Des niveaux faibles de surface et de production en Europe

Les légumineuses à graines sont actuellement peu récoltées en Europe. En 2014, la surface de légumineuses à graines représentait 2% de la surface européenne récoltée (FAOSTAT, 2016). Les surfaces ont souvent évolué sporadiquement en fonction des mesures de la Politique Agricole Commune et d'autres facteurs connexes (Bues et al., 2013 ; Voisin et al., 2014 ; Zander et al., 2016). Des auteurs ont fait l'hypothèse que la variabilité interannuelle des rendements des légumineuses à graines pourrait être un frein à leur introduction dans les systèmes de culture européens (Bues et al., 2013 ; Reckling et al., 2016 ; Voisin et al., 2014 ; Zander et al., 2016). Les pertes de production issues de la variabilité du rendement en grains d'une année à l'autre rendraient les cultures de légumineuses moins compétitives à l'échelle du cycle cultural, particulièrement en termes de retours économiques escomptés (LMC International, 2009a, 2009b). Cette forte variabilité interannuelle du rendement en grains pourrait en conséquence amener les agriculteurs européens à privilégier des cultures de non-légumineuses (Bues et al., 2013 ; Voisin et al., 2014 ; Zander et al., 2016). Les légumineuses à graines sont également peu produites en Europe, représentant actuellement moins de 2% de leur production mondiale (Schneider et al., 2015). Entre 2008 et 2012, la France produisait, en moyenne annuelle, toutes espèces confondues, 29% de la production européenne de légumineuses à graines, devant l'Italie (17%), le Royaume-Uni (15%) et l'Espagne (8%) (Schneider et al., 2015). La production domestique de légumineuses à graines en Europe est principalement destinée aux productions animales (Bues et al., 2013 ; Martin, 2014 ; Schneider et al., 2015 ; Voisin et al., 2014). Récemment, les exportations et certains segments de marché de faibles volumes mais à forte valeur ajoutée se développent à destination de l'alimentation humaine, particulièrement sous forme d'appellations (Magrini et al., 2016 ; Schneider et al., 2015 ; Voisin et al., 2014 ; Zander et al., 2016). Cependant, la consommation de légumineuses à graines en alimentation humaine reste faible en Europe. Actuellement, 12 et 13% de la production européenne de fêverole (*Vicia faba*) et de pois protéagineux (*Pisum sativum*) sont destinées à l'alimentation humaine en Europe (Bues et al., 2013).



**Figure 1.** Surfaces récoltées ( $10^6$  ha) de différentes espèces de légumineuses à graines en Europe entre 1961 et 2014. Les espèces sont le pois protéagineux (*Pisum sativum*), le soja (*Glycine max*), les haricots (*Phaseolus* spp. et *Vigna* spp.), les lupins (*Lupinus* spp.), la fèverole (*Vicia faba*), le pois chiche (*Cicer arietinum*), et la lentille (*Lens culinaris*). Les données sont issues de la base de données FAOSTAT.

Plusieurs espèces de légumineuses à graines peuvent être cultivées en Europe (Figure 1). Entre 1961 et 2014, le pois protéagineux (*Pisum sativum*) était, en moyenne, la légumineuse à graine européenne la plus récoltée (FAOSTAT, 2016). Plus récemment, les surfaces de soja (*Glycine max*) ont fortement augmenté dans l'Est de l'Europe, en particulier au pourtour du Danube (Krön et Bittner, 2015). En 2014, la surface de soja (*Glycine max*) représentait 56% de la surface européenne de légumineuses à graines (FAOSTAT, 2016). Si le pois protéagineux (*Pisum sativum*) et le soja (*Glycine max*) totalisaient 76% de la surface de légumineuses à graines en Europe en 2014 (FAOSTAT, 2016), les autres espèces de légumineuses à graines étaient caractérisées par des proportions de surface récoltée plus faibles. En 2014, seulement 14% de la surface européenne de légumineuses à graines étaient récoltées en pois chiche (*Cicer arietinum*), lentille (*Lens culinaris*), lupins (*Lupinus* spp.), haricots (*Phaseolus* spp. et *Vigna* spp.), fèverole (*Vicia faba*), et vesce commune (*Vicia sativa*) (FAOSTAT, 2016). Ce déséquilibre entre les surfaces des différentes espèces de légumineuses à graines pourrait être la conséquence de deux phénomènes concomitants. Le premier serait qu'un faible nombre de légumineuses à graines – dont le pois protéagineux (*Pisum sativum*) en France – sont utilisées comme espèces modèles dans la recherche (Gepts et al., 2005 ; Magrini et al., 2016 ; Murphy-Bokern et al., 2014 ; Voisin et al., 2014). Le second phénomène serait que la sélection variétale des légumineuses à graines en Europe est effectuée par un faible nombre d'entreprises privées, travaillant sur un nombre d'espèces restreint (LMC International, 2009a, 2009b ; Murphy-Bokern et al., 2014 ; Schläfke et al., 2014 ; Wiggering et al., 2012).

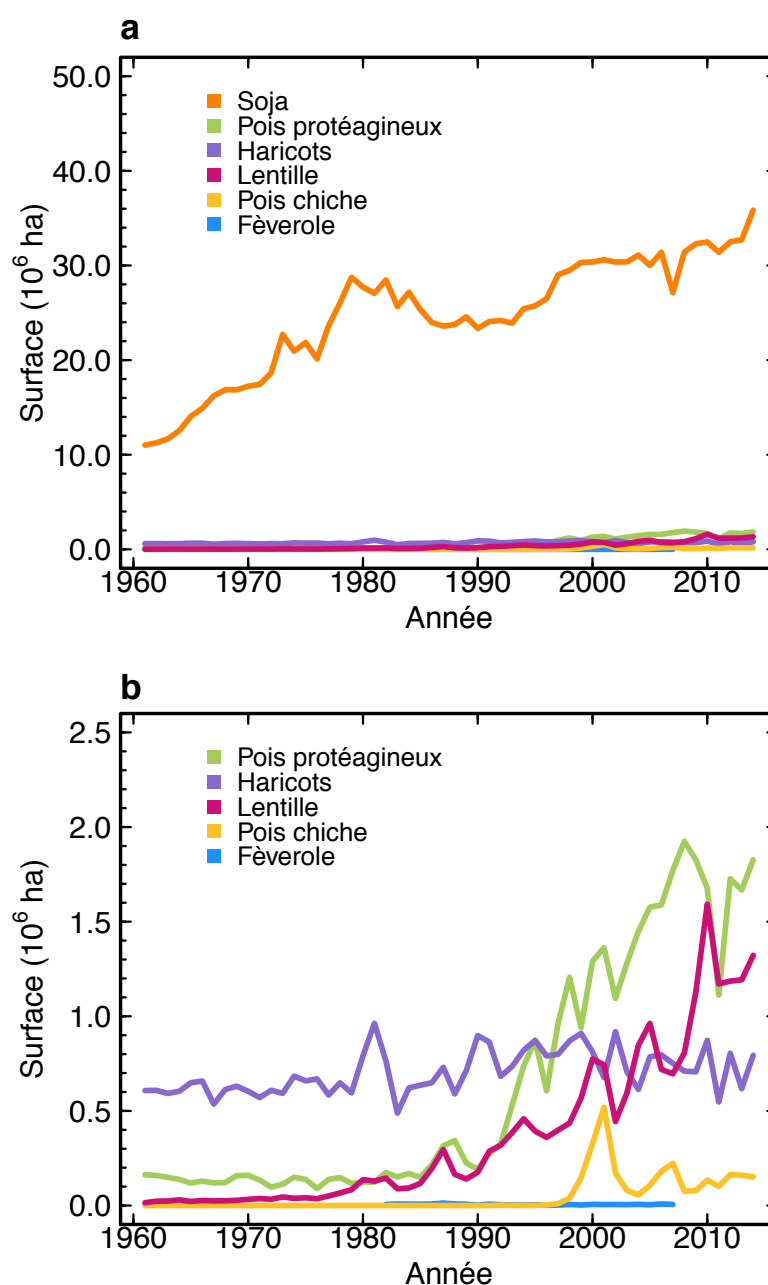


Dans ce contexte, le Parlement Européen a adopté une résolution en 2011 pour augmenter la production de légumineuses à graines en Europe (Haüsling, 2011). Deux principales raisons sont mentionnées. La première est de diminuer (partiellement) les impacts environnementaux négatifs liés à l'intensification des systèmes céréaliers européens en y introduisant une proportion plus élevée de légumineuses à graines (Bues et al., 2013 ; Haüsling, 2011). La seconde raison est de diminuer (partiellement) les importations de soja provenant d'Amérique en les substituant par une production plus élevée de protéines issues de légumineuses à graines, cultivées en Europe (Bues et al., 2013 ; Hache, 2015 ; Haüsling, 2011). Entre 1961 et 2011, l'Europe a importé 63% de sa demande domestique de légumineuses à graines (FAOSTAT, 2014), en raison de sa faible autosuffisance à produire des graines et des tourteaux de soja (*Glycine max*) à destination de l'alimentation de ses filières animales (e.g., son autosuffisance était de 2% entre 2012 et 2013 ; Martin, 2014, 2015 ; European Feed Manufacturers' Federation, 2016). Historiquement, les Etats-Unis présentent des avantages comparatifs pour la production de soja (Hache, 2015 ; Haüsling, 2011). Ces avantages comparatifs ont permis à la production américaine de soja d'être très compétitive sur les marchés internationaux (Bues et al., 2013 ; Hache, 2015 ; Haüsling, 2011 ; Martin, 2014 ; Zander et al., 2016). Dans le cadre de l'Accord Général sur les Tarifs Douaniers et le Commerce (1961-1962), suivi de l'Accord de Blair House (1992), les Etats-Unis ont négocié la libre entrée du soja en Europe (Bues et al., 2013 ; Hache, 2015 ; Haüsling, 2011 ; Martin, 2014 ; Zander et al., 2016). Ces deux facteurs ont permis au soja américain d'être compétitif par rapport aux autres légumineuses sur le marché européen. En conséquence, la production de légumineuses à graines cultivées en Europe est devenue économiquement moins compétitive.

## **2.2. Des niveaux élevés de surface et de production en Amérique du Nord et Océanie**

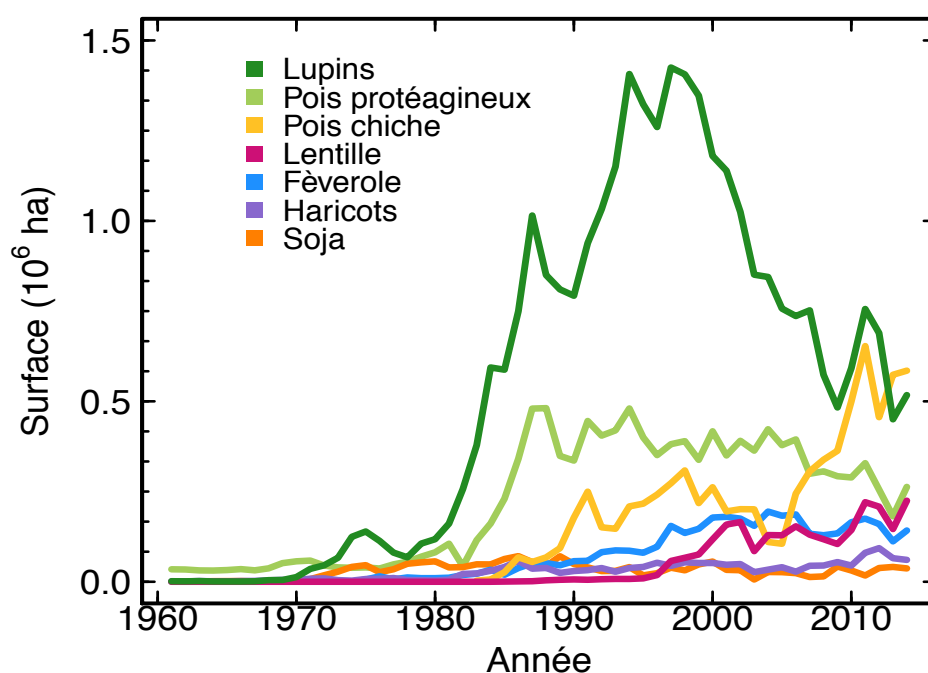
Bien que la gamme des conditions climatiques en Amérique du Nord et Océanie (i.e., définies dans la thèse comme les Etats-Unis et le Canada, et l'Australie et la Nouvelle-Zélande, respectivement ; pour des raisons méthodologiques) soit similaire à celle présente en Europe (Peel et al., 2007), la faible proportion des surfaces européennes de légumineuses à graines contraste avec celles de ces deux régions. En 2014, les surfaces de légumineuses à graines représentaient 4%, 7%, et 13% des surfaces récoltées en Australie, Canada, et Etats-Unis, respectivement (FAOSTAT, 2016). Entre 2003 et 2013, ces deux régions exportaient, en moyenne annuelle, 11 fois plus de quantités de légumineuses à graines que l'Europe (sachant que les quantités produites de légumineuses à graines sont liées aux surfaces récoltées) (FAOSTAT, 2016). Entre 1961 et 2014, le soja (*Glycine max*) était la légumineuse à graines la plus récoltée dans l'agriculture nord-américaine, représentant, en moyenne annuelle, 90% de la surface récoltée en légumineuses à graines aux Etats-Unis et au Canada (FAOSTAT, 2016). Cependant, d'autres espèces de légumineuses, alternatives au soja (*Glycine max*), sont devenues attractives auprès des agriculteurs dans le Nord des Grandes Plaines depuis les années 1990 ;

notamment en tant que source de diversification dans les fréquentes rotations de céréales et de jachères dans cette région (Angadi et al., 2008 ; Gan et al., 2015 ; Miller et al., 2001, 2002 ; Zentner et al., 2002). La surface récoltée de lentille (*Lens culinaris*) a ainsi été multipliée par 3 et 9 aux Etats-Unis et au Canada entre 1990 et 2014 (Figure 2), respectivement (FAOSTAT, 2016) ; celle de pois protéagineux (*Pisum sativum*) a été multipliée par 5 et 12 dans ces deux pays durant la même période (Figure 2), respectivement (FAOSTAT, 2016).



**Figure 2.** Surfaces récoltées (10<sup>6</sup> ha) de différentes espèces de légumineuses à graines en Amérique du Nord entre 1961 et 2014 en (a) incluant, et (b) excluant le soja (*Glycine max*). Les espèces sont le pois protéagineux (*Pisum sativum*), le soja (*Glycine max*), les haricots (*Phaseolus* spp. et *Vigna* spp.), la fève ( *Vicia faba*), le pois chiche (*Cicer arietinum*), et la lentille (*Lens culinaris*). Les données sont issues de la base de données FAOSTAT.

Entre 1961 et 2014, les lupins (*Lupinus* spp.) étaient les légumineuses à graines les plus récoltées dans l'agriculture australienne (Figure 3), représentant, en moyenne annuelle, 37% de la surface de légumineuses à graines (FAOSTAT, 2016). En Australie, les sélectionneurs variétaux, les agronomes et les industriels ont étroitement collaboré depuis les années 1970 et 1980 pour développer la production industrielle d'une large diversité spécifique de légumineuses à graines (Siddique et Sykes, 1997 ; Siddique et al., 2013). De nombreux essais variétaux ont été conduits pour comparer la production d'un grand nombre d'espèces de légumineuses à graines dans plusieurs états australiens (French et Ewing, 1989 ; Gregory, 1998 ; Laurence, 1979 ; Malik et al., 2015 ; Siddique et al., 1993 ; Siddique et Loss, 1999 ; Silsbury, 1990 ; Thomson et al., 1997). Entre 1980 et 2014, la surface récoltée de pois protéagineux (*Pisum sativum*) a ainsi été multipliée par 5 en Australie (Figure 3 ; FAOSTAT, 2016) ; entre 1990 et 2014, celles de pois chiche (*Cicer arietinum*) et de fève (*Vicia faba*) ont été multipliées, chacune, par 3 (Figure 3 ; FAOSTAT, 2016).



**Figure 3.** Surfaces récoltées ( $10^6$  ha) de différentes espèces de légumineuses à graines en Océanie entre 1961 et 2014. Les espèces sont les lupins (*Lupinus* spp.), le pois protéagineux (*Pisum sativum*), le soja (*Glycine max*), les haricots (*Phaseolus* spp. et *Vigna* spp.), la fève (*Vicia faba*), le pois chiche (*Cicer arietinum*), et la lentille (*Lens culinaris*). Les données sont issues de la base de données FAOSTAT.

### 3. Etudes prospectives des légumineuses à graines en Europe

Il existe peu d'études prospectives sur les évolutions possibles des cultures de légumineuses à graines en Europe ; notamment, la faisabilité et les conséquences d'une augmentation de leurs surfaces et de leurs productions respectives.

Dans le cadre du programme européen Legume Futures (2011-2014), Helming et al. (2014) ont analysé les impacts économiques et environnementaux de politiques soutenant la production de légumineuses à graines en Europe d'ici 2020. Les auteurs ont comparé différents scénarios définis à dire d'experts, puis, évalués avec le modèle CAPRI de l'Union Européenne (i.e., « Common Agricultural Policy Regional Impact »), enfin, comparés à un scénario de référence. Le modèle CAPRI est un modèle économique global à équilibre partiel qui calcule les effets de la politique agricole et commerciale de l'Union Européenne sur la production, les revenus, les marchés et l'environnement, aussi bien régionalement que mondialement (Britz et al., 2011 ; Jansson et Heckelei, 2011 ; Weiss et Leip, 2012). Les auteurs ont étudié plusieurs trajectoires prospectives telles que l'effet de l'augmentation mondiale de l'utilisation de soja (*Glycine max*) génétiquement modifié ou l'effet de l'augmentation de la proportion de légumineuses incluses dans des surfaces d'intérêt écologique (i.e., « ecological focus area ») ; à partir de 2015, quand les terres agricoles d'une exploitation couvrent plus de quinze hectares, les agriculteurs déclarent une surface correspondant à au moins 5% de leur surface agricole utile comme surface d'intérêt écologique ; parmi lesquelles, les légumineuses ; EUR-Lex, 2016). Le modèle CAPRI a quantifié, en simulant des scénarios à une maille régionale, la proportion de surface récoltée de légumineuses à graines (toutes espèces confondues) ainsi que leurs impacts environnementaux et économiques. Les auteurs ont utilisé de nombreuses sources de données, qualitatives et quantitatives. Ces données ont été analysées à partir d'un grand nombre de variables d'entrée et de paramètres. Cependant, les auteurs n'ont pas distingué les performances productives et environnementales de différentes espèces de légumineuses à graines dans leur méthodologie, que ce soit en termes de rendement en grains et de variabilité interannuelle ou encore d'effet sur les rendements des cultures suivantes. Les niveaux d'incertitude associée ne sont pas explicités dans les résultats. Dans les variables de sortie, les auteurs regroupent toutes les légumineuses à graines ensemble, sans mentionner les espèces ; en fait, une seule est explicitement spécifiée : le soja (*Glycine max*).

L'expertise AgriMonde (Paillard et al., 2010), Röss et al. (2016), et Westhoek et al. (2014), ont proposé des scénarios plus globaux – souvent exploratoires – de l'impact potentiel de différents changements dans les régimes alimentaires sur l'utilisation des terres agricoles. Ces scénarios racontent des trajectoires prospectives pouvant intégrer des légumineuses à graines. Cependant, ces espèces ne représentent qu'un des maillons des histoires explorées. Ces études combinent des analyses

quantitatives ou qualitatives, et intègrent généralement – comme l'étude de Helming et al. (2014) – des modèles économiques et/ou biophysiques. Dans la majorité des cas, les performances des légumineuses ne sont pas ou peu spécifiquement distinguées. Dans l'expertise Agrimonde (Paillard et al., 2010), deux scénarios sont considérés à l'échelle mondiale à l'horizon 2050 : (1) le scénario Agrimonde GO, et (2) le scénario Agrimonde 1. Le scénario Agrimonde GO est un scénario tendanciel d'évolution des régimes alimentaires et des pratiques d'intensification agricole en se basant sur les hypothèses du scénario Global Orchestration du Millennium Ecosystem Assessment (Carpenter et al., 2005). Au contraire, le scénario Agrimonde 1 est un scénario de rupture qui explore un système alimentaire à vocation plus durable, par réduction, en partie, des excès des apports caloriques. Tout d'abord, des bilans alimentaires régionaux ont été quantifiés à partir de la consommation calorique par habitant, des surfaces récoltées, et des rendements de différentes cultures (i.e., légumineuses, et non-légumineuses). Une fois les hypothèses quantitatives formulées, le modèle de bilan de biomasse Agribiom a permis de tester si les bilans entre les ressources et les emplois sont excédentaires ou déficitaires, dans différentes grandes régions du monde. Dans cette prospective, les légumineuses à graines ne sont qu'un des éléments narratifs des scénarios testés. Par ailleurs, les performances productives des légumineuses à graines ne sont pas comparées entre elles. Dans l'étude de Rööös et al. (2016), les auteurs ont quantifié les besoins théoriques minimaux de terres agricoles ainsi que les émissions associées de gaz à effet de serre afin d'assurer l'autosuffisance alimentaire en Europe d'ici 2050. Ils ont analysé différents scénarios d'approvisionnement alimentaire, explorant plusieurs combinaisons de productions animales et productions végétales. Néanmoins, les légumineuses à graines ne sont pas distinguées par espèce, et n'intègrent que (très) partiellement l'histoire de chaque scénario. Dans l'étude de Westhoek et al. (2014), les auteurs ont examiné les conséquences de substituer 25% et 50% des produits d'origine animale avec des produits d'origine végétale à l'échelle de l'Union Européenne. En utilisant plusieurs modèles globaux dont le modèle CAPRI, les auteurs ont testé l'impact de différents régimes alimentaires alternatifs au régime actuel. Certains de ces régimes sont fortement composés de protéines issues de légumineuses à graines. Cependant, les légumineuses à graines ne sont évoquées que par leur fonction d'apport protéique dans l'alimentation. Dans cette étude, comme dans celle de Rööös et al. (2016), les légumineuses à graines ne sont pas spécifiées, et ne représentent qu'une des caractéristiques du récit prospectif.

Dans différents pays européens, certaines études ont évalué les performances agronomiques et environnementales des légumineuses à graines dans des systèmes de culture (Nemecek et al., 2008 ; Preissel et al., 2015 ; Reckling et al., 2016b). Ces performances sont souvent opposées aux performances de systèmes de culture où des légumineuses à graines ne sont pas incluses. Nemecek et al. (2008) ont analysé les impacts agroenvironnementaux de l'introduction de légumineuses à graines dans des systèmes céréaliers en Allemagne, Espagne, France, et Suisse. Les auteurs ont réalisé une analyse de cycle de vie. Ils ont quantifié, par exemple, le potentiel de réchauffement climatique et le

risque de pertes d'azote vers l'environnement pour des systèmes de culture sans et avec des légumineuses à graines. Les auteurs utilisent différentes sources de données, issues d'expérimentations, des littératures scientifique et technique, d'enquêtes, et de dires d'experts. Ils ont mobilisé, en particulier, les expérimentations et l'expertise de l'action concertée européenne GL-Pro (i.e., une collaboration d'instituts techniques de huit pays européens entre 2003 et 2006, à l'issue de laquelle les aires de culture de principales légumineuses à graines cultivées en Europe ont été cartographiées à dires d'experts). Cependant, Nemecek et al. (2008) ont seulement étudié le pois protéagineux (*Pisum sativum*) et la fève (*Vicia faba*), et n'ont pas analysé les performances pour chacune des deux espèces. Ils n'ont ni inclus les variabilités des niveaux de performances, ni l'incertitude associée en fonction de l'espèce de légumineuses. Preissel et al. (2015) ont synthétisé quantitativement les gains de production de céréales et d'oléagineux cultivés après des légumineuses à graines en utilisant les données de 29 expérimentations agronomiques en Europe. Les auteurs ont analysé l'effet du pois protéagineux (*Pisum sativum*), de la fève (*Vicia faba*), des lupins (*Lupinus* spp.), et du soja (*Glycine max*), en tant que cultures précédentes, sur la compétitivité économique de plusieurs séquences culturales composées de céréales et d'oléagineux. Ils ont également analysé l'effet de différentes techniques culturales en tant que covariables – associés à l'effet des légumineuses précédentes – dans l'estimation des marges comptables à l'échelle de la séquence. Cependant, les performances relatives des quatre espèces de légumineuses à graines ne sont pas comparées entre elles. En Suède et en Allemagne, Reckling et al. (2016b) ont utilisé un modèle décisionnel pour évaluer des indicateurs environnementaux et économiques – générés *in silico* – à partir de séquences culturales incluant ou non des légumineuses à graines. Les espèces de légumineuses à graines étaient le pois protéagineux (*Pisum sativum*), la fève (*Vicia faba*), et les lupins (*Lupinus* spp.). Les indicateurs ont ensuite été évalués à dires d'experts. Aussi, les auteurs n'ont pas comparé les performances productives et environnementales des espèces de légumineuses à graines entre elles.

Enfin, certaines études ont distingué explicitement les performances productives par espèce de légumineuses, mais seulement par des approches qualitatives ou semi-quantitatives. En posant l'objectif d'améliorer l'autosuffisance alimentaire des productions d'animaux monogastriques en Europe, en particulier, en augmentant l'offre domestique de protéines végétales dans cette région, Van Krimpen et al. (2013) ont décrit les teneurs en protéines, les aires de culture et les principales contraintes agronomiques de plusieurs espèces de légumineuses à graines – supposées cultivables en Europe. Ces espèces étaient le pois protéagineux (*Pisum sativum*), le soja (*Glycine max*), la fève (*Vicia faba*), le pois chiche (*Cicer arietinum*), et les lupins (*Lupinus* spp.). En posant l'hypothèse que 50% des importations de soja provenant d'Amérique seraient substituées par les productions individuelles de différentes cultures légumineuses ou non-légumineuses en Europe, De Visser et al. (2014) ont comparé les rendements en grains et les surfaces récoltées sous forme de potentiels que le soja (*Glycine max*), le pois protéagineux (*Pisum sativum*), la fève (*Vicia faba*), et les lupins

(*Lupinus* spp.) devraient atteindre pour être économiquement aussi compétitifs que ceux du blé (*Triticum aestivum*), du maïs (*Zea mays*), du colza (*Brassica napus*), et du tournesol (*Helianthus annuus*). Bien que cette étude ait comparé les performances productives de quatre espèces de légumineuses à graines, les auteurs n'ont pas quantifié l'incertitude associée. Le nombre d'espèces de légumineuses étudiées demeure, par ailleurs, relativement limité.

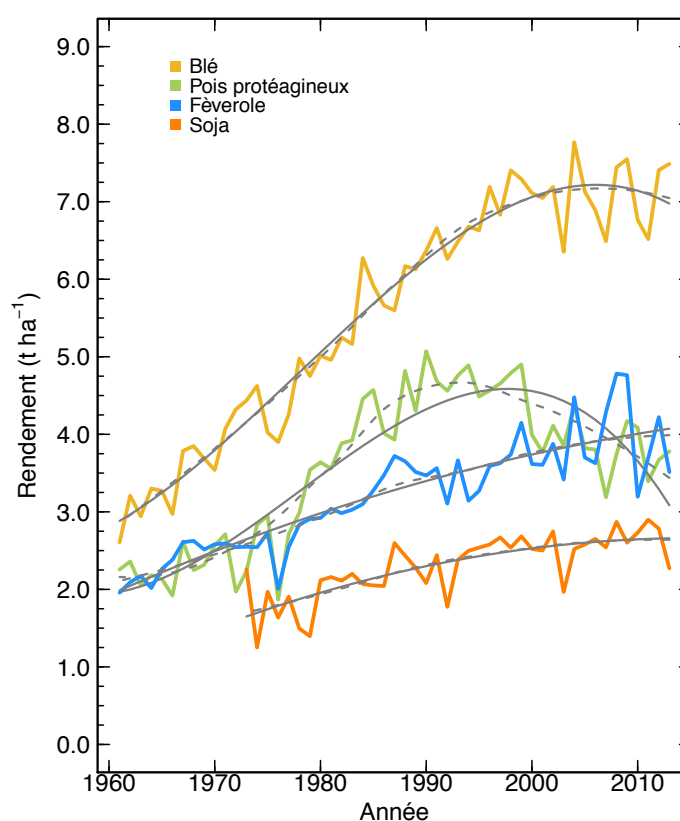
Dans l'ensemble, ces études soulignent le manque de références quantitatives sur les performances productives (et environnementales, dans une moindre mesure) d'une diversité de légumineuses à graines, qu'elles soient cultivées ou non actuellement en Europe. Ces travaux ont relativement peu distingué les espèces selon leurs niveaux de production et leurs effets respectifs sur les productions des cultures suivantes. Par ailleurs, les performances des légumineuses à graines historiquement récoltées en Europe ont très peu été comparées avec celles d'autres espèces de légumineuses à graines, récoltées dans des régions différentes. L'incertitude associée à ces niveaux de performances n'a jamais encore été spécifiquement analysée. Jusqu'à présent, il n'existe pas de comparaisons plurispécifiques concernant les productivités d'un grand nombre d'espèces de légumineuses à graines. Ces comparaisons entre espèces, présentées sous forme de synthèses quantitatives de données à des larges échelles spatiotemporelles, permettraient de proposer une base solide afin de guider des études prospectives sur ce sujet.

## **4. Données disponibles pour estimer les performances des légumineuses à graines à large échelle**

### **4.1. Séries chronologiques de statistique publique**

L'Organisation Mondiale pour l'Agriculture et l'Alimentation (FAO) met à disposition une base de données statistiques, officielles, et publiques (appelée FAOSTAT), pour de nombreuses espèces récoltées, légumineuses et non-légumineuses. Ces données sont sous forme de séries chronologiques de rendement, de surface, et de production, dans 245 pays, et 35 zones régionales, entre 1961 et 2014 (Figures 4 et 5). Les cultures sont catégorisées selon une nomenclature basée sur leurs usages alimentaires. Les légumineuses à graines sont regroupées en 17 catégories culturelles. Si certaines légumineuses à graines ont une catégorie spécifique parce qu'elles sont principalement récoltées dans certaines régions du monde, d'autres espèces sont souvent regroupées dans une seule catégorie parce qu'elles sont moins récoltées. Les données de la base FAOSTAT sont estimées à partir de données préexistantes, collectées elles-mêmes à partir de sources différentes (e.g., certaines données FAOSTAT relatives à l'Union Européenne peuvent provenir de la base de données statistiques officielles de la Commission Européenne (EUROSTAT)).

**Figure 4.** Extraction de données de rendement (en grains) (« Yield ») et de surface récoltée (« Area Harvested ») pour du pois protéagineux (*Pisum sativum*, « Peas, dry ») en Europe en 2014, depuis la base FAOSTAT.



**Figure 5.** Séries chronologiques de rendement de blé (*Triticum aestivum*, ligne jaune continue), pois protéagineux (*Pisum sativum*, ligne verte continue), fèverole (*Vicia faba*, ligne bleue continue) et soja (*Glycine max*, ligne orange continue) en Europe de l'Ouest entre 1961 et 2013. Les données de rendement sont issues de la base de données FAOSTAT. Deux modèles de régression sont ajustés sur les série chronologiques de chaque culture : une régression polynomiale (ligne grise continue) et une régression locale (ligne grise pointillée).



## 4.2. Données issues d'expérimentations agronomiques

De nombreuses expérimentations agronomiques ont comparé les performances productives et environnementales de diverses espèces de légumineuses à graines dans différentes régions du monde. Ces données comparent les performances des légumineuses à graines dans des conditions contextualisées localement par des facteurs environnementaux et des techniques culturales. Lorsque les données sont publiées, elles peuvent être extraites d'articles scientifiques ou techniques (Figure 6). Les données les plus fréquentes sont les rendements en grains (Figures 6 et 7), mais d'autres variables sont également parfois renseignées comme la biomasse aérienne (Figure 7), les quantités (teneurs) en azote de la culture, les quantités (teneurs) en azote dans le sol avant le semis ou après la récolte, ou l'utilisation en eau. Dans les articles scientifiques, ces données sont renseignées dans des tableaux (Figure 6) ou des figures (Figure 7). Les données renseignées dans des tableaux peuvent être extraites manuellement. Les données renseignées dans des figures peuvent être extraites par digitalisation en utilisant un logiciel spécifique (e.g., le logiciel libre d'accès WebPlotDigitizer). Quand la donnée est digitalisée, elle est associée à une erreur liée à la qualité de l'image digitalisée. Cette erreur est souvent inférieure de 5%. Cependant, la qualité de l'image est parfois fortement pixélisée pour des publications parues avant 1990. En conséquence, la précision de la donnée digitalisée est réduite.

**Table 3. Grain yield and number of pods per square meter by crop species and tillage system (CT, conventional tillage; NT, no tillage) in the four growing seasons.**

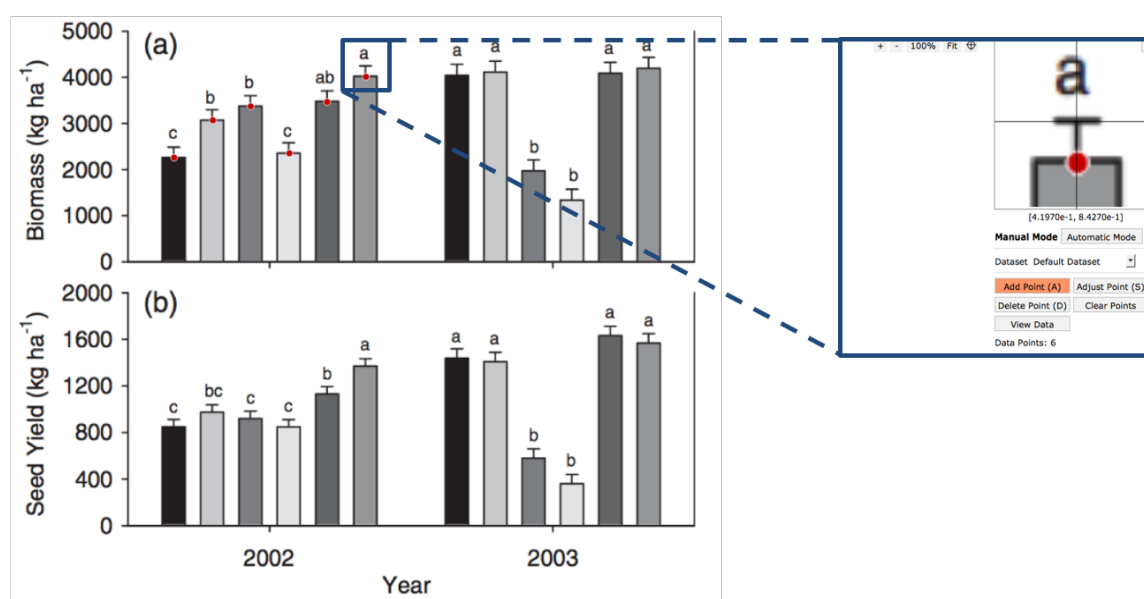
Year	Species	Grain yield			Pods		
		CT	NT	Significance level	CT	NT	Significance level
		Mg ha <sup>-1</sup>			no. m <sup>-2</sup>		
1	chickpea	2.89	2.95	ns†	867	916	ns
	faba bean	3.71	4.01	ns	239	262	ns
	lentil	1.37	1.47	ns	2104	2368	ns
	pea	4.20	4.78	**	397	467	*
2	chickpea	1.42	2.17	***	467	658	**
	faba bean	1.65	1.97	ns	126	138	ns
	lentil	0.66	0.91	ns	996	1309	*
	pea	3.18	4.94	***	302	420	***
3	chickpea	2.69	2.60	ns	785	808	ns
	faba bean	3.17	3.37	ns	155	167	ns
	lentil	1.85	2.18	ns	2536	3072	*
	pea	3.65	3.84	ns	462	518	ns
4	chickpea	3.06	2.84	ns	898	844	ns
	faba bean	2.78	2.45	ns	185	196	ns
	lentil	1.13	0.99	ns	1676	1367	ns
	pea	2.80	2.39	ns	354	320	ns

\* Significant at the 0.05 probability level.  
 \*\* Significant at the 0.01 probability level.  
 \*\*\* Significant at the 0.001 probability level.  
 † ns, not significant.

**Figure 6.** Tableau extrait de l'article de Ruisi et al. (2012). Les données de rendement en grains sont présentées pour différentes espèces de légumineuses à graines cultivées dans un site expérimental durant quatre saisons culturales pour deux types de travail de sol. Ces données expérimentales sont extraites manuellement.

## WebPlotDigitizer

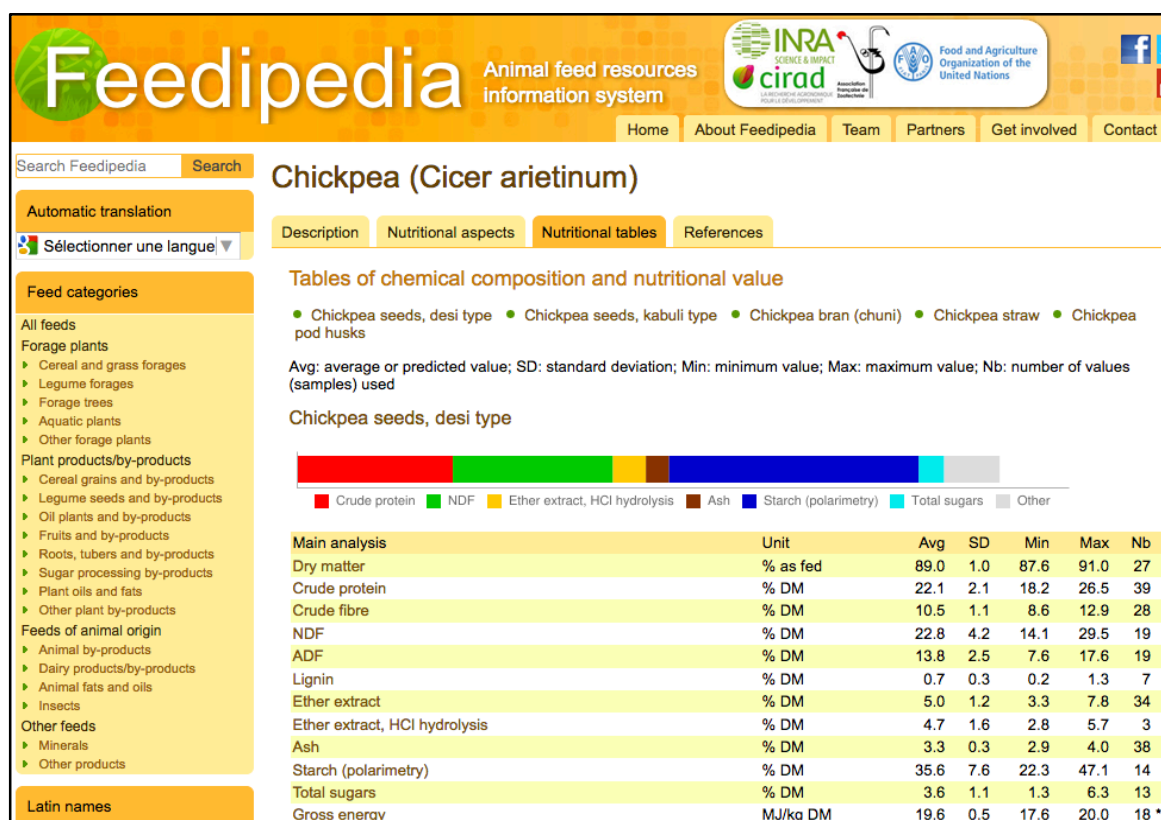
Web based tool to extract data from plots, images, and maps



**Figure 7.** Figure extraite de l'article de Chen et al. (2006). Les données de rendement en grains et de biomasse aérienne sont présentées pour différentes espèces de légumineuses à graines cultivées dans un site expérimental durant deux saisons culturales. Ces données expérimentales sont digitalisées avec le logiciel libre d'accès WebPlotDigitizer. Chaque point rouge indique une donnée digitalisée dont la valeur numérique est extraite.

### 4.3. Données de tables nutritionnelles

Des bases de données libres d'accès contiennent des informations sur les performances productives des légumineuses à graines. Le système d'information FEEDIPEDIA en est un exemple. Cette base de données inclut des informations quantitatives sur les propriétés nutritionnelles d'environ 1400 ingrédients utilisés en alimentation animale, dont des espèces de légumineuses à graines. Parmi les propriétés présentées sous forme de tables nutritionnelles, la teneur moyenne en protéines et la quantité moyenne d'énergie brute dans les graines sont disponibles par espèce de légumineuse (Figure 8). L'incertitude associée à ces valeurs moyennes est également disponible. Ces données sont issues de la littérature scientifique et technique, nationales ou et internationales.



**Figure 8.** Table nutritionnelle extraite de la base de données FEEDIPEDIA. Les données de teneur moyenne en protéines et de quantité d'énergie brute moyenne dans les grains du pois chiche (*Cicer arietinum*, type Desi) sont présentées ainsi que l'incertitude associée.

## 5. Méthodes statistiques appliquées à des données collectées à large échelle

### 5.1. Analyse de séries chronologiques

L'étude des séries chronologiques est utilisée dans de nombreux domaines, aussi bien en économétrie, démographie et agronomie (Ben-Ari et Makowski, 2014 ; Michel et Makowski, 2013). Il s'agit de s'intéresser à l'évolution au cours du temps d'un phénomène dans le but de le décrire, de l'expliquer, et de le prévoir. Des observations sont collectées à des dates différentes. Elles forment une distribution de valeurs numériques indicées par le temps. Elles peuvent être collectées à différentes échelles : de la parcelle à l'exploitation agricole, du pays à la région, du continent au monde. Les rendements renseignés dans la base de données FAOSTAT constituent des exemples de séries chronologiques à un pas de temps annuel. L'analyse d'une série chronologique représente une synthèse quantitative de données à large échelle. Une série chronologique est considérée comme la résultante de différentes composantes : une tendance, une composante résiduelle, et parfois, une composante saisonnière. La tendance représente l'évolution à long terme de la série chronologique étudiée (Figure 5). Cette

composante traduit un comportement moyen de la série. La composante résiduelle correspond à des variations irrégulières de différentes amplitudes autour de la tendance. Pour estimer ces composantes, il est possible d'ajuster des modèles statistiques aux séries chronologiques considérées (Brockwell et Davis, 2002). Différents modèles statistiques peuvent être utilisés, paramétriques ou non-paramétriques. Quand la série chronologique est composée d'une tendance significative, son espérance à une date donnée n'est pas constante et évolue dans le temps. Il est alors nécessaire de l'estimer puis de l'éliminer afin de pouvoir estimer correctement certaines propriétés de la série chronologique, en particulier sa variance. Les séries chronologiques de rendement extraites de la base de données FAOSTAT présentent souvent de fortes tendances (Figure 5) (Ben-Ari et Makowski, 2014 ; Michel et Makowski, 2013). Ces tendances peuvent avoir des origines différentes comme le progrès technologique, le changement climatique, et l'évolution tendancielle de la fréquence de certaines maladies dans le temps. Si la variance interannuelle des rendements est estimée à partir de la série de données brutes (i.e., sans élimination de la tendance), cette variance risque alors d'être fortement surestimée. Elle représentera à la fois l'accroissement tendanciel du rendement et sa variabilité interannuelle.

### 5.2. Méta-analyse

Les revues systématiques sont fréquemment utilisées pour résumer des études autour d'une question de recherche précise (Evans et Foster, 2011). Elles représentent une approche scientifique rigoureuse intégrant la collection, l'évaluation, et la synthèse de l'ensemble des études à propos d'une question de recherche (Bland et al., 1995). Ces revues informent sur ce qui est connu et sur ce qui n'est pas connu à propos d'un sujet étudié (Yuan et Hunt, 2009). Les revues systématiques associées à une analyse statistique sont généralement appelées des méta-analyses (Gurevitch et Hedges, 1993). Dans une méta-analyse, une analyse statistique est appliquée à un jeu de données issues d'articles scientifiques généralement publiés. La méta-analyse représente une synthèse quantitative de données à large échelle. Elle comporte souvent les étapes suivantes (Borenstein et al., 2009) : (1) la définition de l'objectif, (2) la présentation de la recherche de la littérature et/ou du jeu de données incluant les valeurs de la variable de réponse, (3) l'analyse de la qualité des données (e.g., la précision de la variable de réponse), (4) l'évaluation de la variabilité entre les articles sélectionnés, (5) l'évaluation du biais de publication, et (6) la présentation des résultats et des niveaux d'incertitude associée. Philibert et al. (2012) ont proposé huit critères pour évaluer la qualité d'une méta-analyse : (1) la présence d'une procédure reproductible de la sélection des articles issus de la recherche de la littérature, (2) la présence de la liste des références des articles sélectionnés, (3) l'analyse des origines de la variabilité des résultats, (4) l'analyse de la sensibilité des conclusions aux observations ou aux méthodes, (5) l'étude du biais de publication, (6) la pondération des observations en fonction de leurs niveaux de précision dans les modèles statistiques, (7) la disponibilité du jeu de données, et (8) la disponibilité des

codes des analyses statistiques. Doré et al. (2011) et Makowski et al. (2013) ont suggéré l'utilisation plus systématique des méta-analyses en agronomie. Il existe en effet de nombreuses données expérimentales agronomiques disponibles dans des articles publiés. Ces données peuvent être analysées pour répondre à diverses questions d'intérêt agronomique telles que l'estimation de la performance relative des systèmes en agriculture biologique par rapport aux systèmes en agriculture conventionnelle (Seufert et al., 2012), celle des systèmes à bas niveaux d'intrants (Hossard et al., 2016), et celle de l'agriculture de conservation (Pittelkow et al., 2015). La méta-analyse de Laurent et al. (2015) présente un intérêt particulier dans le contexte de cette thèse. Les auteurs comparent les rendements de 36 espèces lignocellulosiques cultivées pour la production d'énergie et hiérarchisent leurs niveaux de production. Ils utilisent les données extraites de 28 articles publiés pour estimer les ratios de rendement de différentes espèces par rapport à une espèce de référence. Si ce ratio est supérieur à un pour une espèce donnée, alors l'espèce comparée est plus productive que l'espèce de référence. Dans le cas contraire, l'espèce de référence est plus productive que l'espèce comparée. Le modèle statistique utilisé par Laurent et al. (2015) est un modèle linéaire à effets aléatoires. Ce modèle permet d'estimer les ratios de rendement en tenant compte à la fois de la variabilité entre les expérimentations, et de la précision des observations dans chaque expérimentation. Ce modèle permet également d'estimer un intervalle de confiance pour chaque ratio d'une espèce comparée à l'espèce de référence, et de déterminer si l'estimation est significativement différente de un ou non. La méthode de Laurent et al. (2015) peut potentiellement être utilisée pour comparer d'autres espèces cultivées, en particulier des légumineuses à graines. Cette méthode doit être appliquée à des expérimentations sur lesquelles au moins deux espèces différentes sont comparées sur les mêmes sites durant sur les mêmes saisons culturales. L'utilisation de ce type d'expérimentation limite les risques de confusion d'effets entre les caractéristiques de l'espèce cultivée, d'une part, et celles du site expérimental et de la saison culturale, d'autre part.

## 6. Problématique et objectif

Les légumineuses à graines sont actuellement peu produites en Europe. Des études récentes soulignent la nécessité d'augmenter la production européenne de ces cultures. D'après ces études, une production domestique plus élevée de légumineuses à graines pourrait non seulement permettre de diminuer partiellement les impacts environnementaux négatifs de la production céréalière en Europe, mais également de diminuer partiellement la dépendance historique de cette région aux importations de soja produit en Amérique du Nord et du Sud, destiné principalement à l'alimentation des productions animales européennes. La question du choix des espèces de légumineuses à graines à privilégier en Europe n'a pas ou peu encore été étudiée. Certaines études prospectives ont été réalisées sur ce sujet, mais elles ont été confrontées à un manque de références quantitatives sur les performances productives (et environnementales, dans une moindre mesure) entre les espèces. Plusieurs sources de

données pourraient être mobilisées pour palier à cette limite : des données statistiques officielles, des données issues d'expérimentations en Europe et dans d'autres régions, et des données sur les propriétés nutritionnelles des légumineuses. L'objectif de ma thèse est d'identifier des espèces de légumineuses à graines potentiellement intéressantes à produire en Europe. Pour répondre à cet objectif, je compare un grand nombre d'espèces en utilisant des critères de production à l'échelle du cycle cultural, mais également leurs effets sur les rendements des céréales qui les suivent à l'échelle de la séquence culturale. Les critères de production étudiés sont le rendement en grains et sa variabilité interannuelle, la biomasse aérienne totale, la quantité d'azote, la quantité de protéines et la teneur en énergie brute dans les graines, et l'utilisation en eau. Les effets sur les rendements des céréales suivantes sont exprimés en fonction de la dose de fertilisation d'azote minéral qui est appliquée sur les céréales. Ces comparaisons plurispécifiques sont réalisées à partir de trois sources de données : (1) des données statistiques officielles à large échelle, (2) des données d'expérimentations situées dans différentes régions du monde, et (3) des données issues de tables nutritionnelles.

## 7. Organisation

Ma thèse est structurée en quatre chapitres rédigés sous forme d'articles scientifiques publiés ou soumis pour publication dans des revues à comité de lecture. Une discussion est ensuite proposée, mettant en perspective les principaux enseignements scientifiques de ce travail.

### 7.1. Comparer la variabilité interannuelle et les pertes de rendement des légumineuses à graines en Europe et Amérique

L'objectif du chapitre 1 est d'estimer la variabilité interannuelle des rendements en grains des principales espèces de légumineuses et non-légumineuses récoltées en Europe et Amérique. Des séries chronologiques de rendement et de surface issues de la base de données FAOSTAT ont été analysées entre 1961 et 2013. Dans ce chapitre, je compare la variabilité interannuelle et les pertes de rendement des cultures de légumineuses à graines à celles des cultures non-légumineuses en Europe. Je tente d'exprimer la variabilité des rendements des légumineuses à graines aux proportions des surfaces récoltées en Europe. Je compare les niveaux de variabilité estimés en Europe à ceux estimés en Amérique du Nord et du Sud – deux régions cultivant d'avantage de légumineuses à graines qu'en Europe. Par conséquent, je pose deux questions de recherche dans ce chapitre :

- (1) Les rendements des cultures de légumineuses à graines sont-ils quantitativement plus variables d'une année sur l'autre que ceux des cultures non-légumineuses en Europe et Amérique ?
- (2) La variabilité interannuelle des rendements des légumineuses à graines peut-elle expliquer leurs faibles surfaces récoltées en Europe ?

## **7.2. Une base de données expérimentales pour comparer les performances productives des légumineuses à graines à l'échelle mondiale**

Si le chapitre 1 caractérise la variabilité de la production des légumineuses à graines à partir de données statistiques officielles, les chapitres suivants caractérisent les niveaux de production de ces espèces et leurs effets sur les cultures suivantes à partir de données expérimentales. L'objectif du chapitre 2 est de construire une base de données expérimentales à l'échelle mondiale pour comparer les productions des différentes espèces de légumineuses à graines et leurs effets sur les cultures qui les suivent. Dans ce chapitre, je décris les différentes étapes de la construction de cette base de données, et je propose des exemples de son utilisation. Présentée sous forme d'un « data paper », la base de données inclut des données extraites de 173 articles publiés couvrant 41 pays et 39 espèces de légumineuses à graines. Elle est utilisée dans les deux méta-analyses présentées dans les chapitres suivants et mise à disposition de la communauté scientifique.

## **7.3. Comparer les niveaux de production des légumineuses à graines en Europe, Amérique du Nord et Océanie**

L'objectif du chapitre 3 est de hiérarchiser les niveaux de production d'une large diversité spécifique de légumineuses à graines en Europe, Amérique du Nord et Océanie. Dans les trois régions, les espèces de légumineuses à graines sont directement comparées au pois protéagineux (*Pisum sativum*), défini comme l'espèce de référence. En réalisant une première méta-analyse de la base de données du chapitre 2, les productions des espèces de légumineuses à graines sont hiérarchisées en fonction de sept critères : (1) le rendement en grains et (2) sa variabilité interannuelle, (3) la biomasse aérienne totale, les quantités (4) d'azote, (5) de protéines et (6) d'énergie brute dans les graines, et (7) l'utilisation en eau. Dans ce chapitre, je compare non seulement plusieurs espèces de légumineuses à graines au pois protéagineux (*Pisum sativum*) en Europe – l'espèce de légumineuses à graines la plus récoltée, en moyenne annuelle, entre 1961 et 2014, dans cette région –, mais également des légumineuses à graines, non encore testées en Europe, et caractérisées par des niveaux de production élevés par rapport à ceux du pois protéagineux (*Pisum sativum*) en Amérique du Nord et en Océanie. Les espèces les plus intéressantes sont proposées comme espèces candidates à inclure potentiellement dans de futures expérimentations agronomiques en Europe. Les résultats de la méta-analyse sont utilisés pour estimer les conséquences d'une substitution de la surface de pois protéagineux (*Pisum sativum*) par d'autres espèces de légumineuses en Europe, en tenant compte des incertitudes. Par conséquent, je pose deux questions de recherche dans ce chapitre :

(1) Quelles espèces de légumineuses à graines sont caractérisées par des niveaux plus faibles, similaires ou plus élevés de production que ceux du pois protéagineux (*Pisum sativum*) en Europe ?

(2) Quelles légumineuses à graines caractérisées par des niveaux élevés de production comparés à ceux du pois protéagineux (*Pisum sativum*) en Amérique du Nord et Océanie sont des espèces candidates pour être potentiellement incluses dans de futures expérimentations en Europe ?

### **7.4. Comparer l'effet des légumineuses à graines sur les rendements des céréales suivantes**

L'objectif du chapitre 4 est de comparer l'effet de différentes espèces de légumineuses sur les rendements des céréales qui les suivent à l'échelle de la séquence culturale. En réalisant une seconde méta-analyse de la base de données du chapitre 2, les rendements de céréales cultivées après différentes légumineuses à graines sont comparés aux rendements de céréales cultivées après des céréales. Le rendement d'une céréale cultivée après une céréale est sélectionné comme le rendement de référence. Dans ce chapitre, je compare les gains de rendement induits par les effets des différentes espèces de légumineuses à graines en tant que cultures précédentes. Les gains de rendement sont également estimés en fonction de la dose de fertilisation en azote minéral appliquée sur les céréales suivantes. Les résultats sont utilisés pour quantifier les conséquences d'une augmentation de la proportion d'une espèce de légumineuses à graines cultivée dans une monoculture d'une céréale sur la production céréalière totale, en tenant compte des incertitudes. Par conséquent, je pose trois questions de recherche dans ce chapitre :

- (1) L'effet des légumineuses à graines sur les rendements des céréales suivantes est-il significativement positif et dépend-il de l'espèce de légumineuses à graines en tant que culture précédente ?
- (2) L'effet des légumineuses à graines sur les rendements des céréales suivantes dépend-il de la dose de fertilisation azotée appliquée sur ces céréales ?
- (3) L'effet des légumineuses à graines sur les rendements des céréales suivantes permet-il de compenser une diminution relative attendue de la production céréalière ?





# Chapitre 1



# Chapitre 1

## **Comparer la variabilité interannuelle et les pertes de rendement des légumineuses à graines en Europe et Amérique**

Ce chapitre est un article scientifique publié dans *Scientific Reports*.

Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M. & Makowski, D. Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* **5**, 11171 (2015).

Les informations supplémentaires sont en Annexes.

La correspondance entre les noms communs et les noms scientifiques des espèces est indiquée en Annexes, dans le Tableau Supplémentaire 2.



# SCIENTIFIC REPORTS

OPEN

## Estimating variability in grain legume yields across Europe and the Americas

Received: 08 December 2014  
Accepted: 30 April 2015  
Published: 08 June 2015

Charles Cernay<sup>1</sup>, Tamara Ben-Ari<sup>2,2</sup>, Elise Pelzer<sup>1,2</sup>, Jean-Marc Meynard<sup>3,4</sup> & David Makowski<sup>1,2</sup>

Grain legume production in Europe has recently come under scrutiny. Although legume crops are often promoted to provide environmental services, European farmers tend to turn to non-legume crops. It is assumed that high variability in legume yields explains this aversion, but so far this hypothesis has not been tested. Here, we estimate the variability of major grain legume and non-legume yields in Europe and the Americas from yield time series over 1961–2013. Results show that grain legume yields are significantly more variable than non-legume yields in Europe. These differences are smaller in the Americas. Our results are robust at the level of the statistical methods. In all regions, crops with high yield variability are allocated to less than 1% of cultivated areas. Although the expansion of grain legumes in Europe may be hindered by high yield variability, some species display risk levels compatible with the development of specialized supply chains.

The European Union (EU) recently stressed the importance of increasing its domestic production of grain legume crops<sup>1</sup>. Two reasons were outlined; the first is the necessity to reduce EU's dependency on soybean imports from the Americas, and the second to reduce negative environmental impacts associated with intensive cereal production. The EU's deficit in grain legume production takes its roots in the General Agreement on Tariffs and Trade in 1947 followed by the Blair House Agreement in 1992. These agreements allowed non-taxable protein imports from the Americas. This resulted in a competitive disadvantage for the EU's grain legume production. Since 1974, grain legume cultivated areas sporadically increased following public support measures (i.e., price supports and subsidies from the Common Agricultural Policy). Legumes cultivated areas represented 1.8% of total European agricultural area in 2013 (compared to 14.5% in North America and 25.5% in South America in the same year)<sup>2</sup>. This is despite grain legumes representing a significant source of protein<sup>3–6</sup> and a mean to reduce the reliance of arable cropping systems on synthetic nitrogen fertilizer and pesticides<sup>7–10</sup>.

Producing grain legumes is often viewed as risky by European farmers, who tend to prefer cultivating non-legume species such as cereals, oilseeds and tubers<sup>11–14</sup>. Several authors have in fact hypothesized that a large adoption of legume crops by European farmers is hampered by frequent losses due to high inter-annual yield variability<sup>14–16</sup>. Still, thus far, grain legume yield variability and risks of yield loss have never been quantitatively analyzed. A rigorous analysis of such risks is crucial for decision makers to define and promote efficient agro-environmental policies for grain legumes in Europe. An accurate estimation of yield loss distribution is also a key element for developing insurance systems against yield loss<sup>17–20</sup>. To date, only qualitative information from either country-level surveys or expertise is available on this topic<sup>12–16,21</sup>.

<sup>1</sup>INRA, UMR 211 Agronomie, F-78850, Thiverval-Grignon, France. <sup>2</sup>AgroParisTech, UMR 211 Agronomie, F-78850, Thiverval-Grignon, France. <sup>3</sup>INRA, UMR 1018 Sciences pour l'Action et le Développement: Activités, Produits, Territoires, F-78850, Thiverval-Grignon, France. <sup>4</sup>AgroParisTech, UMR 1018 Sciences pour l'Action et le Développement: Activités, Produits, Territoires, F-78850, Thiverval-Grignon, France. Correspondence and requests for materials should be addressed to C.C. (email: Charles.Cernay@grignon.inra.fr)

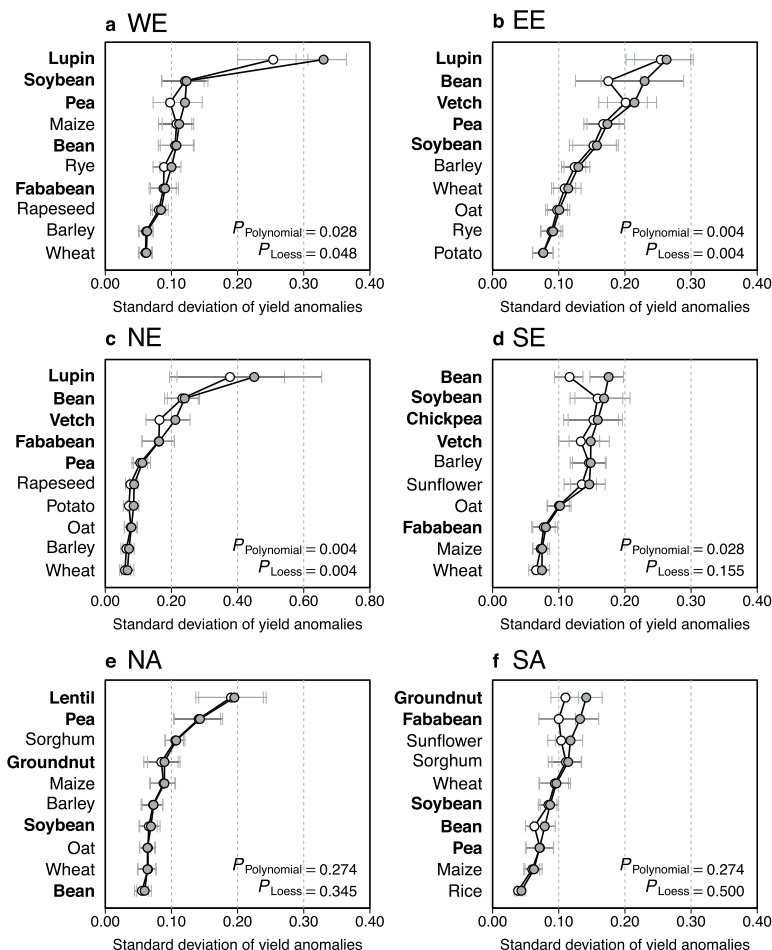
In this study, we quantitatively analyzed inter-annual yield variability and risks of yield loss for the major legume and non-legume yields at the scale of large world regions over 1961–2013. Historical time series were used to compute yield anomalies - defined as normalized yield residuals - for four European and two American regions. Three different risk measures were then estimated from yield anomaly distributions for each species in each region: (i) variance of yield anomalies, (ii) 10<sup>th</sup> percentile of yield anomalies and (iii) expected yield loss. We compared each of these three measures applied to legumes and to non-legumes and classified crops along a risk gradient. We then investigated whether species characterized by higher levels of variability tend to be allocated to smaller proportions of total cultivated areas. Our results reveal that in Europe, the yields of legume crops are generally more variable than non-legumes but that levels of variability vary strongly both between crops and European sub-regions. Importantly, differences between legumes and non-legumes are much smaller in the Americas. These results are robust at the level of the statistical methods used to calculate yield anomalies (polynomial regression or local regression). This study also shows that species characterized by high levels of yield variability are allocated to less than 1% of total cultivated areas (i.e., legumes plus non-legumes). Production of grain legumes in Europe is thus likely to be hindered by high levels of yield variability occurring in this world region. Our analysis however provides a contribution for targeting low risk grain legume species in the prospect of expanding plant protein production in Europe.

## Results

**Grain legume yields are more variable than non-legume yields in Europe.** We analyze yield anomaly distributions for the five most cropped legume and non-legume species for three complementary risk measures in four European regions (see Methods section). A classification of standard deviation of yield anomalies shows that, in Europe, the majority of the five most cropped grain legumes are characterized by higher standard deviation values than the five most cropped non-legume crops (Fig. 1a–d). In all four European regions (i.e., Western, Eastern, Northern and Southern Europe), the three most variable crop yields are grain legumes and, in Eastern and Northern Europe, all five selected legumes have the highest standard deviations in yield. In Southern Europe, four out of five of the most variable crops are grain legumes. Ranks of legume species are overall significantly higher (indicating higher standard deviations) than non-legume ranks in all considered regions ( $P < 0.05$ ), with one exception (Southern Europe with yield anomalies calculated using the local regression method, Fig. 1d). Compared to wheat (the 1<sup>st</sup> European crop in terms of total cultivated area), variance of yield anomalies is systematically and significantly higher for legume species in all European regions, with only one exception (fababean in Southern Europe) (Fig. 2a). In particular, lupin has the highest variance ratio (see Methods section) for all three European regions where it is among the five most cropped legumes (i.e., in Northern, Eastern, and Western Europe); variance of yield anomalies of lupin is 5 to 45 times higher than that of wheat. Bean ranks first in Southern Europe and soybean is the second most variable crop in both Western and Southern Europe. Compared to wheat, variance of soybean yield anomalies is roughly four times higher in Western Europe (almost two times and above five times higher in Eastern and Southern Europe, respectively). Wheat shows the lowest yield standard deviation for all legume and non-legume crops studied in all European regions except Eastern Europe where potato is the least variable crop. Among legume species, the least variable specie is obtained for fababean in Southern and Western Europe, and for soybean and pea in Eastern and Northern Europe respectively. The second risk measure (10<sup>th</sup> percentile of yield anomaly distributions) gives a somewhat equivalent ranking for studied species, with differences in Southern Europe where chickpea, soybean and vetch have smaller 10<sup>th</sup> percentiles (in absolute values) compared to barley and sunflower (Supplementary Fig. 1). Confidence intervals obtained for estimated 10<sup>th</sup> percentiles are larger than those calculated for standard deviations. Species rankings based on this risk measure are thus less reliable. Rankings derived from the third risk measure (expected yield loss, hereby defined as the mean value of yield anomalies below the 10<sup>th</sup> percentile) are almost identical to those derived from yield standard deviations (Supplementary Fig. 2). The only notable difference is observed for Southern Europe bean ranks: 1<sup>st</sup> according to yield standard deviation computed using the polynomial regression technique and 6<sup>th</sup> for expected yield loss. Note, however, that bean is also ranked 6<sup>th</sup> when yield standard deviation is calculated using the local regression. Expected yield losses are greater than 0.25 tons for several legume crops: lupin and soybean in Western Europe, lupin, bean, vetch and fababean in Northern Europe, chickpea, soybean, vetch and bean (both not with local regression) in Southern Europe and all legume species in Eastern Europe (Supplementary Fig. 2). In the above-mentioned cases, expected yield losses are equivalent to more than 25% of expected yield values estimated by fitted yield trends. The highest yield loss values are obtained for lupin; for this species, expected yield losses range from 0.37 to 0.57 tons depending on the region and detrending method used to compute yield anomalies. In comparison, expected yield loss values are much lower for wheat (i.e., 0.09 to 0.20 tons depending upon the region, Supplementary Fig. 2).

**Risk levels of legume and non-legume crops are smaller in the Americas.** In North America, crop species characterized both by the highest and the lowest yield standard deviations are two legume crops, namely lentil and bean (Fig. 1e). In South America, groundnut and sorghum have the highest standard deviations (calculated from polynomial and local regression respectively). In the latter region, the species showing the lowest standard deviation is rice (Fig. 1f). In the Americas, the comparison

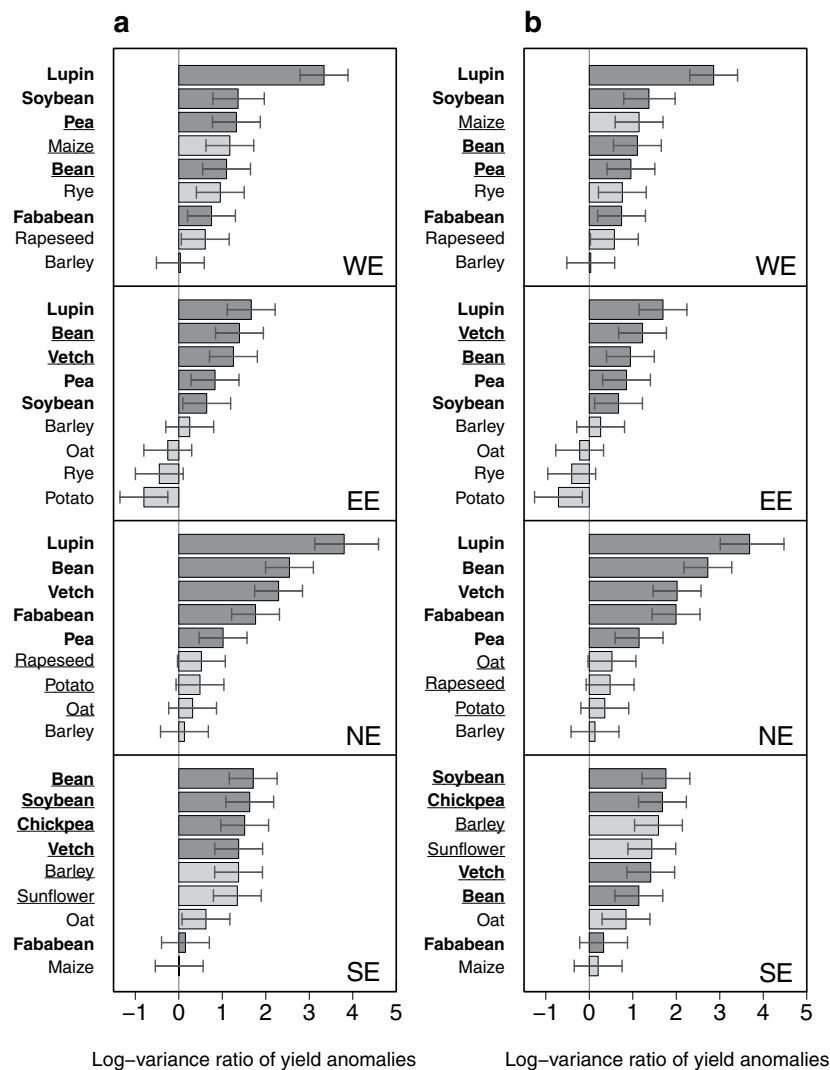
www.nature.com/scientificreports/



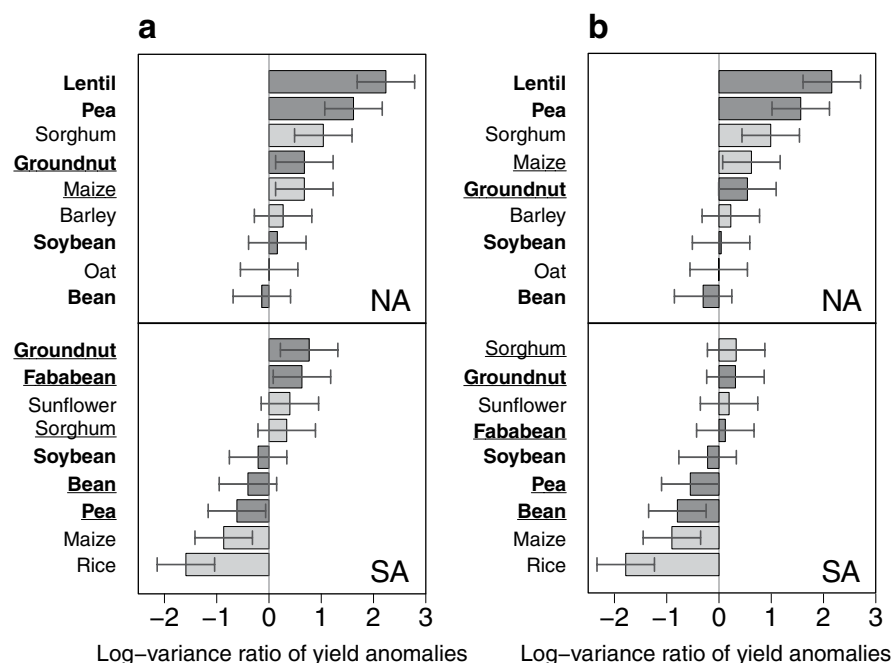
**Figure 1. Standard deviation of yield anomalies for 10 crops in Europe and the Americas over 1961-2013.** Standard deviation of yield anomalies for 10 crops in four European regions (i.e., Western Europe (WE), Eastern Europe (EE), Northern Europe (NE) and Southern Europe (SE)) and two American regions (i.e., North America (NA) and South America (SA)) over 1961-2013. Polynomial (grey points) and local 'loess' (empty points) regression models are used to calculate yield anomalies (relative differences between yield data and yield trend). Horizontal lines correspond to 95% confidence intervals estimated by bootstrap (10,000 samples). Among the 10 crops, 5 are legume crops (bold names) and 5 are non-legume crops (non-bold). All crops are ranked according to standard deviation of yield anomalies calculated using the polynomial model (decreasing order).  $P_{\text{Polynomial}}$  and  $P_{\text{Loess}}$  correspond to the p-value of the Wilcoxon rank test on a hypothesis assuming that ranks of legume and non-legume crops do not differ (against the alternative that legume crop ranking is lower) and are computed using polynomial and loess regression respectively. The number of yield data includes 53 observations for most crops and regions. There are two exceptions: soybean yield data in Western Europe includes 41 observations and lupin yield data in Northern Europe includes 22 observations. Grey vertical dashed lines represent the standard deviation values of 0.10, 0.20 and 0.30 in all regions but 0.20, 0.40 and 0.60 in Northern Europe.



www.nature.com/scientificreports/

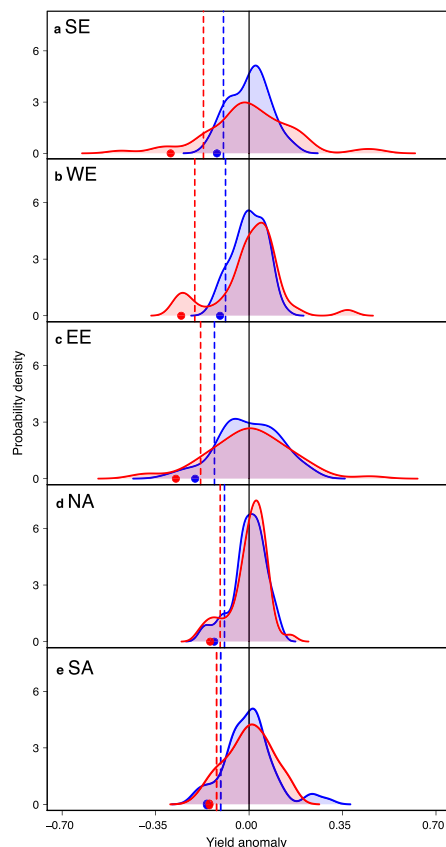


**Figure 2. Log-variance ratio of yield anomalies for 9 crops compared to wheat in four European regions.** Variances are calculated over 1961-2013 and log-transformed in Western Europe (WE), Eastern Europe (EE), Northern Europe (NE) and Southern Europe (SE). Yield anomalies are calculated using both regression models (polynomial (a), loess (b)). Horizontal lines correspond to 95% confidence intervals. Among the 9 crops, 5 are legume crops (bold names and dark grey bars) and 4 are non-legume crops (non-bold names and light grey bars). All crops are ranked according to variance ratio values (decreasing order). Crops that change in ranking between the polynomial and loess regression models used are underlined.



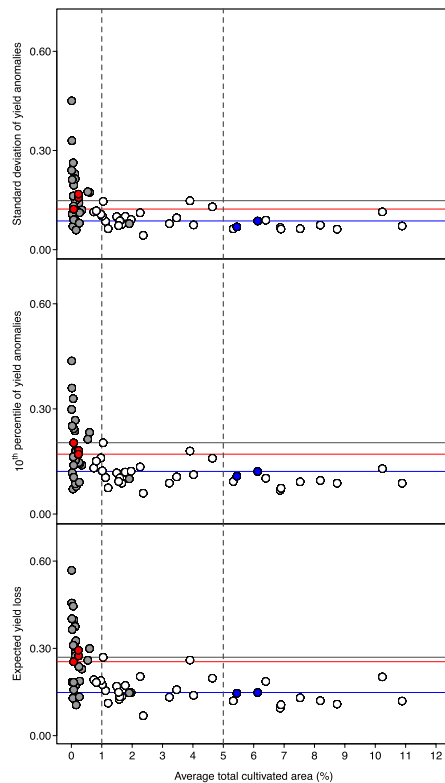
**Figure 3. Log-variance ratio of yield anomalies for 9 crops compared to wheat in two American regions.** Variances are calculated over 1961–2013 and log-transformed in North America (NA) and South America (SA). Yield anomalies are calculated using both regression models (polynomial (a), loess (b)). Horizontal bars correspond to 95% confidence intervals. Among the 9 crops, 5 are legume crops (bold names and dark grey bars) and 4 are non-legume crops (non-bold names and light grey bars). All crops are ranked according to variance ratio values (decreasing order). Crops that change in ranking between the polynomial and loess regression models used are underlined.

between legume and non-legume species is in fact less disadvantageous for legumes. In South (North) America, only two (three) legume species show significantly higher yield variance compared to wheat (lentil, pea and groundnut in North America, and groundnut and fababean in South America, Fig. 3a,b). Importantly, the variance of soybean is close to or not significantly higher than wheat variance in both American regions (Figs. 1e,f and 3a,b). The variance of soybean is also lower than that of maize in North America, but higher in South America (Figs. 1e,f and 3a,b). Pea, maize, and rice show significantly lower yield variance than wheat in South America (Figs. 1f and 3a,b). Overall, whatever the risk measure considered or the detrending method, rankings of grain legumes are never significantly lower than those of non-legumes in either North or South America ( $P > 0.20$ ) (Fig. 1e,f and Supplementary Figs. 1e, f and 2e, f). In fact, in the Americas, soybean risk measures are roughly equal to those calculated for wheat (Figs. 1e,f and 4d,e, Supplementary Figs. 1e,f and 2e,f). Hence, expected losses range from 0.13 to 0.15 tons for soybean, and from 0.13 to 0.16 tons for wheat, depending on the region and the detrending method. The small differences between the estimated risk levels of soybean and wheat are due to the strong similarities in yield anomaly distributions for these two species; in both North and South America, yield anomaly distributions obtained for soybean and wheat are almost the same (Fig. 4d,e). Consequently, soybean and wheat also have very similar values for 10<sup>th</sup> percentile and expected yield losses. In Europe, soybean yield anomaly distributions have longer tails (revealing more extreme yield anomalies) than wheat distributions. This explains the contrasted 10<sup>th</sup> percentile and expected yield loss values observed between soybean and wheat (Fig. 4a–c). Consistent results are obtained using the loess regression method (Supplementary Fig. 3).



**Figure 4. Yield anomaly distributions, 10<sup>th</sup> percentile and expected yield loss for wheat and soybean in Europe and the Americas over 1961-2013.** Probability densities of yield anomalies (curves), 10<sup>th</sup> percentiles of yield anomalies (vertical dashed lines), and expected yield losses (mean values of yield anomalies lower than 10% percentiles; points) for wheat (in blue) and soybean (in red) in three European regions (i.e., Southern Europe (SE), Western Europe (WE) and Eastern Europe (EE)) and two American regions (i.e., North America (NA) and South America (SA)). Yield anomalies (relative differences between yield data and yield trend) are calculated using the polynomial regression models over 1961–2013. Northern Europe is excluded due to insufficient soybean yield data. Regions are ranked according to variance ratio values of soybean yield anomalies compared to wheat (i.e., Southern Europe shows the highest soybean variance ratio compared to wheat). Probability densities are estimated using a Gaussian smoothing kernel.

**Higher volumes of risky crops tend to be grown on small proportions of total cultivated area.** All species\*region combinations with standard deviation above 0.15, with 10<sup>th</sup> percentile above 0.20 or with expected yield loss above 0.27 are grain legumes (Fig. 5). Consistent results are obtained using the loess regression method (Supplementary Fig. 4). The proportions of total cultivated areas reported in Fig. 5 indicate that these species\*region combinations are allocated to less than 1% of regional cultivated area. This is true for all three risk measures. Figure 5 also shows that species\*region combinations allocated to more than 5% of the cultivated area are all characterized by low levels of risk (e.g., soybean in North and South America, Fig. 5). Our results show that species with higher risk levels



**Figure 5. Yield risk measures as a function of the percentage of total cultivated areas in Europe and the Americas over 1961–2013.** Risk measures are calculated over 1961–2013: standard deviation (a), 10<sup>th</sup> percentile (absolute values, b) and expected yield loss (absolute values, c) for all crop\*region combinations. Crop\*region combinations corresponding to legume and non-legume crops are indicated by grey points and empty points, respectively. The percentages reported in the x-axis correspond to the percentage of total cultivated area by a given crop in a given region over 1961–2013. Soybeans grown in the Americas and Europe are indicated in blue and red points. Blue horizontal lines represent the maximal values of risk measures for soybean grown in the Americas. Red horizontal lines represent the minimal values of risk measures for non-legume crops. Grey horizontal lines represent the maximal values of risk measures for non-legume crops. Grey vertical dashed lines represent the thresholds of 1% and 5% of total cultivated area. Yield anomalies are computed using the polynomial regression.

(mostly legumes) are grown on small proportions of total cultivated area, whereas species allocated to high proportions of total cultivated area tend to be characterized by low levels of risk. Hence, there is a consistent pattern in the proportion of risk-cultivated area. However, several species (both legumes and non-legumes) characterized by low risk levels are also allocated to less than 1% of cultivated area. A low level of risk is clearly not a sufficient condition to insure a large proportion of cultivated area.

**Discussion**

Based on European and American historical yield anomalies over the past five decades, we show that yields of legume crops are in general more variable than yields of non-legume crops in Europe, especially in Western, Eastern and Northern Europe. Our results also reveal that the levels of yield variability differ strongly across legume species. In the Americas, legume species also show the highest levels of yield

variability, but the differences between legume and non-legume species are smaller. The trend-removal method (polynomial regression or local regression) used to calculate yield anomalies only marginally affects species rankings but has an effect on estimated values of yield standard deviations: yield anomalies and their standard deviations tend to be smaller when they are estimated using the local regression than when they are estimated with the polynomial regression (Fig. 1). The effect of trend-removal method on estimated yield standard deviations and other risk measures is small in most cases, with the exception of bean in Eastern and Southern Europe. We also investigated the sensitivity of our results to the length of yield time series. All three risk measures were computed over a more recent time period (1983–2013) and the results (Supplementary Figs. 5–8) are, in most cases, very similar to those obtained over the full time period (1961–2013). Reducing the time period increases to some extent lupin yield variance estimates in Western Europe and decreases those of soybean in Southern Europe.

This study is the first quantification of yield risk levels of a large range of legume species compared to non-legume species. Previous studies have used qualitative approaches based on farmers interviews and surveys<sup>12,14,16</sup>. Some studies have addressed variability of inter-annual market prices and incomes<sup>12,15,22,23</sup>. Although our assessment does not take important economic factors (e.g., crop prices, subsidies) into account, our results may help decision makers in defining targeted policies to support growing legumes in Europe. For example, yield insurance systems could be set up in order to compensate high yield losses for risky crops<sup>24–26</sup>. Such systems require accurate estimations of yield anomaly distributions in order to derive reliable risk premiums<sup>17–20</sup>. Our study indicates that, in Europe, risk premiums are likely to be higher for most of the legume species compared to major non-legume species. Note though that the use of aggregated yield data at the scale of large regions may under-estimate yield risk measures derived from less spatially aggregated levels (e.g., from farm data to country data). The larger the scale of yield aggregation, the more likely the effects of factors affecting crop yield may be averaged out, resulting in lower variability and risk of yield loss<sup>27,28</sup>.

The high susceptibility of biological nitrogen fixation – a function specific to legume crops – to both biotic<sup>29</sup> and abiotic stresses<sup>30,31</sup> could be a possible explanation to the observed higher variability of grain legume compared to non-legume yields. However, legume crops do not show the same levels of risk in Europe and in the Americas. Our results show that growing soybean entails lower risks in the Americas than in Europe (Figs. 2, 3 and 4). A broad spectrum of factors may be invoked to explain this difference from the use of specifically adapted varieties in the Americas to the effects of climate variability or pests and diseases. Evidently, there are major differences in cropping systems between the studied regions. The wider use of genetically modified soybean varieties in North and South America is an important feature of these differences and may contribute to the observed risk contrasts, as reported for some non-legume crops<sup>32</sup>. Recent studies outline a comparative lack of breeding investment in Europe to improve grain legume adaptation to local agroclimatic conditions and management techniques<sup>11,16,33,34</sup>, such as crop protection or sowing and harvesting techniques<sup>35–37</sup>. Note that a general lack of specific agronomic references to manage grain legumes may be a barrier for farmers cultivating these crops in Europe<sup>11,16</sup>. Increasing our knowledge about how grain legumes respond to soil and the environment will improve management of these crops across a large diversity of European agroclimatic zones.

We failed to find any clear relationship between the average yield annual growth rates - sometimes proposed as a proxy for technological improvements - and yield variance for legume and non-legume crops (Supplementary Figs. 9 and 10). At the scale of large world regions, it is thus not possible to conclude on a relationship between risk and yield trends. Interestingly, our results indicate that legume species showing high levels of yield variability are always allocated to less than 1% of cultivated areas, suggesting that high yield variability may restrict the expansion of grain legume crops in Europe. Nonetheless, many low-risk crops are cultivated on less than 1% of agricultural areas. Variability in yields is thus only one of the factors behind the variations in cultivated areas. Other external factors may explain this limited expansion, for example, the historical dependence to the Common Agricultural Policy measures<sup>12</sup>; the low level of production of processed products<sup>11,38</sup>; and the competition with soybean imports from the Americas<sup>15</sup> or protein-rich by-products derived from non-legume crops<sup>38,39</sup>. It may also be that European farmers are not incentivized to cultivate these species on soils of good quality in adequate environments, and that they prefer growing more profitable major crops in these environments (e.g., wheat, maize and rapeseed)<sup>11,16,21</sup>.

In average over the 1961–2011 period, Europe imported 63% of its grain legume domestic supply<sup>2</sup>. The European livestock sector dependency on American soybean imports is leading the EU to investigate alternative paths to sustain its meat production<sup>1,38,39</sup>. This could partly be achieved by cropping significantly more legumes in Europe<sup>11,15</sup>, i.e., reducing risk aversion of European farmers for growing legume species. In the coming years, investments in legume breeding and in legume-based cropping system design will be unavoidable if we want to improve the plant protein supply and demand balance in Europe.

## Methods

**Yield data.** Yield time series were retrieved from the Food and Agriculture Organization statistics (FAOSTAT) database<sup>2</sup>. These data had already been used in several studies for analyzing past yield trends and anomalies<sup>40–43</sup>. Crop yield time series were collected from four European regions (i.e., Western, Eastern, Northern and Southern Europe) and for two American regions (i.e., North and South America).

Supplementary Table 1 lists countries belonging to the studied regions and aggregated by the United Nations Statistics Division. In general, the database time interval spans from 1961 to 2013 (i.e.,  $n = 53$  years). There are only two exceptions, for which time intervals are shorter (i.e., soybean in Western Europe from 1973 to 2013 with  $n = 41$  years and lupin in Northern Europe from 1992 to 2013 with  $n = 22$  years). For each region, we selected the five most cropped legume and non-legume species in terms of average harvested area over 1961-2013. A FAOSTAT crop item is defined as a legume when the crop item species belongs to the family of Fabaceae plants, according to the United States Department of Agriculture classification<sup>44</sup>. Concerning grain legumes, we excluded FAOSTAT crop items for which (i) different legume species are pooled into one single crop item (i.e., 'Pulses, nes' and 'String beans' crop items) and (ii) the post-harvest use is for shelling (i.e., 'Beans, green', 'Peas, green' and 'Vegetables, leguminous nes' crop items). Regarding non-legumes, we excluded FAOSTAT crop items that correspond to vegetables and fruits (e.g., 'Grape' and 'Olive' crop items in Southern Europe). For each region, Supplementary Table 2 documents (i) the 10 crop species examined, (ii) the number of observed FAOSTAT yields over 1961-2013 and (iii) the average harvested area in decreasing order.

For each crop and region, a yield trend is estimated using polynomial and local 'loess' regression models fitted to observed data (thereafter, polynomial and loess regression models respectively). The polynomial regression model groups three variants which are linear, quadratic and cubic as follows:

$$Y_t = a + b \times T_t + e_t \tag{1}$$

$$Y_t = a + b \times T_t + c \times T_t^2 + e_t \tag{2}$$

$$Y_t = a + b \times T_t + c \times T_t^2 + d \times T_t^3 + e_t \tag{3}$$

where  $Y_t$  is the yield data in year  $t$  ( $t \text{ ha}^{-1}$ ) and  $t$  is the year index (with  $t = 1$  for the first year of the time series, i.e., 1961). The letters  $a$ ,  $b$ ,  $c$  and  $d$  are parameters of the polynomial yield trend, and  $e_t$  is the residual error which equals the difference between  $Y_t$  and the polynomial trend. The three polynomial variants are systematically fitted to time series using the 'lm' in the R software version 3.0.2<sup>45</sup> and compared with the Akaike Information Criteria (AIC)<sup>46</sup>. The polynomial variant that has the lowest AIC is selected. In addition, we used the loess regression model to assess the suitability of a different calculation method for analyzing yield time series. We used the 'loess' function in the R software with standard arguments (span = 0.75, degree = 2) (Supplementary Fig. 11). The selected polynomial model and the output of the local 'loess' regression are both used to calculate a yield trend further noted  $\hat{Y}_t$ . In a second step, a yield anomaly  $\bar{Y}_t$  is defined as a normalized yield residual, which is obtained by the difference between yield data  $Y_t$  and expected yields  $\hat{Y}_t$  (i.e., yield residual) and then normalized with  $\hat{Y}_t$ :

$$\bar{Y}_t = \frac{Y_t - \hat{Y}_t}{\hat{Y}_t} \tag{4}$$

We thus obtained time series of yield anomalies describing yield variability. Normalizing yield residual enables us to compare crop species and regions with no dimension dependency.

**Estimation of yield risk measures.** To survey yield variability leading to yield loss of legume crops, a set of three classic risk measures<sup>18,19,24,47,48</sup> is evaluated across time series distribution of yield anomalies. The first risk measure is standard deviation expressed as:

$$SD = \sqrt{\frac{1}{T-1} \sum_{i=1}^t \left( \bar{Y}_t - \frac{1}{T} \sum_{i=1}^t \bar{Y}_t \right)^2} \tag{5}$$

Yield variance is the squared value of the standard deviation. The second risk measure is the 10<sup>th</sup> percentile defined by:

$$Pr[\bar{Y}_t \leq q] = 0.10 \tag{6}$$

where  $q$  is the 10<sup>th</sup> percentile of the yield anomalies distribution. It accounts for the value beyond which the yield is considered as a substantial loss. Values of  $q$  are estimated from the yield anomalies using the 'quantile' function in the R software. It would have been difficult to estimate more extreme quantiles (e.g., 5<sup>th</sup> percentile) due to the relatively short yield length of the time series. Supplementary Figure 1 shows that confidence intervals of the estimated 10<sup>th</sup> percentiles are large and the situation would have been even worse with the 5<sup>th</sup> quantile. The third risk measure is the expected yield loss<sup>19,47</sup>. It is equal to the mean of yield anomalies that are lower than the 10<sup>th</sup> percentile  $q$  and is calculated as follows:

$$EYL = E[\bar{Y}_t | \bar{Y}_t < q] \tag{7}$$

It represents the yield loss that is expected to exceed  $q$  with 1 chance of 10. For each of the three risk measures (standard deviation of yield anomalies, 10<sup>th</sup> percentile of yield anomalies, expected yield loss), we calculated the 95% confidence intervals by bootstrap iterations with 10,000 samples. Crop species are then ranked from the highest to the lowest value, using each risk measure in turn. Within a single classification, we then tested the hypothesis that rankings of legume crops are significantly higher than rankings of non-legume crops. This test is performed using the Wilcoxon rank test with the 'wilcox.test' function in the R software.

For each species and each region, we calculated the variance ratio of yield anomalies of a given crop to wheat. We selected wheat as a reference crop in the variance ratio as it is the non-legume crop with the lowest risk measures in three out of four European regions and in North America over the 1961–2013 period. To test whether variance of yield anomalies of a given crop is different from wheat, we used the 'variance.test' function in the R software ( $\alpha = 0.05$ ). The 95% confidence intervals are obtained from the calculations of this function. To appraise the robustness of our calculations of risk measures and of crop rankings, we provided results for a shorter time period that begins in 1983 and ends in 2013 (i.e.,  $n = 31$  years). Note that, however, expected yield loss measure was not provided for this shorter time period due to an insufficient number of yield anomalies below the 10<sup>th</sup> percentile in order to calculate accurately their mean. We calculated the average cultural area of each crop in each region, expressed in percentage of total cultivated area calculated from all FAO items averaged over 1961–2013. We finally calculated the average annual growth rate of yield for each crop in each region over 1961–2013 (i.e.,  $n = 53$  years), expressed in percentage and using the fitted yield trend  $\hat{Y}_t$ , computed from both polynomial and loess regression methods:

$$AAGR = \left( \left( \frac{\hat{Y}_{t=53}}{\hat{Y}_{t=1}} \right)^{\frac{1}{53}} - 1 \right) \times 100 \quad (8)$$

where AAGR is the average annual growth rate of yield expressed in percentage,  $\hat{Y}_{t=1}$  and  $\hat{Y}_{t=53}$  are the yield trends in 1961 and 2013, respectively.

## References

- Häusling, M. The EU protein deficit: what solution for a long-standing problem? (2010/2111(INI)). <http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2011-0026&language=EN> (European Parliament, 2011) Accessed on September 2014.
- Food and Agriculture Organization of the United Nations (FAOSTAT). <http://faostat3.fao.org/home/E> Accessed on September 2014.
- Vaz Panto, M. C. *et al.* Achievements and Challenges in Improving the Nutritional Quality of Food Legumes. *Crit. Rev. Plant Sci.* **34**, 105–143 (2015).
- Jezierny, D., Mosenthin, R. & Bauer, E. The use of grain legumes as a protein source in pig nutrition: A review. *Anim. Feed Sci. Technol.* **157**, 111–128 (2010).
- Davis, J., Sonesson, U., Baumgartner, D. U. & Nemecek, T. Environmental impact of four meals with different protein sources: Case studies in Spain and Sweden. *Food Res. Int.* **43**, 1874–1884 (2010).
- Duranti, M. Grain legume proteins and nutraceutical properties. *FitoTerapia* **77**, 67–82 (2006).
- Preissel, S., Reckling, M., Schläpke, N. & Zander, P. Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: A review. *Field Crops Res.* **175**, 64–79 (2015).
- Nemecek, T. *et al.* Environmental impacts of introducing grain legumes into European crop rotations. *Eur. J. Agron.* **28**, 380–393 (2008).
- Crews, T. E. & Peoples, M. B. Legume versus fertilizer sources of nitrogen: ecological tradeoffs and human needs. *Agric. Ecosyst. Environ.* **102**, 279–297 (2004).
- Drinkwater, L. E., Wagoner, P. & Sarrantonio, M. Legume-based cropping systems have reduced carbon and nitrogen losses. *Nature* **396**, 262–265 (1998).
- Voisin, A.-S. *et al.* Legumes for feed, food, biomaterials and bioenergy in Europe: a review. *Agron. Sustain. Dev.* **34**, 361–380 (2014).
- LMC International. Evaluation of measures applied under the Common Agricultural Policy to the protein crop sector. Main Report. [http://ec.europa.eu/agriculture/eval/reports/protein\\_crops/index\\_en.htm](http://ec.europa.eu/agriculture/eval/reports/protein_crops/index_en.htm) (European Commission, 2009) Accessed on September 2014.
- LMC International. Evaluation of measures applied under the Common Agricultural Policy to the protein crop sector. Case Study Monographs. [http://ec.europa.eu/agriculture/eval/reports/protein\\_crops/index\\_en.htm](http://ec.europa.eu/agriculture/eval/reports/protein_crops/index_en.htm) (European Commission, 2009) Accessed on September 2014.
- Von Röhthofen, J. S. *et al.* What do European farmers think about grain legumes? *Grain Legumes* **45**, 14–15 (2006).
- Bues, A. *et al.* The environmental role of protein crops in the new Common Agricultural Policy. <http://www.legumefutures.de/legumefuturesnews/30052013brussels.html> (European Parliament, 2013) Accessed on June 2014.
- Stoddard, F. L. The case studies of participant expertise in Legume Futures. Legume Futures Report 1.2. <http://www.legumefutures.de/> (Legume Futures, 2013) Accessed on July 2014.
- Sherrick, B. J., Zanini, F. C., Schmitkey, G. D. & Irwin, S. H. Crop Insurance Valuation under Alternative Yield Distributions. *Am. J. Agr. Econ.* **86**, 406–419 (2004).
- Ker, A. P. & Coble, K. Modeling Conditional Yield Densities. *Am. J. Agr. Econ.* **85**, 291–304 (2003).
- Goodwin, B. K. & Ker, A. P. Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *Am. J. Agr. Econ.* **80**, 139–153 (1998).
- Zheng, Q., Wang, H. H. & Shi, Q. H. Estimating bivariate yield distributions and crop insurance premiums using nonparametric methods. *Appl. Econ.* **46**, 2108–2118 (2014).

21. Peltonen-Sainio, P. & Niemi, J. K. Protein crop production at the northern margin of farming: to boost, or not to boost. *Agr. Food Sci. Finland* **21**, 370–383 (2012).
22. Schilizzi, S. G. M. & Kingwell, R. S. Effects of climatic and price uncertainty on the value of legume crops in a Mediterranean-type environment. *Agr. Syst.* **60**, 55–69 (1999).
23. Ghadim, A. K. A., Pannell, D. J. & Burton, M. P. Risk, uncertainty, and learning in adoption of a crop innovation. *Agr. Econ.* **33**, 1–9 (2005).
24. Miranda, M. J. & Glauber, J. W. Systemic risk, reinsurance, and the failure of crop insurance markets. *Am. J. Agr. Econ.* **79**, 206–215 (1997).
25. Antón, J., Cattaneo, A., Kimura, S. & Lankoski, J. Agricultural risk management policies under climate uncertainty. *Global Environ. Chang.* **23**, 1726–1736 (2013).
26. Coble, K. H. & Barnett, B. J. Why Do We Subsidize Crop Insurance? *Am. J. Agr. Econ.* **95**, 498–504 (2013).
27. Popp, M., Rudstrom, M. & Manning, P. Spatial Yield Risk Across Region, Crop and Aggregation Method. *Can. J. Agr. Econ.* **53**, 103–115 (2005).
28. Marra, M. C. & Schurle, B. W. Kansas Wheat Yield Risk Measures and Aggregation: A Meta-Analysis Approach. *J. Agr. Resour. Econ.* **19**, 69–77 (1994).
29. Rubiales, D. *et al.* Achievements and Challenges in Legume Breeding for Pest and Disease Resistance. *Crit. Rev. Plant Sci.* **34**, 195–236 (2015).
30. Araújo, S. S. *et al.* Abiotic Stress Responses in Legumes: Strategies Used to Cope with Environmental Challenges. *Crit. Rev. Plant Sci.* **34**, 237–280 (2015).
31. Aranjuelo, I., Arrese-Igor, C. & Molero, G. Nodule performance within a changing environmental context. *J. Plant Physiol.* **171**, 1076–1090 (2014).
32. Shi, G., Chavas, J.-P. & Lauer, J. Commercialized transgenic traits, maize productivity and yield risk. *Nat. Biotechnol.* **31**, 111–114 (2013).
33. Lizarazo, C. I. *et al.* Nutritive quality and protein production from grain legumes in a boreal climate. *J. Sci. Food Agr.* (2014).
34. Annicchiarico, P. & Iannucci, A. Breeding strategy for faba bean in southern Europe based on cultivar responses across climatically contrasting environments. *Crop Sci.* **48**, 983–991 (2008).
35. Jensen, E. S., Peoples, M. B. & Hauggaard-Nielsen, H. Faba bean in cropping systems. *Field Crop. Res.* **115**, 203–216 (2010).
36. Corre-Hellou, G. & Crozat, Y. N<sub>2</sub> fixation and N supply in organic pea (*Pisum sativum* L.) cropping systems as affected by weeds and pea weevil (*Sitona lineatus* L.). *Eur. J. Agron.* **22**, 449–458 (2005).
37. Flores, F. *et al.* Faba bean adaptation to autumn sowing under European climates. *Agron. Sustain. Dev.* **32**, 727–734 (2012).
38. de Visser, C. L. M., Schreuder, R. & Stoddard, F. The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. *OCL-OL Corps Gras Li.* **21**, D407 (2014).
39. Martin, N. What is the way forward for protein supply? The European perspective. *OCL-OL Corps Gras Li.* **21**, D403 (2014).
40. Michel, L. & Makowski, D. Comparison of Statistical Models for Analyzing Wheat Yield Time Series. *PLoS One* **8**, e78615 (2013).
41. Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918 (2013).
42. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **3**, 1293 (2012).
43. Ben-Ari, T. & Makowski, D. Decomposing global crop yield variability. *Environ. Res. Lett.* **9**, 114011 (2014).
44. United States Department of Agriculture (USDA). Natural Resources Conservation Service. Plants Database. <http://plants.usda.gov> Accessed on May 2014.
45. R. Core Team. R: A language and environment for statistical computing. <http://www.r-project.org> (R Foundation for Statistical Computing, 2013) Accessed on October 2013.
46. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
47. Acerbi, C. & Tasche, D. On the coherence of expected shortfall. *J. Bank. Financ.* **26**, 1487–1503 (2002).
48. Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. Coherent Measures of Risk. *Math. Financ.* **9**, 203–228 (1999).

### Acknowledgments

We thank D. White and T. Parkinson for comments on the manuscript. This work was supported by the French National Research Agency (ANR) under the program Investing in the future Grant n°ANR-10-IDEX-0003-02 and is part of LabEx BASC (ANR-11-LABX-0034).

### Author Contributions

C.C., D.M. and T.B.-A. designed the study. C.C. and D.M. performed the statistical analysis. All authors discussed the results and contributed to writing the paper.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cernay, C. *et al.* Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* **5**, 11171; doi: 10.1038/srep11171 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>





## Chapitre 2



## Chapitre 2

### **Une base de données expérimentales pour comparer les performances productives des légumineuses à graines à l'échelle mondiale**

Ce chapitre est un article scientifique publié dans *Scientific Data*.

Cernay, C., Pelzer, E., & Makowski, D. A global experimental dataset for assessing grain legume production. *Sci. Data* **3**, 160084 (2016).

Le Tableau 3 est en Annexes.

La correspondance entre les noms communs et les noms scientifiques des espèces est indiquée dans le texte.



# SCIENTIFIC DATA

**OPEN****SUBJECT CATEGORIES**

- » Agroecology
- » Environmental sciences
- » Biodiversity
- » Plant ecology

Received: 14 March 2016

Accepted: 17 August 2016

Published: 27 September 2016

## Data Descriptor: A global experimental dataset for assessing grain legume production

Charles Cernay<sup>1</sup>, Elise Pelzer<sup>1</sup> & David Makowski<sup>1</sup>

Grain legume crops are a significant component of the human diet and animal feed and have an important role in the environment, but the global diversity of agricultural legume species is currently underexploited. Experimental assessments of grain legume performances are required, to identify potential species with high yields. Here, we introduce a dataset including results of field experiments published in 173 articles. The selected experiments were carried out over five continents on 39 grain legume species. The dataset includes measurements of grain yield, aerial biomass, crop nitrogen content, residual soil nitrogen content and water use. When available, yields for cereals and oilseeds grown after grain legumes in the crop sequence are also included. The dataset is arranged into a relational database with nine structured tables and 198 standardized attributes. Tillage, fertilization, pest and irrigation management are systematically recorded for each of the 8,581 crop\*field site\*growing season\*treatment combinations. The dataset is freely reusable and easy to update. We anticipate that it will provide valuable information for assessing grain legume production worldwide.

<b>Design Type(s)</b>	database creation objective • data integration objective
<b>Measurement Type(s)</b>	crop production measures
<b>Technology Type(s)</b>	data item extraction from journal article
<b>Factor Type(s)</b>	
<b>Sample Characteristic(s)</b>	Fabaceae

<sup>1</sup>UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France. Correspondence and requests for materials should be addressed to C.C. (email: Charles.Cernay@grignon.inra.fr) or to D.M. (email: David.Makowski@grignon.inra.fr).

www.nature.com/sdata/

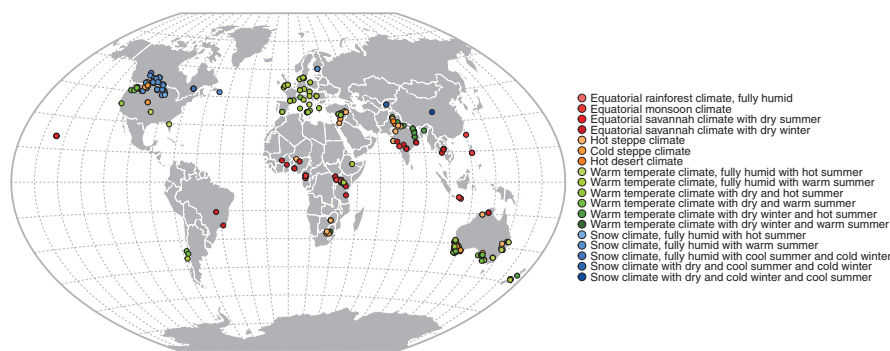
### Background & Summary

The 68th United Nations General Assembly has proclaimed 2016 as the International Year of Pulses. The Food and Agriculture Organization of the United Nations defines 'pulses' as plant species from the Fabaceae family cropped annually, and harvested only for dry grain (hereafter 'grain legume' for unambiguous use<sup>1</sup>). As part of this initiative, grain legumes are being promoted for use as nutritional protein-rich grains, and for their environmental and economic impacts<sup>2-7</sup>. Grain legumes can complement cereals as an affordable source of protein for the human diet<sup>8-10</sup> and for animal feed<sup>11-13</sup>. Through atmospheric nitrogen fixation, grain legumes can significantly increase soil nitrogen supply and the yields of following crops<sup>14-19</sup>. Grain legumes can therefore play a significant role in maintaining global food security and ecosystem resilience.

Fabaceae is one of the largest families of plants worldwide, with 20,000 species growing across a wide range of climatic conditions and soil types<sup>20,21</sup>. Grain legume crops play significant roles in the human diet and animal feed and the environment, but only a fraction of the species in this diverse group of plants is currently exploited in agriculture. From 1961 to 2014, 75 and 90% of the area under legumes was allocated to soybean (*Glycine max*) in South America and North America, respectively<sup>22</sup>. Over the same period, 70, 76 and 78% of the area under legumes was covered by only three species each in Europe, Oceania and Africa: garden pea (*Pisum sativum*), soybean and beans (*Phaseolus* spp. and *Vigna* spp.) in Europe; lupins (*Lupinus* spp.), chickpea (*Cicer arietinum*) and garden pea in Oceania, and groundnut (*Arachis hypogaea*), cowpea (*Vigna unguiculata*) and beans in Africa<sup>22</sup>. In Asia, 76% of the area under legumes was allocated to four species (i.e., soybean, beans, groundnut and chickpea)<sup>22</sup>.

Experimental comparisons of grain legumes can help researchers and decision-makers to identify high-performance species with high yields. Over the last 50 years, many field experiments have assessed the agronomic and environmental performances of grain legumes. These performances vary between field sites and growing seasons, as a function of the climatic conditions and soil types. It would therefore be misleading to draw general conclusions from individual experiments considered separately. A global dataset would provide us with a unique opportunity to analyze variability in grain legume performances across a large spectrum of environmental conditions, and to rank legume species of agricultural and economic interest according to several criteria.

We introduce here a global dataset including the results of field experiments comparing 39 grain legume species grown as sole crops. Most of grain legume species included in the database correspond to species of significant agricultural and economic importance. We have selected only experiments comparing at least two grain legume species grown at the same field site during the same growing season, to prevent any confusion between species characteristics and environmental conditions. We excluded experiments on single grain legume species because, in such experiments, differences between species can be confounded with the effects of environmental factors. Experimental data were extracted from 173 published articles<sup>2-4,6,14-19,23-185</sup>. In total, measurements from 360 field sites were collected across 18 Köppen-Geiger climatic zones<sup>186</sup> in 41 countries (Fig. 1) over five continents (Table 1). The dataset contains 8,581 crop\*field site\*growing season\*treatment combinations. Article references, field site locations, climatic conditions, soil types, yields, crop nitrogen contents, residual soil nitrogen contents



**Figure 1.** Latitude and longitude coordinates of the field sites included in the database. The Köppen-Geiger climatic classification<sup>186</sup> was used to link each field site to a grid size with a resolution of 0.50 degrees of latitude by 0.50 degrees of longitude. Eighteen Köppen-Geiger climatic zones are considered: equatorial climates (red), arid climates (orange), warm temperate climates (green) and snow climates (blue). Within each main Köppen-Geiger climatic zone, each Köppen-Geiger climatic subzone is indicated by a color gradient.

www.nature.com/sdata/

Region	Field site	Field site*growing season	Field site*growing season*treatment	Grain yield	Aerial biomass	Grain nitrogen content	Aerial nitrogen content	Aerial fixed nitrogen content	Residual soil nitrogen content	Water use
Oceania	131 (36.39)	183 (22.45)	2,372 (27.64)	2,324 (28.25)	727 (27.16)	191 (19.47)	107 (12.53)	28 (7.11)	216 (11.71)	142 (16.01)
North America	72 (20.00)	165 (20.25)	2,597 (30.26)	2,524 (30.68)	600 (22.41)	285 (29.05)	178 (20.84)	38 (9.64)	806 (43.69)	474 (53.44)
Asia	65 (18.06)	253 (31.04)	1,475 (17.19)	1,408 (17.11)	598 (22.34)	129 (13.15)	171 (20.02)	87 (22.08)	468 (25.37)	259 (29.20)
Africa	48 (13.33)	101 (12.39)	907 (10.57)	827 (10.05)	243 (9.08)	161 (16.41)	172 (20.14)	145 (36.80)	70 (3.79)	0 (0.00)
Europe	39 (10.83)	102 (12.52)	1,174 (13.68)	1,089 (13.24)	479 (17.89)	181 (18.45)	188 (22.01)	74 (18.78)	255 (13.82)	12 (1.35)
South America	5 (1.39)	11 (1.35)	56 (0.65)	56 (0.68)	30 (1.12)	34 (3.47)	38 (4.45)	22 (5.58)	30 (1.63)	0 (0.00)
Total number (percentage) of available data	360 (100.00)	815 (100.00)	8,581 (100.00)	8,228 (95.89)	2,677 (31.20)	981 (11.43)	854 (9.95)	394 (4.59)	1,845 (21.50)	887 (10.33)
Total number (percentage) of missing data	0 (0.00)	0 (0.00)	0 (0.00)	353 (4.11)	5,904 (68.80)	7,600 (88.57)	7,227 (90.05)	8,187 (95.41)	6,736 (78.50)	7,694 (89.66)
Total number (percentage) of data	360 (100.00)	815 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)	8,581 (100.00)

**Table 1.** Number (percentage) of field sites, field site\*growing season and field site\*growing season\*treatment combinations, and data for grain yield, aerial biomass, grain nitrogen content, aerial nitrogen content, fixed aerial nitrogen content, residual soil nitrogen content and water use, by main world regions. Regions are ranked in descending order of field sites. Grain yield includes data from the 'Crop\_Yield\_Grain' attribute. Aerial biomass includes data from both 'Crop\_Biomass\_Aerial' and 'Crop\_Harvest\_Index' attributes. Grain nitrogen content includes data from both 'Crop\_N\_Quantity\_Grain' and 'Crop\_N\_Percentage\_Grain' attributes. Aerial nitrogen content includes data from the 'Crop\_N\_Quantity\_Aerial', 'Crop\_N\_Percentage\_Aerial' and 'Crop\_N\_Harvest\_Index' attributes. Fixed aerial nitrogen content includes data from both 'Crop\_N\_Fixed\_Quantity\_Aerial' and 'Crop\_N\_Fixed\_Percentage\_Aerial' attributes. Residual soil nitrogen content includes data from both 'Crop\_N\_Soil\_Quantity\_Percentage\_Seeding' and 'Crop\_N\_Soil\_Quantity\_Percentage\_Harvest' attributes. Water use includes data from the 'Crop\_Water\_Use\_Balance', 'Crop\_Water\_Use\_Balance\_Efficiency\_Grain' and 'Crop\_Water\_Use\_Balance\_Efficiency\_Aerial' attributes. The total number (percentage) of available data and the total number (percentage) of missing data are calculated over all considered world regions.

Table	Class attribute	Numerical attribute	Index attribute	Binary attribute	Date attribute	Total
Literature_Search	1	0	1	0	0	2 (1.01)
Article	5	0	2	0	1	8 (4.04)
Site	11	12	2	0	4	29 (14.65)
Crop_Sequence_Trt	3	2	2	1	0	8 (4.04)
Crop	47	41	3	5	10	106 (53.54)
Tillage	7	6	2	4	0	19 (9.60)
Fertilization	4	1	2	0	0	7 (3.54)
Weed_Insect_Fungi	2	1	2	6	2	13 (6.57)
Irrigation	2	1	2	1	0	6 (3.03)
Total	82 (41.41)	64 (32.32)	18 (9.09)	17 (8.59)	17 (8.59)	198 (100.00)

**Table 2.** Number (percentage) of attribute types included in the nine tables of the database. Tables are presented according to the cascade path of the database.

and management practices are systematically recorded for each crop\*field site\*growing season\*treatment combination. When available, data on non-legume species grown at the same field site during the same growing season than grain legume species, and data on non-legume species grown after grain legumes in the crop sequence are also included. Most of these non-legume species correspond to cereals and oilseeds. The data are organized into a relational database with nine structured tables and 198 standardized attributes (Tables 2 and 3 (available online only)).

The dataset can be used for two types of quantitative analysis. First, the dataset can be used to compare the crop production of a broad range of grain legume species, on the basis of experimental data with diverse criteria (e.g., grain yield, aerial biomass and crop nitrogen content). Second, the dataset can be used to assess the crop production of cereal and oilseed species following grain legume species cultivated as preceding crops in the same crop sequences, based on a consideration of field data for various criteria.



The dataset is freely available to facilitate such analyses. It could easily be updated in the future, by adding the results of new experiments not originally included in the dataset. It might also be interesting to expand the dataset to include legumes grown for purposes other than grain production (e.g., forage production) or legumes grown in intercropping systems. The global dataset should prove to be a useful support for experimental assessments of the agronomic and environmental performances of a large diversity of grain legumes.

## Methods

### Literature search

We carried out a systematic search of peer-reviewed journals for articles comparing grain legume yields. We defined a grain legume species as a plant from the Fabaceae family, based on the United States Department of Agriculture Plants Database (<http://plants.usda.gov/java/>), and cropped for grain production. The literature search was completed on February 15, 2016. The equation search was: 'crop\* AND (legum\* OR pulse\*) AND (yield\* OR 'dry matter' OR biomass) AND (compar\* OR assessment OR product\* OR performance\*) AND (trial\* OR factorial OR experiment\* OR treatment\* OR condition\*) NOT (intercrop\* OR catch OR cover OR 'green manure' OR forage OR fodder)'. The search terms were used to query the Institute for Scientific Information Web of Science (<http://wokinfo.com/>), with no restrictions concerning the date and language of publication in the article title, abstract and author keywords.

The initial literature search identified 8,386 articles as of potential interest (Fig. 2). Each article title and article abstract were screened for eligibility according to six criteria: (1) article title and/or article abstract reporting one or several annual grain legume species grown as sole crops, (2) article title and/or article abstract reporting at least two grain legume species grown at the same field site during the same growing season, (3) article title and/or article abstract reporting at least one experiment conducted during one or several growing seasons, from the seeding stage to the harvest stage, (4) article title and article abstract referring to an article published in a peer-reviewed journal, (5) article title or article abstract written in English and (6) full-text article available. We selected 223 eligible full-text articles that met these first six criteria (Fig. 2).

Eligible full-text articles were then examined according to three additional criteria: (7) full-text article reporting raw data not duplicated in other articles or raw data that could be obtained by contacting authors, (8) full-text article reporting individual grain yield for each species and (9) full-text article reporting one or several experiments for which field site location or soil characteristics were precisely stated. We selected 60 full-text articles that met all nine criteria. This search was supplemented by screening the references cited in these 60 full-text articles. We also screened the references included in one meta-analysis about drought effects on food legume production<sup>187</sup> for eligibility. When reviewing the full-text articles identified from references screening, all nine selection criteria defined above had to be met for the new article to be considered eligible. Note that, according to the criterion (2), experiments reporting data for single grain legume species were excluded. This selection criterion was used to ensure the direct comparability of different grain legume species, and avoid confounding effects between species characteristics and environmental factors. Experiments testing single species cannot be used to compare several species due to the effects of field site and growing season characteristics (e.g., climate conditions, soil types and plant diseases) on the growth and development of grain legumes.

We finally selected 173 full-text articles<sup>2–4,6,14–19,23–185</sup> published between 1967 and 2016 that met all nine selection criteria (Fig. 2).

### Database structure

All data are recorded in a relational database (Data Citation 1). The Structured Query Language (SQL) system is used to query and maintain the database. We used the open-access application Sequel Pro version 1.0.2 (<http://www.sequelpro.com/>). The data collected are grouped into nine related tables including 198 standardized attributes of five types: class, numerical, index, binary and date (Fig. 3 and Table 2). Within the database, the tables are organized according to a cascade path: each 'child' table is related to a 'mother' table. For instance, the 'Article' table is the 'mother' table for the 'child' 'Site' table (Fig. 3). The cascade path from each 'mother' table to each 'child' table is structured by a 'primary key' and a 'secondary key' (Fig. 3). A 'primary key' assigns an index to each row of the table, whether the table is a 'mother' table or a 'child' table. A 'secondary key' assigns the 'primary key' of a 'mother' table to each row of a 'child' table. The cardinality from each 'mother' table to each 'child' table is based on 'one-to-one' and 'one-to-many' relationships (Fig. 3).

The database is structured into nine separate but related tables, stored as CSV-formatted files (Data Citation 1). Tables are related to each other via primary and secondary keys, as explained in Fig. 3. The names, types and definitions of attributes included in the nine tables are listed in Table 3 (available online only).

The 'Literature\_Search' table describes each step in the literature search at which each original article was selected (e.g., selection from the initial literature search or from references screening). The corresponding file is entitled 'Literature\_Search.csv' (Data Citation 1), and includes 2 columns and 3 rows (including the row header for the names of attributes).

The 'Article' table describes the references of the 173 selected articles (e.g., the name of the first author and the name of the journal). The corresponding file is entitled 'Article.csv' file (Data Citation 1), and includes 8 columns and 174 rows (including the row header for the names of attributes).

The 'Site' table describes the characteristics of each field site considered in each article (e.g., latitude and longitude coordinates, soil texture, precipitation and temperature). The corresponding file is entitled 'Site.csv' (Data Citation 1), and includes 29 columns and 361 rows (including the row header for the names of attributes).

The 'Crop\_Sequence\_Trtr' table describes each combination of crop sequences and management practices into the treatments studied at each field site (e.g., names of the species and their order in each crop sequence). The corresponding file is entitled 'Crop\_Sequence\_Trtr.csv' (Data Citation 1), and includes 8 columns and 4,560 rows (including the row header for the names of attributes).

The 'Crop' table provides information about each crop (e.g., names of the species, seeding and harvest dates, number of replicates, grain yield, aerial biomass, crop nitrogen content, residual soil nitrogen content, water use, error terms and error types). The main attributes included in this central table are described below in the Data Records section. The corresponding file is entitled 'Crop.csv' (Data Citation 1), and includes 106 columns and 8,582 rows (including the row header for the names of attributes).

The 'Tillage' table describes tillage management for each crop (e.g., tillage tools, incorporation of preceding crop residues, seeding density and legume inoculation). The corresponding file is entitled 'Tillage.csv' (Data Citation 1), and includes 19 columns and 8,582 rows (including the row header for the names of attributes).

The 'Fertilization' table describes nitrogen, phosphate and potassium fertilizer management for each crop (e.g., names and doses of fertilizers). Only the total fertilizer dose is reported for each type of nutrient. The corresponding file is entitled 'Fertilization.csv' (Data Citation 1), and includes 7 columns and 25,744 rows (including the row header for the names of attributes).

The 'Weed\_Insect\_Fungi' table describes weeds, insects, and fungi management for each crop (e.g., mechanical treatment, names and doses of pesticides). The corresponding file is entitled 'Weed\_Insect\_Fungi.csv' (Data Citation 1), and includes 13 columns and 45,002 rows (including the row header for the names of attributes).

The 'Irrigation' table describes irrigation management for each crop (e.g., quantity of water applied and irrigation method). The corresponding file is entitled 'Irrigation.csv' (Data Citation 1), and includes 6 columns and 8,582 rows (including the row header for the names of attributes).

In addition to the nine CSV-formatted files (tables), downloadable from *Dryad Digital Repository* (Data Citation 1), the entire content of the database is also stored in a SQL-formatted file. The corresponding file is entitled 'Database.sql', and is also downloadable from *Dryad Digital Repository* (Data Citation 1). Examples of SQL queries for extracting data for each table are stored in a TXT-formatted file. The corresponding file is entitled 'Examples\_SQL\_Queries.txt', and is also downloadable from *Dryad Digital Repository* (Data Citation 1).

The names, types, and definitions of the 198 attributes included in the nine tables are reported in Table 3 (available online only).

The values (including error terms) and dates reported in graphics were digitized manually with the open-access application WebPlotDigitizer (<http://arohatgi.info/WebPlotDigitizer/>). The maximum error was estimated at 5.0% for the digitization of low-resolution images, generally from articles published before 1990. 'NA' indicates that data were 'Not Available' for the cell concerned. 'NULL' indicates a logical absence of data for attributes included in the 'Crop', 'Tillage', 'Fertilization', 'Weed\_Insect\_Fungi', and 'Irrigation' tables. For example, for the 'Fertilization' table, if no nitrogen fertilizer was applied to the crop (i.e., '0.00' was reported in the 'Fertilization\_NPK\_Dose' attribute), then 'NULL' was reported for the 'Fertilization\_NPK\_Dose\_Product\_Name' attribute.

### Data Records

We describe below the main attributes of the 'Crop' table because this table includes most of the experimental data extracted from the 173 selected articles. Information on other attributes (e.g., articles, field sites, combinations of crop sequences and management practices) is defined in Table 3 (available online only).

In the 'Crop' table, grain yield is by far the attribute including the highest number of data. This high reporting rate reflects the explicit requirement for presence of grain yield data during the article selection process (i.e., criterion 8). Reporting rates are lower for aerial biomass, grain nitrogen content, aerial nitrogen content, fixed aerial nitrogen content, residual soil nitrogen content and water use. Table 1 presents the total number (percentage) of available and missing data for these attributes over all crop\*field site\*growing season\*treatment combinations.

When data were not reported for some attributes (e.g., aerial biomass or water use) in the selected articles, we systematically collected data for related attributes (e.g., harvest index or grain water use efficiency) in order to retrieve the missing data. For examples, aerial biomass can be deduced from grain yield and harvest index, and water use can be deduced from grain yield and grain water use efficiency. When data were not available for any related attributes, we contacted the authors of the selected articles, and we asked them to provide us with additional raw data when available.

**'Crop\_Sequence\_Treatment\_Name' attribute**

The name of each combination of crop sequences and management practices was based on the common names of the species, such as for both 'Crop\_Sequence\_Tr1t\_Name' and 'Crop\_Sequence\_Tr1t\_Species\_Order' attributes in the 'Crop\_Sequence\_Tr1t' table. For instance, the name of a legume-cereal sequence without application of nitrogen fertilizer (0N) could be 'Garden pea-Common wheat, 0N' where 'Garden pea' and 'Common wheat' are the common names listed in the United States Department of Agriculture Plants Database (<http://plants.usda.gov/java/>) for *Pisum sativum* and *Triticum aestivum*, respectively. Malik *et al.*<sup>105</sup> and McEwen *et al.*<sup>108</sup> described several crop sequences including grain legumes and crop sequences including barrelclover (*Medicago truncatula*) or common oat (*Avena sativa*), both preceding common wheat. For these two articles, we excluded the crop sequences including barrelclover and common oat because these crops were grown for forage production.

**'Crop\_Site\_Growing\_Season\_ID' attribute**

This attribute is an index identifying each species grown at a given field site during one or several growing seasons. Identical raw data were found to have been duplicated in two pairs of articles: Muchow *et al.*<sup>114</sup> and Sinclair *et al.*<sup>153</sup> on the one hand, and Heenan *et al.*<sup>71</sup> and Armstrong *et al.*<sup>2</sup> on the other. The duplicated raw data from Sinclair *et al.*<sup>153</sup> and Heenan *et al.*<sup>71</sup> were excluded because the number of crop\*field site\*growing season\*treatment combinations was smaller in these two articles than in their duplicates.

**'Crop\_Species\_Scientific\_Name' and 'Crop\_Species\_Common\_Name' attributes**

These attributes give the scientific and common names of the species. The scientific name of each species was related to the common name listed in the United States Department of Agriculture Plants Database (<http://plants.usda.gov/java/>), to avoid confusion due to the use of different common names for the same species. In the absence of a common name for *Brassica campestris*, *Lupinus atlanticus* and *Triticum sativum*, the scientific names of these species were used as common names. In the presence of fallow period, it was not possible to give a scientific name and a common name, and 'Fallow' was reported.

**'Crop\_Date\_From\_Seeding\_To\_Harvest\_Day\_Number' attribute**

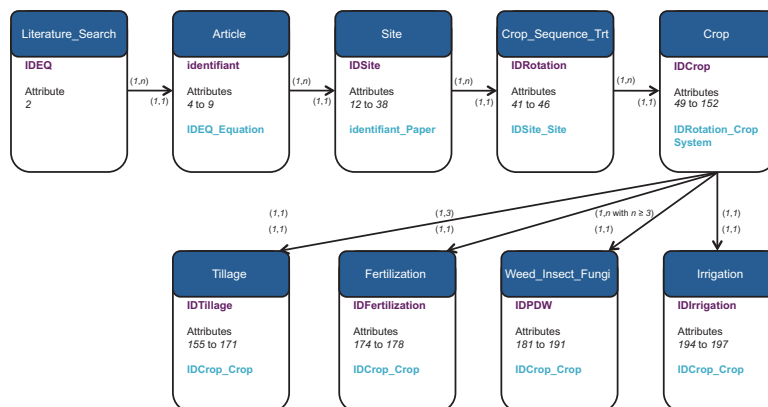
We calculated the number of days from seeding date to harvest date, with the open-access application Time and Date (<http://www.timeanddate.com/>). For data averaged across multiple growing seasons, we calculated the number of days from seeding date to harvest date for each growing season and then obtained the average by dividing by the total number of growing seasons.

Some articles approximated seeding date and harvest date by describing these events as occurring in the 'early', 'middle' or 'late' part of the month. We defined 'early' as the first 15-day period of the month



**Figure 2.** Flowchart of the steps in the literature search. Boxes with solid lines represent the articles identified (orange), excluded (red) or included in the database (green). In these boxes, the number of articles ( $n_i$ ) is indexed according to each step  $i$  of the literature search. Boxes with dashed edges represent the selection process, and selection criteria are indexed in *italics*.

www.nature.com/sdata/



**Figure 3.** Relational model of the database. Each box represents one table. One ‘primary key’ and one ‘secondary key’ are assigned to each table (except for the ‘Literature\_Search’ table, which is exclusively a ‘mother’ table). Each table includes many attributes. For the sake of readability, attributes are indexed in italic from one ‘mother’ table to one or many ‘child’ tables along the cascade path of the database (Table 3 (available online only)). Arrows indicate relationships from one ‘mother’ table to one or many ‘child’ tables. For upward and backward matching between tables, each pair of numbers in brackets indicates the cardinality of the relationships between attributes. The cardinality may involve ‘one-to-one’ (i.e., 1,1) relationship or ‘one-to-many’ (i.e., 1,n) relationship. For upward matching, for instance, the cardinality (1,n) from the ‘Article’ table to the ‘Site’ table indicates that one article may have one or many field sites. For backward matching, the cardinality (1,1) from the ‘Site’ table to the ‘Article’ table indicates that each field site may belong to only one article. Within each table, the names of primary and secondary keys are indicated in purple and blue, respectively.

(1st–15th), ‘middle’ as the 15th day of the month and ‘last’ as the second 15-day period of the month (15th–30th or 15th–31st). In these cases, the number of days from seeding to harvest was calculated by selecting the last day of the period concerned, i.e., the 15th day of the month for ‘early’ and ‘middle’ and the 30th or 31st day of the month for ‘late’.

Some articles reported only the number of days from seeding to harvest, without indicating precise dates or months. In these cases, we reported only the number of days from seeding to harvest. We used the expression ‘NA NA NA’ (i.e., ‘Day Month Year’ formatted expression) for both seeding and harvest dates.

**‘Crop\_Following\_Number’ attribute**

This attribute is used to distinguish preceding crops from following crops in the crop sequence. It takes three values: ‘0’ (i.e., the main crop or the preceding crop, mostly grain legumes), ‘1’ (i.e., the following crop, mostly cereals and oilseeds) and ‘2’ (i.e., the crop after the following crop, mostly cereals and oilseeds).

**‘Crop\_Multiple\_Following\_For\_Same\_Preceding’ attribute**

Some studies reported results for many different crops and management practices following the same preceding crop. The binary ‘Crop\_Multiple\_Following\_For\_Same\_Preceding’ attribute was used to identify data associated with the same preceding crop.

**‘Crop\_Across\_Treatment\_Averaged\_Value’ and ‘Crop\_Across\_Treatment\_Averaged\_Value\_Type’ attributes**

For species grown at the same field site during the same growing season, some articles reported only data averaged over combinations of treatments (e.g., cultivar\*seeding date\*presence of irrigation). We included these data provided that each type of individual treatment was precisely defined in the article. In all cases, we systematically reported whether or not the data were averaged over combinations of

treatments. When data were averaged over combinations of treatments, the total number of replicates was calculated as the sum of the replicates for each of the treatments for which results were averaged.

For articles reporting data for several cultivars of the same species but without data averaging, the data were reported separately for each cultivar. For articles reporting data averaged over several cultivars of the same species, only the averaged data were included in the dataset. The total number of replicates was calculated by multiplying the number of replicates of each cultivar by the total number of cultivars.

#### 'Crop\_Across\_Species\_Same\_Treatment\_Value' and 'Crop\_Across\_Species\_Same\_Treatment\_Value\_Type' attributes

In some articles, different types of treatment were applied to species grown at the same site during the same growing season. Each different type of treatment was reported in this case.

#### 'Crop\_Replicate\_Number' attribute

As mentioned above, when averaged data were reported in the articles, the number of replicates was equal to the sum of the replicates used to calculate each average.

#### 'Crop\_Yield\_Grain' attribute

This attribute corresponds to grain yield data, with a few exceptions. For *Brassica chinensis* (pak choy), *Citrullus lanatus* (watermelon), *Gossypium hirsutum* (upland cotton), *Ipomoea batatas* (sweet potato) and *Solanum lycopersicum* (garden tomato), the yields reported are the economic yields. For *Arachis hypogaea* (peanut), pods are included in grain yields. In all other situations, the yield data given correspond to grain yields. Mutant non-nodulating legume cultivars, shading treatment and under-sowing treatment were excluded from the database. When grain yield data of following crops were confounded between the effect of preceding species and the effect of nitrogen fertilizer dose, these data were also excluded. Data were reported in 96% of all crop\*field site\*growing season\*treatment combinations. Grain yield varied strongly both between grain legume species and between articles for a given species (Fig. 4a). Median grain yield was lowest for *Vigna subterranea* (bambarra groundnut) and highest for *Trigonella foenum-graecum* (sicklefruit fenugreek).

#### 'Crop\_Biomass\_Aerial' attribute

This attribute corresponds to aerial biomass data. Data were reported in 27% of all crop\*field site\*growing season\*treatment combinations. Aerial biomass varied considerably both between grain legume species and between articles for a given species (Fig. 4b). Median aerial biomass was lowest for *Vigna aconitifolia* (moth bean) and highest for *Trifolium repens* (white clover).

#### 'Crop\_Yield\_Grain\_DM\_Percentage' and 'Crop\_Biomass\_Aerial\_DM\_Percentage' attributes

These two attributes correspond to the percentage of dry matter to which grain yield and aerial biomass correspond, respectively. When only the percentage of dry matter corresponding to aerial biomass was available and grains were included in aerial biomass, we assumed that the grains accounted for the same percentage of dry matter as the aerial biomass.

#### 'Crop\_Harvest\_Index' attribute

This attribute was reported in the database to calculate aerial biomass at physiological maturity from grain yield. Data were reported in 4% of all crop\*field site\*growing season\*treatment combinations (Fig. 4c). Median harvest index was lowest for *Vicia villosa* (winter vetch) and highest for *Vicia faba* (fababean).

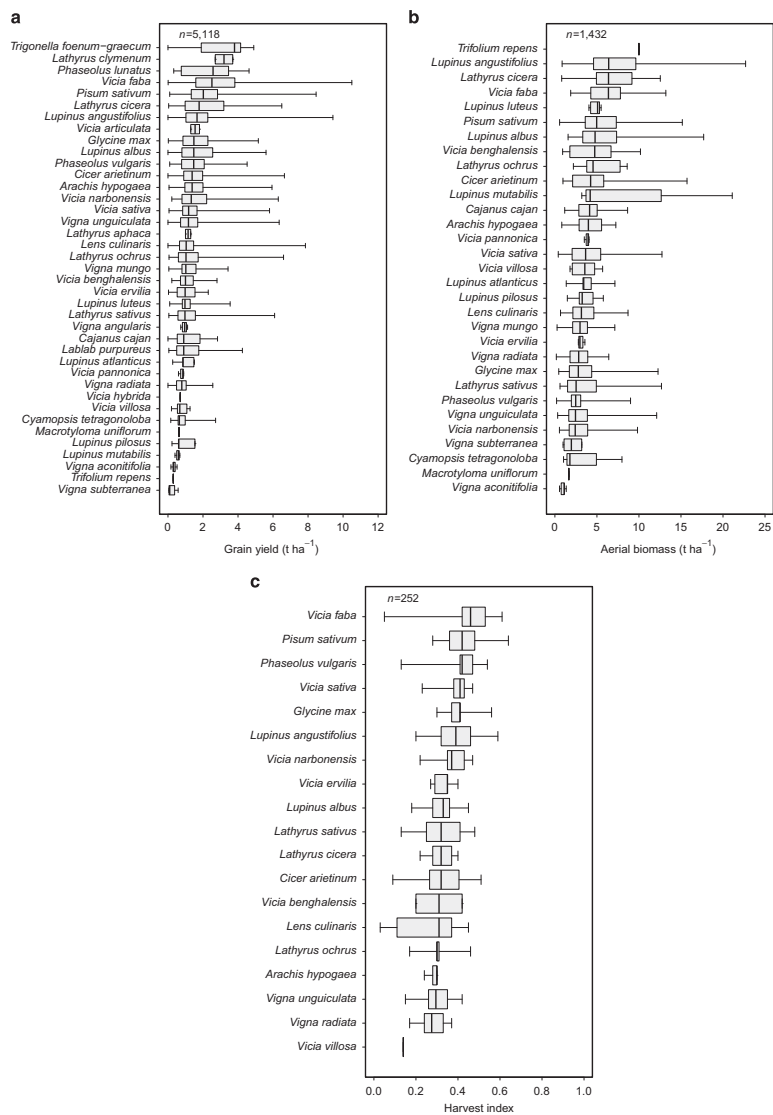
#### 'Crop\_N\_Quantity\_Grain' and 'Crop\_N\_Quantity\_Aerial' attributes

These two attributes correspond to the quantity of nitrogen in grains and aerial components, respectively. For the 'Crop\_N\_Quantity\_Grain' attribute, data were reported in 10% of all crop\*field site\*growing season\*treatment combinations. For the 'Crop\_N\_Quantity\_Aerial' attribute, data were reported in 10% of all crop\*field site\*growing season\*treatment combinations. As previous attributes, grain and aerial nitrogen quantities varied both between grain legume species and between articles for a given species (Fig. 5a,b). Median grain nitrogen quantity was lowest for *Vigna subterranea* (bambarra groundnut) and highest for *Lupinus albus* (white lupine). Median aerial nitrogen quantity was lowest for *Vicia narbonensis* (purple broad vetch) and highest for *Lupinus mutabilis* (sweet tarwi).

#### 'Crop\_N\_Fixed\_Percentage\_Aerial' attribute

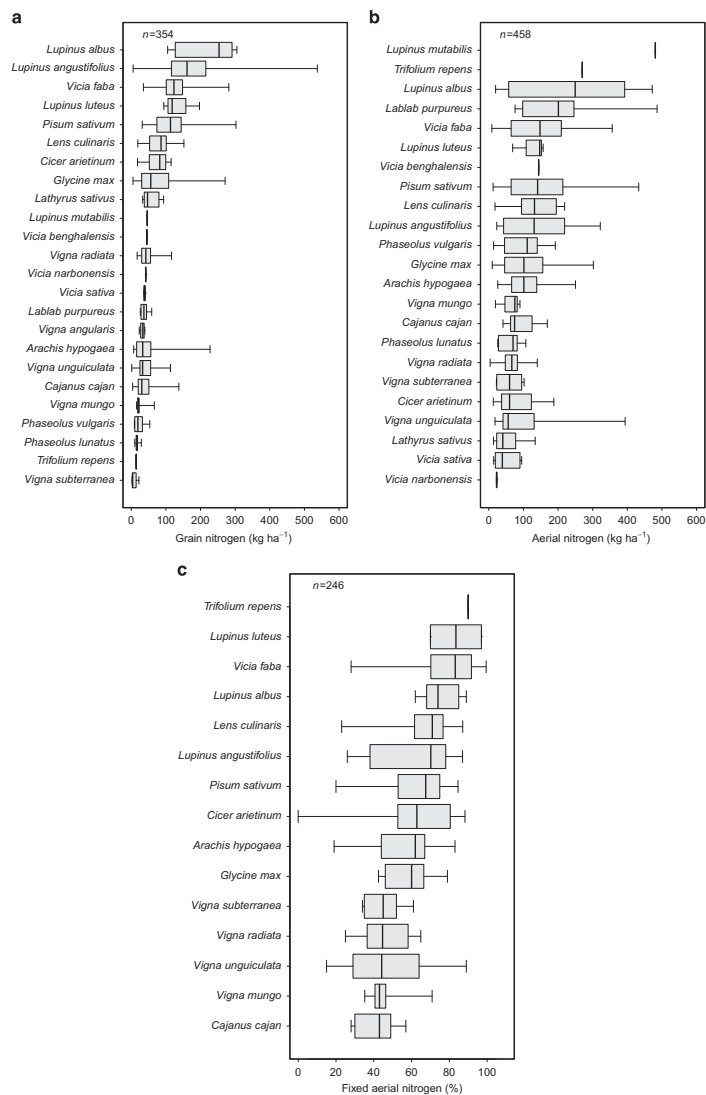
This attribute corresponds to the percentage of aerial nitrogen fixed by legume species. 'NA' was systematically reported for non-legume species. Data were reported in 3% of all crop\*field site\*growing season\*treatment combinations (Fig. 5c). Median fixed aerial nitrogen percentage was lowest for *Cajanus cajan* (pigeonpea) and highest for *Trifolium repens* (white clover).

www.nature.com/sdata/



**Figure 4.** Distribution of grain yield ( $t\ ha^{-1}$ ) for 39 grain legume species (a), aerial biomass ( $t\ ha^{-1}$ ) for 31 grain legume species (b), and harvest index for 19 grain legume species (c). Distributions are derived using data extracted from the database without additional calculations. Intrabox lines indicate medians, box edges indicate 25th and 75th percentiles, and whiskers indicate minimum and maximum values. The number of observations ( $n$ ) is also indicated. The scientific names of the species are ranked in descending order of median values.

www.nature.com/sdata/



**Figure 5.** Distribution of grain nitrogen ( $\text{kg ha}^{-1}$ ) for 24 grain legume species (a), aerial nitrogen ( $\text{kg ha}^{-1}$ ) for 23 grain legume species (b), and fixed aerial nitrogen (%) for 15 grain legume species (c). Distributions are derived using data extracted from the database without additional calculations. Intrabox lines indicate medians, box edges indicate 25th and 75th percentiles, and whiskers indicate minimum and maximum values. The number of observations ( $n$ ) is also indicated. The scientific names of the species are ranked in descending order of median values.

**'Crop\_N\_Fixed\_Percentage\_Aerial\_Method' and 'Crop\_N\_Fixed\_Percentage\_Aerial\_Reference\_Species' attributes**

These two attributes correspond to the method used to determine the percentage of aerial nitrogen fixed by legume species (e.g., the  $^{15}\text{N}$  isotope dilution method or the A-value method), and the scientific name of the non-fixing reference species. Some articles used a legume reference species rather than a non-legume reference species. In all cases, the legume reference species was a mutant non-nodulating legume cultivar that did not fix atmospheric nitrogen.

**'Crop\_Biomass\_Aerial\_Stage\_Detailed', 'Crop\_Biomass\_Aerial\_Stage\_Simplified', 'Crop\_N\_Fixed\_Percentage\_Aerial\_Stage\_Detailed' and 'Crop\_N\_Fixed\_Percentage\_Aerial\_Stage\_Simplified' attributes**

These attributes correspond to the phenological stages at which aerial biomass and the percentage of fixed aerial nitrogen (or the quantity of fixed aerial nitrogen with the 'Crop\_N\_Fixed\_Percentage\_Aerial\_Stage\_Detailed' attribute) were determined. The 'Crop\_Biomass\_Aerial\_Stage\_Detailed' and 'Crop\_N\_Fixed\_Percentage\_Aerial\_Stage\_Detailed' attributes correspond to the detailed phenological stage originally stated in the article. The 'Crop\_Biomass\_Aerial\_Stage\_Simplified' and 'Crop\_N\_Fixed\_Percentage\_Aerial\_Stage\_Simplified' attributes correspond to a simplified phenological stage divided into 'Before physiological maturity' and 'Physiological maturity'.

**'Crop\_Protein\_Quantity\_Percentage\_Grain' attribute**

This attribute corresponds to the percentage or the quantity of protein in grains. In the selected articles, these protein contents were often calculated by multiplying the percentage or the quantity of nitrogen in grains by a constant. However, this constant differed between articles. Note that only a few articles referred to the percentage or the quantity of protein. We reported the percentage or the quantity of protein in grains independently of the percentage or the quantity of nitrogen in grains.

**'Crop\_N\_Balance\_Simplified' attribute**

This attribute corresponds to the simplified nitrogen balance originally calculated in the articles (e.g., the difference between the quantity of nitrogen in grains and the quantity of fixed aerial nitrogen). Nitrogen balance data were only reported if the attributes used to calculate them were not directly available from raw data (e.g., the quantity of nitrogen in grains and the quantity of fixed aerial nitrogen). This was the case for only three articles.

**'Crop\_N\_Soil\_Quantity\_Percentage\_Seeding' and 'Crop\_N\_Soil\_Quantity\_Percentage\_Harvest' attributes**

These two attributes correspond to the percentage or the quantity of soil nitrogen at seeding and at harvest, respectively.

**'Crop\_N\_Soil\_Quantity\_Percentage\_Seeding\_Type', 'Crop\_N\_Soil\_Quantity\_Percentage\_Seeding\_Depth', 'Crop\_N\_Soil\_Quantity\_Percentage\_Seeding\_Date', 'Crop\_N\_Soil\_Quantity\_Percentage\_Harvest\_Type', 'Crop\_N\_Soil\_Quantity\_Percentage\_Harvest\_Depth' and 'Crop\_N\_Soil\_Quantity\_Percentage\_Harvest\_Date' attributes**

These attributes correspond to (i) the type of nitrogen (e.g., nitrogen or nitrate or mineral), (ii) the depth of soil used to determine the percentage or the quantity of soil nitrogen and (iii) the date at which soil measurements were made. These attributes were reported at both seeding and harvest.

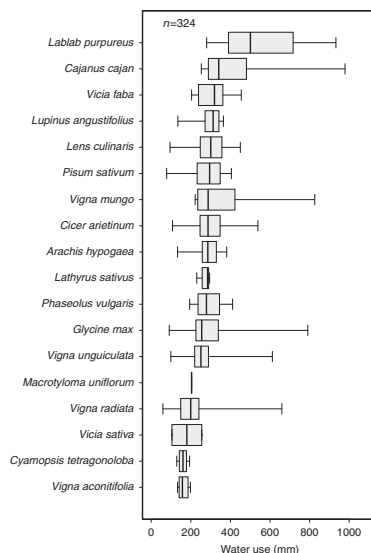
**'Crop\_Water\_Use\_Balance' attribute**

This attribute corresponds to the water use or the water balance, according to the equation given in the selected articles. Data were reported in 6% of all crop\*field site\*growing season\*treatment combinations. Water use (or water balance) varied both between grain legume species and between articles for a given species (Fig. 6). Median water use (or water balance) was lowest for *Vigna aconitifolia* (moth bean) and highest for *Lablab purpureus* (hyacinthbean).

**'Crop\_Harvest\_Index', 'Crop\_N\_Percentage\_Grain', 'Crop\_N\_Percentage\_Aerial', 'Crop\_N\_Harvest\_Index', 'Crop\_N\_Fixed\_Quantity\_Aerial', 'Crop\_Water\_Use\_Balance\_Efficiency\_Grain' and 'Crop\_Water\_Use\_Balance\_Efficiency\_Aerial' attributes**

These seven attributes were reported in the database to calculate missing data: aerial biomass, quantity of nitrogen in grains, quantity of nitrogen in aerial components, percentage of fixed aerial nitrogen, and water use.





**Figure 6.** Distribution of water use (mm) for 18 grain legume species. Water use is calculated using different types of equations, indicated within the 'Crop\_Water\_Use\_Balance\_Equation' attribute. The distribution is derived using data extracted from the database without additional calculations. Intrabox lines indicate medians, box edges indicate 25th and 75th percentiles, and whiskers indicate minimum and maximum values. The number of observations (*n*) is also indicated. The scientific names of the species are ranked in descending order of median values.

**'Crop\_Biomass\_Aerial\_Definition', 'Crop\_N\_Percentage\_Aerial\_Definition', 'Crop\_N\_Quantity\_Aerial\_Definition', 'Crop\_N\_Fixed\_Quantity\_Aerial\_Definition' and 'Crop\_Water\_Use\_Balance\_Efficiency\_Aerial\_Definition' attributes**

Different aerial components were included in the aerial biomass, the percentage or the quantity of aerial nitrogen, and the efficiency of aerial water use or aerial water balance. These five attributes were used to determine the aerial components originally reported in the articles. When the 'shoot', 'straw' and 'stubble' terms were used to define the aerial components in the articles, we assumed that the grains were not included in the aerial components. This information was reported for (i) the aerial biomass in the 'Crop\_Biomass\_Aerial\_Definition' attribute, (ii) the percentage of aerial nitrogen in the 'Crop\_N\_Percentage\_Aerial\_Definition' attribute, (iii) the quantity of aerial nitrogen in the 'Crop\_N\_Quantity\_Aerial\_Definition' attribute, (iv) the quantity of fixed aerial nitrogen in the 'Crop\_N\_Fixed\_Quantity\_Aerial\_Definition' attribute, and (v) the efficiency of aerial water use or aerial water balance in the 'Crop\_Water\_Use\_Balance\_Efficiency\_Aerial\_Definition' attribute.

**'Crop\_N\_Balance\_Simplified\_Equation' and 'Crop\_Water\_Use\_Balance\_Equation' attributes**

For these two attributes, we reported the equations used to calculate simplified nitrogen balance and water use or water balance, respectively.

**Attributes relating to error terms and error types**

When available, we systematically reported error terms and error types associated with data about grain yield, aerial biomass, crop nitrogen content, residual soil nitrogen content and water use. For the 'Crop\_Yield\_Grain' attribute, the 'Crop\_Yield\_Grain\_Error' attribute indicates the error term and the 'Crop\_Yield\_Grain\_Error\_Type' attribute indicates the error type for a given item of grain yield data for a given crop in the 'Crop' table. Error terms and error types were reported as raw data. For instance, when an article reported the error type as Fisher's Least Significant Difference, the data were directly reported as Fisher's Least Significant Difference. Unidentified error bars digitized from graphs were assumed to represent standard errors. When available, the numbers of replicates were also reported. For

48% of grain yields, both error terms and the numbers of replicates were reported. For 47% of grain yields, only the number of replicates was reported.

#### Technical Validation

Each article was read carefully at least three times by the same person, to determine the type and the quantity of data reported by the authors. Once the data had been extracted, all the data reported in the tables were checked at least three times by the same person, to identify possible mistakes. SQL subset queries were systematically performed, to check the structural validity and coherence of class, numerical, index, binary and date attributes within each table, and to check the relationships between 'mother' and 'child' tables. Once the set of data was complete, SQL queries were carried out, to compare the entire content of the database with the original data reported in the selected articles. We systematically and manually checked for outliers in order to detect possible mistakes made during data extraction. We returned to the original articles as many times as needed to check the accuracy of the data. We checked the qualitative and quantitative contents of all class, numerical, index, binary and date attributes by importing each table in turn into the R software (version 3.2, <https://cran.r-project.org/>), and by visualizing data distribution for each attribute in turn. When the meaning of the data reported in the articles was unclear, authors were directly contacted and asked to provide additional information about their experimental protocols. Authors were also asked to provide additional data, particularly if large numbers of treatments had been averaged in their articles. Overall, 17 authors provided us with additional information and raw data (see the Acknowledgements section).

#### Usage Notes

The dataset is based on a compilation of experimental data published in 173 articles over the last 50 years. To our knowledge, this dataset is unique and constitutes the most comprehensive agronomic dataset for grain legume crops worldwide.

The dataset can be analyzed to assess performances for a broad diversity of grain legume species, and to provide global rankings for these species in terms of grain yield, aerial biomass, harvest index, aerial nitrogen fixation, nitrogen content in aerial components, nitrogen balance, and water use. It can also be used to assess the effect of including different grain legumes as preceding crops, before cereals and oilseeds in the same crop sequences. Global species rankings were recently estimated for energy crops<sup>188</sup>, but never for grain legumes. Rankings of grain legume species could be directly derived from our dataset by using standard meta-analysis methods based on random-effect models<sup>188</sup>. Attributes describing environmental factors (e.g., climate conditions and soil types) and management practices (e.g., tillage, fertilization, pest management and irrigation) can be used to analyze the variability of grain legume performances over field sites, growing seasons, and management practices.

Our dataset covers several contrasted geographical areas. It can be used to target suitable grain legume species for cultivation in particular pedoclimatic conditions. In the context of climate change, the database represents a useful resource to assess comparatively the production of grain legume species in drought-prone environments, or to identify innovative agricultural techniques for improving grain legume cultivation under yield-limiting abiotic and biotic stresses.

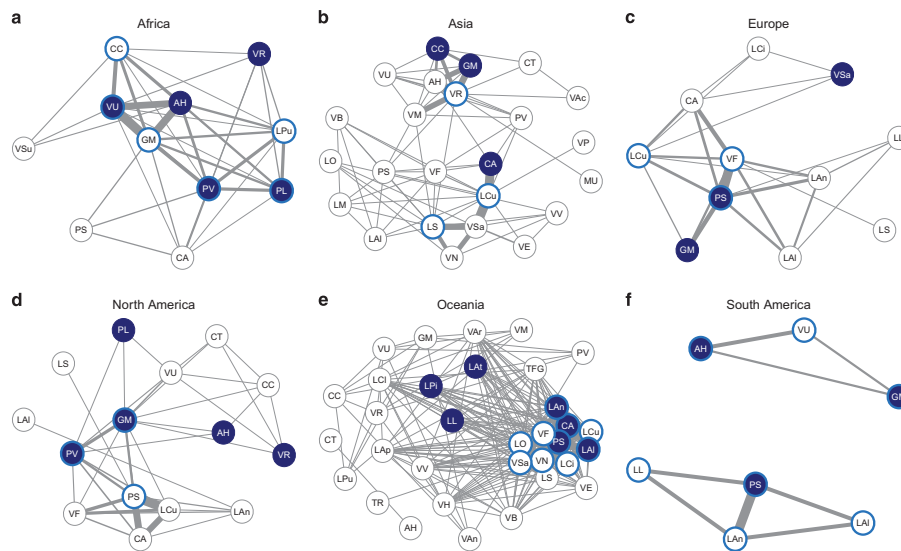
Subsets of the dataset can be used to address regional issues. Figure 7 presents six regional networks including the pairs of grain legume species frequently compared at the same field sites during the same growing seasons, and the grain legume species that were not frequently compared with each other. Such networks can be used to identify the species for which reliable comparisons are feasible, and those for which limited data are available. A quantitative analysis can then be computed to determine regional rankings of grain legume species. This approach could be used to identify highly productive species, and to compare them with major regional grain legume crops (e.g., garden pea in Europe or soybean in North America). Our dataset could thus shed new light on the potential value of as yet underused grain legumes from regional to global scales.

As geographical coordinates of the experiments were systematically reported, our dataset can be connected to large-scale climate and soil maps, and to Geographic Information Systems. An example is shown in Fig. 1 where the Köppen-Geiger climatic classification was indicated for field sites included in the database. Similar maps could be easily produced using other global classification of agroecological zones (e.g., the Global Agro-Ecological Zones Data Portal, <http://gaez.fao.org/Main.html#>), or soil typology (e.g., the Soils Portal of the Food and Agriculture Organization of the United Nations, <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>).

The dataset is also useful for comparing productivity levels of native and non-native grain legume species used as raw materials for food and feed across diverse geographic regions. Grain yield data can be converted into crude protein or energy contents metabolizable for livestock animals (e.g., pigs and poultry) using, for example, the Feedipedia Animal Feed Resources Information System (<http://www.feedipedia.org/>).

In the future, the dataset could be expanded in different ways. Results of new experiments comparing grain legume species can be easily included in our database. So far, we focused on legume species produced for grains, but legume grown for forage can also be included in the database without changing the relational database structure. In many world regions such as Africa, Asia and South America, agricultural grain legumes are frequently intercropped. Data collected in intercropping experiments could

www.nature.com/sdata/



**Figure 7.** Regional networks of grain legume species included in the database. The regions considered are: (a) Africa, (b) Asia, (c) Europe, (d) North America, (e) Oceania, and (f) South America. The links represent the pairs of species grown simultaneously at the same field sites during the same growing seasons. The thickness of the links increases with the number of field sites and the number of growing seasons over which the species are compared. The three most widely cropped grain legume species in each region over the 1961–2014 period, according to the crop classification and crop data from the Statistics Division of Food and Agriculture Organization of the United Nations<sup>22</sup>, are indicated as nodes in dark blue. The three most frequently compared grain legume species in the experimental dataset are indicated, by region, with light blue edges. The scientific names of grain legume species are abbreviated: AH, *Arachis hypogaea*; CA, *Cicer arietinum*; CC, *Cajanus cajan*; CT, *Cyamopsis tetragonoloba*; GM, *Glycine max*; LAI, *Lupinus albus*; LAN, *Lupinus angustifolius*; LAP, *Lathyrus aphaca*; LAT, *Lupinus atlanticus*; LCI, *Lathyrus cicera*; LCL, *Lathyrus clymenum*; LCU, *Lens culinaris*; LL, *Lupinus luteus*; LM, *Lupinus mutabilis*; LO, *Lathyrus ochrus*; LPI, *Lupinus pilosus*; LPU, *Lablab purpureus*; LS, *Lathyrus sativus*; MU, *Macrotyloma uniflorum*; PL, *Phaseolus lunatus*; PS, *Pisum sativum*; PV, *Phaseolus vulgaris*; TFG, *Trigonella foenum-graecum*; TR, *Trifolium repens*; VAc, *Vigna aconitifolia*; VAn, *Vigna angularis*; VAr, *Vicia articulata*; VB, *Vicia benghalensis*; VE, *Vicia ervilia*; VF, *Vicia faba*; VH, *Vicia hybrida*; VM, *Vigna mungo*; VN, *Vicia narbonensis*; VP, *Vicia pannonica*; VR, *Vigna radiata*; VSa, *Vicia sativa*; VSu, *Vigna subterranea*; VU, *Vigna unguiculata*; VV, *Vicia villosa*.

be further included in our dataset. Note that the relational structure of the database is relatively coercive, and should be modified with great care. The addition of a new table can have consequences on the relational framework and the cardinality relationships. But new data or new attributes can be easily incremented in existing tables.

The CSV format is well adapted for analyzing data using standard statistical softwares such as the R software (<https://cran.r-project.org/>). However, because of the cascade path between tables and of the cardinality relationships between attributes (Fig. 3), data extraction can be easily performed using SQL queries. An example of query is presented below for extracting binary data indicating absence ('0') or presence ('1') of tillage management for grain legume species included in the article indexed '29' in our dataset:

```
SELECT IDCrop, Crop_Species_Scientific_Name, IDTillage, Tillage_Presence_Tillage
FROM Article, Site, Crop_Sequence_Trtr, Crop, Tillage
WHERE identifiant = identifiant_Paper
```

www.nature.com/sdata/

```
AND IDSite = IDSite_Site
AND IDRotation = IDRotation_CropSystem
AND IDCrop = Tillage.IDCROP_Crop
AND identifiant = '29'
```

The result of the SQL query is:

IDCROP	Crop_Species_Scientific_Name	IDTillage	Tillage_Presence_Tillage
853	Cicer arietinum	849	1
854	Vicia faba	851	1
857	Lens culinaris	856	1
858	Pisum sativum	858	1
859	Cicer arietinum	860	1
860	Vicia faba	861	1
861	Lens culinaris	862	1
862	Pisum sativum	863	1
864	Cicer arietinum	864	1
865	Vicia faba	865	1
866	Lens culinaris	866	1
867	Pisum sativum	867	1
869	Cicer arietinum	868	1
870	Vicia faba	870	1
871	Lens culinaris	871	1
872	Pisum sativum	872	1
873	Cicer arietinum	873	0
874	Vicia faba	874	0
875	Lens culinaris	875	0
876	Pisum sativum	876	0
877	Cicer arietinum	877	0
878	Vicia faba	878	0
879	Lens culinaris	879	0
880	Pisum sativum	880	0
881	Cicer arietinum	881	0
882	Vicia faba	882	0
883	Lens culinaris	883	0
884	Pisum sativum	884	0
885	Cicer arietinum	885	0
887	Vicia faba	886	0
888	Lens culinaris	887	0
890	Pisum sativum	889	0

Other examples of SQL queries are shown in the TXT-formatted file entitled 'Examples\_SQL\_Queries.txt', downloadable from *Dryad Digital Repository* (Data Citation 1).

## References

- Vollmann, J. Soybean versus other food grain legumes: a critical appraisal of the United Nations International Year of Pulses 2016. *Bodenkultur* **67**, 17–24 (2016).
- Armstrong, E. L., Heenan, D. P., Pate, J. S. & Unkovich, M. J. Nitrogen benefits of lupins, field pea, and chickpea to wheat production in south-eastern Australia. *Aust. J. Agr. Res.* **48**, 39–47 (1997).
- Ahmad, T., Hafeez, F. Y., Mahmood, T. & Malik, K. A. Residual effect of nitrogen fixed by mungbean (*Vigna radiata*) and blackgram (*Vigna mungo*) on subsequent rice and wheat crops. *Aust. J. Exp. Agr.* **41**, 245–248 (2001).
- Bado, B. V., Bationo, A. & Cescas, M. P. Assessment of cowpea and groundnut contributions to soil fertility and succeeding sorghum yields in the Guinean savannah zone of Burkina Faso (west Africa). *Biol. Fert. Soils* **43**, 171–176 (2006).
- Duranti, M. Grain legume proteins and nutraceutical properties. *Fitoterapia* **77**, 67–82 (2006).
- Lizarazo, C. I. *et al.* Nutritive quality and protein production from grain legumes in a boreal climate. *J. Sci. Food Agr.* **95**, 2053–2064 (2014).
- Reckling, M. *et al.* Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. *Front. Plant Sci.* **7**, 669 (2016).
- Asif, M., Rooney, L. W., Ali, R. & Riaz, M. N. Application and opportunities of pulses in food system: a review. *Crit. Rev. Food Sci. Nutr.* **53**, 1168–1179 (2013).
- Mudryj, A. N., Yu, N. & Aukema, H. M. Nutritional and health benefits of pulses. *Appl. Physiol. Nutr. Metab.* **39**, 1197–1204 (2014).
- Temba, M. C., Njobeh, P. B., Adebo, O. A., Olugbile, A. O. & Kayitesi, E. The role of compositing cereals with legumes to alleviate protein energy malnutrition in Africa. *Int. J. Food Sci. Tech.* **51**, 543–554 (2016).
- Jezierny, D., Mosenthin, R. & Bauer, E. The use of grain legumes as a protein source in pig nutrition: a review. *Anim. Feed Sci. Tech.* **157**, 111–128 (2010).
- Martens, S. D., Tiemann, T. T., Bindelle, J., Peters, M. & Lascano, C. E. Alternative plant protein sources for pigs and chickens in the tropics—nutritional value and constraints: a review. *J. Agr. Rural Dev. Trop.* **113**, 101–123 (2012).
- Koivunen, E. *et al.* Digestibility and energy value of pea (*Pisum sativum* L.), faba bean (*Vicia faba* L.) and blue lupin (narrow-leaf) (*Lupinus angustifolius*) seeds in broilers. *Anim. Feed Sci. Tech.* **218**, 120–127 (2016).

14. Badaruddin, M. & Meyer, D. W. Grain legume effects on soil nitrogen, grain yield, and nitrogen nutrition of wheat. *Crop Sci.* **34**, 1304–1309 (1994).
15. Jawale, S. M., Jadhav, A. S. & Patil, V. G. Productivity and economics of legumes-winter sorghum (*Sorghum bicolor*) double cropping systems under tillage practices in western Maharashtra. *Indian J. Agron.* **43**, 396–403 (1998).
16. Bloem, A. A. & Barnard, R. O. Effect of annual legumes on soil nitrogen and on the subsequent yield of maize and grain sorghum. *S. Afr. J. Plant Soil* **18**, 56–61 (2001).
17. Cheruiyot, E. K., Mumera, L. M., Nakhone, L. N. & Mwonga, S. M. Effect of legume-managed fallow on weeds and soil nitrogen in following maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) crops in the Rift Valley highlands of Kenya. *Aust. J. Exp. Agr.* **43**, 597–604 (2003).
18. Espinoza, S. *et al.* Contribution of legumes to wheat productivity in Mediterranean environments of central Chile. *Field Crop Res.* **133**, 150–159 (2012).
19. Williams, C. M. *et al.* Effects of three pulse crops on subsequent barley, canola, and wheat. *Agron. J.* **106**, 343–350 (2014).
20. Yahara, T. *et al.* Global legume diversity assessment: concepts, key indicators, and strategies. *Taxon* **62**, 249–266 (2013).
21. Smýkal, P. *et al.* Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* **34**, 43–104 (2015).
22. Statistics Division of Food and Agriculture Organization of the United Nations (FAOSTAT). <http://faostat3.fao.org/home/E/> (2016).
23. Abd El Moneim, A. M., Khair, M. A. & Cocks, P. S. Growth analysis, herbage and seed yield of certain forage legume species under rainfed conditions. *J. Agron. Crop Sci.* **164**, 34–41 (1990).
24. Abd El Moneim, A. M. Agronomic potential of three vetches (*Vicia* spp.) under rainfed conditions. *J. Agron. Crop Sci.* **170**, 113–120 (1993).
25. Adela, J., Simona, N., Gheorghe, D., Pirşan, P. & Alda, S. The behaviour of some varieties of pea, soybean and lentils concerning the relationship bacterization-sowing period on the yield and quality in Timiş Plain. *Res. J. Agr. Sci.* **42**, 82–87 (2010).
26. Al-Thahabi, S. A., Yasin, J. Z., Abu-Irmaileh, B. E., Haddad, N. I. & Saxena, M. C. Effect of weed removal on productivity of chickpea (*Cicer arietinum* L.) and lentil (*Lens culinaris* Med.) in a Mediterranean environment. *J. Agron. Crop Sci.* **172**, 333–341 (1994).
27. Anderson, R. L., Tanaka, D. L. & Merrill, S. D. Yield and water use of broadleaf crops in a semiarid climate. *Agr. Water Manage.* **58**, 255–266 (2003).
28. Angadi, S. V. *et al.* Adaptation of alternative pulse and oilseed crops to the semiarid Canadian prairie: seed yield and water use efficiency. *Can. J. Plant Sci.* **88**, 425–438 (2008).
29. Annicchiarico, P. Adaptation of cool-season grain legume species across climatically-contrasting environments of southern Europe. *Agron. J.* **100**, 1647–1654 (2008).
30. Avola, G., Tuttobene, R., Gresta, F. & Abbate, V. Weed control strategies for grain legumes. *Agron. Sustain. Dev.* **28**, 389–395 (2008).
31. Ayaz, S., McKenzie, B. A., Hill, G. D. & McNeil, D. L. Variability in yield of four grain legume species in a subhumid temperate environment I. Yields and harvest index. *J. Agr. Sci.* **142**, 9–19 (2004).
32. Bailey, K. L., Gossen, B. D., Derksen, D. A. & Watson, P. R. Impact of agronomic practices and environment on diseases of wheat and lentil in southeastern Saskatchewan. *Can. J. Plant Sci.* **80**, 917–927 (2000).
33. Baroowa, B., Gogoi, N., Paul, S. & Sarma, B. Morphological responses of pulse (*Vigna* spp.) crops to soil water deficit. *Belgrade J. Agr. Sci.* **57**, 31–40 (2012).
34. Beck, D. P., Wery, J., Saxena, M. C. & Ayadi, A. Dinitrogen fixation and nitrogen-balance in cool-season food legumes. *Agron. J.* **83**, 334–341 (1991).
35. Bhambe, P. R., Sondge, V. D., Rodge, R. P. & Oza, S. R. Performance of different crop sequences under various levels of irrigation in Vertisols. *Indian J. Agr. Sci.* **60**, 402–405 (1990).
36. Boddey, R. M., Urquiaga, S., Neves, M. C. P., Suhet, A. R. & Peres, J. R. Quantification of the contribution of N<sub>2</sub> fixation to field-grown grain legumes—A strategy for the practical application of the <sup>15</sup>N isotope dilution technique. *Soil Biol. Biochem.* **22**, 649–655 (1990).
37. Bolland, M. D. A., Siddique, K. H. M., Loss, S. P. & Baker, M. J. Comparing responses of grain legumes, wheat and canola to applications of superphosphate. *Nutr. Cycl. Agroecosys.* **53**, 157–175 (1999).
38. Borowska, M., Prusinski, J. & Kaszkowiak, E. Production results of intensification of cultivation technologies in three lupin (*Lupinus* L.) species. *Plant Soil Environ.* **61**, 426–431 (2015).
39. Brennan, R. F., Bolland, M. D. A. & Shea, G. Comparing how *Lupinus angustifolius* and *Lupinus luteus* use zinc fertilizer for seed production. *Nutr. Cycl. Agroecosys.* **59**, 209–217 (2001).
40. Brennan, R. F. & French, R. J. Grain yield and cadmium concentration of a range of grain legume species grown on two soil types at Merredin, western Australia. *Aust. J. Exp. Agr.* **45**, 1167–1172 (2005).
41. Calderón, F. J., Vigil, M. F., Nielsen, D. C., Benjamin, J. G. & Poss, D. J. Water use and yields of no-till managed dryland grasspea and yellow pea under different planting configurations. *Field Crop Res.* **125**, 179–185 (2012).
42. Chemining'wa, G. N., Muthomi, J. W. & Theuri, S. W. M. Effect of rhizobia inoculation and starter-N on nodulation, shoot biomass and yield of grain legumes. *Asian J. Plant Sci.* **6**, 1113–1118 (2007).
43. Chen, C. *et al.* Winter pea and lentil response to seeding date and micro- and macro-environments. *Agron. J.* **98**, 1655–1663 (2006).
44. Christiansen, S., Ryan, J. & Singh, M. Forage and food legumes in a multi-year, wheat-based rotation under drought-stressed conditions in northern Syria's medium rainfall zone. *J. Agron. Crop Sci.* **197**, 146–154 (2011).
45. Christiansen, S. *et al.* Potential legume alternatives to fallow and wheat monoculture for Mediterranean environments. *Crop Pasture Sci.* **66**, 113–121 (2015).
46. Cutforth, H. W., McConkey, B. G., Ulrich, D., Miller, P. R. & Angadi, S. V. Yield and water use efficiency of pulses seeded directly into standing stubble in the semiarid Canadian prairie. *Can. J. Plant Sci.* **82**, 681–686 (2002).
47. Dakora, F. D., Aboyinga, R. A., Mahama, Y. & Apaseku, J. Assessment of N<sub>2</sub> fixation in groundnut (*Arachis hypogaea* L.) and cowpea (*Vigna unguiculata* L. Walp.) and their relative N contribution to a succeeding maize crop in northern Ghana. *MIRCEN J. Appl. Microb.* **3**, 389–399 (1987).
48. De, R., Rao, Y. Y. & Ali, W. Grain and fodder legumes as preceding crops affecting the yield and N economy of rice. *J. Agr. Sci.* **101**, 463–466 (1983).
49. Debaeke, P. & Hilaire, A. Production of rainfed and irrigated crops under different crop rotations and input levels in southwestern France. *Can. J. Plant Sci.* **77**, 539–548 (1997).
50. Devries, J. D., Bennett, J. M., Boote, K. J., Albrecht, S. L. & Maliro, C. E. Nitrogen accumulation and partitioning by three grain legumes in response to soil water deficits. *Field Crop Res.* **22**, 33–44 (1989).
51. Eaglesham, A. R. J., Ayanaba, A., Rao, V. R. & Eskew, D. L. Mineral N effects on cowpea and soybean crops in a Nigerian soil I. Development, nodulation, acetylene reduction and grain yield. *Plant Soil* **68**, 171–181 (1982).

52. Enyi, B. A. C. An analysis of the effect of weed competition on growth and yield attributes in sorghum (*Sorghum vulgare*), cowpeas (*Vigna unguiculata*) and green gram (*Vigna aureus*). *J. Agr. Sci.* **81**, 449–453 (1973).
53. Estevez de Jensen, C., Kurl, J. E. & Percich, J. A. Integrated management of edaphic and biotic factors limiting yield of irrigated soybean and dry bean in Minnesota. *Field Crop. Res.* **86**, 211–224 (2004).
54. Fernandez, A. L., Sheaffer, C. C., Wyse, D. L. & Michaels, T. E. Yield and weed abundance in early- and late-sown field pea and lentil. *Agron. J.* **104**, 1056–1064 (2012).
55. French, R. J. & Ewing, M. A. Soil type influences the relative yields of different cereals and crop legumes in the western Australian wheat-belt. *Aust. J. Exp. Agr.* **29**, 829–835 (1989).
56. French, R. J., Sweetingham, M. W. & Shea, G. G. A comparison of the adaptation of yellow lupin (*Lupinus luteus* L.) and narrow-leaved lupin (*L. angustifolius* L.) to acid sandplain soils in low rainfall agricultural areas of western Australia. *Aust. J. Agr. Res.* **52**, 945–954 (2001).
57. Gan, Y. T., Miller, P. R., Liu, P. H., Stevenson, F. C. & McDonald, C. L. Seedling emergence, pod development, and seed yields of chickpea and dry pea in a semiarid environment. *Can. J. Plant Sci.* **82**, 531–537 (2002).
58. Gan, Y., Campbell, C. A., Liu, L., Basnyat, P. & McDonald, C. L. Water use and distribution profile under pulse and oilseed crops in semiarid northern high latitude areas. *Agr. Water Manage.* **96**, 337–348 (2009).
59. Ghosh, P. K. & Singh, N. P. Production potential of summer legumes-maize (*Zea mays*) sequence under varying levels of nitrogen. *Indian J. Agron.* **41**, 525–528 (1996).
60. Ghosh, P. K., Dayal, D., Mandal, K. G., Wanjari, R. H. & Hati, K. M. Optimization of fertilizer schedules in fallow and groundnut-based cropping systems and an assessment of system sustainability. *Field Crop. Res.* **80**, 83–98 (2003).
61. Gregory, P. J. Alternative crops for Duplex soils: growth and water use of some cereal, legume, and oilseed crops, and pastures. *Aust. J. Agr. Res.* **49**, 21–32 (1998).
62. Gül, İ., Yildirim, M., Akinci, C., Doran, İ. & Kiliç, H. Response of silage maize (*Zea mays* L.) to nitrogen fertilizer after different crops in a semi arid environment. *Turk. J. Agr. For.* **32**, 513–520 (2008).
63. Gupta, J. P. & Gupta, G. N. Effect of grass mulching on growth and yield of legumes. *Agr. Water Manage.* **6**, 375–383 (1983).
64. Hanbury, C. D., Siddique, K. H. M., Galwey, N. W. & Cocks, P. S. Genotype-environment interaction for seed yield and ODAP concentration of *Lathyrus sativus* L. and *L. cicera* L. in Mediterranean-type environments. *Euphytica* **110**, 45–60 (1999).
65. Hatch, D. J., Joynes, A. & Stone, A. Nitrogen uptake in organically managed spring sown lupins and residual effects on leaching and yield of a following winter cereal. *Soil Use Manage.* **26**, 21–26 (2010).
66. Hauggaard-Nielsen, H., Mundus, S. & Jensen, E. S. Nitrogen dynamics following grain legumes and subsequent catch crops and the effects on succeeding cereal crops. *Nutr. Cycl. Agroecosys.* **84**, 281–291 (2009).
67. Hauggaard-Nielsen, H., Mundus, S. & Jensen, E. S. Grass-clover undersowing affects nitrogen dynamics in a grain legume-cereal arable cropping system. *Field Crop. Res.* **136**, 23–31 (2012).
68. Hayat, R. & Ali, S. Potential of summer legumes to fix nitrogen and benefit wheat crop under rainfed condition. *J. Agron.* **3**, 273–281 (2004).
69. Hayat, R. & Ali, S. Nitrogen fixation of legumes and yield of wheat under legumes-wheat rotation in Pothwar. *Pak. J. Bot.* **42**, 2317–2326 (2010).
70. Haynes, R. J., Martin, R. J. & Goh, K. M. Nitrogen fixation, accumulation of soil nitrogen and nitrogen balance for some field-grown legume crops. *Field Crop. Res.* **35**, 85–92 (1993).
71. Heenan, D. P. Effects of sowing time on growth and grain yield of lupin and field pea in south-eastern New South Wales. *Aust. J. Exp. Agr.* **34**, 1137–1142 (1994).
72. Henry, J. L., Slinkard, A. E. & Hogg, T. J. The effect of phosphorus fertilizer on establishment, yield and quality of pea, lentil and faba bean. *Can. J. Plant Sci.* **75**, 395–398 (1995).
73. Herridge, D. F. & Holland, J. F. Production of summer crops in northern New South Wales. I. Effects of tillage and double cropping on growth, grain and N yields of six crops. *Aust. J. Agr. Res.* **43**, 105–122 (1992).
74. Hodgson, A. S., Holland, J. F. & Rayner, P. Effects of field slope and duration of furrow irrigation on growth and yield of six grain-legumes on a waterlogging-prone Vertisol. *Field Crop. Res.* **22**, 165–180 (1989).
75. Imai, H., Kamata, K., Ma, C.-H. & Yang, Y.-F. Developing improved cropping systems for vegetables and legumes in the tropics 1. Assessment of cropping sequences in terms of yields. *Jpn. J. Trop. Agr.* **33**, 106–115 (1989).
76. Khan, D. F. *et al.* Effects of below-ground nitrogen on N balances of field-grown fababean, chickpea, and barley. *Aust. J. Agr. Res.* **54**, 333–340 (2003).
77. Jat, R. A., Dungrani, R. A., Arvadia, M. K. & Sahrawat, K. L. Diversification of rice (*Oryza sativa* L.)-based cropping systems for higher productivity, resource-use efficiency and economic returns in south Gujarat, India. *Arch. Agron. Soil Sci.* **58**, 561–572 (2012).
78. Jemo, M. *et al.* Phosphorus benefits from grain-legume crops to subsequent maize grown on acid soils of southern Cameroon. *Plant Soil* **284**, 385–397 (2006).
79. Jensen, E. S. Symbiotic N<sub>2</sub> fixation in pea and field bean estimated by <sup>15</sup>N fertilizer dilution in field experiments with barley as a reference crop. *Plant Soil* **92**, 3–13 (1986).
80. Jensen, C. R. *et al.* The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop. *Eur. J. Agron.* **20**, 405–418 (2004).
81. Jessop, R. S. & Mahoney, J. The effect of soil nitrogen on grain legume yield and nitrogen fixation. *J. Agr. Sci.* **105**, 231–236 (1985).
82. Jones, M. J. & Singh, M. Yields of crop dry matter and nitrogen in long-term barley rotation trials at two sites in northern Syria. *J. Agr. Sci.* **124**, 389–402 (1995).
83. Jones, M. J. & Singh, M. Long-term yield patterns in barley-based cropping systems in northern Syria. 2. The role of feed legumes. *J. Agr. Sci.* **135**, 237–249 (2000).
84. Kaizzi, K. C. *et al.* Optimizing smallholder returns to fertilizer use: bean, soybean and groundnut. *Field Crop. Res.* **127**, 109–119 (2012).
85. Kang, S., McKenzie, B. A. & Hill, G. D. Effect of irrigation on growth and yield of *Kabuli* chickpea (*Cicer arietinum* L.) and narrow-leaved lupin (*Lupinus angustifolius* L.). *Agron. N. Z.* **38**, 11–32 (2008).
86. Kar, G. & Kumar, A. Evaluation of post-rainy season crops with residual soil moisture and different tillage methods in rice fallow of eastern India. *Agr. Water Manage.* **96**, 931–938 (2009).
87. Karrou, M. & Oweis, T. Water and land productivities of wheat and food legumes with deficit supplemental irrigation in a Mediterranean environment. *Agr. Water Manage.* **107**, 94–103 (2012).
88. Kaul, H.-P. Pre-crop effects of grain legumes and linseed on soil mineral N and productivity of subsequent winter rape and winter wheat crops. *Bodenkultur* **55**, 95–102 (2004).
89. Kirchof, G. *et al.* The effect of soil puddling on the soil physical properties and the growth of rice and post-rice crops. *Soil Till. Res.* **56**, 37–50 (2000).
90. Kirchof, G. *et al.* Growth and yield response of grain legumes to different soil management practices after rainfed lowland rice. *Soil Till. Res.* **56**, 51–66 (2000).

91. Krupinsky, J. M., Tanaka, D. L., Merrill, S. D., Liebig, M. A. & Hanson, J. D. Crop sequence effects of 10 crops in the northern Great Plains. *Agr. Syst.* **88**, 227–254 (2006).
92. Kumar, K. & Goh, K. M. Biological nitrogen fixation, accumulation of soil nitrogen and nitrogen balance for white clover (*Trifolium repens* L.) and field pea (*Pisum sativum* L.) grown for seed. *Field Crop. Res.* **68**, 49–59 (2000).
93. Lafond, J. & Pageau, D. Nitrogen and non nitrogen benefit of legumes to a subsequent spring barley crop. *Can. J. Soil Sci.* **87**, 445–454 (2007).
94. Lal, R. No-tillage effects on soil properties under different crops in western Nigeria. *Soil Sci. Soc. Am. J.* **40**, 762–768 (1976).
95. Larbi, A. *et al.* Annual feed legume yield and quality in dryland environments in north-west Syria: 2. Grain and straw yield and straw quality. *Anim. Feed Sci. Tech.* **160**, 90–97 (2010).
96. Laurence, R. C. N. A comparison of the grain and protein yield potential of some annual legume species in south Australia. *Aust. J. Exp. Agr. Anim. Husb.* **19**, 495–503 (1979).
97. Lawn, R. J. Response of four grain legumes to water stress in south-eastern Queensland. III Dry matter production, yield and water use efficiency. *Aust. J. Agr. Res.* **33**, 511–521 (1982).
98. Lemerle, D. & Hinkley, R. B. Tolerances of canola, field pea, lupin and faba bean cultivars to herbicides. *Aust. J. Exp. Agr.* **31**, 379–386 (1991).
99. Lensen, A. W., Johnson, G. D. & Carlson, G. R. Cropping sequence and tillage system influences annual crop production and water use in semiarid Montana, USA. *Field Crop. Res.* **100**, 32–43 (2007).
100. López-Bellido, R. J. *et al.* Chickpea and faba bean nitrogen fixation in a Mediterranean rainfed Vertisol: effect of the tillage system. *Eur. J. Agron.* **34**, 222–230 (2011).
101. López-Bellido, L., Benitez-Vega, J., Garcia, P., Redondo, R. & López-Bellido, R. J. Tillage system effect on nitrogen rhizodeposited by faba bean and chickpea. *Field Crop. Res.* **120**, 189–195 (2011).
102. Lutman, P. J. W., Dixon, F. L. & Risotti, R. The response of four spring-sown combinable arable crops to weed competition. *Weed Res.* **34**, 137–146 (1994).
103. Maidl, F. X., Haunz, F. X., Panse, A. & Fischbeck, G. Transfer of grain legume nitrogen within a crop rotation containing winter wheat and winter barley. *J. Agron. Crop Sci.* **176**, 47–57 (1996).
104. Malhi, S. S., Johnston, A. M., Schoenau, J. J., Wang, Z. H. & Vera, C. L. Seasonal biomass accumulation and nutrient uptake of pea and lentil on a black Chernozem soil in Saskatchewan. *J. Plant Nutr.* **30**, 721–737 (2007).
105. Malik, R. S. *et al.* Dynamic crop sequencing in western Australian cropping systems. *Crop Pasture Sci.* **66**, 594–609 (2015).
106. Matus, A., Derksen, D. A., Walley, F. L., Loepky, H. A. & van Kessel, C. The influence of tillage and crop rotation on nitrogen fixation in lentil and pea. *Can. J. Plant Sci.* **77**, 197–200 (1997).
107. McDonagh, J. F., Toomsan, B., Limpinuntana, V. & Giller, K. E. Grain legumes and green manures as pre-rice crops in northeast Thailand I. Legume N<sub>2</sub>-fixation, production and residual nitrogen benefits to rice. *Plant Soil* **177**, 111–126 (1995).
108. McEwen, J., Darby, R. J., Hewitt, M. V. & Yeoman, D. P. Effects of field beans, fallow, lupins, oats, oilseed rape, peas, ryegrass, sunflowers and wheat on nitrogen residues in the soil and on the growth of a subsequent wheat crop. *J. Agr. Sci.* **115**, 209–219 (1989).
109. Miller, P. R., McDonald, C. L., Derksen, D. A. & Waddington, J. The adaptation of seven broadleaf crops to the dry semiarid prairie. *Can. J. Plant Sci.* **81**, 29–43 (2001).
110. Miller, P. R., Gan, Y., McConkey, B. G. & McDonald, C. L. Pulse crops for the northern Great Plains: I. Grain productivity and residual effects on soil water and nitrogen. *Agron. J.* **95**, 972–979 (2003).
111. Miller, P. R. & Holmes, J. A. Cropping sequence effects of four broadleaf crops on four cereal crops in the northern Great Plains. *Agron. J.* **97**, 189–200 (2005).
112. Miller, P. R., Brandt, S. A., McDonald, C. L. & Waddington, J. Chickpea, lentil, and pea response to delayed spring seeding on the northern Great Plains. *Can. J. Plant Sci.* **86**, 1059–1070 (2006).
113. Moschini, V. *et al.* Performance of organic grain legumes in Tuscany. *Ital. J. Agron.* **9**, 38–43 (2014).
114. Muchow, R. C. Phenology, seed yield and water use of grain legumes grown under different soil water regimes in a semi-arid tropical environment. *Field Crop. Res.* **11**, 81–97 (1985).
115. Muchow, R. C., Robertson, M. J. & Pengelly, B. C. Accumulation and partitioning of biomass and nitrogen by soybean, mungbean and cowpea under contrasting environmental conditions. *Field Crop. Res.* **33**, 13–36 (1993).
116. Muhammad, A., Dikko, A. U., Audu, M. & Singh, A. Comparative effects of cowpea and soybean genotypes on N<sub>2</sub>-fixation and N-balance in Sokoto dry sub-humid agro-ecological zone of Nigeria. *Niger. J. Basic Appl. Sci.* **18**, 297–303 (2010).
117. Muthomi, J. W., Otieno, P. E., Chemining'wa, G. N., Nderitu, J. H. & Wagacha, J. M. Effect of chemical spray on insect pests and yield quality of food grain legumes. *J. Entomol.* **5**, 156–163 (2008).
118. Nan, Z. B., Abd El-Moneim, A. M., Larbi, A. & Nie, B. Productivity of vetches (*Vicia* spp.) under alpine grassland conditions in China. *Trop. Grasslands* **40**, 177–182 (2006).
119. Ncube, B., Twomlow, S. J., van Wijk, M. T., Dimes, J. P. & Giller, K. E. Productivity and residual benefits of grain legumes to sorghum under semi-arid conditions in southwestern Zimbabwe. *Plant Soil* **299**, 1–15 (2007).
120. Neuschwandtner, R. W., Wagenstristl, H. & Kaul, H.-P. Nitrogen yield and nitrogen use of chickpea compared to pea, barley and oat in central Europe. *Int. J. Plant Prod.* **9**, 291–304 (2015).
121. Nielsen, D. C. Production functions for chickpea, field pea, and lentil in the central Great Plains. *Agron. J.* **93**, 563–569 (2001).
122. Nijä, S. *et al.* Results obtained for soybean, pea and lentils crops on a Cambic Chernozem in the Banat's Plain during 2008–2010. *Rom. Agr. Res.* **29**, 155–162 (2012).
123. Ojiem, J. O., Vanlauwe, B., de Ridder, N. & Giller, K. E. Niche-based assessment of contributions of legumes to the nitrogen economy of western Kenya smallholder farms. *Plant Soil* **292**, 119–135 (2007).
124. Pala, M., Harris, H. C., Ryan, J., Makboul, R. & Dozom, S. Tillage systems and stubble management in a Mediterranean-type environment in relation to crop yield and soil moisture. *Exp. Agr.* **36**, 223–242 (2000).
125. Pandey, R. K., Herrera, W. A. T. & Pendleton, J. W. Drought response of grain legumes under irrigation gradient: I. Yield and yield components. *Agron. J.* **76**, 549–553 (1984).
126. Pandey, R. K., Morris, R. A. & Whisler, F. D. Water extraction patterns, water use and yield of ten upland crops following rainfed lowland rice in the tropics. *Philipp. J. Crop Sci.* **12**, 163–168 (1987).
127. Payne, W. A., Chen, C. & Ball, D. A. Agronomic potential of narrow-leaved and white lupins in the inland Pacific northwest. *Agron. J.* **96**, 1501–1508 (2004).
128. Phoomthaisong, J., Toomsan, B., Limpinuntana, V., Cadisch, G. & Patanothai, A. Attributes affecting residual benefits of N<sub>2</sub>-fixing mungbean and groundnut cultivars. *Biol. Fert. Soils* **39**, 16–24 (2003).
129. Piha, M. I. & Munns, D. N. Nitrogen-fixation capacity of field-grown bean compared to other grain legumes. *Agron. J.* **79**, 690–696 (1987).
130. Pilbeam, C. J., Wood, M. & Mugane, P. G. Nitrogen use in maize-grain legume cropping systems in semi-arid Kenya. *Biol. Fert. Soils* **20**, 57–62 (1995).
131. Plaza-Bonilla, D., Nolot, J.-M., Passot, S., Raffaillac, D. & Justes, E. Grain legume-based rotations managed under conventional tillage need cover crops to mitigate soil organic matter losses. *Soil Till. Res.* **156**, 33–43 (2016).



132. Pospíšil, A. & Pospíšil, M. Influence of sowing density on agronomic traits of lupins (*Lupinus* spp.). *Plant Soil Environ.* **61**, 422–425 (2015).
133. Pratibha, G., Pillai, K. G. & Satyanarayana, V. Production potential and profitability of some rice (*Oryza sativa*)-based cropping systems involving sequence cropping of pulses and oilseeds in rice fallows. *Indian J. Agr. Sci.* **67**, 98–101 (1997).
134. Prasad, N. K. & Kumar, S. Nutrient harvest and soil fertility as influenced by legume-wheat (*Triticum aestivum*) sequences. *Indian J. Agr. Sci.* **69**, 167–171 (1999).
135. Przednowek, D. W. A., Entz, M. H., Irvine, B., Flaten, D. N. & Thiessen Martens, J. R. Rotational yield and apparent N benefits of grain legumes in southern Manitoba. *Can. J. Plant Sci.* **84**, 1093–1096 (2004).
136. Rao, S. C. & Northup, B. K. Capabilities of four novel warm-season legumes in the southern Great Plains: grain production and quality. *Crop Sci.* **49**, 1103–1108 (2009).
137. Rennie, R. J. & Dubetz, S. Nitrogen-15-determined nitrogen fixation in field-grown chickpea, lentil, fababean, and field pea. *Agron. J.* **78**, 654–660 (1986).
138. Ruisi, P. *et al.* Tillage effects on yield and nitrogen fixation of legumes in Mediterranean conditions. *Agron. J.* **104**, 1459–1466 (2012).
139. Ryan, J., Singh, M. & Christiansen, S. Assessment of long-term barley-legume rotations in a typical Mediterranean agro-ecosystem: grain and straw yields. *Arch. Agron. Soil Sci.* **58**, 233–246 (2012).
140. Sandaña, P. A., Harcha, C. I. & Calderini, D. F. Sensitivity of yield and grain nitrogen concentration of wheat, lupin and pea to source reduction during grain filling. A comparative survey under high yielding conditions. *Field Crop. Res.* **114**, 233–243 (2009).
141. Santonoceto, C., Monti, M. & Anastasi, U. A comparison of the agronomic performance of grasspea and faba bean in a semiarid Mediterranean environment. *Options Méditerranéennes* **60**, 223–228 (2004).
142. Schultz, J. E. Crop production in a rotation trial at Tarlee, south Australia. *Aust. J. Exp. Agr.* **35**, 865–876 (1995).
143. Schulz, S., Keatings, J. D. H. & Wells, G. J. Productivity and residual effects of legumes in rice-based cropping systems in a warm-temperate environment I. Legume biomass production and N fixation. *Field Crop. Res.* **61**, 23–35 (1999).
144. Senthong, C. & Pandey, R. K. Response of five food legume crops to an irrigation gradient imposed during reproductive growth. *Agron. J.* **81**, 680–686 (1989).
145. Sharma, S. N., Prasad, R. & Singh, S. Residual effects of growing mungbean and uridbean on the yield and nitrogen uptake of a succeeding wheat crop. *Fert. Res.* **44**, 163–168 (1996).
146. Sharma, G. L., Mahapatra, B. S., Singh, A. K. & Pandey, B. M. Role of summer legumes on productivity of maize (*Zea mays*)-wheat (*Triticum aestivum*) cropping system on sandy-loam soil of western Uttar Pradesh. *Indian J. Agr. Sci.* **68**, 295–298 (1998).
147. Siddique, K. H. M., Walton, G. H. & Seymour, M. A comparison of seed yields of winter grain legumes in western Australia. *Aust. J. Exp. Agr.* **33**, 915–922 (1993).
148. Siddique, K. H. M. & Loss, S. P. Growth and seed yield of vetches (*Vicia* spp.) in south-western Australia. *Aust. J. Exp. Agr.* **36**, 587–593 (1996).
149. Siddique, K. H. M., Loss, S. P., Regan, K. L. & Jettner, R. L. Adaptation and seed yield of cool season grain legumes in Mediterranean environments of south-western Australia. *Aust. J. Agr. Res.* **50**, 375–387 (1999).
150. Siddique, K. H. M. & Loss, S. P. Studies on sowing depth for chickpea (*Cicer arietinum* L.), faba bean (*Vicia faba* L.) and lentil (*Lens culmaris* Medik.) in a Mediterranean-type environment of south-western Australia. *J. Agron. Crop Sci.* **182**, 105–112 (1999).
151. Sidhu, M. S., Sharma, B. D., Sharma, R. K. & Singh, S. Performance of cropping systems based on winter maize (*Zea mays*) under irrigated condition. *Indian J. Agr. Sci.* **64**, 745–750 (1994).
152. Silsbury, J. H. Grain yield and flowering of some introduced grain legumes in south Australia. *Aust. J. Exp. Agr. Anim. Husb.* **15**, 556–560 (1975).
153. Sinclair, T. R. *et al.* Field and model analysis of the effect of water deficits on carbon and nitrogen accumulation by soybean, cowpea and black gram. *Field Crop. Res.* **17**, 121–140 (1987).
154. Singh, K. & Singh, S. Different wheat (*Triticum aestivum*)-based cropping systems, and their fertilizer requirement, yield and economic returns. *Indian J. Agr. Sci.* **61**, 709–714 (1991).
155. Singh, M. K., Thakur, R., Verma, U. N. & Pal, S. K. Production potential of crop sequences in plateau region of Bihar. *Indian J. Agr. Sci.* **65**, 242–245 (1995).
156. Singh, V. & Deo, R. Productivity and economics of different cropping systems under various levels of irrigation. *Indian J. Agron.* **43**, 419–425 (1998).
157. Singh, R. & Kundu, D. K. Water use by horsegram (*Macrotyloma uniflorum*), greengram (*Phaseolus radiatus*) and sesame (*Sesamum indicum*) under shallow water-table condition in Orissa. *Indian J. Agr. Sci.* **73**, 89–93 (2003).
158. Smith, S. C., Bezdicsek, D. F., Turco, R. F. & Cheng, H. H. Seasonal N<sub>2</sub> fixation by cool-season pulses based on several <sup>15</sup>N methods. *Plant Soil* **97**, 3–13 (1987).
159. Spaner, D., Todd, A. G. & McKenzie, D. B. Pea and soybean performance in Newfoundland. *Can. J. Plant Sci.* **81**, 723–726 (2001).
160. Srivastava, G. P. & Srivastava, V. C. Nitrogen economy and productivity of wheat (*Triticum aestivum*) succeeding grain legumes. *Indian J. Agr. Sci.* **63**, 694–698 (1993).
161. Stringi, L., Giambalvo, D., Trapani, P., Scarpello, C. & Attardo, C. Effect of soil management techniques on performance of different grain legumes in a Mediterranean environment. *Options Méditerranéennes* **60**, 79–83 (2004).
162. Suwanarit, A., Suwanarit, C. & Chotechaungmanirat, S. Quantities of fixed N and effects of grain legumes on following maize, and N and P status of soil as indicated by isotopes. *Plant Soil* **93**, 249–258 (1986).
163. Tanaka, D. L., Anderson, R. L. & Rao, S. C. Crop sequencing to improve use of precipitation and synergize crop growth. *Agron. J.* **97**, 385–390 (2005).
164. Tanaka, D. L., Krupinsky, J. M., Merrill, S. D., Liebig, M. A. & Hanson, J. D. Dynamic cropping systems for sustainable crop production in the northern Great Plains. *Agron. J.* **99**, 904–911 (2007).
165. Tesfaye, K., Walker, S. & Tsubo, M. Radiation interception and radiation use efficiency of three grain legumes under water deficit conditions in a semi-arid environment. *Eur. J. Agron.* **25**, 60–70 (2006).
166. Thakur, H. C. & Sharma, N. N. Production potential and economic profitability of different cropping patterns including cereals, pulses and oilseeds. *Indian J. Agr. Sci.* **58**, 194–201 (1988).
167. Thakur, R. B., Jha, G. & Choudhary, S. K. Production and economics of rice (*Oryza sativa*)-based cropping systems. *Indian J. Agron.* **43**, 572–575 (1998).
168. Thies, J. E., Singleton, P. W. & Bohlool, B. B. Influence of the size of indigenous rhizobial populations on establishment and symbiotic performance of introduced rhizobia on field-grown legumes. *Appl. Environ. Microbiol.* **57**, 19–28 (1991).
169. Thomas, G. A. *et al.* Crop rotations for sustainable grain production on a Vertisol in the semi-arid tropics. *J. Sustain. Agr.* **35**, 2–26 (2011).



170. Thomson, B. D., Siddique, K. H. M., Barr, M. D. & Wilson, J. M. Grain legume species in low rainfall Mediterranean-type environments I. Phenology and seed yield. *Field Crop. Res.* **54**, 173–187 (1997).
171. Toomsan, B. *et al.* Biological N<sub>2</sub> fixation and residual N benefit of pre-rice leguminous crops and green manures. *Neth. J. Agr. Sci.* **48**, 19–29 (2000).
172. Turpin, J. E., Herridge, D. F. & Robertson, M. J. Nitrogen fixation and soil nitrate interactions in field-grown chickpea (*Cicer arietinum*) and fababean (*Vicia faba*). *Aust. J. Agr. Res.* **53**, 599–608 (2002).
173. Vasilakoglou, I., Vlachostergios, D., Dhima, K. & Lithourgidis, A. Response of vetch, lentil, chickpea and red pea to pre- or post-emergence applied herbicides. *Span. J. Agr. Res.* **11**, 1101–1111 (2013).
174. Wall, D. A. Lentil (*Lens culinaris*) and fababean (*Vicia faba*) tolerance to post-emergence applications of Imazethapyr. *Can. J. Plant Sci.* **76**, 525–529 (1996).
175. Wang, X., Gan, Y., Hamel, C., Lemke, R. & McDonald, C. Water use profiles across the rooting zones of various pulse crops. *Field Crop. Res.* **134**, 130–137 (2012).
176. Webber, H. A. *et al.* Water use efficiency of common bean and green gram grown using alternate furrow and deficit irrigation. *Agr. Water Manage.* **86**, 259–268 (2006).
177. Wen, G., Chen, C., Neill, K., Wichman, D. & Jackson, G. Yield response of pea, lentil and chickpea to phosphorus addition in a clay loam soil of central Montana. *Arch. Agron. Soil Sci.* **54**, 69–82 (2008).
178. Wendt, J. W. & Atemkeng, M. F. Soybean, cowpea, groundnut, and pigeonpea response to soils, rainfall, and cropping season in the forest margins of Cameroon. *Plant Soil* **263**, 121–132 (2004).
179. Wetselaar, R. Estimation of nitrogen fixation by four legumes in a dry monsoonal area of north-western Australia. *Aust. J. Exp. Agr. Anim. Husb.* **7**, 518–522 (1967).
180. Withers, N. J. A comparison of several grain legumes at two sowing times I. Seed yield and components. *New Zeal. J. Exp. Agr.* **7**, 361–364 (1979).
181. Wright, A. T. Yield effect of pulses on subsequent cereal crops in the northern prairies. *Can. J. Plant Sci.* **70**, 1023–1032 (1990).
182. Yau, S. K. & Ryan, J. Differential impacts of climate variability on yields of rainfed barley and legumes in semi-arid Mediterranean conditions. *Arch. Agron. Soil Sci.* **59**, 1659–1674 (2013).
183. Yusuf, A. A., Abaidoo, R. C., Iwuafor, E. N. O. & Olufajo, O. O. Genotype effects of cowpea and soybean on nodulation, N<sub>2</sub>-fixation and N balance in the northern Guinea savanna of Nigeria. *J. Agron.* **7**, 258–264 (2008).
184. Zentner, R. P. *et al.* Economics of crop diversification and soil tillage opportunities in the Canadian prairies. *Agron. J.* **94**, 216–230 (2002).
185. Zhang, H., Pala, M., Oweis, T. & Harris, H. Water use and water-use efficiency of chickpea and lentil in a Mediterranean environment. *Aust. J. Agr. Res.* **51**, 295–304 (2000).
186. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **15**, 259–263 (2006).
187. Daryanto, S., Wang, L. & Jacinthe, P.-A. Global synthesis of drought effects on food legume production. *PLoS ONE* **10**, e0127401 (2015).
188. Laurent, A., Pelzer, E., Loyce, C. & Makowski, D. Ranking yields of energy crops: a meta-analysis using direct and indirect comparisons. *Renew. Sust. Energ. Rev.* **46**, 41–50 (2015).

#### Data Citation

1. Cernay, C., Pelzer, E. & Makowski, D. *Dryad Digital Repository* <http://dx.doi.org/10.5061/dryad.mf42f> (2016).

#### Acknowledgements

We thank G. Amato, C. Chen, A. L. Fernandez, Y. T. Gan, K. E. Giller, N. Gogoi, D. Herridge, H.-P. Kaul, G. Kirchhof, M. A. Liebig, P. Miller, R. Neugschwandtner, K. H. M. Siddique, D. Spaner, M. Unkovich, C. S. Wortmann and A. A. Yusuf for providing us with additional information and raw data. We thank A. Bône for assistance with the literature search. We thank D. Beillouin, T. Ben-Ari, G. Corre-Hellou, L. Hossard, M.-H. Jeuffroy, J.-M. Meynard, O. Réchauchère, A. Schneider, J.-M. Teulé and A.-S. Voisin for insightful comments. This work was supported by the French National Research Agency (ANR) under the 'Investments for the Future' program (ANR-10-IDEX-0003-02) as part of the LabEx BASC (ANR-11-LABX-0034).

#### Author Contributions

C.C. compiled data records, performed standardization and technical validation of the database. D.M. and E.P. conceived and supervised the work. C.C., D.M. and E.P. wrote the Data Descriptor.

#### Additional Information

Table 3 is only available in the online version of this paper.

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cernay, C. *et al.* A global experimental dataset for assessing grain legume production. *Sci. Data* **3**:160084 doi: 10.1038/sdata.2016.84 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.

© The Author(s) 2016

## Chapitre 3



## Chapitre 3

### **Comparer les niveaux de production des légumineuses à graines en Europe, Amérique du Nord et Océanie**

Ce chapitre est un article scientifique soumis pour publication dans une revue à comité de lecture.

Cernay, C., Makowski, D. & Pelzer, E. Ranking grain legume productivity in Europe, North America and Oceania. (2016).

Les informations supplémentaires sont en Annexes.

La correspondance entre les noms communs et les noms scientifiques des espèces est indiquée en Annexes, dans le Tableau Supplémentaire 1.

## Ranking grain legume productivity in Europe, North America and Oceania

Charles Cernay, David Makowski & Elise Pelzer

UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

### Abstract

Legumes are grown for their protein-rich grains for both human food and animal feed, and their multiple benefits to the environment. Pea (*Pisum sativum*) is the most widely cultivated grain legume in Europe over the last fifty years, with other species covering smaller areas. The most productive grain legume species need to be selected by farmers to satisfy an increasing protein demand, and to diversify protein sources. Several field experiments have compared the productivity levels of a broad range of grain legumes in contrasting environments, but these experiments have never been synthesized. A quantitative synthetic analysis of these individual experiments would help European plant breeders, agricultural companies, and policy makers, to identify grain legume species of interest for food and feed. We address two key questions: (1) Which grain legume species have productivity levels lower than, similar to, and higher than, that of pea in Europe? (2) Which grain legume species with good productivity performances in North America and Oceania are potential candidates for inclusion in future European experiments? We conduct a meta-analysis comparing 22 grain legume species with pea, according to different criteria based on experimental data collated from 61 peer-reviewed articles, on official agricultural statistics, and on nutritional data. We identify soybean (*Glycine max*), narrowleaf lupin (*Lupinus angustifolius*), and faba bean (*Vicia faba*), as interesting alternative grain legumes to pea in Europe. We show that replacing 25% of the area currently under pea with soybean would increase the mean crude protein production by +28% in Europe. The replacement of pea with narrowleaf lupin and faba bean would also increase the mean crude protein production in this world region, but only up to +4% and +3%, respectively. Based on data collected in North America and Oceania, we suggest assessing the relative productivities of several vetches and lupins (*Lathyrus*, *Lupinus*, and *Vicia* species) in new field European experiments. Our findings reveal new comprehensive insights into the productivity levels of as yet underexploited grain legume species for potential future use in Europe.

## 1. Introduction

Legumes are often cultivated for the grains as a source of protein and energy for human and animal consumption. Legume grains may indeed represent a nutritious complement to cereal grains for human diet and animal feed [1-4]. It is assumed that legumes provide environmental benefits and potential economic returns [5-8]. Legume crops are able to fix atmospheric nitrogen, require generally low nitrogen input conditions [5,6], and produce low greenhouse gas emissions [9,10]. Grain legume cultivation may also enhance soil nitrogen supply through mineralization [11-13], and increase the yields of subsequent cereals and oilseeds in the crop sequence [14-17].

From 2004 to 2014, the area under grain legumes accounted for less than 2% of the cultivated area in Europe [18]. The European Parliament adopted a resolution in 2011 including support for the expansion of grain legume production in this region [19]. Between 1961 and 2014, pea (*Pisum sativum*) was the most widely cultivated grain legume species in Europe [18,20]. More recently, there has been a substantial increase in the cultivated area under soybean (*Glycine max*) in several European countries [18,20]. Other grain legume species occupy smaller proportions of the cultivated area. In 2014, pea and soybean together accounted for 76% of the area under grain legumes in Europe, whereas only 15% of the area under grain legumes was covered by chickpea (*Cicer arietinum*), faba bean (*Vicia faba*), garden vetch (*Vicia sativa*), kidney bean (*Phaseolus vulgaris*), lentil (*Lens culinaris*), and lupins (*Lupinus* species) [18].

This large imbalance in the areas under the different grain legume species is the consequence of two concomitant phenomena in Europe. First, only a small number of grain legume species, including pea, have been investigated as model species in research programs [21–26]. Second, the legume breeding sector is fragmented into a small number of private companies, working on only a few species [23,25,27–29], despite reports of potential environmental and nutritional benefits associated with the cultivation of alternative legumes species [14,30-33].

Europe could learn valuable lessons for the diversification of grain legume crops from the expansion of legume production in North America and Oceania. Both these regions have ranges of climatic conditions similar to those found in Europe [34]. However, North America and Oceania together exported eleven times more legumes than Europe over a period extending from 2003 to 2013 (considering that exports are linked to harvested areas) [18]. Soybean is the predominant grain legume in North American agriculture, but alternative grain legume species have become more popular since the 1980s, for example in the northern Great Plains as a means of diversifying cereal-fallow rotations [35–39]. In Australia, crop breeders, agronomists, and industry, have been working closely together since the 1970s and 1980s, to develop the commercial production of a broad range of grain legume

species [40,41]. This also relied on the establishment of experiments specifically carried out to compare the productivities of a large number of legume species but for a few criteria at a time from a single data source [42-49]. Yet, to the best of our knowledge, there is no multi-criteria comparison of a broad range of grain legume species based on several data sources. Indeed, the productive performances of grain legumes can be assessed from at least three different data sources: (1) official agricultural statistics, (2) results from field experiments comparing grain legume species, and (3) nutritional data extracted from food and feed composition tables.

In this study, we chiefly focus on Europe, but we rely on data collected in both North America and Oceania in order to identify grain legume species that could be cultivated in Europe in the future. We address two key questions: (1) Which grain legume species have productivities lower than, similar to, and higher than, that of pea in Europe? (2) Which grain legume species that have been directly compared with pea in North America and Oceania but not in Europe may be considered as potential candidates for inclusion in new field European experiments?

We perform a meta-analysis comparing 22 grain legume species with pea as a reference species. All grain legumes considered in this study are compared on the basis of seven criteria related to different aspects of their productivity: (1) grain yield (hereafter referred to as grain biomass), (2) total aerial biomass, (3) grain crude protein, (4) grain gross energy, (5) grain nitrogen content, (6) water use, and (7) interannual variability in grain biomass (hereafter referred to as grain biomass interannual variability). We rank each species according to each criterion.

Our analysis provides European farmers, plant breeders, agricultural companies and policy makers with new comprehensive insights into grain legume productivity. In particular, our results identify the grain legumes with the greatest productivity potential for expansion in Europe.

## **2. Methods**

### **2.1. Literature search**

We use a comprehensive query on the Institute for Scientific Information Web of Science [50]. No limitations are imposed concerning the date and language of publication, and we query article titles, abstracts and keywords. The search equation is: ‘crop\* AND (legum\* OR pulse\*) AND (yield\* OR ‘dry matter’ OR biomass) AND (compar\* OR assessment OR product\* OR performance\*) AND (trial\* OR factorial OR experiment\* OR treatment\* OR condition\*) NOT (intercrop\* OR catch OR cover OR ‘green manure’ OR forage OR fodder)’. We identify 8,386 potentially relevant articles in total, and the literature search was completed on February 15, 2016.

## 2.2. Article selection

Articles are selected on the basis of the following nine criteria: (1) at least two grain legume species grown as annual sole crops, (2) at least two grain legume species grown at the same field site during the same growing season, (3) at least one experiment carried out during one growing season, from seeding to harvest, (4) either the title or the abstract of the article written in English, (5) full-text article published in a peer-reviewed journal, (6) full-text article available, (7) either raw data not redundantly reported in other articles or raw data collected after contacting the authors, (8) grain biomass data reported individually for each species, and (9) field site location or soil site characteristics precisely stated. Experiments including only one legume species are excluded because it is not possible to compare this species directly with other species in the same environment (e.g., climate conditions, and soil types) with the same crop management techniques (e.g., tillage, fertilization, pest control, and irrigation). We select 173 articles, in total.

## 2.3. Data extraction

Relevant information is extracted manually or digitally from the text, tables, and figures of the selected 173 articles. Crop management techniques (i.e., tillage, fertilization, pest control, and irrigation) are systematically extracted from each selected article. All the data are included into a global experimental data set. Further information can be found in the Data Descriptor published by Cernay *et al* 2016 [51]. We use only a subset of this global data set for this study. This subset includes data extracted from 61 peer-reviewed articles. The references of these articles are provided in the Supplementary Materials (note S1). All these articles report results from field experiments carried out in Europe, North America or Oceania (figure 1), include data for pea, and for at least one other grain legume species (table S1).



**Figure 1.** Experimental field sites included in the meta-analysis. The study regions are: Europe (a), North America (b), and Oceania (c). Each red bullet point indicates the location of an experimental field site. In Europe, field sites are located in Austria, Denmark, France, Germany, Italy, the United Kingdom, and Romania. In North America, field sites are located in Canada, and the United States of America. In Oceania, field sites are located in Australia, and New Zealand. Basemap from Google Maps 2016.



Data are extracted for the following productivity variables (tables S2 and S3): grain biomass, total aerial biomass, harvest index, grain nitrogen content, percentage of nitrogen in grains, water use, and grain water use efficiency. The names and definitions of all variables included in the data set are provided in table S4. The data set includes data for 22 grain legume species. We choose pea – the grain legume species most widely cultivated, on average, in Europe, over the 1961-2014 period – as a reference species for all regions. Both grain crude protein and grain gross energy are calculated for each legume species from the nutritional tables of the Animal Feed Resources Information System database (hereafter referred to as FEEDIPEDIA database) [52] (table S5).

#### 2.4. Statistical analysis for estimating mean ratios

All grain legume species are directly compared with pea (i.e., a reference species) by estimating mean log-transformed ratios for six productivity criteria: (1) grain biomass, (2) total aerial biomass, (3) grain crude protein, (4) grain gross energy, (5) grain nitrogen content, and (6) water use. For each criterion

$X$ , the log-transformed ratio, as in (1):  $L_{ijk} = \log_e \left( \frac{\bar{X}_{ijk}}{\bar{X}_{rjk}} \right)$ , is calculated for each species  $i$  grown in

each field site\*growing season  $k$  in each article  $j$ .  $\bar{X}_{ijk}$  is the value of  $X$  averaged over crop management techniques (i.e., tillage, fertilization, pest control, and irrigation) for species  $i$  grown in field site\*growing season  $k$  in article  $j$ , and  $\bar{X}_{rjk}$  is the value of  $X$  averaged over the same crop management techniques for pea grown in the same field site\*growing season in the same article. The mean log-transformed ratio is estimated for each criterion  $X$  and each species  $i$  across all field site\*growing season combinations and selected articles, with linear random-effect models. Model variants including one random ‘field site\*growing season’ effect or two random ‘article’ and ‘field site\*growing season’ effects are fitted in turn, with the ‘lme’ function of the ‘nlme’ package (version 3.1.111) [53] of the R software (version 3.0.2) [54]. The best model variant is selected on the basis of the Akaike Information Criterion (hereafter referred to as AIC) [55]. Mean ratios  $R_i$  and 95% confidence intervals are then calculated from the exponential of the estimated mean log-transformed

ratios for each criterion  $X$  and each species  $i$ , as in (2):  $R_i = e^{\left( \frac{\bar{X}_i}{\bar{X}_r} \right)}$ . Estimated mean ratios are considered to be significantly different from one (i.e., the species show significant higher or lower performance levels compared with that of pea for a given criterion) if their 95% confidence intervals do not include one. This analysis is then repeated, weighting the log-transformed ratios  $L_{ijk}$  by their precisions. As the variances of  $\bar{X}_{ijk}$  and  $\bar{X}_{rjk}$  are available for only a small number of the original measurements, the log-transformed ratios  $L_{ijk}$  are weighted on the basis of the number of replicates

$n_{ijk}$  and  $n_{rjk}$  used to calculate  $\bar{X}_{ijk}$  and  $\bar{X}_{rjk}$ , respectively. In situations in which a given criterion  $X$  is averaged over more than one growing season, the number of replicates is considered to be the sum of the number of replicates over all growing seasons. The weight  $\omega_{ijk}$  of each log-transformed ratio

$L_{ijk}$  is calculated, as in (3):  $\omega_{ijk} = \frac{n_{ijk}n_{rjk}}{n_{ijk} + n_{rjk}}$ . This weight is a generalization of the sample size weight

described by Wiebe *et al* (2006) [56]. Species rankings obtained with and without data weighting are compared. We use weighted values to assess sensitivity to model assumptions for the two model variants, separately: (1) inclusion of one random ‘field site\*growing season’ effect (table S6), and (2) inclusion of two random ‘article’ and ‘field site\*growing season’ effects (table S7). We also use weighted values to analyse whether raw residuals by species, standardized residuals by species, and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles are symmetrically distributed in shape for each criterion in each region (figure S1 to figure S15). Mean ratios are estimated independently for Europe, North America and Oceania.

## 2.5. Statistical analysis for estimating the interannual variability in grain biomass

For the identification of species displaying lower or higher levels of interannual variability in grain biomass than that of pea, we estimate interannual variability as described by Cernay *et al* (2015) [57], using time series provided by the Statistics Division of Food and Agriculture Organization of the United Nations database (hereafter referred to as FAOSTAT database) [18]. Grain biomass data are retrieved for Europe, North America and Oceania with the FAOSTAT region items ‘Europe’, ‘Northern America’, and ‘Australia & New Zealand’ (table S8), respectively. Times series are reported only for grain biomass in the FAOSTAT database. It is not, therefore, possible to estimate interannual variability for total aerial biomass, grain nitrogen content, and water use. Most of the time series span the period from 1961 to 2014 (i.e., 54 years), but there are four exceptions: (1) chickpea (i.e., 1992-2014, 22 years), and (2) faba bean (i.e., 1982-2007, 26 years) in North America, (3) chickpea (i.e., 1983-2014, 32 years), and (4) lentil (i.e., 1985-2014, 30 years) in Oceania. For each FAOSTAT region item\*FAOSTAT crop item combination, polynomial regression models are fitted to time series for the estimation of grain biomass trend. Linear, quadratic and cubic models are fitted in turn to each time series, with the ‘lm’ function of the ‘stats’ package of the R software (version 3.0.2) [54]. The model with the lowest AIC is selected, and used to calculate both grain biomass trend and normalized grain biomass residuals. We estimate the ratio of the variance of normalized grain biomass residuals for a given FAOSTAT crop item to that of pea (i.e., the FAOSTAT ‘Peas, dry’ crop item) for each FAOSTAT region item. We use the ‘var.test’ function of the ‘stats’ package of the R software (version 3.0.2) [54], to estimate both variance ratios and 95% confidence intervals. Variance ratios are considered to be significantly different from one (i.e., the species show significant higher or lower levels of interannual

variability in grain biomass compared with that of pea) if their 95% confidence intervals do not include one.

## 2.6. Effect of replacing pea with other grain legume species in Europe

We investigate the effect on crude protein production in Europe of replacing pea with other grain legume species in cultivated areas. Based on estimated mean grain crude protein ratios, we identify the three species with the highest grain crude protein levels relative to that of pea in Europe. We assume that a given percentage  $z$  of the cultivated area under pea ranging from 0% to 25% is instead cultivated with each of these three alternative grain legume species. The expected mean relative percentage change in crude protein production  $\Delta$  resulting from this replacement over a fraction  $z$  of the cultivated area currently under pea is calculated for each alternative grain legume species in turn,

as in (4):  $\Delta = (1 - z) + z \left( \frac{CP_i}{CP_r} \right)$ , where  $\left( \frac{CP_i}{CP_r} \right)$  is the estimated mean ratio of the grain crude protein

$CP_i$  of alternative grain legume species  $i$  to the grain crude protein  $CP_r$  of pea. Mean ratios are estimated with weighted values, as explained above. An uncertainty envelope encompassing the expected relative percentage changes in crude protein production is defined for each alternative grain legume species, using the lower and upper values of the 95% confidence interval of the estimated mean grain crude protein ratio.

## 3. Results

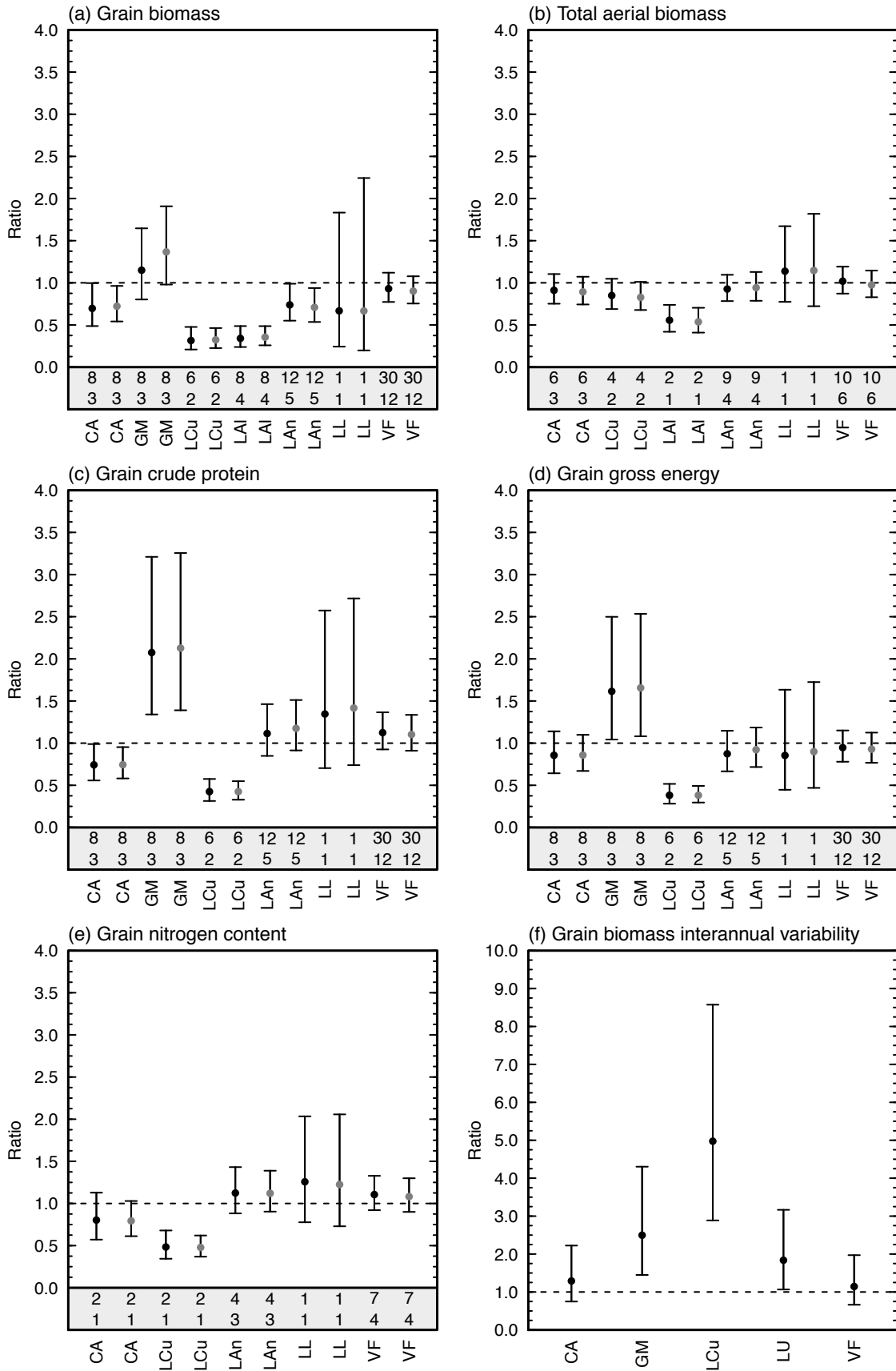
We focus below on the results obtained weighting log-transformed ratios. Very similar results are obtained unweighting log-transformed ratios (figure 2(a) to figure 2(e), figure 4(a) to figure 4(f), and figure 5(a) to figure 5(d)).

### 3.1. Good performances of soybean, narrowleaf lupin, and faba bean in Europe

The mean grain biomass ratios estimated for soybean ( $R=1.37$ ), faba bean ( $R=0.90$ ) and yellow lupin ( $R=0.67$ ) in Europe are not significantly different from one (figure 2(a) and table 1); their 95% confidence intervals include one. The grain biomasses of these three species are, therefore, on average, not significantly different from that of pea for the same field sites during the same growing seasons with the same crop management techniques. The grain biomass data are extracted from twelve articles for faba bean, but from only three articles for soybean, and one article for yellow lupin. These differences in sample size result in soybean having a larger 95% confidence interval, and yellow lupin having a much larger confidence interval than that of faba bean. Thus, the results obtained for faba

bean are more robust than those obtained for the other two species. The estimated mean grain biomasses of chickpea, narrowleaf lupin, and white lupin are significantly lower than that of pea, by -28%, -29%, and -65%, respectively. Lentil also has a significant lower grain biomass than that of pea ( $R=0.32$ ), but only two articles are available for this species. The mean total aerial biomass ratios estimated for chickpea, narrowleaf lupin, and faba bean, are not significantly different from one (figure 2(b) and table 1).

EUROPE



**Figure 2.** Ratios of grain biomass (a), total aerial biomass (b), grain crude protein (c), grain gross energy (d), grain nitrogen content (e), and grain biomass interannual variability (f), estimated for seven grain legume species, relative to those estimated for pea (*Pisum sativum*) in Europe. The scientific names of the species concerned are abbreviated here as follows: *Cicer arietinum* (chickpea, CA), *Glycine max* (soybean, GM), *Lens culinaris* (lentil, LCu), *Lupinus albus* (white lupin, LAI), *Lupinus angustifolius* (narrowleaf lupin, LAN), *Lupinus luteus* (yellow lupin, LL), *Lupinus* species (lupins, LU), and *Vicia faba* (faba bean, VF). All species are ranked alphabetically according to scientific names. Bullet points indicate mean ratios estimated unweighted (black) or weighted (gray) values for each criterion. Error bars indicate 95% confidence intervals. Within the bottom box (gray), the lower and upper rows indicate the number of selected articles and log-transformed ratios, respectively.

**Table 1.** Productivity rankings for all grain legume species relative to pea (*Pisum sativum*) in Europe, North America and Oceania. All species are ranked alphabetically according to scientific names. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (no common name), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch). Mean ratios are estimated from weighted experimental data. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The symbol '-' indicates that the value of the criterion for the given species is significantly lower than that of pea, and that more than two articles are available for the species concerned. The symbol '(-)' indicates that the value of the criterion for the given species is significantly lower than that of pea, and that there are no more than two articles for the species concerned. The symbol '0' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are more than two articles for the species concerned. The symbol '(0)' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are no more than two articles available for the species concerned. The symbol '+' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are more than two articles available for the species concerned. The symbol '(+)' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are no more than two articles available for the species concerned.

	Europe						North America						Oceania					
	Grain biomass ratio <sup>a</sup>	Total aerial biomass ratio <sup>a</sup>	Grain crude protein ratio <sup>a,b</sup>	Grain gross energy ratio <sup>a,b</sup>	Grain nitrogen content ratio <sup>a</sup>	Grain biomass variance ratio <sup>a</sup>	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain nitrogen content ratio	Water use ratio <sup>a</sup>	Grain biomass variance ratio	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain biomass variance ratio
<i>Cicer arietinum</i>	-	0	-	0	(0)	0	-	-	-	(-)	+	-	-	-	-	-	-	+
<i>Glycine max</i>	0		+	+		+	-	(0)	0	-	-	-	(-)	0	0	0	0	+
<i>Lathyrus aphaca</i>														(-)	0	0	0	
<i>Lathyrus cicera</i>														(0)	0	0	0	
<i>Lathyrus cymenium</i>														(0)	0	0	0	
<i>Lathyrus ochrus</i>														(0)	0	0	0	
<i>Lathyrus sativus</i>							(0)	(0)	(0)	(0)	(+)	0	0	0	0	0	0	+
<i>Lens culinaris</i>	(-)	(0)	(-)	(-)	(-)	+	-	-	-	-	0	0	-	-	-	-	-	+
<i>Lupinus albus</i>	-	(-)				+								-	-	-	-	+
<i>Lupinus angustifolius</i>	-	0	0	0	0	+	-	-	-	(-)	(0)	0	0	0	0	0	0	+
<i>Lupinus atlanticus</i>						+								(0)	(0)	(0)	(0)	+
<i>Lupinus luteus</i>	(0)	(0)	(0)	(0)	(0)	+								(-)	(-)	(0)	(0)	+
<i>Lupinus pilosus</i>						+								(0)	(0)	(0)	(0)	+
<i>Phaseolus vulgaris</i>							-	-	-	(-)	(0)	0						
<i>Vicia articulata</i>														(0)	(0)	(0)	(0)	
<i>Vicia benghalensis</i>														(-)	(0)	(0)	(0)	
<i>Vicia ervilia</i>														(-)	(0)	(0)	(0)	
<i>Vicia faba</i>	0	0	0	0	0	0	-	(0)	0	-	(0)	0	0	+	+	+	+	
<i>Vicia hybrida</i>														(-)	0	0	0	+
<i>Vicia narbonensis</i>														0	(0)	0	0	+
<i>Vicia sativa</i>														0	0	0	0	+
<i>Vicia villosa</i>														(-)	(-)	(-)	(-)	

<sup>a</sup> Experimental data from the selected articles are used to estimate the given criterion.

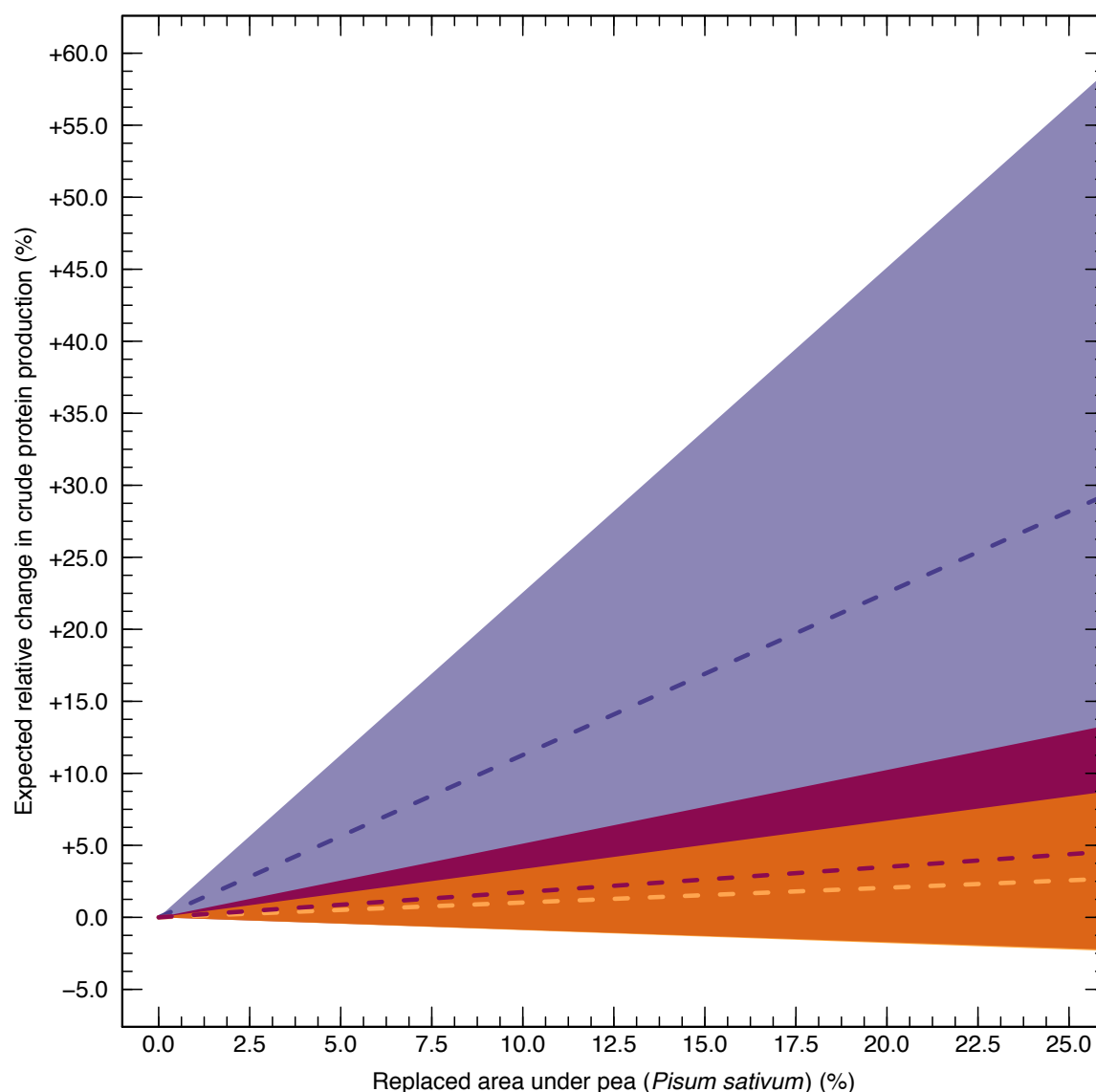
<sup>b</sup> Nutritional data from the FEEDIPEDIA database are used to estimate the given criterion.

<sup>c</sup> Statistical data from the FAOSTAT database are used to estimate the given criterion.

Species rankings are similar for the grain crude protein and grain gross energy criteria (figure 2(c), figure 2(d), and table 1). The mean grain crude protein and mean grain gross energy ratios estimated for soybean are 2.13 and 1.66, respectively. The mean grain crude protein and mean grain gross energy ratios estimated for narrowleaf lupin and faba bean are not significantly different from one. The mean grain crude protein ratio estimated for chickpea is significantly lower than one, indicating a mean grain crude protein level -26% lower than that of pea. The species rankings obtained for grain nitrogen content are similar to those obtained for grain crude protein (figure 2(c), figure 2(e), and table 1), except chickpea.

Replacing 25% of the area currently under pea in Europe with soybean would increase crude protein production by +28%, according to the estimated mean ratio for soybean (figure 3). The 95% confidence interval is large (i.e., ranging from +10% to +56%), but does not overlap zero. The replacement of pea with narrowleaf lupin and faba bean would also increase crude protein production, but to a lesser extent (i.e., +4% with the 95% confidence interval ranging from -2% to +13%, and +3% with the 95% confidence interval ranging from -2% to +8%, respectively; figure 3).





**Figure 3.** Expected relative percentage changes in crude protein production as a function of the percentage of the area under pea (*Pisum sativum*) in Europe replaced with soybean (*Glycine max*, purple), narrowleaf lupin (*Lupinus angustifolius*, red), and faba bean (*Vicia faba*, orange). Dashed lines indicate the expected mean relative percentage changes in crude protein production. Envelopes indicate 95% confidence intervals.

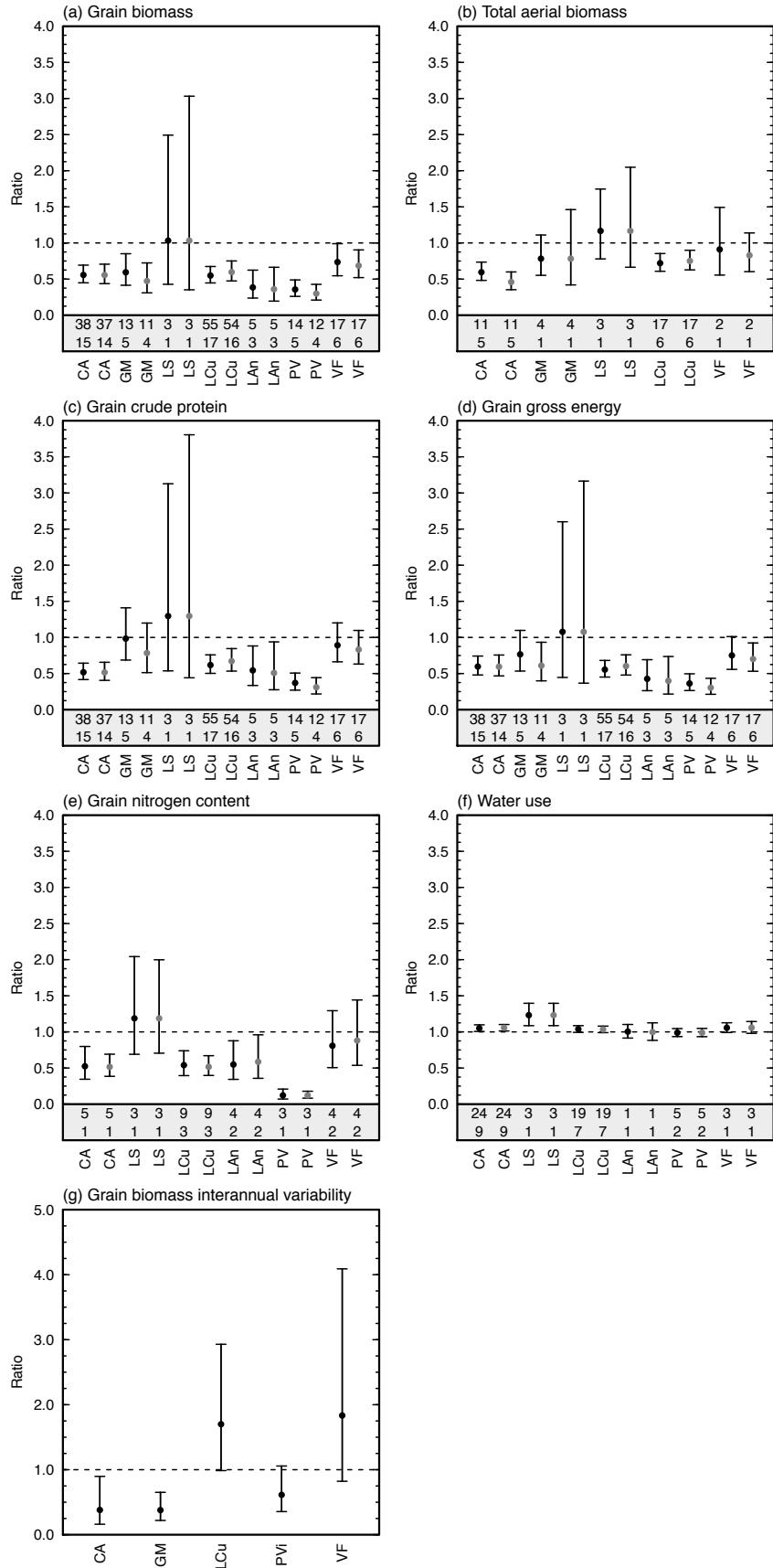
Residuals are symmetrically distributed around zero for all species and criteria (figure S1 to figure S5).

The grain biomass variance ratios estimated for chickpea and faba bean are not significantly different from one, indicating an interannual variability in their grain biomasses similar to that of pea (figure 2(f) and table 1). By contrast, interannual variability in grain biomass is significantly higher in lupins, soybean, and lentil, than in pea, by +84%, +150%, and +398%, respectively.

### 3.2. Most of grain legume species considered do not outperform pea in North America

The other grain legume species grown in North America perform less well than pea for most of the criteria, with few exceptions. Estimated mean grain biomass is significantly lower than that of a reference species for all species, except white pea (*Lathyrus sativus*) (figure 4(a) and table 1). The uncertainty on the estimated mean grain biomass ratio for this species is very high, due to the small number of articles on white pea. All the other species have significantly lower mean grain biomasses than that of pea. The mean grain biomasses estimated for faba bean and lentil are -31% and -40% lower than that of pea, respectively. Mean grain biomass is -45% lower for chickpea, -53% lower for soybean, -64% lower for narrowleaf lupin, and -70% lower for kidney bean. These lower mean grain biomasses probably reflect the location of most of the experiments in the northern Great Plains, where the grain biomass of pea tends to be higher than those of other grain legume species [36]. The mean total aerial biomass ratios estimated for lentil and chickpea are significantly lower than one (figure 4(b) and table 1). The differences in this ratio are not significant for soybean, white pea, and faba bean, but few data are available for these three species.

NORTH AMERICA



**Figure 4.** Ratios of grain biomass (a), total aerial biomass (b), grain crude protein (c), grain gross energy (d), grain nitrogen content (e), water use (f), and grain biomass interannual variability (g), estimated for seven grain legume species, relative to those estimated for pea (*Pisum sativum*) in North America. The scientific names of the species concerned are abbreviated here as follows: *Cicer arietinum* (chickpea, CA), *Glycine max* (soybean, GM), *Lathyrus sativus* (white pea, LS), *Lens culinaris* (lentil, LCu), *Lupinus angustifolius* (narrowleaf lupin, LAn), *Phaseolus vulgaris* (kidney bean, PV), *Phaseolus* and *Vigna* species (beans, PVi), and *Vicia faba* (faba bean, VF). All species are ranked alphabetically according to scientific names. Bullet points indicate mean ratios estimated unweighted (black) or weighted (gray) values for each criterion. Error bars indicate 95% confidence intervals. Within the bottom box (gray), the lower and upper rows indicate the number of selected articles and log-transformed ratios, respectively.

The mean grain crude protein ratios estimated for soybean and faba bean are 0.78 and 0.83 (figure 4(c) and table 1), respectively. Both these species have performances similar to that of pea for this criterion. For the other species, the estimated mean grain crude protein ratios are either significantly lower than one (i.e., chickpea, lentil, narrowleaf lupin, and kidney bean) or very uncertain (i.e., white pea). The mean grain gross energy estimated for all species is significantly lower than that of a reference species (figure 4(d) and table 1), except white pea, but with more uncertainty. The estimated mean grain nitrogen content ratios (figure 4(e) and table 1) are significantly lower than one for all species, with two exceptions: white pea and faba bean.

The water use results show that no species consume significantly less water than pea during the growing season (figure 4(f) and table 1). Chickpea and white pea use +6% and +23% more water, respectively, than pea. Estimates for white pea are very uncertain due to the small number of data available for this species. The estimated mean water use ratios are not significantly different from one for the other grain legume species (i.e., their 95% confidence intervals include one).

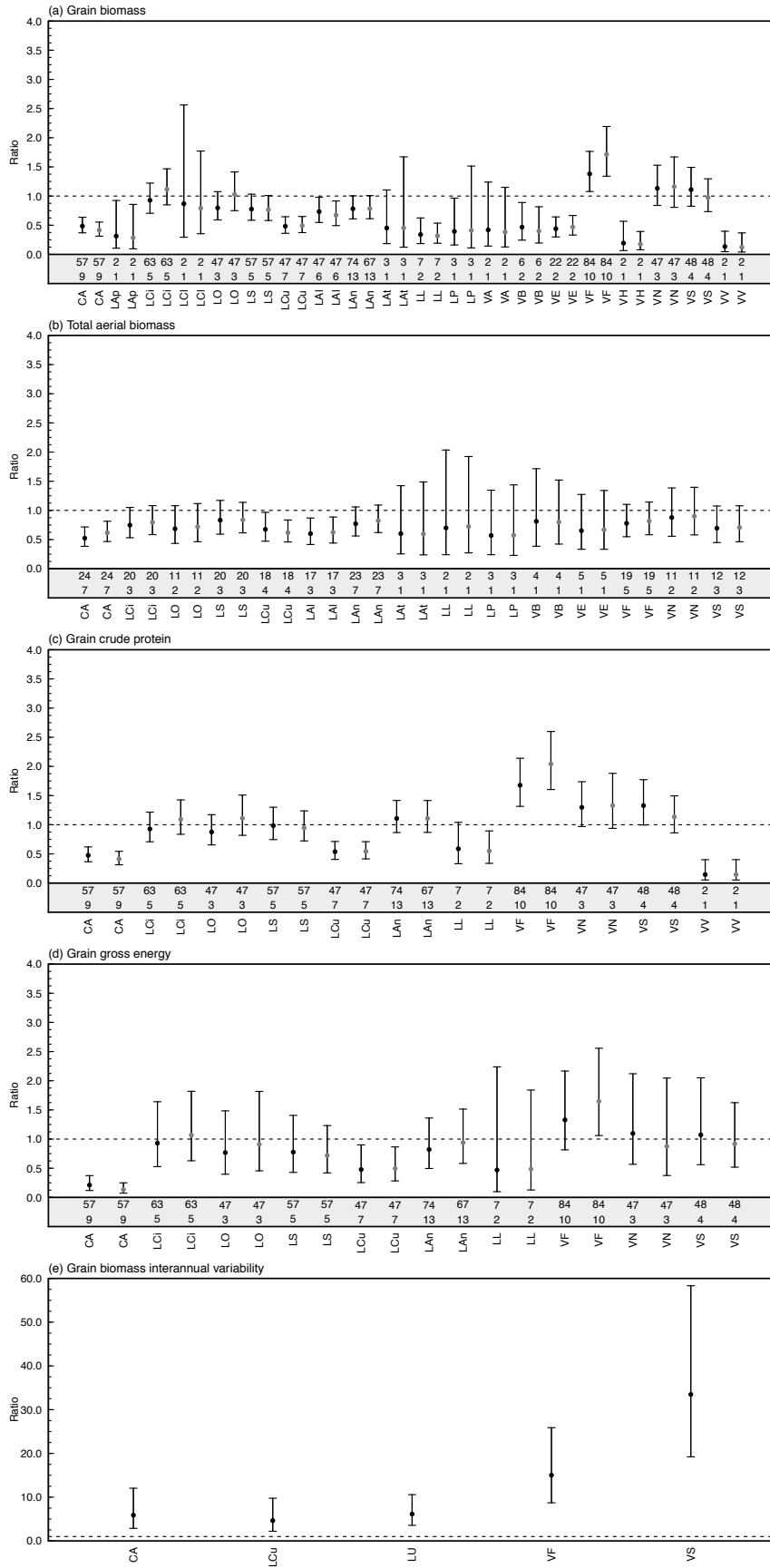
Residuals are symmetrically distributed around zero for all species and criteria (figure S6 to figure S11).

The grain biomass variance ratios estimated from the FAOSTAT database (figure 4(g) and table 1) show that interannual variability in the grain biomasses of chickpea and soybean is significantly lower for than that of pea. For both these species, grain biomass is, on average, -62% less variable than for pea. The grain biomass variance ratios estimated for lentil, kidney bean, and faba bean are not significantly different from one.

### 3.3. Several vetches and lupins tend to outperform pea in Oceania

Faba bean has significantly higher mean grain biomass than that of pea, with estimated ratio of  $R=1.71$  (figure 5(a) and table 1). Several species do not differ significantly from pea in terms of grain biomass: red pea (*Lathyrus cicera*), cicercha purpurina (*Lathyrus clymenum*), cyprus vetch (*Lathyrus ochrus*), white pea, narrowleaf lupin, *Lupinus atlanticus* (no common name), blue lupin (*Lupinus pilosus*), oneflower vetch (*Vicia articulata*), purple broad vetch (*Vicia narbonensis*), and garden vetch. The remaining nine grain legume species have significantly lower grain biomasses than that of pea. The mean total aerial biomasses of chickpea, lentil, and white lupin are significantly lower than that of pea.

### OCEANIA



**Figure 5.** Ratios of grain biomass (a), total aerial biomass (b), grain crude protein (c), grain gross energy (d), and grain biomass interannual variability (e), estimated for twenty grain legume species, relative to those estimated for pea (*Pisum sativum*) in Oceania. The scientific names of the species concerned are abbreviated here as follows: *Cicer arietinum* (chickpea, CA), *Lathyrus aphaca* (yellow pea, LAp), *Lathyrus cicera* (red pea, LCi), *Lathyrus clymenum* (cicercha purpurina, LCI), *Lathyrus ochrus* (cyprus vetch, LO), *Lathyrus sativus* (white pea, LS), *Lens culinaris* (lentil, LCu), *Lupinus albus* (white lupin, LAI), *Lupinus angustifolius* (narrowleaf lupin, LAn), *Lupinus atlanticus* (no common name, LAt), *Lupinus luteus* (yellow lupin, LL), *Lupinus pilosus* (blue lupin, LP), *Vicia articulata* (oneflower vetch, VA), *Vicia benghalensis* (purple vetch, VB), *Vicia ervilia* (blister vetch, VE), *Vicia faba* (faba bean, VF), *Vicia hybrida* (hairy yellow vetch, VH), *Vicia narbonensis* (purple broad vetch, VN), *Vicia sativa* (garden vetch, VS), and *Vicia villosa* (winter vetch, VV). All species are ranked alphabetically according to scientific names. Bullet points indicate mean ratios estimated unweighted (black) or weighted (gray) values for each criterion. Error bars indicate 95% confidence intervals. Within the bottom box (gray), the lower and upper rows indicate the number of selected articles and log-transformed ratios, respectively.

The mean grain crude protein estimated for faba bean is significantly higher than for pea, by +104% (figure 5(c) and table 1). Estimated mean grain crude protein ratio is not significantly different from one in red pea, cyprus vetch, white pea, narrowleaf lupin, purple broad vetch, and garden vetch (i.e., their 95% confidence intervals include one). Estimated mean grain crude protein for chickpea, lentil, yellow lupin, and winter vetch, is significantly lower than for pea. Only faba bean performs significantly better than pea for grain gross energy (figure 5(d) and table 1).

Residuals are symmetrically distributed around zero for all species and criteria (figure S12 to figure S15).

Chickpea, lentil, lupins, faba bean, and garden vetch have significantly higher interannual variability in their grain biomasses than that of pea (figure 5(e) and table1).

#### 4. Discussion

The main conclusions of our meta-analysis are not particularly sensitive to model assumptions (table S6 and table S7). Model variants based on weighted values tend to produce narrower confidence intervals, and more accurate results. Distributions of residuals are symmetric in shape for all species and criteria in all regions (figure S1 to figure S15).

The use of official agricultural statistics enables us to estimate interannual variability in grain biomass on the basis of long time series. However, there is a high inherent risk of confusion between effects of legume species, environmental conditions, and crop management techniques, with this data source. To deal with this issue, we use experimental data collected in agronomic experiments where several grain legume species are directly compared at the same field sites during the same growing seasons with the

same crop management techniques (i.e., tillage, fertilization, pest control, and irrigation). Based on this data source, we are able to rank the productivity levels of grain legume species cultivated in the same agronomic and environmental conditions. Nutritional data extracted from feed and food composition tables enables us to compare grain legume species according to their grain crude protein and gross grain energy; in fact, there is very scant information on legume grain quality criteria in the selected articles. Based on these three data sources, we are able to perform a multi-criteria analysis of the productivity performances of a broad range of grain legume species.

Faba bean has performances similar to those of pea for all criteria in Europe. Faba bean may, therefore, be considered an interesting alternative to pea in this world region. This species is well adapted to the diverse climate conditions and soil types of Europe [58,59]. It is recognized as a valuable source of protein for the human diet and animal feed [31,33,60,61]. Soybean and narrowleaf lupin also perform well relative to pea, especially for grain crude protein and grain gross energy. However, only a small number of field experiments have directly compared these species in Europe. New experiments are thus warranted to identify the geographical areas where both these species perform best under European conditions. It will be useful to compare grain biomass of soybean and lupins at the same field sites during the same growing seasons with the same management treatments, to limit the inherent risk of confusion between effects of legume specific characteristics, environmental conditions, and crop techniques. Grain biomass data could then be converted into grain crude protein and grain gross energy, on the basis of the feed and food composition tables from the FEEDIPEDIA database. Measurements of total aerial biomass (or harvest index), grain nitrogen content (or percentage of nitrogen into grains), and water use (or grain water use efficiency), would also be useful to further refine our estimates, and decrease the uncertainty levels associated with the productivity performances of soybean and lupins in Europe.

Narrowleaf and yellow lupins display good relative performances for most criteria in Europe. Lucas *et al* (2015) [32] stressed the importance of advanced breeding and processing operations to turn European-grown lupins into attractive marketable ingredients for human food and animal feed. Chickpea and lentil have lower overall performances than pea for most criteria. Both chickpea and lentil are suitable for incorporation into animal feed, but both are widely used for human consumption [62–65]. Hence, these grain legume species would provide farmers with high added value market opportunities, which might offset their low productivities [24,26]. The replacement of pea with other legume species, displaying higher levels of grain crude protein, could substantially increase overall crude protein production in European countries. Results show that the replacement of pea with soybean by 25% of the cultivated area currently under pea would increase crude protein production in Europe, by +10% to +56%. The replacement of pea with narrowleaf lupin or faba bean would also increase crude protein production, but to a lesser extent (i.e., +4% or +3%, respectively).



The overall species rankings obtained in North America and Oceania confirm the key findings for Europe. In North America, Miller *et al* (2002) [36] reported that the mean grain biomass of soybean was -47% lower than that of pea. This lower grain biomass is consistent with our results. In North America, soybean has a mean grain biomass -41% (unweighted values) or -53% (weighted values) lower than that of pea (figure 4(a) and table 1). However, soybean performs well for most of the other criteria, and presents lower interannual variability in grain biomass than that of pea.

The results obtained in North America and Oceania also confirm relatively good productivity levels of faba bean in North America, and its interesting performances in Oceania. Siddique *et al* (2013) [41] reported that faba bean could be grown across a broad spectrum of contrasting environments in Australia. White pea has productivity rankings similar to those of pea for most of the criteria considered in both North America and Oceania, but with high levels of uncertainty due to the small number of experiments carried out. Calderón *et al* (2012) [66] and Vaz Patto and Rubiales (2014) [67] argued that further research efforts are required to assess the agronomic potential of white pea. In Europe, this species has not been directly compared with pea. It would be useful to include white pea in new field experiments, to estimate its relative performance; however, the grains of this legume species harbor several undesirable nutritional and neurotoxic factors, decreasing its likely suitability as a food and feed substitute for pea [67–69].

The number of species directly compared with pea is about three times higher in Oceania (20 species) than in North America (7 species) and Europe (7 species). Most of the experiments conducted in Oceania and included in our meta-analysis took place in south-western Australia. Harries *et al* (2015) [70] pointed out that a broad range of grain legume species were compared in this subregion in the past, with a view to increasing their commercial production. Several of the species that have been tested in Oceania but not in Europe perform well with respect to pea in terms of productivity, especially for red pea, cicercha purpurina, cyprus vetch, white pea, *Lupinus atlanticus* (no common name), blue lupin, oneflower vetch, purple broad vetch, and garden vetch. It will be useful to include these grain legume species in future field experiments with the prospect of assessing their relative productive performances across European conditions.

## References

- [1] Tharanathan R N and Mahadevamma S 2003 Grain legumes – a boon to human nutrition *Trends Food Sci. Tech.* **14** 507–18
- [2] Duranti M 2006 Grain legume proteins and nutraceutical properties *Fitoterapia* **77** 67–82

- [3] Vaz Patto M C, Amarowicz R, Aryee A N A, Boye J I, Chung H J, Martín-Cabrejas M A and Domoney C 2015 Achievements and challenges in improving the nutritional quality of food legumes *Crit. Rev. Plant Sci.* **34** 105–43
- [4] Temba M C, Njobeh P B, Adebo O A, Olugbile A O and Kayitesi E 2016 The role of compositing cereals with legumes to alleviate protein energy malnutrition in Africa *Int. J. Food Sci. Tech.* **51** 543-54
- [5] Graham P H and Vance C P 2003 Legumes: importance and constraints to greater use *Plant Physiol.* **131** 872–77
- [6] Sinclair T R and Vadez V 2012 The future of grain legumes in cropping systems *Crop Pasture Sci.* **63** 501–12
- [7] Jensen E S, Peoples M B, Boddey R M, Gresshoff P M, Hauggaard-Nielsen H, Alves B J R and Morrison M J 2012 Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries: a review *Agron. Sustain. Dev.* **32** 329–64
- [8] Reckling M, Bergkvist G, Watson C A, Stoddard F L, Zander P M, Walker R L, Pristeri A, Toncea I and Bachinger J 2016 Trade-offs between economic and environmental impacts of introducing legumes into cropping systems *Front. Plant Sci.* **7** 669
- [9] Jeuffroy M H, Baranger E, Carrouée B, de Chezelles E, Gosme M, Hénault C, Schneider A and Cellier P 2013 Nitrous oxide emissions from crop rotations including wheat, oilseed rape and dry peas *Biogeosciences* **10** 1787–97
- [10] Schmeer M, Loges R, Dittert K, Senbayram M, Horn R and Taube F 2014 Legume-based forage production systems reduce nitrous oxide emissions *Soil Till. Res.* **143** 17–25
- [11] Chalk P M 1998 Dynamics of biologically fixed N in legume-cereal rotations: a review *Aust. J. Agr. Res.* **49** 303–16
- [12] Evans J, McNeill A M, Unkovich M J, Fettell N A and Heenan D P 2001 Net nitrogen balances for cool-season grain legume crops and contributions to wheat nitrogen uptake: a review *Aust. J. Exp. Agr.* **41** 347–59

- [13] Peoples M B *et al* 2009 The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems *Symbiosis* **48** 1–17
- [14] Jensen C R, Joernsgaard B, Andersen M N, Christiansen J L, Mogensen V O, Friis P and Petersen C T 2004 The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop *Eur. J. Agron.* **20** 405–18
- [15] Williams C M, King J R, Ross S M, Olson M A, Hoy C F and Lopetinsky K J 2014 Effects of three pulse crops on subsequent barley, canola, and wheat *Agron. J.* **106** 343–50
- [16] Angus J F, Kirkegaard J A, Hunt J R, Ryan M H, Ohlander L and Peoples M B 2015 Break crops and rotations for wheat *Crop Pasture Sci.* **66** 523–52
- [17] Preissel S, Reckling M, Schläfke N and Zander P 2015 Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: a review *Field Crop. Res.* **175** 64–79
- [18] Statistics Division of Food and Agriculture Organization of the United Nations (FAOSTAT database) 2016 (<http://faostat3.fao.org/home/E>) (accessed 05 Apr 2016)
- [19] Häusling M 2011 The EU protein deficit: what solution for a long-standing problem? (2010/2011(INI), European Parliament, Committee on Agriculture and Rural Development) (<http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2011-0026&language=EN>) (accessed 12 Apr 2016)
- [20] Zander P *et al* 2016 Grain legume decline and potential recovery in European agriculture: a review *Agron. Sustain. Dev.* **36** 1–20
- [21] Gepts P, Beavis W D, Brummer E C, Shoemaker R C, Stalker H T, Weeden N F and Young N D 2005 Legumes as a model plant family Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference *Plant Physiol.* **137** 1228–35
- [22] Weeden N F 2007 Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the “domestication syndrome” for legumes? *Ann. Bot.–London* **100** 1017–25

- [23] Murphy-Bokern D, Watson C A, Stoddard F L, Lindström K, Zander P, Reckling M, Preisel S, Bues A and Torres A 2014 Outlook for knowledge and technology for legume-supported cropping systems Legume Futures Report 5.3 (European Commission, Seventh Framework Programme for Research and Technological Development, Legume-Supported Cropping Systems for Europe (Legume Futures)) ([http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_5.3.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_5.3.pdf)) (accessed 19 Apr 2016)
- [24] Voisin A S, Guéguen J, Huyghe C, Jeuffroy M H, Magrini M B, Meynard J M, Mougél C, Pellerin S and Pelzer E 2014 Legumes for feed, food, biomaterials and bioenergy in Europe: a review *Agron. Sustain. Dev.* **34** 361–80
- [25] Foyer C H *et al* 2016 Neglecting legumes has compromised human health and sustainable food production *Nat. Plants* **2** 16112
- [26] Magrini M B, Anton M, Cholez C, Corre-Hellou G, Duc G, Jeuffroy M H, Meynard J M, Pelzer E, Voisin A S and Walrand S 2016 Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system *Ecol. Econ.* **126** 152–62
- [27] LMC International 2009 Evaluation of measures applied under the Common Agricultural Policy to the protein crop sector (European Commission, Agriculture and Rural Development, Evaluation, Reports on Market and Income Policies) ([http://ec.europa.eu/agriculture/eval/reports/protein\\_crops/index\\_en.htm](http://ec.europa.eu/agriculture/eval/reports/protein_crops/index_en.htm)) (accessed 12 Apr 2016)
- [28] Wiggering H *et al* 2012 The Legumes Expert Forum Science, economy and society — making ecosystem services from legumes competitive A research strategy of the German Agricultural Research Alliance (Braunschweig, Germany: German Agricultural Research Alliance (DAFA)) ([http://www.dafa.de/fileadmin/dam\\_uploads/images/Fachforen/ff\\_leguminosen-en\\_2012.pdf](http://www.dafa.de/fileadmin/dam_uploads/images/Fachforen/ff_leguminosen-en_2012.pdf)) (accessed 12 Apr 2016)

- [29] Schläfke N, Zander P, Reckling M, Hecker J M and Bachinger J 2014 Evaluation of legume-supported agriculture and policies at farm level Legume Futures Report 4.3 (European Commission, Seventh Framework Programme for Research and Technological Development, Legume-Supported Cropping Systems for Europe (Legume Futures)) ([http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_4.3.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_4.3.pdf)) (accessed 19 Apr 2016)
- [30] McEwen J, Darby R J, Hewitt M V and Yeoman D P 1990 Effects of field beans, fallow, lupins, oats, oilseed rape, peas, ryegrass, sunflowers and wheat on nitrogen residues in the soil and on the growth of a subsequent wheat crop *J. Agr. Sci.* **115** 209–19
- [31] Jezierny D, Mosenthin R and Bauer E 2010 The use of grain legumes as a protein source in pig nutrition: a review *Anim. Feed Sci. Tech.* **157** 111–28
- [32] Lucas M M, Stoddard F L, Annicchiarico P, Frías J, Martínez-Villaluenga C, Sussmann D, Duranti M, Seger A, Zander P M and Pueyo J J 2015 The future of lupin as a protein crop in Europe *Front. Plant Sci.* **6** 705
- [33] Multari S, Stewart D and Russell W R 2015 Potential of fava bean as future protein supply to partially replace meat intake in the human diet *Compr. Rev. Food Sci. Food Saf.* **14** 511–22
- [34] Peel M C, Finlayson B L and McMahon T A 2007 Updated world map of the Köppen-Geiger climate classification *Hydrol. Earth Syst. Sci.* **11** 1633–44
- [35] Miller P R, McDonald C L, Derksen D A and Waddington J 2001 The adaptation of seven broadleaf crops to the dry semiarid prairie *Can. J. Plant Sci.* **81** 29–43
- [36] Miller P R, McConkey B G, Clayton G W, Brandt S A, Staricka J A, Johnston A M, Lafond G P, Schatz B G, Baltensperger D D and Neill K E 2002 Pulse crop adaptation in the northern Great Plains *Agron. J.* **94** 261–72
- [37] Zentner R P *et al* 2002 Economics of crop diversification and soil tillage opportunities in the Canadian Prairies *Agron. J.* **94** 216–30
- [38] Angadi S V, McConkey B G, Cutforth H W, Miller P R, Ulrich D, Selles F, Volkmar K M, Entz M H and Brandt S A 2008 Adaptation of alternative pulse and oilseed crops to the semiarid Canadian Prairie: seed yield and water use efficiency *Can. J. Plant Sci.* **88** 425–38

- [39] Gan Y, Hamel C, O'Donovan J T, Cutforth H, Zentner R P, Campbell C A, Niu Y and Poppy L 2015 Diversifying crop rotations with pulses enhances system productivity *Sci. Rep.* **5** 14625
- [40] Siddique K H M and Sykes J 1997 Pulse production in Australia past, present and future *Aust. J. Exp. Agr.* **37** 103–11
- [41] Siddique K H M, Erskine W, Hobson K, Knights E J, Leonforte A, Khan T N, Paull J G, Redden R and Materne M 2013 Cool-season grain legume improvement in Australia — use of genetic resources *Crop Pasture Sci.* **64** 347–60
- [42] Silsbury J H 1975 Grain yield and flowering of some introduced grain legumes in south Australia *Aust. J. Exp. Agr. Anim. Husb.* **15** 556–60
- [43] Laurence R 1979 A comparison of the grain and protein yield potential of some annual legume species in south Australia *Aust. J. Exp. Agr. Anim. Husb.* **19** 495–503
- [44] French R and Ewing M 1989 Soil type influences the relative yields of different cereals and crop legumes in the western Australian wheatbelt *Aust. J. Exp. Agr.* **29** 829–35
- [45] Siddique K H M, Walton G and Seymour M 1993 A comparison of seed yields of winter grain legumes in western Australia *Aust. J. Exp. Agr.* **33** 915–22
- [46] Siddique K H M, Loss S P, Regan K L and Jettner R L 1999 Adaptation and seed yield of cool season grain legumes in Mediterranean environments of south-western Australia *Aust. J. Agr. Res.* **50** 375–88
- [47] Thomson B D, Siddique K H M, Barr M D and Wilson J M 1997 Grain legume species in low rainfall Mediterranean-type environments I. Phenology and seed yield *Field Crop. Res.* **54** 173–87
- [48] Gregory P J 1998 Alternative crops for Duplex soils: growth and water use of some cereal, legume, and oilseed crops, and pastures *Aust. J. Agr. Res.* **49** 21–32
- [49] Malik R S, Seymour M, French R J, Kirkegaard J A, Lawes R A and Liebig M A 2015 Dynamic crop sequencing in western Australian cropping systems *Crop Pasture Sci.* **66** 594–609

- [50] Institute for Scientific Information Web of Science 2016 (<http://wokinfo.com/>) (accessed 15 Feb 2016)
- [51] Cernay C, Pelzer E and Makowski D 2016 A global experimental dataset for assessing grain legume production *Sci. Data* **3** 160084
- [52] Animal Feed Resources Information System 2016 (FEEDIPEDIA database) (<http://www.feedipedia.org/>) (accessed 15 Mar 2016)
- [53] Pinheiro J, Bates D, DebRoy S, Sarkar D and the R Development Core Team 2013 R package ‘nlme’: linear and nonlinear mixed effects models Version 3.1.111
- [54] R Development Core Team 2013 R: a language and environment for statistical computing Version 3.0.2 (Vienna, Austria: R Foundation for Statistical Computing) (<https://www.r-project.org/>) (accessed 15 Oct 2013)
- [55] Akaike H 1974 A new look at the statistical model identification *IEEE T. Automat. Contr.* **19** 716–23
- [56] Wiebe N, Vandermeer B, Platt R W, Klassen T P, Moher D and Barrowman N J 2006 A systematic review identifies a lack of standardization in methods for handling missing variance data *J. Clin. Epidemiol.* **59** 342–53
- [57] Cernay C, Ben-Ari T, Pelzer E, Meynard J M and Makowski D 2015 Estimating variability in grain legume yields across Europe and the Americas *Sci. Rep.* **5** 11171
- [58] Flores F, Nadal S, Solis I, Winkler J, Sass O, Stoddard F L, Link W, Raffiot B, Muel F and Rubiales D 2012 Faba bean adaptation to autumn sowing under European climates *Agron. Sustain. Dev.* **32** 727–34
- [59] Flores *et al* 2013 Adaptation of spring faba bean types across European climates *Field Crop. Res.* **145** 1–9
- [60] Crépon K, Marget P, Peyronnet C, Carrouée B, Arese P and Duc G 2010 Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food *Field Crop. Res.* **115** 329–39

- [61] Koivunen E, Partanen K, Perttilä S, Palander S, Tuunainen P and Valaja J 2016 Digestibility and energy value of pea (*Pisum sativum* L.), faba bean (*Vicia faba* L.) and blue lupin (narrow-leaf) (*Lupinus angustifolius*) seeds in broilers *Anim. Feed Sci. Tech.* **218** 120–27
- [62] Boye J, Zare F and Pletch A 2010 Pulse proteins: processing, characterization, functional properties and applications in food and feed *Food Res. Int.* **43** 414–31
- [63] Erskine W, Sarker A and Kumar S 2011 Crops that feed the world 3. Investing in lentil improvement toward a food secure world *Food Secur.* **3** 127–39
- [64] Jukanti A K, Gaur P M, Gowda C L L and Chibbar R N 2012 Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review *Br. J. Nutr.* **108** S11–26
- [65] Faris M A I E, Takruri H R and Issa A Y 2013 Role of lentils (*Lens culinaris* L.) in human health and nutrition: a review *Med. J. Nutrition Metab.* **6** 3–16
- [66] Calderón F J, Vigil M F, Nielsen D C, Benjamin J G and Poss D J 2012 Water use and yields of no-till managed dryland grasspea and yellow pea under different planting configurations *Field Crop. Res.* **125** 179–85
- [67] Vaz Patto M C and Rubiales D 2014 *Lathyrus* diversity: available resources with relevance to crop improvement – *L. sativus* and *L. cicera* as case studies *Ann. Bot.–London* **113** 895–908
- [68] Hanbury C D, White C L, Mullan B P and Siddique K H M 2000 A review of the potential of *Lathyrus sativus* L. and *L. cicera* L. grain for use as animal feed *Anim. Feed Sci. Tech.* **87** 1–27
- [69] Yan Z Y, Spencer P S, Li Z X, Liang Y M, Wang Y F, Wang C Y and Li F M 2006 *Lathyrus sativus* (grass pea) and its neurotoxin ODAP *Phytochemistry* **67** 107–21
- [70] Harries M, Anderson G C and Hüberli D 2015 Crop sequences in western Australia: what are they and are they sustainable? Findings of a four-year survey *Crop Pasture Sci.* **66** 634–47





## Chapitre 4



## Chapitre 4

### **Comparer l'effet des légumineuses à graines sur les rendements des céréales suivantes**

Ce chapitre est un article scientifique soumis pour publication dans une revue à comité de lecture.

Cernay, C, Pelzer, E. & Makowski, D. Preceding grain legume cultivation increases cereal yields under low nitrogen input conditions. (2016).

Les informations supplémentaires sont en Annexes.

Seulement les noms scientifiques des espèces sont indiqués dans le texte.

## Preceding grain legume cultivation increases cereal yields under low nitrogen input conditions

Charles Cernay, Elise Pelzer & David Makowski

UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

### Abstract

2016 is the International Year of Pulses – a United Nations initiative for highlighting grain legume cultivation worldwide. Several studies have indicated that preceding grain legume cultivation had a positive effect on the yields of subsequent cereals. This argument is frequently used to introduce grain legumes to cereal-based systems. However, no quantitative synthesis of the data has been performed on a global scale to estimate the relative increases in cereal yields arisen from preceding grain legume cultivation. Here we report a meta-analysis of 1,181 yields of cereals encompassing fifteen countries. The results confirm that the yields of cereals cultivated after grain legumes were, on average, +29% significantly higher than the yields of cereals cultivated after cereals; this relative effect was significantly positive for thirteen of sixteen grain legume species. Our study demonstrates that the grain legume effect decreased as a function of the nitrogen (N) fertilization rate applied to subsequent cereals, and became negligible when the mean N fertilization exceeded 150 kg N ha<sup>-1</sup>. Globally, this threshold is often exceeded in conventional cereal systems. Thus, the positive effect of grain legumes on cereal yields is expected to be low if the current N fertilization rate on cereals is not reduced, especially for European agriculture. If low-input systems are widely adopted by crop producers in the future, the introduction of grain legumes to major cereal-producing regions will contribute to the maintenance of high cereal yields.

### 1. Introduction

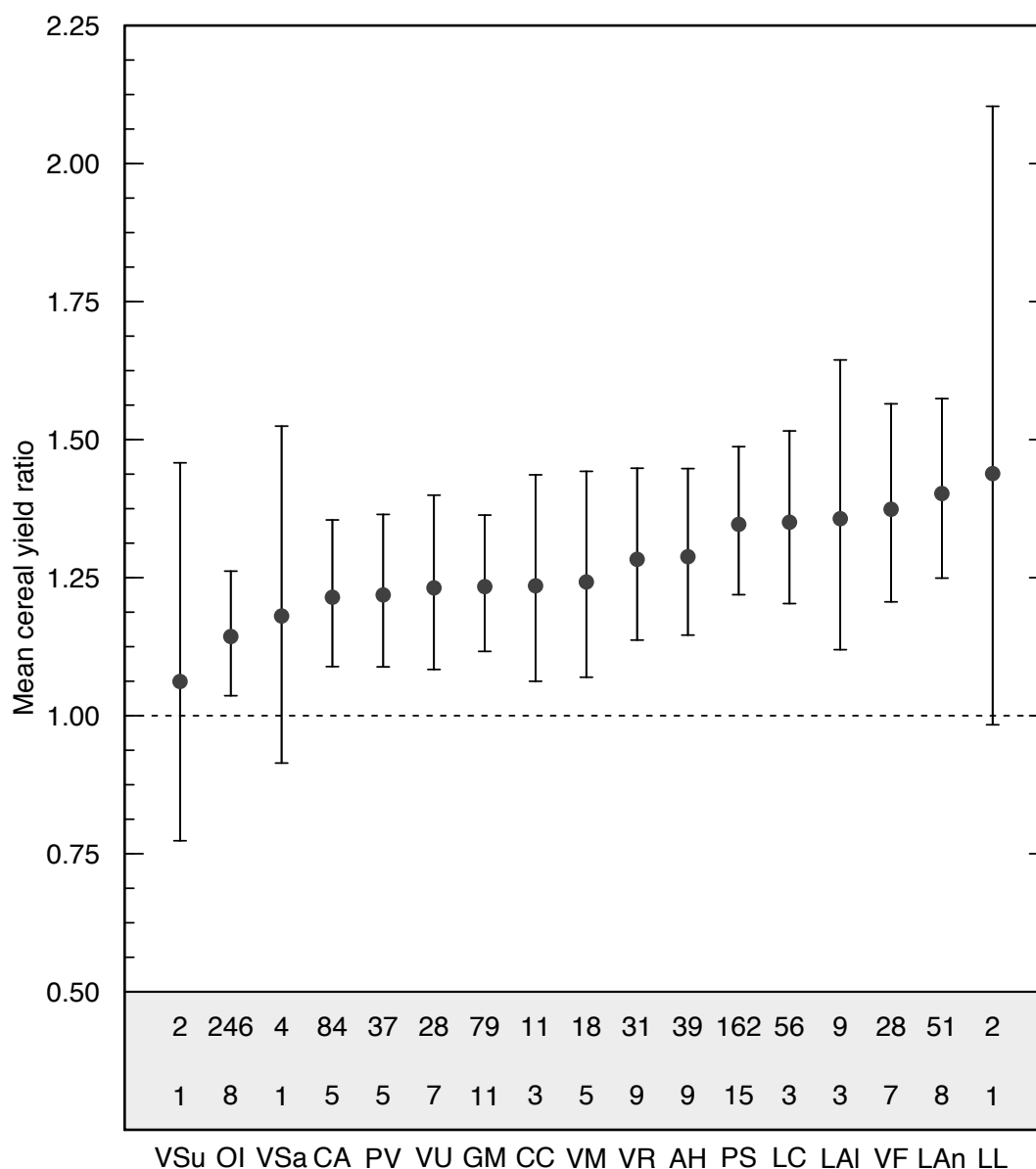
The 68<sup>th</sup> United Nations General Assembly has declared 2016 the International Year of Pulses (hereafter grain legumes). These annual crops belong to the Fabaceae plant family and are mainly harvested for grain production. Legume grains are considered a protein-rich complement to cereal grains for the human diet and animal feed (1), thereby contributing importantly to sustainable food and feed production (2). Grain legume cultivation is encouraged to provide benefits to the environment and potential economic competitive returns (3, 4). Legume crops are able to biologically fix

atmospheric N<sub>2</sub> and require little or no N fertilization (5). Cultivating grain legumes can increase the soil concentration of mineral N for subsequent cereals, through the mineralization of crop residues that are returned to the soil and the N uptake generally lower than those of non-legumes (6). Several individual field experiments have provided evidence that grain legume cultivation increases the yields of subsequent cereals (7, 8, 9, 10, 11). Recent studies have advocated an increase in grain legume cultivation (2), especially in Europe (12). Replacement of cereals with grain legumes is expected to decrease total cereal production, but this decrease could be partially mitigated if grain legume cultivation had a positive effect on the yields of subsequent cereals. Several authors have hypothesized that the effect of grain legumes on cereal yields depends on the N fertilization rate applied to subsequent cereals (7, 10, 13, 14, 15), but no quantitative relationship has been established. To date, no global meta-analysis of field experiments has been performed to investigate this topic. Here we synthesized data extracted from peer-reviewed articles (*i*) to compare the relative effects of sixteen grain legume species on the yields of subsequent cereals, (*ii*) to analyse the relationship between the effect of preceding grain legume cultivation and the N fertilization rate applied to subsequent cereals, and (*iii*) to assess whether the positive effect of grain legumes on cereal yields is able to mitigate a decrease in cereal production from the introduction of grain legumes to a cereal monoculture.

We conducted a meta-analysis of the yields of cereals cultivated after grain legumes using experimental data published in thirty-three articles (Note S1) conducted in six regions of the world (Fig. S1). Each field experiment systematically included several plots in which the yield was measured (*i*) for cereals cultivated after cereal species, and (*ii*) for cereals cultivated after grain legume species (Table S1). All plots that were compared with each other within an experiment were located on the same field site and cultivated during the same growing seasons using the same crop management techniques. In several experiments, the yields of cereals following oilseeds were also available, and these data were included in the dataset to compare the relative effects of oilseeds and grain legumes on the yields of subsequent cereals.

Yield ratios were calculated for all field sites, growing seasons, and crop management techniques. Each yield ratio was expressed as the ratio of the yield of a cereal cultivated after a non-cereal (either grain legume or oilseed) to the yield of a cereal cultivated after a cereal. First, the mean value of the cereal yield ratios was estimated for each preceding non-cereal species using statistical mixed-effect models. The mean yield ratios and the 95% confidence intervals were estimated for each preceding non-cereal species. Second, yield ratios were related to N fertilization rates applied to subsequent cereals using statistical mixed-effect models, including different covariates as the preceding non-cereal species, the differences in N fertilization rates applied on preceding non-cereals and preceding cereals and the N fertilization rate applied to subsequent cereals (see Methods).

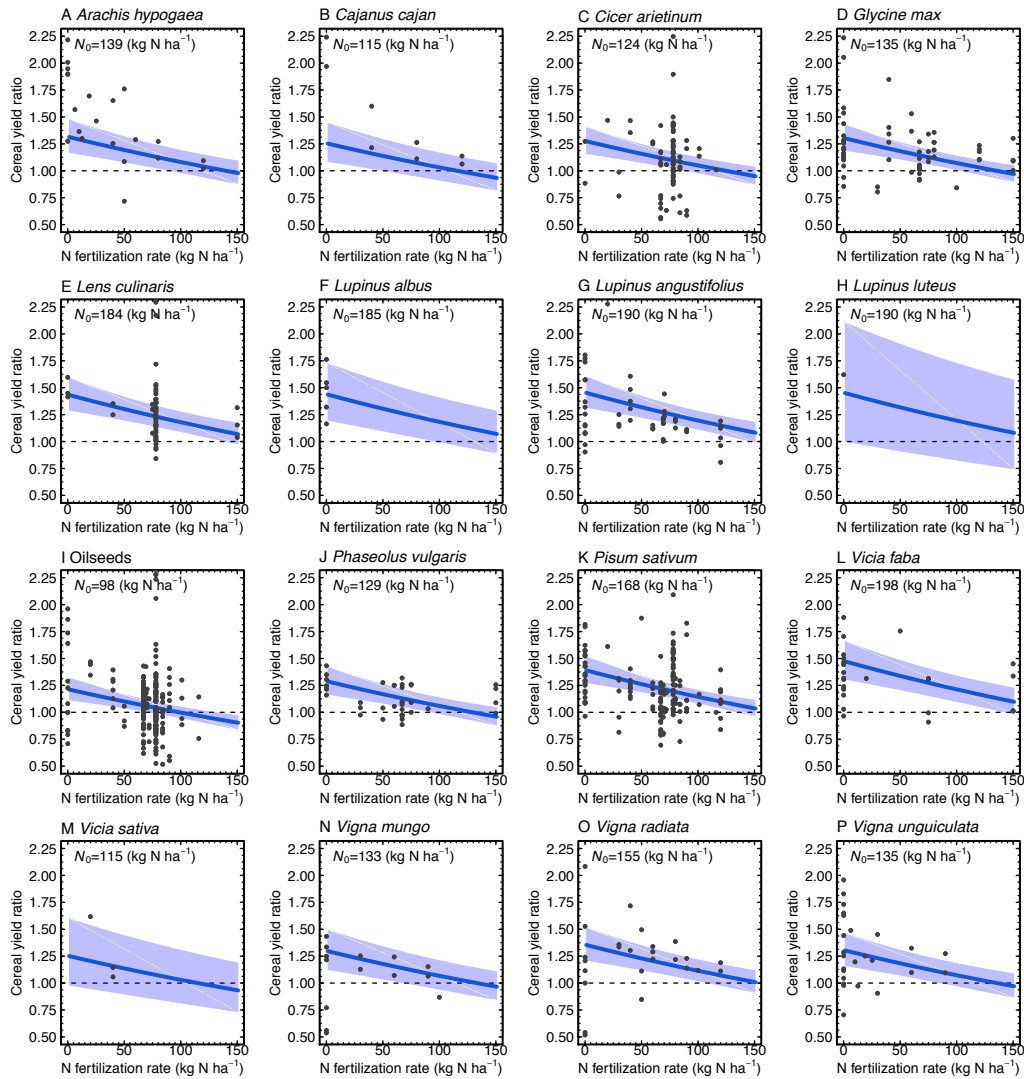
2. Results & Discussion



**Fig. 1.** Estimated mean ratios of the yields of cereals cultivated after non-cereals (either grain legumes or oilseeds) to the yields of cereals cultivated after cereals. The scientific names of the grain legume species preceding cereals are abbreviated as follows: *Arachis hypogaea* (AH), *Cajanus cajan* (CC), *Cicer arietinum* (CA), *Glycine max* (GM), *Lens culinaris* (LC), *Lupinus albus* (LAI), *Lupinus angustifolius* (LAn), *Lupinus luteus* (LL), *Phaseolus vulgaris* (PV), *Pisum sativum* (PS), *Vicia faba* (VF), *Vicia sativa* (VSa), *Vigna mungo* (VM), *Vigna radiata* (VR), *Vigna subterranea* (VSu), and *Vigna unguiculata* (VU). The category for oilseeds (OI) includes six species cultivated as preceding crops, namely, *Brassica juncea*, *Brassica napus*, *Carthamus tinctorius*, *Crambe abyssinica*, *Helianthus annuus*, and *Linum usitatissimum*. Bullet points indicate the mean yield ratios estimated from the weighted data using the statistical model variant M4. Error bars indicate the 95% confidence intervals. The lower and upper rows in the bottom box (grey) indicate the numbers of selected articles and observed cereal yield ratios available for each preceding non-cereal species, respectively.

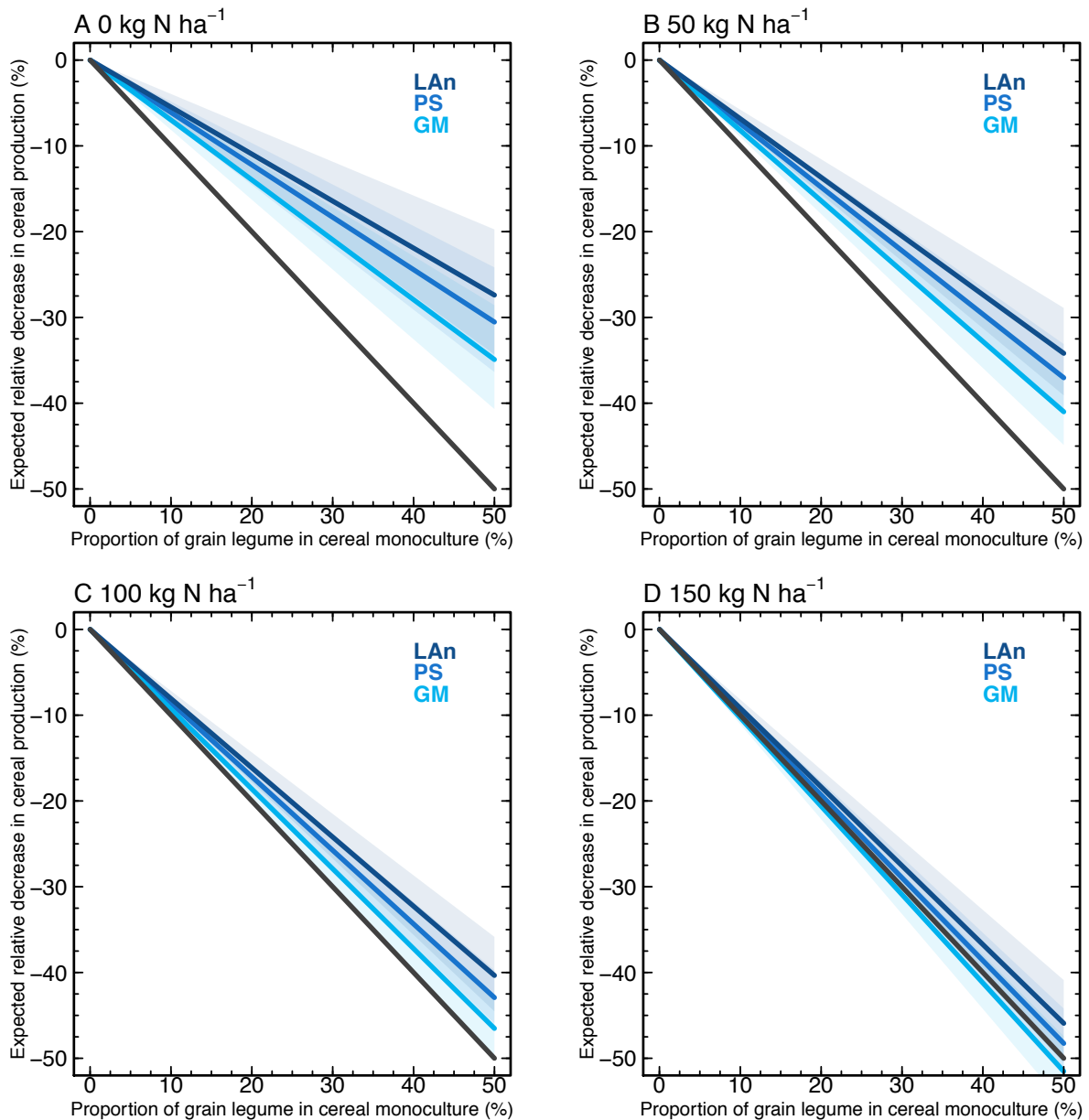
The estimated mean yield ratios were higher than one (yield of cereal after non-cereal > yield of cereal after cereal) for oilseeds and all grain legumes cultivated as preceding crops (Fig. 1). The relative effect was statistically significant ( $p < 0.05$ ) for oilseeds and thirteen of the sixteen grain legume species; the three exceptions were *Lupinus luteus*, *Vicia sativa*, and *Vigna subterranea*. The dearth of available yield data resulted in large 95% confidence intervals for these three species, and less certain results compared with those obtained for oilseeds and other grain legume species. The mean cereal yield increase resulting from preceding grain legume cultivation ranged from +21% (*Cicer arietinum*) to +40% (*Lupinus angustifolius*). The estimated mean yield increases for cereals cultivated after *Glycine max*, *Pisum sativum*, *Lens culinaris*, *Lupinus albus*, and *Vicia faba*, were +23%, +35%, +35%, +36%, and +37%, respectively. *Pisum sativum* was the grain legume species with the highest number of yield ratios (162 yield ratios). The rankings of preceding non-cereal species obtained with the unweighted data was similar to that obtained with the weighted data (Fig. S2). We found that the increased cereal yields derived from preceding oilseed cultivation (+14%) was smaller than those derived from preceding grain legume cultivation (on average, +29% significantly for thirteen of sixteen grain legume species), but significant ( $p < 0.05$ ). This result is consistent with previously reported values about the effect of preceding oilseed cultivation on the yields of subsequent cereals (3, 16, 17, 18), including wheat (*Triticum aestivum*). Preceding grain legume cultivation increased significantly the yields of subsequent cereals by +29%, on average. This value is also consistently reported in previous studies based on quantitative syntheses of the data collected from local field experiments (3, 16, 17, 18).





**Fig. 2.** Cereal yield ratios as a function of the nitrogen (N) fertilization rate ( $\text{kg N ha}^{-1}$ ) applied to cereals cultivated after oilseeds and grain legume species. Panels (A-P) indicate non-cereal species preceding cereals. The category for oilseeds (OI) includes six species cultivated as preceding crops, namely, *Brassica juncea*, *Brassica napus*, *Carthamus tinctorius*, *Crambe abyssinica*, *Helianthus annuus* and *Linum usitatissimum*. Bullet points indicate the observed ratios of the yields of cereals cultivated after non-cereals to the yields of cereals cultivated after cereals. The number of observed cereal yield ratios  $n$  depends on the non-cereal species cultivated as preceding crops: *Arachis hypogaea* ( $n=22$ , extracted from 6 articles), *Cajanus cajan* ( $n=8$ , extracted from 1 article), *Cicer arietinum* ( $n=84$ , extracted from 5 articles), *Glycine max* ( $n=64$ , extracted from 9 articles), *Lens culinaris* ( $n=56$ , extracted from 3 articles), *Lupinus albus* ( $n=9$ , extracted from 3 articles), *Lupinus angustifolius* ( $n=51$ , extracted from 8 articles), *Lupinus luteus* ( $n=2$ , extracted from 1 article), oilseeds ( $n=246$ , extracted from 8 articles), *Phaseolus vulgaris* ( $n=37$ , extracted from 5 articles), *Pisum sativum* ( $n=162$ , extracted from 15 articles), *Vicia faba* ( $n=28$ , extracted from 7 articles), *Vicia sativa* ( $n=4$ , extracted from 1 article), *Vigna mungo* ( $n=14$ , extracted from 4 articles), *Vigna radiata* ( $n=26$ , extracted from 7 articles), and *Vigna unguiculata* ( $n=26$ , extracted from 6 articles). The blue curve indicates the regression trend fitted to the observed cereal yield ratios using the statistical model variant *M31*. Blue uncertainty envelopes indicate the 95% confidence intervals obtained using the statistical model variant *M31*.  $N_0$  indicates the threshold of the nitrogen fertilization rate applied to subsequent cereals above which the estimated mean yield ratios become lower than zero.  $N_0$  is estimated using the statistical model variant *M31*.

The estimated mean yield ratios decreased significantly as a function of the N fertilization rate applied to subsequent cereals ( $p < 0.05$ ) (Fig. 2). The positive effects of oilseeds and grain legumes on yields of subsequent cereals became negligible when the mean N fertilization exceeded 150 kg N ha<sup>-1</sup> (note: the mean N fertilization exceeded 98 kg N ha<sup>-1</sup> when only oilseeds were considered, and 153 kg N ha<sup>-1</sup> when only grain legumes were considered). However, this threshold showed a strong variation with the preceding non-cereal species. This decreasing trend was observed for all preceding non-cereal species, but the estimated responses were more uncertain for some species than others, especially for *Lupinus albus*, *Lupinus luteus*, and *Vicia sativa*. The 95% confidence intervals were large because of a lack of available data for these three species (Figs. 2F, 2G, and 2M). Globally, N fertilization is often higher than 150 kg N ha<sup>-1</sup> in conventional cereal systems (19). Hence, the mean cereal yield increase that arises from preceding grain legume cultivation is expected to be low if current N input conditions remain unchanged, especially for European agriculture. Note that we failed to find a significant effect of the differences in N fertilization rates applied on preceding non-cereals and preceding cereals on the estimated mean yield ratios.



**Fig. 3.** Expected relative decreases in cereal production resulting from the introduction of grain legumes to a cereal monoculture. The proportion of grain legumes varied from 0% to 50% in the cereal monoculture. Four nitrogen (N) fertilization rates were applied to the cereals: 0 kg N ha<sup>-1</sup> (A), 50 kg N ha<sup>-1</sup> (B), 100 kg N ha<sup>-1</sup> (C), and 150 kg N ha<sup>-1</sup> (D). Dark lines indicate the relative decreases in cereal production, assuming that the positive effect of grain legumes on the yields of subsequent cereals is not realized. Blue lines indicate the mean expected relative decreases in cereal production when cereals fertilized at a nitrogen fertilization rate are replaced with grain legumes, assuming that grain legumes increase the yields of subsequent cereals. Calculations were performed using the mean yield ratios estimated with the statistical model variant M31. Blue uncertainty envelopes indicate the 95% confidence intervals obtained using the statistical model variant M31. The scientific names of the grain legume species preceding cereals are abbreviated as follows: *Glycine max* (GM), *Lupinus angustifolius* (LAn), and *Pisum sativum* (PS).

The results of the meta-analysis were used to calculate the expected relative change in cereal production from an increase in the proportion of a grain legume introduced to a cereal monoculture (Fig. 3). Expected relative changes in cereal production were assessed for three different grain legume species (namely, *Glycine max*, *Lupinus angustifolius*, and *Pisum sativum*) and four different N fertilization rates applied to the cereals (0 kg N ha<sup>-1</sup>, 50 kg N ha<sup>-1</sup>, 100 kg N ha<sup>-1</sup>, and 150 kg N ha<sup>-1</sup>). The proportion of grain legumes varied from 0% to 50% in the cereal monoculture. As expected, an increase in grain legume proportions in the cereal monoculture consistently decreased total cereal production. However, the relative decreases in cereal production were partially mitigated by the positive effect of grain legumes on the yields of subsequent cereals when the N fertilization applied to subsequent cereals was less than 150 kg N ha<sup>-1</sup> (Figs. 3A, 3B, and 3C). In this situation, the mean expected relative decrease in cereal production was smaller than the grain legume proportion alone in the cereal monoculture. For example, a grain legume proportion of 25% resulted in a mean loss in cereal production of less than -25%; cereal production was indeed decreased by -15%, -19%, and -22% for cereals fertilized with 0 kg N ha<sup>-1</sup>, 50 kg N ha<sup>-1</sup>, and 100 kg N ha<sup>-1</sup>, respectively, over the considered preceding legumes. The highest and lowest levels of cereal production were obtained with *Lupinus angustifolius* and *Glycine max*, respectively, among the three grain legume species. It should be noted that *Pisum sativum* provided intermediate results. These differences were consistent with the ranking of preceding non-cereal species (Fig. 1). When cereals were fertilized with at least 150 kg N ha<sup>-1</sup>, the mitigation effect of preceding grain legume cultivation was not significant (Fig. 3D). In this situation, the mean expected relative decrease in cereal production was equivalent to the grain legume proportion alone in the cereal monoculture. For example, a grain legume proportion of 25% resulted in a mean loss in cereal production of approximately -25%; cereal production was indeed decreased by -24% for cereals fertilized with 150 kg N ha<sup>-1</sup>, over the considered preceding legumes.

The differences between the grain legume species reported in our study may be partly due to their different abilities to biologically fix atmospheric N<sub>2</sub> (20). Some grain legume species have root systems more nodulated than other legume species, enabling them to potentially fix more N<sub>2</sub> and generate more N-rich above-ground residues that can be returned to the soil for mineralization (6, 21). Based on crop specific models in Australia (6), the estimated mean soil N balance (including N below-ground residues) for *Cicer arietinum*, *Pisum sativum*, and *Lupinus angustifolius*, were 18 kg N ha<sup>-1</sup>, 44 kg N ha<sup>-1</sup>, and 88 kg N ha<sup>-1</sup>, respectively. These differences in mean soil N balance may explain the differences in mean cereal yield increases reported in our study; we estimated that the yield increases for cereals cultivated after *Cicer arietinum*, *Pisum sativum*, and *Lupinus angustifolius*, were, on average, +21%, +35%, and +40%, respectively.

Crop producers often know about the positive effect of grain legumes on the yields of subsequent cereals (3, 17, 18, 22, 23). Agricultural and environmental scientists generally promote this relative effect for the introduction of grain legumes to intensive cereal systems (22, 23), particularly in European countries (3, 4, 12). Wheat was ranked among the most cultivated crops worldwide in 2014 (24), and it was cultivated after a cereal on approximately 40% of the area (18). The results of our meta-analysis aid in targeting the grain legume species with the greatest potential for diversifying cereal-based systems. We show that grain legume species significantly increase the yields of subsequent cereals under low N input conditions, especially *Pisum sativum*, *Lens culinaris*, *Vicia faba*, and *Lupinus* species. Because high levels of N fertilization are currently applied to cereals in many important agricultural areas around the globe (19), the positive effect of grain legumes on the yields of subsequent cereals is likely to be negligible. However, if low-input systems are widely adopted by crop producers in the future, the introduction of grain legumes to major cereal-producing regions will contribute the maintenance of high cereal yields.

### 3. Methods

#### 3.1. Literature search

We conducted a literature search using the Institute for Scientific Information Web of Science (25) with the following search equation in the article titles, abstracts and keywords: ‘crop\* AND (legum\* OR pulse\*) AND (yield\* OR ‘dry matter’ OR biomass) AND (compar\* OR assessment OR product\* OR performance\*) AND (trial\* OR factorial OR experiment\* OR treatment\* OR condition\*) NOT (intercrop\* OR catch OR cover OR ‘green manure’ OR forage OR fodder)’. No specifications were included concerning the date and the language of the article publication. The literature search was completed on February 15, 2016. We identified 8,386 potentially relevant articles, in total. These articles were comprehensively selected based on the following nine criteria: (i) at least one experiment conducted from seeding to harvest, (ii) either the title or the abstract of the article written in English, (iii) the full-text article published in a peer-reviewed journal, (iv) the full-text article available, (v) either primary data reported in one of the identified articles or primary data collected after contacting the authors, (vi) cereals cultivated after cereals and cereals cultivated after at least grain legumes at the same field sites during the same growing seasons using the same crop management techniques (note: if oilseeds were also cultivated as preceding crops, these species were included in the analysis), (vii) at least two grain legume species preceding cereals and cultivated as annual sole crops at the same field sites during the same growing seasons, (viii) yield data individually reported for each species cultivated as preceding and subsequent crops in the same crop sequence, and (ix) site location or soil properties precisely stated. We excluded field experiments that included only single grain legume species cultivated as preceding crop because one of our key objectives was to compare several grain

legume species cultivated at the same field sites during the same growing seasons. Finally, we selected thirty-three articles (Note S1) reporting results from field experiments conducted in fifteen countries (Table S2).

### 3.2. Data collection

Relevant information was extracted manually or digitally from the text, tables and figures of the articles. It should be emphasized that we used a subset of the data gathered from a global experimental dataset including 173 peer-reviewed articles (26). The subset considered here included data extracted from thirty-three selected articles. The data of this subset consisted of 1,181 yields of ten cereal species following either sixteen grain legume species or six oilseed species or nine cereal species. These nine preceding cereal species were used as a reference (Table S1). Wheat (*Triticum aestivum*) considered as a subsequent crop is the cereal species with the higher number of available yields in our dataset (56% of all yields of subsequent cereals). We excluded fallow management technique because of the frequent lack of accurate information in the selected articles. Cereal yields were systematically expressed in tonnes per hectare on a dry matter basis. If the percentage of dry matter was not available for all yields of subsequent cereals cultivated at the same field sites during the same growing seasons, we assumed that the yields of subsequent cereals were expressed on a dry matter basis. If the percentage of dry matter was not available for the yields of a subsequent cereal species but available for the yields of other subsequent cereal species cultivated at the same field sites during the same growing seasons, we assumed that the missing percentage of dry matter was the mean percentage of dry matter of the yields of subsequent cereal species having available information about their percentage of dry matter. Because the differences in N fertilization rates applied on non-cereals and cereals cultivated as preceding crops may influence the yields of subsequent cereals, these differences were categorized into four levels: (i) the same N fertilization rate was applied to both preceding non-cereal (either grain legume or oilseed) and cereal crops (30% of all yields of subsequent cereals), (ii) the N fertilization rate applied to preceding non-cereal crops was higher than that applied to preceding cereal crops (0% of all yields of subsequent cereals), (iii) the N fertilization rate applied to preceding non-cereal crops was lower than that applied to preceding cereal crops (33% of all yields of subsequent cereals), and (iv) the N fertilization rate was not available for either preceding non-cereal crops and/or preceding cereal crops (37% of all yields of subsequent cereals). If no N fertilization was applied to both preceding non-cereal and cereal crops, this information was categorized into the level (i). Information about the N fertilization rate applied to subsequent cereals was available for twenty-seven of the thirty-three selected articles. The range of N fertilization rate applied to a subsequent cereal species was systematically the same whatever the preceding species was (either grain legume or oilseed or cereal) for each field site and each growing season. The N fertilization rate was systematically expressed in kilogrammes of N per hectare. When necessary, the fertilization rates were

converted into N based on the percentage of N extracted from the original fertilizer formula in the article. For a few yield data, we assumed that ammonium sulphate included 21% N (24). The dataset is associated and freely reusable to the manuscript. The names and definitions of the variables included in the dataset are provided in Table S3.

### 3.3. Statistical analyses

Preceding crops were either grain legumes (53% of all yields of subsequent cereals) or oilseeds (19% of all yields of subsequent cereals) or cereals (28% of all yields of subsequent cereals). The yields of cereals cultivated after non-cereals were compared with those of cereals cultivated after cereals at the same field sites during the same growing seasons under the same crop management techniques. We calculated log-transformed cereal yield ratios as follows (Eq. 1):

$$L_{ipjk} = \log_e \left( \frac{\bar{Y}_{ipjk}}{\bar{Y}_{cpjk}} \right) \quad [1]$$

where  $L_{ipjk}$  is the log-transformed cereal yield ratio;  $\bar{Y}_{ipjk}$  is the yield of a cereal species cultivated after a non-cereal species  $i$ , averaged over experimental treatments for N fertilization rate  $p$  applied to the cereal species in field site×growing season combination  $k$  in article  $j$ ; and  $\bar{Y}_{cpjk}$  is the yield of the same cereal species grown after cereal species  $c$ , averaged over the same experimental treatments for the same N fertilization rate in the same field site×growing season combination in the same article. The log-transformed yield ratios calculated for cereals cultivated after *Brassica juncea*, *Brassica napus*, *Carthamus tinctorius*, *Crambe abyssinica*, *Helianthus annuus*, and *Linum usitatissimum* were then categorized as log-transformed yield ratios of cereals cultivated after oilseeds. First, the mean value of the log-transformed cereal yield ratios was estimated for each preceding non-cereal species using statistical mixed-effect models. Four statistical model variants ( $M1$ – $M4$ ) were fitted to the data (Table S4).  $M1$  and  $M2$  were fitted to the unweighted data, whereas  $M3$  and  $M4$  were fitted to the weighted data. As yield variances were available only for a few yield data, the log-transformed cereal yield ratios were weighted on the basis of the numbers of replicates  $n_{ipjk}$  and  $n_{cpjk}$  used to calculate  $\bar{Y}_{ipjk}$  and  $\bar{Y}_{cpjk}$ , respectively. When the cereal yields were averaged over more than one growing season, the number of replicates was considered the sum of the number of replicates over all growing seasons. The weight  $\omega_{ipjk}$  of each log-transformed cereal yield ratio was calculated (27) as follows (Eq. 2):

$$\omega_{ipjk} = \frac{n_{ipjk}n_{cpjk}}{n_{ipjk} + n_{cpjk}} \quad [2]$$

Statistical model variants  $M1$ – $M4$  were fitted to two different datasets because the number of replicates was not available for one of the thirty-three selected articles. Statistical model variants  $M1$

and  $M2$  were thus fitted to the original dataset (951 yield ratios, and thirty-three articles), whereas statistical model variants  $M3$  and  $M4$  were fitted to a subset of the original dataset that excluded the article lacking the number of replicates (887 yield ratios, and thirty-two articles). All statistical model variants were fitted using the ‘lme’ function of the ‘nlme’ (28) (version 3.1.111) of the R software package (29) (version 3.0.2). The statistical model variant leading to the lowest value of the Akaike Information Criterion (30) (hereafter AIC) was selected between  $M1$  and  $M2$  on the one hand and between  $M3$  and  $M4$  on the other hand to estimate the mean yield ratios  $R_i$  and the 95% confidence intervals for each non-cereal species cultivated as preceding crops.  $M2$  was selected for the unweighted data, and  $M4$  for the weighted data. Second, the log-transformed cereal yield ratios  $L_{tpjk}$  were related to the N fertilization rate applied to subsequent cereals using statistical mixed-effect models. Thirty-two statistical model variants ( $M5$ – $M36$ ) were fitted successively to test several types of weighting and several combinations of random effects (Table S4) using the ‘lme’ function of the ‘nlme’ (28) (version 3.1.111) of the R software package (29) (version 3.0.2), including additive covariates as the differences in N fertilization rates applied on preceding non-cereals and preceding cereals and the N fertilization rate applied to subsequent cereals (Table S4). The statistical model variants  $M5$ – $M36$  were fitted in turn to the same subset. This subset was smaller than the two previous datasets in which  $M1$ – $M4$  were fitted. In fact, we excluded six of the thirty-three selected articles because no information was available for the N fertilization rate applied to subsequent cereals in these articles. It should be noted that among these six articles, the number of replicates was also not available for one article. Finally, the statistical model variants  $M5$ – $M36$  were fitted to a subset that included 839 yields ratios, and twenty-seven articles. The number of replicates was systematically available for all yield data included in this subset. The statistical model variant leading to the lowest AIC value,  $M31$ , was then selected to estimate the mean response of the cereal yield ratios and the 95% confidence intervals as a function of the N fertilization rate applied to subsequent cereals. We analysed whether the raw residuals and standardized residuals were symmetrically distributed (Figs S3, S4, and S5).

### 3.4. Replacement of cereals with grain legumes

Three grain legume species were considered: (i) *Glycine max* because it was the most cultivated grain legume species worldwide from 2004 to 2014 (24), (ii) *Pisum sativum* because it had the highest number of yield ratios of all grain legume species in the meta-analysis, and (iii) *Lupinus angustifolius* because it demonstrated the highest significant mean yield ratio of all grain legume species in the meta-analysis. For each of these grain legume species, we considered  $x_i$  as the proportion of grain legume species  $i$  introduced to a cereal monoculture as  $0 \leq x_i \leq \frac{1}{2}$ . The fraction of the cultivated



area under cereal is thus equal to  $1 - x_i$ . This fraction is the sum of the two fractions of the cultivated areas: (i) the fraction of area  $x_i$  cultivated under cereals following grain legume species  $i$ , and (ii) the fraction of area  $1 - 2x_i$  cultivated under cereals following cereals. The mean expected relative change in cereal production compared with the relative change in cereal production of a scenario without grain legumes (in this situation,  $x_i = 0$ ) was calculated as follows (Eq. 3):

$$Z_{ip} = \frac{Y_{ip}x_i + Y_{cp}(1 - 2x_i)}{Y_{cp}} = 1 + x_i(R_{ip} - 2) \quad [3]$$

where  $Y_{ip}$  is the yield of cereals cultivated after grain legume species  $i$  for N fertilization rate  $p$ ;  $Y_{cp}$  is the yield of cereals cultivated after cereal species  $c$  for the same N fertilization rate; and  $R_{ip}$  is the mean ratio of  $Y_{ip}$  to  $Y_{cp}$ . Cereal production decreases as a function of  $x_i$  if  $R_{ip} < 2$  and increases otherwise. The rate of the decrease (increase) is equal to  $R_{ip} - 2$ . Four N fertilization rates applied to the cereals were assessed: 0 kg N ha<sup>-1</sup>, 50 kg N ha<sup>-1</sup>, 100 kg N ha<sup>-1</sup>, and 150 kg N ha<sup>-1</sup>. The mean yield ratios  $R_{ip}$  and the 95% confidence intervals were estimated for each grain legume species and each N fertilization rate using the statistical model variant M31. The values of  $Z_{ip}$  were calculated using the estimated mean yield ratios  $R_{ip}$ . Uncertainty envelopes encompassing all expected relative changes in cereal production were calculated using the estimated lower and upper values of the 95% confidence intervals for the estimated mean yield ratios.

## References

1. Boye J, Zare F, Pletch A (2010) Pulse proteins: processing, characterization, functional properties and applications in food and feed. *Food Res Int* **43**(2):414–431.
2. Foyer CH, et al. (2016) Neglecting legumes has compromised human health and sustainable food production. *Nat Plants* **2**(16112).
3. Preissel S, Reckling M, Schläfke N, Zander P (2015) Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: a review. *Field Crop Res* **175**:64–79.
4. Reckling M, et al. (2016) Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. *Front Plant Sci* **7**(669).

5. Crews TE, Peoples MB (2004) Legume versus fertilizer sources of nitrogen: ecological tradeoffs and human needs. *Agr Ecosyst Environ* **102**(3):279–297.
6. Evans J, McNeill AM, Unkovich MJ, Fettell NA, Heenan DP (2001) Net nitrogen balances for cool-season grain legume crops and contributions to wheat nitrogen uptake: a review. *Aust J Exp Agr* **41**(3):347–359.
7. Jensen CR, et al. (2004) The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop. *Eur J Agron* **20**(4):405–418.
8. Espinoza S, et al. (2012) Contribution of legumes to wheat productivity in Mediterranean environments of central Chile. *Field Crop Res* **133**(11):150–159.
9. Williams CM, et al. (2014) Effects of three pulse crops on subsequent barley, canola, and wheat. *Agron J* **106**(2):343–350.
10. Badaruddin M, Meyer DW (1994) Grain legume effects on soil nitrogen, grain yield, and nitrogen nutrition of wheat. *Crop Sci* **34**(5):1304–1309.
11. Armstrong EL, Heenan DP, Pate JS, Unkovich MJ (1997) Nitrogen benefits of lupins, field pea, and chickpea to wheat production in south-eastern Australia. *Aust J Agr Res* **48**(1):39–47.
12. Zander P, et al. (2016) Grain legume decline and potential recovery in European agriculture: a review. *Agron Sustain Dev* **36**(26):1–20.
13. Dakora FD, Aboyinga RA, Mahama Y, Apaseku J (1987) Assessment of N<sub>2</sub> fixation in groundnut (*Arachis hypogaea* L.) and cowpea (*Vigna unguiculata* L. Walp.) and their relative N contribution to a succeeding maize crop in northern Ghana. *MIRCEN J Appl Microb* **3**(4):389–399.
14. Ghosh PK, Singh NP (1996) Production potential of summer legumes-maize (*Zea mays*) sequence under varying levels of nitrogen. *Indian J Agron* **41**(4):525–528.
15. Cheruiyot EK, Mumera LM, Nakhone LN, Mwonga SM (2003) Effect of legume-managed fallow on weeds and soil nitrogen in following maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) crops in the Rift Valley highlands of Kenya. *Aust J Exp Agr* **43**(6):597–604.

16. Kirkegaard J, Christen O, Krupinsky J, Layzell D (2008) Break crop benefits in temperate wheat production. *Field Crop Res* **107**(3):185–195.
17. Seymour M, Kirkegaard JA, Peoples MB, White PF, French RJ (2012) Break-crop benefits to wheat in western Australia – insights from over three decades of research. *Crop Pasture Sci* **63**(1):1–16.
18. Angus JF, et al. (2015) Break crops and rotations for wheat. *Crop Pasture Sci* **66**(6):523–552.
19. Mueller ND, et al. (2014) A tradeoff frontier for global nitrogen use and cereal production. *Environ Res Lett* **9**(054002).
20. Hardarson G, Atkins C (2003) Optimising biological N<sub>2</sub> fixation by legumes in farming systems. *Plant Soil* **252**(1):41–54.
21. Peoples MB, et al. (2009) The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. *Symbiosis* **48**(1):1–17.
22. Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. *Plant Physiol* **131**(3):872–877.
23. Sinclair TR, Vadez V (2012) The future of grain legumes in cropping systems. *Crop Pasture Sci* **63**(6):501–512.
24. Statistics Division of Food and Agriculture Organization of the United Nations (2016) <http://faostat3.fao.org/home/E>. Accessed on August 2016.
25. Institute for Scientific Information Web of Science (2016) <http://wokinfo.com>. Accessed on February 2016.
26. Cernay C, Pelzer E, Makowski D (2016) A global experimental dataset for assessing grain legume production. *Sci Data* **3**(160084).
27. Wiebe N, et al. (2006) A systematic review identifies a lack of standardization in methods for handling missing variance data. *J Clin Epidemiol* **59**(4):342–353.

28. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Development Core Team (2013) R package ‘nlme’: linear and nonlinear mixed effects models. Version 3.1.111. <http://www.r-project.org>. Accessed on October 2013.
29. R Development Core Team (2013) R: a language and environment for statistical computing. Version 3.0.2. <http://www.r-project.org>. Accessed on October 2013.
30. Akaike H (1974) A new look at the statistical model identification. *IEEE T Automat Contr* **19**(6):716–723.

## Discussion



# Discussion

L'objectif général de ma thèse était d'identifier des espèces de légumineuses à graines potentiellement intéressantes à produire en Europe. Pour répondre à cet objectif, j'ai comparé un grand nombre d'espèces en utilisant plusieurs critères de production à l'échelle du cycle cultural : le rendement en grains et sa variabilité interannuelle, la biomasse aérienne totale, la quantité d'azote, la teneur en protéines et la quantité d'énergie brute dans les graines, et l'utilisation en eau. Ces comparaisons ont été réalisées à partir de (1) données statistiques issues de la base FAOSTAT, (2) données expérimentales situées en Europe, Amérique du Nord et Océanie, et (3) données nutritionnelles issues de la base FEEDIPEDIA. J'ai également comparé les effets de différentes espèces de légumineuses à graines sur les rendements des céréales suivantes à partir de données expérimentales issues de différentes régions du monde.

## **1. Comparer la variabilité interannuelle et les pertes de rendement des légumineuses à graines en Europe et Amérique**

Les légumineuses à graines ont été comparées à partir de leurs niveaux de variabilité interannuelle et de perte de rendement en Europe, Amérique du Nord et Amérique du Sud. Ces comparaisons ont été réalisées entre espèces de légumineuses d'une part, et entre espèces de légumineuses et non-légumineuses d'autre part. Nous avons utilisé des données statistiques de rendement et de surface issues de la base FAOSTAT, entre 1961 et 2013. Tout d'abord, nous nous sommes demandés si les rendements des légumineuses à graines étaient quantitativement plus variables d'une année sur l'autre que celui des non-légumineuses en Europe, Amérique du Nord et Amérique du Sud. Nous avons montré que les légumineuses à graines étaient caractérisées par des niveaux de variabilité interannuelle et de perte de rendement plus élevés que ceux des non-légumineuses en Europe. Au contraire, les différences de variabilité entre légumineuses et non-légumineuses étaient moins élevées en Amérique. Puis, nous nous sommes demandés si la variabilité interannuelle des rendements des légumineuses à graines pouvait expliquer leurs faibles surfaces récoltées en Europe. Nous avons montré que des espèces caractérisées par une variabilité élevée de leurs rendements en grains et récoltées sur des faibles surfaces étaient des légumineuses, alors que des espèces récoltées sur des surfaces élevées et caractérisées par une faible variabilité de leurs rendements en grains étaient des non-légumineuses, à l'exception du soja en Amérique du Nord et Amérique du Sud. Cependant, plusieurs espèces de légumineuses et non-légumineuses étaient caractérisées par une faible variabilité de rendement en grains et récoltées sur des faibles surfaces.

Cette étude représente la première quantification à large échelle de la variabilité interannuelle des rendements des légumineuses à graines et des non-légumineuses. Plusieurs originalités peuvent être identifiées par rapport à des études précédentes sur ce sujet.

Une première originalité est le nombre d'espèces caractérisées. Jusqu'à présent, les niveaux de variabilité étaient seulement comparés entre quelques espèces de légumineuses et non-légumineuses (Bues et al., 2013 ; LMC International, 2009a, 2009b ; Peltonen-Sainio et Niemi, 2012 ; Reckling, 2015). Nous avons comparé les niveaux de variabilité de neuf espèces de légumineuses à ceux de sept espèces de céréales, deux espèces d'oléagineux et une espèce de tubéreux. Une deuxième originalité est notre comparaison interrégionale de la variabilité interannuelle des rendements des légumineuses à graines conduite en Europe et Amérique. Une troisième originalité est la méthodologie employée. Trois mesures de variabilité ont été appliquées à des séries chronologiques de rendement en grains. Pour chaque mesure de variabilité, nous avons utilisé deux méthodes de régression pour enlever la tendance des séries chronologiques. Le choix de la méthode de régression affecte peu les classements des espèces légumineuses et non-légumineuses. La sensibilité des résultats à la durée de la série chronologique a également été testée. Les trois mesures de variabilité ont été calculées entre 1983 et 2013. Dans l'ensemble, les résultats sont similaires à ceux obtenus entre 1961 et 2013. Enfin, la méthode du bootstrap a permis de quantifier l'incertitude associée à chaque estimation.

Notre estimation de la variabilité des rendements en grains en Europe de l'Ouest a été discutée avec celle estimée en France par des collègues à partir de séries chronologiques de statistique nationale pour de mêmes espèces (Schneider et al., 2015). Leurs estimations ont été réalisées sans enlever la tendance temporelle du rendement. Cette analyse statistique des séries chronologiques peut biaiser la comparaison des niveaux de variabilité entre les rendements des légumineuses à graines et ceux des céréales. En effet, le progrès technologique aurait été plus investi envers les céréales que les légumineuses à graines en Europe depuis 50 ans (Bues et al., 2013). Cette différence aurait permis un accroissement plus élevé des rendements des céréales que ceux des légumineuses, ayant potentiellement un effet significatif sur les tendances de leurs séries chronologiques. Une confusion d'effets est alors possible. Une apparente plus grande variabilité des rendements en grains des céréales refléterait un accroissement tendanciel plus élevé de leurs niveaux de rendement que ceux des légumineuses à graines. Pour cette raison, nous avons préféré travailler sur des séries chronologiques corrigées de leurs tendances. Cependant, la littérature européenne compare souvent la variabilité interannuelle des rendements en grains des céréales à celle des légumineuses à graines sans enlever la tendance (Bues et al., 2013 ; LMC International, 2009a, 2009b ; Reckling, 2015 ; Schneider et al., 2015). Peltonen-Sainio et Niemi (2012) avaient comparé la variabilité interannuelle des rendements en grains en excluant la tendance temporelle pour plusieurs espèces de légumineuses en Europe. Cependant, les auteurs avaient peu explicité leurs méthodes de correction appliquées aux séries



chronologiques. Reckling et al. (2016a, 2016b) et Zander et al. (2016) ont mentionné nos résultats avec ceux d'études où les tendances des séries chronologiques n'ont pas été exclues, corroborant ainsi des différences de variabilité entre céréales et légumineuses plus faibles que nos estimations.

Notre estimation à large échelle de la variabilité interannuelle des rendements des légumineuses à graines est complémentaire des estimations d'études précédentes réalisées à des échelles spatiotemporelles moins larges. Reckling (2015) ont par exemple comparé les coefficients de variation des rendements de lupin à feuilles étroites (*Lupinus angustifolius*) et de pois protéagineux (*Pisum sativum*) à ceux des rendements de maïs (*Zea mays*), de seigle (*Secale cereale*), de blé (*Triticum aestivum*), et d'avoine (*Avena sativa*), en Allemagne, à partir d'expérimentations agronomiques. Les auteurs ont utilisé des séries chronologiques courtes et discontinues (i.e., entre 1999 et 2004, entre 2006 et 2009, entre 2011 et 2013). Peltonen-Sainio et Niemi (2012) ont comparé les coefficients de variation des rendements de pois protéagineux (*Pisum sativum*) à ceux des rendements de blé (*Triticum aestivum*) et de colza (*Brassica napus*) dans six pays européens (i.e., Allemagne, Danemark, Espagne, Finlande, France, et Suède) entre 1963 et 2007, à partir de la base de données FAOSTAT. Les auteurs ont également comparé les coefficients de variation des rendements de pois protéagineux (*Pisum sativum*) et de fève (*Vicia faba*) à ceux des rendements d'orge (*Hordeum vulgare*) et de colza (*Brassica napus*) en Finlande en 2011, à partir de données statistiques nationales. LMC International (2009a, 2009b) ont comparé les coefficients de variation des rendements de lupins (*Lupinus albus*, *Lupinus angustifolius*, et *Lupinus luteus*), de fève (*Vicia faba*) et de pois protéagineux (*Pisum sativum*) à des rendements de blé (*Triticum aestivum*), de maïs (*Zea mays*), de colza (*Brassica napus*), d'orge (*Hordeum vulgare*), et de tournesol (*Helianthus annuus*) en Europe entre 2000 et 2007, à partir de la base de données FAOSTAT. Nos estimations sont complémentaires des estimations d'études précédentes qui ont quantifié la variabilité interannuelle de critères différents du rendement en grains. LMC International (2009a, 2009b) ont par exemple comparé les coefficients de variation des marges comptables des cultures de fève (*Vicia faba*) et de pois protéagineux (*Pisum sativum*) à celles des cultures de blé (*Triticum aestivum*), de maïs (*Zea mays*), d'orge (*Hordeum vulgare*), et de colza (*Brassica napus*) entre 2000 et 2007, à partir de données statistiques de petites régions agricoles dans quatre pays européens (i.e., Allemagne, Angleterre, Espagne, et France). Enfin, notre caractérisation quantitative est complémentaire des estimations d'études précédentes qui ont caractérisé qualitativement la variabilité interannuelle des rendements des légumineuses à graines. A partir d'enquêtes réalisées auprès d'agriculteurs dans quatre pays européens (i.e., Allemagne, Belgique, Espagne, et Suisse), Von Richthofen et al. (2006) ont mis en évidence que les pertes de production des légumineuses d'une année sur l'autre représenteraient un inconvénient potentiel à les cultiver. A partir d'entretiens conduits auprès d'agriculteurs au Luxembourg, Zimmer et al. (2016a) ont identifié que les agriculteurs manquaient de différentes informations pour cultiver plusieurs espèces de légumineuses, en particulier leurs niveaux de rendement et de variabilité associée. Dans le

cadre du programme « Legume Futures » (2011-2014), Stoddard (2013) a également reporté ce constat sur la base d'entretiens semi-directifs réalisés auprès de producteurs de légumineuses dans treize pays européens.

Plusieurs facteurs peuvent expliquer les différences de variabilité interannuelle des rendements en grains entre les régions, les sous-régions et les espèces. La forte sensibilité de la fixation symbiotique des légumineuses aux stress biotiques et abiotiques (Aranjuelo et al., 2014 ; Araújo et al., 2015 ; Rubiales et al., 2015) pourraient en partie expliquer ces différences. Cependant, les légumineuses ne présentent pas les mêmes niveaux de variabilité en Europe et Amérique. Les résultats montrent par exemple que le rendement du soja en Amérique est moins variable qu'en Europe. L'utilisation de variétés spécifiquement adaptées en Amérique (Leff et al., 2004), la variabilité des conditions pédoclimatiques (Peel et al., 2007) ou celle des aires de répartition des bioagresseurs (Vadez et al., 2012) peuvent expliquer ces contrastes interrégionaux. Les différences de caractéristiques des systèmes de culture en Europe et Amérique sont importantes à inclure. L'utilisation plus importante du soja génétiquement modifié en Amérique du Nord et Amérique du Sud peut contribuer aux écarts observés des niveaux de variabilité par rapport à ceux estimés en Europe, de manière similaire à ce que Shi et al. (2013) avaient montré à propos du risque de perte de rendement du maïs génétiquement modifié en Amérique du Nord. Certaines études ont mis en évidence un manque d'information à disposition des agriculteurs et des conseillers agricoles pour cultiver des légumineuses à graines dans certains pays européens (Stoddard, 2013 ; Voisin et al., 2014 ; Von Richthofen et al., 2006 ; Zimmer et al., 2016a). Cette absence de références techniques concerne les variétés, les niveaux de rendement, et les techniques culturales (Stoddard, 2013 ; Voisin et al., 2014 ; Von Richthofen et al., 2006 ; Zimmer et al., 2016a), et représentant un frein possible à introduire ces espèces dans les systèmes de culture européens. D'autres études soulignent un déficit en amélioration génétique d'une diversité de légumineuses à graines en Europe (Bues et al., 2013 ; Murphy-Bokern et al., 2014 ; Voisin et al., 2014 ; Wiggering et al., 2012), et identifié comme un facteur probable de la variabilité de leurs rendements en grains plus élevée que celle des rendements en grains des non-légumineuses. Ces études soulignent également la priorité d'augmenter les efforts en sélection variétale pour améliorer l'adaptation des cultures de légumineuses à la mosaïque des conditions pédoclimatiques et à la diversité des techniques culturales en Europe (Häusling, 2011 ; Magrini et al., 2016 ; Stoddard, 2013 ; Voisin et al., 2014 ; Zander et al., 2016). Au-delà des effets de ces facteurs affectant potentiellement la variabilité interannuelle des rendements en grains, Marra et Schurle (1994) et Popp et al. (2005) ont souligné l'importance de l'échelle spatiale d'agrégation des données de rendement. Plus cette échelle est large, plus il est probable que l'effet des facteurs affectant le rendement soit atténué. Dans notre étude, des données de rendement agrégées ont été utilisées à large échelle spatiale (e.g., en Europe de l'Ouest et Amérique du Nord). Il est ainsi probable que l'effet des facteurs affectant la variabilité interannuelle des rendements en grains ait pu être atténué. Par ailleurs, nous avons estimé les niveaux

de variabilité à l'échelle de régions définies par des limites géographiques. Il est probable que les aires de culture des légumineuses ne soient pas définies par ces limites. Dans une même région géographique, il est également possible que les aires de culture des espèces de légumineuses ne se superposent pas ou peu. Ces discontinuités spatialement explicites pourraient ainsi contribuer aux différences de leurs niveaux de variabilité.

Si ces facteurs sont sous-jacents de la variabilité interannuelle des rendements des légumineuses à graines, nous avons mis en évidence que ce phénomène n'est pas une condition suffisante pour expliquer leurs faibles surfaces récoltées en Europe. D'autres mécanismes – dont certains sont économiques – peuvent représenter des sources de variation des surfaces récoltées : le prix de vente de la culture (Bues et al., 2013 ; Zander et al., 2016), la dépendance historique de la production européenne de légumineuses aux mesures de la Politique Agricole Commune (Bues et al., 2013 ; Hache, 2015 ; Voisin et al., 2014 ; Zander et al., 2016), le faible niveau de transformation des graines de légumineuses en Europe (Magrini et al., 2016 ; Voisin et al., 2014 ; Zander et al., 2016), la compétition de la production européenne de légumineuses avec les importations de soja provenant d'Amérique (Bues et al., 2013 ; Hache, 2015 ; Voisin et al., 2014), ou la concurrence de coproduits riches en protéines végétales mais non dérivés de légumineuses (e.g., les tourteaux de colza (*Brassica napus*) ; De Visser et al., 2014 ; Peltonen-Sainio et Niemi, 2012). Entre 1961 et 2013, il est probable que les effets de ces mécanismes ne soient pas négligeables sur la prise de décision des agriculteurs à cultiver des légumineuses en Europe. En Australie, des études ont par exemple mis en évidence une relation entre la variabilité des rendements en grains des légumineuses et certains facteurs économiques affectant leurs productions respectives. Schilizzi et Kingwell (1999) ont analysé l'effet de la variabilité interannuelle du rendement en grains associé à celui de la variabilité de leur prix de vente sur le risque perçu par les agriculteurs australiens à introduire du pois protéagineux (*Pisum sativum*), de la fève (*Vicia faba*) et du pois chiche (*Cicer arietinum*) dans leurs exploitations. Ghadim et al. (2005) ont également mis en évidence une relation entre le risque réel de cultiver du pois chiche (*Cicer arietinum*) – dont ses pertes de rendement d'une année à l'autre – et le risque perçu par les agriculteurs à cultiver cette espèce en Australie de l'Ouest. Aussi, les agriculteurs n'auraient peut-être pas été suffisamment incités à cultiver des légumineuses à graines dans des environnements adaptés, et auraient tendance à cultiver des non-légumineuses à la place (Peltonen-Sainio et Niemi, 2012 ; Stoddard, 2013 ; Voisin et al., 2014 ; Zander et al., 2016). Enfin, certaines études soulignent que la combinaison de plusieurs facteurs techniques, économiques et organisationnels – dont certains mentionnés précédemment – aurait favorisé l'émergence d'un système agroindustriel organisé « en défaveur » des cultures de légumineuses (Bues et al., 2013 ; Hache, 2015 ; Häusling, 2011 ; Magrini et al., 2016 ; Schneider et al., 2015 ; Voisin et al., 2014 ; Zander et al., 2016). D'après ces études, une augmentation de la production des légumineuses à graines en Europe nécessiterait de « déverrouiller le régime sociotechnique » actuel en actionnant un ensemble de leviers et d'acteurs afin d'enclencher des

« rendements croissants d'adoption » favorables aux cultures de légumineuses.

Dans l'ensemble, nos résultats peuvent aider à définir des politiques européennes de soutien à la production de légumineuses à graines en Europe. Dans le cadre de mesures agroenvironnementales adaptées, ces politiques peuvent être modulées en fonction de l'espèce et la sous-région européenne où elle est récoltée. Des contrats d'assurance peuvent par exemple être mis place contre des fortes pertes de rendement en se basant sur nos estimations. Ces contrats pourraient alors proposer des compensations financières qui répondraient aux attentes des agriculteurs. D'après les résultats, les compensations à destination des cultures de légumineuses seraient plus élevées que celles des non-légumineuses en Europe, en particulier pour les lupins (*Lupinus* spp.).

## **2. Une base de données expérimentales pour comparer les performances productives des légumineuses à graines à l'échelle mondiale**

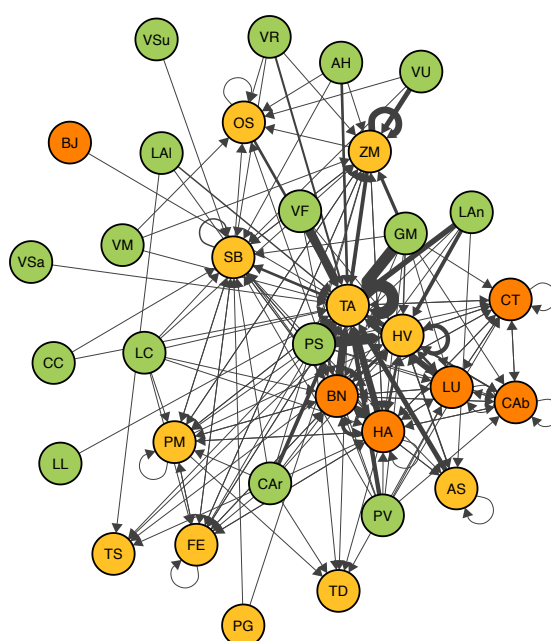
Dans le chapitre 1, nous avons utilisé des données statistiques pour estimer la variabilité interannuelle des rendements des légumineuses à graines. Les données FAOSTAT ont permis d'appréhender cette performance productive à des larges échelles spatiotemporelles. Une appréciation qui n'était pas possible de réaliser uniquement à partir des données issues d'expérimentations agronomiques, où les séries chronologiques de rendement en grains sont très souvent courtes et discontinues (i.e., inférieures à 10 ans). Cependant, l'utilisation de cette source de données statistiques soulève certaines limites méthodologiques pour comparer des espèces de légumineuses à graines entre elles. L'utilisation de la base de données FAOSTAT peut générer un risque de confusion d'effets liés aux caractéristiques des légumineuses (e.g., la fixation symbiotique), l'environnement (e.g., les conditions pédoclimatiques), et les techniques culturales (e.g., le travail du sol, la fertilisation, le contrôle des bioagresseurs, et l'irrigation). Cette confusion est liée à l'utilisation de sources de données statistiques préexistantes dans les calculs de la base de données FAOSTAT. Certaines espèces de légumineuses à graines – souvent récoltées sur des faibles surfaces – sont regroupées dans une catégorie culturelle avec la même dénomination. Cette confusion est créée par la nomenclature de la base de données FAOSTAT. Le rendement est le principal critère de production des différentes cultures dans cette base. La biomasse aérienne, la quantité d'azote dans les graines ou l'utilisation en eau sont des critères absents. Au-delà de l'échelle du cycle cultural, les performances productives des cultures qui suivent les légumineuses à graines ne sont pas présentes. Il n'est donc pas possible d'estimer l'effet des légumineuses sur la production des céréales ou des oléagineux suivants dans la séquence culturale. L'utilisation des données issues d'expérimentations agronomiques (publiées ou non) conduites dans différentes régions du monde permet de contourner en grande partie ces limites lorsque plusieurs espèces sont comparées sur les mêmes expérimentaux durant les mêmes saisons culturales. Les caractéristiques des légumineuses à graines (e.g., les noms scientifiques des espèces), les conditions climatiques, les types

de sols et les techniques culturales sont en général précisés dans les expérimentations. L'identification précise de ces informations permet de limiter le risque de confusion d'effets. Les lupins (*Lupinus* spp.) sont par exemple regroupés dans une seule catégorie culturale dans la base de données FAOSTAT alors qu'il en existe plusieurs espèces (e.g., *Lupinus albus*, *Lupinus angustifolius*, *Lupinus atlanticus*, *Lupinus luteus* et *Lupinus pilosus*), dont les productions ont été mesurées individuellement dans des expérimentations. Cette information spécifique permet également d'identifier des espèces de légumineuses qui ne sont pas catégorisées dans la nomenclature de la base de données FAOSTAT en raison de leurs faibles importances économiques à l'échelle mondiale. Les performances productives des cultures ne sont généralement pas limitées au rendement en grains. La biomasse aérienne totale (ou l'indice de récolte), la quantité d'azote dans les parties aériennes (ou l'indice de récolte en azote), l'utilisation en eau (ou l'efficacité d'utilisation en eau des parties aériennes) peuvent être mesurées. L'effet des légumineuses sur la production des cultures qui les suivent et les flux d'azote sont également quantifiés dans plusieurs expérimentations (e.g., la quantification de l'azote dans le sol après la récolte des légumineuses et/ou avant le semis des cultures suivantes). En ce sens, les données expérimentales sont complémentaires des données statistiques afin de comparer les performances productives (et environnementales) d'une large diversité spécifique de légumineuses à graines.

Dans le chapitre 2, nous avons présenté la procédure de construction et le contenu d'une base de données expérimentales pour comparer les performances productives (et environnementales, dans une moindre mesure) des légumineuses à graines à l'échelle mondiale. Ce chapitre est descriptif et a été valorisé sous la forme d'un « data paper », une nouvelle forme de publication centrée sur les bases de données. Les informations ont été incluses pour 39 espèces de légumineuses à graines. Le périmètre de la base de données dépasse donc le nombre d'espèces de légumineuses principalement récoltées dans le monde. Les données ont été extraites à partir de 173 articles publiés entre 1967 et 2016. Les informations collectées proviennent de 360 sites expérimentaux situés à travers 18 zones climatiques dans 41 pays sur cinq continents. La base de données contient 8581 situations de culture. Pour chaque observation, les références des articles, les coordonnées géographiques, les conditions pédoclimatiques des sites expérimentaux, les rendements en grains, les biomasses aériennes, les utilisations en eau, les quantités en azote dans le sol ont été enregistrés. La production des céréales et des oléagineux cultivés après des légumineuses a également été collectée. L'erreur associée à chaque valeur de production a été extraite. Quand les données étaient renseignées, les techniques culturales ont été enregistrées (i.e., le travail du sol, la fertilisation en azote, phosphore et potassium, le contrôle des adventices, insectes et champignons, et l'irrigation). La base de données est structurée en neuf tables qui suivent une organisation en cascade. La table Culture est la table centrale de la base de données puisqu'elle regroupe l'ensemble des informations des cultures. Dans cette table, le rendement en grains est l'attribut qui a le plus grand nombre de données en raison de la condition systématique de présence de données de rendement dans les articles sélectionnés. Quand les données n'étaient pas disponibles pour

certain attributs concernant des critères de production (e.g., la biomasse aérienne ou l'utilisation en eau), les valeurs d'attributs qui en dépendent (e.g., l'indice de récolte ou l'efficacité en eau des graines) ont été extraites afin d'en déduire les données manquantes. Dans certains cas, les auteurs des articles ont été contactés afin d'obtenir les données brutes.

A partir de notre base de données, des réseaux de légumineuses et non-légumineuses peuvent être obtenus. Ces réseaux permettent non seulement d'identifier les paires d'espèces qui sont fréquemment (ou peu fréquemment) comparées entre elles sur les mêmes sites durant les mêmes saisons culturales, mais également d'identifier les séquences culturales de légumineuses (en tant que cultures précédentes) et de non-légumineuses (en tant que cultures suivantes) qui sont les plus (ou les moins) testées sur les mêmes sites expérimentaux dans différentes régions du monde (Figure 1).



**Figure 1.** Réseau des couples de culture précédente et de culture suivante incluses dans la base de données. Chaque nœud correspond à une espèce cultivée. Chaque flèche indique un couple de culture précédente et de culture suivante. Le sens de la flèche indique l'espèce qui est la culture précédente d'une autre espèce qui est la culture suivante ; les deux espèces étant cultivées sur le même site expérimental durant deux saisons culturales successives. Si la flèche d'un nœud est une boucle, alors l'espèce est cultivée en monoculture. L'épaisseur de la flèche indique le nombre de sites expérimentaux et de saisons culturales pour lesquelles les deux espèces culturales se suivent. L'espèce est soit une légumineuse à graines (vert), une céréale (jaune) ou un oléagineux (orange). Les acronymes indiquent les noms scientifiques des espèces : AH : *Arachis hypogaea*, AS : *Avena sativa*, BJ : *Brassica juncea*, BN : *Brassica napus*, CC : *Cajanus cajan*, CT : *Carthamus tinctorius*, CAr : *Cicer arietinum*, Cab : *Crambe abyssinica*, FE : *Fagopyrum esculentum*, GM : *Glycine max*, HA : *Helianthus annuus*, HV : *Hordeum vulgare*, LC : *Lens culinaris*, LU : *Linum usitatissimum*, LAI : *Lupinus albus*, LAn : *Lupinus angustifolius*, LL : *Lupinus luteus*, OS : *Oryza sativa*, PM : *Panicum miliaceum*, PG : *Pennisetum glaucum*, PS : *Pisum sativum*, PV : *Phaseolus vulgaris*, SB : *Sorghum bicolor*, TA : *Triticum aestivum*, TD : *Triticum durum*, TS : *Triticum sativum*, VF : *Vicia faba*, VM : *Vigna mungo*, VR : *Vigna radiata*, VSa : *Vicia sativa*, VSu : *Vigna subterranea*, VU : *Vigna unguiculata*, ZM : *Zea mays*.

La base de données présente cependant certaines limites. Le nombre d'articles qui étudient les performances des légumineuses à graines est probablement plus élevé que les 173 articles dont nous avons extrait les données. En effet, nous avons inclus des articles où au moins deux espèces de légumineuses étaient systématiquement comparées sur le même site expérimental durant la(es) même(s) saison(s) culturale(s), et dont les données de rendement étaient systématiquement renseignées. Néanmoins, ces deux conditions initiales étaient nécessaires pour comparer le plus grand nombre d'espèces de légumineuses entre elles tout en minimisant les confusions d'effets liés à l'environnement et aux techniques culturales. Le coût en temps de la construction de la base de données n'a pas été négligeable. Cette construction a été effectuée en trois étapes : (1) la définition du modèle relationnel, (2) l'extraction des données, et (3) la standardisation et la vérification des données implémentées. Dans le cadre de la thèse, les étapes (2) et (3) ont duré 15 mois ; l'étape (2) a représenté approximativement 75% de cette durée. De nombreux allers-retours ont été nécessaires entre les articles et la base de données. Si certains articles présentaient des expérimentations assez simples, d'autres décrivaient des expérimentations plus complexes, incluant plusieurs combinaisons de sites et saisons culturales caractérisées par différents niveaux de facteurs liés aux espèces et aux techniques culturales. La durée d'extraction des données a parfois été importante car les observations n'étaient disponibles que sous forme de figures. Leur digitalisation a été par conséquent nécessaire pour extraire individuellement chaque valeur moyenne et son erreur associée en utilisant le logiciel libre d'accès WebPlotDigitizer (Figure 7). L'extraction des données a également requis une durée significative en raison du grand nombre d'attributs liés aux performances des cultures (e.g., la table Culture comprend à elle seule 106 attributs) et aux techniques culturales. Le temps de construction de la base de données aurait pu être diminué de 50% si les informations relatives aux techniques culturales n'avaient pas été enregistrées ; elles représentent cependant une des grandes originalités du jeu de données. Si la construction de la base de données a demandé des moyens humains non négligeables, elle requiert peu de moyens matériels. Une des difficultés à construire la base de données concerne la sélection objective de l'information et de sa standardisation en des attributs génériques. Par exemple, la structure des attributs de la table Culture comprend au minimum quatre attributs pour décrire un critère de production pour une observation culturelle donnée. Ces attributs renseignent (1) la valeur moyenne, (2) l'unité de mesure, (3) la valeur, et (4) le type de l'erreur associée. D'autres attributs sont nécessaires à inclure pour informer la méthodologie utilisée pour mesurer un critère de production. Par exemple, le pourcentage d'azote fixé par la légumineuse est informé par trois attributs distincts dans le jeu de données : (1) la méthode de mesure de l'azote fixé (e.g., la méthode de l'abondance naturelle  $^{15}\text{N}$  ou celle de la dilution isotopique), (2) le nom scientifique de l'espèce non-fixatrice de référence, et (3) le stade phénologique auquel l'azote fixé a été mesuré durant l'expérimentation. Par conséquent, la base de données est limitée dans sa structuration horizontale en raison de la traduction de l'information en des attributs multiples. Le constructeur de la base de données doit faire alors un arbitrage contextuel entre les degrés d'exhaustivité et de précision de l'information à implémenter.

L'hétérogénéité et la disponibilité de cette information entre les articles peuvent également être contraignantes. Par exemple, la présence ou l'absence de travail du sol est l'une des techniques culturales les plus renseignées dans la base de données ; celles relatives au contrôle des bioagresseurs ne sont les plus renseignées qu'à propos de la présence ou l'absence de contrôle mécanique (ou chimique) des adventices. Les doses des produits phytosanitaires sont en effet peu renseignées et très hétérogènes entre les articles. Initialement, nous avons souhaité inclure la présence ou l'absence d'une maladie pour chaque observation culturale afin d'utiliser cette information comme un facteur explicatif de sa possible moindre performance. Très peu d'articles ont renseigné quantitativement cette donnée ; pour cette raison, nous ne l'avons pas inclus. Enfin, la structure relationnelle de la base de données est relativement coercitive : l'inclusion d'une nouvelle table peut avoir des conséquences sur le modèle relationnel de la base et les relations de cardinalité entre les attributs des différentes tables. Cependant, de nouvelles données et de nouveaux attributs peut être facilement être incrémentés à partir des tables existantes.

Depuis cinquante ans, de nombreuses expérimentations ont comparé les performances agronomiques et environnementales de différentes espèces de légumineuses à graines. Ces performances varient entre les sites expérimentaux, les saisons culturales et les techniques culturales. Il est difficile d'établir des comparaisons plurispécifiques à large échelle à partir d'expérimentations individuelles considérées séparément. La base de données permet de contourner cette difficulté. Le jeu de données permet d'analyser la variabilité de la production d'une diversité de légumineuses et de quantifier l'incertitude associée au travers d'un large spectre de conditions environnementales. Le jeu de données permet également de hiérarchiser des espèces de légumineuses en fonction de plusieurs critères de production. Les classements obtenus peuvent permettre de comparer les performances d'espèces faiblement récoltées avec celles d'espèces fortement récoltées. Ces classements peuvent alors apporter un nouvel éclairage sur l'identification d'espèces à forts niveaux de production. La base de données peut également être utilisée pour évaluer l'effet de différentes espèces de légumineuses à graines sur la production des céréales et des oléagineux qui les suivent à l'échelle de la séquence culturale. Cet effet peut alors être exprimé en fonction de l'espèce de légumineuses et de différentes covariables agronomiques et environnementales. Les deux méta-analyses des chapitres 3 et 4 ont valorisé la base de données en ce sens.

Une première valorisation future de la base de données pourrait être de la connecter à d'autres bases existantes, dont certaines se présentent sous forme de systèmes d'information géographique. Les coordonnées de latitude et de longitude de chaque site expérimental ont été systématiquement renseignées. Notre jeu de données peut donc être connecté à des bases de données climatiques ou édaphiques où les coordonnées géographiques sont des données vectorielles. Dans le chapitre 2, nous avons illustré cette approche de manière simple en connectant les 360 sites expérimentaux de la base



de données à la classification climatique de Köppen-Geiger (Kottek et al., 2006). Des cartes similaires peuvent être construites en utilisant des classifications de la FAO concernant des zones agroécologiques ou des types de sol. Ces cartes permettraient d'identifier les cultures de légumineuses les mieux adaptées à certaines régions. Dans le contexte du changement climatique, la base de données pourrait alors représenter un support intéressant pour évaluer comparativement la production d'une large diversité de légumineuses en fonction de gradients de température et de précipitation. Une seconde valorisation possible de la base de données serait d'y ajouter les résultats de nouvelles expérimentations agronomiques où des légumineuses à graines seraient comparées entre elles. Cette mise à jour est facile à réaliser : la base de données a été publiée dans une revue « open source » et mise à disposition de la communauté scientifique. Les légumineuses peuvent être cultivées selon des modalités différentes de celles que nous avons considérées dans la thèse. En Asie, Afrique et Amérique du Sud, les légumineuses à graines sont souvent cultivées en association avec d'autres espèces de légumineuses ou non-légumineuses. Dans le futur, ce mode d'insertion pourrait être inclus dans la base de données sans changement majeur de la structure relationnelle. Dans certaines régions du monde, les légumineuses sont également cultivées pour la production de fourrages. Ce mode d'exploitation peut également être inclus dans la base de données sans modification de l'organisation en cascade entre les tables et des relations de cardinalité entre les attributs.

**3. Comparer les niveaux de production des légumineuses à graines en Europe, Amérique du Nord et Océanie**



## Discussion

Nom scientifique	Europe				Ratio de rendement en grains <sup>b</sup>	Ratio de biomasse aérienne totale <sup>b</sup>	Ratio de protéines dans les graines <sup>b,c</sup>	Ratio d'énergie brute dans les graines <sup>b,c</sup>	Ratio d'azote dans les grains <sup>b</sup>	Ratio de variance de rendement en grains <sup>a</sup>	Effet sur les rendements des céréales suivantes <sup>b</sup>	N <sup>o</sup>
	de l'Ouest de l'Europe	de l'Est de l'Europe	du Nord de l'Europe	du Sud de l'Europe								
<i>Arachis hypogaea</i>											+	139
<i>Cajanus cajan</i>											+	115
<i>Cicer arietinum</i>											+	124
<i>Glycine max</i>	4	1			0	-	0	(0)	0	-	+	135
<i>Lathyrus aphaca</i>												
<i>Lathyrus cicera</i>												
<i>Lathyrus clymenum</i>												
<i>Lathyrus ochrus</i>												
<i>Lathyrus sativus</i>												
<i>Lens culinaris</i>												
<i>Lupinus albus</i>	5	5	5	5	(-)	(0)	(-)	(-)	(-)	-	+	184
<i>Lupinus angustifolius</i>	5	5	5	5	-	(-)	0	0	0	-	+	185
<i>Lupinus atlanticus</i>	5	5	5	5	(0)	(0)	(0)	(0)	(0)	-		
<i>Lupinus luteus</i>	5	5	5	5	(0)	(0)	(0)	(0)	(0)	-		
<i>Lupinus pilosus</i>	5	5	5	5								
<i>Phaseolus vulgaris</i>	2	4	4	4							+	129
<i>Pisum sativum</i>	3	2	1	5							+	168
<i>Vicia articulata</i>												
<i>Vicia berghalensis</i>												
<i>Vicia ervilla</i>												
<i>Vicia faba</i>	1		2	1	0	0	0	0	0	0	+	198
<i>Vicia hybrida</i>												
<i>Vicia narbonensis</i>												
<i>Vicia sativa</i>		3	3	2							(0)	115
<i>Vicia villosa</i>												
<i>Vigna mungo</i>											+	133
<i>Vigna radiata</i>											+	155
<i>Vigna subterranea</i>											(0)	
<i>Vigna unguiculata</i>											+	135



## Discussion

Nom scientifique	Amérique du Nord		Amérique du Nord							Océanie				
	Ecart-type des anomalies <sup>a</sup>	de rendement en grains <sup>a</sup>	Ratio de rendement en grains <sup>b</sup>	Ratio de biomasse aérienne totale <sup>b</sup>	Ratio de protéines dans les graines <sup>b,c</sup>	Ratio d'énergie brute dans les graines <sup>b,c</sup>	Ratio d'azote dans les graines <sup>b</sup>	Ratio d'utilisation en eau <sup>b</sup>	Ratio de variance de rendement en grains <sup>a</sup>	Ratio de rendement en grains <sup>b</sup>	Ratio de biomasse aérienne totale <sup>b</sup>	Ratio de protéines dans les graines <sup>b,c</sup>	Ratio d'énergie brute dans les graines <sup>b,c</sup>	Ratio de variance de rendement en grains <sup>a</sup>
<i>Arachis hypogaea</i>	3	5												
<i>Cajanus cajan</i>														
<i>Cicer arietinum</i>			-	-	-	-	(-)	-						-
<i>Glycine max</i>	2	3	-	(0)	0	-								+
<i>Lathyrus aphaca</i>														
<i>Lathyrus ciceria</i>														
<i>Lathyrus clymenum</i>														
<i>Lathyrus ochrus</i>														
<i>Lathyrus sativus</i>			(0)	(0)	(0)	(0)	(0)	(-)						
<i>Lens culinaris</i>	5		-	-	-	-	-	0						0
<i>Lupinus albus</i>														
<i>Lupinus angustifolius</i>			-		-	-	(-)	(0)						
<i>Lupinus atlanticus</i>														
<i>Lupinus luteus</i>														
<i>Lupinus pilosus</i>														
<i>Phaseolus vulgaris</i>	1	2	-		-	-	(-)	(0)						0
<i>Pisum sativum</i>	4	1												
<i>Vicia articulata</i>														
<i>Vicia benghalensis</i>														
<i>Vicia ervilia</i>														
<i>Vicia faba</i>		4	-	(0)	0	-	(0)	(0)						0
<i>Vicia hybrida</i>														
<i>Vicia narbonensis</i>														
<i>Vicia sativa</i>														
<i>Vicia villosa</i>														
<i>Vigna mungo</i>	1	2												0
<i>Vigna radiata</i>	1	2												0
<i>Vigna subterranea</i>	1	2												0
<i>Vigna unguiculata</i>														



**Figure 2.** Synthèses quantitatives des performances productives (et environnementales, dans une moindre mesure) des légumineuses à graines en Europe, Amérique et Océanie. Les résultats sont issus des chapitres 1, 3, et 4 de la thèse. Dans la première partie du tableau, les écart-types des anomalies du rendement sont issus du chapitre 1, dans quatre sous-régions européennes. Les numéros indiquent le classement des espèces de légumineuses à graines dans un ordre décroissant de variabilité interannuelle du rendement en grains : « 1 » indique la plus faible variabilité interannuelle, alors que « 5 » indique la plus forte variabilité interannuelle, pour une sous-région européenne donnée. Les numéros « 1 » et « 2 » sont en vert, « 3 » et « 4 » en jaune, et « 5 » en orange. Nous faisons l'hypothèse que la variabilité interannuelle du rendement de la catégorie culturelle « Beans, dry » issue de la base de données FAOSTAT (utilisée dans le chapitre 1) et qui regroupe les espèces *Phaseolus* spp. et *Vigna* spp. (à l'exception de *Vigna unguiculata*) ne représente que la variabilité interannuelle du rendement du haricot commun (*Phaseolus vulgaris*) en Europe. Au contraire, nous faisons l'hypothèse que la variabilité interannuelle du rendement de la catégorie culturelle « Beans, dry » représente la variabilité interannuelle des rendements des espèces *Phaseolus* spp. et *Vigna* spp. (à l'exception de *Vigna unguiculata*) en Amérique. Dans la seconde partie du tableau, les ratios de différents critères de production sont issus du chapitre 3, en Europe. Trois couleurs ont été utilisées : la couleur orange indique une performance moins bonne, la couleur jaune indique une performance égale, et la couleur verte indique une performance meilleure. Les symboles « + », « 0 » et « - » indiquent que la performance de l'espèce de légumineuses à graines est soit significativement meilleure, significativement égale ou significativement moins bonne que celle du pois protéagineux (*Pisum sativum*) pour un critère de production donné, respectivement ; et cette performance a été estimée à partir de données issues d'un nombre d'articles strictement supérieur à deux. La ligne pour le pois protéagineux (*Pisum sativum*) est hachurée puisque cette espèce est définie comme référence dans les comparaisons directes de la méta-analyse du chapitre 3. Les symboles « (+) », « (0) » et « (-) » indiquent que la performance de l'espèce de légumineuses à graines a été estimée à partir de données issues d'un nombre d'articles inférieur ou égal à deux. Dans la troisième partie du tableau, les effets des espèces de légumineuse à graines sur les rendements des céréales suivantes sont issus du chapitre 4, à l'échelle mondiale. Les symboles « + », « 0 » et « - » indiquent que l'effet de l'espèce de légumineuses à graines est soit significativement meilleur, significativement non différent (i.e., nul), ou significativement moins bon que l'effet d'une céréale sur les rendements des céréales suivantes.  $N_0$  indique la dose de fertilisation azotée ( $\text{kg N ha}^{-1}$ ) appliquée sur les céréales suivantes, et au-dessus de laquelle l'effet de l'espèce de légumineuses à graines sur les rendements des céréales suivantes devient négligeable. Deux couleurs ont été utilisées : la couleur orange indique que l'effet de l'espèce de légumineuses à graines sur les rendements des céréales suivantes est significativement positif quand la dose de fertilisation azotée des céréales est considérée comme faible (i.e., dans ce cas,  $0 < N_0 < 130 \text{ kg N ha}^{-1}$ ), la couleur jaune indique que l'effet de l'espèce de légumineuses à graines sur les rendements des céréales suivantes est significativement positif quand la dose de fertilisation azotée des céréales est considérée comme intermédiaire (i.e., dans ce cas,  $130 < N_0 < 160 \text{ kg N ha}^{-1}$ ), et la couleur verte indique que l'effet de l'espèce de légumineuses à graines sur les rendements des céréales suivantes est significativement positif quand la dose de fertilisation azotée des céréales est considérée comme élevée (i.e., dans ce cas,  $160 < N_0 < 200 \text{ kg N ha}^{-1}$ ). De la même manière que dans la seconde partie du tableau, les symboles « (+) », « (0) » et « (-) » indiquent l'incertitude associée à partir du nombre d'articles dont les données sont collectées. L'indice « a » indique que les données sont issues de la base de données FAOSTAT. L'indice « b » indique que les données sont issues d'expérimentations agronomiques collectées dans la base de données présentée dans le chapitre 2. L'indice « c » indique que les données sont issues des tables nutritionnelles de la base de données FEEDIPEDIA.



Les niveaux de production de 22 espèces de légumineuses à graines ont été comparés à celui du pois protéagineux (*Pisum sativum*, définie comme espèce de référence) en Europe, Amérique du Nord et Océanie. Sept critères de production ont été analysés : (1) le rendement en grains, (2) la biomasse aérienne totale, (3) la quantité d'azote et (4) la quantité de protéines et (5) la quantité d'énergie brute dans les graines, (6) l'utilisation en eau, et (7) la variabilité interannuelle du rendement en grains. Les données expérimentales de la base de données ont été analysées pour estimer les critères (1) à (6) alors que les données statistiques de la base FAOSTAT ont été analysées pour le critère (7). Tout d'abord, nous nous sommes demandés si des espèces de légumineuses à graines cultivées en Europe étaient caractérisées par des niveaux de production plus faibles, similaires, ou plus élevés, que celui du pois protéagineux (*Pisum sativum*). Nous avons montré que le soja (*Glycine max*), le lupin à feuilles étroites (*Lupinus angustifolius*) et la fêverole (*Vicia faba*) peuvent être des espèces intéressantes en Europe, alternatives au pois protéagineux (*Pisum sativum*). Puis, nous nous sommes demandés si des légumineuses à graines caractérisées par des niveaux de productions élevés par rapport à celui du pois protéagineux (*Pisum sativum*) en Amérique du Nord et Océanie pouvaient être des espèces candidates à inclure potentiellement dans des expérimentations en Europe. En Amérique du Nord, nous avons montré que la plupart des légumineuses à graines – à quelques exceptions spécifiques – ne sont pas plus productives que le pois protéagineux (*Pisum sativum*). En Océanie, nous avons montré que plusieurs espèces de gesses (*Lathyrus* spp.), de lupins (*Lupinus* spp.) et de vesces (*Vicia* spp. excepté *Vicia faba*) sont plus productives que le pois protéagineux (*Pisum sativum*). La fêverole (*Vicia faba*) est l'espèce qui présente un des meilleurs niveaux de production par rapport à celui du pois protéagineux (*Pisum sativum*) dans cette région. Le nombre d'espèces de légumineuses comparées au pois protéagineux (*Pisum sativum*) est trois fois supérieur à celui observé en Europe et Amérique du Nord. Dans les trois régions, les conclusions de la méta-analyse ne sont pas sensibles aux hypothèses des différents modèles testés. La pondération des données par leurs précisions permet d'obtenir des intervalles de confiance plus faibles. Les distributions des résidus sont symétriques pour l'ensemble des critères de production pour toutes les espèces dans toutes les régions.

Les résultats de cette première méta-analyse peuvent servir de support intéressant, notamment pour la prise de décision face à l'intérêt croissant du soja (*Glycine max*) dans l'agriculture européenne. En 2014, les surfaces de soja (*Glycine max*) représentaient plus de 25% et 50% de la surface récoltée de légumineuses à graines en France et en Europe, respectivement (FAOSTAT, 2016). La surface de soja (*Glycine max*) a fortement augmenté en Allemagne ces dernières années (Recknagel, 2015), simultanément à l'Italie et la Roumanie où sa production est historiquement développée (FAOSTAT, 2016). D'après le chapitre 1, la variabilité interannuelle du rendement en grains du soja (*Glycine max*) est la plus faible en Europe de l'Est. Ce résultat pourrait expliquer l'expansion privilégiée de la surface de cette espèce dans la région du Danube (Krön et Bittner, 2015 ; Recknagel, 2015). A l'échelle européenne, plusieurs raisons peuvent expliquer l'intérêt récent du soja (*Glycine max*). Une première

serait la forte augmentation du prix du tourteau de soja (*Glycine max*) ainsi que la prime d'approvisionnement pour du soja (*Glycine max*) non génétiquement modifié (Jouffret et al., 2015 ; Le Cadre et al., 2015 ; Quinsac et al., 2015). Une seconde raison serait le verdissement relatif de la nouvelle Politique Agricole Commune depuis 2015 qui se traduit par la mise place de mesures incitatives à cultiver des légumineuses (Bues et al., 2013 ; Preissel et al., 2015). Bien qu'économiquement faible, l'augmentation de la consommation de soja (*Glycine max*) en alimentation humaine a probablement favorisé l'intérêt de cette espèce en Europe en multipliant le développement de filières locales de production comme en Allemagne (Krön et Bittner, 2015 ; Recknagel, 2015), et en France (Jouffret et al., 2015 ; Le Cadre et al., 2015 ; Maury et al., 2015 ; Quinsac et al., 2015). Dans l'ensemble, le soja (*Glycine max*) reste actuellement la légumineuse à graines la plus utilisée dans la ration alimentaire des monogastriques en Europe (European Feed Manufacturers' Federation, 2016 ; Martin, 2014, 2015). Son équilibre en acides aminés et sa forte teneur en protéines lui confèrent en effet une forte polyvalence en alimentation animale (Dalgaard et al., 2008 ; Hartman et al., 2011). Les graines de soja (*Glycine max*) ne peuvent pas être consommées crues à cause de leurs fortes teneurs en facteurs antitrypsiques. Pour cette raison, elles sont généralement triturées pour en extraire de l'huile, et obtenir un tourteau – en tant que coproduit – riche en protéines (i.e., près de 50% de la matière sèche). Actuellement, les besoins protéiques des animaux monogastriques en Europe sont couverts par l'utilisation de tourteaux de soja (*Glycine max*), dont une grande partie est importée d'Amérique (Bues et al., 2013 ; Martin, 2014, 2015 ; Schreuder et De Visser, 2014). Entre 1961 et 2011, l'Europe a en effet importé 63% de sa demande domestique en légumineuses à graines (FAO, 2014). L'alimentation des productions animales européennes est ainsi progressivement devenue dépendante de ces importations (Bues et al., 2013 ; De Visser et al., 2014 ; Hache, 2015 ; Martin, 2014, 2015) ; l'autonomie protéique est un enjeu d'actualité en Europe.

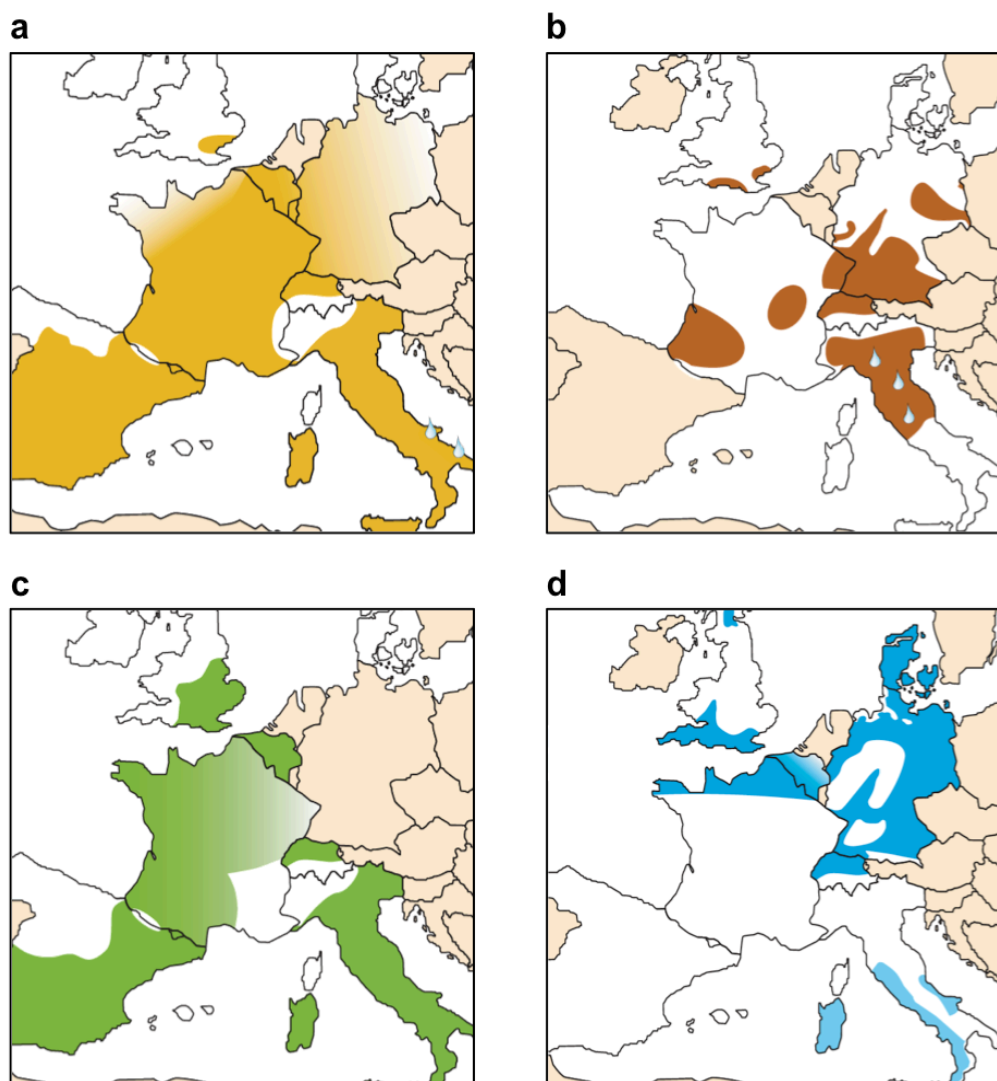
Pour répondre à cet enjeu, plusieurs études suggèrent d'augmenter les surfaces de légumineuses à graines dans cette région (Bues et al., 2013 ; Voisin et al., 2014 ; Zander et al., 2016), permettant d'augmenter la production de protéines issues de légumineuses en Europe, et de diminuer partiellement les importations de soja provenant d'Amérique. Nous pouvons faire l'hypothèse que la récente augmentation de la surface européenne de soja (*Glycine max*) s'inscrit au moins en partie dans cet enjeu. Il est probable qu'une augmentation rapide de la surface de soja (*Glycine max*) favorise des phénomènes de substitution entre les surfaces récoltées de différentes espèces de légumineuses à graines en Europe. J'ai évoqué ce point à Moritz Reckling, agronome au Leibniz Centre for Agricultural Landscape and Land Use Research au Nord-Ouest de l'Allemagne. Dans cette région, la surface de soja (*Glycine max*) a fortement augmenté depuis trois ans (FAOSTAT, 2016). D'après Moritz Reckling, la récente mise en place des surfaces d'intérêt écologique aurait incité les agriculteurs à cultiver cette espèce sur des sols de bonne qualité. Il a remarqué que la culture du soja (*Glycine max*) concurrençait progressivement la culture du lupin à feuilles étroites (*Lupinus*

*angustifolius*), historiquement récolté dans cette région. De plus, il a souligné que des essais variétaux de soja (*Glycine max*) se multipliaient en Allemagne afin d'identifier des variétés adaptées aux faibles températures et à l'absence de souches d'inoculum adéquates (Krön et Bittner, 2015 ; Recknagel, 2015 ; Zimmer et al., 2016b), et répondre à une demande croissante des agriculteurs pour obtenir des références variétales. Depuis 50 ans, le pois protéagineux (*Pisum sativum*) est la légumineuse la plus cultivée, en moyenne annuelle, en Europe (FAOSTAT, 2016). Nous pouvons faire l'hypothèse que l'augmentation de la surface de soja (*Glycine max*) peut se faire actuellement aux dépens de celle de pois protéagineux (*Pisum sativum*) par phénomène de substitution sous contrainte d'une surface de légumineuses constante en Europe. L'intérêt récent du soja (*Glycine max*) soulève en fait la question du choix optimal de l'espèce ou des espèces de légumineuses à produire préférentiellement en Europe. Les éléments contextuels semblent mettre en évidence que le choix de production actuel se spécialiserait sur le soja (*Glycine max*). Cependant, il serait intéressant de connaître si la diversification plurispécifique de la production de légumineuses en Europe représenterait une stratégie moins risquée et plus viable qu'une spécialisation monospécifique de la production.

Les résultats de la méta-analyse peuvent apporter des éléments de réponse à cette question. Nous avons déterminé que la fève (*Vicia faba*) présente des performances productives similaires à celles du pois protéagineux (*Pisum sativum*) pour tous les critères étudiés en Europe. Flores et al. (2012, 2013) ont montré que la fève (*Vicia faba*) était relativement bien adaptée à la diversité des conditions pédoclimatiques européennes. Le lupin à feuilles étroites (*Lupinus angustifolius*) et le lupin jaune (*Lupinus luteus*) présentent également des performances productives relativement intéressantes par rapport à celles du pois protéagineux (*Pisum sativum*). Lucas et al. (2015) ont souligné la nécessité des innovations en sélection variétale et des traitements technologiques afin de mieux valoriser ces espèces en tant que matières premières en Europe. Contrairement à celles du soja (*Glycine max*), les graines de fève (*Vicia faba*) et certaines espèces de lupins (*Lupinus* spp.) peuvent être utilisées crues (Guillamón et al., 2008), et représenter une source de protéines alternative au pois protéagineux (*Pisum sativum*) en alimentation animale (Crépon et al., 2010 ; Jezierny et al., 2010 ; Koivunen et al., 2016 ; Multari et al., 2015).

Nous avons estimé qu'une substitution de la surface européenne de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*), du lupin à feuilles étroites (*Lupinus angustifolius*) et de la fève (*Vicia faba*) augmenterait la production de protéines issues de légumineuses à graines cultivées en Europe. D'après les résultats de la méta-analyse, ces trois espèces sont les légumineuses qui présentent les quantités de protéines dans les graines les plus élevées en Europe. Le gain moyen de production de protéines serait de 28% pour une substitution de 25% de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) alors qu'il serait de 3 et 4% dans les cas de la fève (*Vicia faba*) et du lupin à feuilles étroites (*Lupinus angustifolius*), respectivement. Si une substitution de la surface

de pois protéagineux (*Pisum sativum*) par une autre espèce de légumineuse à graines était donc envisagée, nos résultats tendraient à corroborer le regain d'intérêt actuel du soja (*Glycine max*) récolté en Europe. Cependant, le gain de production issue d'une substitution de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) est plus incertain (i.e., entre 10 et 56% d'après l'intervalle de confiance à 95%) que dans les cas de la fève (*Vicia faba*, entre -2 et 8%) et du lupin à feuilles étroites (*Lupinus angustifolius*, entre -2 à 13%). Cette incertitude est liée au faible nombre de données expérimentales disponibles, comparant directement le pois protéagineux (*Pisum sativum*) au soja (*Glycine max*) dans notre base de données. Par ailleurs, une forte substitution de la surface récoltée de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) semblerait être relativement peu réaliste, d'un point de vue agronomique. Ces deux espèces ne partageraient potentiellement qu'une faible fraction de leurs aires de culture respectives en Europe (Figure 3). Au contraire, la substitution du pois protéagineux (*Pisum sativum*) par de la fève (*Vicia faba*) – et dans une moindre mesure, par du lupin à feuilles étroites (*Lupinus angustifolius*) – semblerait être agronomiquement plus réaliste que dans le cas du soja (*Glycine max*). La fève (*Vicia faba*) partagerait potentiellement une fraction élevée de son aire de culture avec celle du pois protéagineux (*Pisum sativum*) (Figure 3) ; l'aire de culture du lupin à feuilles étroites (*Lupinus angustifolius*) partagerait également une fraction de son aire de culture avec celle du pois protéagineux (*Pisum sativum*). Cependant, il est fortement probable qu'il existe d'autres types de substitution de surface des légumineuses en Europe. La surface de légumineuses à graines autres que le pois protéagineux (*Pisum sativum*) peut être substituée (e.g., la substitution de la surface de lupin à feuilles étroites (*Lupinus angustifolius*) par du soja (*Glycine max*) au Nord-Ouest de l'Allemagne). Plusieurs études soulignent également que la substitution d'une fraction de la surface céréalière par des légumineuses à graines (toutes espèces confondues) en Europe reste le scénario de substitution de surface le plus envisagé (Bues et al., 2013 ; De Visser et al., 2014 ; LMC International, 2009a, 2009b ; Preissel et al., 2015 ; Reckling et al., 2016a ; Zander et al., 2016).



**Figure 3.** Aires de culture pour (a) le pois protéagineux (*Pisum sativum*, culture d'hiver), (b) le soja (*Glycine max*), (c) la fève (Vicia faba, culture d'hiver), et (d) le lupin à feuilles étroites (*Lupinus angustifolius*, culture de printemps) en Europe. Les aires de culture ont été définies à dire d'experts dans le cadre de l'action concertée européenne GL-Pro (2003-2006).

La méta-analyse a montré que le pois chiche (*Cicer arietinum*) et la lentille (*Lens culinaris*) ont des performances productives moindres que celle du pois protéagineux (*Pisum sativum*) en Europe. Bien que ces deux espèces puissent être utilisées dans l'alimentation animale, elles sont généralement destinées à l'alimentation humaine. Leurs prix de vente sont souvent plus rémunérateurs que celui du pois protéagineux (*Pisum sativum*) (Schneider et al., 2015). Le pois chiche (*Cicer arietinum*) et la lentille (*Lens culinaris*) peuvent constituer des opportunités de marché intéressantes, en termes de retours économiques et de valorisation par l'intermédiaire d'appellations d'origine contrôlée, permettant ainsi de compenser leurs faibles niveaux de production (Erskine et al., 2011 ; Jukanti et al., 2012 ; Magrini et al., 2016 ; Voisin et al., 2014).

Dans l'ensemble, les classements des espèces en Amérique du Nord et Océanie confirment les principaux résultats en Europe. En Amérique du Nord, les faibles productions des espèces de légumineuses par rapport à celui du pois protéagineux (*Pisum sativum*) peuvent être expliqués par la situation géographique des expérimentations nord-américaines incluses dans la méta-analyse, principalement situées au Nord des Grandes Plaines. Dans cette région, Miller et al. (2002) ont en effet indiqué que le rendement du pois protéagineux (*Pisum sativum*) était supérieur à celui des autres légumineuses à graines. Les auteurs ont montré par exemple que le rendement du soja (*Glycine max*) était plus faible de 47% à celui du pois protéagineux (*Pisum sativum*). Cette diminution de rendement est similaire avec celle mesurée dans la méta-analyse (i.e., 41 et 53% sans et avec pondération des données, respectivement). Malgré un niveau moins élevé, le rendement du soja (*Glycine max*) est 62% moins variable que celui du pois protéagineux (*Pisum sativum*) dans cette région. Ce dernier résultat n'a pu être obtenu qu'en utilisant la base de données FAOSTAT, soulignant ainsi la complémentarité entre les données statistiques et les données expérimentales pour quantifier la production des espèces de légumineuses à large échelle.

En Océanie, la fêverole (*Vicia faba*) présente une production significativement supérieure à celle du pois protéagineux (*Pisum sativum*). Siddique et al. (2013) ont indiqué que cette espèce était adaptée à une plus grande diversité de conditions environnementales que le pois protéagineux (*Pisum sativum*) en Australie. L'ubiquité de la fêverole (*Vicia faba*) dans cette région pourrait ainsi expliquer ses forts niveaux de production.

La gesse blanche (*Lathyrus sativus*) présente également des performances productives similaires à celles du pois protéagineux (*Pisum sativum*) en Amérique du Nord et Océanie. Cependant, la production estimée pour cette espèce est associée à une forte incertitude en raison d'un faible nombre de comparaisons directes au pois protéagineux (*Pisum sativum*). Calderón et al. (2012) et Vaz Patto et Rubiales (2014) ont souligné l'importance de poursuivre des efforts en amélioration génétique et en agronomie afin de mieux évaluer la production de cette espèce. Cependant, les graines de la gesse blanche (*Lathyrus sativus*) – ainsi que celles d'autres espèces de gesses (*Lathyrus* spp.) et vesces (*Vicia* spp. excepté *Vicia faba*) – contiennent des facteurs nutritionnels indésirables (Gilani et al., 2005 ; Hanbury et al., 2000 ; Vaz Patto et Rubiales, 2014), voire neurotoxiques (Yan et al., 2006). Ces facteurs peuvent réduire la valeur azotée des protéines des graines de ces légumineuses dans l'alimentation des animaux (Jeziorny et al., 2010 ; Koivunen et al., 2016 ; Longstaff et al., 1991). Bien que des traitements technologiques permettent de les inactiver (D'Mello, 1992 ; Singh, 1988), le coût d'application de telles opérations reste, souvent, déterminant (Gilani et al., 2005).

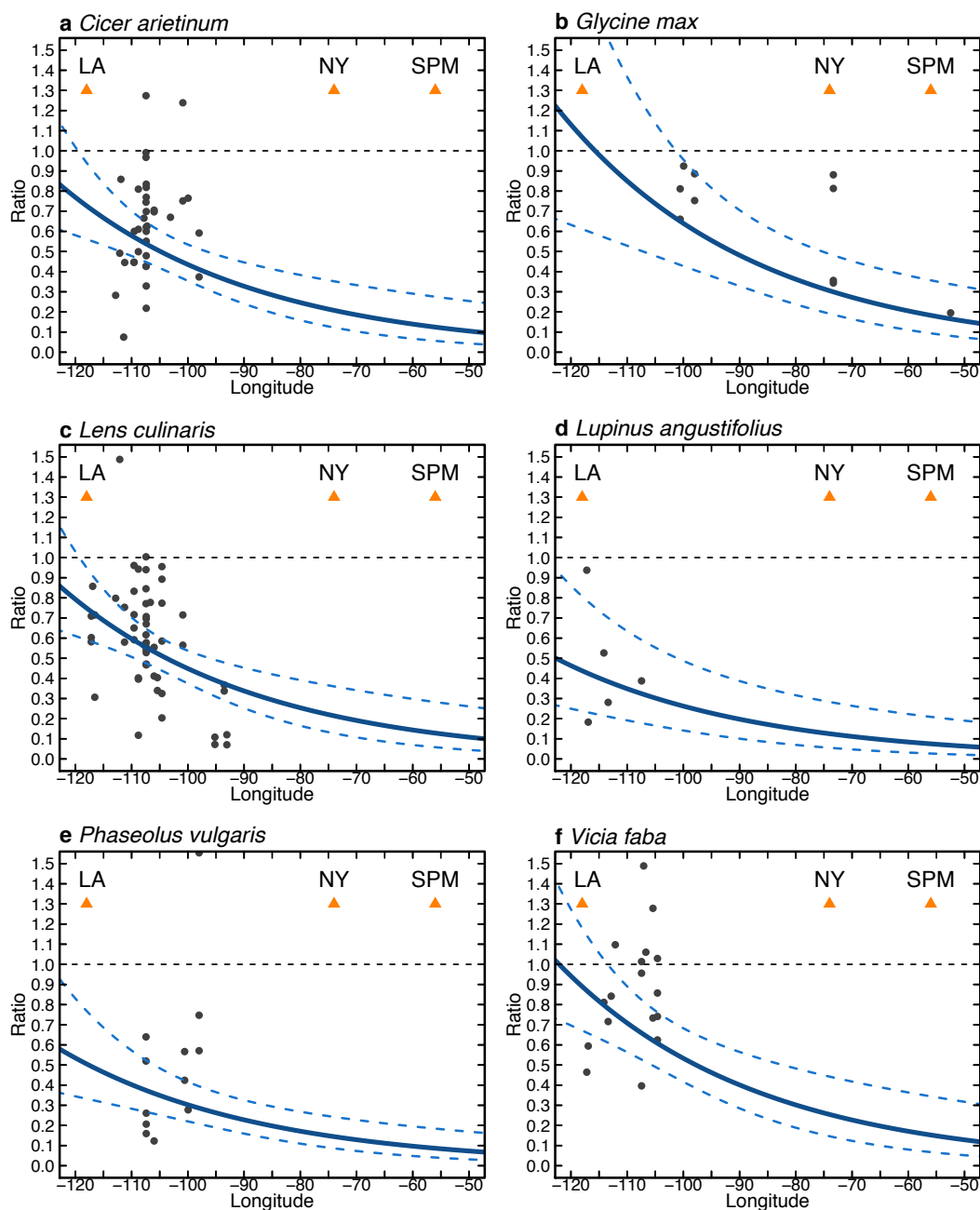
Le nombre observé d'espèces de légumineuses à graines comparées au pois protéagineux (*Pisum sativum*) en Océanie est trois fois supérieur à celui observé en Europe et Amérique du Nord. Dans la méta-analyse, la majorité des expérimentations en Océanie sont situées dans le Sud-Ouest de l'Australie. Harries et al. (2015) ont souligné qu'un grand nombre d'espèces de légumineuses avaient été comparées entre elles dans cette région afin de développer leurs productions commerciales. J'ai évoqué ce sujet avec Kadambot H. M. Siddique, professeur en agronomie à l'université d'Australie de l'Ouest. D'après lui, le fort soutien économique du Grains Research & Development Corporation, associé au gouvernement et à des industries de transformation développées vers l'export, avait eu un rôle important dans l'expansion de la production des légumineuses à graines en Australie.

La méta-analyse a montré la nécessité de réaliser d'autres expérimentations agronomiques comparant des légumineuses entre elles en Europe. Ces futures expérimentations peuvent inclure certaines espèces cultivées actuellement, mais pour lesquelles nos estimations sont associées d'une forte incertitude en raison du faible nombre de données expérimentales incluant du soja (*Glycine max*), du lupin à feuilles étroites (*Lupinus angustifolius*), et lupin jaune (*Lupinus luteus*). Étant donné l'intérêt croissant du soja (*Glycine max*) dans l'agriculture européenne, réaliser des futures expérimentations incluant cette espèce permettrait de la considérer comme une nouvelle espèce de référence dans des comparaisons directes de performances avec celles d'autres légumineuses. Ces futures expérimentations peuvent également inclure des espèces non comparées au pois protéagineux (*Pisum sativum*) en Europe mais comparées dans d'autres régions du monde – comme en Océanie – et caractérisées par des niveaux élevés de production. Ces espèces regroupent des espèces de gesses (*Lathyrus* spp.): *Lathyrus cicera*, *Lathyrus clymenum*, *Lathyrus ochrus*, et *Lathyrus sativus*; des espèces de lupins (*Lupinus* spp.): *Lupinus atlanticus*, et *Lupinus pilosus*; des espèces de vesces (*Vicia* spp. excepté *Vicia faba*): *Vicia articulata*, *Vicia narbonensis*, et *Vicia sativa*. Bien que certaines de ces espèces présentent des facteurs nutritionnels indésirables dans leurs graines, il serait néanmoins intéressant de les étudier dans de futures expérimentations en Europe afin de mieux déterminer leurs productions relatives au travers d'un gradient de conditions pédoclimatiques.

Pour préciser les résultats de la méta-analyse, nous avons testé l'effet des coordonnées géographiques des sites expérimentaux sur les ratios des rendements de différentes espèces de légumineuses à celui du pois protéagineux (*Pisum sativum*). Nous n'avons sélectionné que les espèces pour lesquelles le nombre de données était suffisant (i.e., les données sont issues de plus de deux articles). En Europe, nous n'avons pas mis en évidence d'effet significatif des coordonnées géographiques à cause du manque de données expérimentales. Au contraire, nous avons identifié un effet significativement négatif ( $p < 0.05$ ) de la longitude sur les rendements de six espèces comparées au pois protéagineux (*Pisum sativum*) en Amérique du Nord (Figure 4). Plus les expérimentations sont situées vers l'Ouest (i.e., vers Los Angeles), plus les espèces de légumineuses présentent des rendements élevés par

rapport à celui du pois protéagineux (*Pisum sativum*). Plus les expérimentations sont situées vers l'Est (vers New York, et Saint-Pierre-et-Miquelon), plus le rendement du pois protéagineux (*Pisum sativum*) augmente, et ceux des autres espèces de légumineuses tendent à diminuer. Outre la longitude, les effets d'autres covariables environnementales et agronomiques sur la production des légumineuses à graines, cultivées en particulier dans le Nord des Grandes Plaines, ont été synthétisés dans des revues telles que celle de Miller et al. (2002). Les auteurs ont, par exemple, indiqué les effets de la température et de la pluviométrie des sites expérimentaux sur les rendements des légumineuses à graines qui y étaient cultivées. En utilisant notre base de données, les effets de ces covariables environnementales pourraient être testés dans les modèles statistiques. Cette analyse permettrait d'identifier des facteurs environnementaux explicatifs des différences de production entre les espèces. Les effets de covariables agronomiques pourraient également être testés. A ce titre, la présence ou l'absence de travail du sol (e.g., le semis direct) serait une covariable intéressante car le nombre de données disponibles est élevé pour des expérimentations conduites en Amérique du Nord, et extraites dans la base de données. L'effet de cette technique est également discuté dans plusieurs synthèses sur les systèmes de culture dans les Grandes Plaines (Cutforth et al., 2007 ; Miller et al., 2001, 2002, 2006 ; Tanaka et al., 2002, 2007 ; Zentner et al., 2002). Il serait alors intéressant de confronter les conclusions des deux analyses. Il est toutefois nécessaire de rappeler que l'introduction de covariables – qu'elles soient environnementales ou agronomiques – dans les méta-analyses peut être une approche risquée, conduisant à de fausses découvertes dues à l'identification d'effets significatifs, qui ne peuvent pas être interprétés comme des relations causales, en raison des confusions d'effets.





**Figure 4.** Effet de la longitude sur les ratios de rendement en grains de six espèces de légumineuses à celui du pois protéagineux (*Pisum sativum*) en Amérique du Nord ( $n$  est le nombre de ratios de rendement en grains) : (a) *Cicer arietinum* (pois chiche,  $n=37$  issus de 14 articles), (b) *Glycine max* (soja,  $n=11$  issus de 4 articles), (c) *Lens culinaris* (lentille,  $n=54$  issus de 16 articles), (d) *Lupinus angustifolius* (lupin à feuilles étroites,  $n=5$  issus de 3 articles), (e) *Phaseolus vulgaris* (haricot commun,  $n=12$  issus de 4 articles), (f) *Vicia faba* (fêverole,  $n=17$  issus de 6 articles). Les triangles oranges avec les acronymes « LA », « NY » et « SPM » indiquent les longitudes de Los Angeles, New York, et Saint-Pierre-et-Miquelon, respectivement. Les points indiquent les ratios observés de rendement. La courbe bleue continue indique la tendance de la régression appliquée aux ratios des différentes espèces de légumineuses à graines. Un modèle linéaire à effets aléatoires a été utilisé avec l'espèce et la longitude en effets additifs fixes, et la combinaison site expérimental\*saïson culturale comme effet aléatoire. Les données ont été pondérées par leurs précisions. Les courbes bleues pointillées indiquent les intervalles de confiance à 95%, estimés à partir du modèle linéaire à effets aléatoires.

#### 4. Comparer l'effet des légumineuses à graines sur les rendements des céréales suivantes

L'effet des légumineuses à graines sur les rendements des céréales suivantes a été quantifié en réalisant une seconde méta-analyse à partir des données expérimentales collectées dans la base de données du chapitre 2. Les rendements des céréales cultivées après différentes espèces de légumineuses ont été systématiquement comparés aux rendements des céréales cultivées après des céréales, sur les mêmes sites expérimentaux durant les mêmes saisons culturales et les mêmes techniques culturales. Les rendements des céréales après des céréales ont été définis comme des rendements de référence. Ces comparaisons ont été conduites à l'échelle mondiale pour 16 espèces de légumineuses à graines en tant que cultures précédentes. L'effet des cultures précédentes d'oléagineux sur les rendements des céréales suivantes a également été étudié. La dose de fertilisation en azote minéral des céréales suivantes a été analysée en covariable dans les modèles statistiques. D'abord, nous nous sommes demandés si l'effet des légumineuses à graines sur les rendements des céréales suivantes était significativement positif et dépendait de l'espèce de légumineuses. Nous avons montré que les cultures précédentes de 13 espèces de légumineuses augmentaient significativement, de 29% en moyenne, les rendements des céréales suivantes. La pondération des données par leurs précisions change peu les classements des espèces de légumineuses en tant que cultures précédentes. Les espèces de légumineuses à graines qui permettent les plus forts gains de rendement des céréales suivantes sont – par ordre croissant de gain – le pois protéagineux (*Pisum sativum*), la lentille (*Lens culinaris*), la fève (*Vicia faba*), et les lupins (*Lupinus* spp.). Ensuite, nous nous sommes demandés si l'effet des légumineuses à graines sur les rendements des céréales suivantes dépendait de la dose de fertilisation en azote minéral qui leur était appliquée. Nous avons montré que l'effet positif résultant des cultures précédentes de légumineuses décroissait significativement en fonction de la dose de fertilisation. En fait, les effets positifs des oléagineux et des légumineuses à graines devenaient négligeables quand les céréales suivantes étaient fertilisées en moyenne au-dessus de 150 kg N ha<sup>-1</sup>. Cette dose représentait un seuil, variant toutefois fortement entre les espèces de légumineuses. Enfin, nous nous sommes demandés si l'effet des légumineuses sur les rendements des céréales permettait de compenser une diminution relative attendue de production céréalière, résultant de l'introduction d'une légumineuse dans une monoculture d'une céréale. Nous avons montré que la diminution relative attendue de production céréalière était partiellement compensée par l'effet positif de la légumineuse quand la dose de fertilisation azotée appliquée sur la céréale était inférieure à 150 kg N ha<sup>-1</sup>. Dans cette situation, la diminution relative attendue de production céréalière était inférieure à la proportion de la légumineuse dans la monoculture d'une céréale. Quand la dose de fertilisation azotée sur la céréale était supérieure à 150 kg N ha<sup>-1</sup>, l'effet de compensation n'était plus significatif. Dans cette situation, la diminution relative attendue de production céréalière était similaire à la proportion de la légumineuse dans la monoculture d'une céréale.

L'augmentation moyenne significative des rendements des céréales, résultant des cultures précédentes d'oléagineux et de légumineuses à graines, a été estimée à 14 et 29%, respectivement. Ces valeurs sont relativement cohérentes avec des études qui avaient également quantifié l'effet d'espèces d'oléagineux et de légumineuses à graines sur les rendements des céréales qui les suivent dans la séquence culturale, principalement le blé (*Triticum aestivum*) (Angus et al., 2015 ; Kirkegaard et al., 2008 ; Preissel et al., 2015 ; Seymour et al., 2012). Ce gain de rendement varie toutefois entre les espèces de légumineuses. Nous avons estimé que les rendements moyens des céréales suivantes étaient plus élevés de 21% pour une culture précédente de pois chiche (*Cicer arietinum*), de 35% pour le pois protéagineux (*Pisum sativum*), et de 40% pour le lupin à feuilles étroites (*Lupinus angustifolius*). Ces différences observées entre les espèces de légumineuses à graines peuvent être dues aux différences spécifiques d'efficacité à fixer du N<sub>2</sub> atmosphérique (Hardarson et Atkins, 2003). Certaines légumineuses ont des systèmes racinaires plus nodulés que ceux d'autres espèces de légumineuses, permettant de fixer potentiellement plus de N<sub>2</sub>, et de générer plus de résidus aériens riches en azote (Angus et al., 2015 ; Peoples et al., 2009 ; Seymour et al., 2012). Ces résidus aériens peuvent retourner au sol, et être minéralisés. En se basant sur des modèles de culture en Australie, Evans et al. (2001) ont calculé que la fourniture moyenne potentielle en azote (incluant les résidus souterrains riches en azote) était de 18 kg N ha<sup>-1</sup> pour une culture de pois chiche (*Cicer arietinum*), de 44 kg N ha<sup>-1</sup> le pois protéagineux (*Pisum sativum*), et de 88 kg N ha<sup>-1</sup> pour le lupin à feuilles étroites (*Lupinus angustifolius*). Pour ces trois espèces, ces différences de fourniture potentielle pourraient expliquer les différences d'augmentation des rendements des céréales estimées dans la méta-analyse. Par ailleurs, ces fournitures potentielles en azote peuvent être recalculées en utilisant certaines variables de la base de données du chapitre 2 (e.g., le pourcentage en azote fixé et les quantités d'azote dans le sol à la récolte (au semis) des légumineuses (des céréales et des oléagineux suivants). Ces données sont principalement renseignées en Océanie où le plus grand nombre d'expérimentations ont été réalisées pour comparer les fournitures potentielles en azote de différentes espèces de légumineuses. Par ailleurs, ces fournitures potentielles sont calculables à partir de différentes équations. Nous avons commencé à analyser certaines variables d'entrée, mais le nombre de données disponibles dans la base de données reste limitant. Plusieurs synthèses quantitatives australiennes ont bien documenté l'analyse de la fourniture potentielle en azote issue des cultures de légumineuses à graines ainsi que leurs effets sur les cultures non-légumineuses suivantes. Ces synthèses ont proposé des éléments d'analyse intéressants à considérer si la quantification de la fourniture potentielle en azote des légumineuses est investie à large échelle. Dans le futur, il serait également nécessaire que les variables d'entrée de la fourniture potentielle en azote puissent être plus amplement renseignées en incluant de nouveaux articles dans la base de données.

A l'échelle mondiale, la dose de fertilisation azotée appliquée sur les céréales est souvent supérieure à 150 kg N ha<sup>-1</sup> dans les systèmes céréaliers conventionnels intensifs (Mueller et al., 2014), principalement en Europe (Lassaletta et al., 2014). En France, la dose moyenne de fertilisation en azote minéral appliquée sur l'orge (*Hordeum vulgare*), le blé (*Triticum aestivum*), le blé dur (*Triticum durum*), et le maïs (*Zea mays*), est estimée à 160 kg N ha<sup>-1</sup> à partir de données statistiques nationales en 2011 (AGRESTE, 2014). En Europe, la dose moyenne de fertilisation azotée appliquée aux systèmes céréaliers à bas niveau d'intrants est de 126 kg N ha<sup>-1</sup> (Hossard et al., 2016), inférieure à la dose seuil de 150 kg N ha<sup>-1</sup> au-dessus de laquelle nous avons estimé que l'effet positif moyen des légumineuses devenait significativement négligeable. Pour affiner les résultats de la méta-analyse, nous avons quantifié l'effet de certaines espèces de légumineuses à graines (cultivées actuellement en Europe) sur les rendements des céréales suivantes fertilisées à différentes doses d'azote minéral (Tableau 1). Nous avons sélectionné trois doses de fertilisation azotée : (1) la dose moyenne de fertilisation azotée appliquée actuellement en France sur les céréales (i.e., 160 kg N ha<sup>-1</sup>), (2) la dose moyenne de fertilisation azotée appliquée aux systèmes céréaliers à bas niveaux d'intrants en Europe (i.e., 126 kg N ha<sup>-1</sup>), et (3) la moitié de la dose moyenne de fertilisation azotée en France sur les céréales (i.e., 80 kg N ha<sup>-1</sup>). Si nous considérons la dose moyenne actuelle de 160 kg N ha<sup>-1</sup> appliquée sur les céréales en France, l'effet positif moyen des légumineuses à graines sur les rendements des céréales suivantes est inférieur à 10%. Le gain moyen de rendement est le plus élevé (8%) pour une culture précédente de fèverole (*Vicia faba*) alors que le gain est entre 5% et 6% pour la lentille (*Lens culinaris*) ou les lupins (*Lupinus* spp.), et de 1% pour le pois protéagineux (*Pisum sativum*). Pour cette dose de fertilisation azotée, les effets des cultures précédentes de pois chiche (*Cicer arietinum*), de soja (*Glycine max*), de haricot commun (*Phaseolus vulgaris*), et de vesce commune (*Vicia sativa*) sont négligeables sur les rendements des céréales suivantes. Si nous considérons la dose actuelle de fertilisation azotée de 126 kg N ha<sup>-1</sup> appliquée sur des céréales à bas niveaux d'intrants en Europe, l'effet positif moyen des légumineuses à graines sur les rendements des céréales suivantes est également faible. Cet effet est au maximum de 15% dans le cas d'une culture précédente de fèverole (*Vicia faba*), approximativement de 13% dans les cas de la lentille (*Lens culinaris*) et des lupins (*Lupinus* spp.), de 8% dans le cas du pois protéagineux (*Pisum sativum*), et de 2% dans le cas du soja (*Glycine max*). Il faudrait en fait réduire de 50% la dose moyenne de fertilisation azotée (i.e., 80 kg N ha<sup>-1</sup>) appliquée sur les céréales en France pour que le gain de rendement des céréales suivantes résultant des légumineuses à graines (en tant que cultures précédentes) soit supérieur de 15%. Le gain de rendement des céréales serait alors de 19% pour une culture précédente de pois protéagineux (*Pisum sativum*), et de 11% dans le cas d'une du soja (*Glycine max*). Les gains de rendement pour des cultures précédentes de pois chiche (*Cicer arietinum*), de soja (*Glycine max*), de lentille (*Lens culinaris*), de lupin à feuilles étroites (*Lupinus angustifolius*), et de pois protéagineux (*Pisum sativum*) ont été estimés à partir d'un nombre de ratios de rendement en grains supérieur à 50. Les gains de rendement pour des cultures précédentes de haricot commun (*Phaseolus vulgaris*) et de fèverole (*Vicia*

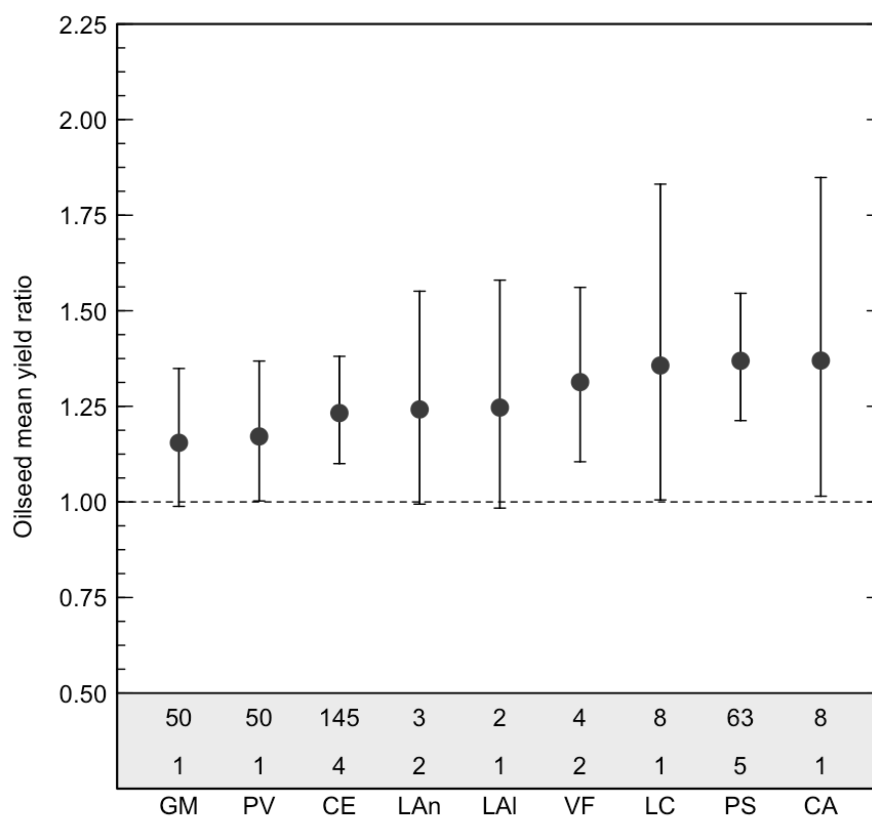
*faba*) ont été estimés à partir d'un nombre de ratios de rendement en grains compris entre 25 et 40. Les gains de rendement pour des cultures précédentes de lupin blanc (*Lupinus albus*) et de lupin jaune (*Lupinus luteus*) ont été estimés à partir d'un nombre de ratios de rendement en grains inférieur à 10.

Nom scientifique	$N_0$ (kg N ha <sup>-1</sup> )	Changement relatif (%) de rendement des céréales suivantes fertilisées en azote minéral (N)								
		160 kg N ha <sup>-1</sup>		126 kg N ha <sup>-1</sup>		80 kg N ha <sup>-1</sup>				
		Moyen	Intervalle de confiance <sup>a</sup>	Moyen	Intervalle de confiance	Moyen	Intervalle de confiance			
<i>Cicer arietinum</i>	124	-7	-15	+2	0	-8	+8	+9	+2	+17
<i>Glycine max</i>	135	-5	-13	+4	+2	-5	+9	+11	+5	+18
<i>Lens culinaris</i>	184	+5	-5	+16	+12	+3	+22	+22	+13	+33
<i>Lupinus albus</i>	185	+5	-13	+27	+12	-6	+34	+23	+3	+46
<i>Lupinus angustifolius</i>	190	+6	-3	+17	+13	+5	+23	+24	+15	+33
<i>Lupinus luteus</i>	190	+6	-27	+55	+13	-22	+64	+24	-14	+80
<i>Phaseolus vulgaris</i>	129	-6	-14	+3	0	-7	+9	+10	+2	+18
<i>Pisum sativum</i>	168	+1	-6	+10	+8	+2	+16	+19	+13	+25
<i>Vicia faba</i>	198	+8	-4	+21	+15	+4	+28	+26	+14	+39
<i>Vicia sativa</i>	115	-9	-29	+17	-2	-23	+24	+7	-16	+36

<sup>a</sup>Intervalle de confiance à 95%

**Tableau 1.** Changement relatif (%) de rendement en grains des céréales fertilisées en azote minéral (N) et cultivées après dix espèces de légumineuses. Les espèces de légumineuses sont les suivantes (*n* est le nombre de ratios de rendement en grains) : *Cicer arietinum* (pois chiche, *n*=84), *Glycine max* (soja, *n*=64), *Lens culinaris* (lentille, *n*=56), *Lupinus albus* (lupin blanc, *n*=9), *Lupinus angustifolius* (lupin à feuilles étroites, *n*=51), *Lupinus luteus* (lupin jaune, *n*=2), *Phaseolus vulgaris* (haricot commun, *n*=37), *Pisum sativum* (pois protéagineux, *n*=162), *Vicia faba* (féverole, *n*=28), *Vicia sativa* (vesce commune, *n*=4). Les changements relatifs moyens de rendement des céréales sont présentés. Les changements relatifs estimés pour les bornes inférieures et supérieures des intervalles de confiance à 95% sont présentés. Ces changements sont des gains relatifs (+), des pertes relatives (-), ou sans changement (0) de rendement par rapport aux rendements des céréales cultivées après des céréales. Les valeurs moyennes et les bornes des intervalles de confiance à 95% ont été estimées en utilisant le modèle statistique M31 présenté dans le chapitre 4.  $N_0$  est la dose de fertilisation azotée appliquée sur les céréales suivantes, et au-dessus de laquelle les ratios moyens de rendement sont inférieurs à zéro.  $N_0$  a été estimé en utilisant le modèle statistique M31. Trois doses de fertilisation azotée appliquées sur les céréales suivantes sont considérées : (1) 160 kg N ha<sup>-1</sup>, (2) 126 kg N ha<sup>-1</sup>, et (3) 80 kg N ha<sup>-1</sup>.

Les effets des cultures précédentes de légumineuses à graines et d'oléagineux sur les rendements des céréales ont été quantifiés dans le chapitre 4. Les rendements d'oléagineux qui suivent différentes légumineuses à graines ont également été inclus dans notre base de données. Nous avons analysé ces rendements de la même manière que ceux des céréales suivantes. Nous avons estimé que l'effet sur les rendements des oléagineux est significatif pour 6 des 9 espèces de non-oléagineux, et 5 des 8 espèces de légumineuses considérées en tant que cultures précédentes (Figure 5). Ces espèces (i.e., céréales et légumineuses à graines) augmentent significativement de 31% en moyenne les rendements des cultures suivantes d'oléagineux. L'effet du pois protéagineux (*Pisum sativum*) et de la fève (*Vicia faba*) sur les rendements des oléagineux suivants est fortement significatif (Figure 5). L'effet est également significatif pour d'autres légumineuses (e.g., le pois chiche (*Cicer arietinum*)). Cependant, l'incertitude associée est plus élevée en raison d'un nombre très faible d'articles à partir desquels les données expérimentales ont été extraites. La fève (*Vicia faba*) et le pois protéagineux (*Pisum sativum*) augmentent significativement de 31 et 37% les rendements des oléagineux suivants, soit 8 et 14% de plus, respectivement, que les céréales en cultures précédentes (Figure 5). Les ratios de rendement en grains des oléagineux ont été exprimés en fonction de la dose de fertilisation azotée, mais le nombre de données est trop faible pour étudier cette relation de manière robuste.



**Figure 5.** Ratios de rendement en grains moyens estimés d'oléagineux cultivés après des espèces de non-oléagineux (i.e., céréales ou légumineuses à graines). Les noms scientifiques des légumineuses à graines sont : CA : *Cicer arietinum*, GM : *Glycine max*, LC : *Lens culinaris*, LAI : *Lupinus albus*, LAn : *Lupinus angustifolius*, PV : *Phaseolus vulgaris*, PS : *Pisum sativum*, VF : *Vicia faba*. La catégorie pour les céréales (CE) regroupe différentes espèces de céréales cultivées en tant que cultures précédentes : *Avena sativa*, *Fagopyrum esculentum*, *Hordeum vulgare*, *Oryza sativa*, *Panicum miliaceum*, *Pennisetum glaucum*, *Sorghum bicolor*, *Triticum aestivum*, *Triticum durum*, et *Triticum sativum*. Les points indiquent les ratios de rendement en grains moyens estimés à partir du modèle statistique M3 présenté dans le chapitre 4. Les données sont pondérées par leurs précisions. Les barres d'erreur indiquent des intervalles de confiance à 95%. La ligne inférieure et la ligne supérieure dans le rectangle gris indiquent le nombre d'articles et le nombre de ratios de rendement en grains disponibles pour chaque espèce de non-oléagineux (i.e., céréales ou légumineuses à graines), respectivement.



L'effet des oléagineux et de différentes espèces de légumineuses a été estimé sur les rendements de dix espèces de céréales suivantes dans le chapitre 4. Les espèces de céréales étaient (1) l'avoine (*Avena sativa*, 1% des données de rendement des céréales suivantes dans notre jeu de données), (2) le sarrasin (*Fagopyrum esculentum*, 2%), (3) l'orge (*Hordeum vulgare*, 19%), (4) le riz (*Oryza sativa*, 3%), (5) le millet commun (*Panicum miliaceum*, 2%), (6) le sorgho commun (*Sorghum bicolor*, 6%), (7) le blé (*Triticum aestivum*, 56%), (8) le blé dur (*Triticum durum*, 2%), (9) le blé touselle (*Triticum sativum*, 1%), et (10) le maïs (*Zea mays*, 8%). Nous avons regroupé les espèces de céréales suivantes en raison du faible nombre de données disponibles pour chacune d'entre elles. Le plus grand nombre de données de rendement de céréales était disponible pour le blé (*Triticum aestivum*). Il était donc difficile d'estimer l'effet de différentes légumineuses en fonction de l'espèce de céréale suivante. Cet effet spécifique serait plus facilement quantifiable si de nouvelles expérimentations avec des données de rendement d'espèces de céréales autres que du blé (*Triticum aestivum*) sont incluses dans la base de données.

Nous avons testé l'effet de la dose de fertilisation azotée sur les céréales suivantes comme une covariable dans les modèles statistiques de la méta-analyse ; nous avons mis en évidence un effet significatif de cette covariable. Nous avons également testé si la dose de fertilisation azotée sur les cultures précédentes pouvait avoir un effet sur les rendements des céréales suivantes ; cependant, nous n'avons pas estimé d'effet significatif. Des covariables autres que celles associées à la fertilisation azotée pourraient être incluses dans des travaux futurs. L'effet de la date de semis de la culture pourrait, par exemple, être étudié. Nous avons en effet strictement comparé des rendements de céréales suivantes qui étaient cultivées avec les mêmes techniques culturales, et en particulier les mêmes dates de semis. Pour des expérimentations en zone tempérée, les rendements des céréales de printemps n'ont ainsi été comparés qu'entre elles (de même pour des céréales d'hiver). Il est probable que l'effet d'une légumineuse en tant que culture précédente dépende de la date de semis de la céréale qui la suit (pour une dose de fertilisation connue). Bien que les dates de semis aient été extraites dans la base de données, cette relation entre l'effet de la culture précédente de légumineuses et la date de semis de la céréale suivante nous paraît difficile à analyser car le nombre de données n'est pas assez élevé dans le jeu de données actuel pour analyser rigoureusement cette relation. Une deuxième covariable possible serait la gestion de l'interculture entre la légumineuse comme culture précédente et la non-légumineuse comme culture suivante ; mais le jeu de données reste à compléter avec des données d'autres articles pour analyser finement cette relation. Une troisième covariable intéressante serait l'effet d'une diminution de la présence des bioagresseurs résultant de la culture précédente de légumineuse à graines sur les rendements des cultures suivantes. Cependant, les données relatives aux bioagresseurs sont peu présentes et souvent incomplètes dans les articles – comme il a été souligné précédemment – pour les étudier de manière standardisée. De plus, l'effet des bioagresseurs se mesurerait préférentiellement à une échelle supérieure que celle d'une séquence culturale composée

d'un couple de cultures précédente et suivante. Enfin, une quatrième covariable possible serait la production de la légumineuse considérée en tant que culture précédente. Angus et al. (2015) et Seymour et al. (2012) ont proposé cette covariable comme un facteur explicatif du niveau de rendement du blé (*Triticum aestivum*) qui suit la légumineuse dans la séquence culturale. A partir de données expérimentales (situées principalement en Australie), les auteurs ont suggéré que plus la production de la légumineuse à graines est élevée, plus le rendement du blé (*Triticum aestivum*) suivant serait élevé. Les niveaux de production incluent non seulement le rendement mais également d'autres critères tels que le pourcentage d'azote fixé. Cette relation entre les niveaux de production des cultures précédentes et ceux des cultures suivantes pourrait être assez facilement à partir de notre base de données. Nous avons en effet systématiquement collecté les rendements des différentes espèces de légumineuses qui précèdent des céréales ou des oléagineux dont nous avons également collecté les rendements. Si cette analyse est envisagée dans le futur, il serait alors intéressant de confronter les résultats avec ceux issus de nombreux travaux d'équipes de recherche australiennes sur ce sujet.

## **5. Intérêt de la méta-analyse pour tester la sensibilité des conclusions aux sources de données dans une démarche de scénarisation**

Les résultats des méta-analyses des chapitres 3 et 4 ont été utilisés pour explorer deux scénarios : (1) l'effet de la substitution d'une proportion de la surface actuelle de pois protéagineux (*Pisum sativum*) par d'autres légumineuses sur la production de protéines issues de légumineuses à graines en Europe (chapitre 3), et (2) l'effet d'une augmentation de la proportion d'une légumineuse à graines introduite dans une monoculture d'une céréale sur la production céréalière (chapitre 4). Nous proposons d'appliquer les résultats issus de la méta-analyse du chapitre 3 sous forme d'une démarche de scénarisation simple. Nous avons exploré la possibilité de couvrir les besoins protéiques des volailles en Europe à partir de protéines issues d'espèces de légumineuses à graines européennes. Nous avons fait l'hypothèse que la production de volailles soit une des productions animales en Europe qui connaisse la plus forte augmentation de sa demande protéique dans les prochaines années pour couvrir ses besoins nutritionnels (European Commission, 2015). Actuellement, la demande protéique des volailles en Europe est principalement couverte par l'importation de matières premières riches en protéines, essentiellement du tourteau de soja d'Amérique (European Feed Manufacturers' Federation, 2016 ; Martin, 2014, 2015). Comme il a été souligné précédemment, l'autonomie des productions animales est un enjeu d'actualité en Europe. Pour répondre à cet enjeu, plusieurs études ont suggéré d'augmenter la production (surface) de protéines de légumineuses (cultures de légumineuses à graines) en Europe. Une augmentation de la surface de légumineuses peut favoriser des phénomènes de substitution de surface entre les espèces de légumineuses. Dans cette perspective, nous avons analysé l'effet d'une substitution de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) sur la couverture des besoins protéiques des volailles en Europe. Nous avons fait l'hypothèse que la

production de volailles soit constante en Europe. Nous avons fait également l'hypothèse que les besoins protéiques des volailles ainsi que les surfaces et les rendements des cultures de pois protéagineux (*Pisum sativum*) et de soja (*Glycine max*) soient ceux observés en moyenne en Europe entre 2003 et 2013. Nous avons considéré deux productions de volailles : (1) les volailles pondeuses, et (2) les volailles de chair. Ces deux productions ont des besoins protéiques différents ; leurs demandes protéiques sont donc distinguées. Nous avons utilisé des données statistiques de la base FAOSTAT en Europe entre 2003 et 2013.

La base de données FAOSTAT ne renseignait pas directement le nombre de volailles de chair. Il a été nécessaire de l'estimer à partir de : (1) du nombre de volailles de chair et de volailles pondeuses, et (2) du nombre de volailles pondeuses. A partir de la base de données FAOSTAT, nous avons collecté le nombre de volailles de chair et de volailles pondeuses  $N_T$  (1000 têtes) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Livestock Primary ; Meat indigenous, chicken » et « Producing Animals ; Slaughtered ». Nous avons collecté le nombre de volailles pondeuses  $N_L$  (1000 têtes) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Livestock Primary ; Egg, hen, in shell » et « Producing Animals ; Slaughtered ». Nous avons en déduit le nombre de volailles de chair  $N_B$  (1000 tête) puis en  $N_B$  (tête) en Europe entre 2003 et 2013 à partir de l'équation (1) :

$$N_B = N_T - N_L \quad (1)$$

A partir de la base de données FAOSTAT, nous avons collecté le rendement de volailles de chair (et de volailles pondeuses)  $Q_T$  ( $10^{-1}$  g tête $^{-1}$ ) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Livestock Primary ; Meat indigenous, chicken » et « Yield ». Nous avons en déduit la quantité de viande produite de volailles de chair  $M_B$  (tonne carcasse) en Europe entre 2003 et 2013 à partir de l'équation (2) :

$$M_B = N_B \times 10^{-7} \times Q_T \quad (2)$$

Nous avons convertis  $M_B$  en poids vif (i.e., de tonne carcasse à tonne poids vif) par un facteur de 73% (FranceAgriMer, 2016). Le taux de conversion alimentaire  $FCR$  est le rapport entre le poids sec d'aliments distribués et le gain de production obtenu. Appliqué aux volailles de chair, nous avons fait l'hypothèse que le taux de conversion alimentaire  $FCR_B$  est tel que  $FCR_B = 1.6$ , signifiant qu'il faut 1.6 kg d'aliments distribués pour un gain de production obtenu de 1 kg de poids vif de volailles de chair. Nous avons fait l'hypothèse que les besoins protéiques  $PRT_B$  dans la ration alimentaire des volailles de chair est tel que  $PRT_B = 0.200$ , signifiant que les besoins protéiques des volailles de chair représentent 20% de la ration alimentaire. Les constantes  $FCR_B$  et  $PRT_B$  ont été estimées à partir des recommandations nutritionnelles de deux grandes entreprises avicoles mondiales : (1) Aviagen

(<http://en.aviagen.com/>), et (2) Cobb-Vantress (<http://www.cobb-vantress.com/>). Sachant  $FCR_B$  et  $PRT_B$ , nous en avons déduit le besoin protéique  $PR_B$  (tonne) nécessaire pour produire  $M_B$  en Europe entre 2003 et 2013 à partir de l'équation (3) :

$$PR_B = M_B \times FCR_B \times PRT_B \quad (3)$$

A partir de la base de données FAOSTAT, nous avons collecté la quantité d'œufs produite  $E_L$  (tonne) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Livestock Primary ; Egg, hen, in shell » et « Production Quantity ». A partir des recommandations nutritionnelles de deux grandes entreprises avicoles mondiales : (1) Lohman GB (<http://www.lohmanngb.co.uk/>), et (2) Hendrix Genetics (<http://www.hendrix-genetics.com/>), nous avons fait l'hypothèse que le taux de conversion alimentaire  $FCR_L$  appliqué aux volailles pondeuses et les besoins protéiques  $PRT_L$  dans la ration alimentaire des volailles pondeuses sont tels que  $FCR_L = 2.1$ , et  $PRT_L = 0.184$ . Sachant  $FCR_L$  et  $PRT_L$ , nous en déduisons le besoin protéique  $PR_L$  (tonne) nécessaire pour produire  $E_L$  en Europe entre 2003 et 2013 à partir de l'équation (4) :

$$PR_L = E_L \times FCR_L \times PRT_L \quad (4)$$

A partir de la base de données FAOSTAT, nous avons collecté les surfaces récoltées de pois protéagineux (*Pisum sativum*)  $A_P$  (ha), et de soja (*Glycine max*)  $A_S$  (ha) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Crop ; Peas, dry ; Soybeans » et « Area Harvested ». Nous avons collecté les rendements des cultures de pois protéagineux (*Pisum sativum*)  $Y_P$  (t ha<sup>-1</sup>) et de soja (*Glycine max*)  $Y_S$  (t ha<sup>-1</sup>) en Europe entre 2003 et 2013. Les catégories utilisées ont été : « Crop ; Peas, dry ; Soybeans » et « Yield ». A partir de la base de données FEEDIPEDIA, nous avons collecté les pourcentages moyens de matière sèche dans les graines de pois protéagineux (*Pisum sativum*)  $DM_P$  et dans les graines de soja (*Glycine max*)  $DM_S$ , ainsi que les pourcentages moyens de protéines dans les graines de pois protéagineux (*Pisum sativum*)  $CP_P$  et dans les graines de soja (*Glycine max*)  $CP_S$ . Les catégories utilisées ont été : « Pea seeds ; Soybean seeds ». Nous avons déduit que  $DM_P = 0.865$ ,  $DM_S = 0.888$ ,  $CP_P = 0.239$ ,  $CP_S = 0.396$ . Nous en déduisons le rendement protéique de la culture du pois protéagineux (*Pisum sativum*)  $GPY_P$  (t ha<sup>-1</sup>) et le rendement protéique de la culture de soja (*Glycine max*)  $GPY_S$  (t ha<sup>-1</sup>) en Europe entre 2003 et 2013 à partir des équations (5) et (6), respectivement :

$$GPY_P = Y_P \times DM_P \times CP_P \quad (5)$$

$$GPY_S = Y_S \times DM_S \times CP_S \quad (6)$$

Les besoins protéiques des volailles de chair  $PR_B$  et des volailles pondeuses  $PR_L$ , les surfaces récoltées de pois protéagineux (*Pisum sativum*)  $A_P$  et de soja (*Glycine max*)  $A_S$ , les rendements protéiques de pois protéagineux (*Pisum sativum*)  $GPY_P$  et de soja (*Glycine max*)  $GPY_S$ , ont été estimées par année ; elles ont ensuite été moyennées entre 2003 et 2013. Ces valeurs moyennes sont utilisées dans les équations (7) à (12).

Nous avons estimé la couverture des besoins protéiques des volailles par des protéines de pois protéagineux (*Pisum sativum*) et de soja (*Glycine max*), en testant la sensibilité des résultats à deux sources de données pour déterminer l'offre protéique totale suite à la substitution de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*).

Dans un premier temps, nous avons seulement utilisé les données issues de la base de données FAOSTAT. Nous avons alors remplacé le rendement moyen protéique de la culture de pois protéagineux (*Pisum sativum*) – calculé à partir de la base de données FAOSTAT – par celui de la culture de soja (*Glycine max*) – calculé également à partir de la base de données FAOSTAT. Cette hypothèse est souvent réalisée dans des démarches de scénarisation. La surface de pois protéagineux (*Pisum sativum*) a été substituée par du soja (*Glycine max*) sans facteur de conversion lié aux effets des conditions environnementales et des techniques culturales. A partir de la base de données FAOSTAT, nous en avons déduit l'offre protéique moyenne totale  $PS_{FAOSTAT}$  (tonne) suite à une proportion de substitution  $\lambda$  de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) à partir de l'équation (7) :

$$PS_{FAOSTAT} = (GPY_S \times A_S) + (GPY_S \times \lambda \times A_P) + (GPY_P \times (1 - \lambda) \times A_P) \quad (7)$$

Nous en avons déduit les couvertures moyennes des besoins protéiques des volailles de chair  $PP_{B(FAO)}$  (%) et des volailles pondeuses  $PP_{L(FAO)}$  (%) à partir des équations (8) et (9), respectivement :

$$PP_{B(FAOSTAT)} = PS_{FAOSTAT} / PR_B \quad (8)$$

$$PP_{L(FAOSTAT)} = PS_{FAOSTAT} / PR_L \quad (9)$$

Dans un second temps, nous avons considéré que les rendements protéiques de la culture de pois protéagineux (*Pisum sativum*) et de la culture de soja (*Glycine max*) n'étaient pas interchangeables. Nous avons alors utilisé le ratio moyen de la quantité de protéines dans les graines de soja (*Glycine max*) sur la quantité de protéines dans les graines de pois protéagineux (*Pisum sativum*). Ce ratio a été estimé à l'aide de données expérimentales dans la méta-analyse du chapitre 3. Dans cette méta-analyse, nous avons comparé directement des cultures de soja (*Glycine max*) à des cultures de pois

protéagineux (*Pisum sativum*) sur les mêmes sites expérimentaux durant les mêmes saisons culturales avec les mêmes techniques culturales. En prenant en compte les estimations de la méta-analyse, nous avons ainsi limité le risque de confusion d'effets résultant des conditions pédoclimatiques et des techniques culturales.

A partir de la base de données FAOSTAT avec l'utilisation des estimations de la méta-analyse du chapitre 3, nous en avons déduit l'offre protéique moyenne totale  $PS_{MA}$  (tonne) suite à une proportion de substitution  $\lambda$  de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*), et le ratio moyen  $R_M$  de la quantité de protéines dans les graines de soja (*Glycine max*) sur la quantité de protéines dans les graines de pois protéagineux (*Pisum sativum*) à partir de l'équation (10) :

$$PS_{MA} = (GPY_S \times A_S) + (GPY_P \times R_M \times \lambda \times A_P) + (GPY_P \times (1 - \lambda) \times A_P) \quad (10)$$

Nous en avons déduit les couvertures moyennes des besoins protéiques des volailles de chair  $PP_{B(MA)}$  (%) et des volailles pondeuses  $PP_{L(MA)}$  (%) à partir des équations (11) et (12), respectivement :

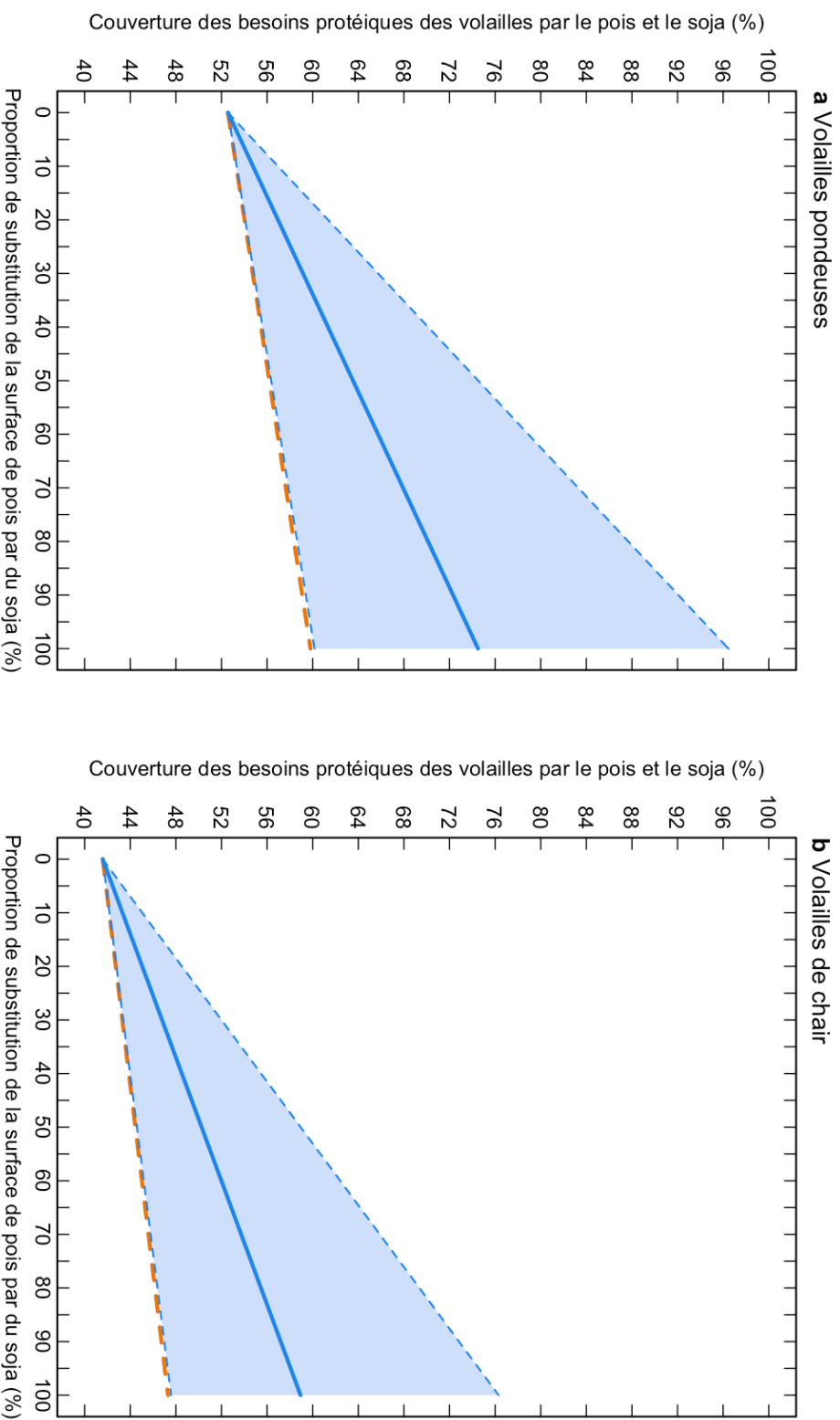
$$PP_{B(MA)} = PS_{MA} / PR_B \quad (11)$$

$$PP_{L(MA)} = PS_{MA} / PR_L \quad (12)$$

L'intervalle de confiance à 95% a été estimé de la même manière que l'offre protéique moyenne totale  $PS_{MA}$ , mais en remplaçant le ratio moyen  $R_M$  dans l'équation (10) par les ratios des bornes inférieures et supérieures de l'intervalle de confiance à 95%.

Les résultats obtenus à partir de la base de données FAOSTAT sans utilisation des résultats de la méta-analyse montrent un gain moyen de 6% de couverture des besoins protéiques des volailles de chair (6%) et des volailles pondeuses (7%), en considérant une substitution totale de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) en Europe (Figure 6). Au contraire, les résultats obtenus à partir de la base de données FAOSTAT avec utilisation des résultats de la méta-analyse montrent un gain moyen de 20% de couverture des besoins protéiques des volailles de chair (17%) et des volailles pondeuses (22%), en considérant une substitution totale de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*) en Europe (Figure 6). La méthode avec utilisation des résultats de la méta-analyse permet de quantifier l'incertitude associée au gain moyen de couverture des besoins protéiques des volailles en Europe. Cette incertitude est forte en raison du faible nombre de données expérimentales disponibles où le pois protéagineux (*Pisum sativum*) et le soja (*Glycine max*) sont directement comparés sur les mêmes sites expérimentaux durant les mêmes saisons culturales avec les mêmes techniques culturales en Europe. Cette démarche de scénarisation,

aussi simple soit-elle, a montré que les conclusions peuvent être sensibles à la source de données utilisée, soulignant ainsi la nécessité d'utiliser différentes sources de données pour quantifier l'incertitude associée aux conclusions qui en émanent.



**Figure 6.** Couverture (%) des besoins protéiques des volailles en Europe par le pois protéagineux (*Pisum sativum*) et le soja (*Glycine max*) en fonction de la proportion (%) de substitution de la surface de pois protéagineux (*Pisum sativum*) par du soja (*Glycine max*). Deux productions de volailles sont considérées : (a) les volailles pondueuses et (b) les volailles de chair. La courbe orange pointillée indique la couverture moyenne des besoins protéiques des volailles, estimée à partir de la base de données FAOSTAT sans utilisation des résultats de la méta-analyse. La courbe bleue continue indique la couverture moyenne des besoins protéiques des volailles, estimée à partir de la base de données FAOSTAT avec utilisation des résultats de la méta-analyse. Les deux courbes bleues pointillées indiquent l'intervalle de confiance à 95% associé à la couverture moyenne des besoins protéiques des volailles, estimées à partir de la base de données FAOSTAT avec utilisation des résultats de la méta-analyse.





## **Conclusion**



## Conclusion

L'objectif de ma thèse était d'identifier des espèces de légumineuses à graines potentiellement intéressantes à produire en Europe. La Figure 2 (i.e., présentée dans la Discussion) synthétise les résultats de mon travail, et permet d'évaluer 29 espèces de légumineuses à partir de plusieurs critères de production.

D'après mes résultats, la fève (*Vicia faba*) est une des espèces de légumineuses à graines les plus performantes en Europe. La variabilité interannuelle de son rendement en grains est une des plus faibles de l'ensemble des rendements des légumineuses à graines, dans la majorité des sous-régions européennes. Les performances productives de la fève (*Vicia faba*) sont similaires ou supérieures à celles du pois protéagineux (*Pisum sativum*) pour l'ensemble des critères étudiés en Europe. L'incertitude associée à ses performances est faible. Le classement obtenu pour la fève (*Vicia faba*) en Europe est confirmé en Amérique du Nord et Océanie, d'après les résultats de la méta-analyse du chapitre 3. Similairement à l'ensemble des légumineuses à graines, l'effet de la fève (*Vicia faba*) sur les rendements des céréales suivantes est positif. Cet effet est celui qui est le moins impacté par la dose de fertilisation azotée appliquée sur les céréales suivantes ; nous avons en effet estimé que cet effet ne devenait négligeable qu'à partir d'une dose de fertilisation sur les céréales supérieure à 198 kg N ha<sup>-1</sup>.

Le pois protéagineux (*Pisum sativum*) est également une espèce performante en Europe. Ses niveaux de production sont similaires à celles de la fève (*Vicia faba*) pour tous les critères. Cependant, la quantité de protéines et la teneur en énergie brute dans ses graines sont significativement inférieures à celles du soja (*Glycine max*). La variabilité interannuelle du rendement en grains du pois protéagineux (*Pisum sativum*) est assez faible dans certaines régions européennes, notamment en Europe du Nord. L'effet de cette légumineuse en tant que culture précédente sur les rendements des céréales devient négligeable à partir d'une dose de fertilisation azotée supérieure à 168 kg N ha<sup>-1</sup> sur les céréales.

Les résultats révèlent également que le soja (*Glycine max*) et les lupins (*Lupinus* spp.) ont de bonnes performances productives en Europe pour plusieurs critères. Les bonnes performances productives du soja (*Glycine max*) sont principalement dues à ses graines riches en protéines et en énergie brute. Cependant, la variabilité interannuelle de son rendement en grains est plus élevée que celle du pois protéagineux (*Pisum sativum*). Le rendement en grains du soja (*Glycine max*) est le moins variable en Europe de l'Est, d'après nos résultats. Cette sous-région européenne apparaît donc la plus adaptée pour

cultiver du soja (*Glycine max*), alors que cette espèce semble présenter des rendements plus instables en Europe du Nord et du Sud, d'après ce critère. L'effet du soja (*Glycine max*) sur les rendements des céréales suivantes est limité, et devient négligeable pour une dose de fertilisation appliquée sur les céréales assez faible (i.e., supérieure à 135 kg N ha<sup>-1</sup>).

Bien que certaines espèces de lupins (e.g., le lupin à feuilles étroites (*Lupinus angustifolius*)) présentent des performances productives intéressantes par rapport à celles du pois protéagineux (*Pisum sativum*) en Europe (en particulier, pour leurs quantités de protéines dans les graines), ces espèces sont caractérisées par une très forte variabilité interannuelle de leurs rendements en grains dans la majorité des sous-régions européennes. L'incertitude est assez élevée pour certaines espèces de lupin (e.g., les lupins jaunes et blancs (*Lupinus luteus*, et *Lupinus albus*, respectivement) à cause d'un nombre faible de données expérimentales où elles sont comparées directement au pois protéagineux (*Pisum sativum*) en Europe. Cependant, avec la fêverole (*Vicia faba*), les lupins (*Lupinus* spp.) sont les espèces de légumineuses qui ont l'effet le plus élevé sur les rendements des céréales suivantes. Leur effet tend à devenir négligeable pour une dose élevée de fertilisation azotée sur les céréales (i.e., supérieure entre 185 et 190 kg N ha<sup>-1</sup>).

La lentille (*Lens culinaris*) et le pois chiche (*Cicer arietinum*) ont des performances productives moins bonnes que celles du pois protéagineux (*Pisum sativum*) en Europe pour plusieurs critères de production. Les niveaux de variabilité interannuelle de leurs rendements en grains sont plus élevés que celui du pois protéagineux (*Pisum sativum*). Leurs rendements en grains et leurs quantités en protéines sont plus faibles que ceux du pois protéagineux (*Pisum sativum*). L'effet du pois chiche sur les rendements des céréales suivantes devient négligeable dès que la dose de fertilisation azotée de la céréale devient supérieure à 124 kg N ha<sup>-1</sup>. Bien que nous n'ayons pas pu collecter des données expérimentales sur la production du haricot commun (*Phaseolus vulgaris*) par rapport au pois protéagineux (*Pisum sativum*) en Europe, cette espèce présente une forte variabilité interannuelle de son rendement en grains dans la majorité des sous-régions européennes. L'effet du haricot commun (*Phaseolus vulgaris*) sur les rendements des céréales suivantes devient négligeable quand les céréales sont fertilisées à des doses de fertilisation azotée supérieures à 129 kg N ha<sup>-1</sup>, d'après les résultats de la méta-analyse du chapitre 4. Malgré leurs performances productives relativement faibles, la lentille (*Lens culinaris*), le pois chiche (*Cicer arietinum*), et le haricot commun (*Phaseolus vulgaris*) ont des propriétés nutritionnelles intéressantes qui permettent de les utiliser préférentiellement en alimentation humaine – un débouché généralement plus rémunérateur que celui de l'alimentation animale.

Les classements obtenus en Amérique et Océanie montrent qu'il serait intéressant de tester plusieurs espèces de légumineuses dans de futures expérimentations agronomiques en Europe afin de préciser leurs intérêts potentiels pour l'agriculture européenne, notamment les lupins (*Lupinus* spp.), les gesses

(*Lathyrus* spp.), et les vesces (*Vicia* spp. excepté *Vicia faba*). Cependant, les graines de certaines de ces espèces contiennent des facteurs nutritionnels indésirables. L'amélioration variétale et des traitements technologiques seront donc indispensables pour inactiver ces facteurs, et envisager, dans le futur, de les utiliser comme des substituts possibles au pois protéagineux (*Pisum sativum*) en Europe.

Globalement, ma thèse met en exergue la forte complémentarité de différentes sources de données pour réaliser des comparaisons d'une large diversité spécifique de légumineuses à graines sur la base de nombreux critères de performances productives (et environnementales). Les données statistiques ont permis de caractériser la production des légumineuses à de larges échelles spatiotemporelles, mais elles sont associées à un risque de confusion d'effets entre les caractéristiques des espèces, les conditions environnementales, et les techniques culturales. Les données issues d'expérimentations agronomiques ont permis de contourner en grande partie cette limite lorsque plusieurs légumineuses sont comparées sur les mêmes sites expérimentaux durant les mêmes saisons culturales. Les données expérimentales ont permis également de prendre en compte d'autres critères de productivité que le rendement, et de tester les effets de certaines covariables environnementales et agronomiques sur les niveaux de performances. Les données nutritionnelles ont permis de comparer des espèces de légumineuses selon des critères tels que la quantité de protéines et la teneur en énergie brute dans les graines ; ces deux critères étant très peu mesurés dans les expérimentations agronomiques. Les conclusions issues de démarches de scénarisation peuvent être sensibles à la source de données utilisée, soulignant ainsi la nécessité d'utiliser plusieurs sources données pour quantifier l'incertitude associée aux résultats.

A partir des différentes comparaisons plurispécifiques d'espèces de légumineuses à graines, mon travail de thèse permet de conclure que le choix de l'espèce de légumineuses à graines sera probablement un levier significatif pour augmenter la production de protéines issues de ces cultures en Europe, et compenser partiellement les diminutions relatives attendues de production céréalière suite à l'introduction de légumineuses à graines dans les systèmes céréaliers européens.



## Références





## Références

Les références de l'Introduction et de la Discussion sont hiérarchisées alphabétiquement, puis chronologiquement.

1. AGRESTE, 2014. Enquêtes sur les pratiques culturales 2011. La fertilisation, Les Dossiers AGRESTE, **21**.
2. Akibode, S., Maredia, M., 2011. Global and regional trends in production, trade and consumption of food legume crops. Michigan State University. [impact.cgiar.org/sites/default/files/images/Legumetrendsv2.pdf](http://impact.cgiar.org/sites/default/files/images/Legumetrendsv2.pdf). Accès en Janvier 2016.
3. Angadi, S.V., McConkey, B.G., Cutforth, H.W., Miller, P.R., Ulrich, D., Selles, F., Volkmar, K.M., Entz, M.H., Brandt, S.A., 2008. Adaptation of alternative pulse and oilseed crops to the semiarid Canadian Prairie: seed yield and water use efficiency. *Can. J. Plant Sci.* **88**, 425–38.
4. Angus, J.F., Kirkegaard, J.A., Hunt, J.R., Ryan, M.H., Ohlander, L., Peoples, M.B., 2015. Break crops and rotations for wheat. *Crop Pasture Sci.* **66**, 523–552.
5. Aranjuelo, I., Arrese-Igor, C., Molero, G., 2014. Nodule performance within a changing environmental context. *J. Plant Physiol.* **171**, 1076–1090.
6. Araújo, S.S., Beebe, S., Crespi, M., Delbreil, B., González, E.M., Gruber, V., Lejeune-Henaut, I., Link, W., Monteros, M.J., Prats, E., Rao, I., Vadez, V., Patto, M.C.V., 2015. Abiotic stress responses in legumes: strategies used to cope with environmental challenges. *Crit. Rev. Plant Sci.* **34**, 237–280.
7. Armstrong, E.L., Heenan, D.P., Pate, J.S., Unkovich, M.J., 1997. Nitrogen benefits of lupins, field pea, and chickpea to wheat production in south-eastern Australia. *Aust. J. Agr. Res.* **48**, 39-48.
8. Asif, M., Rooney, L.W., Ali, R., Riaz, M.N., 2013. Application and opportunities of pulses in food system: a review. *Crit. Rev. Food Sci. Nutr.* **53**, 1168–1179.
9. Badaruddin, M., Meyer, D.W., 1994. Grain legume effects on soil nitrogen, grain yield, and nitrogen nutrition of wheat. *Crop Sci.* **34**, 1304-1309.
10. Bassett, C., Boye, J., Tyler, R., Oomah, B.D., 2010. Molecular, functional and processing characteristics of whole pulses and pulse fractions and their emerging food and nutraceutical applications. *Food Res. Int.* **43**, 397–398.

11. Ben-Ari, T., Makowski, D., 2014. Decomposing global crop yield variability. *Environ. Res. Lett.* **9**, 114011.
12. Bennett, A.J., Bending, G.D., Chandler, D., Hilton, S., Mills, P., 2012. Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. *Biol. Rev. Camb. Philos. Soc.* **87**, 52–71.
13. Bland, C.J., Meurer, L.N., Maldonado, G., 1995. A systematic approach to conducting a non-statistical meta-analysis of research literature. *Acad. Med. J. Assoc. Am. Med. Coll.* **70**, 642–653.
14. Borenstein M., Hedges L.V., Higgins J.P.T., Rothstein H.R., 2009. Introduction to meta-analysis. Wiley, Hoboken.
15. Britz, W., Verburg, P.H., Leip, A., 2011. Modelling of land cover and agricultural change in Europe: combining the CLUE and CAPRI-Spat approaches. *Agr. Ecosyst. Environ.* **142**, 40–50.
16. Brockwell, P.J., Davis, R.A., 2002. Introduction to time series and forecasting. Springer, New York.
17. Bues, A., Preissel, S., Reckling, M., Zander, P., Kuhlman, T., Topp, K., Watson, C., Lindström, K., Stoddard, F.L., Murphy-Bokern, D., 2013. The environmental role of protein crops in the new Common Agricultural Policy. European Parliament. <http://www.legumefutures.de/legumefuturesnews/30052013brussels.html>. Accès en Octobre 2014.
18. Calderón, F.J., Vigil, M.F., Nielsen, D.C., Benjamin, J.G., Poss, D.J., 2012. Water use and yields of no-till managed dryland grasspea and yellow pea under different planting configurations. *Field Crop. Res.* **125**, 179–185.
19. Carpenter, S.R., Pingali, P.L., Bennett, E.M., Zurek, M.B., 2005. Ecosystems and human well-being: scenarios. **2**, The Millennium Ecosystem Assessment, Washington DC.
20. Chalk, P.M., 1998. Dynamics of biologically fixed N in legume-cereal rotations: a review. *Aust. J. Agr. Res.* **49**, 303–316.
21. Chen, C., Miller, P., Muehlbauer, F., Neill, K., Wichman, D., McPhee, K., 2006. Winter pea and lentil response to seeding date and micro- and macro-environments. *Agron. J.* **98**, 1655–1663.
22. Cheruiyot, E.K., Mumera, L.M., Nakhone, L.N., Mwonga, S.M., 2003. Effect of legume-managed fallow on weeds and soil nitrogen in following maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) crops in the Rift Valley highlands of Kenya. *Aust. J. Exp. Agr.* **43**, 597–604.
23. European Commission, 2015. EU Agricultural Outlook. Prospects for EU agricultural markets and income 2015-2025. European Commission. [http://ec.europa.eu/agriculture/markets-and-prices/medium-term-outlook/index\\_en.htm](http://ec.europa.eu/agriculture/markets-and-prices/medium-term-outlook/index_en.htm). Accès en Juin 2016.
24. Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., Duc, G., 2010. Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. *Field Crop. Res.* **115**, 329–339.

25. Crotty, F.V., Fychan, R., Scullion, J., Sanderson, R., Marley, C.L., 2015. Assessing the impact of agricultural forage crops on soil biodiversity and abundance. *Soil Biol. Biochem.* **91**, 119–126.
26. Cutforth, H.W., McGinn, S.M., McPhee, K.E., Miller, P.R., 2007. Adaptation of pulse crops to the changing climate of the northern Great Plains. *Agron. J.* **99**, 1684–1699.
27. Dakora, F. D., Aboyinga, R. A., Mahama, Y., Apaseku, J., 1987. Assessment of N<sub>2</sub> fixation in groundnut (*Arachis hypogaea* L.) and cowpea (*Vigna unguiculata* L. Walp.) and their relative N contribution to a succeeding maize crop in northern Ghana. *MIRCEN J. Appl. Microb.* **3**, 389–399.
28. Dalgaard, R., Schmidt, J., Halberg, N., Christensen, P., Thrane, M., Pengue, W.A., 2008. LCA of soybean meal. *Int. J. Life Cycle Assess.* **13**, 240–254.
29. Davis, J., Sonesson, U., Baumgartner, D.U., Nemecek, T., 2010. Environmental impact of four meals with different protein sources: case studies in Spain and Sweden. *Food Res. Int.* **43**, 1874–1884.
30. De Visser, C.L.M., Schreuder, R., Stoddard, F., 2014. The EU's dependency on soya bean import for the animal feed industry and potential for EU produced alternatives. *OCL* **21**, D407.
31. D'Mello, J.P.F., 1992. Chemical constraints to the use of tropical legumes in animal nutrition. *Anim. Feed Sci. Tech.* **38**, 237–261.
32. Doré, T., Makowski, D., Malézieux, E., Munier-Jolain, N., Tchamitchian, M., Tittone, P., 2011. Facing up to the paradigm of ecological intensification in agronomy: revisiting methods, concepts and knowledge. *Eur. J. Agron.* **34**, 197–210.
33. Doyle, P.T., Francis, S.A., Stockdale, C.R., 2005. Associative effects between feeds when concentrate supplements are fed to grazing dairy cows: a review of likely impacts on metabolisable energy supply. *Aust. J. Agr. Res.* **56**, 1315–1329.
34. Duranti, M., 2006. Grain legume proteins and nutraceutical properties. *Fitoterapia* **77**, 67–82.
35. Erskine, W., Sarker, A., Kumar, S., 2011. Crops that feed the world 3. Investing in lentil improvement toward a food secure world. *Food Secur.* **3**, 127–139.
36. EUR-Lex, 2016. EUR-Lex. [http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L\\_.2013.347.01.0608.01.FRA](http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L_.2013.347.01.0608.01.FRA). Accès en Septembre 2016.
37. Evans, J., McNeill, A.M., Unkovich, M.J., Fettell, N.A., Heenan, D.P., 2001. Net nitrogen balances for cool-season grain legume crops and contributions to wheat nitrogen uptake: a review. *Aust. J. Exp. Agr.* **41**, 347–359.
38. Evans, J.A., Foster, J.G., 2011. Metaknowledge. *Science* **331**, 721–725.
39. FAOSTAT, 2014. Statistics Division of Food and Agriculture Organization of the United Nations (FAOSTAT). <http://faostat3.fao.org/home/E>. Accès en Septembre 2014.

40. FAO STAT, 2016. Statistics Division of Food and Agriculture Organization of the United Nations (FAO STAT). <http://faostat3.fao.org/home/E>. Accès en Septembre 2016.
41. European Feed Manufacturers' Federation, 2016. Annual report 2015-2016. European Feed Manufacturers' Federation. <http://www.fefac.eu/>. Accès en Septembre 2016.
42. Flores, F., Nadal, S., Solis, I., Winkler, J., Sass, O., Stoddard, F.L., Link, W., Raffiot, B., Muel, F., Rubiales, D., 2012. Faba bean adaptation to autumn sowing under European climates. *Agron. Sustain. Dev.* **32**, 727–734.
43. Flores, F., Hybl, M., Knudsen, J.C., Marget, P., Muel, F., Nadal, S., Narits, L., Raffiot, B., Sass, O., Solis, I., Winkler, J., Stoddard, F.L., Rubiales, D., 2013. Adaptation of spring faba bean types across European climates. *Field Crop. Res.* **145**, 1–9.
44. Foyer, C.H., Lam, H.-M., Nguyen, H.T., Siddique, K.H.M., Varshney, R.K., Colmer, T.D., Cowling, W., Bramley, H., Mori, T.A., Hodgson, J.M., Cooper, J.W., Miller, A.J., Kunert, K., Vorster, J., Cullis, C., Ozga, J.A., Wahlqvist, M.L., Liang, Y., Shou, H., Shi, K., Yu, J., Fodor, N., Kaiser, B.N., Wong, F.-L., Valliyodan, B., Considine, M.J., 2016. Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* **2**, 16112.
45. FranceAgriMer, 2016. Données et bilans. Les filières animales terrestres et aquatiques. Bilan 2015. Perspectives 2016. FranceAgriMer. [http://www.franceagrimer.fr/Stockage-Actualites/Publication-Bilan-2015-Perspectives-2016-des-filieres-animales-terrestres-et-aquatiques/\(filiere\)/539/\(nodeActu\)/546](http://www.franceagrimer.fr/Stockage-Actualites/Publication-Bilan-2015-Perspectives-2016-des-filieres-animales-terrestres-et-aquatiques/(filiere)/539/(nodeActu)/546). Accès en Septembre 2016.
46. French, R., Ewing, M., 1989. Soil type influences the relative yields of different cereals and crop legumes in the western Australian wheat-belt. *Aust. J. Exp. Agr.* **29**, 829–835.
47. Gan, Y., Hamel, C., O'Donovan, J.T., Cutforth, H., Zentner, R.P., Campbell, C.A., Niu, Y., Poppy, L., 2015. Diversifying crop rotations with pulses enhances system productivity. *Sci. Rep.* **5**, 14625.
48. Gatlin, D.M., Barrows, F.T., Brown, P., Dabrowski, K., Gaylord, T.G., Hardy, R.W., Herman, E., Hu, G., Krogdahl, Å., Nelson, R., Overturf, K., Rust, M., Sealey, W., Skonberg, D., J Souza, E., Stone, D., Wilson, R., Wurtele, E., 2007. Expanding the utilization of sustainable plant products in aquafeeds: a review. *Aquac. Res.* **38**, 551–579.
49. Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., Young, N.D., 2005. Legumes as a model plant family. Genomics for food and feed. Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiol.* **137**, 1228–1235.
50. Ghadim, A.K.A., Pannell, D.J., Burton, M.P., 2005. Risk, uncertainty, and learning in adoption of a crop innovation. *Agr. Econ.* **33**, 1–9.
51. Ghosh, P.K., Singh, N.P., 1996. Production potential of summer legumes-maize (*Zea mays*) sequence under varying levels of nitrogen. *Indian J. Agron.* **41**, 525–528.

52. Gilani, G.S., Cockell, K.A., Sepehr, E., 2005. Effects of antinutritional factors on protein digestibility and amino acid availability in foods. *J. AOAC Int.* **88**, 967–987.
53. Graham, P.H., Vance, C.P., 2003. Legumes: importance and constraints to greater use. *Plant Physiol.* **131**, 872–877.
54. Gregory, P.J., 1998. Alternative crops for Duplex soils: growth and water use of some cereal, legume, and oilseed crops, and pastures. *Aust. J. Agr. Res.* **49**, 21–32.
55. Guillamón, E., Pedrosa, M.M., Burbano, C., Cuadrado, C., Sánchez, M. de C., Muzquiz, M., 2008. The trypsin inhibitors present in seed of different grain legume species and cultivar. *Food Chem.* **107**, 68–74.
56. Gurevitch, J., Hedges, L.V., 1993. Meta-analysis: combining the results of independent experiments. Design and analysis of ecological experiments. Chapman and Hall, New York.
57. Hache, E., 2015. Géopolitique des protéines. *Rev. Int. Strat.* **97**, 36–46.
58. Hanbury, C.D., White, C.L., Mullan, B.P., Siddique, K.H.M., 2000. A review of the potential of *Lathyrus sativus* L. and *L. cicera* L. grain for use as animal feed. *Anim. Feed Sci. Tech.* **87**, 1–27.
59. Hardarson, G., Atkins, C., 2003. Optimising biological N<sub>2</sub> fixation by legumes in farming systems. *Plant Soil* **252**, 41–54.
60. Harries, M., Anderson, G.C., Hüberli, D., 2015. Crop sequences in western Australia: what are they and are they sustainable? Findings of a four-year survey. *Crop Pasture Sci.* **66**, 634–647.
61. Hartman, G.L., West, E.D., Herman, T.K., 2011. Crops that feed the World 2. Soybean – worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **3**, 5–17.
62. Häusling, M., 2011. The EU protein deficit: what solution for a long-standing problem? (2010/2111(INI)). European Parliament.  
<http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A7-2011-0026&language=EN>. Accès en Septembre 2014.
63. Helming, J., Kuhlman, T., Linderhof, V., Oudendag, D., 2014. Impacts of legume-related policy scenarios. Legume Futures Report 4.5. European Commission.  
[http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_4.5.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_4.5.pdf). Accès en Septembre 2014.
64. Hossard, L., Archer, D.W., Bertrand, M., Colnenne-David, C., Debaeke, P., Ernfors, M., Jeuffroy, M.-H., Munier-Jolain, N., Nilsson, C., Sanford, G.R., Snapp, S.S., Jensen, E.S., Makowski, D., 2016. A meta-analysis of maize and wheat yields in low-input vs. conventional and organic systems. *Agron. J.* **108**, 1155–1167.
65. Jansson, T., Heckeley, T., 2011. Estimating a primal model of regional crop supply in the European Union. *J. Agr. Econ.* **62**, 137–152.

66. Jensen, C.R., Joernsgaard, B., Andersen, M.N., Christiansen, J.L., Mogensen, V.O., Friis, P., Petersen, C.T., 2004. The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop. *Eur. J. Agron.* **20**, 405–418.
67. Jeuffroy, M.H., Baranger, E., Carrouée, B., de Chezelles, E., Gosme, M., Hénault, C., Schneider, A., Cellier, P., 2013. Nitrous oxide emissions from crop rotations including wheat, oilseed rape and dry peas. *Biogeosciences* **10**, 1787–1797.
68. Jezierny, D., Mosenthin, R., Bauer, E., 2010. The use of grain legumes as a protein source in pig nutrition: a review. *Anim. Feed Sci. Tech.* **157**, 111–128.
69. Jouffret, P., Labalette, F., Parachini, E., 2015. Analyse multicritère de la production de soja dans des exploitations agricoles contrastées du Sud-Ouest de la France. *OCL* **22**, D505.
70. Jukanti, A.K., Gaur, P.M., Gowda, C.L.L., Chibbar, R.N., 2012. Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. *Br. J. Nutr.* **108**, 11–26.
71. Kirkegaard, J., Christen, O., Krupinsky, J., Layzell, D., 2008. Break crop benefits in temperate wheat production. *Field Crop. Res.* **107**, 185–195.
72. Koivunen, E., Partanen, K., Perttilä, S., Palander, S., Tuunainen, P., Valaja, J., 2016. Digestibility and energy value of pea (*Pisum sativum* L.), faba bean (*Vicia faba* L.) and blue lupin (narrow-leaf) (*Lupinus angustifolius*) seeds in broilers. *Anim. Feed Sci. Tech.* **218**, 120–127.
73. Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **15**, 259–263.
74. Krön, M., Bittner, U., 2015. Danube Soya – improving European GM-free soya supply for food and feed. *OCL* **22**, D509.
75. Laleg, K., Cassan, D., Barron, C., Prabhasankar, P., Micard, V., 2016. Structural, culinary, nutritional and anti-nutritional properties of high protein, gluten free, 100% legume pasta. *PLOS ONE* **11**, e0160721.
76. Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J., Garnier, J., 2014. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ. Res. Lett.* **9**, 105011.
77. Laurence, R., 1979. A comparison of the grain and protein yield potential of some annual legume species in South Australia. *Aust. J. Exp. Agr. Anim. Husb.* **19**, 495–503.
78. Laurent, A., Pelzer, E., Loyce, C., Makowski, D., 2015. Ranking yields of energy crops: a meta-analysis using direct and indirect comparisons. *Renew. Sustain. Energy Rev.* **46**, 41–50.
79. Le Cadre, P., Pressenda, F., Labalette, F., Peyronnet, C., 2015. Valorisation de tourteaux de soja issus d'une production locale non OGM chez les fabricants d'aliments. *OCL* **22**, D507.

80. Leff, B., Ramankutty, N., Foley, J.A., 2004. Geographic distribution of major crops across the world. *Glob. Biogeochem. Cycles* **18**, 1009.
81. LMC International, 2009a. Evaluation of measures applied under the Common Agricultural Policy to the protein crop sector. Main Report.  
[http://ec.europa.eu/agriculture/eval/reports/protein\\_crops/index\\_en.htm](http://ec.europa.eu/agriculture/eval/reports/protein_crops/index_en.htm). European Commission. Accès en Septembre 2014.
82. LMC International, 2009b. Evaluation of measures applied under the Common Agricultural Policy to the protein crop sector. Case Study Monographs.  
[http://ec.europa.eu/agriculture/eval/reports/protein\\_crops/index\\_en.htm](http://ec.europa.eu/agriculture/eval/reports/protein_crops/index_en.htm). European Commission. Accès en Septembre 2014.
83. Longstaff, M., McBain, B., McNab, J.M., 1991. The antinutritive effect of proanthocyanidin-rich and proanthocyanidin-free hulls from field beans on digestion of nutrients and metabolisable energy in intact and caecotomised cockerels. *Anim. Feed Sci. Tech.* **34**, 147–161.
84. Lucas, M.M., Stoddard, F.L., Annicchiarico, P., Frias, J., Martínez-Villaluenga, C., Sussmann, D., Duranti, M., Seger, A., Zander, P.M., Pueyo, J.J., 2015. The future of lupin as a protein crop in Europe. *Front. Plant Sci.* **6**, 705.
85. Magrini, M.-B., Anton, M., Cholez, C., Corre-Hellou, G., Duc, G., Jeuffroy, M.-H., Meynard, J.-M., Pelzer, E., Voisin, A.-S., Walrand, S., 2016. Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. *Ecol. Econ.* **126**, 152–162.
86. Makowski, D., Nesme, T., Papy, F., Doré, T., 2013. Global agronomy, a new field of research. A review. *Agron. Sustain. Dev.* **34**, 1–15.
87. Malik, R.S., Seymour, M., French, R.J., Kirkegaard, J.A., Lawes, R.A., Liebig, M.A., 2015. Dynamic crop sequencing in western Australian cropping systems. *Crop Pasture Sci.* **66**, 594–609.
88. Marra, M.C., Schurle, B.W., 1994. Kansas wheat yield risk measures and aggregation: a meta-analysis approach. *J. Agr. Resour. Econ.* **19**, 69–77.
89. Martens, S.D., Tiemann, T.T., Bindelle, J., Peters, M., Lascano, C.E., 2012. Alternative plant protein sources for pigs and chickens in the tropics – nutritional value and constraints: a review. *J. Agr. Rural Dev. Trop.* **113**, 101–123.
90. Martin, N., 2014. What is the way forward for protein supply? The European perspective. *OCL* **21**, D403.
91. Martin, N., 2015. Domestic soybean to compensate the European protein deficit: illusion or real market opportunity? *OCL* **22**, D502.



92. Maury, P., Andrianasolo, F.N., Alric, F., Berger, M., Beugniet, G., Chambert, C., Champolivier, L., Doumenc, A., Estragnat, A., Gras, A., Jeanson, P., Jouffret, P., Labalette, F., Thomas, R., Justes, E., Debaeke, P., 2015. Le semis très précoce : une stratégie agronomique pour améliorer les performances du soja en France ? *OCL* **22**, D503.
93. Michel, L., Makowski, D., 2013. Comparison of statistical models for analyzing wheat yield time series. *PLOS ONE* **8**, e78615.
94. Miller, P.R., McDonald, C.L., Derksen, D.A., Waddington, J., 2001. The adaptation of seven broadleaf crops to the dry semiarid prairie. *Can. J. Plant Sci.* **81**, 29–43.
95. Miller, P.R., McConkey, B.G., Clayton, G.W., Brandt, S.A., Staricka, J.A., Johnston, A.M., Lafond, G.P., Schatz, B.G., Baltensperger, D.D., Neill, K.E., 2002. Pulse crop adaptation in the northern Great Plains. *Agron. J.* **94**, 261–72.
96. Miller, P.R., Brandt, S.A., McDonald, C.L., Waddington, J., 2006. Chickpea, lentil, and pea response to delayed spring seeding on the northern Great Plains. *Can. J. Plant Sci.* **86**, 1059–1070.
97. Mudryj, A.N., Yu, N., Aukema, H.M., 2014. Nutritional and health benefits of pulses. *Appl. Physiol. Nutr. Metab.* **39**, 1197–1204.
98. Mueller, N.D., West, P.C., Gerber, J.S., MacDonald, G.K., Polasky, S., Foley, J.A., 2014. A tradeoff frontier for global nitrogen use and cereal production. *Environ. Res. Lett.* **9**, 54002.
99. Multari, S., Stewart, D., Russell, W.R., 2015. Potential of fava bean as future protein supply to partially replace meat intake in the human diet. *Compr. Rev. Food Sci. Food Saf.* **14**, 511–522.
100. Murphy-Bokern, D., Watson, C., Stoddart, F.L., Lindstrom, K., Zander, P., Reckling, M., Preissel, S. Bues, A., 2014. Outlook for knowledge and technology for legume-supported cropping systems Legume Futures Report 5.3. European Commission. [http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_5.3.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_5.3.pdf). Accès en Avril 2016.
101. Nedumaran, S., Abinaya, P., Shraavya, B., Rao, P.P., Bantilan, M.C.S., 2013. Grain legumes production, consumption and trade trends in developing countries-an assessment and synthesis. Socioeconomics Discussion Paper Series. **3**, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT).
102. Nemecek, T., Von Richthofen, J.-S., Dubois, G., Casta, P., Charles, R., Pahl, H., 2008. Environmental impacts of introducing grain legumes into European crop rotations. *Eur. J. Agron.* **28**, 380–393.
103. Paillard, S., Treyer, S., Dorin, B., 2010. Agrimonde. Scénarios et défis pour nourrir le monde en 2050. Quæ, Versailles.
104. Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644.

105. Peltonen-Sainio, P., Niemi, J.K., 2012. Protein crop production at the northern margin of farming: to boost, or not to boost. *Agr. Food Sci.* **21**, 370–383.
106. Peoples, M.B., Brockwell, J., Herridge, D.F., Rochester, I.J., Alves, B.J.R., Urquiaga, S., Boddey, R.M., Dakora, F.D., Bhattarai, S., Maskey, S.L., Sampet, C., Rerkasem, B., Khan, D.F., Hauggaard-Nielsen, H., Jensen, E.S., 2009. The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. *Symbiosis* **48**, 1–17.
107. Philibert, A., Loyce, C., Makowski, D., 2012. Assessment of the quality of meta-analysis in agronomy. *Agr. Ecosyst. Environ.* **148**, 72–82.
108. Pittelkow, C.M., Liang, X., Linquist, B.A., Van Groenigen, K.J., Lee, J., Lundy, M.E., Van Gestel, N., Six, J., Venterea, R.T., Van Kessel, C., 2015. Productivity limits and potentials of the principles of conservation agriculture. *Nature* **517**, 365–368.
109. Popp, M., Rudstrom, M., Manning, P., 2005. Spatial yield risk across region, crop and aggregation method. *Can. J. Agr. Econ.* **53**, 103–115.
110. Preissel, S., Reckling, M., Schläfke, N., Zander, P., 2015. Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: a review. *Field Crop. Res.* **175**, 64–79.
111. Quinsac, A., Labalette, F., Carré, P., Parachini, E., Jouffret, P., 2015. Simulation technico-économique d'une filière locale de valorisation du soja en alimentation animale dans le Sud-Ouest de la France. *OCL* **22**, D506.
112. Reckling, M., Döring, T.F., Stein-Bachinger, K., Bachinger, J., 2015. Yield stability of grain legumes in an organically managed monitoring experiment. *Asp. Appl. Biol.* **128**, 57-62.
113. Reckling, M., Bergkvist, G., Watson, C.A., Stoddard, F.L., Zander, P.M., Walker, R.L., Pristeri, A., Toncea, I., Bachinger, J., 2016a. Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. *Front. Plant Sci.* **7**, 669.
114. Reckling, M., Hecker, J.-M., Bergkvist, G., Watson, C.A., Zander, P., Schläfke, N., Stoddard, F.L., Eory, V., Topp, C.F.E., Maire, J., Bachinger, J., 2016b. A cropping system assessment framework – evaluating effects of introducing legumes into crop rotations. *Eur. J. Agron.* **76**, 186-197.
115. Recknagel, J., 2015. Le soja local en Allemagne, une demande soutenue, une offre qui s'étoffe. *OCL* **22**, D508.
116. Rööß, E., Bajželj, B., Smith, P., Patel, M., Little, D., Garnett, T., 2016. Protein futures for western Europe: potential land use and climate impacts in 2050. *Reg. Environ. Change.* **1013**, 1–11.
117. Rubiales, D., Fondevilla, S., Chen, W., Gentzbittel, L., Higgins, T.J.V., Castillejo, M.A., Singh, K.B., Rispaill, N., 2015. Achievements and challenges in legume breeding for pest and disease resistance. *Crit. Rev. Plant Sci.* **34**, 195–236.

118. Ruisi, P., Giambalvo, D., Di Miceli, G., Frenda, A.S., Saia, S., Amato, G., 2012. Tillage effects on yield and nitrogen fixation of legumes in Mediterranean conditions. *Agron. J.* **104**, 1459–1466.
119. Schilizzi, S.G.M., Kingwell, R.S., 1999. Effects of climatic and price uncertainty on the value of legume crops in a Mediterranean-type environment. *Agr. Syst.* **60**, 55–69.
120. Schläfke, N., Zander, P., Reckling, M., Hecker, J.M., Bachinger, J., 2014. Evaluation of legume-supported agriculture and policies at farm level. Legume Futures Report 4.3. European Commission. [http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_4.3.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_4.3.pdf). Accès en Avril 2016.
121. Schmeer, M., Loges, R., Dittert, K., Senbayram, M., Horn, R., Taube, F., 2014. Legume-based forage production systems reduce nitrous oxide emissions. *Soil Tillage Res.* **143**, 17–25.
122. Schneider, A., *et al.* 2015. Les légumineuses pour des systèmes agricoles et alimentaires durables. Quæ, Versailles.
123. Schreuder, R., De Visser, C., 2014. EIP-AGRI Focus Group Protein Crops. Final Report. European Commission. <http://ec.europa.eu/eip/agriculture/en/content/focus-groups>. Accès en Septembre 2014.
124. Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. *Nature* **485**, 229–232.
125. Seymour, M., Kirkegaard, J.A., Peoples, M.B., White, P.F., French, R.J., 2012. Break-crop benefits to wheat in western Australia – insights from over three decades of research. *Crop Pasture Sci.* **63**, 1–16.
126. Shi, G., Chavas, J.-P., Lauer, J., 2013. Commercialized transgenic traits, maize productivity and yield risk. *Nat. Biotech.* **31**, 111–114.
127. Siddique, K.H.M., Walton, G., Seymour, M., 1993. A comparison of seed yields of winter grain legumes in western Australia. *Aust. J. Exp. Agr.* **33**, 915–922.
128. Siddique, K.H.M., Sykes, J., 1997. Pulse production in Australia past, present and future. *Aust. J. Exp. Agr.* **37**, 103–111.
129. Siddique, K.H.M., Loss, S.P., 1999. Studies on sowing depth for chickpea (*Cicer arietinum* L.), faba bean (*Vicia faba* L.) and lentil (*Lens culinaris* Medik) in a Mediterranean-type environment of south-western Australia. *J. Agron. Crop Sci.* **182**, 105–112.
130. Siddique, K.H.M., Erskine, W., Hobson, K., Knights, E.J., Leonforte, A., Khan, T.N., Paull, J.G., Redden, R., Materne, M., 2013. Cool-season grain legume improvement in Australia – use of genetic resources. *Crop Pasture Sci.* **64**, 347–360.
131. Silsbury, J., 1990. Grain yield of wheat in rotation with pea, vetch or medic grown with three systems of management. *Aust. J. Exp. Agr.* **30**, 645–649.

132. Sinclair, T.R., Vadez, V., 2012. The future of grain legumes in cropping systems. *Crop Pasture Sci.* **63**, 501–512.
133. Singh, U., 1988. Antinutritional factors of chickpea and pigeonpea and their removal by processing. *Plant Foods Hum. Nutr.* **38**, 251–261.
134. Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W., Berger, J., Greene, S.L., Nelson, M.N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R.K., Roorkiwal, M., Pandey, M.K., Hu, J., Li, Y.H., Wang, L.X., Guo, Y., Qiu, L.J., Redden, R.J., Varshney, R.K., 2015. Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* **34**, 43–104.
135. Stoddard, F.L., 2013. The case studies of participant expertise in Legume Futures. Legume Futures Report 1.2. European Commission.  
[http://www.legumefutures.de/images/Legume\\_Futures\\_Report\\_1.2.pdf](http://www.legumefutures.de/images/Legume_Futures_Report_1.2.pdf). Accès en Septembre 2014.
136. Tanaka, D.L., Krupinsky, J.M., Liebig, M.A., Merrill, S.D., Ries, R.E., Hendrickson, J.R., Johnson, H.A., Hanson, J.D., 2002. Dynamic cropping systems: an adaptable approach to crop production in the Great Plains. *Agron. J.* **94**, 957–961.
137. Tanaka, D.L., Krupinsky, J.M., Merrill, S.D., Liebig, M.A., Hanson, J.D., 2007. Dynamic cropping systems for sustainable crop production in the northern Great Plains. *Agron. J.* **99**, 904–911.
138. Temba, M.C., Njobeh, P.B., Adebo, O.A., Olugbile, A.O., Kayitesi, E., 2016. The role of compositing cereals with legumes to alleviate protein energy malnutrition in Africa. *Int. J. Food Sci. Tech.* **51**, 543–554
139. Tharanathan, R.N., Mahadevamma, S., 2003. Grain legumes – a boon to human nutrition. *Trends Food Sci. Tech.* **14**, 507–518.
140. Thomson, B.D., Siddique, K.H.M., Barr, M.D., Wilson, J.M., 1997. Grain legume species in low rainfall Mediterranean-type environments I. Phenology and seed yield. *Field Crop. Res.* **54**, 173–187.
141. Vadez, V., Berger, J.D., Warkentin, T., Asseng, S., Ratnakumar, P., Rao, K.P.C., Gaur, P.M., Munier-Jolain, N., Larmure, A., Voisin, A.-S., Sharma, H.C., Pande, S., Sharma, M., Krishnamurthy, L., Zaman, M.A., 2012. Adaptation of grain legumes to climate change: a review. *Agron. Sustain. Dev.* **32**, 31–44.
142. Van Krimpen, M.M., Bikker, P., Van Der Meer, I.M., Van Der Peet-Schwering, C.M.C., Vereijken, J.M., 2013. Cultivation, processing and nutritional aspects for pigs and poultry of European protein sources as alternatives for imported soybean products. Report. **662**, Wageningen UR Livestock Research and Central Veterinary Institute of Wageningen UR.
143. Vaz Patta, M.C., Rubiales, D., 2014. Lathyrus diversity: available resources with relevance to crop improvement – *L. sativus* and *L. cicera* as case studies. *Ann. Bot.* **113**, 895–908.

144. Vaz Patto, M.C., Amarowicz, R., Aryee, A.N.A., Boye, J.I., Chung, H.-J., Martín-Cabrejas, M.A., Domoney, C., 2015. Achievements and challenges in improving the nutritional quality of food legumes. *Crit. Rev. Plant Sci.* **34**, 105–143.
145. Voisin, A.-S., Guéguen, J., Huyghe, C., Jeuffroy, M.-H., Magrini, M.-B., Meynard, J.-M., Mougél, C., Pellerin, S., Pelzer, E., 2014. Legumes for feed, food, biomaterials and bioenergy in Europe: a review. *Agron. Sustain. Dev.* **34**, 361–380.
146. Von Richthofen, J.S., Pahl, H., Bouttet, D., Casta, P., Cartryse, C., Charles, R., Lafarga, A., 2006. What do European farmers think about grain legumes? *Grain Legumes* **45**, 14–15.
147. Weiss, F., Leip, A., 2012. Greenhouse gas emissions from the EU livestock sector: a life cycle assessment carried out with the CAPRI model. *Agr. Ecosyst. Environ.* **149**, 124–134.
148. Westhoek, H., Lesschen, J.P., Rood, T., Wagner, S., De Marco, A., Murphy-Bokern, D., Leip, A., Van Grinsven, H., Sutton, M.A., Oenema, O., 2014. Food choices, health and environment: Effects of cutting Europe's meat and dairy intake. *Glob. Environ. Change.* **26**, 196–205.
149. Wiggering, H., Finckh, M., Heß, J., Wehling, P., Michaelis, T., 2012. The Legumes Expert Forum. Science, economy and society – making ecosystem services from legumes competitive A research strategy of the German Agricultural Research Alliance. German Agricultural Research Alliance, Braunschweig. <http://kore.dekodes.no/the-legumes-expert-forum-science-economy-and-society-making-ecosystem-services-from-legumes-competitive-a-research-strategy-of-the-german-agricultural-research-alliance>. Accès en Avril 2016.
150. Williams, C.M., King, J.R., Ross, S.M., Olson, M.A., Hoy, C.F., Lopetinsky, K.J., 2014. Effects of three pulse crops on subsequent barley, canola, and wheat. *Agron. J.* **106**, 343–350.
151. Yahara, T., Javadi, F., Onoda, Y., de Queiroz, L.P., Faith, D.P., Prado, D.E., Akasaka, M., Kadoya, T., Ishihama, F., Davies, S., Slik, J.W.F., Yi, T., Ma, K., Bin, C., Darnaedi, D., Pennington, R.T., Tuda, M., Shimada, M., Ito, M., Egan, A.N., Buerki, S., Raes, N., Kajita, T., Vatanparast, M., Mimura, M., Tachida, H., Iwasa, Y., Smith, G.F., Victor, J.E., Nkonki, T., 2013. Global legume diversity assessment : concepts, key indicators, and strategies. *Taxon* **62**, 249–266.
152. Yan, Z.-Y., Spencer, P.S., Li, Z.-X., Liang, Y.-M., Wang, Y.-F., Wang, C.-Y., Li, F.-M., 2006. *Lathyrus sativus* (grass pea) and its neurotoxin ODAP. *Phytochemistry* **67**, 107–121.
153. Yuan, Y., Hunt, R.H., 2009. Systematic reviews: the good, the bad, and the ugly. *Am. J. Gastroenterol.* **104**, 1086–1092.
154. Zander, P., Amjath-Babu, T.S., Preissel, S., Reckling, M., Bues, A., Schläfke, N., Kuhlman, T., Bachinger, J., Uthes, S., Stoddard, F., Murphy-Bokern, D., Watson, C., 2016. Grain legume decline and potential recovery in European agriculture: a review. *Agron. Sustain. Dev.* **36**, 1–20.

155. Zentner, R.P., Wall, D.D., Nagy, C.N., Smith, E.G., Young, D.L., Miller, P.R., Campbell, C.A., McConkey, B.G., Brandt, S.A., Lafond, G.P., Johnston, A.M., Derksen, D.A., 2002. Economics of crop diversification and soil tillage opportunities in the Canadian Prairies. *Agron. J.* **94**, 216–230.
156. Zimmer, S., Liebe, U., Didier, J.-P., Heß, J., 2016a. Luxembourgish farmers' lack of information about grain legume cultivation. *Agron. Sustain. Dev.* **36**, 1–10.
157. Zimmer, S., Messmer, M., Haase, T., Piepho, H.-P., Mindermann, A., Schulz, H., Habekuß, A., Ordon, F., Wilbois, K.-P., Heß, J., 2016b. Effects of soybean variety and Bradyrhizobium strains on yield, protein content and biological nitrogen fixation under cool growing conditions in Germany. *Eur. J. Agron.* **72**, 38–46.



# Annexes

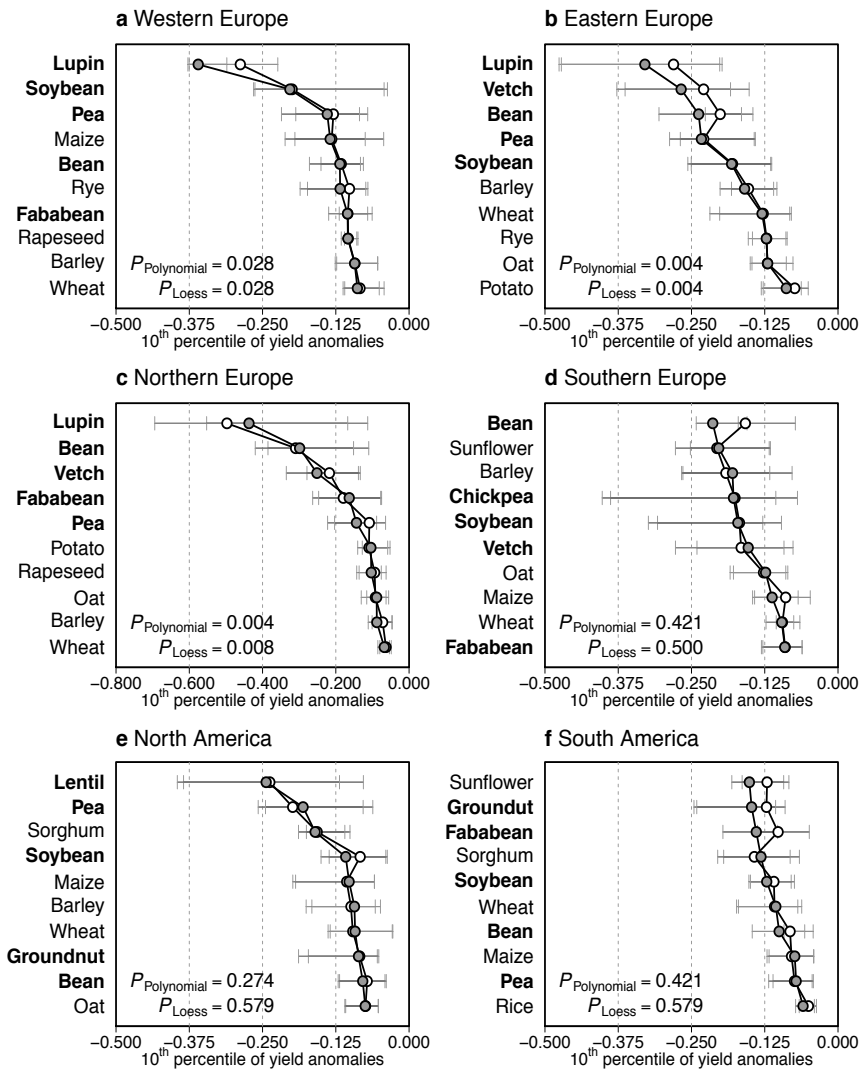




# **Annexes du Chapitre 1**



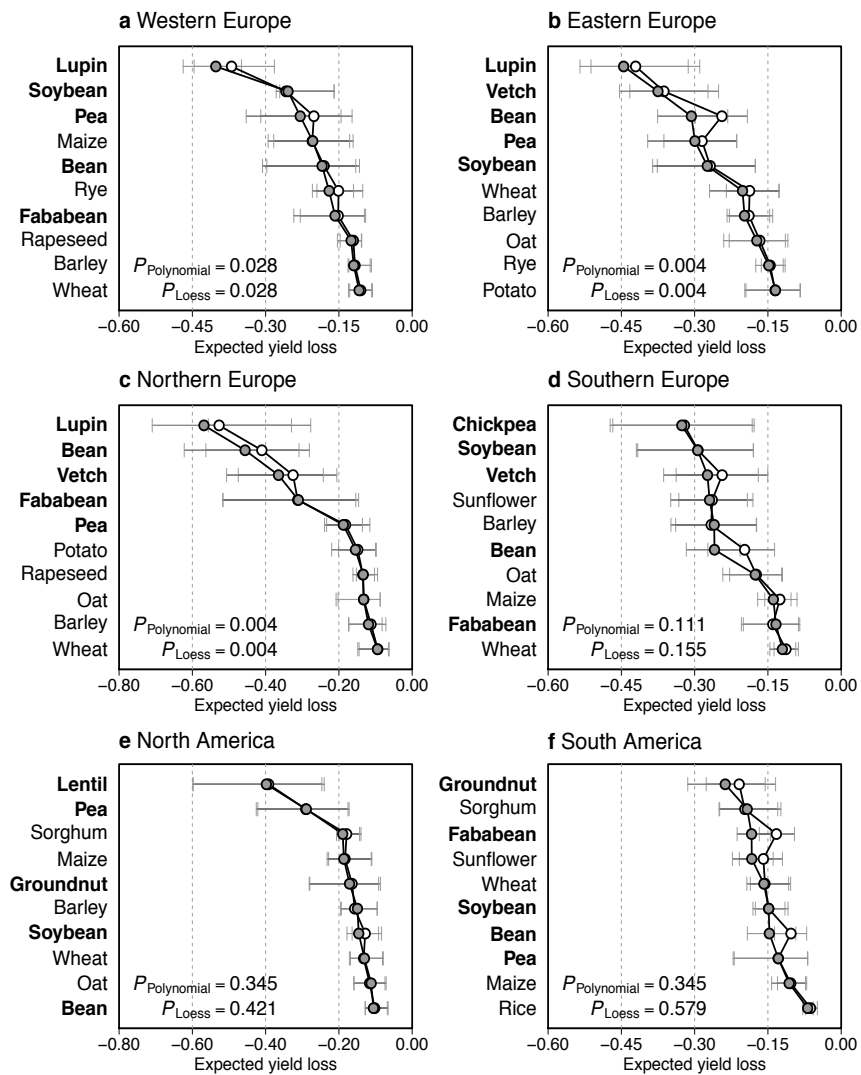




Estimating variability in grain legume yields across Europe and the Americas  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 1: 10<sup>th</sup> percentile of yield anomalies for 10 crops in Europe and the Americas over 1961-2013.**

Values of 10<sup>th</sup> percentile of yield anomalies for 10 crops in four European and two American regions over 1961-2013. Polynomial (grey points) and local 'loess' (empty points) regression models are used to calculate yield anomalies (relative differences between yield data and yield trend). Horizontal lines correspond to 95% confidence intervals estimated by bootstrap (10,000 samples). Among the 10 crops, 5 are legume crops (bold names) and 5 are non-legume crops (non-bold). All crops are ranked according to standard deviation of yield anomalies calculated using the polynomial model (decreasing order).  $P_{\text{Polynomial}}$  and  $P_{\text{Loess}}$  correspond to the p-value of the Wilcoxon rank test on a hypothesis assuming that ranks of legume and non-legume crops do not differ (against the alternative that legume crop ranking is lower) and are computed using the polynomial and loess regression respectively. The number of yield data includes 53 observations for most of crops and regions. There are two exceptions: soybean yield data in Western Europe includes 41 observations and lupin yield data in Northern Europe includes 22 observations. Grey vertical dashed lines represent the 10<sup>th</sup> percentile values of -0.375, -0.250, -0.125 in all regions but -0.600, -0.400 and -0.200 in Northern Europe.

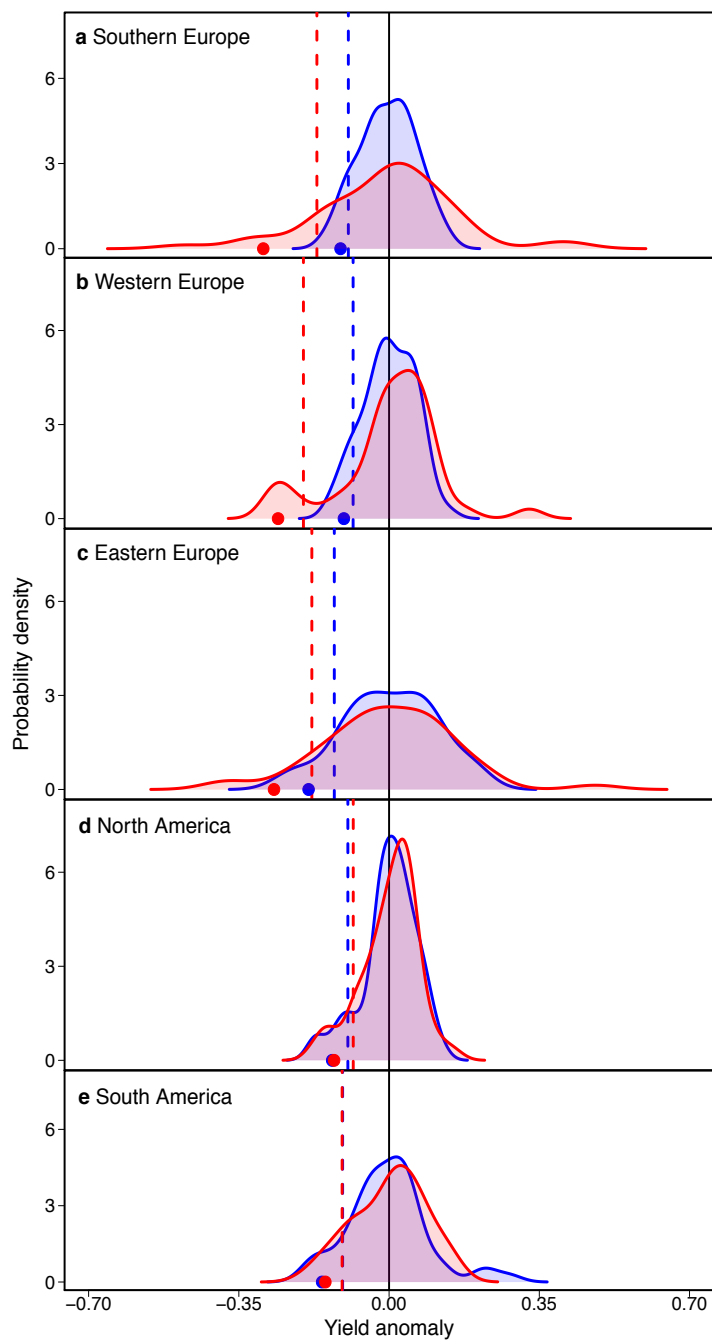


Estimating variability in grain legume yields across Europe and the Americas  
Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 2: Expected yield loss for 10 crops in Europe and the Americas over 1961-2013.**

Expected yield loss (mean values of yield anomalies lower than the 10<sup>th</sup> percentile) for 10 crops in four European and two American regions over 1961-2013. Polynomial (grey points) and local 'loess' (empty points) regression models are used to calculate yield anomalies (relative differences between yield data and yield trend). Horizontal lines correspond to 95% confidence intervals estimated by bootstrap (10,000 samples). Among the 10 crops, 5 are legume crops (bold names) and 5 are non-legume crops (non-bold). All crops are ranked according to standard deviation of yield anomalies calculated using the polynomial model (decreasing order).  $P_{\text{Polynomial}}$  and  $P_{\text{Loess}}$  correspond to the p-value of the Wilcoxon rank test on a hypothesis assuming that ranks of legume and non-legume crops do not differ (against the alternative that legume crop ranking is lower) and are computed using the polynomial and loess regression respectively. The number of yield data includes 53 observations for most of crops and regions. There are two exceptions: soybean yield data in Western Europe includes 41 observations and lupin yield data in Northern Europe includes 22 observations. Grey vertical dashed lines represent the values of expected yield loss of -0.45, -0.30, -0.15 in all regions but -0.60, -0.40 and -0.20 in Northern Europe and North America.



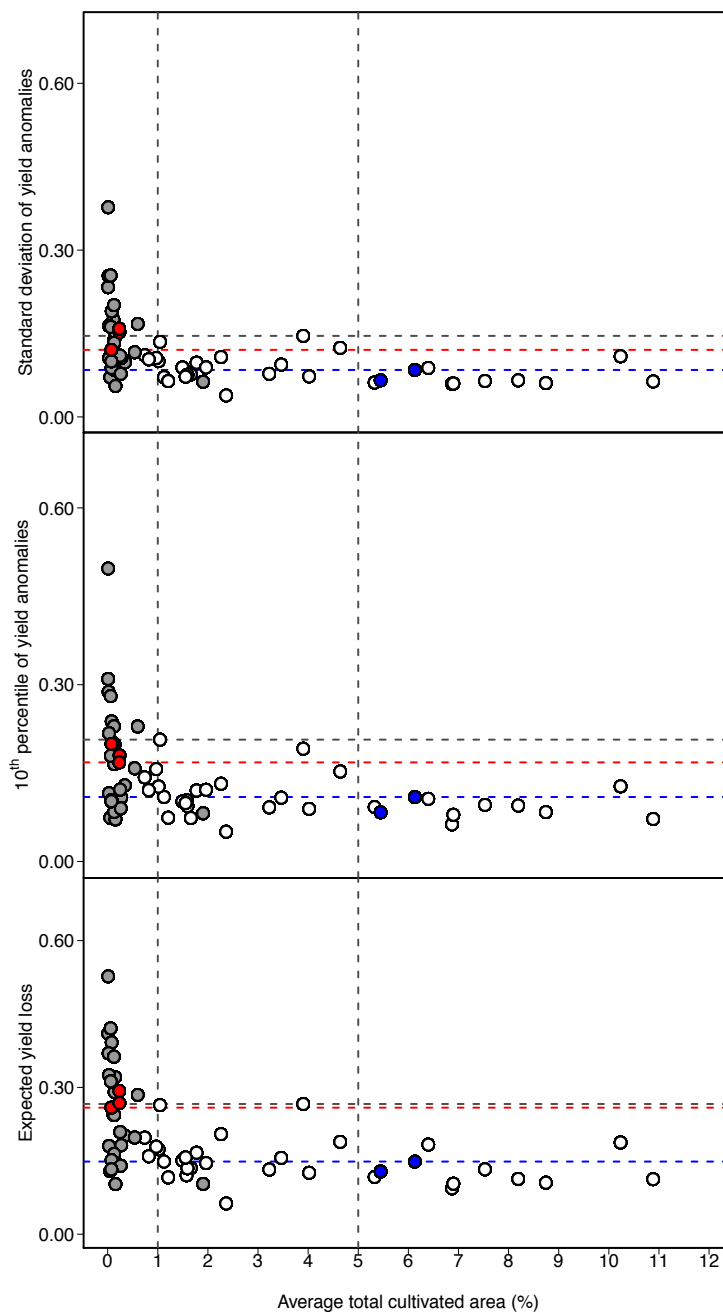


**Estimating variability in grain legume yields across Europe and the Americas**  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

6

**Supplementary Figure 3: Yield anomaly distributions, 10<sup>th</sup> percentile and expected yield loss for wheat and soybean in Europe and the Americas over 1961-2013 using the loess regression.**

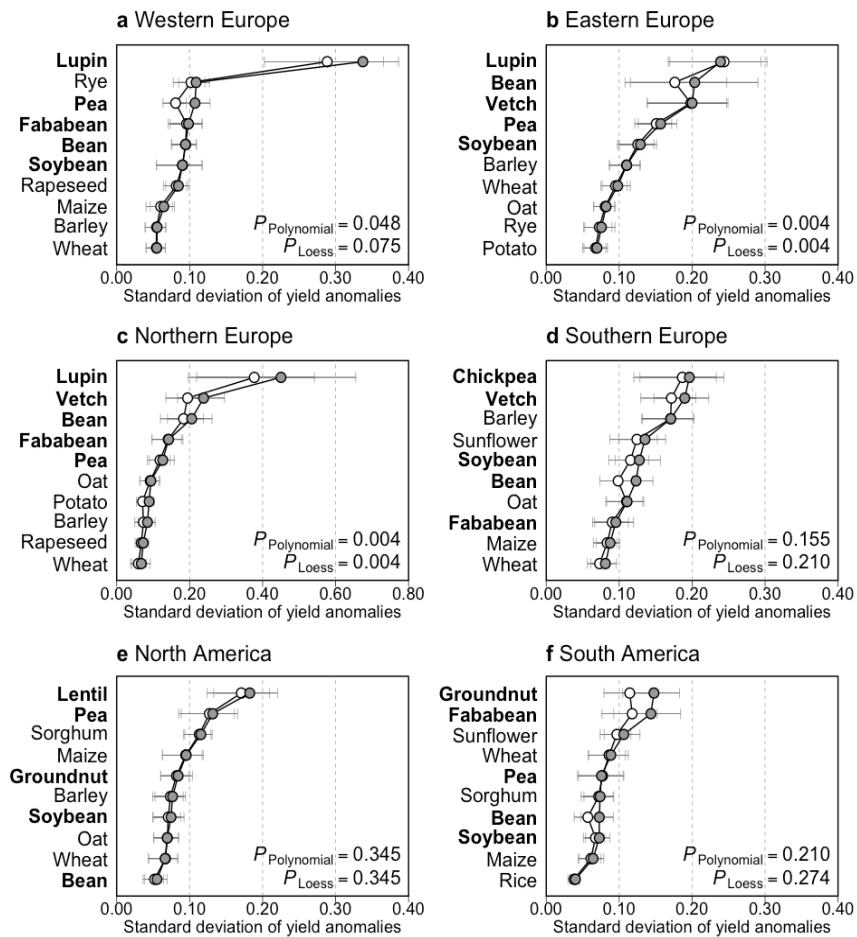
Probability densities of yield anomalies (curves), 10<sup>th</sup> percentiles of yield anomalies (vertical dashed lines), and expected yield losses (mean values of yield anomalies lower than 10% percentiles; points) for wheat (in blue) and soybean (in red) in three European regions and two American regions. Yield anomalies (relative differences between yield data and yield trend) are calculated using the loess regression model over 1961-2013. Northern Europe is excluded due to insufficient soybean yield data. Regions are ranked according to the variance ratio of soybean yield anomalies compared to wheat (i.e., Southern Europe shows the highest soybean variance ratio compared to wheat). Probability densities are estimated using a Gaussian smoothing kernel.



**Estimating variability in grain legume yields across Europe and the Americas**  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 4: Yield risk measures as a function of the percentage of total cultivated areas in Europe and the Americas over 1961-2013 using the loess regression.**

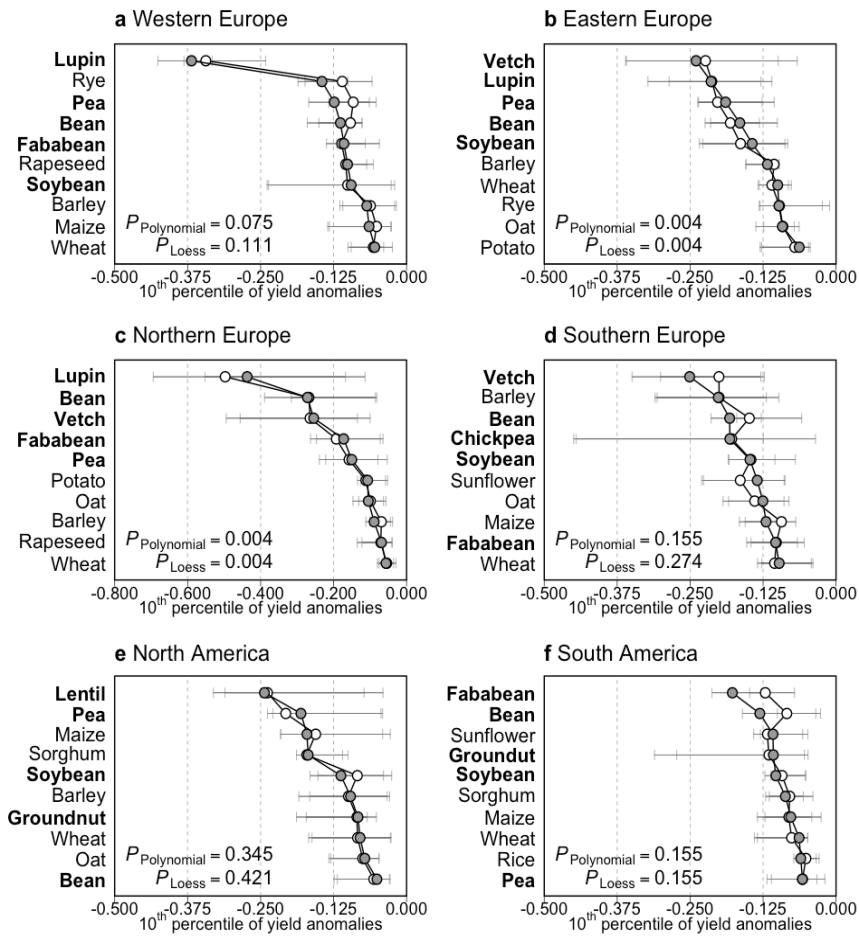
Risk measures are calculated over 1961-2013: standard deviation (a), 10<sup>th</sup> percentile (absolute values, b) and expected yield loss (absolute values, c) for all crop\*region combinations. Crop\*region combinations corresponding to legume and non-legume crops are indicated by grey points and empty points, respectively. The percentages reported in the x-axis correspond to the percentage of total cultivated areas by a given crop in a given region over 1961-2013. Soybeans grown in the Americas and Europe are indicated in blue and red points. Blue dashed horizontal lines represent the maximal values of risk measures for soybean grown in the Americas. Red dashed horizontal lines represent the minimal values of risk measures for soybean grown in Europe. Grey horizontal dashed lines represent the maximal values of risk measures for non-legumes. Grey vertical dashed lines represent the thresholds of 1% and 5% of total cultivated areas. Yield anomalies are computed using the loess regression.



Estimating variability in grain legume yields across Europe and the Americas  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 5: Standard deviation of yield anomalies for 10 crops in Europe and the Americas over 1983-2013.**

Standard deviation of yield anomalies for 10 crops in four European and two American regions over 1983-2013. Polynomial (grey points) and local 'loess' (empty points) regression models are used to calculate yield anomalies (relative differences between yield data and yield trend). Horizontal lines correspond to 95% confidence intervals estimated by bootstrap (10,000 samples). Among the 10 crops, 5 are legume crops (bold names) and 5 are non-legume crops (non-bold). All crops are ranked according to standard deviation of yield anomalies calculated using the polynomial model (decreasing order).  $P_{\text{Polynomial}}$  and  $P_{\text{Loess}}$  correspond to the p-value of the Wilcoxon rank test on a hypothesis assuming that ranks of legume and non-legume crops do not differ (against the alternative that legume crop ranking is lower) and are computed the polynomial and loess regression respectively. The number of yield data includes 31 observations for all crops and regions but lupin yield data in Northern Europe includes 22 observations. Grey vertical dashed lines represent the standard deviation values of 0.10, 0.20 and 0.30 in all regions but 0.20, 0.40 and 0.60 in Northern Europe.

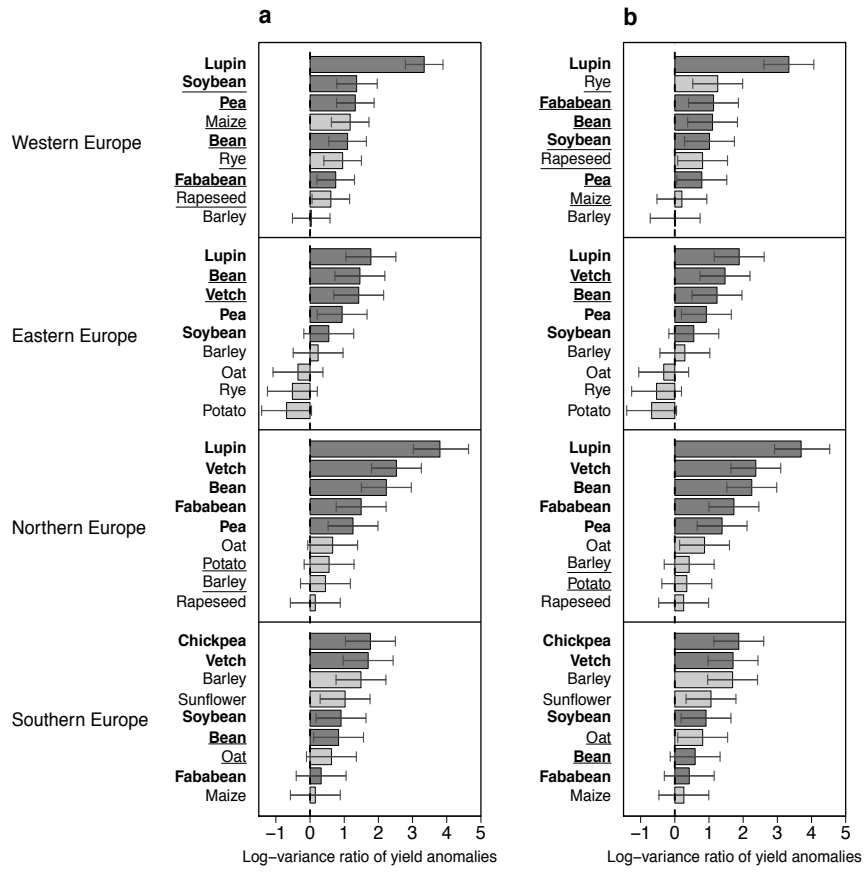


Estimating variability in grain legume yields across Europe and the Americas  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 6: 10<sup>th</sup> percentile of yield anomalies for 10 crops in Europe and the Americas over 1983-2013.**

Values of 10<sup>th</sup> percentile of yield anomalies for 10 crops in four European and two American regions over 1983-2013. Polynomial (grey points) and local 'loess' (empty points) regression models are used to calculate yield anomalies (relative differences between yield data and yield trend). Horizontal lines correspond to 95% confidence intervals estimated by bootstrap (10,000 samples). Among the 10 crops, 5 are legume crops (bold names) and 5 are non-legume crops (non-bold). All crops are ranked according to standard deviation of yield anomalies calculated using the polynomial model (decreasing order).  $P_{\text{Polynomial}}$  and  $P_{\text{Loess}}$  correspond to the p-value of the Wilcoxon rank test on a hypothesis assuming that ranks of legume and non-legume crops do not differ (against the alternative that legume crop ranking is lower) and are computed using the polynomial and loess regression respectively. The number of yield data includes 31 observations for all crops and regions but lupin yield data in Northern Europe includes 22 observations. Grey vertical dashed lines represent the standard deviation values of 0.10, 0.20 and 0.30 in all regions but 0.20, 0.40 and 0.60 in Northern Europe.

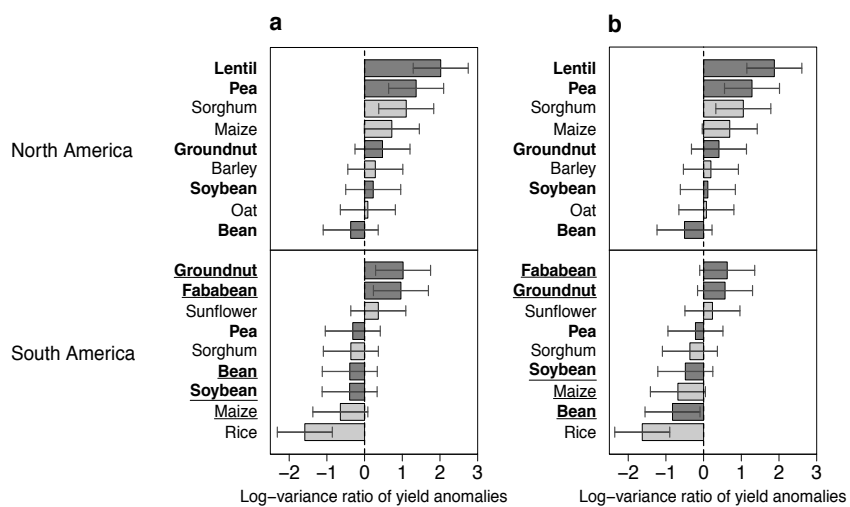




Estimating variability in grain legume yields across Europe and the Americas  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

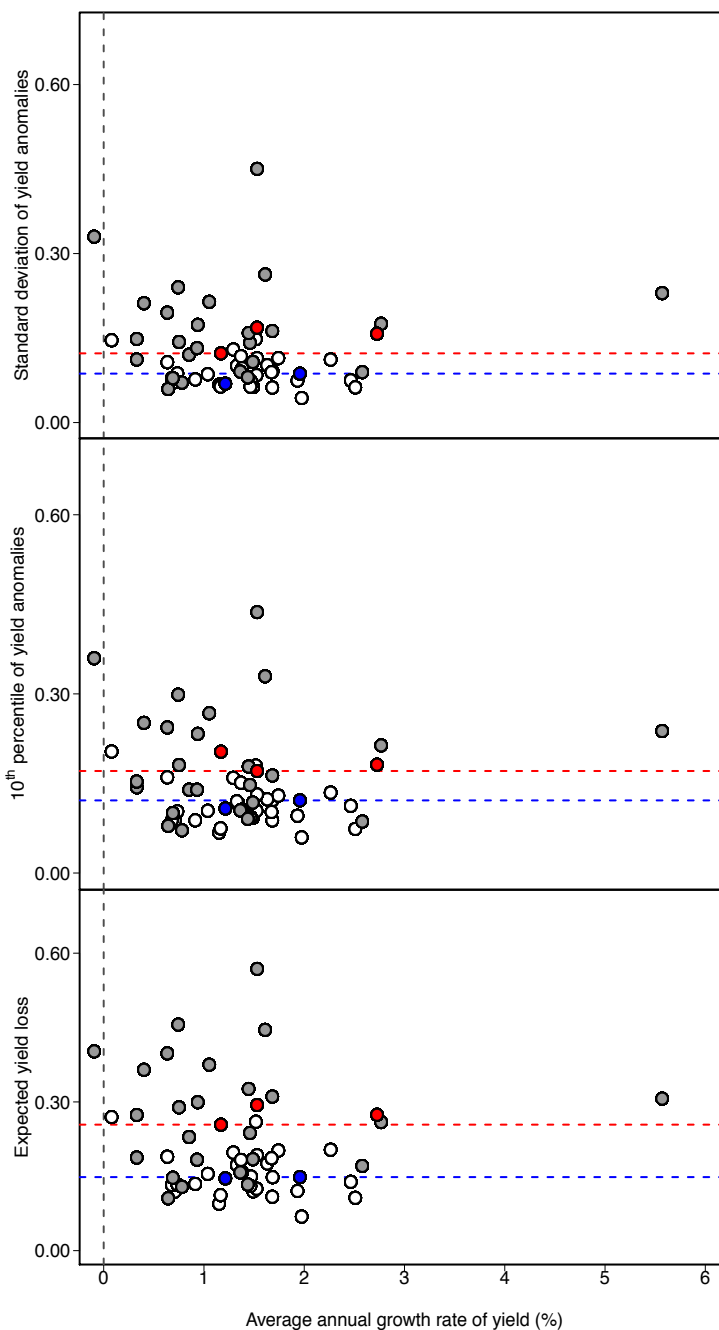
**Supplementary Figure 7: Log-variance ratio of yield anomalies for 9 crops compared to wheat in four European regions over 1983-2013.**

Variances are calculated over 1983-2013 and log-transformed. Yield anomalies are calculated using both regression models (polynomial (a), loess (b)). Horizontal lines correspond to 95% confidence intervals. Among the 9 crops, 5 are legume crops (bold names and dark grey bars) and 4 are non-legume crops (non-bold names and light grey bars). All crops are ranked according to variance ratio values (decreasing order). Crops that change of ranking between polynomial and loess regression models used are underlined.



**Supplementary Figure 8: Log-variance ratio of yield anomalies for 9 crops compared to wheat in two American regions over 1983-2013.**

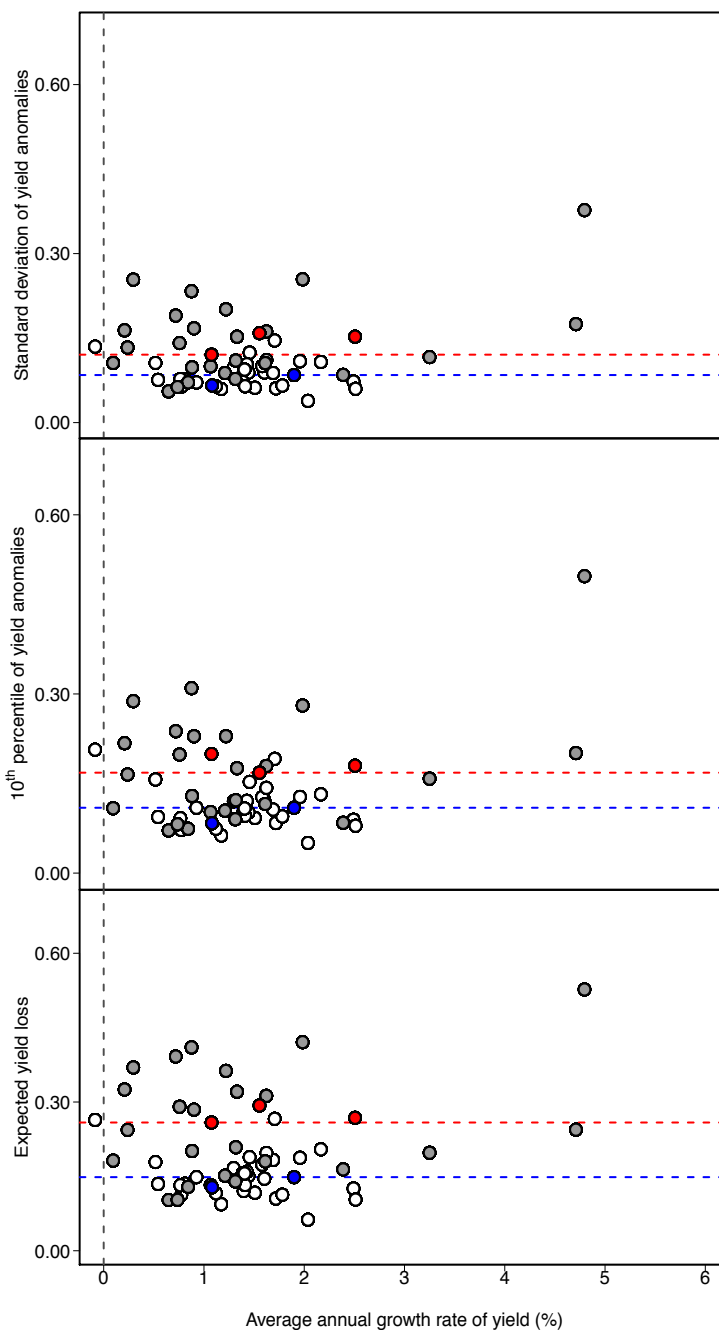
Variances are calculated over 1983-2013 and log-transformed. Yield anomalies are calculated using polynomial regression models (polynomial (a), loess (b)). Horizontal bars correspond to 95% confidence intervals. Among the 9 crops, 5 are legume crops (bold names and dark grey bars) and 4 are non-legume crops (non-bold names and light grey bars). All crops are ranked according to variance ratio values (decreasing order). Crops that change of ranking between polynomial and loess regression models used are underlined.



**Estimating variability in grain legume yields across Europe and the Americas**  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 9: Yield risk measures as a function of the average annual growth rate of yield in Europe and the Americas over 1961-2013 using the polynomial regression.**

Risk measures are calculated over 1961-2013: standard deviation (a), 10<sup>th</sup> percentile (absolute values, b) and expected yield loss (absolute values, c) for all crop\*region combinations. Crop\*region combinations corresponding to legume and non-legume crops are indicated by grey points and empty points, respectively. The percentages reported in the x-axis correspond to average annual growth rate of yield (expressed in percentage) by a given crop in a given region over 1961-2013. Soybeans grown in the Americas and Europe are indicated in blue and red points. Blue dashed horizontal lines represent the maximal values of risk measures for soybean grown in the Americas. Red dashed horizontal lines represent the minimal values of risk measures for soybean grown in Europe. Grey vertical dashed lines represent the thresholds of 0% of average annual growth rate of yield. Yield risk measures and the average annual growth rate of yield are derived from yield anomalies calculated with the polynomial regression.

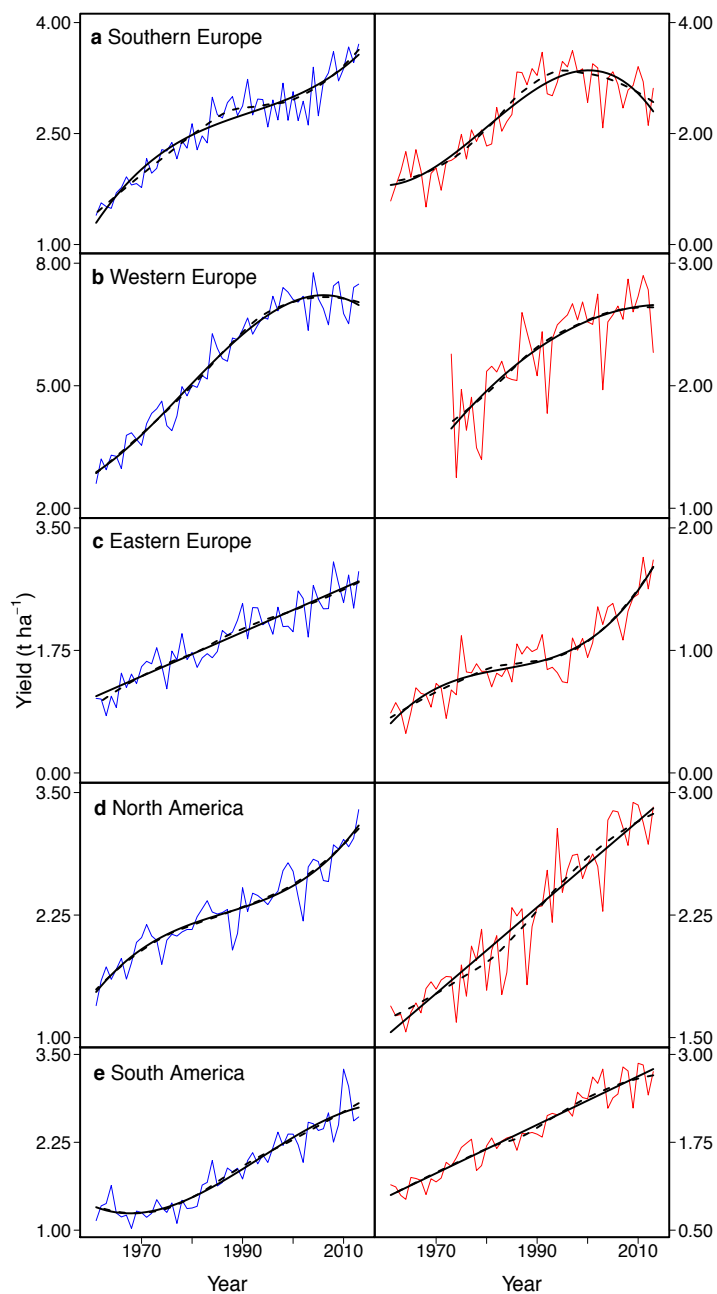


**Estimating variability in grain legume yields across Europe and the Americas**  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 10: Yield risk measures as a function of the average annual growth rate of yield in Europe and the Americas over 1961-2013 using the loess regression.**

Risk measures are calculated over 1961-2013: standard deviation (a), 10<sup>th</sup> percentile (absolute values, b) and expected yield loss (absolute values, c) for all crop\*region combinations. Crop\*region combinations corresponding to legume and non-legume crops are indicated by grey points and empty points, respectively. The percentages reported in the x-axis correspond to average annual growth rate of yield (expressed in percentage) by a given crop in a given region over 1961-2013. Soybeans grown in the Americas and Europe are indicated in blue and red points. Blue dashed horizontal lines represent the maximal values of risk measures for soybean grown in the Americas. Red dashed horizontal lines represent the minimal values of risk measures for soybean grown in Europe. Grey vertical dashed lines represent the thresholds of 0% of average annual growth rate of yield. Yield risk measures and the average annual growth rate of yield are derived from yield anomalies calculated with the loess regression.





**Estimating variability in grain legume yields across Europe and the Americas**  
 Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

**Supplementary Figure 11: Observed and fitted yield times series of wheat and soybean in Europe and the Americas over 1961-2013.**

Observed yield time series for wheat (blue continuous line) and soybean (red continuous line) in three European regions and two American regions over 1961-2013. Polynomial (black continuous line) and local 'loess' (black dashed line) regressions models are fitted to yield time series. Northern Europe is excluded due to insufficient soybean yield data. Regions are ranked according to values of variance ratios of soybean yield anomalies compared to wheat (i.e., Southern Europe shows the highest soybean variance ratio compared to wheat).

Region	Country
Western Europe	Austria
	Belgium
	France
	Germany
	Luxembourg
	Netherlands
Eastern Europe	Switzerland
	Belarus
	Bulgaria
	Czech Republic
	Hungary
	Poland
	Republic of Moldova
	Romania
	Russian Federation
	Slovakia
Ukraine	
Northern Europe	Denmark
	Estonia
	Finland
	Iceland
	Ireland
	Latvia
	Lithuania
	Norway
	Sweden
	United Kingdom of Great Britain and Northern Ireland
Southern Europe	Albania
	Bosnia and Herzegovina
	Croatia
	Greece
	Italy
	Montenegro
	Portugal
	Serbia
	Slovenia
	Spain
The former Yugoslav Republic of Macedonia	
North America	Canada
	United States of America
South America	Argentina
	Bolivia (Plurinational State of)
	Brazil
	Chile
	Colombia
	Ecuador
	French Guiana
	Guyana
	Paraguay
	Peru
	Suriname
Uruguay	
Venezuela (Bolivarian Republic of)	

**Supplementary Table 1: List of countries in the studied regions.**

List of countries in the four European and two American studied regions. Countries are aggregated by the United Nations Statistics Division.

Region	Legumes			Non-legumes		
	Crop	Scientific name	Average cultivated area (10 <sup>6</sup> ha)	Crop	Scientific name	Average cultivated area (10 <sup>6</sup> ha)
Western Europe	Pea	<i>Pisum sativum</i>	53 0.32	Wheat	<i>Triticum</i> spp.	53 7.95
	Fababean	<i>Vicia faba</i>	53 0.07	Barley	<i>Hordeum</i> spp.	53 4.83
	Soybean	<i>Glycine max</i>	41 0.07	Maize	<i>Zea mays</i>	53 2.06
	Bean	<i>Phaseolus</i> spp.	53 0.02	Rapeseed	<i>Brassica napus</i>	53 1.44
	Lupin	<i>Lupinus</i> spp.	53 0.02	Rye	<i>Secale cereale</i>	53 1.34
	Pea	<i>Pisum sativum</i>	53 3.40	Wheat	<i>Triticum</i> spp.	53 54.65
Eastern Europe	Soybean	<i>Glycine max</i>	53 1.13	Barley	<i>Hordeum</i> spp.	53 24.84
	Veitch	<i>Vicia sativa</i>	53 0.72	Rye	<i>Secale cereale</i>	53 10.93
	Bean	<i>Phaseolus</i> spp.	53 0.67	Oat	<i>Avena</i> spp.	53 9.70
	Lupin	<i>Lupinus</i> spp.	53 0.35	Potato	<i>Solanum tuberosum</i>	53 8.86
	Pea	<i>Pisum sativum</i>	53 0.12	Barley	<i>Hordeum</i> spp.	53 4.37
	Fababean	<i>Vicia faba</i>	53 0.03	Wheat	<i>Triticum</i> spp.	53 2.85
Northern Europe	Veitch	<i>Vicia sativa</i>	53 0.01	Oat	<i>Avena</i> spp.	53 1.27
	Lupin	<i>Lupinus</i> spp.	22 0.00	Rapeseed	<i>Brassica napus</i>	53 0.67
	Bean	<i>Phaseolus</i> spp.	53 0.00	Potato	<i>Solanum tuberosum</i>	53 0.45
	Bean	<i>Phaseolus</i> spp.	53 0.59	Wheat	<i>Triticum</i> spp.	53 8.52
	Fababean	<i>Vicia faba</i>	53 0.28	Maize	<i>Zea mays</i>	53 4.13
	Soybean	<i>Glycine max</i>	53 0.23	Barley	<i>Hordeum</i> spp.	53 3.99
Southern Europe	Chickpea	<i>Cicer arietinum</i>	53 0.16	Oat	<i>Avena</i> spp.	53 1.05
	Veitch	<i>Vicia sativa</i>	53 0.13	Sunflower	<i>Helianthus annuus</i>	53 1.05
	Soybean	<i>Glycine max</i>	53 24.45	Wheat	<i>Triticum</i> spp.	53 33.80
North America	Bean	<i>Phaseolus</i> spp.	53 0.70	Maize	<i>Zea mays</i>	53 28.58
	Pea	<i>Pisum sativum</i>	53 0.61	Barley	<i>Hordeum</i> spp.	53 6.98
	Groundnut	<i>Arachis hypogaea</i>	53 0.59	Oat	<i>Avena</i> spp.	53 5.28
	Lentil	<i>Lens culinaris</i>	53 0.35	Sorghum	<i>Sorghum</i> spp.	53 4.33
	Soybean	<i>Glycine max</i>	53 17.56	Maize	<i>Zea mays</i>	53 17.24
	Bean	<i>Phaseolus</i> spp.	53 4.75	Wheat	<i>Triticum</i> spp.	53 8.56
South America	Groundnut	<i>Arachis hypogaea</i>	53 0.57	Rice	<i>Oryza</i> spp.	53 5.83
	Fababean	<i>Vicia faba</i>	53 0.17	Sunflower	<i>Helianthus annuus</i>	53 2.11
	Pea	<i>Pisum sativum</i>	53 0.13	Sorghum	<i>Sorghum</i> spp.	53 1.84

Estimating variability in grain legume yields across Europe and the Americas  
Charles Cernay, Tamara Ben-Ari, Elise Pelzer, Jean-Marc Meynard & David Makowski

26

**Supplementary Table 2: List of studied crops and their average cultivated areas over 1961-2013.**

Five major legume and five major non-legume crops and their average cultivated area in four European and two American regions over the 1961-2013 period. Crops are ranked according to the values of average cultivated area (decreasing order).  $n$  is the number of years of yield data available for each crop in each region.



## **Annexes du Chapitre 2**





**Table 3.** Tables and attributes included in the dataset. The number of attributes, their names, types and definitions are presented for each table.

Table number	Table name	Attribute number	Attribute name	Attribute type	Attribute definition
1	Literature_Search	1	IDEQ	Index	Index of the step of literature search. Primary key of the 'Literature_Search' table.
1	Literature_Search	2	Literature_Search_Origin	Class	Step of the literature search.
2	Article	3	identifiant	Index	Index of each article from each step of the literature search. Primary key of the 'Article' table.
2	Article	4	Article_Author_First	Class	Name of the first author.
2	Article	5	Article_Title	Class	Article title.
2	Article	6	Article_Year_Publication	Date	Publication year or 'NA'.
2	Article	7	Article_Journal	Class	Journal name or 'NA'.
2	Article	8	Article_Volume	Class	Journal volume or 'NA'.
2	Article	9	Article_Page	Class	First and last journal pages ('First journal page-Last journal page') or 'NA'.
2	Article	10	IDEQ_Equation	Index	Corresponding index from the 'Literature_Search' table. Secondary key of the 'Article' table.
3	Site	11	IDSite	Index	Index of each site from each article. Primary key of the 'Site' table.
3	Site	12	Site_Name	Class	Site name.
3	Site	13	Site_Country	Class	Site country.

3	Site	14	Site_City_State_Region	Class	City, state and/or region name(s) where the site is precisely located or the nearest located or 'NA'.
3	Site	15	Site_Latitude	Numerical	Site latitude coordinate.
3	Site	16	Site_Longitude	Numerical	Site longitude coordinate.
3	Site	17	Site_Coordinate_Source	Class	Source of latitude and longitude coordinates. The source is from the article when latitude and longitude coordinates are originally reported or from the National Aeronautics and Space Administration (NASA) finder ( <a href="http://mydasdata.larc.nasa.gov/latitude-longitude-finder">http://mydasdata.larc.nasa.gov/latitude-longitude-finder</a> ) when coordinates are not originally reported.
3	Site	18	Site_Soil_Depth_Variable_m	Numerical	Soil depth layer (in m) at which soil attributes are determined or 'NA'. When soil attributes are reported for many soil depth layers, soil attributes are averaged over many soil depth layers.
3	Site	19	Site_Soil_Classification_Name	Class	Soil classification name(s) or 'NA'.
3	Site	20	Site_Soil_Texture_Name	Class	Soil texture name(s) or 'NA'.
3	Site	21	Site_Soil_Sand_Percentage	Numerical	Soil average percentage of sand or 'NA'. When many percentages of sand are reported for a given soil depth layer, all percentages of sand are added for the given soil depth layer.
3	Site	22	Site_Soil_Silt_Percentage	Numerical	Soil average percentage of silt or 'NA'.
3	Site	23	Site_Soil_Clay_Percentage	Numerical	Soil average percentage of clay or 'NA'.
3	Site	24	Site_Soil_pH	Numerical	Soil average pH or 'NA'.
3	Site	25	Site_Soil_pH_Basis	Class	Soil chemical basis at which soil average pH ('Ca'/CaCl <sub>2</sub> /'H <sub>2</sub> O'/'KCl') is determined or 'NA'.

3	Site	26	Site_Soil_Organic_Matter_Percentage	Numerical	Soil average percentage of organic matter or 'NA'.
3	Site	27	Site_Soil_N_Percentage	Numerical	Soil average percentage of nitrogen (N) or 'NA'.
3	Site	28	Site_Soil_N_Percentage_Type	Class	Type of soil average percentage of nitrogen ('Total'/'Organic') or 'NA'. This attribute may be completed with soil nitrogen quantity or soil nitrogen percentage at seeding from the 'Crop_N_Soil_Quantity_Percentage_Seeding' attribute or at harvest from the 'Crop_N_Soil_Quantity_Percentage_Harvest' attribute.
3	Site	29	Site_Soil_N_Percentage	Numerical	Soil average percentage of nitrogen (N) or 'NA'.
3	Site	29	Site_Soil_C_Percentage	Numerical	Soil average percentage of carbon (C) or 'NA'.
3	Site	30	Site_Soil_C_Percentage_Type	Class	Type of soil average percentage of carbon ('Total'/'Organic') or 'NA'.
3	Site	31	Site_Precipitation_mm	Numerical	Site average precipitation (in mm) or 'NA'.
3	Site	32	Site_Precipitation_Period	Class	Period ('Annual'/'Growing season') at which site average precipitation is determined or 'NA'.
3	Site	33	Site_Precipitation_Period_Month	Date	First and last calendar months of the period ('First month-Last month') at which site average precipitation is determined or 'NA'. Months are abbreviated: 'Jan.': January, 'Feb.': February, 'Mar.': March, 'Apr.': April, 'May': May, 'Jun.': June, 'Jul.': July, 'Aug.': August, 'Sep.': September, 'Oct.': October, 'Nov.': November, 'Dec.': December.
3	Site	34	Site_Precipitation_Period_Year	Date	First and last calendar years of the period ('First year-Last year') at which site average precipitation is determined or 'NA'.
3	Site	35	Site_Temperature_Celsius	Numerical	Site average temperature (in Celsius) or 'NA'.

3	Site	36	Site_Temperature_Period	Class	Period ('Annual'/'Growing season') at which site average temperature is determined or 'NA'.
3	Site	37	Site_Temperature_Period_Month	Date	First and last calendar months of the period ('First month-Last month') at which site average temperature is determined or 'NA'. Months are abbreviated: 'Jan.': January, 'Feb.': February, 'Mar.': March, 'Apr.': April, 'May': May, 'Jun.': June, 'Jul.': July, 'Aug.': August, 'Sep.': September, 'Oct.': October, 'Nov.': November, 'Dec.': December.
3	Site	38	Site_Temperature_Period_Year	Date	First and last calendar years of the period ('First year-Last year') at which site average temperature is determined or 'NA'.
3	Site	39	identifiant_Paper	Index	Corresponding index from the 'Article' table. Secondary key of the 'Site' table.
4	Crop_Sequence_Trtr	40	IDRotation	Index	Index of each crop sequence and/or treatment from each site. Primary key of the 'Crop_Sequence_Trtr' table.
4	Crop_Sequence_Trtr	41	Crop_Sequence_Trtr_Name	Class	Crop sequence and/or treatment name(s). Common names of the 'Crop_Species_Common_Name' attribute from the 'Crop' table are reported. See Data Records section for further information.
4	Crop_Sequence_Trtr	42	Crop_Sequence_Trtr_Species_Order	Class	Species order. Common names of the 'Crop_Species_Common_Name' attribute from the 'Crop' table are reported. Each common name is separated by a '!'. See Data Records section for further information.
4	Crop_Sequence_Trtr	43	Crop_Sequence_Trtr_Species_Number	Numerical	Species number. Monoculture accounts for one species. Fallow accounts for zero species.
4	Crop_Sequence_Trtr	44	Crop_Sequence_Trtr_Species_Legume_Harvested	Binary	There is ('1') or there is not ('0') at least one harvested legume species in the crop sequence. Fallow is reported as a non-legume species.
4	Crop_Sequence_Trtr	45	Crop_Sequence_Trtr_Cultivar_Name	Class	Cultivar name(s) of each species in the crop sequence or 'NA'. Cultivar names of preceding and following species in the crop sequence are separated by a '!'. 'NA' is reported for fallow.

4	Crop_Sequence_Trtr	46	Crop_Sequence_Trtr_Growing_Season_Number	Numerical	Number of consecutive growing season(s) in the crop sequence.
4	Crop_Sequence_Trtr	47	IDSite_Site	Index	Corresponding index from the 'Site' table. Secondary key of the 'Crop_Sequence_Trtr' table.
5	Crop	48	IDCrop	Index	Index of each crop from each crop sequence and/or treatment. Primary key of the 'Crop' table.
5	Crop	49	Crop_Sequence_Treatment_Name	Class	Crop sequence and/or treatment name(s). Common names of the 'Crop_Species_Common_Name' attribute from the 'Crop' table are reported. See Data Records section for further information.
5	Crop	50	Crop_Site_Growing_Season_ID	Index	Index for each crop grown at the same field site during the same growing seasons.
5	Crop	51	Crop_Growing_Season_Year_First	Date	First calendar year at which the crop is seeded and/or the growing season starts or 'NA'. When values are averaged over many growing seasons, only the calendar year of the first growing season is reported. For instance, if values are averaged over 5 growing seasons from 2005 to 2010, then only 2005 is reported.
5	Crop	52	Crop_Growing_Season_Year_Last	Date	Last calendar year at which the crop is harvested and/or the growing season ends or 'NA'. When values are averaged over many growing seasons, only the calendar year of the last growing season is reported. For instance, if values are averaged over 5 growing seasons from 2005 to 2010, then only 2010 is reported.
5	Crop	53	Crop_Growing_Season_Number	Numerical	Number of growing season(s). When values are averaged over many growing seasons, the number of growing seasons is reported. For instance, if values are averaged from 2005 to 2010, then 5 growing seasons are reported.
5	Crop	54	Crop_Species_Scientific_Name	Class	Species scientific name. See Data Records section for further information.
5	Crop	55	Crop_Species_Common_Name	Class	Species common name. See Data Records section for further information.

## Annexes du Chapitre 2

5	Crop	56	Crop_Species_Legume	Binary	The species is ('1') or is not ('0') a legume species. Fallow is reported as a non-legume species.
5	Crop	57	Crop_Date_Seeding	Date	Average seeding date ('Day Month Year') or 'NA NA NA'. When values are averaged over many growing seasons, seeding dates for each growing season are reported. Months are abbreviated: 'Jan.': January, 'Feb.': February, 'Mar.': March, 'Apr.': April, 'May': May, 'Jun.': June, 'Jul.': July, 'Aug.': August, 'Sep.': September, 'Oct.': October, 'Nov.': November, 'Dec.': December.
5	Crop	58	Crop_Date_Harvest	Date	Average harvest date ('Day Month Year') or 'NA NA NA'. When values are averaged over many growing seasons, harvest dates for each growing season are reported. Months are abbreviated: 'Jan.': January, 'Feb.': February, 'Mar.': March, 'Apr.': April, 'May': May, 'Jun.': June, 'Jul.': July, 'Aug.': August, 'Sep.': September, 'Oct.': October, 'Nov.': November, 'Dec.': December.
5	Crop	59	Crop_Date_From_Seeding_To_Harvest_Day_Number	Numerical	Average number of Julian days from seeding to harvest dates or 'NA'. See Data Records section for further information.
5	Crop	60	Crop_Following_Number	Binary	The species is ('1') or is not ('0') a following species. See Data Records section for further information.
5	Crop	61	Crop_Multiple_Following_For_Same_Preceding	Binary	The row is ('1') or is not ('0') a duplicated row when values are averaged over a same crop preceding different crops. See Data Records section for further information.
5	Crop	62	Crop_Across_Treatment_Averaged_Value	Binary	The values are ('1') or are not ('0') averaged over many treatments. See Data Records section for further information.
5	Crop	63	Crop_Across_Treatment_Averaged_Value_Type	Class	If values are averaged over many treatments, then the type of treatment(s) is reported. If values are not averaged over many treatments, 'NULL' is reported.
5	Crop	64	Crop_Across_Species_Same_Treatment_Value	Binary	The species shares ('1') or does not share ('0') the same treatment(s) tested on other species grown at the same field site during the same growing seasons. See Data Records section for further information.

5	Crop	65	Crop_Across_Species_Same_Treatment_Value_Type	Class	If the species does not share the same treatment(s) tested on other species grown at the same field site during the same growing seasons, then the type of different treatments is reported. If the species shares the same treatment(s) tested on other species grown at the same field site during the same growing seasons, then 'NULL' is reported. See Data Records section for further information.
5	Crop	66	Crop_Replicate_Number	Numerical	Number of replicates or 'NA'. See Data Records section for further information.
5	Crop	67	Crop_Yield_Grain	Numerical	Grain yield or 'NA'. Shells are included but pods are not, except for <i>Arachis hypogaea</i> (peanut). See Data Records section for further information.
5	Crop	68	Crop_Yield_Grain_Unit	Class	Unit of grain yield or 'NA'.
5	Crop	69	Crop_Yield_Grain_Error	Numerical	Error term of grain yield or 'NA'.
5	Crop	70	Crop_Yield_Grain_Error_Type	Class	Error type of grain yield or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	71	Crop_Yield_Grain_DM_Percentage	Numerical	Dry matter percentage of grain yield or 'NA'.
5	Crop	72	Crop_Biomass_Aerial	Numerical	Aerial biomass or 'NA'.
5	Crop	73	Crop_Biomass_Aerial_Unit	Class	Unit of aerial biomass or 'NA'.



5	Crop	74	Crop_Biomass_Aerial_Error	Numerical	Error term of aerial biomass or 'NA'.
5	Crop	75	Crop_Biomass_Aerial_Error_Type	Class	Error type of aerial biomass or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	76	Crop_Biomass_Aerial_DM_Percentage	Numerical	Dry matter percentage of aerial biomass or 'NA'.
5	Crop	77	Crop_Biomass_Aerial_Definition	Class	Definition of aerial components in aerial biomass or 'NA'.
5	Crop	78	Crop_Biomass_Aerial_Stage_Detailed	Date	Detailed phenology stage (i.e. originally stated in the article) at which aerial biomass is determined or 'NA'.
5	Crop	79	Crop_Biomass_Aerial_Stage_Simplified	Date	Simplified phenology stage ('Before physiological maturity'/'Physiological maturity') at which aerial biomass is determined or 'NA'.
5	Crop	80	Crop_Harvest_Index	Numerical	Harvest index or 'NA'. See Data Records section for further information.
5	Crop	81	Crop_Harvest_Index_Error	Numerical	Error term of harvest index or 'NA'.

5	Crop	82	Crop_Harvest_Index_Error_Type	Class	Error type of harvest index or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	83	Crop_N_Quantity_Grain	Numerical	Grain nitrogen quantity or 'NA'.
5	Crop	84	Crop_N_Quantity_Grain_Unit	Class	Unit of grain nitrogen quantity or 'NA'.
5	Crop	85	Crop_N_Quantity_Grain_Error	Numerical	Error term of grain nitrogen quantity or 'NA'.
5	Crop	86	Crop_N_Quantity_Grain_Error_Type	Class	Error type of grain nitrogen quantity or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	87	Crop_N_Quantity_Aerial	Numerical	Aerial nitrogen quantity or 'NA'.
5	Crop	88	Crop_N_Quantity_Aerial_Unit	Class	Unit of aerial nitrogen quantity or 'NA'.

5	Crop	89	Crop_N_Quantity_Aerial_Error	Numerical	Error term of aerial nitrogen quantity or 'NA'.
5	Crop	90	Crop_N_Quantity_Aerial_Error_Type	Class	Error type of aerial nitrogen quantity or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	91	Crop_N_Quantity_Aerial_Definition	Class	Definition of aerial components in aerial nitrogen quantity or 'NA'.
5	Crop	92	Crop_N_Percentage_Grain	Numerical	Grain nitrogen percentage or 'NA'. See Data Records section for further information.
5	Crop	93	Crop_N_Percentage_Grain_Error	Numerical	Error term of grain nitrogen percentage or 'NA'.
5	Crop	94	Crop_N_Percentage_Grain_Error_Type	Class	Error type of grain nitrogen percentage or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.

5	Crop	95	Crop_N_Percentage_Aerial	Numerical	Aerial nitrogen percentage or 'NA'. See Data Records section for further information.
5	Crop	96	Crop_N_Percentage_Aerial_Error	Numerical	Error term of aerial nitrogen percentage or 'NA'.
5	Crop	97	Crop_N_Percentage_Aerial_Error_Type	Class	Error type of aerial nitrogen percentage or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	98	Crop_N_Percentage_Aerial_Definition	Class	Definition of aerial components in aerial nitrogen percentage or 'NA'.
5	Crop	99	Crop_N_Harvest_Index	Numerical	Nitrogen harvest index or 'NA'. See Data Records section for further information.
5	Crop	100	Crop_N_Harvest_Index_Error	Numerical	Error term of nitrogen harvest index or 'NA'.

5	Crop	101	Crop_N_Harvest_Index_Error_Type	Class	Error type of nitrogen harvest index or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	102	Crop_N_Fixed_Quantity_Aerial	Numerical	Aerial fixed nitrogen quantity or 'NA'. See Data Records section for further information.
5	Crop	103	Crop_N_Fixed_Quantity_Aerial_Unit	Class	Unit of aerial fixed nitrogen quantity or 'NA'.
5	Crop	104	Crop_N_Fixed_Quantity_Aerial_Error	Numerical	Error term of aerial fixed nitrogen quantity or 'NA'.
5	Crop	105	Crop_N_Fixed_Quantity_Aerial_Error_Type	Class	Error type of aerial fixed nitrogen quantity or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	106	Crop_N_Fixed_Quantity_Aerial_Definition	Class	Definition of aerial components in aerial fixed nitrogen quantity or 'NA'.

5	Crop	107	Crop_N_Fixed_Percentage_Aerial	Numerical	Aerial fixed nitrogen percentage or 'NA'.
5	Crop	108	Crop_N_Fixed_Percentage_Aerial_Error	Numerical	Error term of aerial fixed nitrogen percentage or 'NA'.
5	Crop	109	Crop_N_Fixed_Percentage_Aerial_Error_Type	Class	Error type of aerial fixed nitrogen percentage or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	110	Crop_N_Fixed_Percentage_Aerial_Method	Class	Method at which aerial fixed nitrogen percentage is determined or 'NA'.
5	Crop	111	Crop_N_Fixed_Percentage_Aerial_Reference_Species	Class	Species scientific name(s) of non-fixing reference species at which aerial fixed nitrogen percentage is determined or 'NA'.
5	Crop	112	Crop_N_Fixed_Percentage_Aerial_Stage_Detailed	Date	Detailed phenology stage (i.e. originally stated in the article) at which aerial fixed nitrogen percentage is determined or 'NA'.
5	Crop	113	Crop_N_Fixed_Percentage_Aerial_Stage_Simplified	Date	Simplified phenology stage ('Before physiological maturity'/'Physiological maturity') at which aerial fixed nitrogen percentage is determined or 'NA'.
5	Crop	114	Crop_Protein_Quantity_Percentage_Grain	Numerical	Grain protein quantity or grain protein percentage or 'NA'.

5	Crop	115	Crop_Protein_Quantity_Percentage_Grain_Unit	Class	Unit of grain protein quantity or grain protein percentage or 'NA'.
5	Crop	116	Crop_Protein_Quantity_Percentage_Grain_Error	Numerical	Error term of grain protein quantity or grain protein percentage or 'NA'.
5	Crop	117	Crop_Protein_Quantity_Percentage_Grain_Error_Type	Class	Error type of grain protein quantity or grain protein percentage or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	118	Crop_N_Balance_Simplified	Numerical	Simplified nitrogen balance or 'NA'. See Data Records section for further information.
5	Crop	119	Crop_N_Balance_Simplified_Unit	Class	Unit of simplified nitrogen balance or 'NA'. Simplified nitrogen balance is only reported in nitrogen.
5	Crop	120	Crop_N_Balance_Simplified_Error	Numerical	Error term of simplified nitrogen balance or 'NA'.

5	Crop	121	Crop_N_B alance_Si mplified_Er ror_Type	Class	Error type of simplified nitrogen balance or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	122	Crop_N_B alance_Si mplified_E quation	Class	Equation of simplified nitrogen balance or 'NA'. See Data Records section for further information.
5	Crop	123	Crop_N_S oil_Quantit y_Percenta ge_Seedin g	Numerical	Soil nitrogen quantity or soil nitrogen percentage at seeding or 'NA'.
5	Crop	124	Crop_N_S oil_Quantit y_Percenta ge_Seedin g_Unit	Class	Unit of soil nitrogen quantity or soil nitrogen percentage at seeding or 'NA'.
5	Crop	125	Crop_N_S oil_Quantit y_Percenta ge_Seedin g_Type	Class	Type of soil nitrogen quantity or soil nitrogen percentage at seeding ('Mineral'/'Nitrate'/'Nitrogen') or 'NA'. Mineral nitrogen is defined as ammonium plus nitrate.
5	Crop	126	Crop_N_S oil_Quantit y_Percenta ge_Seedin g_Error	Numerical	Error term of soil nitrogen quantity or soil nitrogen percentage at seeding or 'NA'.



5	Crop	127	Crop_N_Soil_Quantity_Percentage_Seeding_Error_Type	Class	Error type of soil nitrogen quantity or soil nitrogen percentage at seeding or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	128	Crop_N_Soil_Quantity_Percentage_Seeding_Depth	Numerical	Soil depth layer at which soil nitrogen quantity or soil nitrogen percentage at seeding is determined or 'NA'.
5	Crop	129	Crop_N_Soil_Quantity_Percentage_Seeding_Depth_Unit	Class	Unit of soil depth layer at which soil nitrogen quantity or soil nitrogen percentage at seeding is determined or 'NA'.
5	Crop	130	Crop_N_Soil_Quantity_Percentage_Seeding_Date	Date	Date at which soil nitrogen quantity or soil nitrogen percentage at seeding is determined or 'NA'.
5	Crop	131	Crop_N_Soil_Quantity_Percentage_Harvest	Numerical	Soil nitrogen quantity or soil nitrogen percentage at harvest or 'NA'.
5	Crop	132	Crop_N_Soil_Quantity_Percentage_Harvest_Unit	Class	Unit of soil nitrogen quantity or soil nitrogen percentage at harvest or 'NA'.
5	Crop	133	Crop_N_Soil_Quantity_Percentage_Harvest_Type	Class	Type of soil nitrogen quantity or soil nitrogen percentage at harvest ('Mineral'/'Nitrate'/'Nitrogen') or 'NA'. Mineral nitrogen is defined as ammonium plus nitrate.

5	Crop	134	Crop_N_Soil_Quantity_Percentage_Harvest_Error	Numerical	Error term of soil nitrogen quantity or soil nitrogen percentage at harvest or 'NA'.
5	Crop	135	Crop_N_Soil_Quantity_Percentage_Harvest_Error_Type	Class	Error type of soil nitrogen quantity or soil nitrogen percentage at harvest or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	136	Crop_N_Soil_Quantity_Percentage_Harvest_Depth	Numerical	Soil depth layer at which soil nitrogen quantity or soil nitrogen percentage at harvest is determined or 'NA'.
5	Crop	137	Crop_N_Soil_Quantity_Percentage_Harvest_Depth_Unit	Class	Unit of soil depth layer at which soil nitrogen quantity or soil nitrogen percentage at harvest is determined or 'NA'.
5	Crop	138	Crop_N_Soil_Quantity_Percentage_Harvest_Date	Date	Date at which soil nitrogen quantity or soil nitrogen percentage at harvest is determined or 'NA'.
5	Crop	139	Crop_Water_Use_Balance	Numerical	Water use or water balance or 'NA'.
5	Crop	140	Crop_Water_Use_Balance_Unit	Class	Unit of water use or water balance or 'NA'.
5	Crop	141	Crop_Water_Use_Balance_Error	Numerical	Error term of water use or water balance or 'NA'.

5	Crop	142	Crop_Water_Use_Balance_Error_Type	Class	Error type of water use or water balance or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	143	Crop_Water_Use_Balance_Equation	Class	Equation of water use or water balance or 'NA'. See Data Records section for further information.
5	Crop	144	Crop_Water_Use_Balance_Efficiency_Grain	Numerical	Grain water use efficiency or grain water balance efficiency or 'NA'. See Data Records section for further information.
5	Crop	145	Crop_Water_Use_Balance_Efficiency_Grain_Unit	Class	Unit of grain water use efficiency or grain water balance efficiency or 'NA'.
5	Crop	146	Crop_Water_Use_Balance_Efficiency_Grain_Error	Numerical	Error term of grain water use efficiency or grain water balance efficiency or 'NA'.
5	Crop	147	Crop_Water_Use_Balance_Efficiency_Grain_Error_Type	Class	Error type of grain water use efficiency or grain water balance efficiency or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.

5	Crop	148	Crop_Water_Use_Balance_Efficiency_Aerial	Numerical	Aerial water use efficiency or aerial water balance efficiency or 'NA'. See Data Records section for further information.
5	Crop	149	Crop_Water_Use_Balance_Efficiency_Aerial_Unit	Class	Unit of aerial water use efficiency or aerial water balance efficiency or 'NA'.
5	Crop	150	Crop_Water_Use_Balance_Efficiency_Aerial_Error	Numerical	Error term of aerial water use efficiency or aerial water balance efficiency or 'NA'.
5	Crop	151	Crop_Water_Use_Balance_Efficiency_Aerial_Error_Type	Class	Error type of aerial water use efficiency or aerial water balance efficiency or 'NA'. Error types are abbreviated: 'CD0.05': Confidence Distribution at the probability level of 5%, 'CV': Coefficient of Variation (%), 'DMR0.05': Duncan's Multiple Range Test at the probability level of 5%, 'LSD0.01': Fisher's Least Significant Difference Test at the probability level of 1%, 'LSD0.05': Fisher's Least Significant Difference Test at the probability level of 5%, 'LSD0.10': Fisher's Least Significant Difference Test at the probability level of 10%, 'SD': Standard Deviation, 'SE': Standard Error, 'SED': Standard Error of the Difference, 'SEDM': Standard Error of the Difference between Means, 'SEM': Standard Error of the Mean, 'HSD0.05': Tukey's Honest Significant Difference Test at the probability level of 5%, 'HSD0.10': Tukey's Honest Significant Difference Test at the probability level of 10%.
5	Crop	152	Crop_Water_Use_Balance_Efficiency_Aerial_Definition	Class	Definition of aerial components in aerial water use efficiency or aerial water balance efficiency or 'NA'.
5	Crop	153	IDRotation_CropSystem	Index	Corresponding index from the 'Crop_Sequence' table. Secondary key of the 'Crop' table.
6	Tillage	154	IDTillage	Index	Index of each tillage management from each crop. Primary key of the 'Tillage' table.
6	Tillage	155	Tillage_Presence_Tillage	Binary	There is ('1') or there is not ('0') tillage management or 'NA'. When the crop is 'seeded directly', then '0' is reported in the 'Tillage_Presence_Tillage' attribute, and '0.00' and 'm' are reported in the 'Tillage_Presence_Tillage_Depth' and 'Tillage_Presence_Tillage_Depth_Unit' attributes, respectively.

6	Tillage	156	Tillage_Presence_Tillage_Tool	Class	Tillage management tool(s) or 'NA'.
6	Tillage	157	Tillage_Presence_Tillage_Depth	Numerical	If there is tillage management, then the tillage depth is reported. If there is no tillage management, 'NULL' is reported. Elsewhere, 'NA' is reported.
6	Tillage	158	Tillage_Presence_Tillage_Depth_Unit	Class	If there is tillage management, then the unit of tillage depth is reported. If there is no tillage management, 'NULL' is reported. Elsewhere, 'NA' is reported.
6	Tillage	159	Tillage_Incorporation_Preceding_Residue	Binary	There is ('1') or there is no ('0') incorporation of the preceding crop residues in soil. Elsewhere, 'NA' is reported. Incorporation of the preceding crop residues in soil is independently reported from the presence or the absence of tillage management. Except when the crop is 'seeded directly' or the preceding crop residues are 'burned', then '0' is reported in the 'Tillage_Incorporation_Preceding_Residue' attribute.
6	Tillage	160	Tillage_Preceding_Species_Scientific_Name	Class	Species scientific name of the preceding crop or 'NA'.
6	Tillage	161	Tillage_Seeding_Depth	Numerical	Soil seeding depth or 'NA'.
6	Tillage	162	Tillage_Seeding_Depth_Unit	Class	Unit of soil seeding depth or 'NA'.
6	Tillage	163	Tillage_Seeding_Delay_Day	Binary	There is ('1') or there is not ('0') seeding delay or 'NA'.
6	Tillage	164	Tillage_Seeding_Delay_Day_Number	Numerical	If there is seeding delay, then the number of seeding delayed days is reported. Elsewhere, 'NA' is reported.
6	Tillage	165	Tillage_Seeding_Row_Inter	Numerical	Inter-row spacing at seeding or 'NA'.

6	Tillage	166	Tillage_Seeding_Row_Inter_Unit	Class	Unit of inter-row spacing at seeding or 'NA'.
6	Tillage	167	Tillage_Seeding_Row_Intra	Numerical	Intra-row spacing at seeding or 'NA'.
6	Tillage	168	Tillage_Seeding_Row_Intra_Unit	Class	Unit of intra-row spacing at seeding or 'NA'.
6	Tillage	169	Tillage_Seeding_Density	Numerical	Seeding density or 'NA'. Initial seeding density is reported but plant density after seeding is not.
6	Tillage	170	Tillage_Seeding_Density_Unit	Class	Unit of seeding density or 'NA'.
6	Tillage	171	Tillage_Seeding_Inoculation	Binary	There is ('1') or there is not ('0') inoculation of legume species at seeding or 'NA'. 'NA' is reported for non-legume species.
6	Tillage	172	IDCrop_Crop	Index	Corresponding index from the 'Crop' table. Secondary key of the 'Tillage' table.
7	Fertilization	173	IDFertilization	Index	Index of each fertilization management from each crop. Primary key of the 'Fertilization' table.
7	Fertilization	174	Fertilization_NPK	Class	Fertilization nutrient (Nitrogen ('N')/Phosphate ('P')/Potassium ('K')) or 'NA'.
7	Fertilization	175	Fertilization_NPK_Dose	Numerical	Fertilization dose or 'NA'. When many fertilization doses are reported for a given fertilization nutrient, all fertilization doses are added for the given fertilization nutrient.
7	Fertilization	176	Fertilization_NPK_Dose_Unit	Class	Unit of fertilization dose or 'NA'.
7	Fertilization	177	Fertilization_NPK_Dose_Unit_Type	Class	Type of unit of fertilization dose or 'NA'.

7	Fertilization	178	Fertilization_NPK_Dose_Product_Name	Class	Product name(s) of fertilization dose or 'NA'.
7	Fertilization	179	IDCrop_Crop	Index	Corresponding index from the 'Crop' table. Secondary key of the 'Fertilization' table.
8	Weed_Insect_Fungi	180	IDPDW	Index	Index of each weed and/or insects and/or fungi management from each crop. Primary key of the 'Weed_Insect_Fungi' table.
8	Weed_Insect_Fungi	181	Weed_Insect_Fungi_Presence_Weed	Binary	There is ('1') or there is not ('0') weed management or 'NA'. This attribute is mutually exclusive with the 'Weed_Insect_Fungi_Presence_Insect' and 'Weed_Insect_Fungi_Presence_Fungi' attributes.
8	Weed_Insect_Fungi	182	Weed_Insect_Fungi_Presence_Insect	Binary	There is ('1') or there is not ('0') insect management or 'NA'. This attribute is mutually exclusive with the 'Weed_Insect_Fungi_Presence_Weed' and 'Weed_Insect_Fungi_Presence_Fungi' attributes.
8	Weed_Insect_Fungi	183	Weed_Insect_Fungi_Presence_Fungi	Binary	There is ('1') or there is not ('0') fungi management or 'NA'. This attribute is mutually exclusive with the 'Weed_Insect_Fungi_Presence_Weed' and 'Weed_Insect_Fungi_Presence_Insect' attributes.
8	Weed_Insect_Fungi	184	Weed_Insect_Fungi_Presence_Treatment_Mechanical	Binary	If there is weed or insect or fungi management, then there is ('1') or there is not ('0') mechanical treatment. Elsewhere, 'NA' is reported.
8	Weed_Insect_Fungi	185	Weed_Insect_Fungi_Presence_Treatment_Mechanical_Date	Date	If there is mechanical treatment, then the date is reported or 'NA'. If there is no mechanical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	186	Weed_Insect_Fungi_Presence_Treatment_Chemical	Binary	If there is weed or insect or fungi management, then there is ('1') or there is not ('0') chemical treatment. Elsewhere, 'NA' is reported.

8	Weed_Insect_Fungi	187	Weed_Insect_Fungi_Presence_Treatment_Chemical_Date	Date	If there is chemical treatment, then the date is reported or 'NA'. If there is no chemical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	188	Weed_Insect_Fungi_Presence_Treatment_Chemical_Dose	Numerical	If there is chemical treatment, then the chemical dose is reported or 'NA'. If there is no chemical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	189	Weed_Insect_Fungi_Presence_Treatment_Chemical_Dose_Unit	Class	If there is chemical treatment, then the unit of chemical dose is reported or 'NA'. If there is no chemical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	190	Weed_Insect_Fungi_Presence_Treatment_Chemical_Dose_Unit_Active	Binary	If there is chemical dose, then the unit of chemical dose is reported ('1') or is not reported ('0') in active ingredients or 'NA'. If there is no chemical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	191	Weed_Insect_Fungi_Presence_Treatment_Chemical_Dose_Product_Name	Class	If there is chemical treatment, then the product name(s) of chemical dose is(are) reported or 'NA'. If there is no chemical treatment, then 'NULL' is reported.
8	Weed_Insect_Fungi	192	IDCrop_Crop	Index	Corresponding index from the 'Crop' table. Secondary key of the 'Weed_Insect_Fungi' table.
9	Irrigation	193	IDIrrigation	Index	Index of each irrigation management from each crop. Primary key of the 'Irrigation' table.
9	Irrigation	194	Irrigation_Presence_Irrigation	Binary	There is ('1') or there is not ('0') irrigation management or 'NA'.
9	Irrigation	195	Irrigation_Presence_Irrigation_Dose	Numerical	If there is irrigation management, then the irrigation dose is reported. If there is no irrigation management, then '0' is reported. Elsewhere, 'NA' is reported.



9	Irrigation	196	Irrigation_P resence_Irr igation_Do se_Unit	Class	If there is irrigation dose, then the unit of irrigation dose is reported. If there is no irrigation dose, then 'mm' is reported. Elsewhere, 'NA' is reported.
9	Irrigation	197	Irrigation_P resence_Irr igation_Met hod	Class	If there is irrigation management, then the irrigation method is reported. If there is no irrigation management, then 'NULL' is reported. Elsewhere, 'NA' is reported.
9	Irrigation	198	IDCrop_Cr op	Index	Corresponding index from the 'Crop' table. Secondary key of the 'Irrigation' table.

## **Annexes du Chapitre 3**



**Note S1.** References for articles used in the meta-analysis.

- Adela J, Simona N, Gheorghe D, Pîrşan P and Alda S 2010 The behaviour of some varieties of pea, soybean and lentils concerning the relationship bacterization-sowing period on the yield and quality in Timiș Plain *Res. J. Agr. Sci.* **42** 82–7
- Angadi S V, McConkey B G, Cutforth H W, Miller P R, Ulrich D, Selles F, Volkmar K M, Entz M H and Brandt S A 2008 Adaptation of alternative pulse and oilseed crops to the semiarid Canadian Prairie: seed yield and water use efficiency *Can. J. Plant Sci.* **88** 425–38
- Annicchiario P 2008 Adaptation of cool-season grain legume species across climatically-contrasting environments of southern Europe *Agron. J.* **100** 1647–54
- Ayaz S, McKenzie B A, Hill G D and McNeil D L 2004 Variability in yield of four grain legume species in a subhumid temperate environment I. Yields and harvest index *J. Agr. Sci.* **142** 9–19
- Beck D, Wery J, Saxena M and Ayadi A 1991 Dinitrogen fixation and nitrogen-balance in cool-season food legumes *Agron. J.* **83** 334–41
- Brennan R F and French R J 2005 Grain yield and cadmium concentration of a range of grain legume species grown on two soil types at Merredin, western Australia *Aust. J. Exp. Agr.* **45** 1167–72
- Calderón F J, Vigil M F, Nielsen D C, Benjamin J G and Poss D J 2012 Water use and yields of no-till managed dryland grasspea and yellow pea under different planting configurations *Field Crop. Res.* **125** 179–85
- Chen C, Miller P, Muehlbauer F, Neill K, Wichman D and McPhee K 2006 Winter pea and lentil response to seeding date and micro- and macro-environments *Agron. J.* **98** 1655–63
- Cutforth H W, McConkey B G, Ulrich D, Miller P R and Angadi S V 2002 Yield and water use efficiency of pulses seeded directly into standing stubble in the semiarid Canadian prairie *Can. J. Plant Sci.* **82** 681–6
- Fernandez A L, Sheaffer C C, Wyse D L and Michaels T E 2012 Yield and weed abundance in early- and late-sown field pea and lentil *Agron. J.* **104** 1056–64

- French R and Ewing M 1989 Soil type influences the relative yields of different cereals and crop legumes in the western Australian wheatbelt *Aust. J. Exp. Agr.* **29** 829–35
- Gan Y T, Miller P R, Liu P H, Stevenson F C and McDonald C L 2002 Seedling emergence, pod development, and seed yields of chickpea and dry pea in a semiarid environment *Can. J. Plant Sci.* **82** 531–7
- Gan Y, Campbell C A, Liu L, Basnyat P and McDonald C L 2009 Water use and distribution profile under pulse and oilseed crops in semiarid northern high latitude areas *Agr. Water Manage.* **96** 337–48
- Gregory P J 1998 Alternative crops for Duplex soils: growth and water use of some cereal, legume, and oilseed crops, and pastures *Aust. J. Agr. Res.* **49** 21–32
- Hanbury C D, Siddique K H M, Galwey N W and Cocks P S 1999 Genotype-environment interaction for seed yield and ODAP concentration of *Lathyrus sativus* L. and *L. cicera* L. in Mediterranean-type environments *Euphytica* **110** 45–60
- Hatch D J, Joynes A and Stone A 2010 Nitrogen uptake in organically managed spring sown lupins and residual effects on leaching and yield of a following winter cereal *Soil Use Manage.* **26** 21–6
- Haugaard-Nielsen H, Mundus S and Jensen E S 2009 Nitrogen dynamics following grain legumes and subsequent catch crops and the effects on succeeding cereal crops *Nutr. Cycl. Agroecosys.* **84** 281–91
- Haugaard-Nielsen H, Mundus S and Jensen E S 2012 Grass-clover undersowing affects nitrogen dynamics in a grain legume-cereal arable cropping system *Field Crop. Res.* **136** 23–31
- Haynes R J, Martin R J and Goh K M 1993 Nitrogen fixation, accumulation of soil nitrogen and nitrogen balance for some field-grown legume crops *Field Crop. Res.* **35** 85–92
- Heenan D 1994 Effects of sowing time on growth and grain yield of lupin and field pea in south-eastern New South Wales *Aust. J. Exp. Agr.* **34** 1137–42
- Henry J, Slinkard A and Hogg T 1995 The effect of phosphorus fertilizer on establishment, yield and quality of pea, lentil and faba bean *Can. J. Plant Sci.* **75** 395–8

- Jensen E S 1986 Symbiotic N<sub>2</sub> fixation in pea and field bean estimated by <sup>15</sup>N fertilizer dilution in field experiments with barley as a reference crop *Plant Soil* **92** 3–13
- Jensen C R, Joernsgaard B, Andersen M N, Christiansen J L, Mogensen V O, Friis P and Petersen C T 2004 The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop *Eur. J. Agron.* **20** 405–18
- Jessop R S and Mahoney J 1985 The effect of soil nitrogen on grain legume yield and nitrogen fixation *J. Agr. Sci.* **105** 231–6
- Kaul H P 2004 Pre-crop effects of grain legumes and linseed on soil mineral N and productivity of subsequent winter rape and winter wheat crops *Bodenkultur* **55** 95–102
- Krupinsky J M, Tanaka D L, Merrill S D, Liebig M A and Hanson J D 2006 Crop sequence effects of 10 crops in the northern Great Plains *Agr. Syst.* **88** 227–54
- Lafond J and Pageau D 2007 Nitrogen and non nitrogen benefit of legumes to a subsequent spring barley crop *Can. J. Soil Sci.* **87** 445–54
- Laurence R 1979 A comparison of the grain and protein yield potential of some annual legume species in south Australia *Aust. J. Exp. Agr. Anim. Husb.* **19** 495–503
- Lemerle D and Hinkley R 1991 Tolerances of canola, field pea, lupin and faba bean cultivars to herbicides *Aust. J. Exp. Agr.* **31** 379–86
- Lutman P J W, Dixon F L and Risiott R 1994 The response of four spring-sown combinable arable crops to weed competition *Weed Res.* **34** 137–46
- Maidl F X, Haunz F X, Panse A and Fischbeck G 1996 Transfer of grain legume nitrogen within a crop rotation containing winter wheat and winter barley *J. Agron. Crop Z. Acker Pflanzenbau* **176** 47–57
- Malhi S S, Johnston A M, Schoenau J J, Wang Z H and Vera C L 2007 Seasonal biomass accumulation and nutrient uptake of pea and lentil on a black Chernozem soil in Saskatchewan *J. Plant Nutr.* **30** 721–37

- Malik R S, Seymour M, French R J, Kirkegaard J A, Lawes R A and Liebig M A 2015 Dynamic crop sequencing in western Australian cropping systems *Crop Pasture Sci.* **66** 594–609
- McEwen J, Darby R, Hewitt M and Yeoman D 1989 Effects of field beans, fallow, lupins, oats, oilseed rape, peas, ryegrass, sunflowers and wheat on nitrogen residues in the soil and on the growth of a subsequent wheat crop *J. Agr. Sci.* **115** 209–19
- Miller P R, McDonald C L, Derksen D A and Waddington J 2001 The adaptation of seven broadleaf crops to the dry semiarid prairie *Can. J. Plant Sci.* **81** 29–43
- Miller P R, Gan Y, McConkey B G and McDonald C L 2003 Pulse crops for the northern Great Plains: I. Grain productivity and residual effects on soil water and nitrogen *Agron. J.* **95** 972–9
- Miller P R and Holmes J A 2005 Cropping sequence effects of four broadleaf crops on four cereal crops in the northern Great Plains *Agron. J.* **97** 189–200
- Miller P R, Brandt S A, McDonald C L and Waddington J 2006 Chickpea, lentil, and pea response to delayed spring seeding on the northern Great Plains *Can. J. Plant Sci.* **86** 1059–70
- Moschini V, Casella G, Vivoli R, Vazzana C, Martini A, Lotti C and Migliorini P 2014 Performance of organic grain legumes in Tuscany *Ital. J. Agron.* **9** 38–43
- Neugschwandtner R W, Wagentristl H and Kaul H P 2015 Nitrogen yield and nitrogen use of chickpea compared to pea, barley and oat in central Europe *Int. J. Plant Prod.* **9** 291–303
- Nielsen D C 2001 Production functions for chickpea, field pea, and lentil in the central Great Plains *Agron. J.* **93** 563–9
- Niță S, Tabără V, David G, Niță L D, Simion A, Dragoș M and Borcean A 2012 Results obtained for soybean, pea and lentils crops on a Cambic Chernozem in the Banat's Plain during 2008-2010 *Rom. Agr. Res.* **29** 155–62
- Plaza-Bonilla D, Nolot J M, Passot S, Raffailac D and Justes E 2016 Grain legume-based rotations managed under conventional tillage need cover crops to mitigate soil organic matter losses *Soil Till. Res.* **156** 33–43

- Przednowek D W A, Entz M H, Irvine B, Flaten D N and Martens J R T 2004 Rotational yield and apparent N benefits of grain legumes in southern Manitoba *Can. J. Plant Sci.* **84** 1093–6
- Rennie R and Dubetz S 1986 Nitrogen-15-determined nitrogen fixation in field-grown chickpea, lentil, fababean, and field pea *Agron. J.* **78** 654–60
- Ruisi P, Giambalvo D, Di Miceli G, Frenda A S, Saia S and Amato G 2012 Tillage effects on yield and nitrogen fixation of legumes in Mediterranean conditions *Agron. J.* **104** 1459–66
- Schultz J E 1995 Crop production in a rotation trial at Tarlee, south Australia *Aust. J. Exp. Agr.* **35** 865–76
- Siddique K H M, Walton G and Seymour M 1993 A comparison of seed yields of winter grain legumes in western Australia *Aust. J. Exp. Agr.* **33** 915–22
- Siddique K H M and Loss S P 1999 Studies on sowing depth for chickpea (*Cicer arietinum* L.), faba bean (*Vicia faba* L.) and lentil (*Lens culinaris* Medik) in a Mediterranean-type environment of south-western Australia *J. Agron. Crop Sci.* **182** 105–12
- Silisbury J H 1975 Grain yield and flowering of some introduced grain legumes in south Australia *Aust. J. Exp. Agr. Anim. Husb.* **15** 556–60
- Smith S C, Bezdicsek D F, Turco R F and Cheng H H 1987 Seasonal N<sub>2</sub> fixation by cool-season pulses based on several <sup>15</sup>N methods *Plant Soil* **97** 3–13
- Spaner D, Todd A G and McKenzie D B 2001 Pea and soybean performance in Newfoundland *Can. J. Plant Sci.* **81** 723–6
- Stringi L, Giambalvo D, Trapani P, Scarpello C and Attardo C 2004 Effect of soil management techniques on performance of different grain legumes in a Mediterranean environment *Options Méditerranéennes* **60** 79–83
- Tanaka D L, Anderson R L and Rao S C 2005 Crop sequencing to improve use of precipitation and synergize crop growth *Agron. J.* **97** 385–90
- Tanaka D L, Krupinsky J M, Merrill S D, Liebig M A and Hanson J D 2007 Dynamic cropping systems for sustainable crop production in the northern Great Plains *Agron. J.* **99** 904–11



- Thomson B D, Siddique K H M, Barr M D and Wilson J M 1997 Grain legume species in low rainfall Mediterranean-type environments I. Phenology and seed yield *Field Crop. Res.* **54** 173–87
- Wang X, Gan Y, Hamel C, Lemke R and McDonald C 2012 Water use profiles across the rooting zones of various pulse crops *Field Crop. Res.* **134** 130–7
- Wen G, Chen C, Neill K, Wichman D and Jackson G 2008 Yield response of pea, lentil and chickpea to phosphorus addition in a clay loam soil of central Montana *Arch. Agron. Soil Sci.* **54** 69–82
- Williams C M, King J R, Ross S M, Olson M A, Hoy C F and Lopetinsky K J 2014 Effects of three pulse crops on subsequent barley, canola, and wheat *Agron. J.* **106** 343–50
- Wright A T 1990 Yield effect of pulses on subsequent cereal crops in the northern Prairies *Can. J. Plant Sci.* **70** 1023–32
- Zentner R P *et al* 2002 Economics of crop diversification and soil tillage opportunities in the Canadian Prairies *Agron. J.* **94** 216–30

**Table S1.** Scientific and common names for species by region, adapted from the United States Department of Agriculture Plants database (<http://plants.usda.gov/java/>) (accessed 29 Mar 2016). All species are ranked alphabetically according to scientific names.

Scientific name	Common name	Region
<i>Cicer arietinum</i>	Chickpea	Europe, North America, Oceania
<i>Glycine max</i>	Soybean	Europe, North America
<i>Lathyrus aphaca</i>	Yellow pea	Oceania
<i>Lathyrus cicera</i>	Red pea	Oceania
<i>Lathyrus clymenum</i>	Cicercha purpurina	Oceania
<i>Lathyrus ochrus</i>	Cyprus vetch	Oceania
<i>Lathyrus sativus</i>	White pea	North America, Oceania
<i>Lens culinaris</i>	Lentil	Europe, North America, Oceania
<i>Lupinus albus</i>	White lupin	Europe, Oceania
<i>Lupinus angustifolius</i>	Narrowleaf lupin	Europe, North America, Oceania
<i>Lupinus atlanticus</i>	NA <sup>a</sup>	Oceania
<i>Lupinus luteus</i>	Yellow lupin	Europe, Oceania
<i>Lupinus pilosus</i>	Blue lupin	Oceania
<i>Phaseolus vulgaris</i>	Kidney bean	North America
<i>Pisum sativum</i>	Pea	Europe, North America, Oceania
<i>Vicia articulata</i>	Oneflower vetch	Oceania
<i>Vicia benghalensis</i>	Purple vetch	Oceania
<i>Vicia ervilia</i>	Blister vetch	Oceania
<i>Vicia faba</i>	Faba bean	Europe, North America, Oceania
<i>Vicia hybrida</i>	Hairy yellow vetch	Oceania
<i>Vicia narbonensis</i>	Purple broad vetch	Oceania
<i>Vicia sativa</i>	Garden vetch	Oceania
<i>Vicia villosa</i>	Winter vetch	Oceania

<sup>a</sup>Not Available.

**Table S2.** Number of available experimental data for productivity variables by region.

Region	Grain biomass <sup>a,c</sup>	Total aerial biomass <sup>a,b,c</sup>	Harvest index <sup>d</sup>	Grain nitrogen content <sup>e</sup>	Percentage of nitrogen in grains <sup>f</sup>	Water use <sup>g</sup>	Grain water use efficiency <sup>h,i</sup>
Europe	265	82	21	41	0	0	0
North America	917	176	7	22	46	89	113
Oceania	1,014	215	121	NA <sup>j</sup>	0	0	0
Total	2,196	473	149	63	46	89	113

<sup>a</sup>Grain biomass and total aerial biomass are systematically expressed on a dry matter basis. If the percentage of dry matter is not available for grain biomass and/or total aerial biomass of all legume species cultivated at the same field sites during the same growing seasons, we assume that their grain biomass and/or total aerial biomass are expressed on a dry matter basis. If the percentage of dry matter is not available for the grain biomass and/or total aerial biomass of a legume species but available for the grain biomass and/or total aerial biomass of other legume species cultivated at the same field sites during the same growing seasons, we assume that the missing percentage of dry matter is the mean percentage of dry matter of the grain biomass and/or total aerial biomass of legume species having available information about their percentage of dry matter.

<sup>b</sup>Total aerial biomass is defined as the sum of grain biomass and non-grain aerial biomass at physiological maturity. The aerial biomass originally termed as either 'stover' or 'stubble' or 'straw' is assumed to not include grain biomass. The total aerial biomass including components not precisely stated is systematically excluded.

<sup>c</sup>All calculations were made with grain biomass and total aerial biomass expressed in t ha<sup>-1</sup>.

<sup>d</sup>Harvest index is defined as the ratio of grain biomass to total aerial biomass.

<sup>e</sup>All calculations were made with grain nitrogen content expressed in kg N ha<sup>-1</sup>, and 'N' is the abbreviation of nitrogen.

<sup>f</sup>Percentage of nitrogen in grains is multiplied by grain biomass to be converted into grain nitrogen content, and used as such.

<sup>g</sup>All calculations were made with water use expressed in mm.

<sup>h</sup>Grain water use efficiency is defined as the ratio of grain biomass to water use.

<sup>i</sup>All calculations were made with grain water use efficiency expressed in kg ha<sup>-1</sup> mm<sup>-1</sup>.

<sup>j</sup>Not Available.

**Table S3.** Equation of water use for each index of field site\*growing season included in the data set.

Index of field site* growing season	Equation of water use
15	Calculated as soil water at planting-soil water at harvest+precipitation during growing season (termed as water use).
16	Calculated as soil water at planting-soil water at harvest+precipitation during growing season (termed as water use).
17	Calculated as soil water at planting-soil water at harvest+precipitation during growing season (termed as water use).
26	Calculated as soil moisture at planting-soil moisture at harvest+precipitation received during growing season+evaporation from the soil during this period (termed as water use).
27	Calculated as soil moisture at planting-soil moisture at harvest+precipitation received during growing season+evaporation from the soil during this period (termed as water use).
56	Calculated as difference in soil water between spring and fall soil sampling plus the sum of all rainfall between soil sampling dates (termed as water use).
58	Calculated as difference in soil water between spring and fall soil sampling plus the sum of all rainfall between soil sampling dates (termed as water use).
62	Calculated as difference in soil water between spring and fall soil sampling plus the sum of all rainfall between soil sampling dates (termed as water use).
64	Calculated as difference in soil water between spring and fall soil sampling plus the sum of all rainfall between soil sampling dates (termed as water use).
82	Calculated as growing season precipitation+irrigation+soil water at seeding-soil water at harvest (termed as water use).
82	Calculated as growing season precipitation+soil water at seeding-soil water at harvest (termed as water use).
83	Calculated as growing season precipitation+irrigation+soil water at seeding-soil water at harvest (termed as water use).
83	Calculated as growing season precipitation+soil water at seeding-soil water at harvest (termed as water use).
305	Calculated as soil water at seeding-soil water at harvesting+growing season precipitation (evapotranspiration, ET) (termed as water use).
325	Calculated as pre-seeding soil profile moisture plus precipitation between pre-seeding and post-harvest sampling dates, minus post-harvest soil moisture. Deep percolation, soil surface evaporation and surface runoff were not measured. Deep percolation can be considered negligible; however, deep percolation losses at Swift Current were likely substantial prior to soil sampling in the fall of 1993 and after seeding in the spring of 1994. In 1993 and 1996 at Swift Current, evaporation losses between crop harvest and fall soil sampling were likely great. Runoff was considered insignificant because the sites were nearly level. Failure to account for these soil water losses would have the effect of underestimating water use efficiency (termed as water use).

---

326	Calculated as pre-seeding soil profile moisture plus precipitation between pre-seeding and post-harvest sampling dates, minus post-harvest soil moisture. Deep percolation, soil surface evaporation and surface runoff were not measured. Deep percolation can be considered negligible; however, deep percolation losses at Swift Current were likely substantial prior to soil sampling in the fall of 1993 and after seeding in the spring of 1994. In 1993 and 1996 at Swift Current, evaporation losses between crop harvest and fall soil sampling were likely great. Runoff was considered insignificant because the sites were nearly level. Failure to account for these soil water losses would have the effect of underestimating water use efficiency (termed as water use).
327	Calculated as pre-seeding soil profile moisture plus precipitation between pre-seeding and post-harvest sampling dates, minus post-harvest soil moisture. Deep percolation, soil surface evaporation and surface runoff were not measured. Deep percolation can be considered negligible; however, deep percolation losses at Swift Current were likely substantial prior to soil sampling in the fall of 1993 and after seeding in the spring of 1994. In 1993 and 1996 at Swift Current, evaporation losses between crop harvest and fall soil sampling were likely great. Runoff was considered insignificant because the sites were nearly level. Failure to account for these soil water losses would have the effect of underestimating water use efficiency (termed as water use).
328	Calculated as pre-seeding soil profile moisture plus precipitation between pre-seeding and post-harvest sampling dates, minus post-harvest soil moisture. Deep percolation, soil surface evaporation and surface runoff were not measured. Deep percolation can be considered negligible; however, deep percolation losses at Swift Current were likely substantial prior to soil sampling in the fall of 1993 and after seeding in the spring of 1994. In 1993 and 1996 at Swift Current, evaporation losses between crop harvest and fall soil sampling were likely great. Runoff was considered insignificant because the sites were nearly level. Failure to account for these soil water losses would have the effect of underestimating water use efficiency (termed as water use).
416	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
417	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
418	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
419	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
420	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
421	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
422	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
423	Calculated as pre-seeding soil water minus post-harvest soil water plus precipitation from spring to fall soil sampling dates (termed as water use).
429	Calculated as difference between spring (i.e. just after seeding) and fall (i.e. just after harvesting) soil water measurements to 122 cm plus all rain received between spring and fall sampling dates. Runoff was considered insignificant because the field sites were level. Deep percolation between spring and fall soil sampling was not measured, but if it occurred, it was reasonably considered to affect all crops equally. Rainfall received between crop maturity and fall soil sampling was presumed to be stored in the soil (termed as water use).

---

---

611	Calculated as evapotranspiration with the water balance method using soil water measurements and assuming runoff and deep percolation were negligible. Soil water measurements were made at seeding and at harvest (termed as water use).
633	Calculated as evapotranspiration (ETc) (mm) of each plot was determined using water balance equation as follows: $ETc=Pe+DS+U-R-Dw$ where Pe is the growing season precipitation (mm), DS is the change of soil water in the 0-120 cm soil layer (mm) between the spring and harvest sampling dates, U is the upward capillary flow into the root zone (mm), R is the surface runoff (mm), Dw is the downward drainage out the root zone (mm). The upward flow and downward flow were negligible at the experimental site, similar to the findings from previous studies conducted at the experimental site. Surface runoff was also negligible during the whole growing seasons. No surface runoff was noticed (termed as water use).
634	Calculated as evapotranspiration (ETc) (mm) of each plot was determined using water balance equation as follows: $ETc=Pe+DS+U-R-Dw$ where Pe is the growing season precipitation (mm), DS is the change of soil water in the 0-120 cm soil layer (mm) between the spring and harvest sampling dates, U is the upward capillary flow into the root zone (mm), R is the surface runoff (mm), Dw is the downward drainage out the root zone (mm). The upward flow and downward flow were negligible at the experimental site, similar to the findings from previous studies conducted at the experimental site. Surface runoff was also negligible during the whole growing seasons. No surface runoff was noticed (termed as water use).
635	Calculated as evapotranspiration (ETc) (mm) of each plot was determined using water balance equation as follows: $ETc=Pe+DS+U-R-Dw$ where Pe is the growing season precipitation (mm), DS is the change of soil water in the 0-120 cm soil layer (mm) between the spring and harvest sampling dates, U is the upward capillary flow into the root zone (mm), R is the surface runoff (mm), Dw is the downward drainage out the root zone (mm). The upward flow and downward flow were negligible at the experimental site, similar to the findings from previous studies conducted at the experimental site. Surface runoff was also negligible during the whole growing seasons. No surface runoff was noticed (termed as water use).

---

**Table S4.** Name and definition for variables included in the data set.

Name	Definition
Article_Index	Index of each article.
Article_Author_First	Name of the first author.
Article_Title	Article title.
Article_Year_Publication	Publication year or 'NA' <sup>a</sup> .
Article_Journal	Journal name or 'NA'.
Site_Index	Index of each site from each article.
Site_Name	Site name.
Site_Country	Site country.
Crop_Index	Index of each crop.
Crop_Site_Growing_Season_Index	Index for each crop grown at the same field site during the same growing seasons.
Crop_Growing_Season_Year_First	First calendar year at which the crop is seeded and/or the growing season starts or 'NA'. When values are averaged over more than one growing season, only the calendar year of the first growing season is reported.
Crop_Growing_Season_Year_Last	Last calendar year at which the crop is harvested and/or the growing season ends or 'NA'. When values are averaged over more than one growing season, only the calendar year of the last growing season is reported.
Crop_Growing_Season_Number	Number of growing seasons. When values are averaged over more than one growing season, the number of growing seasons is reported.
Crop_Species_Scientific_Name	Species scientific name.
Crop_Replicate_Number	Number of replicates or 'NA'.
Crop_Biomass_Grain	Grain biomass or 'NA'.
Crop_Biomass_Grain_Unit	Unit of grain biomass or 'NA'.
Crop_Biomass_Grain_DM_Percentage	Dry matter (DM) percentage of grain biomass or 'NA'.
Crop_Biomass_Aerial	Aerial biomass or 'NA'.
Crop_Biomass_Aerial_Unit	Unit of aerial biomass or 'NA'.
Crop_Biomass_Aerial_DM_Percentage	Dry matter (DM) percentage of aerial biomass or 'NA'.
Crop_Biomass_Aerial_Definition	Definition of components included in aerial biomass or 'NA'.
Crop_Biomass_Aerial_Stage_Detailed	Detailed phenology stage (i.e., originally reported in the selected article) at which aerial biomass is measured or 'NA'.

---

Crop_Biomass_Aerial_Stage_Simplified	Simplified phenology stage (i.e., either ‘Before physiological maturity’ or ‘Physiological maturity’) at which aerial biomass is measured or ‘NA’.
Crop_Harvest_Index	Harvest index or ‘NA’.
Crop_N_Content_Grain	Grain nitrogen (N) content or ‘NA’.
Crop_N_Content_Grain_Unit	Unit of grain nitrogen (N) content or ‘NA’.
Crop_N_Percentage_Grain	Percentage of nitrogen (N) in grains or ‘NA’.
Crop_Water_Use_Balance	Water use or water balance or ‘NA’.
Crop_Water_Use_Balance_Unit	Unit of water use or water balance or ‘NA’.
Crop_Water_Use_Balance_Equation	Equation of water use or water balance or ‘NA’.
Crop_Water_Use_Balance_Efficiency_Grain	Grain water use efficiency or grain water balance efficiency or ‘NA’.
Crop_Water_Use_Balance_Efficiency_Grain_Unit	Unit of grain water use efficiency or grain water balance efficiency or ‘NA’.

---

<sup>a</sup>Not Available.



**Table S5.** Mean grain crude protein (unit) and mean grain gross energy (unit) according to the FEEDIPEDIA database (<http://www.feedipedia.org/>) (accessed 15 Mar 2016). All species are ranked alphabetically according to scientific names. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

Scientific name	Mean grain crude protein (%) <sup>a</sup>	Mean grain gross energy (10 <sup>9</sup> J t <sup>-1</sup> ) <sup>a</sup>	FEEDIPEDIA database item
<i>Cicer arietinum</i> <sup>b</sup>	22.20	19.60	Chickpea seeds, desi type Chickpea seeds, kabuli type
<i>Glycine max</i>	39.60	23.60	Soybean seeds
<i>Lathyrus aphaca</i>	NA <sup>c</sup>	NA	NA
<i>Lathyrus cicera</i>	23.80	18.80	Chick vetch ( <i>Lathyrus cicera</i> ), seeds
<i>Lathyrus clymenum</i>	NA	NA	NA
<i>Lathyrus ochrus</i>	26.10	18.90	Ochre vetch ( <i>Lathyrus ochrus</i> ), seeds
<i>Lathyrus sativus</i>	30.00	19.10	Grass pea ( <i>Lathyrus sativus</i> ), seeds
<i>Lens culinaris</i>	26.90	18.50	Lentil seeds
<i>Lupinus albus</i>	NA	NA	NA
<i>Lupinus angustifolius</i>	33.80	20.30	Lupin ( <i>Lupinus angustifolius</i> ), blue, seeds
<i>Lupinus atlanticus</i>	NA	NA	NA
<i>Lupinus luteus</i>	43.00	20.90	Lupin ( <i>Lupinus luteus</i> ), yellow, seeds
<i>Lupinus pilosus</i>	NA	NA	NA
<i>Phaseolus vulgaris</i>	24.80	18.60	Common bean seeds
<i>Pisum sativum</i>	23.90	18.30	Pea seeds
<i>Vicia articulata</i>	NA	NA	NA
<i>Vicia benghalensis</i>	NA	NA	NA
<i>Vicia ervilia</i>	NA	NA	NA
<i>Vicia faba</i> <sup>d</sup>	29.00	18.70	Faba bean ( <i>Vicia faba</i> ), all cultivars
<i>Vicia hybrida</i>	NA	NA	NA
<i>Vicia narbonensis</i>	27.20	19.00	NA
<i>Vicia sativa</i>	28.40	18.90	Common vetch ( <i>Vicia sativa</i> ), seeds
<i>Vicia villosa</i>	29.30	NA	Hairy vetch ( <i>Vicia villosa</i> ), seeds

<sup>a</sup>Mean grain crude protein and mean grain gross energy are expressed on a dry matter basis.

<sup>b</sup>Mean grain crude protein and mean grain gross energy are determined from the FEEDIPEDIA database items 'Chickpea seeds, desi type' and 'Chickpea seeds, kabuli type'.

<sup>c</sup>Not Available.

<sup>d</sup>Crude protein and gross energy are for the grains.

**Table S6.** Productivity rankings for all grain legume species relative to pea in Europe, North America and Oceania. All species are ranked alphabetically according to scientific names. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch). Mean ratios are estimated from weighted experimental data. Only the model variant including one random 'field site\*growing season' effect is used. The symbol '-' indicates that the value of the criterion for the given species is significantly lower than that for pea, and that more than two articles are available for the species concerned. The symbol '(-)' indicates that the value of the criterion for the given species is significantly lower than that of pea, and that there are no more than two articles for the species concerned. The symbol '0' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are more than two articles for the species concerned. The symbol '(0)' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are no more than two articles available for the species concerned. The symbol '+' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are more than two articles available for the species concerned. The symbol '(+)' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are no more than two articles available for the species concerned.

	Europe					North America					Oceania						
	Grain biomass ratio <sup>a</sup>	Total aerial biomass ratio <sup>a</sup>	Grain crude protein ratio <sup>a,b</sup>	Grain gross energy ratio <sup>a,b</sup>	Grain nitrogen content ratio	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain nitrogen content ratio	Water use ratio <sup>a</sup>	Grain biomass variance ratio	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain biomass variance ratio
<i>Cicer arietinum</i>	-	0	-	0	(0)	0	-	-	(-)	+	-	-	-	-	-	-	+
<i>Glycine max</i>	0		+	+			(0)	0	-					(-)	0	0	0
<i>Lathyrus aphaca</i>														(-)	0	0	0
<i>Lathyrus cicera</i>														(0)	0	0	0
<i>Lathyrus cymenium</i>														(0)	0	0	0
<i>Lathyrus ochrus</i>														(0)	0	0	0
<i>Lathyrus sativus</i>						(0)	(0)	(0)	(0)	(+)	0	0	-	-	-	0	+
<i>Lens culinaris</i>	(-)	(0)	(-)	(-)	(-)	+	-	-	-	0	0	0	-	-	-	-	+
<i>Lupinus albus</i>	-	(-)				+								-	-	-	+
<i>Lupinus angustifolius</i>	-	0	0	0	0	+	-	-	(-)	(0)			-	0	0	0	+
<i>Lupinus atlanticus</i>						+								(0)	(0)	(0)	+
<i>Lupinus luteus</i>	(0)	(0)	(0)	(0)	(0)	+								(-)	(-)	(0)	+
<i>Lupinus pilosus</i>						+								(0)	(0)	(0)	+
<i>Phaseolus vulgaris</i>							-	-	(-)	(0)	0			(-)	(-)	(-)	
<i>Vicia articulata</i>														(-)	(-)	(-)	
<i>Vicia benghalensis</i>														(-)	(0)	(0)	
<i>Vicia ervilia</i>														(-)	(0)	(0)	
<i>Vicia faba</i>	0	0	0	0	0	0	-	(0)	(0)	(0)	0	0	+	+	+	+	
<i>Vicia hybrida</i>														(-)	(-)	(-)	
<i>Vicia narbonensis</i>														0	(0)	0	0
<i>Vicia sativa</i>														-	0	0	+
<i>Vicia villosa</i>														(-)	(-)	(-)	

<sup>a</sup> Experimental data from the selected articles are used to estimate the given criterion.

<sup>b</sup> Nutritional data from the FEEDIPEDIA database are used to estimate the given criterion.

<sup>c</sup> Statistical data from the FAOSTAT database are used to estimate the given criterion.

**Table S7.** Productivity rankings for all grain legume species relative to pea in Europe, North America and Oceania. All species are ranked alphabetically according to scientific names. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch). Mean ratios are estimated from weighted experimental data. Only the model variant including two random 'article' and 'field site\*growing season' effects is used. The symbol '-' indicates that the value of the criterion for the given species is significantly lower than that for pea, and that more than two articles are available for the species concerned. The symbol '(-)' indicates that the value of the criterion for the given species is significantly lower than that of pea, and that there are no more than two articles for the species concerned. The symbol '0' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are more than two articles for the species concerned. The symbol '(0)' indicates that the value of the criterion for the given species is not significantly different from that of pea, and that there are no more than two articles available for the species concerned. The symbol '+' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are more than two articles available for the species concerned. The symbol '(+)' indicates that the value of the criterion for the given species is significantly higher than that of pea, and that there are no more than two articles available for the species concerned.

	Europe						North America						Oceania					
	Grain biomass ratio <sup>a</sup>	Total aerial biomass ratio <sup>a</sup>	Grain crude protein ratio <sup>a,b</sup>	Grain gross energy <sup>c</sup> ratio <sup>a,b</sup>	Grain nitrogen content ratio <sup>a</sup>	Grain biomass variance ratio <sup>a</sup>	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain nitrogen content ratio	Water use ratio <sup>a</sup>	Grain biomass variance ratio	Grain biomass ratio	Total aerial biomass ratio	Grain crude protein ratio	Grain gross energy ratio	Grain biomass variance ratio
<i>Cicer arietinum</i>	-	0	-	0	(0)	0	-	-	-	(-)	+	-	-	-	-	-	-	+
<i>Glycine max</i>	0		+	+		+	-	(0)	0	-	-	-	(-)	0	0	0	0	+
<i>Lathyrus aphaca</i>													(-)	0	0	0	0	
<i>Lathyrus cicera</i>													(0)	0	0	0	0	
<i>Lathyrus cymenium</i>													(0)	0	0	0	0	
<i>Lathyrus ochrus</i>													(0)	0	0	0	0	
<i>Lathyrus sativus</i>							(0)	(0)	(0)	(0)	(+)	0	0	0	0	0	0	
<i>Lens culinaris</i>	(-)	(0)	(-)	(-)	(-)	+	-	0	-	-	0	0	-	-	-	-	-	+
<i>Lupinus albus</i>	-	(-)				+							-	-	-	-	-	+
<i>Lupinus angustifolius</i>	-	0	0	0	0	+	-	-	-	(-)	(0)	0	0	0	0	0	0	+
<i>Lupinus atlanticus</i>						+							(0)	0	0	0	0	+
<i>Lupinus luteus</i>	(0)	(0)	(0)	(0)	(0)	+							(-)	(0)	(-)	(0)	+	+
<i>Lupinus pilosus</i>						+							(0)	0	0	0	0	+
<i>Phaseolus vulgaris</i>							-	-	-	(-)	(0)	0						
<i>Vicia articulata</i>													(0)	0	0	0	0	
<i>Vicia benghalensis</i>													(-)	0	0	0	0	
<i>Vicia ervilia</i>													(-)	0	0	0	0	
<i>Vicia faba</i>	0	0	0	0	0	0	-	(0)	0	-	(0)	0	+	0	+	+	+	+
<i>Vicia hybrida</i>													(-)	0	0	0	0	+
<i>Vicia narbonensis</i>													0	0	0	0	0	+
<i>Vicia sativa</i>													0	0	0	0	0	+
<i>Vicia villosa</i>													(-)	0	0	0	0	+

<sup>a</sup>Experimental data from the selected articles are used to estimate the given criterion.

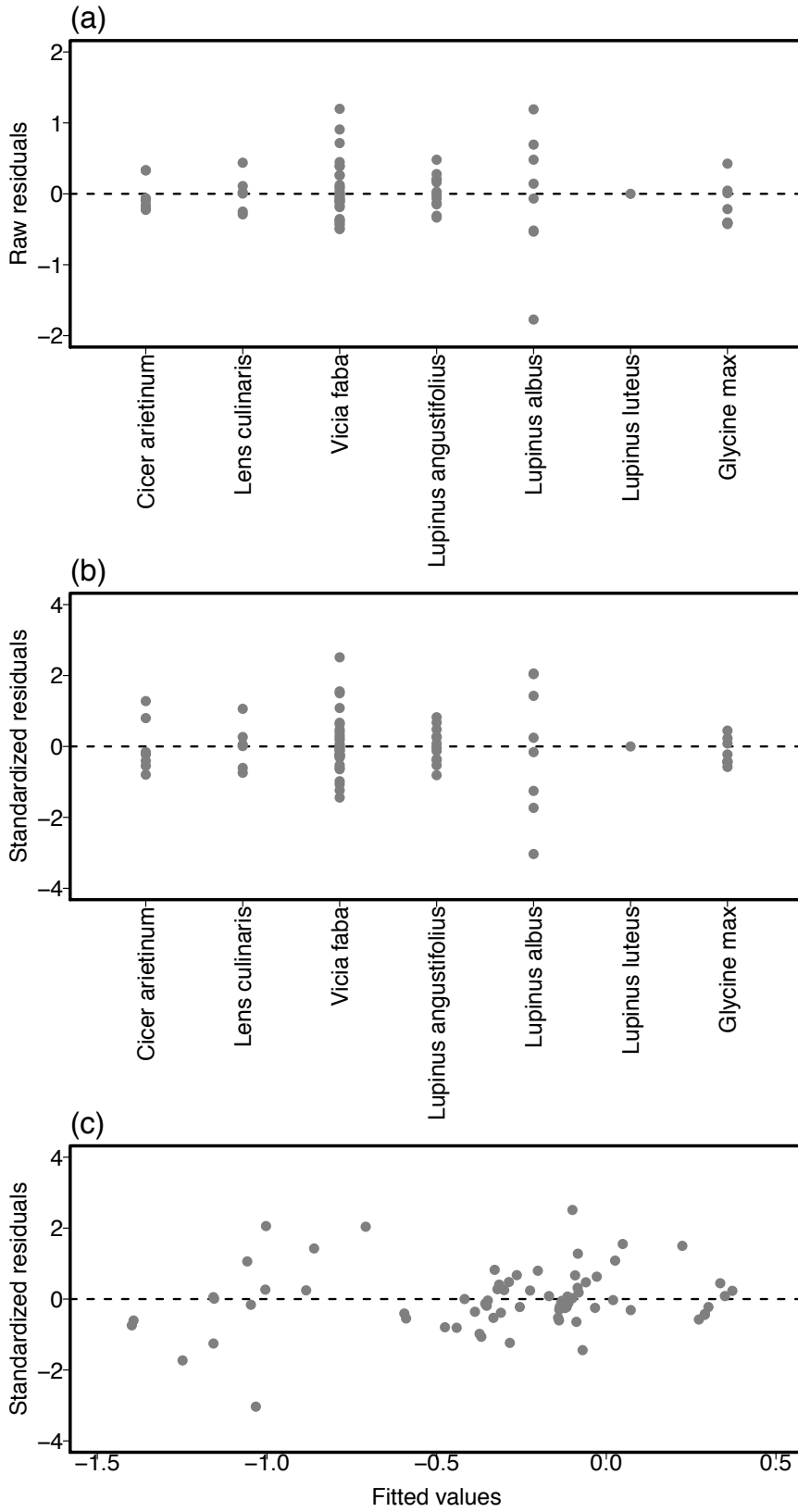
<sup>b</sup>Nutritional data from the FEEDIPEDIA database are used to estimate the given criterion.

<sup>c</sup>Statistical data from the FAOSTAT database are used to estimate the given criterion.

**Table S8.** Countries for FAOSTAT region items included in the analysis.

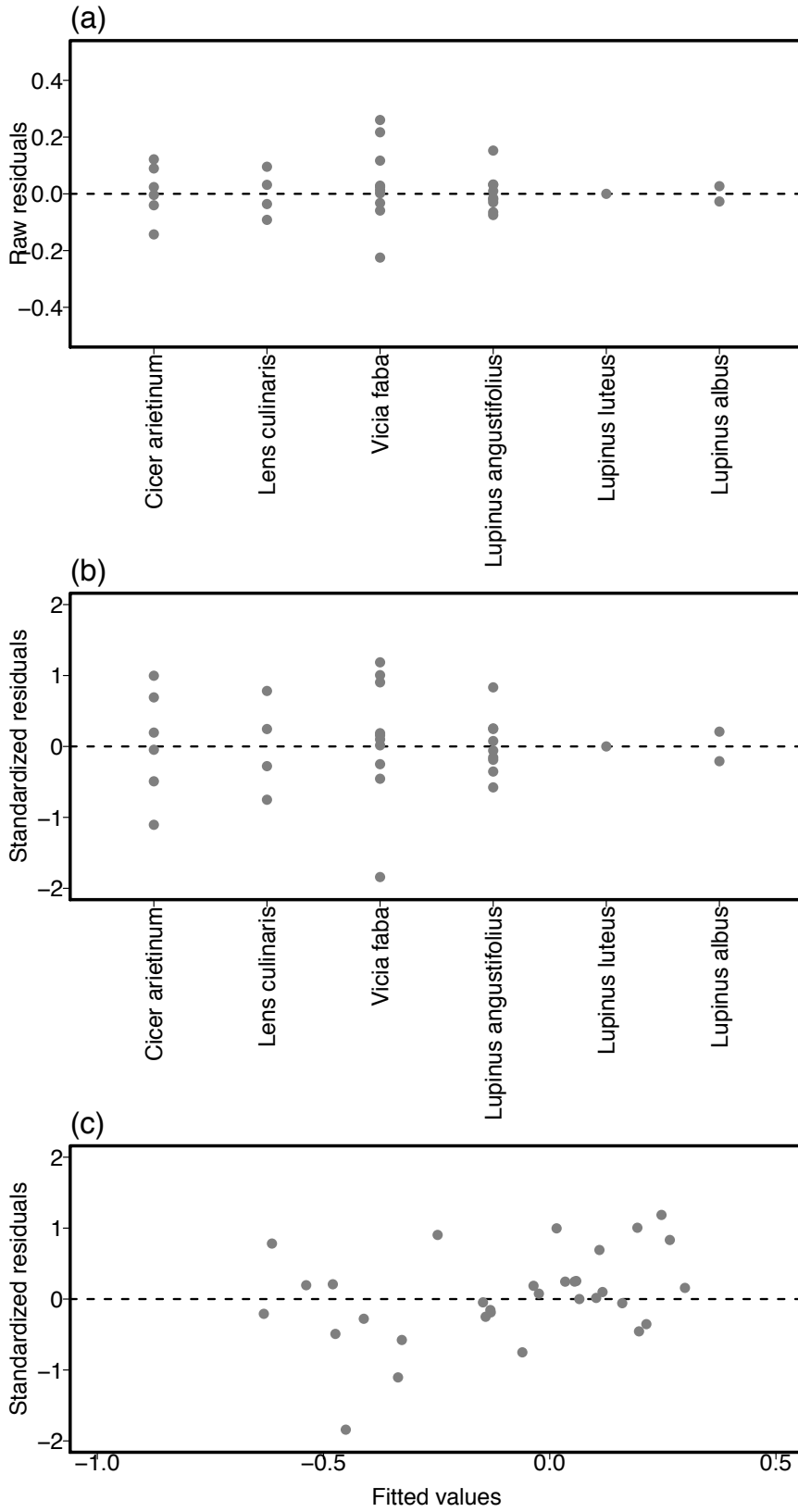
FAOSTAT region item	Country
Europe	Albania
	Austria
	Belarus
	Belgium
	Bosnia and Herzegovina
	Bulgaria
	Croatia
	Czech Republic
	Denmark
	Estonia
	Finland
	France
	Germany
	Greece
	Hungary
	Iceland
	Ireland
	Italy
	Latvia
	Lithuania
	Luxembourg
	Montenegro
	Netherlands
	Norway
	Poland
	Portugal
	Republic of Moldova
	Romania
	Russian Federation
	Serbia
	Slovakia
	Slovenia
	Spain
Sweden	
Switzerland	
The former Yugoslav Republic of Macedonia	
Ukraine	
United Kingdom of Great Britain and Northern Ireland	
Northern America	Canada
	United States of America
Australia & New Zealand	Australia
	New Zealand

**Figure S1.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain biomass in Europe. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

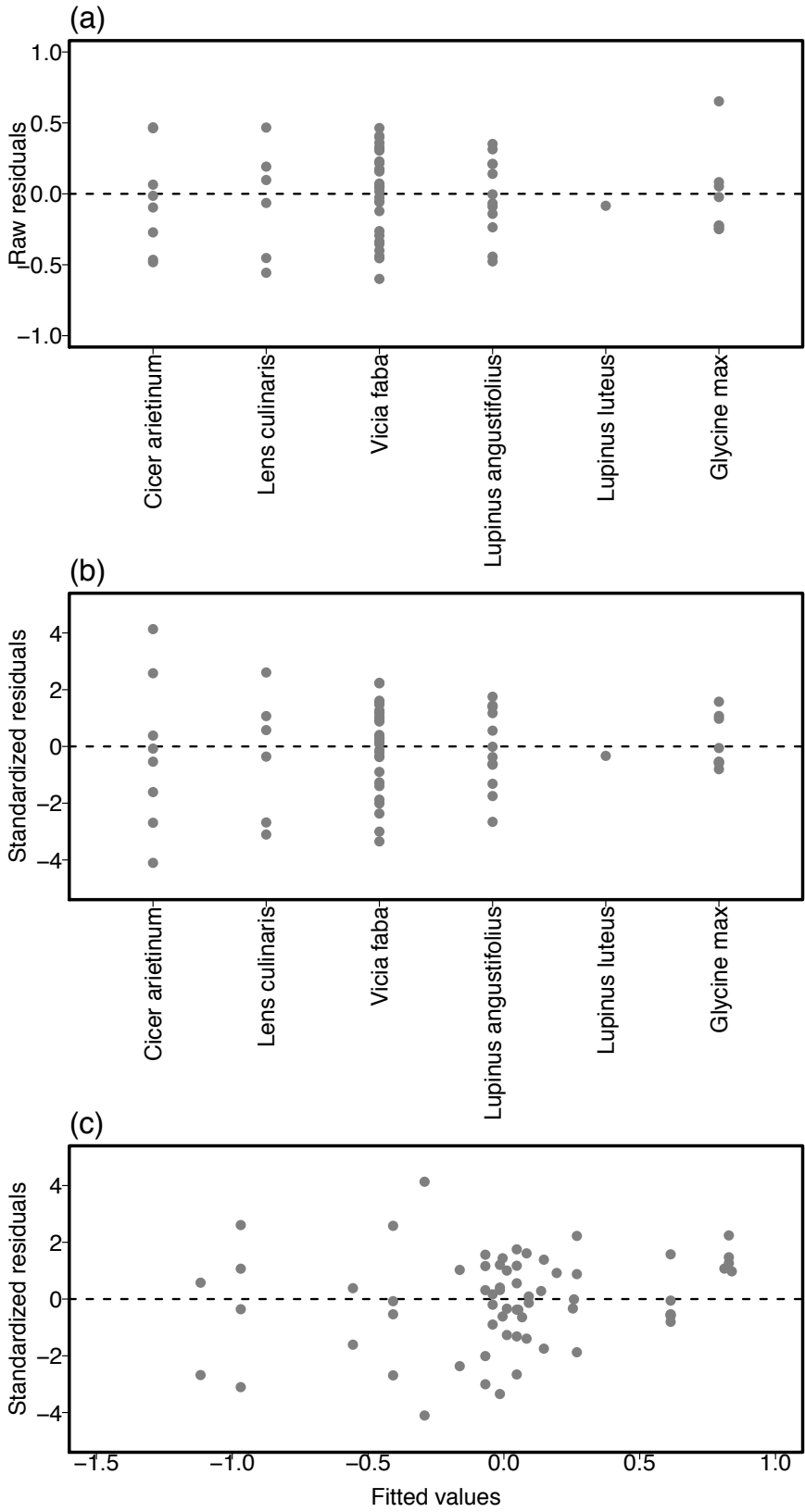




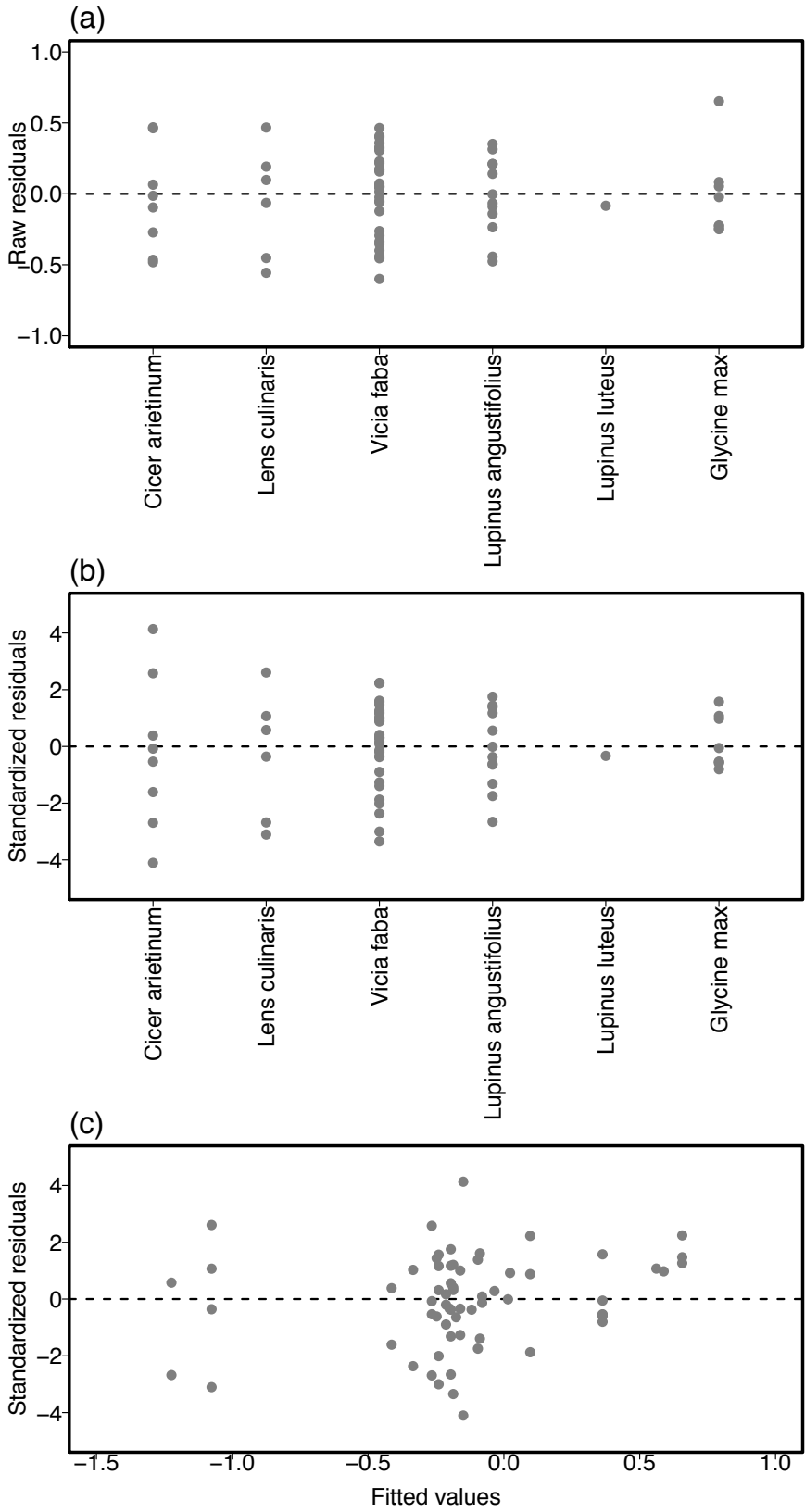
**Figure S2.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for total aerial biomass in Europe. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



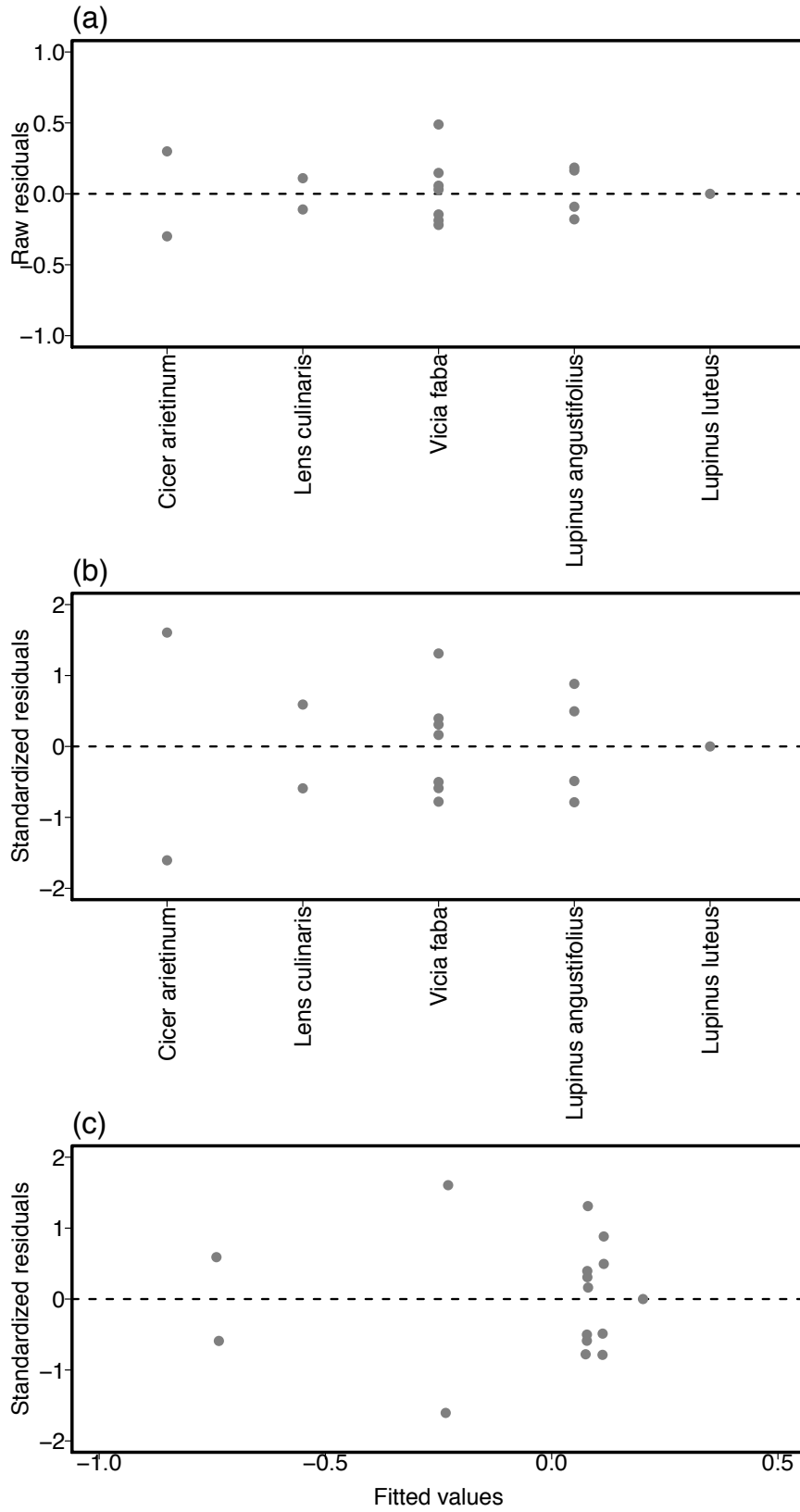
**Figure S3.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain crude protein in Europe. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



**Figure S4.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain gross energy in Europe. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

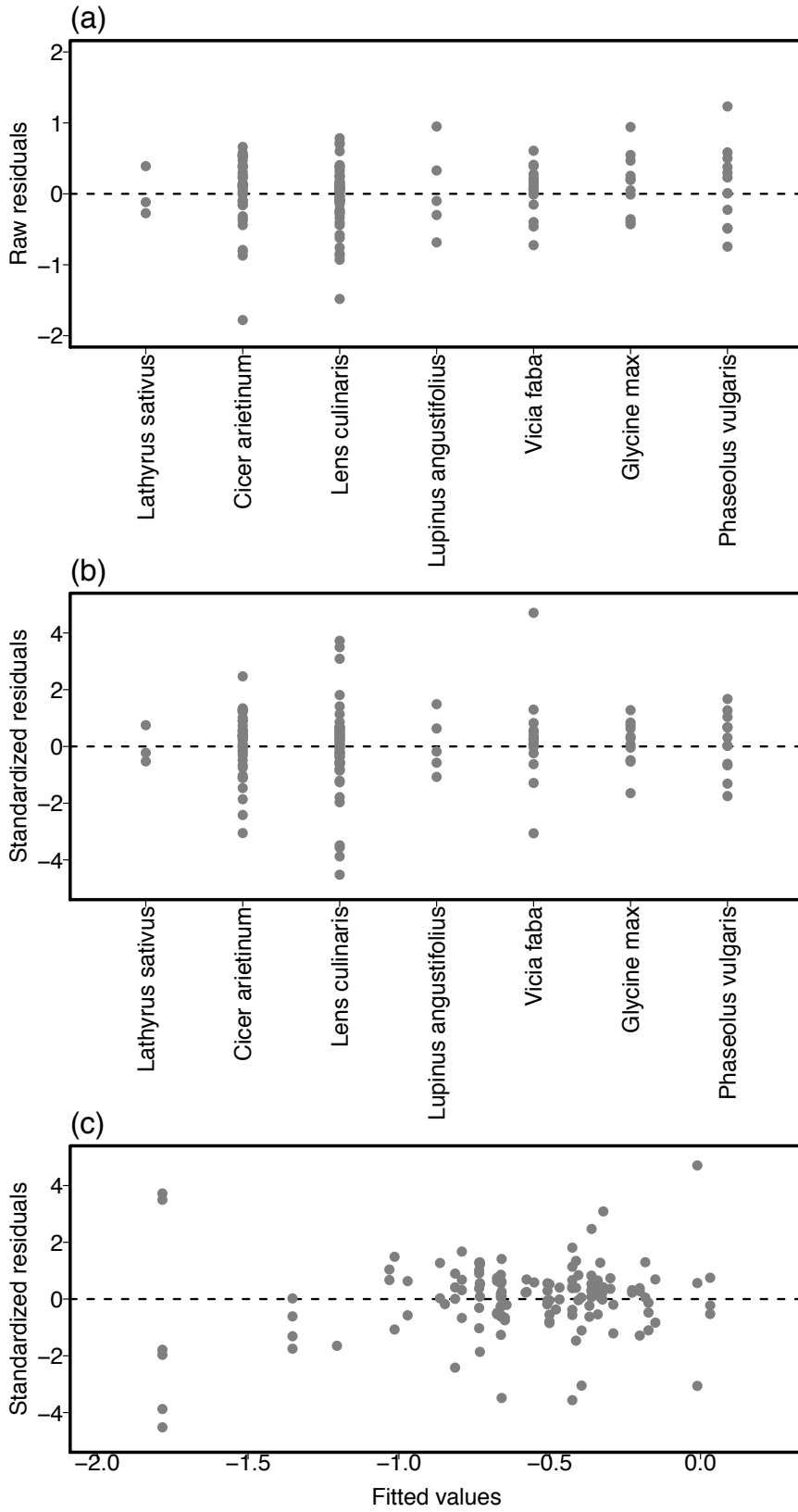


**Figure S5.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain nitrogen content in Europe. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

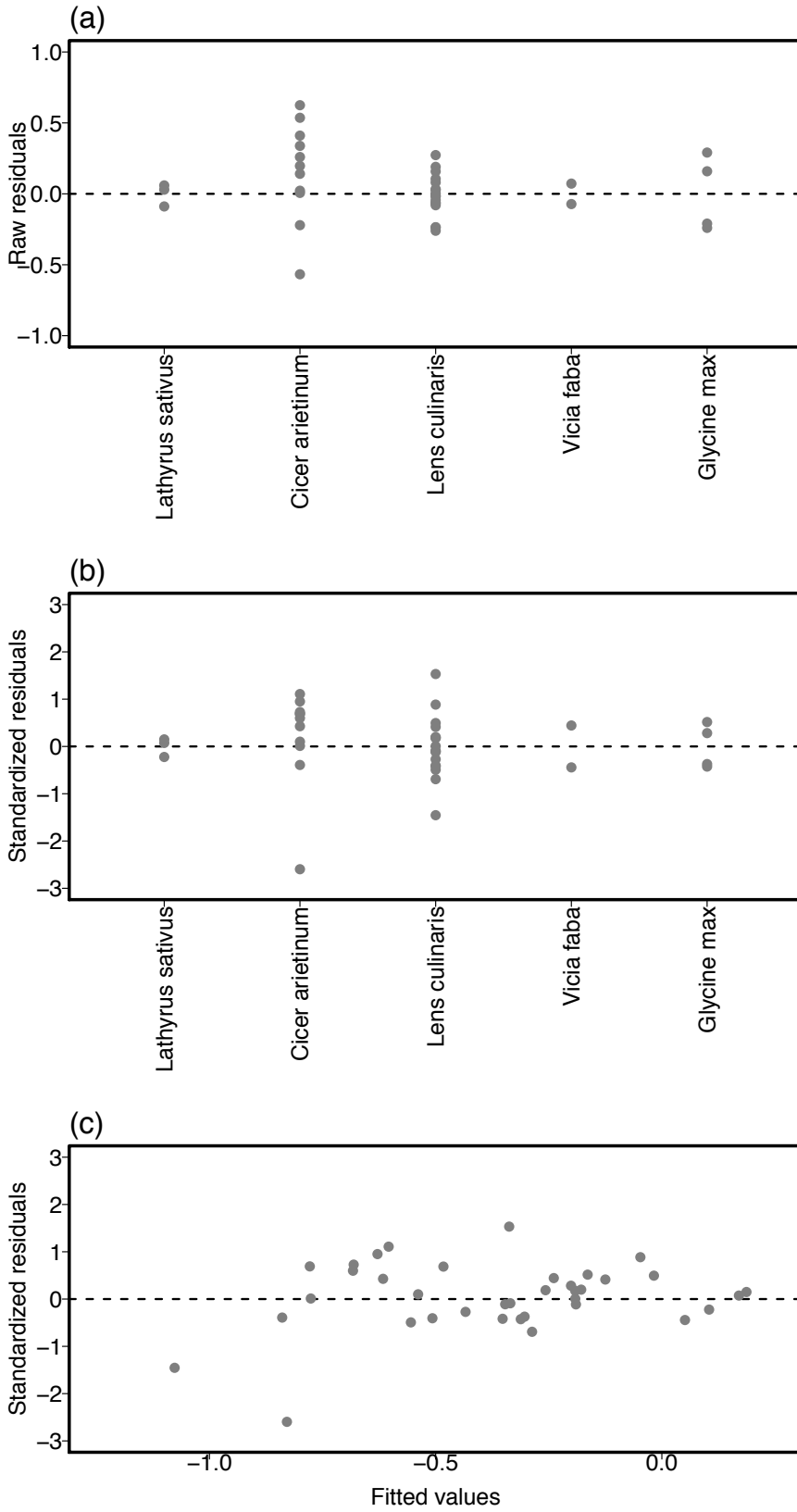




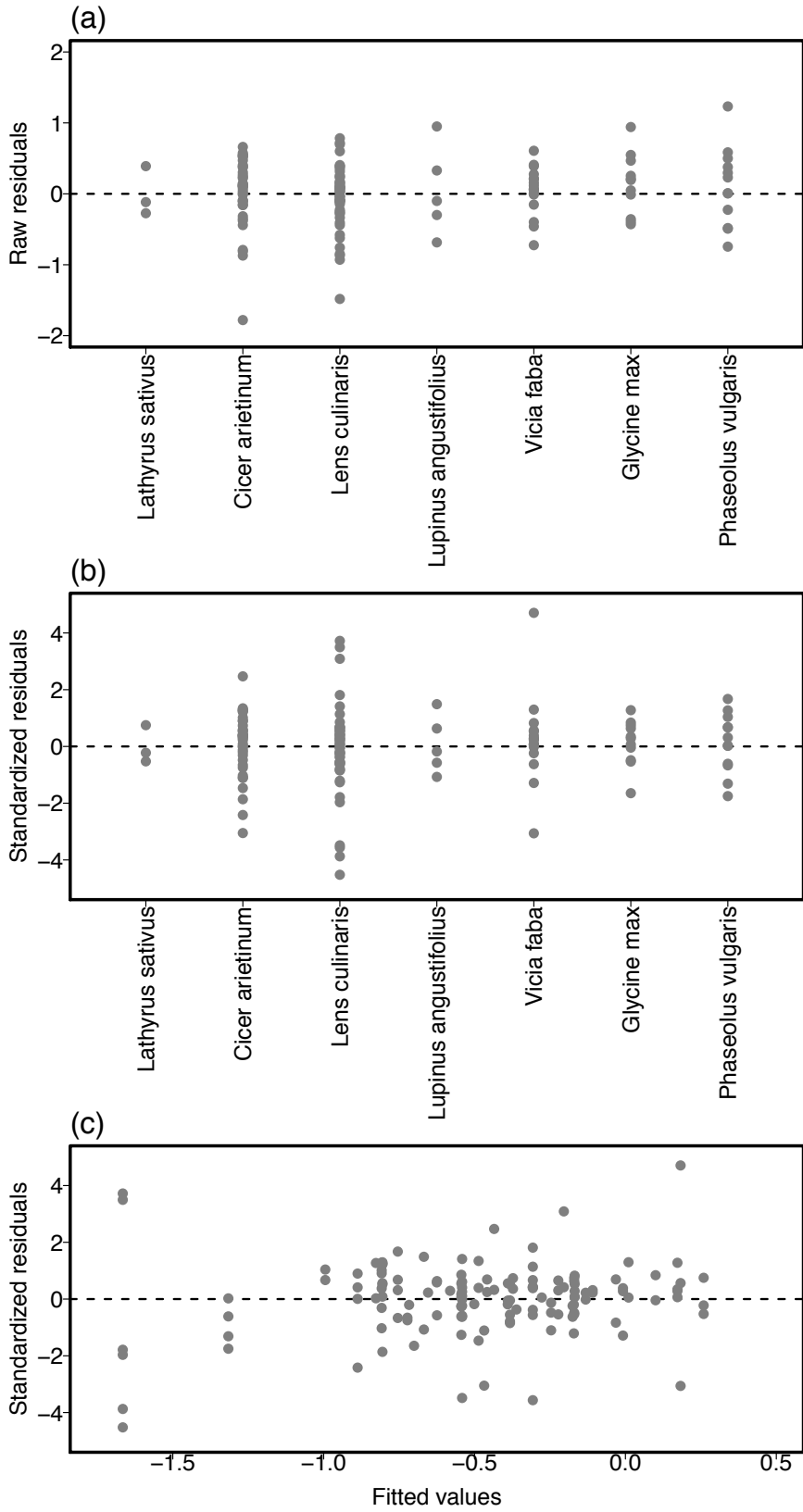
**Figure S6.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain biomass in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



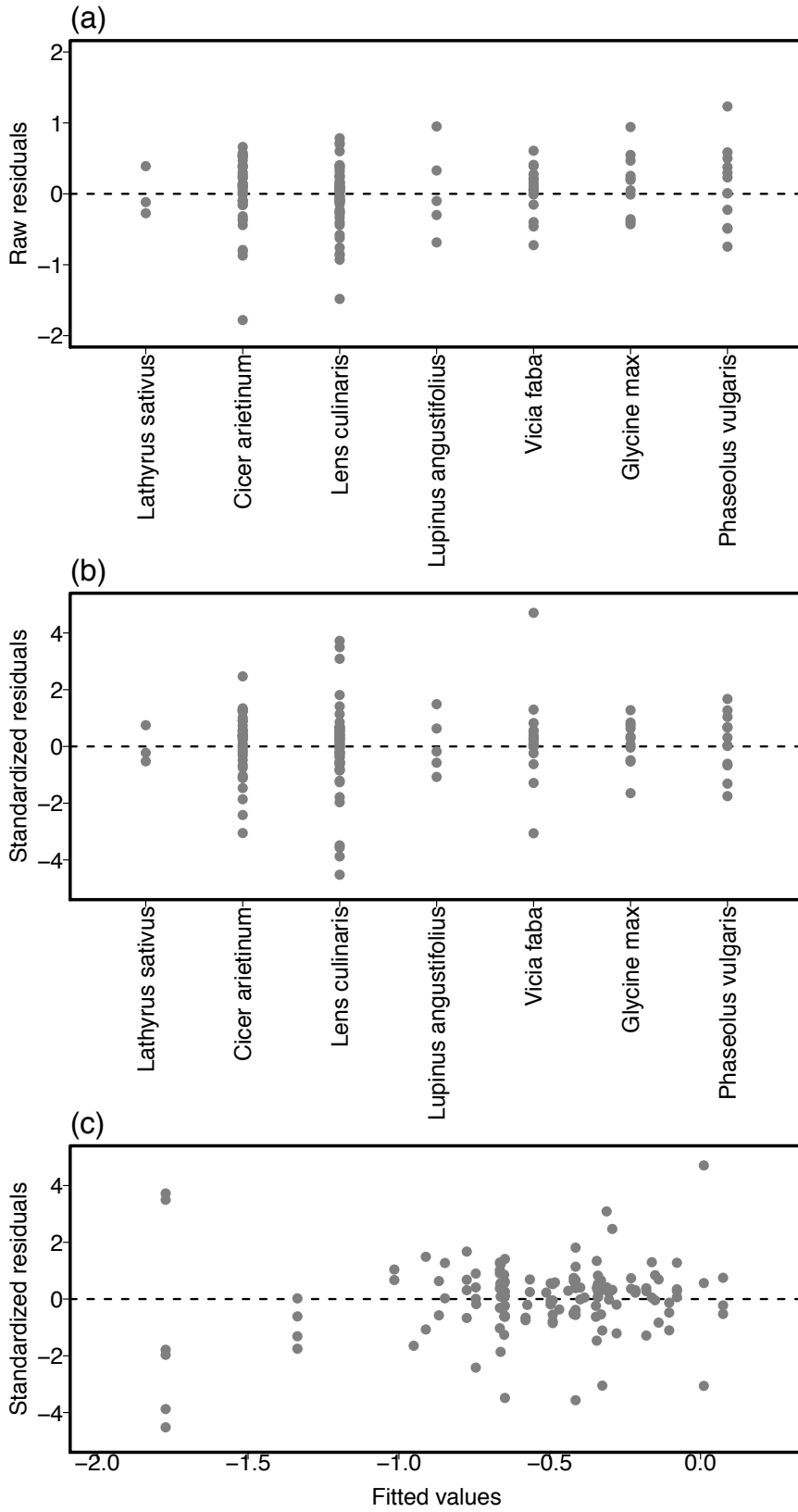
**Figure S7.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for total aerial biomass in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



**Figure S8.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain crude protein in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

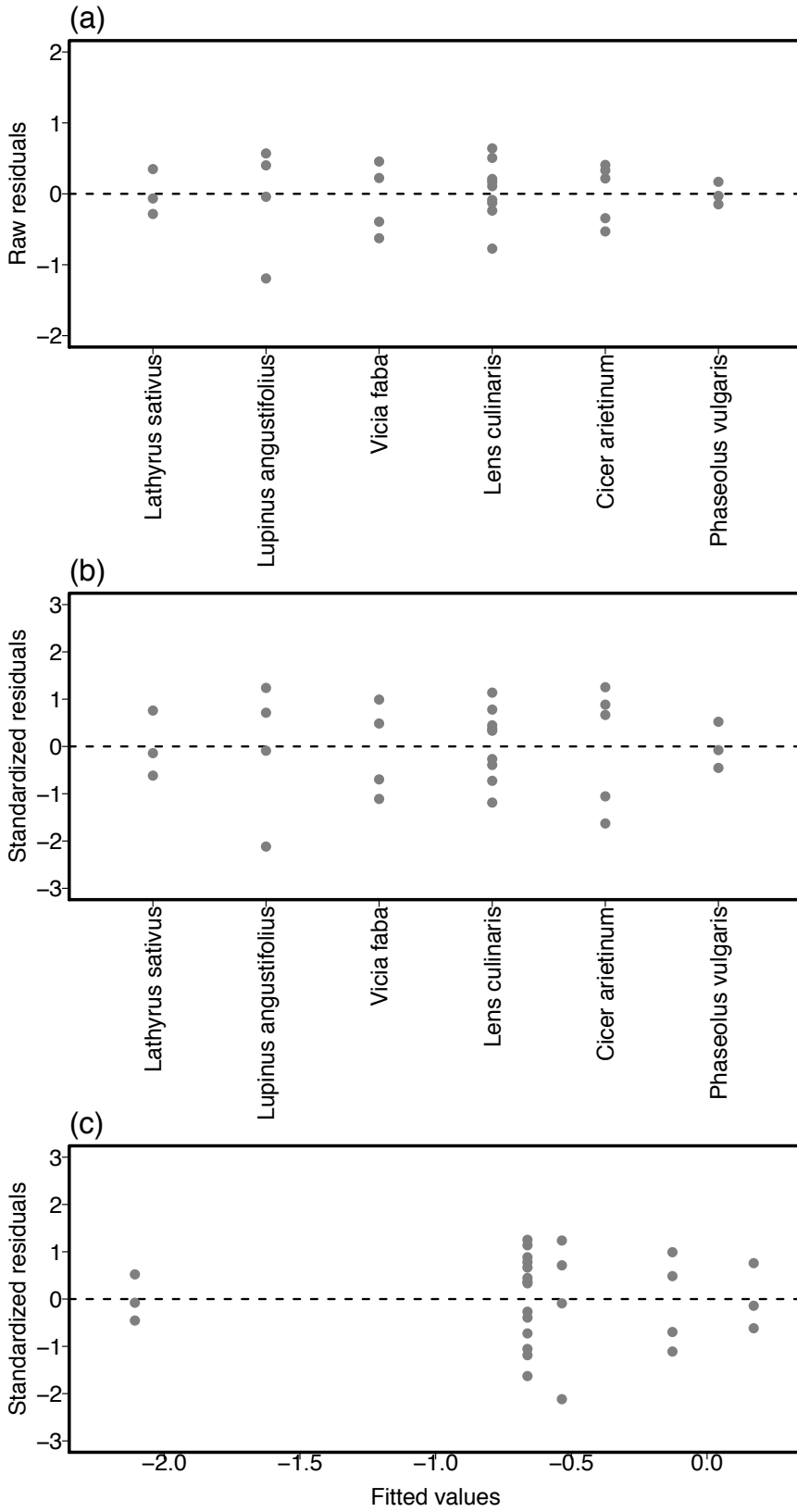


**Figure S9.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain gross energy in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

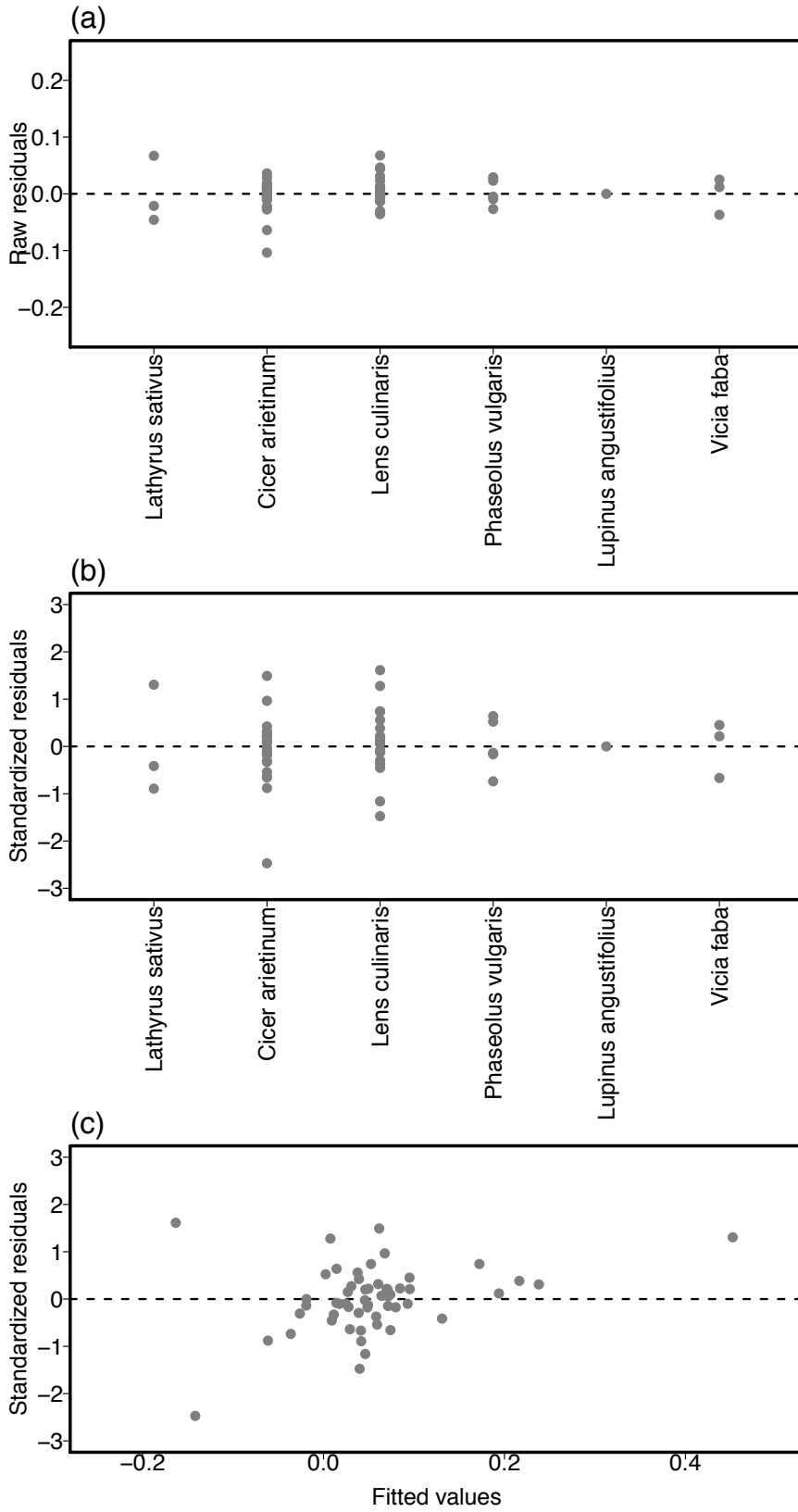




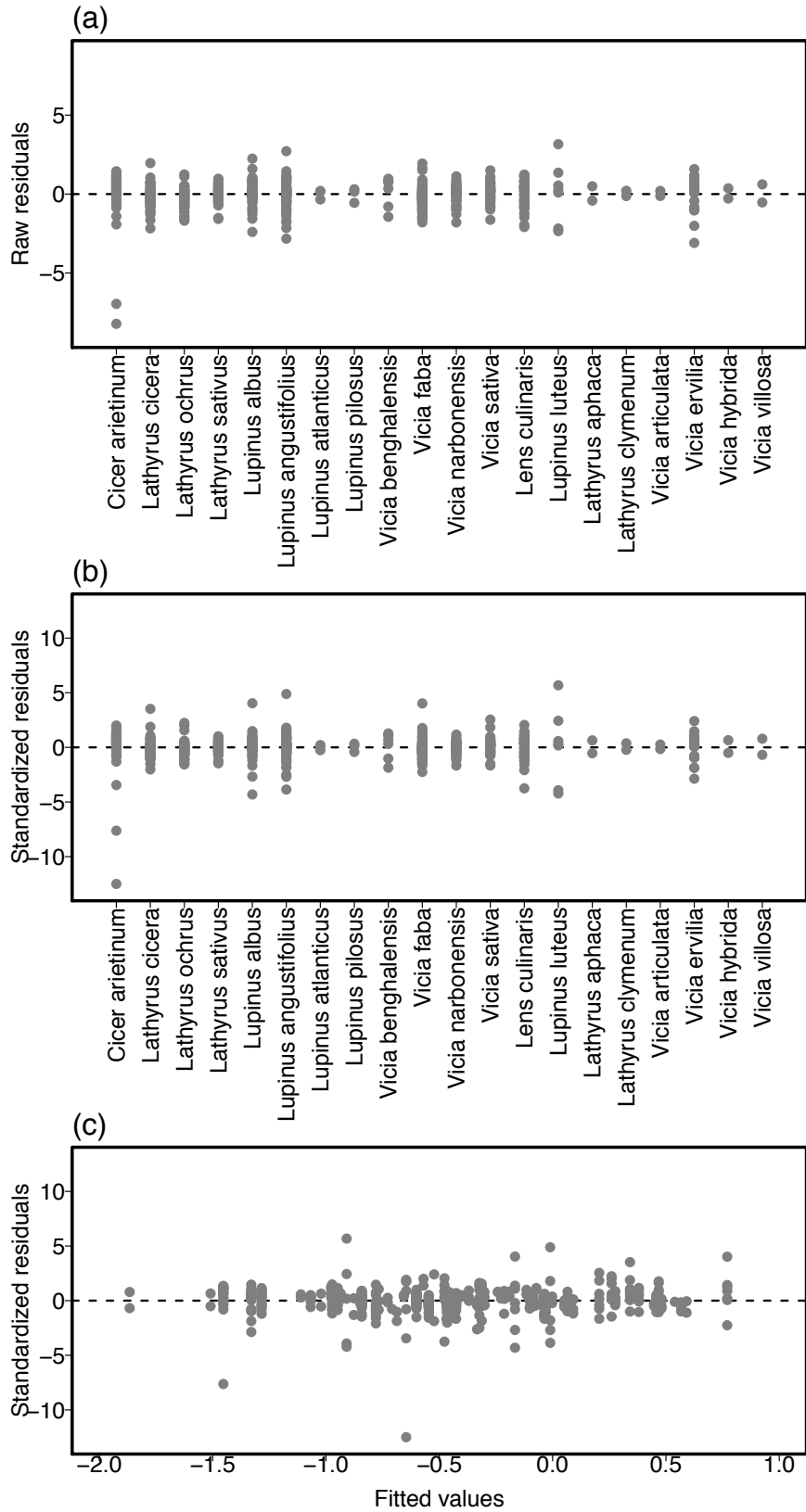
**Figure S10.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain nitrogen content in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



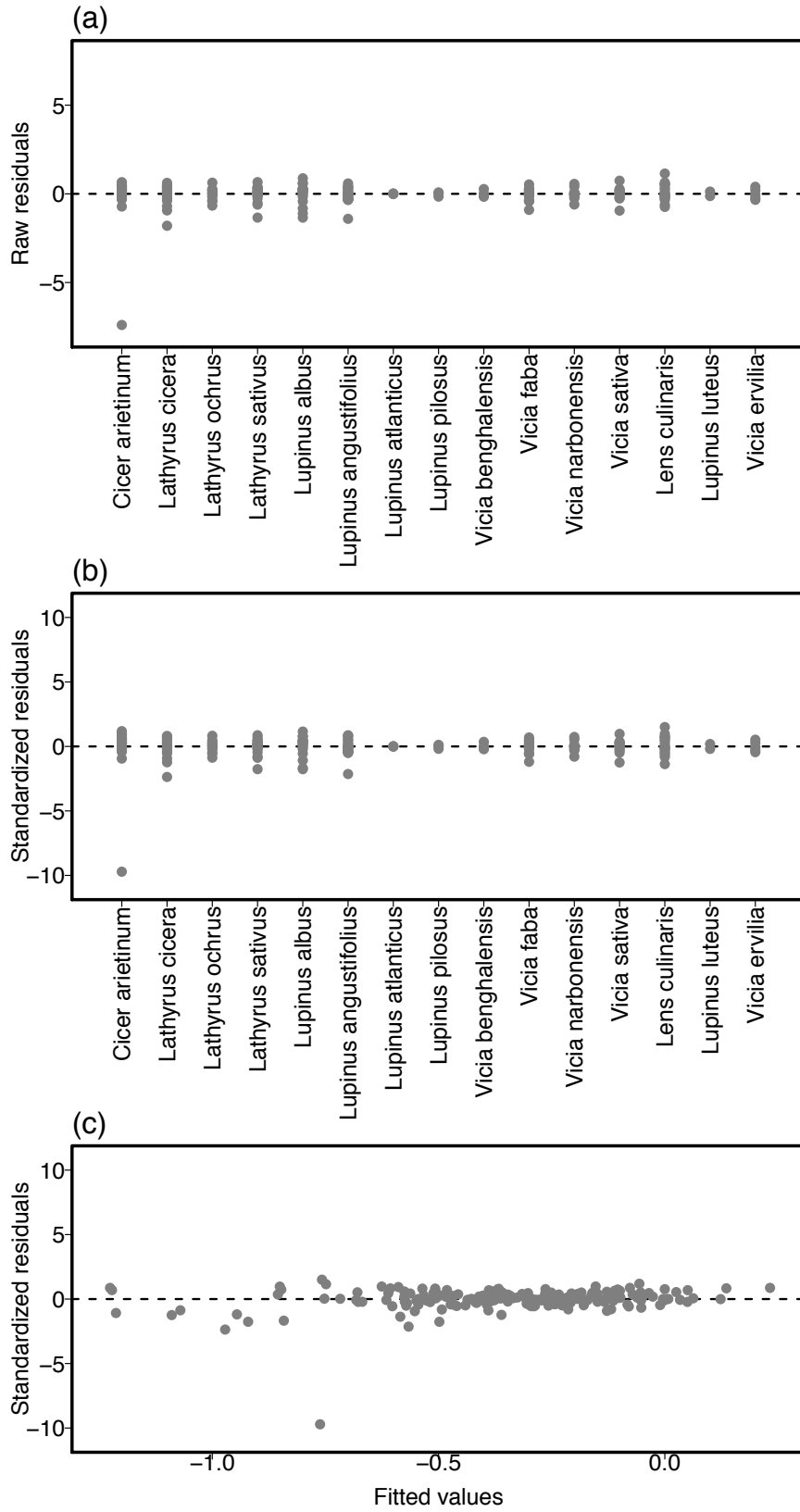
**Figure S11.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for water use in North America. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



**Figure S12.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain biomass in Oceania. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

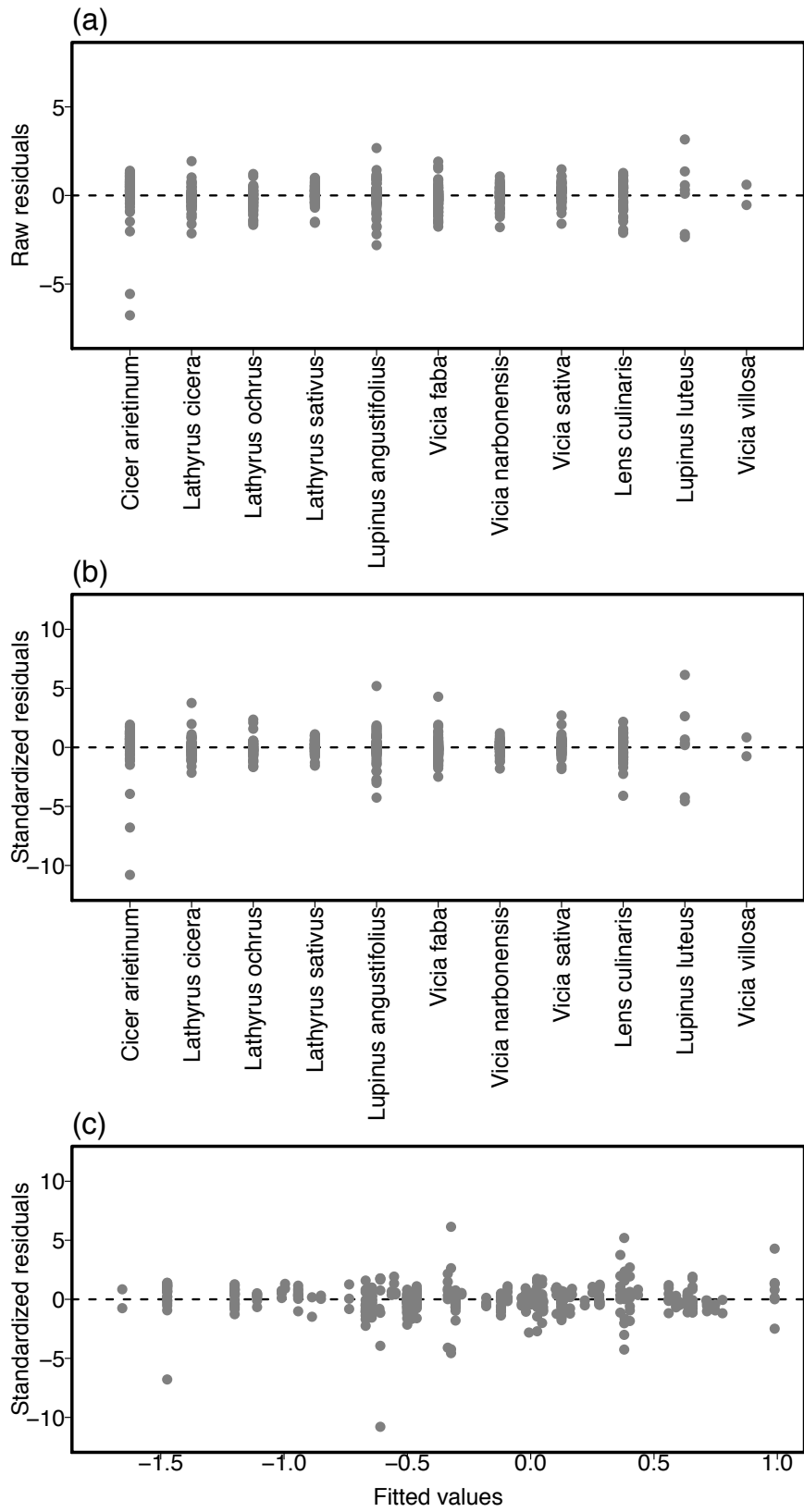


**Figure S13.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for total aerial biomass in Oceania. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).

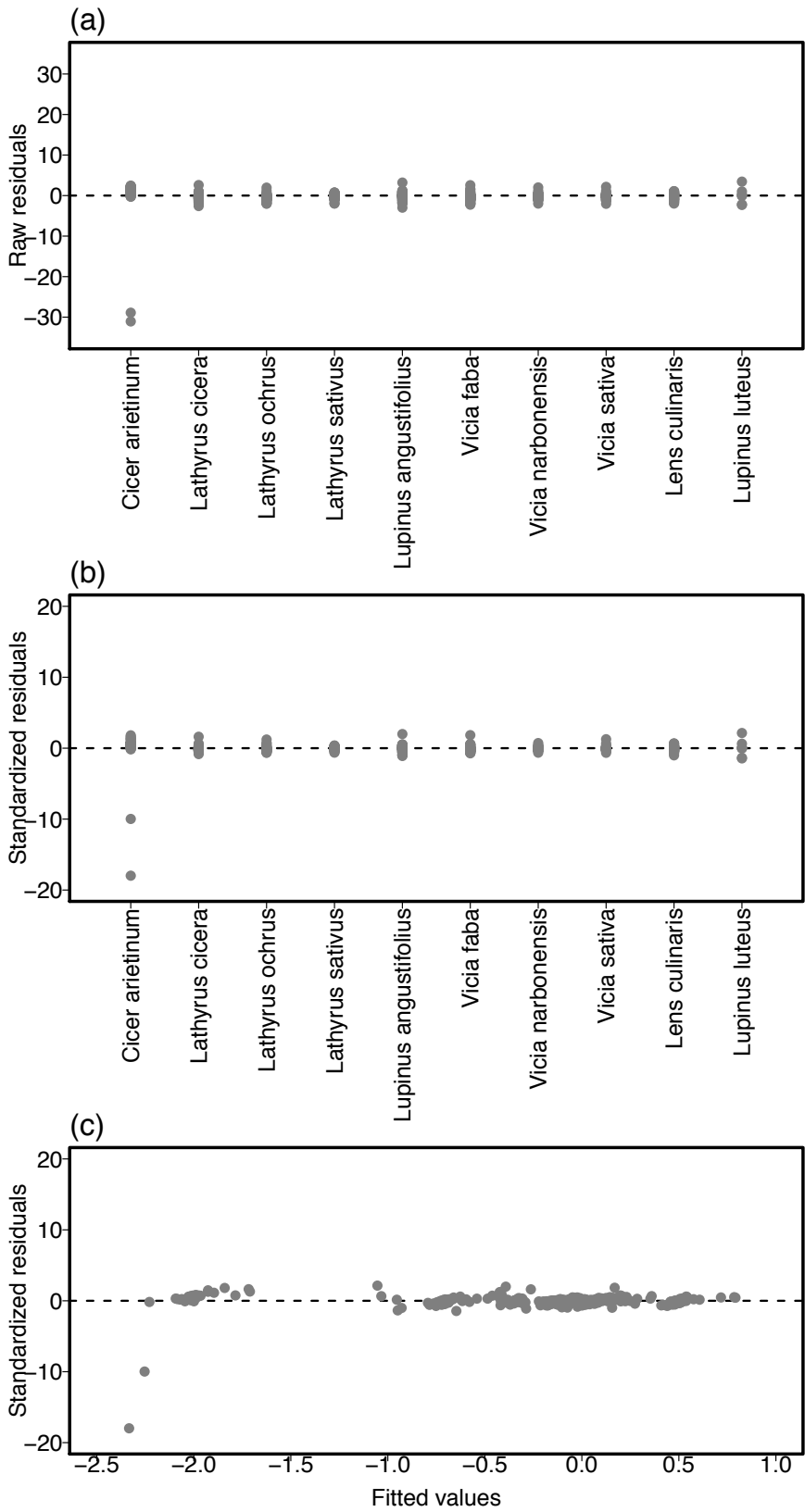




**Figure S14.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain crude protein in Oceania. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).



**Figure S15.** Distribution of raw residuals by species (a), standardized residuals by species (b), and standardized residuals versus fitted values across all field site\*growing season combinations and selected articles (c) for grain gross energy in Oceania. Weighted experimental values are used. The best model variant including either one random 'field site\*growing season' effect or two random 'article' and 'field site\*growing season' effects is used. The scientific names of the species correspond to the common names between parentheses: *Cicer arietinum* (chickpea), *Glycine max* (soybean), *Lathyrus aphaca* (yellow pea), *Lathyrus cicera* (red pea), *Lathyrus clymenum* (cicercha purpurina), *Lathyrus ochrus* (cyprus vetch), *Lathyrus sativus* (white pea), *Lens culinaris* (lentil), *Lupinus albus* (white lupin), *Lupinus angustifolius* (narrowleaf lupin), *Lupinus atlanticus* (Not Available), *Lupinus luteus* (yellow lupin), *Lupinus pilosus* (blue lupin), *Phaseolus vulgaris* (kidney bean), *Pisum sativum* (pea), *Vicia articulata* (oneflower vetch), *Vicia benghalensis* (purple vetch), *Vicia ervilia* (blister vetch), *Vicia faba* (faba bean), *Vicia hybrida* (hairy yellow vetch), *Vicia narbonensis* (purple broad vetch), *Vicia sativa* (garden vetch), and *Vicia villosa* (winter vetch).





## **Annexes du Chapitre 4**



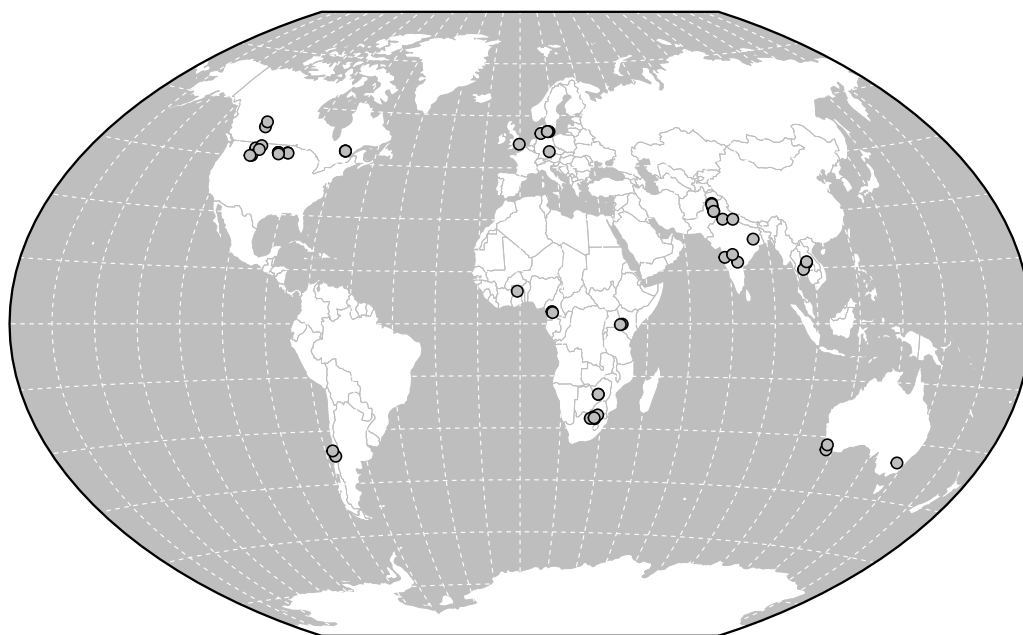
**Note S1.** References for the articles included in the dataset.

1. Ahmad T, Hafeez FY, Mahmood T, Malik KA (2001) Residual effect of nitrogen fixed by mungbean (*Vigna radiata*) and blackgram (*Vigna mungo*) on subsequent rice and wheat crops. *Aust J Exp Agr* **41**(2):245–248.
2. Armstrong EL, Heenan DP, Pate JS, Unkovich MJ (1997) Nitrogen benefits of lupins, field pea, and chickpea to wheat production in south-eastern Australia. *Aust J Agr Res* **48**(1):39–47.
3. Badaruddin M, Meyer DW (1994) Grain legume effects on soil nitrogen, grain yield, and nitrogen nutrition of wheat. *Crop Sci* **34**(5):1304–1309.
4. Bharambe PR, Sondge VD, Rodge RP, Oza SR (1990) Performance of different crop sequences under various levels of irrigation in Vertisols. *Indian J Agr Sci* **60**(6):402–405.
5. Bloem AA, Barnard RO (2001) Effect of annual legumes on soil nitrogen and on the subsequent yield of maize and grain sorghum. *S Afr J Plant Soil* **18**(2):56–61.
6. Cheruiyot EK, Mumera LM, Nakhone LN, Mwonga SM (2003) Effect of legume-managed fallow on weeds and soil nitrogen in following maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.) crops in the Rift Valley highlands of Kenya. *Aust J Exp Agr* **43**(6):597–604.
7. Dakora FD, Aboyinga RA, Mahama Y, Apaseku J (1987) Assessment of N<sub>2</sub> fixation in groundnut (*Arachis hypogaea* L.) and cowpea (*Vigna unguiculata* L. Walp.) and their relative N contribution to a succeeding maize crop in northern Ghana. *MIRCEN J Appl Microb* **3**(4):389–399.
8. De R, Rao YY, Ali W (1983) Grain and fodder legumes as preceding crops affecting the yield and N economy of rice. *J Agr Sci* **101**(2):463–466.
9. Espinoza S, et al. (2012) Contribution of legumes to wheat productivity in Mediterranean environments of central Chile. *Field Crop Res* **133**(11):150–159.
10. Ghosh PK, Singh NP (1996) Production potential of summer legumes-maize (*Zea mays*) sequence under varying levels of nitrogen. *Indian J Agron* **41**(4):525–528.
11. Gregory PJ (1998) Alternative crops for Duplex soils: growth and water use of some cereal, legume, and oilseed crops, and pastures. *Aust J Agr Res* **49**(1):21–32.
12. Hauggaard-Nielsen H, Mundus S, Jensen ES (2009) Nitrogen dynamics following grain legumes and subsequent catch crops and the effects on succeeding cereal crops. *Nutr Cycl Agroecosys* **84**(3):281–291.
13. Hauggaard-Nielsen H, Mundus S, Jensen ES (2012) Grass-clover undersowing affects nitrogen dynamics in a grain legume-cereal arable cropping system. *Field Crop Res* **136**:23–31.
14. Hayat R, Ali S (2004) Potential of summer legumes to fix nitrogen and benefit wheat crop under rainfed condition. *J Agron* **3**(4):273–281.
15. Hayat R, Ali S (2010) Nitrogen fixation of legumes and yield of wheat under legumes-wheat rotation in Pothwar. *Pak J Bot* **42**(4):2317–2326.
16. Jawale SM, Jadhav AS, Patil VG (1998) Productivity and economics of legumes-winter sorghum (*Sorghum bicolor*) double cropping systems under tillage practices in western Maharashtra. *Indian J Agron* **43**(3):396–403.



17. Jemo M, et al. (2006) Phosphorus benefits from grain-legume crops to subsequent maize grown on acid soils of southern Cameroon. *Plant Soil* **284**(1):385–397.
18. Jensen CR, et al. (2004) The effect of lupins as compared with peas and oats on the yield of the subsequent winter barley crop. *Eur J Agron* **20**(4):405–418.
19. Krupinsky JM, Tanaka DL, Merrill SD, Liebig MA, Hanson JD (2006) Crop sequence effects of 10 crops in the northern Great Plains. *Agr Syst* **88**(2–3):227–254.
20. Lafond J, Pageau D (2007) Nitrogen and non nitrogen benefit of legumes to a subsequent spring barley crop. *Can J Soil Sci* **87**(4):445–454.
21. Maidl FX, Haunz FX, Panse A, Fischbeck G (1996) Transfer of grain legume nitrogen within a crop rotation containing winter wheat and winter barley. *J Agron Crop Sci* **176**(1):47–57.
22. Malik RS, et al. (2015) Dynamic crop sequencing in western Australian cropping systems. *Crop Pasture Sci* **66**(6):594–609.
23. McDonagh JF, Toomsan B, Limpinuntana V, Giller KE (1995) Grain legumes and green manures as pre-rice crops in northeast Thailand I. Legume N<sub>2</sub>-fixation, production and residual nitrogen benefits to rice. *Plant Soil* **177**(1):111–126.
24. McEwen J, Darby RJ, Hewitt MV, Yeoman DP (1989) Effects of field beans, fallow, lupins, oats, oilseed rape, peas, ryegrass, sunflowers and wheat on nitrogen residues in the soil and on the growth of a subsequent wheat crop. *J Agr Sci* **115**(2):209–219.
25. Miller PR, Holmes JA (2005) Cropping sequence effects of four broadleaf crops on four cereal crops in the northern Great Plains. *Agron J* **97**(1):189–200.
26. Ncube B, Twomlow SJ, Van Wijk MT, Dimes JP, Giller KE (2007) Productivity and residual benefits of grain legumes to sorghum under semi-arid conditions in southwestern Zimbabwe. *Plant Soil* **299**(1):1–15.
27. Singh K, Singh S (1991) Different wheat (*Triticum aestivum*)-based cropping systems, and their fertilizer requirement, yield and economic returns. *Indian J Agr Sci* **61**(10):709–714.
28. Singh MK, Thakur R, Verma UN, Pal SK (1995) Production potential of crop sequences in plateau region of Bihar. *Indian J Agr Sci* **65**(4):242–245.
29. Suwanarit A, Suwannarat C, Chotechaungmanirat S (1986) Quantities of fixed N and effects of grain legumes on following maize, and N and P status of soil as indicated by isotopes. *Plant Soil* **93**(2):249–258.
30. Tanaka DL, Anderson RL, Rao SC (2005) Crop sequencing to improve use of precipitation and synergize crop growth. *Agron J* **97**(2):385–390.
31. Tanaka DL, Krupinsky JM, Merrill SD, Liebig MA, Hanson JD (2007) Dynamic cropping systems for sustainable crop production in the northern Great Plains. *Agron J* **99**(4):904–911.
32. Toomsan B, et al. (2000) Biological N<sub>2</sub> fixation and residual N benefit of pre-rice leguminous crops and green manures. *Neth J Agr Sci* **48**(1):19–29.
33. Williams CM, et al. (2014) Effects of three pulse crops on subsequent barley, canola, and wheat. *Agron J* **106**(2):343–350.

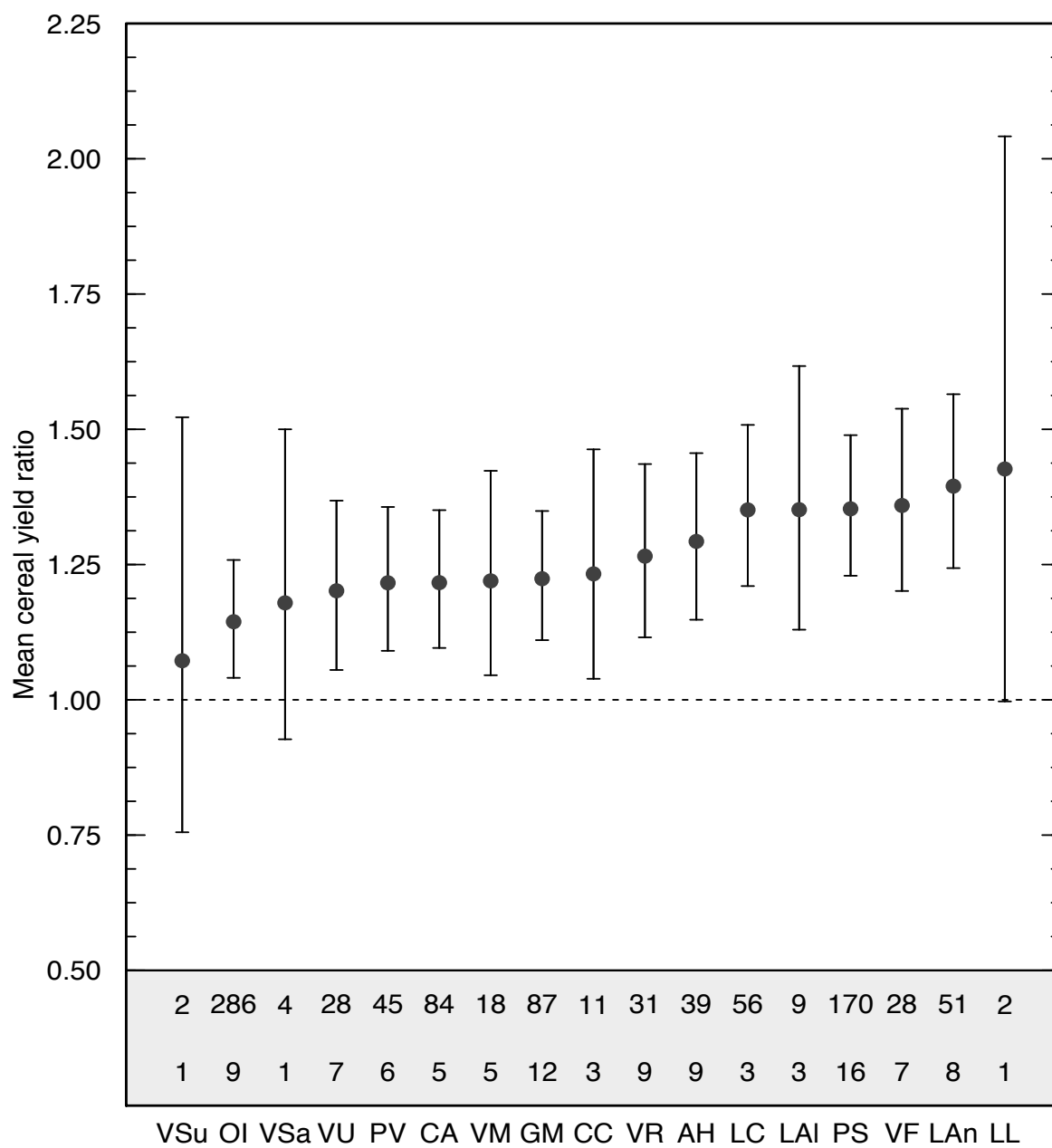
Fig. S1. Field sites (indicated by filled circles) included in the meta-analysis.



**Table S1.** Scientific names (ranked alphabetically) and crop categories for each species grown as either preceding and/or subsequent crop.

Scientific name	Crop category	Preceding crop	Subsequent crop
<i>Arachis hypogaea</i>	Grain legume	Yes	No
<i>Avena sativa</i>	Cereal	Yes	Yes
<i>Brassica juncea</i>	Oilseed	Yes	No
<i>Brassica napus</i>	Oilseed	Yes	No
<i>Cajanus cajan</i>	Grain legume	Yes	No
<i>Carthamus tinctorius</i>	Oilseed	Yes	No
<i>Cicer arietinum</i>	Grain legume	Yes	No
<i>Crambe abyssinica</i>	Oilseed	Yes	No
<i>Fagopyrum esculentum</i>	Cereal	Yes	Yes
<i>Glycine max</i>	Grain legume	Yes	No
<i>Helianthus annuus</i>	Oilseed	Yes	No
<i>Hordeum vulgare</i>	Cereal	Yes	Yes
<i>Lens culinaris</i>	Grain legume	Yes	No
<i>Linum usitatissimum</i>	Oilseed	Yes	No
<i>Lupinus albus</i>	Grain legume	Yes	No
<i>Lupinus angustifolius</i>	Grain legume	Yes	No
<i>Lupinus luteus</i>	Grain legume	Yes	No
<i>Oryza sativa</i>	Cereal	Yes	Yes
<i>Panicum miliaceum</i>	Cereal	Yes	Yes
<i>Pennisetum glaucum</i>	Cereal	Yes	No
<i>Phaseolus vulgaris</i>	Grain legume	Yes	No
<i>Pisum sativum</i>	Grain legume	Yes	No
<i>Sorghum bicolor</i>	Cereal	Yes	Yes
<i>Triticum aestivum</i>	Cereal	Yes	Yes
<i>Triticum durum</i>	Cereal	No	Yes
<i>Triticum sativum</i>	Cereal	No	Yes
<i>Vicia faba</i>	Grain legume	Yes	No
<i>Vicia sativa</i>	Grain legume	Yes	No
<i>Vigna mungo</i>	Grain legume	Yes	No
<i>Vigna radiata</i>	Grain legume	Yes	No
<i>Vigna subterranea</i>	Grain legume	Yes	No
<i>Vigna unguiculata</i>	Grain legume	Yes	No
<i>Zea mays</i>	Cereal	Yes	Yes

**Fig. S2.** Estimated mean ratios of the yields of cereals cultivated after non-cereals (either grain legumes or oilseeds) to the yields of cereals cultivated after cereals. The scientific names of the grain legume species preceding cereals are abbreviated as follows: *Arachis hypogaea* (AH), *Cajanus cajan* (CC), *Cicer arietinum* (CA), *Glycine max* (GM), *Lens culinaris* (LC), *Lupinus albus* (LAI), *Lupinus angustifolius* (LAn), *Lupinus luteus* (LL), *Phaseolus vulgaris* (PV), *Pisum sativum* (PS), *Vicia faba* (VF), *Vicia sativa* (VSa), *Vigna mungo* (VM), *Vigna radiata* (VR), *Vigna subterranea* (VSu), and *Vigna unguiculata* (VU). The category for oilseeds (OI) includes six species cultivated as preceding crops, namely, *Brassica juncea*, *Brassica napus*, *Carthamus tinctorius*, *Crambe abyssinica*, *Helianthus annuus*, and *Linum usitatissimum*. Bullet points indicate the mean yield ratios estimated from the unweighted data using the statistical model variant *M2* (note: estimation from the weighted data using the statistical model variant *M4* is presented in Fig. 1). Error bars indicate the 95% confidence intervals. The lower and upper rows in the bottom box (grey) indicate the numbers of selected articles and observed cereal yield ratios available for each preceding non-cereal species, respectively.



**Table S2.** Regions and countries included in the dataset.

Region	Country
Africa	Cameroon
	Ghana
	Kenya
	South Africa
	Zimbabwe
Asia	India
	Pakistan
	Thailand
Europe	Denmark
	Germany
	United Kingdom
North America	Canada
	United States of America
Oceania	Australia
South America	Chile

**Table S3.** Names and definitions of the variables included in the dataset. 'N' is the abbreviation of nitrogen. 'NA' indicates data not available in the selected articles.

Name	Definition
Article_Index	Index of each article.
Article_Author_First	Name of the first author.
Article_Title	Article title.
Article_Year_Publication	Publication year or 'NA'.
Article_Journal	Journal name or 'NA'.
Site_Index	Index of each field site from each article.
Site_Name	Site name.
Site_Country	Site country.
Crop_Sequence_Index	Index of each crop sequence from each field site.
Crop_Sequence_Name	Crop sequence name.
Crop_Index	Index of each crop from each field site.
Crop_Site_Growing_Season_Index	Index of each crop grown at the same field site during the same growing season(s).
Crop_Species_Scientific_Name	Species scientific name.
Crop_Replicate_Number	Number of replicates or 'NA'.
Crop_Yield_Grain	Grain yield or 'NA'.
Crop_Yield_Grain_Unit	Unit of grain yield or 'NA'.
Crop_Yield_Grain_DM_Percentage	Dry matter percentage of grain yield or 'NA'.
Fertilization_N_Rate	N fertilization rate or 'NA'.
Fertilization_N_Rate_Unit	Unit of N fertilization rate or 'NA'.
Fertilization_N_Rate_Unit_Type	Type of unit of N fertilization rate or 'NA'.
Crop_Species_Scientific_Name_Preceding	Species scientific name of the preceding crop or 'NA'.
Crop_Species_Fertilization_N_Preceding	Differences in N fertilization rates applied on preceding non-cereals (NC) and preceding cereals (C). There are four levels: (i) 'FertilizationNCSameC' indicates that the same N fertilization rate is applied to both preceding non-cereals (either grain legumes or oilseeds) and preceding cereals, (ii) 'FertilizationNCSupC' indicates that the N fertilization rate applied to preceding non-cereal crops was higher than that applied to preceding cereal crops, (iii) 'FertilizationNChnC' indicates that the N fertilization rate applied to preceding non-cereal crops was lower than that applied to preceding cereal crops, and (iv) 'NA' indicates that the N fertilization rate was not available for either preceding non-cereal crops and/or preceding cereal crops.

**Table S4.** Statistical model variants tested for the meta-analysis. 'N' is the abbreviation of nitrogen.

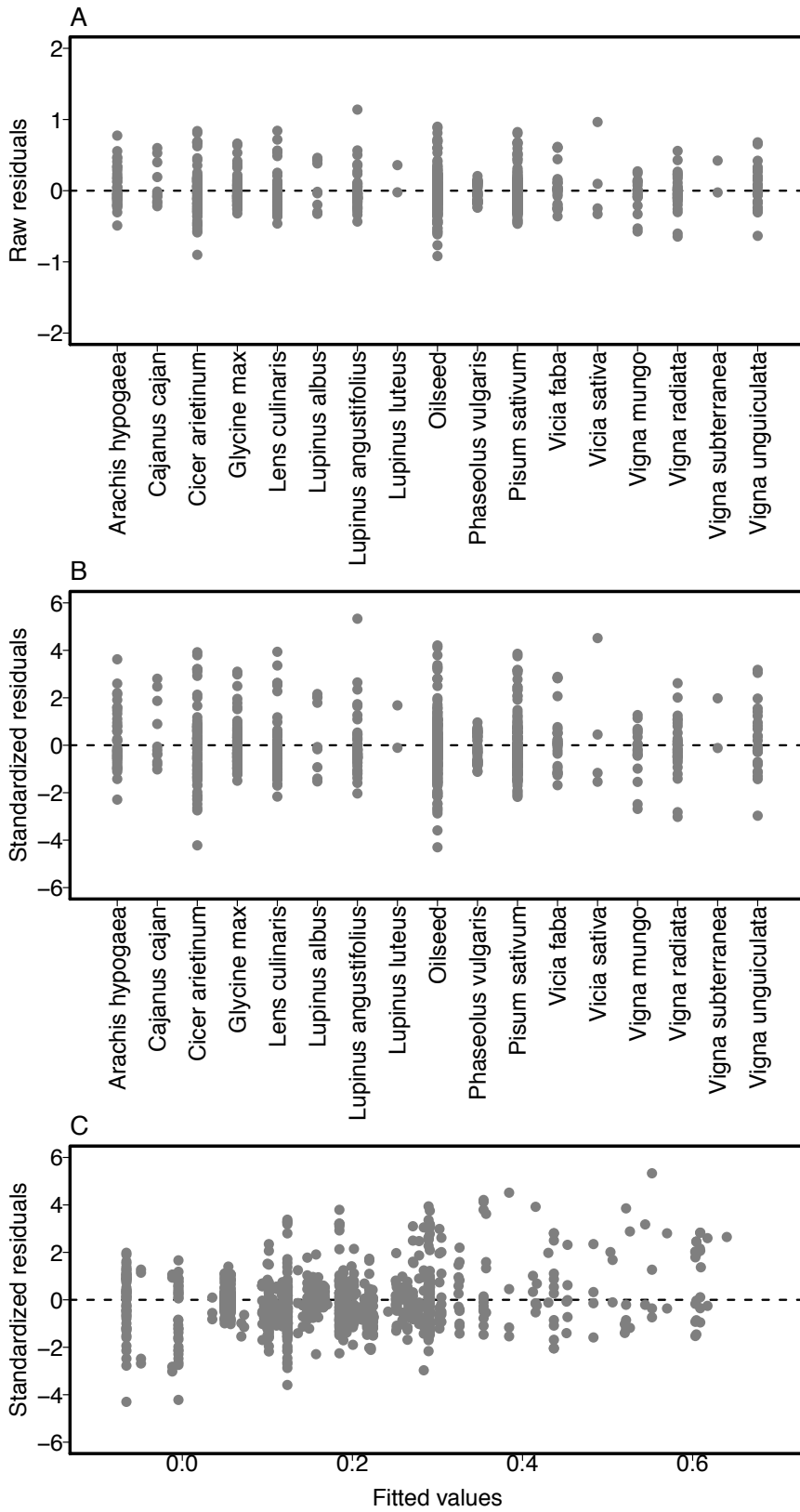
Model	Fixed effect	Random effect	Weight
M1	Preceding non-cereal	Field site×growing season	None
M2	Preceding non-cereal	Article	None
M3	Preceding non-cereal	Field site×growing season	Precision of yield ratios
M4	Preceding non-cereal	Article	Precision of yield ratios
M5	Preceding non-cereal	Field site×growing season	None
M6	Preceding non-cereal	Article	None
M7	Preceding non-cereal	Field site×growing season	None
M8	Preceding non-cereal	Article	None
M9	Preceding non-cereal	Field site×growing season	Fitted values
M10	Preceding non-cereal	Field site×growing season	N fertilization rate for subsequent cereal
M11	Preceding non-cereal	Field site×growing season	Fitted values
M12	Preceding non-cereal	Field site×growing season	N fertilization rate for subsequent cereal
M13	Preceding non-cereal	Article	Fitted values
M14	Preceding non-cereal	Article	N fertilization rate for subsequent cereal
M15	Preceding non-cereal	Article	Fitted values
M16	Preceding non-cereal	Article	N fertilization rate for subsequent cereal



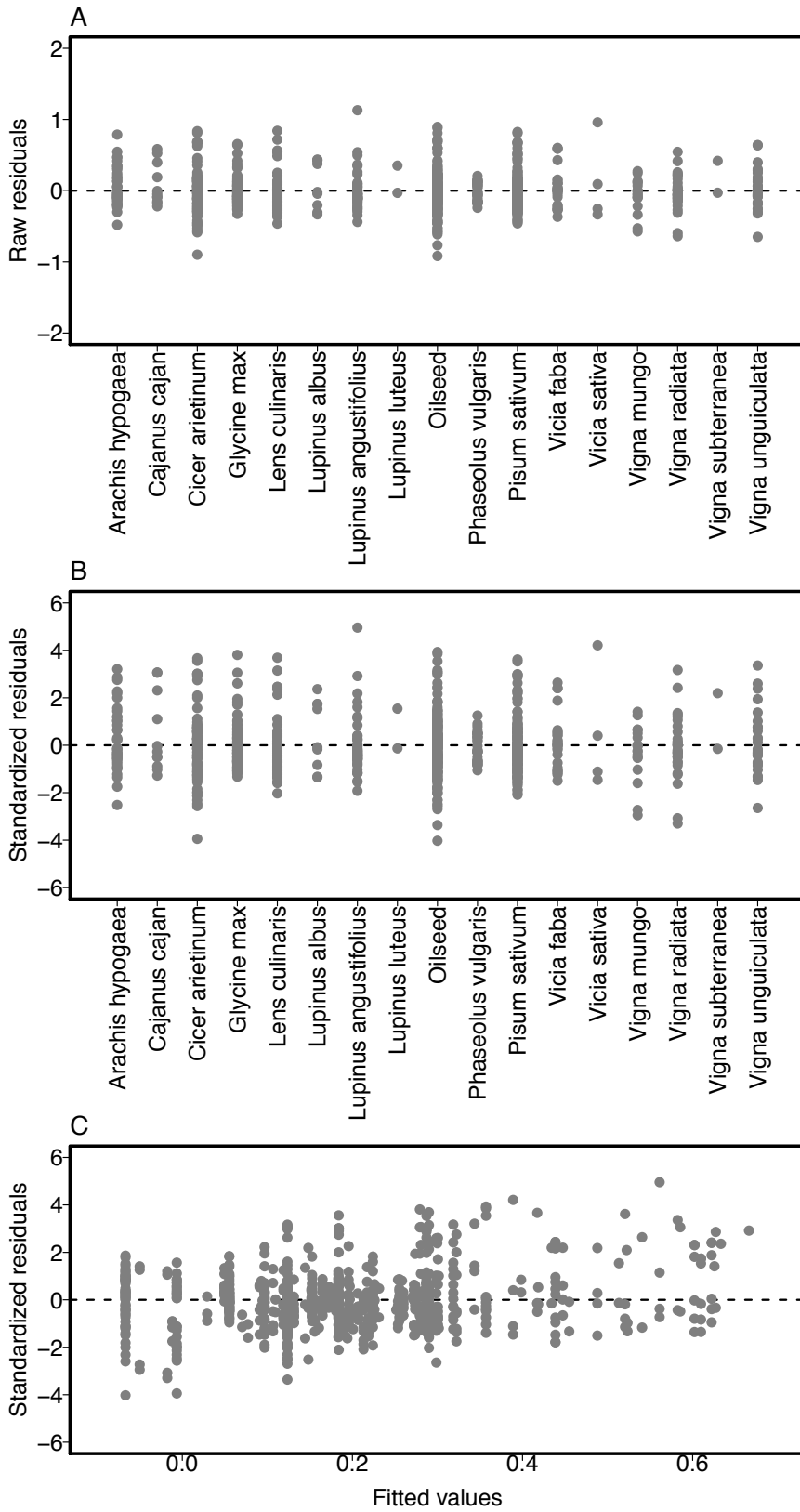
M17	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season	None
M18	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season	None
M19	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season N fertilization rate for subsequent cereal	None
M20	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season N fertilization rate for subsequent cereal	None
M21	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season	Fitted values
M22	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season	N fertilization rate for subsequent cereal
M23	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season N fertilization rate for subsequent cereal	Fitted values
M24	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season N fertilization rate for subsequent cereal	N fertilization rate for subsequent cereal
M25	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season	Fitted values
M26	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season	N fertilization rate for subsequent cereal
M27	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season N fertilization rate for subsequent cereal	Fitted values
M28	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season N fertilization rate for subsequent cereal	N fertilization rate for subsequent cereal
M29	Preceding non-cereal N fertilization rate for subsequent cereal	Field site×growing season	Precision of yield ratios

M30	Preceding non-cereal N fertilization rate for subsequent cereal	Article Field site×growing season	Precision of yield ratios
M31	Preceding non-cereal N fertilization rate for subsequent cereal	Field site×growing season N fertilization rate for subsequent cereal	Precision of yield ratios
M32	Preceding non-cereal N fertilization rate for subsequent cereal	Article Field site×growing season N fertilization rate for subsequent cereal	Precision of yield ratios
M33	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season	Precision of yield ratios
M34	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season	Precision of yield ratios
M35	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Field site×growing season N fertilization rate for subsequent cereal	Precision of yield ratios
M36	Preceding non-cereal N fertilization rate for subsequent cereal N fertilization rate for preceding non-cereal and cereal	Article Field site×growing season N fertilization rate for subsequent cereal	Precision of yield ratios

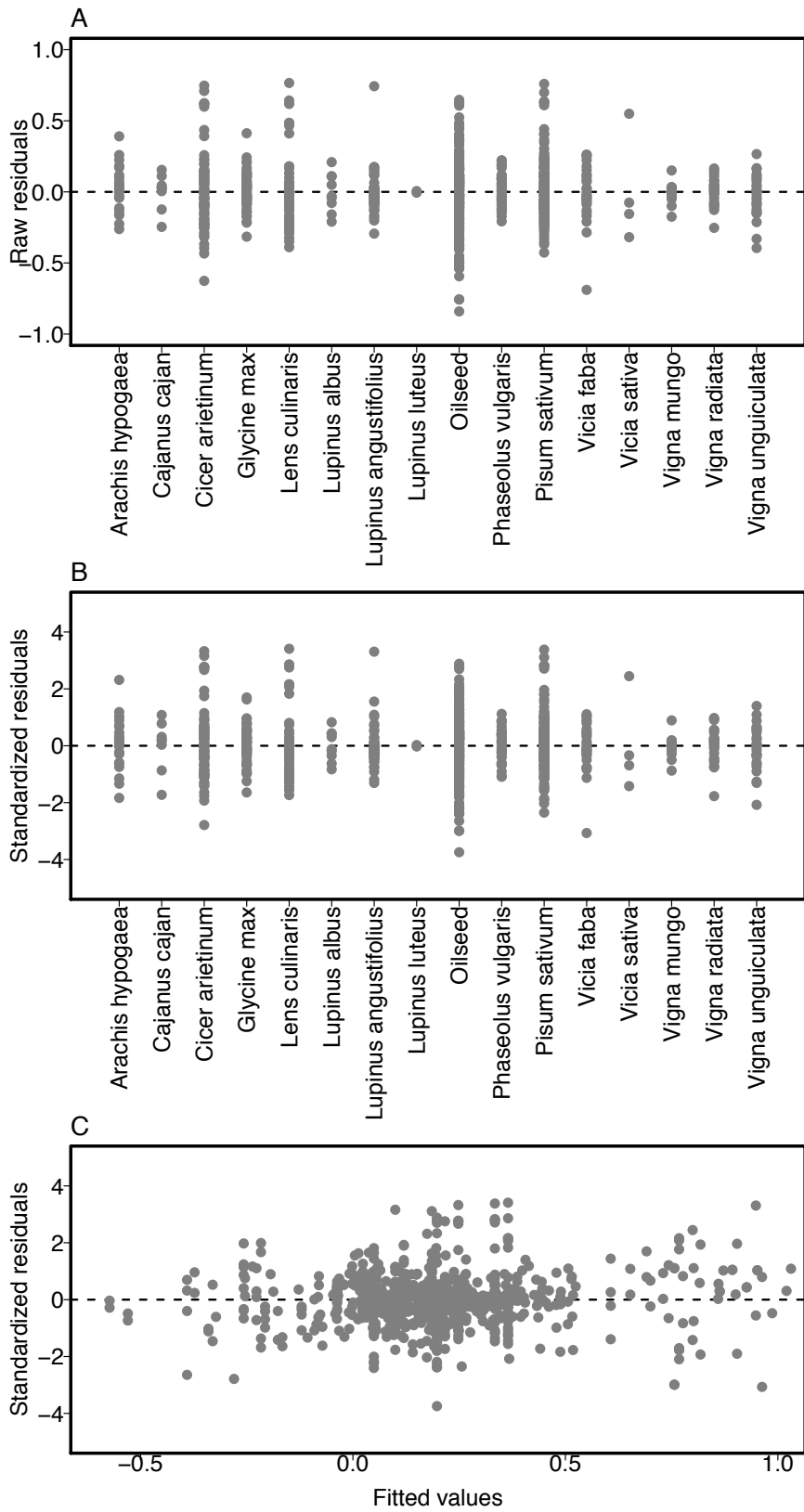
**Fig. S3.** Distribution of residuals using the statistical model variant *M2*. Distribution of raw residuals by non-cereal species preceding cereals (*A*), standardized residuals by non-cereal species preceding cereals (*B*), and standardized residuals versus fitted values across all field site×growing season combinations and selected articles (*C*).



**Fig. S4.** Distribution of residuals using the statistical model variant *M4*. Distribution of raw residuals by non-cereal species preceding cereals (*A*), standardized residuals by non-cereal species preceding cereals (*B*), and standardized residuals versus fitted values across all field site×growing season combinations and selected articles (*C*).



**Fig. S5.** Distribution of residuals using the statistical model variant *M31*. Distribution of raw residuals by non-cereal species preceding cereals (*A*), standardized residuals by non-cereal species preceding cereals (*B*), and standardized residuals versus fitted values across all field site×growing season combinations and selected articles (*C*).









## Identifier des légumineuses à graines productives en Europe par synthèses quantitatives de données à large échelle

**Mots clés** légumineuse à graines, production, synthèse quantitative de données, Europe

**Résumé** Plusieurs études ont souligné la nécessité d'augmenter la production des légumineuses à graines en Europe. Jusqu'à présent, il n'existait pas de synthèses quantitatives de données qui comparaient les performances productives (et environnementales) de différentes légumineuses à graines dans cette région. L'objectif de la thèse était d'identifier des espèces de légumineuses à graines caractérisées par des niveaux élevés de production en Europe. Trois sources de données ont été utilisées à large échelle : des données statistiques, des données expérimentales en Europe et dans d'autres régions du monde, et des données sur les propriétés nutritionnelles des légumineuses à graines. Au total, 29 espèces ont été comparées à partir de leurs niveaux de production, et de leurs effets sur les rendements des céréales suivantes. Nous avons estimé la variabilité interannuelle des rendements des légumineuses à graines en Europe et Amérique. Les résultats montrent que les rendements des légumineuses à graines sont significativement plus variables que les rendements des non-légumineuses en Europe. Ces différences sont plus faibles en Amérique. Nous avons construit un jeu de données expérimentales global incluant 173 articles publiés, 41 pays, et 8581 situations de culture. Une première méta-analyse a été conduite à partir de ce jeu de données expérimentales. Les résultats montrent que le soja (*Glycine max*), le lupin à feuilles étroites (*Lupinus angustifolius*), et la fève (*Vicia faba*), présentent, en général, des niveaux de production similaires, et parfois supérieurs, comparés à ceux du pois protéagineux (*Pisum sativum*) en Europe. D'après les résultats de cette méta-analyse, nous avons estimé qu'une substitution de 25% de la surface actuelle de pois protéagineux (*Pisum sativum*) par de la fève (*Vicia faba*), du lupin à feuilles étroites (*Lupinus angustifolius*), et du soja (*Glycine max*), augmenterait la production de protéines de +3%, +4%, et +28%, en Europe, respectivement. Une seconde méta-analyse a été conduite à partir du même jeu de données expérimentales. Les résultats montrent que les rendements des céréales cultivées après des légumineuses à graines sont, en moyenne, +29% significativement plus élevés que les rendements des céréales cultivées après des céréales ; cet effet positif est significatif pour 13 des 16 espèces de légumineuses à graines. L'effet des cultures précédentes de légumineuses à graines décroît en fonction de la dose de fertilisation azotée (N) appliquée sur les céréales suivantes, et devient négligeable quand la dose moyenne de fertilisation azotée est supérieure à 150 kg N ha<sup>-1</sup>. D'après les résultats de cette méta-analyse, nous avons estimé que la diminution relative attendue de production céréalière, résultant d'une augmentation de la proportion d'une légumineuse à graines dans une monoculture d'une céréale, est partiellement compensée par l'effet positif de la légumineuse à graines sur le rendement de la céréale suivante peu fertilisée en azote. Globalement, la thèse identifie la fève (*Vicia faba*) comme une espèce candidate intéressante en Europe, suivie du pois protéagineux (*Pisum sativum*), du soja (*Glycine max*), et des lupins (*Lupinus* spp.). La lentille (*Lens culinaris*), le pois chiche (*Cicer arietinum*), et le haricot commun (*Phaseolus vulgaris*), présentent des niveaux faibles de production. Cependant, ces espèces sont souvent reconnues pour leurs bénéfices nutritionnels en alimentation humaine. En croisant les regards depuis des expérimentations en Amérique du Nord et Océanie, nous suggérons d'évaluer les niveaux de production de plusieurs gesses (*Lathyrus* spp.), lupins (*Lupinus* spp.), et vesces (*Vicia* spp. excepté *Vicia faba*), dans de futures expérimentations agronomiques en Europe.

## Identifying productive grain legumes in Europe by large-scale quantitative data syntheses

**Keywords** grain legume, crop productivity, quantitative data synthesis, Europe

**Abstract** Several studies have stressed the importance of increasing grain legume production in Europe. To date, no quantitative data syntheses have been conducted to compare the productive (and environmental) performances of different grain legumes in this region. The objective of the PhD thesis was to identify grain legume species displaying high productivity levels in Europe. Three data sources were used on a large scale: statistical data, experimental data across Europe and other world regions, and food and feed composition data for grain legumes. In total, 29 species were compared on the basis of their productivity levels, and on their effects on the yields of the subsequent cereals. We estimated the interannual variability in grain legume yields across Europe and the Americas. Results show that grain legume yields are significantly more variable than non-legume yields in Europe. These differences are smaller in the Americas. We built a global experimental dataset including 173 published articles, 41 countries, and 8,581 crop observations. A first meta-analysis was conducted using this experimental dataset. Results show that soybean (*Glycine max*), narrow-leafed lupin (*Lupinus angustifolius*), and faba bean (*Vicia faba*), display, in general, similar productivity levels, and sometimes higher, compared with those of pea (*Pisum sativum*) in Europe. Based on the results of this meta-analysis, we estimated that a replacement of 25% of the area currently under pea (*Pisum sativum*) with faba bean (*Vicia faba*), narrow-leafed lupin (*Lupinus angustifolius*), and soybean (*Glycine max*), would increase protein production by +3%, +4%, and +28%, in Europe, respectively. A second meta-analysis was conducted using the same experimental dataset. Results show that the yields of cereals cultivated after grain legumes are, on average, +29% significantly higher than the yields of cereals cultivated after cereals; this positive effect is significant for 13 of 16 grain legume species. The effect of preceding grain legume cultivation decreases as a function of the nitrogen (N) fertilization rate applied to subsequent cereals, and becomes negligible when the mean nitrogen fertilization rate exceeds 150 kg N ha<sup>-1</sup>. Based on the results of this meta-analysis, we estimated that the expected relative decrease in cereal production, resulting from an increase in the proportion of a grain legume in a cereal monoculture, is partially mitigated by the positive effect of the grain legume on the yield of the subsequent cereal under low nitrogen input conditions. Globally, the PhD thesis identifies faba bean (*Vicia faba*) as an interesting candidate species in Europe, followed by pea (*Pisum sativum*), soybean (*Glycine max*), and lupins (*Lupinus* spp.). Lentil (*Lens culinaris*), chickpea (*Cicer arietinum*), and kidney bean (*Phaseolus vulgaris*), display low productivity levels. However, these species are often promoted for their nutritional benefits for the human diet. Based on comparative insight gained from experiments in North America and Oceania, we suggest assessing the productivity levels of several vetches and lupins (i.e., *Lathyrus*, *Lupinus*, and *Vicia* species excluding *Vicia faba*), in future field experiments in Europe.