Méthodes de méta-analyse pour l'estimation des émissions de N2O par les sols agricoles

Aurore Philibert

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Statistique appliquée

présentée et soutenue publiquement par

Aurore PHILIBERT

le 16 novembre 2012

Méthodes de méta-analyse

pour l'estimation des émissions de N$_2$O par les sols agricoles

Directeur de thèse : **David MAKOWSKI**
Co-encadrement de la thèse : **Chantal LOYCE**

Jury

M. Jean-Baptiste DENIS,	Directeur de Recherche, INRA Jouy-en-Josas	Rapporteur
M. Sylvain PELLERIN,	Directeur de Recherche, INRA Bordeaux	Rapporteur
M. Benoît GABRIELLE,	Professeur, AgroParisTech	Examinateur
M. Philippe NAVEAU,	Chargé de Recherche, LSCE	Examinateur
M. David MAKOWSKI,	Directeur de Recherche, INRA Grignon	Examinateur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries
du Vivant et de l'Environnement

(AgroParisTech)

Spécialité : Statistique appliquée

présentée et soutenue publiquement par

Aurore PHILIBERT

le 16 novembre 2012

Méthodes de méta-analyse

pour l'estimation des émissions de N$_2$O par les sols agricoles

Directeur de thèse : **David MAKOWSKI**
Co-encadrement de la thèse : **Chantal LOYCE**

Jury
M. Jean-Baptiste DENIS, Directeur de Recherche, INRA Jouy-en-Josas   Rapporteur
M. Sylvain PELLERIN,   Directeur de Recherche, INRA Bordeaux   Rapporteur
M. Benoît GABRIELLE,   Professeur, AgroParisTech   Examinateur
M. Philippe NAVEAU,   Chargé de Recherche, LSCE   Examinateur
M. David MAKOWSKI,   Directeur de Recherche, INRA Grignon   Examinateur

« On ne fait jamais attention à ce qui a été fait ; on ne voit que ce qui reste à faire. »

Marie Curie

Remerciements

Je tiens tout d'abord à remercier David Makowski et Chantal Loyce de m'avoir fait confiance pour mener à bien cette thèse, m'avoir encouragée, guidée et conseillée pendant ces trois années. Merci à toi, David, pour la qualité de ton encadrement et de tes conseils. J'ai énormément appris à tes côtés que ce soit en statistique mais aussi en termes d'encadrement ou de gestion de tes activités professionnelles. Ton calme olympien a permis de ne pas amplifier mon stress et j'espère pouvoir m'inspirer de ton sens de l'organisation pour mes prochaines activités professionnelles. Merci Chantal de m'avoir transmis une partie des tes connaissances agronomiques et de m'avoir quotidiennement rappelé que je faisais des statistiques appliquées, qui devaient donc être accessibles aux agronomes. Cela m'a forcé à expliciter tous les éléments de ma démarche et m'a été très profitable.

Mes remerciements vont également à Jean-Baptiste Denis et Sylvain Pellerin pour avoir accepté d'évaluer mon travail ainsi que Benoît Gabrielle et Philippe Naveau pour avoir accepté de l'examiner. Merci à Benoît Gabrielle d'avoir été mon professeur référent pendant ces trois années et d'avoir accepté d'être président du jury de ma soutenance.

Les membres de mon comité de pilotage m'ont permis d'enrichir et de mettre en perspective mon travail grâce à de fructueuses discussions. Je remercie donc Pierre Cellier, Eric Gozé, Marie-Hélène Jeuffroy et Jacques Wéry.

J'ai eu la chance de rencontrer Fernando Miguez (Professeur assistant, Iowa State University) lors d'un séjour aux États-unis. Je le remercie pour son accueil, son écoute et ses judicieux conseils.

Je tiens également à remercier Isabelle Albert, qui a pris de son temps pour m'aider à utiliser le logiciel WinBUGS.

Je remercie aussi Philippe Rochette (Agri-Food Canada) d'avoir accepté de me rencontrer et d'avoir été présent pour répondre à mes questions.

Antoine Cornuéjols et Christine Martin m'ont permis de réaliser ma première expérience d'enseignement en encadrant des étudiants de première année d'AgroParisTech pour un projet de fouille de données, je leur en suis reconnaissante.

Je tiens à remercier aussi Françoise Launay et Corinne Fiers de l'école doctorale ABIES ainsi que Donald White et Eric Birlouez pour leurs formations toujours très enrichissantes, que ce soit pour l'écriture d'articles dans la langue de Shakespeare ou bien pour appréhender la gestion de projet. Ces formations ont aussi été très utiles pour les échanges qu'elles permettaient entre doctorants.

Un grand merci à Liliane Bel qui m'a permis de présenter mes travaux aux Journées de Statistique.

Merci à la société Alex Edelman qui a relu tous les chapitres de cette thèse rédigés en anglais, ce qui a permis d'améliorer la qualité de ce manuscrit.

Je remercie bien évidemment tous les membres de l'unité d'Agronomie, permanents et temporaires qui se sont succédés, pour leur accueil, leur sympathie et la bonne ambiance qui règne à Grignon. Les barbecues de l'été et les pots vont me manquer.

Je tiens à remercier particulièrement Marie-Hélène Jeuffroy, directrice de l'unité pour m'avoir suivie pendant ses trois années.

Merci à Caroline Colnenne-David pour m'avoir permis de faire une journée de terrain sur les mesures d'émissions de N_2O, me permettant ainsi de sortir la tête des chiffres et de mieux comprendre et appréhender la réalité des mesures.

Je remercie aussi Elise Pelzer d'avoir relu ma thèse et de m'avoir donné de précieux conseils pour ma soutenance.

Mes remerciements vont à toutes les personnes qui ont assisté à mes deux répétitions de soutenance pour les conseils et remarques qu'ils ont pu me faire, ce qui a grandement amélioré la qualité de ma présentation. Un merci particulier à Diana, Marie, Michel et Safia pour avoir assisté à la première répétition dans des circonstances particulières.

Merci à Michel pour les grandes discussions que nous avons eu, ainsi que pour tes arrivées inopinées dans notre bureau, c'était toujours très intéressant et j'ai beaucoup appris de choses avec toi !

Merci à Jean-Marc pour le soutien informatique indispensable à la réalisation de cette thèse dont il a su faire preuve.

Je ne t'oublie pas Alain pour les très nombreux articles que tu as pu récupérer avec rapidité mais aussi pour les encouragements à aller « bader ».

Je tiens à remercier Michèle et Joëlle d'avoir su m'épauler dans les nombreuses démarches administratives. Michèle tu seras toujours pour moi la reine de l'administration, moi qui n'y comprends pas grand-chose.

Je ne peux pas repenser à ces trois années sans remercier la team des *Girls in Agronomy* : Aïcha, Caroline, Céline, Claire, Diana, Laure et Tatiana pour tous les très bons moments partagés, les déjeuners à la cafet, les sorties shopping, les séances de bad et autres moments privilégiés qui m'ont permis de passer ces trois années dans les meilleures conditions. Sans

vous je ne sais pas si j'aurais supporté ces trajets interminables jusqu'au petit coin de verdure qu'est Grignon.

Un remerciement spécial pour Claire la meilleure collègue/ amie de bureau que j'ai connu. Nos discussions tant sur la thèse que sur nos « petits » problèmes personnels ont toujours été très fructueuses. Nous nous sommes soutenues pendant les moments difficiles et avons profité ensemble des bons moments. Que dire à part que j'aimerais bien encore maintenant n'avoir qu'à tourner la tête pour te poser mes questions existentielles ☺

Je n'oublie pas les deux petits Anges arrivés en cours de route, Kintana et Noémie, qui sont devenues nos petites lumières à nous.

Une pensée pour *PhD* la mascotte officielle qui a n'a pas tenu le coup jusqu'au bout mais a rendu les journées à Grignon pleines de vie.

Je tiens à remercier mes amis Alban, Charlotte, Cindy, Guillaume, Nathalie, Raphaëlle et Stéphanie pour avoir été présent dans les bons mais aussi les mauvais moments.

Je remercie enfin toute ma famille pour leur relecture et les nombreux encouragements prodigués pendant ces trois années. Ils ont été d'un grand soutien.

Une pensée à ma belle-famille pour le soutien et l'intérêt qu'ils ont porté à l'avancement de ma thèse.

Merci au Caracole de cette expérience pour son soutien sans faille, ses précieux conseils et bien plus encore…

Table des Matières

Publications scientifiques et valorisation

Publications issues des travaux de thèse

Philibert A, Loyce C, Makowski D (2012a) Assessment of the quality of meta-analysis in agronomy. *Agriculture, Ecosystems & Environment* 148: 72– 82.

Philibert A, Loyce C, Makowski D (2012b) Quantifying uncertainties in N_2O emission due to N fertilizer application in cultivated areas. *PLoS ONE* 7 (11).

Philibert A, Loyce C, Makowski D (2013) Prediction of N_2O emissions from local information with Random Forest. *Environmental Pollution* 177: 156-163.

Publication hors thèse

Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, Makowski D (2011) Predicting invasion success of forest pathogenic fungi from species traits. *Journal of applied ecology* 48: 1381-1390.

Communications à des manifestations nationales ou internationales :

Communications orales

Philibert A, Loyce C, Makowski D (2010) Hierarchical models for estimating risk of nitrous oxide emission in agricultural systems. Society of Risk Analysis 2010 Annual Meeting, Salt Lake City.

Philibert A, Loyce C, Makowski D (2012) Etude de la sensibilité des émissions de N_2O à la méthode statistique utilisée. 44è journée de Statistique SFdS, Bruxelles.

Posters

Philibert A, Loyce C, Makowski D (2010) Predicting invasion success from species traits using machine learning techniques. Society of Risk Analysis 2010 Annual Meeting, Salt Lake City.

Philibert A, Loyce C, Makowski D (2011) Estimating risk of nitrous oxide (N_2O) emissions in legume crops. Journée de l'école doctorale ABIES- Agriculture, Alimentation, Biologie, Environnements et Santé, Paris.

Philibert A, Loyce C, Makowski D (2012) Sensitivity analysis of estimated global N_2O emission in cultivated areas. Planet Under Pressure, Londres.

Liste des Figures

Chapitre 4

Encadrés

Liste des Tables

<u>Chapitre 3</u>

<u>Chapitre 4</u>

« Il faut toujours viser la Lune car même en cas d'échec on atterrit dans les étoiles. »

Oscar Wilde

Introduction générale

1.   La méta-analyse : historique d'une méthode

Le terme de méta-analyse a été proposé en 1976 par le statisticien américain Gene V Glass pour désigner l'analyse statistique d'un large ensemble de résultats provenant d'études individuelles (Glass, 1976). Pour un sujet donné, la méta-analyse permet d'obtenir une conclusion générale ainsi qu'une évaluation de la cohérence entre les différents résultats (Hedges et Olkin, 1985). Travaillant dans le domaine de la psychologie de l'éducation et des sciences sociales, GV Glass a donné un nom à une méthode déjà étudiée depuis le XVIII$^{\text{ème}}$ siècle. C'est en effet à cette période que la distinction entre résultats provenant d'une étude individuelle et résultats provenant de plusieurs études est apparue (O'Rourke, 2007). Ce sont les mathématiciens et astronomes (*e.g.* Stigler, 1986) qui ont abordé les premiers l'analyse de combinaison de résultats, mais c'est Karl Pearson qui, en 1904, présente la première méta-analyse, basée sur cinq essais cliniques testant l'effet d'un vaccin contre la fièvre typhoïde et ses conséquences sur l'infection et la mortalité (Pearson, 1904). Au cours du XX$^{\text{ème}}$ siècle, la méta-analyse est utilisée en sciences sociales, psychologie et éducation, et son utilisation se répand plus largement. La méta-analyse propose en effet un cadre quantitatif qui permet de synthétiser les données de manière objective : elle apporte ainsi une plus-value par rapport à des synthèses qualitatives sur un sujet donné, telles que celles publiées dans des articles de revues (Arnqvist et Wooster, 1995 ; Petticrew, 2001 ; Roberts et al., 2006). Elle permet de répondre à des questions dans de très nombreux domaines allant de l'astronomie à la zoologie (Petticrew, 2001).

Dans les années 70, les premières méta-analyses analysant des résultats d'essais cliniques apparaissent (Stjernswärd, 1974), dans la continuité des travaux de Pearson (1904), permettant de juger de l'efficacité d'interventions médicales. Les données utilisées y sont très standardisées, sous forme d'essais cliniques randomisés (*Randomized Control Trials*). En 1992, une organisation internationale, appelée *Cochrane Collaboration*, a été créée dans le but de préparer, mettre à jour et promouvoir la méta-analyse dans le domaine de la santé humaine (http://www.cochrane.org/). Une reconnaissance accrue des méta-analyses dans le domaine de la santé humaine a fait augmenter le nombre de publications dans ce domaine (Sutton et Higgins, 2008). De 1990 à 2006, ce nombre est passé de 300 à 2300 soit une multiplication par 8 des publications de méta-analyses en seulement 16 ans (PubMed ; Sutton et Higgins, 2008). La recherche méthodologique connait aussi une augmentation des publications puisque pour la même période le nombre de papiers s'y rapportant a augmenté d'un facteur 10. Sutton et Higgins (2008) présentent une revue des récents développements méthodologiques en termes de méta-analyse dans la recherche médicale sur de nombreux sujets (par exemple la méthode Bayésienne, l'évaluation de l'hétérogénéité) qui nous montrent la considérable activité de recherche dans ce domaine ces dernières années. Ils présentent notamment pour la synthèse complexe de données les comparaisons de traitements

mixtes (*Mixed Treatment Comparison*, MTC), méthode permettant de comparer des traitements médicaux non plus par rapport à un placebo (approche standard) mais par rapport à un autre traitement. C'est un sujet de recherche méthodologique récent et pas encore abouti mais qui promet de nombreuses applications.

Dans le domaine de l'écologie, les premières méta-analyses sont apparues au début des années 1990 (Arnqvist et Wooster, 1995 ; Gurevitch et al., 1992) le nombre de papiers utilisant ces méthodes a depuis explosé. De 1991 à 2008, le nombre de publications en écologie a augmenté de 9% par an tandis que le nombre de méta-analyses effectuées dans ce domaine a augmenté de 28% par an pour cette même période (Cadotte et al., 2012). Un numéro spécial a été consacré à la méta-analyse dans la revue *Evolutionary Ecology* en 2012 (Cadotte et al., 2012). De nombreux problèmes méthodologiques ont été résolus ces dernières années grâce à une recherche dynamique dans ce domaine (Nakagawa et Santos, 2012). La méta-analyse est reconnue pour être adaptée aux problématiques écologiques (Cadotte et al., 2012), entre autres pour traiter les caractéristiques spatiales et temporelles multi-échelles d'une dynamique de population, de communauté ou d'écosystème (Levin, 1992). Elle permet aussi de fournir une évaluation critique d'hypothèses écologiques reconnues comme la réponse des espèces au changement climatique (Cadotte et al., 2012).

En agronomie, c'est Ronald Fisher qui s'est le premier intéressé à l'application de la méta-analyse. Il propose en 1935 une analyse combinant plusieurs résultats d'expérimentations sur les effets de la fertilisation qui varient selon l'année et le lieu (Fisher, 1935). Une recherche sur la plateforme Web of Knowledge (au 21 septembre 2012) nous révèle que la présence du mot « *meta-analysis* » dans le titre des articles a été multipliée par neuf ces dix dernières années. En sélectionnant les articles du domaine « *agriculture* », 103 publications respectent ce critère en 2012, contre 55 en 2007 et seulement 12 en 2002. La première application de modèle à effets aléatoires pour traiter des données agricoles issues de plusieurs expérimentations date de 1938 (Yates et Cochran, 1938). Les modèles à effets aléatoires sont des modèles couramment utilisés en méta-analyse (Borenstein et al., 2009). Avec des modèles à effets fixes, on fait l'hypothèse que la « vraie » réponse est la même pour toutes les mesures individuelles. Les différences entre mesures correspondent alors à une erreur aléatoire inhérente à chaque mesure. Les modèles à effets aléatoires font l'hypothèse que chaque mesure individuelle a une « vraie » réponse différente de celles des autres résultats mais toutes issues d'une distribution commune (Borenstein, 2009). La deuxième hypothèse est plus réaliste face aux données disponibles en écologie (Nakagawa et Santos, 2012) mais aussi en agronomie.

$$\text{EQUATION 11.1}$$
$$\text{DIRECT N}_2\text{O EMISSIONS FROM MANAGED SOILS (TIER 1)}$$

$$N_2O_{Direct}-N = N_2O-N_{N\,inputs} + N_2O-N_{OS} + N_2O-N_{PRP}$$

Where:

$$N_2O-N_{N\,inputs} = \begin{bmatrix} \left[\left(F_{SN} + F_{ON} + F_{CR} + F_{SOM} \right) \bullet EF_1 \right] + \\ \left[\left(F_{SN} + F_{ON} + F_{CR} + F_{SOM} \right)_{FR} \bullet EF_{1FR} \right] \end{bmatrix}$$

$$N_2O-N_{OS} = \begin{bmatrix} \left(F_{OS,CG,Temp} \bullet EF_{2CG,Temp} \right) + \left(F_{OS,CG,Trop} \bullet EF_{2CG,Trop} \right) + \\ \left(F_{OS,F,Temp,NR} \bullet EF_{2F,Temp,NR} \right) + \left(F_{OS,F,Temp,NP} \bullet EF_{2F,Temp,NP} \right) + \\ \left(F_{OS,F,Trop} \bullet EF_{2F,Trop} \right) \end{bmatrix}$$

$$N_2O-N_{PRP} = \left[\left(F_{PRP,CPP} \bullet EF_{3PRP,CPP} \right) + \left(F_{PRP,SO} \bullet EF_{3PRP,SO} \right) \right]$$

Where:

$N_2O_{Direct}-N$ = annual direct N_2O–N emissions produced from managed soils, kg N_2O–N yr^{-1}

$N_2O-N_{N\,inputs}$ = annual direct N_2O–N emissions from N inputs to managed soils, kg N_2O–N yr^{-1}

N_2O-N_{OS} = annual direct N_2O–N emissions from managed organic soils, kg N_2O–N yr^{-1}

N_2O-N_{PRP} = annual direct N_2O–N emissions from urine and dung inputs to grazed soils, kg N_2O–N yr^{-1}

F_{SN} = annual amount of synthetic fertiliser N applied to soils, kg N yr^{-1}

F_{ON} = annual amount of animal manure, compost, sewage sludge and other organic N additions applied to soils (Note: If including sewage sludge, cross-check with Waste Sector to ensure there is no double counting of N_2O emissions from the N in sewage sludge), kg N yr^{-1}

F_{CR} = annual amount of N in crop residues (above-ground and below-ground), including N-fixing crops, and from forage/pasture renewal, returned to soils, kg N yr^{-1}

F_{SOM} = annual amount of N in mineral soils that is mineralised, in association with loss of soil C from soil organic matter as a result of changes to land use or management, kg N yr^{-1}

F_{OS} = annual area of managed/drained organic soils, ha (Note: the subscripts CG, F, Temp, Trop, NR and NP refer to Cropland and Grassland, Forest Land, Temperate, Tropical, Nutrient Rich, and Nutrient Poor, respectively)

F_{PRP} = annual amount of urine and dung N deposited by grazing animals on pasture, range and paddock, kg N yr^{-1} (Note: the subscripts CPP and SO refer to Cattle, Poultry and Pigs, and Sheep and Other animals, respectively)

EF_1 = emission factor for N_2O emissions from N inputs, kg N_2O–N (kg N input)$^{-1}$ (Table 11.1)

EF_{1FR} is the emission factor for N_2O emissions from N inputs to flooded rice, kg N_2O–N (kg N input)$^{-1}$ (Table 11.1) [5]

EF_2 = emission factor for N_2O emissions from drained/managed organic soils, kg N_2O–N ha^{-1} yr^{-1}; (Table 11.1) (Note: the subscripts CG, F, Temp, Trop, NR and NP refer to Cropland and Grassland, Forest Land, Temperate, Tropical, Nutrient Rich, and Nutrient Poor, respectively)

EF_{3PRP} = emission factor for N_2O emissions from urine and dung N deposited on pasture, range and paddock by grazing animals, kg N_2O–N (kg N input)$^{-1}$; (Table 11.1) (Note: the subscripts CPP and SO refer to Cattle, Poultry and Pigs, and Sheep and Other animals, respectively)

Encadré 1 : Equation pour les émissions de N_2O du niveau 1 du GIEC (Source : IPCC, 2006).

17

La méta-analyse est donc une approche très utilisée et étudiée de nos jours. De nombreux livres ont été publiés récemment sur le sujet (Borenstein et al., 2009, Spiegelhalter et al., 2010) et des formations sur la méta-analyse se développent (Université de Géorgie : http://www.coe.uga.edu/events/2011/10/03/may-22-23-2012-workshop-of-meta-analysis/; Université Catholique du Louvain : http://www.uclouvain.be/419731.html).

2.   La méta-analyse : un outil utilisé pour estimer les émissions de N_2O

Le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC / IPCC), a été créé en 1988. Il a pour mission de fournir un état des lieux des connaissances sur le changement climatique et d'en évaluer ses impacts potentiels au plan environnemental et socio-économique. Il effectue un travail de synthèse des informations scientifiques et techniques sur le changement climatique au niveau mondial (http://www.ipcc.ch).

Le N_2O est un puissant gaz à effet de serre avec un pouvoir de réchauffement 298 fois plus puissant que le CO_2 sur une période de 100 ans et une durée de vie dans l'atmosphère de 114 ans (IPCC, 2007). En plus d'être un puissant gaz à effet de serre, le N_2O est aussi une substance jouant un rôle majeur dans l'appauvrissement de la couche d'ozone (Ravishankara et al., 2009). C'est au cours des deux processus de nitrification (transformation de l'ammoniac en nitrite puis de nitrite en nitrate) et de dénitrification (réduction du nitrate) que le N_2O est émis (Hénault et al., 2012). Les émissions de N_2O dues à l'agriculture représentent approximativement 60% des émissions de N_2O anthropiques (IPCC, 2007). Les facteurs modifiant les processus de nitrification et de dénitrification influent sur les émissions de N_2O. En agriculture, les facteurs environnementaux liés au sol (pH, texture), aux conditions climatiques (pluie, température) ainsi qu'aux pratiques culturales (labour, fertilisation) forment un ensemble de facteurs influant sur les émissions de N_2O. La principale source reconnue est l'apport d'engrais azoté aux sols cultivés (Mosier et al., 1998). Toutefois, l'influence des autres facteurs sur les émissions de N_2O n'est pas à négliger et est largement étudiée (Hénault et al., 2012 ; Leip et al., 2011 ; Stehfest et Bouwman, 2006). Les émissions de N_2O sont caractérisées par une forte variabilité spatiale et temporelle, les facteurs évoqués plus haut variant considérablement entre sites d'expérimentations (Gu et al., 2012).

L'équation générale d'estimation des émissions de N_2O par le GIEC est $N_2O = EF*DA$ où DA correspondent à une mesure d'activité humaine reconnue comme ayant un impact sur les émissions de N_2O (par exemple la dose d'engrais azoté) et où EF est un facteur d'émission représentant le pourcentage de données d'activités transformées en émission de N_2O.

Le GIEC considère trois approches, appelées niveaux (Tier), pour estimer les émissions de N_2O selon la disponibilité des données d'EF et de DA par pays.

Tier 2

If more detailed emission factors and corresponding activity data are available to a country than are presented in Equation 11.1, further disaggregation of the terms in the equation can be undertaken. For example, if emission factors and activity data are available for the application of synthetic fertilisers and organic N (F_{SN} and F_{ON}) under different conditions *i*, Equation 11.1 would be expanded to become [6]:

EQUATION 11.2
DIRECT N$_2$O EMISSIONS FROM MANAGED SOILS (TIER 2)

$$N_2O_{Direct}-N = \sum_i \left(F_{SN} + F_{ON}\right)_i \bullet EF_{1i} + \left(F_{CR} + F_{SOM}\right) \bullet EF_1 + N_2O-N_{OS} + N_2O-N_{PRP}$$

Where:

EF_{1i} = emission factors developed for N$_2$O emissions from synthetic fertiliser and organic N application under conditions *i* (kg N$_2$O–N (kg N input)$^{-1}$); *i* = 1, …n.

Equation 11.2 may be modified in a variety of ways to accommodate any combination of N source-, crop type-, management-, land use-, climate-, soil- or other condition-specific emission factors that a country may be able to obtain for each of the individual N input variables (F_{SN}, F_{ON}, F_{CR}, F_{SOM}, F_{OS}, F_{PRP}).

Encadré 2 : Equation pour les émissions de N$_2$O du niveau 2 du GIEC (Source : IPCC, 2006).

Si le pays ne détient aucune mesure de facteur d'émission spécifique à chaque donnée d'activité (au minimum celle d'apport d'engrais azoté) et détient ou non des mesures de données d'activités, alors ses émissions de N_2O sont estimées grâce à des valeurs de paramètres par défaut disponible au niveau 1 (voir Encadré 1). Les pays utilisant ce niveau pour calculer leurs émissions de N_2O provenant des sols agricoles représentent 56% des pays développés de la convention-cadre des Nations Unies sur les changements climatiques (CCNUCC / UNFCCC) (Lokupitiya et Paustian, 2006) et la moitié des publications sur les inventaires de N_2O (Berdanier et Conant, 2012).

Un pays peut passer au niveau 2 (ou 3) quand il a à sa disposition des valeurs de facteurs d'émissions spécifiques (Voir Encadré 2 pour l'équation du niveau 2). C'est le cas par exemple du Canada, qui est passé au niveau 2 (Rochette et al., 2008). L'estimation des émissions de N_2O requiert la même équation qu'au niveau 1 mais les valeurs des paramètres sont celles calculées pour le pays ; la valeur du facteur d'émission peut alors être estimée pour différents types d'environnement.

Le niveau 3 est une approche par modélisation et mesures expérimentales. Le niveau 3 se base généralement sur des modèles de culture mécanistes qui ont besoin de données d'activités géographiquement détaillées. De telles méthodes peuvent fournir des estimations plus robustes et plus précises mais deux contraintes majeures en font le niveau le moins utilisé. Premièrement, ces méthodes nécessitent un effort important de documentation et de transparence pour assurer la comparabilité entre les pays et deuxièmement les données doivent être spatialement explicites (Lokupitiya et Paustian, 2006). Il semble que le seul pays a avoir proposé une estimation d'émission de N_2O à ce niveau est l'Italie avec le modèle DNDC (Lugato et al., 2010).

Le GIEC réalise des études spécifiques pour estimer les émissions de N_2O à partir de données publiées. Ces estimations sont effectuées à partir de bases de données incluant des mesures d'émission issues d'expérimentations réalisées dans différentes régions du monde. Le GIEC utilise la méta-analyse pour estimer les émissions de N_2O sans citer explicitement ce terme. Trois exemples sont présentés ci-dessous.

Le premier exemple concerne les émissions de N_2O provenant des légumineuses. Les légumineuses sont une famille de plantes qui ont la propriété de fixer l'azote présent dans l'atmosphère et par conséquent n'ont pas besoin d'apport additionnel d'azote par le biais de la fertilisation. La fixation de l'azote atmosphérique se fait grâce à des bactéries de l'espèce *Rhizobium*, au niveau de nodules présents sur les racines de la plante. Le GIEC considérait jusqu'en 1999 qu'une fraction des émissions de N_2O était produite durant ce processus de fixation symbiotique (IPCC, 1996). Rochette et Janzen (2005) ont compilé 79 mesures de N_2O provenant de 33 études portant sur plusieurs espèces de légumineuses en Amérique du Nord, Europe et Nouvelle-Zélande. Grâce à un traitement quantitatif de ces mesures, les

auteurs ont conclu qu'il n'y avait pas d'émissions de N_2O dues au processus de fixation symbiotique de l'azote atmosphérique pendant la phase de croissance des légumineuses. Ces émissions de N_2O liées à la fixation symbiotique ont ensuite été supprimées dans le calcul du GIEC (IPCC, 2006).

Les deuxième et troisième exemples concernent les émissions de N_2O dues uniquement à la fertilisation azotée. Le facteur d'émission spécifique à la dose d'engrais azoté minéral et organique apportée issu de l'équation du niveau 1 du GIEC (F_{SN} et F_{ON}, Encadré 1) a tout d'abord été fixé à la valeur de la pente d'une régression linéaire sur 20 mesures de N_2O en fonction des doses d'engrais azoté appliquées (Bouwman, 1996), égale à 1,25%. Cette valeur a servi pour tous les calculs d'émission de N_2O jusqu'en 2006, date à laquelle Stehfest et Bouwman (2006) ont publié une nouvelle valeur de facteur d'émission (1%) obtenue à partir d'une analyse statistique d'une base de données beaucoup plus grande comprenant 1008 mesures de N_2O provenant de 204 études publiées (disponible à l'adresse suivante : http://www.pbl.nl/en/publications/2006/N2OAndNOEmissionFromAgriculturalFieldsAndSoilsUnderNaturalVegetation).

3.   Evaluation de la qualité d'une méta-analyse

Du fait de l'utilisation croissante de la méta-analyse en agronomie et dans les sciences de l'environnement, il apparait nécessaire de définir des critères de qualité et des règles de bonnes pratiques. Dans le domaine de la santé humaine, plusieurs guides d'évaluation ont déjà été proposés (Higgins et Green, 2005 ; Moher et al., 2001), de même qu'en écologie (Gates, 2002 ; Pullin et Stewart, 2006). Mais comment évaluer la qualité d'une méta-analyse réalisée en agronomie ? Quelle est la qualité de ces méta-analyses ? Peut-on l'améliorer ? Le premier chapitre de ce mémoire de thèse s'efforce de répondre à ces questions en proposant une liste de critères pour évaluer la qualité d'une méta-analyse. Ces critères sont ensuite appliqués à l'ensemble des méta-analyses réalisées jusqu'au 16 août 2011 en agronomie. Ils sont ensuite illustrés à l'aide d'un exemple concernant les émissions de N_2O des cultures de légumineuses mobilisant les données de Rochette et Janzen (2005).

Un des résultats marquant de ce premier chapitre est que très peu de méta-analyses étudient la robustesse de leurs conclusions aux données et aux méthodes statistiques utilisées. Ce résultat nous a conduits à analyser en détail ce problème dans la suite de la thèse.

4.   Etude de la robustesse des conclusions issues de méta-analyse

L'étude de la robustesse des conclusions issues de méta-analyses se fait par le biais d'analyse de sensibilité. Un éclaircissement est nécessaire concernant la définition donnée à l'analyse de sensibilité dans le cadre de la méta-analyse. En agronomie l'utilisation la plus courante du terme « analyse de sensibilité » se rattache à l'étude des modèles de culture mécanistes. Un modèle de culture mécaniste (*process-based crop model*) est un modèle qui décrit et quantifie les mécanismes et processus du système plante-sol-atmosphère à partir d'hypothèses formulées sur le fonctionnement de ce système. Dans ce contexte, l'analyse de sensibilité consiste à étudier la sensibilité des sorties du modèle en faisant varier les paramètres incertains et/ou variables qui le composent (Monod et al., 2006). Les sorties du modèle peuvent être évaluées dans un objectif explicatif ou bien prédictif. Une autre utilisation du terme d'analyse de sensibilité se réfère à tester la sensibilité d'une inférence statistique aux données manquantes (Kenward et al., 2001), cette dernière n'est cependant pas ici l'objet d'étude.

Dans le cadre de la méta-analyse, une analyse de sensibilité peut être conduite pour étudier la sensibilité d'une variable de réponse (ex : efficacité d'un médicament, rendement d'une culture, émission de N_2O) à deux types d'éléments (Borenstein et al., 2009) :

(i)   la méthode statistique (notamment le type de modèle et la méthode d'inférence) (Egger et al., 1997 ; Oxman, 1994),

(ii)   les données ou variables explicatives utilisées (Gates, 2002 ; Oxman, 1994 ; Roberts et al., 2006).

D'une manière plus générale, l'analyse de sensibilité permet d'étudier la sensibilité des résultats d'une méta-analyse aux différents éléments définissant la démarche utilisée. Si des modifications de ces éléments remettent en cause certaines conclusions d'une méta-analyse, on peut conclure que ses résultats ne sont pas robustes. L'analyse de sensibilité permet d'augmenter la transparence de la démarche et de préciser son domaine de validité.

Différents modèles statistiques peuvent conduire à des résultats très différents puisque derrière le choix d'un modèle se cachent des hypothèses relatives à la structure des données (Nakagawa et Santos, 2012).

Une approche très souvent utilisée en médecine consiste à enlever une partie des données individuelles définies comme de moindre qualité et de vérifier la stabilité des résultats (Egger et al., 1997). Une généralisation de cette approche peut être envisagée en enlevant les observations les unes après les autres tout en regardant l'impact sur les résultats (Tudoreanu et Phillips, 2004, Annexe 1).

5. Objectif et plan de la thèse

L'objectif de cette thèse est de démontrer et d'illustrer l'importance d'étudier la robustesse des conclusions issues de méta-analyses. Je m'appuierai sur un exemple précis : l'estimation globale des émissions de N_2O provenant des sols agricoles. Deux bases de données seront utilisées : la base de données de Rochette et Janzen (2005) et celle de Stehfest et Bouwman (2006). Elles recensent de nombreuses mesures d'émissions de N_2O réparties dans le monde provenant d'études publiées et ont joué un rôle important lors des estimations d'émission de N_2O réalisées par le GIEC.

Ce manuscrit est basé sur quatre chapitres centraux rédigés sous forme d'articles.

Le chapitre 1 recense les méta-analyses en agronomie et les évalue à l'aide de **huit critères de qualité.** Un cas d'étude basé sur les données d'émissions de N_2O par les légumineuses issues de Rochette et Janzen (2005) permet d'illustrer en détail l'ensemble de ces critères.

Dans le chapitre 2, j'étudie la sensibilité du facteur d'émission de N_2O estimé par Stehfest et Bouwman (2006) (et utilisé par le GIEC) à la **méthode statistique** utilisée pour analyser les données et à la **forme de la relation** entre émission de N_2O et dose d'engrais azoté.

Le chapitre 3 présente une analyse de sensibilité aux **données** et aux **variables prédictives** grâce à une méthode d'apprentissage automatique *Random Forest* qui permet de gérer les données manquantes. Des relations entre les émissions de N_2O et plusieurs variables caractérisant le milieu et les pratiques culturales sont établies et comparées en utilisant la base de données de Stehfest et Bouwman (2006). J'étudie également dans ce chapitre la sensibilité du niveau de précision des prédictions de N_2O à la **méthode de prédiction utilisée**

Le chapitre 4 présente une analyse de sensibilité au **plan d'expérience** utilisé pour réaliser les mesures d'émissions de N_2O afin de pouvoir prédire des émissions de N_2O pour des doses de fertilisation azotée non testée. J'étudie également la sensibilité des prédictions à **la méthode statistique** utilisée pour estimer les paramètres des modèles.

Enfin, les principaux résultats obtenus dans les quatre chapitres sont discutés et mis en perspective.

Chapitre 1

Evaluation de la qualité des méta-analyses en agronomie

Assessment of the quality of meta-analysis in agronomy

Aurore Philibert[b,a], Chantal Loyce[b,a], David Makowski[a,b]

[a] INRA, UMR 211 Agronomie, F-78000 Thiverval Grignon, France
[b] AgroParisTech, UMR 211 Agronomie, F-78000 Thiverval Grignon, France

Abstract

A meta-analysis is a statistical treatment of a dataset derived from a literature review. Meta-analysis appears to be a promising approach in agricultural and environmental sciences, but its implementation requires special care. We assessed the quality of the meta-analyses carried out in agronomy, with the intent to formulate recommendations, and we illustrate these recommendations with a case study relative to the estimation of nitrous oxide emission in legume crops. Eight criteria were defined for evaluating the quality of 73 meta-analyses from major scientific journals in the domain of agronomy. Most of these meta-analyses focused on production aspects and the impact of agriculture activities on the environment or biodiversity. None of the 73 meta-analyses reviewed satisfied all eight quality criteria and only three satisfied six criteria. Based on this quality assessment, we formulated the following recommendations: i) the procedure used to select papers from scientific databases should be explained, ii) individual data should be weighted according to their level of precision when possible, iii) the heterogeneity of data should be analyzed with random-effect models, iv) sensitivity analysis should be carried out and v) the possibility of publication bias should be investigated. Our case study showed that meta-analysis techniques would be beneficial to the assessment of environmental impacts because they make it possible to study between site-year variability, to assess uncertainty and to identify the factors with a potential environmental impact. The quality criteria and recommendations presented in this paper could serve as a guide to improve future meta-analyses made in this area.

Keywords: meta-analysis, nitrous oxide, mixed model, Bayesian statistics, sensitivity analysis, legume

1.   Introduction

Systematic reviews are frequently carried out to compile research studies on a specific subject (Evans and Foster, 2011). They involve a rigorous scientific approach comprising the collection, evaluation and synthesis of all studies on a given topic, sometimes contradictory ones, while limiting the introduction of bias (Bland et al., 1995). Systematic reviews may be qualitative if they provide a synthesis of research studies (*e.g.*, Robson et al., 2002), or quantitative, if they involve the processing of a set of data gathered from previous publications. These two kinds of approaches are useful for summarizing large numbers of papers and for the objective establishment (Gurevitch and Hedges, 1993) of what is known and unknown in a specific field (Yuan and Hunt, 2009).

Quantitative systematic reviews are generally referred to as ―meta-analyses‖ when a statistical treatment is applied to a dataset derived from a literature review. The term ―meta-analysis‖ was first coined by Glass in 1976, in the field of educational science, and is defined as a ―statistical analysis of a large collection of results from individual studies―(Glass, 1976). A meta-analysis includes typically the following steps (Borenstein et al., 2009; Doré et al., 2011): (i) definition of the objective of the meta-analysis and of the response variable to be estimated from the data. For example, in Miguez and Bollero (2005), the response variable is the ratio of maize yield after a winter cover crop to maize yield in the absence of a cover crop, (ii) systematic review of the literature and/or of the dataset reporting values of the response variable, (iii) analysis of data quality (*i.e.*, quality of experimental design and measurement techniques, precision of the response variable), (iv) assessment of between-study variability and heterogeneity, (v) assessment of publication bias, (vi) presentation of the results and of the level of uncertainty.

To date, most of the meta-analyses carried out concerned medical science (Normand, 1999; Sutton et al., 2000). In this field, meta-analysis aims (i) to detect an overall treatment effect, (ii) to evaluate the variability between studies, or (iii) to identify study characteristics associated with really effective treatments (Normand, 1999).

Meta-analysis has become an essential technique in human health, and an international organization, the Cochrane Organization, was created in 1993 to prepare, update and promote meta-analyses in this domain (http://www.cochrane.org/). In human health, meta-analyses have long been considered as a field of research in their own right (Cucherat et al., 1997). Meta-analysis has been also applied to other areas of science (although less systematically than in human health), such as ecology (*e.g.*, Arnqvist and Wooster, 1995; Cardinale et al., 2006; Stewart, 2010), plant pathology (Madden and Paul, 2011; Rosenberg et al., 2004) and animal science (Sauvant et al., 2008).

Table 1.1: Definition of the eight criteria used to assess the quality of meta-analyses

Criterion	Definition
Repeatable procedure	A repeatable procedure for the selection of papers for the meta-analysis is presented
References	A list of the references used for the meta-analysis is provided
Heterogeneity	The origins of the variability of the results are analyzed
Sensitivity analysis	The sensitivity of the conclusion to observations or methods is analyzed
Investigation of publication bias	The publication bias is studied
Weighting	Observations are weighted according to their level of accuracy in the statistical model
Availability of the dataset	The dataset is available in an electronic format or published directly in the paper
Availability of the program	The program used for statistical analysis is made available

Doré et al. (2011) recommended the more systematic use of meta-analysis in agronomy. A considerable amount of experimental data is available from papers published in agronomic journals, and such data could be reviewed, combined and analyzed with statistical techniques to rank cropping systems (within a given environment) according to their impact on crop production and on key environmental variables, such as water nitrate content, the emission of greenhouse gases (*e.g.*, N_2O) or the presence/absence of species of ecological interest (*e.g.*, earthworms, birds). According to Doré et al. (2011), the meta-analysis framework provides an interesting alternative to dynamic crop models (*e.g.*, Brisson et al., 2003; Jones et al., 2003; Keating et al., 2003; Stöckle et al., 2003; van Ittersum et al., 2003) because these models include several sources of uncertainty (Monod et al., 2006) and their predictions are not always reliable (*e.g.*, Barbottin et al., 2008; Makowski et al., 2009).

Meta-analysis appears to be a promising approach for assessing the agronomic and environmental performances of cropping systems, but its implementation requires special care and the value of a meta-analysis may be greatly decreased by the use of inappropriate techniques. Indeed, there is a risk of biased estimation, misinterpretation and incorrect conclusions in meta-analyses performed without sufficient quality control (Sutton et al., 2000). Several authors have proposed quality criteria that could be used to assess the quality of a meta-analysis (Borenstein et al., 2009; Gates, 2002; Roberts et al., 2006; Sutton et al., 2000), but these criteria have not yet been used to assess the quality of the meta-analyses carried out in agronomy.

We therefore assessed the quality of the meta-analyses carried out in agronomy, with the intent to formulate recommendations. We illustrate these recommendations with a case study on the estimation of the emission by legume crops of nitrous oxide, a very potent greenhouse gas with a global warming potential 298 times greater than that of CO_2 (IPCC, 2007).

2. Materials and methods

2.1. Criteria for quality assessment

We defined eight criteria (Table 1.1), based on the findings of previous studies (Borenstein et al., 2009; Roberts et al., 2006; Gates, 2002), for assessment of the various steps in meta-analyses carried out in agronomy:

1) Correct description of the bibliographic search procedures used by the authors to select the individual studies (*i.e.*, papers) and the repeatability of these procedures;

2) Listing of the references of the selected individual studies used in the meta-analysis;

3) Analysis of the variability of the results of individual studies, including checking to see whether the results vary between the selected individual studies and, when relevant, investigation of the sources of between-study variability (*e.g.*, using random-effects model).

Evaluation of the between-study variability of the response variable and of differences in the accuracy of individual estimates is an important step in a meta-analysis and several statistical methods have been proposed for the estimation of between- and within-study variances (Borenstein et al., 2009);

4) Analysis of the sensitivity of the conclusions to any change in the dataset and/or in the statistical method used to analyze the data. Sensitivity analyses should be carried out to identify influential data and to assess the robustness of the main conclusions of a meta-analysis to the assumptions made in the statistical analysis;

5) Assessment of the publication bias, which occurs when only studies with highly significant results are published. In this case, a meta-analysis can lead to a biased conclusion and an overestimation of the effect of a given factor. Publication bias is a predominant issue in meta-analysis and several methods such as funnel plots (*e.g.*, Borenstein et al., 2009; Light and Pillemer, 1984) have been developed to detect the presence of such bias in datasets including published results;

6) Data weighting. When the results reported in the individual studies differ in their levels of accuracy, weighting of the data according to their levels of precision is recommended, based, for example, on the inverse of the variance of the measurements, as suggested by Hedges and Olkin (1985);

7) Availability of the dataset;

8) Availability of the program used for statistical analysis.

These last two criteria are used to determine whether the meta-analysis could easily be re-run.

2.2. Assessment of the quality of the meta-analyses carried out in agronomy

The quality of the meta-analyses carried out in agronomy was assessed with the eight criteria listed above. One hundred and thirty six scientific journals publishing papers in agronomy were selected for this purpose. These journals were referred to in the Journal of Citation Report (JCR) as journals publishing papers in Agronomy, Agriculture Multidisciplinary, Agricultural Engineering, or Environmental Sciences. Journals belonging to the first three categories are further referred to as Agronomy and Agricultural journals. The scopes of these journals were analyzed and found to be consistent with either the American or European definition of agronomy. The American Society of Agronomy (ASA) definition is "the application of soil and plant sciences to crop production that incorporates the wise use of natural resources and conservation practices to produce food, feed, fuel, fiber, and pharmaceutical crops while maintaining and improving the environment". The definition of the European Society of Agronomy (ESA) is "the relationships between crops, soils, climates and agricultural practices, and between agriculture and the environment".

The papers published in these journals were screened with a systematic literature search (until August 16 2011) using the key-word "meta-analysis" (or the equivalent in other languages,
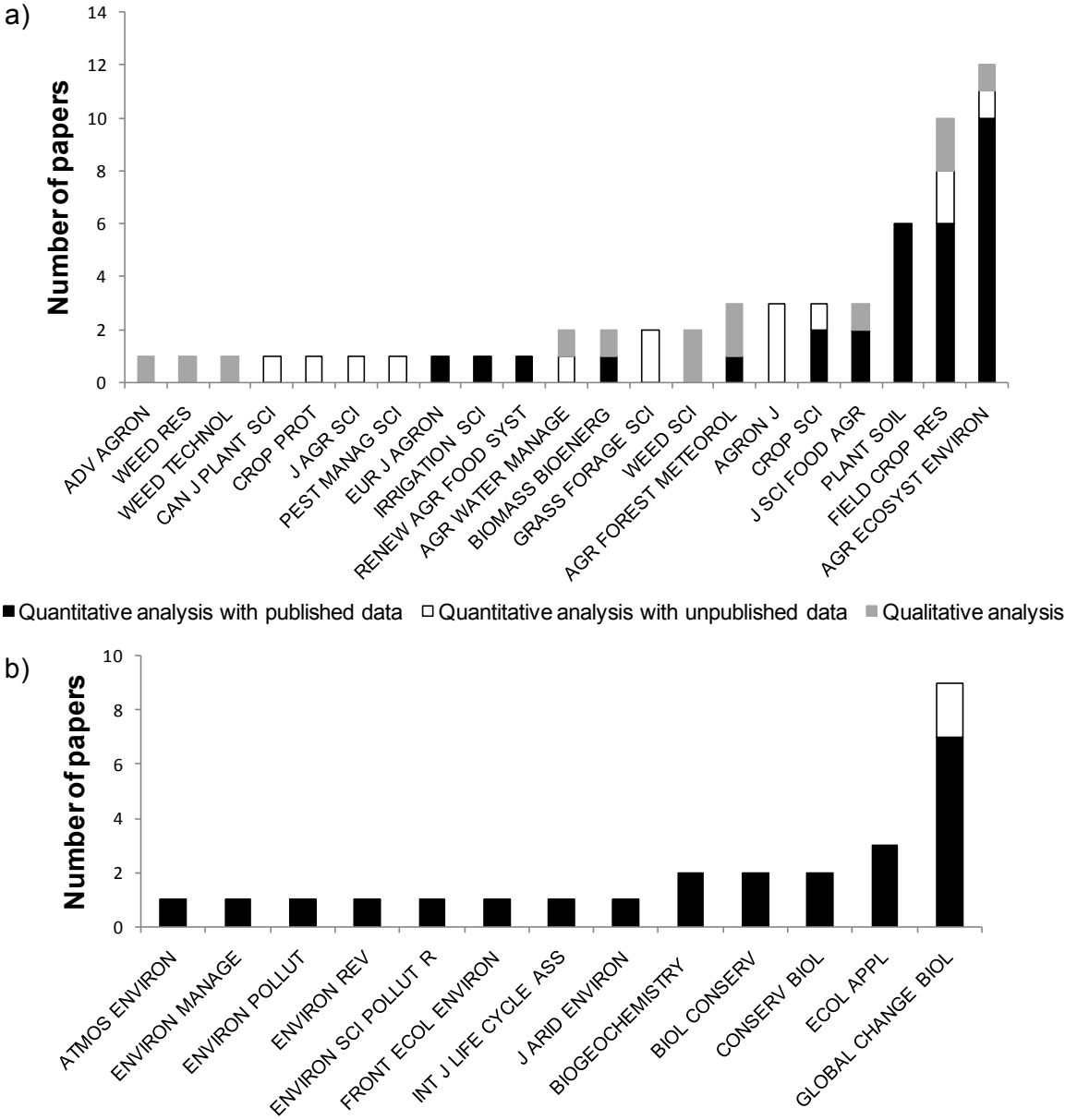
Figure 1.1 : Number of papers reporting results of meta-analyses in agronomy published in a) Agronomy, Agriculture Multidisciplinary and Agricultural Engineering journals, and in b) Environmental Sciences journals.

according to the language of the journal) except for Environmental Sciences journals where the key-words ―meta-analysis‖ and crop* were used. The ScienceDirect, WileyInterScience and SpringerLink databases and the websites of the 136 journals were analyzed and 1729 articles were identified in 94 of the 136 selected journals.

The 1729 papers were studied, and 1645 were excluded for the following reasons:
- The word ―meta-analysis‖ appeared only in the references of 1444 papers;
- The topic concerned was related to a discipline other than agronomy (*e.g.*, social science, animal science, focus on forest crop only, plant physiology) in 125 papers;
- Meta-analysis was mentioned but not carried out in 44 papers;
- 32 papers were not research articles (index, table of contents);

Thus, 84 papers in total were found to report the results of a meta-analysis. Thirteen of these papers reported a qualitative analysis of the dataset, 16 reported the results of a quantitative analysis of unpublished data, and 55 reported the results of a quantitative analysis of published data (Fig. 1.1).

We focused on the 55 papers reporting the results of quantitative analyses of published data. Thirty-one were published in Agronomy and Agricultural journals (A&A journals) and 24 in Environmental Sciences journals (ES journals). As several meta-analyses were conducted in some of these papers, there were 73 quantitative meta-analyses in total; 45 published in A&A journals and 28 in ES journals. Each of these meta-analyses was assessed according to the eight criteria listed in Table 1.1. Recommendations concerning statistical procedures in meta-analysis were then formulated, based on the results of the quality assessment.

2.3. Case study on the estimation of nitrous oxide emission by legume crops

The recommendations formulated after the quality assessment were illustrated by a case study of nitrous oxide emission by legume crops during the growing season.

2.3.1. Data

The dataset was extracted from the paper by Rochette and Janzen (2005). It included seventeen values for nitrous oxide emission measured in ―pure legume forage stands‖- alfalfa (*Medicago sativa*) and clover (*Trifolium pretense* and *Trifolium repens*) - published in nine papers. These papers, published from 1982 to 2004, reported the results of experiments carried out at nine experimental sites. Alfalfa experiments were located in North America and clover experiments in Europe and New Zealand.

The response variable of the meta-analysis was nitrous oxide emission, expressed in kg N ha^{-1} year^{-1}. We used the following explanatory variables: (i) location of the experiment, (ii) type of crop (alfalfa or clover), (iii) number of days of the experiment.

Table 1.2: Description and assessment of the 16 statistical models for the analysis of N_2O emission data: use of a crop effect, use of a random site effect, use of observation weights, frequentist or Bayesian estimation method, values of the Akaïke, Schwartz, and Deviance Information criteria (AIC, BIC, DIC).

Model number	Crop effect	Random site effect	Weight	Frequentist / Bayesian	AIC	BIC	DIC
1	yes	yes	no	Bayesian	-	-	24.23
2	no	yes	no	Bayesian	-	-	25.96
3	yes	no	no	Bayesian	-	-	35.16
4	no	no	no	Bayesian	-	-	37.1
5	yes	yes	no	Frequentist	32.67	35.51	-
6	no	yes	no	Frequentist	34.6	36.92	-
7	yes	no	no	Frequentist	36.94	39.06	-
8	no	no	no	Frequentist	38.78	40.33	-
9	yes	yes	yes	Bayesian	-	-	23.65
10	no	yes	yes	Bayesian	-	-	24.62
11	yes	no	yes	Bayesian	-	-	36.43
12	no	no	yes	Bayesian	-	-	39.01
13	yes	yes	yes	Frequentist	33.84	36.67	-
14	no	yes	yes	Frequentist	35.28	37.6	-
15	yes	no	yes	Frequentist	38.8	40.92	-
16	no	no	yes	Frequentist	40.77	42.32	-

2.3.2. Statistical models

Sixteen models were fitted to data, to determine the relationship between nitrous oxide emission and the explanatory variables. These models differed in terms of their explanatory variables (with/without crop effect, with/without location effect), the probability distribution of the residual error of the model (with/without weighting by the number of days of the experiment) and the estimation method used (frequentist/Bayesian) (Table 1.2).

The type of crop was defined as a fixed effect and the location of the experiment was defined as a random effect describing between-site variability.

Specifically, model 5 is defined by: $\log(Y_{ij}) = \mu + \alpha X_{ij} + b_i + \varepsilon_{ij}$ with $b_i \sim N(0, \sigma^2)$ and $\varepsilon_{ij} \sim N(0, \tau^2)$, where Y_{ij} is the nitrous oxide emission in kg N ha^{-1} year^{-1}, μ is emission from alfalfa and $\mu + \alpha$ is emission from clover (X_{ij} is 1 for clover observations and zero for alfalfa observations), $i=1,\ldots,9$ is the number of experimental sites (equal to the number of papers) and $j=1,\ldots,n_i$ with n_i the number of observations per location; $n_i \in [1;5]$. b_i and ε_{ij} are two random terms corresponding to the site effect and to the residual error of the model, respectively, both of which are assumed to be independent and normally distributed. The variability of the measurements within a given location is described by the residual error ε_{ij}. We carried out a log transformation of the response variable to normalize the residuals.

Models 6, 7, and 8 are simplified versions of model 5. Model 6 includes a random location effect but no crop effect, and is defined by $\log(Y_{ij}) = \mu + b_i + \varepsilon_{ij}$. Model 7 includes a crop effect but no location effect, and is defined by $\log(Y_{ij}) = \mu + \alpha X_{ij} + \varepsilon_{ij}$. Model 8 is simply defined by $\log(Y_{ij}) = \mu + \varepsilon_{ij}$.

Models 13-16 are based on the same equations as models 5-8 except that the residual error variance of the model is defined by $Var(\varepsilon_{ij}) = \dfrac{\tau^2}{Z_{ij}}$, with Z_{ij} the number of days of each experiment at each location. Thus, according to models 13-16, the higher the number of days of experiment, the lower the variance is.

The parameters of models 5-8 and models 13-16 were estimated from data using a frequentist method (maximum likelihood).

Models 1-4 and models 9-12 are based on the equations of models 5-8 and of models 13-16, respectively, but their parameters were estimated using a Bayesian method (Markov chain Monte Carlo, MCMC). Prior parameter probability distributions were defined with Gamma and Normal distributions, as σ^{-2}, $\tau^{-2} \sim \Gamma(0.001, 0.001)$ and $\mu, \alpha \sim N(0, 1000)$. μ and α had a
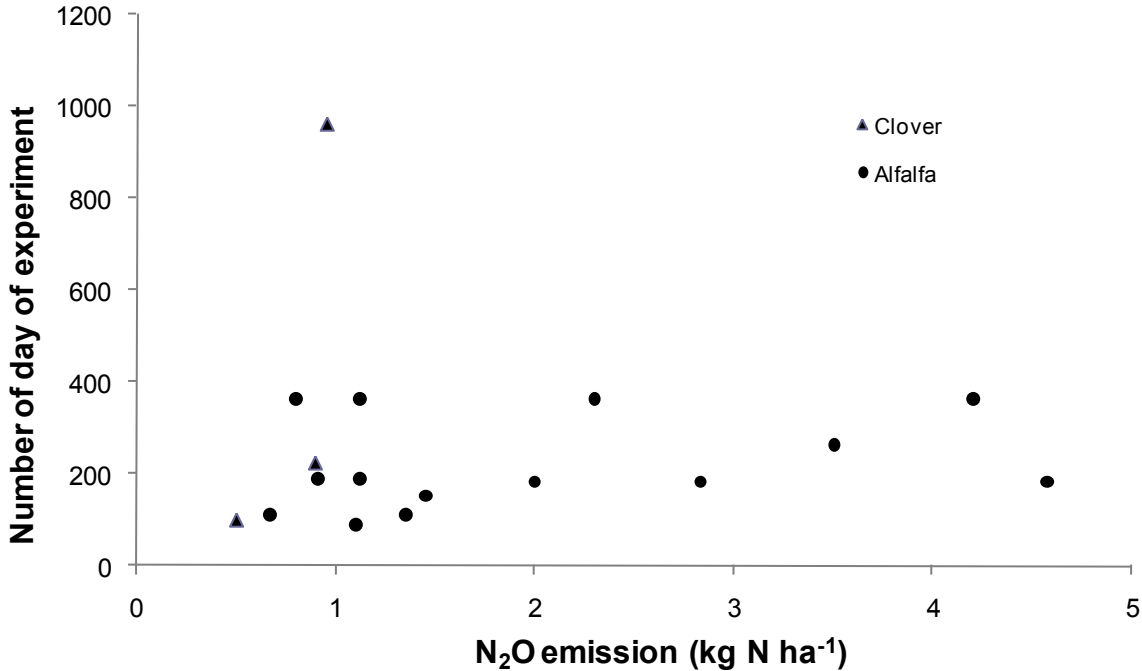
Figure 1.2: Funnel plot of nitrous oxide emission and number of days of experiment.

mean of zero and a standard deviation of 32, which is large given the range of response variable values, which extended from 0.5 to 5. Along the same lines, σ^{-2} and τ^{-2} had a mean of 1 and a standard deviation of 1000. These distributions represent broad *a priori* distribution when compared with the data.

Frequentist models were assessed by calculating the Akaike Information Criterion (AIC) and Schwartz Criterion (BIC) (Akaike, 1974; Burnham and Anderson, 2002). Bayesian models were assessed by calculating the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). The best models are those with the lowest AIC, BIC and DIC.

The Jackknife technique (Tudoreanu and Phillips, 2004) was used to study the sensitivity of estimates of N_2O emission to each observation of the dataset. Each observation was removed from the dataset in turn, and the models were fitted to the remaining data. The relationship between the values of the measured emission and their influence on the estimates was analyzed.

We used R software for frequentist models and Winbugs software (implementing MCMC algorithms) for Bayesian models. For MCMC simulation, three chains were run, with 20,000 to 200,000 iterations, depending on the model.

2.3.3. Quality assessment

The dataset used in our case study was published in the paper by Rochette and Janzen (2005) (criterion 1); the list of references and the dataset (quality criteria 2 and 7) were presented in this paper.

In our case study, the heterogeneity (criterion 3) of nitrous oxide emissions between individual studies was analyzed with the statistical models presented above. A sensitivity analysis (criterion 4) was performed to identify influential observations and to assess the effect of the statistical models on the estimated values (by testing sixteen different models). Publication bias (criteria 5) was investigated by generating a funnel plot (Fig. 1.2). A funnel plot usually presents the precision of the effect size (inverse of its standard error) *versus* the effect size (Borenstein et al., 2009). As the variances of the N_2O measures were not available in our dataset, the lengths of the experiments (numbers of days) were plotted in function of the measured N_2O emissions, and no publication bias was found based on this figure (Fig. 1.2). Individual data were weighted (criterion 6) by the number of days of the experiment (Model 9-16), as the variances for nitrous oxide emissions were not available. Finally, the code used for statistical analysis is available, on request, from the corresponding author (criterion 8).
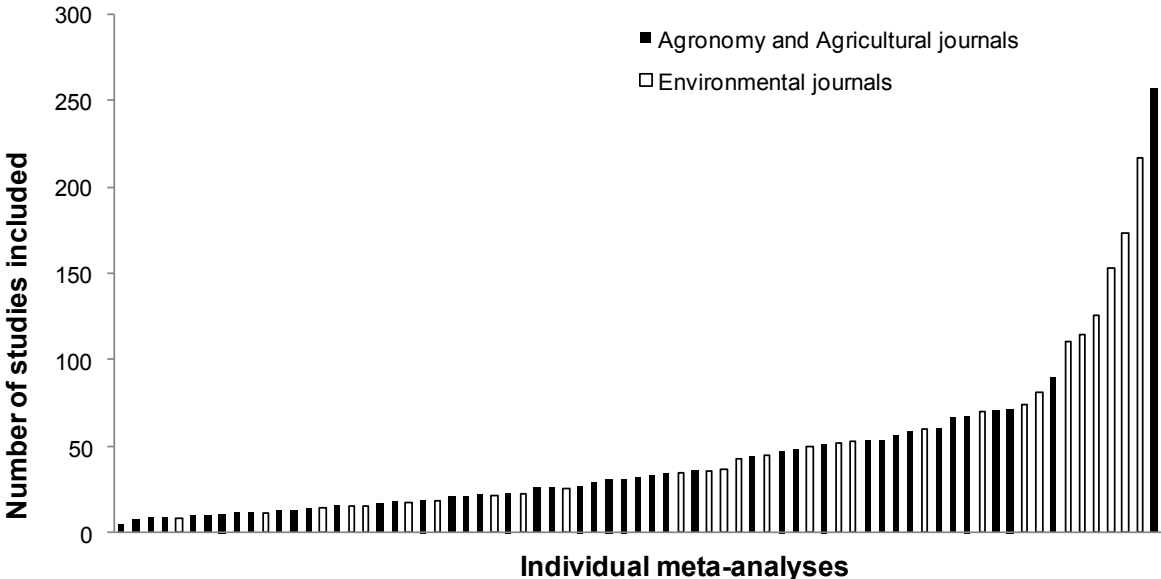
Figure 1.3 : Number of individual studies used in each of the 73 quantitative meta-analyses. The median number was 32. Results for meta-analyses published in A&A journals were presented in black whereas results for meta-analyses published in ES journals were presented in white.

3.   **Results**

3.1. Description of the 73 selected meta-analyses

The 55 papers (Annex 1) reporting meta-analyses of published data were found in 23 journals (Fig. 1.1): Agriculture, Ecosystems & Environment (10 papers), Global Change Biology (7 papers), Field Crops Research (6 papers), Plant and Soil (6 papers), Ecological Applications (3 papers), Biogeochemistry (2 papers), Biological conservation (2 papers), Conservation Biology (2 papers), Crop Science (2 papers), Journal of the Science of Food and Agriculture (2 papers), Agricultural and Forest Meteorology (1 paper), Atmospheric Environment (1 paper), Biomass and Bioenergy (1 paper), Environmental Management (1 paper), Environmental Pollution (1 paper), Environmental reviews (1 paper), Environmental Science and Pollution Research (1 paper), European Journal of Agronomy (1 paper), Frontiers in Ecology and the Environment (1 paper), International Journal of Life Cycle Assessment (1 paper), Irrigation Science (1 paper), Journal of Arid Environments (1 paper) and Renewable Agriculture and Food Systems (1 paper). These papers were published from 2001 to 2011. Eleven papers were published between 2001 and 2007, the others being published after 2007.

Twelve of the 55 papers reported results from more than one meta-analysis, bringing the total number of meta-analyses published in these papers to 73, as explained above. The number of individual studies analyzed in these meta-analyses ranged from 5 to 257, with a median of 32 for the 73 meta-analyses, 27 for meta-analyses published in A&A journals and 44 for meta-analyses published in ES journals (Fig. 1.3).

Crop yield and production quality were used as response variables respectively in 40 and 3 meta-analyses. In fifteen meta-analyses, environmental characteristics: soil carbon (10 meta-analyses), gas emissions (4) and nitrate leaching (1) were considered as response variables. Pests or natural enemies of pests were used as response variables respectively in 5 and 4 meta-analyses. The response variable was a biodiversity indicator in 8 meta-analyses. In the remaining two meta-analyses, the response variable was related to cropping system (crop rotation, Thenail et al., 2009) and plant response (Kaschuk et al., 2011). In four meta-analyses, several types of response variable were considered.
Explanatory variables identified to explain the variability of the response variables were management practices (46 meta-analyses), environmental characteristics (climatic variables in 11 meta-analyses and soil characteristics in 5 meta-analyses), land use (land use characteristics in 9 meta-analyses, land use change in 6 meta-analyses), plant characteristics (3 meta-analyses), and pest natural enemy (1 meta-analysis). Several types of explanatory variables were considered in 5 meta-analyses.

Table 1.3 : Statistical methods and software used in the 73 meta-analyses reporting the results of quantitative analysis of published data.

Statistical methods	Software											Total
	MetaWin	R	SAS	Blossom	Genstat	Irene	MIX	SPLUS	SPSS	Statistica	Unknown	
Mean effect size	30		2								8	40
Random coefficients regression model		6	5		1		1	1	1		6	21
Regression model			2	2		1				1	2	8
Statistical tests		4										4
Total	30	10	9	2	1	1	1	1	1	1	16	73
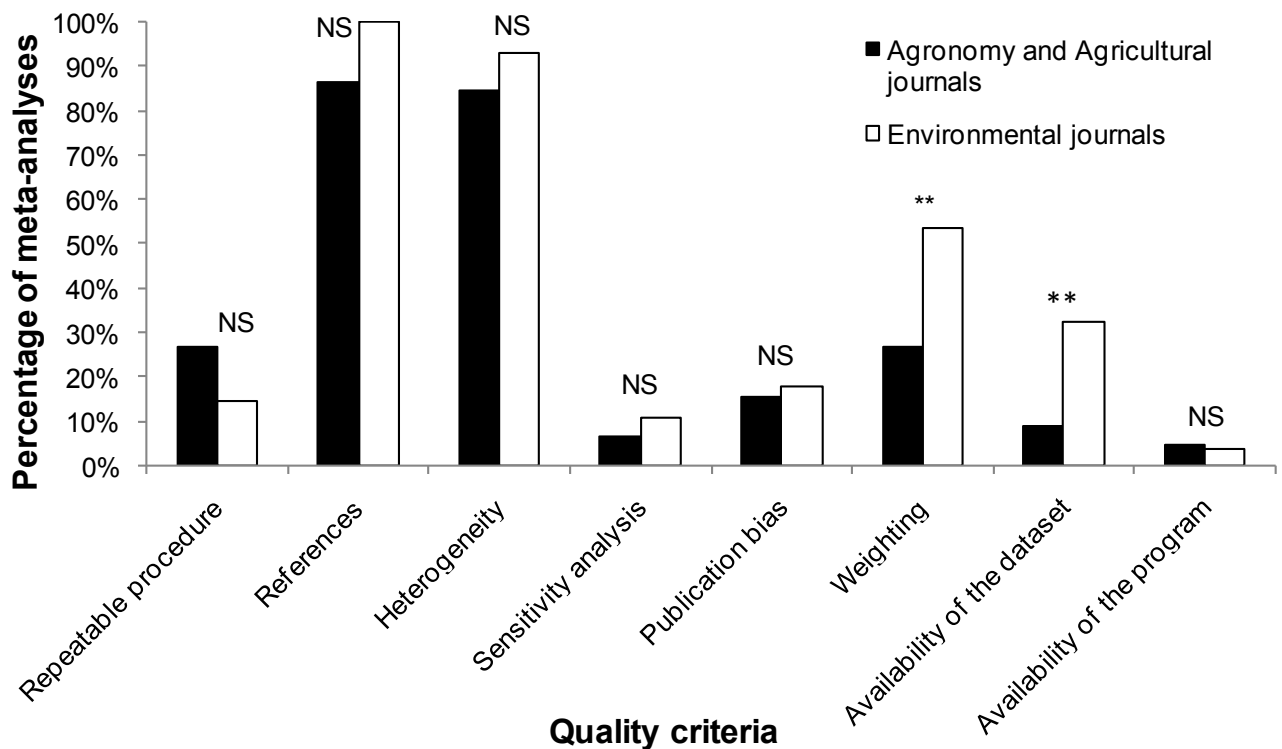
Figure 1.4 : Number of meta-analyses (out of 73 studied) satisfying each of the 8 quality criteria. NS for a non-significant difference for this criterion, between meta-analyses published in A&A and ES journals and ** for a significant difference with a p value < 0.05. Results for meta-analyses published in A&A journals were presented in black whereas results for meta-analyses published in ES journals were presented in white.

Lastly, in 30 of the 73 meta-analyses, the response variables dealing with crop production were related to explanatory variables dealing with management practices.

The statistical methods used by the authors are summarized in Table 1.3. Mean effect size method (*i.e.*, mean individual estimated effect size weighted by the inverse of the variance of the effect size) was used in 40 of the 73 meta-analyses. In the other meta-analyses, statistical analysis was based on a regression model with random coefficients (21) or fixed coefficients (8) or statistical tests (4). MetaWin, a specialized software for meta-analyses, was used in 30 meta-analyses, R in 10 meta-analyses and SAS® in 9 meta-analyses. Software and statistical methods were found to be significantly related (Chi-squared test, p value=$2.2*10^{-10}$): random coefficients regression models were generally implemented in SAS® and R, and mean effect size was systematically computed with MetaWin. Bayesian statistical techniques were not used in any of these meta-analyses.

3.2. Assessment of the quality of the meta-analyses reviewed

The ―References" criterion (criterion 2) was satisfied by 92% of the meta-analyses (87% of the meta-analyses published in A&A journals and 100% of the meta-analyses published in ES journals). The ―Heterogeneity" criterion (criterion 3) was satisfied by 88% of the meta-analyses; 84% of the meta-analyses published in A&A journals and 93% of the meta-analyses published in ES journals (Fig. 1.4). The ―Repeatable procedure" criterion (criterion 1) was satisfied by 22% of the meta-analyses (27% published in A&A journals and 14% in ES journals) and publication bias in the selection of papers (criterion 5) was studied in 16% of the meta-analyses (16% published in A&A journals and 18% in ES journals). Sensitivity analysis (criterion 4) was performed in 8% of the meta-analyses (7% published in A&A journals and 11% in ES journals) and the program was made available (criterion 8) in 4% of the meta-analyses (4% published in A&A journals and 4% in ES journals). Observations were weighted (criterion 6) in 37% of the meta-analyses (27% of the meta-analysis published in A&A journals and 54% of the meta-analyses published in ES journals, p value < 0.05). The dataset was made available in 18% of the meta-analyses (9% and 32 % of the meta-analyses for A&A and ES journals respectively, p value < 0.05).

None of the meta-analyses fulfilled all eight criteria, but three meta-analyses satisfied at least six criteria (Philpott et al., 2008; Kiær et al., 2009; Akiyama et al., 2010).

Publication bias, sensitivity analysis, and availability of the program were the criteria with the lowest scores for both A&A and ES journals. The fourth criterion with the lowest score was the repeatable procedure for meta-analyses published in ES journals, and the availability of the dataset for meta-analyses published in A&A journals.

Table 1.4 : Mean, standard deviation, 2.5% and 97.5% percentiles of nitrous oxide emission calculated with models 1, 5, 9 and 13 for alfalfa and clover crops.

Model number	Crop	N$_2$O emission (kg N ha^{-1} year^{-1})				P(N$_2$O emission≥1.8)[b]
		Mean	Sd[a]	Percentile 2.5%	Percentile 97.5%	
1	Alfalfa	2	1.10	0.75	4.36	0.49
5	Alfalfa	2.03	1.08	0.68	4.78	0.50
9	Alfalfa	2.02	1.10	0.71	4.32	0.49
13	Alfalfa	2.04	1.14	0.64	4.96	0.49
1	Clover	0.89	0.63	0.24	2.38	0.06
5	Clover	0.86	0.45	0.28	2.01	0.04
9	Clover	0.97	0.88	0.24	2.61	0.08
13	Clover	0.92	0.51	0.29	2.23	0.06

[a] sd=standard deviation

[b] The probability of exceeding the value of 1.8 kg N ha^{-1} year^{-1} was calculated with each model for each crop.
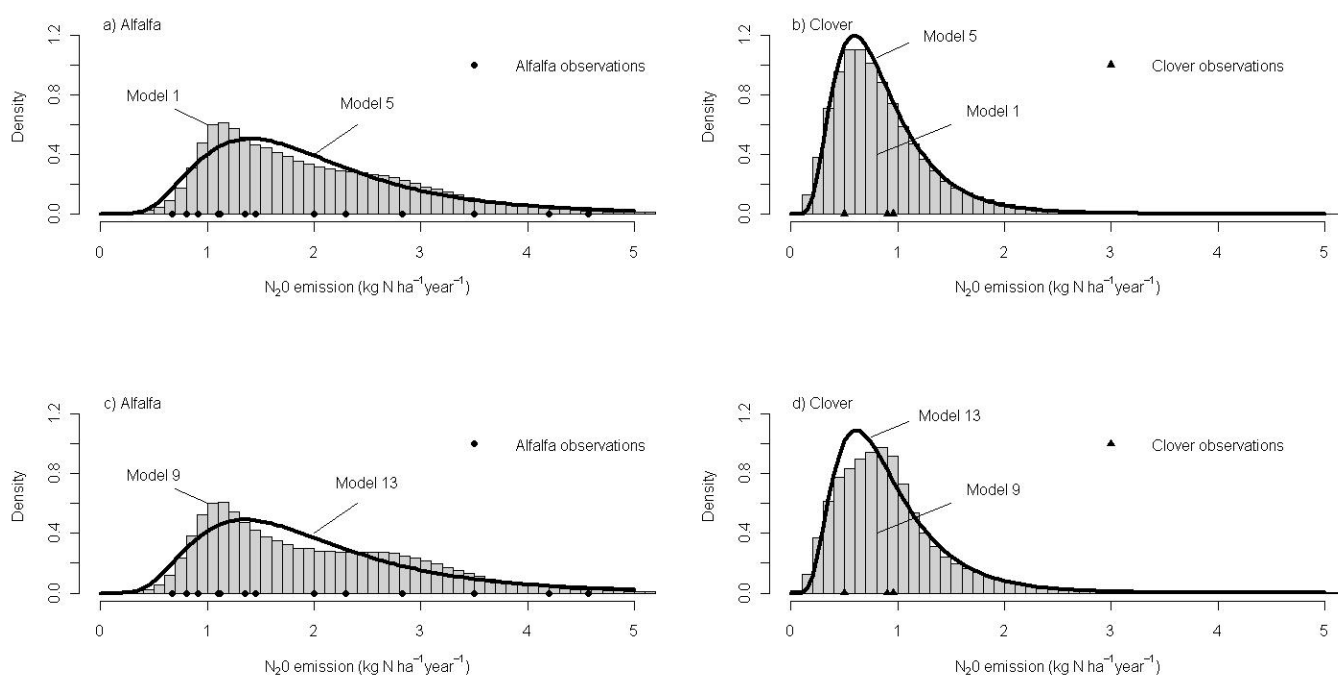
Figure 1.5 : Probability distribution of N$_2$O emission obtained with frequentist models 5 and 13 (black curves) and with Bayesian models 1 and 9 (gray histograms). The seventeen observations used for the statistical analysis are represented in black circles for alfalfa crops and in black triangles for clover crops.

The use of sensitivity analysis, weighting and Bayesian techniques is illustrated in the case study below.

3.3. Case study of the estimation of nitrous oxide emission by legume crops

3.3.1. Selection of statistical models

AIC and BIC values showed that models including random site effects gave better results; significant between-site variability of N_2O emission was observed. AIC, BIC, and DIC values also showed that the crop effect was significant in both frequentist and Bayesian models. Finally, the use of a model with a weighted residual variance decreased DIC, but not AIC and BIC.

Model 9 (*i.e.*, the model with a weighted residual variance) was the Bayesian model with the lowest DIC. The frequentist model giving the lowest AIC and BIC values was model 5. However, the DIC of model 1 was close to the value obtained for model 9, and the AIC and BIC values of model 13 were close to the values obtained for model 5. Hence, only the results obtained with models 1, 5, 9, and 13 were considered further for the estimation of nitrous oxide emission, since these models showed the best performances (*i.e.*, the lowest AIC, BIC or DIC).

3.3.2. Estimation of nitrous oxide emission

Results were initially obtained in log units and were back-transformed for expression of the model estimates in kg N ha^{-1} year^{-1} units. All the selected models included a random site effect, so their output was the probability distribution describing the between-site variability of N_2O emissions. These distributions are presented for alfalfa and clover separately, in Figure 1.5. The means, standard deviations and percentiles of these distributions were similar (Table 1.4).

All models showed that emissions were much higher in alfalfa than in clover. Estimated mean nitrous oxide emission ranged from 2 to 2.04 kg N ha^{-1} year^{-1} in alfalfa, depending on the model used, and from 0.86 to 0.97 kg N ha^{-1} year^{-1} in clover. The 95% confidence intervals obtained were slightly larger with frequentist models than with Bayesian models for alfalfa N_2O emission, and larger for Bayesian models than for frequentist models for clover N_2O emission. However, overall, the two types of model gave similar results. The differences between the 2.5 and 97.5 percentiles were large, with all models revealing considerable between-site variability for N_2O emissions (Table 1.4, Figure 1.5).

We compared the mean values estimated with our models with the estimate used by Rochette and Janzen (2005) (1.8 kg N ha^{-1} year^{-1} for clover and alfalfa), by calculating the probability of exceeding this value based on the probability distributions shown in Figure 1.5. The value of
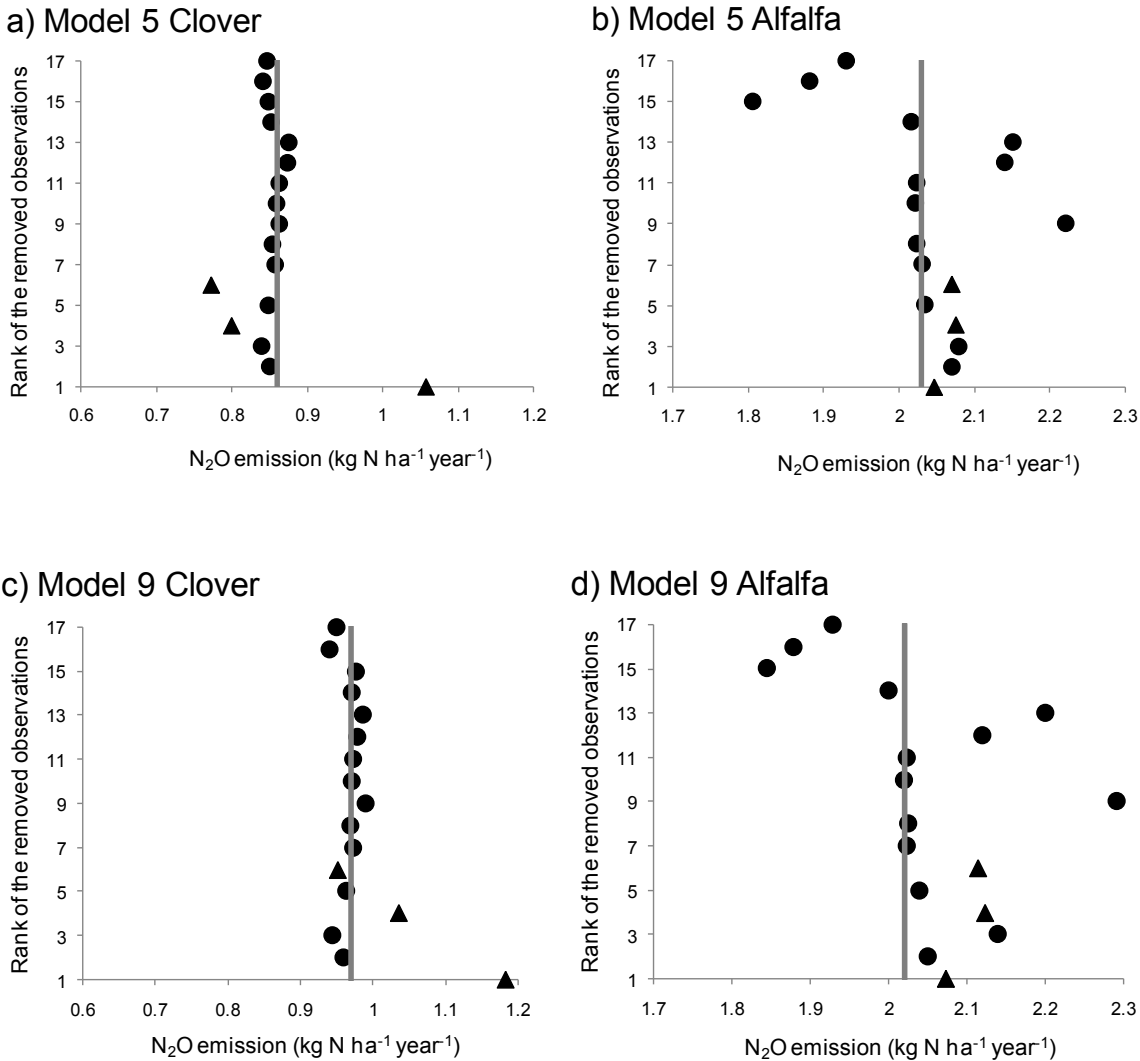
Figure 1.6 : Sensitivity analysis for models 5 and 9. Each point indicates N_2O emission estimated (on the x axis) with model 5 (a for clover and b for alfalfa) and model 9 (c for clover and d for alfalfa), when a given observation (on the y axis) is removed from the dataset. The x axis shows estimated nitrous oxide emissions. They y axis shows the rank of the information removed. Observations were ranked in decreasing order of N_2O emission (observation number 1 has the lowest value of nitrous oxide emission). Gray vertical lines represent the estimate obtained with the model for all observations. Black circles show estimates of nitrous oxide emission when one alfalfa observation is removed and black triangles show estimates when one clover observation is removed.

1.8 kg N ha^{-1} year^{-1} was exceeded with a probability close to 0.5 for alfalfa and a probability of 0.04-0.08 for clover, depending on the model (Table 1.4). As the between-site variability of N$_2$O emissions is strong, it is more relevant to present confidence intervals for N$_2$O emissions (Table 1.4) than absolute-number estimates (*e.g.*, 1.8 kg N ha^{-1} year^{-1}).

3.3.3. Analysis of sensitivity to individual data

The results of the Jacknife procedure are presented in Figure 1.6. The circles and triangles at some distance from the gray vertical line indicate influential data respectively for alfalfa and clover. The estimated mean level of N$_2$O emission in clover was sensitive to the data collected in clover fields, but not to the data collected in alfalfa fields (Figure 1.6ac). For N$_2$O emissions in clover, the most influential observation was #1; the removal of this observation increased the estimated mean emission from 0.86 to 1.06 kg ha^{-1} year^{-1} with model 5, and from 0.97 to 1.18 kg ha^{-1} year^{-1} with model 9.

Estimated mean N$_2$O emission in alfalfa was sensitive to data #9, #12, #13, #15, #16 and #17 (Figure 1.6bd). These data were collected in alfalfa fields. The strong influence of data #15, #16 and #17 may reflect the association of these data with high N$_2$O emissions. However, data #9, #12 and #13 were also influential but corresponded to only moderate levels of N$_2$O emission.

The sensitivity analysis identified several influential data, but mean emissions in clover crops never exceeded 1.18 kg N ha^{-1} year^{-1}, whereas alfalfa emissions were never lower than 1.8 kg N ha^{-1} year^{-1}; mean emissions were thus consistently higher in alfalfa crops located in North America than in clover crops located in Europe and New Zealand.

4.   Discussion

It is useful to compare our quality assessment with similar quality assessments carried out in the past for meta-analyses in the fields of ecology and medical science. In our discussion of the quality of meta-analyses in agronomy we consider two previous studies for this purpose: Roberts et al. (2006), comparing the characteristics of 73 meta-analyses in ecology with those of 73 meta-analyses in medical science (*i.e.*, the same total number of meta-analyses as the number considered in our study) and Gates (2002), analyzing 29 meta-analyses in ecology.

The results of these previous studies show that criterion 1 (description of the procedure used for literature search) is more frequently satisfied in medical science (meta-analyses described in 2006: 100%) than in ecology (meta-analyses described in 2002: 66%, p value < 0.05; meta-analyses described in 2006: 0%, p value < 0.05) and in agronomy (meta-analysis described

here: 22%, p value < 0.05). The references of individual studies were presented in details in all fields (medical science: 100%; ecology, 2002: 93%; ecology, 2006: 100% and agronomy: 92%). However, the existence of a possible publication bias was rarely investigated in any of the fields considered (medical science: 23%; ecology, 2002: 34%; ecology, 2006: 8% and agronomy: 16%) despite the development of several methods for dealing with this issue (*e.g.*, funnel plot). Sensitivity analyses were more frequently carried out in medical science (32%) than in ecology (2006: 0%, p value < 0.05) or agronomy (8%, p value < 0.05). Sensitivity analysis was performed in less than one third of the reviewed meta-analyses in all fields, despite the work of Oxman (1994) stressing the importance of sensitivity analysis for systematic reviews. According to Oxman, sensitivity analysis is a useful method for –testing how robust the results of a review are relative to key decisions and assumptions that were made in the process of conducting the review".

The heterogeneity of individual results was more frequently studied, in all fields (medical science: 49%; ecology, 2002: 76%; ecology, 2006: 16%; agronomy: 88%). Sources of heterogeneity were analyzed almost systematically in agronomy, probably because it is important for agronomists to identify the crop management techniques and environmental characteristics yielding the best performances. As shown above, most of the analyses carried out in agronomy did in fact explain crop production variations according to crop management techniques.

Overall, more quality criteria were satisfied in past meta-analyses in medical science than in agronomy, with one exception: the criterion concerning the heterogeneity of the individual studies. The meta-analyses performed in ecology and in agronomy were more similar in terms of quality. None of the 73 meta-analyses in agronomy satisfied all eight criteria and only three satisfied six of these criteria.

The higher quality of meta-analyses in medical science is probably due to the intensive research conducted on meta-analysis in this area. In 1997, the Cochrane Collaboration (van Tulder et al., 1997 updated in 2003; van Tulder et al., 2003; Furlan et al., 2009), introduced methodological guidelines for systematic reviews, suggesting a framework to help researchers with quantitative synthesis. Other groups, such as the QUORUM (Quality of reporting meta-analyses) group, have provided advice about ways of carrying out high-quality meta-analyses. This group, consisting of 30 epidemiologists, clinicians, statisticians, editors and researchers, produced a checklist of standard items to be included in meta-analyses (Moher et al., 1999).

An important difference between medical and agronomical meta-analyses is due to the fact that, in medicine, results are almost entirely constituted of Randomized Control Trials including a control treatment (Begg et al., 1996; Gates, 2002) whereas in agronomy experiment designs are more diverse and do not systematically include a control treatment.

The between site-year variability of the agronomical response variable is usually high, and it is recommended to analyze this variability using a mixed-effect model as shown in our case study. However, results obtained with such models should be interpreted with care due to the possible effect of confounding variables. For example, in our case study, we found that emissions were significantly higher in alfalfa than in clover, but this difference may be as well due to a geographical effect because all the alfalfa experiments were located in North America and all the clover experiments were carried out in Europe and New Zealand.

Another interesting finding of the study was that Bayesian methods are entirely absent from the meta-analyses carried out in agronomy, despite the implementation of such methods in meta-analyses in medical science since the early 1990s (Smith et al., 1995; Sutton and Abrams, 2001) and more recently in ecology (Myers, 2001; Stewart, 2010). However, frequentist and Bayesian statistical methods were compared in our case study and the two types of methods were found to give similar results. The conclusions were thus insensitive to the type of statistical method used for parameter estimation. As this comparative study is the first one performed with agronomic data, further studies are required to determine the value of Bayesian statistics in meta-analysis.

Based on our quality assessment, we were able to formulate several recommendations for improving future meta-analyses in agronomy:
- The procedure used to search for papers in scientific databases should be presented.
- Individual data should be weighted according to the level of precision (like in ecology, as mentioned by Stewart (2010)). A typical approach involves weighting data by the inverse of the standard error of the measurements. In our case study, we used the number of days of the experiment, because standard error values were not available.
- The heterogeneity of data in individual studies should be analyzed with random-effect models (Madden and Paul, 2011). These models are useful for analyzing the between site-year variability of the response variables and for identifying relevant explanatory variables (*i.e.*, crop effect in our case study).
- The sensitivity of the estimated values to i) the statistical method (frequentist vs. Bayesian), explanatory variables and residual error distribution and ii) individual data should be analyzed. Sensitivity analysis is useful for assessing the robustness of the main conclusions of a meta-analysis.
- Efforts should be made to check for the publication bias.
- The design generated by merging data from different studies should be checked for confounding effects (*e.g.* continent effect in our case study).

5.   Conclusions

Finally, the 73 meta-analyses reviewed in this paper focused largely on production aspects. The response variable was crop production in most cases. Our case study showed that agronomical studies would benefit from meta-analysis techniques, because these techniques would make it possible to study the between site-year variability of the response variable (N_2O emission in our case study), to assess uncertainty and to identify the factors potentially influencing environmental impacts. The quality criteria and recommendations presented here could serve as a guide to improve future meta-analyses made in this area.

Chapitre 2

Quantifier les incertitudes relatives aux émissions de N$_2$O due à la fertilisation azotée des zones cultivées

Quantifying uncertainties in N_2O emission due to N fertilizer application in cultivated areas

Aurore Philibert[b,a], Chantal Loyce[b,a], David Makowski[a,b]

[a] INRA, UMR 211 Agronomie, F-78000 Thiverval Grignon, France
[b] AgroParisTech, UMR 211 Agronomie, F-78000 Thiverval Grignon, France

Abstract

Nitrous oxide (N_2O) is a greenhouse gas with a global warming potential approximately 298 times greater than that of CO_2. In 2006, the Intergovernmental Panel on Climate Change (IPCC) estimated N_2O emission due to synthetic and organic nitrogen (N) fertilization at 1% of applied N.

We investigated the uncertainty on this estimated value, by fitting 13 different models to a published dataset including 985 N_2O measurements. These models were characterized by (i) the presence or absence of the explanatory variable ―applied N", (ii) the function relating N_2O emission to applied N (exponential or linear function), (iii) fixed or random background (*i.e.* in the absence of N application) N_2O emission and (iv) fixed or random applied N effect. We calculated ranges of uncertainty on N_2O emissions from a subset of these models, and compared them with the uncertainty ranges currently used in the IPCC-Tier 1 method.

The exponential models outperformed the linear models, and models including one or two random effects outperformed those including fixed effects only. The use of an exponential function rather than a linear function has an important practical consequence: the emission factor is not constant and increases as a function of applied N. Emission factors estimated using the exponential function were lower than 1% when the amount of N applied was below 160 kg N ha^{-1}. Our uncertainty analysis shows that the uncertainty range currently used by the IPCC-Tier 1 method could be reduced.

Keywords: nitrous oxide, IPCC-Tier 1, N fertilization, mixed model, uncertainty analysis

1.   Introduction

Nitrous oxide (N$_2$O) is a greenhouse gas (GHG) with a global warming potential approximately 298 times greater than that of CO$_2$ (IPCC, 2007). N$_2$O emissions increased by almost 17% from 1990 to 2005 (Smith et al., 2007). The nitrogen (N) cycle is complex and N$_2$O emissions are determined by many factors (Galloway et al., 2004). Natural and anthropogenic N$_2$O is emitted as a result of nitrification (oxidation of ammonia) and denitrification (nitrate reduction), and these processes are influenced by applications of mineral N fertilizer and manure to agricultural soils (IPCC, 2001; Mosier et al., 1998). N applications are recognized as the major source of anthropogenic nitrous oxide emission (Davidson, 2009; Snyder et al., 2009). N$_2$O emissions are also influenced by other management practices (*e.g.*, tillage (Rochette, 2008)), soil and climate characteristics (*e.g.* soil water content) (Leip et al., 2011; Lesschen et al., 2011; Rochette et al., 2008; Rochette et al., 2010; Stehfest and Bouwman, 2006).

For countries unable to provide local statistics, N$_2$O emission can be estimated by the IPCC-Tier 1 method. In this approach, direct N$_2$O emission from N inputs is calculated as $Y_{N\ inputs} = (F_{SN} + F_{ON} + F_{CR} + F_{SOM})*EF_1 + (F_{SN} + F_{ON} + F_{CR} + F_{SOM})_{FR}*EF_{1FR}$, where F_{SN} is the annual amount of synthetic N fertilizer applied to soils, F_{ON} is the annual amount of organic N applied to soils, F_{CR} is the annual amount of N in crop residues, F_{SOM} is the annual amount of N in mineral soils, EF_1 is the emission factor for N$_2$O emissions from N inputs and FR indicates that the value concerned is for flooded rice (IPCC, 2006). For all crops other than flooded rice, the relationship between N$_2$O emission from N fertilizer and the dose of N applied can be expressed as $Y = EF*X$, where Y represents N$_2$O emissions due solely to N fertilization, X is the amount of synthetic and organic N applied and EF (emission factor) is the amount of N$_2$O emitted per unit of applied N. In the United Nations Framework Convention on Climate Change (Lokupitiya and Paustian, 2006), 56% of developed countries reported using the Tier 1 method of the IPCC to estimate N$_2$O emission from agricultural soils in 2006, and half the published N$_2$O emission inventories are based on this approach (Berdanier and Conant, 2012).

The EF value of 1.25%, set in 1999 (IPCC, 1999), was calculated from the following linear regression: $Y = 0.0125 * X$, where Y is the emission rate (in kg N$_2$O-N ha^{-1} yr^{-1}) and X is the fertilizer application rate (in kg N ha^{-1} yr^{-1}), based on 20 experiments (Bouwman, 1996). A background emission of 1 kg N$_2$O-N ha^{-1} yr^{-1} (*i.e.*, emission for X=0) was obtained in five experiments. The new value of EF used by the IPCC after 2006 (1%; IPCC, 2006) was estimated from a larger dataset, including N$_2$O emission measurements from studies on both crops and grassland (Stehfest and Bouwman, 2006).

Several recent studies have improved the estimation of N$_2$O emission further. Process-based models, such as the DNDC model (Leip et al., 2011) have been used to calculate the N$_2$O

emission factor as a function of the organic carbon content of the soil, fertilizer type and weather conditions, and the DAYCENT model (Del Grosso et al., 2009) has been used to calculate N_2O emissions as a function of soil class, daily weather, historical vegetation cover and land management practices, such as the type of crop grown, fertilizer additions and cultivation events. As these models describe the nitrogen cycle in detail, they may require long computation times and many input variables and are therefore difficult to implement (Roelandt et al., 2005). Various statistical models have also recently been proposed for the estimation of N_2O emission from global datasets. For example, linear regression models have been used (Freibauer and Kaltschmitt, 2003; Roelandt et al., 2005) and a nonlinear model based on an exponential function was proposed in another study (Lesschen et al., 2011).

The IPCC-Tier 1 method used for the estimation of N_2O emissions due to N fertilization includes three main sources of uncertainty on N_2O emission: (i) the uncertainty concerning the equation relating N_2O emission to applied N, (ii) the uncertainty concerning the equation parameters and (iii) the uncertainty about the amount of applied N (X).

In the IPCC-Tier 1 method, N_2O emission is assumed to be linearly related to applied N, but this assumption has been challenged; some authors (Hoben et al., 2011; McSwiney and Robertson, 2005) suggest that N_2O emission may instead increase exponentially as a function of applied N, and an exponential relationship between Y and X was also considered in the N_2O mitigation protocol proposed by Millar et al., (2010). A nonlinear relationship between Y and X was also considered by Stehfest and Bouwman (2006). There is currently no consensus concerning the most appropriate function for describing the relationship between N_2O emission and applied N at the global scale.

Uncertainty about the true value of the model parameter EF is another source of concern, for two reasons. First, N_2O emission measurements are known to be highly variable, both within a given site-year and between site-years. For a given site-year, N_2O emission varies principally due to climatic conditions, such as variations in the timing and intensity of rainfall, which modify microbial activity and the rates of gaseous emission (Skiba and Smith, 2000). For example, N_2O emissions must be measured after a period of rain to detect peaks in emission. Many factors may be responsible for variability between site-years, including differences in management practices (*e.g.* type of N fertilizer), soil characteristics and weather conditions between sites and years (Roelandt et al., 2005), by modifying chemical exchanges in agricultural soils. Duration of the experiment (Stehfest and Bouwman, 2006) and method used to measure emissions (Rochette et al., 2008) may also affect N_2O emission measurements. Second, the emission factor can be estimated by several different statistical methods, some based on fixed-parameter models (*i.e.*, classical regression) and others based on mixed-effect models or Bayesian methods. The sensitivity of EF to the statistical method used for its estimation has never been evaluated.

Table 2.1 : Minimal, maximal, median and mean values of nitrous oxide (N₂O) and amount of applied N (N rate) for the world and for North America, South America, Asia, Europe and Oceania.

Variable	Continent / world	min	median	mean	max	Number of data
	World	0.003	1.07	2.4	46.44	985
	Asia	0.01	0.53	1.11	15.60	124
N_2O	Europe	0.004	1.25	2.53	31.73	453
$(kg\ N\ ha^{-1}\ yr^{-1})$	North America	0.004	0.93	2.16	26.9	306
	Oceania	0.016	1.39	2.45	15	26
	South America	0.003	1.56	4.67	46.44	76
	World	0	100	124	500	985
	Asia	0	120	139.8	423	124
N rate	Europe	0	100	132	500	453
$(kg\ N\ ha^{-1})$	North America	0	92	115.3	450	306
	Oceania	0	66	108.6	500	26
	South America	0	0	90.96	360	76

Finally, the amounts of N applied can be estimated from regional and national statistics and from interviews with farmers (FAO, 2011; Spiertz, 2010; Stehfest and Bouwman, 2006), but the actual amounts of N applied are not perfectly known and vary from year to year.

In this study, we focused on the first two of these sources of uncertainty: the equation of the model and the values of the model parameters. We fitted 13 different models to the dataset of Stehfest and Bouwman (2006), and calculated uncertainty ranges on average N$_2$O emissions from a subset of these models, comparing our ranges with those currently used by the IPCC.

2. Materials and Methods

2.1. Database

The dataset is a global compilation of nitrous oxide (N$_2$O) and nitric oxide (NO) emissions extracted from peer-reviewed publications appearing between 1979 and 2004, established by Stehfest and Bouwman (2006). Readers should refer to the original paper by Stehfest and Bouwman for a more complete presentation of the data.

The dataset (available from http://www.pbl.nl/en/publications/2006/N2OAndNOEmissionFromAgriculturalFieldsAndSoilsUnderNaturalVegetation) includes 1891 measurements of N$_2$O and NO emissions in natural and agricultural fields from 387 publications. As we focused on calculation of the emission factor associated with fertilizer applications in agricultural fields (EF), we excluded the following experiments from the initial dataset: (i) 418 experiments carried out in natural areas, (ii) 360 experiments including measurements of NO emission only, (iii) 57 experiments on organic soils (not concerned by EF), (iv) 25 experiments including the use of chemicals or additives considered to inhibit nitrification (also excluded by Stehfest and Bouwman (2006)), (v) 8 experiments in grazing systems (also excluded by Stehfest and Bouwman (2006)), (vi) 38 experiments in which the amounts of applied N exceeded 500 kg N ha^{-1} yr^{-1} (given that the maximum amounts of N applied to agricultural fields has been estimated at 400 kg N ha^{-1} (Roelandt et al., 2005; Spiertz 2010; Tilman et al., 2002)).

We finally worked with a dataset including 985 measurements of N$_2$O emission in agricultural fields extracted from 203 publications, corresponding to a set of experiments encompassing various soil and climatic characteristics and types of fertilization (Fig. 2.1 and 2.2).

The distribution of N$_2$O measurements and amounts of applied N are presented in Table 2.1 for the entire dataset and for each continent separately. The largest amount of data was available for the temperate-continental climate (460), followed by the temperate-oceanic

Table 2.2 : Characteristics of the 13 statistical models for N$_2$O emission. Models were characterized by their response function (linear or exponential), the use of the explanatory variable _amount of applied N', the use of random effects for the intercept and/or the effect of the amount of N applied, values of the Akaïke and Schwartz criteria (AIC and BIC), and of the deviance information criterion (DIC) for Bayesian models. % AIC and % BIC indicate the percentage increase in AIC and BIC with respect to the best linear and nonlinear models.

Model name	Linear	Amount of N applied	Intercept	Effect of the amount of N applied	AIC	% AIC	BIC	% BIC	DIC
NL-N-FF	No	Yes	Fixed	Fixed	5513.1	23.0	5527.8	22.6	-
NL-0-R	No	No	Random	-	5091.9	13.6	5106.5	13.3	-
NL-N-RF	No	Yes	Random	Fixed	4553.9	1.6	4573.5	1.5	-
NL-N-FR	No	Yes	Fixed	Random	4598.9	2.6	4618.5	2.5	-
NL-N-RR	No	Yes	Random	Random	4482.7	0	4507.1	0	-
NL-N-RR-B	No	Yes	Random	Random	-	-	-	-	4196.71
L-0-F	Yes	No	Fixed	-	5653.9	20.5	5663.7	20.1	-
L-N-FF	Yes	Yes	Fixed	Fixed	5512.1	17.4	5526.8	17.2	-
L-0-R	Yes	No	Random	-	5268.5	12.3	5283.2	12.0	-
L-N-RF	Yes	Yes	Random	Fixed	5117.4	9.0	5136.9	8.9	-
L-N-FR	Yes	Yes	Fixed	Random	4698.0	0.1	4717.5	0	-
L-N-RR	Yes	Yes	Random	Random	4693.2	0	4717.6	0.002	-
L-N-RR-B	Yes	Yes	Random	Random	-	-	-	-	4421.63

climate (258) and the tropics-warm humid climate (104). Only 80, 44, 21, 12 and 6 data were collected for the subtropical-summer rains, subtropical-winter rains, tropic-seas dry, boreal and cool tropics climates, respectively.

2.2. Statistical analysis

2.2.1. Statistical models

Thirteen models relating N₂O emission to the amount of applied N were fitted to the data (Table 2.2). These models were characterized by (i) the presence or absence of the explanatory variable ―applied N" (X), (ii) the function relating emission to applied N (an exponential or linear function), (iii) fixed or random background emission (*i.e.*, emission for X=0), and (iv) fixed or random N effect.

The first 11 models (with L, NL, N, 0, F, and R standing for linear, nonlinear, nitrogen effect, no nitrogen effect, fixed parameter and random parameter, respectively) can be expressed as:

Model NL-N-FF: (1) $Y_{ijk} = \exp(\mu_0 + \mu_1 X_{ij}) + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$

Model NL-0-R: (2) $Y_{ijk} = \exp(\alpha_{0i}) + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$

Model NL-N-RF: (3) $Y_{ijk} = \exp(\alpha_{0i} + \mu_1 X_{ij}) + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$

Model NL-N-FR: (4) $Y_{ijk} = \exp(\mu_0 + \alpha_{1i} X_{ij}) + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{1i} \sim N(\mu_1, \sigma_1^2)$

Model NL-N-RR: (5) $Y_{ijk} = \exp(\alpha_{0i} + \alpha_{1i} X_{ij}) + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$, $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$ and $\alpha_{1i} \sim N(\mu_1, \sigma_1^2)$

Model L-0-F: (6) $Y_{ijk} = \mu_0 + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$

Model L-N-FF: (7) $Y_{ijk} = \mu_0 + \mu_1 X_{ij} + \varepsilon_{ijk}$

with $\varepsilon_{ijk} \sim N(0, \tau^2)$

Model L-0-R: (8) $Y_{ijk} = \alpha_{0i} + \varepsilon_{ijk}$

                    with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$

Model L-N-RF: (9) $Y_{ijk} = \alpha_{0i} + \mu_1 X_{ij} + \varepsilon_{ijk}$

                    with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$

Model L-N-FR: (10) $Y_{ijk} = \mu_0 + \alpha_{1i} X_{ij} + \varepsilon_{ijk}$

                    with $\varepsilon_{ijk} \sim N(0, \tau^2)$ and $\alpha_{1i} \sim N(\mu_1, \sigma_1^2)$

Model L-N-RR: (11) $Y_{ijk} = \alpha_{0i} + \alpha_{1i} X_{ij} + \varepsilon_{ijk}$

                    with $\varepsilon_{ijk} \sim N(0, \tau^2)$, $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$ and $\alpha_{1i} \sim N(\mu_1, \sigma_1^2)$

where Y_{ijk} is the N₂O emission (kg N ha⁻¹ yr⁻¹) measured in the ith published experiment (i=1 ... 203), the jth applied N dose (j=1 ... N_i), and the kth replicate (k=1 ... K_{ij}), X_{ij} is the jth applied N dose (kg N ha⁻¹) in the ith published experiment, μ_0 is the mean background emission, α_{0i} is the published experiment-specific background emission (random), μ_1 is the mean applied N effect, α_{1i} is the published experiment-specific applied N effect (random), and ε_{ijk} is the residual error term. The random terms α_{0i}, α_{1i} and ε_{ijk} were assumed to be independent and normally distributed. Models including correlated α_{0i} and α_{1i} were also fitted to the data but, as their outputs were very similar to the outputs of the models with independent random parameters, they were not considered further. Note that, in nonlinear models (1-5), the N₂O response does not follow a normal distribution, even if its parameters α_{0i} and α_{1i} do, due to the use of an exponential function to relate emissions to model parameters.

In the linear models (6-11), the parameter μ_1 corresponds to the emission factor EF used by the IPCC. In the nonlinear models based on an exponential function (1-5), N₂O emission per unit of applied N is not constant; instead, it increases as a function of X if μ_1 is positive. In the models including one or two random parameters (2-5 and 8-11), the response of N₂O to the amount of applied N is assumed to follow the same function (linear or exponential) in all experiments, but the parameters of these models (background emission α_{0i}, effect of applied N α_{1i}, or both) were assumed to vary between experiments. Distributions of α_{0i} and α_{1i} describe the between-experiment variability of background emission and N fertilizer effect. An intercept was included in all statistical models to account for background anthropogenic N₂O emission (Bouwman, 1996). The values of the $\mu_0, \mu_1, \sigma_0, \sigma_1$ and τ parameters of models 1-

11 were estimated by an approximate maximum likelihood method, with the nlme R statistical package (Pinheiro and Bates, 2000).

Two additional models, NL-N-RR-B and L-N-RR-B, were defined. These models were based on the equations of models NL-N-RR and L-N-RR, respectively, but their parameters were estimated by a Bayesian method implemented with a Markov chain Monte Carlo algorithm (MCMC). Normal and independent prior probability distributions were defined for μ_0 and μ_1; μ_0, $\mu_1 \sim N(0,1000)$. Uniform and independent prior probability distributions were defined for τ, σ_0, σ_1; τ, σ_0, $\sigma_1 \sim U(0,100)$. Under these assumptions, μ_0 and μ_1 had a prior mean of zero and a prior standard deviation of 32, which is quite large given the measured values, which ranged from 0.003 to 46.44 in our dataset. These distributions represent a broad *a priori* distribution with respect to the data obtained. For example, the 95% credibility interval derived from the prior distributions ranged from -6272.3 to 6331.8 N₂O kg N ha^{-1} yr^{-1} for X=100 kg N ha^{-1}. Posterior distributions of the parameters of models NL-N-RR-B and L-N-RR-B were calculated with WinBUGS software (Lunn et al., 2000), with three chains of 100,000 MCMC iterations. Convergence was checked with the Gelman-Rubin method (Brooks and Gelman, 1998).

2.2.2. Model assessment and uncertainty analysis

The Akaike information criterion (AIC) and the Schwartz criterion (BIC) (Akaike, 1974; Burnham and Anderson, 2002) were calculated for the first 11 models, and the deviance information criterion (DIC) (Spiegelhalter et al., 2002) was calculated for the two Bayesian models. Lower values of AIC, BIC or DIC are considered to indicate better models. Note that the weighting of the experiments according to their lengths did not reduce AIC, BIC or DIC.

We calculated the 95% confidence intervals for each model by a bootstrap method (Efron and Tibshirani, 1986; 1993); data were sampled, with replacement, 500 times, and each model was fitted to each of the generated samples. For the two Bayesian models, 95% credibility intervals for the predicted N₂O emissions were calculated from the parameter values generated by the MCMC algorithm.

The predictions generated by the three best non-Bayesian linear models, the three best non-Bayesian exponential models (selected with AIC and BIC criteria) and the two Bayesian models were compared with the N₂O emissions calculated by the IPCC-Tier1 method: Y = EF*X, where EF is taken as 0.01 (IPCC, 2006). The range of uncertainty on predicted N₂O emissions for the IPCC method was calculated from the minimum and maximum values of EF (0.003 and 0.03, respectively) reported by the IPCC (IPCC, 2006). The emissions due to applied N calculated with the IPCC method were compared with the predictions of the eight selected models minus the values predicted at X=0.

Table 2.3 : Estimated values of the parameters of the 13 models. The standard deviations of the estimators of μ_0 and μ_1 are indicated in brackets.

Model name	μ_0	σ_0	μ_1	σ_1	τ
NL-N-FF	0.25 (0.096)	-	0.0042 (0.0003)	-	3.96
NL-0-R	0.87 (0.077)	0.84	-	-	2.84
NL-N-RF	-0.068 (0.092)	0.83	0.0050 (0.0003)	-	2.09
NL-N-FR	0.31 (0.068)	-	0.0033 (0.0005)	0.0043	2.13
NL-N-RR	0.19 (0.09)	0.72	0.0037 (0.0004)	0.0025	1.94
NL-N-RR-B	-0.21 (0.13)	0.92	0.0038 (0.0005)	0.0032	1.91
L-0-F	2.40 (0.14)	-	-	-	4.26
L-N-FF	0.69 (0.19)	-	0.0138 (0.0011)	-	3.96
L-0-R	2.78 (0.27)	3.44	-	-	2.91
L-N-RF	0.99 (0.28)	3.16	0.0130 (0.0010)	-	2.67
L-N-FR	1.09 (0.11)	-	0.0113 (0.0017)	0.0195	2.12
L-N-RR	1.04 (0.13)	0.70	0.0117 (0.0017)	0.0187	2.08
L-N-RR-B	1.04 (0.14)	0.76	0.0117 (0.0017)	0.0189	2.08

This uncertainty range was then compared with each of the confidence intervals for the eight selected models. We also compared the lower limit of the IPCC uncertainty range with the lowest of the eight 2.5 percentiles calculated for the eight selected models, and the upper limit of the IPCC uncertainty range with the highest of the eight 97.5 percentiles of the eight selected models. The most extreme 2.5 and 97.5 percentiles obtained with the eight selected models can be interpreted as best-case and worst-case emission scenarios, respectively. They correspond to the lowest and highest limits of the confidence intervals calculated for the eight models.

The code used for statistical analysis is available, on request, from the corresponding author (see Annex 2 and 3 for a part of the Winbugs and R code).

3.   Results

3.1. Parameter values

The estimated value of parameter μ_0 (mean background emission) ranged from -0.21 to 0.88 for nonlinear models and from 0.69 to 2.78 for linear models (Table 2.3). The between-model variability of the estimated values of μ_1 (mean applied N effect) was small: estimated values ranged from 0.0033 to 0.0050 for nonlinear models and from 0.0113 to 0.0138 for linear models. For both linear and nonlinear models, the estimated values of μ_1 were lower when the effect of applied N was considered a random effect (Table 2.3). For example, the estimated value of μ_1 was 0.005 for NL-N-RF, but only 0.0037 for NL-N-RR.

μ_0 was less accurately estimated than μ_1; the coefficient of variation (standard deviation/estimated value) was lower for μ_1 than for μ_0. For a given type of function (linear or exponential) estimates of σ_0 and σ_1 (between-experiment standard deviation of background emission and applied N effects, respectively) were similar between models. The estimated values of τ (standard deviation of model residuals) were lower for models with random parameters and for those containing the explanatory variable X.

3.2. Model selection

The lowest AIC and BIC values were obtained for the nonlinear model including two random effects (NL-N-RR) (Table 2.2). Thus, models based on an exponential function outperformed models based on a linear function. This result was confirmed by the DIC values obtained for the two Bayesian models: DIC was lower with the exponential function.

AIC and BIC values were much higher in models in which applied N (X) was not included as an explanatory variable. The AIC and BIC values of the NL-0-R model were 5091.9 and
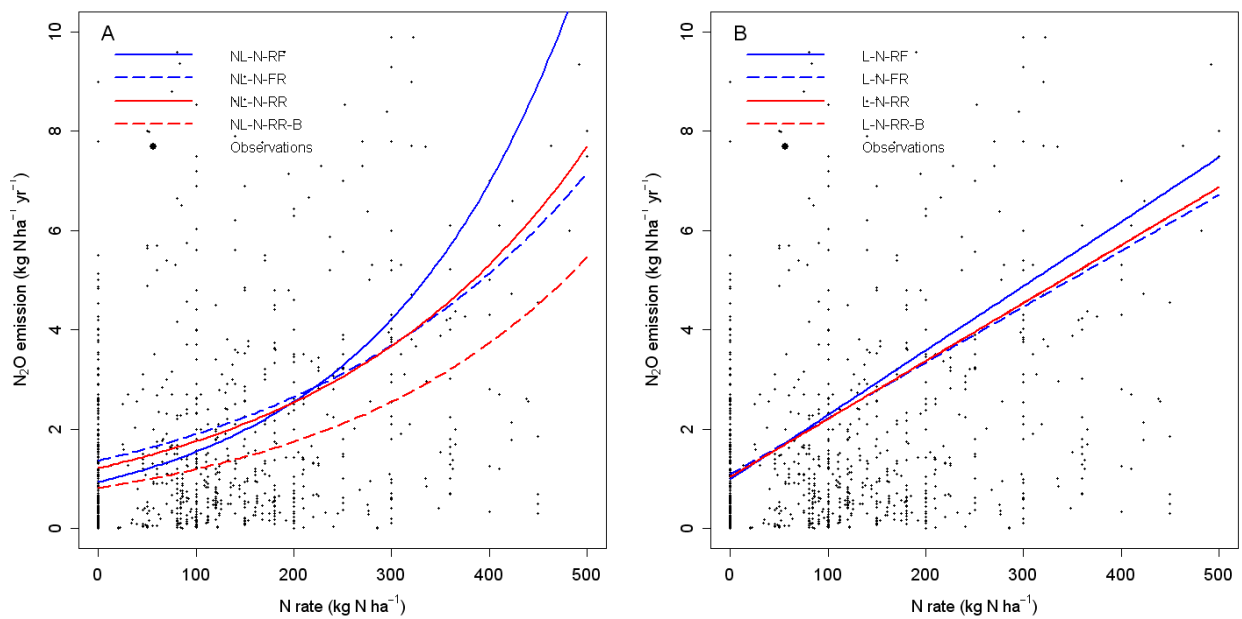
Figure 2.1 : Fitted response curves obtained with the four selected nonlinear models (A) and the four selected linear models (B). Black points correspond to N₂O data (96.04% of available observations are displayed; the other data are too extreme for graphical presentation).
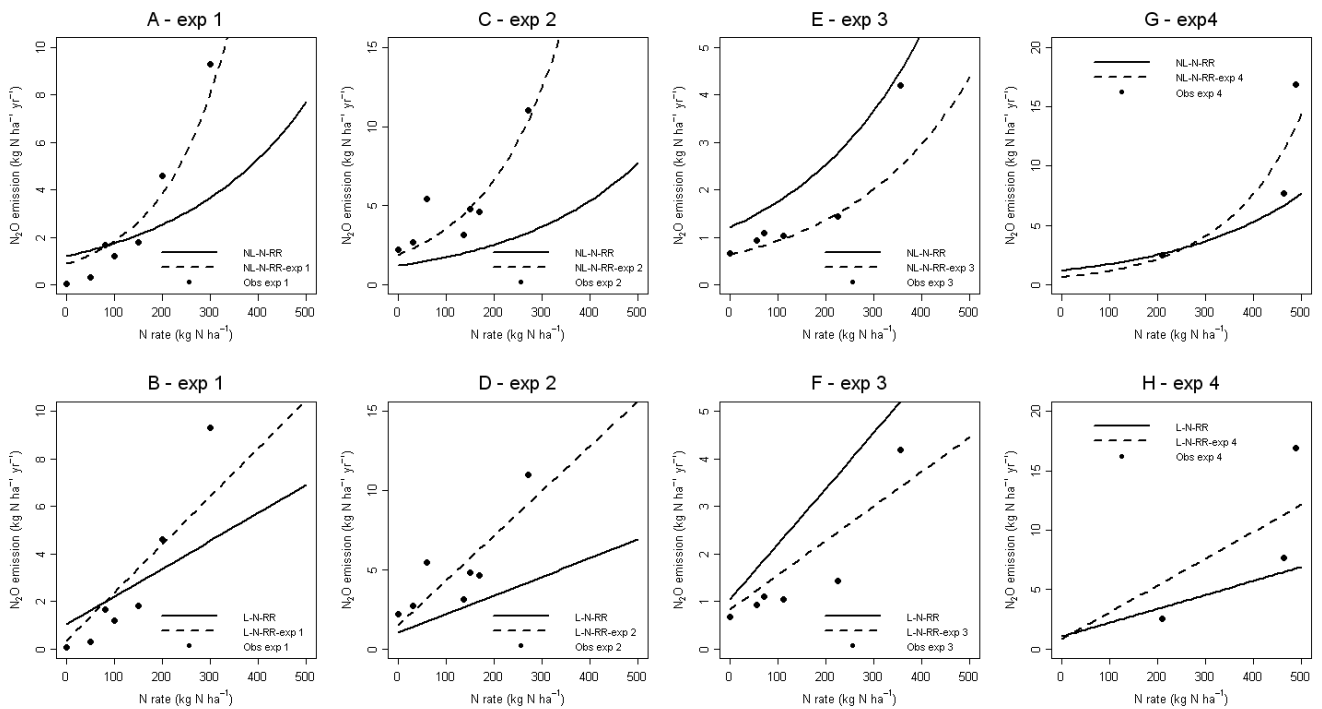
Figure 2.2 : Fitted response curves for four experiments (exp 1: (A-B), exp 2: (C-D), exp 3: (E-F) and exp 4: (G-H)). For each published-experiment, mean response (solid black line) and experiment-specific response (dotted black line) were calculated with model NL-N-RR (A, C, E, G) and model L-N-RR (B, D, F, H). Black points represent N₂O data averaged over replicates.

5106.5, respectively, whereas the AIC and BIC values of the NL-N-RF model were 4553.9 and 4573.5, respectively (Table 2.2).

Models including one or two random effects outperformed those including only fixed effects. The best linear model was L-N-RR on the basis of AIC, and L-N-FR, on the basis of BIC. The NL-N-FF (no random effect) model had an AIC of 5513.1 and a BIC of 5527.8, whereas both these values were much lower (AIC=4482.7 and BIC=4507.1) for the NL-N-RR (two random parameters) model. Models including one or two random effects had similar AIC and BIC values; the use of one random effect rather than two did not increase AIC and BIC by more than 2.6% and 9% for the nonlinear and linear models, respectively (Table 2.2).

The AIC and BIC values of models (1), (2), (6), (7) and (8) were more than 10% higher than those for the best nonlinear and linear models, and were therefore not considered further. We therefore considered only models (3), (4), (5), (9), (10) and (11) and the two Bayesian models in our subsequent estimations of N_2O emission.

3.3. Estimation of N₂O emissions with the selected models

Figure 2.1 shows N_2O emissions estimated with the eight selected models. The rate of increase of N_2O emissions with the amount of N applied was greater with the NL-N-RF model (Fig. 2.1A) than with the other two nonlinear models (NL-N-FR and NL-N-RR). The predicted increase in the amount of N_2O emitted per unit increase in the amount of applied N was thus lower when the effect of applied N was defined as a random effect. Similar results were obtained for linear models, for which the highest rate of increase in N_2O emissions with the amount of N applied was obtained for the L-N-RF model, which had a fixed slope (Fig. 2.1B). These results are consistent with the estimated parameter values reported in Table 2.3.

The emissions predicted by the Bayesian model NL-N-RR-B were the lowest for all values of applied N (Fig. 2.1A), due to the low estimated value of the intercept for this model (Table 2.3). The amounts of emission predicted by the L-N-RR model and its Bayesian counterpart (L-N-RR-B) were very similar and were essentially undistinguishable.

Figure 2.2 shows the fitted response curves obtained with the best linear and nonlinear models, NL-N-RR and L-N-RR, for four experiments. Considering experiment-specific responses, the nonlinear model better fitted the emissions measured at high N doses in experiments 1, 2, and 4, and the emissions measured at low N doses in experiments 3 and 4. Between-experiment variability was high for N_2O emissions (Figure 2.2) and could be accounted for by the experiment-effects included in the mixed-effect models. The residual standard error was lower with NL-N-RR than with L-N-RR (see values of τ in Table 2.3).
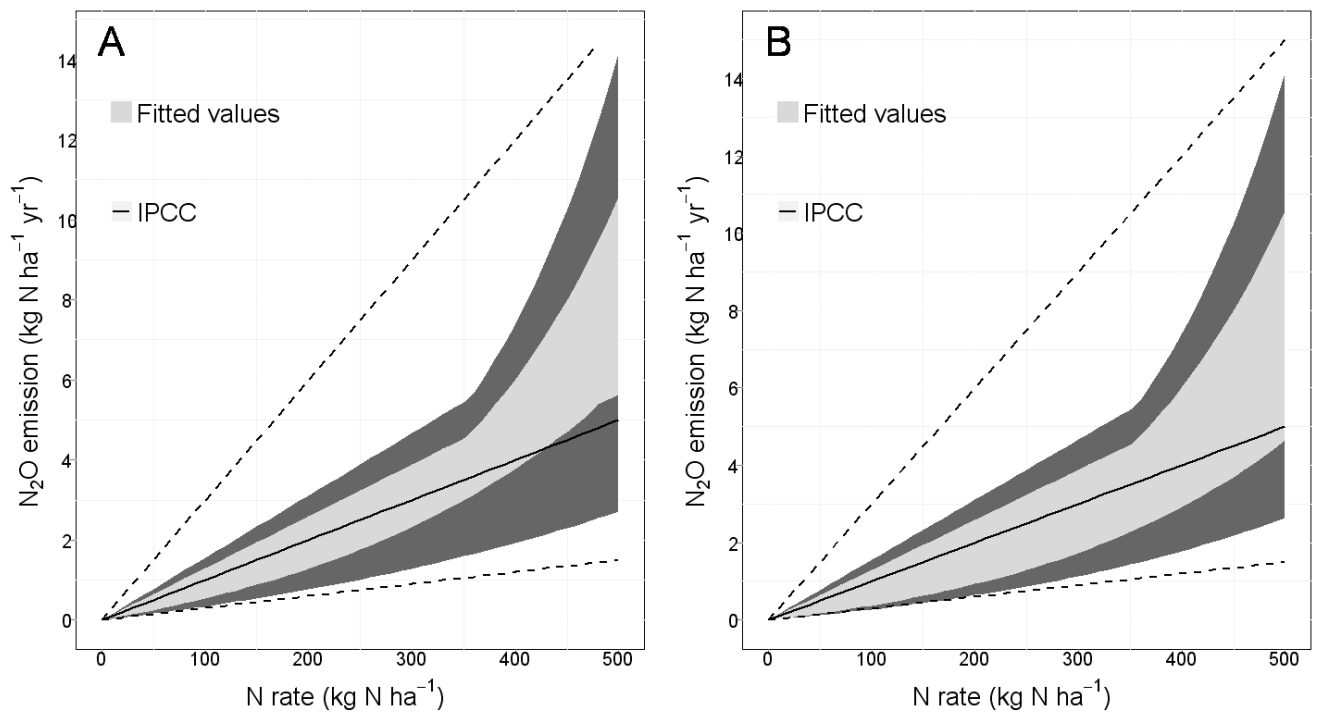
Figure 2.3: Predicted N$_2$O emissions and uncertainty ranges for our eight selected models and the IPCC-Tier 1 method. The light gray area represents the uncertainty in the model equations and includes the mean values predicted by six models (A) or by eight models (6 non-Bayesian models + 2 Bayesian models) (B). The dark gray area represents the uncertainty in model equations and parameter values. The upper and lower limits of the dark gray area indicate the worst-case and best-case scenarios, respectively, defined from six models (A) or from eight models (6 non-Bayesian models + 2 Bayesian models) (B). The solid black line and the dotted lines indicate the N$_2$O emissions predicted with an EF of 1% and the uncertainty range of the IPCC-Tier1 method, respectively.

3.4. Comparison with the emissions estimated with the IPCC-Tier 1 method

We determined the ranges of N_2O emissions (Fig. 2.3) covered by the eight models considered in Fig. 2.1, either taking into account the uncertainty on the estimated parameter values (dark gray area) or not taking this uncertainty into account (light gray area). The final values predicted by the models were calculated by subtracting the predicted value at X=0 (background emission) from the value actually predicted for a given amount of applied N. This graphical presentation made it possible to compare our models with the N_2O emissions predicted with the IPCC-Tier 1 method. The estimates of N_2O emission obtained with an emission factor of 1% (as used by the IPCC) were within the range of values covered by the eight selected models (Fig. 2.3B), but the range of uncertainty for emissions estimated with the IPCC-Tier 1 method was larger than that for the eight selected models. The upper limit of the uncertainty range for the IPCC method was much higher than that defined by the highest value of the eight 97.5 percentiles of the eight selected models, particularly for N applications below 300 kg ha^{-1}, as generally practiced in farmers' fields (Fig. 2.3). The lower limits of the uncertainty ranges for the IPCC method and for our models were more similar.

This result was confirmed (Fig. 2.4) by comparing the estimates of N_2O emissions due to applied N obtained with the eight models with those obtained by the IPCC-Tier 1 method for four different amounts of applied N. These amounts of applied N correspond to the average amounts applied in western, eastern and southern Africa, worldwide, Europe and eastern Asia (Stehfest and Bouwman, 2006). The purpose of Figure 2.4 was to compare model predictions for contrasted applied N doses, not to calculate average emissions at the continental scale. The uncertainty ranges obtained with the IPCC method were indeed larger than those defined by the highest of the 97.5 percentiles and the lowest of the 2.5 percentiles for the eight selected models (Fig. 2.4). The upper limits of the IPCC uncertainty ranges were systematically higher than the highest 97.5 percentile obtained with our models. We also found that the emissions predicted by the IPCC method were very similar to those obtained with the linear models, but systematically higher than the emissions predicted by the nonlinear models (Fig. 2.4).

4.   Discussion

Our analysis was carried out with the dataset of Stehfest and Bouwman (2006) because this dataset includes a large number of data of N$_2$O emissions in agricultural soils. These data were collected under various conditions characterized by different measurement methods (*e.g.* from short to long periods of measurements), different soils, climates, and crops. The variability of these conditions and their effects on N$_2$O emission were taken into account in our analysis using random parameter models. In these models, the N$_2$O emission was related to applied N using linear or nonlinear functions including two random parameters (α_{0i} and α_{1i}). The probability distribution of these parameters describes the between-study variability of the parameter values due to i) the heterogeneity of the experimental protocols (*i.e.* length of the experiment and measurement frequency) and ii) the variability of soil, climate, and crop characteristics. This approach accounts for both the heterogeneity of the measurement protocols and the variability of the environments.

Exponential models outperformed linear models for the three statistical criteria considered: AIC, BIC and DIC. However, the differences were small. The AIC of the best linear model was only 4.7% higher than the AIC of the best exponential model. The use of an exponential function rather than a linear model has an important practical consequence: EF is not constant and increases as a function of applied N.

Our results indicate that EF is lower than the estimated value used by the IPCC-Tier 1 (*i.e.,* 1% of applied N) if the amount of N applied is below 160 kg N ha^{-1} for the NL-N-RF model, 240 kg N ha^{-1} for the NL-N-FR model and 220 kg N ha^{-1} for the NL-N-RR model. According to Spiertz (2010), farmers may apply amounts of nitrogen fertilizer below these thresholds in ecological low-input cropping systems and in some technological high-input systems. Consequently, the use of an exponential model rather than a linear model is likely to decrease estimates of N$_2$O emissions in many cases. According to Hoben et al. (2011), the current IPCC-Tier 1 method could lead to an underestimation of N$_2$O emission if the true response is exponential. Our results suggest that this is the case only for the application of large amounts of N fertilizer.
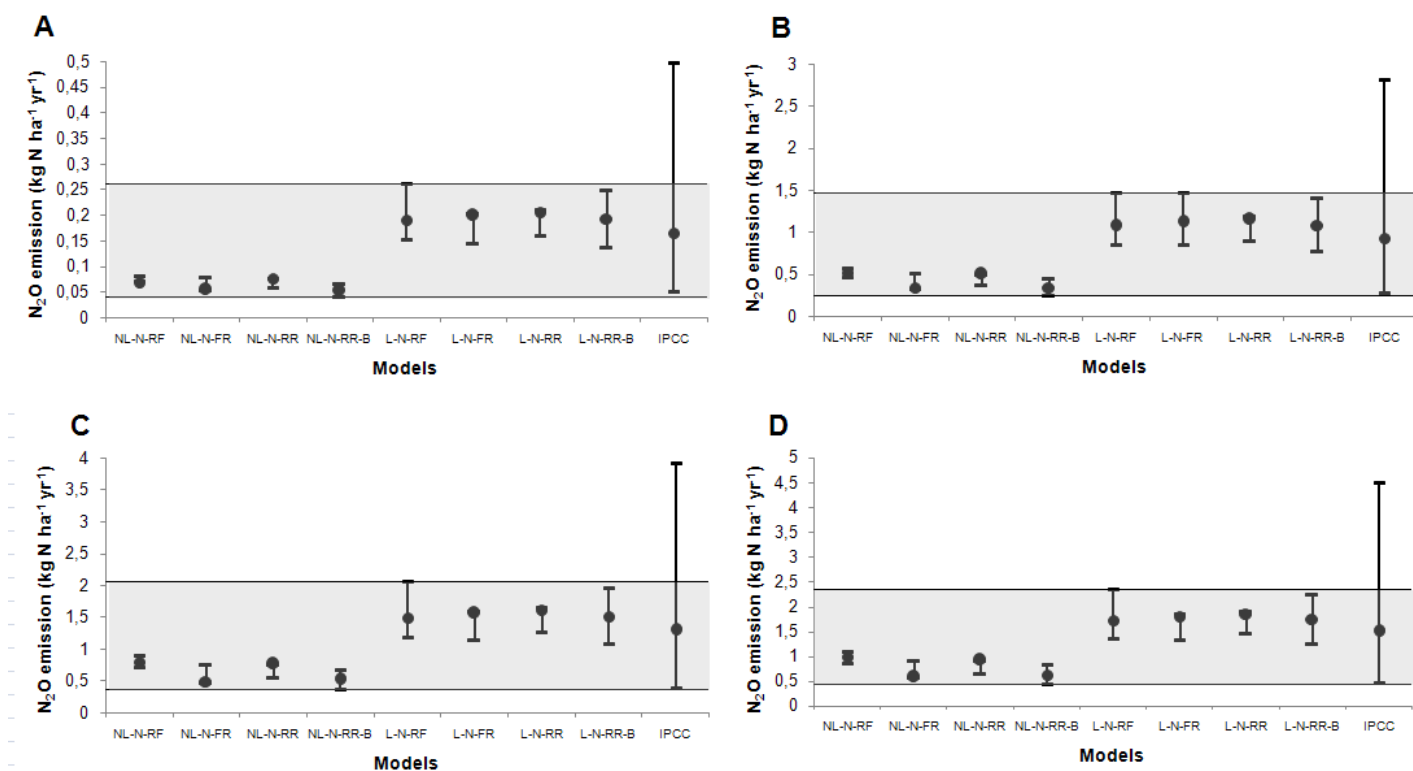
Figure 2.4: Predicted N$_2$O emissions due to N fertilization and 95% confidence intervals (CI) for each model, and predicted values and uncertainty ranges for the IPCC-Tier 1 method. The light gray area corresponds to the values covered by the 95% CI of our eight models. The amounts of N applied were A) 16.62 kg N ha^{-1}, B) 93.6 kg N ha^{-1}, C) 130.74 kg N ha^{-1}, D) 149.58 kg N ha^{-1} (average amounts of applied N for western, eastern and southern Africa, worldwide, Europe, and eastern Asia respectively). N$_2$O emissions were estimated by subtracting the value corresponding to the application of no N from the value for each amount of N applied.

McSwiney and Robertson (2005) suggested that the use of a nonlinear model instead of a linear model leads to a greater estimated reduction in N_2O emission for a moderate reduction in the amount of applied N, with little or no yield penalty. Our results do not entirely support this statement, because little difference was observed between the two types of models for doses of up to about 200 kg N ha^{-1} (Fig. 2.1). For example, if the amount of applied N is decreased from 150 kg N to 120 kg N (minus 20%), the resulting reduction of N_2O calculated with the NL-N-RR model is 0.22 kg N ha^{-1} yr^{-1}, slightly less than that calculated with the current IPCC emission factor (0.01*30=0.3 kg N ha^{-1} yr^{-1}). With the same model, the reduction induced by a decrease from 350 kg N ha^{-1} to 280 kg N ha^{-1} (minus 20%) is much larger, reaching 1 kg N ha^{-1} yr^{-1}; this value is higher than the reduction calculated with the IPCC emission factor (0.01*70=0.7 kg N ha^{-1} yr^{-1}). The estimated reduction of N_2O emission induced by a decrease in the amount of applied N is greater with the nonlinear model than with the linear model only for high N doses.

According to the AIC and BIC values obtained, models including one or two random effects outperformed models including fixed effects only. Mixed-effect models are commonly used in meta-analysis studies (Philibert et al., 2012a) and are recommended for the analysis of repeated measurements on the same individuals (Davidian and Giltinan, 1995). In the dataset of Stehfest and Bouwman (2006), N_2O emissions were measured for several amounts of N applied in the same published-experiment. It was therefore appropriate to estimate N_2O emissions with mixed-effect models including one or two random effects in our study (Fig. 2.2). Models including one or two random effects performed similarly (less than 10% difference in AIC and BIC values), but the estimated effect of the amount of N applied on the amount of N_2O emitted tended to be lower when the amount of N applied was considered as a random effect.

Several models had very similar performances. We therefore used an ensemble approach based on eight models for the estimation of N_2O emissions and the definition of uncertainty ranges. The confidence intervals obtained with the models were used to define lower and upper limits, corresponding to the best-case and worst-case scenarios, respectively. These confidence intervals represent the uncertainty in average N_2O emissions over all experiments, but they do not describe the between-experiment variability of N_2O emission. The range of uncertainty defined here is relevant for the Tier 1 method and useful for explorations of the consequences of N applications for average N_2O emissions, taking into account the uncertainty due to model equations and parameter estimations. The lower limit of our uncertainty range is close to that defined by the IPCC-Tier 1, although our lower limit is slightly higher than that of the IPCC for applications of large amounts of N. Our upper limit is much lower than the upper limit of the IPCC range, particularly for total N applications below 300 kg ha^{-1}, as commonly used in agriculture. Thus, the upper limit of the IPCC range gives

an estimated N$_2$O emission of 9 kg N ha^{-1} yr^{-1} for a dose of 300 kg N ha^{-1}, whereas the upper limit of our uncertainty range (*i.e.*, the highest upper limit of the confidence intervals of the eight models considered) gave an estimated emission value of only 4.7 kg N ha^{-1} yr^{-1}. This result is consistent with the findings of Leip et al. (2011), suggesting that the uncertainty on estimates of N$_2$O emissions was overestimated when derived from experimental data variances, which largely compensate at large scales.

It is useful to compare our uncertainty ranges with other ranges calculated with process-based models (Del Grosso et al., 2010), top-down methods (Crutzen et al., 2008) and hierarchical Bayesian models (Berdanier and Conant, 2012).

Our uncertainty range for the average N dose applied in North America — 0.49-1.88 kg N ha^{-1} yr^{-1} — is similar to the 95% confidence interval proposed by Del Grosso et al. (2010) for the United States (133-304 Gg N yr^{-1} *i.e.* 0.99-2.27 kg N ha^{-1} yr^{-1} with the cropland area of North America reported by Stehfest and Bouwman (2006)).

Our uncertainty range for the average N dose applied at the world scale (93.6 kg ha^{-1} of applied N, as reported by Stehfest and Bouwman, 2006) (Fig. 2.4B) — 0.25-1.48 kg N ha^{-1} yr^{-1} — is lower and narrower than the interval proposed by Crutzen et al. (2008) (2.8-4.68 kg N ha^{-1} yr^{-1}). However, it is difficult to compare these intervals, due to the use of a top-down method by Crutzen et al. (2008). Furthermore, these authors did not consider direct emission due to N fertilizer only, instead also taking into account indirect emissions from leaching and atmospheric deposition (Lesschen et al., 2011).

The 95% confidence interval calculated by Berdanier and Conant (2012) with a hierarchical Bayesian linear model is similar to our uncertainty ranges for the four regions of the world presented in Figure 2.4. The two intervals overlap in all four regions, but our intervals tend to have lower upper and lower limits. For example, Berdanier and Conant (2012) reported an interval of 0.05-0.46 for Africa, for a N fertilizer dose of 16.62 kg N ha^{-1} (Stehfest and Bouwman, 2006), whereas our interval was 0.04-0.26 kg N ha^{-1} yr^{-1} for the average N fertilizer dose reported for West, East and Southern Africa by Stehfest and Bouwman (2006).

When between-experiment variability was taken into account, the experiment-specific N$_2$O estimated with our models covered a wider range of values. Thus, for applied N levels of 100 kg N ha^{-1} and with the NL-N-RR model, the 90% percentile for N$_2$O emission was 1.79 kg N ha^{-1} yr^{-1}, the 95% percentile was 2.52 kg N ha^{-1} yr^{-1} and the 99% percentile was 5.03 kg N ha^{-1} yr^{-1}, all these values being higher than the 1 kg N ha^{-1} yr^{-1} of the IPCC-Tier 1 method. Thus, N$_2$O emission has 1% chance to exceed 5 kg N ha^{-1} yr^{-1} for an N fertilizer dose of 100 kg ha^{-1}.

The nonlinear models presented in this paper should be used with caution for estimating average N$_2$O emissions at the country and continental scales. The average output value of a

nonlinear model is not strictly equal to the output value obtained with the average input value. In order to calculate the average N$_2$O emission in a given country with a nonlinear model, the best approach is i) to determine the distribution of applied N fertilizer doses in this country, ii) to run the model for all doses, and iii) to take the average of all the model outputs. However, this approach requires the knowledge of the distribution of N fertilizer dose.

We focused on the Tier 1 approach of the IPCC, but the proposed exponential models could be extended to take several other environmental variables, such as climatic characteristics, soil types and fertilizer type, into account. This possibility has already been explored by Lesschen et al. (2011), who took several variables into account (type of fertilizer, crop residues, atmospheric deposition, land use, soil type and precipitation) and by Leip et al. (2011), who calculated the stratified emission factor as a function of soil organic carbon content, fertilizer type (mineral fertilizer or manure) and weather conditions. Such variables could be included in our models, for the estimation of region-specific N$_2$O emissions, taking local characteristics into account.

Chapitre 3

Prédire les émissions de N$_2$O à partir d'informations locales avec Random Forest

Prediction of N$_2$O emission from local information with Random Forest

Aurore Philibert[b,a], Chantal Loyce[b,a], David Makowski[a,b]

[a] INRA, UMR 211 Agronomie, F-78000 Thiverval Grignon, France
[b] AgroParisTech, UMR 211 Agronomie, F-78000 Thiverval Grignon, France

Abstract:

Nitrous oxide is a potent greenhouse gas, with a global warming potential 298 times greater than that of CO$_2$. In agricultural soils, N$_2$O emissions are influenced by a large number of environmental characteristics and crop management techniques that are not systematically reported in experiments. Random Forest (RF) is a machine learning method that can handle missing data and ranks input variables on the basis of their importance. We aimed to predict N$_2$O emission on the basis of local information, to rank environmental and crop management variables according to their influence on N$_2$O emission, and to compare the performances of RF with several regression models. RF outperformed the regression models for predictive purposes, and this approach led to the identification of three important input variables: N fertilization, type of crop and experiment duration. This method could be used in the future for prediction of N$_2$O emissions from local information.

1.   Introduction

Nitrous oxide (N_2O) is a major greenhouse gas produced by agricultural activities: it has a global warming potential approximately 298 times higher than that of CO_2 and a lifetime of 114 years in the atmosphere (IPCC, 2007). In 2005, agriculture accounted for 60% of all anthropogenic N_2O emissions (Smith et al., 2007). N_2O emission from agricultural soils depends on nitrification and denitrification processes (Mosier et al., 1998), which are influenced by a large number of environmental characteristics (*e.g.* soil organic content) and crop management techniques (*e.g.* soil tillage). The amount of nitrogen (N) fertilizer applied has been identified as the main factor directly influencing N_2O emission (Mosier et al., 1998; Stehfest and Bouwman, 2006). However, a number of other variables, including soil tillage, soil drainage and rainfall, also determine N_2O emission levels (Gu et al., 2011; Snyder et al., 2009).

The Intergovernmental Panel on Climate Change (IPCC) defined N_2O emission as 1% of the applied N in the Tier 1 level (IPCC, 2006). As an alternative, several studies have suggested using nonlinear models based on exponential functions to calculate emissions as a function of applied N (Hoben et al., 2011; Ma et al., 2010; Millar et al., 2010; Philibert et al., 2012b). In these models, the amount of N applied is the only input variable, but the emission coefficient is no longer constant.

Many studies have focused on the calculation of N_2O emissions not only as a function of applied N, but also as a function of several environmental and crop management variables. Stehfest and Bouwman (2006) developed linear regression models relating N_2O emissions to several variables describing climate and soil conditions, and agricultural practices. Leip et al. (2011) developed spatially stratified emission factors as a function of soil organic content, fertilizer type (mineral fertilizer or manure) and weather conditions, using the process-based model DNDC-EUROPE. Lesschen et al. (2011) developed an inference scheme taking into account several variables, including fertilizer type, crop residues, atmospheric deposition, land use, soil type and precipitation. Berdanier and Conant (2012) used a hierarchical Bayesian model and took into account climate, crops, soils, type of fertilizer (synthetic or manure), crop residues and crop management, for each region of the world.

The main problem with models including several environmental and crop management characteristics as input variables is that they require the input of large amounts of data. This makes it difficult to estimate the parameters of the model, because only a few of the published N_2O experiments have included detailed information about local pedoclimatic conditions and agricultural practices. The dataset of Stehfest and Bouwman (2006) is one of the largest available datasets concerning N_2O in agricultural soils: it reports N_2O emission data extracted from 206 papers. In this dataset, 29% of the data are missing for soil carbon content, 12% for
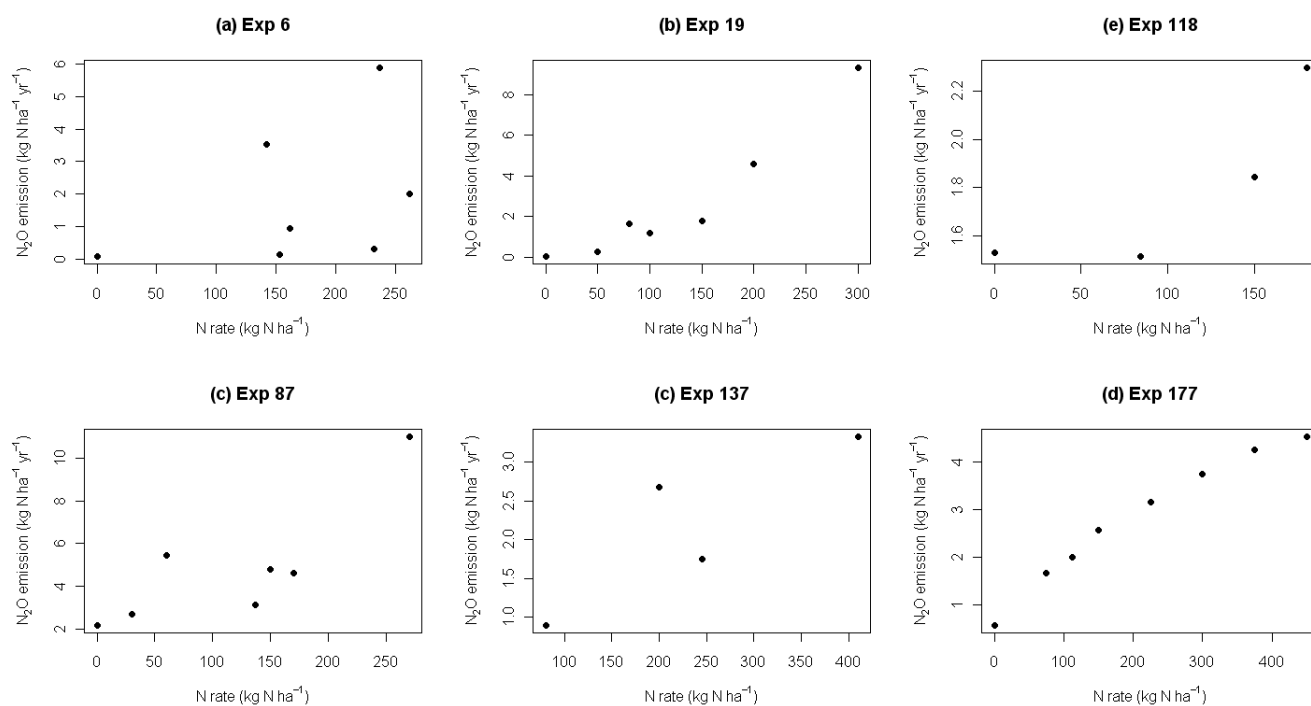
Figure 3.1 : N₂O emissions measured for several amounts of applied N in six experiments of the Stehfest and Bouwman's dataset. Black points indicate mean N₂O values averaged over replicates.

soil texture and 94% for annual precipitation. As pointed out by Lesschen et al. (2011), "(…) the empirical model of Stehfest and Bouwman could only be used for 133 sites because of missing data for several factors." Process-based models are also costly and difficult to use in practice, due to the large numbers of inputs required (Villa-Vialaneix et al., 2012).

Random Forest (RF) is a machine learning method developed by Breiman (2001) that has been shown to be of value in many situations. It is frequently used in bioinformatics and genetics (Cutler and Stevens, 2006), and its capacity for dealing with large numbers of explanatory variables for which only small numbers of observations are available and for handling missing data have made RF an indispensable tool in ecology (Cutler et al., 2007; Lapointe and Light, 2012; Philibert et al., 2011; Prasad et al., 2006; Stohlgren et al., 2010). RF can be used for quantitative predictions and for classification purposes, in addition to ranking input variables in order of importance for the prediction of a variable of interest.

RF is a potentially useful tool for predicting N_2O emissions, due to (i) its ability to deal with missing data, (ii) its ability to identify the most important variables, and (iii) its computation time for large datasets, which is much shorter than that for process-based models. This method was used by Villa-Vialaneix et al. (2012) to build a metamodel of the process-based DNDC-Europe model (Leip et al., 2008; Li et al., 1992; Li, 2000; Follador and Leip, 2009). In this context, RF was compared with seven other techniques, which it outperformed in terms of computation time, with a large training dataset and an acceptable level of accuracy. However, despite this demonstration that RF outperforms other techniques, RF has not since been used for the prediction of real N_2O emissions. As the study of Villa-Vialaneix et al. (2012) was based on simulated data and focused on European corn crops only, it would be useful to evaluate RF performance with a large dataset including real N_2O emission measurements.

Our objective was to predict N_2O emissions and to rank environmental and crop management (input) variables with the Random Forest method. We compared the performances of RF and several regression models for both quantitative predictions and classification.

2. Materials and methods

2.1. Data

The dataset used was a global compilation of the results of N_2O experiments extracted from published papers by Stehfest and Bouwman (2006). Each experiment involved one or several

Table 3.1 (First part): Name, type, number of categories (for categorical variables), description and number of measurements available for the two response variables and the 19 input variables. The first column indicates the group to which each variable belongs.

Group	Name	Type*	Number of categories	Description	Number of measurements
-	N_2O	Q	-	nitrous oxide emission ($kg\ N\ ha^{-1}\ year^{-1}$)	594
-	N_2O_cat	C	2	> or < 0.931 ($kg\ N\ ha^{-1}\ year^{-1}$)	594
1	Carbon content	C	3	soil organic content < 1%, $1\% \leq SOC \leq 3\%$, SOC > 3%	594
1	pH	C	3	pH< 5.5, $5.5 \leq pH \leq 7.3$, pH >7.3	594
1	Exp duration	C	2	$\leq$ or > to 300 days of experiment	594
1	Crop_type	C	6	cereal, grass, wetland rice, legume, other, none	594
1	Climate	C	4	temperate continental, temperate oceanic, subtropical, tropical	594
1, 2, 3	N rate	Q	-	amount of N applied ($kg\ N\ ha^{-1}$)	594
1, 2, 3	Texture class	C	3	soil texture (coarse, medium or fine)	594
2, 3	Carbon content	Q	-	soil organic content (%)	594
2, 3	pH	Q	-	soil pH	594
2, 3	Exp duration	Q	-	number of days of experiment	594
2, 3	Crop type	C	10	cereal, grass, grass-clover, irrigated crop, wetland rice, legume, row crop, rotation, other, none	594

cultivated plots, with one or several N fertilizer treatments for which N$_2$O emissions were measured (examples of data taken from Stehfest and Bouwman's dataset are shown in Figure 3.1).

We removed N$_2$O measurements obtained with N fertilizer doses of more than 500 kg N ha^{-1} from the dataset, because such high doses of fertilizer are rarely applied in farmers' fields (Roelandt et al., 2005; Spiertz, 2010; Tilman et al., 2002). Only experiments including data for the seven input variables of the regression model of Stehfest and Bouwman (2006) were selected (see Group 1 in Table 3.1). This selection was necessary to facilitate a comparison of the performance of this regression model with RF. Twelve observations from sites with a boreal climate were also removed. The final dataset included 594 observations extracted from 114 publications.

The response variable was the emission of nitrous oxide (N$_2$O Table 3.1), and its value varied from 0.004 to 31.73 kg N ha^{-1} year^{-1}, with a mean of 2.13 kg N ha^{-1} year^{-1} and a median of 0.93 kg N ha^{-1} year^{-1}. A binary N$_2$O response variable was created (N$_2$O_cat Table 3.1) to assess the utility of RF for classification; the two categories corresponded to N$_2$O emissions above and below the median value in the dataset. As a result, each category included 297 observations.

Three groups of input (predictor) variables were created:

- A group containing the seven input variables of the regression model of Stehfest and Bouwman (2006) (Group 1, Table 3.1).
- A group containing the same input variables, but with different types and higher number of categories (Group 2, Table 3.1). Carbon content, experiment duration and pH were considered as quantitative rather than categorical variables; the number of categories for climate was increased (six rather than four categories) and the number of crop types was also increased (ten rather than six); we also distinguished between grass-clover, grass, irrigated, row crop, and rotation crop.
- A group containing the seven input variables, as defined for the second group, plus seven additional variables for which values were available from the dataset of Stehfest and Bouwman (2006). Some data were missing for these additional variables, but the proportion of missing data was always below 50% (Group 3, Table 3.1).

The use of three different groups of input variables made it possible to assess the performance of RF with different types of predictors, with or without missing data, and to compare RF (i) with the regression model of Stehfest and Bouwman (2006) and (ii) with a simple nonlinear model predicting N$_2$O emission as a function of applied N (Philibert et al., 2012b).

Table 3.1 (Second part): Name, type, number of categories (for categorical variables), description and number of measurements available for the two response variables and the 19 input variables. The first column indicates the group to which each variable belongs.

Group	Name	Type*	Number of categories	Description	Number of measurements
2, 3	Climate	C	6	temperate continental, temperate oceanic, subtropical summer rains, subtropical winter rains, tropical warm humid, tropical seas dry	594
3	Fertilizer type	C	13	AA, AN, AP, AS, CA, CAN, KN, Mix, NP, O, OS, U, UAN**	434
3	Method of N application	C	4	broadcast, broadcast at panicle initiation, incorporated, solution	348
3	Timing of N application	C	3	single, single but part of split, split	408
3	Drainage	C	2	hydromorphic (or no) properties	557
3	Method N₂O	C	5	closed chamber, soil core method, soil gradient, micrometeorological, open chamber	594
3	Freq N₂O	C	6	more than 1 per day, daily, every 2 d-3 d, every 3 d-week, less than 1 per week, not known	594
3	Country	C	22	Germany, Belgium, Brazil, Canada, China, Costa Rica, Denmark, Spain, France, India, Indonesia, Italy, Japan, New Zealand, UK, Pakistan, Netherlands, Philippines, Puerto Rico, Switzerland, USA, Venezuela	594

* C = categorical variable, Q = quantitative variable

** AA= anhydrous ammonia including aqueous ammonia, AN= ammonium (sulfate) nitrate, AS= ammonium sulfate/chloride, CAN= calcium ammonium nitrate, KN= potassium nitrate/sodium nitrate/calcium nitrate, Mix= combination of various synthetic fertilizers, NP= ammonium phosphate and other NP fertilizers, O= organic, OS= combination of organic and synthetic fertilizers, U= urea/urine, UAN= urea ammonium nitrate

2.2. Main principles of Random Forest

Random Forest (RF) is a machine learning technique that builds an ensemble of classification (or regression) trees (Breiman and Cutler, 2003). With this technique, no precise information is required about the form of the relationship between response and input variables. RF creates a set of binary decision rules based on the predictor (input) variables (Breiman et al., 1984).

The first step of the method is the random selection, with replacement, of a bootstrap sample of observations (called the ―in-bag") of equal size to the total number of observations (Efron and Tibshirani, 1986, 1993). About one third the initial number of observations are not selected and these observations are referred to as ―out-of-bag" data (OOB).

The second step involves the random selection of explanatory variables, *mtry* (Ghattas 2000; Prasad et al., 2006).

The third step involves the building of a tree based on the in-bag data and *mtry* variables selected. The tree is built by recursively partitioning the initial data into smaller groups, called nodes, through binary splits based on a single predictor variable. All splits with all predictor variables are examined and the best split at each step is chosen with different criteria for classification and regression trees. Tree growth is considered to have ended when the number of observations in the terminal nodes reaches one for classification trees and is five or less for regression trees. This third-step process is repeated *ntree* times, to generate *ntree* bootstrap samples and trees. At each iteration, the OOB data are used to calculate a misclassification rate for classification trees and an $MSE_{OOB(t)}$ value for regression trees, defined as

$$MSE_{OOB(t)} = \frac{1}{n_{OOB,t}} \sum_{\substack{i=1 \\ i \in OOB_t}}^{n_{OOB,t}} \left(y_i - \hat{y}_{i,t}\right)^2 \quad \text{where } n_{OOB,t} \text{ is the number of OOB data for tree } t, \ y_i \text{ the}$$

observation i and $\hat{y}_{i,t}$ the prediction of the observation i for the tree t.

For regression trees, the predicted OOB observation i in tree t is the mean N₂O value for the in-bag observations present in the same terminal node. For classification trees, the predicted value is the category accounted for the largest number of in-bag observations in the same terminal node.

RF handles missing values by replacing them with the mean/mode value for observations without missing data defined as being close by a measurement of proximity (Breiman and Cutler, 2003). The first step consists in replacing missing data by the mean/mode values of the input variable leading to a new dataset without missing data. Secondly, a matrix of proximity (n*n where n is the total number of observations) is computed using Random Forest with the new dataset. Each element of this matrix is set equal to the fraction of trees where two

Table 3.2 : Percentages of misclassification by Random Forest (RF) for out-of-bag samples (OOB error rate) and by ten-fold cross-validation (CV). RF was implemented with three different groups of input variables.

	RF Group 1	RF Group 2	RF Group 3
OOB error rate	23%	24%	21%
ten-fold CV error rate	22%	22%	20%
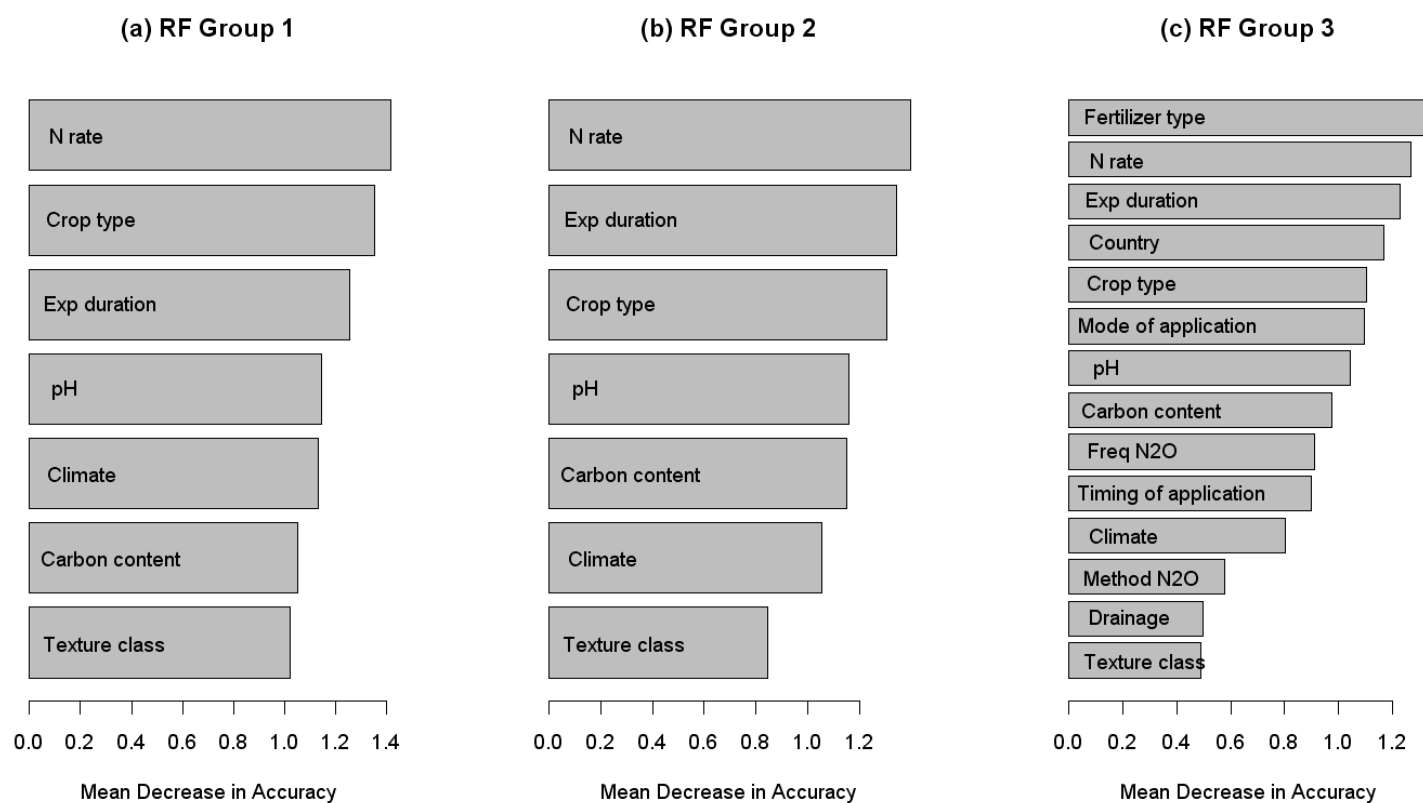
Figure 3.2: Importance of the input variable for N₂O classification, as determined by mean decrease in accuracy (MDA) with Random Forest for both categories of N₂O emission. The variables were sorted in decreasing order of MDA.

observations fall in the same terminal node. The third step consists in replacing the mean/mode values defined at the first step by new mean/mode values calculated by weighting the observations (without missing data) with the elements of the proximity matrix. Missing data are estimated by repeating the above procedure *m* times and, thus, using a total number of trees equal to *ntree*m*. Once the missing data have been replaced, RF is run again to generate *ntree* new trees, used for subsequent predictions (here, of N$_2$O emission) and for ranking inputs according to their importance.

The global error rate for the total algorithm is, for classification trees the mean frequency of misclassified OOB data over the *ntree* trees (OOB error rate) (Liaw and Wiener, 2002). For regression trees, the global error rate is the root mean of square residuals defined as

$RMSE_{OOB} = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{\hat{y}}_i^{OOB}\right)^2}$ where *n* is the total number of observations, $\bar{\hat{y}}_i^{OOB}$ is the mean

predicted value for observation *i* over all trees in which this observation is in the OOB data and y_i the *i*[th] observation (Liaw and Wiener, 2002).

2.3. Application of Random Forest to the prediction of N$_2$O emission

Random forest was used to predict the response variables N$_2$O and N$_2$O_cat from the input variables included in groups 1, 2 and 3 successively (Table 3.1). For the variable N$_2$O, regression trees were built by choosing, at each step of the algorithm, the split maximizing the between-group sum of squares (Therneau and Atkinson, 1997). The number of input variables used to build individual trees (*mtry*) was set at one-third the total number of variables (Breiman, 2001; Liaw and Wiener, 2002), *i.e.* 2 for groups 1 and 2, and 4 for group 3. For N$_2$O_cat, classification trees were generated with the Gini index criterion used to determine the best split at each step of the algorithm (Breiman et al., 1984; Therneau and Atkinson, 1997). The number of input variables used to build individual trees (*mtry*) was set to the integer part of the square root of the total number of variables (Breiman, 2001; Liaw and Wiener, 2002), *i.e.* 2 for groups 1 and 2, and 3 for group 3. As recommended by Liaw and Wiener (2002), we also tested values half and twice that of *mtry* for classification and regression trees, but this gave no improvement in the results. For the replacement of missing data, the number of trees was set to *ntree*=1000, with *m*=5 iterations. For prediction and input variable ranking purposes, the number of bootstrap samples (equal to the number of trees) was set to *ntree*=500. Predictions were not improved by the use of higher values.

Input variables were ranked according to two measurements of importance: mean decrease in accuracy (MDA) for classification trees and percentage increase in mean square error (%MSE increase) for regression trees. For N$_2$O_cat, the MDA of a given input variable was defined as the relative increase in the misclassification rate calculated with OOB data after a random

(a) RF Group 1 **(b) RF Group 2** **(c) RF Group 3**
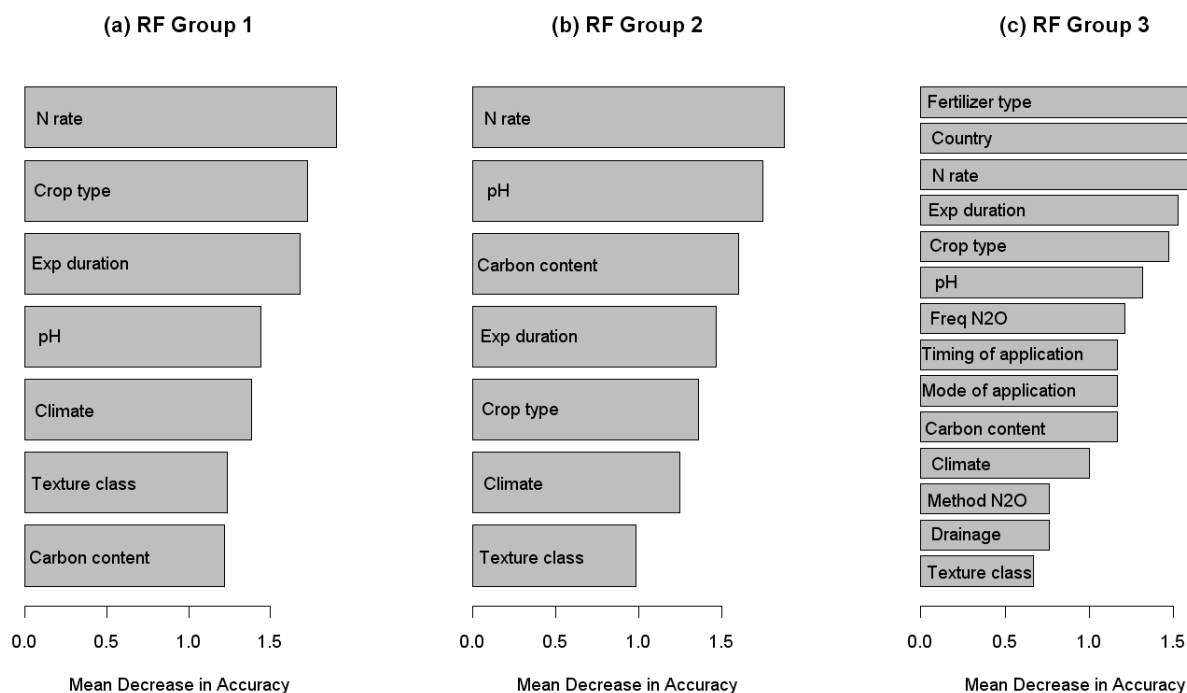
Figure 3.3: Importance of the input variable for N₂O classification, as determined by calculating the mean decrease in accuracy (MDA) with Random Forest for the category ―low N₂O emissions" (< 0.93 kg N ha⁻¹ year⁻¹). The variables were sorted in decreasing order of MDA.

(a) RF Group 1 **(b) RF Group 2** **(c) RF Group 3**
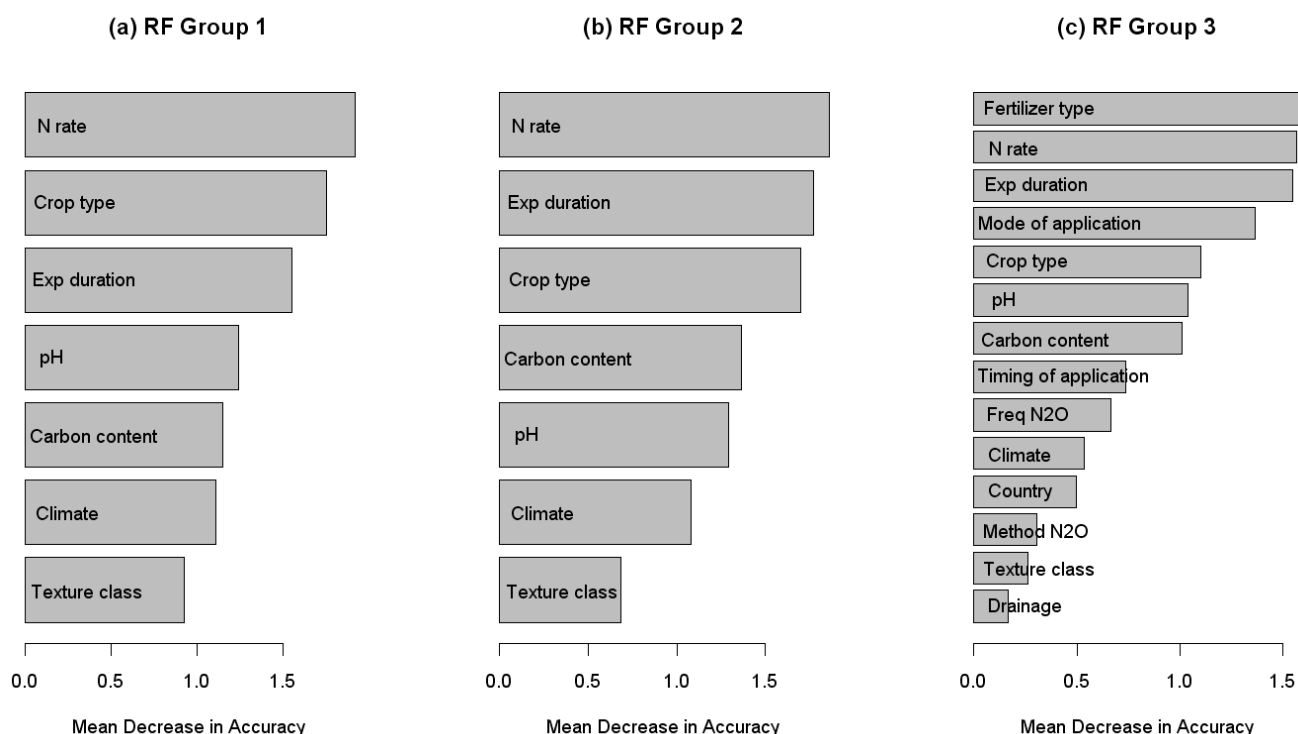
Figure 3.4: Importance of the input variable for N₂O classification measured by determining the mean decrease in accuracy (MDA) with Random Forest for the category ―high N₂O emissions" (> 0.93 kg N ha⁻¹ year⁻¹). The variables were sorted in decreasing order of MDA.

permutation of the values taken by this input variable, the other variables remaining unchanged. The misclassification rate was averaged over all the trees in which this variable was involved in at least one split. Three types of misclassification were considered successively: misclassification of the ―low N₂O emission" category (< 0.93 kg ha^{-1} year^{-1}), misclassification of the ―high N₂O emission" category (> 0.93 kg ha^{-1} year^{-1}), and misclassification of both categories.

For the response variable N₂O, the % MSE increase was defined as the relative increase in $MSE_{OOB(t)}$ induced by a random permutation of the input values, with the other variables remaining unchanged. Results were averaged over all the trees in which this variable was involved in at least one split (Liaw and Wiener, 2002).

All computations were carried out with the randomForest package of R software (R 2.9.2) (Liaw and Wiener, 2002). The code is available, on request, from the corresponding author (an example of R code with randomForest was presented in Annex 4).

2.4. Regression models

Two regression models were considered. The first one was that presented by Stehfest and Bouwman (2006). It is a linear model relating log-transformed N₂O emission values to seven explanatory variables (Group 1 in Table 3.1). Parameter values were set to the values reported by Stehfest and Bouwman (2006) (Table 5 of their paper).

The second model was a nonlinear random-effects model relating N₂O emission to N fertilizer dose via an exponential function (Philibert et al., 2012b). The two parameters of the exponential function were assumed to differ between published experiments. Compared to its version with fixed effects, the random-effects model showed lower values for the Akaïke Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model can be defined as follows:

$Y_{ijk} = \exp\left(\alpha_{0i} + \alpha_{1i} X_{ij}\right) + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \sim N\left(0, \tau^2\right)$, $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$ and $\alpha_{1i} \sim N\left(\mu_1, \sigma_1^2\right)$. Y_{ijk} is the N₂O emission (kg N ha^{-1} year^{-1}) measured in the ith published experiment (i=1 ... 114), for the jth applied N dose (j=1 ... N_i) and the kth replicate (k=1 ... K_{ij}). X_{ij} is the jth applied N dose (kg N ha^{-1}) in the ith published experiment. μ_0 is the log mean background emission, α_{0i} is the log published experiment-specific background emission (random), μ_1 is the log mean applied N effect, α_{1i} is the log published experiment-specific applied N effect (random), and ε_{ijk} is the residual error term. The random terms α_{0i}, α_{1i} and ε_{ijk} were assumed to be independent and normally distributed. The values of μ_0, μ_1, σ_0, σ_1, and τ were estimated from the dataset described in 2.1 by an approximate maximum likelihood method implemented in the nlme package of R statistical software (Pinheiro and Bates, 2000).
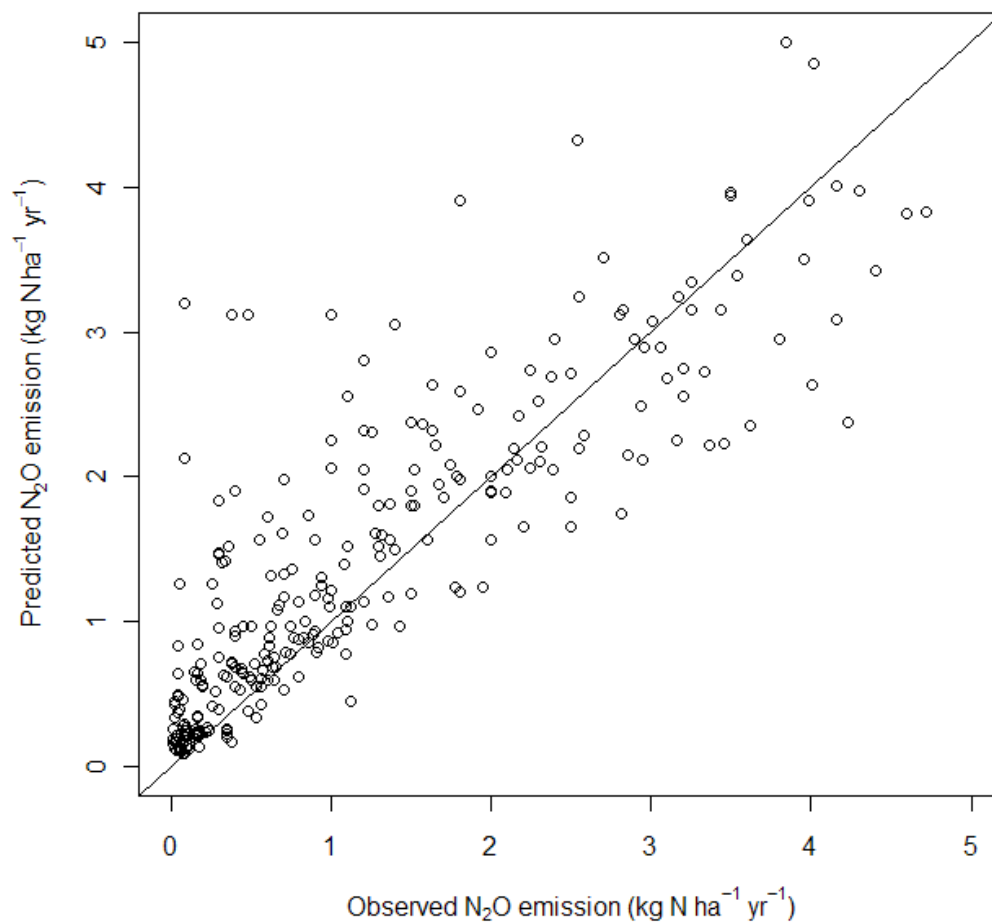
Figure 3. 5: Observed N$_2$O emissions versus predicted emissions obtained with Random Forest and Group 3 input variables.

2.5. Evaluation of predictions

The regression model of Stehfest and Bouwman (2006) was evaluated by calculating the root mean square error of N_2O predictions with the parameter values reported by the authors. Predictions of N_2O emissions derived with RF and with the random-effects nonlinear regression model were evaluated by ten-fold cross validation. The dataset was divided into ten subsets of equal size, and nine tenths of the dataset were used for model fitting and RF. N_2O emission for the remaining data was then predicted with the fitted model and with RF. The results were used to calculate the root mean square error of N_2O predictions (RMSEP) and the percentage misclassification of the binary variable N_2O_cat. For RF, predictions for one tenth of the dataset were obtained from the 500 trees fitted to the other nine tenths of the dataset. RF predictions were assessed for each of the three groups of input variables.

In addition to the ten-fold cross-validation described above, RF predictions were assessed by calculating $RMSE_{OOB}$, as defined above, and the percentage misclassification of the OOB samples (OOB error rate) obtained with the R function randomForest.

3.   Results

3.1. Classification of nitrous oxide emissions

For RF, the misclassification rate ranged from 20 to 24%, depending on the group of input variables considered and the assessment method used (Table 3.2). Thus, more than 75% of the data were correctly classified with RF, which clearly outperformed a random decision rule based on the choice between two categories with a probability of 0.5, giving a misclassification rate of 50%. The misclassification rate was lower for group 3 (*i.e.*, with 14 input variables) than for groups 1 and 2, but the differences were small. This result indicates that the use of a higher number of input variables, including missing data, did not decrease RF performance. Misclassification rates were slightly higher when estimated with OOB samples, but the two types of assessment (OOB and ten-fold cross validation) gave very similar results.

When MDA was calculated with both categories of N_2O emission (*i.e.* high and low), "N rate", "crop type", and "experiment duration" were the three most important input variables in groups 1 and 2 (Figure 3.2ab). These three variables were also ranked among the five most important input variables in group 3 but, in this group, "type of fertilizer" was the highest ranked variable and "country" was ranked fourth (Figure 3.2c).

The ranking of the input variables was only slightly modified by the calculation of MDA with data from only the "low emission" category (Figure 3.3) or the "high emission" category (Figure 3.4). The first three variables in groups 1 and 2 were still "N rate", "crop type" and

Table 3.3 : Root mean square error of prediction (RMSEP, in kg N ha^{-1} year^{-1}) for the random-effects model and Random Forest (RF), obtained with out-of-bag samples (OOB) and by ten-fold cross-validation (CV).

	Random-effects model	RF Group 1	RF Group 2	RF Group 3
RMSE$_{OOB}$	-	3.12	2.99	2.91
ten-fold CV RMSEP	3.35	3.17	3.02	2.97
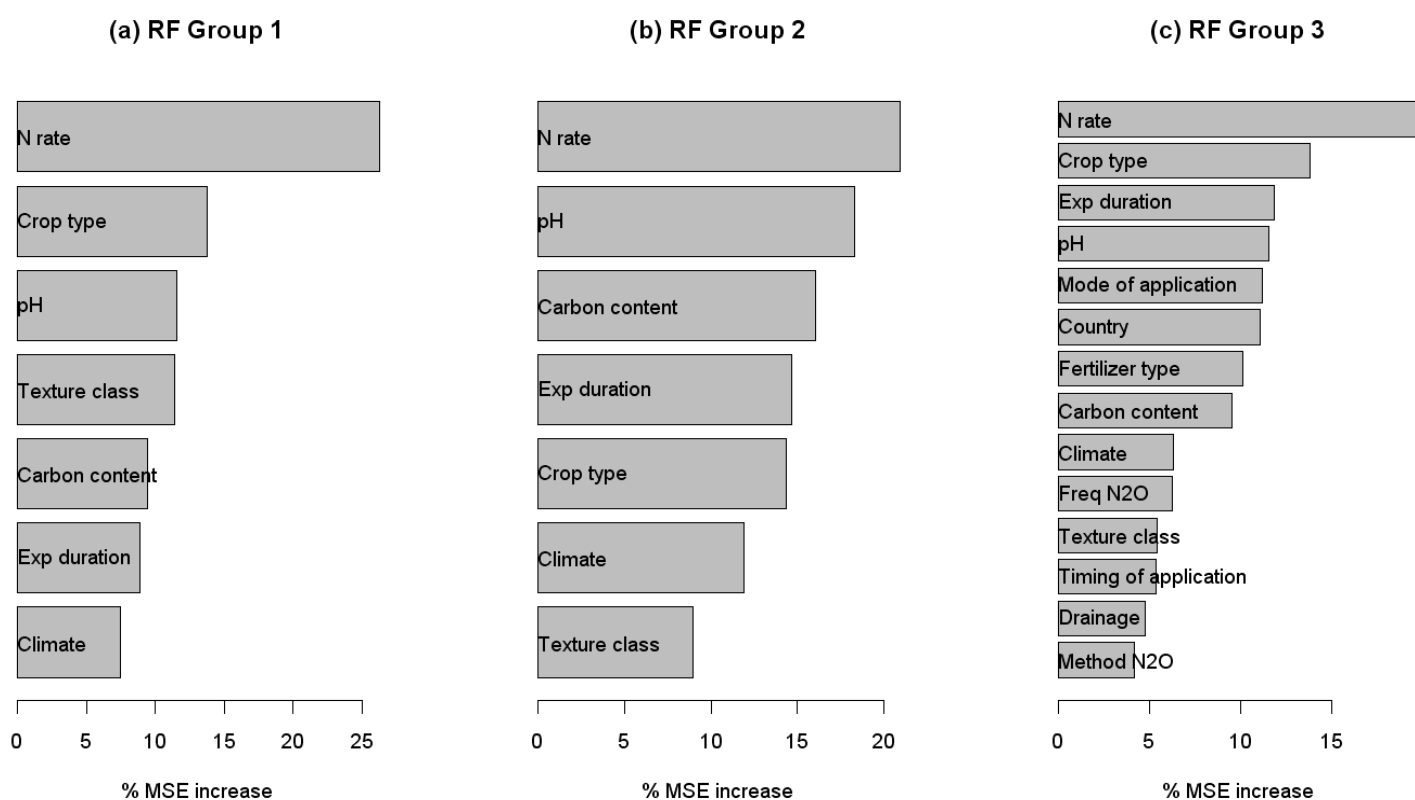
Figure 3. 6 : Importance of the input variable for quantitative N$_2$O predictions, as measured by determining the percentage increase in mean square error (%MSE increase) with Random Forest. The variables were sorted in decreasing order of the importance value.

―experiment duration" (Figures 3.3ab, 3.4ab). In group 3, the variable ―country" was more highly ranked when MDA was calculated with data from the ―low emission" category than when calculated with data from the ―high emission" category only (Figures 3.3c, 3.4c). ―Country" is therefore more important for the prediction of low N$_2$O emissions than for the prediction of high N$_2$O emissions.

3.2. Quantitative predictions of nitrous oxide emissions

RF systematically outperformed the random-effects model (Table 3.3); RMSEP ranged from 2.91 to 3.17 kg ha^{-1} year^{-1} for RF, and 3.35 kg ha^{-1} year^{-1} for the random-effects model. The RMSEP of the regression model of Stehfest and Bouwman (2006) was 4.05 kg ha^{-1} year^{-1} and was thus higher than the values obtained for either the random-effects model or RF.

The lowest RMSEP (and thus the best predictions) were obtained with RF and group 3 input variables. Observed versus predicted N$_2$O emissions are shown in Figure 3.5. The inclusion of a larger number of input variables (14 rather than 7), including missing data, slightly improved the performance of RF. We found no major difference between the two methods used to estimate RMSEP (OOB and ten-fold cross-validation) although the use of OOB samples resulted in slightly lower RMSEP values.

Figure 3.6 shows the rankings of the input variables obtained with RF. The variable ―N rate" was ranked first in the three groups of input variables. This result confirms that ―N rate" is a key variable for predicting N$_2$O emissions. The MDA value for ―N rate" was much higher than those of the other input variables of groups 1 and 3, but the differences between variables were smaller in group 2. The second most important input variable was ―crop type" for groups 1 and 3 (Figure 3.6ac), and ―soil pH" for group 2 (Figure 3.6b). The variable ―soil pH" was the third-ranked input variable in group 1 and the fourth-ranked variable in group 3.

4. Discussion

The IPCC has proposed three methods for estimating the emissions of N$_2$O due to agricultural activities (IPCC, 2006). The Tier 1 method involves calculating N$_2$O emission by multiplying the amounts of nitrogen coming from several sources (mineral and organic fertilizer, crop residues and mineralization of organic matter) by emission factors (fixed default values). This method requires very few inputs and is easy to implement in practice. However, it does not take into account crop type, soil type, climatic conditions or crop management practices. This oversimplicity may lead to inaccurate estimations of N$_2$O emission (Flynn et al., 2005; Grace et al., 2011; Skiba and Smith, 2000).

The Tier 2 and Tier 3 methods may replace the Tier 1 method if country- or region-specific data are available (IPCC, 2006; Leip et al., 2011; Lugato et al., 2010; Rochette et al., 2008). The Tier 2 method is similar to the Tier 1 method, but makes use of country-specific emission factors. The Tier 3 method is based on model and inventory measurement approaches. Tier 2 and Tier 3 methods may relate soil, environmental and management variables to N$_2$O emissions.

Various types of model have been proposed for the estimation of N$_2$O emissions. Process-based models, such as the DNDC-Europe model (Follador et al., 2011; Leip et al., 2008; Li et al., 1992; Li, 2000), describe the process underlying N$_2$O emissions. They include large numbers of input variables and, thus, require a large number of measurements. They also frequently require large amounts of computation time, making it difficult and costly to implement such models over large scales (Villa-Vialaneix et al., 2012). An interesting alternative consists in using meta-models that reproduce process-based models using statistical techniques (Britz and Leip, 2009; Villa-Vialaneix et al., 2012).

Statistical regression models constitute an alternative approach. They relate N$_2$O emissions to several input variables through simple equations and are not particularly demanding in terms of computation. However, it is necessary to estimate their parameters from a large dataset including measurements of the input variables and N$_2$O emissions. Datasets including N$_2$O emission measurements are available (*e.g.* Stehfest and Bouwman, 2006), but they include few data for environmental and management characteristics. It is thus difficult to build robust statistical relationships between these characteristics and N$_2$O emissions.

Although prediction errors were large with all techniques, we found that RF gave more accurate predictions than two statistical regression models. We found that the performance of the RF method was robust, with the missing data handled well. This finding is consistent with those of previous ecological studies showing that RF outperforms other techniques (Cutler et

al., 2007; Philibert et al., 2011; Prasad et al., 2006). RF has also recently been shown to be of value for meta-modeling (Villa-Vialaneix et al., 2012).

One of the disadvantages of RF with respect to statistical regression models is that RF cannot be used to analyze precisely the effects of different input variables on N_2O emissions. It is thus difficult to use RF to determine whether a given input variable (*e.g.* a given crop type) has a positive or negative effect on N_2O emission. However, RF can be used to rank input variables according to their degree of importance. We used two different measurements: the mean decrease in accuracy (for the classification of N_2O emission into two categories) and the relative increase in mean square error (for quantitative prediction). Analyses based on these two measurements indicated that the three most important factors were N fertilization, type of crop and experiment duration. The importance of the amount of N applied is not surprising. This variable is the main input variable for the Tier 1 and Tier 2 methods and its effect on N_2O emissions has been acknowledged in many papers (Davidson, 2009; Mosier et al., 1998; Snyder et al., 2009; Stehfest and Bouwman, 2006).

The type of crop has been shown to affect N_2O emission in previous studies. For example, according to Rochette and Janzen (2005), legume crops generally result in lower emission levels than other crop types. The effect of the type of crop may also be partly due to other factors, because several management practices are adjusted as a function of the type of crop grown (*e.g.* soil tillage) and because some crops are cultivated only on particular soil types or in particular climatic conditions (*e.g.* wetland rice mostly grown in poorly drained soil). The effect of experiment duration was discussed by Stehfest and Bouwman (2006), who found that measurements of N_2O emission increased with experiment duration and that measurements over a period of less than one year might be biased. The input variable ‚experiment duration' could be fixed at a high value (365 days) for practical application of Random Forest in Tier 2 or Tier 3 methods. N_2O emissions will then be simulated for one year.

Given its ability to predict N_2O emissions and to handle missing data, we believe that Random Forest could be used by the IPCC for the implementation of its Tier 2 and Tier 3 methods. Due to the lack of availability for some input variables (*e.g.* mode and timing of application of N fertilizer), RF based on Group 2 input variables seems to be the most appropriate technique for practical application.

Chapitre 4
Prédire les émissions de N$_2$O avec un modèle à effets aléatoires

Ce chapitre correspond à un article soumis à la revue Environmental Modelling and Software en mars 2013.

Predicting N$_2$O emission with a random-effects model

Aurore Philibert[b,a], Chantal Loyce[b,a], David Makowski[a,b]

[a] INRA, UMR 211 Agronomie, F-78000 Thiverval Grignon, France

[b] AgroParisTech, UMR 211 Agronomie, F-78000 Thiverval Grignon, France

Abstract

Nitrous oxide (N$_2$O) is a greenhouse gas with a global warming potential 298 times greater than that of CO$_2$. The main source of anthropogenic N$_2$O emission is the application of nitrogen fertilizer to agricultural soils. We present an approach for predicting N$_2$O emissions based on a statistical random-effects model: the N$_2$O emission response to applied nitrogen is described by an exponential function, the parameters of which are assumed to vary randomly between locations. One of the advantages of this model is that its parameters are easily adjusted to one or several location-specific N$_2$O measurements. The adjusted model can then be used to predict N$_2$O emissions for nitrogen fertilizer doses other than those applied at the location considered. We evaluated the accuracy of model prediction, with real and simulated data. The use of location-specific rather than average predictions reduced prediction errors in most cases. Location-specific predictions could be used to estimate background emission in on-farm studies.

Key-words: nitrous oxide, experimental design, random-effects model, Bayesian model, nitrogen fertilization, model error

1.   Introduction

N_2O is a potent greenhouse gas with a lifetime in the atmosphere of 114 years and a global warming potential 298 times greater than that of CO_2 for a 100-year horizon (IPCC, 2007). Atmospheric N_2O concentration has risen from 270 parts per billion (ppb) in the pre-industrial period to 319 ppb in 2005 (IPCC, 2007). N_2O is emitted due to nitrification and denitrification processes. Thus, all factors modifying these physicochemical exchanges affect N_2O emission. The main source of anthropogenic emissions is nitrogen (N) fertilizer application to agricultural soils (Davidson, 2009; Mosier et al., 1998; Snyder, 2009; Stehfest and Bouwman, 2006). Many other factors may also modify N_2O emissions in cultivated areas. These factors are of three main types (Freibauer and Kaltschmitt, 2003): (i) soil characteristics (*e.g.* soil organic content, soil texture), (ii) climate characteristics (*e.g.* rainfall, freeze-thaw cycle (Freibauer and Kaltschmitt, 2003)) and (iii) management practices (*e.g.* N fertilization, tillage (Rochette, 2008)).

The amount of N applied and other management practices are often reported in N_2O emission experiments, but information about soil and climate characteristics is often missing. For example, in the dataset of Stehfest and Bouwman (2006), the amount of applied N was systematically reported, but 29% of soil organic content data, 12% of soil texture data and 94% of annual precipitation data were missing. Datasets including N_2O measurements have been used to develop simple statistical models relating N_2O emissions to the amount of applied N through linear or exponential functions (Bouwman, 1996; Hoben et al., 2011; Philibert et al., 2012b). These simple models can be used to predict N_2O emission as a function of the amount of N applied, but they cannot describe the variability of the N_2O response.

Several authors have suggested that N_2O emission could be predicted from several input variables relating to soil characteristics, climate characteristics and management practices (Berdanier and Conant, 2012; Leip et al., 2011; Lesschen et al., 2011). However, as mentioned above, little information about soil and climate characteristics is generally given and it is often difficult to establish robust relationships between N_2O emissions and environmental variables (Lesschen et al., 2011). Mechanistic models have been developed for the prediction of N_2O emissions as a function of soil and climate characteristics and management practices (DNDC model; Li, 2000). These models are highly attractive but require large numbers of input data that are not frequently available in practice, particularly for farmers' fields, and are often too costly in terms of computation time (Villa-Vialaneix et al., 2012).

We present here a new approach for predicting N_2O emissions in experiment and farmers' fields. Our approach involves relating N_2O emissions to the amount of applied N through an

exponential function, the parameters of which are assumed to vary between locations. One of the major advantages of the proposed model (a statistical random-effects model) is that its parameters are easily adjusted to one or several location-specific N_2O measurements, when such measurements are available. The adjusted model can then be used to predict N_2O emissions for N fertilizer doses other than those applied at the location considered or in the farmer's field. This method does not require any information about soil and climate characteristics; these characteristics are taken into account through the single or small number of location-specific N_2O measurements used to adjust the model.

The approach described here could be used in several applications. It could, for example, be applied to a farmer's field, to estimate background N_2O emission (*i.e.*, the emission occurring in the absence of N fertilizer application), or to estimate the consequences of decreasing or increasing the amount of N fertilizer applied by the farmer. Our approach can also be used at experimental sites at which a limited number of N doses are tested, to estimate N_2O emission for other amounts of applied N, smaller or larger than those applied in the trial.

We aimed to assess the ability of two different random-effects models (one Bayesian, one frequentist) to predict N_2O emission from one or several location-specific N_2O measurements for the locations for which N_2O predictions were performed. We assessed the accuracy of model predictions and used the results to evaluate the benefits of adjusting random-effects models on the basis of location-specific measurements.

2.   Materials and methods

2.1. Data

We used the dataset of Stehfest and Bouwman (2006), discarding observations for amounts of applied N greater than 500 kg N ha^{-1}, because such high doses of fertilizer are not usually applied to farmers' fields (Roelandt et al., 2005; Spiertz, 2010; Tilman et al. 2002). The resulting dataset includes 985 N_2O emission measurements, extracted from 203 publications. Each publication reported a set of N_2O emission measurements for several N fertilizer treatments at a single location. N_2O emission values ranged from 0.003 to 46.44 kg N ha^{-1} year^{-1}, with a mean of 2.4 kg N ha^{-1} year^{-1} and a median of 1.07 kg N ha^{-1} year^{-1}. These data were obtained for N fertilizer doses of 0 to 500 kg N ha^{-1}, with a mean of 124 kg N ha^{-1} and a median of 100 kg N ha^{-1}. The number of doses tested at each location ranged from 1 to 13, with a mean of 2.54 and a median of 2. In total, 92 of the publications came from Europe, 57 from North America, 25 from Asia, 21 from Latin America and eight from Oceania.

2.2. Statistical models

Two random-effects models were considered (Philibert et al., 2012b). Both used an exponential function to relate N₂O to applied N and both assumed that the two parameters of this function varied randomly between locations. The two models used the same equation but differed in terms of the method used to estimate model parameters. One model was a frequentist method based on an approximate maximum likelihood method, whereas the parameters of the second model were estimated by a Bayesian method.

Both models are based on the following equations:

$$Y_{ijk} = \exp\left(\alpha_{0i} + \alpha_{1i}X_{ij}\right) + \varepsilon_{ijk} \tag{1}$$

with $\varepsilon_{ijk} \sim N\left(0, \tau^2\right)$, $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$ and $\alpha_{1i} \sim N\left(\mu_1, \sigma_1^2\right)$.

Y_{ijk} is the N₂O emission (kg N ha^{-1} yr^{-1}) measured at the ith location (confounded with the publication) (i=1 ... 203), for the jth applied N dose (j=1 ... N_i), and the kth replicate (k=1 ... K_{ij}). μ_0 is the log mean background emission, α_{0i} is the log location-specific background emission (random), μ_1 is the log mean applied N effect, α_{1i} is the log location-specific applied N effect (random), and ε_{ijk} is the residual error term. The random terms α_{0i}, α_{1i} and ε_{ijk} were assumed to be independent and normally distributed.

In the first model, the values of the μ_0, μ_1, σ_0, σ_1, and τ parameters were estimated by an approximate maximum likelihood method, with the nlme R statistical package (Pinheiro and Bates, 2000). In the second model, the parameters were estimated by a Bayesian method implemented with a Markov chain Monte Carlo algorithm (MCMC). In this second model, τ, σ_0 and σ_1 had uniform and independent prior probability distributions — τ, σ_0, $\sigma_1 \sim U(0,100)$ — whereas μ_0 and μ_1 had normal and independent prior probability distributions: μ_0, $\mu_1 \sim N(0,1000)$. The posterior distributions of the parameters were calculated with WinBUGS software (Lunn et al., 2000), with three chains of 100,000 MCMC iterations (see Annex 2). Convergence was checked with the Gelman-Rubin method (Brooks and Gelman, 1998). Random-effects models were compared to fixed-effects models using AIC (Akaïke Information Criterion), BIC (Bayesian Information Criterion), and DIC (Deviation Information Criterion). All criteria were lower for the random-effects models and showed that the random-effects model performed better.

Two types of N₂O prediction can be performed with each of these two models: average predictions and location-specific predictions. Average predictions are expressed as:

$$Y_{avg}(X) = \exp\left(\hat{\mu}_0 + \hat{\mu}_1 X\right) \tag{2}$$

where $\hat{\mu}_0$ and $\hat{\mu}_1$ are the values of μ_0 and μ_1 estimated by frequentist or Bayesian methods (posterior means).

Table 4.1 : Root mean square error of location-specific N_2O predictions (RMSEP) (kg N ha^{-1} year^{-1}) calculated for the frequentist model. RMSEP was calculated separately for different groups of locations with different numbers of applied N doses per location. The percentage decrease/increase in RMSEP with respect to average predictions is shown in brackets.

Number of N fertilizer doses per location used for location-specific predictions	RMSEP (kg N ha^{-1} year^{-1})	Number of predictions
1	4.77 (9.15)	140
2	2.71 (-8.45)	108
3, 4, or 5	3.77 (-7.82)	129
6 or more	1.38 (-24.59)	75

Table 4.2 : Root mean square error of location-specific N_2O predictions (RMSEP) (kg N ha^{-1} year^{-1}) calculated for the frequentist model. RMSEP was calculated separately for different groups of locations with different ranges of N fertilizer doses. The percentage decrease/increase in RMSEP with respect to average predictions is shown in brackets.

Ranges of N fertilizer dose (kg N ha^{-1})	RMSEP (kg N ha^{-1} year^{-1})	Number of predictions
0	0.97 (-5.83)	99
(0; 100]	2.00 (-13.79)	123
(100; 200]	1.78 (-30.47)	124
(200; 500]	6.85 (5.71)	106

Location-specific predictions are calculated from one or several location-specific N_2O measurements, as follows:

$$Y_{loc}(X) = \exp(\hat{\alpha}_{0loc} + \hat{\alpha}_{1loc}X) \tag{3}$$

where $\hat{\alpha}_{0loc}$ and $\hat{\alpha}_{1loc}$ are the values of α_{0i} and α_{1i} estimated for the locations considered, from one or several N_2O measurements obtained at the location concerned. In the first model, $\hat{\alpha}_{0loc}$ and $\hat{\alpha}_{1loc}$ are equal to the best linear unbiased prediction (BLUP) of α_{0i} and α_{1i} (Pinheiro and Bates, 2000). In the second model, $\hat{\alpha}_{0loc}$ and $\hat{\alpha}_{1loc}$ are equal to the estimated expected values of the posterior distribution. Note that, in both cases, $\hat{\alpha}_{0loc}$ and $\hat{\alpha}_{1loc}$ can be calculated if at least one N_2O value for one N fertilizer dose is available for the location considered.

2.3. Evaluation of model predictions with real data

Cross-validation was used to assess the prediction capacity of the frequentist random-effect model. Cross-validation was performed as follows:

- Step i: A location/dose combination was chosen and the corresponding data were removed from the dataset (*i.e.* all measurements of N_2O emission corresponding to a given amount of applied N at a given location),

- Step ii: The parameters of equation (1) were estimated with nlme, without the data excluded at step i,

- Step iii: The values of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\alpha}_{0loc}$ and $\hat{\alpha}_{1loc}$ obtained in step ii were used to calculate average and locations-specific predictions (Eqs.(2-3)) for the location/dose combination for which the data were removed in step i.

We did not considered locations for which data were available for only one amount of applied N, because location-specific predictions could not be calculated for such locations by the cross-validation procedure described above.

Prediction accuracy was measured by calculating the root mean squared error of prediction (RMSEP) for both average and location-specific predictions. RMSEP was calculated separately, for four groups of locations differing in terms of the number of N fertilizer doses tested per location (Table 4.1). RMSEP values were also calculated for several ranges of N fertilizer doses used to predict N_2O emission (Table 4.2).

The procedure described above was not used in the Bayesian model, due to the long computation times required (1200 seconds for each prediction). Predictions for a small

Table 4.3 : Description of the three groups of locations used to evaluate prediction accuracy with real data

Group	Number of locations	Number of data available per location	Number of data used for location-specific predictions	Range of N fertilizer doses considered per prediction (kg N ha^{-1})
1	9	7 to 13	6 to 12	0 - 450
2	10	4	3	0 - 410
3	10	2	1	0 - 336

number of location/dose combinations were calculated and evaluated, for comparison of the predictions of Bayesian and frequentist random-effects models. Three groups of locations extracted from the dataset of Stehfest and Bouwman (2006) were considered successively (Table 4.3):

- Group 1: nine locations for which N$_2$O data were obtained for at least seven different N fertilizer doses

- Group 2: 10 locations for which N$_2$O data were obtained for four different N fertilizer doses

- Group 3: 10 locations for which N$_2$O data were obtained for two different N fertilizer doses

The range of amounts of fertilizer applied was quite large in each of the various groups (Table 4.3). A cross-validation procedure was applied for each of these three groups. Average and location-specific predictions were obtained for each location in each group, for two to four different amounts of applied N. In group 1, four different N fertilizer doses were selected successively: the minimum and maximum N doses at each location, and two N doses close to the first and third quartiles. Only two N doses were selected in groups 2 and 3: the minimum and maximum N doses at each location. RMSEP was calculated for each group of locations separately, and a specific RMSEP was also calculated for N$_2$O predictions for N dose=0.

2.4. Evaluation of model predictions with simulated data

Model predictions were also assessed with simulated data, to investigate more precisely the relationship between dataset characteristics and prediction accuracy. We assessed the influence of three different factors: (i) the size of the dataset used to estimate the probability distribution of the random parameters of the model, (ii) the number of N$_2$O measurements used to obtain location-specific predictions and (iii) the fertilizer dose considered for N$_2$O prediction.

Our procedure involved the generation of data for locations with a random-effects model. Each location i was characterized by specific values of α_{0i} and α_{1i} randomly drawn from two normal distributions, $\alpha_{0i} \sim N(\mu_0, \sigma_0^2)$ and $\alpha_{1i} \sim N(\mu_1, \sigma_1^2)$. N$_2$O data were then generated with Eq.(1) for five different amounts of applied N: 0, 50, 100, 200 and 300 kg N ha^{-1}. This procedure was implemented with the parameter values estimated from the dataset of Stehfest and Bouwman (2006) with the R function nlme; $\mu_0 = 0.1932$, $\mu_1 = 0.0037$, $\sigma_0 = 0.7194$, $\sigma_1 = 0.0025$ and $\tau = 1.9364$. These parameter values were considered to be the true parameter values and were used for data generation.
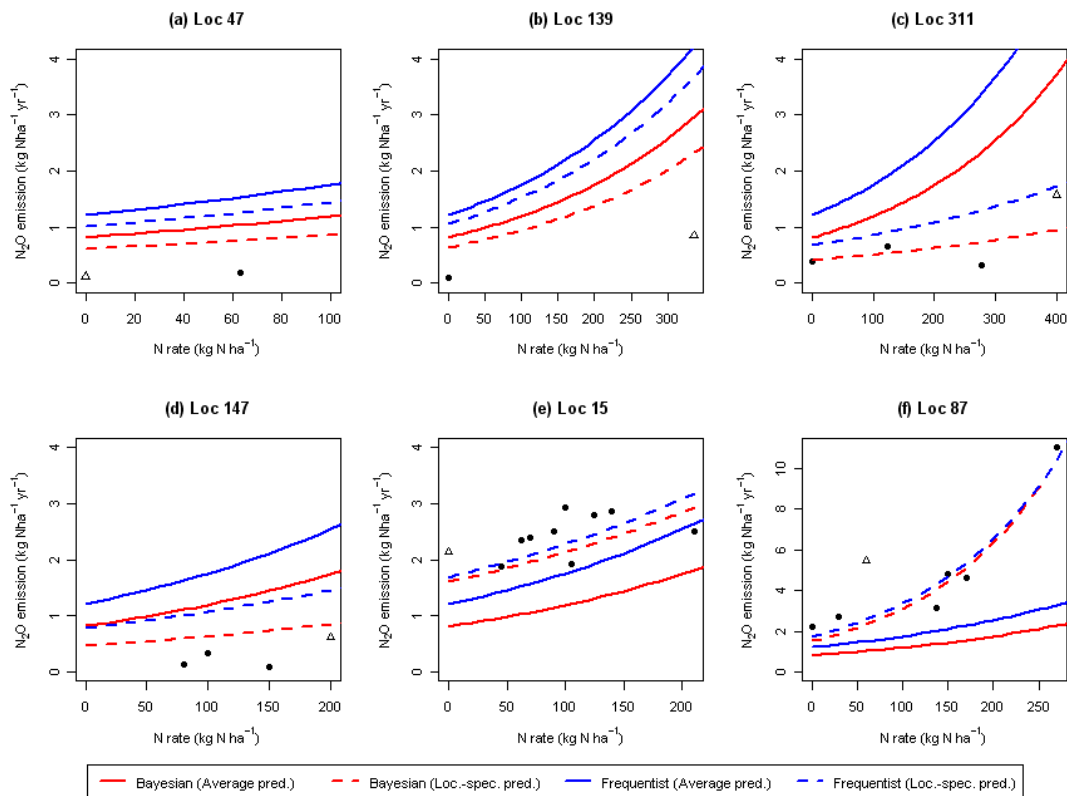
Figure 4.1: Predicted response curves for six locations for which real data were available. Average predictions (solid lines) and location-specific predictions (dashed lines) were calculated with the Bayesian model (red lines) and the frequentist model (blue lines). Black points indicate the N₂O emission data used to fit the model. Triangles indicate independent data not used for parameter estimation.
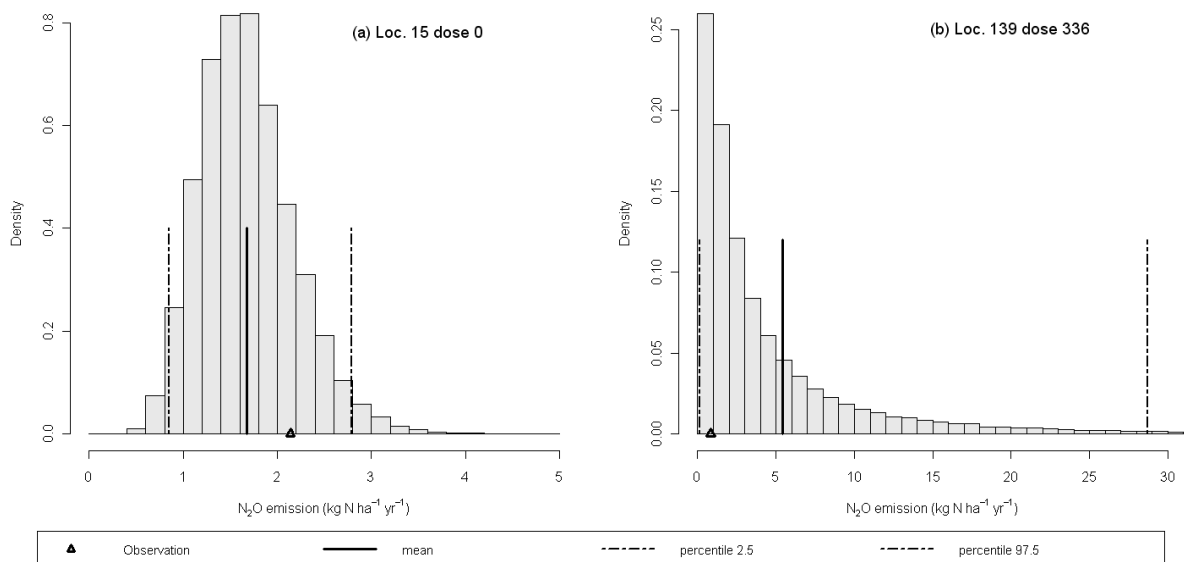
Figure 4.2: Posterior probability distribution for N₂O emission obtained with the Bayesian model for two locations (loc. 15 and loc. 139) and two N fertilizer doses (0 and 336 kg ha⁻¹). Measured N₂O emission values (triangle) and distribution means (solid black line), percentiles 2.5 and 97.5 (dotted black lines) are shown.

Six hundred datasets including M = 200, 300, 400 or 500 locations were generated successively. Each dataset contained (M – 1) "complete locations", each including N₂O data for five different N fertilizer doses (0, 50, 100, 200 and 300 kg N ha^{-1}), and an additional location with "missing data" including N₂O measurements for either one N fertilizer dose (100 kg N ha^{-1}) or three doses (50, 100, and 200 kg N ha^{-1}). The parameters of the random-effects model (Eq.(1)) were estimated with the nlme R statistical package for each dataset in turn, and N₂O predictions were derived for the location with "missing data" with Eq.(2) (average prediction) and Eq.(3) (location-specific prediction), for two different N fertilizer doses, X_p=0 and X_p=300 kg N ha^{-1}. The accuracy of N₂O prediction was assessed by calculating RMSEP with N₂O data generated with the true parameter values. The R code with simulated data is presented in Annex 5.

3.   Results

3.1. Examples of the predictions obtained with the models

The location-specific and average predictions obtained with the two models are compared with real data for six locations in Figure 4.1. Location-specific predictions were derived from a single item of data (black points) at the first two locations, from three items of data for the next two locations and from six or more items of data for the last two locations. Location-specific predictions were closer than average predictions to the actual data, indicated by triangles. As these data were not used to obtain location-specific predictions, location-specific predictions were more accurate than average predictions.

The predictions obtained with the frequentist model were slightly higher than those obtained with the Bayesian model (Figure 4.1). An interesting feature of the Bayesian model is illustrated in Figure 4.2. This figure shows two histograms representing posterior probability distributions for N₂O emission calculated with the Bayesian model for two different locations and two different N fertilizer doses. These distributions describe the uncertainty associated with model predictions. They can be used to compute N₂O percentiles and credibility intervals. Note that the N₂O distributions calculated with the Bayesian model may diverge markedly from a Gaussian distribution. One of the two distributions displayed in Figure 4.2 (obtained for a dose of 336 kg N ha^{-1}) is asymmetric, has a long tail and was close to the observed emission; the 97.5% percentile of this distribution is 28.69 kg N ha^{-1} year^{-1}, whereas the mean is only 5.4 N ha^{-1} year^{-1} (Figure 4.2b). The observed N₂O values were in the 95% credibility interval and were close to the means shown on the histograms.

Table 4.4 : Root mean square error of N_2O predictions (RMSEP) (kg N ha^{-1} year^{-1}) calculated for the Bayesian and frequentist models on the basis of average predictions (Avg. pred.) or location-specific predictions (Loc. spec. pred.). RMSEP was calculated separately for three groups of locations for which at least 7 (group 1), 4 (group 2), or no more than one item (group 3) of data were available. We calculated RMSEP both averaged over all the N fertilizer doses (1-all, 2-all, 3-all) considered, and for the application of zero N only (1-dose 0, 2-dose 0, 3-dose 0). Relative differences in RMSEP between location-specific predictions and average predictions are expressed as percentages.

Group	RMSEP for the Bayesian model			RMSEP for the frequentist model			Number of predictions
	Avg. pred.	Loc. spec. pred.	Difference (%)	Avg. pred.	Loc. spec. pred.	Difference (%)	
1-all	2.30	1.50	-35.78	2.10	1.54	-26.67	36
1-dose 0	0.96	0.92	-4.17	1.00	1.07	7.00	6
2-all	4.82	1.29	-73.24	4.75	1.14	-76.00	20
2-dose 0	0.77	0.59	-23.38	0.76	0.63	-17.11	6
3-all	1.95	1.78	-8.72	1.90	1.64	-13.68	20
3-dose 0	1.04	0.83	-20.19	1.10	0.96	-12.73	10

3.2. Evaluation of model predictions with real data

Tables 4.1-4.2 show RMSEP values for location-specific predictions for the frequentist model. RMSEP values varied from 1.38 to 4.77 kg N ha^{-1} year^{-1} (Table 4.1), depending on the number of N fertilizer doses for which data were available for the calculation of location-specific predictions. The use of location-specific rather than average predictions decreased the RMSEP by 8 to 25% for locations for which data were available for at least two different N fertilizer doses (Table 4.1). For locations with data for only one N fertilizer dose, RMSEP values were lower for average than for location-specific predictions (Table 4.1). Location-specific predictions were more accurate and gave lower RMSEP values for amounts of applied N between 0 and 200 kg N ha^{-1} (Table 4.2). For amounts of applied N exceeding 200 kg N ha^{-1}, RMSEP was 6% lower for average than for location-specific predictions (Table 4.2).

The average and location-specific predictions of the Bayesian and frequentist models were compared (Table 4.4) for a smaller number of location/dose combinations. Overall, for all the N fertilizer doses considered, location-specific predictions were more accurate than average predictions. The use of location-specific predictions instead of average predictions decreased RMSEP by 8.72 to 73.24% for the Bayesian model, and by 13.68% to 76% for the frequentist model. The smallest decrease in RMSEP was that for group 3, in which location-specific predictions were derived from a single measurement of N$_2$O emission; in this group, RMSEP was 9 to 14% lower for location-specific predictions than for average predictions.

For an N fertilizer dose of zero, RMSEP values were also lower for location-specific predictions than for average predictions. Group 1 was the only exception, but is the findings for this group were uncertain due to the limited number of data used for RMSEP calculation. We found no systematic difference between the RMSEP values obtained with the Bayesian and frequentist models. The frequentist model gave a lower RMSEP in some cases and a higher RMSEP in others (Table 4.4). The prediction accuracies of the two models were similar.

3.3. Evaluation of model predictions with simulated data

Tables 4.5-4.6 show the RMSEP values for location-specific predictions calculated for N fertilizer doses of 0 and 300 kg N ha^{-1}, respectively. RMSEP was systematically higher for the higher dose than for the lower dose. This is not particularly surprising, as the variability of N$_2$O emission data was greater for the higher dose than for the lower dose of N; for a dose of 0 kg N ha^{-1}, N$_2$O emission is not dependent on the random fertilizer effect α_{1i} (see Eq.(1)) and is therefore less variable. The number of locations, M, included in the dataset had no systematic

Table 4.5 : Root mean square error of location-specific prediction for a fertilizer dose of 0 kg N ha^{-1}. Location-specific predictions were derived for *M* locations, *M-1* ―complete locations" for which data were available for five fertilizer doses, and one location with ―missing data" for which data were available for *Q* fertilizer doses (100 kg N ha^{-1} for *Q*=1; 50, 100, and 200 kg N ha^{-1} for *Q*=3). Predictions were derived for the location with ―missing data" only. Each RMSEP (kg N ha^{-1} year^{-1}) was calculated from 600 simulated data. The percentage decrease in RMSEP reduction with respect to average predictions is shown in brackets.

Number of locations	RMSEP (kg N ha^{-1} year^{-1})	
	Q=1	*Q*=3
M=200	2.14 (-7.36)	2.16 (-6.09)
M=300	2.13 (-4.05)	2.16 (-2.70)
M=400	2.16 (-6.09)	2.18 (-5.22)
M=500	2.16 (-7.69)	2.17 (-7.26)

Table 4.6 : Root mean square error of location-specific predictions for a fertilizer dose of 300 kg N ha^{-1}. Location-specific predictions were derived for *M* locations, *M-1* ―complete locations" for which data were available for five fertilizer doses, and one location with ―missing data" for which data were available for *Q* fertilizer doses (100 kg N ha^{-1} for *Q*=1; 50, 100, and 200 kg N ha^{-1} for *Q*=3). Predictions were derived for the location with ―missing data" only. Each RMSEP (kg N ha^{-1} year^{-1}) was calculated from 600 simulated data. The percentage decrease in RMSEP with respect to average predictions is shown in brackets.

Number of locations	RMSEP (kg N ha^{-1} year^{-1})	
	Q=1	*Q*=3
M=200	6.13 (-24.79)	3.66 (-55.09)
M=300	6.66 (-21.18)	3.98 (-52.90)
M=400	9.02 (-28.75)	3.71 (-70.74)
M=500	6.39 (-21.69)	4.04 (-50.49)

effect on RMSEP and, thus, on prediction accuracy; no clear relationship was found between M and RMSEP (Tables 4.5 and 4.6).

For an N fertilizer dose of 0 kg N ha^{-1}, the RMSEP values obtained with $Q=1$ (one piece of data for the location with ‒missing data") and with $Q=3$ (three data for the location with ‒missing data") were similar (Table 4.5); RMSEP ranged from 2.13 to 2.16 kg N ha^{-1} year^1 for $Q=1$, and from 2.16 to 2.18 kg N ha^{-1} year^{-1} for $Q=3$. Thus, the use of two additional pieces of information for N fertilizer dose=50 kg ha^{-1} and for N fertilizer dose=200 kg ha^{-1} ($Q=3$ instead of $Q=1$) did not improve predictions of N₂O emission at an N fertilizer dose of 0 kg N ha^{-1}. By contrast, for a fertilizer dose of 300 kg N ha^{-1}, the RMSEP values obtained with $Q=1$ were much higher than those obtained with $Q=3$ (Table 4.6). The use of two additional items of data greatly improved predictions of N₂O emission for an N fertilizer dose of 300 kg N ha^{-1}. This is because N₂O emission data were generated for an N fertilizer dose of 100 kg N ha^{-1} for $Q=1$, N; this amount of fertilizer is much smaller than 300 kg N ha^{-1} and the addition of information for a fertilizer dose of 200 kg N ha^{-1}, which is closer to 300 kg N ha^{-1}, improved the prediction of N₂O emission at an N fertilizer dose of 300 kg N ha^{-1}.

Location-specific predictions were more accurate than average predictions (Tables 4.5 and 4.6), giving a 3 to 8% lower RMSEP for an N fertilizer dose of 0 kg N ha^{-1} and a 21 to 71% lower RMSEP for an N fertilizer dose of 300 kg N ha^{-1}. The improvement due to the use of location-specific predictions was thus greater for applications of large amounts of N fertilizer than for the application of 0 kg N ha^{-1}.

4. Discussion

The two random-effects models presented here can be used to calculate average and location-specific predictions (derived from one or several location-specific N₂O emission measurements). Location-specific predictions can be calculated even if only a few N₂O measurements are available. Our models predicted the response of N₂O emission to the amount of applied N and the predicted response can be adjusted on the basis of even a single N₂O measurement. This opens up several interesting possibilities in terms of practical applications. For instance, our models could be used in farmers' fields in which N₂O emissions have been determined for the amounts of N fertilizer applied by the farmers. They could be used to predict emissions for other amounts of N fertilizer, including a total absence of N fertilizer application, to predict background N₂O emissions. Such fields, as considered by Gu et al. (2011), for example, could be incorporated into national inventories and taken into account for the calculation of specific emission factors (IPCC, 2006). Our models could also be used at locations for which N₂O emission has been measured for a limited number of N fertilizer doses (*e.g.*, one, two or three doses), to predict the full curve for the response of

N$_2$O emission to applied N. Examples of predicted response curves for locations for which N$_2$O emission has been measured for a single fertilizer dose are shown in Figure 4.1. Finally, our models could be used to estimate missing data, such as N$_2$O emissions not measured for some N fertilizer treatments due to technical issues (*e.g.* closed chambers damaged by management practices, Osborne et al., 2010).

We found, with both models, that location-specific predictions were more accurate than average predictions, with two exceptions: locations with N$_2$O data for a single dose of N fertilizer and N$_2$O predictions for amounts of N fertilizer exceeding 200 kg N ha^{-1}. This result was obtained with real data but not with simulated data, for which the use of location-specific prediction systematically resulted in an improvement. With real data, we found that these two exceptions were caused by only two data for single dose predictions and by only four data for N fertilizer doses exceeding 200 kg N ha^{-1}. It indicates that location-specific predictions are not systematically more accurate, particularly if data are available for only one dose of applied N.

The overall benefit gained from the use of location-specific rather than average predictions results from their adjustment to local environmental characteristics through the use of location-specific N$_2$O measurements. When possible, it is thus useful to measure N$_2$O emission for at least one N fertilizer dose or more at the location of interest, to facilitate the prediction of emissions for other amounts of applied N at the same location.

The frequentist and Bayesian models presented here have similar levels of accuracy. However, both have advantages and disadvantages. The frequentist model has a very short computation time but provides point-value predictions only, and there is no obvious technique for determining the level of uncertainty associated with its predictions. The Bayesian model requires longer computation times (*e.g.*, 1200 seconds were required for each prediction with the Bayesian model, and this must be multiplied by the number of predictions for calculation of the RMSEP reported in Table 4.4, whereas only 30 seconds were required for the calculation of each RMSEP with the frequentist model, with a 3.19 GHz processor). However, the Bayesian model is not restricted to point-value prediction, but can also be used to generate a probability distribution for N$_2$O emission values that can be used to derive credibility intervals.

The models were evaluated with both real and simulated data. Simulated data are not often used for the assessment of models in agronomy. This technique is useful for studying the relationship between experimental design and model performance. It was used by Makowski and Wallach (2001) to determine the number of experiments required to optimize N fertilizer dose on the basis of yield-response models. We used simulated data to assess model predictions for two different doses (0 and 300 kg N ha^{-1}), as a function of the number of N$_2$O emission measurements available. We showed that the use of location-specific predictions

rather than average predictions reduced prediction errors in all the cases considered, but that the extent of this reduction depended on the amount of fertilizer applied and the number of measurements used to predict N_2O emission.

5.   Conclusion

We show here that random-effects models are simple and useful tools for predicting N_2O emissions. They can be used to obtain both average and location-specific predictions, by taking into account one or several location-specific N_2O measurements. We showed that the use of location-specific rather than average predictions decreased prediction errors in most cases. Location-specific predictions could be used in various applications, particularly for estimating background emission in on-farm studies and for estimating the values of missing data.

Acknowledgements

« Je déteste les discussions, elles vous font parfois changer d'avis. »

Oscar Wilde

Discussion générale

L'objectif de cette thèse était de montrer le rôle important joué par l'étude de la robustesse des conclusions issues de méta-analyses et de l'illustrer au travers d'une application : l'estimation des émissions de N_2O provenant des sols agricoles, en mobilisant la base de données constituée par Stehfest et Bouwman (2006). A notre connaissance aucune analyse de sensibilité n'a été réalisée pour étudier la robustesse des estimations d'émission de N_2O obtenues à partir de cette base de données. Le travail réalisé par Stehfest et Bouwman (2006) a pourtant joué un rôle clé puisqu'il a permis de fixer le facteur d'émission du niveau 1 du GIEC à 1%, largement utilisé aujourd'hui: en effet, les pays utilisant ce facteur d'émission pour estimer leurs émissions de N_2O provenant des sols agricoles représentent 56% des pays développés de la convention-cadre des Nations Unies sur les changements climatiques (CCNUCC / UNFCCC) (Lokupitiya et Paustian, 2006) et la moitié des publications sur les inventaires de N_2O (Berdanier et Conant, 2012).

Mon travail a fourni des résultats à deux niveaux, relatifs à (i) l'estimation des émissions de N_2O issues des sols agricoles et à (ii) la mise en place de critères de qualité pour évaluer les méta-analyses.

1.   Retour sur le cas d'étude : estimation des émissions de N_2O issues des sols agricoles

Ce mémoire de thèse met en évidence un résultat majeur en termes de choix de modèle, en montrant la très bonne adéquation des modèles à effets aléatoires pour estimer des émissions de N_2O dans le cadre d'une méta-analyse. Les modèles à effets aléatoires apparaissent être bien adaptés à la structure de nos données (à savoir des observations répétées sur un même site pour différentes doses d'engrais, et différents sites considérés). Ils sont sélectionnés comme étant les modèles s'ajustant le mieux aux données, par les critères statistiques AIC, BIC et DIC (chapitre 1 et 2), et ils peuvent être ajustés localement même lorsqu'un petit nombre de mesures est disponible (chapitre 4). Ces modèles sont aussi plus précis que le modèle linéaire avec log transformation de la variable de N_2O présenté par Stehfest et Bouwman (2006) (chapitre 3). Un autre intérêt des modèles à effets aléatoires est qu'ils décrivent la variabilité inter-site des émissions de N_2O à l'aide de distributions de probabilité (chapitre 1, 2 et 4). Au vu de ces nombreux intérêts démontrés dans ce mémoire, l'utilisation de modèles à effets aléatoires ne peut être que fortement conseillée pour l'analyse de base de données d'émissions de N_2O à large échelle.

La relation entre les émissions de N_2O et son principal facteur reconnu, la dose d'engrais azoté a fait l'objet de nombreuses publications. Une fonction exponentielle a été proposée dans plusieurs publications sur la base d'expérimentations locales (Hoben et al., 2011 ; Ma et

al., 2010). Dans ce mémoire, l'analyse de la forme de la relation « Emission de N_2O – Dose d'engrais N » à l'aide d'une base de données globale a confirmé que la relation exponentielle est plus adaptée. Il apparaitrait ainsi souhaitable de remplacer le facteur d'émission constant du GIEC (1% d'émission quelque soit la dose d'engrais N) par un facteur variable qui augmenterait en fonction de la dose. Cette nouvelle approche conduirait à une diminution des estimations d'émission de N_2O pour des doses d'engrais inférieures à 160 kg N ha^{-1} mais augmenterait les émissions pour des doses supérieures.

L'estimation des émissions de N_2O et de ses incertitudes pour différentes cultures de légumineuses à partir de la base de données de Rochette et Janzen (2005) a montré qu'une différence d'émissions existe probablement entre espèces (trèfle vs. luzerne). Cette différence d'émission a cependant été mise en évidence à partir d'un jeu de données de taille réduite. Il serait utile de réaliser des expérimentations supplémentaires pour estimer plus précisément les émissions de N_2O de différentes espèces de légumineuses.

L'analyse d'incertitude des émissions de N_2O a été réalisée sur les estimations d'émissions moyennes (chapitre 2). Comme elle ne tient pas compte de la variabilité inter-expérimentation, elle a été conçue pour être utilisée dans le cadre du niveau 1 du GIEC. Une originalité de notre travail est que l'intervalle d'incertitude que nous proposons est basé sur un ensemble de modèles plutôt que sur un modèle unique. Cet intervalle inclut l'incertitude associée au choix de modèle et permet d'analyser la sensibilité des estimations au modèle choisi. La borne inférieure de cet intervalle est très proche de celle proposée par le GIEC tandis que la borne supérieure est nettement inférieure à celle du GIEC, réduisant ainsi la gamme d'incertitude par rapport à celle proposée par le GIEC.

Nous avons testé l'apport de variables explicatives additionnelles à celle de la dose d'engrais azoté dans les modèles de régression utilisés pour les estimations de N_2O au niveau 1 du GIEC. Cette approche permet de passer aux niveaux 2 et 3 du GIEC. Le principal obstacle à cet ajout de variables explicatives est le nombre important de données manquantes pour ces variables décrivant le milieu (climat, sol) et les pratiques culturales. Ce problème est lié au mode de construction de la base de données. Cette base de données recense les mesures de N_2O publiées, elle est donc dépendante des données disponibles dans les articles. Comme le soulignent Freibauer et Kaltschmitt (2003), les variables caractérisant les pratiques agricoles et le milieu ne sont pas systématiquement mises à disposition dans les articles et ces auteurs proposent un cadre minimal d'informations à publier sur les caractéristiques du site d'expérimentation ainsi que des pratiques agricoles. L'intégration de ces variables dans les modèles du chapitre 2 ne peut être réalisée qu'au prix d'une réduction de la taille de la base de données de manière à éliminer les données manquantes. Il devient alors difficile d'estimer les paramètres des modèles avec suffisamment de précision.

Pour palier à ce problème, nous avons proposé deux solutions. La première consiste à utiliser la méthode d'apprentissage automatique *Random Forest*, qui gère les données manquantes et peut prendre en compte un très grand nombre de variables en construisant un grand nombre d'arbres de régression. Cette méthode permet de gérer les données manquantes et de conserver la totalité des expérimentations, même celles qui sont incomplètes. Nos résultats montrent par ailleurs que *Random Forest* conduit à des prédictions plus précises que les modèles de régression (chapitre 3). *Random Forest* nous a également permis d'identifier les trois variables les plus importantes pour estimer les émissions de N_2O : la dose de fertilisation azotée, le type de culture ainsi que la durée de l'expérimentation.

La deuxième solution est moins intuitive : elle consiste à prendre en compte les variables décrivant le milieu et les pratiques agricoles de manière indirecte, par le biais d'une ou plusieurs mesures d'émission de N_2O disponibles pour un site donné grâce au modèle à effets aléatoires (chapitre 4). Cette approche permet de prédire des émissions de N_2O à un niveau local sans connaître les caractéristiques du site donné. De telles prédictions peuvent être utilisées soit dans des expérimentations pour estimer les émissions de N_2O pour des doses d'engrais azoté non testées, soit en parcelles agricoles pour estimer les émissions pour des doses différentes de celles appliquées par les agriculteurs (ex : émissions de N_2O issues d'un sol non fertilisé).

En alternative aux modèles statistiques ou à l'apprentissage automatique présentés dans ce mémoire de thèse, les modèles de culture mécanistes peuvent également être mobilisés pour estimer les émissions de N_2O. Ils permettent en effet de modéliser le processus sol-plante-atmosphère à partir d'hypothèses formulées sur le fonctionnement de ce système. Ces modèles peuvent être utilisés pour estimer les émissions de N_2O au niveau 3 du GIEC. Ils présentent toutefois l'inconvénient de demander des temps de calcul assez importants dus à leur complexité et requièrent un nombre élevé de variables d'entrée. Nous présentons ici des méthodes à temps de calcul très court et facile à mettre en œuvre avec les logiciels gratuits R et Winbugs. D'après un article récent de Villa-Vialaneix (2012) la méthode de *Random Forest* serait également d'un grand intérêt pour « méta-modéliser » l'un des modèles de cultures mécaniste le plus utilisé pour estimer les émissions de N_2O, le modèle DNDC (Leip et al., 2008 ; Li et al., 1992 ; Li, 2000 ; Follador et Leip, 2009). A cause des temps de calculs très importants, les auteurs proposent de comparer plusieurs méthodes permettant de « méta-modéliser » le modèle DNDC, c'est-à-dire d'estimer les sorties de DNDC avec des modèles plus simples. Sur les huit méthodes testées, c'est la méthode de *Random Forest* qui présente le temps de calcul le plus court et qui fournit une précision d'un niveau acceptable.

2.   Les critères de qualité des méta-analyses proposés et testés

Dans ce mémoire, nous avons proposé huit critères permettant d'évaluer la qualité des méta-analyses effectuées en agronomie relatifs (i) à la récupération des données (procédure répétable, étude du biais de publication), (ii) au traitement des données (pondération, étude de l'hétérogénéité, analyse de sensibilité) et (iii) à la reproductibilité du processus entier (disponibilité des données et du programme ainsi que de la liste des références utilisées) (chapitre 1). En lien avec le critère de « mise à disposition des bases de données, des programmes et des références» nous avons mis en ligne sur le site internet de l'unité (http://www5.versailles-grignon.inra.fr/agronomie/Meta-analysis-in-agronomy) une page web recensant les méta-analyses présentées dans le chapitre 1 ainsi que les liens sur les bases de données et les programmes. Ce site a pour vocation de permettre aux agronomes de (i) faire une nouvelle méta-analyse en utilisant les mêmes données, (ii) vérifier les résultats de méta-analyses publiées, et (iii) compléter un jeu de données disponible pour le mettre à jour et actualiser les résultats d'une méta-analyse. Un des objectifs de ce site est de réduire le temps de récupération des données qui est le principal facteur limitant de la mise en place d'une méta-analyse.

A l'heure actuelle, la création d'une base de données pour effectuer une méta-analyse requiert un temps considérable excepté pour des problématiques où très peu de publications sont disponibles. La démarche la plus fréquemment décrite dans les papiers de méta-analyses en agronomie recensés dans le chapitre 1 consiste en quatre étapes. Premièrement une recherche par mot-clé sur les plateformes de recherche est effectuée ce qui sélectionne un nombre d'articles souvent très important (par exemple pour l'équation de recherche *"nitrous oxide" AND crop** sur la plateforme Web of Knowledge, nous obtenons 2279 articles au 24 septembre 2012). Puis tous les titres et résumés sont lus pour écarter les « mauvais » articles et en retenir un plus petit nombre. L'intégralité des articles retenus est ensuite lue pour ne garder que ceux où les données sur le sujet considéré sont disponibles. Enfin, les données sont extraites des articles afin de créer la base de données finale. Une fois la base de données créée, le traitement des données demande beaucoup moins de temps quelque soit la taille du jeu de données, que la collecte des données. En collaboration avec le département MMIP d'AgroParisTech (dans le cadre du Projet *ExtraEx* avec Antoine Cornuéjols, Antonin Duroy, Chantal Loyce, David Makowski, Christine Martin et Aurore Philibert), nous travaillons sur la mise en place d'un outil informatique qui permettrait d'automatiser au maximum les recherches et les tris d'articles scientifiques pour de nouveaux sujets de méta-analyse. Le principe est de fournir à l'outil un échantillon d'apprentissage constitué d'un petit nombre d'articles déjà trié. L'idée est d'utiliser cet échantillon pour établir des règles de classement qui pourront ensuite être appliquées à un nombre important de résumés et d'articles qui seront alors triés de manière automatique.
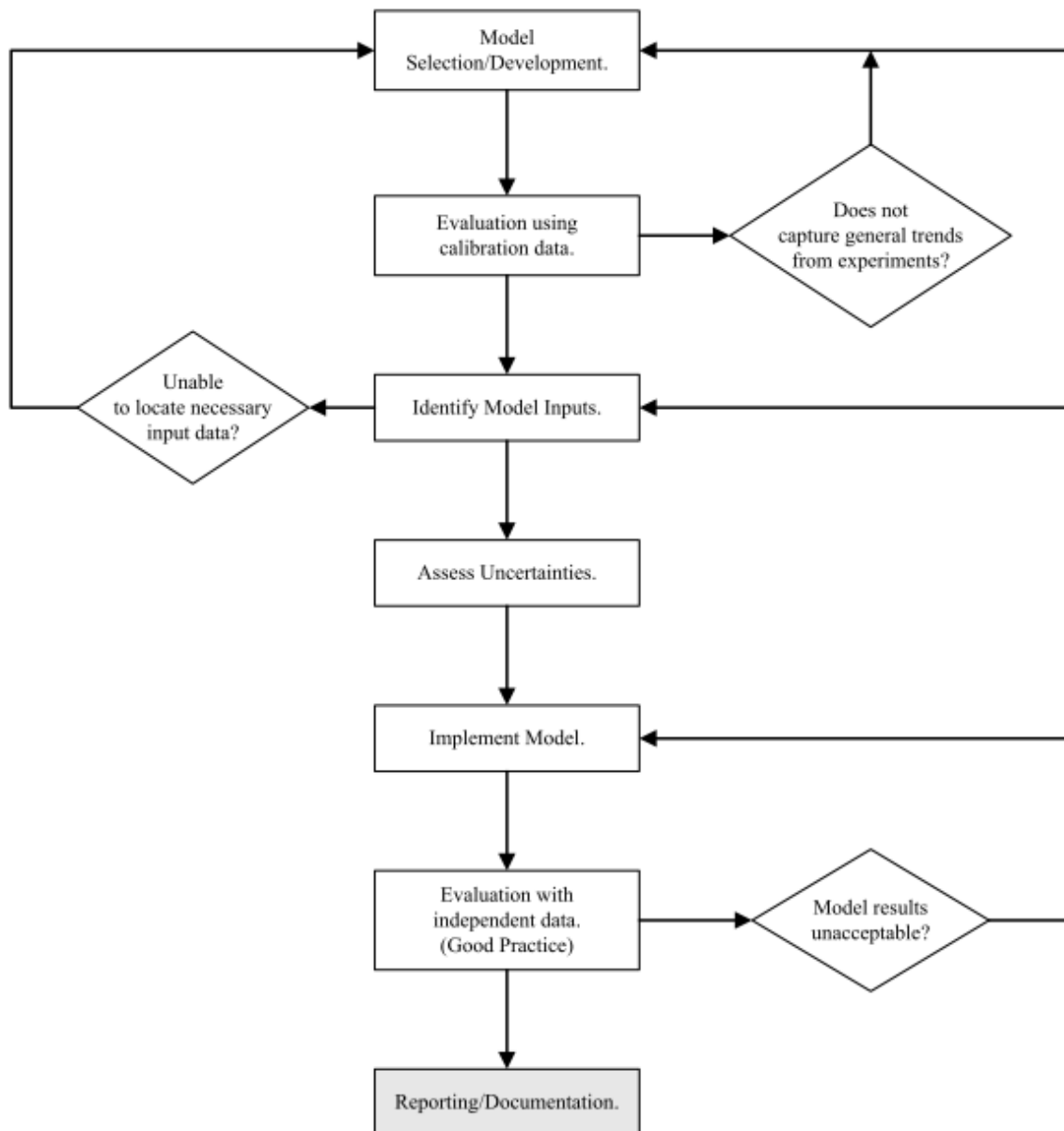
Encadré 3: Présentation des étapes à suivre pour mettre en œuvre le niveau 3 du GIEC afin d'estimer les émissions de N_2O (Source : IPCC, 2006)

En ce qui concerne les critères de qualité relatifs au traitement des données, l'utilisation de la pondération des observations individuelles est présentée uniquement dans le chapitre 1. La méthode la plus usuelle en méta-analyse (Hedges et Olkin, 1985) consiste à pondérer grâce à l'écart-type de la mesure individuelle. Dans le cadre des mesures d'émissions de N_2O les écart-types relatifs aux mesures sont rarement disponibles ce qui rend difficile une pondération par la précision de la mesure. L'utilisation de la durée de l'expérimentation comme substitut à cet écart-type n'est pas concluante : en effet, dans le chapitre 1, cette pondération n'apporte que très peu d'amélioration au modèle (selon les critères d'ajustement aux données AIC, BIC et DIC), tandis que dans le chapitre 2, les modèles sont apparus être moins bons (selon AIC, BIC et DIC) avec une pondération des observations.

Nous avons constaté dans le chapitre 1 que seul un faible nombre de méta-analyses en agronomie effectuaient une analyse de sensibilité des résultats. Le faible intérêt suscité par cette technique dans le cadre de la méta-analyse est probablement dû à une méconnaissance des techniques disponibles et de leur intérêt. Ce mémoire a illustré l'intérêt d'analyser la sensibilité des résultats d'une méta-analyse aux données et aux méthodes statistiques utilisées.

Pour l'analyse de sensibilité à la méthode statistique, les deux inférences (fréquentiste et Bayésienne) ont été comparées. Très peu de différences de résultats ressortent de la comparaison sans doute parce que nous avons utilisé des distributions *a priori* de paramètres peu informatives. L'inférence Bayésienne pourrait être utilisée en définissant des distributions *a priori* plus informatives en reprenant des données issues de la littérature ou bien par dires d'experts. Une comparaison entre modèle fréquentiste et modèle Bayésien utilisant ces distributions *a priori* permettrait de mieux valoriser les capacités de l'inférence Bayésienne. La similitude des résultats obtenus dans ce mémoire de thèse entre les deux inférences confirme cependant la robustesse des résultats vis-à-vis du type d'inférence. Dans le chapitre 1, pour le cas d'étude sur l'émission de N_2O par les légumineuses, la comparaison entre les deux méthodes d'inférence ne montre pas de différence notable. Dans le chapitre 2, le modèle Bayésien à effets aléatoires présente cependant une courbe d'émission avec des valeurs d'émission de N_2O légèrement inférieures à celles obtenues en fréquentiste. Par exemple pour une dose d'engrais azoté de 200 kg N ha^{-1}, on aboutit avec le modèle fréquentiste à une émission de N_2O de 2,53 kg N ha^{-1} an^{-1} tandis que le modèle Bayésien donne une valeur de 1,73 kg N ha^{-1} an^{-1}. Dans le chapitre 4, la différence se ressent plus particulièrement en termes de temps de calcul, beaucoup plus long pour le Bayésien que pour le fréquentiste. Toutefois, le modèle Bayésien fournit, en plus d'une prédiction ponctuelle d'émission de N_2O, une distribution de probabilité, très utile pour décrire l'incertitude.

3.   Perspectives

Les résultats du chapitre 2 ont montré que la forme de la relation « Emission de N$_2$O-Dose d'engrais N » la mieux ajustée aux données ne semblait pas correspondre à celle utilisée au niveau 1 du GIEC. Dans ce cadre, il serait intéressant d'analyser l'influence d'un facteur d'émission croissant avec la dose, comme définie au chapitre 2, sur les émissions de N$_2$O françaises. Ce travail pourrait être fait, par exemple en collaboration avec le centre interprofessionnel technique d'études de la pollution atmosphérique (CITEPA, http://www.citepa.org/fr/) chargé de la réalisation des inventaires d'émissions de polluants atmosphériques et de gaz à effet de serre en France. Actuellement, le CITEPA recense les données d'activités utilisées dans le niveau 1 du GIEC pour estimer les émissions de N$_2$O provenant des sols agricoles.

Les résultats obtenus avec la méthode d'apprentissage automatique *Random Forest* pourraient être utilisés dans le niveau 3 du GIEC dont le principe est résumé dans l'Encadré 3. Les analyses présentées au chapitre 3 ont rempli toutes les étapes d'acceptation du modèle excepté celles concernant l'acquisition de données d'activités spatio-temporelles ainsi que la mise en œuvre et la documentation du modèle (IPCC, 2006). Ces étapes doivent donc être effectuées pour pouvoir valider l'utilisation de *Random Forest* comme méthode de niveau 3. Toutefois, cette méthode d'apprentissage automatique présente déjà de réelles capacités de prédiction d'émissions de N$_2$O avec des variables explicatives nombreuses et avec des données manquantes.

Les résultats présentés dans ce mémoire posent la question de l'intérêt de complexifier les modèles d'estimation d'émissions de N$_2$O en utilisant d'autres variables explicatives que la dose d'engrais azoté. La réduction des erreurs moyennes d'estimation des émissions de N$_2$O peut être estimée à 0,91 kg N ha^{-1} an^{-1}. En effet l'écart-type résiduel du modèle constant est de 4,26 kg N ha^{-1} an^{-1} (chapitre 2, Table 2.3) tandis que la racine de l'erreur quadratique moyenne de prédiction (RMSEP) en validation croisée est de 3,35 kg N ha^{-1} an^{-1} pour le modèle à effets aléatoires avec la dose d'engrais azoté (chapitre 3, Table 3.3). Environ 21% ((4,26-3,35)/4,26) de la variabilité des émissions de N$_2$O est donc expliquée par la dose d'engrais azotée. Le meilleur des modèles utilisant des variables explicatives autres que la dose d'engrais azoté obtient un RMSEP de 2,97 kg N ha^{-1} an^{-1} (chapitre 3, Table 3.3) avec 14 variables explicatives. Ce modèle permet donc de réduire l'erreur moyenne de prédiction de 9% ((3,35-2,97)/4,26) par rapport au modèle n'incluant que la dose d'engrais, soit 0,38 kg N ha^{-1} an^{-1}. Le modèle à effet aléatoire nous fournit une autre information intéressante : la variabilité intra-site (écart-type intra-site) qui est égale à 1,94 kg N ha^{-1} an^{-1} et correspond à l'erreur restante si on suppose qu'on connait parfaitement la réponse des émissions de N$_2$O à la dose d'engrais pour chaque site d'expérimentation. Il est intéressant de remarquer que cette erreur résiduelle (1,94 kg N ha^{-1} an^{-1}) n'est pas tellement plus faible que celle du modèle

Random Forest incluant 14 variables explicative (2,97 N ha^{-1} an^{-1}). La différence est égale à 1,03 kg N ha^{-1} an^{-1} (24%, (2,97-1,94)/4,26). Elle indique que, par rapport à *Random Forest*, on ne peut pas espérer réduire de plus de 24% l'erreur de prédiction en améliorant la modélisation de la variabilité inter-sites. Le bénéfice potentiel n'est donc pas si grand.

De nombreux développements méthodologiques liés à l'analyse de sensibilité pourraient être illustrés dans le cadre de la méta-analyse. Deux perspectives me semblent intéressantes à étudier et concernent la gestion des données manquantes des variables explicatives. La première serait d'utiliser la méthode d'imputation multiple qui fonctionne en trois étapes : (i) remplacer l'ensemble des données manquantes *m* fois en tirant ces valeurs dans une distribution (qui peut être différente pour chaque observation manquante), (ii) faire une analyse pour chacune des *m* bases de données complètes générées, puis (iii) combiner les résultats des *m* analyses en un résultat final (Rubin, 1987). Cette méthode est de plus en plus souvent utilisée (http://www.multiple-imputation.com/) et est implémentée dans le package R mice.

La deuxième serait d'utiliser l'inférence Bayésienne pour remplacer les données manquantes. Il faut pour cela définir une distribution a priori pour la variable avec données manquantes (par exemple une loi de Bernouilli pour une variable binaire). Le modèle Bayésien utilise ensuite alternativement les données quand elles sont présentes ou bien tire dans cette loi *a priori* quand les données sont manquantes. Il serait intéressant d'évaluer et de comparer les ces deux approches.

Pour l'estimation des paramètres de nos modèles en inférence fréquentiste, c'est le package R nlme qui a été utilisé. Il serait intéressant de ré-estimer ces paramètres grâce au logiciel *Monolix* (http://software.monolix.org/sdoms/software/, MOdèles NOn LInéaires à effets miXtes) spécialement conçu pour les modèles non linéaires à effets mixtes. Cet outil permet une meilleure estimation des paramètres car il calcule l'estimateur du maximum de vraisemblance des paramètres sans faire d'approximation du modèle. Nous avons tenté de l'utiliser mais n'avons pas été capable d'obtenir la convergence dans tous les cas. Une étude plus approfondie de l'utilisation de *Monolix* sur des données d'émissions de N$_2$O permettrait de comparer les estimations obtenues avec celles présentées dans ce mémoire.

L'intérêt des modèles à effets aléatoires présentés dans le cadre des émissions de N$_2$O dans ce mémoire peut être généralisé au domaine de l'agronomie. La structuration des données dans la base de données de Stehfest et Bouwman (2006) est sous forme d'observations répétées sur un même site pour différentes doses d'engrais, et différents sites considérés. Cette structure est typique des données agronomiques et ne peut pas être résumé par un modèle moyen. L'apport des ces modèles est d'un grand intérêt pour toutes les problématiques agronomiques.

Un autre exemple de données agronomiques dans le cadre d'une méta-analyse (Miguez et al., 2008 ; Lesur et al., soumis) concerne des données de rendement de la plante pérenne à vocation énergétique *Miscanthus* x *giganteus* au cours du temps. Ces rendements proviennent de plusieurs expérimentations en Europe et l'utilisation de modèles à effets aléatoires permet comme dans le chapitre 4 d'avoir une courbe de rendement différente selon la localisation de l'expérimentation. Ces courbes de type « *site-spécifique* » pourraient permettre de prédire pour une plantation de *Miscanthus* x *giganteus* le rendement pour une année future grâce à des données de rendement disponibles pour quelques années.

Même si les résultats du chapitre 4 sont prometteurs, l'intérêt pratique de prédire des émissions de N_2O pour d'autres doses que celles appliquées dans l'expérimentation (ou par l'agriculteur) reste à tester. Dans cette optique, on pourrait utiliser une base de données d'émissions de N_2O mesurées en parcelles agricoles où une seule dose de fertilisation azotée est testée (Gu et al., 2012). Le fait de pouvoir prédire grâce à un modèle à effets aléatoires l'émission de N_2O en cas de dose nulle permettrait d'estimer le potentiel d'émission des sols sans témoin non fertilisé (Hénault et al., 2001).

Références

bibliographiques

Références

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.

Arnqvist G, Wooster D (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution* 10: 236-240.

Barbottin A, Makowski D, Le Bail M, Jeuffroy M, Bouchard C, Barrier C (2008) Comparison of models and indicators for categorizing soft wheat fields according to their grain protein contents. *European Journal of Agronomy* 29: 175-183.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *The Journal of the American Medical Association* 276: 637-639.

Berdanier AB, Conant RT (2012) Regionally differentiated estimates of croplands N_2O emissions reduce uncertainty in global calculations. *Global Change Biology* 18: 928-935.

Bland CJ, Meurer LN, Maldonado G (1995) A systematic approach to conducting a non-statistical meta-analysis of research literature. *Academic Medicine* 70: 642-653.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to Meta-Analysis, first ed. Wiley, Hoboken NJ.

Bouwman AF (1996) Direct emission of nitrous oxide from agricultural soils. *Nutrient Cycling in Agroecosystems* 46: 53-70.

Breiman L, Freidman J, Olshen R, Stone C (1984) Classification and regression trees. Belmont (CA): Wadsworth, 358 p.

Breiman L (2001) Random forests. *Machine Learning* 45: 15-32.

Breiman L, Cutler A (2003) Setting up, using, and understanding Random Forests v4.0. http://www.stat.berkeley.edu/~breiman/RandomForests/.

Brisson N, Gary C, Justes E, Roche R, Mary B, Ripoche D, Zimmer D, Sierra J, Bertuzzi P, Burger P, Bussière F, Cabidoche Y, Cellier P, Debaeke P, Gaudillère J, Hénault C, Maraux F, Seguin B, Sinoquet H (2003) An overview of the crop model STICS. *European Journal of Agronomy* 18: 309-332.

Britz W, Leip A (2009) Development of marginal emission factors for N losses from agricultural soils with the DNDC-CAPRI meta-model. *Agriculture Ecosystems & Environment* 133: 267-279.

Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7: 434-455.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: A practical information-theoretic approach, second ed. Springer-Verlag, New York.

Cadotte MW, Mehrkens LR, Menge DNL (2012) Gauging the impact of meta-analysis on ecology. *Evolutionary Ecology* 26: 1153-1167.

Cardinale BJ, Srivastava DS, Duffy JE, Wright JP, Downing AL, Sankaran M, Jouseau C (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* 443: 989-992.

Crutzen PJ, Mosier AR, Smith KA, Winiwarter W (2008) N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmospheric Chemistry and Physics* 8: 389-395.

Cucherat M, Boissel JP, Leizorovicz A (1997) Manuel pratique de méta-analyse des essais thérapeutiques. Available at http://www.spc.univ-lyon1.fr/livreMA/frame.htm.

Cutler A, Stevens J (2006) Random forests for microarrays. *Methods in Enzymology* 411: 422-432.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88: 2783-2792.

Davidian M, Giltinan DM (1995) Non linear mixed effect models for repeated measurement data. Chapman & Hall. 359p.

Davidson EA (2009) The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nature Geoscience* 2: 659-662.

Del Grosso SJ, Ojima DS, Parton WJ, Stehfest E, Heistemann M (2009) Global scale DAYCENT model analysis of greenhouse gas emissions and mitigation strategies for cropped soils. *Global and Planetary Change* 67: 44-50.

Del Grosso SJ, Ogle SM, Parton WJ, Breidt FJ (2010) Estimating uncertainty in N_2O emissions from US cropland soils. *Global Biogeochemical Cycles* 24: 12pp.

Doré T, Makowski D, Malézieux E, Munier-Jolain N, Tchamitchian M, Tittonell P (2011) Facing up to the paradigm of ecological intensification in agronomy: Revisiting methods, concepts and knowledge. *European Journal of Agronomy* 34: 197-210.

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science* 1: 54-75.

Efron B, Tibshirani R (1993) An Introduction to the Bootstrap (Chapman & Hall, London).

Egger M, Smith GD, Phillips AN (1997) Meta-analysis: principles and procedures. *British Medical Journal* 315: 1533−1537.

Evans JA, Foster JG (2011) Metaknowledge. *Science* 331: 721-725.

Fisher RA (1935). The design of experiments. Edinburgh: Oliver and Boyd.

Flynn HC, Smith J, Smith KA, Wright J, Smith P, Massheder J (2005) Climate- and crop-responsive emission factors significantly alter estimates of current and future nitrous oxide emissions from fertilizer use. *Global Change Biology* 11: 1522-1536.

Follador M, Leip A (2009) Derivation of DNDC Meta-models to Evaluate the Impact of Cross Compliance Measures on Nitrogen N Surplus, N Leaching, N_2O Emissions at European Scale. European Commission, D.G Joint Research Centre, Institute for Environment and Sustainability, Climate Change Unit, Ispra, Italy. Technical Report Deliverable 4.2.3 of EU STREP 44423-CCA.

Follador M, Leip A, Orlandini L (2011) Assessing the impact of Cross Compliance measures on nitrogen fluxes from European farmlands with DNDC-EUROPE. *Environmental pollution* 159: 3233-3242.

Food and Agricultural Organisation (2011) FAO statistic database (FAOSTAT). Roma. Available at http://faostat.fao.org.

Furlan AD, Pennick V, Bombardier C, van Tulder M (2009) 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* 34: 1929-1941.

Freibauer A, Kaltschmitt M (2003) Controls and models for estimating direct nitrous oxide emissions from temperate and sub-boreal agricultural mineral soils in Europe. *Biogeochemistry* 63: 93-115.

Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW , Seitzinger SP, Asner GP, Cleveland CC, Green PA, Holland EA, Karl DM, Michaels AF, Porter JH, Townsend AR, Vörösmarty CJ (2004) Nitrogen cycles: past, present and future. *Biogeochemistry* 70: 153-226.

Gates S (2002) Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology* 71: 547-557.

Ghattas B (2000) Agrégation d'arbres de classification. *Revue de Statistique Appliquée* 48: 85-98.

Glass GV (1976) Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3-8.

Grace PR, Robertson GP, Millar N, Colunga-Garcia M, Basso B, Gage SH, Hoben J (2011) The contribution of maize cropping in the Midwest USA to global warming: a regional estimate. *Agricultural Systems* 104: 292-296.

Gu J, Nicoullaud B, Rochette P, Pennock DJ, Hénault C, Cellier P, Richard G (2011) Effect of topography on nitrous oxide emissions from winter wheat fields in Central France. *Environmental Pollution* 159: 3149-3155.

Gurevitch J, Morrow LL, Wallace A, Walsh JS (1992) A meta-analysis of competition in field experiments. *The American Naturalist* 140: 539-572.

Gurevitch J, Hedges L (1993) Meta-analysis: combining the results of independent experiments, in: Scheiner, S.M., Gurevitch, J, Design and analysis of ecological experiments. Chapman and Hall, pp. 378-398.

Hedges L, Olkin I (1985) Statistical methods for meta-analysis. Academic Press, New York.

Hénault C, Chèneby D, Heurlier K, Garrido F, Perez S, Germon J-C (2001) Laboratory kinetics of soil denitrification are useful to discriminate soils with potentially high levels of N_2O emission on the field scale. *Agronomie* 21: 713-723.

Hénault C, Grossel A, Mary B, Roussel M, Léonard J (2012) Nitrous oxide emission by agricultural soils: a review of spatial and temporal variability for mitigation. *Pedosphere* 22: 426-433.

Higgins JPT, Green S (2005) Cochrane Handbook for systematic reviews of interventions 4.2.5 [updated may 2005]. The Cochrane Library, Issue 3. Eds Wiley: Chichester, U.K.

Hoben JP, Gehl RJ, Millar N, Grace PR, Robertson GP (2011) Nonlinear nitrous oxide (N_2O) response to nitrogen fertilizer in on-farm corn crops of the US Midwest. *Global Change Biology* 17: 1140-1152.

IPCC (1996) Revised 1996 IPCC guidelines for national greenhouse gas inventories, Workbook (vol 2) Module 4 Agriculture IGES, Japan.

IPCC (1999) N_2O: Direct Emissions from Agricultural Soils. In: Background Papers: IPCC Expert Meetings on Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. pp. 361-380.

IPCC (2001) Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the IPCC. Cambridge: Cambridge University Press. 881p.

IPCC (2006) Agriculture, Forestry and Other Land Use, Volume 4. In: 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Japan: Institute for Global Environmental Strategies.

IPCC (2007) Climate Change 2007: The Physical Science Basis Contribution of Working Group I to the Fourth Assessment Report of the IPCC Cambridge: Cambridge University Press.

Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT (2003) The DSSAT cropping system model. *European Journal of Agronomy* 18: 235-265.

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z, McLean G, Verburg K, Snow V, Dimes JP, Silburn M, Wang E, Brown S, Bristow KL, Asseng S, Chapman S, McCown RL, Freebairn DM, Smith CJ (2003) An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* 18: 267-288.

Kenward MG, Goetghebeur EJT, Molenberghs G (2001) Sensitivity analysis for incomplete categorical tables. *Statistical Modelling* 1: 31-48.

Lapointe NWR, Light T (2012) Landscape-scale determinants of non-native fish communities. *Diversity and Distributions* 18: 282-293.

Leip A, Marchi G, Koeble R, Kempen M, Britz W, Li C (2008) Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. *Biogeosciences* 5: 73-94.

Leip A, Busto M, Winiwarter W (2011) Developing spatially stratified N_2O emission factors for Europe. *Environmental Pollution* 159: 3223-3232.

Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967.

Lesschen JP, Velthof GL, de Vries W, Kros J (2011) Differentiation of nitrous oxide emission factors for agricultural soils. *Environmental Pollution* 159: 3215-3222.

Lesur C, Jeuffroy M-H, Makowski D, Riche AB, Shield I, Yates N, Fritz M, Formowitz B, Grunert M, Jorgensen U, Laerke PE, Loyce C. Modeling long-term yield trends of Miscanthus x giganteus using experimental data from across Europe. Submitted.

Li C, Frolking S, Frolking T (1992) Model of nitrous oxide evolution from soil driven by rainfall events: 1, model structure and sensitivity. *Journal of Geophysical Research* 97: 9759-9776.

Li C (2000) Modeling trace gas emissions from agricultural ecosystems. *Nutrient Cycling in Agroecosystems* 58: 259-273.

Liaw A, Wiener M (2002) Classification and regression by Random Forest. *R News* 2: 18-22.

Light R, Pillemer D (1984) Summing up: The Science of reviewing research. Harvard University Press, Cambridge.

Lokupitiya E, Paustian K (2006) Agricultural soil greenhouse gas emissions. *Journal of Environment Quality* 35: 1413-1427.

Lugato E, Zuliani M, Alberti G, Delle Vedove G, Gioli B, Miglietta F, Peressotti A (2010) Application of DNDC biogeochemistry model to estimate greenhouse gas emissions from Italian agricultural areas at high spatial resolution. *Agriculture, Ecosystems & Environment* 139: 546-556.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10: 325-337.

Ma BL, Wu, TY, Tremblay WD, Morrison MJ, McLaughlin NB, Gregorich EG, Stewart G (2010) Nitrous oxide fluxes from corn-fields: on –farm assessment of the amount and timing of nitrogen fertilizer. *Global Change Biology* 16: 156-170.

Madden LV, Paul PA (2011) Meta-analysis for evidence synthesis in plant pathology: An overview. *Phytopathology* 101: 16-30.

Makowski D, Wallach D (2001) How to improve model-based decision rules for nitrogen fertilization. *European Journal of Agronomy* 15: 197-208.

Makowski D, Tichit M, Guichard L, Van Keulen H, Beaudoin N (2009) Measuring the accuracy of agro-environmental indicators. *Journal of Environmental Management* 90: S139-S146.

McSwiney CP, Robertson GP (2005) Nonlinear response of N_2O flux to incremental fertilizer addition in a continuous maize (*Zea mays* L.) cropping system. *Global Change Biology* 11: 1712–1719.

Millar N, Robertson GP, Grace PR, Gehl RJ, Hoben JP (2010) Nitrogen fertilizer management for nitrous oxide (N_2O) mitigation in intensive corn (maize) production: an emissions reduction protocol for US Midwest agriculture. *Mitigation and Adaptation Strategies for Global Change* 15: 185–204.

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP (1999) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technology Assessment* 3: 1-98.

Moher D, Schulz KF, Altman DG for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *The Lancet* 357: 1191-1194.

Monod H, Naud C, Makowski D (2006) Chapter 3: Uncertainty and sensitivity analysis for crop models, in: Wallach, D., Makowski, D., Jones, J., Working with dynamic crop models. Eds, Elsevier, pp. 55- 100.

Mosier A, Kroeze C, Nevison C, Oenema O, Seitzinger S, van Cleemput O (1998) Closing the global atmospheric N_2O budget: nitrous oxide emissions through the agricultural nitrogen cycle; OECD/IPCC/IEA Phase II Development of IPCC Guidelines for National Greenhouse Gas Inventories. *Nutrient Cycling in Agroecosystems* 52: 225-248.

Myers RA (2001) Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science* 58: 937-951.

Nakagawa S, Santos ESA (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* 26: 1253-1274.

Normand SL (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* 18: 321-359.

O'Rourke K (2007) A historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine* 100: 579-582.

Osborne B, Saunders M, Walmsley D, Jones M, Smith P (2010) Key questions and uncertainties associated with the assessment of the cropland greenhouse gas balance. *Agriculture, Ecosystems & Environment* 139: 293-301.

Oxman AD (1994) Checklists for review articles. *British Medical Journal* 309: 648-651.

Pearson K (1904) Report on certain enteric fever inoculation statistics. *British Medical Journal* 3: 1243-1246.

Petticrew M (2001) Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal* 322: 98-101.

Philibert A, Desprez-Loustau ML, Fabre B, Frey P, Halkett F, Husson C, Lung-Escarmant B, Marçais B, Robin C, Vacher C, Makowski D (2011) Predicting invasion success of forest pathogenic fungi from species traits. *Journal of Applied Ecology* 48: 1381-1390.

Philibert A, Loyce C, Makowski D (2012a) Assessment of the quality of meta-analysis in agronomy. *Agriculture, Ecosystems & Environment* 148: 72–82.

Philibert A, Loyce C, Makowski D (2012b) Quantifying uncertainties in N_2O emission due to N fertilizer application in cultivated areas. *PLoS ONE* 7 (11).

Pinheiro J, Bates D (2000) Mixed-effects Models in S and S-PLUS. NewYork: Springer. 2nd Ed.

Prasad A, Iverson L, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181-199.

Pullin AS, Stewart GB (2006) Guidelines for systematic review in conservation and environmental Management. *Conservation Biology* 20: 1647-1656.

Ravishankara AR, Daniel JS, Portmann RW (2009) Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st century. *Science* 326: 123-125.

Roberts PD, Stewart GB, Pullin AS (2006) Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. *Biological Conservation* 132: 409-423.

Robson M, Fowler S, Lampkin N, Leifert C, Leitch M, Robinson D, Watson C, Litterick A (2002) The agronomic and economic potential of break crops for ley/arable rotations in temperate organic agriculture. *Advances in Agronomy* 77: 369-427.

Rochette P, Janzen HH (2005) Towards a revised coefficient for estimating N_2O emissions from legumes. *Nutrient Cycling in Agroecosystems* 73: 171-179.

Rochette P (2008) No-till only increases N_2O emissions in poorly-aerated soils. *Soil and Tillage Research* 101: 97-100.

Rochette P, Worth DE, Lemke RL, McConkey BG, Pennock DJ, Wagner-Riddle C, Desjardins RJ (2008) Estimation of N_2O emissions from agricultural soils in Canada. I. Development of a country-specific methodology. *Canadian Journal of Soil Science* 88: 641–654.

Rochette P, Tremblay N, Fallon E, Angers DA, Chantigny MH, MacDonald JD, Bertrand N, Parent L-E (2010) N_2O emissions from an irrigated and non-irrigated organic soil in eastern Canada as influenced by N fertilizer addition. *European Journal of Soil Science* 61: 186-196.

Roelandt C, Van Wesemael B, Rounsevell M (2005) Estimating annual N_2O emissions from agricultural soils in temperate climates. *Global Change Biology* 11: 1701-1711.

Rosenberg MS, Garrett KA, Su Z, Bowden RL (2004) Meta-analysis in plant pathology: synthesizing research results. *Phytopathology* 94: 1013-1017.

Rubin DB (1987) Multiple imputation for nonresponse in surveys. Wiley, New York.

Sauvant D, Schmidely P, Daudin JJ, St-Pierre NR (2008) Meta-analyses of experimental data in animal nutrition. *Animal* 2: 1203-1214.

Skiba U, Smith KA (2000) The control of nitrous oxide emissions from agricultural and natural soils. *Chemosphere Global Change Science* 2: 379-386.

Smith T, Spiegelhalter D, Thomas A (1995) Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* 14: 2685-2699.

Smith PD, Martino Z, Cai D, Gwary H, Janzen P, Kumar P, McCarl B, Ogle S, O'Mara F, Rice C, Scholes B, Sirotenko O (2007) Agriculture. In: Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. [Metz B, Davidson OR, Bosch PR, Dave R, Meyer LA (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Snyder CS, Bruulsema TW, Jensen TL, Fixen PE (2009) Review of greenhouse gas emissions from crop production systems and fertilizer management effects. *Agriculture, Ecosystems & Environment* 133: 247-266.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64: 583-639.

Spiegelhalter DJ, Abrams KR, Myles JP (2010) Bayesian approaches to clinical trials and health-care evaluation. Eds Wiley Statistics in Practice.

Spiertz JHJ (2010) Nitrogen, sustainable agriculture and food security. A review. *Agronomy for Sustainable Development* 30: 43-55.

Stehfest E, Bouwman L (2006) N_2O and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems* 74: 207-228.

Stewart G (2010) Meta-analysis in applied ecology. *Biology Letters* 6: 78-81.

Stigler SM (1986) The history of statistics: The measurement of uncertainty before 1900. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Stjernswärd J (1974) Decreased survival related to irradiation postoperatively in early breast cancer. *Lancet* 304: 1285-1286.

Stöckle CO, Donatelli M, Nelson R (2003) CropSyst, a cropping systems simulation model. *European Journal of Agronomy* 18: 289-307.

Stohlgren TJ, Ma P, Kumar S, Rocca M, Morisette JT, Jarnevich CS, Benson N (2010) Ensemble habitat mapping of invasive plant species. *Risk Analysis* 30: 224−235.

Sutton A, Abrams K, Jones D, Sheldon T, Song F (2000) Methods for meta-analysis in medical research. Wiley, Chichester.

Sutton AJ, Abrams KR (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 10: 277 -303.

Sutton AJ, Higgins PT (2008) Recent developments in meta-analysis. *Statistics in medicine* 27: 625-650.

Therneau TM, Atkinson EJ (1997) An introduction to recursive partitioning using the RPART routines. *Technical Report*, 61, Mayo Clinic, Section of Statistics, Rochester, 52 p.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature* 418: 671-677.

van Ittersum M, Leffelaar P, van Keulen H, Kropff M, Bastiaans L, Goudriaan J (2003) On approaches and applications of the Wageningen crop models. *European Journal of Agronomy* 18: 201-234.

van Tulder MW, Assendelft WJ, Koes BW, Bouter LM 1(997) Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine* 22: 2323-2330.

van Tulder M, Furlan A, Bombardier C, Bouter L (2003) Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine* 28: 1290-1299.

Villa-Vialaneix N, Follador M, Ratto M, Leip A (2012) A comparison of eight metamodeling techniques for the simulation of N_2O fluxes and N leaching from corn crops. *Environmental Modelling & Software* 34: 51-66.

Yates F, Cochran WG (1938) The analysis of groups of experiments. *Journal of Agricultural Science* 28: 556-580.

Yuan Y, Hunt RH (2009) Systematic reviews: the good, the bad, and the ugly. *American Journal of Gastroenterology* 104: 1086-1092.

Annexes

Annexe 1 : Références des 55 papiers faisant une méta-analyse en agronomie(chapitre 1)

Ainsworth EA (2008) Rice production in a changing climate: a meta-analysis of responses to elevated carbon dioxide and elevated ozone concentration. *Global Change Biology* 14: 1642-1650.

Ainsworth EA, Davey PA, Bernacchi CJ, Dermody OC, Heaton EA, Moore DJ, Morgan PB, Naidu SL, Yoo Ra H, Zhu X, Curtis PS, Long SP (2002) A meta-analysis of elevated [CO_2] effects on soybean (Glycine max) physiology, growth and yield. *Global Change Biology* 8: 695-709.

Akiyama H, Yan X, Yagi K (2010) Evaluation of effectiveness of enhanced-efficiency fertilizers as mitigation options for N_2O and NO emissions from agricultural soils: meta-analysis. *Global Change Biology* 16: 1837-1846.

Breseghello F, de Morais OP, Pinheiro PV, Silva ACS, da Maia de Castro E, Guimarães ÉP, de Castro AP, Pereira JA, de Matos Lopes A, Utumi MM, de Oliveira JP (2011) Results of 25 years of upland rice breeding in Brazil. *Crop Science* 51: 914-923.

Carrubba A (2009) Nitrogen fertilisation in coriander (Coriandrum sativum L.): a review and meta-analysis. *Journal of the Science of Food and Agriculture* 89: 921-926.

Chivenge P, Vanlauwe B, Six J (2010) Does the combined application of organic and mineral nutrient sources influence maize productivity? A meta-analysis. *Plant and Soil* 342: 1-30.

Conant RT, Paustian K, Elliott ET (2001) Grassland management and conversion into grassland: effects on soil carbon. *Ecological Applications* 11: 343-355.

Danielsen F, Beukema H, Burgess ND, Parish F, Brühl CA, Donald PF, Murdiyarso D, Phalan B, Reijnders L, Struebig M, Fitzherbert EB (2009) Biofuel plantations on forested lands: double jeopardy for biodiversity and climate. *Conservation Biology* 23: 348-358.

Don A, Schumacher J, Freibauer A (2011) Impact of tropical land-use change on soil organic carbon stocks – a meta-analysis. *Global Change Biology* 17: 1658-1670.

Felton A, Knight E, Wood J, Zammit C, Lindenmayer D (2010) A meta-analysis of fauna and flora species richness and abundance in plantations and pasture lands. *Biological Conservation* 143: 545-554.

Feng Z, Kobayashi K (2009) Assessing the impacts of current and future concentrations of surface ozone on crop yield with meta-analysis. *Atmospheric Environment* 43: 1510-1519.

Feng Z, Kobayashi K, Ainsworth EA (2008) Impact of elevated ozone concentration on growth, physiology, and yield of wheat (Triticum aestivum L.): a meta-analysis. *Global Change Biology* 14: 2696-2708.

Feng Z, Wang S, Szantoi Z, Chen S, Wang X (2010) Protection of plants from ambient ozone by applications of ethylenediurea (EDU): a meta-analytic review. *Environmental Pollution* 158 : 3236-3242.

Fletcher RJ, Robertson BA, Evans J, Doran PJ, Alavalapati JR, Schemske DW (2011) Biodiversity conservation in the era of biofuels: risks and opportunities. *Frontiers in Ecology and the Environment* 9: 161-168.

Gardner JB, Drinkwater LE (2009) The fate of nitrogen in grain cropping systems: a meta-analysis of ^{15}N field experiments. *Ecological Applications* 19: 2167-2184.

Garratt MPD, Wright DJ, Leather SR (2011) The effects of farming system and fertilisers on pests and natural enemies: A synthesis of current research. *Agriculture, Ecosystems & Environment* 141: 261-270.

Gobbi M, Fontaneto D (2008) Biodiversity of ground beetles (Coleoptera: Carabidae) in different habitats of the Italian Po lowland. *Agriculture, Ecosystems & Environment* 127: 273-276.

Grassini P, Hall AJ, Mercau JL (2009) Benchmarking sunflower productivity in semiarid environments. *Field Crops Research* 110: 251-262.

Guo LB, Gifford RM (2002) Soil carbon stocks and land use change: a meta analysis. *Global Change Biology* 8: 345-360.

Heaton E, Voigt T, Long SP (2004) A quantitative review comparing the yields of two candidate C_4 perennial biomass crops in relation to nitrogen, temperature and water. *Biomass and Bioenergy* 27: 21-30.

Kaschuk G, Alberton O, Hungria M (2010) Quantifying effects of different agricultural land uses on soil microbial biomass and activity in Brazilian biomes: inferences to improve soil quality. *Plant and Soil* 338: 467-481.

Kiær LP, Skovgaard IM, Østergård H (2009) Grain yield increase in cereal variety mixtures: A meta-analysis of field trials. *Field Crops Research* 114: 361-373.

Koellner T, Scholz RW (2006) Assessment of land use impacts on the natural environment. *The International Journal of Life Cycle Assessment* 13: 32-48.

Letourneau DK, Armbrecht I, Rivera BS, Lerma JM, Carmona EJ, Daza MC, Escobar S, Galindo V, Gutiérrez C, López SD, Mejía JL, Rangel AMA, Rangel JH, Rivera L, Saavedra CA, Torres AM, Trujillo AR (2011) Does plant diversity benefit agroecosystems? A synthetic review. *Ecological Applications* 21: 9-21.

Linquist B, Groenigen KJ, Adviento-Borbe MA, Pittelkow C, Kessel C (2012) An agronomic assessment of greenhouse gas emissions from major cereal crops. *Global Change Biology* 18: 194-209.

Lu M, Zhou X, Luo Y, Yang Y, Fang C, Chen J, Li B (2011) Minor stimulation of soil carbon storage by nitrogen addition: A meta-analysis. *Agriculture, Ecosystems & Environment* 140: 234-244.

Luo Z, Wang E, Sun OJ (2010) Can no-tillage stimulate carbon sequestration in agricultural soils? A meta-analysis of paired experiments. *Agriculture, Ecosystems & Environment* 139: 224-231.

Miguez F, Bollero G (2005) Review of corn yield response under winter cover cropping systems using meta-analytic methods. *Crop Science* 45: 2318-2329.

Miguez FE, Villamil MB, Long SP, Bollero GA (2008) Meta-analysis of the effects of management factors on Miscanthus x giganteus growth and biomass production. *Agricultural and Forest Meteorology* 148: 1280-1292.

Ogle SM, Breidt FJ, Paustian K (2005) Agricultural management impacts on soil organic carbon storage under moist and dry climatic conditions of temperate and tropical regions. *Biogeochemistry* 72: 87-121.

Philpott SM, Arendt WJ, Armbrecht I, Bichier P, Diestch TV, Gordon C, Greenberg R, Perfecto I, Reynoso-santos R, Soto-pinto L, Tejeda-cruz C, Williams-linera G, Valenzuela J, Zolotoff JM (2008) Biodiversity loss in latin American coffee landscapes: review of the evidence on ants, birds, and trees. *Conservation Biology* 22: 1093-1105.

Prieto-Benítez S, Méndez M (2011) Effects of land management on the abundance and richness of spiders (Araneae): A meta-analysis. *Biological Conservation* 144: 683-691.

Roberts PD, Pullin AS (2007) The effectiveness of management interventions used to control Ragwort species. *Environmental Management* 39: 691-706.

Rotundo JL, Westgate ME (2009) Meta-analysis of environmental effects on soybean seed composition. *Field Crops Research* 110: 147-156.

Rui W, Zhang W (2010) Effect size and duration of recommended management practices on carbon sequestration in paddy field in Yangtze Delta Plain of China: A meta-analysis. *Agriculture, Ecosystems & Environment* 135: 199-205.

Sadras VO (2009) Does partial root-zone drying improve irrigation water productivity in the field? A meta-analysis. *Irrigation Science* 27: 183-190.

Salvagiotti F, Cassman KG, Specht JE, Walters DT, Weiss A, Dobermann A (2008) Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *Field Crops Research* 108: 1-13.

Sileshi G, Akinnifesi FK, Ajayi OC, Place F (2008) Meta-analysis of maize yield response to woody and herbaceous legumes in sub-Saharan Africa. *Plant and Soil* 307: 1-19.

Sileshi G, Akinnifesi FK, Debusho LK, Beedy T, Ajayi OC, Mong'omba S (2010) Variation in maize yield gaps with plant nutrient inputs, soil type and climate across sub-Saharan Africa. *Field Crops Research* 116: 1-13.

Sullivan TP, Sullivan DS (2003) Vegetation management and ecosystem disturbance: impact of glyphosate herbicide on plant and animal diversity in terrestrial systems. *Environmental Reviews* 11: 37-59.

Taub DR, Miller B, Allen H (2008) Effects of elevated CO_2 on the protein concentration of food crops: a meta-analysis. *Global Change Biology* 14: 565-575.

Thenail C, Joannon A, Capitaine M, Souchère V, Mignolet C, Schermann N, Di Pietro F, Pons Y, Gaucherel C, Viaud V, Baudry J (2009) The contribution of crop-rotation organization in farms to crop-mosaic patterning at local landscape scales. *Agriculture, Ecosystems & Environment* 131: 207-219.

Tirol-Padre A, Ladha JK (2006) Integrating rice and wheat productivity trends using the SAS mixed-procedure and meta-analysis. *Field Crops Research* 95: 75-88.

Tonitto C, David M, Drinkwater L (2006) Replacing bare fallows with cover crops in fertilizer-intensive cropping systems: A meta-analysis of crop yield and N dynamics. *Agriculture, Ecosystems & Environment* 112: 58-72.

Tudoreanu L, Phillips CJ (2004) Empirical models of cadmium accumulation in maize, rye grass and soya bean plants. *Journal of the Science of Food and Agriculture* 84: 845-852.

Turmel M-S, Turner BL, Whalen JK (2011) Soil fertility and the yield response to the system of rice intensification. Renew. *Renewable Agriculture and Food Systems* 26: 185-192.

Unkovich MJ, Baldock J, Peoples MB (2010) Prospects and problems of simple linear models for estimating symbiotic N_2 fixation by crop and pasture legumes. *Plant and Soil* 329: 75-89.

Valkama E, Uusitalo R, Ylivainio K, Virkajärvi P, Turtola E (2009) Phosphorus fertilization: A meta-analysis of 80 years of research in Finland. *Agriculture, Ecosystems & Environment* 130: 75-85.

Van den Putte A, Govers G, Diels J, Gillijns K, Demuzere M (2010) Assessing the effect of soil tillage on crop growth: A meta-regression analysis on European crop yields under conservation agriculture. *European Journal of Agronomy* 33: 231-241.

Vanlauwe B, Kihara J, Chivenge P, Pypers P, Coe R, Six J (2010) Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management. *Plant and Soil*. 339: 35-50.

Veres A, Petit S, Conord C, Lavigne C (2012) Does landscape composition affect pest abundance and their control by natural enemies? A review. *Agriculture, Ecosystems & Environment* In Press, Corrected Proof. Available at http://www.sciencedirect.com/science/article/pii/S0167880911001824 (verified september 13 2012).

Veresoglou SD, Menexes G (2010) Impact of inoculation with Azospirillum spp. on growth properties and seed yield of wheat: a meta-analysis of studies in the ISI Web of Science from 1981 to 2008. *Plant and Soil* 337: 469-480.

Virto I, Barré P, Burlot A, Chenu C (2012) Carbon input differences as the main factor explaining the variability in soil organic C storage in no-tilled compared to inversion tilled agrosystems. *Biogeochemistry* 108: 17-26.

Wang S (2011) Management and land use change effects on soil carbon in northern China's grasslands: a synthesis. *Agriculture, Ecosystems & Environment* 142: 329-340.

Yahdjian L, Gherardi L, Sala OE (2011) Nitrogen limitation in arid-subhumid ecosystems: A meta-analysis of fertilization studies. *Journal of Arid Environments* 75: 675-680.

Annexe 2: Code WinBUGS du modèle non linéaire à effets aléatoires Bayésien (chapitre 2 et 4)

 # Nreference: number of published experiments (=203)
 # Nobs: number of observations (=985)
 # reference: vector composed of all numbers of published experiments repeated T times,
 # with T the number of N$_2$O measurements per published experiment
 # muu0 = μ_0, muu1 = μ_1, sigma1 = σ_1^2, sigma0 = σ_0^2, tau = τ^2
 # mu0[reference[i]] = α_{0i}, mu1[reference[i]] = α_{1i}, NRate = X

 # Model definition
MODEL {

 # Precisions for normal distribution
```
        prec1 <- 1/sigma1 ;
        prec0 <- 1/sigma0 ;
        prectau <- 1/tau;

        for (k in 1:Nreference) {
                mu0[k] ~ dnorm (muu0, prec0);
                mu1[k] ~ dnorm (muu1, prec1);
        }

        for( i in 1:Nobs) {
                N20[i] ~ dnorm (mu[i], prectau);
                mu[i] <- exp (mu0[reference[i]]+mu1[reference[i]]*NRate[i]);
        }
```

 # Prior definitions
```
        muu0 ~ dnorm (0, 0.0001);
        muu1 ~ dnorm (0, 0.0001);
        tau ~ dunif (0, 100);
        sigma0 ~ dunif (0, 100);
        sigma1 ~ dunif (0, 100);
}
```

Annexe 3: Code R des modèles fréquentistes présentés en Table 2.2 au chapitre 2

```
# N20 = Y, N_rate = X, Ref_num = numbers of published experiments
# Package for non linear mixed-effects models
library(nlme)
```

 # Loading and formatting data
```
data <- read.csv2(…)
data.T <- groupedData (N20 ~ N_rate | Ref_num, data = data)
```

 # Nonlinear models:
 # model NL-N-FF with gnls function
```
model-NL-N-FF <- gnls(N20~exp(alpha1+alpha2*N_rate), data=data,
start=c(alpha1=1, alpha2=0.004))
```
 # model NL-0-R with nlme function
```
model-NL-0-R <- nlme(N20~exp(alpha1), data=data, fixed=alpha1~1,
random=alpha1~1|Ref_num, start=c(alpha1= -0.009 ))
```
 # model NL-N-RF with nlme function
```
model-NL-N-RF <- nlme(N20~exp(alpha1+alpha2*N_rate),
data=data,fixed=alpha1+alpha2~1, random=alpha1~1|Ref_num,
start=c(alpha1= -0.6, alpha2= 0.003 ))
```
 # model NL-N-FR with nlme function
```
model-NL-N-FR <- nlme(N20~exp(alpha1+alpha2*N_rate), data=data,
fixed=alpha1+alpha2~1, random=alpha2~1|Ref_num,
start=c(alpha1= -0.6, alpha2= 0.004 ))
```
 # model NL-N-RR with nlme function
```
model-1NL-N-RR <- nlme(N20~exp(alpha1+alpha2*N_rate), data=data.T,
fixed=alpha1+alpha2~1, random=pdDiag(alpha1+alpha2~1),
start=c(alpha1= 1.46, alpha2= 0.002 ))
```

 # Linear models:
 # model L-0-F with glm function
```
model-L-0-F <- glm (N20~1, data=data)
```
 # model L-N-FF with glm function
```
model-L-N-FF <- glm (N20~1+N_rate, data=data)
```
 # model L-0-R with lme function
```
model-L-0-R <- lme (N20~1, data = data, random=~1|Ref_num)
```
 # model L-N-RF with lme function
```
model-L-N-RF <- lme (N20~1+N_rate, data=data, random=~1|Ref_num)
```
 # model L-N-FR with lme function
```
model-L-N-FR <- lme (N20~1+N_rate, data = data, random=~N_rate-1|Ref_num)
```
 # model L-N-RR with lme function
```
model-L-N-RR <- lme (N20~1+N_rate, data=data.T, random=pdDiag(~1+N_rate))
```

Annexe 4: Exemple de code R avec la fonction randomForest pour le groupe 3 de variables explicatives et la variable N₂O quantitative (chapitre 3)

 # Loading data
```
data <- read.csv2(…)
```

 # Replacement of missing data with Random Forest as explained page 73-74.
 # 1000 trees built five times
```
data.impute <-
rfImpute(N2O~Texture_class+Carbon_content+pH+Drainage+N_rate+Fertilizer_type
+Method_of_N_application+Timing_of_N_application+Exp_duration+Method_N2O+
Freq_N2O+Climate+Country+Crop_type, data=data, ntree=1000)
```

 # 500 trees built with Random Forest to compute measurement of importance for
 # input variables and global error rate using data created above
```
data.rf <-
randomForest(N2O~Texture_class+Carbon_content+pH+Drainage+N_rate+
Fertilizer_type+Method_of_N_application+Timing_of_N_application+Exp_duration+
Method_N2O+Freq_N2O+Climate+Country+Crop_type, data=data.impute,
importance=TRUE, ntree=500)
```

 # Display of input variables by decreasing order of importance
```
sort(importance(data.rf)[,1], decreasing=TRUE)
```

Annexe 5 : Code R pour les prédictions avec données simulées (chapitre 4)

 # Example with M = 500 locations and Q = 3 amounts of N fertilizer tested for the
 # 'incomplete' location

```r
M <- 500
Q <- 3
```

 # Initialization of the simulation

```r
set.seed(1)
Moyenne0 <- integer(0)
Site0 <- integer(0)
Moyenne300 <- integer(0)
Site300 <- integer(0)
```

 # Simulation of the N_2O measurements for the M locations with parameter values from
 # the dataset of Stehfest and Bouwman

```r
while (length(Site0)!=600) {
        alpha0 <- rnorm (M, 0.1931946, 0.7194441)
        alpha1 <- rnorm (M, 0.0036925, 0.002511928)
```

 # Amounts of N fertilizer tested for the M-1 'complete' locations

```r
        doses.temp <- c(0,50,100,200,300)
        totN2O <- integer(0)
```

 # Simulation of the N_2O measurements for the M-1 'complete' locations

```r
        for (i in 1:M) {
                N2O <- exp(alpha0[i]+alpha1[i]*doses.temp)+ rnorm(5,0,1.936393)
                N2O[N2O<0] <- 0
                totN2O <- c(totN2O,N2O)
        }
```

 # Simulation of the N_2O measurements for the Mth 'incomplete' location

```r
        reference <- rep(1:M,each=5)
        referenceplus <- rep(M+1,Q)
        if (Q==1) {doseplus <- 100
                } else if (Q==3) {
                doseplus <- c(50,100,200) }
        alpha0plus <- rnorm(1, 0.1931946, 0.7194441)
        alpha1plus <- rnorm(1, 0.0036925, 0.002511928)
        residplus <- rnorm(1, 0, 1.936393)
        N2Oplus <- exp(alpha0plus+alpha1plus*doseplus)+residplus
```

 # Compilation of N_2O measurements for the M locations

```r
        baseM <- cbind(reference, rep(doses.temp,M), totN2O)
        baseM1 <- cbind(referenceplus, doseplus, N2Oplus)
        base <- rbind(baseM, baseM1)
        colnames(base) <- c("reference","doses","N2O")
```

```
# Formatting data
base.frame <- data.frame(reference=base[,1],doses=base[,2],N2O = base[,3])
# Nonlinear model with the dataset created above
library(nlme)
base.T <- groupedData(N2O ~ doses | reference, data = base.frame)
modele <-try(nlme(N2O~exp(alpha1+alpha2*doses), data=base.T,
fixed=alpha1+alpha2~1, random=pdDiag(alpha1+alpha2~1),
start=c(alpha1= 1.46,alpha2= 0.002 )))

# Predictions computation
if(length(class(modele)) == 2) {
        dosepredi <- c(0,300)
        # Observed N2O measurements
        vraiesite <- exp(alpha0plus+alpha1plus*dosepredi)

        # Average prediction of N2O emissions simulated
        alpha0 <- modele$coefficients$fixed[1]
        alpha1 <- modele$coefficients$fixed[2]
        predimoyenne <- exp(alpha0+alpha1*dosepredi)

        # Location-specific prediction of N2O emissions simulated
        data.pred <- data.frame(doses = dosepredi, reference = M+1)
        predisite <- predict(modele,newdata = data.pred)

        # RMSEP computation
        moyenne0 <- (predimoyenne[1] - vraiesite[1])^2
        site0 <- (predisite[1] - vraiesite[1])^2
        Moyenne0 <- c(Moyenne0,moyenne0)
        Site0 <- c(Site0,site0)
        moyenne300 <- (predimoyenne[2] - vraiesite[2])^2
        site300 <- (predisite[2] - vraiesite[2])^2
        Moyenne300 <- c(Moyenne300,moyenne300)
        Site300 <- c(Site300,site300)
    }
}
# RMSEP results for average and location-specific predictions
rmseMoyenne0 <- sqrt(mean(Moyenne0))
rmseSite0 <- sqrt(mean(Site0))
rmseMoyenne300 <- sqrt(mean(Moyenne300))
rmseSite300 <- sqrt(mean(Site300))
```

Abstract

Meta-analysis methods to estimate N$_2$O emissions from agricultural soils.

The term meta-analysis refers to the statistical analysis of a large set of results coming from individual studies about the same topic. This approach is increasingly used in various areas, including agronomy. In this domain however, a bibliographic review conducted by this thesis, showed that meta-analyses were not always of good quality. Meta-analyses in agronomy very seldom study the robustness of their findings relative to data quality and statistical methods.

The objective of this thesis is to demonstrate and illustrate the importance of sensitivity analysis in the context of meta-analysis and as an example this is based on the estimation of N$_2$O emissions from agricultural soils. The estimation of emissions of nitrous oxide (N$_2$O) is made at the worldwide level by the Intergovernmental Panel on Climate Change (IPCC). N$_2$O is a potent greenhouse gas with a global warming power 298 times greater than the one of CO$_2$ over a 100 year period. The key characteristics of N$_2$O emissions are a significant spatial and time variability.

Two databases are used for this work: the database of Rochette and Janzen (2005) and the one of Stehfest and Bouwman (2006). They collect numerous worldwide N$_2$O emissions measurements from published studies and have played a significant role in the estimation of N$_2$O emissions produced by the IPCC.

The results show the value of random effects models in order to estimate N$_2$O emissions from agricultural soils. They are well suited to the structure of the data (repeated observations on the same site for different doses of fertilizers, with several sites considered). They allow to differentiate the inter-site and intra-site variability and to estimate the effect of the rate of nitrogen fertilizer on the N$_2$O emissions. In this paper, the analysis of the sensitivity of the estimations to the shape of the relationship "Emission of N$_2$O / N fertilizer dose" has shown that an exponential relationship would be the most appropriate. Therefore it would be appropriate to replace the constant emission factor of the IPCC (1% emission whatever the dose of nitrogen fertilizer) by a variable factor which would increase with the dose. On the other hand we did not identify significant differences between frequentist and Bayesian inference methods.

Two approaches have been proposed to include environmental variables and cropping practices in the estimates of N$_2$O. The first one using the *Random Forest* method allows managing missing data and provides the best N$_2$O emissions predictions. The other one, based on random effects models allow to take into account these explanatory variables via one or several measurements of N$_2$O. They allow predicting N$_2$O emissions for non-tested doses in unfertilized farmer's field. However their results are sensitive to the experimental design used locally to measure N$_2$O emissions.

Key words: sensitivity analysis, meta-analysis, nitrous oxide, climate change, nitrogen fertilization, mixed model, Bayesian statistics, Random Forest, IPCC, uncertainty analysis.

Résumé

Méthodes de méta-analyse pour l'estimation des émissions de N_2O par les sols agricoles

Le terme de méta-analyse désigne l'analyse statistique d'un large ensemble de résultats provenant d'études individuelles pour un même sujet donné. Cette approche est de plus en plus étudiée dans différents domaines, notamment en agronomie. Dans cette discipline, une revue bibliographique réalisée dans le cadre de la thèse a cependant montré que les méta-analyses n'étaient pas toujours de bonne qualité. Les méta-analyses effectuées en agronomie étudient ainsi très rarement la robustesse de leurs conclusions aux données utilisées et aux méthodes statistiques.

L'objectif de cette thèse est de démontrer et d'illustrer l'importance des analyses de sensibilité dans le cadre de la méta-analyse en s'appuyant sur l'exemple de l'estimation des émissions de N_2O provenant des sols agricoles. L'estimation des émissions de protoxyde d'azote (N_2O) est réalisée à l'échelle mondiale par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). Le N_2O est un puissant gaz à effet de serre avec un pouvoir de réchauffement 298 fois plus puissant que le CO_2 sur une période de 100 ans. Les émissions de N_2O ont la particularité de présenter une forte variabilité spatiale et temporelle.

Deux bases de données sont utilisées dans ce travail : la base de données de Rochette et Janzen (2005) et celle de Stehfest et Bouwman (2006). Elles recensent de nombreuses mesures d'émissions de N_2O réparties dans le monde provenant d'études publiées et ont joué un rôle important lors des estimations d'émission de N_2O réalisées par le GIEC.

Les résultats montrent l'intérêt des modèles à effets aléatoires pour estimer les émissions de N_2O issues de sols agricoles. Ils sont bien adaptés à la structure des données (observations répétées sur un même site pour différentes doses d'engrais, avec plusieurs sites considérés). Ils permettent de distinguer la variabilité inter-sites de la variabilité intra-site et d'estimer l'effet de la dose d'engrais azoté sur les émissions de N_2O. Dans ce mémoire, l'analyse de la sensibilité des estimations à la forme de la relation « Emission de N_2O / Dose d'engrais azoté » a montré qu'une relation exponentielle était plus adaptée. Il apparait ainsi souhaitable de remplacer le facteur d'émission constant du GIEC (1% d'émission quelque soit la dose d'engrais azoté) par un facteur variable qui augmenterait en fonction de la dose. Nous n'avons par contre pas identifié de différence importante entre les méthodes d'inférence fréquentiste et bayésienne.

Deux approches ont été proposées pour inclure des variables de milieu et de pratiques culturales dans les estimations de N_2O. La méthode *Random Forest* permet de gérer les données manquantes et présente les meilleures prédictions d'émission de N_2O. Les modèles à effets aléatoires permettent eux de prendre en compte ces variables explicatives par le biais d'une ou plusieurs mesures d'émission de N_2O. Cette méthode permet de prédire les émissions de N_2O pour des doses non testées comme le cas non fertilisé en parcelles agricoles. Les résultats de cette méthode sont cependant sensibles au plan d'expérience utilisé localement pour mesurer les émissions de N_2O.