Analyse des erreurs des modèles utilisés en agronomie

David Makowski

Université Paris-Sud Orsay

Ecole doctorale Sciences du Végétal

Mémoire présenté en vue d'obtenir

L'Habilitation à diriger des recherches

Analyse des erreurs des modèles utilisés en agronomie

par David Makowski

Chargé de Recherche à l'Institut National de la Recherche Agronomique

Le 25 octobre 2007

devant la commission d'examen composée de :

Marc Lavielle	Professeur, Universités Paris 5 et Paris 11, INRIA	Rapporteur
Robert Faivre	Directeur de Recherche, INRA	Rapporteur
Jacques Wery	Professeur, SupAgro Montpellier	Rapporteur
Jacques-Eric Bergez	Chargé de Recherche HDR, INRA	Examinateur
Jean-Jacques Daudin	Professeur, AgroParisTech	Examinateur
Benoît Gabrielle	Chargé de Recherche HDR, INRA	Examinateur

Avant-propos

Mes dix années de recherche ont été essentiellement consacrées à un sujet, *les erreurs des modèles utilisés en agronomie*. Ce choix a rendu ma vie professionnelle parfois facile, parfois difficile, toujours intéressante. Facile parce que les modèles sont nombreux en agronomie et parce que les données expérimentales utilisées pour leur évaluation sont abondantes. Difficile parce que les performances des modèles ne sont pas toujours à la hauteur des attentes de leurs concepteurs et de leurs utilisateurs. Quelles que soient les difficultés rencontrées, mon sujet de recherche m'a toujours permis d'avoir des échanges d'idées très riches avec de nombreuses personnes : des agronomes et biologistes réalisant des expérimentations et développant des modèles, mais aussi des statisticiens, des mathématiciens et des économistes.

Mes travaux mêlent deux ingrédients : des modèles et des données. Finalement, je n'ai que peu participé à leur production. Certains des modèles utilisés dans mes travaux sont issus de mes propres activités, mais beaucoup ont été développés par d'autres personnes, lors de travaux antérieurs. J'ai également bénéficié des nombreuses expérimentations réalisées pas mes collègues agronomes et biologistes.

Parmi les personnes rencontrées au cours de ces années de recherche, beaucoup viennent de l'INRA, notamment de l'UMR Agronomie de Grignon où je travaille aujourd'hui, mais aussi de l'UMR Arche de Toulouse et de l'UR Mathématiques et Informatique Appliquées de Jouy-en-Josas où j'ai eu plaisir à travailler dans le passé et avec lesquelles j'entretiens encore des relations étroites. Mes travaux m'ont également conduit à collaborer avec d'autres unités INRA, notamment l'UMR Climat Sol Environnement d'Avignon, l'UMR Environnement et Grande Culture de Grignon, l'UMR Agronomie de Laon, le laboratoire d'économie appliquée de Grenoble, l'UMR Sciences pour l'action et le développement de Grignon, l'UMR de biologie des organismes et des populations appliquées à la protection des plantes de Rennes.

J'ai également eu l'opportunité de travailler à de nombreuses reprises avec d'autres organismes professionnels, notamment ARVALIS-Institut du végétal, CETIOM, Yara, le CEMAGREF, Université d'Orsay, Wageningen University, University of Florida. Je remercie chaleureusement toutes les personnes avec qui j'ai pu collaborer pour leurs données, leurs modèles, leurs méthodes et, surtout, pour les échanges d'idées que nous avons eus.

La réalisation de ce mémoire a bénéficié du soutien de plusieurs personnes. Je tiens à remercier en particulier Thierry Doré et Marie-Hélène Jeuffroy, respectivement professeur à AgroParisTech et directeur de recherche à l'INRA, qui m'ont encouragé à soutenir mon Habilitation. Merci aussi à Jacqui Shykoff, directeur de recherche au CNRS, qui m'a aidé à réaliser les formalités administratives, et à Hervé Monod qui a accepté de relire une version provisoire de ce rapport. Merci bien sûr aux membres du jury qui m'ont fait l'honneur d'évaluer ce travail.

Le mémoire est organisé en quatre parties. La première présente mon cursus. La deuxième partie présente mes principaux travaux de recherche sur la période 1997-2007. La troisième partie présente quelques perspectives pour les années à venir. Une conclusion est finalement proposée.

Bonne lecture,

David.

Résumé

Les modèles sont utilisés dans différentes disciplines pour prédire l'état de systèmes naturels ou humains, notamment en météorologie, hydrologie, finance, biologie. La modélisation est pratiquée depuis le début du 20ème siècle en agronomie. Les modèles se sont rapidement multipliés et diversifiés dans cette discipline. Ils peuvent potentiellement être utilisés pour répondre à de nombreuses questions pratiques, à l'échelle de la parcelle agricole, de la région, voire du continent. Ils permettent, par exemple, d'optimiser les doses d'engrais appliquées aux cultures, de raisonner les traitements fongicides, de quantifier les risques de pollution de l'eau au niveau d'un bassin versant, d'optimiser la répartition de différents types de culture au sein d'une région.

Ce mémoire s'intéresse à l'analyse des erreurs des modèles utilisés en agronomie, c'est-à-dire à l'évaluation de ces erreurs et à leur réduction. Je considère ici à la fois les erreurs de prédiction des modèles et les erreurs de décision induites par leur utilisation.

La première section a pour objectif de présenter l'intérêt pratique d'une évaluation quantitative des erreurs des modèles. Je montre, en m'appuyant sur plusieurs cas d'étude, que l'évaluation des erreurs des modèles est utile, non seulement pour les utilisateurs de ces modèles mais aussi pour leurs concepteurs : une quantification précise des erreurs des modèles permet de faciliter le choix d'un modèle pour un usage donné, de donner aux utilisateurs une information sur les risques encourus, et de raisonner l'orientation des travaux de recherche. Dans la deuxième section, je présente plusieurs méthodes statistiques pour quantifier les erreurs des modèles. L'intérêt de plusieurs critères d'évaluation est discuté (MSEP, sensibilité, spécificité, perte économique, risque). La section suivante est dédiée aux méthodes statistiques permettant de réduire les erreurs des modèles utilisés en agronomie. Les méthodes proposées permettent soit d'améliorer l'estimation des paramètres des modèles (prise en compte des corrélations des résidus, régression quantile, analyse de sensibilité globale), soit de corriger les variables simulées en cours de saison (filtrage particulaire). Je montre que ces méthodes peuvent permettre aux agronomes d'améliorer sensiblement les performances de leurs modèles.

Deux axes de recherche sont ensuite proposés pour les prochaines années. Le premier axe porte sur l'étude de l'incertitude dans les procédures de sélection des modèles. Cet axe repose sur le constat que de nombreuses procédures ont été développées pour choisir un modèle parmi une série de modèles candidats, mais que ces procédures sont susceptibles de faire des erreurs, c'est-à-dire de sélectionner un modèle qui n'est pas réellement le meilleur pour un usage donné. Je propose de quantifier ces erreurs pour différents types d'applications agronomiques et d'étudier l'intérêt de combiner plusieurs modèles candidats plutôt que d'utiliser un modèle unique. Le deuxième axe de recherche porte sur la généralisation de l'utilisation de modèles stochastiques en agronomie. L'utilisation de modèles stochastiques permettrait aux agronomes, d'une part, de présenter les résultats des simulations sous forme de distributions de probabilité et, d'autre part, d'utiliser plus facilement des techniques de type filtrage dont l'intérêt pratique a été illustré dans le cadre de mes travaux. En conclusion, je montre comment mes activités peuvent contribuer à enrichir la «boîte à outils statistique» des agronomes.

Mots-clés : agronomie, assimilation de données, erreur, estimation, fertilisation, filtrage, modèle, protection des cultures, risque, statistique.

Table des matières

Présentation du candidat

David Makowski

Né le 3 décembre 1972 à Paris.

Nationalité française. Marié, deux enfants.

Formation

- 2001. Docteur AgroParisTech (INA P-G).

  Thèse: *Modèles et méthodes statistiques pour optimiser la fertilisation azotée.*
  Mention très honorable, félicitations du jury, Médaille d'argent de l'Académie d'Agriculture.
- 1996-97. Formation complémentaire en Statistique (DEA Modèles stochastiques et statistique).
- 1996. DEA Biologie, diversité, adaptation des plantes cultivées. AgroParisTech.
- 1996. Ingénieur AgroParisTech.

Eléments de carrière

- Depuis nov. 2006. Chargé de recherche 1ère classe INRA. UMR Agronomie, Grignon (78).
- Nov. 2005 – Oct. 2006. Chargé de recherche 1ère classe INRA, détaché à l'UR Mathématique et Informatique appliquées. Jouy-en-Josas (78).
- 2001-2005. Chargé de recherche 2ème classe INRA. UMR Agronomie, Grignon (78).
- 1999-2001. Attaché scientifique contractuel INRA. UMR Arche, Toulouse (31).
- Déc. 1997 – Avril 1999. Research Associate, department of Theoretical Production Ecology et subdepartment of Mathematics, Wageningen University (Pays-Bas).
- Nov. 1996 – Déc. 1997. Attaché scientifique contractuel INRA. UMR Arche, Toulouse (31).

Encadrement, animation, participation à des projets

- *Co-encadrement de trois thèses* : S. Ennaïfar (thèse soutenue en 2006, co-encadrement avec J-M. Meynard et Ph. Lucas), C. Naud (thèse soutenue en 2007, co-encadrement avec M-H. Jeuffroy), M. Lamboni (depuis novembre 2006, co-encadrement avec H. Monod).
- *Participation à six comités de pilotage de thèse (*M. Tremblay, V. Houlès, C. Lauvernet, Z. Assaghir, H. Varella, S. Lehuger).
- *Co-encadrement de quatre étudiants en Master* (M. Tremblay, A. Maltas, M. Ducarne, C. Naud).
- *Encadrement d'un post-doc* en 2007 (Claire Cadet).
- *Coordinateur du projet PICSEL* (prise en compte de l'incertitude dans la sélection des modèles) financé par l'ANR de 2006 à 2009 (appel d'offre *jeunes chercheuses et jeunes chercheurs*).
- *Membre du projet Franco-Néerlandais WUR-INRA* « Multifunctionality of agriculture » de 2003 à 2006.
- *Membre du projet « Bâtir un outil de pilotage et de prévision de la qualité des blés pour les coopératives* (2004-2006) *».* Programme de recherche action INRA - AGRALYS. Dir M. Le Bail de l'UMR SAD-APT, Partenaire : Agralys (Union de coopératives de Beauce).
- *Responsable du projet « Modèles et Mesures »* de l'action transversale INRA-CIRAD « Aide à la décision » (2003-2005) E. de Turckheim et B. Hubert animateurs.
- *Co-responsable Assurance Qualité Recherche* de l'UMR Agronomie de Grignon depuis 2002.
- Membre de la *Société française de Statistique* et de *l'International Biometric Society.*

Expertises

- *Membre du panel Plant Health* de la *European Food Safety Authority* (Parme, Italie) depuis juin 2006. Réalisation d'expertises pour la commission européenne sur les espèces invasives (environ 10 réunions de 2 jours par an).
- *Reviewer* pour des revues scientifiques, 2-3 par an (Environmental modelling and software, Agronomy for Sustainable Development, European journal of Agronomy, Weed Research, Journal of Hydrology).
- *Reviewer* en mai 2007 pour *l'ANR,* évaluation de projets de recherche pour l'appel d'offre JCJC.
- *Reviewer* en avril 2006 et en avril 2007 pour *Department for Environment Food and Rural Affairs,* London, évaluation des programmes *Atmospheric pollution* et *Organic Farming and Non-food Crops*, notamment sur les aspects statistiques.
- *Reviewer* en mai 2005 et 2006 pour *Science and Research group of the Scottish Executive* (évaluation de projets de recherche).

Enseignement

- Enseignant vacataire
  - Module « Statistique » de la spécialité AGER AgroParisTech (36 heures/an depuis 2006).
  - Module « Modélisation » du Master STV AgroParisTech (10 heures/an depuis 2005).
  - Analyse ROC pour l'ESA Angers (4 heures/an depuis 2005).
- Formations professionnelles en modélisation (ACTA-ICTA-INRA). Environ 5 heures par an depuis 2002.

Problématique de recherche

- Evaluation des erreurs des modèles mathématiques utilisés en agronomie (erreurs de prédiction et de classement, analyse de sensibilité).
- Réduction des erreurs de prédiction des modèles à l'aide de méthodes d'estimation de paramètres et de filtrage.
- Amélioration du diagnostic agronomique à l'aide de méthodes statistiques.

Publications

ARTICLES DANS DES REVUES A COMITE DE LECTURE

1. **Makowski, D., D. Wallach, J-M. Meynard.** 1999. Models of yield, grain protein, and residual mineral nitrogen responses to applied nitrogen for winter wheat. *Agronomy Journal* 91:377-385.

2. **Makowski, D., E.M.T. Hendrix, M.K. van Ittersum, W.A.H. Rossing.** 2000. A framework to study nearly optimal solutions of linear programming models developed for agricultural land use exploration. *Ecological Modelling* 131:65-77.

3. **Makowski, D., E.M.T. Hendrix, M.K. van Ittersum, W.A.H. Rossing.** 2001. Generation and presentation of nearly optimal solutions for mixed-integer linear programming, applied to a case in farming system design. *European Journal of Operational Research* 132:425-438.

4. **Makowski, D., D. Wallach, J-M. Meynard.** 2001. Statistical methods for predicting responses to applied nitrogen and for calculating optimal nitrogen rates. *Agronomy Journal* 93:531-539.

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

6. **Makowski, D., D. Wallach.** 2002. It pays to base parameter estimation on a realistic description of model errors. *Agronomy for Sustainable Development* 22:179-189.

7. **Monod, H., D. Makowski, M. Sahmoudi, D. Wallach.** 2002. Optimal experimental designs for estimating model parameters, applied to yield response models. *Agronomy for Sustainable Development* 22:229-338.

8. **Makowski, D., D. Wallach, M. Tremblay.** 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomy for Sustainable Development* 22:191-203.

9. **Makowski, D.** 2002. Modèle non linéaire mixte pour simuler la réponse du blé à la dose d'engrais azoté. *Journal de la Société Française de Statistique* 143:219-227.

10. **Meynard J-M., M. Cerf, L. Guichard, M-H. Jeuffroy, D. Makowski**. 2002. Which decision support tools for the Environmental Management of nitrogen ? *Agronomy for Sustainable Development* 22:817-829.

11. **Le Bail, M., D. Makowski.** 2004. A model-based approach for optimizing segregation of soft wheat in country elevators. *European Journal of Agronomy* 21:171-180.

12. **Houlès, V., B. Mary, M. Guérif, D. Makowski, E. Justes**. 2004. Evaluation of the crop model STICS to recommend nitrogen fertilization rates according to agro-environmental criteria. *Agronomy for Sustainable Development* 24:339-349.

13. **Makowski, D., A. Maltas, M. Morison, R. Reau.** 2005. Calculating N fertilizer rates for oil-seed rape using plant and soil data. *Agronomy for Sustainable Development* 25:159-161.

14. **Makowski, D., M. Taverne, J. Bolomier, M. Ducarne.** 2005. Comparison of risk indicators for sclerotinia control in oilseed rape. *Crop Protection* 24:527-531.

15. **Lacroix, A., N. Beaudoin, D. Makowski.** 2005. Agricultural water nonpoint pollution control under uncertainty and climate variability. *Ecological Economics* 53:115-127.

16. **Ennaïfar, S., P. Lucas, J.M. Meynard, D. Makowski.** 2005. Effects of Summer-fallow Management on Take-all of Winter Wheat caused by *Gaeumannomyces graminis* var. *tritici*. *European Journal of Plant Pathology* 112:167-181.

17. **Hillier, J., D. Makowski, B. Andrieu.** 2005. Maximum likelihood inference and bootstrap methods for plant organ growth via multi-phase kinetic models and their application to maize. *Annals of Botany* 96:137-148.

18. **Primot, S., M. Valantin-Morison, D. Makowski**. 2006. Predicting the risk of weed infestation in winter oilseed rape crops. *Weed Research* 46:22-33.

19. **Makowski, D., M. Lavielle.** 2006. Using SAEM (stochastic approximation of EM) to estimate parameters of models of response to applied fertilizer. *Journal of Agricultural, Biological, and Environmental Statistics* 11 (1):45-60.

20. **Makowski, D., C. Naud, M-H. Jeuffroy, A. Barbottin, H. Monod.** 2006. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model predictions. *Reliability Engineering and System Safety* 91:1142-1147.

21. **Makowski D., T. Doré, H. Monod**. 2007. A new method to analyse relationships between yield components with boundary lines. *Agronomy for Sustainable Development* 27:119-128.

22. **Makowski D, T. Doré, N. Munier-Jolain, J. Gasquez.** 2007. Modelling land use strategies to optimize crop production and protection of ecologically important weed species. *Weed research* 47:202-211.

23. **Ennaïfar, S., D. Makowski, J-M. Meynard, Ph. Lucas**. 2007. Evaluation of models to predict take-all incidence on winter wheat as a function of cropping practices, soil, and climate. *European Journal of Plant Pathology* 118:127-143.

24. **Naud C, Makowski D, Jeuffroy MH.** 2007**.** An interacting particle filter to improve model-based predictions of nitrogen nutrition index for winter wheat. *Ecological modelling* 207:251-263.

25. **Naud C, Makowski D, Jeuffroy MH.** Is it useful to combine measurements taken during the growing season with a dynamic model to predict the nitrogen status of winter wheat ? *European journal of Agronomy*, in press.

Articles soumis

26. **Makowski D., Tichit M., Guichard L., van Keulen H.** Measuring the accuracy of agro-environmental indicators. Soumis.

27. **Makowski D.**, **Denis J-B., Ruck L., Penaud A**. A Bayesian approach to assess the accuracy of a diagnostic test based on plant disease measurement. Soumis.

LIVRES ET CHAPITRES D'OUVRAGES

28. **Wallach D., Makowski D., Jones J.** 2006. *Working with dynamic crop models*, Elsevier.

29. Guest **co-editor** of the special issue on the multifunctionality of agriculture (10 papers currently under review) in collaboration with the Wageningen University.

30. **Makowski, D., M-H. Jeuffroy, M. Guérif**. 2004. Bayesian methods for updating crop model predictions, applications for predicting biomass and grain protein content. *In : « Bayesian Statistics and quality modelling in the agro-food production chain ».* Van Boekel *et al*. (eds). Kluwer Academic Publishers, Dordrecht. p.57-68.

31. **Beaudoin, N., V. Parnaudeau, B. Mary, D. Makowski, J.-M. Meynard.** 2004. Simulation de l'impact de différents scénarios agronomiques sur les pertes de nitrate à l'échelle d'un bassin hydrologique. *In : Organisation spatiale des activités agricoles et processus environnementaux.* P. Monestiez, S. Lardon et B. Seguin Eds, Coll. Science Update, INRA Editions, p. 117-141.

32. **Makowski, D., J. Hillier, D. Wallach, B. Andrieu, MH. Jeuffroy.** 2006. Parameter estimation for crop models. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 101-150.

33. **Monod, H., C. Naud, D. Makowski.** 2006. Uncertainty and sensitivity analysis for crop models. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 55-100.

34. **Makowski, D., M. Guérif, J. Jones., W. Graham.** 2006. Data assimilation with crop models. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 151-172.

35. **Guérif, M., V. Houlès, D. Makowski, C. Lauvernet.** 2006. Data assimilation and parameter estimation for precision agriculture using the crop model STICS. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 391-398.

36. **Houlès V., B. Mary, M. Guérif, D. Makowski, E. Justes, J.-M. Machet.** *2007. Critères agro-*environnementaux fondés sur le modèle de culture STICS pour la modulation intra-parcellaire de la fertilisation azotée du blé. *In: Agriculture de précision*, Guérif M. and King D. (eds), QUAE, Paris, France. p. 199-224.

37. **Makowski D., M. Le Bail, A. Barbottin, M-H. Jeuffroy, C. Barrier, Ch. Bouchard, C. Pasquier**. 2007. Utilisation de modèles pour prédire la qualité du blé. A paraître.

VULGARISATION

38. **Beaudoin, N., D. Makowski, V. Parnaudeau, B. Parisseaux, D. Wallach, B. Mary, J-M. Meynard.** 1998. Evaluation de l'impact économique et environnemental de la mesure agri-environnementale "réduction d'intrants" au moyen de modèles agronomiques. *Rapport Ministère de l'Agriculture*, 79 pages + annexes.

39. **Makowski, D., M. Tremblay, D. Debroize, F. Laurent.** 2000. Apports d'engrais hétérogènes, quel impact économique et environnemental ? *Perspectives Agricoles* 263:56-61.

40. **Makowski, D.** 2001. Evaluation et optimisation des recommandations de doses d'engrais azoté à l'aide de modèles agronomiques simples. Dans « *Les nouveaux défis de la fertilisation raisonnée* ». GEMAS-COMIFER. G. Thévenet et A. Joubet (eds.), p.269-279.

41. **Makowski, D., P. Castillon.** 2002. Fertilisation azotée du blé dur. Rémunérer la teneur en protéines du grain. *Perspectives Agricoles* 277 :56-58.

42. **Makowski, D., A. Maltas, M. Morison, R. Reau.** 2004. Réglette azote colza : même revenu, moins d'engrais, plus d'huile. *Oléoscope* 74 :12-16.

43. **Bouthier A, Makowski D.** 2005. Fertilisation azotée de l'orge de printemps : concilier marge du producteur avec objectifs de qualité. *Perspectives Agricoles* 318 : 62-63.

44. **Makowski, D., L. Ruck**. 2007. Des perspectives prometteuses pour la PCR quantitative. *Oléoscope* 88, 14-15.

CONFERENCES

Beaudoin, N., D. Makowski, V. Parnaudeau, B. Mary. 1999. Impact of agricultural scenarios on nitrate pollution at the catchment scale. *10th Nitrogen Workshop* 23-26 August 1999, KVL Copenhague, vol.2, 4p, presentation orale + résumé.

Meynard, J-M., M. Cerf, L. Guichard, M-H. Jeuffroy, D. Makowski. 2001. Nitrogen, decision support and environmental management. 11th Nitrogen Workshop 9-12 September 2001, Reims, 389-390, presentation orale + résumé.

Makowski, D., D. Wallach, J-M. Meynard. 2001. Random parameter response model as a generalization of the balance-sheet method. 11th Nitrogen Workshop 9-12 September 2001, Reims, 389-390, poster + résumé.

Makowski, D., D. Wallach. 2002. Modèle non linéaire mixte pour optimiser la fertilisation azotée du blé tendre. Journées « Modèles Mixtes et Biométrie », Société Française de Biométrie, Paris, 24-25 janvier 2002, présentation orale + résumé.

Houlès, V., M. Guérif, B. Mary, J-M. Machet, D. Makowski. 2003. A method for optimising nitrogen fertilisation of wheat within a field based on a crop model approach. 4th European Conference on Precision Agriculture, Berlin, june 14-19 2003, in: Wermer A., Jarfe A. (Eds.), pp.437-438, presentation orale + résumé.

Taverne, M., F. Dupeuble, J-M. Meynard, D. Makowski. 2003. Sclerotinia attacks risk assessment on oilseed rape crops at field level. 11th International Rapeseed Congress, Copenhague, july 6-10 2003, p165 (résumé).

Houlès, V., M. Guérif, M. Nowakowski, J. Demarty, D. Makowski, B. Mary. 2004. Comparison of three methods of data assimilation into a crop model, based on remote sensing observations. VIIIth ESA Congress, Copenhague, 11-15 juillet 2004, poster + résumé.

Houlès, V., M. Guérif, B. Mary, D. Makowski, J.M. Machet, S. Moulin, N. Beaudoin. 2004. A tool devoted to recommend spatialised nitrogen rates at the field scale, based on a crop model and remote sensing data assimilation. VIIIth ESA Congress, Copenhague, 11-15 juillet 2004, présentation orale + résumé.

Makowski, D., C. Naud, H. Monod, M-H. Jeuffroy, A. Barbottin. 2004. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model predictions. Fourth International Conference on Sensitivity Analysis of Modeling Output, Santa Fe (USA), march 8-11 2004, présentation orale + article complet disponible sur CD et sur le site : webfarm.jrc.cec.eu.int/uasa/events/SAMO2004/Papers/samo04-28%20(D_Makowski).pdf).

Guichard, L., A. Aveline, C. Bockstaller, D. Makowski. 2004. A model-based approach to assess nitrogen losses in crop rotations including grain legumes. International Workshop on the Methodology for Environmental assessment of Grain Legumes, Zurick, 18-19 novembre 2004, poster+résumé.

Ennaïfar, S., P. Lucas, J-M. Meynard, D. Makowski. 2005. Summer-fallow management to reduce take-all disease (Gaeumannomyces graminis var. tritici) on winter wheat. 9th International workshop on plant disease epidemiology, Landerneau, 11-15 avril 2005, poster+résumé, p.104.

Makowski D, Guichard L, Aveline A, Laurent F. 2005. A method to compare the accuracy of indicators of water pollution by nitrates. 14th N workshop, Maastricht, 24-26 octobre 2005, poster+article.

Naud C., D. Makowski, MH Jeuffroy. 2005. An interacting particle filter for improving the predictions of a dynamic crop model with real-time measurements. Applied Statistics, september 18-21, Ribno, Slovénie.

Makowski D., J-B. Denis, L. Ruck. 2006. Un modèle bayésien pour réaliser une analyse ROC avec un indicateur de risque de sclérotinia du colza. Journées de la société française de statistique, Clamart. Résumé + présentation orale.

Naud C., D. Makowski, MH Jeuffroy. 2006. An interacting particle filter for improving the predictions of NNI. International Biometrics Conference, July 2006, Montreal, Canada, présentation orale.

Jeuffroy M.H., Barbottin A., Barrier C., Bouchard C., Le Bail M., Makowski D. 2006. Performances of five crop models simulating yield and grain protein content in farmers fields. European Society for Agronomy, 2006. présentation orale + résumé.

Lamboni M., Monod H., Makowski D. 2007. Indice de sensibilité pour les modèles à sortie multidimensionnelles: application à un modèle de culture dynamique. 39 ième Journées de la Société Française de Statistique, Anger, France, 11-15 juin 2007, pp 34-35 pour résumé court et CD pour résumé long.

Lamboni M., Monod H., Makowski D. 2007. Global sensitivity analysis for simulation models with multiple outputs: an application to a dynamic crop models. Fifth International Conference on sensitivity analysis of model output SAMO, 18-22 June 2007, pp 115-116, Budapest, Hungary.

Makowski D., Denis J-B., Ruck L., Penaud A. 2007. A Bayesian approach to assess the accuracy of diagnostic test based on plant disease measurement. 2007 American Phytopathological Society Annual meeting, July 28-August 1, San Diego, California, USA. Phytopathology 97 (7) S69.

Prediction is very difficult, especially if it’s about the future.

- *Neils Bohr*

Activités de recherche

1. Introduction

Un modèle est défini ici comme une fonction qui permet de calculer au moins une variable de sortie à partir d'une variable d'entrée ou plus. Des modèles sont utilisés dans différentes disciplines pour prédire l'état de systèmes naturels ou humains, notamment en météorologie, hydrologie, finance, biologie (Baudrillard *et al*., 1996[1]). D'autres modèles sont développés, non pas pour faire des prédictions, mais pour aider leurs concepteurs à formaliser un problème ou à mieux comprendre le fonctionnement d'un système complexe. Bien qu'ils jouent un rôle important dans l'élaboration des théories scientifiques, ce deuxième type de modèles n'est pas considéré dans ce mémoire. Je ne m'intéresse ici qu'aux modèles qui, d'une façon ou d'une autre, sont utilisés pour réaliser des prédictions.

La modélisation est pratiquée depuis longtemps en agronomie. Le plus ancien modèle agronomique est probablement celui de Mitscherlich développé au tout début du 20ème siècle pour relier le rendement d'une culture à la dose d'engrais (Mitscherlich, 1909). Depuis, les modèles utilisés en agronomie se sont multipliés et diversifiés. Le tableau 1 présente un petit nombre d'exemples. Certains modèles sont linéaires, d'autres sont des modèles linéaires généralisés (logistique notamment). Il existe également des modèles non linéaires statiques ou dynamiques. Les modèles utilisés en agronomie diffèrent non seulement par leur nature mais aussi par le nombre de leurs paramètres. Il existe des modèles à un seul paramètre et des modèles incluant plusieurs centaines de paramètres.

Tableau 1. Exemples de modèles utilisés en agronomie. *x* représente une variable d'entrée ou un vecteur de variables d'entrée, *y* une variable de sortie éventuellement dynamique, θ un paramètre ou un vecteur de paramètres, *f* une fonction.

Type de modèle	Equation	Exemple d'utilisation	Référence
Linéaire	$y=\theta_1 x_1 + \theta_2 x_2 + ... + \theta_p x_p$	Prédiction de l'incidence d'une maladie.	[23]
Linéaire généralisé	$y=\frac{\exp\left(\theta_1 x_1 + \theta_2 x_2 + ... + \theta_p x_p\right)}{1+\exp\left(\theta_1 x_1 + \theta_2 x_2 + ... + \theta_p x_p\right)}$	Raisonnement du désherbage.	[18]
Non linéaire statique	$y=\min\left[\theta_1, \theta_1 + \theta_2\left(x-\theta_3\right)\right]$	Optimisation de la dose d'engrais.	[19]
Dynamique	$y_t = y_{t-1} + f\left(x_t, \theta\right)$	Prédiction de la biomasse.	[34]

Les modèles peuvent être utilisés pour répondre à de nombreuses questions en agronomie (e.g Boote *et al*, 1996 ; Passioura, 1996 ; Rossing *et al*. 1997) comme, par exemple : Quelle dose d'engrais azoté apporter à une parcelle de blé ? A quelle date semer une parcelle de colza ? Quel itinéraire technique pour obtenir un rendement élevé, une qualité satisfaisante et limiter les risques de pollution ? Doit-on traiter cette parcelle avec un fongicide ? Quelle méthode pour lutter contre les mauvaises herbes ? Doit-on implanter une culture « piège à nitrate » sur cette parcelle ?

Pilkey et Pilkey-Jarvis (2007) montrent, à travers une intéressante liste de cas d'étude, que l'utilisation de modèles mathématiques a conduit, dans le passé, à de nombreuses erreurs de prédiction et de décision dans divers domaines (économie, écologie, érosion…). En agronomie, les valeurs des variables de sortie prédites par les modèles (e.g. rendement, teneur en protéines des grains, teneur en nitrates de l'eau…) peuvent être éloignées des vraies valeurs. Les modèles agronomiques peuvent également être à l'origine de recommandations techniques (e.g traitement fongicide) différentes des

[1] Les références aux publications de l'auteur de ce mémoire sont présentées sous la forme de numéros entre [.] renvoyant à la partie « Présentation du candidat ». Les autres références sont présentées de manière classique.

recommandations qu'il aurait réellement fallu appliquer sur une parcelle donnée pour un objectif donné. De telles erreurs peuvent être dues à des équations inappropriées, des paramètres mal estimés, des variables d'entrée inconnues ou connues de façon imprécise.

La puissance de calcul actuelle des ordinateurs et les progrès récents de la statistique offrent de nouvelles possibilités aux agronomes pour analyser les erreurs des modèles, c'est-à-dire pour quantifier ces erreurs et essayer de les réduire. Ces avancées techniques sont pour l'instant sous-valorisées en agronomie. Un des objectifs de mes travaux est d'illustrer leurs intérêts pratiques et de favoriser leur diffusion.

Dans ce mémoire, je présente des éléments de réponse à trois questions:

i) Pourquoi s'intéresser aux erreurs des modèles en agronomie ?

ii) Comment évaluer les erreurs de ces modèles ?

iii) Comment les réduire ?

Dans la section 2, je montre que l'évaluation des erreurs des modèles est utile, non seulement pour les utilisateurs de ces modèles mais aussi pour leurs concepteurs. Des méthodes statistiques sont ensuite présentées dans la section 3 pour quantifier les erreurs des modèles. Plusieurs critères d'évaluation sont proposés dans cette section et leur intérêt pratique est discuté. La section 4 est dédiée aux méthodes permettant de réduire les erreurs des modèles en agronomie. Les méthodes statistiques présentées dans cette section permettent soit d'améliorer l'estimation des paramètres des modèles, soit de corriger les variables simulées en cours de saison. Toutes les méthodes présentées dans ce mémoire sont illustrées par des applications à des problèmes agronomiques.

2. Pourquoi s'intéresser aux erreurs des modèles ?

L'évaluation des erreurs des modèles facilite le choix d'un modèle pour un usage donné, donne une information sur le risque d'erreur de prédiction ou de décision, et permet d'orienter les travaux de recherche. Ces trois aspects sont discutés ci-dessous à partir de plusieurs cas d'étude.

2.1. Evaluer les erreurs pour faciliter le choix d'un modèle

Il est fréquent que plusieurs modèles soient disponibles pour un usage donné. C'est notamment le cas pour le raisonnement de la dose totale d'engrais azoté. Certains des modèles proposés pour ce type d'application simulent la réponse du rendement à la dose d'engrais (Mombiela *et al*., 1981 ; Sain and Jauregui, 1993 ; [1] ; [4]). D'autres sont des modèles de type bilan (Stanford, 1973 ; Rémy et Hébert, 1977 ; [13]). De nombreux modèles ont également été développés pour traiter d'autres problèmes pratiques, par exemple pour prédire le niveau de contamination d'une culture par une maladie ([14] ; [23]) ou encore pour évaluer les risques de pollution de l'eau par les nitrates [10].

Tous les modèles font des erreurs et il est important de les évaluer pour choisir le modèle le plus approprié pour un usage donné. Le choix d'un modèle ne doit cependant pas être basé exclusivement sur l'importance de ses erreurs. D'autres considérations doivent être prises en compte comme, par exemple, le coût d'utilisation des modèles et leur temps de calcul.

L'évaluation des erreurs est d'autant plus nécessaire qu'il est difficile de déduire le niveau d'erreur d'un modèle directement à partir de sa structure ou de son niveau de complexité. Il est bien connu en statistique qu'un modèle complexe pourra, dans certains cas, conduire à des niveaux d'erreur identiques, voire plus grands, que ceux d'un modèle plus simple même si ce dernier néglige certains facteurs (e.g Miller 2002). Makowski *et al* [37] ont comparé les valeur de Root Mean Squared Error of Prediction (RMSEP) de cinq modèles permettant de prédire le rendement et la teneur en protéines du blé (*Triticum aestivum* L) en utilisant des mesures réalisées sur 35 parcelles agricoles en 2004 et 2005 (Tableau 2). Le critère d'évaluation RMSEP permet d'évaluer l'erreur moyenne de prédiction d'un modèle à partir de données expérimentales. C'est un critère souvent utilisé en agronomie pour évaluer

des modèles de façon individuelle (e.g Colson *et al*., 1995 ; Wallach *et al*., 2001), mais plus rarement dans un contexte de comparaison de modèles.

Les quatre modèles testés dans [37] avaient des niveaux de complexité très différents. Un des modèles était dynamique et utilisait de nombreuses variables d'entrée telles que des variables climatiques journalières et des caractéristiques du sol. Les autres modèles étaient plus simples et n'incluaient que deux variables d'entrée. Ce sont ces modèles simples qui se sont avérés être les moins imprécis pour prédire la teneur en protéines. Les résultats présentés dans le Tableau 2 étaient difficilement prévisibles avant cette étude. Ils illustrent ainsi l'intérêt de comparer les niveaux d'erreurs des modèles disponibles pour un usage donné.

Tableau 2. Evaluation des erreurs de prédiction de modèles pour le rendement et la teneur en protéines des grains (RMSEP=root mean square error of prediction ; RRMSEP : RMSEP relatif). Calculs réalisés avec des données obtenues en 2004 et 2005 [37].

Modèles testés	Prédictions de rendement		Prédictions de teneur en protéines	
	RMSEP ($t.ha^{-1}$)	RRMSEP (%)	RMSEP (%)	RRMSEP (%)
Modèle 1	1.5	17.36	2.53	21.58
Modèle 2	1.8	20.50	3.24	27.62
Modèle 3	2.8	32.37	-	-
Modèle 4	1.4	16.12	1.37	11.65
Modèle 5	1.6	18.21	1.37	11.67

Modèle 1 : modèle dynamique Azodyn utilisant des variables d'entrée mesurées à la sortie de l'hiver (Jeuffroy et Recous, 1999 ; David et al., 2005). Modèle 2 : modèle dynamique Azodyn utilisant des variables d'entrée mesurées à la floraison. Modèle 3 : modèle proposé par Le Bail et al., 2005 utilisant une mesure réalisée avec le chlorophyll meter et le nombre d'épi par mètre carré au stade GS71. Modèle 4 : modèle proposé par Le Bail et al., 2005 utilisant uniquement la mesurée réalisée par le chlorophyll meter au stade GS71. Modèle 5 : modèle statique proposé par [4] utilisant la dose totale d'engrais et le reliquat sortie hiver.

La thèse de S. Ennaïfar [16, 23] (Ennaïfar, 2006) montre également tout l'intérêt d'analyser les erreurs des modèles quand plusieurs modèles sont disponibles pour prédire une même variable. La variable prédite était ici l'incidence d'une maladie racinaire du blé appelée piétin échaudage (*Gaeumannomyces graminis* var. *tritici*). Dans cette étude, les erreurs de 16 modèles simulant l'incidence du piétin échaudage ont été analysées sur divers critères, notamment le RMSEP. Avant l'étude, le modèle privilégié par les agronomes était un modèle non linéaire incluant des relations multiplicatives entres les variables d'entrée et certaines variables intermédiaires. L'étude réalisée par Ennaïfar *et al*. [23] a révélé que les prédictions des valeurs d'incidence étaient plus précises avec des modèles basés sur des relations linéaires (Tableau 3). Les résultats montrent, par ailleurs, qu'il n'y a pas de lien clair entre précision des paramètres et nombre de paramètres, c'est-à-dire entre précision et complexité (Tableau3).

L'étude réalisée par Houlès *et al*. [12] constitue un autre exemple intéressant. Cette étude a permis d'évaluer les erreurs de prédiction du modèle dynamique STICS (Brisson *et al*., 1998) pour deux modalités d'usage : (i) prédiction du rendement et de la teneur en protéines en supposant le climat de l'année connu, (ii) prédiction du rendement et de la teneur en protéines en supposant le climat de l'année inconnu. Dans le deuxième cas, le modèle a été utilisé avec 30 années de données climatiques passées et les 30 prédictions résultantes ont été moyennées pour chacune des parcelles expérimentales considérées. Les résultats révèlent que les prédictions ne sont pas plus précises lorsqu'on suppose le

climat connu. Ce résultat montre qu'il n'est pas pertinent, dans le contexte de l'étude, d'utiliser le climat de l'année pour prédire le rendement et la teneur en protéines avec STICS.

Tous ces résultats montrent que, dans de nombreuses situations, plusieurs modèles sont disponibles pour un usage donné et que les niveaux d'erreur de ces modèles sont difficilement prévisibles sans étude spécifique. Il apparaît donc nécessaire d'évaluer les erreurs des modèles aussi précisément que possible afin de faciliter le choix du modèle le plus approprié pour une application donnée.

Tableau 3. Evaluation de 16 modèles prédisant l'incidence du piétin échaudage du blé au stade GS33 sur les racines nodales [23]. Les noms des modèles s'interprètent de la façon suivante. « Dyn / Sta » = modèle dynamique / statique, « L / M » = modèle intégrant des relations linéaires / multiplicatives, « 0 / 1 » = modèle n'incluant pas / incluant une mesure précoce d'incidence de la maladie, « -/+ » = pas de sélection / sélection des variables avec une méthode statistique.

	Nom du modèle	Nombre de paramètres	RMSEP (%)	Biais (%)
1	$DynL0^-$	29	38.1	1.17
2	$DynL0^+$	21	34.3	1.54
3	$DynM0^-$	27	37.0	5.44
4	$DynM0^+$	23	39.7	5.96
5	$StaL0^-$	19	41.3	0.41
6	$StaL0^+$	16	32.1	0.64
7	$StaM0^-$	19	40.1	7.08
8	$StaM0^+$	17	33.2	7.15
9	$DynL1^-$	28	16.9	0.54
10	$DynL1^+$	23	16.3	0.48
11	$DynM1^-$	29	23.2	4.53
12	$DynM1^+$	24	22.8	4.87
13	$StaL1^-$	20	19.4	0.24
14	$StaL1^+$	17	16.4	0.20
15	$StaM1^-$	20	65.9	2.96
16	$StaM1^+$	17	58.2	2.82

2.2. Donner aux utilisateurs une information sur les risques d'erreurs

L'évaluation des erreurs d'un modèle permet également de donner une information sur les risques induits par l'utilisation de ce modèle. Comme tous les modèles font des erreurs, l'utilisation d'un modèle pour réaliser une application agronomique pourra conduire à des prédictions erronées ou à de mauvaises préconisations. Il est donc essentiel de fournir à tout utilisateur (scientifiques, conseillers techniques, ingénieurs des pouvoirs publics ou des instituts techniques…) une information sur le type d'erreur résultant de l'utilisation d'un modèle donné.

La prise en compte des erreurs est d'autant plus nécessaire que les modèles agronomiques jouent un rôle de plus en plus important dans les expertises scientifiques à diverses échelles : à l'échelle de l'exploitation [3], à l'échelle régionale [11] [15], et même à l'échelle d'un continent [2]. Le projet européen *System for Environmental and Agricultural Modelling ; Linking European Science and Society* (SEAMLESS, http://www.seamless-ip.org/) vise explicitement à développer une plateforme de modélisation pour analyser les conséquences des politiques agricoles et environnementales européennes. La prise en compte des erreurs des modèles est essentielle dans un tel contexte. Cet aspect est souligné par Lacroix *et al*. [15] qui montre l'importance de tenir compte des erreurs des modèles lorsque ceux-ci sont utilisés pour évaluer le coût et l'efficacité de pratiques agricoles proposées par les agronomes pour réduire les risques de pollution de l'eau par les nitrates.

2.3. Orienter les recherches futures

Une évaluation des erreurs d'un modèle permet également de quantifier l'intérêt d'une amélioration de ce modèle et ainsi contribuer à l'orientation des recherches futures. Si un modèle fait des erreurs très faibles, toute étude supplémentaire ne pourra conduire qu'à une amélioration marginale de ses performances. Inversement, si les erreurs s'avèrent être importantes, des travaux complémentaires pourront, peut-être, améliorer sensiblement les résultats.

L'évaluation des erreurs permet également de faire des hypothèses sur les points faibles des modèles et sur la meilleure façon de les améliorer. Par exemple, le tableau 3 montre que l'utilisation d'une mesure précoce de l'incidence du piétin échaudage permet de réduire substantiellement les erreurs de prédiction des modèles. Ce résultat plaide en faveur du développement de méthode de mesure précoce et rapide de l'incidence de la maladie, à partir d'analyses PCR par exemple.

Une autre illustration est présentée par Makowski et Wallach [5]. Ces auteurs ont calculé l'intérêt économique d'améliorer les recommandations de doses d'engrais calculées avec un modèle statique. Le modèle considéré permettait de calculer des doses d'engrais azoté maximisant la marge brute de l'agriculteur en fonction de l'azote minéral disponible dans le sol avant l'apport d'engrais, du prix de vente du blé et du prix de l'engrais. Une estimation de la perte économique liée à l'utilisation de ce modèle (voir section 3.1 et 3.4 pour une définition mathématique de ce critère) a montré qu'une amélioration du modèle pouvait conduire à une augmentation du revenu de l'agriculteur au maximum égale à 34 euros ha^{-1} pour un prix de l'engrais égal à 0.3 euro kg^{-1}, et au maximum égale à 50.5 euros ha^{-1} pour un prix de l'engrais égal à 0.61 euro kg^{-1}. Ce type d'information permet d'évaluer l'intérêt de mettre en place de nouveaux travaux dans le but d'améliorer un modèle pour un usage donné. La méthode utilisée est détaillée dans la section suivante. Elle permet, notamment, de déterminer si la réalisation d'expérimentations supplémentaires pourrait conduire à une amélioration de l'estimation des paramètres du modèle [5].

3. Comment évaluer les erreurs des modèles ?

Je présente ici une série de critères permettant d'évaluer les erreurs des modèles utilisés en agronomie. Ces critères sont présentés dans le cadre de la théorie de la décision (e.g. Lindberg BW 1971 ; Lindsay, 2004). Leur intérêt est discuté ci-dessous à partir d'une série de cas d'étude.

3.1. Fonction de coût et risque

Les notions de fonction de coût et de risque sont définies dans cette section. Ces notions, issues de la théorie de la décision, seront reprises ici pour évaluer les erreurs des modèles en agronomie. Elles permettent d'évaluer des erreurs portant sur des variables très diverses comme, par exemple, le rendement d'une culture, la qualité d'une production, ou des variables décisionnelles.

Notons z la vraie valeur d'une variable d'intérêt pour un site-année donné. z représentera, par exemple, le vrai rendement d'une culture sur un site-année particulier. z pourra également représenter une variable décisionnelle comme, par exemple, une dose optimale d'engrais pour un site-année donné.

Nous nous plaçons dans la situation où un modèle est utilisé pour prédire la valeur z d'un site-année en fonction d'un vecteur *x* décrivant les caractéristiques de ce site-année (type de sol, climat…). Notons $f(x,\hat{\theta})$ la prédiction obtenue avec le modèle lorsque les paramètres θ du modèle sont fixés à une valeur estimée $\hat{\theta}$. La *fonction de coût* $L[z, f(x,\hat{\theta})]$ permet d'évaluer l'erreur résultant de l'utilisation de $f(x,\hat{\theta})$ pour prédire z.

En pratique, on ne s'intéresse pas à un seul site-année mais à plusieurs, par exemple à un ensemble de sites-années cultivés avec du blé dans le nord de la France. Dans ce cas, il est utile d'utiliser un *coût moyen* défini par :

$$C(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} L\left[z_i, f(x_i,\hat{\theta})\right] \quad (1)$$

où *i* est l'indice désignant le *i*ème site-année et *N* est le nombre de sites-années considérés. Le coût moyen (1) est calculé pour une valeur donnée $\hat{\theta}$ des paramètres du modèle. Il dépend donc de la valeur estimée des paramètres. Le concept de coût moyen se généralise de la façon suivante :

$$C(\hat{\theta}) = E\left\{ L[z, f(x,\hat{\theta})] \,\middle|\, \hat{\theta} \right\} \quad (2)$$

où *E*(.) est l'espérance prise sur la distribution des valeurs de *z* et de *x* possibles. Le calcul du coût moyen (2) nécessite de définir des distributions de probabilité pour *x* et *z* décrivant les caractéristiques d'une population de sites-années virtuelles.

D'autres critères que le coût moyen peuvent être calculés. Par exemple, il est parfois intéressant de calculer le *coût maximal* résultant de l'utilisation d'un modèle pour prédire z ce qui revient à se placer dans le pire des cas pour évaluer le modèle. Ce critère est défini par :

$$C_{\max}(\hat{\theta}) = \max_{x,z}\left\{ L[z, f(x,\hat{\theta})] \right\} \quad (3)$$

Les critères (1), (2) et (3) sont déterminés pour une valeur de paramètres notée $\hat{\theta}$. En pratique, $\hat{\theta}$ résulte d'une estimation des paramètres du modèle à partir de données expérimentales et peut donc être considérée comme une variable dépendant des données et de la méthode d'estimation utilisée. Une autre valeur de paramètres conduira à des valeurs différentes de coût, coût moyen ou coût maximal. Il peut donc être intéressant d'utiliser un critère plus global appelé *risque* pour évaluer les erreurs d'un modèle, non pas pour une valeur unique de paramètres mais pour une procédure plus globale incluant la méthode d'estimation des paramètres. Le risque est défini par :

$$R = E[C(\hat{\theta})] \quad (4)$$

L'espérance de (4) est prise sur l'ensemble des estimations possibles pour les paramètres du modèle considéré. L'intérêt de *R* est que ce critère ne dépend pas des valeurs estimées, mais seulement de la fonction *f* utilisée dans le modèle et de la méthode utilisée pour estimer les paramètres.

Les critères décrits ci-dessus sont tous définis à partir d'une fonction de coût. Cette fonction doit être choisie en fonction des objectifs de l'analyse. Parmi les fonctions de coût classiques, on peut citer le coût quadratique, le coût 0-1, et la perte économique. Les paragraphes suivants montrent comment ces fonctions de coût permettent d'évaluer les erreurs des modèles en agronomie.

3.2. Erreur quadratique moyenne de prédiction

L'erreur quadratique moyenne de prédiction ou *mean squared error of prediction* (MSEP) d'un modèle est souvent utilisée pour évaluer la précision des prédictions des modèles en agronomie (e.g Colson *et al*., 1995 ; Wallach et Goffinet, 1987). Ce critère donne une information sur la qualité globale des prédictions d'un modèle. Il permet de comparer facilement la précision des prédictions de différents modèles pour une même variable.

La MSEP d'un modèle dont les paramètres sont fixés à la valeur $\hat{\theta}$ correspond au coût moyen (1) avec une fonction de coût *quadratique*, $L[z, f(x,\hat{\theta})] = [z - f(x,\hat{\theta})]^2$. La MSEP est ainsi définie par :

$$C(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\left[z_i - f(x_i,\hat{\theta})\right]^2 \qquad (5)$$

si on considère *N* sites-années. Bien souvent, c'est la racine carrée de la MSEP qui est utilisée pour se ramener à l'unité de mesure de z (on utilise alors la RMSEP). La MSEP peut être exprimée comme la somme de deux termes, le biais au carré et la variance (Bunke et Droge, 1984 ; Wallach et Goffinet, 1987 ; Miller, 2002 ; Wallach, 2006). Un biais différent de zéro révèlera une erreur de prédiction systématique, une sous-estimation ou une surestimation. Une variance élevée indique que la valeur de *z* varie beaucoup pour une valeur de *x* fixée.

Un critère plus global est nécessaire si on s'intéresse aux erreurs de prédiction moyennées sur l'ensemble des valeurs des paramètres pouvant être obtenues avec une méthode d'estimation donnée. Dans ce cas, il est intéressant de calculer la valeur du risque (4) qui devient, avec la fonction de coût quadratique,

$$R = E\left\{\frac{1}{N}\sum_{i=1}^{N}\left[z_i - f(x_i,\hat{\theta})\right]^2\right\} \qquad (6)$$

où l'espérance est sur les valeurs des paramètres estimées avec une méthode d'estimation donnée. Le calcul de (6) nécessite une modélisation du processus à l'origine des données.

La MSEP peut être estimée avec une série de mesures de z obtenues sur différentes parcelles agricoles. Si ces mesures sont indépendantes de celles utilisées pour estimer les paramètres du modèle, alors la MSEP peut être estimée en utilisant le modèle pour prédire les valeurs mesurées de z et en calculant directement la moyenne des écarts quadratiques entre mesures et prédictions. Cette approche a été utilisée par Le Bail et Makowski [11] et Makowski *et al* [37] pour évaluer les erreurs de prédiction de modèles simulant le rendement et la teneur en protéines du blé (Tableau 2). Ceci n'a été possible que parce que des expérimentations spécifiques, indépendantes de celles utilisées pour développer les modèles, avaient été réalisées spécifiquement pour évaluer les erreurs des modèles.

Souvent, des données indépendantes ne sont pas disponibles. Il est alors nécessaire de réutiliser les données ayant déjà servi à estimer les paramètres du modèle. Plusieurs approches sont possibles, notamment la validation croisée (e.g Efron, 1983), déjà utilisée depuis les années 1990 pour estimer la MSEP des modèles en agronomie (e.g Colson *et al*., 1995). Dans la plupart des études publiées, seul un modèle est évalué. Plus récemment, la validation croisée a été utilisée pour comparer 16 modèles permettant de prédire l'incidence d'une maladie racinaire, le piétin échaudage du blé (Tableau 3) [23].

Les résultats ont permis de comparer plusieurs équations et de plusieurs types de variables d'entrée pour modéliser cette maladie.

L'utilisation de la validation croisée nécessite cependant une bonne formalisation des procédures utilisées pour estimer les paramètres des modèles. Lorsque la méthode d'estimation est bien identifiée (e.g. moindres carrés ordinaires), l'utilisation de la validation croisée est généralement possible. Par contre, lorsque la méthode d'estimation des paramètres résulte d'une procédure ad hoc (e.g mélange d'ajustement manuel des paramètres, d'expertise et de techniques statistiques), l'utilisation de la validation croisée est délicate voire impossible.

3.3. Probabilités de bonnes et de mauvaises décisions

Supposons que $f(x,\hat{\theta})$ représente une variable décisionnelle binaire prenant deux valeurs, 0 ou 1. Un tel cas de figure se rencontre, notamment, lorsque $f(x,\hat{\theta})$ est une règle de décision définie par « Si $x \geq \hat{\theta}$, alors $f(x,\hat{\theta})$=1 et si $x < \hat{\theta}$ alors $f(x,\hat{\theta})$=0 ». Ce type de règle peut être utilisé pour raisonner le traitement contre les maladies. La variable *x* correspond alors à un indicateur de risque de maladie, par exemple un pourcentage de plantes malades dans un site-année, et $\hat{\theta}$ à un *seuil de décision*. Lorsque la valeur de l'indicateur est supérieure au seuil, la règle de décision recommande un traitement (cas $f(x,\hat{\theta})$=1) pour éviter des pertes de rendement et/ou de revenu. Dans le cas inverse, aucun traitement n'est recommandé (cas $f(x,\hat{\theta})$=0). L'indicateur utilisé peut être plus complexe qu'une simple mesure et peut représenter lui-même une prédiction issue d'un modèle.

L'erreur associée à la règle peut être évaluée à l'aide la fonction de coût 0-1 définie par :

$L[z, f(x,\hat{\theta})] = 1 - f(x,\hat{\theta})$ si z=1

$L[z, f(x,\hat{\theta})] = f(x,\hat{\theta})$ si z=0

où z représente une variable de référence égale à 1 si un traitement est réellement nécessaire et à zéro sinon.

Le coût moyen (2) est défini par :

$$\begin{aligned} C(\hat{\theta}) &= E\left\{ L[z, f(x,\hat{\theta})] \,\middle|\, \hat{\theta} \right\} \\ &= P[f(x,\hat{\theta}) = 0 | z = 1] \times P(z = 1) + P[f(x,\hat{\theta}) = 1 | z = 0] \times P(z = 0) \end{aligned} \quad (7)$$

Cette décomposition du coût moyen dépend des probabilités de prendre une mauvaise décision, c'est-à-dire des probabilités que $f(x,\hat{\theta})$=0 si z=1 et que $f(x,\hat{\theta})$=1 si z=0. Lorsque $f(x,\hat{\theta})$ représente une règle utilisée pour raisonner un traitement contre une maladie, ces probabilités correspondent respectivement à la probabilité de ne pas recommander de traitement sur le site-année alors qu'un traitement est nécessaire et à la probabilité de recommander un traitement alors que celui-ci est inutile sur le site-année considéré. Ces deux probabilité sont souvent appelées *probabilité de faux négatif* et *probabilité de faux positif* respectivement. On définit également la probabilité que $f(x,\hat{\theta})$=0 si z=0 comme la *spécificité* et la probabilité que $f(x,\hat{\theta})$=1 si z=1 comme la *sensibilité* de la règle de décision.

La valeur du coût moyen dépend de la valeur $\hat{\theta}$. Lorsque $\hat{\theta}$ représente un seuil de décision, une pratique courante consiste à considérer différentes valeurs de $\hat{\theta}$ successivement, et à calculer les probabilités de faux négatif et de faux positif pour chacune des valeurs de $\hat{\theta}$ considérée. Les probabilités ainsi obtenues permettent de déterminer l'importance des deux types d'erreurs de décision pour une série de seuils de décision et non pas un seuil unique. Les résultats peuvent alors être utilisées pour identifier un seuil de décision ne conduisant ni à une trop grande probabilité de faux

positif, ni à une trop grande probabilité de faux négatifs. Le seuil choisi dépendra des niveaux d'erreur que l'utilisateur est prêt à accepter. Ce type d'analyse est appelé *Receiver Operating Characteristics* (ROC). L'analyse ROC est très utilisée en médecine pour évaluer les tests de diagnostic (Swets, 1988) et, dans une moindre mesure, en agronomie depuis la fin des années 1990 (Hughes *et al*., 1999). Les résultats sont généralement présentés sous la forme d'une courbe reliant « 1- prob. faux négatif » à « prob. faux positif » ou encore *sensibilité* à 1 – *specificité*.

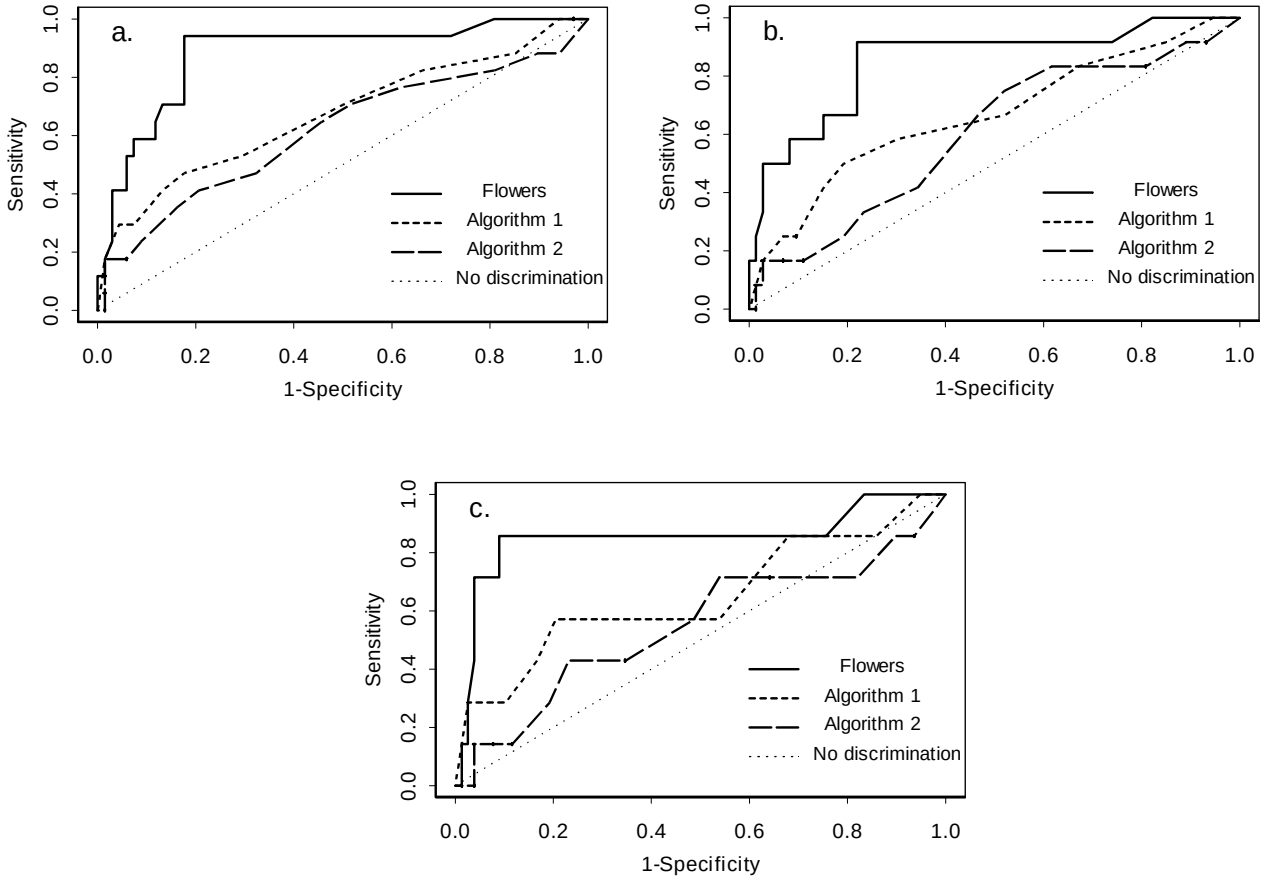

Figure 1. Sensibilités et spécificités de trois règles de décision utilisées pour raisonner le traitement du sclérotinia du colza ([14]).

Chacune des trois règles est du type "Traitement si l'indicateur est supérieur à un seuil de décision" et est basée sur un indicateur spécifique, soit le pourcentage de fleurs malade (Flower), soir une grille de risque ne prenant pas en compte le climat (Algorithm 1), soit une grille de risque intégrant le climat (Algorithm 2). Chaque point de chacune des courbes correspond à un seuil de décision particulier. Chaque figure correspond à un seuil de nuisibilité différent : incidence du sclérotinia égale à 10% (a), 20% (b), 30%(c). La règle est d'autant meilleure que sa courbe passe près du point (0, 1). La diagonale correspond à la règle « décision aléatoire ». Les résultats montrent que la règle « Flower » est nettement meilleure que les deux autres.

Makowski *et al*. [14] ont comparé la probabilité d'obtenir des recommandations erronées lorsque des indicateurs de risque sont utilisés pour raisonner les traitements fongicides contre le sclerotinia (*Sclerotinia sclerotiorum*, Lib., de Bary) du colza (*Brassica napus* L.). Les résultats de l'analyse ROC (Figure 1) ont montré que la probabilité d'erreur était plus élevée avec des indicateurs de type « grille de risque » qu'avec l'indicateur « mesure de pourcentage de fleurs malades ». Parmi les deux indicateurs de type « grille de risque » testés, celui ne prenant pas en compte le climat s'est avéré être le plus performant. De telles informations permettent aux utilisateurs de tenir compte des erreurs associées à l'utilisation de tels indicateurs pour raisonner les traitements fongicides.

Une méthode bayésienne a également été proposée pour réaliser des analyses ROC avec des indicateurs de risque de maladie basés sur des mesures de nombre/proportion d'organes malades [27] (encadré 1). Cette méthode permet de quantifier l'intérêt potentiel d'une amélioration de la précision de ces mesures. Elle permet aussi de déterminer un seuil optimal de décision en fonction des valeurs de sensibilité et de spécificité (Figure 2). La méthode proposée dans l'encadré 1 ci-dessous pourrait être généralisée pour traiter les situations où des données sont manquantes, notamment les cas où la mesure de référence (ici une mesure du niveau de la maladie à la récolte) n'est pas disponible.

Encadré 1 - Analyse ROC d'indicateurs basés sur un nombre d'organes malades.

Supposons que des données ont été obtenues sur m_D sites-années avec un niveau élevé de maladie à la récolte et sur $m_{\overline{D}}$ sites-années avec un faible niveau de maladie. y_{Di} est le nombre d'organes malades mesurés parmi n_i organes collectés avant la date du traitement chimique dans le *i*ème site-année avec un fort niveau de maladie, *i*=1, …, m_D, et $y_{\overline{D}j}$ représente le nombre d'organes malades parmi les n_j organes collectés dans le *j*ème site-année avec un faible niveau de maladie, *j*=1, …, $m_{\overline{D}}$. Un traitement est recommandé si le nombre d'organes malades *y* est supérieur à un seuil de décision.

Niveau 1: intra site-année. On suppose que les distributions de y_{Di} et $y_{\overline{D}j}$ sont binomiales.

$$y_{Di}\left|\theta_{Di}\right. \sim Bin\left(n_i, \theta_{Di}\right) \qquad i=1, \ldots, m_D,$$

$$y_{\overline{D}j}\left|\theta_{\overline{D}j}\right. \sim Bin\left(n_j, \theta_{\overline{D}j}\right) \qquad j=1, \ldots, m_{\overline{D}}$$

avec θ_{Di} et $\theta_{\overline{D}j}$ les probabilités qu'un organe soit malade dans les *i*ème site-année et *j*ème site-année respectivement. y_{Di}, *i*=1, …, m_D, et $y_{\overline{D}j}$, *j*=1, …, $m_{\overline{D}}$ sont supposés indépendant.

Niveau 2: inter sites-années. On suppose que les distributions des logit de θ_{Di} et $\theta_{\overline{D}j}$ sont gaussiennes.

$$\text{logit}\left(\theta_{Di}\right)\left|\mu_D, \sigma_D^2\right. \sim N\left(\mu_D, \sigma_D^2\right) \text{ iid} \qquad i=1, \ldots, m_D,$$

$$\text{logit}\left(\theta_{\overline{D}j}\right)\left|\mu_{\overline{D}}, \sigma_{\overline{D}}^2\right. \sim N\left(\mu_{\overline{D}}, \sigma_{\overline{D}}^2\right) \text{ iid} \qquad j=1, \ldots, m_{\overline{D}}.$$

Niveau 3: *distributions a priori*. On suppose des distributions a priori indépendantes pour μ_D, $\mu_{\overline{D}}$, $1/\sigma_D^2$, et $1/\sigma_{\overline{D}}^2$. On utilise $N(0, 10^6)$ pour μ_D et $\mu_{\overline{D}}$, et gamma(0.001, 0.001) pour $1/\sigma_D^2$ et $1/\sigma_{\overline{D}}^2$.

Ce modèle bayésien permet de calculer la sensibilité et la spécificité des décisions basées sur *y* en fonction du seuil de décision et de la taille de l'échantillon (nombre total d'organes prélevés).

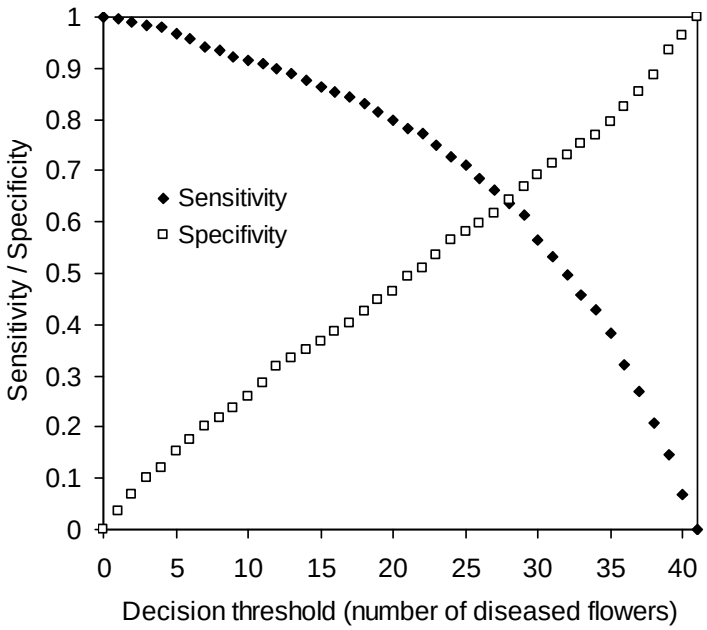

Figure 2. Valeurs de sensibilité et de spécificité obtenues pour une règle de decision utilisée pour raisonner le traitement du sclérotinia du colza en fonction du seuil de décision [27].

La règle est définie par "Traitement si l'indicateur est supérieur à un seuil de décision". L'indicateur est le nombre de fleurs malades parmi 40 fleurs prélevées dans une parcelle de colza. Le seuil correspond à un nombre de fleurs malades. Il est compris entre 0 et 41. Une valeur de sensibilité et une valeur de spécificité ont été calculées à partir de 571 données expérimentales par une méthode bayésienne (*encadré 1*).

Ces applications et développement méthodologiques montrent que l'analyse ROC constitue un outil très puissant pour les agronomes. Ce type d'analyse statistique permet, en effet, à la fois de comparer un grand nombre de modèles et de raisonner le choix d'un seuil de décision en tenant compte des deux types d'erreur possibles (faux positifs, faux négatifs). L'analyse ROC a ainsi été récemment utilisée pour évaluer une série d'indicateurs de risque d'invasion par les mauvaises herbes [18]. L'analyse ROC pourrait à terme être appliquée à d'autres types de modèles ayant des sorties binaires ou pouvant être transformées de façon à devenir binaires comme, par exemple, les modèles utilisés pour prédire les risques de pollution de l'eau (e.g teneur en nitrates supérieure à 50 mg/l ou non), pour prédire la qualité des produits (e.g teneur en protéines des grains supérieure à 11.5% ou non), ou pour prédire les risques de contamination des grains par des mycotoxines. D'une façon plus générale, l'analyse ROC pourrait aider les agronomes à identifier le type d'information (mesures « plante », mesures « sol », techniques agricoles…) le plus efficace pour classer des parcelles agricoles en deux catégories.

3.4. Perte économique

Lorsque z représente une variable décisionnelle (e.g une dose d'engrais, une décision de traiter contre une maladie), il est intéressant d'évaluer la prédiction $f(x,\hat{\theta})$ de la valeur optimale de cette variable en utilisant une autre fonction de coût, la *perte économique*. Cette fonction de coût peut être définie de la façon suivante :

$$L[z, f(x,\hat{\theta})] = J(z) - J[f(x,\hat{\theta})] \quad (8)$$

où J est une fonction objectif. Une fonction objectif usuelle est la marge brute définie par

$$J(z) = p \times y(z) - c \times z \quad (9)$$

avec $y(z)$ le rendement obtenu avec la décision z, p le prix d'une unité de rendement et c le coût lié à la décision. Par exemple, si z correspond à une dose d'engrais appliquée sur une parcelle, c représente le coût d'une unité d'engrais.

Le coût (8) correspond à la perte économique qui résulte de l'application d'une approximation de la décision optimale, $f(x,\hat{\theta})$, plutôt que de la vraie décision optimale z. Si l'approximation est parfaite, $f(x,\hat{\theta}) = z$ et $L[z, f(x,\hat{\theta})]$=0. Sinon, la perte est non nulle et est d'autant plus grande que l'approximation est mauvaise. La perte (8) permet de quantifier l'intérêt d'améliorer la prédiction $f(x,\hat{\theta})$. Comme indiqué dans la section 3.1, l'équation (8) peut être généralisée au cas où plusieurs sites-années sont considérées. Il est également possible de définir un *risque économique* qui correspond à l'espérance de la perte lorsque $\hat{\theta}$ est aléatoire.

La valeur de la fonction objectif $J[f(x,\hat{\theta})]$, obtenue en appliquant les décisions calculées avec le modèle, peut être déterminée à partir de données expérimentales. Par exemple, Makowski *et al*. [13] ont comparé deux versions de la méthode du bilan utilisée pour déterminer les doses d'engrais azoté à appliquer sur les cultures de colza. La première version permet de calculer la dose d'engrais en fonction de la biomasse de colza mesurée à la fin de l'hiver. La deuxième est plus complexe et permet de calculer la dose en tenant compte à la fois de la biomasse de colza et d'une mesure d'azote minéral du sol mesurée à la même date. Des données expérimentales issues de 53 essais ont été utilisées pour estimer la marge brute résultant de l'application des recommandations issues de ces deux types de bilan. Les résultats montrent que, selon le prix de vente du colza, la marge brute moyenne est comprise entre 596 et 722 euro ha^{-1} et entre 600 et 728 euros ha^{-1} pour les versions complexe et simple du bilan respectivement. Ces résultats montrent qu'il n'est pas intéressant sur le plan économique d'utiliser la version complexe, d'autant plus que les calculs réalisés ne tiennent pas compte du coût de la mesure d'azote du sol.

Le calcul de la perte économique (8) nécessite de pouvoir calculer également la performance des vraies décisions optimales $J(z)$. Celle-ci est inconnue mais il est possible d'estimer $J(z)$ en utilisant un modèle et une description statistique de son erreur. Cette approche a été appliquée par Makowski et Wallach [5] pour évaluer les pertes économiques liées à l'utilisation de doses d'engrais optimales calculées avec un modèle statique simulant la réponse du rendement du blé. Dans ce travail, le modèle de réponse a été défini comme un modèle à paramètres aléatoires ce qui a permis de décrire ses erreurs puis d'estimer $J(z)$, la perte économique et le risque. Une application de cette approche a déjà été présentée dans la section 2.3.

4. Comment réduire les erreurs des modèles ?

Lorsqu'on se place dans le cadre de la théorie de la décision, réduire les erreurs des modèles revient à réduire les valeurs des fonctions de coût et à réduire le risque. Un modèle comporte quatre éléments : i) des équations reliant des variables d'entrée à des variables de sortie, ii) des paramètres dont les valeurs sont inconnues et doivent donc être estimées, iii) des variables d'entrée renseignées pour chacune des situations où le modèle est utilisé, iv) des variables de sortie simulées par le modèle. Il est possible de jouer sur chacun de ces éléments pour essayer d'améliorer les performances d'un modèle. Mes travaux ont essentiellement porté sur l'utilisation de données expérimentales pour estimer les paramètres des modèles et pour corriger les prédictions en cours de saison.

4.1. Problèmes spécifiques posés par l'estimation des paramètres des modèles utilisés en agronomie

L'estimation des paramètres des modèles pose plusieurs problèmes, notamment la prise en compte des corrélations des mesures (induites par des mesures répétées sur un même site-année), la sélection des paramètres à estimer dans le cas des modèles complexes, l'estimation des paramètres des courbes enveloppes. Des méthodes ont été comparées pour traiter chacun de ces problèmes.

4.1.1. Prise en compte des corrélations entre observations

Les données disponibles pour estimer les paramètres sont souvent des mesures répétées sur un même ensemble de parcelles expérimentales. Ce sont des données de type *longitudinales*. Par exemple, des mesures de biomasse sont souvent réalisées par les agronomes à différentes dates dans le cycle de la culture sur un site-année donné. Dans ce cas, les mesures obtenues ne sont pas indépendantes car les caractéristiques du site et de l'année ont une influence sur l'ensemble des mesures réalisées. Cette corrélation des mesures complique l'estimation des paramètres. Des résidus obtenus avec le modèle Azodyn sont présentés sur la figure 3 à titre d'illustration [32]. Les résidus présentés sur cette figure sont tous positifs pour l'année 1995 et tous négatifs pour l'année 1992. Ils ne peuvent donc pas être raisonnablement considérés comme issus d'erreurs indépendantes.

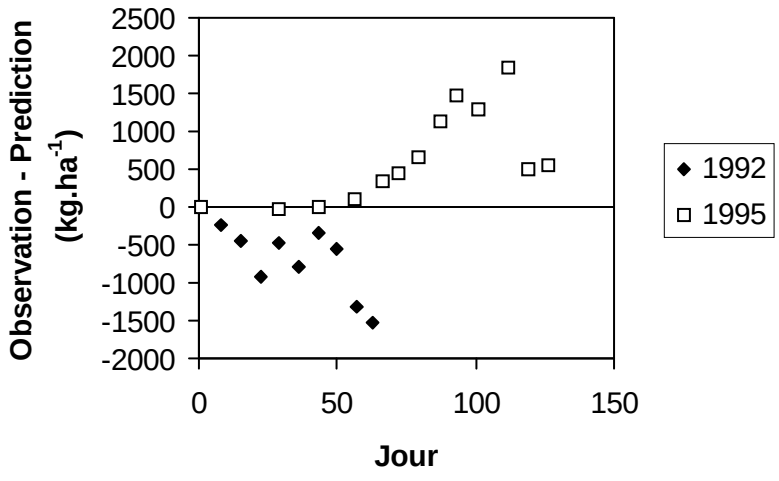

Figure 3. Résidus obtenus pour les prédictions de biomasse du blé obtenues avec Azodyn pour deux années sur le site de Grignon [32].

Ces résidus ont été obtenus après estimation de certains paramètres du modèle par la méthode des moindres carrés pondérés.

Il est important de prendre en compte les corrélations des erreurs pour estimer les paramètres d'un modèle. En effet, la qualité des estimateurs dépend de la distribution des erreurs. Si une hypothèse irréaliste est faite sur cette distribution, l'estimation des paramètres n'est plus optimale et la variance des estimateurs risque d'être sous-estimée. Ce problème peut se rencontrer si une méthode comme les moindres carrés pondérés est utilisée pour estimer les paramètres d'un modèle. Ce type de méthode donne des résultats optimaux si les résidus sont indépendants mais, comme le montre la figure 3, ce n'est pas toujours le cas, notamment lorsque des données longitudinales sont utilisées.

La prise en compte des corrélations est un problème souvent difficile [32]. Dans le cas de modèles relativement simples, incluant un nombre de paramètres inférieur à 10, il est possible d'estimer les paramètres en définissant le modèle comme un modèle non linéaire mixte (Davidian et Giltinan, 2003 ; Kuhn et Lavielle, 2004 ; Pinheiro and Bates, 2000). Cette approche a été appliquée pour estimer les paramètres de modèles non linéaires simulant la réponse à la dose d'engrais ([4], [9]) et utilisés pour calculer des doses d'engrais optimales (encadré 2, et figure 4).

Encadré 2 – Modèle mixte pour simuler la réponse du rendement à la dose d'engrais

Soit y_{ij} la *j*ème mesure de rendement obtenue sur le *i*ème site-année pour une dose d'engrais d_{ij} ;

i=1, …, *n*, *j*=1, …, m_i, où m_i est le nombre total de mesures du site-année *i*. Un modèle de réponse peut être défini comme un modèle hiérarchique à deux niveaux:

Niveau 1: intra site-année. $y_{ij} = f(d_{ij};\phi_i) + \varepsilon_{ij}$, *j*=1, …, m_i.

La fonction f décrit la réponse intra site-année à la dose d'engrais et inclut un vecteur de paramètres ϕ_i spécifique du site-année *i*. L'erreur intra site-année, $\varepsilon_{ij} = y_{ij} - f\left(d_{ij};\phi_i\right)$, est supposée distribuée selon une loi normale $\varepsilon_{ij} \sim N(0,\sigma_y^2)\ i.i.d$

Niveau 2: inter site-année. $\phi_i = A_i\mu + \eta_i$ avec $\eta_i \sim N\left(0,\Gamma\right)\ i.i.d$, *i*=1, …, *n*,

où A_i est une matrice spécifique du *i*ème site-année, supposée connue, μ est un vecteur de paramètres fixe, η_i est un vecteur d'effets aléatoires, Γ est une matrice de variance-covariance.

Ces deux niveaux du modèle peuvent éventuellement être complétés par un troisième niveau lorsqu'on souhaite se placer dans un cadre bayésien. Le troisième niveau permet, dans ce cas, de définir des distributions a priori pour les paramètres à estimer μ , Γ , σ_y (e.g Carlin et Louis, 2000).

Makowski et Wallach [6] ont évalué la qualité des doses optimales calculées avec un modèle de réponse lorsque les paramètres sont estimés soit avec la méthode des moindres carrés, soit avec un modèle mixte. Un critère de type marge brute moyenne, estimé avec une méthode non paramétrique à l'aide de données expérimentales, a été utilisé pour comparer les doses optimales. Les résultats ont montré que la marge moyenne était plus élevée lorsque la deuxième méthode d'estimation était utilisée : l'utilisation d'un modèle mixte plutôt que la méthode des moindres carrés a ainsi permis d'augmenter la marge brute moyenne de 8 à 49 euros ha^{-1} selon le prix considéré.

Makowski et Lavielle [19] ont, par ailleurs, comparé deux méthodes pour estimer les paramètres d'un modèle non linéaire mixte, une approche basée sur une linéarisation du modèle non linéaire et une approche basée sur une version stochastique de l'algorithme EM proposée par Kuhn et Lavielle (2005). Les résultats de cette étude ont montré que les estimations étaient plus précises avec la seconde méthode. Une application de la version stochastique de l'algorithme EM est présentée sur la figure 4.

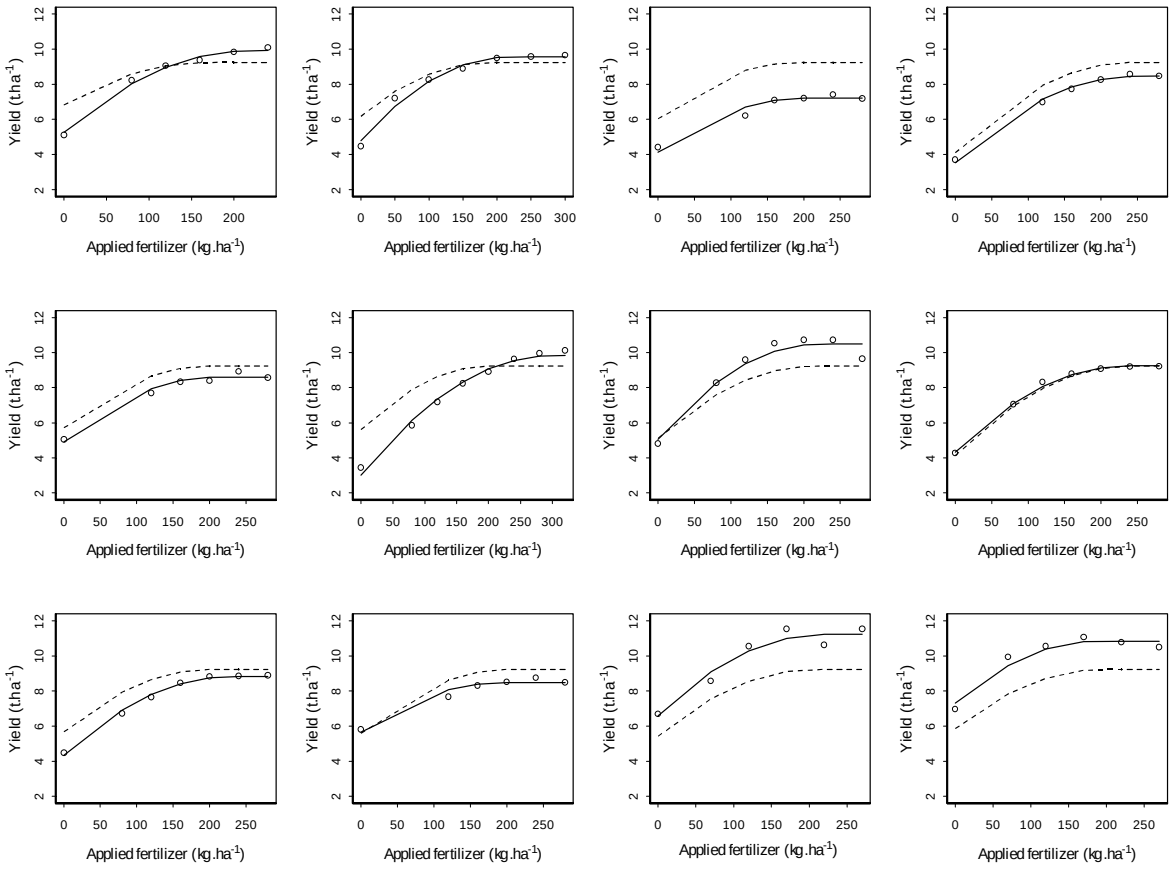

Figure 4. Résultat de l'ajustement d'un modèle quadratique-plus-plateau à douze essais expérimentaux décrivant la réponse du rendement du blé à la dose d'engrais azoté [19]. Les paramètres du modèle ont été estimés avec la méthode proposée par Kuhn et Lavielle (2005).

Les points correspondent aux observations. Le modèle décrit dans l'encadré 2 a été utilisé avec une fonction f quadratique-plus-plateau. L'ajustement obtenu pour la population de parcelles correspond aux courbes en pointillé. Ces ajustements varient légèrement d'un essai à l'autre car le modèle inclut le reliquat d'azote mineral sortie hiver en entrée. Les courbes continues correspondent aux ajustements essai par essai. Des simulations numériques ont montré que les estimateurs des paramètres étaient plus précis avec la méthode de Kuhn et Lavielle (2005) qu'avec la fonction S+/R nlme [19].

4.1.2. Sélection des paramètres à estimer

Il est généralement impossible d'estimer simultanément tous les paramètres d'un modèle non linéaire dès que ceux-ci sont nombreux (>20). Il est alors nécessaire de sélectionner les paramètres qu'il est important d'estimer à partir de données et de fixer les autres à des valeurs de références. Parmi les diverses méthodes possibles, l'analyse de sensibilité globale est une approche attractive car elle permet de hiérarchiser les paramètres d'un modèle selon leurs effets sur les variables de sortie en tenant compte de l'existence d'interactions éventuelles (encadré 3).

Encadré 3 – Le principe de l'analyse de sensibilité globale

Notons Y une variable de sortie d'un modèle déterministe. La variance totale de Y, induite en faisant varier simultanément *p* paramètres du modèle dans leurs gammes d'incertitude, est décomposée de la façon suivante (*e.g* Saltelli *et al*. 2000): $V(Y) = \sum_{i=1}^{p} V_i + \sum_{i<j} V_{ij} + ... + V_{1,2,...,p}$, où $V(Y)$ est la variance totale, $V_i = V\left[E\left(Y|x_i\right)\right]$ mesure l'effet principal du paramètre x_i, *i*=1, …, *p*, et les autres termes mesurent les interactions. Deux types d'indices peuvent être définis à partir de cette décomposition: $S_i = \frac{V_i}{V(Y)}$ et $S_{Ti} = \frac{V(Y) - V_{-i}}{V(Y)}$, où V_{-i} est la somme de tous les termes de variance qui n'incluent pas l'indice *i*.

S_i est *l'indice de sensibilité de premier ordre* pour le $i^{ème}$ paramètre. Cet indice représente l'effet principal du paramètre x_i sur la variable de sortie Y et mesure la réduction de variance qui serait obtenue en fixant ce paramètre. S_{Ti} est *l'indice total* pour le $i^{ème}$ paramètre x_i. Cet indice prend en compte les interactions entre paramètres. Il correspond à la fraction moyenne de variance qui resterait si seul le paramètre x_i restait inconnu.

Deux méthodes ont été comparées pour estimer les indices de sensibilité avec le modèle Azodyn [20] : la méthode *FAST étendue* et la méthode *Winding stairs*. Les résultats ont montré que les deux méthodes conduisaient à des indices de sensibilité similaires mais que la méthode *FAST étendue* était plus économique en temps de calcul et donc plus précise pour un temps de calcul donné. Une application de cette méthode à un modèle dynamique simulant la biomasse du blé est présentée sur la figure 5 [33].

Les méthodes d'analyse de sensibilité globales constituent des outils séduisants pour identifier les paramètres ayant une forte influence sur les variables de sortie. Elles nécessitent cependant la définition de gammes d'incertitude pour les valeurs paramètres ce qui n'est pas toujours facile en pratique D'autre part, leur intérêt réel pour identifier les paramètres à estimer dans un modèle complexe n'a été que peu étudié. La thèse de M. Lamboni, co-encadrée par Hervé Monod (unité MIA INRA Jouy-en-Josas) et par moi-même depuis novembre 2006, a pour objectif de déterminer si les méthodes d'analyse de sensibilité globale peuvent réellement conduire à une meilleure estimation des paramètres des modèles dynamiques complexes utilisés en agronomie.

Les méthodes d'analyse de sensibilité globale ne doivent pas faire oublier les analyses de sensibilité locales. Celles-ci peuvent également être utiles. Elles permettent de déterminer l'influence d'un paramètre sur les variables de sortie localement autour d'une valeur particulière du paramètre. Ces techniques locales sont utiles pour comprendre le fonctionnement du modèle mais elles sont peu opérationnelles pour analyser l'effet simultané d'un grand nombre de paramètres. Makowski *et al* [22] ont illustré l'intérêt de l'analyse de sensibilité locale avec un modèle permettant d'optimiser des stratégies de gestion de la biodiversité. Cette approche a permis de comparer l'intérêt de deux

stratégies : cultiver une grande surface de manière extensive (moins de pesticides, densités de semis plus faibles…) ou réduire la surface cultivée tout en utilisant des techniques agricoles plus intensives.

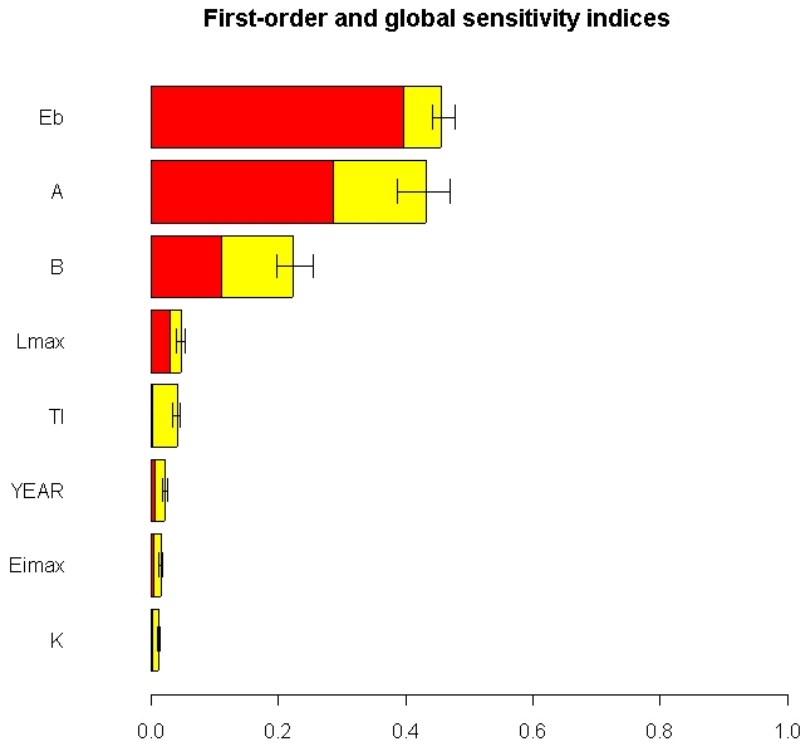

Figure 5. Indices de sensibilité de premier ordre et totaux estimés avec la méthode FAST étendu pour les 8 paramètres du modèle WWDM simulant la biomasse du blé en fonction du climat (Year) [33].

Ces indices ont été estimés vingt fois de suite avec 8×1000 *simulations à chaque fois; la première partie de chaque barre correspond à la moyenne (sur les 20 estimations) des indices de premier ordre et chaque barre horizontale complète correspond à la moyenne des indices totaux. Les lignes noires indiquent les gammes de variation des indices totaux sur les 20 estimateurs.*

4.1.3. Estimation des paramètres des courbes enveloppes

Les courbes enveloppes sont souvent utilisées en agronomie et en écologie (Webb, 1972; Fleury, 1991; Doré *et al*., 1997; Brancourt-Hulmel *et al*., 1999; Cade *et al*., 1999 ; Casanova *et al*., 1999). Elles correspondent à des fonctions $f(x;\theta)$ reliant une variable d'intérêt agronomique y (e.g un nombre de grains par m² ou un rendement) à une autre variable x (e.g un facteur limitant ou une composante du rendement autre que y) (Figure 6), avec θ un vecteur de paramètres.

Quand y est mesurée sans erreur, la courbe enveloppe $f(x;\theta)$ correspond à la plus grande valeur possible de y pour une valeur x donnée. La différence entre y et $f(x;\theta)$ est alors due à un ou plusieurs facteurs limitants (maladie, stress azoté…) autres que x. Quand y est mesurée avec erreur, y peut être supérieure à $f(x;\theta)$ mais seulement à cause d'une erreur de mesure.

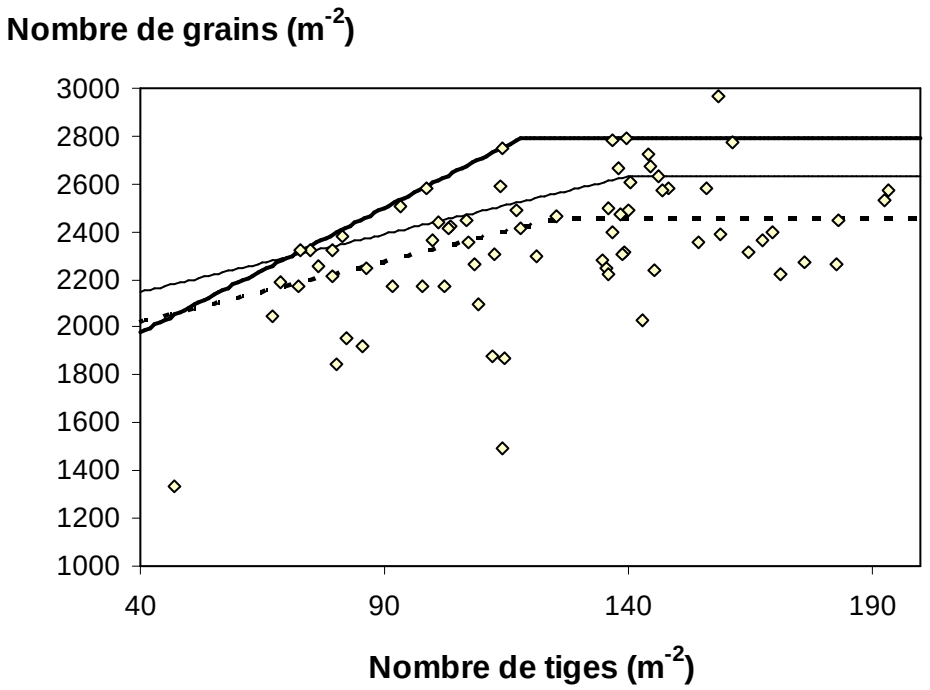

Figure 6. Ajustement d'une courbe enveloppe linéaire-plus-plateau pour les quantiles 0.5 (pointillés), 0.80 (ligne continue), et 0.96 (en gras) [21]. Données pois de Doré *et al*. (1998) : 71 parcelles de pois localisées dans le bassin parisien en 1988, 1989, et 1990.

L'estimation des paramètres θ des courbes enveloppes est un problème important car ces courbes sont souvent utilisées dans le cadre du diagnostic agronomique (Doré *et al*., 1997) et, également, dans des modèles simulant le développement de culture (e.g Gonzalez Montaner *et al*., 1997). Dans [21], je montre que les techniques de régression quantile peuvent être utilisées pour estimer ces paramètres. Le principe est de définir un quantile τ tel que $\tau = P[Y < f(X;\theta)]$, puis d'estimer θ en minimisant la fonction suivante (Koenker et Basset, 1978) :

$$L(\theta) = \sum_{i=1}^{N} \rho_\tau [y_i - f(x_i;\theta)] \qquad (10)$$

où y_i et x_i sont les mesures obtenues dans la $i^{ème}$ parcelle, i=1, …, N.

La fonction $\rho_\tau[.]$ est définie par :

$$\rho_\tau[y_i - f(x_i;\theta)] = \tau \times [y_i - f(x_i;\theta)] \times 1_{\{y_i \geq f(xi;\theta))\}} - (1-\tau) \times [y_i - f(x_i;\theta)] \times 1_{\{y_i < f(xi;\theta))\}} \qquad (11)$$

où $1_{\{v\}} = 1$ si la condition v est vraie et $1_{\{v\}} = 0$ sinon. En général, $f(x;\theta)$ est non linéaire. Dans ce cas, les paramètres peuvent être estimés avec l'algorithme de Koenker et Park (1996). Une application est présentée sur la figure 6 [21].

Cette méthode a été comparée à d'autres techniques d'estimation et les résultats ont montré que, dans de nombreuses situations, la régression quantile permet d'obtenir des estimateurs de θ ayant un biais et une variance plus faibles [21]. Cependant, pour pouvoir appliquer cette technique il est nécessaire de définir le quantile τ . Makowski *et al*. [21] ont proposé de l'estimer en modélisant la relation entre *x* et *y* de la façon suivante :

$$y = w \times f(x;\theta) + \varepsilon \qquad (12)$$

où *w* correspond à l'effet global des facteurs limitants autre que *x*. Ainsi, *w* est compris entre 0 et 1 et sa variabilité peut être décrite par une loi beta : *w* ~ Beta(α, β). Ce type de loi est très flexible. Le terme résiduel peut être modélisé par $\varepsilon \sim N(0,\sigma^2)$. La valeur du quantile τ associée à la courbe enveloppe est directement liée aux distributions de *w* et ε . La valeur de τ est d'autant plus élevée que ε est petit par rapport à *w*. L'équation (12) peut être utilisée pour estimer τ en faisant des hypothèses sur α, β et σ. J'envisage de perfectionner cette méthode en développant une approche bayésienne pour estimer θ , α, β et σ simultanément.

4.2. Filtrage des modèles dynamiques

Les modèles de culture dynamiques permettent de prédire l'état d'un couvert végétal et de son environnement à un pas de temps journalier. L'objectif de la thèse de C. Naud [24, 25] était d'étudier l'intérêt de corriger les prédictions d'un de ces modèles avec des mesures expérimentales obtenues en cours de saison. En raison du développement des techniques de télédétection, il est de plus en plus facile d'obtenir des mesures caractérisant l'état du couvert végétal à une date donnée. Ces mesures peuvent permettre à l'utilisateur d'ajuster les variables d'état simulées par un modèle aux caractéristiques du milieu et, ainsi, de réduire les erreurs de prédiction.

Le problème pratique considéré dans la thèse concernait la prédiction de l'indice de nutrition azotée (INN) du blé par le modèle dynamique Azodyn (un indice INN<1 indique une carence azotée). Une méthode statistique bayésienne, appelée *filtrage particulaire avec interaction* (Del Moral *et al*., 1999 ; Doucet *et al*, 2001), a été utilisée pour corriger séquentiellement trois variables d'état simulées par Azodyn en utilisant des mesures d'azote absorbé et/ou de biomasse à différentes dates dans le cycle de la culture (encadré 4). Cette méthode consiste à estimer la distribution a posteriori des variables d'état par échantillonnage d'importance à chaque date de mesure. Contrairement à de nombreuses autres méthodes d'assimilation de données, elle ne nécessite pas de linéarisation du modèle et peut donc être appliquée directement à un modèle non linéaire. Son application nécessite cependant la réalisation d'un grand nombre de simulations de Monte Carlo pouvant conduire à des temps de calcul élevés. Cette méthode n'avait jamais été appliquée à des modèles de culture dynamiques complexes.

La première étape du travail a consisté à définir Azodyn comme un modèle stochastique en ajoutant un terme aléatoire aux variables d'état simulées par le modèle (Encadré 4, Figure 7). Le filtre particulaire avec interaction a ensuite été programmé afin de pouvoir être appliqué à Azodyn avec un nombre variable de mesures. Ce programme a été utilisé pour étudier i) la sensibilité du filtre au modèle d'erreur ajouté au modèle dynamique, ii) le nombre de simulations requis pour obtenir des résultats fiables, iii) l'intérêt du filtre pour réduire les erreurs de prédiction du modèle en fonction du nombre et du type de mesures disponibles (biomasse, azote absorbé, transmittance du HN-Tester).

Les résultats obtenus avec le filtre particulaire sont sensibles aux variances des termes d'erreur du modèle [24]. Des variances élevées rendent le filtre plus performant car elles permettent de générer

des valeurs mieux réparties dans l'espace des variables d'état. Les résultats montrent également que 10000 simulations de Monte Carlo sont suffisantes pour obtenir des estimations stables [24].

Une autre série de simulations a révélé qu'il était possible de réduire substantiellement les erreurs de prédiction en corrigeant les variables d'état du modèle soit avec une mesure d'azote absorbé (Figure 8) [24], soit avec une mesure de transmittance [25]. La RMSE (Root Mean Squared Error) des prédictions d'INN a ainsi pu être réduite de 18 à 53% selon les cas par rapport aux prédictions d'Azodyn non corrigées.

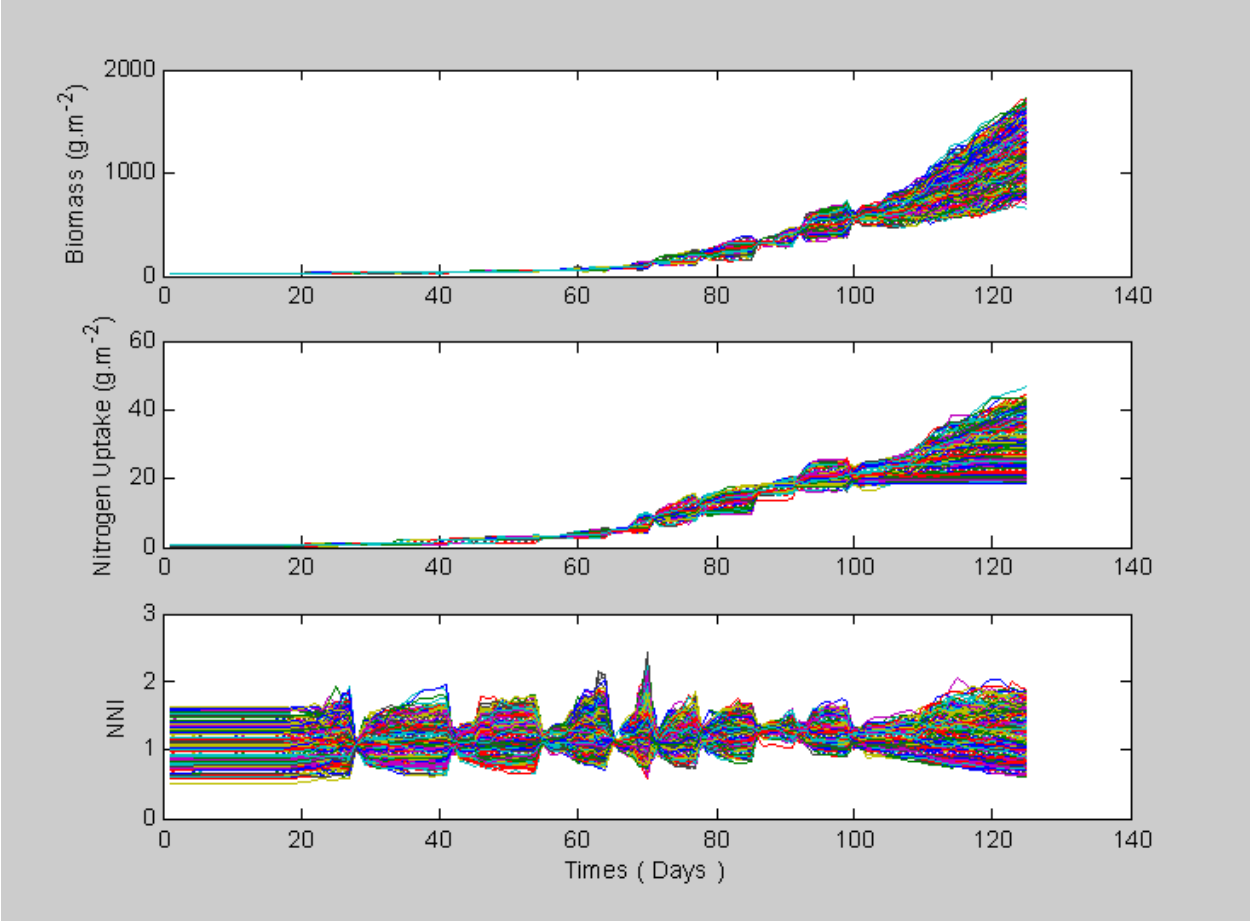

Figure 7. Ensemble des simulations réalisées avec une version stochastique du modèle Azodyn pour une parcelle de blé de Grignon en 1998 [24].

Des termes d'erreur aléatoires ont été ajoutés à trois variables d'état du modèle simulées à un pas de temps journalier. Les variables d'état présentées sur cette figure sont la biomasse du blé, l'azote absorbé et l'indice de nutrition azoté (NNI). Les étranglements visibles sur la figure correspondent aux dates de mesure. A ces dates, des mesures expérimentales sont utilisées pour filtrer les simulations du modèle et seules les simulations relativement proches des mesures sont conservées.

Encadré 4 – Le filtre particulaire appliqué à trois variables d'état du modèle Azodyn.

Une étape préalable a consisté à transformer le modèle Azodyn en modèle stochastique. La version stochastique d'Azodyn est définie par : $X_t = \begin{pmatrix} B_t \\ NU_t \\ CumN_t \end{pmatrix} = X_{t-1} + \begin{bmatrix} f_1(B_{t-1}) \\ f_2(NU_{t-1}) \\ f_3(t-1) \end{bmatrix} + \varepsilon_t$

$= X_{t-1} + f(X_{t-1}) + \varepsilon_t$

avec f la fonction utilisée par Azodyn et ε_t la composante stochastique, $\varepsilon_t \sim N[(0,0,0)^T, \Sigma]$. B_t, NU_t, $CumN_t$ correspondent à trois variables (biomasse, azote absorbé, azote minéral cumulé du sol) simulées par le modèle à un pas de temps journalier t. Le filtre particulaire a été appliqué de la façon suivante :

1. *Initialisation*. Soit N le nombre de simulations de Monte Carlo. Pour $i = 1,...,N$, générer aléatoirement N valeurs pour le vecteur des variables d'état à t=0, $X_0^{(i)} = \left(B_0^{(i)}, NU_0^{(i)}, CumN_0\right)^T$ avec $\begin{pmatrix} B_0^{(i)} \\ NU_0^{(i)} \end{pmatrix} \sim N(\mu_0, \Sigma_0)$, où $\mu_0 = \begin{pmatrix} m_0^B \\ m_0^{NU} \end{pmatrix}$ et $CumN_0$ sont fixés. Chaque valeur générée est appelée 'particule' et $\hat{w}_0^{(i)} = \frac{1}{N}$ est le poids initial de la $i$ème particule.
2. *Propagation*. Les N particules sont propagées avec les équations du modèle Azodyn jusqu'à la date de la prochaine mesure. Pour $i = 1,...,N$, générer aléatoirement $\tilde{X}_t^{(i)} \sim N[X_{t-1}^{(i)} + f(X_{t-1}^{(i)}), \Sigma_t]$ et définir $\hat{w}_t^{(i)} = \frac{1}{N}$.
3. *Correction*. Les poids des particules sont mises à jour avec les mesures. Soit $t = t_j$, $j \in \{1,...,k\}$, la date de la $j$ème mesure m_{tj}. Pour $i = 1,...,N$, calculer $w_{t_j}^{(i)} = \hat{w}_{t_j-1}^{(i)} P\left(m_{t_j} \mid \tilde{X}_{t_j}^{(i)}\right)$ où $P\left(m_{t_t} \mid \tilde{X}_{t_j}^{(i)}\right)$ représente la vraisemblance de la mesure c'est-à-dire la densité de probabilité de la mesure conditionnellement à une valeur des variables d'état du modèle. Calculer les poids normalisés, $\hat{w}_{t_j}^{(i)} = \frac{w_{t_j}^{(i)}}{\sum_{l=1,..,N} w_{t_j}^{(l)}}$, i=1, …, N.
4. *Ré-échantillonnage*. Générer $X_{t_j}^{(i)}$ selon une distribution de probabilité discrète définie par $\left(\tilde{X}_{t_j}^{(i)}, \hat{w}_{t_j}^{(i)}\right)$ i=1, …, N. A cette étape, les particules ayant des poids faibles ont tendance à être supprimées.
5. Définir $\hat{w}_{t_j}^{(i)} = \frac{1}{N}$. Retourner à l'étape 2.

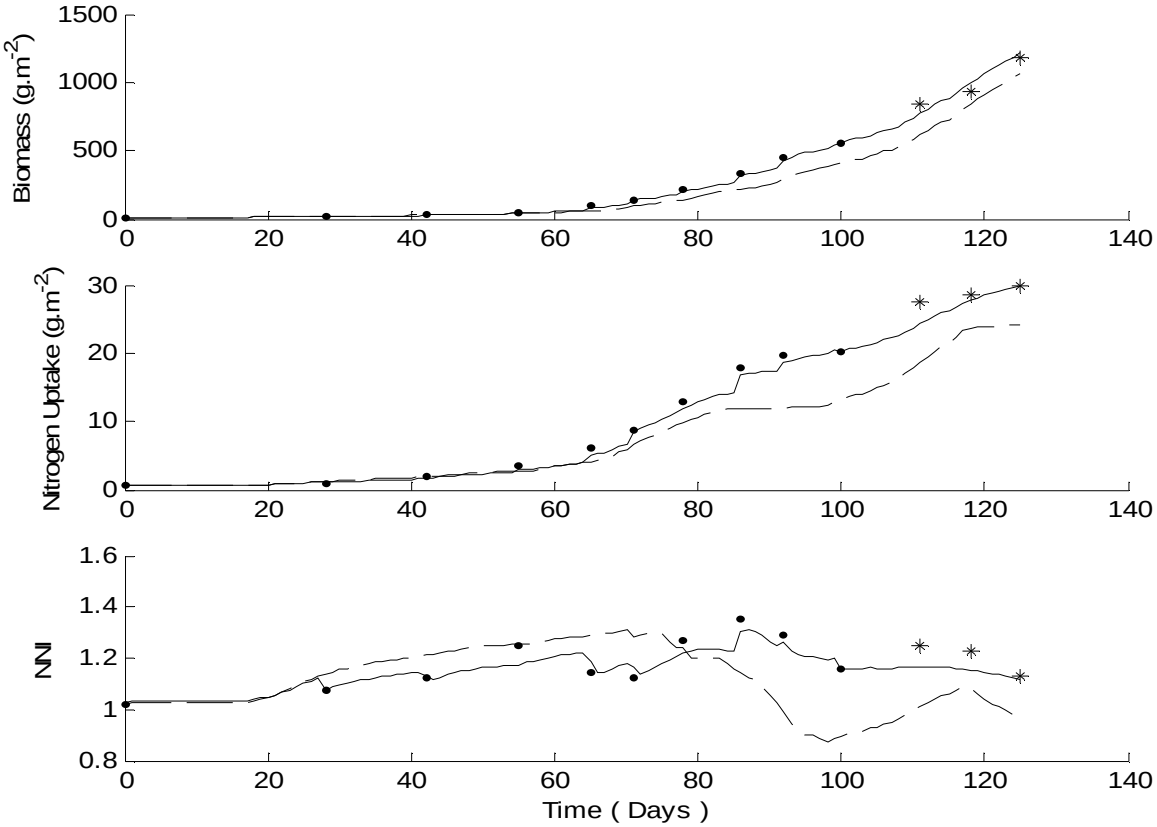

Figure 8. Prédictions initiales du modèle Azodyn en 1998 (ligne en tirés) et prédictions obtenues après correction du modèle par filtrage (moyenne de 10000 particules) (ligne continue). Les 3 dernières mesures (*) n'ont pas été utilisées pour corriger le modèle et permettent de juger de la réduction des erreurs [24].

Science is the belief in the ignorance of the experts.

- *Richard P. Feynman*.

Perspectives

1. Le projet PICSEL[2]

1.1. Contexte

Il est rare qu'un seul modèle soit disponible pour un usage donné. Bien souvent, il est possible de considérer une série de modèles M_1, ..., M_N, pour prédire une variable ou optimiser une technique. Les modèles disponibles peuvent être basés sur différentes fonctions mathématiques (linéaire, non linéaire, statique, dynamique…), différentes variables d'entrée, différentes méthodes pour estimer les paramètres (moindres carrés, maximum de vraisemblance, méthode bayesienne…). Le nombre de modèles peut être très grand. Par exemple, si on suppose qu'un agronome souhaite prédire un risque d'invasion par des mauvaises herbes dans une parcelle cultivée avec un modèle logistique (*e.g*. Primot *et al*. 2006 [18]) et si K variables explicatives sont disponibles, alors 2^K modèles peuvent être développés. En pratique, un sous-ensemble de tous les modèles possibles est considéré du fait de problèmes de calcul mais aussi de diverses considérations théoriques. Ainsi, 11 variables explicatives ont été identifiées dans l'étude [18] mais seulement 15 des 2^{11} modèles possibles ont été développés et évalués.

Lorsque plusieurs modèles sont disponibles, une approche classique consiste à sélectionner un modèle parmi les N modèles possibles. Différentes méthodes peuvent être utilisées pour sélectionner un modèle parmi N en utilisant des données expérimentales : estimation de la MSEP par validation croisée, utilisation du critère d'Akaïke, stepwise… Ces méthodes permettent de choisir un modèle parmi N en calculant la valeur d'un critère (e.g MSEP, AIC) pour chaque modèle à l'aide de données et en choisissant le modèle qui optimise le critère considéré.

Le projet PICSEL part du constat que l'incertitude dans les méthodes de sélection de modèles est généralement ignorée une fois le modèle final sélectionné (Chatfield, 1995 ; Draper, 1995). L'interprétation des valeurs estimées des paramètres et les prédictions du modèle ne sont basées que sur le modèle sélectionné. Pourtant, dans certains cas, l'instabilité des résultats de la procédure de sélection est grande. Ainsi, Yuan et Yang (2005) ont montré que, quand les erreurs des modèles sont grandes, une procédure de sélection a de grandes chances de conduire à un modèle complètement différent dès qu'une base de données légèrement différente est utilisée.

Des statisticiens (e.g Buckland *et al.* 1997 ; Burnham et Anderson, 2002) ont suggéré que, dans certains cas, il serait préférable de mélanger tous les modèles possibles plutôt que d'utiliser uniquement le modèle sélectionné. L'idée de base est d'utiliser une somme pondérée des prédictions issues des modèles disponibles plutôt que d'utiliser la prédiction du modèle supposé être le meilleur. Plusieurs méthodes ont été récemment développées pour estimer les pondérations associées à chaque modèle à partir d'une base de données (Buckland *et al.* 1997 ; Burnham et Anderson, 2002 ; Hoeting *et al.* 1999 ; Yang 2003 ; Raftery *et al* 2005 ; Yuan and Yang 2005). Ces méthodes peuvent être appliquées à une grande diversité de modèles, linéaires, logistiques, non linéaires, dynamiques.

Le mélange de modèles peut améliorer la précision des prédictions et conduire à des intervalles de confiances plus réalistes (Chatfield, 1995 ; Draper, 1995). Selon une étude récente (Yuan and Yang, 2005), le mélange de modèles est plus performant que la sélection lorsque les erreurs des modèles sont grandes. En agronomie, les erreurs des modèles peuvent représenter 20% des mesures moyennes (Naud, 2007). Le mélange de modèles pourrait donc constituer une approche intéressante dans cette discipline.

1.2. Mesurer l'instabilité des méthodes de sélection de modèles

Le premier volet de ce projet a pour but d'analyser l'instabilité des résultats des procédures de sélection de modèles lorsqu'on modifie la base de données utilisée. Trois méthodes seront utilisées

[2] Le projet PICSEL (prise en compte de l'incertitude dans la sélection des modèles) est un projet financé par l'ANR depuis novembre 2006 (appel d'offre Jeunes Chercheuses et Jeunes Chercheurs) sur trois ans. J'en suis l'initiateur et le coordinateur.

dans cette optique : (i) application d'une série de méthodes de sélection à différentes bases de données, (ii) bootstrap, et (iii) estimation du critère Perturbation Instability in Estimation (PIE).

La première méthode est très simple. Elle consiste à appliquer des procédures de sélection (par exemple, sélection parmi *N* modèles non linéaires avec la MSEP) à une série de bases de données correspondant à plusieurs populations de parcelles agricoles (e.g. différentes régions, différents climats, or différentes cultures). Chaque base de données doit inclure des mesures expérimentales de variables simulées par les modèles considérés (*e.g* rendement, teneur en protéines). Les modèles sélectionnés sont ensuite comparés. Cette approche est utile pour analyser les relations entre modèles sélectionnés, procédures de sélection et structure des bases de données. Cependant, elle ne permet pas de distinguer facilement l'effet échantillonnage de l'effet de la population.

La deuxième méthode (bootstrap) a été décrite par Buckland *et al.* (1997) et par Burnham et Anderson (2002). Le principe de base est de générer *P* bases de données de tailles identiques à partir d'une vraie base de données expérimentales par tirages aléatoires avec remise des mesures. Une ou plusieurs méthodes de sélection sont ensuite appliquées à chacune des *P* bases de données. L'instabilité des procédures de sélection est ensuite analysée en déterminant notamment la proportion des bases de données conduisant à un même modèle.

La troisième méthode (PIE) a été développée par Yuan and Yang (2005). PIE est un indice qui mesure l'instabilité des procédures de sélection lorsque des termes aléatoires sont ajoutés aux observations. L'idée est que, si une procédure de sélection est stable, une faible modification des données ne doit pas changer les prédictions des modèles sélectionnés. Une valeur élevée du PIE révèle un manque de robustesse de la procédure de sélection. Une caractéristique intéressante de cette méthode est qu'elle permet de quantifier l'effet d'une procédure de sélection sur les prédictions des modèles.

1.3. Méthodes pour mélanger les modèles

Le Bayesian Model Averaging (BMA) est une approche séduisante pour mélanger les modèles (voir Hoeting *et al.* 1999 pour une revue). Supposons que Δ est une quantité d'intérêt comme un futur rendement ou un paramètre représentant l'effet d'un facteur limitant. Le BMA a pour but de déterminer la distribution a posteriori de Δ définie par:

$$P(\Delta|D) = \sum_{i=1}^{N} P(\Delta|D, M_i) P(M_i|D) \tag{13}$$

où $M_1,...,M_N$ sont des modèles disponibles, *D* est une base de données, et *P*(.|.) désigne des distributions de probabilité conditionnelles. Les espérance et variance a posteriori de Δ sont définies par:

$$E(\Delta|D) = \sum_{i=1}^{N} w_i \hat{\Delta}_i \tag{14}$$

$$\text{var}(\Delta|D) = \sum_{i=1}^{N} w_i \, \text{var}(\Delta|D, M_i) + \sum_{i=1}^{N} w_i \left(\hat{\Delta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_i \right)^2 \tag{15}$$

= variance intra modèle + variance inter modèles

où $\hat{\Delta}_i = E(\Delta|D, M_i)$ (estimation/prédiction de la quantité d'intérêt avec le modèle M_i) et $w_i = P(M_i|D)$ (poids associé au modèle M_i).

L'équation (14) montre que la prédiction de Δ obtenue avec BMA est une somme pondérée des prédictions individuelles et l'équation (15) montre que la variance de Δ dépend de deux composantes, une représentant la variance intra (incertitude pour un modèle donné) et une représentant la variabilité entre modèles de la prédiction de Δ.

Des algorithmes ont été développés pour mettre en oeuvre BMA avec des modèles linéaires (Raftery *et al*. 1997), logistiques (Viallefond *et al.* 2001), dynamiques (Raftery *et al*. 2005). Des approches non bayésiennes (Buckland *et al*., 1997; Yang, 2003; Yuan and Yang, 2005) ont aussi été proposées.

Toutes ces approches offrent des possibilités intéressantes aux agronomes car elles permettent d'estimer l'effet de variables explicatives (e.g effet d'un facteur limitant sur le rendement) ou de prédire une variable d'intérêt (e.g probabilité que l'incidence d'une maladie soit supérieure à un seuil de nuisibilité) en combinant plusieurs modèles. Les méthodes de mélange de modèles comme l'approche BMA constituent donc des alternatives intéressantes aux méthodes de sélection de modèles.

1.4. Cas d'étude

Les techniques présentées ci-dessus seront appliquées à une série de cas d'étude couvrant la diversité des modèles utilisés en agronomie (linéaire, linéaire généralisé, non linéaire, dynamique). Le premier cas d'étude a déjà débuté. Il porte sur l'utilisation de modèles linéaires de type régression multiple pour identifier les principaux facteurs limitants du blé d'hiver. Ce type de technique est souvent utilisé dans le cadre du Yield Gap Analysis (e.g Casanova *et al*., 1999). Le principe est d'utiliser une base de données pour sélectionner un petit nombre de variables explicatives du rendement, parmi une série de variables candidates représentant des facteurs limitants potentiels.

L'objectif de ce premier d'étude est de comparer les résultats des méthodes de sélection classiques de type *stepwise* avec ceux obtenus sans sélection (estimation systématique de tous les paramètres) et avec l'approche BMA. Plusieurs types de plans d'expérience basés sur différents traitements et un nombre plus ou moins grand de mesures sont considérés. Le travail est réalisé avec Lorène Prost, actuellement en thèse à l'UMR d'agronomie de Grignon (MH Jeuffroy et M. Cerf, directeurs de thèse). Des résultats sont présentés ci-dessous pour un type de plan d'expérience (Tableaux 4-5).

Le modèle considéré ici est linéaire et défini par $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_5 x_5 + \varepsilon$ où y est une perte de rendement mesurée sur un site-année pour la variété de blé Soissons, x_1-x_5 sont des variables explicatives représentant des facteurs limitants climatiques (exemple, x_1=nombre de jours de gel), $\theta_0, ..., \theta_5$ sont des paramètres à estimer, ε est un terme d'erreur.

Une base de données incluant au total 160 sites-années de blé a été considérée. Des mesures de y et de x_1-x_5 ont été réalisées sur chaque site-année. Dix milles échantillons de 40 sites-années ont été générés par tirage aléatoire avec remise dans la base de données initiale. Les valeurs des paramètres ont été estimées pour chacun des 10000 échantillons en utilisant successivement quatre méthodes :

i) pas de sélection des variables (tous les paramètres sont estimés systématiquement),

ii) sélection des variables par stepwise avec le critère AIC en partant du plus petit modèle (Stepwise AIC),

iii) sélection des variables par stepwise avec le critère BIC en partant du plus petit modèle (Stepwise BIC),

iv) Bayesian Model Averaging (BMA).

Avec les méthodes i) et iv), tous les paramètres sont systématiquement estimés. Avec les méthodes ii) et iii), seuls les paramètres des variables sélectionnées sont estimés. Le nombre de paramètres estimés est donc susceptible de varier d'un échantillon à l'autre avec les méthodes ii) et iii). Les paramètres sont estimés par les moindres carrés ordinaires pour les méthodes i) à iii) et avec une approche bayésienne pour la méthode iv) (Raftery *et al*., 1997).

Le tableau 4 présente les fréquences de sélection de chaque variable du modèle (nombre de fois sur 10000 où les variables sont sélectionnées et les paramètres correspondant estimés). Les méthodes de type stepwise conduisent à des fréquences de sélection inférieures à un. Ces fréquences dépendent à la fois du critère de sélection (AIC ou BIC). Par exemple, la variable 3 est sélectionnée eulement pour 5% des échantillons avec stepwise BIC. Le tableau 5 présente les moyennes des valeurs estimées des

paramètres sur les 10000 échantillons. Les moyennes calculées pour les méthodes de type stepwise correspondent aux moyennes des valeurs estimées lorsque les variables ont été sélectionnées.

Tableau 4. Fréquences de sélection des cinq variables explicatives du modèle avec six méthodes statistiques. Résultats obtenus avec 10000 échantillons bootstrap.

	Fréquences de sélection des variables pour les six méthodes			
Variable	Pas de sélection	Stepwise AIC 2	Stepwise BIC 2	BMA
1	1.00	0.46	0.29	1.00
2	1.00	0.23	0.10	1.00
3	1.00	0.14	0.05	1.00
4	1.00	0.51	0.32	1.00
5	1.00	0.79	0.61	1.00

Le tableau 5 montre que les moyennes des valeurs estimées après sélection stepwise sont plus grandes en valeurs absolues que les moyennes obtenues sans sélection. Ceci indique que les méthodes de type stepwise ont tendance à accentuer l'effet des facteurs limitants si on compare leurs résultats à ceux obtenus sans sélection. Cette accentuation est due à un biais de sélection (Miller, 2002) qui peut conduire les agronomes à surestimer l'effet des facteurs limitants. La différence entre les moyennes des estimations est très importante pour certains paramètres. Par exemple, la moyenne des valeurs estimées du paramètre θ_3 est neuf fois plus grande avec la méthode stepwise BIC que sans sélection (Tableau 5).

Les résultats obtenus avec BMA sont très différents. Cette technique estime tous les paramètres en considérant successivement tous les modèles possibles. Les fréquences de sélection sont donc égales à 1 ; aucun facteur limitant n'est exclu. Par ailleurs, le tableau 5 montrent que les moyennes des estimations sont plus faibles en valeur absolue avec BMA qu'avec la méthode sans sélection. Ainsi, contrairement aux méthodes stepwise, la méthode BMA a tendance à atténuer l'effet des facteurs limitants. Ceci est dû au fait que BMA considère tous les modèles linéaires possibles, y compris ceux où les facteurs limitants n'ont aucun effet (Raftery *et al*., 1997), et que les valeurs estimées des paramètres résultent d'une somme pondérée des estimations obtenues avec tous ces modèles (eq.14).

Les résultats de ce premier cas d'étude montrent que l'interprétation des résultats d'une démarche de type Yield Gap Analysis est très dépendante de la méthode statistique utilisée pour analyser les données. Les techniques de sélection stepwise conduisent à des fréquences de sélection très différentes selon le critère utilisé pour comparer les modèles et selon le point de départ de la procédure. Ces techniques de sélection peuvent par ailleurs conduire à une surestimation de l'effet des facteurs limitants.

Tableau 5. Moyennes des valeurs estimées des paramètres pour quatre méthodes statistiques. Résultats obtenus avec 10000 échantillons bootstrap.

	Moyennes des valeurs estimées des paramètres pour les quatre méthodes			
Paramètre	Pas de sélection	Stepwise AIC	Stepwise BIC	BMA
1	-3.36	-5.99	-7.34	-2.29
2	-1.12E-02	-2.53E-02	-3.33E-02	-4.87E-03
3	-1.73E-05	-1.08E-04	-1.60E-04	-8.65E-06
4	-5.36E-02	-8.42E-02	-9.76E-02	-3.30E-02
5	1.39E-01	1.60E-01	1.69E-01	1.03E-01

2. Généralisation de l'utilisation de modèles stochastiques

2.1. Contexte

L'approche stochastique est assez bien passée dans le grand public : des indices de confiance sont associés aux prévisions météorologiques depuis quelques années. L'utilisation de modèles stochastiques apparaît particulièrement justifiée en agronomie du fait des erreurs souvent importantes des modèles développés par les agronomes, de l'utilisation de ces modèles pour répondre à des questions pratiques et du fait de la complexité des phénomènes biologiques étudiés. Pourtant, en agronomie, les modèles restent, pour la plupart, déterministes.

L'utilisation de modèles stochastiques permettrait aux agronomes de présenter des prédictions sous la forme de distributions de probabilité plutôt que sous la forme de valeurs ponctuelles. Les distributions de probabilité générées par de tels modèles seraient utiles pour l'analyse des risques (Vose, 2000). La figure 10 présente une illustration de cette approche [15].

L'utilisation de modèles stochastiques facilite également l'application de technique de type assimilation de données comme le filtre de Kalman, le filtre de Kalman ensemble et le filtre particulaire. L'intérêt de ces techniques pour corriger les prédictions des modèles en cours de saison a déjà été illustré dans le cadre de mes travaux ([24], [30], [34]).

Il est donc intéressant de généraliser l'utilisation de ces modèles et d'étudier leurs intérêts pour d'autres applications que celles déjà testées, par exemple pour le raisonnement de l'irrigation ou pour analyser l'impact des politiques agricoles.

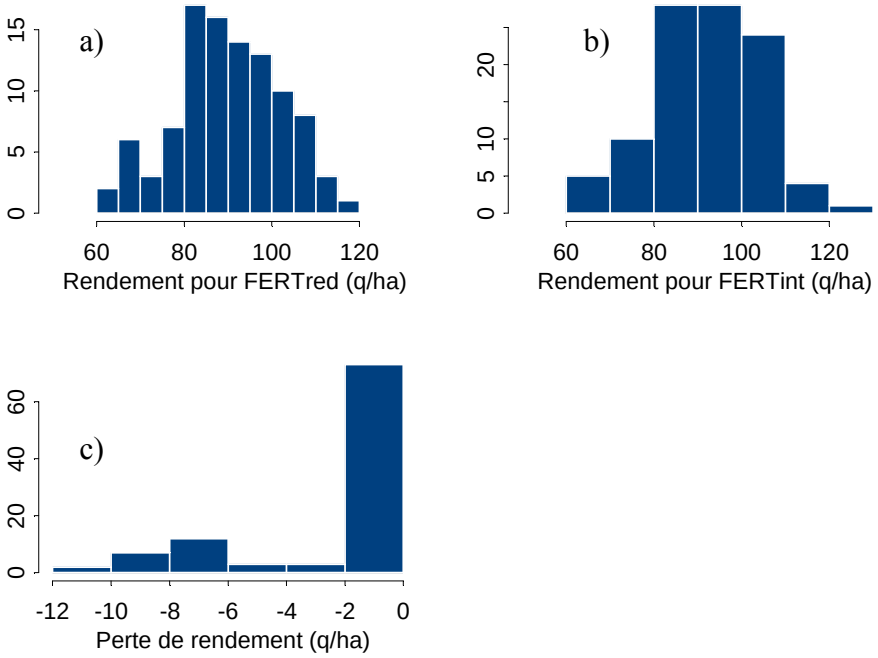

Figure 10. Distributions de valeurs de rendement simulées par un modèle stochastique pour une parcelle de blé en Picardie en fonction de la dose d'engrais azoté appliquée et des caractéristiques du sol [15].

Figure 10a présente la distribution de 100 valeurs de rendement générées avec le modèle pour une parcelle de blé ayant reçu une dose d'engrais réduite de 20% par rapport à la dose recommandée (FERTred). Figure 10b présente la distribution des valeurs de rendement générées avec le modèle pour la même parcelle mais avec une dose non réduite (FERTint). Figure 10c présente la distribution des pertes de rendement induite par la réduction de 20%.

2.2. Problèmes méthodologiques

Le développement et l'utilisation de modèles stochastiques posent plusieurs problèmes méthodologiques. Deux de ces problèmes sont brièvement discutés ci-dessous.

- Distributions des erreurs des modèles

J'ai déjà utilisé deux grands types d'approche pour décrire les erreurs des modèles. La première est basée sur une description de l'incertitude et de la variabilité dans les valeurs des paramètres à l'aide de *modèles hiérarchiques*, soit de type modèles mixtes ([4], [19]), soit de type bayésien ([8], [35]). La deuxième consiste à définir le modèle comme *un processus stochastique* ([24], [32]). Cette deuxième approche semble bien adaptée aux modèles de culture dynamiques comme STICS, AZODYN ou CERES.

Plusieurs aspects mériteraient d'être approfondis. Un de ces aspects concerne la définition des distributions a priori des paramètres lors de la mise en œuvre des méthodes bayésiennes. Ces méthodes nécessitent en effet la définition de distributions de probabilité décrivant l'état de connaissance initial dans les valeurs des paramètres, avant utilisation des données. Il serait très utile de développer une méthodologie pour définir de telles distributions en agronomie. Une approche prometteuse serait d'utiliser des méta-analyses sur des articles scientifiques.

Un autre aspect concerne la définition des distributions des termes d'erreurs que l'on pourrait ajouter aux modèles dynamiques pour les transformer en processus stochastique. Naud *et al* [24] ont défini la distribution des termes d'erreurs de la version stochastique d'Azodyn en appliquant des méthodes apparentées à l'analyse de sensibilité. Une alternative consisterait à estimer la distribution de ces termes aléatoires à partir des mesures disponibles en cours de saison. Les filtres non linéaires par

noyaux de convolution (Rossi, 2004) sont particulièrement séduisants. Ce sont des méthodes non paramétriques et les travaux de Rossi (2004) montrent qu'elles donnent des résultats très satisfaisants. Un intérêt de ces techniques est qu'elles permettent d'estimer conjointement les variances et covariances des termes d'erreur et les variables d'état des modèles dynamiques avec des données expérimentales.

- Evaluation des prédictions probabilistes

Les prédictions d'un modèle stochastique se présentent sous la forme d'une distribution de probabilité (figure 10). Des prédictions ponctuelles peuvent être déduites de ces distributions, en prenant par exemple la moyenne ou la médiane. Il est alors possible d'évaluer la précision de ces prédictions avec les techniques statistiques présentées dans la partie précédente de ce mémoire (e.g MSEP…). Il peut cependant être intéressant d'évaluer l'ensemble de la distribution de probabilité et pas seulement la prédiction ponctuelle qui en est déduite. Plusieurs approches ont été proposées par les statisticiens pour évaluer des prédictions probabilistes avec des données expérimentales (e.g Gneiting *et al*., 2007). Une des approches proposées consiste à analyser la distribution des valeurs de la fonction de répartition obtenues pour les différentes mesures disponibles. Il serait intéressant d'étudier la pertinence de ces approches dans un contexte agronomique.

Conclusion

La diminution des temps de calcul des ordinateurs et le développement de nouvelles méthodes ont fait évoluer les méthodes statistiques au cours des 20 dernières années (e.g Efron, 2005). Les innovations produites par les statisticiens ont été à l'origine de nombreux logiciels, souvent libres et téléchargeables depuis internet (**www.r-project.org**, **www.monolix.org, www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml**..).

Le travail des agronomes des instituts de recherche, des instituts techniques, et des administrations s'est également transformé au cours de la même période. La réalisation d'expertises régionales, nationales et européennes occupe dorénavant une place importante dans les activités des agronomes. Ces expertises visent le plus souvent à analyser les risques environnementaux (pollution de l'eau et de l'air) et les risques alimentaires (quantitatifs, qualitatifs, sanitaires) induits pas les activités agricoles. La mise en place en 2006 d'un groupe d'experts sur la santé des plantes financé par la Commission Européenne et les expertises « Pesticides » et « Sécheresse » réalisées en 2005-2006 par l'INRA et le CEMAGREF illustrent cette tendance.

Pour répondre à ces demandes, les agronomes sont souvent amenés à manipuler des bases de données complexes (e.g données du Service Central des Enquêtes et Études Statistiques (SCEES) du Ministère de l'Agriculture) et des modèles de natures variées dans leur travail quotidien (modèles statiques ou dynamiques). Ce phénomène est accentué par l'évolution des entreprises privées du secteur agricole qui cherchent dorénavant à associer leurs produits (pesticides, engrais azoté…) à des outils d'aide à la décision et à des modèles plus ou moins complexes, développés à partir de données expérimentales (e.g modèle POSITIF® pour anticiper les attaques de maladie). L'INRA a également fortement investi ces dernières années dans le développement de modèles simulant la croissance de diverses cultures.

Mes travaux me conduisent à formuler les conclusions et recommandations suivantes :

- Actuellement, beaucoup de projets de modélisation sont centrés sur un modèle unique. Mes résultats montrent qu'il est préférable d'évaluer et de comparer les performances de plusieurs modèles pour un usage donné, en utilisant des données expérimentales.
- Les modèles de culture dynamiques sont globalement peu performants par rapport aux modèles statiques ou aux mesures directes d'indicateurs simples. Il ne faut donc pas tout miser sur les modèles dynamiques.
- Les méthodes « analyse ROC » et « calcul de risque » sont adaptées à de nombreux problèmes pratiques et leur utilisation devrait être généralisée pour évaluer les performances des modèles et indicateurs agronomiques.
- La méthode utilisée pour estimer les paramètres d'un modèle a parfois une forte influence sur les performances de celui-ci. L'estimation des paramètres devrait donc recevoir une attention toute particulière dans les projets de modélisation.
- La correction des modèles avec des mesures obtenues en cours de saison est une approche qui donne des résultats prometteurs.

La double évolution de la statistique et de l'agronomie pourrait conduire à une multiplication des travaux à l'interface de ces deux disciplines au cours des prochaines années. Pourtant, si la demande est assez forte du côté agronomie, il n'est pas sûr que l'offre puisse être à la hauteur du côté statistique. En effet, les statisticiens sont déjà fortement sollicités par les scientifiques travaillant en génomique, dans les sciences médicales, l'écologie et l'économie. Il n'est pas sûr que les disciplines proches du secteur agricole réussissent à être assez attractives. Une option serait d'augmenter les compétences des agronomes en statistique en accentuant l'offre de formations dans ce domaine. C'est probablement une solution à privilégier.

Je tente d'identifier, ci-dessous, plusieurs domaines qui pourraient bénéficier d'une collaboration entre agronomes et statisticiens au cours des prochaines années.

Modèles stochastiques

Il est important de tenir compte des différentes sources d'incertitudes et de variabilité lorsqu'un modèle est utilisé pour l'aide à la décision (e.g Vose, 2000). La généralisation de l'utilisation des modèles stochastiques, proposée dans la partie *Perspectives* de ce mémoire, permettrait aux agronomes de le faire de manière plus systématique.

Les modèles hiérarchiques (modèles mixtes, modèles bayésiens) et les processus stochastiques, présentés dans partie *Activités de recherche* de ce mémoire, semblent être bien adaptés aux problèmes pratiques des agronomes. Les modèles hiérarchiques sont économes en paramètres et permettent de tenir compte des différentes sources de variabilités des données expérimentales (intra site-année, inter sites-années…) lors de l'estimation des paramètres. L'utilisation de processus stochastique dans les modèles dynamiques permet de corriger les prédictions de ces modèles à partir de mesures en cours de saison. Enfin, les techniques de combinaison de modèles, présentées dans la partie *Perspectives* du mémoire, ouvrent des possibilités intéressantes, notamment les approches bayésiennes de type BMA.

Grâce aux moyens informatiques modernes, il est maintenant possible d'utiliser des modèles stochastiques pour réaliser des applications agronomiques avec des temps de calcul raisonnables. L'effort de programmation informatique nécessaire à la mise en œuvre de ces techniques constitue cependant un frein. Une meilleure formation dans ce domaine ainsi que le développement de plateformes informatiques comme, par exemple, la plateforme RECORD (//record.toulouse.inra.fr/), pourrait résoudre ce problème.

Diagnostic agronomique

Il existe plusieurs types de diagnostic en agronomie. Le premier correspond au *yield gap analysis* dont l'objectif est d'identifier les principaux facteurs limitants ayant affecté une culture (e.g Casanova *et al*., 1999). Il s'agit en fait d'un diagnostic *a posteriori* réalisé avec des données expérimentales obtenues sur plusieurs dizaines de sites-années, généralement en situations agricoles. Cette approche a été perfectionnée en France, notamment par l'analyse séquentielle de plusieurs composantes du rendement (nombre de plantes, nombre de grains, poids d'un grain…) (Doré *et al*., 1997).

Un autre type de diagnostic concerne le diagnostic en cours de saison. L'objectif est de prévoir l'impact d'un facteur limitant assez tôt dans la saison pour que l'agriculteur puisse ajuster ses techniques culturales pendant le développement de la culture. Il s'agit ici d'un diagnostic *a priori*. Parmi les nombreux exemples existants, on peut noter la prévision d'une carence azotée ou encore le diagnostic précoce d'une maladie. Ce type de diagnostic nécessite l'utilisation d'indicateurs de risque issus soit de mesures directes (% fleurs malades, PCR, réflectance…), soit de prédictions réalisées avec des modèles.

Les techniques utilisés dans mes travaux sont susceptibles d'enrichir certaines des méthodes utilisées dans le cadre des diagnostics *a priori* et *a posteriori*. Les principales innovations sont résumées dans le tableau 6. Elles portent sur l'évaluation d'indicateurs de risque, sur la correction des modèles avec des mesures en cours de saison, l'estimation des paramètres des courbes enveloppes, et l'analyse de l'incertitude dans l'identification des facteurs limitants. Toutes ces méthodes ont été décrites dans ce

mémoire. Certaines pourraient, à terme, être utilisées en routine par les agronomes utilisant le diagnostic dans le cadre de leurs activités professionnelles.

Tableau 6. Méthodes statistiques proposées pour le diagnostic.

Méthode proposée	Objectif	Type de diagnostic concerné	Référence
Analyse ROC	Evaluer les probabilités d'erreur d'indicateurs de risque.	*a priori* (ex : diagnostic de maladie)	[14] [18] [26] [27] [44]
Filtrage	Corriger une prédiction en cours de saison.	*a priori* (ex : diagnostic de carence azotée)	[24] [25] [30] [35]
Estimation des paramètres des courbes enveloppes	Méthode reproductible pour ajuster les courbes enveloppes à des données expérimentales.	*a posteriori* (ex : identification de parcelles ayant subis des facteurs limitants)	[21]
Analyse de l'incertitude dans la sélection des modèles.	Déterminer la stabilité des résultats des procédures utilisées pour identifier les facteurs limitant.	*a posteriori* (ex : quantification de l'effet de certains facteurs limitants)	Projet PICSEL

Utilisation, sélection et combinaison des indicateurs agro-environnementaux

De nombreux indicateurs ont été proposés pour évaluer l'impact des pratiques agricoles sur l'environnement, notamment l'impact de l'utilisation des pesticides et des engrais (Girardin *et al*., 2000 ; Girardin *et al*., 2005). Il est fort probable que des indicateurs de biodiversité s'ajoutent à cette liste dans les toutes prochaines années.

Ces indicateurs sont de natures variées : mesures directes, variables calculées à partir des pratiques agricoles, simulations de modèles. Leur utilisation à grande échelle conduit souvent à générer des bases de données de tailles importantes. Les calculs réalisés par Champeaux (2006) pour analyser l'évolution de l'usage des pesticides en France illustrent ce phénomène.

Les moyens informatiques actuels permettent de calculer facilement différents types d'indicateurs sur un grand nombre de parcelles. Le recours aux indicateurs agro-environnementaux risque de s'accentuer dans un proche avenir. Plusieurs questions se posent dans un tel contexte :

- Comment analyser les masses de données générées par le calcul d'indicateurs à l'échelle nationale et européenne ? Comment analyser l'évolution spatiale et temporelle de ces indicateurs ?
- Quels indicateurs utiliser ? Comment les sélectionner ?
- Comment combiner les informations fournies par des indicateurs de natures différentes ?

Des travaux à l'interface statistique/agronomie seraient utiles pour répondre à ces questions. Certains des travaux présentés dans ce mémoire ouvrent des pistes comme l'analyse ROC pour comparer les niveaux de précision d'indicateurs, l'approche BMA pour combiner plusieurs modèles, ou encore la régression quantile pour analyser les queues de distributions.

Références

Baudrillard J., Bertaux C., Cazenave M., Dubois-Gance M., de Jouvenel H., Lachièze-Rey M., Le Bras H., Lévy-Leblond J-M., Madariaga R., Piettre B., Ramade F., de Ricqulès A., Ruffié J., Sadourny R., Walter C. 1996. Les sciences de la prévision. Editions du Seuil.

Boote K.J., Jones J.W., Pickering N.B. 1996. Potential uses and limitations of crop models. Agronomy Journal 88, 704-716.

Brancourt-Hulmel M., Lecomte C., Meynard J-M. 1999. A diagnosis of yield limiting factors on probe genotypes for characterizing environments in winter wheat trials. Crop Science 39, 1798-1808.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.-H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.-M., Meynard, J.-M., and Delécolle, R., 1998. STICS: a generic model for the simulation crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 8, 311-346.

Buckland S.T., Burnham K.P., Augustin N.H. 1997. Model selection: an integral part of inference. Biometrics 53, 603-618.

Bunke O., Droge B. 1987. Estimators of the mean squared error of prediction in linear regression. Technometrics 26, 145-155.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference. Second Edition. Springer.

Cade B.S., Terrel J.W., Schroeder R.L. 1999. Estimating the effects of limiting factors with regression quantiles, Ecology 80, 311-323.

Casanova D., Goudriaan J., Bouma J., Epema G.F. 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice, Geoderma 91, 191-216.

Carlin B.P., Louis T.A. 2000. Bayes and empirical Bayes methods for data analysis. Chapman & Hall/CRC, New-York.

Champeaux C. 2006. Recours à l'utilisation de pesticides en grandes cultures. INRA, Ministère de l'agriculture et de la pêche, Paris, Grignon.

Chatfield C. 1995. Model uncertainty, data mining, and statistical inference. *J.R. Statist. Soc. A.* 158, 419-466.

Colson J., Wallach D., Bouniols D., Denis J-B., Jones J.W. 1995. Mean squared error of yield prediction by SOYGRO. Agronomy Journal 87, 397-402.

David, C., Jeuffroy, M.-H., Laurent, F., Mangin, M., Meynard, J.-M., 2005. The assessment of Azodyn-Org model for managing nitrogen fertilization of organic winter wheat. European Journal of Agronomy 23, 225-242.

Davidian M, Giltinan D.M. 2003. Nonlinear models for repeated measurement data: an overview and update. JABES 8, 387-419.

Del Moral P., Guionnet, A., 1999. Central limit theorem for nonlinear filtering and interacting particle systems. The Annals of Applied Probability 9, 275-297.

Doré T., Meynard J.-M., Sebillotte M. 1998. The role of grain number, nitrogen nutrition and stem number in limiting pea crop (Pisum sativum) yields under agricultural conditions. European Journal of Agronomy 8, 29-37.

Doré T., Sebillotte M., Meynard J.-M. 1997. A diagnostic method for assessing regional variation for crop yield. Agricultural Systems 54, 169-188.

Doucet, A., de Freistas, N., Gordon, N. (Eds.), 2001. Sequential Monte Carlo Methods in Practice. Statistics for Engineering and Information Science. Springer, New York.

Draper D. 1995. Assessment and propagation of model uncertainty. *J.R. Statist. Soc. B.* 57, 45-97.

Efron B. 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association 78, 316-331.

Efron B. 2005. Bayesians, Frequentists, and Scientists. Journal of the American Statistical Association 1000, 1-5.

Ennaïfar S. 2006. Protection intégrée contre le piétin échaudage du blé d'hiver. Thèse de doctorat. AgroCampus Rennes, France.

Fleury A. 1991. Méthodologies de l'analyse de l'élaboration du rendement, in: Picard, D., (Ed), Physiologie et production du maïs, INRA, Paris, pp. 279-290.

Girardin Ph., Bockstaller C., Van der Werf H. 2000. Evaluation of relationship between the environment and agricultural practices – the AGRO-ECO method. Environmental impact assessment review 20, 227-239.

Girardin Ph., Guichard L., Bockstaller C. 2005. Indicateurs et tableaux de bord. Guide pratique pour l'évaluation environnementale. Tec&Doc, Lavoisier, Paris.

Gneiting T., Balabdaoui F., Raftery A.E. 2007. Probabilistic forecasts, calibration and sharpness. J.R. Statist. Soc. B. 69, 243-268.

Gonzales Montaner J.H., Maddoni G.A., DiNapoli M.R. 1997. Modeling grain yield and grain yield response to nitrogen in spring wheat crops in the Argentinean Southern Pampa. Field Crops research 51, 241-252.

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: a tutorial. Statistical Science 14, 382-417.

Hughes G., McRoberts N., Burnett F.J. 1999. Decision-making and diagnosis in disease management. Plant Pathology 48, 147-153.

Jeuffroy M.-H., Recous S., 1999. Azodyn: a simple model simulating the date of nitrogen deficiency for decision support in wheat fertilization. European Journal of Agronomy 10, 129-144.

Kalman R.E. 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME – Journal of Basic Engineering, 35-45.

Koenker R., Basset G. 1978. Regression quantiles. Econometrica 46, 33-50.

Koenker R., Park B.J. 1996. An interior point algorithm for nonlinear quantile regression. Journal of Econometrics 71, 265-283.

Kuhn E., Lavielle M. 2004. Coupling a stochastic approximation version of EM with a MCMC procedure. ESAIM Probability and Statistics 8, 114-131.

Le Bail M., Jeuffroy M. H., Bouchard C., Barbottin A. 2005. Is it possible to forecast the grain quality and yield of different varieties of winter wheat from Minolta SPAD meter measurements? European Journal of Agronomy 23, 379-391.

Lindberg B.W. 1971. Elements of decision theory. Macmillan, New York

Lindsay B.G. 2004. Statistical distances as loss functions in assessing model adequacy. In: The nature of scientific evidence. Taper M.L. & Subhash R. L. (eds.). The University of Chicago Press, Chicago.

Miller A. 2002. Subset selection in regression. Second Edition. Chapman & Hall/CRC, London.

Mitscherlich E.A. 1909. Das gesetz des minimums und das gesets des abnehmenden bodenertrages. Landwirtsch. Jahrb. 38, 537-552.

Mombiela F., Nicholaides J.J. III., Nelson L.A. 1981. A method to determine the appropriate mathematical form for incorporating soil test levels in fertilizer response models for recommendation purposes. Agronomy Journal 73, 937-941.

Naud C. 2007. Améliorer les prédictions de l'indice de nutrition azotée en combinant modèle dynamique et mesures expérimentales. Thèse de doctorat. AgroParisTech, Paris, France.

Passioura J.R. 1996. Simulation models: Science, Snake oil, Education or Engineering. Agronomy Journal 88, 690-694.

Pilkey O.H., Pilkey-Jarvis L. 2007. Useless arithmetic. Why Environmental Scientists can't predict the future. Columbia University Press, New-York.

Pinheiro J.C., Bates D.M. 2000. Mixed-effects models in S and S-PLUS. Springer-Verlag, New-York.

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review 133, 1155-1174.

Raftery AE, Madigan D, Hoeting JA. 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92, 179-191.

Rémy JC, Hébert J. 1977. Le devenir des engrais azoté dans le sol. C.R. Acad. Agric. Fr. 63, 700-710.

Rossi V. 2004. Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique. Thèse ENSAM, Montpellier, France.

Rossing W.A.H., Meynard J-M., van Ittersum M. K. 1997. Model-based explorations to support development of sustainable farming systems: case studies from France and the Netherlands. European Journal of Agronomy 7, 271-283.

Sain G.E., Jauregui M.A. 1993. Deriving fertilizer recommandations with a flexible functional form. Agronomy Journal 85, 934-937.

Saltelli A., Tarantola S., Campolongo F. 2000. Sensitivity analysis as an ingredient of modeling. Statistical Science 45, 377-395.

Stanford G. 1973. Rationale for optimum N fertilization in corn production. J. Environn. Qual. 2, 159-166.

Swets J.A. 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285-1293.

Viallefont V., Raftery A.E., Richardson S. 2001. Variable selection and Bayesian model averaging in case-control studies. Statistics in medicine 20, 3215-3230.

Vose D. 2000. Risk analysis, a quantitative guide. Wiley, Chichester.

Wallach D. 2006. Evaluating crop models. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 11-53.

Wallach D., Goffinet B. 1987. Mean squared error of prediction in models for studying ecological and agronomic systems. Biometrics 43, 561-573.

Wallach D., Goffinet B., Bergez J-E., Debaeke Ph., Leenhardt D., Aubertot J-N. 2001. Parameter estimation for crop models: a new approach and application to a corn model. Agronomy Journal 93, 757-766.

Webb R.A. 1972. Use of boundary line in the analysis of biological data. J. hort. Sci. 47, 309-319.

Yang Y. 2003. Regression with multiple candidate models: selecting or mixing ? Statistica Sinica 13, 783-809.

Yuan Z, Yang Y. 2005. Combining linear regression models : when and how ? Journal of the American Statistical Association 100, 1202-1214.