

COMMUNAUTE FRANCAISE DE BELGIQUE
ACADEMIE UNIVERSITAIRE WALLONIE-EUROPE
UNIVERSITE DE LIEGE – GEMBLoux AGRO-BIO TECH

**INCERTITUDES LIÉES À LA MODÉLISATION
AGRO-ENVIRONNEMENTALE EN VUE DE
DÉVELOPPER DES OUTILS D'AIDE À LA DÉCISION**

Benjamin DUMONT

Dissertation originale présentée en vue de l'obtention du grade
de docteur en sciences agronomiques et ingénierie biologique.

Promoteur :
Pr. Marie-France Destain

Année Académique
2013-2014

COMMUNAUTE FRANCAISE DE BELGIQUE
ACADEMIE UNIVERSITAIRE WALLONIE-EUROPE
UNIVERSITE DE LIEGE – GEMBLoux AGRO-BIO TECH

**INCERTITUDES LIÉES À LA MODÉLISATION
AGRO-ENVIRONNEMENTALE EN VUE DE
DÉVELOPPER DES OUTILS D'AIDE À LA DÉCISION**

Benjamin DUMONT

Dissertation originale présentée en vue de l'obtention du grade
de docteur en sciences agronomiques et ingénierie biologique.

Promoteur :
Pr. Marie-France Destain

Année Académique
2013-2014

Dumont Benjamin (2014). Incertitudes liées à la modélisation agro-environnementale en vue de développer des outils d'aide à la décision.¹

Thèse de doctorat défendue en Français.

Université de Liège - Gembloux Agro-Bio Tech

Gembloux, Belgique

193 p., 6 tabl., 20 fig.

Résumé :

La directive Européenne 91/676/EEC vise à protéger la qualité de l'eau en prévenant la pollution des eaux souterraines et superficielles par les nitrates provenant de sources agricoles et en promouvant l'usage de bonnes pratiques. Si la mise en œuvre de cette directive s'avère efficace, il apparaît cependant que l'utilisation d'azote a augmenté de 6 % au cours des quatre dernières années dans les 27 Etats membres et que l'agriculture est toujours à l'origine de plus de 50 % de la quantité totale d'azote déversée dans les eaux superficielles (http://ec.europa.eu/environment/water/watzer-nitrates/index_en.html). En Wallonie, la directive «Nitrates» est transposée sous la forme du Programme de Gestion Durable de l'Azote (PGDA). Entré en vigueur en 2002, celui-ci comporte un ensemble de mesures comme, par exemple, la fixation de règles au niveau de l'épandage des fertilisants, la mise en place de modes de gestion spécifiques des cultures en zones vulnérables, la mesure des teneurs en azote potentiellement lessivable (APL) dans les sols, etc.

La thèse de doctorat s'inscrit dans ce contexte général qui vise à optimiser la quantité de fertilisants azotés à apporter à une culture de blé d'hiver (*Triticum aestivum* L.). Elle s'appuie sur un modèle de culture ou modèle écophysiological, qui décrit la croissance et le développement d'une culture en interaction avec le sol et l'atmosphère. La difficulté majeure dans l'utilisation d'un tel modèle en tant qu'outil d'aide à la décision est liée aux nombreuses incertitudes qui interviennent à différents niveaux. En effet, les modèles de culture dynamiques sont constitués d'un grand nombre d'équations différentielles non-linéaires, comportant de nombreux paramètres à estimer pour que les sorties se rapprochent le plus possible de séquences observées. Une première source d'incertitude existe donc au niveau des paramètres du modèle. Une fois le modèle robustement calibré, il peut être utilisé à des fins prédictives. Toutefois, dans un contexte agronomique, le délai entre la date de semis et celle de récolte est important. Comme il est souhaitable de prédire le rendement final d'un point de vue quantitatif et qualitatif, différentes hypothèses sur les scénarios climatiques qui interviennent pendant la saison culturale doivent être posées. Une deuxième source d'incertitude est donc liée au climat. Enfin, le passage d'un modèle à un outil fonctionnel de gestion des fertilisants azotés nécessite de franchir une étape supplémentaire car l'impact réel des fertilisants n'est connu qu'en fin de saison. Une troisième source d'incertitude est donc liée à la quantification du niveau d'erreur toléré sur la gestion des fertilisants, compte tenu des interactions existant entre la pratique envisagée et la réalisation climatique. Cette thèse de doctorat a pour objectif d'étudier les différentes sources

¹ Copyright : Aux termes de la loi belge du 30 juin 1994, sur le droit d'auteur et les droits voisins, seul l'auteur a le droit de reproduire partiellement ou complètement cet ouvrage de quelque façon et forme que ce soit ou d'en autoriser la reproduction partielle ou complète de quelque manière et sous quelque forme que ce soit. Toute photocopie ou reproduction sous autre forme est donc faite en violation de la dite loi et de des modifications ultérieures.

d'incertitudes afin d'aboutir à un outil d'aide à la décision efficace. Elle comporte cinq parties.

Dans la première partie, une approche Bayésienne d'identification des paramètres du modèle de culture STICS est présentée. Elle repose sur l'algorithme DREAM (DiffeRential Evolution Adaptative Metropolis). Les distributions *a posteriori* de plusieurs paramètres du modèle STICS sont échantillonnées en vue d'améliorer les simulations de la croissance et du développement du blé d'hiver. Par rapport aux algorithmes de type Simplex habituellement utilisés dans les modèles de culture, le couplage de STICS avec DREAM fournit une approximation de la distribution *a posteriori* des paramètres, une évaluation des corrélations qui existent entre eux et surtout quantifie l'incertitude dans l'estimation des rendements. Par ailleurs, une fonction de vraisemblance faisant explicitement mention du coefficient de variation est proposée. Les résultats montrent que cette dernière permet de prendre en compte le bruit sur les mesures et l'hétéroscédasticité régulièrement rencontré dans les modèles de culture.

En parallèle ou en complément aux approches d'identification paramétrique, des techniques de filtrage permettent d'améliorer les simulations des modèles de culture. Au moyen d'algorithmes couplés au modèle de culture, ces techniques visent à intégrer dans le modèle des mesures réalisées en cours de saison, comme par exemple l'indice de développement foliaire LAI ou la teneur en eau du sol, de manière à corriger la trajectoire modélisée de l'évolution des variables d'état. La capacité d'un algorithme récemment développé et connu sous le nom de filtre variationnel est évaluée et sa supériorité en termes d'estimation paramétrique et de réduction des incertitudes sur les variables de sortie est démontrée par rapport à celle de méthodes plus conventionnelles, basées sur le filtre de Kalman.

La troisième partie de cette étude se focalise sur la prédiction du rendement à l'aide du modèle STICS en s'attachant aux incertitudes liées aux entrées climatiques. La méthodologie utilisée consiste à construire des '*ensembles de matrices*' climatiques évoluant au cours de la saison, constitués d'une part de données climatiques réelles mesurées au fur et à mesure de la croissance de la culture, et d'autre part de données climatiques moyennes calculées au départ d'une base de données locale historique. Au fur et à mesure de l'avancement de la saison, la part des premières données devient prépondérante par rapport à celle des secondes. Il est montré que cette méthodologie, que nous appellerons 'climat moyen' permet de prédire les rendements environ un mois avant la date de récolte avec un intervalle de confiance de 10%. En outre, cette approche permet de quantifier les potentialités restantes de rendement d'une culture donnée dans des conditions agro-environnementales données et de détecter l'occurrence de stress lors de la croissance des cultures.

Dans la quatrième partie, une méthode reposant sur la génération d'un grand nombre de réalisations climatiques stochastiques (LARS-Weather Generator) est mise en œuvre. En se basant sur le théorème de la convergence en loi, il est démontré que les approches stochastiques et 'climat moyen' présentent des critères statistiques comme la RRMSE et la déviation normalisée inférieurs à 10% et sont dans 90% des situations climatiques équivalentes en termes de simulation du rendement. Elles offrent par ailleurs le même délai prédictif. La distinction entre les méthodes se situe surtout au niveau de la finalité visée, l'une permettant de simuler rapidement le potentiel restant à la culture, l'autre assurant la quantification du niveau d'incertitude sur la prédiction.

Dans la cinquième et dernière partie, la réponse du modèle STICS est étudiée sous des

réalisations climatiques stochastiques en vue de quantifier le niveau d'incertitude associé à différentes modalités d'apports de fertilisants azotés. Il est montré que, sous nos latitudes, en l'absence de fertilisant azoté, les distributions de rendements sont normales. Avec l'accroissement des niveaux de fertilisation, un degré d'asymétrie croissant est observé. Dès l'instant où un fertilisant azoté est appliqué, non seulement les rendements augmentent, mais la fréquence des rendements au moins supérieurs à la moyenne croît également. Ceci réduit inéluctablement le risque pour l'agriculteur d'obtenir des rendements bas. Pour synthétiser les études menées au long de cette thèse, un outil de gestion stratégique de l'azote, c'est-à-dire sans connaissance a priori des conditions climatiques à venir, est mis au point. D'une manière générale, pour ce qui concerne nos régions limoneuses, la supériorité des fertilisations en trois fractions est mise en évidence. Un système de gestion de l'azote, basé sur un apport de deux modalités de 60 kilos d'azote par hectare appliqués aux stades tallage et redressement et offrant la possibilité de moduler la fraction de dernière feuille, semble optimal. L'incertitude associée aux prédictions et issue de la variabilité climatique peut être caractérisée finement et il est possible de quantifier le risque encouru par l'agriculteur qui envisage différentes pratiques afin de procéder à une optimisation de celles-ci.

D'une manière générale, les recherches pourraient être poursuivies en étudiant de façon plus fondamentale et plus systématique différentes situations agro-environnementales. En particulier, il serait intéressant de développer l'étude de l'interaction $\text{Genotype} \times \text{Environnement} \times \text{Pratiques}$ culturelles pour garantir la sécurité alimentaire dans un contexte de changement climatique.

Dumont Benjamin (2014). Uncertainty linked to crop modelling in order to develop decision support tools.

PhD thesis defended in French.

University of Liege - Gembloux Agro-Bio Tech

Gembloux, Belgium

193 p., 6 tabl., 20 fig.

Summary:

The European Nitrates Directive 91/676/EEC aims to ensure water quality by preventing pollution of surface and groundwater induced by nitrates originating from agricultural sources and by promoting agronomical good practices. While the implementation of this Directive seems effective, it appears however that the use of nitrogen has still increased by 6% over the last four years in 27 European countries. Furthermore, agricultural sources would be still at the origin of 50% of the total amount of nitrogen discharged into surface waters (http://ec.europa.eu/environment/water/water-nitrates/index_en.html). In Wallonia (Belgium), the Nitrates Directive has been transposed under the Sustainable Nitrogen Management in Agriculture Program (PGDA). Launched in 2002, it involves different sets of actions, like rules definitions concerning fertilizers application, specific and appropriate crop management in vulnerable areas, the control of potentially leachable nitrogen (APL) levels in soils, etc.

This is the global context in which lies the present thesis. The main aim is to optimise the nitrogen fertiliser practices to ensure that the needs of a winter wheat culture (*Triticum aestivum* L.) could be met while reducing the environmental pressure. It relies on the use of crop models, which describe the growth and the development of a culture interacting with its environment, namely the soil and the atmosphere. The major difficulty while working with crop models and model-based decision support tools lies in the fact that different sources of uncertainties have an impact on the modelled phenomena. Indeed, crop models are constituted by a consequent number of differential non-linear equations, involving a lot of parameters which need to be determined as accurately as possible in order to match as close as possible observed sequences of measurements. The first source of uncertainty is thus constituted by the parameters definition. Once the model has been correctly and robustly calibrated it can be used to perform predictions. However, in an agronomical context, the time-delay between sowing and harvest is consequent. As the end-season yield is often the expected output, the uncertainty linked to the non-knowledge of the future implies for the modeller to refer to different hypothesis concerning upcoming climatic scenarios. Finally, moving from models to decision systems dealing with N management involves a last source of uncertainty. Indeed the main problem is that the impact of a given practice is delayed in time from its realisation. In addition to the uncertainty linked to climatic projections themselves, it is highly important to consider the interactions between the practices and the climate. Furthermore, in a decision-making process, it could be highly relevant to know the uncertainty's estimation that could be tolerated on the decision.. Therefore, the present thesis aims to study these different sources of uncertainty in order to design an efficient decision support system. It is divided into five parts.

In the first part, a Bayesian sampling algorithm, known as DREAM (DiffeRential Evolution Adaptative Metropolis) will be presented. It was successfully coupled with the STICS soil-crop model

used in this study. The *a posteriori* probability density function of many parameters was sampled in order to improve the simulations of the growth of a winter wheat culture (*Triticum aestivum* L.). The DREAM algorithm offers different advantages in comparison to usual methods. Among these, it is possible to study *i*) the most probable *a posteriori* parameters distributions, *ii*) the parameters correlations, and *iii*) the uncertainties impacted on model outputs. Furthermore, a new version of the likelihood function was proposed, making an explicit use of the coefficient of variation. Results showed that it allowed the noise existing on measurements to be considered, but also the heteroscedasticity phenomenon usually encountered in biological growth processes.

In parallel, assimilation data is another way to improve models simulations. These techniques allow considering measurements performed in real-time (*e.g.* remote measures of LAI or soil water content) in order to correct and adjust the possible drift of model simulations. In particular, a recently developed algorithm, known as variational filter, was evaluated. Its superiority, both in term of state variables simulations improvement and parameter resampling, was demonstrated.

The third part of the research focuses on the real-time end-season yield prediction. It involves building climate *matrix ensembles*, combining different time ranges of projected mean data and real measured weather originating from the historical records. As the crop growing season progresses, the effects of real monitored data plays a greater role and the prediction reliability increases. Our results demonstrated that a reliable predictive delay of 3-4 weeks before harvest could be obtained. Finally, using real-time data acquired with a micrometeorological station enabled to *(i)* predict, daily, potential yield at the local level, *(ii)* detect stress occurrence, and *(iii)* quantify yield losses (or gains).

Being based on projected seasonal norms, this methodology is in opposition to another technique that consists to offer a panel of solution for what concerns the future. Such probabilistic technique relies on the use of stochastic weather generator (LARS-WG in this case). However, in the fourth part of this thesis, on the basis of the convergence in law theorem, it was demonstrated that in 90% of the climatic situations, both approaches were equivalent, exhibiting RRMSE and normalised deviation criteria inferior to 10%. Furthermore the two approaches offered similar predictive delay-time. The main difference between techniques lies in the finality. The first allows to quickly simulate the remaining yield potential, while the second aims to quantify the uncertainty level associated to the predictions.

In the fifth and last part of this thesis, in order to quantify the uncertainty level associated to different modalities of N applications, the STICS model answers were studied under stochastic climatic realisations. It was demonstrated that, if no N was applied, under our temperate climatic conditions, the yield distribution could be considered as normal. However, with increasing N practices, the asymmetry level was found itself increasing. As soon as N was applied, not only were the yields higher, but also was the probability to achieve yields that were at least superior to the mean of the distribution. This undoubtedly reduced the risk for the farmer to achieve low yields levels. To summary all the researches conducted in this thesis, a N strategic decision support system was developed. In a general way, for what concerns the Hesbaye Region, the superiority of three fractions N protocols was demonstrated. In addition, the three rates fertilisation management based on the systematic applications of 60 kgN.ha⁻¹ at tillering and stem extension stages and offering the possibility to adapt the flag-leaf fraction in real-time appeared as an optimal strategy. Within this tool,

the uncertainty associated to climatic variability could be finely characterised, and the risk encountered by the farmer was quantified for different investigated practices. But far more important, it was demonstrated that N management could be optimised in real-time.

In a general way, the research should be pursued by studying more fundamentally and systematically a wide range of different agro-environmental situations. In particular, it would be interesting to study of the Genotype \times Environment \times Cultural practices interactions to ensure food security in a climatic changing world.

Remerciements

J'adresse mes remerciements au SPW (DGO-3) pour le financement du projet intitulé '*Suivi en temps réel de l'environnement d'une parcelle agricole par un réseau de microcapteurs en vue d'optimiser l'apport en engrais azotés*', en parallèle duquel s'inscrit cette thèse de doctorat.

J'aimerais également remercier toutes les personnes ayant contribué à l'aboutissement de la recherche en tant que telle et des publications qui y sont associées, principalement mon promoteur de thèse, la Prof. Marie-France Destain, pour son soutien au long de mes recherches, ainsi que les autres promoteurs du projet de recherche "Microcapteurs", les Prof. Bernard Bodson et Prof. Jean-Pierre Destain pour les discussions fructueuses sur l'analyse des données ; et finalement les collaborateurs scientifiques, MM. Benoit Seutin, Rodrigo Meza, Vincent Leemans, Salvador Ferrandis, Rudy Schartz. Merci aussi à tous ceux qui, sur de plus petits détails, m'ont aidé et encouragé à mener ce travail à son terme.

Merci à mes collègues et amis de la fac., pour les sessions de mess, les parties de molkky, de ping-pong, les discussions et les pauses café. J'aimerais aussi remercier tous les modélisateurs de renom qui, certes, ne me connaissent pas tous, mais qui ont su m'inspirer, et en particulier l'Équipe du Projet STICS dans son intégralité pour m'avoir invité à prendre part à leur belle aventure, ainsi que le Pr. Bruno Basso, qui a toujours répondu présent que ce soit pour une discussion via e-mail ou pour m'accueillir dans son laboratoire.

Je souhaite remercier tous mes proches, ceux auprès de qui j'ai pu me plaindre lorsque mon travail n'avancait pas, qui m'ont écouté ou ont suffisamment fait semblant pour que j'y croie, et ceux avec qui j'ai pu discuter, peut-être parfois davantage pour m'auto-convaincre des avancées et de l'intérêt de mon travail.

Merci aussi, surtout et avant tout à Lorraine, qui me porte, me supporte et m'emporte depuis plus de 5 ans. Merci à ma famille dans son intégralité qui a contribué à faire de moi qui je suis aujourd'hui, et en particulier à ma sœur Elisabeth et mes parents, Béatrice et Jean-François. Merci à eux tout particulièrement pour m'avoir montré le bon chemin et m'avoir apporté leur soutien durant les petits et grands projets de vie.

Et finalement, merci encore à tous mes amis (que je ne citerai pas au risque d'en oublier) pour les discussions, les blagues, les films, les soirées, le sport, le ski, la pêche, la rénovation,, et tous les instants partagés.

Comme me le faisait remarquer un homme sage, Nelson Mandela n'avait-il pas mentionné qu' "*Aucun de nous, en agissant seul, ne peut atteindre le succès*".

Table des matières

CHAPITRE 1 : INTRODUCTION GÉNÉRALE	- 2 -
1. ENJEUX AGRO-ENVIRONNEMENTAUX ET AGRICULTURE DE PRÉCISION	- 2 -
2. LES MODÈLES DE CULTURE	- 4 -
2.1. Développement des modèles	- 4 -
2.2. Emploi des modèles	- 4 -
2.3. Contraintes associées à l'emploi des modèles agro-environnementaux	- 5 -
2.4. Principales limitations associées à l'emploi des modèles	- 6 -
3. DÉFINITION DU CADRE ET DES OBJECTIFS DE LA THÈSE	- 7 -
4. ORGANISATION DE LA THÈSE	- 8 -
CHAPITRE 2 : PROTOCOLE EXPÉRIMENTAL	- 12 -
1. PROTOCOLE EXPÉRIMENTAL	- 12 -
1.1. Les essais phytotechniques	- 12 -
1.2. Caractérisation géopédologique	- 14 -
1.3. Mesures agronomiques	- 15 -
1.4. Mesures environnementales	- 15 -
1.5. Le réseau de capteurs sans fil	- 15 -
1.5.1. Choix des micro-capteurs	- 15 -
1.5.2. Relais de l'information à la station de base	- 17 -
2. LE MODÈLE DE CULTURE STICS	- 19 -
2.1. Description générale	- 19 -
2.2. Méthodologie de paramétrisation du modèle STICS	- 19 -
2.3. <i>OptimiSTICS</i> , vers l'emploi de librairie	- 22 -
2.4. Evaluation du modèle	- 22 -
CHAPITRE 3 : ETAT DE L'ART ET SYNTHÈSE BIBLIOGRAPHIQUE	- 26 -
1. OBJECTIF	- 26 -
2. PUBLICATION ASSOCIÉE	- 26 -
CHAPITRE 4 : ESTIMATION PARAMÉTRIQUE	- 39 -
1. OBJECTIF	- 39 -
2. CONTRIBUTION ORIGINALE À L'ESTIMATION DES PARAMÈTRES DE STICS	- 39 -
2.1. La sélection des paramètres à estimer	- 39 -
2.2. Les méthodes d'estimation et d'échantillonnage des paramètres	- 40 -
2.3. La qualité des mesures en complément aux méthodes d'échantillonnages	- 41 -
3. DISCUSSIONS INTERMÉDIAIRES	- 43 -
3.1. Echantillonnage paramétrique	- 43 -
3.2. Analyse des distributions <i>a posteriori</i>	- 43 -
3.3. Incertitude sur les sorties du modèle	- 45 -
3.4. Qualité globale des simulations	- 46 -
4. CONCLUSIONS INTERMÉDIAIRES	- 47 -
5. PUBLICATION ASSOCIÉE	- 48 -
CHAPITRE 5 : ASSIMILATION DE DONNÉES	- 65 -
1. OBJECTIF	- 65 -
2. CONTRIBUTION ORIGINALE À L'ASSIMILATION DE DONNÉES	- 65 -
2.1. Le filtre variationnel	- 65 -
2.2. Enonciation du problème	- 66 -
3. DISCUSSIONS INTERMÉDIAIRES	- 67 -
3.1. Enonciation du problème et opérationnalité du modèle	- 67 -
3.2. Avantages offerts par le filtre variationnel	- 68 -
3.3. Limites liées à l'approche proposée	- 69 -
4. CONCLUSIONS INTERMÉDIAIRES	- 70 -
5. PUBLICATIONS ASSOCIÉES	- 70 -

CHAPITRE 6 : PRÉDICTION DES RENDEMENTS : MÉTHODOLOGIE POUR LA MISE AU POINT D'UN OUTIL FONCTIONNEL.	- 88 -
1. OBJECTIF	- 88 -
2. CONTRIBUTION ORIGINALE À LA PRÉDICTION DES RENDEMENTS EN TEMPS-RÉEL	- 89 -
2.1. <i>Base de données historique</i>	- 89 -
2.2. <i>Calcul du climat moyen</i>	- 89 -
2.3. <i>Génération d'une matrice ensemble représentative du climat étudié</i>	- 90 -
3. DISCUSSIONS INTERMÉDIAIRES	- 90 -
3.1. <i>Analyse des données de simulations d'une matrice ensemble</i>	- 90 -
3.2. <i>Analyse sur des années multiples et prédiction du rendement</i>	- 93 -
4. CONCLUSIONS INTERMÉDIAIRES	- 94 -
5. PUBLICATION ASSOCIÉE	- 95 -
CHAPITRE 7 : PRÉDICTION DES RENDEMENTS : ETUDE DU COMPORTEMENT DU MODÈLE ET COMPARAISON DE DEUX APPROCHES VALIDÉES.	- 115 -
1. OBJECTIF	- 115 -
2. CONTRIBUTION ORIGINALE À L'ÉTUDE DU COMPORTEMENT DU MODÈLE	- 116 -
2.1. <i>Le générateur stochastique d'aléa climatique</i>	- 116 -
2.2. <i>Le théorème de la convergence en loi</i>	- 116 -
2.3. <i>Application de la démarche au cas d'étude</i>	- 116 -
3. DISCUSSIONS INTERMÉDIAIRES	- 116 -
3.1. <i>Analyse des distributions de probabilités expérimentales</i>	- 117 -
3.2. <i>Comparaison des deux approches</i>	- 118 -
4. CONCLUSIONS INTERMÉDIAIRES	- 119 -
5. PUBLICATION ASSOCIÉE	- 120 -
CHAPITRE 8 : COMPORTEMENT DES MODÈLES DYNAMIQUES EN RÉPONSE À LA FERTILISATION AZOTÉE ET LA VARIABILITÉ CLIMATIQUE	- 143 -
1. OBJECTIF	- 143 -
2. CONTRIBUTION ORIGINALE À LA GESTION STRATÉGIQUE DE LA FERTILISATION AZOTÉE	- 144 -
2.1. <i>Le système de Pearson appliqué à l'étude des distributions</i>	- 144 -
2.2. <i>Les coefficients de Pearson</i>	- 144 -
3. DISCUSSIONS INTERMÉDIAIRES	- 145 -
3.1. <i>Réponse de la culture à différents niveaux de fertilisation azotée et sous des conditions climatiques variées.</i>	- 145 -
3.2. <i>Gestion stratégique de l'azote : à la recherche d'une fertilisation optimale</i>	- 146 -
4. CONCLUSIONS INTERMÉDIAIRES	- 149 -
5. PUBLICATIONS ASSOCIÉES	- 150 -
CHAPITRE 9 : CONCLUSIONS GÉNÉRALES ET PERSPECTIVES	- 160 -
1. CONCLUSIONS GÉNÉRALES	- 160 -
2. ENJEUX AGRONOMIQUES ET ENVIRONNEMENTAUX	- 164 -
3. PERSPECTIVES	- 165 -
CHAPITRE 10 : REFERENCES BIBLIOGRAPHIQUES GÉNÉRALES	- 169 -

Table des figures

FIGURE 1 : ORGANIGRAMME DU PLAN DE LA THÈSE.	- 9 -
FIGURE 2 : IDENTIFICATION DU CHAMP EXPÉRIMENTAL PAR RAPPORT AU CENTRE WALLON DE RECHERCHES AGRONOMIQUES CRA-W, SITUÉ RUE DU BORDIA, GEMBOUX. IDENTIFICATION DES BLOCS D'ESSAIS "SOL" EN ROUGE.	- 12 -
FIGURE 3 : SCHÉMA DU DISPOSITIF EXPÉRIMENTAL IMPLANTÉ DANS LES DIFFÉRENTES ZONES DU CHAMP, ICI PRÉSENTÉ POUR LE BLOC 2 (LIGNES 7 À 12).	- 13 -
FIGURE 4 : DESCRIPTION DU PROFIL DU BLOC 2 DU PETIT BORDIA RÉALISÉ PAR L'UNITÉ DE GÉOPÉDOLOGIE DE GEMBOUX AGRO-BIO TECH (ULG).	- 14 -
FIGURE 5 : SCHÉMA DU DISPOSITIF GÉNÉRAL D'IMPLÉMENTATION SUR LE TERRAIN	- 18 -
FIGURE 6 : MÂTS EN ALUMINIUM SUR LESQUELS REPOSENT LES NŒUDS ET LES CAPTEURS (À GAUCHE ET AU CENTRE) - IMPLANTATION DES CAPTEURS SOL, ICI DECAGON EC5, FIXÉE À UNE PROFONDEUR DE 20 ET 50 CM (DROITE).	- 18 -
FIGURE 7 : ORGANISATION MODULAIRE DE STICS – EXTRAIT DE BRISSON <i>ET AL.</i> (2008)	- 19 -
FIGURE 8 : REPRÉSENTATION SCHÉMATIQUE DU FONCTIONNEMENT DU LOGICIEL STICS - ADAPTATION D'UN SCHÉMA DE RUGET <i>ET AL.</i> (2002)	- 20 -
FIGURE 9: FONCTION DE DENSITE DE PROBABILITE MARGINALE (MARGINAL PDF) POUR QUATRE PARAMETRES DU MODELE STICS : <i>efcroiveg</i> , <i>efcroijuv</i> , <i>stlevamf</i> ET <i>psisto</i> . LES HISTOGRAMMES CORRESPONDENT AUX RESULTATS OBTENUS AU TERME DU PROCESSUS D'ECHANTILLONNAGE FAISANT EMPLOI D'UNE FONCTION DE VRAISEMBLANCE TYPE SOMME DES CARRÉS DES ECARTS PONDERES (WLS).	- 44 -
FIGURE 10: SIMULATIONS OBTENUS APRES ECHANTILLONNAGE PARAMETRIQUE SUR LA SITUATION CULTURALE DE L'ANNEE 2008-2009 (POINTS NOIRS). LE TRAIT NOIR REPRESENTE LA SIMULATION CONDUITE AVEC L'ESTIMATEUR PARAMETRIQUE MOYEN, OBTENU RESPECTIVEMENT (DE GAUCHE A DROITE PUIS DE HAUT EN BAS) AVEC LES FONCTION DE VRAISEMBLANCE SLS, ET CV (0.05 - 0.0147 - 0.50). LES NIVEAUX D'INCERTITUDE (A 95%) ASSOCIES A L'ESTIMATION PARAMETRIQUE SUR LES SIMULATIONS SONT REPRESENTES PAR L'AIRES GRISE.	- 46 -
FIGURE 11: RMSE'S EF'S ET ND'S OBTENUS AU-TERME DES DIFFERENTS PROCESSUS D'ECHANTILLONNAGE (SLS - WLS - CV) EN COMPARAISON DE LA SIMULATION SUR LE JEU DE PARAMETRE INITIAL (TRAIT HORIZONTAL NOIR). L'EVALUATION DU MODELE PEUT-ETRE CONDUITE SUR LES SIMULATIONS REALISEES AU MOYEN DE L'ESTIMATEUR PARAMETRIQUE MOYEN (HISTOGRAMMES GRIS) OU SUR LA MEDIANE DES SIMULATIONS ((HISTOGRAMMES NOIRS).	- 47 -
FIGURE 12: REPRESENTATION SCHEMATIQUE D'UNE BASE DE DONNEES CLIMATIQUES ENSEMBLE (ICI POUR UNE VARIABLE CLIMATIQUE DONNEE CV). LA DIMENSION 1 A <i>Nd</i> REPRESENTE LE NOMBRE DE JOURS NECESSAIRES A LA SIMULATION D'UNE SAISON DE CULTURE. L'INDEX CV REPRESENTE UNE VARIABLE CLIMATIQUE DONNEE. LES DIFFERENTES ANNEES DE LA BASE DE DONNEES SONT CONTENUES DANS LA TROISIEME DIMENSION (VE1 A VE30) ET REPRESENTÉES SOUS L'INDICE <i>i</i> .	- 90 -
FIGURE 13: SIMULATION DE LA MATIERE SECHE TOTALE AU COURS DE LA SAISON (GAUCHE) ET RENDEMENT TOTAL FINAL CORRESPONDANT, EN FONCTION DU NOMBRE DE JOURS DE CLIMAT REEL CONNU (DROITE). RENDEMENTS OBTENUS AVEC LE CLIMAT MOYEN (TRAIT NOIR CONTINU) ET AVEC LE CLIMAT REEL (TRAIT NOIR POINTILLE) POUR L'ANNEE 2005-06. LES REPRESENTATIONS GRISES REPRESENTENT TOUTES LES COMBINAISONS DE CES DEUX CLIMATS (MATRICE ENSEMBLE), AVEC UN TAUX DE REMPLACEMENT DE 10 JOURS. LES TRAIT EN TIRETS NOIRS CORRESPONDENT A UNE ERREUR DE PREDICTION DE 10 % DU RENDEMENT FINAL.	- 91 -
FIGURE 14: A GAUCHE : EVOLUTION DU RENDEMENT GRAIN POUR TOUTES LES "MATRICES-ENSEMBLES" ET TOUTES LES ANNEES DE LA BASE DE DONNEES. LE TRAIT NOIR CONTINU CORRESPOND A LA SIMULATION AVEC LE CLIMAT MOYEN. A DROITE : DISTRIBUTION DE PROBABILITE CUMULEE RELATIVE A LA PREDICTION DU RENDEMENT GRAIN.	- 93 -
FIGURE 15 : DISTRIBUTION DE PROBABILITE (HISTOGRAMME GRIS) ET FONCTION DE DENSITE CUMULEE (TRAIT GRIS) DES RENDEMENTS SIMULEES POUR 300 SERIES TEMPORELLES STOCHASTIQUEMENT GENEREES. DISTRIBUTIONS LOG-NORMALES (TRAIT NOIR) CALCULEES SUR BASE DES VALEURS CARACTERISTIQUES DES OBSERVATIONS.	- 117 -
FIGURE 16 : COMPARAISON DES PREDICTIONS DES SORTIES DU MODELE POUR LES DEUX METHODOLOGIES DE PREDICTION DU RENDEMENT APPLIQUEES AUX ENTREES CLIMATIQUES DE LA SAISON 1981-1982. L'INTERVALLE DE CONFIANCE A 95% AUTOUR DE LA PREDICTION EST PRESENTE EN TRAIT GRIS DISCONTINU. L'ESPERANCE MATHÉMATIQUE (MOYENNE) DES OBSERVATIONS EST PRESENTEE EN TRAIT GRIS CONTINU. LA SORTIE DE LA METHODOLOGIE REPOSANT SUR LA NORMALE SAISONNIERE EST PRESENTEE EN TRAIT NOIR.	- 118 -
FIGURE 17 : CONFRONTATION GRAPHIQUE DES DEUX APPROCHES DE PREDICTION DU RENDEMENT.	- 119 -
FIGURE 18 : COMPARAISON DES RESULTATS DE L'EXPERIMENTATION NUMERIQUE (TRAIT ET HISTOGRAMMES GRIS) AVEC LA DISTRIBUTION DE PEARSON TYPE I (TRAIT NOIRS) AJUSTEE SUR BASE DE LA METHODE DES	

MOMENTS. A GAUCHE : FONCTION DE DENSITE DE PROBABILITE CUMULEE POUR UN NIVEAU DE FERTILISATION 0-0-0 kgN.ha^{-1} . A DROITE : FONCTION DE DENSITE DE PROBABILITE CUMULEE POUR UN NIVEAU DE FERTILISATION 60-60-60 kgN.ha^{-1} . - 145 -

FIGURE 19 : ETUDE DU NIVEAU D'ASYMETRIE DANS LES SIMULATIONS CONDUITES SOUS DIFFERENTS NIVEAUX DE FERTILISATION, VARIANT DE 0-0-0 kgN.ha^{-1} A 100-100-100 kgN.ha^{-1} APPLIQUES AUX STADES 'TALLAGE', 'REDRESSEMENT' ET 'DERNIERE FEUILLE', ET POUR DES INCRÉMENTS DE 30 kgN.ha^{-1} PAR APPORT (10-10-10 kgN.ha^{-1}). A GAUCHE : SKEWNESS DES DISTRIBUTIONS DES TRAITEMENTS 0-0-0 kgN.ha^{-1} (T1) A 100-100-100 kgN.ha^{-1} (T11). A DROITE : ETUDE DU RENDEMENT EN GRAINS A DES NIVEAUX DE FERTILISATION AZOTEE CROISSANTS (TRAITS DE NUANCE GRISE CROISSANTE) ET SOUS DIFFERENTS (#300) SCENARIOS CLIMATIQUES GENERES STOCHASTIQUEMENT (CDF). LES COURBES DONT L'INTENSITE DE GRIS AUGMENTE SUR LE GRAPHE DE DROITE REPRÉSENTENT LES TRAITEMENTS N CROISSANTS. LA MOYENNE(--*--), LA MÉDIANE(--o--) ET LE MODE (--o--) DES DISTRIBUTIONS ONT ÉTÉ NUMÉRIQUEMENT DÉRIVÉS. - 146 -

FIGURE 20 : FONCTION DE DENSITE CUMULEE (CDF) DU REVENU MARGINAL NET (MNR) EXPRIME EN FONCTION DU NIVEAU DE FERTILISATION AZOTE ET TRACE POUR 300 SCENARIIS CLIMATIQUES. LES DIFFERENTS PROTOCOLES AZOTES CORRESPONDENT A UNE APPLICATION SYSTEMATIQUE DE 60 kgN.ha^{-1} AUX STADES TALLAGE ET REDRESSEMENT, MAIS POUR DES NIVEAUX CROISSANTS D'AZOTE APPLIQUES AU STADE DERNIERE FEUILLE VARIANT DE 0 kgN.ha^{-1} (M60-1) A 100 kgN.ha^{-1} (M60-11) PAR PAS DE 10 kgN.ha^{-1} . POUR CHAQUE NIVEAU DE PROBABILITE (1%, 5% 10%, ..., 95%, 99%) LA FRACTION MAXIMISANT LE MNR EST REPRESENTEE PAR LE POINT NOIR (•) - 148 -

Liste des tableaux

TABLEAU 1 : TRAITEMENTS PHYTOSANITAIRES APPLIQUÉS EN GRANDE PARCELLE EN 2009.	- 13 -
TABLEAU 2 : MODALITÉ D'APPORT DE LA FERTILISATION AZOTÉE APPLIQUÉE EN PETITES PARCELLES	- 13 -
TABLEAU 3 : CLASSEMENT DES DONNÉES MESURÉES SELON LEUR ORIGINE ET LEUR UTILISATION DANS LE MODÈLE.	- 21 -
TABLEAU 4 : VALEURS ATTENDUES DES PARAMÈTRES D'ÉVALUATION D'UN MODÈLE, DANS LE CAS DU MODÈLE PARFAIT.	- 24 -
TABLEAU 5 : PRESENTATION DES RMSE ENTRE VALEURS SIMULEES ET OBSERVEES (LAI, HUR1, HUR2) ET, LE CAS ECHEANT, DE LA MOYENNE DE LA VALEUR DES PARAMETRES EN REGIME ETABLI FINAL, POUR L'APPLICATION DES FILTRES EKF ET VF SUR DEUX PROBLEMES DIFFERENTS (EQ. 12 ET 13). NC = NON CONVERGENCE DU PARAMETRE	- 68 -
TABLEAU 6 : COMPARAISON DES DISTRIBUTIONS DU REVENU MARGINAL NET EN UTILISANT UN TEST DE WILCOXON, POUR DIFFERENTS TRAITEMENTS DE FERTILISANT AZOTE APPLIQUES, BASE SUR UNE MODULATION DE LA DOSE DE DERNIERE FEUILLE (60-60-XX kgN.ha ⁻¹). LES RESULTATS DANS LE TABLEAU REPRESENTENT LA P-VALEUR DU TEST DE WILCOXON. LES NIVEAUX DE SIGNIFICATION SONT REPRESENTES PAR LE NOMBRE D'ETOILES ('***' : $P < 0.001$; '**' : $P < 0.01$; '*' : $P < 0.05$).	- 149 -

Introduction générale

"À force d'annoncer les événements, on en provoque l'accomplissement."

*Louise Ackermann
Pensées d'une solitaire - 1903.*

Chapitre 1 : Introduction générale

1. Enjeux agro-environnementaux et agriculture de précision

Depuis 1950, la rapide intensification des systèmes de production agricole a eu pour conséquence une augmentation importante des intrants et, en particulier, des fertilisants azotés (Van Alphen and Stoorvogel, 2000). Pour s'assurer de l'obtention du rendement potentiel, c'est-à-dire le rendement limité exclusivement par les variables atmosphériques et la disponibilité en eau (Reid, 2002; Robertson *et al.*, 2008), les agriculteurs ont bien souvent appliqué des quantités d'azote excédentaires par rapport aux besoins des cultures (Lemaire *et al.*, 2008).

Or, à l'heure actuelle, sur le plan économique, le renchérissement de l'énergie nécessaire à la fabrication des engrais azotés et la demande croissante en engrais dans les pays émergents entraînent une augmentation spectaculaire du prix des engrais azotés.

Au plan environnemental, l'excès d'engrais azoté par rapport aux besoins stricts de la culture entraîne un accroissement du reliquat azoté à moyen terme. A moins que des mesures ne soient prises, comme par exemple l'implantation d'une Culture Intermédiaire Piège À Nitrate (CIPAN), l'excès d'engrais azoté conduit à une pollution des eaux superficielles et des aquifères (Basso and Ritchie, 2005; Mary *et al.*, 2002). A ces pertes sont associés divers effets néfastes sur l'environnement comme l'eutrophisation ou la diminution de biodiversité. En outre, certaines études suggèrent l'existence d'effets négatifs sur la santé humaine liés à la consommation d'eaux trop fortement chargées en nitrate. Le temps de transfert des nitrates étant très variable en fonction de la profondeur à laquelle est située l'aquifère, de la nature du matériau, ..., on estime que, dans certains cas, les effets néfastes de pratiques culturales inadaptées pourraient encore se manifester plusieurs années après l'arrêt de toute fertilisation.

Par ailleurs, à l'échelle annuelle, l'accroissement de la fertilisation azotée favorise les émissions de gaz nuisibles (NH_3 , N_2O et NO_x) vers l'atmosphère (Germon and Hénault, 2002). Ainsi, 65 % des émissions de N_2O , gaz contribuant à l'effet de serre, seraient attribuées aux processus de nitrification et dénitrification dans le sol. Pour le secteur agricole, les causes principales de l'accroissement de ces émissions sont dues à l'application d'engrais azotés et de matières organiques, résidus de l'élevage d'animaux. Des études menées en 2005 (McSwiney and Robertson) ont rapporté que les flux de N_2O étaient faibles jusqu'à ce que l'apport en azote excède les besoins de la culture, instant à partir duquel les flux peuvent presque doubler.

Parmi les objectifs de l'agriculture de demain, accroître l'efficacité de la fertilisation azotée tout en réduisant ses impacts environnementaux tient donc une place majeure. C'est pourquoi la Communauté Européenne a lancé plusieurs directives visant à réduire la pollution des eaux induite par les nitrates provenant de l'agriculture (EC-Council Directive, 1991). En 2002, le gouvernement Wallon a transposé la directive Nitrate 91/676/EEC dans la législation Wallonne en initiant le Programme de Gestion Durable de l'Azote (PGDA) (Vandenberghé *et al.*, 2011). Ce plan vise à promouvoir de bonnes pratiques agricoles en termes de gestion de la fertilisation azotée à l'échelle de la parcelle, afin de maintenir de hauts niveaux de rendement tout en réduisant les impacts environnementaux.

De nombreux auteurs voient ainsi dans le concept d'*agriculture de précision*, défini comme un

« système de gestion basé sur la connaissance et qui doit permettre d'optimiser les profits tout en minimisant les impacts de l'agriculture sur l'environnement » (Wright *et al.*, 2009), la possibilité d'offrir une réponse à la double problématique que constituent les enjeux économiques et environnementaux liés à la gestion de la fertilisation azotée.

Guérif *et al.* (2001) distinguent trois phases dans la mise en œuvre de l'agriculture de précision. La première phase porte sur la caractérisation et la quantification de la variabilité du milieu, incluant à la fois celle du peuplement végétal et celle du sol. La seconde phase repose sur l'établissement de règles de modulation, focalisées ici sur la fertilisation azotée. Finalement, la troisième étape est constituée par les techniques de mise en œuvre de cette modulation.

Au sein même de la seconde phase, les auteurs distinguent deux grands types d'approches, visant à établir des règles décisionnelles au départ de la caractérisation de l'état de la culture et de celui de l'environnement. La première approche, dite approche "indicateur" s'appuie le plus souvent sur des mesures à réaliser en de nombreux points de la parcelle, caractérisant l'état du couvert végétal ou du sol. La seconde approche, de type "modèle de culture", bien que plus complexe, présente plusieurs avantages. Tout d'abord, la raison majeure pour l'adoption des modèles de culture dans les systèmes décisionnels vient de leur capacité à prendre en compte explicitement les caractéristiques du sol, du climat, de la plante et des pratiques culturales, mais aussi et surtout de leurs possibles interactions. De plus, tandis que les formalismes sont construits sur base d'équations différentielles permettant de simuler de façon dynamique les phénomènes à l'œuvre, la plupart des processus modélisés sont physiquement basés et reliés à des variables atmosphériques indépendantes. Les simulations peuvent ainsi être conduites au jour le jour, afin de tenir compte de l'évolution dans le temps du système sol-plante-atmosphère. Cela rend possible l'établissement d'un diagnostic continu du système, mais permet aussi de réaliser des pronostics sur l'état futur de la culture. Finalement, fait important, plusieurs variables du système sol-plante-atmosphère sont simulées de façon concomitante, ce qui permet d'élaborer des critères de type agro-environnementaux prenant en compte simultanément le rendement, en termes quantitatifs et qualitatifs, mais aussi les éventuelles pertes nuisibles vers l'environnement.

Pour toutes ces raisons, les modèles de culture apparaissent comme des approches particulièrement prometteuses pour servir de base au processus décisionnel lié à la planification des opérations agricoles, pour étudier l'effet de différentes techniques culturales ou pour optimiser ces pratiques (Basso *et al.*, 2011; Ewert *et al.*, 2011).

McBratney *et al.* (2005) reconnaissent et mettent en évidence tout le travail dédié à la cartographie des rendements (Arslan and Colvin, 2002) ou à la quantification de la variabilité entre zones (Basso *et al.*, 2012). Cependant, analysant les directions possibles des recherches futures à mener dans le domaine de l'agriculture de précision, McBratney *et al.* (2005) pointent différentes thématiques sur lesquelles l'attention devrait se focaliser en priorité. Afin d'amener le concept évolutif d'agriculture de précision à son potentiel maximal, les auteurs stipulent que l'analyse des variations temporelles, tout à la fois à l'échelle inter- et intra-annuelle, devrait désormais faire l'objet de plus d'attention. Ce constat rejoint celui de Basso *et al.* (2012) qui, sur base d'une approche "modèle de culture", cherchent à déterminer les pratiques azotées les plus performantes. Les auteurs affirment que, s'il est évident que la gestion azotée doit être réfléchie d'année en année au gré des conditions climatiques (variation inter-annuelle), la complexité du processus décisionnel en cours de saison vient

de la non-connaissance des conditions climatiques futures (variation intra-annuelle), tandis que l'effet réel de la quantité de fertilisant apportée ne sera connu qu'au terme de la saison.

2. Les modèles de culture

2.1. Développement des modèles

Depuis bientôt 40 ans, les modèles de culture ou écophysiologicals fournissent une représentation conceptuelle du système sol-plante dans différentes conditions agro-environnementales. Ces modèles décrivent les processus physiques et biologiques qui caractérisent la croissance et le développement d'une culture en interaction avec son environnement, à l'aide d'équations différentielles, linéaires ou non (Campbell and Norman, 1989; Semenov and Porter, 1995). Ainsi, la majorité des relations qui unissent la dynamique de croissance d'une plante aux variables climatiques sont non-linéaires et interdépendantes, une croissance favorable étant assurée par l'existence concomitante d'une humidité adéquate du sol, d'une température adaptée et d'un ensoleillement suffisant. Par ailleurs, l'absence ou l'excès d'un ou de plusieurs des facteurs de croissance conduit à divers degrés de stress abiotiques pouvant eux-mêmes affecter variablement les différentes composantes du développement d'une culture (appareil végétatif et reproducteur).

Plusieurs familles de modèles ont été créées, selon des concepts souvent proches, mais avec des approches plus ou moins empiriques ou mécanistes et des spécificités qui varient selon les objectifs. On distingue des modèles spécifiques dédiés à l'étude de la croissance du blé ou du maïs, comme CERES-Wheat (Ritchie *et al.*, 1984 ; Singh *et al.*, 2008), ARCWHEAT (Weir *et al.*, 1984), CERES-Maize (Jones *et al.*, 1986) et des modèles plus génériques tels que EPIC (William *et al.*, 1989), WOFOST (Van Diepen *et al.*, 1989), DAISY (Hansen *et al.*, 1990), STICS (Brisson *et al.*, 1998 ; 2003, 2008) et SALUS (Basso *et al.*, 2005 ; 2009).

En parallèle à cela, certains modèles sont davantage axés sur la compréhension du fonctionnement écophysiological, comme SUCROS (De Wit, 1978) ou ARCWHEAT. D'autres sont initialement dédiés à l'analyse des implications des pratiques agricoles, tels que CERES ou CROPGRO (Boote *et al.*, 1998). D'autres modèles, tel EPIC ou DAISY, ont par contre comme vocation de prendre en compte les enjeux environnementaux. Finalement, depuis les années 90, des modèles plus complets, comme STICS ou SALUS, intègrent à la fois des objectifs agronomiques et environnementaux. Grâce à la prise en compte de formalismes génériques, offrant ainsi la capacité de simuler la croissance d'un grand nombre de cultures, STICS offre quant à lui la possibilité d'analyser les systèmes de cultures.

2.2. Emploi des modèles

Traditionnellement, les modèles écophysiologicals jouent un double rôle. Ils constituent une approche privilégiée pour étudier, voire pour mettre au point les systèmes de culture en présence de combinaisons complexes de techniques et de successions de cultures, face à la variabilité du sol et du climat (actuel et futur). Outre le rendement, ils permettent de modéliser des critères de qualité comme la teneur en protéines des grains de blé qui est directement liée au fonctionnement carboné et azoté de la culture (Génard *et al.*, 2006). Ils permettent ainsi globalement de comprendre l'influence des conditions agro-environnementales sur le système sol-plante.

D'autre part, ils sont capables de fournir des indicateurs d'impacts environnementaux tels que la lixiviation des nitrates, la salinisation des sols ou encore la séquestration du carbone dans le sol (Loyce *et al.*, 2006 ; Beaudoin *et al.*, 2008). Etant donné les exigences de la directive européenne 91/6/76, les indicateurs relatifs à l'azote sont apparus les premiers dans les modèles. La quantité d'azote non prélevée par les racines au cours du cycle cultural et l'azote minéral contenu dans le sol après la récolte sont des variables privilégiées.

Les modèles de culture ont également été utilisés à différentes échelles spatio-temporelles, que ce soit à des fins agronomiques, économiques ou environnementales. A l'échelle internationale ou nationale, les modèles peuvent être utilisés pour réaliser le suivi des productions agricoles (Rugé *et al.*, 2001). En s'intéressant à différentes échelles temporelles, les modèles ont été utilisés pour étudier les impacts du changement climatique (Ewert *et al.*, 1999, Semenov and Doblans-Reyes, 2007). A des échelles locales, allant du petit bassin versant à la parcelle, les modèles sont utilisés dans le cadre de l'agriculture de précision pour suivre l'impact des pratiques agricoles sur la qualité de l'eau (Beaudoin *et al.*, 2008) ou sur la distribution spatiale des rendements (Basso *et al.*, 2009) ou encore pour optimiser spatialement les pratiques agricoles telles que la fertilisation azotée (Houles *et al.*, 2004), de façon à maximiser le rendement ou le revenu net de l'exploitant tout en minimisant le lessivage des nitrates vers la nappe (Basso *et al.*, 2010).

Plus récemment, des adaptations ont été apportées aux modèles afin d'offrir une aide à la décision en temps réel, prenant en compte à la fois des aspects agronomiques et environnementaux. Les applications concernent par exemple la recommandation de doses de fertilisants adaptées à des critères agro-environnementaux (Houlès *et al.*, 2004), la simulation de l'état de prairies pour anticiper la prise de décision en présence de fortes sécheresses (Rugé *et al.*, 2009), etc. Les modèles de culture peuvent ainsi se révéler être des outils précieux pour les acteurs de terrains.

Finalement, il convient d'apporter une dernière remarque générale liée à l'emploi des modèles. Un modèle est avant tout un compromis judicieux entre le niveau de "réalisme recherché" obtenu le plus souvent au prix d'une complexité accrue et son opérationnalité. Les modèles de culture les plus récents, tels que STICS, permettent de simuler les systèmes de culture en intégrant des interactions entre les composantes agronomiques et environnementales. Ce niveau de représentation des systèmes réels n'est obtenu qu'au prix de formalismes complexes se traduisant par un nombre important d'équations et de paramètres représentatifs des différents systèmes et phénomènes. De ce fait, la mise en œuvre réelle des modèles de culture et *a fortiori* le développement d'outils d'aide à la décision qui en sont dérivés relèvera ou sera largement influencé par la connaissance et le savoir-faire du modélisateur. Dans l'extrapolation du modèle sous d'autres conditions que celle du développement initial, le modélisateur devra rester prudent et éviter de tirer trop vite des conclusions qui pourraient être davantage imputables au formalisme du modèle qu'aux conditions d'entrée.

2.3. Contraintes associées à l'emploi des modèles agro-environnementaux

Dans la recherche en modélisation agro-environnementale, les mesures de terrain resteront toujours une référence. Elles permettent de quantifier l'état du système étudié mais aussi l'incertitude liée à celui-ci, qu'elle soit associée à la stratégie de gestion, à la variété étudiée ou aux conditions pédoclimatiques.

La pertinence de l'utilisation des modèles de culture repose sur la qualité des simulations

réalisées (Varella, 2011). Ces dernières pourront être obtenues (i) sur base d'un modèle de qualité par la pertinence des formalismes sur lesquels il repose, (ii) via une détermination précise des nombreuses entrées du modèle et (iii) au travers d'erreurs associées aux simulations les plus faibles possible.

Les entrées concernent les variables climatiques journalières (température, pluie, rayonnement, ...), les variables caractérisant le sol (teneurs en eau aux points de flétrissement et aux capacités au champ, densité apparente, ...), les variables déterminant les grands processus comme le cycle de minéralisation-déminéralisation, le transfert d'eau et d'azote dans la matrice du sol, ..., les caractéristiques écophysiologiques de la plante relevant de l'espèce ou de la variété transcrites sous formes de paramètres (besoin en température, efficacité d'absorption du rayonnement solaire, ...) et les itinéraires techniques mis en place par l'agriculteur (dates de semis et de fertilisation, ...).

Les variables climatiques sont couramment qualifiées de variables de forçage du modèle car elles déterminent la croissance au gré de la succession des saisons, des séries temporelles de jours secs et humides, chauds ou froids, ... Dans le cadre de l'expérimentation agronomique, elles sont en général mesurées automatiquement et quotidiennement, soit sur le site expérimental (stations locales), soit dérivées par extrapolation de données issues d'un réseau d'observation national.

Les itinéraires techniques agricoles sont quant à eux relativement faciles à connaître, et ce d'autant plus que l'échelle d'application est petite. Ils correspondent aux choix qui ont été faits par l'agriculteur en cours de saison ou aux modalités techniques dont on cherche à quantifier l'impact sur les sorties du modèle.

Les propriétés du sol sont d'une importance capitale en modélisation agronomique. Le sol est le lieu de prélèvement de l'eau et des nutriments nécessaires au bon développement de la plante. Si plusieurs auteurs procèdent à l'estimation ou la ré-évaluation des propriétés de sol par calibration et inversion d'un modèle de culture (Braga and Jones, 2004, Varella *et al.*, 2010b), il est préférable que les paramètres du sol soient déterminés uniquement sur base de mesures. Si les mesures sont de piètre qualité, le modélisateur devra soit s'en accommoder, soit discuter de leur pertinence et éventuellement les ré-évaluer selon l'avis d'un expert pédologue, par exemple. En dernier recours, le modélisateur pourrait les ajuster via un processus d'échantillonnage ou d'optimisation paramétrique.

Il convient également de discuter du cas des paramètres représentatifs de la plante et de certains paramètres généraux. Ceux-ci, ou du moins une première évaluation de ceux-ci, sont en général fournis par les concepteurs de modèles, qui les établissent à partir de la bibliographie, à partir d'expériences menées sur des processus spécifiques de la plante, ou bien à partir d'une calibration préalablement établie sur une large base de données expérimentales (Flenet *et al.*, 2004, Singh *et al.*, 2008). Avant d'entamer la calibration des paramètres du modèle, il est généralement recommandé de procéder à un diagnostic de leur bonne adaptation au contexte d'étude. Ce diagnostic prend bien souvent la forme d'une analyse de sensibilité permettant de mettre en évidence le/les paramètre(s) pour lequel la réponse du modèle à une faible variation du/des paramètre(s) est la plus forte (Ruguet *et al.*, 2002).

2.4. Principales limitations associées à l'emploi des modèles

Beaudoin *et al.* (2008) ont mis l'accent sur l'importance d'avoir des modélisations des pertes en azote de qualité sur des sols différents, des conditions climatiques variées et sur de longues périodes,

étayées par des mesures fiables. La prédiction des besoins en azote des plantes reste une des priorités agronomiques majeures. A celle-ci est étroitement liée la prédiction sur le long terme des pertes en eau et en nitrates des systèmes de culture. Dans ce contexte, Beaudoin *et al.* (2008) soulignent que les modèles de culture ont jusqu'à présent été majoritairement employés sur des expériences annuelles et qu'il existe un intérêt considérable à les mettre en œuvre sur de longues périodes de simulation.

Dans leur étude sur la réponse du blé à la fertilisation azotée, Basso *et al.* (2010) insistent également sur la nécessité de réaliser des études de terrain à long terme, afin d'améliorer la compréhension des mécanismes de croissance d'une culture. Cela doit permettre de définir la meilleure stratégie de gestion de la culture (ici, la fertilisation azotée). En effet, la réponse d'une culture pourrait ne pas se refléter dans sa globalité si le nombre d'années de mesures était insuffisant du fait de la variation des conditions climatiques et principalement celle des précipitations. De plus, lorsqu'ils ne sont pas le résultat de protocoles expérimentaux adéquats ou de mesures fiables, les résultats expérimentaux, issus d'essais en champs très coûteux, peuvent ne pas être transférables à travers le temps ou l'espace.

Dès lors, il convient de définir des protocoles adéquats, présentant le juste compromis entre les coûts d'expérimentations de terrain et la quantité d'informations suffisante pour calibrer et valider un modèle sur plusieurs saisons culturales. Une fois ce stade atteint, la communauté scientifique s'accorde à dire que les modèles de culture, dits « processus orientés » (à voir ici comme un modèle qui simule le développement réel d'une culture, en intégrant les interactions de celle-ci avec son environnement), sont une alternative prometteuse à la simulation des effets à long terme de l'eau et de l'azote sur les rendements d'une culture et leur interaction à travers le temps (Basso *et al.*, 2010). Il a ainsi été prouvé que si un modèle a pu être testé et calibré dans une région particulière, il est possible de l'utiliser pour explorer le potentiel de croissance d'une culture sous différentes stratégies de gestion culturale dans cette région.

3. Définition du cadre et des objectifs de la thèse

Nous nous proposons d'approfondir un certain nombre de questions qui se posent autour de l'utilisation des modèles de culture et plus particulièrement celles relatives à la caractérisation de diverses sources d'incertitudes qui y sont associées dans le cadre plus vaste du développement d'outils d'aide à la décision.

Il est clairement admis qu'une gestion raisonnée de l'azote à l'échelle de la parcelle agricole montre des limites techniques en raison de la variabilité du milieu (Power *et al.*, 2000) qu'il convient donc de caractériser au mieux. De plus, dans le cadre d'une étude à caractère méthodologique sur la mise en œuvre d'une fertilisation azotée différenciée au sein d'une parcelle agricole, Beaudoin *et al.* (2007) ont mis en évidence l'intérêt d'une paramétrisation indépendante du modèle selon la partie étudiée du champ.

Dans cette thèse, le choix a été fait de se concentrer sur le cas d'étude que constitue une parcelle culturale implantée en Hesbaye (Belgique), sur un sol limoneux, en région tempérée. La Hesbaye représente 240.000 ha en Belgique. Les zones cultivées y couvrent plus de 80% de l'espace non artificialisé (jusqu'à 90% dans certains villages). Par son épaisse couverture de limon nivéo-éolien, elle offre un des meilleurs sols pour l'agriculture, ce qui lui vaut souvent l'appellation de "grenier à blé" de

la Belgique. La plupart des propriétés du sol sont mesurées afin d'assurer la paramétrisation du module sol. Au sein de cette parcelle, une culture de froment d'hiver (*Triticum aestivum* L.) est suivie, le cultivar Julius étant choisi parce qu'il offre des caractéristiques similaires à celles des variétés largement plantées à l'heure actuelle. Dans une certaine mesure, on peut donc estimer que les résultats obtenus sur ce cas d'étude seront généralisables à d'autres situations agro-environnementales similaires.

Par ailleurs, comme le signalent McBratney *et al.* (2005) et Basso *et al.* (2012), le développement d'outils d'aide à la décision est désormais étroitement lié à l'étude des variabilités inter- et intra-annuelles. Avant d'entamer la saison culturale, rien ne permet de prédire à long terme et avec une précision suffisante les conditions climatiques qui prévaudront entre le moment du semis et celui de la moisson. Or, développer des techniques visant à pallier ce manque apparaît primordial afin d'utiliser le modèle à des fins réellement prédictives plutôt qu'*a posteriori* à partir d'une expérimentation, comme cela est souvent le cas. Dans cette thèse, l'accent sera donc clairement porté sur la mise au point d'outils d'aide à la décision reposant sur des méthodologies génériques et robustes, qui soient transférables et transposables à différents modèles et applicables à d'autres cas d'étude (type de sol ou culture).

La généricité de l'emploi d'un modèle de culture donné, ici le modèle STICS, sera ainsi recherchée à tous les niveaux, qu'il s'agisse de la manière de paramétrer les modèles, d'assurer la qualité de ses sorties ou d'utiliser ses capacités prédictives.

Dans un premier temps, l'étude portera sur l'incertitude des simulations liée à l'estimation paramétrique. Comme nous le discuterons, la qualité des mesures de terrain est de la plus haute importance dans la mise au point d'un modèle robuste. En effet, les mesures de terrain sont l'interface entre l'utilisateur-modélisateur, elles reflètent les conditions réelles de croissance rencontrées par la plante.

Une fois les mécanismes de fonctionnement du modèle compris et le modèle correctement calibré, il convient d'en faire une utilisation concrète et d'exploiter ses capacités prédictives. Nous verrons ce qu'un modèle robuste peut apporter à l'agriculteur en tant qu'outil d'aide à la décision dans une région donnée. Plus particulièrement, nous verrons comment l'étude du passé peut nous aider à anticiper l'avenir et comment il est possible de prédire les rendements de fin de saison en prélude à l'adaptation des stratégies de gestion culturale.

Pour estimer le rendement en grain de fin de saison ou mettre au point un outil d'aide à la décision sur la gestion de la fertilisation azotée, le modélisateur-développeur sera amené à poser des hypothèses de travail, parfois fortes, comme l'emploi de projections climatiques hypothétiques. Le comportement des modèles et l'implication de telles hypothèses seront analysés.

4. Organisation de la thèse

Cette thèse est donc axée sur la caractérisation et la prise en compte de l'incertitude associée à l'emploi des modèles agro-environnementaux dans une optique de mise au point d'outils d'aide à la décision. La Figure 8 schématise le raisonnement qui sera suivi et le domaine d'expertise mobilisé afin de caractériser l'incertitude associée aux prévisions des modèles.

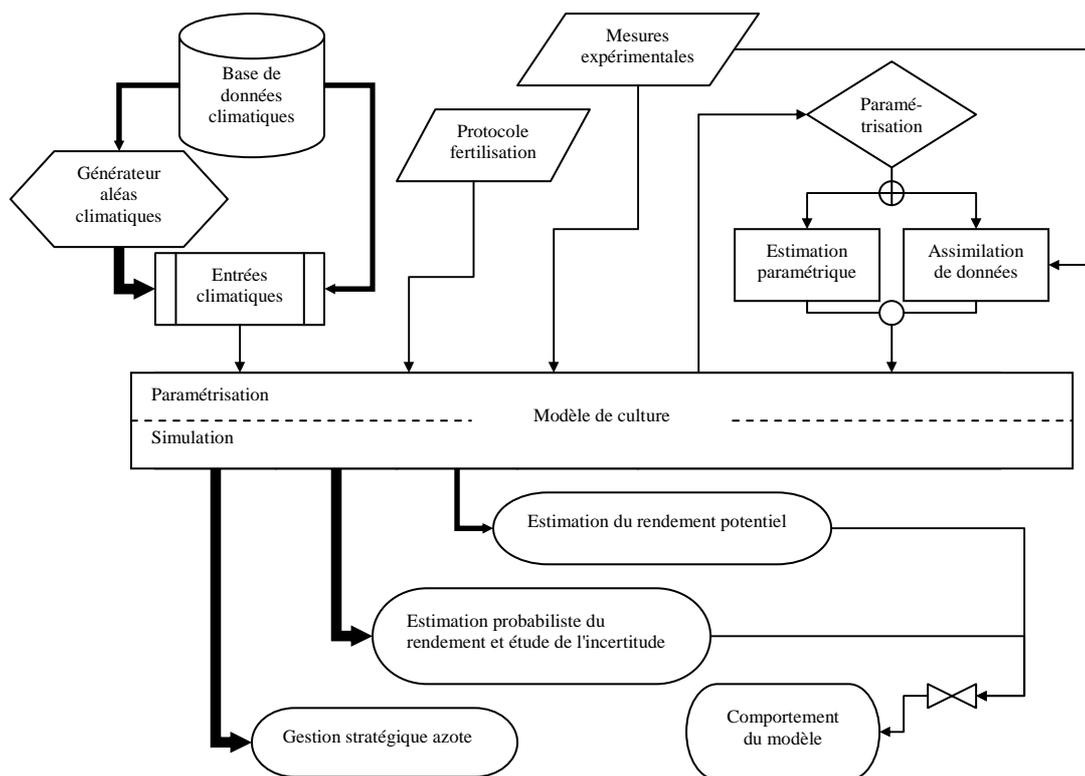


Figure 1 : Organigramme du plan de la thèse.

La structure du document est la suivante :

- Le Chapitre 2 décrit les outils mis en place et relatifs au protocole expérimental, c'est-à-dire les mesures expérimentales qui sont liées à la croissance de la culture sous différents niveaux de fertilisation, ainsi que l'acquisition de données relatives aux variables climatiques.
- Le Chapitre 3 est consacré à une synthèse bibliographique relative à la construction mathématique des modèles de cultures, aux méthodes d'optimisation des sorties de ces modèles par la voie de l'optimisation paramétrique ou de l'assimilation de données et enfin les limitations liées à leur emploi.
- Le Chapitre 4 est dédié à l'étude des méthodes d'estimation paramétrique. Il se focalise plus particulièrement sur les méthodes d'échantillonnage basées sur le théorème de Bayes et une approche formelle du modèle statistique liée à la représentation des résidus entre le modèle et les mesures de terrain.
- Le Chapitre 5 fait l'objet de la caractérisation des méthodes d'assimilation de données.
- Le Chapitre 6 est dédié à l'emploi du modèle calibré. Ce dernier est ici utilisé pour ses capacités prédictives. Dans ce contexte, en prélude au développement d'outils d'aide à la décision, nous verrons comment il est possible de transformer un modèle déterministe en un outil de prédiction du rendement statistiquement fiable.
- Le Chapitre 7 est axé sur le comportement du modèle en tant qu'outil mathématique. Sur base des grandes lois de la statistique, nous rechercherons dans quelle mesure l'outil développé au chapitre précédent et son efficacité s'accordent avec le comportement attendu du modèle.
- Le Chapitre 8 investigate la notion d'incertitude liée à la non-connaissance des conditions climatiques futures, tout en visant à optimiser les pratiques actuelles de fertilisation azotée.

- Finalement, le chapitre 9 synthétise les notions abordées et fournit une conclusion générale aux recherches menées dans le cadre de cette thèse de doctorat.

Du chapitre 4 au chapitre 8, c'est-à-dire pour ce qui concerne le développement scientifique proprement dit, la structure des chapitres est systématiquement la même. La première section introduit la recherche présentée dans le chapitre via une remise en contexte et une description des objectifs poursuivis. La seconde section, communément appelée "*Contributions originales à ...*" présente les points novateurs apportés aux différentes recherches en vue de la mise au point d'outils d'aide à la décision. Ensuite, deux sections générales nommées "*Discussions intermédiaires*" et "*Conclusions intermédiaires*" sont l'occasion de discuter les apports spécifiques et/ou généraux de la recherche menée au sein de chaque chapitre. Le cas échéant, les points faibles et points forts des différentes contributions, ainsi que les éventuelles restrictions ou conditions d'utilisation des méthodologies proposées, sont abordés. Bien que certaines Figures et Tableaux de résultats soient fournis pour étayer les propos, dans la mesure du possible, cette section est généraliste et n'entre pas dans le détail de la recherche si cela n'est pas indispensable aux discussions. L'intérêt de ces chapeaux est clairement d'illustrer et d'analyser avec autant de recul que possible les différentes contributions. Car, *in fine*, chaque chapitre est clôturé par la présentation dans son intégralité d'un article scientifique original, publié ou soumis

Protocole expérimental

*"You can't manage
what you don't observe
or don't measure"*

Dr Maarten Stapper

Chapitre 2 : Protocole expérimental

1. Protocole expérimental

Les expériences de terrain ont été conduites sur une culture céréalière de blé (*Triticum aestivum* L.) implantée au lieu-dit du *Petit Bordia*, à Gembloux (Coordonnées 50.568°N, 4.711°E) (Figure 2). Cet emplacement a été retenu car il offre deux types de sol bien contrastés au niveau de leur horizon labouré, à savoir un limon classique et un limon sableux. Sur chacun des sols, une grande parcelle expérimentale a été implantée. Chaque grande parcelle est constituée de petites parcelles où 7 protocoles d'application de fertilisants azotés sont mis en œuvre, en faisant varier à la fois le moment et la dose de l'apport.



Figure 2 : Identification du champ expérimental par rapport au Centre Wallon de Recherches Agronomiques CRA-W, situé rue du Bordia, Gembloux. Identification des blocs d'essais "sol" en rouge.

1.1. Les essais phytotechniques

La culture retenue est le blé d'hiver (*Triticum aestivum* L.), cultivar Julius. Cette variété a été sélectionnée pour son niveau de rendement élevé, mais aussi pour sa tolérance aux maladies et à la verse. Ces caractéristiques ont été jugées comme étant proches de celles que les agriculteurs seront amenés à privilégier lorsqu'ils choisiront une variété à cultiver à l'avenir. La première campagne de mesure a eu lieu durant la période culturale 2008-2009 et trois autres campagnes de mesure ont suivi.

Après un travail du sol conventionnel, le blé est en général semé entre la mi-octobre et la mi-novembre. Le semis a été effectué selon des modalités classiques, à savoir un interligne de 14.6 cm, une profondeur de semis de 2 cm, et une densité de semis croissante lorsque le blé est planté plus tard dans la saison. La culture a été conduite de manière classique (Tableaux 1 et 2). Les traitements phytosanitaires (Tableau 2) sont conventionnels et appliqués de manière généralisée sur

l'ensemble des parcelles. La récolte de la culture est réalisée avec une moissonneuse-batteuse pour parcelles expérimentales.

Tableau 1 : Traitements phytosanitaires appliqués en grande parcelle en 2009.

Traitement	Stade	Date	Caractéristique de et à l'application
Désherbage		14-avr	Lexus xpe 30g + platform S 1kg - vent : 0m/s - soleil - 20d°C
Raccourcisseur		22-avr	Ccc 1L - vent modéré nord - 20d°C
Fongicide	2 Noeuds	7-mai	Opus 0,6L + sportak 1L - 17d°C - vent fort - AD 6kg pression
	Epi	12-juin	Swing gold 1,25 + caramba 0,8
Insecticide		12-juin	Karaté 0,05 L/ha

Le protocole expérimental est conçu de manière à mesurer les réponses du blé sous différentes conditions de nutrition azotée, en faisant varier aussi bien la dose totale que le rythme d'apport (Tableau 2), pour chacun deux types de sol sur lequel il a été semé.

Tableau 2 : Modalité d'apport de la fertilisation azotée appliquée en petites parcelles

Trait. # Zadoks	Quantité d'azote apportée [kgN.ha ⁻¹]				
	Tallage 23	TR 29	Redressement 30	Dernière feuille 39	Total
Exp 1	0	/	0	0	0
Exp 2	30	/	30	60	120
Exp 3	/	60	/	60	120
Exp 4	60	/	60	60	180
Exp 5	/	90	/	90	180
Exp 6	60	/	60	120	240
Exp 7	/	120	/	120	240

Dans chacune de ces deux zones, un bloc de 114 parcelles a été implanté de façon identique selon un dispositif en blocs aléatoires complets (Figure 3), incluant 4 répétitions (zones blanches et grisées sur la Figure 3). Chaque parcelle est numérotée selon un identifiant ligne (ici pour le bloc 2, allant de 700 à 1200) et un identifiant colonne (allant de 1 à 19). Les deux lignes situées aux extrémités (ici, 700, 800, 1100 et 1200) sont dédiées aux mesures destructives réalisées en cours de saison, et les lignes centrales (ici, 900 et 1000) sont conservées intactes pour mesurer le rendement de fin de saison.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1200	1201 3	1202 1	1203 7	1204 N	1205 N	1206 4	1207 6	1208 2	1209 5	1210 C	1211 3	1212 4	1213 5	1214 N	1215 N	1216 6	1217 1	1218 7	1219 2
1100	1101 3	1102 7	1103 2	1104 S	1105 S	1106 5	1107 4	1108 6	1109 1	1110 C	1111 4	1112 7	1113 3	1114 S	1115 S	1116 1	1117 5	1118 6	1119 2
1000	1001 1	1002 6	1003 4	1004 S	1005 S	1006 3	1007 2	1008 5	1009 7	1010 C	1011 6	1012 5	1013 2	1014 S	1015 S	1016 3	1017 1	1018 7	1019 4
900	901 5	902 2	903 7	904 S	905 S	906 4	907 6	908 1	909 3	910 C	911 3	912 7	913 5	914 S	915 S	916 6	917 1	918 2	919 4
800	801 4	802 6	803 3	804 S	805 S	806 7	807 1	808 5	809 2	810 C	811 6	812 3	813 1	814 S	815 S	816 2	817 4	818 7	819 5
700	701 1	702 2	703 3	704 N	705 N	706 4	707 5	708 6	709 7	710 C	711 7	712 5	713 6	714 N	715 N	716 1	717 2	718 3	719 4

Figure 3 : Schéma du dispositif expérimental implanté dans les différentes zones du champ, ici présenté pour le bloc 2 (lignes 7 à 12).

Finalement, les parcelles sont classées en trois catégories:

- Parcelles destinées à la mise en place de microcapteurs [C] et recevant deux modalités de fumure,

à savoir 0 ou 180 kilos d'azote par hectare [kgN/ha], réparties en trois applications égales ;

- Parcelles pour le suivi de la production d'azote minéral du sol (parcelle « nue » [N] et « cultivée » [S]), cette dernière recevant 180 kgN/ha en 3 fractions ;
- Parcelle de suivi de développement de la biomasse et de mesure du rendement, en 4 répétitions. Ces parcelles numérotées de 1 à 7 se différencient par le niveau total d'azote apporté (de 0 à 240 uN) et par le rythme d'apport (2 ou 3 apports), tel que décrit dans le Tableau 2.

1.2. Caractérisation géopédologique

Grâce à la collaboration du service de Géopédologie de Gembloux Agro-Bio Tech, des fosses pédologiques ont été réalisées au niveau des deux types de sols sélectionnés dans le cadre de ce projet (Figure 4). Outre l'examen visuel du sol, la détermination des particularités de chacune des couches constitutives du sol a été réalisée (triangles texturaux, densité, ...).

Cette caractérisation du sous-sol a permis la modélisation des différences entre sols et d'ainsi rendre compte de leur influence respective sur la croissance de la culture. La description du matériau parental est similaire pour les deux types de sol. Elle peut être lue sur la Figure 4, qui, à titre d'illustration, fournit la description complète du profil du bloc 2.

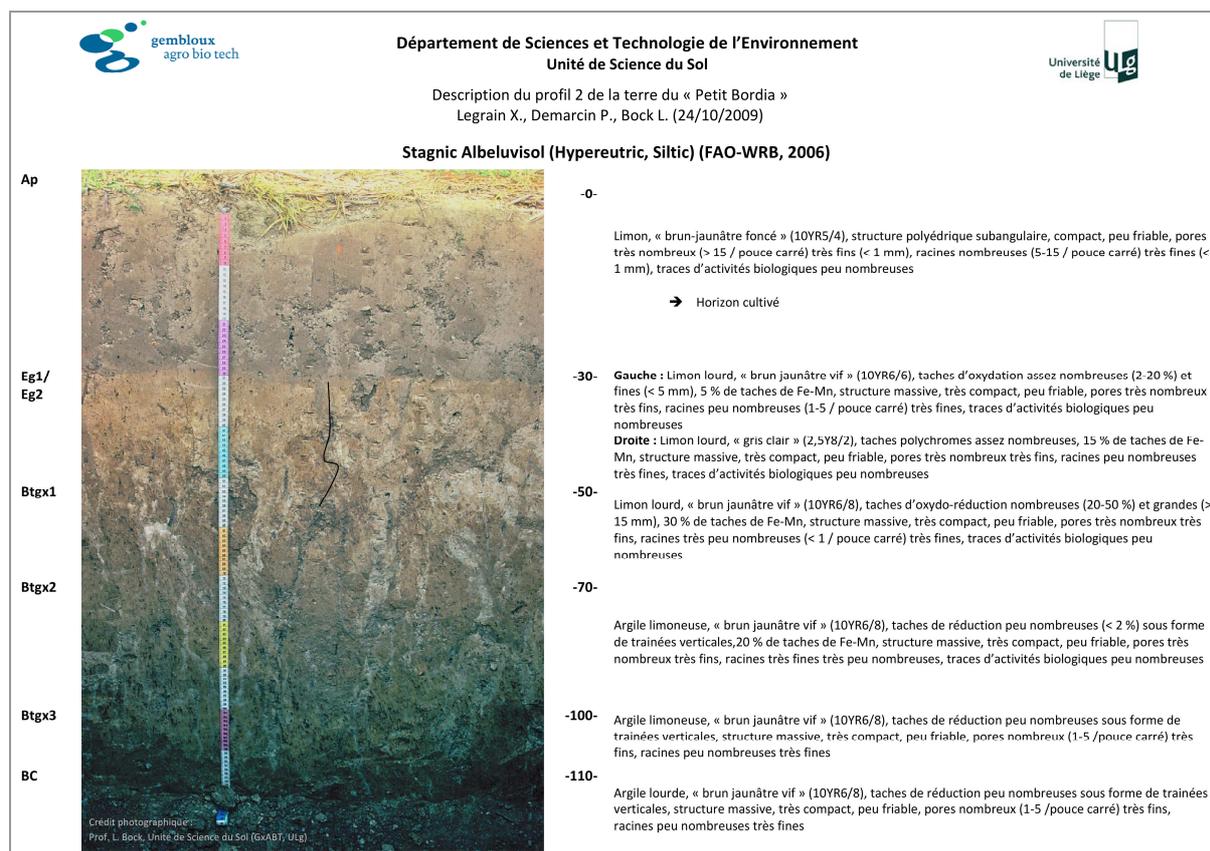


Figure 4 : Description du profil du Bloc 2 du Petit Bordia réalisé par l'unité de Géopédologie de Gembloux Agro-Bio Tech (ULg).

Les deux types de sol se distinguent principalement par leur horizon de surface, explicité ci-après selon la classification FAO-WRB de 2006 :

- Le bloc 1 est un Stagnic Albeluvisol (Hypereutric, Siltic) (FAO-WRB, 2006), dont la première

couche (30 cm) est un limon « brun-jaunâtre foncé » (10YR5/4), avec une structure polyédrique subangulaire, compacte, peu friable. Les pores sont très nombreux (> 15/pouce carré), très fins (< 1 mm), les racines sont nombreuses (5-15 / pouce carré), très fines (< 1 mm) et les traces d'activités biologiques sont peu nombreuses.

- Le bloc 2 est un Luvisol (Hypereutric, Siltic), dont l'horizon de surface (35 cm) un limon sableux lourd « brun » (10YR4/4), avec une structure polyédrique subangulaire, compacte, friable, où les pores sont très nombreux (> 15/pouce carré), très fins (< 1 mm) et où les racines sont nombreuses (5-15/pouce carré) et très fines (< 1 mm). L'horizon présente des traces d'activités biologiques peu nombreuses à nombreuses (Figure 4).

1.3. Mesures agronomiques

Des mesures de la biomasse ont été réalisées à intervalles de 15 jours. Elles permettent un suivi de la production de biomasse sèche totale et en grains. Au moment de la récolte, les rendements grain ont également été mesurés, normalisés à 15 % de teneur en eau. Ces mesures ont été réalisées par l'Unité de Phytotechnie de GxABT (ULg). Il est par ailleurs à signaler que les mesures collectées seront, dans ce document, exprimées la plupart du temps en jours juliens sommés depuis le premier janvier de l'année de semis. Ainsi, à titre d'exemple, une mesure effectuée le 02 janvier correspondra au jour julien 367.

Sur base des échantillons collectés au long de la saison, des mesures de la teneur en azote de la culture, la distinction étant faite entre les tiges et les épis, sont réalisées. La teneur en protéine des grains récoltés est également déterminée. Ces deux types de mesures visent à évaluer la qualité des récoltes, mais aussi à faire le lien avec l'azote exporté du sol.

En parallèle à ces mesures, des mesures régulières du LAI (Leaf Area Index ou indice de développement foliaire) sont pratiquées. Elles sont réalisées à une fréquence d'une mesure par mois environ, selon le développement atteint par la culture. Elles visent à suivre l'expansion foliaire de la culture, qui reste l'interface entre le rayonnement solaire et la production de matière sèche.

1.4. Mesures environnementales

Des mesures de la teneur en azote du sol sont également réalisées par le département Agriculture et Milieu Naturel du CRA-W. Les mesures sont effectuées sous sol nu et sous sol cultivé, au sein des deux grandes parcelles. Ces mesures sont réalisées deux fois par mois, par horizon de 15 cm, jusqu'à 150 cm lorsque cela est possible.

Les mesures de la teneur en azote du sol concernent l'azote ammoniacal NH_4^+ et nitrique NO_3^- . Les mesures sont réalisées dès la sortie de l'hiver (début du mois de mars) jusqu'à la récolte. Sous sol couvert, les mesures sont effectuées sous la fumure classique de 180uN, en 3 apports. L'objectif est ici de suivre la minéralisation de l'azote du sol (sol nu), sa migration dans le sol au grès des événements pluvieux, ainsi que l'épuisement du profil en azote sous sol cultivé. Ce dernier permettra par ailleurs d'avoir une idée de la dynamique de la descente du système racinaire de la culture, celle-ci n'étant pas directement mesurée.

1.5. Le réseau de capteurs sans fil

1.5.1. Choix des micro-capteurs

La société *Crossbow* a récemment mis au point le système *eKo* (*eKo pro series system*, old

Crossbow Technology, now Memsic company, USA), pour l'instrumentation et la recherche agricole et environnementale. Couplé à un système de gestion autonome de l'énergie acquise au moyen de mini panneaux solaires et à un système de transmission sans fil de l'information, ce dispositif inclut un panel de micro-capteurs dit "*plug-and-play*", ainsi que la possibilité d'adjoindre au réseau n'importe quel type de capteur.

Le dispositif est constitué des éléments suivants :

- **Les *éKo Nodes*** : Ces boîtiers constituent les dispositifs de transmission et de relais de l'information acquise au niveau des capteurs. Quatre capteurs peuvent être connectés sur chacun d'eux. Chaque *Node* intègre un dispositif processeur / émetteur radio, et l'antenne qui l'accompagne, ainsi que le dispositif d'alimentation, batteries et/ou panneau solaire (2 heures d'ensoleillement par jour pour assurer le fonctionnement). Ces *Nodes* sont capables d'émettre sur une distance de 150 à 450 mètres (exceptionnellement 600 mètres) selon les conditions locales de dégagement de l'environnement. Chaque *Node* est configuré avec le protocole *XMesh* développé par *Crossbow*. De plus, chacun peut servir de relais, qu'il soit ou non implémenté d'un capteur.
- **L' *éKo Base Radio*** : Dispositif assurant la connexion entre les *éKo Nodes* d'une part, et l'*éKo Gateway* d'autre part. Ce dispositif est modulable selon l'utilisation outdoor ou indoor, avec diverses antennes, de manière à améliorer la capacité du système à relayer l'information.
- **L' *éKo Gateway*** : Dispositif « passerelle », assurant le transfert des données vers un organe de stockage et/ou de traitement de l'information acquise au niveau des capteurs. Elle permet par ailleurs le contrôle de la station radio, la communication et la mise à jour du protocole *XMesh*, ainsi que l'opportunité de visualiser les données au moyen d'une interface web appelée *éKoView*.

Par ailleurs, hormis ce dispositif de relais des données, la société *Crossbow* propose une série de capteurs basés sur diverses technologies, et empruntés à diverses sociétés, mais possédant la particularité d'être connectables facilement aux *éKo Nodes*. Ces capteurs ont été choisis par la société *Crossbow* pour leur robustesse et leur fiabilité.

- **Capteur d'humidité et de température du sol** : Il s'agit d'un capteur *Watermark* (*Watermark company, USA*), qui fournit une mesure du potentiel hydrique du sol. Une mesure de la résistance électrique entre deux électrodes permet de définir la tension d'eau dans le sol (suction), image directe de la difficulté qu'éprouve le système racinaire à extraire l'eau du sol. Ce capteur intègre un thermomètre qui permet de compenser la mesure de la tension d'eau en fonction de la température du sol. Ce capteur peut couvrir la gamme de 0 à 200 kpa, c'est-à-dire l'entièreté de la plage de tension rencontrée dans les sols généralement cultivés.
- **Capteur de mesure de la teneur en eau du sol** : Il s'agit d'un capteur *Decagon-EC-5* (*Decagon company, USA*), qui permet d'obtenir la teneur en eau volumétrique d'un sol, par la mesure de sa constante diélectrique. L'information fournie est exprimée en pourcentage de teneur, mais une étude parallèle menée avec une sonde TDR permettrait d'obtenir la permittivité du milieu.
- **Capteur de température et d'humidité de l'air** : Comme son nom l'indique, ce capteur permet de mesurer la température et l'humidité relative de l'air environnant. Il s'agit en fait d'une sonde *Sensirion SHT7x* (*Sensirion company, Switzerland*) encapsulée sous une coque de protection. Les données qu'elle acquiert sont également utilisées pour mesurer le point de rosée. Pour assurer l'exactitude des mesures en conditions extérieures, le capteur doit être mis à l'abri de

l'ensoleillement direct et de toute autre source de chaleur réfléchi ou rayonnante. Ce capteur tolère la plage de mesure -40° à $+60^{\circ}\text{C}$.

- **Capteur de radiation solaire :** Il s'agit d'un capteur *Davis 6450* (Davis company, USA). Il mesure l'énergie solaire et la radiation globale, c'est-à-dire qu'il intègre à sa mesure les composantes directe et diffuse du rayonnement. L'élément senseur est une photodiode en silicone à large réponse spectrale, capable de détecter les radiations de longueurs d'onde entre 300 et 1100nm. Combinées aux données issues d'un anémomètre et à celles relatives à la température et l'humidité de l'air, les données fournies permettent de calculer l'évapotranspiration des sols.

Disponible depuis peu sous l'interface *eS9200*, la société *Crossbow* a également intégré à son panel de capteurs des senseurs à caractère météorologique :

- **Pluviomètre auto-videur.** Il s'agit d'un capteur *Davis 7852* (Davis company, USA). Utilisé pour afficher les précipitations journalières et cumulées, ce pluviomètre a été créé pour répondre aux directives de l'Organisation Météorologique Mondiale. Ce capteur basculant auto-vidant est d'une conception exceptionnellement précise, au pas de mesure de 0,2 mm. Les augets du pluviomètre sont recouverts d'une mince couche d'aluminium (plus résistant à la corrosion). Les données pluviométriques ainsi acquises permettront de relier l'information à la teneur en eau du sol au niveau de chaque parcelle, en fonction de leur positionnement et de leurs caractéristiques pédologiques.
- **Anémomètre-Girouette :** Il s'agit d'un capteur *Davis 7911* (Davis company, USA) qui permet d'afficher la direction et la vitesse du vent. Cet anémomètre-girouette de précision possède des composants robustes résistant aux ouragans, mais toutefois sensibles aux brises les plus légères. Le moulinet est équilibré à la main pour une stabilité et une précision optimales. Des roulements à billes étanches, en inox, garantissent la longévité de l'appareil. Ce capteur peut être installé sur un tube, un poteau en bois, un mât d'antenne, ... La girouette donne une variation résistive comprise entre 0 et 20 kOhms. 0 Ohms pour une direction de 0° et 20 kOhms pour 360° avec une précision de 5 %. L'anémomètre est un ILS (interrupteur à lames souples) qui réagit au passage d'un aimant. La formule de calcul est basée sur le nombre de tours effectués par unité de temps. Les données relatives au vent permettront, en collaboration d'autres données, d'estimer l'évapotranspiration des différentes parcelles.

1.5.2. Relais de l'information à la station de base

De manière à les conserver à l'abri des intempéries, entraînant un vieillissement prématuré du dispositif, et de tous vols éventuels, il a été décidé de conserver l'*éKobase radio* et l'*éKo gateway* à l'intérieur des bâtiments du CRA-W. Les signaux sont donc envoyés vers la station de base, via une antenne à haut gain (*Doradus SD27 24dBi*, Stella Doradus Ireland Ltd.), permettant de couvrir une distance de plus de 800 m. Celle-ci constitue une amélioration apportée au réseau initial. Les nœuds sont uniquement capables de communiquer à une distance de 150 à 200 mètres (protocole de type *Zigbee*). L'antenne directionnelle limite toutefois le champ de perception du signal à un ou deux nœuds sur le champ (Figure 5).

Afin d'orienter au mieux l'antenne et ainsi améliorer la communication entre le(s) nœud(s) « vu(s) » sur champs et la station de base, une cartographie de la qualité de réception du signal sur champ a été réalisée au cours de la première année. De manière pratique, les *éKo Nodes* ont été montés

sur des mâts en aluminium, afin d'être en condition de visibilité et de communication maximale les uns par rapport aux autres, en vertu de la typologie du terrain. Ces mâts sont montés sur un pied articulé, de manière à assurer leur basculant lors de l'application des traitements phytosanitaires (Figure 6). Les profondeurs choisies pour investiguer les données du sol sont de 20 et 50 cm.

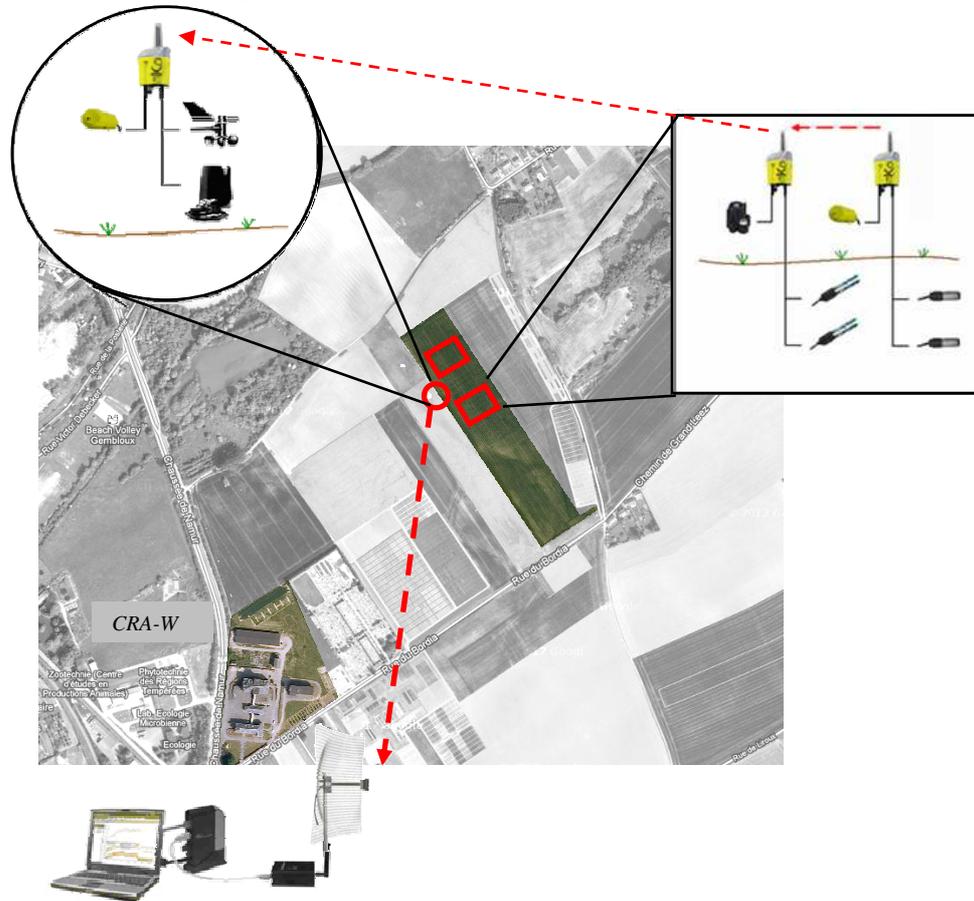


Figure 5 : Schéma du dispositif général d'implémentation sur le terrain



Figure 6 : Mâts en aluminium sur lesquels reposent les nœuds et les capteurs (à gauche et au centre) - Implantation des capteurs sol, ici Decagon EC5, fixée à une profondeur de 20 et 50 cm (droite).

Le réseau de microcapteurs et relais communique ainsi tout au long de la journée avec la station de base. Le temps moyen qui s'écoule entre deux écoutes est d'environ 15 minutes. L'observation des séries de données en provenance du réseau nous a permis de conclure que la fréquence d'écoute était suffisante, bien que toutes les données ne soient pas envoyées à la station de base (intempéries, ...).

2. Le modèle de culture STICS

2.1. Description générale

Le choix du modèle de croissance s'est orienté sur le logiciel STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard, Brisson *et al.*, 2003, 2008), logiciel open source mis au point par l'INRA (Avignon, France).

La présentation du modèle STICS, volontairement non exhaustive et simplificatrice, se borne à ne donner que les éléments essentiels à la compréhension des travaux présentés dans la suite de la thèse. Pour une présentation plus détaillée du modèle, on se référera à l'ouvrage de Brisson *et al.*, 2008.

C'est actuellement la dernière version connue en date qui est employée, la version 6.9. Le modèle est utilisé pour simuler les conséquences des variations du milieu et du système de culture sur la production d'une parcelle agricole. Il s'agit d'un modèle mixte, fonctionnant au pas de temps journalier ou décadaire et dont le formalisme des sous-systèmes s'est voulu tantôt empirique, tantôt mécaniste en vue d'optimiser la robustesse, la genericité et la modularité offerte à ses utilisateurs (Brisson *et al.*, 1998 ; 2003).

Requérant des données climatiques journalières, et basé sur la connaissance initiale de certains paramètres relatifs au sol (profils,..) et aux techniques culturales, le modèle simule la croissance des cultures aussi bien que les balances hydriques et azotées du sol. Il prédit et caractérise à la fois du point de vue qualitatif et quantitatif des variables culturales (rendement, consommation des intrants, ...) et environnementales (perte et prélèvement en eau et nitrate, ...). Pour cela, le modèle simule les bilans de carbone, d'azote et d'eau, les deux derniers dépendants partiellement du premier. L'un des avantages de STICS est sa capacité d'adaptation à la majorité des cultures, par le choix et l'emploi de différents formalismes.

STICS offre une organisation modulaire qui rassemble des processus plus fondamentaux en grands processus biologiques (Figure 7). Cependant, l'une de ses principales qualités réside dans l'interaction constante des différents modules les uns sur les autres.

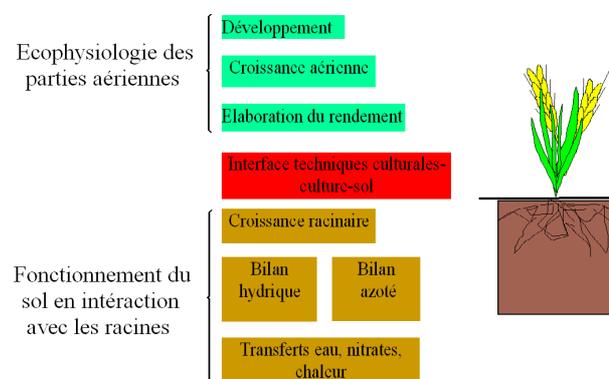


Figure 7 : Organisation modulaire de STICS – extrait de Brisson *et al.* (2008)

2.2. Méthodologie de paramétrisation du modèle STICS

Le modèle STICS requiert beaucoup de paramètres. Le nombre total de paramètres d'entrée est

en effet relativement grand, plus de 200 paramètres dans la version 6.9. Par paramètres d'entrée, il convient d'entendre les conditions initiales du modèle, les paramètres propres aux différents modules de STICS (*e.g.* sol, plante, ...), mais aussi les paramètres définissant les fonctions de stress et les relations entre les différentes étapes de l'élaboration du rendement (Figure 8). Leur attribuer une valeur suffisamment représentative de la réalité est aussi important que d'employer le formalisme le plus à même de représenter le phénomène étudié.

L'adaptation de certains paramètres, moins génériques, nécessite parfois de mettre en place des expériences spécifiques, qui peuvent s'avérer fastidieuses et coûteuses. C'est pourquoi certains auteurs se sont attachés à la mise au point de méthodes d'analyse de sensibilité, en vue d'identifier les paramètres à déterminer en priorité (Ruguet *et al.*, 2002).

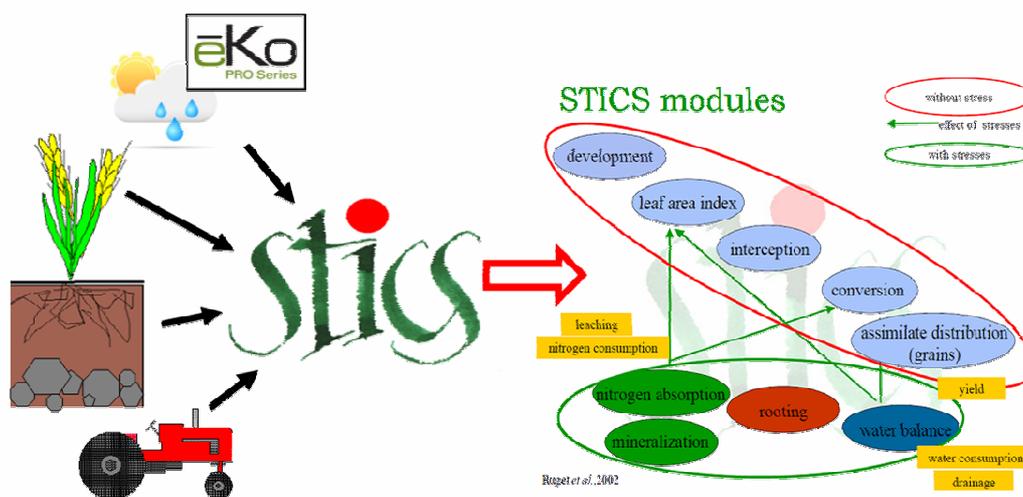


Figure 8 : Représentation schématique du fonctionnement du logiciel STICS - Adaptation d'un schéma de Ruguet *et al.* (2002)

Par ailleurs il est également possible de réaliser une paramétrisation des inputs du modèle. Flenet *et al.*, (2004) et Launay *et al.*, (2005) proposent plusieurs voies pour réaliser la paramétrisation:

- La recherche dans la bibliographie existante. Dans ce cas, la valeur du paramètre est directement obtenue à partir des travaux d'écophysiologie ou à partir d'autres modèles déjà existants sur la culture. Cependant, il est important de s'assurer que le paramètre trouvé dans la bibliographie correspond bien à la même définition que celle de STICS ;
- Le calcul à partir de données expérimentales spécifiques. On distinguera ici deux cas :
 - Soit le paramètre peut être mesuré plus ou moins directement. Les observations ponctuelles réalisées sur le terrain, la réalisation d'échantillonnage sur profil, la récolte continue de données, ... constituent autant d'exemple de mesures pseudo directes de paramètres. Dans ce cas, il convient de s'assurer de la bonne adéquation entre ce qui est mesuré et ce qui est requis par la modèle ;
 - Soit le paramètre ne peut être mesuré directement. Si toutes les autres variables du formalisme considéré ont pu être mesurées, il est alors possible d'estimer le paramètre par un calcul analytique. Il s'agit ici d'une forme d'optimisation paramétrique sur une sous-routine du modèle ;

- Par analogie avec d'autres cas. La recherche bibliographique ou la mesure expérimentale sont ici utilisées à titre indirect.
- Par optimisation. Dans certains cas, la valeur du paramètre ne peut être obtenue par les méthodes citées précédemment. C'est notamment le cas si le paramètre a une valeur purement statistique ou s'il est très difficile de mettre en place des expérimentations pour le mesurer. De tels paramètres sont alors déterminés par optimisation. Une gamme de valeurs possibles du paramètre est examinée et la valeur qui minimise la différence entre les simulations et les mesures réalisées sur le terrain est retenue.

Le choix des paramètres à optimiser peut par ailleurs être réalisé à partir de l'analyse de sensibilité de STICS réalisée par Ruget *et al.* (2002). Une fois les paramètres identifiés, diverses techniques peuvent être mises en place pour procéder à leur optimisation directe (méthode fréquentiste) ou à l'échantillonnage de leur valeur la plus probable (méthode Bayésienne). Ces notions seront discutées en détail dans le chapitre correspondant (Chapitre 4). Au vu du grand nombre de paramètres du modèle, ainsi que du nombre de paramètres recalibrés, ceux-ci ne seront pas explicités en détail dans cet ouvrage. Certains sont toutefois décrits dans les différents articles faisant partie de cette thèse. Nous renvoyons à ceux-ci, ainsi qu'à l'ouvrage de référence (Brisson *et al.*, 2008) pour un descriptif complet des paramètres du modèle.

Il importe ensuite de traiter de l'utilisation faite des différentes données mesurées vis-à-vis du processus de paramétrisation du modèle. Le Tableau 3 synthétise toutes les données mesurées ainsi que l'emploi qui en est fait. Comme on peut le constater, dans la mesure où elles servent à piloter les simulations journalières, seules les données météorologiques ne sont pas employées pour paramétrer le modèle (initialisation ou calibration).

Tableau 3 : Classement des données mesurées selon leur origine et leur utilisation dans le modèle.

Classe de donnée	Variable mesurée	Utilisation
Météorologique	Température (min. et max.)	Variable de forçage
Météorologique	Pression de vapeur	Variable de forçage
Météorologique	Radiation solaire	Variable de forçage
Météorologique	Pluviométrie	Variable de forçage
Météorologique	Anémométrie	Variable de forçage
Environnementale	Température du sol	Calibration/validation du modèle
Environnementale	Potentiel hydrique	Calibration/validation du modèle
		Initialisation du modèle
Environnementale	Teneur en eau du sol	Calibration/validation du modèle
		Initialisation du modèle
Environnementale	Teneur en N (NO_3^- et NH_4^+) du sol	Calibration/validation du modèle
		Initialisation du modèle
Environnementale	Description du profil de sol	Calibration/validation du modèle
		Initialisation du modèle
Agronomique	Opérations culturales (semis, fertil., ...)	Initialisation du modèle
Agronomique	LAI	Calibration/validation du modèle
Agronomique	Biomasse totale	Calibration/validation du modèle
Agronomique	Rendement grain	Calibration/validation du modèle
Agronomique	Teneur en protéines du grain	Calibration/validation du modèle

Finalement, il convient d'apporter quelques précisions sur le choix des variables de sorties pour procéder à l'optimisation paramétrique et aux calculs des critères de qualité du modèle. Guillaume *et al.* (2011) ont démontré qu'une procédure de calibration en un seul temps était à favoriser en comparaison d'une approche de calibration itérative. Considérer simultanément tous les paramètres à optimiser et toutes les variables de sorties permettrait d'éviter les phénomènes de sur-paramétrisation

pouvant être rencontrés dans le cas de procédures itératives. Selon cette dernière manière de procéder, de plus petits sets de paramètres sont optimisés successivement, au gré des différents formalismes dans lesquels ils sont impliqués et des variables de sorties correspondantes sélectionnées. Toutefois, une paramétrisation figée d'un formalisme donné tendrait à limiter les possibilités de calibration des formalismes situés en aval de celui-ci (*e.g.* l'élaboration du rendement, dépendant de la biomasse simulée, elle même fonction du LAI), ce qui, en comparaison d'une procédure de calibration en un seul temps, aboutit à une qualité prédictive moins bonne du modèle sur ces derniers.

2.3. *OptimiSTICS, vers l'emploi de librairie*

Comme le souligne Justes *et al.* (2009), du fait de la complexité liée à l'emploi de modèles de cultures, il est important de disposer d'outils pour automatiser les simulations de tels modèle, analyser leur comportement, évaluer les résultats et estimer les paramètres. C'est cette raison qui a incité les auteurs à développer l'outil *OptimiSTICS*, qui rend possible l'automatisation de diverses procédures de STICS, telles que l'optimisation successive de plusieurs paramètres, la simulation de multiples années culturales, ..., ainsi que la comparaison des résultats du modèle avec des valeurs observées. Toutes ces opérations sont très laborieuses et consommatrices de temps lorsqu'elles sont réalisées au cas par cas. Entièrement écrit sous le langage de programmation MATLAB®, la source principale d'innovation de cet outil par rapport à la version *WinSTICS* (interface utilisateur de STICS) réside dans l'emploi de librairie : *MultiSimLib*, *SensiLib*, et *OptiLib*.

Dans le cadre de ce doctorat, il est ainsi à noter que nous sommes repartis des codes *d'OptimiSTICS* pour développer nos propres codes d'échantillonnage de paramètres et de multi-simulations.

2.4. *Evaluation du modèle*

Comme le souligne Brisson *et al.* (2002), le défi principal est d'obtenir un modèle suffisamment robuste, *i.e.* qui, sous des scénarios climatiques et des types de sols bien contrastés, présente la capacité de simuler les sorties du modèle sans erreur systématique.

Bien qu'il existe une importante littérature en liaison avec les modèles mathématiques, relativement peu de choses ont été écrites sur les procédures d'évaluation de la performance des modèles (Loague and Green, 1991). Un modèle n'est une bonne représentation de la réalité que s'il peut être utilisé pour prédire un phénomène observable avec suffisamment d'exactitude à l'intérieur d'une gamme de valeurs pour laquelle il est calibré et validé.

Une première évaluation des performances d'un modèle consiste simplement à comparer les valeurs statistiques élémentaires (moyenne et écart-type) des données mesurées et simulées. Ensuite, vient la comparaison, au moyen d'un test statistique, de ces moyennes. Les performances d'un modèle pourront être jugées acceptables s'il n'est pas possible de rejeter l'hypothèse de non-différence des valeurs observées et prédites. En utilisant ce type de test, deux types d'erreurs peuvent se produire :

- Erreur de type I : C'est l'erreur portant sur le développement du modèle, elle correspond au rejet d'une hypothèse vraie.
- Erreur de type II : C'est l'erreur portant sur l'utilisation du modèle, et elle correspond à l'acceptation d'une hypothèse fausse.

Finalement, un troisième type d'évaluation des performances est souvent considéré. L'analyse

des erreurs résiduelles entre les valeurs simulées et observées peut ainsi être employée. Il est à noter que ces critères peuvent être calculés pour s'assurer d'une bonne calibration du modèle, mais aussi d'une validation correcte de celui-ci. Plusieurs mesures sont disponibles dans la littérature; seules les principales sont ici décrites. Il s'agit de l'erreur maximale (ME), du coefficient de détermination (CD), de l'erreur quadratique moyenne (RMSE), de la déviation normalisée (ND) et de l'efficacité du modèle (EF) (Loague and Green, 1991, Brisson *et al.*, 2002; Beaudouin *et al.*, 2008). Leurs équations sont présentées ci après.

$$ME = \text{Max} |Y_i - \hat{Y}_i|_{i=1}^n \quad (\text{Eq. 1})$$

$$CD = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2} \quad (\text{Eq. 2})$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (\text{Eq. 3})$$

$$ND = \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{Y}_i}{\sum_{i=1}^n Y_i} \quad (\text{Eq. 4})$$

$$EF = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (\text{Eq. 5})$$

où n est le nombre de mesures utilisées pour calculer le critère (il s'agit ici des moyennes des réplicats), \hat{Y}_i est la valeur simulée, Y_i est la valeur observée, \bar{Y} est la moyenne des valeurs observées.

La limite inférieure pour les critères ME, RMSE et CD est zéro. La valeur maximale pour EF est 1. Aussi bien EF que ND peuvent être négatifs. Si EF est proche de zéro ou négatif, alors il est préférable d'employer les valeurs moyennes des observations que celles prédites par le modèle. Au final, l'utilisateur d'un modèle s'attendra à avoir des valeurs de ces critères aussi proches que possible de celles présentées dans le tableau suivant (Tableau 4).

Ces valeurs n'étant d'application que dans le cas du modèle parfait, Beaudouin *et al.* (2008) définissent des bornes pour la déviation normalisée et l'efficacité du modèle à l'intérieur desquelles la calibration des paramètres d'un modèle pourra être jugée adéquate. Ils considèrent qu'un set de paramètres donnés fournit un modèle calibré/validé si, les valeurs de ND et EF calculées sur les variables d'intérêt satisfont au double critère :

$$|ND| \leq 0.1 \quad (\text{Eq. 6})$$

$$EF \geq 0.5 \quad (\text{Eq. 7})$$

Si ces conditions ne sont pas remplies, un autre set de paramètres ou d'autres valeurs des paramètres sélectionnés doivent être choisis. Dans l'hypothèse où ces conditions ne sont jamais remplies, on retiendra la combinaison et les valeurs de paramètres optimisant ces critères.

Tableau 4 : Valeurs attendues des paramètres d'évaluation d'un modèle, dans le cas du modèle parfait.

Paramètre	Valeur optimale	Valeur attendue
ME	0	-
CD	1	-
RMSE	0	minimum
ND	0	< 0.1
EF	1	> 0.5

En complément à l'étude du coefficient de détermination CD, les paramètres de la régression linéaire entre valeurs observées et simulées peuvent être employés comme critères de comparaison. Ces paramètres sont la pente a , et l'ordonnée interceptée b de la régression. Afin de terminer l'étude sur les critères d'évaluation des modèles, il faut savoir qu'il est possible de décomposer l'étude du RMSE entre d'une part l'erreur quadratique moyenne systématique $RMSE_s$, et l'erreur quadratique moyenne non-systématique $RMSE_u$, comme l'ont fait Houlès *et al.* (2004). Cette décomposition permet de faire la distinction entre la composante systématique de l'erreur due à un biais du modèle, et une erreur aléatoire, expliquée par exemple par la variabilité génétique :

$$RMSE_s = \sqrt{\frac{1}{n} \sum_{i=1}^n (b + a.Y_i - Y_i)^2} \quad (\text{Eq. 8})$$

$$RMSE_u = \sqrt{\frac{1}{n} \sum_{i=1}^n (b + a.Y_i - \hat{Y}_i)^2} \quad (\text{Eq. 9})$$

Il est à noter que cette distinction repose sur l'hypothèse d'un modèle de régression linéaire entre les valeurs simulées et observées. La distinction entre $RMSE_s$ et $RMSE_u$ ne pourra être faite que lorsque cette hypothèse sera validée.

Etat de l'art

*"The illiterate of the 21st century
will not be those who cannot read and write,
but those who cannot learn, unlearn, and relearn"*

Alvin Toffler

Chapitre 3 : Etat de l'art et synthèse bibliographique

1. Objectif

Si elle est conceptuellement simple, l'utilisation des modèles peut se révéler parfois très complexe. Le présent article synthétise la structure mathématique générale des modèles de cultures et les principaux critères permettant de juger de leur qualité en phase de calibration ou de validation. Par ailleurs, à des fins d'amélioration des capacités prédictives des modèles, un aperçu général des différentes méthodes d'estimation paramétrique fréquentielle et Bayésienne, ainsi que des techniques d'assimilation de données est également fourni et discuté.

Il est à noter que cet article ne porte pas sur tous les sujets développés dans la présente étude. Les notions relatives aux incertitudes liées à la variabilité climatique, la gestion de la fertilisation azotée et la mise au point de systèmes d'aide à la décision seront abordés dans les chapitres correspondants. L'article qui suit se veut avant tout être général et introductif sur l'emploi des modèles, tout en apportant au lecteur les considérations nécessaires portant sur l'évaluation et l'amélioration de la qualité de ceux-ci, en prélude à leur emploi dans la mise au point de systèmes opérationnels.

2. Publication associée

Dumont, B., Vancutsem, F., Seutin, B., Bodson, B., Destain, J.-P., Destain, M.-F., 2012 Simulation de la croissance du blé à l'aide de modèles écophysologiques : Synthèse bibliographique des méthodes, potentialités et limitations. *Biotechnologie, Agronomie, Société et Environnement - BASE*, 16(3), 382-392.

Voir publication associée :
<http://hdl.handle/2268/129589>

Estimation paramétrique

*"It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts."*

*Sir Arthur Conan Doyle,
British mystery author & physician (1859 - 1930).
The Adventures of Sherlock Holmes,
A Scandal in Bohemia.*

Chapitre 4 : Estimation paramétrique

1. Objectif

Une paramétrisation correcte et robuste du modèle s'avère primordiale et est sous-jacente à la mise au point d'un outil fonctionnel de prédiction des rendements ou, plus généralement, d'aide à la décision. Différentes techniques d'optimisation paramétrique existent et reposent pour la plupart sur la minimisation d'un critère objectif et explicite quantifiant les différences entre les valeurs observées et simulées.

Or, bien souvent, on considère que les paramètres d'un modèle, étant des valeurs soit fixées par l'utilisateur, soit issues d'un processus d'estimation, conduisent à des erreurs indépendantes sur les résultats. Cependant, selon Makowski and Wallach (2002), cette hypothèse n'est pas réaliste. Ils affirment qu'il n'est pas raisonnable de faire l'hypothèse d'indépendance des erreurs de modélisation pour les mêmes combinaisons site-année. Ils estiment que, d'une manière générale, l'estimation des paramètres d'un modèle fait l'objet de trop peu d'attention.

Ainsi, les méthodes d'optimisation paramétrique dites fréquentistes et les méthodes d'échantillonnage de paramètres basés sur le théorème de Bayes et utilisant une simple somme de carrés des écarts comme fonction objectif ne dérogent pas à cette règle. Celles-ci considèrent des résidus entre mesures et simulations qui sont indépendants et distribués de façon Gaussienne.

Dans ce chapitre, diverses hypothèses sur le modèle statistique représentatif des résidus seront étudiées. La première hypothèse faite sur l'erreur sera de la considérer comme indépendante et Gaussienne pour les combinaisons site-année. Nous étudierons ensuite l'impact de la distribution des erreurs, en considérant une erreur nominale sur chaque mesure telle que mesurée en champ, et toujours en considérant l'indépendance de celle-ci. Finalement afin de traiter à la fois le cas d'indépendance (hétéroscédasticité observée) et de distribution des résidus réduite mais non-normalisée ($\sigma^2 \neq 1$), nous proposons d'inclure dans la fonction de vraisemblance une valeur explicite du coefficient de variation, représentatif de la variabilité naturelle observée, incluant la variabilité génétique naturelle, la variabilité imputable à l'hétérogénéité de la parcelle, et dans une moindre proportion, l'erreur sur la mesure. Par une analyse *a posteriori* des résidus nous verrons dans quelle mesure ces hypothèses restent valables.

Cette section vise également à mettre en exergue la remarque émise en début de dissertation selon laquelle les mesures de terrain sont d'une importance capitale. Pouvoir en extraire l'information d'intérêt, permettant d'obtenir un modèle robuste et générique, s'avère capital et primordial à l'emploi du modèle pour ses capacités prédictives.

2. Contribution originale à l'estimation des paramètres de STICS

2.1. La sélection des paramètres à estimer

Comme nous l'avons énoncé précédemment, les paramètres les plus influents sur la réponse du modèle peuvent être déterminés par une analyse de sensibilité. Mais une fois ces paramètres identifiés, il convient de tenir compte d'autres impératifs lors du passage au processus de sélection finale du set

de paramètres à estimer et de la détermination du set final de paramètres estimés.

Il est important de ne pas oublier, lorsque l'on procède à la ré-estimation de paramètres, qu'une hypothèse implicite consiste à considérer que les autres paramètres sont connus, à leur valeur optimale, ce qui n'est évidemment pas toujours le cas. Ainsi, dans les modèles complexes, les paramètres sont en général organisés en sous-groupe faisant référence à différents formalismes. Mais bien souvent ces paramètres dépendent les uns des autres au travers des équations du modèle. Dès lors, une légère erreur sur les paramètres non estimés entraînera inévitablement un biais sur l'estimation des paramètres sélectionnés pour être estimés. Ce biais est appelé biais d'*omission* (Miller, 2002).

Par ailleurs, indépendamment de la méthode d'estimation choisie, et pour des raisons de temps de calcul ou de manque de données, il est impossible d'estimer simultanément tous les paramètres d'un modèle, et ce en vertu d'un manque d'*identifiabilité*, principalement dû à la structure des équations du modèle (Makowski *et al.*, 2006). Un manque d'identifiabilité apparaît lorsque plusieurs valeurs de paramètres aboutissent aux mêmes valeurs des variables observées, par exemple comme dans le cas d'une équation où deux paramètres apparaissent l'un au dénominateur et l'autre au numérateur. Il est alors difficile d'estimer correctement ces paramètres à partir d'un seul type d'observations.

Finalement, comme nous le verrons dans le chapitre suivant, lorsque la méthode d'optimisation n'explore pas l'entièreté de l'espace des paramètres, et/ou présente une forte propension à converger vers des minima locaux, éloignés de la solution réelle, et que dès lors on accorde trop de confiance dans la paramétrisation réalisée, on se trouve dans un cas de *sur-paramétrisation* (dans ce cas le terme anglais d'*over-confidence* serait plus approprié). Cette sur-paramétrisation se répercutera inévitablement sur l'emploi du modèle dans un cadre prévisionnel.

2.2. Les méthodes d'estimation et d'échantillonnage des paramètres

Lorsque le set de paramètres que l'on veut réévaluer a été déterminé, il convient d'en procéder à l'estimation par inversion du modèle. A cette fin, il existe un large panel de méthodes pouvant être regroupées en deux principales familles : l'approche fréquentiste et l'approche Bayésienne.

Ainsi, dans le domaine de la modélisation, deux types d'informations sont disponibles pour estimer la valeur d'un paramètre : (i) les mesures collectées via l'expérimentation, et (ii) la connaissance dite *a priori* que l'on possède sur les paramètres, bornant la gamme de valeurs pouvant être prises par ce paramètre (bibliographie, limites physiques, connaissance d'experts, ...). La mise en œuvre des méthodes fréquentistes ne nécessite que l'information provenant du jeu d'observations, alors que l'approche Bayésienne utilise les deux sources d'informations.

Les méthodes fréquentistes considèrent les paramètres du modèle comme des variables fixées à une certaine valeur, dont il convient d'estimer celles minimisant les erreurs entre les données issues du modèle et les mesures. L'application d'une méthode fréquentiste permet alors de déterminer une valeur particulière de chaque paramètre, appelée *estimateur* du paramètre. Cependant, les méthodes fréquentistes sont sujettes aux problèmes majeurs, que sont la divergence des estimateurs de la solution optimale (minima locaux) et son impact sur le phénomène de sur-estimation.

A côté de cela, les approches dites Bayésiennes (Eq. 10) permettent de prendre en compte non seulement les mesures de terrain, mais aussi l'information *a priori* des paramètres. Les paramètres du modèle sont ici considérés comme des variables stochastiques, définies par une distribution *a priori*.

L'intérêt des méthodes Bayésiennes consiste alors à estimer une nouvelle densité de probabilité pour la distribution, dite densité *a posteriori*, des valeurs des paramètres. Le passage de la densité *a posteriori* à partir de l'information de la densité *a priori* se fait en utilisant le théorème de Bayes, et en définissant la fonction de vraisemblance la plus à même de représenter les différences observées entre mesures et simulations.

$$\pi(\theta|Y) = \frac{\pi(Y|\theta)\pi(\theta)}{\pi(Y)} \quad (\text{Eq. 10})$$

Dans cette expression θ and Y représentent respectivement le vecteur des paramètres et des mesures, et $\pi(\theta|Y)$ représente la fonction de densité de probabilité (pdf) des paramètres sous les conditions des valeurs des données observées. Cette pdf constitue la probabilité *a posteriori* des paramètres après échantillonnage. $\pi(\theta)$ et la pdf dite *a priori* des paramètres (considéré le plus souvent uniforme lors de l'implémentation numérique des algorithmes génétiques). Elle correspond à la connaissance initiale que nous avons des paramètres, et des bornes dans lesquelles il peut varier (littérature, connaissance de l'expert, ...). Et $\pi(Y|\theta)$ représente ce que l'on appelle la fonction de vraisemblance, qui constitue une mesure de la similitude entre les simulations et les mesures.

Cette nouvelle densité de probabilité contient alors les valeurs des paramètres les plus à même de fournir la bonne réponse sur les simulations du modèle. Pour cette raison, ces algorithmes sont davantage considérés comme échantillonneurs que comme optimiseurs de paramètres. L'information contenue dans la nouvelle distribution peut être utilisée à diverses fins, comme estimer la valeur la plus probable de chaque paramètre (ici on parlera de la notion de meilleur estimateur), calculer l'incertitude associée à l'estimation des paramètres, ou encore celle associée aux simulations.

Les approches Bayésiennes sont devenues de plus en plus populaires et utilisées ces dernières années pour échantillonner les valeurs de paramètres de modèles complexes suite aux décuplements des vitesses de calcul des CPU et le développement de nouveaux algorithmes.

Vrugt et ses collaborateurs (Vrugt *et al.*, 2008a,b, 2009a,b, 2010) ont ainsi récemment développé un algorithme novateur basé sur une approche Bayésienne formelle, de type MCMC (Markov chain Monte Carlo). Cette méthode permet de scruter efficacement l'espace paramétrique multi-dimensionnel afin de déterminer les distributions des probabilités *a posteriori* des paramètres du modèle. Cette méthode est appelée DREAM (DiffeRential Evolution Adaptive Metropolis).

Dans le cadre de cette thèse, l'algorithme a été couplé avec succès à l'outil OptimisTICS, afin d'utiliser les capacités de multi-simulations de ce dernier, et d'ainsi tirer parti au mieux de l'outil DREAM, sans doute l'un des algorithmes MCMC les plus aboutis à l'heure actuelle. La suite de ce chapitre est dédiée, sous forme de publication, à la présentation de la recherche faisant état de ce couplage.

2.3. La qualité des mesures en complément aux méthodes d'échantillonnages

L'emploi des méthodes Bayésiennes, outre le fait de reposer sur le théorème de Bayes, nécessite la définition de la fonction de vraisemblance. Cette dernière fait non seulement le lien entre les mesures et les simulations, mais permet également de tenir compte du modèle statistique représentatif des erreurs entre ces deux grandeurs, autrement appelées résidus. Depuis quelques années maintenant, un débat a lieu sur l'emploi d'approches formelles ou informelles dans la détermination de la fonction

de vraisemblance, et plus particulièrement sur la représentation des erreurs (Beven *et al.*, 2008; Schoups and Vrugt, 2010; Vrugt *et al.* 2008b, 2009b).

Les fonctions de vraisemblance dites informelles reposent sur une représentation pragmatique de l'incertitude associée à la présence de résidus complexes. Offrant la possibilité de gérer l'échantillonnage de paramètres là où les fonctions de vraisemblance théoriques simples ne peuvent pas être appliquées, les algorithmes issus de la famille *Importance Sampling* sont devenus de plus en plus populaire (Beven and Binley, 1992; Beven, 2008; Vrugt *et al.*, 2009b). Cependant, cette approche implique de discrétiser l'espace paramétrique, et de ce fait, peut aboutir à une représentation imparfaite de la distribution *a posteriori* lorsque les paramètres à échantillonner sont trop nombreux (Makowski *et al.*, 2002a,b).

A côté de cela, les approches formelles reposent sur la détermination d'un modèle statistique représentatif des résidus. Ce modèle déterminé *a priori* est ensuite employé dans la fonction de vraisemblance. Couplée aux simulations MCMC, cette représentation de l'erreur permet de séparer les solutions dites '*comportementales*', *i.e.* celles améliorant statistiquement parlant les simulations, de celles qui ne le sont pas.

Typiquement, les erreurs peuvent être classées en trois catégories, selon qu'elles se rapportent à (i) la variance des résidus, (ii) à la forme de leur distribution, et (iii) à la corrélation qu'elles peuvent avoir. L'avantage offert par les approches formelles réside dans la capacité de pouvoir *vérifier a posteriori* la validité des hypothèses faites sur les erreurs (e.g., Schoups and Vrugt, 2010).

L'hypothèse la plus simple que l'on puisse poser sur les résidus consiste à les considérer comme indépendants et identiquement distribués, selon une normale de moyenne zéro et de variance constante. Dans ce cas, le modèle statistique de représentation des erreurs se résume à une somme de carré des écarts, couramment défini par l'acronyme SLS (Sum of Least Square) (Box and Tiao, 1973). Cependant, dans bien des cas, spécialement en recherche agronomique, les erreurs sont corrélées pour une même combinaison site-année, non stationnaire (hétéroscédasticité, impliquant une croissance de l'erreur avec une croissance de la mesure), et non Gaussienne. Il est possible de considérer la valeur nominale de l'écart-type sur chaque mesure pour pondérer le modèle d'erreur, aboutissant au modèle de résidus WLS (Weighted Least Square).

Dans ce chapitre, nous nous proposons d'étudier l'effet de différents modèles statistiques de représentation des résidus. En particulier, une fonction de vraisemblance faisant un emploi explicite du coefficient de variation (CV) est proposée (Eq. 11), afin de tenir compte de l'hétéroscédasticité des mesures, le dispositif expérimental mis en place ayant d'ores-et-déjà permis de réduire les problèmes de corrélation et de distribution des mesures.

$$\pi_{\log}(Y|\theta) = -\frac{n}{2} \cdot \ln(2\pi \cdot CV^2) - \sum_{i=1}^n \ln(y_i) - \frac{1}{2} \sum_{i=1}^n \left(\frac{\left(\frac{\hat{y}_i(\theta, X)}{y_i} - 1 \right)}{CV} \right)^2 \quad (\text{Eq. 11})$$

où n représente le nombre d'observations, $\hat{y}_i(\theta, X) - y_i$ sont les résidus et CV correspond au coefficient de variation, ratio entre σ^2_i , qui est la variance nominale sur la mesure et y_i qui est la mesure.

3. Discussions intermédiaires

3.1. Échantillonnage paramétrique

Il est avant tout important de bien insister sur le fait que DREAM est un algorithme d'échantillonnage paramétrique. Les résultats obtenus correspondent à la distribution dite *a posteriori* du/des paramètre(s), indiquant à l'utilisateur les valeurs les plus probables du/des paramètre(s). Il ne s'agit pas à proprement parler d'une optimisation paramétrique. Ce n'est qu'une fois le processus d'échantillonnage abouti que l'optimisation aura lieu, via la sélection de la valeur la plus représentative de la distribution paramétrique (*e.g.* estimateur moyen).

L'algorithme DREAM synthétise différentes innovations, et se base sur différents critères statistiques (*e.g.* critère de Metropolis), qui lui confèrent des propriétés assurant une exploration maximale de l'espace pluri-paramétrique et une convergence optimale vers les solutions comportementales. L'algorithme DREAM n'est pas limitatif sur le nombre de paramètres pouvant être échantillonnés en même temps, et s'est montré performant sur des problèmes dont la dimensionnalité était importante. Par ailleurs, DREAM est également à même de s'adapter à la non-linéarité des modèles.

Les propriétés de l'algorithme lui permettent de maintenir l'*ergodicité* et la *balance détaillée*. Ces deux propriétés font références au processus d'échantillonnage et d'évolution des chaînes MCMC (Markov Chain Monte Carlo). La première, l'*ergodicité*, consiste à admettre que l'information d'un ensemble de réalisations (ici, les chaînes) apporte au cours du temps la même information qu'un signal aléatoire. La seconde, appelée *balance détaillée*, assure que la probabilité d'évoluer vers une région de l'espace paramétrique est la même dans toutes les directions, incluant la probabilité de revenir au point de départ une fois l'évolution vers une position donnée réalisée. Toutes deux veillent à maintenir un niveau d'exploration de l'espace pluri-paramétrique qui soit le plus aléatoire possible, afin de n'omettre aucune solution possible et d'éviter la convergence vers les minima locaux.

Ces propriétés permettent de faire face aux contraintes liées à la modélisation de phénomènes agronomiques, ce qui a justifié le couplage de DREAM au modèle STICS. Par ailleurs, les sorties offertes par l'algorithme offrent plusieurs avantages majeurs dans l'analyse des simulations agronomiques, décrits dans les sections suivantes.

3.2. Analyse des distributions *a posteriori*

A titre d'illustration, la Figure 9, extraite de la publication associée à ce chapitre, présente un exemple de résultats obtenus après échantillonnage de l'espace paramétrique au moyen de l'outil DREAM. Les résultats ici présentés correspondent au terme d'un processus d'échantillonnage réalisé avec une somme des carrés des écarts pondérés par les variances nominales de chacune des mesures (WLS). Les distributions *a posteriori* de quatre paramètres du modèle sont ici présentés, respectivement, *efcroiveg*, l'efficacité de conversion du rayonnement en phase végétative, *efcroijuv*, l'efficacité de conversion du rayonnement en phase juvénile, *stamflax*, la somme de degré-jours nécessaire pour passer du stade où la vitesse de développement foliaire est maximale (*IAMF*) à celui où le développement foliaire maximal est atteint (*ILAX*), et *psisto*, le potentiel hydrique auquel les stomates se ferment.

Au terme du processus d'échantillonnage, la première source d'information est obtenue via

l'analyse visuelle de ces graphiques. Dans cet exercice, quatre comportements différents ont pu être observés. Le comportement unimodal (*efcroiveg*) se caractérise par un mode bien marqué, et une distribution normale autour de la valeur la plus probable du paramètre. Le paramètre *efcroijuv* se caractérise par un comportement unimodal avec dissymétrie gauche, traduisant une asymétrie de la réponse du modèle par rapport au paramètre : les plus faibles valeurs du paramètre n'ont que peu de chance d'améliorer les simulations, alors que des valeurs plus élevées aboutissent à des réponses plausibles. Le comportement à tendance bimodale (*stamflax*) se traduit par l'existence de différentes valeurs possibles pouvant être prises par le paramètre, et aboutissant à des simulations proches. Finalement, l'absence de mode marqué (*psisto*) aboutit à une distribution *a posteriori* très proche de la distribution *a priori* uniforme.

L'analyse visuelle de ces graphiques nous offre un premier aperçu de la qualité de l'échantillonnage et de la pertinence des paramètres sélectionnés pour échantillonnage. En effet, plus le mode est marqué, plus le paramètre a une influence sur la réponse du modèle. A l'opposé, plus la distribution *a posteriori* est proche du *prior*, moins le paramètre a d'effet sur la réponse. Cela traduit un paramètre pour lequel le modèle peut ne pas être sensible de façon générale, ou un paramètre pour lequel les conditions de simulations n'en ont pas permis l'expression. Dans le cas du paramètre *psisto*, on peut aisément envisager qu'en Belgique les conditions de stress hydriques sont limitées à des épisodes de stress discrets. De ce fait, le paramètre ne s'exprime que peu souvent.

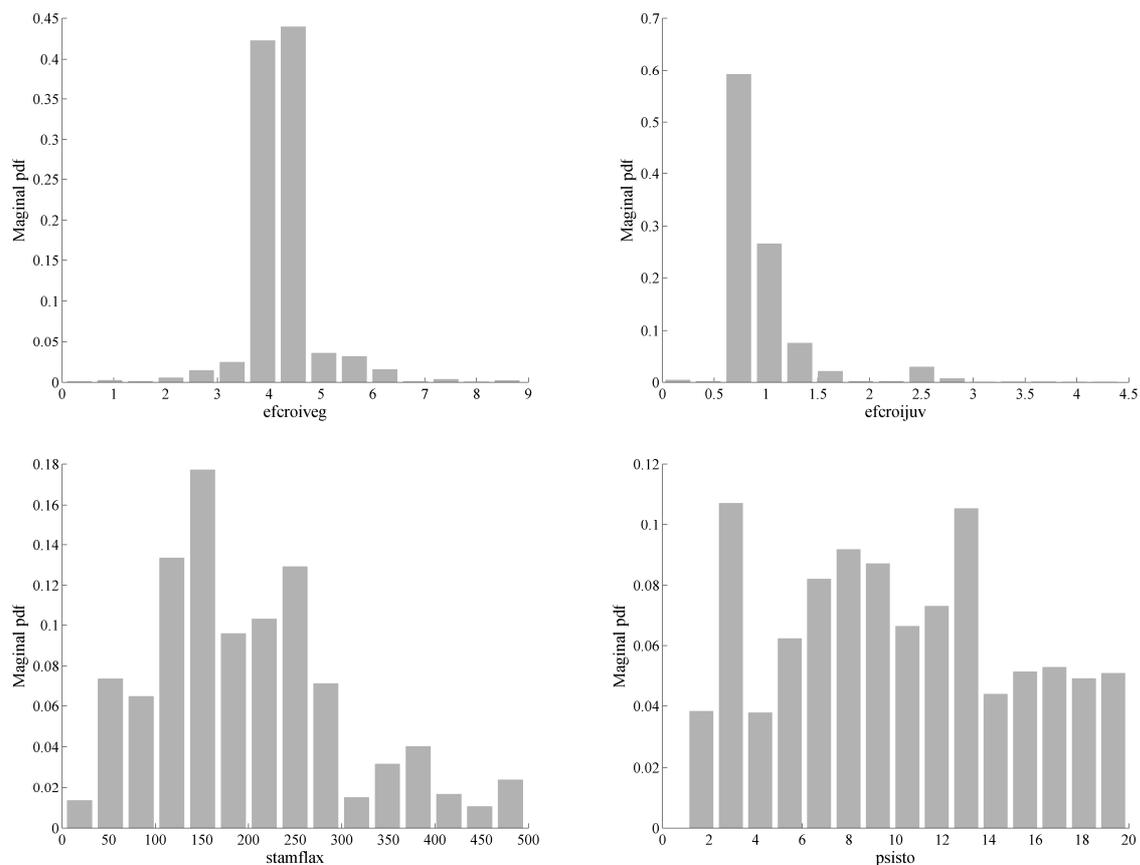


Figure 9: Fonction de densité de probabilité marginale (marginal pdf) pour quatre paramètres du modèle STICS : *efcroiveg*, *efcroijuv*, *stamflax* et *psisto*. Les histogrammes correspondent aux résultats obtenus au terme du processus d'échantillonnage faisant emploi d'une fonction de vraisemblance type somme des carrés des écarts pondérés (WLS).

Finalement, l'outil DREAM est un algorithme dit *génétique*, pour lequel les paramètres sont échantillonnés et évoluent comme une population (sélection, croisement, ...). Lors du processus, les paramètres sont échantillonnés en sets, ce qui implique que la réponse du modèle tient compte de la corrélation entre ces paramètres. Ainsi un des avantages offerts par l'algorithme réside dans l'étude possible de la corrélation paramétrique. Plus de détails à ce sujet peuvent être trouvés dans la publication associée à ce chapitre.

3.3. Incertitude sur les sorties du modèle

Afin de procéder à l'optimisation paramétrique du modèle, il convient de sélectionner une valeur unique qui soit la plus représentative de chaque paramètre. Une technique classique est de calculer l'estimateur moyen de chacune des distributions. Une autre approche consiste à sélectionner dans la population de sets paramétriques, le set aboutissant à la plus faible RMSE entre valeurs simulées et observées. Ces deux approches sont les plus couramment employées. L'intérêt offert est de n'avoir à traiter qu'un seul set de paramètres, ce qui s'avère particulièrement pertinent pour diminuer les temps de simulation, idéal lors de la mise au point d'outils d'aide à la décision.

A l'opposé, l'un des avantages offert par DREAM réside dans la possibilité d'étudier les incertitudes sur les simulations issues du processus d'échantillonnage. En effet, les sorties de l'algorithme sont sous la forme d'une population paramétrique, où les paramètres ont été traités en sets, corrélés donc. Or, il vient d'être démontré que tous les paramètres ne répondent pas à la même densité de fonction. Une approche intéressante consiste ainsi à étudier l'entière des solutions paramétriques d'un nombre donné d'éléments (sets paramétriques) issus des différentes chaînes et obtenus au terme du processus d'échantillonnage. On obtient une distribution de simulations, dont on peut étudier les valeurs particulières, entre autre la moyenne ou certains percentiles donnés (*e.g.* percentiles 2.5, 50 et 95%, *i.e.* médiane et intervalle de confiance à 95%).

A titre illustratif, la Figure 10 présente différents cas de simulations, respectivement obtenus au terme du processus d'échantillonnage avec une fonction de vraisemblance type SLS, et trois fonctions de vraisemblance avec des valeurs croissantes de CV, 0.05, 0.147 et 0.50 (Eq. 11). Les graphiques correspondants à la SLS et à la valeur CV de 0.147 sont issus de l'article associé à ce chapitre. Les deux autres graphiques ont été générés pour cette section de l'étude, sur base du même set paramétrique échantillonné, mais en modifiant la valeur du CV.

Bien que faiblement perceptible sur la Figure 10, il semble que la fonction de vraisemblance type SLS ne soit pas à même de tenir compte de la réalité physique des paramètres échantillonnés pour la phase de croissance en stade juvénile. En effet, une sous-estimation des conditions de croissances en sortie d'hiver (du semis au jour Julian 470) peut être observée, compensée plus tard par une surestimation de la dynamique de croissance. Le cas présenté ici est typiquement un cas de sur-paramétrisation (ou d'*over-confidence*'). Or, il est important de simuler correctement la croissance de la culture en sortie d'hiver. Cet aspect est d'autant plus important et compréhensible dans le cas d'études exploratoires (*e.g.* étude de l'impact du changement climatique) ou dans le cadre de la mise au point d'un outil opérationnel, fonctionnant sur des données acquises en temps réel.

A côté de cela, la formule proposée (Eq. 11) conduit à des valeurs de paramètres qui soient physiquement réalistes. Dans chacun des cas, les simulations en début de phase végétative sont correctes et présentent des intervalles de confiance proches des simulations réalisées au moyen de

l'estimateur paramétrique moyen (contrairement au cas SLS). La suite de la courbe de croissance (au-delà du jour Julian 525) semble quant à elle sous-estimée. La remise en question à laquelle il convient ici de procéder ne relève pas de la qualité de l'échantillonnage ou du modèle, mais davantage du choix des paramètres, ainsi que des situations culturales sélectionnées pour la calibration (cfr. publication associée). Dans notre cas, l'échantillonnage paramétrique a été réalisé simultanément sur des situations agronomiques où 0 et 180 kgN.ha⁻¹ ont été fertilisés.

Finalement, on peut constater que les estimateurs moyens conduisent à peu près aux mêmes simulations pour différentes valeurs de CV, mais que l'intervalle de confiance associé aux simulations est quant à lui croissant avec la valeur du coefficient de variation.

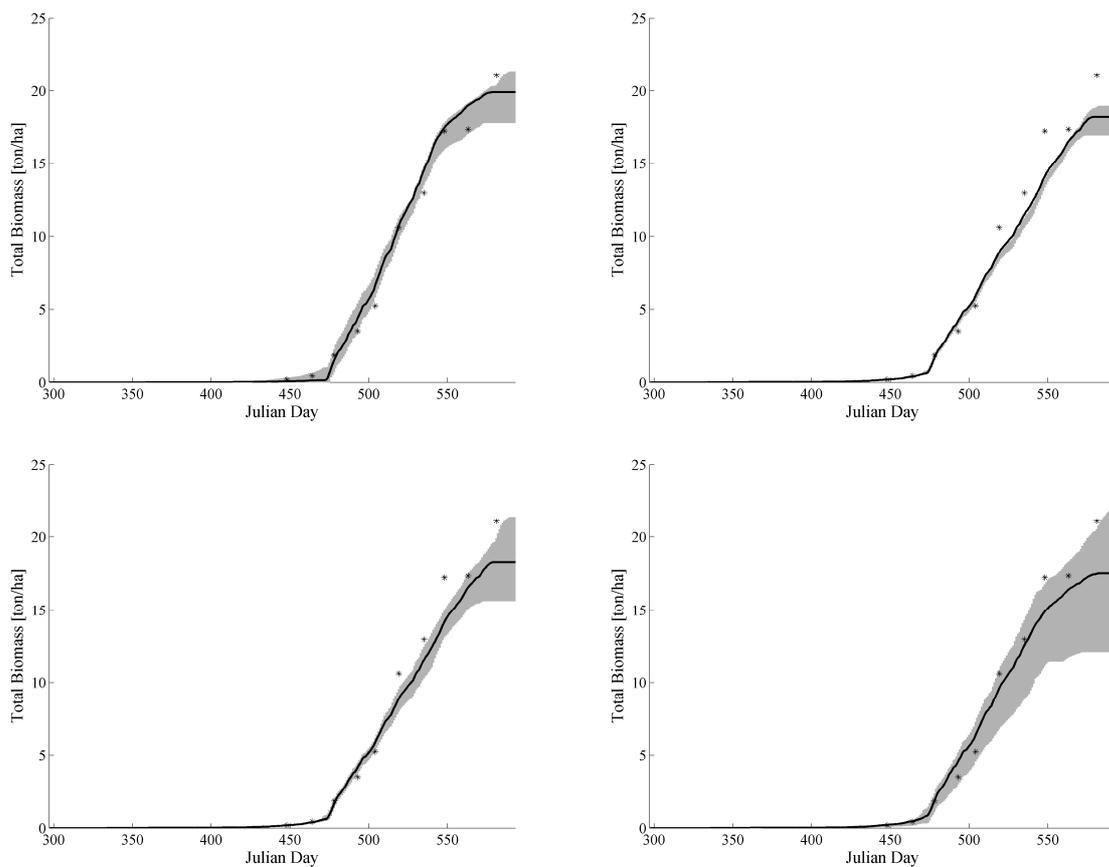


Figure 10: Simulations obtenus après échantillonnage paramétrique sur la situation culturale de l'année 2008-2009 (points noirs). Le trait noir représente la simulation conduite avec l'estimateur paramétrique moyen, obtenu respectivement (de gauche à droite puis de haut en bas) avec les fonction de vraisemblance SLS, et CV (0.05 - 0.0147 - 0.50). Les niveaux d'incertitude (à 95%) associés à l'estimation paramétrique sur les simulations sont représentés par l'aire grise.

3.4. Qualité globale des simulations

Finalement, la question se pose de savoir quel est l'impact des différents choix sur la qualité globale des simulations. En effet, d'une part, il est possible de procéder soit au choix d'un estimateur moyen, soit à la sélection de celui minimisant la RMSE, ou encore de calculer la moyenne d'une distribution de simulations correspondant à une population paramétrique. D'autre part, comme on vient

de le voir, le choix du modèle représentatif des résidus (cas SLS, WLS ou CV) peut aboutir à des simulations différant par les processus simulés.

Ainsi, à titre illustratif, la Figure 11 présente trois critères d'évaluation de la qualité des simulations en comparaison des observations, à savoir la RMSE, l'efficacité de modèle (EF) et la déviation normalisée (ND). Comme au chapitre précédent, les résultats correspondants à la SLS, à la WLS et à la valeur CV de 0.147 sont issus de l'article associé à ce chapitre. Les autres grandeurs ont été calculées pour cette section de l'étude, sur base des mêmes sets paramétriques échantillonnés, mais en modifiant la valeur du CV. Par ailleurs, ces résultats illustrent les différences entre les critères calculés sur base des simulations obtenues avec l'estimateur moyen, ou via la moyenne des simulations de la population paramétrique.

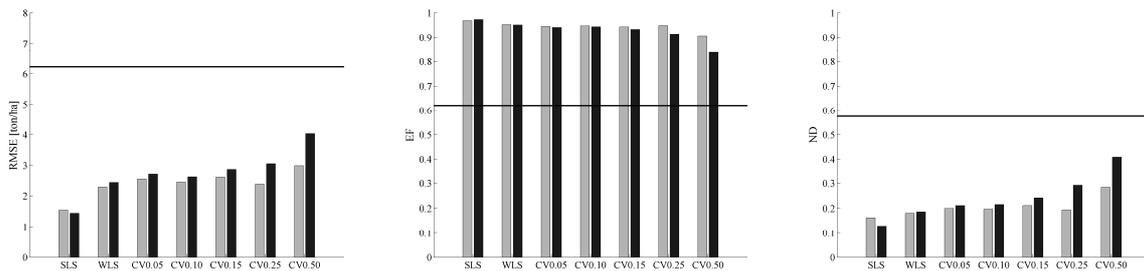


Figure 11: RMSE's EF's et ND's obtenus au-terme des différents processus d'échantillonnage (SLS - WLS - CV) en comparaison de la simulation sur le jeu de paramètre initial (trait horizontal noir). L'évaluation du modèle peut-être conduite sur les simulations réalisées au moyen de l'estimateur paramétrique moyen (histogrammes gris) ou sur la médiane des simulations ((histogrammes noirs).

En comparaison à la situation de référence, il apparaît tout d'abord que l'échantillonnage des paramètres au moyen de l'algorithme DREAM est pertinent. Quelle que soit la fonction de vraisemblance employée, dans tous les cas, les simulations sont améliorées. La RMSE est au moins divisée par deux, l'EF jouxte la valeur optimale de 1, et la déviation normalisée est proche du critère d'acceptation de 10%. Au vu des situations culturelles contrastées sur lesquelles les paramètres ont été optimisés, ces résultats ont été jugés plus que satisfaisants.

Les meilleurs critères sont systématiquement obtenus avec la fonction de vraisemblance SLS. Mais rappelons que celle-ci aboutit à des valeurs de paramètres et des simulations non réalistes en phase juvénile de croissance. Les simulations obtenues avec une fonction de type WLS ou CV permettent d'obtenir un modèle global qui soit de qualité. Toutefois, lorsque la valeur du CV croît trop fortement, la qualité des simulations se dégrade. De plus, il semble que l'estimateur paramétrique moyen permette d'aboutir à des simulations dont les qualités globales sont meilleures que celles obtenues avec la moyenne d'une population de simulations.

4. Conclusions intermédiaires

Dans un cadre idéal, où les mesures seraient parfaites, le coefficient de variation devrait être à l'image exclusive de la variabilité génétique naturelle de la culture. Cependant, la prise d'échantillons en champs et la réalisation de mesures sont entachées d'erreurs liées à l'utilisateur, à l'instrument de mesure et à l'hétérogénéité des parcelles, dont il est important de tenir compte.

De façon plus générale, cette section de la thèse s'est concentrée sur la mise au point d'un outil permettant d'étudier l'impact de différentes sources d'incertitudes sur le processus d'échantillonnage seul. En effet, différentes questions se posent lors de cette étape cruciale préalable à la mise au point d'outils d'aide à la décision. En particulier l'outil composé du couplage de STICS et de DREAM, associé à la nouvelle fonction de vraisemblance proposée, a permis d'aborder deux sources d'erreurs : *i)* faut-il prendre en compte les erreurs sur les mesures en champs lors de l'échantillonnage paramétrique; *ii)* a-t-on intérêt à utiliser un set unique d'estimateurs paramétriques (*e.g.* estimateur moyen) ou à utiliser la distribution des simulations correspondant à une population paramétrique.

Il est ainsi apparu primordial d'établir correctement le modèle des résidus. Considérer uniquement la moyenne des réplicats, sans tenir compte de leur variabilité (écart-type ou CV), peut aboutir à de mauvais estimateurs paramétriques. En outre, la qualité des mesures est également primordiale. Si les mesures sont de mauvaise qualité, ou si l'hétérogénéité du milieu induit des mesures dont la variabilité est trop grande, le processus d'échantillonnage peut aboutir à un modèle incorrectement paramétré.

La nouvelle fonction de vraisemblance proposée permet ainsi à la connaissance de l'expert de s'exprimer non seulement au niveau de la définition de la connaissance *a priori* sur les paramètres, mais aussi lors de chaque boucle du processus d'échantillonnage, et ce via la prise en compte du CV dans la fonction de vraisemblance calculée à chaque itération du processus d'échantillonnage.

5. Publication associée

Dumont B., Leemans V., Mansouri M., Bodson B., Destain J.-P., Destain M.-F., 2014. Parameter optimisation of the STICS crop model, with an accelerated formal MCMC approach. *Environmental Modelling and Software*, 52, 121-135.

Voir publication associée :
<http://hdl.handle/2268/160225>

**Voir publication associée :
<http://hdl.handle/2268/160225>**

Voir publication associée :
<http://hdl.handle/2268/160225>

**Voir publication associée :
<http://hdl.handle/2268/160225>**

Voir publication associée :
<http://hdl.handle/2268/160225>

**Voir publication associée :
<http://hdl.handle/2268/160225>**

Voir publication associée :
<http://hdl.handle/2268/160225>

Voir publication associée :
<http://hdl.handle/2268/160225>

Voir publication associée :
<http://hdl.handle/2268/160225>

**Voir publication associée :
<http://hdl.handle/2268/160225>**

Voir publication associée :
<http://hdl.handle/2268/160225>

Voir publication associée :
<http://hdl.handle/2268/160225>

Assimilation de données

"L'erreur n'annule pas la valeur de l'effort accompli"

Pensée populaire.

Chapitre 5 : Assimilation de données

1. Objectif

D'une manière générale, la modélisation se heurte à différents problèmes aboutissant à des erreurs de prédictions qui peuvent parfois s'avérer importantes. Par exemple, une estimation paramétrique imprécise, des erreurs d'estimation des valeurs initiales des variables d'état ou une mesure imprécise des variables d'entrées peuvent se traduire par une qualité prédictive médiocre. Par ailleurs, les mesures disponibles pour étalonner le modèle sont parfois peu nombreuses, alors que la variabilité spatiale et climatique peut être relativement importante. Le degré d'exactitude du modèle peut de ce fait être très difficile à estimer.

Par ailleurs, les modèles sont en général, du moins pour une partie des formalismes sur lesquels ils reposent, une simplification parfois grossière des phénomènes réellement mis en œuvre. Ainsi, il arrive qu'une partie parfois importante de l'information contenue dans les mesures, celle qui ne peut être exactement modélisée, soit alors considérée comme un simple bruit.

Pour ces raisons, en parallèle ou en complément à l'estimation paramétrique, une autre famille d'approches, reprise sous l'appellation générale de "*méthode de filtrage*" rencontre un certain succès dans la problématique de l'estimation des variables d'état, mais aussi dans l'estimation conjointe des variables d'état et des paramètres. Ce type d'approche consiste à mettre à jour récursivement la distribution *a posteriori* des variables d'état modélisées non-encore observées, sur base d'une séquence de données observées. La notion de "récursivité" est ici évoquée car la base de données des mesures est en général croissante au cours de la saison, l'intérêt des méthodes de filtrages étant d'optimiser les variables d'état et les paramètres à chaque fois qu'une nouvelle observation est réalisée.

Dans ce chapitre, nous nous proposons d'étudier une méthode d'assimilation de données ayant récemment vu le jour et connue sous le nom de "*filtre variationnel*". Principalement appliquée jusqu'à présent appliquée pour le suivi et la localisation d'objets en mouvement, cette méthode s'avère plus efficace que celles qui l'ont précédée.

Sans entrer dans le descriptif général de toutes les méthodes de filtrage, dont il existe énormément de variantes, signalons que le filtre de Kalman, de par la solution Bayésienne optimale qu'il offre, est souvent employé comme base de comparaison lors du test de nouvelles méthodes, alors qu'il doit être réservé aux problèmes linéaires. La publication associée à ce chapitre vise ainsi à démontrer l'avantage de la méthode du filtre variationnel, en le comparant au filtre de Kalman Ensemble et au filtre de Kalman Étendu (optimisé par rapport au filtre original pour l'implémentation sur modèle différentiable).

2. Contribution originale à l'assimilation de données

2.1. Le filtre variationnel

Le problème de l'estimation des variables d'état (et en parallèle des paramètres) a été approché selon différentes méthodes. L'une d'elle particulièrement connue est celle du filtre de Kalman (Kalman, 1960), qui offre une solution Bayésienne optimale. La méthode est efficace d'un point de vue

informatique, mais elle s'avère limitée par le caractère non universel des hypothèses sur lesquelles elle repose.

Le filtre de Kalman Ensemble (EnKF) et le filtre de Kalman Étendu (EKF) (Julier and Uhlmann, 1997) ont été proposés comme alternative au filtre original, afin d'en augmenter la flexibilité. L'utilisation de ces filtres requiert que le modèle, et plus particulièrement les variables d'états simulées, soient différentiables. Cependant, lorsque le modèle est hautement non-linéaire, ces filtres montrent une performance limitée.

Pour surmonter cet inconvénient, une approche non-paramétrique, basée sur l'échantillonnage selon une méthode de Monte Carlo séquentielle, appelée le filtre particulaire (PF), a récemment vu le jour (Arulampalam *et al.*, 2002). Contrairement aux méthodes de Monte Carlo par chaînes de Markov (MCMC) qui se basent sur les densités de probabilités 'jointes' (et sont donc qualifiées de hors-ligne), les méthodes particulières 'en-ligne' estiment la distribution *a posteriori* dites *de filtrage*, c'est-à-dire en se basant seulement sur les données observées et connu à un instant donné, via un set d'échantillons pondérés (les 'particules'). Le filtre particulaire offre l'avantage de s'adapter plus facilement à l'échantillonnage paramétrique et à celui de variables de modèles dynamiques non-linéaires et multimodaux. En contrepartie, ce filtre est plus couteux en temps de simulation, plus gourmand en stockage d'informations, et nécessite un grand nombre de 'particules' pour obtenir un résultat optimal.

Récemment, un filtre variationnel (VF) a été proposé (Mansouri *et al.*, 2009) pour résoudre les problèmes non-linéaires d'estimation paramétrique. Contrairement au filtre particulaire qui repose sur l'emploi d'un grand nombre de particules, la dépendance temporelle du filtre variationnel est réduite à une chaîne simple. Le choix optimal de la distribution d'échantillonnage se fait en minimisant le critère de divergence de Kullback-Leibler.

2.2. Enonciation du problème

Les méthodes d'assimilation de données nécessitent la définition des variables sur lesquelles portent le filtrage, s'agissant ici des sorties du modèle pour lesquelles une amélioration de la qualité prédictive est recherchée. Il convient pour cela de définir le modèle d'évolution différentielle des variables pour lesquelles l'on dispose de mesures en cours de saison.

A titre illustratif, un système de trois variables d'état est ici proposé, à savoir l'indice de développement foliaire (LAI) et les teneurs en eau (HUR) des deux premiers horizons :

$$\begin{cases} LAI(t) = [g_1(\theta)]\Delta t + LAI_{t-1} + w_t^1 \\ HUR1(t) = [g_2(\theta)]\Delta t + HUR1_{t-1} + w_t^2 \\ HUR2(t) = [g_3(\theta)]\Delta t + HUR2_{t-1} + w_t^3 \end{cases} \quad (\text{Eq. 12})$$

Expression dans laquelle $g_{j,j \in \{1,2,3\}}$ fait référence au modèle ou au formalisme simulant la variable d'intérêt basée sur le vecteur de paramètres θ ; t et Δt sont respectivement le temps et le pas de temps de simulations; et $w_{j,j \in \{1,2,3\}}^j$ est le bruit Gaussien sur la mesure de moyenne zéro et de variance connue.

L'application des techniques de filtrage a un tel système repose sur l'hypothèse que les paramètres sont fixés ou ont été déterminés précédemment. Cependant, les modèles incluent souvent des paramètres qui ne sont pas en général exactement connus ou qui nécessitent une réestimation. Ainsi, les méthodes d'assimilation de données peuvent être étendues à l'optimisation de paramètres

afin de simplifier la tâche de modélisation conventionnelle. Il convient alors d'étendre le système sur lequel s'applique le filtre, selon :

$$\begin{cases} LAI(t) = [g_1(\theta_{t-1})]\Delta t + LAI_{t-1} + w_t^1 \\ HUR1(t) = [g_2(\theta_{t-1})]\Delta t + HUR1_{t-1} + w_t^2 \\ HUR2(t) = [g_3(\theta_{t-1})]\Delta t + HUR2_{t-1} + w_t^3 \\ ADENS_t = ADENS_{t-1} + \gamma_t^1 \\ DLAIMAX_t = DLAIMAX_{t-1} + \gamma_t^2 \\ PSISTURG_t = PSISTURG_{t-1} + \gamma_t^3 \end{cases} \quad (\text{Eq. 13})$$

Dans cette expression, *adens*, *dlaimax* et *psisturg* sont trois paramètres du modèle que l'on cherche à réévaluer et $\gamma^j, j \in \{1,2,3\}$ est un bruit Gaussien sur le processus de moyenne zéro et de variance connue. *Adens* est un paramètre jouant sur le phénomène de compétition lié à la densité de la culture, *dlaimax* est un paramètre contrôlant la vitesse de la croissance foliaire et *psisturg* est un paramètre de sensibilité de la plante au stress hydrique influençant directement la croissance et indirectement le prélèvement d'eau dans le sol.

3. Discussions intermédiaires

3.1. Enonciation du problème et opérationnalité du modèle

De plus amples détails sur les variables d'état et paramètres sélectionnés dans le cadre de ces exemples, ainsi que sur le modèle dont ils sont originaires pourront être trouvés dans la publication associée à ce chapitre et dans les références mentionnées en fin de chapitre. Dans cette section de l'ouvrage, les discussions s'axeront sur les deux éléments que sont la sélection du filtre et l'énonciation du problème.

Comme présenté ci-avant, l'emploi de techniques d'assimilation de données requiert la définition préalable du système à optimiser. On distinguera deux familles de systèmes, ceux ne visant qu'à filtrer les variables d'états du système (Eq. 12) et ceux visant à filtrer un système mixte, composé de variables d'état et de paramètres (Eq. 13). Le choix de l'un ou l'autre système correspondra à deux stratégies différentes.

Dans le premier cas, on ne cherchera à améliorer le modèle qu'au niveau de ses sorties. Dans cette optique, il sera possible de pallier des conditions de croissance ou de rendre compte d'un phénomène que le modèle n'est pas à même de traiter. A titre d'exemple, la chute du rendement due à une maladie ou une verse constitue bien souvent un phénomène pour lequel l'assimilation de données trouve toute son utilité. Il est important de signaler que dans ce cas l'utilisateur accorde toute sa confiance d'une part au paramétrage et d'autre part aux formalismes du modèle. Mais, il reste conscient des limites de ceux-ci, et du fait que le modèle ne puisse rendre compte de l'entièreté des processus à l'œuvre dans le monde réel.

La réestimation des paramètres conjointement au filtrage des sorties du modèle répond à une conception différente du problème. Tout d'abord, il conviendra à l'utilisateur de répondre aux deux questions suivantes : *i)* Peut-on avoir confiance dans le paramétrage ? *ii)* Peut-on avoir confiance dans le modèle ?

On peut aisément imaginer des situations où l'utilisateur a confiance dans son modèle, mais pas dans son paramétrage. Ainsi, si l'utilisateur a toute confiance en son paramétrage plante, le déploiement d'un modèle tel que STICS à l'échelle d'un bassin versant soulèvera toutefois la nécessité d'un paramétrage correct du type de sol. Or, la réalisation de telles mesures s'avère très coûteux à la fois en temps et en argent. La mise au point d'un outil fonctionnel de prédiction des rendements à une telle échelle pourrait donc être réalisée en couplant le modèle à des techniques d'assimilation de données. En parallèle au filtrage des sorties du modèle, on envisagera une réestimation des paramètres liés au sol, afin, d'une part, d'affiner la réponse spatialisée du modèle, et d'autre part, d'affiner notre connaissance des paramètres du sol.

Le passage à une échelle encore plus vaste, par exemple celle d'une région, d'un pays ou d'un continent, induira chez le modélisateur des doutes sur le paramétrage mais aussi sur le modèle employé et la pertinence des formalismes utilisés pour modéliser les systèmes de culture à de tels niveaux. Mais avant tout, la mise au point d'un outil fonctionnel à de telles échelles renvoie au compromis fait sur le choix du modèle, en termes d'opérationnalité et de réalisme. Dans ce cas de figure, le modélisateur pourrait avoir intérêt à utiliser des modèles/formalismes plus simples, tout en sachant que les variables d'état et/ou les paramètres les plus importants seront filtrés sur des échelles spatiales importantes, afin de rendre compte de toute la diversité des systèmes de culture rencontrés (type de sol, variabilité climatique, cultivars différents, ...).

3.2. Avantages offerts par le filtre variationnel

A titre d'exemple extrait de la publication associée à ce chapitre, le Tableau 5 synthétise les performances d'estimations des filtres EKF et VF en terme de RMSE entre valeurs simulées et observées pour les deux systèmes définis précédemment (Eq. 12 et Eq. 13). Dans le cas d'une réestimation en parallèle des paramètres du système (seconde partie du Tableau 5), l'estimateur paramétrique moyen au terme de la simulation est présenté si la convergence sur le paramètre a atteint un régime stationnaire.

Tableau 5 : Présentation des RMSE entre valeurs simulées et observées (LAI, HUR1, HUR2) et, le cas échéant, de la moyenne de la valeur des paramètres en régime établi final, pour l'application des filtres EKF et VF sur deux problèmes différents (Eq. 12 et 13). NC = Non convergence du paramètre

Techniques	RMSE			Paramètre moyen		
	LAI [m ² /m ²]	HUR1 [%]	HUR2 [%]	<i>adens</i>	<i>dlaimax</i>	<i>psisturg</i>
EKF	0.063	0.060	0.030	-0.8	NC	NC
VF	0.019	0.019	0.012	-0.8	0.0078	4

Pour obtenir ces résultats, des simulations ont été conduites avec le modèle [mini-]STICS (*cfr.* appendices de la publication) sur les données climatiques de la saison de culture 2008-2009. Pour les simulations correspondantes, on consultera la publication associée à ce chapitre. Toutefois, il est à signaler que la fréquence d'échantillonnage pour l'application du filtre correspond à une mesure par jour. Par ailleurs, ces mesures sont générées au départ d'une simulation conduite sur la saison de

culture, auquel un bruit blanc gaussien, de moyenne nulle et d'écart-type égal à 0.1, est adjoint. C'est sur base de ces données que l'efficacité des différents filtres est comparée.

Comme on peut le voir, les RMSEs relatives aux différentes variables d'état du système sont considérablement plus faibles lors de l'emploi du filtre VF par rapport au filtre EKF. Par ailleurs, le filtre VF est le seul capable d'estimer les trois paramètres en plus d'améliorer les simulations sur les variables d'état. Finalement, on notera que les RMSEs sont plus faibles, et ce quelle que soit la variable d'état considérée, lorsque les paramètres du modèle sont ré-estimés en parallèle à l'application du filtre sur les sorties du modèle.

3.3. Limites liées à l'approche proposée

Ce volet des recherches constitue pour l'instant une voie exploratoire visant à mettre en évidence le meilleur filtre à appliquer en parallèle à l'emploi d'un modèle de culture. Il est évident que le nouveau filtre variationnel proposé surpasse les autres filtres classiquement utilisés en agronomie.

Il conviendrait désormais de poursuivre les recherches en se focalisant sur ce filtre. En guise de discussions complémentaires aux recherches proposées en dernière partie de ce chapitre, différentes considérations sont émises dans cette section de l'étude.

Tout d'abord, on peut se poser la question de la pertinence de la fréquence proposée. Des mesures disponibles à l'échelle journalière semblent en effet peu réalistes, particulièrement pour les variables d'état relatives à la croissance de la culture (*e.g.* LAI et biomasse). En effet pour ces variables, les mesures de références sont réalisées la plupart du temps manuellement, ce qui explique leur espacement dans le temps. Toutefois, dans un futur proche, avec l'avènement des drones en agriculture de précision, on pourrait imaginer avoir des fréquences d'échantillonnage plus importantes pour des mesures indirectes du LAI, sans toutefois avoir accès à des mesures quotidiennes. Finalement, signalons que cette hypothèse est toutefois réaliste pour la teneur en eau du sol, dont des mesures en continue peuvent être rendues disponibles via un système de monitoring environnemental. Il conviendrait d'orienter le futur de ces recherches vers une étude de la fréquence minimale de données requise pour assurer la convergence du filtre variationnel.

Le second point qu'il convient de discuter concerne la précision requise sur les données. En effet, dans les recherches conduites jusqu'à présent, un bruit blanc est adjoint aux simulations conduites sur les différentes entrées climatiques. Ce bruit a jusqu'à présent été considéré gaussien et uniforme. Or, si cette hypothèse est à nouveau réaliste sur les mesures des variables environnementales (*e.g.* teneur en eau du sol), comme il a été discuté au chapitre précédent, les phénomènes de croissance sont quant à eux caractérisés par une hétéroscédasticité affectant les mesures. Il conviendrait dans le futur de ces recherches de considérer ce phénomène.

Finalement, ces recherches exploratoires ont été conduites sur des sets paramétriques prédéfinis. Seules les entrées climatiques enregistrées au cours de deux saisons de culture (2008-09 et 2009-10) ont pour l'instant été utilisées en entrée du modèle. Il conviendrait désormais de compléter l'étude en testant le filtre variationnel sur un nombre plus important d'années de simulations. Il serait également pertinent de confronter les simulations aux mesures de terrain réalisées dans le cadre de l'expérimentation. Cela permettrait de pallier au manque de réalisme des mesures employées discuté ci-avant.

4. Conclusions intermédiaires

Les techniques d'assimilation de données peuvent se révéler des adjuvants particulièrement utiles pour améliorer les simulations d'un modèle, et ré-estimer au besoin certains paramètres du modèle, surtout dans le cadre de la mise au point d'un outil opérationnel à large échelle. Les techniques les plus récentes de filtrage, faisant emploi d'estimateurs Bayésiens, semblent relativement efficaces dans l'accomplissement de cette tâche. En particulier, le filtre variationnel offre des performances qui surpassent les autres filtres étudiés.

Toutefois, il est important de signaler que, dans la mesure du possible, les méthodes d'assimilation de données ne devraient intervenir qu'après un processus estimation paramétrique complet. En effet, seul un processus d'échantillonnage (cfr Chapitre 4) ou d'estimation paramétrique statistiquement fiable permet de calibrer le modèle de façon robuste et de s'assurer de la confiance que l'on peut avoir dans les processus modélisés. Cette condition est primordiale pour employer un modèle pour ses capacités prédictives.

5. Publications associées

Mansouri M., **Dumont B.**, Destain M.-F., 2012. Modeling and prediction of nonlinear environmental system using Bayesian methods. Computers and Electronics in Agriculture, 92, 16-31. <http://hdl.handle.net/2268/137399>

Mansouri M., **Dumont B.**, Leemans, V., Destain M.-F., 2014. Bayesian methods for predicting LAI and soil water content. Precision agriculture 15(2), 184-201. Doi 10.007/s11119-013-9332-7. <http://hdl.handle.net/2268/156796>

Mansouri M., **Dumont B.**, Destain M.-F., 2013. Modeling and Prediction of Time-Varying Environmental Data Using Advanced Bayesian Methods. Exploring Innovative and Successful Applications of Soft Computing, 112-137. IGI Global editions. <http://hdl.handle.net/2268/146909>

Voir publication associée :
<http://hdl.handle/2268/137399>

***Prédiction des rendements :
Méthodologie pour la mise au
point d'un outil fonctionnel***

"What one man can invent another can discover."

*Sir Arthur Conan Doyle,
British mystery author & physician (1859 - 1930).
The Return of Sherlock Holmes,
The Adventure of the Dancing Men.*

Chapitre 6 : Prédiction des rendements : Méthodologie pour la mise au point d'un outil fonctionnel.

1. Objectif

A l'échelle de la parcelle, les modèles agro-environnementaux sont des outils très puissants pour évaluer l'effet des variables d'entrée des systèmes de culture, telle que les pratiques culturales, les conditions agro-environnementales et les effets du climat sur les récoltes. Si de nombreuses études ont déjà été consacrées à l'analyse *a posteriori* des diverses variables d'entrées, très peu cependant ont été vouées au développement d'outils physiquement et/ou statistiquement basés à des fins réellement prévisionnelles. Ainsi, développer une méthodologie capable de prédire les rendements, à l'échelle locale ou régionale, serait d'un intérêt majeur.

A l'opposé des méthodes purement statistiques de l'USDA (Allen *et al.*, 1994), le rendement en grain d'une culture peut être prédit au départ de modèles physiquement basés, tels que les modèles agro-environnementaux. Dans le cadre d'une telle approche, les données climatiques futures non-observées restent les éléments clés d'une bonne prédiction. Même si la précision sur les prévisions climatiques est de plus en plus grande, celles-ci restent limitées dans le temps. Impossible de dire au-delà de 15 jours quel sera exactement le climat à venir. Il convient alors de poser des hypothèses sur l'avenir, autrement dit prédire l'imprévisible.

Ainsi, des simulations basées sur des données climatiques moyennes (Semenov and Porter, 1995) ou des données climatiques générées à l'aide d'un générateur stochastique, *e.g.* LARS-WG Weather Generator (Lawless and Semenov, 2005) sont deux voies possibles dans l'optique de pallier au manque de données futures. Alors que la seconde approche a le désavantage d'être lourde en temps de calcul et qu'il a été démontré que le générateur stochastique d'aléas climatiques pouvait être sous-représentatif de la réalité (Mavromatis and Hansen, 2001), la première approche repose quant à elle sur des hypothèses pouvant être inadaptées au comportement non-linéaire des modèles (Lawless and Semenov, 2005) et offrant un séquençage des combinaisons de variables climatiques inadéquat à une bonne représentation des phénomènes de stress (Riha *et al.*, 1996).

Dans une perspective de développement d'outil d'aide à la décision, le choix a été fait dans cette recherche d'opter pour une approche climat moyen, qui s'avérerait bien moins lourde en terme de temps de calcul. Par ailleurs, cette recherche vise également à mettre en évidence les limites, mais surtout les potentialités de réaliser des simulations sur base d'un climat moyen et la pertinence de celles-ci, en considérant que les conditions climatiques tempérées d'un climat tel que la Belgique induisent des stress, notamment hydriques, relativement limités en comparaison d'autres situations. Finalement, avant toute discussion sur les inconvénients liés à l'emploi d'un climat moyen, il semblait nécessaire de concevoir, d'optimiser et d'évaluer une démarche complète reposant sur cette hypothèse.

Dans ce chapitre, il sera ainsi exposé comment il est possible de mettre au point un outil de prédiction du potentiel de rendement restant à une culture.

2. Contribution originale à la prédiction des rendements en temps-réel

2.1. Base de données historique

La méthodologie qui va être présentée repose sur la disponibilité d'une base de données météorologiques représentative de la région à étudier. Dans cette recherche, la base de données météorologiques MDB (Meterological DataBase) est issue de la station d'Ernage qui fait partie du réseau de l'IRM (Institut Royal Météorologique Belge) et se compose des données climatiques enregistrées depuis 30 ans. La base de données et la station météorologique sont gérées et entretenues par le Centre Wallon de Recherches Agronomiques wallon (CRA-W), département Agriculture et Milieu Naturel.

La station d'Ernage est située à 4 km du champ expérimental. Les informations transmises par le réseau sans fil de micro-capteurs implémenté sur le site ont été comparées à celle de la MDB et jugées équivalentes, même si quelques différences ponctuelles peuvent apparaître principalement au niveau de la pluviométrie.

2.2. Calcul du climat moyen

Dans un premier temps, la MDB a été utilisée pour calculer les moyennes journalières de chacune des variables météorologiques nécessaires aux simulations du modèle. La MDB peut ainsi être vue comme une matrice de données dont la première dimension correspond au jour de l'année d , ce dernier variant de 1 à N_d , 1 et N_d pouvant indifféremment et respectivement représenter le premier et le dernier jour de l'année ou le jour de semis et de récolte ; la seconde dimension correspond à la variable climatique V étudiée (variant de 1 à N_{cv}), et la troisième dimension est à l'image de l'année de mesure, dont l'indice i varie de 1 à N_y , N_y étant la dernière année de mesure disponible (égale à 30 dans ce cas-ci).

Les valeurs moyennes de chaque variable $V_{mean_{cv}}$, celles-ci étant les températures minimale et maximale, la pression de vapeur, la vitesse du vent, l'intensité de la pluie et la radiation solaire, soit 6 variables dans ce cas-ci, sont calculées jour après jour comme la moyenne des N_y données disponibles pour un jour d donné (Eq. 12). Il est à noter à ce stade qu'une variable climatique est donc vue comme un vecteur issu de la matrice climatique.

$$V_{mean_{cv}}(d) = \frac{1}{N_y} \sum_{i=1}^{N_y} V_{i,cv}(d) \quad (\text{Eq. 14})$$

Une base de données climatiques moyennes est finalement créée (MBC - Mean Belgian Climate) et peut également se représenter sous la forme d'une nouvelle matrice de données, suite à la réagrégation des vecteurs moyens des variables climatiques. Etant propre à la base de données originelle, ici le climat tempéré enregistré à la station d'Ernage (Belgique), la MBC correspond dans ce cas à des conditions qui n'induiraient que peu de stress au niveau de la plante. En effet, un tel climat assurerait à la plante des conditions de croissance où un minimum de température ambiante serait assuré chaque jour, où de l'énergie solaire serait apportée en suffisance, et surtout où une quantité minimale de pluie quotidienne arroserait la culture, limitant le stress en eau de celle-ci. De telles conditions peuvent être considérées comme pseudo-optimales, ou encore comme des conditions de croissance presque non-limitatives.

2.3. Génération d'une matrice ensemble représentative du climat étudié

Dans un second temps, une "matrice-ensemble" ME est générée. Le point de départ de la ME, c'est-à-dire la première matrice de la ME, est la matrice MBC. En sélectionnant une matrice climatique réelle et en remplaçant les 10 premiers jours du MBC par les données réellement mesurées, on génère la seconde matrice de la ME. En considérant que les données sont remplacées à un taux donné (*e.g.* à une fréquence décadaire dans cette étude), dont l'incrément est représenté par j , il est possible de construire l'entièreté de la Matrice Ensemble.

Afin de mieux se représenter ce "*rafraichissement*" décadaire des matrices, l'exercice est tout d'abord réalisé pour une seule des variables climatiques. L'équation 15 et la Figure 12 représente le processus correspondant à la génération du vecteur ensemble au départ d'une variable climatique donné, cv :

$$VE_{i,cv}(d, j) = \begin{cases} Vmean_{cv}(d) & \text{if } d > 10 \times j \\ V_{i,cv} & \text{if } d \leq 10 \times j \end{cases} \quad (\text{Eq. 15})$$

La matrice ensemble correspondra quant à elle à l'agrégation de ces vecteurs ensembles. Au final, la première matrice représente un climat moyen pur, la dernière matrice représente un climat réel pur. Toutes les matrices intermédiaires correspondent à une combinaison de ces deux sets de données. Finalement, les simulations peuvent enfin être réalisées, avec l'entièreté de la matrice-ensemble comme entrée de STICS.

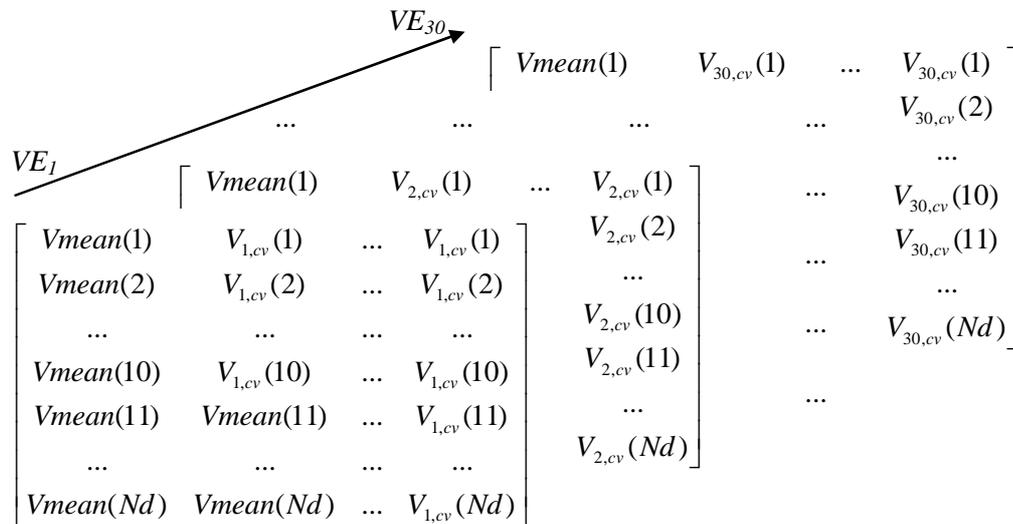


Figure 12: Représentation schématique d'une base de données climatiques ensemble (ici pour une variable climatique donnée cv). La dimension 1 à Nd représente le nombre de jours nécessaires à la simulation d'une saison de culture. L'index cv représente une variable climatique donnée. Les différentes années de la base de données sont contenues dans la troisième dimension ($VE1$ à $VE30$) et représentées sous l'indice i .

3. Discussions intermédiaires

3.1. Analyse des données de simulations d'une matrice ensemble

Dans le cadre de la mise au point d'outils d'aide à la décision, ce qui intéresse l'homme de science ou l'homme de terrain reste, *in fine*, la prédiction du rendement de fin de saison. Pour ce faire,

il convient d'avoir un bon aperçu du climat à venir. Et quelle meilleure représentation que celle offerte par la normale saisonnière.

Par ailleurs, le rendement final obtenu n'est finalement que la résultante des interactions de la plante avec son environnement tout au long de son cycle de croissance. De ce fait, dans un soucis de prendre en compte toutes les informations rendues disponibles au fur et à mesure que la saison s'écoule, il est possible d'affiner notre connaissance du rendement de fin de saison. Pour cela il convient simplement de procéder à un rafraichissement de notre connaissance de l'état instantané de la culture, qui n'est autre que la résultante des conditions de croissance passées qu'elle aura rencontrées.

Ces deux constats relativement simples justifient à eux seuls la mise au point de la Matrice Ensemble telle que proposée à la section précédente. A titre d'illustration, la Figure 13 (graphique de gauche), extraite de la publication associée à ce chapitre, montre les différentes simulations de biomasse obtenues avec comme entrée du modèle le MBC, le climat réel et toutes les situations intermédiaires. Le cas présenté ici correspond à la croissance d'une culture de blé sous les conditions d'une bonne pratique agricole, à savoir un amendement azoté de trois fois 60 kilogrammes d'azote appliqués par hectare, aux trois stades que sont le tallage, le redressement et la dernière feuille. La Figure 13 (graphique de droite) représente quant à elle les valeurs finales de biomasse simulée, en fonction du nombre de jours d'observation de climat réel depuis le semis.

La méthodologie repose donc sur le principe conceptuellement simple du rafraichissement décadaire des données climatiques. Plus on avance dans le temps, plus la proportion de données réellement mesurées augmente et plus les résultats de la simulation permettent de s'approcher du rendement qui sera réellement observé et simulé. Autrement dit, sur base de l'hypothèse d'un climat moyen en tant que climat hypothétique futur, chaque valeur du diagramme correspond au potentiel de rendement restant à la culture, c'est-à-dire le rendement qu'il serait possible d'obtenir si le climat futur venait à se rétablir en se rapprochant de la normale saisonnière.

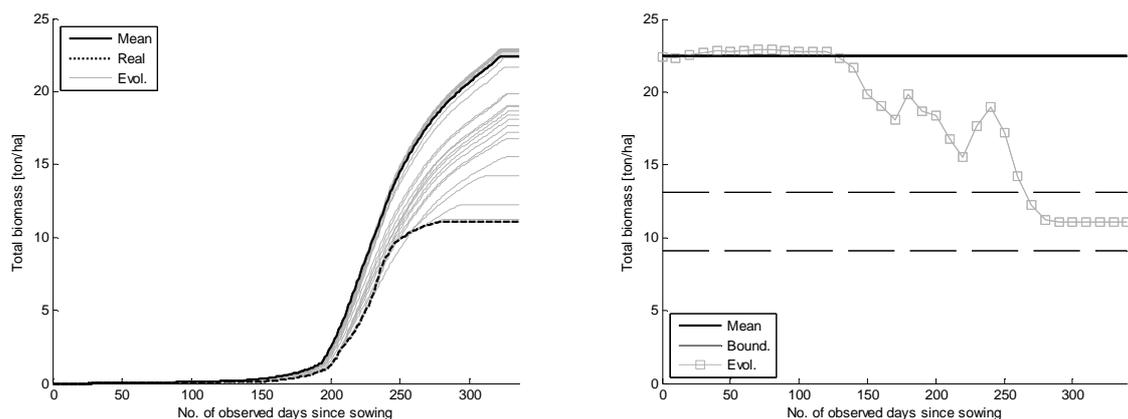


Figure 13: Simulation de la matière sèche totale au cours de la saison (gauche) et rendement total final correspondant, en fonction du nombre de jours de climat réel connu (droite). Rendements obtenus avec le climat moyen (trait noir continu) et avec le climat réel (trait noir pointillé) pour l'année 2005-06. Les représentations grises représentent toutes les combinaisons de ces deux climats (matrice ensemble), avec un taux de remplacement de 10 jours. Les traits en tirets noirs correspondent à une erreur de prédiction de 10 % du rendement final.

A titre d'illustration concrète des résultats offerts par une telle approche, des extraits du Livre

Blanc publié en février 2007 sont ajoutés à cette section de l'étude². Le parallèle peut ainsi être fait avec les prédictions des rendements finaux réalisées en cours de saison. Il s'agit ici de mettre en évidence comment les conditions climatiques instantanées influenceront le rendement final de la culture. Les observations météorologiques extraites du Livre Blanc seront donc ici comparées aux simulations de la Figure 13, graphique de droite, pour lequel le semis a été initié au jour Julien 295. A titre de référence, le 01 Janvier 2006 correspond ainsi au soixantième jour d'observation après le semis.

"La moyenne des températures automnales fut plus élevée que la normale saisonnière. [...]L'automne 2005 a été plus ensoleillé que la normale et fortement déficitaire en eau. [...]Les conditions d'implantations ont été particulièrement favorables. [...]" Les simulations du graphique de la Figure 13 sont clairement en accord avec ces observations. En effet, un bon niveau de rendement potentiel restant est simulé durant la période automnale, c'est-à-dire en première partie de graphique.

"D'un point de vue général, l'hiver fut légèrement plus frais, mais normalement ensoleillé. [...]Malgré un mois de Février normal, l'hiver 2005-2006 a présenté un important déficit hydrique. [...] Suite à un prolongement des conditions hivernales jusque la mi-Mars, la reprise de végétation a été tardive pour les froments." Le prolongement des conditions hivernales est correctement rendu au niveau des simulations, pour lesquelles le potentiel de rendement élevé est maintenu jusqu'au jour d'observation 120 (± 01 mars). Les conditions plus fraîches et le déficit hydrique n'ont encore que peu d'impact sur cette végétation qui attend des conditions plus favorables pour sa reprise.

" Un mois de Mars frais, suivi par un mois d'Avril normal, et précédant un mois de Mai doux ont doté le printemps 2006 d'une température moyenne à peine supérieure aux normales. [...] Le dernier mois de printemps 2006, particulièrement sombre, conféra à la saison un bilan de durée d'ensoleillement largement en deçà de la normale. [...] Les mesures révèlent un printemps particulièrement arrosé. [...]" Les conditions fraîches du mois de Mars ont initié la première phase de chute du potentiel de rendement restant. Les conditions normales d'avril ont permis de maintenir voire de relever ce potentiel (pic observé entre les jours 170 et 180 après semis). Cependant l'important déficit de lumière du mois de Mai aboutira à une seconde phase de chute du potentiel de rendement, qui pourra être noté entre les jours d'observations 180 à 220.

"L'été 2006 fut marqué par deux phases de température extrêmement élevées (deuxième décade de Juin, et Juillet dans son ensemble). [...] Les mois de Juin et Juillet ont bénéficié d'un excès d'insolation important, mais la saison s'est achevée sur un mois d'Août très sombre. [...] Une période de déficit hydrique importante est survenue en période estivale,(Juin et Juillet). Au mois d'Août les précipitations furent anormalement abondantes". La fin de printemps fortement pluvieuse, reconstituant le stock hydrique, les températures élevées et le niveau d'insolation important de Juin permettront à la culture de se rétablir (Jours d'observations 220 à 240) avec un potentiel de rendement qui se relancera. Toutefois le maintien des conditions de températures et d'ensoleillement favorables, couplé à une absence de précipitation feront chuter les rendements en Juillet. Le mois de Juillet très chaud réduira l'activité des deux dernières feuilles, accélérant le moment de la récolte. Le potentiel de

² Livre Blanc "Céréales" F.U.S.A.Gx et CRA-W, Gembloux - Février 2007 - Les extraits sont puisés principalement dans les Chap. 1. *Aperçu climatologique*, Chap. 2. *Implantation des cultures* et Chap. 3. *La fumure azotée*.

rendement n'évoluera plus en Août.

Comme on peut le constater, la dynamique du potentiel de rendement simulé suit correctement le déroulement de la saison observée, témoignant d'une bonne robustesse d'une part du modèle et d'autre part de l'approche prédictive. De plus, l'incertitude liée à la non-connaissance des conditions climatiques à venir a pu être réduite via leur remplacement par des conditions climatiques hypothétiques futures moyennes et la méthodologie de rafraichissement des données acquises en cours de saison.

3.2. Analyse sur des années multiples et prédiction du rendement

La présentation des résultats à la Figure 13 fournit ainsi les distributions de rendement au cours de la saison et à travers les années, ce qui nous permet (i) de détecter et (ii) de quantifier instantanément les pertes ou les gains de rendement dus aux conditions climatiques des 10 derniers jours en comparaison du potentiel restant à la culture (hypothèse d'un climat moyen en projection future). Mais ces résultats sont bien entendu propres à une année donnée. Afin de quantifier le niveau d'incertitude associé aux conditions climatiques d'une région donnée, il convient d'étendre la méthodologie.

Celle-ci a donc été appliquée aux trente années de la MBC. L'étude des rendements sur base des 30 ME générées permet de construire le graphique de gauche de la Figure 14. Ce graphique permet de mettre en évidence les périodes les plus sensibles pour la culture. Ainsi, dès la frontière des 200 jours passée depuis le semis, la culture semble plus sensible aux stress climatiques et les réductions de rendements sont les plus fortes. Cette date clef correspond en fait au stade de la dernière feuille et à la mise en place du rendement grain. Si une culture telle que le blé présente un fort potentiel de compensation sur son cycle de vie, une fois le nombre de grains mis en place par la plante (qui rappelons-le à une croissance déterminée), aucun retour en arrière n'est possible pour la culture. Le nombre de grains mis en place pour un cultivar donné est fonction (i) du développement atteint par la culture, donc de ses conditions de vie passée, et (ii) des conditions climatiques au stade de l'anthèse. La suite de l'élaboration du rendement sera donc uniquement fonction du remplissage des grains, dépendant lui-même des conditions climatiques ultérieures.

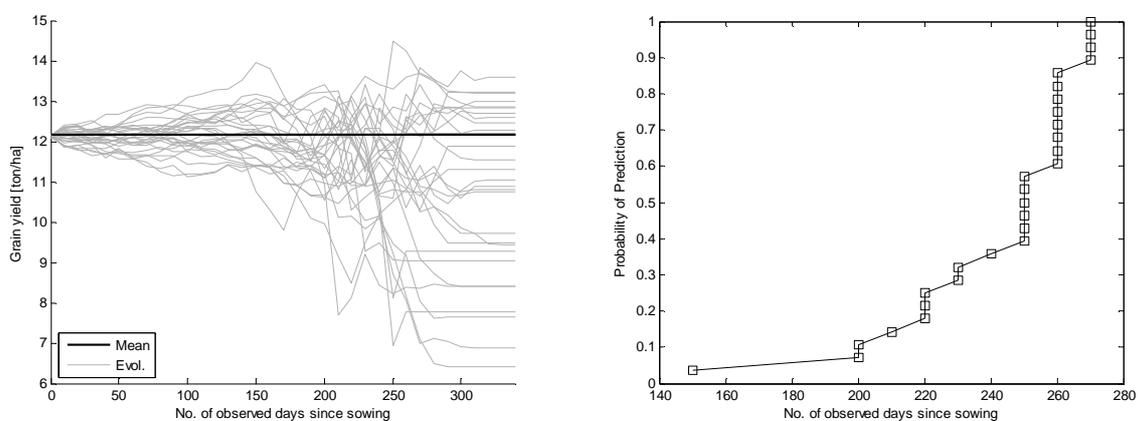


Figure 14: A gauche : Evolution du rendement grain pour toutes les "matrices-ensembles" et toutes les années de la base de données. Le trait noir continu correspond à la simulation avec le climat moyen. A droite : Distribution de probabilité cumulée relative à la prédiction du rendement grain.

Par ailleurs, comme on peut le voir sur la Figure 14, le MBC correspond bien à un climat pseudo-optimal, avec une prédiction de rendement de l'ordre de 12 t/ha. Les trente années de la MDB couvrent quant à elles une gamme de rendement allant de 6,5 à 14 t/ha.

La réalisation d'une prédiction du rendement statistiquement valable repose sur l'analyse des données de la Figure 14 (graphique de gauche). Il convient de déterminer, pour chaque année, la première combinaison de matrices de données climatiques (*i.e.* données historiques et projection climatique moyenne) à partir de laquelle la prédiction aurait été possible. Pour ce faire, on associe au rendement final obtenu sur le climat réel (Figure 14 - graphique de droite) un intervalle de confiance, c-à-d une erreur tolérée sur la prédiction. En fixant une erreur de 10 % (trait en tirets '-' noirs), le premier jour pour lequel la prédiction est possible correspond au premier jour (depuis le semis) inclus dans cet intervalle de confiance. Finalement, l'information correspondant à tous les premiers jours pour lesquels la prédiction est possible peut être synthétisée en traçant la fonction expérimentale de densité cumulée (CDF - cumulative density function) (Figure 14 - graphique de droite). C'est ce dernier graphique qui permettra au modélisateur de quantifier le niveau d'erreur associé à un délai prédictif donné.

4. Conclusions intermédiaires

La présentation des résultats sous la forme proposée fournit ainsi les distributions des rendements au cours de la saison et à travers les années. Cela permet (*i*) de mettre en évidence les périodes les plus sensibles à l'induction de stress, (*ii*) de quantifier instantanément les pertes de rendement dues aux conditions climatiques des 10 derniers jours en comparaison du potentiel restant à la culture et enfin (*iii*) de procéder à une estimation significative (CDF > 90%) du rendement (à une erreur de $\pm 10\%$). Grâce à cette méthodologie, il est possible de prédire le rendement grain environ un mois avant la date de récolte. Appliquée en étude préliminaire, cette approche constitue un outil de diagnostic puissant des potentialités d'une culture sur un site donné (où le sol est déterminé et les données climatiques sont connues sur plusieurs décennies).

La construction de la Matrice Ensemble proposée offre différents avantages. L'hypothèse d'un climat moyen en suppléance du climat futur confère au système une propriété prospective en permettant de simuler le cycle complet de croissance de la plante. A l'opposé, le rafraîchissement des entrées par des données réellement enregistrées permet de qualifier ce système de rétroactif, autorisant la prise en compte des stress passés dans la simulation de la croissance de la plante. Le traitement proposé des données sur les 30 années climatiques autorise la mesure du délai prédictif. Cela permet de quantifier et d'associer l'erreur de prédiction au niveau de fiabilité de cette prédiction. Finalement, la méthodologie offre la possibilité d'être interfacée, non pas avec des données historiques, mais avec un système d'enregistrement en temps réel des données météorologiques.

L'emploi d'un climat moyen en guise de climat hypothétique futur offre donc des perspectives intéressantes. Toutefois, afin de continuer le débat, il serait intéressant de comparer ces simulations avec celles obtenues en utilisant un climat médian. Si pour la plupart des variables climatiques les distributions quotidiennes sont normales, induisant une égalité de la moyenne et de la médiane, cela n'est pas le cas pour la pluviométrie pour laquelle la distribution est fortement asymétrique. D'une façon très générale, nos conditions climatiques sont caractérisées par beaucoup de faibles pluies, et

peu de très grosses pluies. De ce fait, la pluie médiane quotidienne serait bien plus faible que la pluie moyenne. L'impact sur les phénomènes de limitation de la croissance liés aux stress hydriques s'en trouverait accru. Même si ce n'est que pour l'exercice théorique, cette approche mériterait d'être explorée.

5. Publication associée

Dumont B., Leemans V., Ferrandis S., Vancutsem F., Seutin B., Bodson B., Destain J.-P., Destain M.-F., 2014. Assessing the potential of predicting wheat yield using a daily mean climate database. Accepté pour publication dans la revue Precision Agriculture. DOI : 10.1007/s11119-014-9346-9.

Voir publication associée :
<http://hdl.handle/2268/163330>

**Voir publication associée :
<http://hdl.handle/2268/163330>**

Voir publication associée :
<http://hdl.handle/2268/163330>

Voir publication associée :
<http://hdl.handle/2268/163330>

***Prédiction des rendements :
Etude du comportement du
modèle et comparaison de
deux approches validées***

*"Never argue with stupid people,
They will drag you down to their level and
Then beat you with experience."*

Mark Twain

Chapitre 7 : Prédiction des rendements : Etude du comportement du modèle et comparaison de deux approches validées.

1. Objectif

Dans le chapitre précédent, une méthodologie générale visant la mise au point d'un algorithme de prédiction du rendement a été proposée, implémentée et validée. Toutefois, cette recherche allait à l'encontre des affirmations de Semenov and Porter (1995) et Lawless and Semenov (2005) qui stipulaient que l'emploi d'un climat moyen en entrée d'un modèle de culture pouvait ne pas être adapté à la prédiction de rendement.

L'approche précédemment développée qui repose sur l'emploi des normales saisonnières en guise de projection pour le futur (approche dite "climat moyen") s'oppose donc à une approche à caractère stochastique, telle que l'approche proposée par Lawless et Semenov (2005). Celle-ci, faisant emploi d'un générateur d'aléas stochastiques, repose sur l'étude de la réponse du modèle sous une grande diversité de conditions climatiques induisant différents stress variant aussi bien par leur intensité que par le moment de leur occurrence.

La question de la pertinence de ces deux approches ne se pose pas, car elles desservent deux buts différents. De par la nature de son hypothèse climatique moyenne, n'induisant de ce fait que peu de stress, la première permet d'évaluer le potentiel de rendement restant à tout moment de la saison. La seconde quantifie quant à elle le rendement le plus probable, ainsi que l'intervalle de confiance qui l'accompagne. Mais il conviendrait toutefois de se poser la question du niveau de divergence entre ces deux approches.

Pour les besoins de cette recherche, deux voies d'études ont été empruntées. La première s'est attachée à la caractérisation des distributions des rendements. En 1965, Day affirmait que les distributions des rendements de cultures céréalières étaient en général non-normales et non-log-normales. N'ayant procédé qu'à une simple transformation logarithmique des données, il recommandait toutefois de s'intéresser à différents cas d'études de la distribution log-normale, dont notamment une réduction des données par un rendement maximal théorique pouvant être atteint par la culture avant d'appliquer la transformation logarithmique. A l'heure actuelle, l'étude des distributions de rendements fait encore l'objet de nombreuses recherches, notamment dans le secteur de la finance. Au sein des recherches les plus récentes, l'emphase est souvent portée sur le rassemblement des théories d'analyse des distributions des rendements et de celles des grandes lois de la statistique. Hennesy (2009a,b, 2011) a ainsi tenté d'expliquer les distributions des rendements observés en se référant à la loi des grands nombres ou encore à la loi du minimum de Liebig. C'est également cette seconde direction qu'empruntera la présente recherche.

L'objectif de cette recherche sera donc d'évaluer les deux méthodologies de prédictions du rendement évoquées précédemment sur une même base de données. C'est en se basant sur le théorème de la convergence en loi que leur capacités prédictives seront comparées.

2. Contribution originale à l'étude du comportement du modèle

2.1. Le générateur stochastique d'aléa climatique

La base de données d'Ernage (MDB) va être utilisée cette fois à deux reprises. Tout d'abord, elle sera employée afin de calculer les données climatiques journalières moyennes, comme illustré précédemment. Dans un second temps, elle sera analysée et utilisée via le générateur d'aléas stochastique LARS-WG (Weather Generator) (Racsco *et al.*, 1991; Semenov and Barrow, 1997)

La MBD a donc tout premièrement été analysée afin de produire les valeurs caractéristiques représentatives du site expérimental; à savoir les valeurs journalières minimales, maximales, moyennes et l'écart-type de chacune des variables climatiques. La distribution de fréquence de la pluie ainsi que celle des séries sèches et humides sont également fournies.

Le générateur d'aléas stochastiques est ensuite utilisé pour produire des séries climatiques synthétiques qui présenteront les mêmes caractéristiques que la série historique analysée. Sur base des travaux de Lawless et Semenov, ce sont 300 séries temporelles qui ont été générées dans cette étude. Les séries temporelles peuvent alors être utilisées directement en entrée du modèle STICS ou, dans ce cas-ci, être utilisées au travers de la méthodologie de prédiction du rendement.

2.2. Le théorème de la convergence en loi

Le théorème de la convergence en loi (\rightarrow_L) est considéré comme un des théorèmes les plus faibles parmi les lois de la convergence, mais il est cependant à la base d'un bon nombre de démonstrations d'autres théorèmes. Ce théorème est la pierre angulaire de cette étude autour duquel s'articulera toute l'analyse des résultats et de la comparaison des deux approches.

Considérons $\{X_n\}$ et $\{X\}$ comme deux variables aléatoires. Dénommons f une fonction de distribution d'une variable aléatoire. Le théorème de la convergence en loi stipule que $X_n \rightarrow_L X$ lorsque $n \rightarrow \infty$ si il existe une fonction f de l'espace réel ($R \rightarrow R$), continue et bornée, telle que

$$E[f(X_n)] \rightarrow E[f(X)] \quad (\text{Eq. 16})$$

2.3. Application de la démarche au cas d'étude

La validation de l'approche proposée se scinde en études de plus petites ampleurs. L'ensemble climatique généré à l'aide de l'outil LARS-WG sera représentée par la variable aléatoire X_n , et la normale saisonnière sera représentée par la variable aléatoire X . Les propriétés du logiciel LARS-WG impliquent que la génération d'un nombre suffisant de séries temporelles climatiques permettra de stabiliser la moyenne des données ainsi produites et que cette moyenne correspondra à la normale saisonnière.

Ensuite, il conviendra de voir dans quelle mesure la réponse offerte par le modèle STICS sous un grand nombre de réalisations climatiques, pourra être approximée par une fonction générale f . Finalement, dans la mesure où un tel comportement serait mis en évidence, les espérances mathématiques des simulations obtenues sous les deux méthodologies de prédiction du rendement pourront être comparées.

3. Discussions intermédiaires

3.1. Analyse des distributions de probabilités expérimentales

Afin de voir si un comportement du modèle peut être mis en évidence, la première question à laquelle il convient de répondre concerne la forme de la distribution offerte par le modèle. A titre illustratif, un exemple de résultat est donné à la Figure 15. Ces diagrammes présentent les distributions des rendements obtenus suite aux simulations du modèle sur 300 séries temporelles stochastiquement générées. Il n'y pas encore à ce stade de rafraîchissement des données. Les rendements simulés sont obtenus sur 300 climats indépendants.

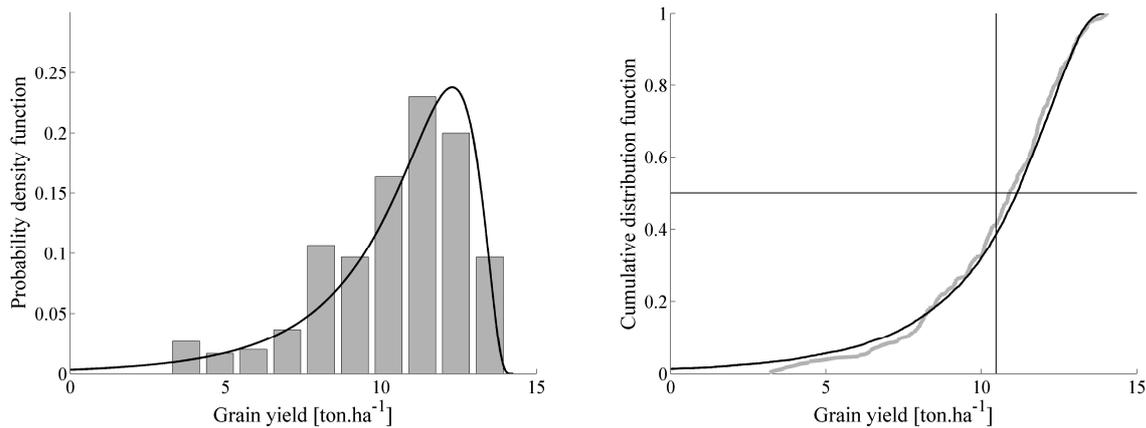


Figure 15 : Distribution de probabilité (histogramme gris) et fonction de densité cumulée (trait gris) des rendements simulés pour 300 séries temporelles stochastiquement générées. Distributions log-normales (trait noir) calculées sur base des valeurs caractéristiques des observations.

Tout d'abord, il apparaît que les discontinuités dans la courbe de distribution des rendements cumulés sont très faibles, traduisant une bonne exploration de l'espace des variables d'entrées. Ensuite, les différences entre la courbe cumulée des rendements simulés et la courbe log-normale théorique sont très faibles. Un test de Kolmogorov-Smirnov révèle qu'aucune différence statistique ne peut être mise en évidence entre ces deux courbes (p -value = 0.837).

Le fait qu'une courbe théorique s'ajuste si bien aux rendements simulés laisse suggérer que le modèle STICS dans son ensemble puisse être représenté plus simplement par une simple fonction f log-normale. En d'autres mots, un comportement sous-jacent au modèle peut donc être défini.

Il est ici important d'insister sur le fait que les 300 scénarios climatiques générés présentent les mêmes propriétés que les conditions climatiques de la base de données historiques. La courbe des rendements observés n'est donc ni plus ni moins que l'expression phénotypique du génome de la plante en interaction avec son environnement. Autrement dit, cette courbe correspond à la distribution des rendements à laquelle on peut s'attendre si on cultivait la même variété durant 300 ans. Dans la mesure où un seul scénario climatique se réalisera par saison de culture, cette distribution de rendement permettra donc de quantifier le niveau de probabilité des différents rendements auxquels on peut s'attendre en début de saison.

Il est finalement à signaler que ce comportement a par ailleurs été confirmé pour la majeure partie de la saison de culture, lorsque qu'une méthodologie de rafraîchissement des données telle que

celle proposée par Lawless et Semenov (2005) est appliquée sur les différentes entrées climatiques historiques du modèle STICS. A tout moment de la saison, la réponse du modèle concernant le rendement de fin de saison peut donc être anticipé.

3.2. Comparaison des deux approches

L'existence d'un comportement sous-jacent au modèle ayant été mis en évidence, la seconde question qui intéressera le modélisateur consiste à déterminer dans quelle mesure les deux approches de prédiction du rendement auront tendance à diverger ou à converger au niveau de leurs prédictions.

En guise d'exemple, la Figure 16 illustre une comparaison des deux approches de prédiction du rendement. Pour cet exercice, le taux de rafraîchissement des données a été fixé à 10 jours, et les deux méthodologies sont appliquées aux entrées climatiques de la saison 1981-1982. La RRMSE et la ND sont respectivement de 0.044 et -0.034 pour les résultats présentés. Ils traduisent une erreur bien inférieure à 10% des valeurs du rendement final simulé.

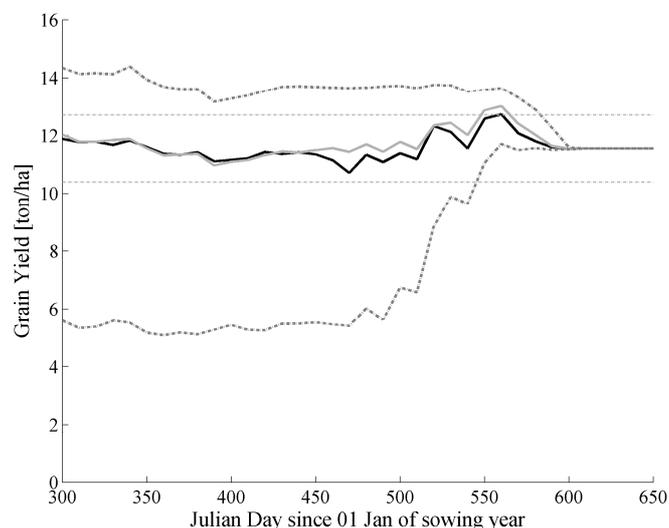


Figure 16 : Comparaison des prédictions des sorties du modèle pour les deux méthodologies de prédiction du rendement appliquées aux entrées climatiques de la saison 1981-1982. L'intervalle de confiance à 95% autour de la prédiction est présenté en trait gris discontinu. L'espérance mathématique (moyenne) des observations est présentée en trait gris continu. La sortie de la méthodologie reposant sur la normale saisonnière est présentée en trait noir.

Ce comportement a été observé pour 27 des 30 années de simulation. Dans 10% des cas, soit pour 3 saisons simulées les RRMSE et ND étaient supérieures à 10%. Pour ces situations, l'explication viendrait sans doute des conditions climatiques particulières lors de l'implantation de la culture.

Finalement, tout comme ce fut le cas au chapitre précédent, après avoir appliqué la méthodologie à toutes les années de la base de données climatiques, il est possible de calculer les délais prédictifs offerts par les deux approches. Nous renvoyons à la publication associée à ce chapitre pour plus de détails sur ces résultats. Toutefois, il est important de signaler que les 10% des situations pour lesquelles les simulations divergent n'ont que peu d'impacts sur les délais de prédiction offerts par les deux approches.

Finalement, la Figure 17 présente l'ensemble des rendements finaux (à un taux de remplacement décadaire) simulés par les deux approches, appliqués aux trente années de la base de données climatiques originale. La RRMSE et la ND correspondant à ce graphique, c'est-à-dire à la globalité des

données fusionnées sont respectivement égales à 0.112 et -0.058.

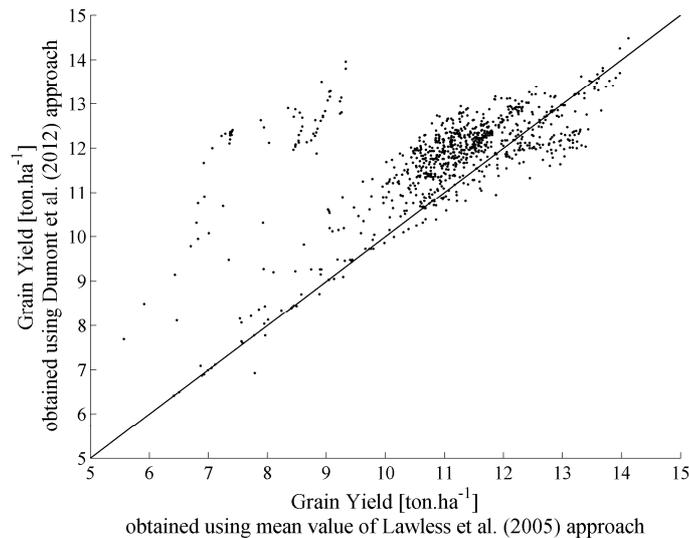


Figure 17 : Confrontation graphique des deux approches de prédiction du rendement.

A ce stade, il est intéressant de se pencher sur une analyse globale du graphique. Un biais global, correspondant à une surestimation d'environ 6%, existe au niveau de la méthodologie reposant sur le climat moyen. Comme discuté précédemment, l'emploi de conditions climatiques moyennes en guise de projections futures correspond à des conditions de croissance presque non-limitatives. A l'opposé l'approche probabiliste offre un panel de situations climatiques contrastées, auxquelles différents stress (hydrique, échaudant, ...) et différents niveaux de stress induits sont associés. En contrepartie, de par sa nature même et sa construction mathématique, l'approche probabiliste explore aussi des situations climatiques plus favorables que la normale saisonnière. *In fine*, La moyenne des rendements correspondant aux scénarii proposés par l'approche probabiliste semble toutefois relativement proche des rendements obtenus avec l'approche climatique moyenne.

4. Conclusions intermédiaires

Lors de la mise au point d'outils d'aide à la décision à vocation agronomique, le développeur se trouve face une décision aux conséquences antagonistes. D'un côté, il apparait extrêmement important de mettre au point des algorithmes aussi rapides que possible tout en fournissant une réponse fiable pour l'agriculteur, lorsqu'une décision doit être posée. A l'opposé, l'homme de science, en respect avec la statistique, voudra quantifier la qualité et l'incertitude des prédictions liées aux différentes situations climatiques. Pour ce faire, seule la voie de la multi-simulation est envisageable, mais s'avère être un processus coûteux en temps et en ressource machine.

Dans ce chapitre, l'approche prédictive sur le rendement présentée au chapitre précédent a donc été comparée à une approche climatique stochastique telle que proposée en 2005 par Lawless et Semenov, et ce, pour la première fois, sur un même set expérimental (modèle, climat, données, ...).

A ce stade de l'étude, il y a de fortes présomptions de trouver dans l'emploi des modèles de culture un cas d'applicabilité du Théorème de la Convergence en Loi. En effet, un comportement

log-normal sous-jacent au modèle est indéniable. Il existe donc bel et bien une loi f qui permette de synthétiser le fonctionnement du modèle et de lier les entrées climatiques aléatoires X_n aux rendements correspondants. Dans un second temps, l'équivalence de $E[f(X_n)]$ et $E[f(X)]$ a pu être validée avec succès dans la majorité des situations climatiques envisagées. Il reste toutefois à comprendre pourquoi certaines situations induisent un comportement différent. A l'heure actuelle nos hypothèses se portent sur les conditions climatiques d'implantation de la culture juste après le semis, qui pourraient être trop défavorables à l'implantation ou alors être hors des conditions de calibration du modèle (*e.g.* les conditions de température et d'humidité affectant le taux de germination).

Avant toute généralisation, il est par ailleurs à signaler que ces résultats sont propres au climat étudié, c'est-à-dire un climat tempéré. Toutefois, si elle ne permet pas de la réduire, cette étude permet bien de quantifier l'incertitude associée aux prédictions du rendement de fin de saison, et plus particulièrement à évaluer celle liée à l'emploi de normales saisonnières en guise de projection climatique.

Au travers de cette étude, il a été démontré que les normales saisonnières pouvaient, employées au sein d'un algorithme approprié, et au même titre qu'une approche probabiliste, être utilisées pour mettre au point un outil efficace de prédiction du rendement. Celui-ci offre l'avantage de ne reposer que sur une projection climatique future, ce qui en fait donc un outil relativement plus rapide en termes de temps de calcul, 300 fois dans ce cas-ci.

5. Publication associée

Dumont, B., Basso, B., Leemans, V., Ferrandis S., Bodson, B., Destain, J.-P., Destain, M.-F., 2014. A comparison of within season yield prediction algorithms : On the analysis of the crop model behaviour. Soumis à la revue Agricultural and Forest Meteorology.

Publication en révision

Comportement des modèles : Fertilisation azotée & Incertitude climatique

*"I've always believed that nothing is withheld from us
what we have conceived to do !
God said it and there were only two people who believed it.
You know who ? God and me,
So I went out and did it."*

" Do things that have never been done before !"

Russell A. Kirsch

The man who invented the first programmable computer

Extract from <http://joelrunyon.com/two3/an-unexpected-ass-kicking/>

Chapitre 8 : Comportement des modèles dynamiques en réponse à la fertilisation azotée et la variabilité climatique

1. Objectif

Les modèles de cultures constituent des outils indispensables pour comprendre l'influence des conditions agro-environnementales sur le système sol-plante, et peuvent par exemple être employés pour estimer les impacts du climat ou du changement climatique sur la production (Varella, 2011). En parallèle à cela, dans certains domaines, par exemple celui du marché des assurances liées aux catastrophes naturelles et aux pertes de rendement, des études s'attachent depuis plus de 50 ans à caractériser les distributions empiriques de rendements, et à comprendre les mécanismes qui les gouvernent. A l'heure actuelle, nous sommes d'avis que les modèles de cultures pourraient être employés à une telle fin. Cependant, personne ne s'est pour l'instant attaché à l'étude du comportement intrinsèque du modèle.

Parmi les études de terrain dédiées à l'étude des distributions des rendements, celle de Richard Day (Day, 1965), portant sur les cultures de maïs, d'avoine et de coton, est sans doute la plus complète et la plus pertinente. S'attachant à la définition de ce qu'il appelle sa connaissance *a priori* des cultures, Day énonçait qu' "*il était clair que la distribution des rendements au champ des cultures présentait une limite inférieure qui, dans le pire des cas, serait le rendement zéro. De la même façon, il était tout aussi clair que la culture d'une espèce donnée, sous des pratiques culturales constantes, mais sous des conditions climatiques variables, devait offrir une limite de rendement supérieur, et ce, même sous les conditions de croissance les plus favorables*". En considérant ce fait, Richard Day s'est proposé d'analyser le comportement des distributions en utilisant d'une part le système de Pearson. Il a par ailleurs conclu que (i) les distributions de rendements sont en général non normales et que (ii) l'asymétrie et l'aplatissement des distributions dépend d'une part de la culture étudiée, et d'autre part, de son niveau de nutrition azotée.

Poursuivant son analyse, Day souligne que le mode ou la médiane des distributions sont bien plus adaptés que sa moyenne dans une optique de prédiction. Il étend par ailleurs ces considérations à la prédiction du risque associé au niveau de fertilisation azoté qu'il conviendrait d'appliquer aux cultures. Ses conclusions furent corroborées par le travail de Du *et al.* (2012), qui énonçaient que le développement d'une théorie complète sur la manière dont les entrées des systèmes sols plantes affectent la symétrie des distributions des rendements correspondantes devrait attendre le développement et la mise en place d'expériences plus systématiques portant leur étude sur différentes cultures et différentes conditions de croissance.

Nous sommes d'avis que cette "théorie complète" attendue par Du *et al.* (2012) pourrait être l'expérimentation numérique à grande échelle offerte par l'emploi des modèles de culture. Dans ce chapitre, nous transposons les théories de Richard Day (Day, 1965) à l'étude du comportement des modèles. En particulier, et en lien avec le chapitre précédent, nous proposons d'appliquer ses théories à l'étude du comportement des modèles couplés à leur emploi à des fins prédictives (Lawless and Semenov, 2005).

2. Contribution originale à la gestion stratégique de la fertilisation azotée

2.1. Le système de Pearson appliqué à l'étude des distributions

Pearson a développé un système alternatif pour l'étude des fonctions de densité (Day, 1965 ; Pearson, 1894) capable de prendre une large variété de formes. Le système de Pearson contient entre autre cas particuliers la distribution normale, la *t*-Student, la log-normale, et bien d'autres. Le système de Pearson est originellement présenté sous la forme différentielle :

$$\frac{df(x)}{dx} = \frac{x-a}{c_0 + c_1x + c_2x^2} f(x) \quad (\text{Eq. 17})$$

Expression dans laquelle x est la variable (aléatoire) et $f(x)$ sa fonction de densité. Par intégration de cette équation, les constantes a , c_0 , c_1 , c_2 peuvent être utilisées pour caractériser les différentes distributions du système. Par ailleurs, il est possible de réécrire l'équation selon :

$$\frac{df(x)}{dx} = \frac{x-a}{c_2(x-\alpha_1)(x-\alpha_2)} f(x) \quad (\text{Eq. 18})$$

Expression dans laquelle α_1 et α_2 sont les racines du terme quadratique. En intégrant cette nouvelle forme de l'équation, et en sélectionnant des valeurs positives de α_1 et α_2 , on retombe sur les distributions de type I du système.

$$f(x) \begin{cases} = k.(x-\alpha_1)^{m_1}(x-\alpha_2)^{m_2} & , \alpha_1 < y < \alpha_2 \\ = 0 & , y \leq \alpha_1 \text{ or } \alpha_2 \leq y \end{cases} \quad (\text{Eq. 19})$$

Le choix de valeurs positives pour α_1 et α_2 se justifie par rapport à la définition *a priori* de notre connaissance du système de culture, tel qu'énoncé par Day. Les distributions de type I sont donc celles qui devraient correspondre à notre cas d'étude. Elles offrent en effet la propriété de présenter une distribution bornée par une valeur inférieure et une valeur supérieure. Or il est logique de considérer que les rendements sont naturellement et physiquement limités inférieurement à un rendement nul. Par ailleurs, même sous un très haut niveau de nutrition (éléments minéraux) et des conditions climatiques favorables, la culture offrira un rendement maximal limité par sa génétique.

2.2. Les coefficients de Pearson

En parallèle à la mise au point de son système, Pearson a développé deux coefficients permettant de caractériser le degré d'asymétrie et d'aplatissement des distributions, et ce, indépendamment de la moyenne et de la variance de la distribution. Il définit ainsi les paramètres β_1 et β_2 comme le rapport des moments centrés m_p , d'ordre p , eux-mêmes définis par rapport à la moyenne des observations m_x .

$$m_p = \frac{1}{n} \sum_{i=1}^n (X_i - m_x)^p \quad (\text{Eq. 20})$$

$$\beta_1 = \frac{m_3^2}{m_2^3} = \frac{m_3^2}{\sigma^6} \quad (\text{Eq. 21})$$

$$\beta_2 = \frac{m_4}{m_2^2} = \frac{m_4}{\sigma^4}$$

Le paramètre β_1 est également connu sous le nom de *skewness* au carré, et le paramètre β_2 est autrement appelé le *kurtosis* de la distribution. Le *skewness* d'une distribution, contrairement au paramètre β_1 a l'avantage d'être signé, ce qui permet de connaître le sens de la dissymétrie. Rappelons également que la moyenne est égale au moment d'ordre 1, et que sa variance est égale au moment d'ordre 2. Par ailleurs, dans la continuité de ses travaux, Pearson a établi le lien entre d'une part les paramètres β_1 et β_2 , et d'autre part, les coefficients c_0, c_1, c_2 .

Il est ainsi possible de caractériser entièrement une distribution expérimentale au travers du calcul des quatre grandeurs que sont les moments des quatre premiers ordres. Et finalement, sur base des grandeurs caractéristiques des coefficients de l'équation 21, il est possible de définir le type de distribution le plus à même de représenter la distribution expérimentale.

3. Discussions intermédiaires

3.1. Réponse de la culture à différents niveaux de fertilisation azotée et sous des conditions climatiques variées.

Comme présenté au chapitre précédent, l'emploi d'un générateur d'aléas climatiques permet de quantifier l'incertitude associée à différentes réalisations climatiques dans une région donnée. Repartant de cet outil, ce chapitre vise à étudier l'interaction qui existe entre les pratiques liées à la gestion azotée avec et la variabilité climatique d'une région donnée.

A titre illustratif, la Figure 18 présente les rendements obtenus pour deux expérimentations numériques réalisées toutes deux sous 300 entrées stochastiquement générées et pour des niveaux de fertilisation d'une part nul (graphique de gauche) et d'autre part à la fumure de référence ($180 \text{ kgN}\cdot\text{ha}^{-1}$), fractionnée en doses égales ($60 \text{ kgN}\cdot\text{ha}^{-1}$) appliquées respectivement aux stades 'tallage', 'redressement' et 'dernière feuille' (graphique de droite).

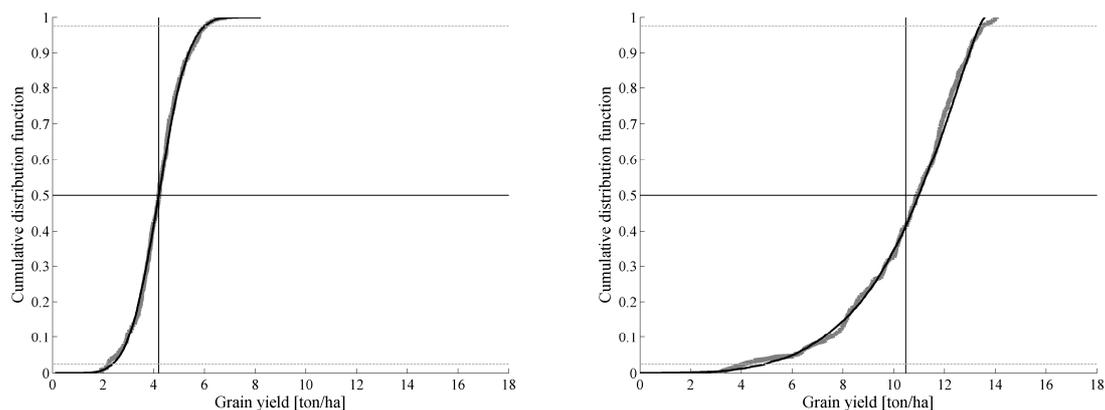


Figure 18 : Comparaison des résultats de l'expérimentation numérique (trait et histogrammes gris) avec la distribution de Pearson Type I (traits noirs) ajustée sur base de la méthode des moments.

A gauche : Fonction de densité de probabilité cumulée pour un niveau de fertilisation $0-0-0 \text{ kgN}\cdot\text{ha}^{-1}$.

A droite : Fonction de densité de probabilité cumulée pour un niveau de fertilisation $60-60-60 \text{ kgN}\cdot\text{ha}^{-1}$.

Comme on peut le constater, à première vue, la distribution de type I ajustée par la méthode des moments semble particulièrement adaptée à la représentation des rendements simulés par le modèle.

Les distributions des rendements sous les deux niveaux de fumure semblent quant à elles bien différentes. Afin d'étudier en détail la façon dont évoluent les distributions sous des niveaux croissants d'apports d'azote, différents protocoles ont été testés avec des incréments réguliers de 10 kgN.ha^{-1} appliqué à chaque apport ('tallage', 'redressement' et 'dernière feuille'), soit un incrément de 30 kgN.ha^{-1} par sets de 300 simulations. Les niveaux d'asymétrie, caractérisés par leur skewness, ont également été étudiés (Figure 19). Une bonne représentation des niveaux d'asymétrie peut être observée sur le graphique de droite pour lequel la moyenne ('-*--', courbe la plus basse), la médiane ('--□--', trait horizontal) et le mode ('--o--', courbe la plus haute) des distributions divergent sous des niveaux de fertilisation croissants.

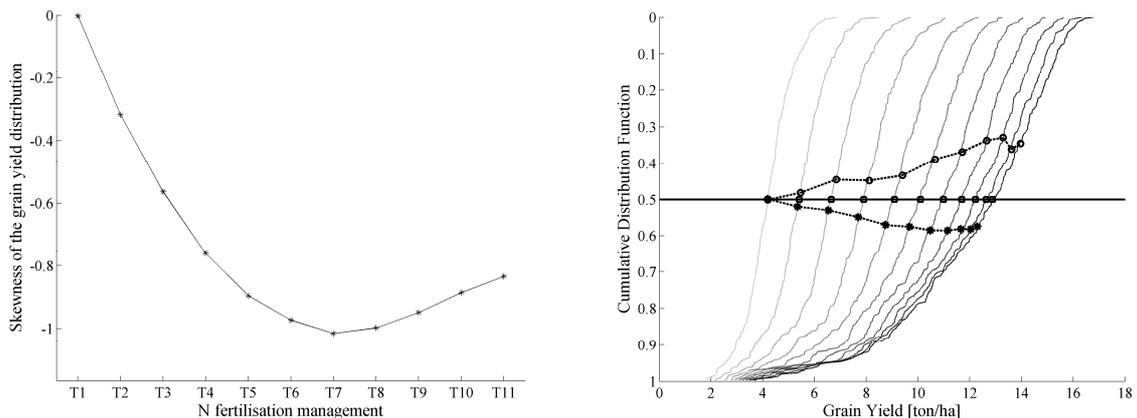


Figure 19 : Etude du niveau d'asymétrie dans les simulations conduites sous différents niveaux de fertilisation, variant de $0-0-0 \text{ kgN.ha}^{-1}$ à $100-100-100 \text{ kgN.ha}^{-1}$ appliqués aux stades 'tallage', 'redressement' et 'dernière feuille', et pour des incréments de 30 kgN.ha^{-1} par apport ($10-10-10 \text{ kgN.ha}^{-1}$).

A gauche : Skewness des distributions des traitements $0-0-0 \text{ kgN.ha}^{-1}$ (T1) à $100-100-100 \text{ kgN.ha}^{-1}$ (T11).

A droite : Etude du rendement en grains à des niveaux de fertilisation azotée croissants (traits de nuance grise croissante) et sous différents (#300) scénarios climatiques générés stochastiquement (CDF). Les courbes dont l'intensité de gris augmente sur le graphe de droite représentent les traitements N croissants. La moyenne(*--), la médiane(--□--) et le mode (--o--) des distributions ont été numériquement dérivés.

Il est ainsi intéressant de signaler que la distribution des rendements observés en l'absence de fertilisation s'avère être normale, impliquant que la moyenne, le mode et la médiane sont confondus, et que le niveau d'asymétrie est égal à zéro. Sous des niveaux croissants d'apports d'azote, le niveau d'asymétrie augmente. Les skewness négatifs signifient une dissymétrie droite, caractérisée par une queue de distribution située à gauche de la médiane.

La fraction azotée offrant le niveau de dissymétrie le plus fort est la fumure de référence $60-60-60 \text{ kgN.ha}^{-1}$. Cela signifie qu'à ce niveau de fertilisation, la probabilité pour le fermier d'obtenir un rendement au moins supérieur à la moyenne est la plus grande. En d'autres mots, le risque est minimal. Avec un accroissement des apports d'azote, les rendements ne deviennent donc pas seulement plus élevés, mais les rendements élevés deviennent aussi plus fréquents !

3.2. Gestion stratégique de l'azote : à la recherche d'une fertilisation optimale

Les recherches menées sur la prédiction du rendement en temps-réel nous avaient déjà permis de mettre en évidence qu'avant le stade de la dernière feuille, l'influence de mauvaises conditions climatiques était réduite. Les capacités de compensation d'une culture à croissance déterminée comme

le blé d'hiver sont telles qu'avant la mise en place du grain, les pertes de rendement liées aux conditions de croissance passées restent limitées. Par contre, une fois le nombre de grains mis en place, ceux-ci "ne peuvent plus qu'être remplis", à concurrence des conditions climatiques du moment.

Par ailleurs, à la section précédente, il a été démontré que le protocole d'application d'azote le plus courant en Belgique (60-60-60 kgN.ha⁻¹) offrait le risque minimal dans l'obtention d'un rendement relativement élevé.

De ce fait, il a été décidé de tester une nouvelle série de protocoles d'application d'azote. Tout d'abord, il a été décidé de ne surtout jamais réduire les fractions de tallage et redressement afin de ne pas induire de carence azotée précoce. Sur cette base, le set de pratique azoté consiste uniquement à moduler la fraction de dernière feuille. Ce protocole sera appelé modulo-60 (M60-X). Des incréments de 10 kgN.ha⁻¹ ont été testés entre les bornes 0 et 100 kgN.ha⁻¹, soit des niveaux totaux variant de 120 (M60-1) à 220 kgN.ha⁻¹ (M60-11) appliqués sur la saison.

Finalement, l'agriculteur sera davantage intéressé par l'obtention d'un revenu maximal, plus que d'un rendement maximal. Dans cette optique, un critère économique basique, connu comme le revenu marginal net (MNR) a été calculé.

$$MNR = (Y_N \cdot G_p) - (N \cdot N_p) \quad \text{Eq. 1}$$

où MNR est le revenu marginal net (€.ha⁻¹), Y_N est le rendement grain simulé (ton.ha⁻¹), G_p est le prix de vente du grain (€.ton⁻¹), N est la quantité totale d'azote appliquée durant la saison (kgN.ha⁻¹), N_p est le prix d'achat de l'azote (€.kgN⁻¹). Pour cette partie de l'étude des prix raisonnables observés au cours des saisons 2011 et 2012 ont été considérés, 180 et 300 €.ton⁻¹ respectivement pour les prix de vente du grain et d'achat de l'azote.

Afin de déterminer la fraction N optimale, la manipulation de la Figure 19 n'est pas la plus évidente. En effet, celle-ci présente, pour chaque niveau de fertilisation, la distribution des rendements ou des revenus nets correspondants. Or, pour optimiser les apports d'azote, c'est la représentation opposée qui s'avère la plus intéressante; il conviendrait, pour chaque niveau de probabilité, de confronter les rendements ou revenus nets en fonction des fractions azotées.

Toutefois, dans l'optique de la mise au point d'un outil d'aide à la décision, afin de synthétiser ces deux représentations, cette étude propose une représentation en trois dimensions (Figure 20). Une surface de réponse présente ainsi les rendements ou revenus nets en fonction des différents protocoles azotés et des niveaux de probabilité correspondant aux temps de retour des différents scénarii climatiques.

Dans la mesure où l'avenir est incertain et imprévisible, il est impossible de déterminer notre position sur ce graphique en terme climatique (quel sera le climat de l'année en cours ?). Pour cette raison, il est tout aussi impossible de déterminer la fraction azotée optimale correspondant au climat de l'année en cours.

Ainsi, la représentation en trois dimensions permet de contourner le problème de l'optimisation de la fraction azotée. La question qui se cache derrière le code d'analyse est la suivante : Quelle est le niveau de fertilisation qui donnera les meilleurs résultats et surpassera les autres dans la majorité des simulations climatiques ?

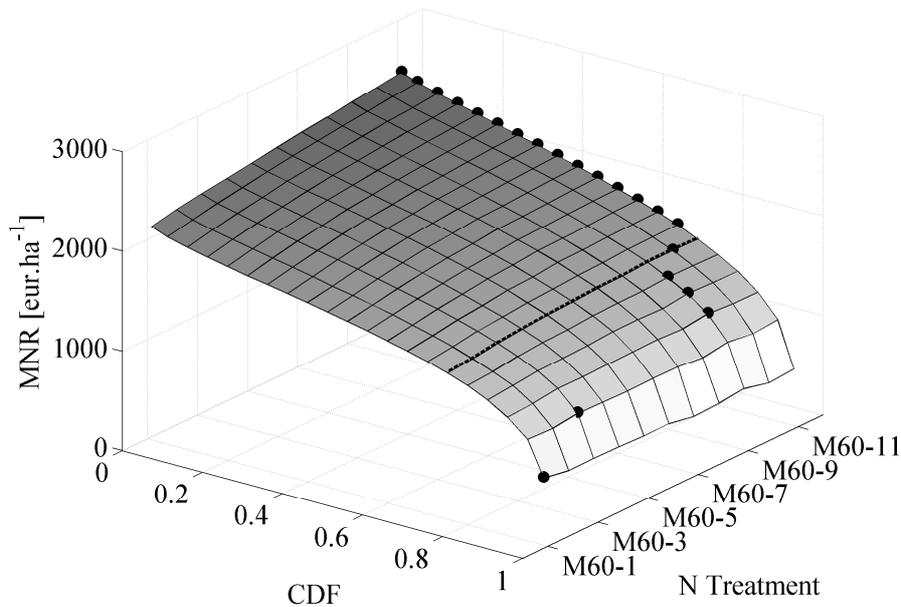


Figure 20 : Fonction de densité cumulée (CDF) du revenu marginal net (MNR) exprimé en fonction du niveau de fertilisation azoté et tracé pour 300 scénarii climatiques. Les différents protocoles azotés correspondent à une application systématique de 60 kgN.ha⁻¹ aux stades tallage et redressement, mais pour des niveaux croissants d'azote appliqués au stade dernière feuille variant de 0 kgN.ha⁻¹ (M60-1) à 100 kgN.ha⁻¹ (M60-11) par pas de 10 kgN.ha⁻¹. Pour chaque niveau de probabilité (1%, 5% 10%, ..., 95%, 99%) la fraction maximisant le MNR est représentée par le point noir (•)

Pour chaque climat à venir, c'est-à-dire pour chaque niveau de probabilité, il faut donc repérer sur le graphique le niveau de fertilisation qui permettrait de maximiser le revenu net (points noirs "•" présentés sur la surface de réponse à la Figure 20). On observe assez logiquement d'après la Figure 20 que si le climat à venir est de type défavorable (CDF tendant vers 1), il conviendra de diminuer le niveau de fertilisation. Pour des climats de type favorable (CDF tendant vers 0), on pourra augmenter le niveau de fertilisation.

Il reste finalement à déterminer ce qu'on appelle "*la majorité du temps*". Suivant les recommandations de Basso *et al.* (2012), la fertilisation optimale doit être sélectionnée comme étant la fertilisation qui surpasse les autres 75% du temps. Ce niveau significatif est représenté sur la Figure 20 par le trait noir discontinu. Comme on peut le voir, le protocole 60-60-90 kgN.ha⁻¹ est celui qui répond à cette définition.

Cependant, un test de Wilcoxon (Tableau 6) visant l'inter-comparaison des protocoles de fertilisation révèle que cette pratique ne peut être considérée statistiquement différente d'une fertilisation où 60-60-60 kgN.ha⁻¹ sont appliqués, soit la fertilisation de référence.

Il est également à noter que 95 % du temps, soit un peu plus de 9 années sur 10, l'application de 20 kgN.ha⁻¹ à la dernière feuille (M60-3) s'avèrerait suffisante pour maximiser le revenu marginal net de l'agriculteur.

Tableau 6: Comparaison des distributions du revenu marginal net en utilisant un test de Wilcoxon, pour différents traitements de fertilisant azoté appliqués, basé sur une modulation de la dose de dernière feuille (60-60-XX kgN.ha⁻¹). Les résultats dans le tableau représentent la *p*-valeur du test de Wilcoxon. Les niveaux de signification sont représentés par le nombre d'étoiles ('*' : $p < 0.001$; '**' : $p < 0.01$; '*': $p < 0.05$).**

Treat.	60-60-100 (M60-11)	60-60-90 (M60-10)	60-60-80 (M60-9)	60-60-70 (M60-8)	60-60-60 (M60-7)	60-60-50 (M60-6)	60-60-40 (M60-5)	60-60-30 (M60-4)	60-60-20 (M60-3)	60-60-10 (M60-2)
60-60-90 (M60-10)	0.682									
60-60-80 (M60-9)	0.401	0.613								
60-60-70 (M60-8)	0.161	0.287	0.523							
60-60-60 (M60-7)	0.043*	0.090	0.210	0.466						
60-60-50 (M60-6)	0.007**	0.015*	0.044*	0.146	0.398					
60-60-40 (M60-5)	0.000***	0.001**	0.004**	0.019*	0.087	0.334				
60-60-30 (M60-4)	0.000***	0.000***	0.000***	0.001***	0.007**	0.052	0.282			
60-60-20 (M60-3)	0.000***	0.000***	0.000***	0.000***	0.000***	0.002**	0.022*	0.187		
60-60-10 (M60-2)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.005**	0.105	
60-60-0 (M60-1)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.039*

4. Conclusions intermédiaires

Le comportement du modèle tel que nous l'avons décrit semble ainsi pouvoir être à tout moment représenté par la distribution théorique de type I du système de Pearson. Cette distribution semble en effet particulièrement adaptée à la représentation des simulations de la croissance des modèles de culture. En effet, elle est physiquement limitée par deux extrema, ce qui correspond à la connaissance intuitive que nous avons des systèmes de culture. Par ailleurs, la distribution de type I permet de tenir compte du niveau d'asymétrie et d'aplatissement des distributions.

Poussant davantage encore l'analyse du comportement des modèles, il a été démontré que la simulation du rendement d'une culture sous un niveau correspondant à la fertilisation de référence, offrait le maximum d'asymétrie, soit le risque minimal d'obtenir des rendements élevés pour l'agriculteur.

Sur cette base, un nouveau set de pratiques de fertilisation a été conçu, visant à étudier l'effet unique d'une modulation de la fraction de dernière feuille. Ce nouveau protocole correspond par ailleurs davantage à la réalité de terrain. D'une part, le blé d'hiver est à même de compenser tout au long de sa croissance; par exemple, une faible densité de plantes levées ou ayant survécu à l'hiver sera compensée par un nombre important de talles. Avant la mise en place du grain, l'effet de conditions climatiques défavorables est donc le plus souvent réduit.

Par ailleurs, le développement atteint par la culture au moment de la troisième application d'azote, influera sur sa capacité future à remobiliser ses réserves éventuelles. En parallèle à cela, si les prix d'achat de l'azote sont connus, au même instant, les prix du marché liés à la vente des grains commencent à se préciser. C'est donc l'instant crucial pour l'agriculteur dans la recherche d'un niveau de fertilisation lui assurant non seulement des rendements élevés, mais avant tout un revenu maximal !

La méthodologie de gestion stratégique de l'azote, entendons par là, sans connaissance *a priori* des conditions climatiques futures, nous permet d'offrir un aperçu des scénarii climatiques à venir, et donc d'optimiser la pratique de fertilisation. Cette méthodologie n'est toutefois pas limitative et laisse à l'agriculteur le choix du niveau de risque qu'il tolère.

Bien que la méthodologie proposée offre plusieurs avantages et semble prometteuse, il convient pour l'heure d'être encore prudent avant toute généralisation des résultats. Ceux qui sont proposés sont pour l'instant propres au site étudié, à la fois en termes de conditions climatiques et de type de sol. Par ailleurs, les résultats des simulations correspondent à des situations d'azote natif, *i.e.* l'azote du sol en début de saison, basées sur les mesures effectuées. Il conviendrait d'évaluer les résultats de la méthodologie proposée pour toute une série de conditions initiales différentes, référant à différents précédents culturaux et des types de sol variables.

Finalement, il importe de discuter des conclusions de cette étude basée sur les simulations. Il faut être extrêmement prudent avant de conclure sur le niveau de fertilisation optimal. Il peut en effet sembler étrange que le fractionnement optimal soit celui habituellement pratiqué. Il faut tout d'abord se rappeler que la fonction de réponse à la fumure implémentée dans le modèle a été calibrée sur base des observations de terrain pour la variété étudiée. Il est important de garder à l'esprit que le modèle représente une réalité, mais qu'il ne constitue pas celle-ci ! Toutefois, et cela est indéniable, le modèle et la méthodologie proposée permettent d'estimer le risque lié aux différents scénarii climatiques. Or, si pour chaque simulation le rendement est bien dépendant de la fonction calibrée de réponse à la fertilisation, la méthodologie proposée détermine quant à elle la fraction la plus intéressante en fonction du risque climatique.

Ainsi, afin d'apporter des éléments de réponse nouveau à ce débat, il serait intéressant de tester les réponses de l'algorithme, pour d'autres types de sols et d'autres conditions climatiques, mais pour la même variété. Les résultats obtenus pourraient ensuite être comparés à ceux d'expérimentations menées en champs en parallèle.

5. Publications associées

Dumont, B., Basso, B., Leemans, V., Bodson, B., Destain, J.P., Destain, M.F., 2013. Yield variability linked to climate uncertainty and nitrogen fertilisation, In: Stafford, J. (Ed.), Precision Agriculture '13. Wageningen Academic Publishers, pp. 427-434. <http://hdl.handle.net/2268/152077>

Dumont B., Basso B., Leemans V, Bodson B., Destain J.-P., Destain M.-F., 2013. Systematic analysis of site-specific yield distributions resulting from nitrogen management and climatic variability interactions. Soumis à la revue Precision Agriculture.

Voir publication associée :
<http://hdl.handle/2268/152075>

Conclusions générales & Perspectives

*"Entre : Ce que je pense, Ce que je veux dire, Ce que je crois dire,
Ce que je dis, Ce que vous avez envie d'entendre,
Ce que vous croyez entendre, Ce que vous entendez,
Ce que vous avez envie de comprendre, Ce que vous comprenez,
Il y a dix possibilités qu'on ait des difficultés à communiquer.
Mais essayons quand même..."*

*Bernard Werber,
Le père de nos pères - 1998.*

Chapitre 9 : Conclusions générales et perspectives

1. Conclusions générales

Les modèles de culture, encore appelés modèles écophysio-logiques, décrivent la croissance et le développement des cultures en interaction avec leurs conditions agro-environnementales, *i.e.* le sol, le climat et l'environnement proche (*e.g.* compétition trophique au sein de la culture) dans lequel croît une plante.

Il s'agit de modèles dynamiques non-linéaires constitués d'un grand nombre d'équations différentielles interdépendantes. Au vu du grand nombre de variables d'entrée et de paramètres requis, la mise en œuvre de tels modèles peut s'avérer délicate. Pour ces raisons, il est fréquent que des écarts, parfois importants, existent entre les valeurs mesurées et simulées. Le premier objectif de cette thèse a été d'étudier l'incertitude sur les sorties d'un modèle liée à la paramétrisation des différents modules qui le constituent. Une fois calibré, et pour autant que les conditions climatiques à venir puissent être anticipées et/ou supplées, un modèle peut s'avérer très pertinent en vue de mettre au point des outils de prédiction ou d'aide à la décision. Dans un second temps, cette thèse de doctorat s'est donc focalisée sur la prise en compte de l'incertitude inhérente aux données climatiques pilotant les simulations.

Des méthodes stochastiques d'estimation paramétrique reposant sur le théorème de Bayes ont tout d'abord été mises en œuvre. Parmi celles-ci, nous nous sommes focalisés sur les approches formelles qui autorisent la prise en compte d'un modèle statistique lié à la représentation des résidus entre le modèle et les mesures. L'algorithme d'échantillonnage DREAM (DiffeREntial Evolution Adaptive Metropolis) a été interfacé avec le modèle de culture STICS. Il est important de souligner que ce type d'algorithme constitue une méthode d'échantillonnage et n'est en aucun cas une méthode d'optimisation paramétrique à proprement parler. En effet, au terme du processus numérique, ce sont les distributions *a posteriori* des valeurs les plus probables des paramètres qui sont obtenues. Une fois cet échantillonnage établi, l'incertitude liée à la non-connaissance absolue de la valeur du paramètre peut être étudiée en analysant son impact sur les simulations. Par ailleurs, une optimisation paramétrique peut également avoir lieu en calculant par exemple la valeur moyenne ou modale de la distribution *a posteriori*. En parallèle, une nouvelle fonction de vraisemblance basée sur un emploi explicite du coefficient de variation (CV) a été développée. L'emploi du coefficient de variation au sein du modèle formel de représentation statistique des résidus offre l'avantage de pouvoir prendre en compte la variabilité génétique propre à la culture étudiée et l'hétéroscédasticité qui y est associée et ce, à chaque étape du processus d'échantillonnage. De plus, en augmentant la valeur du coefficient de variation au-delà des bornes représentatives de la variabilité naturelle seule, la fonction de vraisemblance CV permet d'étudier l'effet d'un bruit croissant sur les mesures. En outre, il a été démontré que cette fonction permettait, au même titre que la fonction WLS (Weigthed Least Square), et contrairement à la SLS (Standard Least Square), de tenir compte de la réalité physique des paramètres. Il semble ainsi que la fonction SLS qui n'intègre pas le bruit sur les mesures conduit à une forme de sur-paramétrisation du modèle, donnant lieu à de mauvaises estimations des faibles valeurs de biomasse. Il a également été mis en évidence qu'un bruit associé aux mesures et inférieur à 50 % ne menait pas nécessairement à un mauvais estimateur paramétrique moyen mais que l'incertitude répercutée sur les simulations était croissante avec la valeur du bruit. Ainsi, la nouvelle fonction de

vraisemblance CV permet à la connaissance de l'expert de s'exprimer non seulement au niveau de la définition de la connaissance *a priori* sur les paramètres, mais aussi à chaque boucle du processus d'échantillonnage. Cette première étape de la recherche nous a ainsi permis d'aboutir à un modèle entièrement calibré, et ce à tous les niveaux des variables de sortie simulées (LAI, biomasse totale, rendement en grain, exportation et prélèvement d'azote).

La suite de nos recherches s'est portée sur les méthodes de filtrage reposant sur des approches Bayésiennes. Différentes méthodes, telles que le filtre de Kalman étendu (EKF), le filtre particulaire (PF) ou le filtre variationnel (VF) ont été considérées en vue de résoudre le problème d'estimation concomitante du LAI (indice de développement foliaire), celui des teneurs en eau des deux premiers horizons de sol et de ré-estimation éventuelle simultanée de différents paramètres. A partir de séries temporelles synthétiques de ces variables, les différentes techniques ont été évaluées sur base de leur capacité à optimiser des paramètres du système et des valeurs de RMSE obtenues entre les sorties filtrées et le signal original (non bruité). Au vu du caractère non-linéaire des modèles, le filtre EKF a montré de médiocres performances. Le filtre PF, qui approxime la distribution de probabilité au moyen d'un set d'échantillons pondérés, a montré des résultats intermédiaires. Toutefois, ce type de filtre est extrêmement gourmand en temps de calcul et en capacité de stockage mémoire. Le filtre VF s'est montré le plus efficace, à la fois en termes de RMSE et de convergence. Cependant, si elles peuvent se révéler particulièrement utiles pour améliorer les résultats des simulations d'un modèle, les méthodes de filtrage ne peuvent se substituer complètement à un processus d'estimation paramétrique qui permet de calibrer le modèle de façon robuste et générique. Elles s'avèrent extrêmement intéressantes lorsqu'il s'agit de pallier des conditions de croissance non prises en compte par le modèle, comme la chute du rendement due à une maladie ou une verse. Dans de telles situations, la réalisation de mesures acquises en cours de saison couplée à l'application des méthodes d'assimilation de données permet d'améliorer, parfois largement, les simulations. Par ailleurs, les techniques d'assimilation de données devraient se révéler particulièrement utiles lors de la mise au point d'outils fonctionnels afin d'adapter le modèle de culture à des contextes différents.

Une fois calibré de façon robuste et générique, le modèle peut être utilisé pour ses capacités prédictives. Nous avons montré comment il était possible de développer un modèle prévisionnel reposant sur une hypothèse simple mais justifiée dans la pratique. La méthodologie utilisée consiste à construire des '*ensembles de matrice*' climatiques évoluant au cours de la saison, constitués d'une part de données climatiques réelles mesurées au fur et à mesure de la croissance de la culture, et d'autre part de données climatiques moyennes calculées au départ d'une base de données historique locale. Au fur et à mesure de l'avancement de la saison, la part des premières données devient prépondérante par rapport à celle des secondes. L'hypothèse d'un climat moyen en guise de projection pour le futur peut être considérée comme une hypothèse forte. En effet, le climat ainsi utilisé dans les simulations est un climat qui n'induit que peu de stress dans la mesure où, chaque jour, les besoins de la plante sont assurés par des conditions minimales au niveau de chacune des variables climatiques (e.g. ensoleillement régulier, pluie quotidienne). Toutefois, l'originalité de l'approche dite '*climat moyen*' réside dans la possibilité de simuler des valeurs de rendements représentatives du potentiel restant à la culture, à tout moment de la saison, *i.e.* si les conditions climatiques consécutives à un stress devaient à l'avenir se rapprocher à nouveau des normales saisonnières. Il a été démontré que cette méthodologie permet (*i*) de mettre en évidence les périodes les plus sensibles à l'induction ou l'apparition de stress,

(ii) de quantifier instantanément les pertes de rendement potentiel dues aux conditions climatiques des 10 derniers jours et (iii) de prédire les rendements environ un mois avant la date de moisson avec un intervalle de confiance de 10%. Par ailleurs, cette approche constitue un outil de diagnostic puissant des potentialités d'une culture donnée (pour laquelle le modèle a été calibré) sur un site donné, c'est-à-dire où le sol est connu et où les données climatiques sont disponibles sur plusieurs décennies. Mais plus que tout, cette méthodologie présente un réel potentiel pour être interfacée avec un système de mesure en temps réel des conditions climatiques, ce qui en fait un outil précieux pour le développement de systèmes d'aide à la décision.

Finalement la question s'est posée de connaître l'impact de l'hypothèse climat moyen sur la qualité prédictive de la méthodologie qui en découle. Lors de la mise au point d'outils d'aide à la décision, le développeur fait face à deux comportements antagonistes liés au temps de calcul. D'une part, il est primordial de développer des outils aussi rapides que possible afin de fournir une réponse immédiate lorsqu'une décision doit être prise. D'autre part, pour quantifier le degré de confiance dans la décision, il est important d'associer à la réponse un intervalle de confiance. Au vu des nombreuses interactions existant entre les formalismes du modèle, le calcul d'un tel indice repose le plus souvent sur une approche de multi-simulations, accroissant largement le temps de calcul. Il a donc été décidé de comparer l'approche climat moyen à une approche probabiliste des entrées climatiques. D'une manière plus générale, la question s'est posée de savoir si un comportement global sous-jacent du modèle pouvait être mis en évidence et quelle serait la place occupée par une hypothèse de type climat moyen au sein de ce comportement. La notion de comportement sous-jacent fait ici référence à la nature mathématique du modèle. En effet, contrairement à la réalité, où des phénomènes complexes, dont nous ne percevons pas toujours la nature ou l'impact, peuvent se produire, les modèles de culture sont des représentations imparfaites de la réalité, mais constituent des entités mathématiques entièrement déterminées. Il devrait donc être possible, sur base de la statistique, d'en dégager un certain comportement.

L'approche climat moyen a donc été comparée à une approche probabiliste. Cette dernière, prouvée efficace pour prédire les rendements d'une culture en temps réel, repose sur la génération d'un grand nombre de réalisations climatiques stochastiques à l'aide du modèle LARS-WG (Weather Generator). En se basant sur le théorème de la convergence en loi et le théorème central limite généralisé, une méthodologie qui se présente comme un outil généralisable à l'étude du comportement des modèles a été développée. Il a ainsi été démontré que, pour autant que le modèle prenne en compte un nombre suffisant de situations climatiques représentatives d'une région donnée, les distributions des rendements obtenues peuvent être considérées comme répondant à une loi de type log-normale. Par ailleurs, il a été démontré que les différences entre les simulations conduites avec les approches climat moyen et probabiliste présentaient des critères statistiques (RRMSE et déviation normalisée) inférieurs à 10%. Les délais nécessaires à l'obtention de prédictions fiables et obtenus avec les deux méthodes sont similaires. Finalement, le théorème de la convergence en loi s'est trouvé validé dans 90% des situations climatiques. Ceci permet d'affirmer que, dans ces situations, la moyenne des simulations obtenues sur une importante base de données climatiques est égale à la simulation obtenue en utilisant la moyenne journalière de cette base de données. Il semble donc raisonnable d'affirmer que les deux approches peuvent se substituer l'une à l'autre dans une optique de prédiction des rendements. La distinction se trouve surtout au niveau de la finalité, l'une permettant de simuler rapidement le

potentiel restant à la culture, l'autre offrant une gamme d'incertitudes sur cette prédiction. Toutefois, l'approche proposée reposant sur les normales saisonnières nécessite autant de fois moins de temps de calcul que le nombre de climats utilisés dans une approche stochastique (300 dans ce cas-ci). Signalons cependant que cette analyse a été réalisée sur base d'une expérimentation numérique où le niveau de fertilisant azoté était maintenu suffisant pour éviter toute carence nutritionnelle et n'étudier que l'effet de la variabilité climatique. Or, rien ne prouve que le comportement mis en évidence précédemment soit maintenu sous des niveaux d'intrants azotés limités.

Dans l'optique de la mise au point d'un outil d'aide à la décision visant une gestion de l'apport de fertilisants azotés, la suite de l'analyse s'est donc intéressée au couplage de l'approche probabiliste climatique avec différentes modalités de gestion de l'azote en cours de saison, afin de déterminer les apports optimaux. Il est important de signaler qu'il s'agit de ce fait d'une gestion stratégique de l'azote, c'est-à-dire une gestion sans connaissance *a priori* des conditions climatiques à venir : toutes les simulations sont réalisées sur des sets climatiques où chaque climat synthétique est utilisé sur l'entièreté de la saison de culture, depuis le semis jusqu'à la récolte. Par ailleurs, l'approche probabiliste maintenue à ce stade semble opportune afin de laisser à l'utilisateur final la capacité de gérer lui-même le risque tolérable sur la décision à prendre. Il a ainsi été démontré que les distributions des variables de sortie du modèle représentant la biomasse ou le rendement en grains peuvent à tout moment être représentées par la distribution théorique de type I du système de Pearson, et ce quel que soit le niveau de fertilisant azoté. Etant physiquement limité à ses deux extrema et reposant sur des paramètres capables de tenir compte d'un degré d'asymétrie et d'aplatissement, cette distribution est en effet particulièrement adaptée à la représentation de la croissance telle que simulée par les modèles de culture. L'ajustement quasi parfait de la distribution de type-I à tous les cas de gestion de l'azote envisagés laisse donc suggérer que les quatre grandeurs caractéristiques sur lesquelles elle repose, à savoir la moyenne, l'écart-type, l'asymétrie (skewness) et l'aplatissement (kurtosis), sont suffisantes pour caractériser toute distribution simulée.

L'étude s'est alors penchée sur l'analyse indépendante du degré d'asymétrie des distributions simulées. Lorsque l'azote est fertilisé en trois fractions égales, apportées respectivement aux stades tallage, redressement et dernière feuille, les résultats ont démontré que le niveau de dissymétrie le plus fort était obtenu pour la pratique 60-60-60 kgN.ha⁻¹. Cela signifie qu'à ce niveau de fertilisation, la probabilité pour l'agriculteur d'obtenir un rendement au moins supérieur à la moyenne de la distribution est la plus grande. En d'autres mots, le risque d'obtenir de faibles rendements est minimal. Avec un accroissement des apports d'azote, il est généralement admis que les rendements ont tendance à s'accroître jusqu'à atteindre un plateau. En complément à cette observation, notre étude révèle que, sous un régime climatique probabiliste, avec l'augmentation de la dose de fertilisants, les rendements élevés, c'est-à-dire supérieurs à la moyenne, deviennent aussi plus fréquents. Cette méthodologie d'analyse a ensuite été étendue à l'inter-comparaison des distributions des rendements obtenus sous des pratiques de fertilisation en deux et trois fractions. Il est à noter que lors d'une fertilisation en deux fractions, la première dose est apportée à un stade intermédiaire entre le tallage et le redressement (plus ou moins 15 jours après le tallage) et à nouveau à la dernière feuille. A même dose totale et pour une dose totale inférieure à 120 kgN.ha⁻¹, il y a absence de différence significative entre les distributions des rendements obtenus sous des fractionnements en deux et trois apports (e.g. 30-30 et 20-20-20 kgN.ha⁻¹). Au-delà de cette valeur seuil de 120 kgN.ha⁻¹, 30 kgN.ha⁻¹ peuvent

systématiquement être économisés en fertilisant en trois fractions, et de ce fait en suivant au plus près la phénologie de la plante. Finalement, notre étude a été étendue à l'analyse des distributions du revenu marginal net correspondant aux rendements observés. Il a été démontré que, 75% du temps, un niveau de fertilisation correspondant à la fumure de référence (60-60-60 kgN.ha⁻¹) permet d'obtenir le revenu marginal net maximal. Toutefois, 90 % du temps, soit 9 années sur 10, l'application de 20 kgN.ha⁻¹ à la dernière feuille s'avère suffisante pour maximiser le revenu marginal net de l'agriculteur. D'une manière générale, la méthodologie, générale et générique (non inféodée au modèle STICS en particulier) proposée dans le cadre de cette thèse permet donc à l'utilisateur final de gérer lui-même le risque qu'il prend pour les différentes pratiques envisagées.

Pour conclure, il est important de rappeler que les apports de la présente thèse se situent à différents niveaux. Tout d'abord, et pour la première fois, le modèle de culture STICS a été couplé à un algorithme d'échantillonnage paramétrique Bayésien reposant sur une représentation formelle des résidus de simulation. Les avantages de la méthode employée sont nombreux, et permettent de caractériser aussi bien l'incertitude sur les paramètres du modèle que celle répercutée sur les simulations. Dans un second temps, cette thèse a exploré le niveau d'incertitude associé à l'emploi d'un modèle de culture à des fins réellement prédictives, c'est-à-dire en cherchant à suppléer le futur inconnu par différentes hypothèses posées sur les entrées climatiques du modèle. Finalement, en couplant des approches probabilistes sur les entrées climatiques et en explorant différentes pratiques de gestion de la fertilisation azotée, cette thèse a démontré qu'il était possible de mettre au point un outil d'aide à la décision effectif et efficace visant une gestion stratégique de l'azote.

2. Enjeux agronomiques et environnementaux

Les différents outils présentés, qu'il s'agisse de la prédiction du rendement ou de la gestion de l'azote, imposent clairement une vision et une utilisation du modèle de culture à des fins de production. Cette section de l'étude vise à étendre la discussion en lien avec l'enjeu environnemental.

Les différents enjeux agronomiques présentent souvent une synergie positive, *e.g.* l'apport d'azote à une culture céréalière permet l'augmentation du rendement mais aussi celle de la teneur en protéines des grains. A l'opposé, les enjeux environnementaux peuvent être antagonistes par rapport aux enjeux agronomiques. Ainsi, un apport d'azote supérieur aux besoins stricts ou aux possibilités d'assimilation de la culture donne lieu à un reliquat azoté important, accroissant le risque de pollution. A l'heure où les enjeux écologiques sont plus que jamais au cœur des priorités, tout l'art du modélisateur consiste, sur base des variables de sorties simulées par le modèle, à déterminer le critère le plus approprié, *i.e.* celui qui présente le juste compromis entre rentabilité et respect de l'environnement.

Un premier point de discussion porte sur l'implémentation de la fertilisation azotée au sein des modèles de culture. Depuis la directive Européenne 91/676/EEC sur la réduction des pollutions par les nitrates, le cycle de l'azote en est un des éléments clefs. Toutefois, on peut se demander si les formalismes liés aux transformations des stocks d'azote dans le sol sont suffisamment aboutis. Ainsi, à l'heure actuelle, aucun modèle ne prend en compte l'impact de l'azote sur la biodiversité, en particulier celle du sol.

Le second point qu'il convient de discuter porte sur la définition d'un critère conciliant enjeux

agronomiques et environnementaux. L'environnement est avant tout déterminé par sa qualité, ce qu'il convient de traduire sous une forme quantitative modélisable. Le reliquat azoté après la récolte est une première forme de quantification, traduisant la sensibilité du système cultural à induire un risque de pollution azoté. Mais le fait même de traduire un critère qualitatif en une donnée quantitative implique une vision productiviste du système. Or la question est posée, peut-on se permettre de faire des compromis sur l'environnement?

Comme cela a été rappelé précédemment, le modèle reste une représentation imparfaite et simplifiée de la réalité. Le modèle représente la réalité telle qu'elle a été mesurée, ce qui implique une délimitation de celle-ci aux conditions prévalant au moment de l'observation. Lors de l'extrapolation à d'autres conditions agro-pédo-climatiques, les simulations des modèles de culture sont ainsi sujettes à des incertitudes et les impacts environnementaux n'échappent pas à cette règle.

Partant des définitions de Hénin et Sebillotte, l'agronomie pourrait être définie comme l' "*étude, menée simultanément dans le temps et l'espace, des relations régissant le système éco-physiologique que constitue le champ cultivé, y incluant le peuplement végétal, le sol en et l'environnement en tant que milieux physiques, chimiques et biologiques, sur lequel l'homme peut appliquer un raisonnement mais également une action afin d'en obtenir une production*". Cette définition ramène le raisonnement et l'action de l'agronome de terrain au sein d'un système plus vaste. C'est pourquoi la gestion, et *a fortiori* l'optimisation de la fertilisation azotée, devraient avant tout être envisagées dans des approches systémiques. Pour raisonner une technique, il est nécessaire de prendre en compte l'ensemble du système et non pas seulement ce qui paraît *a priori* être directement lié à la technique. Or, le modèle de culture en tant qu'entité mathématique et représentation de la réalité n'offre pas pour l'heure cette vision d'ensemble. Faut-il par là conclure que les développements et les résultats proposés dans cette thèse soient caducs? Certainement pas ! Au sein des recherches conduites dans cet ouvrage, un caractère générique et non inféodé à un modèle a été conservé autant que possible lors de la mise au point des différentes méthodologies proposées. Le paradigme que constitue le modèle de culture peut ainsi évoluer vers la représentation de systèmes plus complexes sans que la méthodologie d'analyse ne soit nécessairement à revoir. Les approches ainsi proposées pourraient être transférées à l'étude de systèmes modélisés plus globaux, régis par des interactions plus complexes, dont les enjeux agronomiques et environnementaux mériteraient d'être envisagés par des raisonnements systémiques multifactoriels.

A l'heure où l'agronomie prend la direction de l'*écologiquement intensif*, les modèles de culture ont encore de beaux jours devant eux. L'optimisation de la gestion azotée devrait s'adapter pour prendre en compte les bénéfices de l'association culturale, étudier l'impact de pratiques (*i.a.* date de semis, densité de semis) ou d'environnements en constante évolution (*i.a.* changement climatique), ou encore favoriser la sélection de variétés mieux adaptées via l'*ideotyping* (terme anglais pouvant être défini ici comme la sélection de caractères génétiques d'intérêt sur base d'observations phénotypiques dans des environnements contrastés). Ainsi, l'utilisation des modèles de culture pour accroître les synergies culturales et les synergies cultures-environnement est une voie prometteuse.

3. Perspectives

Les résultats proposés dans le cadre de cette thèse restent liés à un type de sol (sol limoneux de

Hesbaye), à un climat tempéré et à un génotype donné (*Triticum aestivum* L., cultivar Julius et apparentés). Toutefois, *i*) au vu des choix posés lors de la détermination du cas d'étude (type de sol, choix du cultivar), *ii*) compte tenu de la mise en place du protocole expérimental qui inclut plusieurs saisons de culture et niveaux de fertilisation, *iii*) considérant la robustesse du processus de calibration et *iv*) étant donné la généricité des méthodologies de prédiction des rendements et d'optimisation des apports azotés, les résultats présentés sont vraisemblablement généralisables, si ce n'est de manière absolue, au moins dans leurs tendances. Sur base de ce constat, cette section de l'étude vise à mettre en évidence différentes perspectives de recherches.

Tout d'abord, différentes sources d'incertitudes ont été étudiées et quantifiées. Elles ont toutes fait l'objet d'une analyse indépendante. D'une façon générale, il serait intéressant de mettre au point une méthodologie permettant de coupler ces différentes sources d'incertitudes pour analyser de façon fondamentale les interactions qui existent entre elles.

Dans un contexte d'agriculture de précision, il faudrait se pencher sur l'étude des variations intra-parcellaires et leur impact dans le développement d'outils fonctionnels à l'échelle du champ. Une approche intéressante consisterait à caractériser finement les profils pédologiques concernés par le modèle. Sur base de ces données, les résultats obtenus par multi-simulation pourraient être comparés à des mesures acquises à l'aide d'une moissonneuse-batteuse expérimentale géo-référençant les rendements. D'un autre côté, il serait intéressant de réaliser une simulation à l'échelle de la parcelle agricole sur base d'une paramétrisation du sol reposant sur des données moyennes. Les deux approches pourraient être comparées afin de déterminer la plus pertinente en fonction de la variabilité spatiale du sol et de la finalité de l'outil, à savoir prédiction du rendement ou optimisation de la gestion azotée.

Tout cela conduit finalement aux questions ayant trait à l'utilisation opérationnelle du modèle dans des contextes agropédoclimatiques différents ou amenés à changer (*e.g.* émergence de nouvelles variétés ou impacts du changement climatique). Cela implique que la validité du modèle soit régulièrement assurée par la prise en compte de nouvelles observations et mesures. Des procédures standardisées d'acquisition devraient être définies pour faciliter le caractère opérationnel de l'outil. Elles trouveraient d'ailleurs un avantage supplémentaire dans les techniques d'assimilation de données. En ce qui concerne ces dernières, la question est alors posée de savoir s'il faut maintenir des modèles aussi complexes que ceux qui sont utilisés actuellement? Le problème est d'assurer un compromis entre la capacité du modèle à travailler en temps réel à de larges échelles, tout en anticipant autant que possible les conditions climatiques qui pourraient se produire jusqu'à la moisson.

L'amélioration de la gestion azotée a été discutée sur base de critères objectifs, à savoir la forme et la gamme de valeurs couvertes par la distribution de probabilité d'une variable d'intérêt. Deux cas ont été traités, à savoir l'impact de différentes pratiques azotées sur d'une part les distributions des rendements et d'autre part sur le revenu marginal net. En particulier, les revenus marginaux nets sont apparus être des fonctions à maxima, pour lesquelles une pratique optimale pouvait être associée à différents temps de retour d'évènements climatiques. Mais l'approche n'est pas limitée à l'étude des distributions de revenus ou de rendement, ni restrictive dans le critère d'optimisation. En effet, l'intérêt des méthodologies proposées repose aussi sur l'étude du risque. On pourrait utiliser ces approches dans l'amélioration du système d'assurances récoltes/revenus via la définition de barèmes plus équitables pour l'agriculteur et/ou plus adaptés aux conditions spécifiques de régions données. Une autre

perspective consisterait à définir des critères tenant compte de considérations environnementales, comme le suivi des reliquats d'azote dans le sol après récolte, connus sous le nom d'azote potentiellement lessivable (APL).

On pourrait envisager de déterminer d'autres critères à satisfaire, comme la recherche d'une variabilité minimale des distributions des rendements. Dans ce cas, la seule étude des rendements annuels serait insuffisante et il conviendrait d'intégrer l'effets de différentes pratiques culturales comme le choix variétal, la date et densité de semis, l'influence du précédent cultural, la présence de cultures intermédiaires pièges à nitrates (CIPAN),

Les résultats obtenus, à savoir la quantification du niveau d'incertitude associé à différentes modalités d'apports de fertilisants azotés, font suite à des simulations visant à optimiser les pratiques de fertilisation azotée d'un point de vue *stratégique*, c'est-à-dire sans connaissance *a priori* du climat à venir. Cette méthodologie mériterait d'être couplée aux algorithmes de prédiction du rendement en temps réel développés. Ce couplage aboutirait à un outil d'aide à la décision visant cette fois-ci une approche *tactique* de la gestion azotée, c'est-à-dire appliquée, au cas par cas, en temps réel et en cours de saison. Cela permettrait de se rapprocher davantage de la réalité de terrain et de répondre aussi près que possible à la demande de la culture, tout en minimisant les pertes financières et les impacts environnementaux.

D'une manière générale, les recherches devraient être poursuivies en étudiant de façon plus fondamentale et plus systématique différentes situations agro-environnementales, comme l'effet de la disponibilité naturelle de l'azote, des précédents culturaux, des types de sol, etc. En particulier, il serait intéressant de développer l'étude de l'interaction Genotype × Environnement × Pratiques culturales pour garantir la sécurité alimentaire dans un contexte de changement climatique.

Références bibliographiques

"Adapte-toi au monde, car ta tête est trop petite pour que le monde s'y adapte."

"Tout apprendre, non point pour l'afficher, mais pour s'en servir."

Georg Christoph Lichtenberg

Physicien et écrivain allemand - 1742-1799

Chapitre 10 : Références bibliographiques générales

Aidala, V., 1977. Parameter estimation via the Kalman filter. *IEEE Transactions on Automatic Control* 22, 471–472.

Allen, R., Hanuschak, G. A., & Craig, M. E. (1994). Forecasting crop acreages and yield in the face of and in spite of floods, *Crop Yield Forecasting Methods. Proceedings of the Seminar Villefranche- sur-Mer: 24-27 October, Villefranche- sur-Mer*, 87-110.

Andrews, B., Yi, T., Iglesias, P., 2006. Optimal noise filtering in the chemotactic response of *Escherichia coli*. *PLoS Computational Biology* 2, e154.

Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50, 174–188.

Arslan, S., Colvin, T., 2002. Grain Yield Mapping: Yield Sensing, Yield Reconstruction, and Errors. *Precision Agriculture*, 3(2), 135-154

Balaji, B., Friston, K., 2011. Bayesian state estimation using generalized coordinates. *Proceedings of SPIE* 8050.

Baret, F., 1986. Contributions au suivi radiométrique de cultures de céréales, Ph.D. Thesis. Univ. Paris-Sud, Orsay, France, 182 p.

Barndorff-Nielsen, O., 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London* 353, 401–419.

Basso B., Ritchie J.T., 2005. Impact of animal manure, compost and inorganic fertilizer on nitrate leaching and yield in a six-year maize alfalfa rotation. *Agri. Ecosys. Environ.*, 108, 329 341.

Basso, B., Bertocco, M., Sartori, L., & Martin, E. C., 2007. Analyzing the effects of climate variability on spatial pattern of yield in a maize–wheat–soybean rotation. *European Journal of Agronomy*, 26(2), 82-91, doi:10.1016/j.eja.2006.08.008.

Basso, B., Cammarano, D., Chen, D., Cafiero, G., Amato, M., Bitella, G., Rossi, R., Basso, F., 2009. Landscape position and precipitation effects on spatial variability of wheat yield and grain protein in Southern Italy. *J. Agron. Crop Sci.*, 195, 301–312.

Basso B., Cammarano D., Troccoli A, Chen D., Ritchie J.T., 2010. Long-term wheat response to nitrogen in a rainfed Mediterranean environment: Field data and simulation analysis. *Eur. J. Agron.*, 33, 132-138.

Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L., 2011. A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, 35(4), 215-222, doi:10.1016/j.eja.2011.06.004.

Basso, B., Fiorentino, C., Cammarano, D., Cafiero, G., & Dardanelli, J., 2012. Analysis of rainfall distribution on spatial and temporal patterns of wheat yield in Mediterranean environment. *European Journal of Agronomy*, 41(0), 52-65, doi:10.1016/j.eja.2012.03.007.

Beal, M., 2003. Variational Algorithms for Approximate Bayesian Inference, Ph.D. Thesis, Gatsby Computational Neuroscience Unit, University College London.

Beaudoin N., Nicoulaud B., Houles, V., 2007. Etablissement et validation de classes de pédotransfert pour un modèle de culture à l'échelle parcellaire : application au modèle STICS. In : Guérif M. and King D. (Eds.). *Agriculture de précision*, Editions Quae, Collection Update Sciences

and Technologies, 25-42.

Beaudoin N., Launay M., Sauboua E., Ponsardin G. & Mary B., 2008. Evaluation of the soil crop model STICS over 8 years against the “on farm” database of Bruyères catchment. *Eur. J. Agron.*, 29, 46-57.

Bechini L., Bocchi S., Maggiore T., Confalonieri R., 2006. Parameterization of a crop growth and development simulation model at sub model component level. An example for winter wheat (*Triticum aestivum* L.). *Environ. Modell. Softw.*, 21, 1042-1054.

Bellocchi, G., Rivington, M., Donatelli, M., & Matthews, K., 2010. Validation of biophysical models: issues and methodologies. A review. *Agronomy for Sustainable Development*, 30(1), 109-130.

Beven, K., 1989. Changing ideas in hydrology. The case of physically based models. *J. Hydrol.* 105, 157-172.

Beven, K., 2008. Comment on “Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?” by Vrugt JA, ter Braak CJF, Gupta HV, Robinson BA, 2008. *Stoch. Environ. Res. Risk Assess.*

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.* 6, 279-298.

Beven, K., Smith, P., Freer, J., 2008. So just why would a modeller choose to be incoherent? *J. Hydrol.*, 354, 15-32.

Boote, K.J., J.W. Jones, G. Hoogenboom, and N.B. Pickering., 1998. The CROPGRO model for grain legumes. p. 99–128. In G.Y. Tsuji et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publ., Boston.

Box, G., Cox, D., 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26(2), 211-252.

Box, G., Tiao, G., 1973. *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, PA.

Braga, R.P., Jones, J.W., 2004. Using Optimization to Estimate Soil Inputs of Crop Models for Use in Site-Specific Management. *Transactions of the ASAE* 47 1821- 1831

Brisson N., Mary B., Ripoche D., Jeuffroy M.-H., Ruget F., Nicoulaud 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., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. Theory and parameterization applied to wheat and corn. *Agronomie*, 18, 311-346.

Brisson N., Ruget F., Gate P., Lorgeau J., Nicoulaud B., Tayo X., Plenet D., Jeuffroy M.H., Bouthier A., Ripoche D., Mary ., Justes E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie*, 22, 69-82.

Brisson N., C. Gara, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussièrre, Y. M. Cabidoche, P. Cellier, P. Debaeke, J. P. Gaudillère, C. Hénault, F. Maraux, B. Seguin and Sinoquet H., 2003. An overview of the crop model STICS. *Eur. J. Agron.*, 18, 309-332.

Brisson N., Launay M., Mary B., Beaudoin N., editors, 2008. Conceptual basis, formalisations and parameterization of the STICS crop model. Editions Quae, Collection Update Sciences and Technologies, 297 p.

Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2009. Conceptual basis, formalisations and

parameterization of the STICS crop model. Editions Quae, Collection Update Sciences and Technologies, p. 297.

Brooks, S.P., 1998. Markov chain Monte Carlo and its application, *The Statistician*. J. Royal Statist. Soc. Series D.

Campbell, G. S., & Norman, J. M., 1989. The description and measurement of plant canopy structure. In G. Russell, B. Marshall, & P. G. Jarvis (Eds.), *Plant Canopies : their Growth, Form and Function*: Cambridge University Press.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environ. Model. Softw.* 22, 1509-1518.

Cantelaube, P., & Terres, J. M., 2005. Seasonal weather forecasts for crop yield modelling in Europe. *Tellus*, 57(3), 476-487, doi:10.1111/j.1600-0870.2005.00125.x.

Cantelaube, P., & Terres, J. M., 2003. Use of Seasonal Forecasts on Crop Yield Modelling. Publication of the European Commission, EUR 20996 EN.

Challinor, A. J., Slingo, J. M., Wheeler, T. R., & Doblas-Reyes, F. J., 2005. Probabilistic simulations of crop yield over Western India using the DEMETER seasonal hindcast ensembles. *Tellus*, 57(3), 498-512, doi:10.1111/j.1600-0870.2005.00126.x.

Chen, G., Xie, Q., Shieh, L., 1998. Fuzzy Kalman filtering. *Journal of Information Science* 109, 197-209.

Confalonieri, R., Bechini, L., 2004. A preliminary evaluation of the simulation model CropSyst for alfalfa. *Eur. J. Agron.* 21, 223-237.

Corduneanu, A., Bishop, C., 2001. Variational bayesian model selection for mixture distribution. In: *Artificial Intelligence and Statistics*.

Day, R.H., 1965. Probability Distributions of Field Crop Yields. *J. Farm Econ.* 47, 713-741.

de Moivre, A., 1756. *The Doctrine of Chances*, 3d ed. (London: Millar, 1756).

De Wit A.J.W., Van Diepen C.A., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agric. For. Meteorol.*, 146, 38-56.

Diepen, C., Wolf, J., Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. *Soil Use and Management* 5, 16-24.

Dietzel, A., Reichert, P., 2012. Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model. *Environ. Model. Softw.* 38(0) 129-146.

Doucet, A., Johansen, A., 2009. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 656-704.

Doucet, A., Tadic, V., 2003. Parameter estimation in general state-space models using particle methods. *Annals of the Institute of Statistical Mathematics* 55, 409-422.

Du, X., Hennessy, D., Yu, C., 2012. Testing Day's Conjecture that More Nitrogen Decreases Crop Yield Skewness. *Am. J. Agric. Econ.* 94, 225-237.

Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J-P., & Destain, M-F., 2014a. Parameter identification of the STICS crop model, with an accelerated formal MCMC approach. *Environmental Modelling and Software*, 52, 121-135.

Dumont, B., Leemans, V., Ferrandis, S., Vancutsem, F., Bodson, B., Destain, J., Destain, M., 2014b. Assessing the potential of predicting wheat yield using a daily mean climate database. *Precision agriculture (In press)*. <http://link.springer.com/article/10.1007/s11119-014-9346-9>.

EC-Council Directive, 1991. Council Directive 91/676/EEC concerning the protection of waters against pollution caused by nitrates from agricultural sources.

Ewert, F., van Oijen, M., Porter, J.R., 1999. Simulation of growth and development processes of spring wheat in response to CO₂ and ozone for different sites and years in Europe using mechanistic crop simulation models. *European Journal of Agronomy* 10 231-247.

Ewert, F., van Ittersum, M. K., Heckelei, T., Therond, O., Bezlepkina, I., & Andersen, E., 2011. Scale changes and model linking methods for integrated assessment of agri-environmental systems. *Agriculture, Ecosystems & Environment*, 142(1–2), 6-17, doi:10.1016/j.agee.2011.05.016.

Flenet F., Villon P., Ruget F., 2004. Methodology of adaptation of the STICS model to a new crop: spring linseed (*Linum usitatissimum*, L.). *Agronomie*, 24: 367-381.

Gelman, A., Carlin, J.B., Stren, H.S., Rubin, D.B., 1997. *Bayesian Data Analysis*. Chapman and Hall, London, p. 526.

Génard M., Jeuffroy MH., Jullien A., Quilot B., 2006. Les modèles en écophysologie pour l'action en agronomie et la création variétale. *L'agronomie aujourd'hui*, Editions Quae, ISBN 10 : 2-7592-0000-0

Germon J.C., Heanult C., 2003. Surfaces agricoles, bilan de gaz à effet de serre et climat. In *Stocker du carbone dans les sols agricoles en France? Arrouays D. et al., (Eds.). Rapport d'expertise MEDD*, 76-88.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Introducing Markov Chain Monte Carlo*, in *Markov Chain Monte Carlo in Practice*: Chapman & Hall, London, pp. 1-19.

Grewal, M., Andrews, A., 2008. *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley and Sons.

Guérif, M., Beaudoin, N., Dürr, C., Houlès, V., Machet, J.M., Mary, B., Moulin, S., Richard, G., Bruchou, C., Michot, D. ; Nicoullaud, B., 2001. Designing a field experiment for assessing soil and crop spatial variability and defining site-specific management strategies. *Proceedings of the 3re European conference on Precision Agriculture, Montpellier, June*, p677-682

Guillaume, S., Bergez, J.E., Wallach, D., Justes, E., 2011. Methodological comparison of calibration procedures for durum wheat parameters in the STICS model. *Eur. J. Agron.* 35, 115-126.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., Nordlund, P., 2002. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing* 50, 425–437.

Hansen S., Jensen H.E., Nielsen N.E., Swenden H., 1990. Daisy, a soil plant system model. Danish simulation model for transformation and transport of energy and matter in the soil plant atmosphere system. *The National Agency for Environment Protection, Copenhagen*, pp. 369.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97-109.

Hennessy, D.A., 2009a. Crop Yield Skewness and the Normal Distribution. *J. Agric. Resour. Econ.* 34, 34-52.

Hennessy, D.A., 2009b. Crop Yield Skewness Under Law of the Minimum Technology. *Am. J. Agric. Econ.* 91, 197-208.

Hennessy, D.A., 2011. Modeling Stochastic Crop Yield Expectations with a Limiting Beta Distribution. *Journal of Agricultural and Resource Economics* 36, 177-191.

Hewitt, C. D., 2004. Ensembles-based predictions of climate changes and their impacts. *Eos*

Trans. AGU, 85(52), doi:10.1029/2004eo520005.

Houlès, V., Mary, B., Guérif, M., Makowski, D., Justes, E., 2004. Evaluation of the ability of the crop model STICS to recommend nitrogen fertilisation rates to agro-environmental criteria. *Agronomie*, 24, 339–349.

Jamieson, P. D., Semenov, M. A., Brooking, I. R., & Francis, G. S., 1998. Sirius: a mechanistic model of wheat response to environmental variation. *European Journal of Agronomy*, 8(3 4), 161-179, doi:10.1016/s1161-0301(98)00020-3.

Jansen, M., Hagenaars, T., 2004. Calibration in a Bayesian modelling framework, In: van Boekel, M.A.J.S., Stein, A., van Bruggen, A.H.C. (Ed.), *Bayesian Statistics and Quality Modelling in the Agro-food Production Chain*: Kluwer Academic, Dordrecht.

Jeremiah, E., Sisson, S.A., Sharma, A., Marshall, L., 2012. Efficient hydrological model parameter optimization with Sequential Monte Carlo sampling. *Environ. Model. Softw.* 38(0) 283-295.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology* 121, 19–35.

Jones C.A., Kiniry J.R., 1986. *CERES-Maize*, a simulation model of maize growth and development, Texas A&M University Press, College Station, TX.

Jongschaap R.E.E., 2007. Sensitivity of a crop growth simulation model to variation in LAI and canopy nitrogen used for run-time calibration. *Ecol. Modell.*, 200, 89-98.

Joseph, J.F., Guillaume, J.H.A., 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environ. Model. Softw.* 46(0) 292-298.

Julier, S., Uhlmann, J., 1997. New extension of the Kalman filter to nonlinear systems. *Proceedings of SPIE* 3, 182–193.

Just, R.E., Weninger, Q., 1999. Are Crop Yields Normally Distributed? *Am. J. Agric. Econ.* 81, 287-304.

Justes E., Mary B., Nicolardot B., 2009. Quantifying and modeling C and N mineralization kinetics of catch crop residues in soil : parametrization of the residue decomposition module of STICS model for mature and non mature residues. *Plant soil* 325, 171-185.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problem. *Journal of Basic Engineering, Transactions ASME, Series D* 82, 34–45.

Kim, Y., Sul, S., Park, M., 1994. Speed sensorless vector control of induction motor using extended Kalman filter. *IEEE Transactions on Industrial Applications* 30, 1225–1233.

Kotecha, J., Djuric, P., 2003. Gaussian particle filtering. *IEEE Transactions on Signal Processing* 51, 2592–2601.

Laloy, E., Fasbender, D., Biielders, C.L., 2010. Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach. *J. Hydrol.* 380, 82-93.

Laloy, E., Vrugt, J.A., High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(ZS) and high-performance computing, *Water Resources Research*, 48, W01526, doi:10.1029/2011WR010608, 2012.

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H., 2009. Multivariate global sensitivity analysis for dynamic crop models. *Field Crop. Res.* 113, 312–320.

Launay M., Flenet F., Ruget F., Garcia de Cortazar Aauri I., 2005. Généricité et méthodologie

d'adaptation de STICS à de nouvelles cultures. Séminaire STICS, Carry-le-Rouet (mars 2005). 55-57 pp.

Lawless, C., & Semenov, M. A., 2005. Assessing lead-time for predicting wheat growth using a crop simulation model. *Agricultural and Forest Meteorology*, 135(1-4), 302-313, doi:10.1016/j.agrformet.2006.01.002.

Lawless C., Semenov M. A. and Jamieson P. D., 2008. Quantifying the effect of uncertainty in soil moisture characteristics on plant growth using a crop simulation model. *Field Crops Research*, Vol. 106 (2), 138-147.

Lemaire, G., Jeuffroy, M.H., Gastal, F., 2008. Diagnosis tool for plant and crop N status in vegetative stage. Theory and practices for crop N management. *Eur. J. Agron.* 28(4) 614-624.

Lee, J., Ricker, N., 1994. Extended Kalman filter based nonlinear model predictive control. *Industrial & Engineering Chemistry Research* 33, 1530-1541.

Livre Blanc "Céréales", 2007. F.U.S.A.Gx et CRA-W (Eds.). Gembloux - Février 2007. <http://www.gembloux.ulg.ac.be/pt/LIVREBLANC/LB/>

Liu, J., Chen, R., 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 1032-1044.

Ljung, L., 1979. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control* 24, 36-50.

Loague, K., Green, R.E., 1991. Statistical and graphical method for evaluating solute transport models: overview and application. *J. Contam. Hydrol.*, 7, 51-73.

Loyce C. et Wery J., 2006. Les outils des agronomes pour l'évaluation et la conception des systèmes de culture. *L'agronomie aujourd'hui*, Editions Quae, ISBN 1 0 : 2-7592-0000-0

Makowski D., Wallach D., 2001a. How to improve model-based decision rules for nitrogen fertilization, *Eur. J. Agron.*, 15 (3), 197-208.

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

Makowski D., Wallach D., 2002a. It pays to base parameter estimation on a realistic description of model errors, *Agronomie*, 22, 179-189.

Makowski, D., Wallach, D., Tremblay, M., 2002b. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agron.* 22, 191-203.

Makowski D., Hillier J., Wallach D., Andrieu B., Jeuffroy MH., 2006. Parameter estimation for crop models. In: *Working with Dynamic Crop Models. Evaluation, Analysis, Parameterization, and Applications*. Wallach D., Makowski D., Jones, J.W., 462 pp., Elsevier. Amsterdam.

Mansouri, M., Snoussi, H., Richard, C., 2009. A nonlinear estimation for target tracking in wireless sensor networks using quantized variational filtering, in: *Proceedings of 3rd International Conference on Signals, Circuits and Systems*, pp. 1-4.

Mansouri, M., Dumont, B., Destain, M.-F., 2013. Modeling and prediction of nonlinear environmental system using Bayesian methods. *Computers and Electronics in Agriculture* 92(0) 16-31.

Mary B., Laurent F., Beaudoin N., 2002. La gestion durable de la fertilisation azotée. Proc. 65e Congrès IIRB, février 2002, Bruxelles (BEL), 59-65

Matthies, L., Kanade, T., Szeliski, R., 1989. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision* 3, 209-238.

Mavromatis, T., & Hansen, J. W., 2001. Interannual variability characteristics and simulated crop response of four stochastic weather generators. *Agricultural and Forest Meteorology*, 109(4), 283-296, doi:10.1016/s0168-1923(01)00272-6.

Mavromatis, T., & Jones, P. D., 1998. Comparison of climate change scenario construction methodologies for impact assessment studies. *Agricultural and Forest Meteorology*, 91(1-2), 51-67, doi:10.1016/s0168-1923(98)00063-x.

McBratney, A., Whelan, B., Ancev, T., & Bouma, J., 2005. Future directions of precision agriculture. *Precision Agriculture*, 6(1), 7-23, doi:10.1007/s11119-005-0681-8.

McSwiney, C. P., and G. P. Robertson. 2005. Non-linear response of N₂O flux to incremental fertilizer addition in a continuous maize (*Zea mays* sp.) cropping system. *Global Change Biology* 11: 1712-1719.

Merwe, R.V.D., Wan, E., 2001. The square-root unscented Kalman filter for state and parameter-estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing* 6, 3461-3464

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087-1092.

Meyer, G., Neto, J., 2008. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture* 63, 282- 293.

Miller, A., 2002. *Subset Selection in Regression*. Chapman & Hall.

Minunno, F., van Oijen, M., Cameron, D.R., Cerasoli, S., Pereira, J.S., Tomé, M., 2013. Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. *Environ. Model. Softw.* 42(0) 99-115.

Monod, H., Naud, C., Makowski, D., 2006. Uncertainty and sensitivity analysis for crop models. In: Wallach, D., Makowski, D., Jones, J. W. (eds.) *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications*, Chapter 4. Elsevier, pp. 55-100.

Monteith, J.L., Moss, C.J., 1977. Climate and the Efficiency of Crop Production in Britain [and Discussion]. *Phil. Trans. R. Soc. Lond. B., Biol. Sci.*, vol. 281, no. 980, 277-294.

Monteith, J., 1996. The quest for balance in crop modeling. *Agron. J.* 88, 695-697.

Naud C., Makowski D., Jeuffroy M.H. 2008. Is it useful to combine measurements taken during the growing season with a dynamic model to predict the nitrogen status of winter wheat. *Eur. J. Agron.*, 28, 291-300.

Nonhebel, S., 1994a. The effects of use of average instead of daily weather data in crop growth simulation models. *Agricultural Systems*, 44(4), 377-396, doi:10.1016/0308-521x(94)90194-k.

Nonhebel, S., 1994b. Inaccuracies in weather data and their effects on crop growth simulation results: 1. Potential production. *Climate Research*, 4, 47-60.

Nounou, H., Nounou, M., 2006. Multiscale fuzzy Kalman filtering. *Engineering Applications of Artificial Intelligence* 19, 439-450.

Palmer, T. N., Doblus-Reyes, F. J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., et al., 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction, DEMETER. *Bulletin of the American Meteorological Society*, 85(6), 853-872, doi:10.1175/bams-85-6-853.

Palmer, T. N., Doblus-Reyes, F. J., Hagedorn, R., & Weisheimer, A., 2005. Probabilistic

prediction of climate using multi-model ensembles : from basics to applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463), 1991-1998.

Palosuo, T., Kersebaum, K. C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J. E., et al., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *European Journal of Agronomy*, 35(3), 103-114

Pearson, K., 1894. Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London. (A.)*, 185, 71-110, doi:10.1098/rsta.1894.0003.

Pearson, K., 1895. Contributions to the Mathematical Theory of Evolution, II : Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London - ARRAY*, 186, 343-414.

Porter, J. R., & Semenov, M. A., 1999. Climate variability and crop yields in Europe. *Nature*, 400(6746), 724-724.

Porter, J. R., & Semenov, M. A., 2005. Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463), 2021-2035.

Power, J. F., R. Wiese, and D. Flowerday. 2000. Managing nitrogen for water quality—Lessons from Management Systems Evaluation Area. *J. Environ. Qual.* 29:355-366.

Poyiadjis, G., Doucet, A., Singh, S., 2005. Maximum likelihood parameter estimation in general state-space models using particle methods. In: *Proceedings of the American Statistical Associations*.

Preece, D., 1990. R.A. Fisher and experimental design : a review. *Biometrics*, 46(4), 925-935.

Racsko, P., Szeidl, L., & Semenov, M., 1991. A serial approach to local stochastic weather models. *Ecological Modelling*, 57(1-2), 27-41, doi:10.1016/0304-3800(91)90053-4.

Ramirez, O.A., Misra, S.K., Field, J.E., 2001. Are Crop Yields Normally Distributed? *Am. Agric. Econ. Assoc. 2001 Annual meeting*, August 5-8, Chicago, IL.

Rasmussen, R., Hamilton, G., 2012. An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata *Environ. Model. Softw.* 29(1) 1-10.

Reid, J.B., 2002. Yield response to nutrient supply across a wide range of conditions: 1. Model derivation. *Field Crop. Res.* 77(2-3) 161-171.

Riha, S. J., Wilks, D. S., & Simoens, P., 1996. Impact of temperature and precipitation variability on crop model predictions. *Climatic Change*, 32(3), 293-311, doi:10.1007/bf00142466.

Ritchie J.T., Otter S. 1984. Description and performance of CERES-Wheat, a user oriented wheat yield model, USDA-ARS-SR Grassland Soil and Water Research Laboratory, Temple, TX, pp 159-177.

Robertson, M.J., Lyle, G., Bowden, J.W., 2008. Within-field variability of wheat yield and economic implications for spatially variable nutrient management. *Field Crop. Res.* 105(3) 211-220.

Ruget, F., Delecolle, R., Le Bas, C., Duru, M., Bonneville, N., Rabaud, V., Donet, I., Perarnaud, V., Panaguia, S., 2001. L'estimation regionale des productions fourrageres. In: Malezieux, E., Trebuil, G., Jaeger, M., (Eds.), *Modelisation des Agroecosystemes*. CIRAD & INRA.

Ruget F., Brisson N., Delécolle R., Faivre R., 2002. Sensitivity analysis of a crop simulation (STICS) in order to determine accuracy needed for parameters. *Agronomie*, 22:133-158.

Ruget F., Stager S., Volaire F., Lelievre F., 2009. Modelling Tiller Density, Growth, and Yield of Mediterranean Perennial Grasslands with STICS. *Crop Sci.*, 49(6): 2379-2385.

Sarkka, S., 2007. On unscented kalman filtering for state estimation of continuous time

nonlinear systems. *IEEE Transactions on Automatic Control* 52, 1631–1641.

Schoups, G. and Vrugt, J., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.*, 46, W10531, doi:10.1029/2009WR008933

Semenov, M.A., 2008. Simulation of weather extreme events by stochastic weather generator. *Climate Research*, 35, 203-212.

Semenov, M. A., & Barrow, E. M., 1997. Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, 35(4), 397-414, doi:10.1023/a:1005342632279.

Semenov, M.A., Barrow, E.M., 2002. LARS-WG - A stochastic weather generator for use in climate impact studies. User manual, version 3.0, August 2002. Tech. rep., Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK.

Semenov, M. A., & Doblas-Reyes, F. J., 2007. Utility of dynamical seasonal forecasts in predicting crop yield. *Climate Research*, 34(1), 71-81.

Semenov, M., & Porter, J., 1995. Climatic variability and the modelling of crop yields. *Agricultural and Forest Meteorology*, 73(3–4), 265-283, doi:10.1016/0168-1923(94)05078-k.

Simon, D., 2003. Kalman filtering of fuzzy discrete time dynamic systems. *Applied Soft Computing*, vol. 3, pp. 191–207.

Simon, D., 2006. *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*. John Wiley and Sons, 2006.

Sinclair, T.R., Seligman, N.A., 2000. Criteria for publishing papers on crop modeling. *Field Crops Res.* 68, 165–172.

Singh A.K., Tripathy R., Chopra U.K. 2008. Evaluation of CERES-Wheat and CropSyst models for water–nitrogen interactions in wheat crop. *Agric. Water Manage.*, 95, 776-786.

Singh, U., & Thornton, P. K., 1992. Using crop models for sustainability and environmental quality assessment. *Outlook on agriculture*, 21(3) : (Eds.) Turpin, Herts, ROYAUME-UNI, 1956.

Storvik, G., 2002. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing* 50, 281–289.

Ter Braak, C., 2006. A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. *Stat. Comput.* 16, 239-249.

Tremblay, M., Wallach, D., 2004. Comparison of parameter estimation methods for crop models. *Agron.* 24, 351-365.

Tey, Y., & Brindal, M., 2012. Factors influencing the adoption of precision agricultural technologies: a review for policy implications. *Precision Agriculture*, 13(6), 713-730, doi:10.1007/s11119-012-9273-6.

Van Alphen, B.J., Stoorvogel, J.J., 2000. A Methodology for Precision Nitrogen Fertilization in High-Input Farming Systems. *Precis. Agric.* 2(4) 319-332.

Vandenbergh, C., Marcoen, J., Sohier, C., Degre, A., Hendrickx, C., Paulus, P., 2011. Monitoring networks and modelling systems for assessing effectiveness of the EU Nitrates Directive Action Programmes: Approach by the Walloon Region (Belgium). Results of the second International Workshop, 10-11 June 2009.

van der Ploeg, R.R., Bohm, W., Kirkham, M.B., 1999. On the Origin of the Theory of Mineral Nutrition of Plants and the Law of the Minimum. *Soil Sci. Soc. Am. J.* 63, 1055-1062.

Van Diepen C.A., Wolf J., Van Keulen H. Rappoldt C. 1989. WOFOST : a simulation model of crop production. *Soil Use Manage*, 5, 16-24.

Varella, H., 2011. Inversion d'un modèle de culture pour estimer spatialement les propriétés des sols et améliorer la prédiction de variables agro-environnementales. PhD thesis. Univ. of Avignon, France, 207 p.

Varella H., Guérif M., Buis S., 2010a. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environ. Modell. Softw.*, 25, 310-319.

Varella, H., Guérif, M., Buis, S., Beaudoin, N., 2010b. Soil properties estimation by inversion of a crop model and observations on crops improves the prediction of agro-environmental variables. *Eur. J. Agron.* 33(2), 139-147.

Vermaak, J., Lawrence, N., Perez, P., 2003. Variational Inference for visual tracking. In: Conference on Computer Vision and Pattern Recognition.

Vrugt, J., Gupta, V., Bouten, W., Sorooshian, S., 2003. A shuffled Complex Evolution Metropolis algorithm for optimisation and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* 39(8), 1201.

Vrugt, J.A., Gupta, H.V., Nualláin, B., Bouten, W., 2006. Real-Time Data Assimilation for Operational Ensemble Streamflow Forecasting. *J. Hydrometeor.* 7(3) 548-565.

Vrugt, J., ter Braak, C., Clark, M., Hyman, J., Robinson, B., 2008a. Treatment of input uncertainty in hydrologic modeling using adaptive Markov Chain Monte Carlo sampling. *Water Resour. Res.* 44.

Vrugt, J., ter Braak, C., Gupta, H., Robinson, B., 2008b. Response to comment by Keith Beven on 'Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?' *Stoch. Environ. Res. Risk Assess.*, 23, 7p.

Vrugt, J., ter Braak, C., Diks, C., Robinson, B., Hyman, J., Higdon, D., 2009a. Accelerating Markov chain Monte Carlo simulation by self adaptive differential evolution with randomized subspace sampling. *Int. J. Nonlinear Sci. Numer. Simul.* 10, 271-288.

Vrugt, J., ter Braak, C., Gupta, H., Robinson, B., 2009b. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling. *Stochastic environmental research and risk assessment*, 23.

Vrugt, J.A., Laloy, E., and ter Braak, C.J.F., 2010. DiffeREntial Evolution Adaptive Metropolis with Sampling from past states. American Geophysical Union, Fall Meeting 2010

Wallach D., Goffinet M., Bergez J.E., Debaeke P., Leenhardt D., Aubertot J.N., 2001. Parameter estimation for crop models: a new approach and application to a corn model. *Agron. J.*, 93, 757-766.

Wallach D., Makowski D, Jones J.W., 2006. Working with dynamics crop models. Evaluation, Analysis, Parametrization and applications. 462 pp., Elsevier. Amsterdam.

Wallach, D., Van Evert, F., Adam, M., 2009. Parameter estimation software for crop models. Proceedings of the AgSAP Conference 2009: Egmond aan Zee, The Netherlands, pp. 342-343.

Wallach D., Buis S., Lecharpentier P., Bourges J., Clastre P., Launay M., Bergez J.E., Guérif M., Soudais J., Justes E., 2011. A package of parameter estimation methods and implementation for the STICS crop-soil model. *Environ. Modell. Softw.*, 26, 386-394.

Wan, E., Merwe, R.V.D., 2000. The unscented Kalman filter for nonlinear estimation. Adaptive

Systems for Signal Processing, Communications, and Control Symposium, 153–158.

Weir A.H., Bragg P.L., Porter J.R., Rayner J.H., 1984. A winter wheat crop simulation model without water or nutrient limitations, *J. Agric. Sci.*, 102, 371-382.

Weiss A., Wilhem W.W., 2006. The circuitous path to the comparison of simulated values from crop models with field observations. *J. Agric. Sci.*, 144, 475-488.

William J.R., Jones C.A., Kiniry J.R., Spanel D.A., 1989. The EPIC crop growth model, *Trans ASAE*, 32, 497-511.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. *Geoderma* 90, 169–185.

Wright, J.S., Sobel A.H., Schmidt G.A., 2009: Influence of condensate evaporation on water vapor and its stable isotopes in a GCM. *Geophys. Res. Lett.*, 36, L12804, doi:10.1029/2009GL038091.

Wu, Y., Liu, S., 2012. Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example. *Environ. Model. Softw.* 31(0) 99-109.

Yang, N., Tian, W., Jin, Z., Zhang, C., 2005. Particle filter for sensor fusion in a land vehicle navigation system. *Measurement Science and Technology* 16, 677.

Yuping M., Shili W., Li Z., Yingyu H., Liwei Z., Yanbo H., Futang W., 2008. Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 10, 426-437.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. *Weed research*, 14, 415-421.

Zhang, Z., 1997. Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing* 15, 59–76