



**HAL**  
open science

# Inversion d'un modèle de culture pour estimer spatialement les propriétés des sols et améliorer la prédiction de variables agro-environnementales

Hubert Vincent Varella

► **To cite this version:**

Hubert Vincent Varella. Inversion d'un modèle de culture pour estimer spatialement les propriétés des sols et améliorer la prédiction de variables agro-environnementales. Sciences agricoles. Université d'Avignon, 2009. Français. NNT : 2009AVIG0638 . tel-00626100

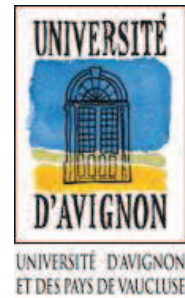
**HAL Id: tel-00626100**

**<https://theses.hal.science/tel-00626100>**

Submitted on 23 Sep 2011

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

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



ACADEMIE D'AIX-MARSEILLE  
UNIVERSITE D'AVIGNON ET DES PAYS DE VAUCLUSE

## THESE DE DOCTORAT

Présentée à l'Université d'Avignon et des Pays de Vaucluse

Ecole doctorale Sibaghe  
Spécialité Sciences Agronomiques

Par **Hubert-Vincent VARELLA**

---

# **Inversion d'un modèle de culture pour estimer spatialement les propriétés des sols et améliorer la prédiction de variables agro-environnementales**

---

Soutenue publiquement le 15 décembre 2009 devant le jury composé de :

Prof. Olivier BANTON  
Dr. Guy RICHARD  
Dr. Daniel WALLACH  
Dr. Hervé MONOD  
Dr. Selma CHERCHALI

Université d'Avignon  
INRA Orléans  
INRA Toulouse  
INRA Jouy-en-Josas  
CNES Toulouse

Président  
Rapporteur  
Rapporteur  
Examineur  
Examinatrice

Directrice de thèse : Dr. Martine Guérif, INRA Avignon

Préparée à l'UMR INRA/UAPV EMMAH

# Table des matières

<b>Introduction générale</b>	<b>5</b>
<b>Chapitre 1. Problématique et état de l'art</b>	<b>10</b>
1.1. Position du problème	10
1.1.1. Les modèles de culture sont des modèles dynamiques complexes avec un grand nombre de paramètres	10
1.1.2. Le problème posé par le grand nombre de paramètres à estimer	12
1.1.3. Les observations dont on dispose sont généralement peu nombreuses et imprécises	14
1.2. Les méthodes d'estimation et de sélection des paramètres à estimer	15
1.2.1. Les méthodes d'estimation de paramètres	15
1.2.2. Les méthodes de sélection des paramètres à estimer	18
1.2.3. Les méthodes d'analyse de sensibilité	20
1.3. Les questions posées par l'estimation des paramètres sol et la prédiction de variables agroenvironnementales	23
1.3.1. Lien entre analyse de sensibilité et quantité d'information disponible dans les observations	23
1.3.2. La prédiction des variables agroenvironnementales dépend de l'estimation des paramètres du sol	26
1.3.3. La distribution spatiale des paramètres du sol dans la parcelle est structurée et la prise en compte de cette structure peut améliorer l'estimation des paramètres	28
1.3.4. Les questions de recherche prises en compte et organisation de la thèse	30
<b>Chapitre 2. Outils et données</b>	<b>32</b>
2.1. Le modèle de culture STICS	32
2.1.1. Les différents modules de STICS, ses entrées et ses sorties	32
2.1.2. La représentation du sol, ses principaux processus et paramètres associés	36

2.2.	La méthode d'analyse de sensibilité Extended FAST	38
2.2.1.	La décomposition de la variance	38
2.2.2.	Les indices de sensibilité	39
2.2.3.	Estimation des indices de sensibilité par la méthode Extended FAST	40
2.3.	Présentation du théorème de Bayes et de la méthode Importance Sampling	43
2.3.1.	Le théorème de Bayes	43
2.3.2.	La méthode Importance Sampling	44
2.4.	Présentation des données utilisées	46
2.4.1.	Le contexte pédologique et les mesures du sol	48
2.4.2.	Les mesures du couvert végétal	55
2.4.3.	Les mesures climatiques	58
<b>Chapitre 3. Sélection des paramètres du sol à estimer</b>		<b>59</b>
3.1.	Objectif	59
3.2.	Méthodes	59
3.2.1.	Choix d'un modèle simplifié	60
3.2.2.	Réduction des problèmes d'identifiabilité	61
3.2.3.	Sélection des principaux paramètres par analyse de sensibilité	62
3.3.	Résultats	63
3.4.	Article 1 : "Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined for simulating agro-environmental variables"	65
<b>Chapitre 4. L'analyse de sensibilité globale mesure la qualité d'estimation des paramètres</b>		<b>90</b>
4.1.	Objectif	90
4.2.	Méthodes	91
4.3.	Résultats	92
4.4.	Article 2 : "Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model"	95
<b>Chapitre 5. Amélioration de la qualité de prédiction des variables d'intérêt à partir de l'estimation des paramètres du sol</b>		<b>106</b>

5.1.	Objectif	106
5.2.	Méthodes	107
5.3.	Résultats	108
5.4.	Article 3 : “The estimation of soil properties using observations on crop biophysical variables and the crop model STICS improves the predictions of agro-environmental variables”	111
<b>Chapitre 6. Prise en compte d’une structure spatiale dans l’estimation des paramètres du sol</b>		<b>147</b>
6.1.	Position du problème et objectif	147
6.2.	Méthodes et données	150
6.2.1.	Estimation et prédiction sans prise en compte d’une structure spatiale sur les paramètres (observations réelles)	150
6.2.2.	Estimation avec prise en compte d’une structure spatiale sur les paramètres (observations synthétiques)	152
6.3.	Résultats et discussion	162
6.3.1.	Estimation et prédiction sans prise en compte d’une structure spatiale sur les paramètres (données réelles)	162
6.3.2.	Estimation avec prise en compte d’une structure spatiale sur les paramètres (données synthétiques)	167
6.4.	Conclusion	173
<b>Conclusion générale</b>		<b>176</b>
<b>Annexe</b>		<b>182</b>
<b>Glossaire</b>		<b>194</b>
<b>Références générales</b>		<b>196</b>

## Introduction générale

Les modèles de cultures représentent le fonctionnement dynamique du système sol-plante en interaction avec les conditions climatiques, pédologiques et les pratiques agricole. Depuis les années 1970, 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 visés : compréhension du fonctionnement écophysologique pour les modèles hollandais comme SUCROS (De Wit, 1978) ou anglais comme AFRCWHEAT (Weir et al., 1984), analyse des implications des pratiques agricoles pour les modèles américains CERES (Ritchie and Otter, 1984) ou CROPGRO (Boote et al., 1998), prise en compte d'enjeux environnementaux pour les modèles comme EPIC (Williams et al., 1984) ou DAISY (Hansen et al., 1990). Dans les années 1990, en France, le modèle STICS (Brisson et al., 2002; 2008) intègre des objectifs à la fois agronomiques et environnementaux, et prend en compte grâce à des formalismes génériques un grand nombre de cultures, ouvrant la possibilité d'analyser les systèmes de culture. Les modèles de cultures constituent des outils indispensables pour comprendre l'influence des conditions agropédoclimatiques sur le système sol-plante. Ils ont été utilisés à différentes échelles spatiales temporelles pour servir des objectifs agronomiques, économiques ou environnementaux. A des échelles globales, les modèles sont utilisés pour réaliser un suivi objectif des productions agricoles aux échelles européenne (Supit et al., 1994) ou française (Ruget et al., 2001), ou pour estimer les impacts du changement climatique sur la production (Ewert et al., 1999; Reidsma et al., 2009) et l'offre agricole (Godard et al., 2008). A l'échelle régionale des terroirs viticoles français, Garcia De Cortazar Atauri (2006) propose une adaptation des pratiques viticoles dans un contexte de changement climatique en étudiant l'effet des pratiques sur le développement de la vigne. A des échelles locales, allant du petit bassin versant à la zone homogène au sein de la parcelle, les modèles sont utilisés dans le cadre de l'agriculture de précision pour suivre l'impact de contraintes agroenvironnementales sur la qualité de l'eau (Beaudoin et al., 2008), ou pour optimiser spatialement les pratiques agricoles telles que l'irrigation (Pang et al., 1998) ou la fertilisation (Engel et al., 1997; Paz et al.,

1999; Booltink et al., 2001; Houlès et al., 2004), de façon à maximiser le rendement ou le revenu tout en en minimisant le lessivage des nitrates vers la nappe.

La pertinence de l'utilisation des modèles de cultures repose sur la qualité des simulations réalisées. Cette dernière est basée à la fois sur un modèle de qualité, avec une erreur associée la plus faible possible, mais aussi sur une bonne détermination des entrées du modèle qui sont généralement très nombreuses. Ces entrées concernent les variables climatiques journalières (température, pluie, rayonnement,...), les paramètres généraux qui caractérisent les grands processus (minéralisation de la matière organique du sol, absorption d'eau et d'azote, bilan d'énergie,...), les caractéristiques écophysologiques de la plante et de la variété (exigences en température, efficacité de croissance, morphologie,...), les propriétés du sol (capacité de rétention d'eau, teneur en matière organique,...) et les techniques agricoles employées (travail du sol, irrigation, fertilisation,...). La connaissance des entrées du modèle peut être plus ou moins précise selon l'échelle d'application du modèle. Les variables climatiques, qui constituent les variables de forçage du modèle, sont mesurées automatiquement soit localement sur le lieu d'expérimentation, soit par les réseaux météorologiques nationaux. Les paramètres généraux et ceux caractéristiques de la plante sont généralement fournis par les concepteurs des modèles qui les établissent à partir de la bibliographie (vitesse de minéralisation de la matière organique du sol, efficacité d'interception de la lumière, efficacité de croissance en biomasse...), à partir d'expériences menées sur des processus spécifiques de la plante (courbe de dilution de l'azote, nombre de grains maximal,...) ou bien à partir d'une calibration préalable faite sur une large base de données expérimentales (Flenet et al., 2003; Hadria et al., 2007; Singh et al., 2008). Pour limiter les biais liés à la fixation, quelquefois un peu arbitraire, de ces paramètres, il est généralement recommandé de faire un diagnostic de leur bonne adaptation au contexte de l'utilisateur pour ensuite procéder à un éventuel travail de calibration sur des données représentatives de ce contexte. Les techniques agricoles sont plus ou moins faciles à connaître selon l'échelle d'application : elles correspondent soit à des estimations issues d'enquêtes ou de données déclaratives, soit aux choix qui ont été fait par l'agriculteur, soit aux modalités techniques dont on souhaite tester l'impact sur les réponses du modèle. La connaissance des propriétés

des sols, quant à elle, devient une étape délicate lorsque les modèles de cultures sont utilisés de manière spatialisée, car il est difficile de connaître les propriétés des sols en tout point l'espace (Calmon et al., 1999b; Leenhardt et al., 2006).

Le problème posé par la connaissance des propriétés des sols se pose évidemment à toutes les échelles. Aux échelles larges, des cartographies des sols existent généralement (selon les régions du monde considérées). La question qui se pose est alors de transformer l'information analytique et descriptive contenue dans ces cartes (succession de différents horizons et leurs texture, profondeur, teneur en cailloux,...) en valeurs de paramètres d'entrée du modèle (humidité à la capacité au champ et au point de flétrissement, réserve utile,...). C'est une problématique complexe, qui nécessite le recours à des fonctions de pédotransfert dont la définition constitue des domaines de recherche très fréquentés (Bastet et al., 1998; Bruand et al., 2003; Al Majou et al., 2007). Aux échelles régionales, la couverture en terme de cartographie pédologique est généralement beaucoup moins importante (voir l'état de l'inventaire pour la France : <http://www.gissol.fr/programme/igcs/igcs.php>). Aux échelles plus locales, la couverture cartographique en termes d'inventaire des sols est très hétérogène et il existe très peu de cartographies à une échelle compatible avec le renseignement intra-parcellaire des propriétés des sols d'une parcelle donnée. A cette échelle, les propriétés des sols peuvent être déterminées par des mesures directes, sur des observations de terrain complétées par des analyses au laboratoire d'échantillons prélevés sur une grille de sondage, permettant de déterminer certains constituants (contenu en argile, calcaire, eau, azote, matière organique,...). Cette technique permet de déterminer avec précision les propriétés des sols mais n'est cependant pas envisageable d'un point de vue opérationnel car l'analyse des sols peu devenir très coûteuse lorsque le nombre de points d'intérêt devient important. Quant aux méthodes géophysiques, elles permettent d'accéder à la cartographie de la résistivité électrique du sol d'une parcelle donnée et renseignent sur certaines propriétés, comme la teneur en eau du sol (Samouelian et al., 2005; Bourennane et al., 2007). Cependant, ces méthodes ne fournissent pas un renseignement suffisant pour déterminer de manière précise les propriétés du sol, car elles sont limitées par un manque d'interprétation robuste du signal électrique.



Il existe cependant, lorsque l'on dispose d'observations sur l'état du système sol-plante, une alternative intéressante pour estimer les propriétés des sols à l'échelle de la parcelle : elle consiste à inverser le modèle de culture à partir d'observations du système sol-plante pour retrouver les valeurs des propriétés des sols. Certains auteurs (Calmon et al., 1999b; Braga and Jones, 2004) ont montré que l'estimation des propriétés des sols par inversion d'un modèle de culture est plus performante à partir d'observations de contenu en eau du sol qu'à partir de cartes de rendement. Cependant, la mesure du contenu en eau du sol par des méthodes automatisées et donc indirectes, telles que les méthodes géophysiques ou la télédétection, n'est pas actuellement atteignable. Par contre, des données automatisées de différentes natures acquises sur la culture sont disponibles et interprétées en termes de variables d'état du couvert. L'utilisation de ces observations nous permet, par inversion du modèle de culture, d'accéder aux propriétés des sols. Plusieurs auteurs ont emprunté cette voie (Irmak et al., 2001; Timlin et al., 2001; Braga and Jones, 2004; Houlès, 2004; Ferreyra et al., 2006; Guérif et al., 2006) en utilisant des informations issues d'images de télédétection (Irmak et al., 2001; Timlin et al., 2001; Braga and Jones, 2004; Houlès, 2004; Ferreyra et al., 2006; Guérif et al., 2006) ou bien des cartes de rendement (Blackmore and Moore, 1999). En effet, les mesures de réflectance des couverts végétaux depuis l'espace sont largement utilisées pour accéder à l'estimation du couvert (indice foliaire, angles foliaires,...) et à la composition biochimique (teneur en chlorophylle, en eau, en azote) des feuilles (Weiss and Baret, 1999; Moulin et al., 2007; Houborg and Boegh, 2008). Par ailleurs, les systèmes de pilotage de la récolte avec capteurs de rendement permettent d'obtenir des cartographies précises du rendement (Machet et al., 2007). Ces observations spatialisées sur les variables d'état de la culture sont disponibles chez un nombre croissant d'agriculteurs impliqués dans l'agriculture de précision (le programme Farmstar concernant les images de télédétection peut par exemple être cité : [http://www.arvalisinstitutduvegetal.fr/fr/fichier/communique/481\\_Dossier.pdf](http://www.arvalisinstitutduvegetal.fr/fr/fichier/communique/481_Dossier.pdf)). La couverture exhaustive des parcelles de ces données, avec une résolution spatiale fine (de l'ordre de 5 mètres pour les cartes de rendement et 20 mètres pour les images issues de la télédétection) et une répétitivité de leur acquisition à différentes dates pendant la saison de culture et sur plusieurs saisons de culture, permettent

d'offrir un potentiel intéressant pour l'estimation inverse des propriétés des sols. Houllès (2004) et Guérif (2006) ont cependant montré, dans un contexte d'agriculture de précision et sur du blé d'hiver, que les résultats obtenus en termes d'estimation des propriétés du sol pouvaient être décevants, même s'ils permettaient d'améliorer la prédiction spatialisée du rendement du blé d'hiver.

Nous nous proposons dans cette thèse d'approfondir un certain nombre de questions qui se posent autour de l'utilisation de ce type d'observations pour estimer les propriétés des sols par inversion du modèle STICS et pour améliorer les prédictions réalisées par le modèle. Au delà du choix et de l'implémentation de méthodes appropriées, nous évaluerons la faisabilité de l'estimation des propriétés des sols, grâce à la mesure de la quantité d'information contenue dans les jeux d'observations. L'analyse des performances d'estimation et de prédiction en fonction du potentiel contenu dans les jeux d'observations disponibles est rarement entreprise dans le cadre des modèles de cultures. Nous y attacherons dans ce travail une attention particulière. Enfin, nous proposerons une voie possible pour l'utilisation d'une source d'information extérieure sur la structure spatiale de la parcelle, afin d'améliorer spatialement les performances d'estimation des propriétés des sols.

# Chapitre 1.      **Problématique et état de l'art**

Nous allons analyser dans ce chapitre les différentes questions soulevées par les objectifs poursuivis, faire un rapide état de l'art et préciser les questions qui sont prises en charge dans ce travail.

## **1.1. Position du problème**

L'estimation des propriétés du sol par inversion d'un modèle de culture avec des observations a déjà été testée par quelques auteurs (Irmak et al., 2001; Timlin et al., 2001; Braga and Jones, 2004; Guérif et al., 2006). Par exemple, Timlin et al. (2001) cherchent à retrouver l'humidité à la capacité au champ (*HCC*) de trois couches de sol en utilisant un modèle simple et des observations de cartes de rendement du maïs ; Braga et Jones (2004) estiment cinq paramètres de neuf couches de sol en inversant le modèle de culture CERES avec des observations de rendement du maïs et de teneur en eau du sol ; Guérif et al. (2006) estiment douze paramètres du sol en inversant le modèle de culture STICS avec des observations d'indice foliaire (*LAI*) et de teneur en azote de la plante (*QN*). Cependant, les résultats de l'estimation des paramètres du sol montrent qu'il est difficile d'obtenir une bonne qualité d'estimation, impliquant de ce fait une difficulté d'utilisation des valeurs estimées pour l'amélioration des prédictions. La question est en effet délicate pour plusieurs raisons que nous présenterons ici.

### **1.1.1. Les modèles de culture sont des modèles dynamiques complexes avec un grand nombre de paramètres**

Les modèles de culture sont des outils particulièrement intéressants pour valoriser les informations recueillies sur les cultures, en estimant les paramètres d'entrée du modèle par inversion. Ces observations représentent en effet des observations de variables d'état simulées par ces modèles. Il est donc possible de

les inverser pour retrouver les valeurs de leurs paramètres d'entrée tels que les caractéristiques du sol. Par ailleurs, ces modèles simulent des variables de sortie pertinentes pour la prévision des consommations en eau et en azote de la plante, du rendement et de la qualité de la récolte, lesquelles variables peuvent être optimisées pour la mise au point d'itinéraires techniques.

Ce sont des modèles complexes qui impliquent le sol, la plante et l'atmosphère. Ces trois compartiments sont reliés entre eux par des flux de matières qui sont pilotés par des bilans énergétiques. Par exemple, le flux de carbone dans le système de culture est piloté par un bilan d'énergie du rayonnement (Brisson et al., 2006). Ces modèles décrivent, sous forme d'équations mathématiques, les processus physiques et biologiques qui caractérisent le fonctionnement d'une culture, en interaction avec son environnement. Les simulations issues de ces modèles sont réalisées à un pas de temps journalier. Ces modèles comportent deux types de variables : des variables d'entrée et des variables de sortie. Les variables d'entrée sont représentées par les variables climatiques (pluies, températures, rayonnement global,...) qui sont généralement mesurées chaque jour et qui sont imposées au modèle (sous la forme d'un forçage climatique). Les variables de sortie sont calculées par le modèle et sont des variables agroenvironnementales qui décrivent l'état du système sol-plante. Leur nombre diffère selon les modèles, mais les principales variables généralement modélisées sont l'indice foliaire (*LAI*), la biomasse aérienne, la quantité d'azote contenu dans la plante, la quantité et la qualité de la récolte ainsi que le contenu en eau et en azote du sol.

Ces modèles peuvent comporter un grand nombre de paramètres qui règlent les différentes lois d'action et équations du modèle. Ils concernent les paramètres généraux, les caractéristiques de la plante cultivée, les techniques agricoles employées et les propriétés du sol. Le modèle STICS (Brisson et al., 2008), avec lequel nous travaillons, considère 227 paramètres pour le cas du blé tendre qui sont répartis en trois groupes : 129 paramètres liés aux caractéristiques de la plante, 23 paramètres des techniques agricoles et 75 paramètres de propriétés des sols. Les valeurs des paramètres liés aux caractéristiques de la plante proviennent d'études antérieures décrites dans la littérature, issues de mesures expérimentales précises

ou bien d'une calibration préalable sur une base de données (Flenet *et al.*, 2003; Hadria *et al.*, 2007; Singh *et al.*, 2008). L'incertitude liée à ce premier groupe de paramètres peut être importante car les valeurs issues des mesures ou de la calibration sont généralement entachées d'erreurs. Les paramètres des techniques agricoles appliquées au système cultural renseignent, entre autres, la date de semis, la nature des résidus de la culture précédente et le type de travail du sol, la date et la dose de fertilisant (ou d'eau) apportée au système. Les valeurs de ces paramètres sont généralement bien renseignées au niveau de la parcelle car elles correspondent aux décisions techniques prises par l'agriculteur. Dans le cas de l'agriculture de précision, la variabilité spatiale des modalités techniques est cependant difficile à appréhender et peut également être entachée d'erreurs. Les propriétés des sols, qui font l'objet central de ce travail, peuvent être déterminées à partir d'analyses ou de cartes de sols, mais cela n'est pas adapté, nous le rappelons, au contexte de l'agriculture de précision au niveau intra-parcellaire. Les paramètres du sol sont les plus difficiles à connaître en chaque point de l'espace et nous proposons dans cette étude de les estimer par inversion du modèle STICS.

### **1.1.2. Le problème posé par le grand nombre de paramètres à estimer**

Nous visons dans ce travail l'estimation des paramètres descriptifs du sol, en considérant que les autres paramètres sont connus. Même avec cette restriction, cela n'est pas chose aisée (Tremblay and Wallach, 2004; Launay and Guérif, 2005). La principale raison, indépendamment de la méthode d'estimation choisie, est qu'il est impossible d'estimer simultanément tous les paramètres du sol de STICS car une grande partie des paramètres n'est pas *identifiable* ; ceci est dû à la structure des équations du modèle (Niu and Fisher, 1997; Makowski *et al.*, 2006a). Un manque d'identifiabilité apparaît lorsque plusieurs valeurs de paramètres aboutissent aux mêmes valeurs des variables observées : il est difficile d'estimer correctement ces paramètres à partir de ces observations. Par exemple, une des équations du modèle STICS (voir Brisson *et al.*, 2008) permet de calculer la quantité d'azote organique actif dans le sol ( $NHUM, t\ ha^{-1}$ ), provenant de la minéralisation de la matière organique du sol, de la façon suivante :

$$NHUM = Norg \times PROFHUM \times DA(1) \times (1 - FINERT) \quad (1-1)$$

où  $Norg$  (%) est le contenu en azote organique du sol,  $profhum$  (cm) est la profondeur de minéralisation,  $DA(1)$  (en  $g\ cm^{-3}$ ) est la densité volumique de la première couche de sol et  $FINERT$  représente la proportion d'azote inactif (fixée à 0.65). Ainsi, lorsque des observations reliées à la quantité en azote organique actif dans le sol sont disponibles, l'équation ci-dessus ne permet pas d'estimer simultanément  $Norg$ ,  $profhum$  et  $DA(1)$  : il y a donc un problème d'identifiabilité. Pour cette raison, il est conseillé de sélectionner un sous-groupe de paramètres à estimer et de fixer les autres à une valeur convenable (appelée valeur nominale).

Quand bien même un sous-groupe de paramètres à estimer a été sélectionné, il subsiste un autre problème. Les trois groupes de paramètres (liés à la plante, aux techniques agricoles et aux propriétés des sols) dépendent les uns des autres à travers les équations de STICS. Cela a pour conséquence que les valeurs estimées des paramètres du sol dépendent des valeurs auxquelles les paramètres des deux autres groupes sont fixés. Une mauvaise valeur donnée à un paramètre technique ou plante entraîne donc un biais sur l'estimation des paramètres du sol. Ce biais est appelé biais d'omission (Miller, 2002). Voici un exemple d'équations de STICS où un paramètre plante (sensibilité de la plante à la sécheresse  $SEN SRSEC$ ) et des paramètres du sol (humidité à la capacité au champ  $HCC$  et au point de flétrissement  $HMIN$ ) donnent ensemble la valeur de l'effet de la sécheresse du sol à la date  $t$  ( $HUMIRAC_t$ ) sur la germination :

$$HUMIRAC_t = SEN SRSEC + (1 - SEN SRSEC) \frac{HUMSOL_t - HMIN}{HCC - HMIN}, \quad HUMSOL_t > HMIN \quad (1-2)$$

où  $HUMSOL_t$  est le contenu en eau du sol à la date  $t$ . A partir d'observations (directes ou indirectes) des variables  $HUMIRAC_t$  et  $HUMSOL_t$ , l'estimation de  $HCC$  et de  $HMIN$  peut être biaisée si  $SEN SRSEC$  n'est pas fixé à une bonne valeur. Il est donc important dans ce cas de calibrer correctement les paramètres liés aux caractéristiques de la plante afin d'estimer les paramètres du sol en minimisant au plus le biais d'omission.

### **1.1.3. Les observations dont on dispose sont généralement peu nombreuses et imprécises**

Nous avons vu que les observations sur les sols par des mesures indirectes n'étaient pas encore exploitables pour estimer les propriétés permanentes des sols ainsi que pour fournir des informations sur des variables d'état des sols (contenu en eau et en azote) qui pourraient elles-mêmes être utilisées en mode inverse pour accéder aux propriétés des sols. Seules sont facilement exploitables les observations sur les couverts végétaux obtenues par télédétection ou par les capteurs de rendement, que les agriculteurs capitalisent année après année dans les contextes d'agriculture de précision. Ces données ne sont toutefois pas exemptes de problèmes. Pour les images de télédétection à haute résolution il s'agit de :

(i) la faible répétitivité temporelle, croisée avec une haute probabilité de nuages, qui, dans les régions septentrionales comme la Picardie ne permet souvent pas d'avoir plus de 5 images exploitables par année,

(ii) la variabilité de la richesse spectrale, qui dépend du capteur disponible : celle des capteurs aéroportés de type CASI peut être grande (mais chère et donc rare) et donne accès au *LAI* et à la teneur en azote de la plante *QN* ; celle des capteurs satellites de type SPOT est faible et ne donne accès qu'au *LAI*,

(iii) le problème de l'inversion des mesures de réflectance pour estimer le *LAI* et le *QN* et des erreurs associées, de l'ordre de 17% pour *LAI* et 30% pour *QN* (Moulin et al., 2007). A cette erreur, s'ajoute celle liée à la mauvaise connaissance du contenu en aérosols de l'atmosphère utilisé pour obtenir les réflectances de surface à partir des mesures au niveau du satellite (Launay et al., 2000).

En ce qui concerne les cartes de rendement, établies à l'aide de capteurs embarqués sur la moissonneuse, le principal problème d'obtention réside dans la difficulté à étalonner ces capteurs de manière précise sur l'ensemble de la parcelle, ce qui engendre une erreur de l'ordre de 9% (Machet et al., 2007).

Compte tenu de la complexité du problème posé, il est nécessaire de mettre en œuvre une méthodologie adaptée, qui consiste dans une première étape à cibler les paramètres qui sont les plus pertinents à estimer et dans une seconde étape à

appliquer la méthode d'estimation la mieux adaptée à notre contexte. Nous présenterons dans la suite une manière de répondre à ces objectifs.

## **1.2. Les méthodes d'estimation et de sélection des paramètres à estimer**

### **1.2.1. Les méthodes d'estimation de paramètres**

Il existe un large panel de méthodes pour estimer les paramètres d'un modèle complexe. Ces méthodes peuvent être regroupées en deux principales familles : l'approche fréquentiste et l'approche Bayésienne (Makowski et al., 2006a). La mise en œuvre de l'approche fréquentiste ne nécessite qu'un jeu d'observations alors que l'approche Bayésienne utilise en plus, une information sur la distribution des paramètres à estimer. Les approches Bayésiennes sont devenues de plus en plus utilisées ces dernières années pour estimer les paramètres de modèles complexes car elles permettent de mieux prendre en compte les incertitudes, aussi bien sur les paramètres d'entrée que sur les simulations du modèle. Par ailleurs, leur utilisation a été largement facilitée par le décuplement des vitesses de calcul des ordinateurs et le développement de nouveaux algorithmes.

Les méthodes fréquentistes ne considèrent pas les paramètres du modèle comme étant des variables aléatoires, comme le font les méthodes Bayésiennes, mais plutôt comme étant des variables fixées à une certaine valeur, à estimer. L'application d'une méthode fréquentiste permet alors de déterminer une valeur particulière de chaque paramètre à partir d'un jeu d'observations, et cette valeur est appelée estimateur du paramètre. Parmi ces méthodes fréquentistes, il existe celle du maximum de vraisemblance (Aldrich, 1997; Hald, 1999), des moindres carrés (Seber and Wild, 2003), les algorithmes génétiques (Mitchell, 1998) ou les méthodes variationnelles (Bouttier and Courtier, 1999) qui permettent, en plus d'estimer les paramètres, de contrôler les variables d'état du modèle (lissage, prédiction à court terme, ...).



Les méthodes Bayésiennes, quant à elles, utilisent une information supplémentaire sur la distribution des paramètres, dite information *a priori*. Les paramètres étant ici considérés comme des variables aléatoires définies par une densité de probabilité *a priori*, le résultat de l'application d'une approche Bayésienne sur un jeu de données est une nouvelle densité de probabilité appelée densité *a posteriori* des paramètres. L'application d'une méthode Bayésienne peut être abordée en deux étapes. La première consiste à déterminer l'information *a priori* des paramètres à estimer, à partir de différentes sources qui peuvent être constituées par des mesures, la littérature ou bien des dires d'experts. Cette information *a priori* peut se limiter à de simples bornes sur les valeurs des paramètres sans être plus informative sur la distribution de ces valeurs (densité uniforme) ; elle peut aussi être plus précise quant à cette distribution (densité normale, de Poisson, Gamma, ...). La seconde étape consiste à déterminer la densité *a posteriori* à partir de la densité *a priori* et du jeu d'observations, en utilisant le théorème de Bayes (Makowski et al., 2006a). Cette densité *a posteriori* peut alors être utilisée à diverses fins comme estimer la valeur la plus probable de chaque paramètre (en considérant le mode ou la moyenne), calculer l'incertitude sur l'estimation des paramètres ou encore calculer l'incertitude sur les variables simulées par le modèle à partir de l'incertitude sur l'estimation des paramètres. Parmi les méthodes Bayésiennes, on peut citer MCMC (Metropolis et al., 1953; Hastings, 1970), Importance Sampling (Beven and Binley, 1992; Beven and Freer, 2001) ou bien certaines méthodes de filtrage (Hilgert et al., 2005; Rossi and Vila, 2005) qui permettent en plus, comme les méthodes fréquentistes variationnelles, de contrôler les variables d'état du modèle.

En ce qui concerne les paramètres du sol, on considère qu'il est possible d'obtenir de l'information *a priori* sur ces paramètres à l'échelle de la parcelle. Cette information est disponible à partir de différentes sources et sont associées à différentes précisions. La première source d'information provient des cartes de sols couplées à des fonctions de pédotransfert qui permettent de fournir, pour une parcelle donnée, des valeurs de paramètres. Cette première source est, nous le rappelons, associée à une faible précision car les cartes sont généralement établies à une échelle plus grande que celle de la parcelle agricole. Cependant, il est par exemple possible de définir l'information *a priori* au niveau intra-parcellaire comme

étant une distribution statistique centrée sur ces valeurs et avec une certaine variance. La seconde source d'information provient de mesures de paramètres issues d'analyses d'échantillons de sols pouvant être réalisées de manière plus ou moins fréquentes dans l'espace. A partir de ces mesures, une distribution statistique peut ensuite être proposée pour chaque paramètre, définissant ainsi l'information *a priori*. Cette seconde source peut amener à déterminer une information *a priori* très précise lorsque les mesures issues des analyses de sols deviennent très fréquentes dans l'espace. L'approche Bayésienne est donc préférée à l'approche fréquentiste dans notre cas. Comme les observations dont on dispose sont généralement peu nombreuses, les méthodes de type filtrage, gourmandes en nombre d'observations, sont donc épargnées. Du point de vue du temps de calcul de la distribution *a posteriori*, la méthode Importance Sampling est une méthode très économique lorsqu'il s'agit d'estimer spatialement (i.e. sur un grand nombre de points) les paramètres du sol dans un domaine comme la parcelle agricole, ce qui n'est pas le cas des méthodes MCMC ou de filtrage. De plus, Importance Sampling s'avère être d'une performance comparable à MCMC (Makowski et al., 2002). Cette méthode est donc celle que nous avons retenue pour faire l'estimation des paramètres sol dans toute notre étude.

Nous avons déjà évoqué au Chapitre 1.1.2, que le principal problème posé par le grand nombre de paramètres à estimer, indépendamment de la méthode choisie, était le problème d'identifiabilité. A cela il faut ajouter d'autres problèmes liés au choix de la méthode fréquentiste ou Bayésienne. Pour les méthodes fréquentistes, il faut aussi compter sur le problème du modèle *sur-paramétré* (l'estimation d'un grand nombre de paramètres conduit à une grande variance des estimateurs) ainsi que sur le problème de la divergence des estimateurs par rapport à la solution optimale (le nombre de minimas locaux est accru par le nombre de paramètres à estimer). Concernant les méthodes Bayésiennes, l'autre principal problème est que l'estimation d'un grand nombre de paramètres est synonyme d'autant de définition d'informations *a priori* que de paramètres à estimer. Ce travail constitue une étude très lourde.

### 1.2.2. Les méthodes de sélection des paramètres à estimer

Compte tenu de la quantité de problèmes causés par l'estimation d'un grand nombre de paramètres, il est recommandé de sélectionner un sous-groupe de paramètres à estimer et de fixer les autres à une valeur nominale. Pour cela, Makowski et al. (2006a) proposent quatre méthodes pour sélectionner ces paramètres :

- sélection basée sur la littérature,
- sélection pour éviter les problèmes d'identifiabilité,
- analyse de sensibilité,
- choix statistique des paramètres à estimer.

Le principe de la première méthode est de sélectionner les paramètres pour lesquels aucune valeur n'est fournie par la littérature. Par exemple, Bonesmo et Bélanger (2002) fixent 4 paramètres, parmi ceux liés aux caractéristiques de la plante, à des valeurs définies par la littérature et en estiment 17 autres. Cependant, cette méthode n'est pas valable pour l'estimation des paramètres du sol car leurs valeurs ne peuvent pas être définies par la littérature en tout point de l'espace.

L'analyse des équations du modèle STICS permet d'éviter en partie des problèmes d'identifiabilité. Reprenons l'exemple de l'Equation (1-1) du Chapitre 1.1.2. Nous avons vu pour cette équation qu'il n'était pas possible d'estimer simultanément les paramètres *Norg*, *profhum* et *DA(1)*. La méthode proposée ici suggère alors d'estimer soit le produit  $Norg \times profhum \times DA(1)$  dans sa globalité, soit un seul des trois paramètres du produit en fixant les autres à une valeur nominale.

Il est possible de sélectionner les paramètres à estimer, pour un jeu d'observations donné, en appliquant des méthodes statistiques. Pour ce faire, il faut proposer à la méthode une liste de sous-groupes candidats de paramètres à estimer (soit en scrutant toutes les possibilités du groupe de paramètres, soit en faisant une proposition raisonnée de liste) afin de fournir des critères permettant de décider du sous-groupe optimal de paramètres à estimer, compte tenu du jeu d'observations.

Parmi les nombreux critères qui existent, le critère *BIC* est un des plus performants (Tremblay and Wallach, 2004) :

$$BIC = -2\log(Lik) + P\log(N)$$

où  $\log(Lik)$  est le logarithme de la fonction vraisemblance,  $N$  est le nombre d'observations et  $P$  est le nombre de paramètres du sous-groupe candidat. La fonction vraisemblance (Makowski et al., 2006a) doit être calculée à partir des estimations des paramètres du sous-groupe, ce qui sous-entend le choix préalable d'une méthode d'estimation de paramètres adaptée. Cependant, l'application de cette méthode pour sélectionner les paramètres à estimer peut être très coûteuse en temps de calcul. Par exemple, la sélection optimale d'un sous-groupe de paramètres à estimer parmi un groupe de 13 paramètres mène à proposer  $2^{13}=8192$  sous-groupes candidats et à effectuer autant de procédures d'estimation de paramètres.

L'analyse de sensibilité est une des méthodes les plus utilisées pour la sélection des paramètres, car elle permet en même temps d'accéder à la compréhension du fonctionnement du modèle (Saltelli et al., 2000b). Elle consiste en effet à détecter les paramètres dont les incertitudes ont un effet significatif sur les variables observables. En quantifiant l'effet de l'incertitude, associée à chaque paramètre, sur les variables observables (sous la forme d'un indice par exemple) et en définissant un seuil en deçà duquel l'effet est considéré comme étant insignifiant, il est ainsi possible de restreindre le groupe de paramètres analysés à une sélection des principaux paramètres à estimer.

Afin de procéder à la sélection des paramètres du sol à estimer, nous avons choisi dans notre étude : (i) de réduire les problèmes d'identifiabilité, en analysant les équations du modèle et en fixant certains paramètres sol à une valeur nominale, (ii) d'utiliser l'analyse de sensibilité des variables observables du modèle STICS aux paramètres. Au détriment des méthodes statistiques telles que *BIC*, l'analyse de sensibilité permet de quantifier la quantité d'information disponible dans les observations pour estimer les paramètres (voir Chapitre 1.3.1) et permet ainsi de répondre à un de nos objectifs. De plus, le nombre de procédures d'estimation, que

la méthode *BIC* suggère d'effectuer pour la sélection, implique un temps de calcul trop important lorsque cette méthode est appliquée au modèle *STICS*. De ce fait, l'analyse de sensibilité est ici préférée aux méthodes statistiques pour sélectionner les paramètres à estimer. Ce travail de sélection de paramètres, à travers les points (i) et (ii), est décrit de manière plus précise dans le Chapitre 3, chapitre dans lequel la question de l'analyse de sensibilité est traitée sous la forme d'un article. Avant cela, il est utile de présenter les différentes méthodes d'analyse de sensibilité d'un modèle complexe.

### 1.2.3. Les méthodes d'analyse de sensibilité

Les modèles de simulations deviennent de plus en plus complexes et la compréhension du fonctionnement du modèle et du comportement des sorties en fonction des entrées devient de plus en plus difficile. L'analyse de sensibilité du modèle est un moyen intéressant pour aider à cette compréhension. Un certain nombre d'auteurs (dont Campolongo and Saltelli, 1997; Brun *et al.*, 2002; Cariboni *et al.*, 2004; Ratto *et al.*, 2007; Manache and Melching, 2008) qualifient l'objectif de l'analyse de sensibilité comme étant une réponse aux questions du type :

1. Quels sont les paramètres d'entrée  $\theta_i, i=1, \dots, P$ , dont l'incertitude influence le plus celle sur la sortie  $Y = f(\theta)$  du modèle  $f$  ?
2. Quels paramètres ont un effet négligeable sur l'incertitude de la sortie de sorte qu'il est possible de les fixer à une valeur nominale ?

Répondre à ces questions permet de connaître, parmi un sous-groupe de paramètres, lesquels sont à estimer en priorité (1) et lesquels sont à fixer à une valeur nominale (2). Pour répondre à ces questions, diverses méthodes d'analyse de sensibilité existent et le choix de l'une d'entre elles doit être fait en fonction du cas d'étude et des propriétés du modèle. Cariboni (2004) dresse un schéma décisionnel concernant le choix de la méthode (voir Figure 1-1) et permet de discriminer les différentes méthodes d'analyse de sensibilité entre elles suivant 3 principaux critères :

- le nombre  $P$  de paramètres d'entrée considérés,
- le temps d'exécution du modèle  $f$ ,
- la linéarité du modèle.

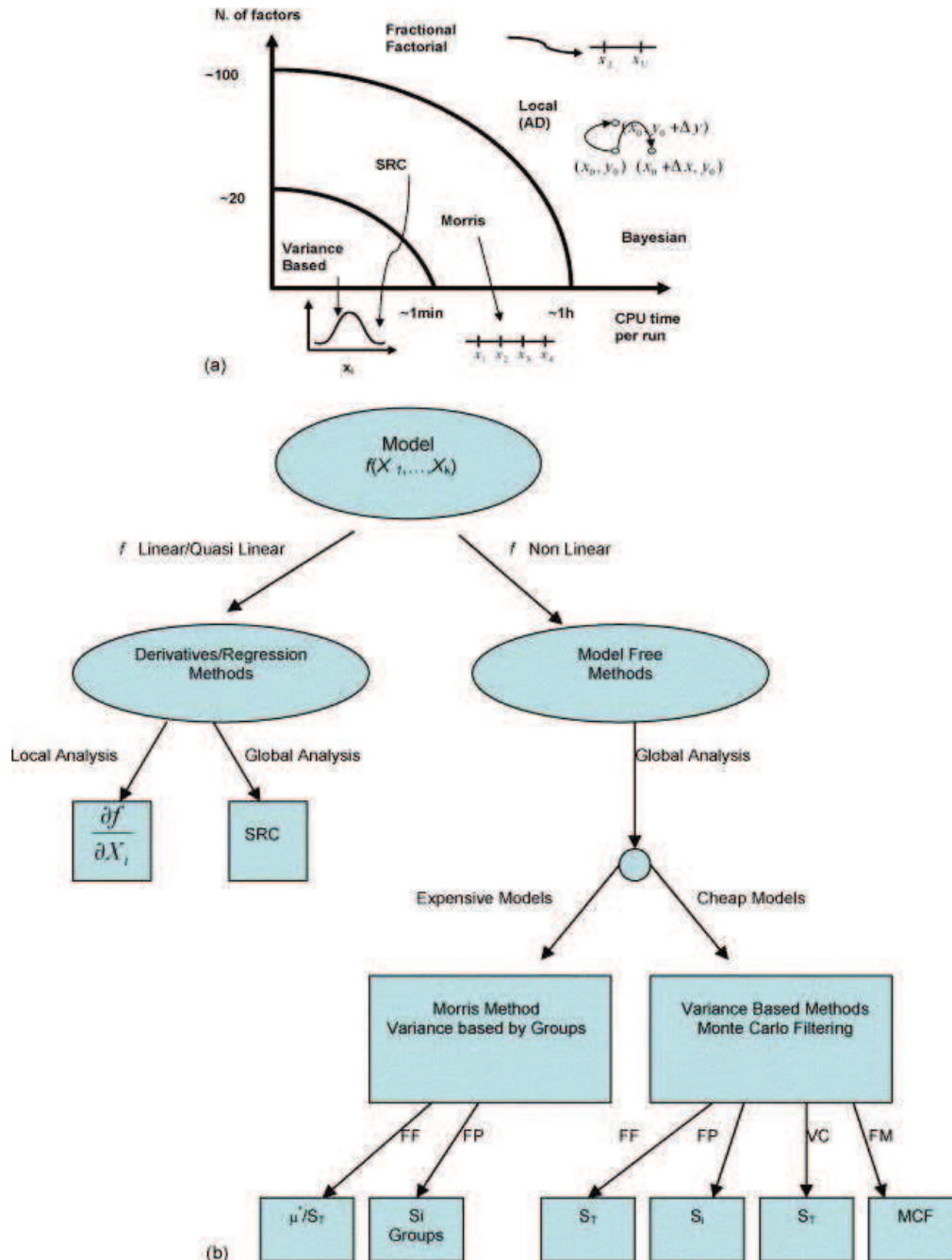


Figure 1-1. Les différents types d'analyses de sensibilité (Cariboni *et al.*, 2004).

Pour un nombre  $P$  inférieur à 20 et un temps d'exécution inférieur à 1 minute, il est conseillé d'utiliser une méthode basée sur la variance ("Variance Based") (Chan *et al.*, 2000; Makowski *et al.*, 2006b) ; pour un nombre  $P$  compris entre 20 et 100 et

un temps d'exécution compris entre 1 minute et 1 heure, une méthode "Standardized Regression Coefficients" (SRCs) ou de Morris (Morris, 1991; Campolongo et al., 2006) serait plus adaptée ; lorsque  $P$  est supérieur à 100 ou bien que le temps d'exécution est supérieur à 1 heure, des méthodes locales type dérivées (Varma et al., 1999; Grievank, 2000), "Fractional Factorial" ou encore Bayésiennes (Saltelli et al., 2000a) seraient plutôt à envisager. La linéarité du modèle  $f$  est le dernier critère permettant de discriminer les méthodes entre elles, comme le montre la Figure 1-1b. Lorsque le modèle a un comportement non-linéaire par rapport à l'incertitude des paramètres d'entrée, seules des méthodes de type globales peuvent être utilisées parmi lesquelles on retrouve la méthode de Morris et les méthodes basées sur la variance. Par contre, si le modèle se comporte linéairement, le choix se partage entre les méthodes locales, de type dérivées, et les méthodes globales SRCs ; choix qui doit être affiné en fonction des deux autres critères. En réponse à la question (1), l'utilisation de l'une de ces méthodes permet de quantifier l'influence de l'incertitude des paramètres  $\theta_i, i=1, \dots, P$ , sur celle de la sortie  $Y = f(\theta)$  et de connaître ceux dont il est important d'éliminer l'incertitude pour réduire au minimum celle sur la sortie. Les méthodes locales de type dérivées donnent des résultats informatifs seulement lorsque le modèle est linéaire ou si la gamme d'incertitude est petite ; si la gamme est grande et le modèle proche du linéaire, les méthodes SRCs sont des outils intéressants pour répondre à ces questions ; sinon d'autres types de méthodes globales doivent être utilisés, comme la méthode de Morris ou les méthodes basées sur la variance. Ces deux dernières méthodes permettent également de quantifier l'importance que les paramètres peuvent avoir sur la sortie lorsqu'ils agissent en interaction entre eux. Pour un paramètre donné, il est possible que son incertitude n'influence celle sur la sortie qu'à travers son interaction avec un ou plusieurs autres paramètres. Si l'incertitude de ce paramètre n'influence pas celle sur la sortie, même à travers son interaction, il est donc possible de le fixer à une valeur nominale. En ce sens, la méthode de Morris ou les méthodes basées sur la variance permettent de répondre à la question (2).

Dans notre cas d'étude appliqué aux paramètres sol du modèle de culture STICS, on ne s'intéressera qu'à un nombre  $P$  de paramètre toujours inférieur à 20. De plus, le modèle STICS est un modèle complexe non linéaire et une exécution du



modèle est de l'ordre de la seconde. Les méthodes basées sur la variance semblent donc être les plus adaptées à notre problème. Parmi elles, deux méthodes sont couramment utilisées : la méthode de Sobol' (Sobol, 1993) et Extended FAST (Saltelli et al., 1999). Makowski (2006b) montre sur un modèle peu coûteux, non linéaire et sur 13 paramètres que les résultats fournis par la méthode Extended FAST convergent plus rapidement vers des solutions stables que ceux fournis par la méthode de Sobol'. Ce comportement a également été constaté par d'autres auteurs (Saltelli and Bolado, 1998; Saltelli et al., 1999). La méthode globale Extended FAST sera donc celle considérée dans toute notre étude.

### **1.3. Les questions posées par l'estimation des paramètres sol et la prédiction de variables agroenvironnementales**

#### **1.3.1. Lien entre analyse de sensibilité et quantité d'information disponible dans les observations**

Une certaine quantité d'information contenue dans les observations est nécessaire pour estimer les paramètres sélectionnés. Cependant, les observations ne fournissent pas la même quantité d'information sur tous les paramètres, ce qui implique une disparité dans les performances d'estimation des paramètres. Comme nous le verrons dans ce chapitre, la quantité d'information peut être mesurée via l'analyse de sensibilité des variables observables aux paramètres. Aussi, la linéarité du modèle par rapport à l'incertitude sur les paramètres est un aspect primordial à considérer avant de choisir une méthode d'analyse de sensibilité. L'analyse de sensibilité est classiquement utilisée dans le cas linéaire pour mesurer la quantité d'information, traduisant la performance d'estimation, alors que dans le cas non linéaires, peu de travaux ont été réalisés sur l'utilisation de l'analyse de sensibilité globale pour quantifier cette quantité. Cette question a été abordée dans ce travail de thèse. Dans ce chapitre, nous commencerons par rappeler comment la quantité d'information est mesurée dans le cas linéaire, pour ensuite présenter les bases de la mesure de cette quantité dans le cas non linéaire.



### La quantité d'information au sens de Fisher

La question posée dans cette section est de savoir ce que nous apprend un jeu d'observations sur les paramètres. Plus précisément, quelle quantité d'information le jeu d'observations fournit-il sur les paramètres ? C'est ce que la théorie de l'information au sens de Fisher (Kauffmann, 1994) permet de déterminer. Ainsi, il est possible de quantifier ce que le jeu d'observations donne comme information pour estimer les paramètres, permettant de distinguer ceux qui peuvent être estimés de ceux qui ne le peuvent pas.

Soit un jeu d'observations composé de  $K$  observations, les sorties observées associées du modèle,  $f_k$ ,  $k=1, \dots, K$ , et un sous-groupe de  $P$  paramètres  $\theta_p$ ,  $p=1, \dots, P$ . La théorie de l'information au sens de Fisher permet de construire une matrice d'information, appelée *FIM*, telle que  $FIM = Q^T W Q$  où  $W$  est une matrice de covariance d'erreur d'observation et  $Q$  est de la forme suivante :

$$Q = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \dots & \frac{\partial f_1}{\partial \theta_p} \\ \vdots & \dots & \vdots \\ \frac{\partial f_K}{\partial \theta_1} & \dots & \frac{\partial f_K}{\partial \theta_p} \end{bmatrix}$$

$\partial f_k / \partial \theta_p$  est la dérivée de la fonction  $f_k$  par rapport au paramètre  $\theta_p$ . Dès lors, deux cas de figure peuvent se présenter en ce qui concerne la linéarité des sorties :

- soit les sorties sont linéaires par rapport à l'incertitude sur les paramètres,
- soit les sorties ne sont pas linéaires par rapport à l'incertitude sur les paramètres.

Dans le cas linéaire, la matrice  $Q$  peut être vue comme une matrice d'indices de sensibilité où  $\partial f_k / \partial \theta_p$  est l'indice de sensibilité de  $f_k$  par rapport à  $\theta_p$ . Les paramètres ayant un indice de sensibilité élevé sont ceux qui peuvent être estimés avec précision, grâce au jeu d'observations. De plus, chaque composante  $C_{pq}$ ,  $p=1, \dots, P$ ,  $q=1, \dots, P$  de la matrice *FIM* quantifie la quantité d'information qui est fournie par le jeu d'observations sur le couple de paramètres  $(\theta_p, \theta_q)$ . Pour finir, le calcul d'un critère d'optimalité à partir de *FIM* tel que le critère A, D ou E (Rodriguez-

Fernandez *et al.*, 2006; Pronzato, 2008) permet de quantifier la quantité d'information fournie sur l'ensemble du sous-groupe de paramètres.

Dans le cas non linéaire, la matrice *FIM* n'a de sens que si elle est calculée pour les valeurs estimées des paramètres. Dans ce cas, la matrice *FIM* permet, comme pour le cas linéaire, de quantifier la quantité d'information qui a été apportée par le jeu d'observations pour estimer les paramètres. Cependant, il n'est pas possible de calculer les indices de sensibilité des sorties du modèle par rapport à l'incertitude sur les paramètres, ainsi que de quantifier l'information apportée par le jeu d'observations pour réduire l'incertitude sur les paramètres. Pour répondre à cette question, d'autres types d'analyses que celle proposée par Fisher peuvent être entreprises. Parmi elles, les méthodes d'analyse de sensibilité globales.

#### *Le cas non linéaire et l'utilisation de l'analyse de sensibilité globale*

L'application d'une méthode d'analyse de sensibilité globale à un modèle non linéaire permet seulement d'apprécier qualitativement la quantité d'information apportée par une observation pour estimer les paramètres. Lorsqu'un paramètre agit directement sur la variable observable de sortie de manière importante et non via ses interactions avec les autres paramètres, on peut dire que l'observation de la variable fournit une grande quantité d'information pour estimer ce paramètre et qu'il est alors possible d'éliminer son incertitude. Ratto *et al.* (2007) qualifie de "Factor Prioritization" (FP sur la Figure 1-1b) le fait de s'intéresser à un paramètre pour en éliminer son incertitude. Lorsqu'un paramètre n'agit d'aucune manière (ni directement, ni en interaction) sur la variable observable de sortie, on peut dire que l'observation de la variable ne fournit aucune information pour estimer ce paramètre et qu'il peut être alors fixé à une valeur nominale. Ratto *et al.* (2007) qualifie de "Factor Fixing" (FF sur la Figure 1-1b) le fait de fixer à une valeur nominale un paramètre n'ayant aucune influence sur la sortie. Par exemple, Ruget (2002) montre sur le modèle de culture STICS qu'il est possible de choisir les principaux paramètres à estimer en étudiant la sensibilité des principales variables de sortie de STICS à 28 paramètres d'entrée (plante, sol et technique). Par conséquent, son travail montre qu'il est possible de qualifier la quantité d'information contenue dans l'observation de

chaque variable de sortie du modèle, afin de choisir les principaux paramètres à estimer.

Cependant, aucune étude n'a permis à ce jour de quantifier la quantité d'information qui est apportée par l'ensemble d'un jeu d'observations sur l'estimation des paramètres sol d'un modèle complexe comme STICS. Nous proposons alors de répondre à cette question au Chapitre 4 en proposant des critères, basés sur les indices de sensibilité globaux, qui mesurent cette quantité d'information, elle-même liée à la performance d'estimation des paramètres.

### **1.3.2. La prédiction des variables agroenvironnementales dépend de l'estimation des paramètres du sol**

L'estimation des paramètres du sol à partir d'observations sur le couvert végétal est un objectif en soi, permettant d'accéder à la connaissance de propriétés difficilement mesurables sur des espaces étendus. Comme nous l'avons vu au chapitre précédent, cette connaissance dépend de la quantité d'information qui est fournie par les observations. En outre, l'estimation des paramètres du sol peut également être considérée comme un moyen d'affiner l'utilisation du modèle de culture, en améliorant le renseignement des paramètres du sol, visant à améliorer les prédictions des variables agroenvironnementales.

La liaison étroite qui existe entre les variables observées et les variables d'intérêt, car corrélées entre elles d'une certaine manière, implique que la sensibilité des variables observables et celle des variables à prédire aux paramètres du sol sont similaires. Dans ce cas, la méthodologie d'estimation des paramètres peut s'avérer être efficace pour améliorer les prédictions. Par exemple, la teneur en protéines est directement liée au rendement et à la quantité d'azote contenu dans la plante, ce qui signifie que leurs sensibilités respectives aux paramètres doivent être semblables. En ce sens, il est donc possible d'améliorer les prédictions des variables d'intérêt, en tout point de l'espace, à partir de l'estimation des paramètres du sol issue de l'inversion de STICS avec des observations du couvert végétal. Cette réflexion n'est pas nouvelle, elle est sous-jacente au processus connu sous le nom de

calibration/validation, où un certain nombre de paramètres sont estimés à partir d'un jeu d'observations (étape de calibration) et où l'impact de cette estimation est testé et analysé en prédiction sur un jeu d'observations indépendant (étape de validation). Un grand nombre de travaux liés à la calibration/validation existent dans la littérature (dont Hadria *et al.*, 2007; Tonitto *et al.*, 2007; Beaudoin *et al.*, 2008; Heidmann *et al.*, 2008; Dimokas *et al.*, 2009). Par exemple, Heidmann (2008) propose de calibrer des paramètres de croissance d'un modèle de culture de pomme de terre avec des observations sur deux années de matière sèche (liée à l'indice foliaire) et de quantité d'azote contenue dans la plante, pour ensuite valider cette approche sur une autre année en prédisant de la matière sèche et de la quantité d'azote contenue dans la plante. Un autre exemple, Tonitto (2007) cale des paramètres chimiques et de sol, qu'il qualifie de mal connus *a priori*, avec des observations de drainage et de lessivage des nitrates sur plusieurs années, pour ensuite améliorer la prédiction du rendement du maïs et du soja, par rapport à celle issue des valeurs *a priori* des paramètres. Dans ces travaux, il y a une réelle volonté d'améliorer les prédictions grâce à l'estimation de paramètres, ce qui est généralement soldé par un succès, car les variables prédites et les variables observées présentent des sensibilités aux paramètres estimés comparables. Dans notre cas d'étude, l'estimation des paramètres du sol pourrait donc permettre d'améliorer la prédiction des variables d'intérêt agroenvironnemental.

Dans notre étude, nous nous sommes intéressés à l'amélioration de la prédiction de variables agroenvironnementales de STICS à partir de l'estimation des paramètres du sol sous un large panel de jeux d'observations (croisant différentes configurations agropédoclimatiques), lequel n'a jamais été aussi exhaustif dans les précédentes études. Le lien existant entre les variables observables et les variables à prédire nous a donc permis d'expliquer les performances de prédiction en fonction des performances d'estimation, compte tenu d'un jeu d'observations donné. Ce travail est présenté au Chapitre 5.

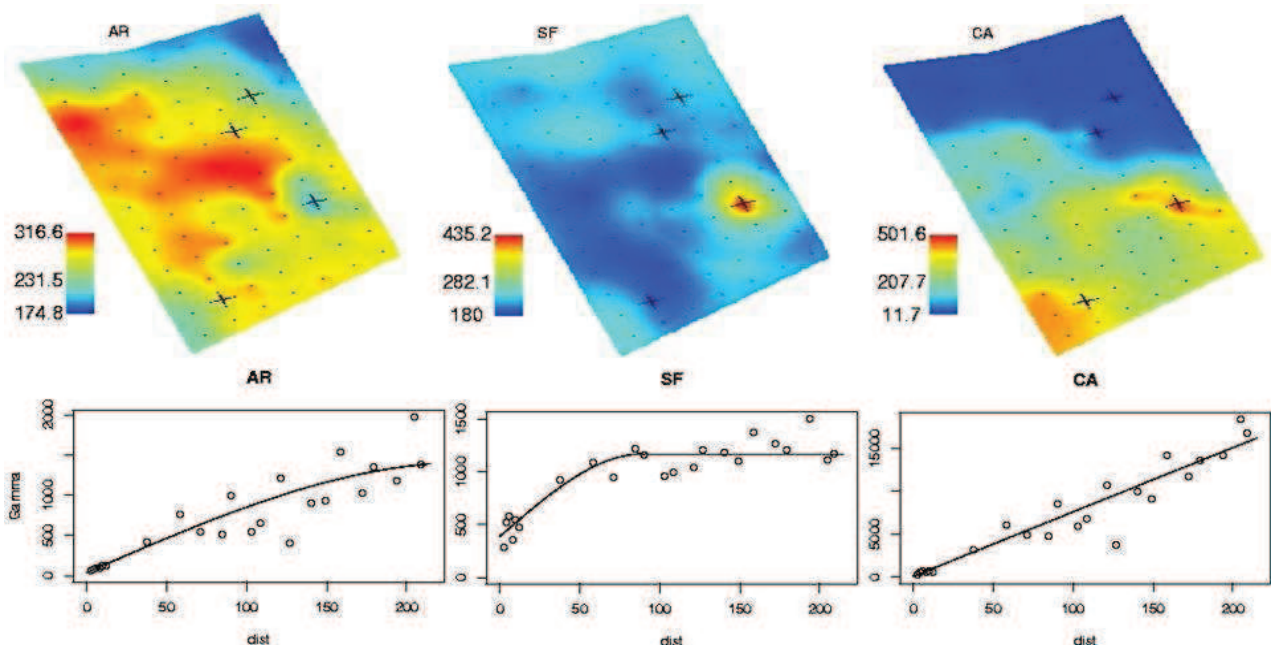
### **1.3.3. La distribution spatiale des paramètres du sol dans la parcelle est structurée et la prise en compte de cette structure peut améliorer l'estimation des paramètres**

Dans les travaux précédents, l'estimation des paramètres du sol par inversion d'un modèle de culture est entreprise sur chaque point de la parcelle agricole sans tenir compte de la structure spatiale des valeurs de paramètres (Houlès, 2004; Guérif et al., 2006). Certains auteurs travaillant à une échelle plus large (telle que le bassin de production) ont développé des approches utilisant des contraintes spatiales. Par exemple, Lauvernet et al. (2005) forcent les valeurs des paramètres liés aux caractéristiques de la plante à être égales sur chaque point de l'espace couvert par la même variété. Pour notre cas d'application, il pourrait être envisageable de forcer les valeurs des paramètres du sol à être égales sur chaque point de l'espace couvert par le même type de sol, mais les cartes de sol, nous le rappelons, sont très rares à l'échelle de la parcelle et sont très coûteuses. Il semble plus intéressant de considérer une information spatiale qui renseignerait sur la structure spatiale des propriétés des sols et que l'on pourrait introduire dans le processus d'estimation.

La disponibilité de nouveaux outils d'investigation pour l'étude du proche sous-sol s'appuyant sur des méthodes géophysiques ouvre des perspectives intéressantes dans ce domaine. Parmi elles, la mesure de la résistivité électrique des sols (Samouelian et al., 2005; Bourennane et al., 2007) qui représente la capacité d'un horizon de sol à limiter le passage du courant électrique, et est étroitement liée à ses caractéristiques intrinsèques. L'interprétation du signal fait l'objet de nombreux travaux et même s'il est encore très difficile de retrouver les propriétés des sols précisément (Samouelian *et al.*, 2005), il n'en demeure pas moins que cette méthode donne accès de manière automatisée à la structure spatiale de certaines caractéristiques du sol de la parcelle.

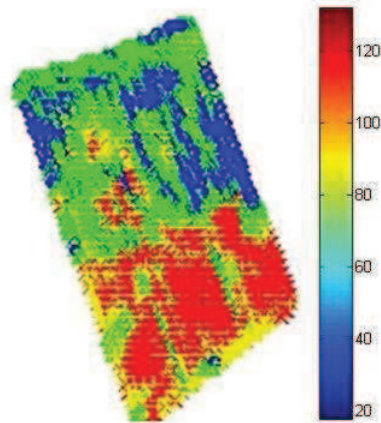
Prenons le cas d'une parcelle expérimentale du site agriculture de précision de Chambry (Guérif *et al.*, 2001), située dans l'Aisne et décrite par Nicoulaud et al. (2004). Cette parcelle présente une topographie relativement accentuée (7 mètres de dénivelé). Elle est située en position haute, favorisant ainsi l'érosion des matériaux. La répartition des sols dans le paysage est étroitement liée à la

topographie : le sommet de la parcelle et le versant orienté sud - sud ouest, soumis à l'érosion, sont occupés de sols calcaires développés sur la craie ou la craie sableuse magnésienne. Le versant opposé (nord-nord est) est occupé par des sols profonds moins calcaires développés sur la craie cryoturbée et de sols limoneux profonds développés sur des formations sablo-calcaires. Dans le cadre du projet agriculture de précision, un grand nombre de mesures du sol ont été faites sur cette parcelle, permettant d'accéder à sa structure spatiale. La Figure 1-2 illustre les types de structure spatiale rencontrés pour trois propriétés de l'horizon de surface de la parcelle en question, avec la carte des valeurs interpolées et le semi-variogramme associé (Chilès and Delfiner, 1999) : soit les valeurs sont corrélées quelle que soit la distance (pour la teneur en calcaire *CA* et dans une moindre mesure pour le taux d'argile *AR*), exprimant l'existence d'un gradient dans la parcelle, soit elles sont corrélées jusqu'à une certaine distance (jusqu'à environ 80 m pour le taux de sable fin *SF*), et indépendantes ensuite. La comparaison avec la Figure 1-3 montre qu'il existe une similitude de structure spatiale entre les valeurs de la teneur en calcaire *CA* et les valeurs de la résistivité électrique.



**Figure 1-2.** Semi-variogrammes de la teneur en argile (*AR*), en sable fin (*SF*) et en calcaire (*CA*), obtenus à partir de mesures du sol de la parcelle P2 de Chambry.

Résistivité électrique (Ohm.m)



**Figure 1-3.** Mesures de résistivité électrique (en Ohm.m) sur la parcelle P2 de Chambry.

La résistivité électrique du sol permet donc de renseigner sur la structure spatiale de certaines de ses propriétés. Dans le Chapitre 6, nous proposons une voie d'utilisation de cette information dans le processus d'estimation des paramètres et analysons les possibilités d'amélioration de l'estimation.

#### **1.3.4. Les questions de recherche prises en compte et organisation de la thèse**

L'estimation des paramètres du sol d'un modèle de culture afin d'améliorer les prédictions de variables agroenvironnementales n'est donc pas chose aisée. Nous avons vu, à travers l'état de l'art dressé dans ce chapitre, qu'un certain nombre de questions de recherche ont été révélées et méritent notre attention. Les réponses à ces questions permettraient alors de faire un meilleur usage du modèle de culture STICS pour la prédiction des variables d'intérêt.

Les quatre questions de recherche prises en compte dans cette thèse sont alors les suivantes :

1. Quel est le sous-groupe optimal de paramètres du sol à sélectionner pouvant être estimé convenablement ?
2. L'analyse de sensibilité globale permet-elle de mesurer la quantité d'information contenue dans les observations et la qualité d'estimation des paramètres ?

3. La prédiction des variables d'intérêt est-elle améliorée à partir de l'estimation des paramètres du sol ?
4. La prise en compte d'une structure spatiale dans l'estimation des paramètres du sol permet-elle d'améliorer leur estimation ?

Après un chapitre consacré à la présentation des outils et des données utilisés dans l'ensemble des travaux (Chapitre 2), la suite de la thèse s'articule autour de ces 4 questions. Elle s'achève par une conclusion générale, incluant un bilan des résultats, une discussion et la présentation de perspectives.



## **Chapitre 2. Outils et données**

L'objectif de ce chapitre est de présenter les principaux outils et données qui sont utilisés pour répondre aux questions de recherche prises en compte dans cette thèse. En premier lieu, les outils seront décrits : le modèle de culture STICS, une méthode d'analyse de sensibilité et une méthode d'estimation de paramètres. En second lieu, les données sur lesquelles les outils sont appliqués seront présentées.

### **2.1. Le modèle de culture STICS**

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, l'ouvrage de référence pourra être consulté (Brisson et al., 2008). Nous présenterons d'abord les différents modules de STICS et ensuite la manière dont le sol, siège des transferts d'eau et de nitrates, est décrit dans le modèle.

#### **2.1.1. Les différents modules de STICS, ses entrées et ses sorties**

Comme la majorité des modèles de culture, STICS fonctionne à un pas de temps journalier. Il simule le comportement du système sol-plante, en une dimension spatiale, à travers un ou plusieurs cycles culturaux. La limite supérieure du système est l'atmosphère, caractérisée par des variables climatiques (température, pluie, rayonnement,...) et la limite inférieure correspond au sol, lui-même défini par plusieurs couches. Les principales variables d'état du système sont la surface foliaire, la biomasse (carbone, eau, azote) totale et des organes récoltés (graines, fruits ou tubercules). Les organes végétatifs (feuilles, tiges, branches, talles ou racines) sont fonctionnellement séparés en termes de capteurs de rayonnement, d'eau, de nitrates ou de réservoir. Le sol et la plante interagissent via les racines,

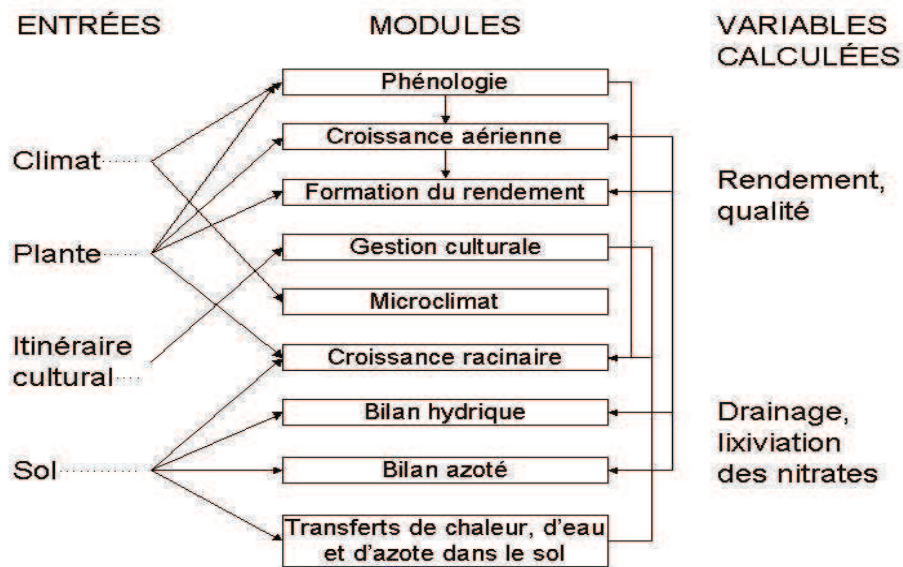
lesquelles sont définies en termes de distribution de densité de racines dans le profil de sol

La croissance de la culture est pilotée par l'accumulation de carbone dans la plante : le rayonnement solaire intercepté par le couvert végétal est d'abord transformé en croissance aérienne puis ensuite dirigé vers les organes récoltés pendant la phase finale du cycle cultural. Le contenu en azote de la plante dépend de l'accumulation de carbone et de la disponibilité en azote du sol. Suivant le type de plante simulé, le développement phénologique de la culture est soit piloté par un indice thermique (qui comptabilise le nombre de degré-jours), un indice photo-thermique (qui prend en compte les exigences photopériodiques), ou encore (dans le cas du blé) un indice vernalo-photo-thermique (qui prend également en compte les besoins en froid de la plante). Le module de développement de STICS détermine les stades clés de la dynamique de l'indice de surface foliaire et de celle des racines, et ceux de la mise en place du remplissage des organes récoltés. Les éventuels stressés (état de manque) en eau et en azote de la plante réduisent la croissance des feuilles et des racines et augmentent leur sénescence, réduisant aussi le taux de production de biomasse. D'autres types de stressés, tels que le stress thermique, sont également pris en compte par le modèle. Une attention particulière est placée sur l'effet de la gestion de la culture, via les techniques agricoles employées, sur la dynamique du système sol-plante.

Le modèle STICS est organisé en plusieurs modules, chacun composé de sous-modules qui traitent de mécanismes spécifiques. La Figure 2-1 présente les neuf modules de STICS ainsi que les groupes de paramètres pris en compte par chacun d'entre eux. Un premier groupe composé de trois modules traite de l'écophysiologie des parties aériennes de la plante : phénologie, croissance aérienne et formation du rendement. Un deuxième groupe composé de quatre modules traite de la façon dont le sol interagit avec les parties souterraines de la plante : croissance des racines, bilan hydrique, bilan azoté et transferts dans le sol, calcul des indices de stress hydrique et azoté. Le module de gestion de la culture traite des interactions qu'il existe entre les techniques agricoles employées (travail du sol, gestion des résidus de la culture précédente, fertilisation azotée, irrigation,..) et le système sol-

plante. Le module microclimat simule les effets combinés du climat et du bilan hydrique sur la température et l'humidité de l'air à l'intérieur du couvert végétal. A l'intérieur de chaque module, il existe des options de simulation qui peuvent être utilisées afin d'étendre l'utilisation du modèle STICS à des systèmes de culture variés. Ces options sont liées à l'écophysiologie de la plante et à la gestion de la culture, comme par exemple :

- la compétition entre les organes végétatifs et les organes récoltés en terme de croissance (ces deux types d'organes ne sont pas en compétition pour le blé alors qu'ils le sont pour la betterave),
- le choix ou non de la mise en place d'un paillage particulier (paillage qui peut être plastique ou bien végétal),
- le choix ou non de considérer un sol avec macroporosité (sol caractérisé par des pores de grande taille), avec des fissures ou avec des cailloux.



**Figure 2-1.** Les principaux modules du modèle de culture STICS, avec les groupes de paramètres associés, avec les entrées (variables et paramètres) et les variables de sortie. Les flèches de gauche traduisent les principales influences entre entrées et modules. Les flèches centrales symbolisent la croissance potentielle. Les flèches de droite représentent l'effet des stressés sur la croissance potentielle (Houlès, 2004).

L'entité élémentaire nécessaire à une simulation de STICS correspond à une zone géographique considérée comme homogène par rapport à tous les types d'entrées (voir Figure 2-1). Les valeurs moyennes parcellaires ont souvent été

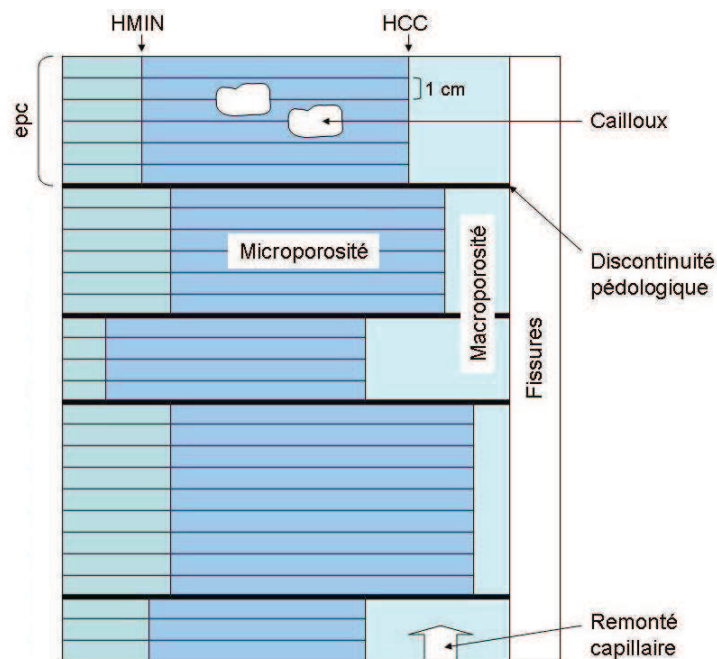
utilisées dans la mise au point du modèle pour estimer les paramètres des équations. Dans ce cas, l'unité de simulation correspond à la parcelle. Le modèle peut être cependant appliqué à des entités spatiales plus précises, définies par l'hétérogénéité des entrées sol dans la parcelle, comme cela est le cas pour l'agriculture de précision (Timlin et al., 2001; Braga and Jones, 2004; Houlès, 2004).

Les principales variables de sortie du modèle STICS sont des variables agroenvironnementales qui décrivent le comportement du système sol-plante au pas de temps journalier. Parmi elles, l'indice foliaire *LAI*, la teneur en azote de la plante *QN*, le rendement, la teneur en protéine du grain et la teneur en azote du sol. Les trois premières variables peuvent être observables par des méthodes automatiques (dérivées de la télédétection pour le *LAI* et *QN* et système embarqué sur tracteur pour le rendement). Pour qu'il soit possible de simuler les variables de sortie de STICS, toutes les entrées du modèle doivent être renseignées. Ces entrées sont constituées de variables climatiques mesurées au pas de temps journalier (pluies, températures, rayonnement global,...) et de paramètres regroupés en 3 grands groupes. Le premier groupe concerne les paramètres liés au génotype de la plante cultivée (incluant celles propres à la variété de la plante), le second groupe concerne tous les paramètres qui définissent le travail technique de l'agriculteur (travail du sol, semis, irrigation, fertilisation, techniques particulières,...) et le troisième groupe caractérise les propriétés du sol considéré, avec notamment les paramètres liés au processus de minéralisation et à la disponibilité en eau du sol. Dans cette thèse, nous rappelons que les paramètres liés au génotype de la plante sont calibrés au préalable (voir Annexe pour le blé et Launay et al., 2009 pour la betterave) et que ceux définissant le travail technique sont supposés être correctement renseignés. Les paramètres qui décrivent le profil de sol, via ses propriétés minérales et hydriques, sont le plus souvent mal connus en tout point de l'espace et font l'objet central de cette thèse.

## 2.1.2. La représentation du sol, ses principaux processus et paramètres associés

### *Les transferts d'eau et de nitrate*

Les transferts d'eau et de nitrates dans le sol sont décrits dans le modèle STICS selon une analogie de type réservoir, par des flux essentiellement descendants (des flux ascendants par capillarité en conditions sèches ou par saturation des couches profondes en cas d'excès d'eau peuvent être décrits également). La capacité de réservoir du sol, définie par son espace poral, peut être représentée (Figure 2-2) par un ensemble allant jusqu'à quatre compartiments : la microporosité, la macroporosité, les fissures et les cailloux. La microporosité du sol, dont la description seule est obligatoire, constitue la base de calcul pour les transferts d'eau et d'azote dans le sol. Dans cette thèse, seul ce compartiment a été considéré dans notre étude.



**Figure 2-2.** Description du profil de sol en compartiments (d'après Brisson et al., 2008)

Le transfert d'eau dans la microporosité du sol est calculé par couche élémentaire de 1 cm, considérée comme un réservoir. L'eau, apportée par les pluies ou l'irrigation, remplit successivement les couches élémentaires de sol avec un flux descendant, sachant que la limite du réservoir correspond à l'humidité à la capacité

au champ (*HCC*). Le sol est décrit par une succession couches de sol (divisées elles-mêmes en couches élémentaires), de différentes épaisseurs (*epc*) et délimitées par une discontinuité pédologique. Ces couches sont caractérisées par des propriétés permanentes du sol, comme les humidités seuils qui définissent la capacité de rétention d'eau du réservoir constitué par la microporosité : l'humidité à la capacité au champ *HCC* (correspondant à un potentiel faible de l'eau dans le sol de l'ordre de 50 à 100 hPa), l'humidité au point de flétrissement *HMIN* (correspondant à un potentiel élevé de l'eau dans le sol de l'ordre de  $1.5 \times 10^4$  hPa), ainsi que la densité volumique du sol (*DA*). Le transfert de nitrates est également décrit suivant le principe du réservoir. Tout nitrate qui arrive dans la couche élémentaire par convection avec l'eau est mélangé avec le nitrate déjà présent, et l'eau en excès descend vers la couche élémentaire suivante avec la nouvelle concentration du mélange. Les quantités d'eau drainée (naturellement ou artificiellement) et d'azote lessivé qui s'échappent de la base du profil de sol ne sont donc plus disponibles pour la culture.

#### *La minéralisation de la matière organique*

Il est considéré dans le modèle que l'essentiel de la matière organique du sol (*C* et *N*) constituée par l'humus et les résidus de culture (et des processus de sa décomposition) se situe dans la couche de surface du sol. L'épaisseur de cette couche (*profhum*) varie en fonction de l'histoire de la parcelle et du travail du sol, mais est généralement de l'ordre de 30 cm. Les processus de minéralisation sont essentiellement sous la dépendance des pools de matière organique présents (certains pools sont à évolution lente, comme celui lié à l'humus *Norg* considéré dans notre étude comme étant une caractéristique permanente, et d'autres à évolution rapide comme la quantité de résidus incorporés), de la quantité d'azote minéral disponible, mais aussi des régimes hydriques et thermiques du sol. Quant aux teneurs en argile (*argi*) et en calcaire (*calc*) de la couche de surface, elles contribuent à diminuer la vitesse de minéralisation de la matière organique.

Pour finir, il faut noter que d'autres propriétés permanentes du sol interviennent dans les processus du sol et correspondent également à des paramètres du modèle :

- l'albédo *albedo* du sol sec nu, qui joue un rôle dans le bilan d'énergie lié à l'évaporation de l'eau du sol (qui dépend essentiellement de la composition minéralogique et de la teneur en matière organique du sol),
- le paramètre  $q_0$ , qui représente le seuil d'évaporation cumulée au-delà duquel le taux d'évaporation de l'eau du sol nu en fonction de l'évaporation potentielle est réduit.

## 2.2. La méthode d'analyse de sensibilité Extended FAST

Le modèle STICS étant un modèle complexe non linéaire avec un temps d'exécution de l'ordre de la seconde, nous avons choisi de mettre en œuvre la méthode Extended FAST, basée sur la décomposition de la variance des sorties du modèle générée par les paramètres du sol. On retrouvera une discussion de ce choix dans les articles qui constituent les Chapitre 3 et 4. Nous commencerons par présenter l'écriture de la décomposition de la variance, à partir de laquelle les indices de sensibilité (mesures de l'effet d'un paramètre sur la variable de sortie) sont issus. Nous finirons par la présentation de la méthode Extended FAST, permettant d'estimer ces indices.

### 2.2.1. La décomposition de la variance

L'analyse de variance (ANOVA) est basée sur la décomposition de la variabilité de la réponse du modèle  $f$  entre les différentes contributions des paramètres d'entrée  $\theta_i, i=1, \dots, P$ , et des interactions qui interviennent entre eux. Cette décomposition est basée sur la théorie statistique des modèles linéaires. La décomposition de la variance présentée ici est une extension de l'ANOVA aux modèles non-linéaires (Monod et al., 2006).

Soit  $Y = f(\theta)$  la variable de sortie du modèle  $f$ ,  $\theta$  appartenant au sous-espace  $\Theta$  de  $\mathfrak{R}^P$  et l'on suppose que les paramètres sont indépendants entre eux. La décomposition de Sobol' de  $f$  (Sobol, 1993) s'écrit ainsi :

$$Y = \mu + \sum_i f_i(\theta_i) + \sum_i \sum_{j \neq i} f_{i,j}(\theta_i, \theta_j) + \dots + f_{1,2,\dots,P}(\theta_1, \theta_2, \dots, \theta_P) \quad (2-1)$$

où  $\mu = E(Y)$  représente l'espérance mathématique de  $Y$ . Si les paramètres sont indépendants entre eux, alors les termes de l'Equation (2-1) sont orthogonaux et la décomposition est unique. En faisant passer  $\mu$  à gauche, en élevant au carré et en intégrant par rapport à  $\theta \in \Theta$ , l'Equation (2-1) devient :

$$\int_{\Theta} (Y - \mu)^2 d\theta = \sum_i \int_{\Theta} f_i^2(\theta_i) d\theta + \sum_i \sum_{j \neq i} \int_{\Theta} f_{i,j}^2(\theta_i, \theta_j) d\theta + \dots + \int_{\Theta} f_{1,2,\dots,P}^2(\theta_1, \theta_2, \dots, \theta_P) d\theta \quad (2-2)$$

où  $\int_{\Theta} (Y - \mu)^2 d\theta$  est la variance de  $Y$ . L'Equation (2-2) s'écrit encore :

$$\text{var}(Y) = \sum_i D_i + \sum_i \sum_{j \neq i} D_{i,j} + \dots + D_{1,2,\dots,P} \quad (2-3)$$

où  $D_i$  est la contribution principale du paramètre  $i$  à la variance de  $Y$ ,  $D_{i,j}$  est la contribution de l'interaction entre les paramètres  $\theta_i$  et  $\theta_j$  et  $D_{1,2,\dots,P}$  est la contribution de l'interaction entre tous les paramètres. La quantité  $\text{var}(Y)$  représente la variabilité de  $Y$  par rapport à toute l'incertitude sur les paramètres d'entrée, ce qui permet une meilleure prise en compte des irrégularités et des non-linéarités du modèle  $f$  qu'une simple ANOVA.

### 2.2.2. Les indices de sensibilité

Les indices de sensibilité sont issus de la décomposition (3) en divisant les contributions par  $\text{var}(Y)$  :

$$S_i = \frac{D_i}{\text{var}(Y)}$$

$$S_{i,j} = \frac{D_{i,j}}{\text{var}(Y)}$$

...

Par conséquent, les indices de sensibilité satisfont  $\sum_i S_i + \sum_i \sum_{j \neq i} S_{i,j} + \dots + S_{1,2,\dots,P} = 1$

et sont interprétés comme une part de variance de  $Y$  expliquée.

Pour chaque paramètre, deux types d'indices de sensibilité peuvent être définis : l'indice principal  $S_i$ , représentant l'effet principal de  $\theta_i$  sur la variance de  $Y$ ,



et l'indice total  $ST_i$ , représentant la somme de tous les effets (principal et interactions) impliquant  $\theta_i$  sur la variance de  $Y$ . L'indice total  $ST_i$  s'écrit alors :

$$ST_i = S_i + \sum_{j \neq i} S_{i,j} + \dots + S_{1,2,\dots,P}$$

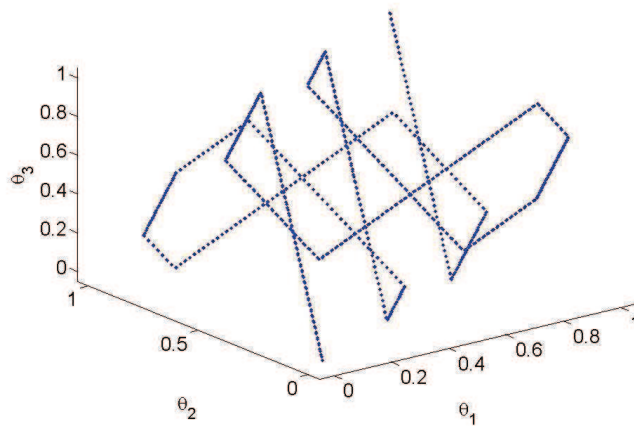
Notons que  $ST_i$  peut être écrit comme  $1 - S_{-i}$  où  $S_{-i} = D_{-i} / \text{var}(Y)$  représente la somme de tous les indices qui n'impliquent pas  $\theta_i$ .

### 2.2.3. Estimation des indices de sensibilité par la méthode Extended FAST

L'idée de la méthode Fourier Amplitude Sensitivity Test (Cukier et al., 1973; Cukier et al., 1975; Cukier et al., 1978) est de transformer l'intégrale multidimensionnelle  $E(Y) = \int_{\Theta} f(\theta) d\theta$ , servant à calculer l'espérance et la variance de la variable de sortie  $Y$ , en une intégrale unidimensionnelle  $E(Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds$ . Pour ce faire, on procède au changement de variable  $\theta_i(s) = G_i(\sin(\omega_i s + \varphi_i))$ ,  $i=1, \dots, P$ , où  $s \in [-\pi, \pi]$ ,  $\omega_i$  est un jeu de fréquences angulaires entières dont le choix sera discuté plus loin,  $\varphi_i$  est un décalage de phase tiré aléatoirement dans  $[0, 2\pi[$  et  $G_i$  est appelée fonction de transformation. Le décalage  $\varphi_i$  est optionnel et permet d'introduire de l'aléatoire dans le parcours déterministe de  $\Theta$  défini par  $G_i$ . Concernant cette fonction de transformation, plusieurs formalismes ont été proposés (Cukier et al., 1973; Koda et al., 1979; Saltelli et al., 1999) ; nous adopterons tout au long de ce travail celui qui est le plus utilisé et qui est proposé par Saltelli (1999) :

$$\theta_i(s) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s + \varphi_i)) \quad (2-4)$$

Les variations de  $s$  dans  $[-\pi, \pi]$  forment ainsi une courbe fermée dans  $\Theta$  appelée "search curve" pouvant passer arbitrairement près de tout point de  $\Theta$ . La Figure 2-3 montre l'exemple d'une courbe définie par (2-4) pour 3 paramètres  $(\theta_1, \theta_2, \theta_3)$  suivant une loi uniforme dans  $\Theta = [0,1]^3$ , avec la transformation ci-dessus et  $\omega_1 = 3$ ,  $\omega_2 = 7$ ,  $\omega_3 = 11$ ,  $\varphi_1 = \varphi_2 = \varphi_3 = 0$ .



**Figure 2-3.** Exemple de courbe définie par l'Equation (2-4) pour 3 paramètres suivant une loi uniforme.

Par la transformation (2-4), le modèle  $f$  devient une fonction périodique en  $s$  et peut ainsi être décomposé en séries de Fourier telles que :

$$f(s) = \sum_{r=-\infty}^{+\infty} (A_r \cos(rs) + B_r \sin(rs)) \quad (2-5)$$

où  $A_r$  et  $B_r$  sont les coefficients de Fourier. D'après l'analyse de Fourier, la décomposition (2-5) est unique quelque soit  $s$  et les spectres de la série calculés suivant  $\Lambda_r = A_r^2 + B_r^2$  caractérise l'importance des oscillations du modèle  $f$  pour la fréquence  $r/2\pi$ . Le théorème de Parseval permet d'exprimer la variance totale de  $Y$  comme la somme des spectres de la série de Fourier :

$$\text{var}(Y) = 2 \sum_{r=1}^{+\infty} \Lambda_r \quad (2-6)$$

Le principe de l'analyse de sensibilité selon la méthode FAST est de regarder l'importance des oscillations de  $f$  lorsque l'on fait osciller les paramètres d'entrée  $\theta_i$ ,  $i=1, \dots, P$ , à différentes fréquences  $\omega_i$ . Cette importance représente la contribution de chaque paramètre  $i$  à la variance de  $Y$  et elle est calculée comme la somme des éléments du spectre  $\Lambda_r$  correspondant à la fréquence  $\omega_i$  de  $\theta_i$  et à ses plus grands harmoniques (ie :  $q\omega_i$ ,  $q$  entier) :

$$D_i = 2 \sum_{q=1}^{+\infty} \Lambda_{q\omega_i} \quad (2-7)$$

L'Equation (2-7) découle du théorème de Parseval (2-6) et permet ainsi de calculer les indices de sensibilité principaux tels que  $S_i = D_i / \text{var}(Y)$ .

En ce qui concerne le calcul de l'indice de sensibilité total du paramètre, Saltelli (1999) propose une extension de la méthode EFAST en affectant une fréquence forte à  $\theta_i$  et une fréquence faible aux autres paramètres  $\theta_{\sim i}$ . Le choix d'une telle fréquence sera discuté plus loin. D'après Saltelli, l'interaction entre  $\theta_i$  et  $\theta_{\sim i}$  est ainsi concentrée autour des pics de fréquences  $\omega_i$  et il considère que l'information sur tous les effets d'interaction impliquant  $\theta_{\sim i}$  et non  $\theta_i$  est contenue dans les fréquences inférieures au seuil  $\omega_i / 2$  :

$$\hat{D}_{\sim i} = 2 \sum_{r=1}^{\omega_i / 2} \Lambda_r \quad (2-8)$$

Les effets d'interaction impliquant  $\theta_{\sim i}$  et non  $\theta_i$  sont donc négligeables au-delà de  $\omega_i / 2$ . L'Equation (2-8) permet alors de calculer  $S_{\sim i} = \hat{D}_{\sim i} / \text{var}(Y)$  et  $ST_i = 1 - S_{\sim i}$ .

D'un point de vue pratique, le calcul des indices de sensibilité est fait de la manière suivante, où l'on commence par explorer de façon discrète l'espace de  $s \in [-\pi, \pi]$  :

$$s_k = -\pi + (2k - 1)\pi / N_s, \quad k=1, \dots, N_s$$

afin de calculer les coefficients de Fourier :

$$A_r = \frac{1}{N_s} \sum_{k=1}^{N_s} f(s_k) \cos(rs_k) \\ B_r = \frac{1}{N_s} \sum_{k=1}^{N_s} f(s_k) \sin(rs_k) \quad , \quad r \in \left[ -\frac{N_s - 1}{2}, \frac{N_s - 1}{2} \right] \text{ entier naturel}$$

et le spectre de la série  $\Lambda_r = A_r^2 + B_r^2$ , nécessaire au calcul de la variance totale

$$\text{var}(Y) \approx 2 \sum_{r=1}^{(N_s - 1)/2} \Lambda_r \quad \text{et des contributions} \quad \hat{D}_i = 2 \sum_{q=1}^M \Lambda_{q\omega_i} \quad , \quad \text{où l'on suppose que les}$$

harmoniques supérieurs à  $M$  sont négligeables. Dans la littérature, la valeur de  $M$  est le plus souvent prise égale à 4.

Pour finir, on définit le choix des fréquences  $\omega_i$ . Afin d'éviter le phénomène de confusion entre les différentes fréquences, Saltelli (1999) propose un algorithme automatique permettant de choisir ces fréquences. Ainsi, pour chaque paramètre  $\theta_i$  :

- (i) on choisit une valeur  $\omega_i$  telle que sa valeur maximale possible  $\omega_i^{\max}$  satisfasse  $N_s \geq 2M\omega_i^{\max} + 1$  (formule de Nyquist-Shannon),
- (ii) on détermine  $\omega_{\sim i}$  pour les autres paramètres  $\theta_{\sim i}$  de façon que la valeur maximale affectée  $\omega_{\sim i}^{\max}$  satisfasse  $\omega_{\sim i}^{\max} = \omega_i / 2\pi$  et que ces fréquences puissent varier entre 1 et  $\omega_{\sim i}^{\max}$  avec le pas le plus grand possible.

L'algorithme sera appliqué tour à tour sur chaque paramètre d'entrée  $i$  afin de déterminer un jeu de fréquences  $\{\omega_i, \omega_{\sim i}\}$  et d'estimer les indices de sensibilité  $S_i$  et  $ST_i$  associés à  $\theta_i$ . Le nombre de simulations du modèle  $f$  nécessaires à la méthode pour calculer les indices  $S_i$  et  $ST_i$  de chaque paramètre est de  $P \times N_s$ . La méthode Extended FAST a été implémentée sous Matlab<sup>®</sup> dans le cadre d'un projet informatique *MultisimLib* du laboratoire EMMAH de l'INRA d'Avignon, projet qui vise entre autre à proposer des méthodes d'analyse de sensibilité pour le modèle STICS (bibliothèque *SensiLib*).

## 2.3. Présentation du théorème de Bayes et de la méthode Importance Sampling

La méthode d'estimation retenue pour estimer les paramètres du sol par inversion du modèle STICS est la méthode Importance Sampling, basée sur l'approche Bayésienne.

### 2.3.1. Le théorème de Bayes

La densité a posteriori des paramètres est calculée à partir du théorème de Bayes :

$$\pi(\theta / Y) = \frac{\pi(Y / \theta)\pi(\theta)}{\pi(Y)} \quad (2-9)$$

où  $Y$  est le vecteur des observations, de taille  $K$ ,  $\pi(\theta/Y)$  est la densité a posteriori des paramètres,  $\pi(\theta)$  est la densité a priori,  $\pi(Y)$  est une constante qui permet à l'intégrale  $\pi(\theta/Y)$  de valoir 1 sur l'espace des paramètres, et  $\pi(Y/\theta)$  est la fonction de vraisemblance du jeu de données  $Y$  conditionnellement aux paramètres  $\theta$ . Sa valeur est déterminée à partir de la densité de probabilité des écarts entre données simulées et données observées. La vraisemblance dépend à la fois de la densité a priori des paramètres et du jeu de données  $Y$ .

### 2.3.2. La méthode Importance Sampling

Le principe de la méthode Importance Sampling (Beven and Binley, 1992; Beven and Freer, 2001) est d'approximer la densité a posteriori des paramètres  $\pi(\theta/Y)$ , donnée par l'Equation (2-9), par une densité de probabilité discrète telle que :

$(\theta_n, p_n), n=1, \dots, N, \sum_{n=1}^N p_n$ , où  $p_n$  est la probabilité associée au vecteur de paramètres  $\theta_n$ . Pour ce faire, la méthode procède en 3 étapes :

- (i) génération aléatoire de  $N$  vecteurs  $\theta_n, n=1, \dots, N$ , à partir de la densité a priori des paramètres  $\pi(\theta)$ , puis réalisation avec le modèle de  $N$  simulations des variables observées,
- (ii) calcul de la valeur de la vraisemblance  $\pi(Y/\theta_n)$  ainsi que de la valeur de la densité a priori  $\pi(\theta_n), n=1, \dots, N$ , associées aux  $N$  vecteurs de paramètres  $\theta_n$  générés,
- (iii) calcul de la probabilité a posteriori associée au vecteur  $\theta_n$  :

$$p_n = \frac{\pi(Y/\theta_n)\pi(\theta_n)}{\sum_{m=1}^N \pi(Y/\theta_m)\pi(\theta_m)} \quad (2-10)$$

Lorsqu'on considère une densité a priori uniforme, l'Equation (2-10) devient

$$p_n = \frac{\pi(Y/\theta_n)}{\sum_{m=1}^N \pi(Y/\theta_m)}. \text{ Les couples } (\theta_n, p_n), n=1, \dots, N, \text{ approchent ainsi la densité a}$$

*posteriori* et peuvent être utilisés afin d'en déterminer diverses caractéristiques, comme :

- la moyenne de la densité a posteriori du paramètre  $\theta_i$ ,  $\bar{\theta}_i^{post} = \sum_{n=1}^N \theta_{i,n} p_n$

- la variance de la densité a posteriori du paramètre  $\theta_i$ ,  $\delta_i^2 = \sum_{n=1}^N \theta_{i,n}^2 p_n - (\bar{\theta}_i^{post})^2$

- la corrélation entre deux paramètres  $\theta_i$  et  $\theta_j$ ,  $\frac{\sum_{n=1}^N (\theta_{i,n} - \bar{\theta}_i^{post})(\theta_{j,n} - \bar{\theta}_j^{post}) p_n}{\delta_i \cdot \delta_j}$

La moyenne de la densité a posteriori est considérée comme estimateur de la valeur du paramètre  $\theta_i$  étant donné le jeu de données  $Y$ , la variance représente la certitude que l'on peut avoir en l'estimation du paramètre (une petite variance révèle une grande certitude) et le coefficient de corrélation permet d'évaluer la structure d'interaction d'ordre deux (entre deux paramètres), non observable à partir des indices de sensibilité (Ratto et al., 2001), et qui complète la représentation de l'interaction donnée par une méthode d'analyse de sensibilité globale. Par exemple, lorsqu'un coefficient positif est détecté, le couple de paramètres  $(\theta_i, \theta_j)$  s'exprime dans le modèle sous la forme d'un quotient ou d'une différence, et lorsqu'il est négatif, le couple s'exprime sous la forme d'un produit ou d'une somme (Ratto et al., 2001; Kanso et al., 2004).

L'écriture de la vraisemblance  $\pi(Y/\theta)$  est basée sur la définition d'un modèle d'erreurs entre les observations et les sorties du modèle. Le modèle d'erreurs  $\varepsilon_k$  que l'on considère dans ce travail est le suivant :

$$\varepsilon_k = y_k - f_k(\theta, x) \quad (2-11)$$

où  $y_k$  est la donnée observée de la variable  $k$ ,  $x$  est le vecteur de variables explicatives et  $f_k$  est la simulation du modèle pour la variable  $k$ . On suppose que les  $\varepsilon_k$  sont indépendants entre les différentes variables et qu'ils suivent une loi de distribution normale de moyenne 0 et d'écart-type  $\sigma_k$ . A partir de l'Equation (2-11) et des hypothèses précédentes, la fonction de vraisemblance  $\pi(Y/\theta)$  du jeu de données  $Y$  conditionnellement à  $\theta$  s'écrit :

$$\pi(Y / \theta) = \prod_{k=1}^K \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2} [y_k - f_k(\theta, x)]^2\right\} \quad (2-12)$$

On suppose de plus que les informations a priori sont indépendantes, ce qui revient à supposer que les paramètres sont indépendants entre eux. Il n'est pas nécessaire de fournir une information a priori conjointe sur les paramètres, difficile à définir dans notre cas d'application. Ainsi, la densité a priori  $\pi(\theta)$  est le produit des densités a priori marginales. Pour finir, l'application de la méthode Importance Sampling requiert la connaissance du nombre  $N$  de vecteurs générés : ce nombre est croissant en fonction du nombre de paramètres à estimer et est généralement fixé de façon empirique. La méthode Importance Sampling a été implémentée sous Matlab<sup>®</sup>, parallèlement à Extended FAST, afin d'estimer des paramètres de STICS à partir d'observations.

#### 2.4. Présentation des données utilisées

Notre objectif consiste à estimer de façon spatialisée les propriétés des sols à partir de l'inversion du modèle STICS et d'observations sur les couverts végétaux, ainsi que d'évaluer l'intérêt de ces estimations pour la prédiction de variables agroenvironnementales. Les données nécessaires pour réaliser ce travail sont les suivantes :

- des observations du couvert végétal, de type *LAI*, teneur en azote de la plante *QN*, rendement *Yld* et teneur en protéine des grains *Prot*, ainsi que des observations de la teneur en azote minéral du sol *Nit* à la récolte,
- des mesures ou des estimations des caractéristiques des sols correspondant aux paramètres sol du modèle STICS et des données auxiliaires sur la résistivité électrique qui seront utilisées au Chapitre 6.

Dans ce travail, les mesures utilisées ont été acquises à 2 échelles différentes :

- à haute résolution : sur un dispositif d'agriculture de précision situé à Chambry, près de Laon (Aisne), deux parcelles d'agriculteur de 10 ha chacune (Guérif, 2001) ont fait l'objet de mesures intensives sur des points de grille espacés de

36 m, ou des surfaces élémentaires de l'ordre de 5 à 20 mètres. L'unité de simulation est ici le pixel, représentant environ 0.04 ha. Les cultures étudiées sur ce dispositif sont des cultures de blé tendre d'hiver et de betterave sucrière,

- à moyenne résolution : sur un ensemble de 21 parcelles constituant un bassin versant de 145 ha situé à Bruyères (Aisne), 31 sites ont été définis en fonction de leur appartenance à une parcelle et à une unité de sol données. Chacun des sites a fait l'objet de mesures ciblées. L'unité de simulation est ici le site, représentant en moyenne  $145/31=4.68$  ha.

Les sites de Chambry et de Bruyères sont proches et on y retrouve des formations pédologiques voisines. Les sols sont majoritairement des luvisols, calcosols et colluviosols (AFES, 1995). Des phénomènes périglaciaires de cryoturbation et de solifluxion ont remanié localement ces matériaux. Les limons éoliens ont évolué en sols bruns lessivés tout en étant soumis à l'érosion : ceux situés sur les pentes convexes eux ont été amincis et ceux situés dans les bases de pente ont été épaissis. Les parcelles agricoles sont donc hétérogènes avec des variations notoires apparaissant à l'échelle hectométrique, voire décamétrique.

Ces différentes observations nous ont servi :

- (i) à définir les gammes de variation des paramètres du sol pour les analyses de sensibilité (Chapitre 3 et 4) ainsi que leur information a priori pour l'estimation par inversion (Chapitre 4, 5 et 6),
- (ii) à valider les approches d'estimation de paramètres et de prédiction des variables d'intérêt (ponctuellement sur Bruyères au Chapitre 5 et spatialement sur Chambry au Chapitre 6),
- (iii) à déterminer une structure spatiale sur les parcelles de Chambry, considérée ensuite dans la procédure d'estimation des paramètres du sol (Chapitre 6),
- (iv) à générer des observations synthétiques utilisées pour étudier le lien entre l'analyse de sensibilité globale et la qualité d'estimation des paramètres (Chapitre 4), pour expliquer la qualité de prédiction des variables par la qualité d'estimation des paramètres (Chapitre 5) ainsi que pour estimer spatialement les paramètres (Chapitre 6).



Les observations synthétiques sont des simulations bruitées (simulations perturbées par de l'erreur d'observation supposée connue) du modèle, à partir de valeurs connues de paramètres. Ce type d'observation permet d'évaluer la méthode utilisée (estimation de paramètre, analyse de sensibilité, ...) tout en négligeant l'erreur modèle, très délicate à évaluer pour les modèles de culture.

Nous allons à présent présenter les différents types de données acquises sur les propriétés des sols et sur les couverts végétaux.

#### **2.4.1. Le contexte pédologique et les mesures du sol**

##### *Les propriétés des sols sur deux parcelles de Chambry*

La brève présentation des mesures des propriétés des sols faite ici est extraite de Nicoullaud (2007) et Beaudoin (2007), lesquels travaux peuvent être consultés pour plus d'informations. La parcelle 1, présente une topographie peu accentuée (3 m de dénivelé). Le substrat (base sur laquelle repose le sol) est constitué de craie cryoturbée (modification du sol sous l'effet du gel) remaniée par un diluvium crayeux présentant fréquemment une grève calcaire (horizon formé d'un mélange de graviers et de sables calcaires). Ces niveaux ont été recouverts par des dépôts éoliens de limons et de sables de Sissonne. Il résulte de cette histoire géologique complexe une forte hétérogénéité spatiale des dépôts, d'ordre métrique. La parcelle 2, présente une topographie plus nette (7 m de dénivelé). Le substrat est composé de craie blanche avec un niveau supérieur de craie magnésienne. La craie a également subi une cryoturbation et elle est recouverte par des sables issus des sables du Thanétien (ère tertiaire) et par des limons éoliens. Le sommet de la parcelle et le versant sud - sud ouest, soumis à l'érosion éolienne, sont occupés par des sols calcaires. En revanche, le versant opposé est occupé par des sols profonds limoneux et moins calcaires. Pour une présentation plus complète de ces deux parcelles, il est possible de se référer à Nicoullaud (2007).

Paramètre	Signification	Acquisition
<i>argi</i>	Teneur en argile (après décarbonatation) de l'horizon de surface (%)	Mesures
<i>Norg</i>	Teneur en azote organique de l'horizon de surface (%)	Mesures
<i>calc</i>	Teneur en calcaire de l'horizon de surface (%)	Mesures
<i>albedo</i>	Albedo (énergie solaire réfléchie / énergie solaire incidente) de l'horizon de surface	Dérivé de <i>calc</i> et <i>argi</i>
<i>q0</i>	Seuil de cumul d'évaporation journalière (mm)	Dérivé de <i>argi</i>
<i>epc(h)</i>	Epaisseur de l'horizon <i>h</i> (cm)	Carte
<i>DA(h)</i>	Densité volumique de l'horizon <i>h</i> (g/cm <sup>3</sup> )	Carte + FPT
<i>HCC(h)</i>	Teneur en eau à la capacité au champ de l'horizon <i>h</i> (g/g)	Carte + FPT
<i>HMIN(h)</i>	Teneur en eau au point de flétrissement de l'horizon <i>h</i> (g/g)	Carte + FPT
<i>Hinit(h)</i>	Teneur en eau dans l'horizon <i>h</i> au début de la simulation (%)	Mesures
<i>NO3init(h)</i>	Azote présent dans l'horizon <i>h</i> au début de la simulation (kg/ha)	Mesures

**Tableau 2-1.** Présentation des paramètres sol de STICS retenus et leur mode d'acquisition.

Le Tableau 2-1 reprend les différents paramètres caractéristiques du sol de STICS que nous avons retenus pour la première analyse (voir Chapitre 3). La troisième colonne précise le mode d'acquisition de la valeur du paramètre en question, valeur qui peut être mesurée par analyse de sol, ou bien obtenue à partir d'une carte de sols couplée à des fonctions de pédotransfert (notées FPT) qui permettent de faire le lien entre certaines propriétés mesurées du sol (la granulométrie, la teneur en matière organique, en calcaire,...) et le paramètre d'intérêt. Une large bibliographie (voir revue de Bastet et al., 1998) existe sur les fonctions de pédotransfert les plus travaillées qui concernent le lien entre la granulométrie et les humidités caractéristiques hydriques et hydrodynamiques des

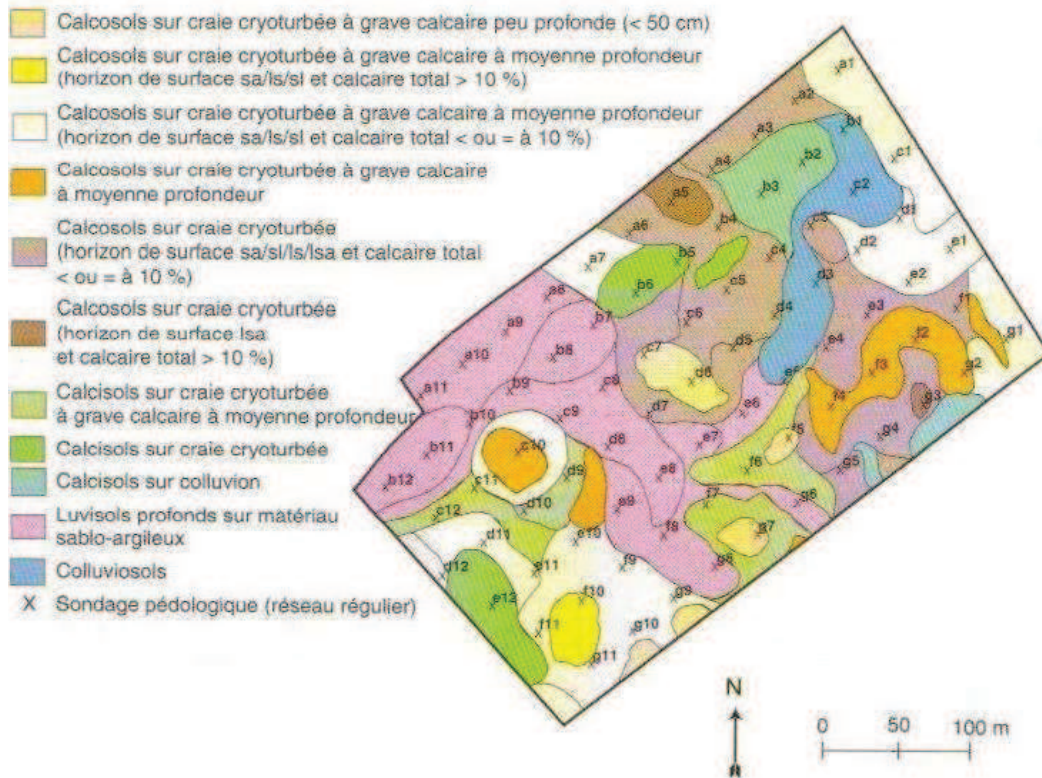
sols (dans notre cas, *HCC* et *HMIN*). Nous avons pour notre part utilisé des FPT établies localement par rapprochement entre unités de sol et propriétés mesurées au laboratoire (*HMIN*) ou in situ (*HCC*, *DA*) (Beaudoin et al., 2007). D'autres fonctions de pédotransfert ont été utilisées pour le calcul des paramètres *albedo* et *q0* de STICS (Brisson et al., 2008).

Des mesures de granulométrie et de contenu en matière organique et calcaire de l'horizon de surface ont été réalisées, permettant de déterminer les paramètres *argi*, *Norg* et *calc*. Elles ont été acquises par une analyse physico-chimique (en laboratoire) des prélèvements de sol réalisés sur les nœuds d'une grille espacés de 36 m (voir la Figure 2-4 : 81 nœuds pour la parcelle 1, 85 pour parcelle 2). Les teneurs en argile après décarbonatation ont été déterminées sur quelques échantillons de sol sur les deux parcelles, permettant d'établir une relation empirique locale entre teneur en argile avant et après décarbonatation (Beaudoin et al., 2007), pour ensuite déterminer *argi* sur l'ensemble de la grille.

Des mesures des variables d'état des sols (teneur en eau et en azote) ont été réalisées à différentes dates (semis, sortie d'hiver pour le blé, post-récolte) sur les nœuds de la grille : des prélèvements ont été réalisés par tranches verticales de 30 cm de sol, jusqu'à une profondeur de 150 cm (lorsque possible) puis analysés au laboratoire. Ces mesures permettent de définir les paramètres *Hinit* et *NO3init* pour chaque horizon de sol. Elles permettent également de définir la valeur de la variable observable *Nit* (teneur en azote du sol) à la récolte, laquelle constitue une variable d'intérêt agroenvironnemental de notre étude.

Une prospection pédologique a également été réalisée permettant de décrire les différents horizons de sols par détermination de différentes caractéristiques : couleur, texture, épaisseur, teneur en calcaire, pierrosité, compacité du sol, et de dresser la carte des sols à l'échelle du 1/3 000<sup>ème</sup> pour les deux parcelles (Figure 2-4). Des fonctions de pédotransfert locales ont alors été établies sur les deux parcelles pour déterminer *HCC*, *HMIN* et *DA*, à partir d'une classification des horizons de sol et des mesures disponibles, soit sur la grille, soit sur des profils de sol représentatifs des classes. Ainsi, chaque classe est finalement renseignée par une valeur des paramètres *HCC*, *HMIN* et *DA*, nécessaire à STICS. La profondeur à laquelle existe un obstacle à l'enracinement est notée, ce qui permet ensuite de définir la profondeur de sol utile pour le modèle, via le paramètre *epc(2)*.

### Carte pédologique de la parcelle Chambry 1.



### Carte pédologique de la parcelle Chambry 2.

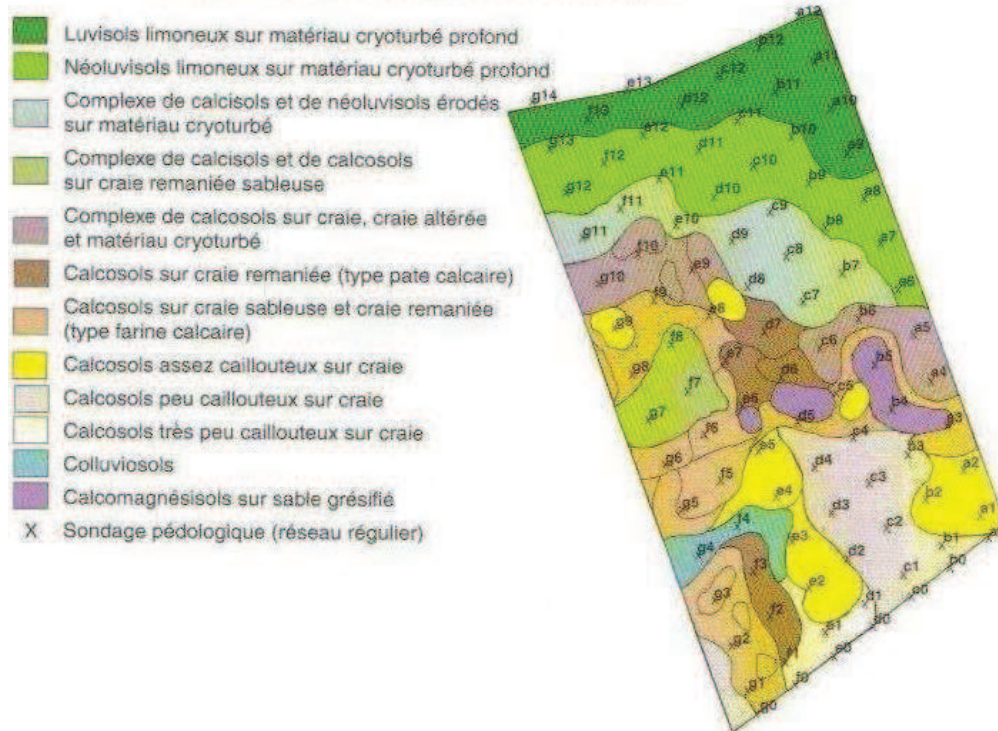


Figure 2-4. Cartes pédologiques à l'échelle du 1/3 000<sup>ème</sup> des deux parcelles de Chambry. Les emplacements des nœuds de la gille (sondages pédologiques) y sont renseignés (Machet et al., 2007).



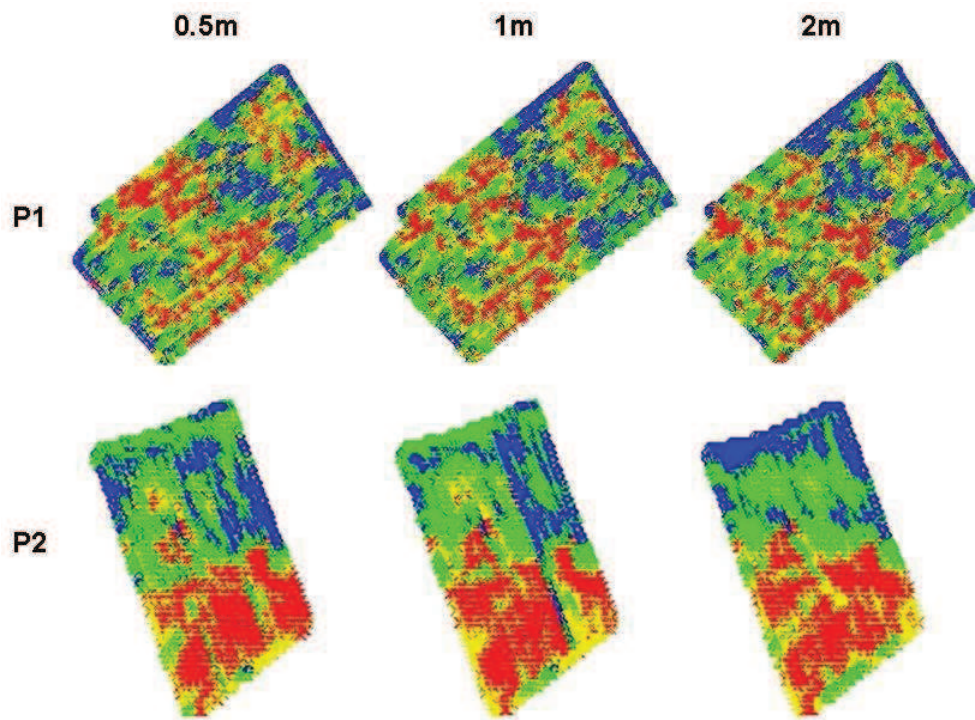
### *Les mesures de résistivité électrique sur les deux parcelles de Chambry*

Dans le cadre du projet d'agriculture de précision, et en vue de caractériser les sols à partir de leurs propriétés électriques, des travaux ont été réalisés par l'Unité INRA d'Orléans afin de tester la résistivité des sols (Bourennane et al., 2007). Cette méthode permet d'obtenir une cartographie de la variabilité des sols à l'échelle de la parcelle. La mesure consiste à injecter un courant électrique dans le sol par des électrodes afin d'estimer la différence de potentiel électrique. Celle-ci est différente selon les caractéristiques du sol (type, texture et structure) et selon l'humidité et la salinité. La méthode utilisée est non destructive, et procure une mesure facile à interpréter, représentable dans l'espace et reproductible dans le temps. Elle permet de visualiser en trois dimensions les différents matériaux présents au sein de la parcelle étudiée et présente ainsi la résistivité électrique comme un outil d'analyse de la structure spatiale des sols. La technique employée sur les deux parcelles de Chambry pour mesurer la résistivité électrique est le MUCEP (MUIri Continuous Electrical Profiling), dont le dispositif est présenté en Figure 2-5 par Samouëlian (2005). Il permet une mesure quasi-continue de la résistivité électrique dans le sol. Sa configuration en "vol de canard" permet d'acquérir les mesures de résistivité à trois profondeurs de sol. Pour ce faire, trois paires d'électrodes d'écartement respectif 50, 100 et 200 cm, mesurent une différence de potentiel à partir de laquelle la valeur de la résistivité électrique est calculée. Ces électrodes sont disposées à la suite de deux autres électrodes qui injectent un courant dans le sol de l'ordre de 2 à 3 ampères. L'ensemble du dispositif est tracté par un véhicule et une mesure de différence de potentiel est déclenchée tous les 10 cm. Comme la profondeur d'investigation est égale à l'écartement inter-électrodes, on réalise ainsi trois cartes de résistivité électrique.



**Figure 2-5.** MUCEP : MUIlti Continuous Electrical Profiling (Samouëlian et al., 2005).

Les cartes fournies par l'application de la méthode MUCEP aux deux parcelles de Chambry, nommées P1 et P2, sont présentées en Figure 2-6. Elles montrent bien la capacité de ces mesures à faire ressortir l'organisation spatiale des sols. Bourennane (2007) a montré que l'utilisation de cette information permettait d'améliorer la précision des prédictions de la teneur en eau du sol par des méthodes d'interpolations de type géostatistiques.

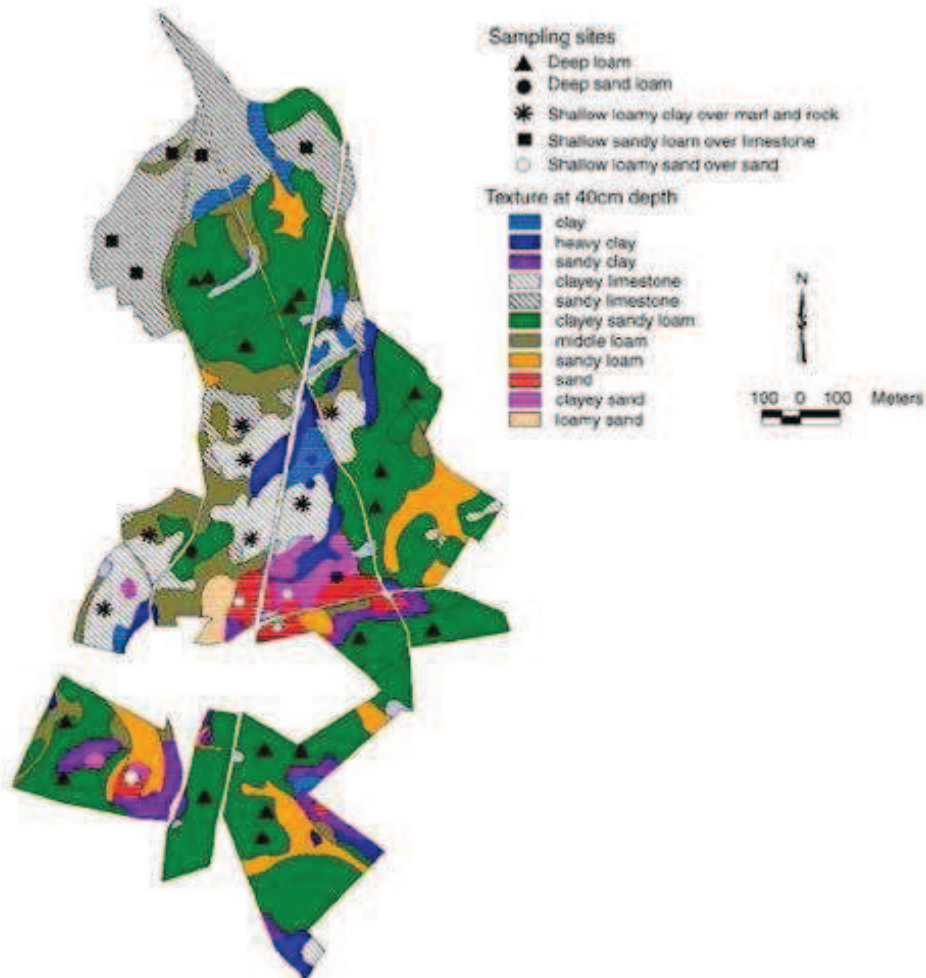


**Figure 2-6.** Cartes de résistivité électrique des deux parcelles P1 et P2 pour les trois profondeurs de sol.

#### *Les parcelles du site de Bruyères*

Le site de Bruyères est un plateau sédimentaire des assises tertiaires du Bassin Parisien. Ses caractéristiques particulières (la nappe, localisée dans le calcaire du Lutétien, est presque isolée hydrologiquement et elle est alimentée presque exclusivement par des zones agricoles de grande culture) en ont fait un lieu privilégié d'étude de l'impact des pratiques agricoles sur la pollution nitrique des aquifères (Beaudoin, 2006; Beaudoin et al., 2008). Une cartographie des sols a été réalisée à l'échelle du 1/5 000<sup>ème</sup> (voir Figure 2-7) et fait apparaître quatre types de sols: limon profond, argile limoneuse peu profonde, limon sableux peu profonds et sable

limoneux peu profond. Un jeu de 31 sites a été défini pour représenter ces types de sol et les différentes parcelles. Les valeurs des paramètres du sol, nécessaires à STICS, ont finalement été déterminées pour chacun des 31 sites du bassin de Bruyères suivant une méthodologie très proche de celle décrite pour Chambry. Pour plus de détail, il est possible de se référer aux publications citées précédemment.



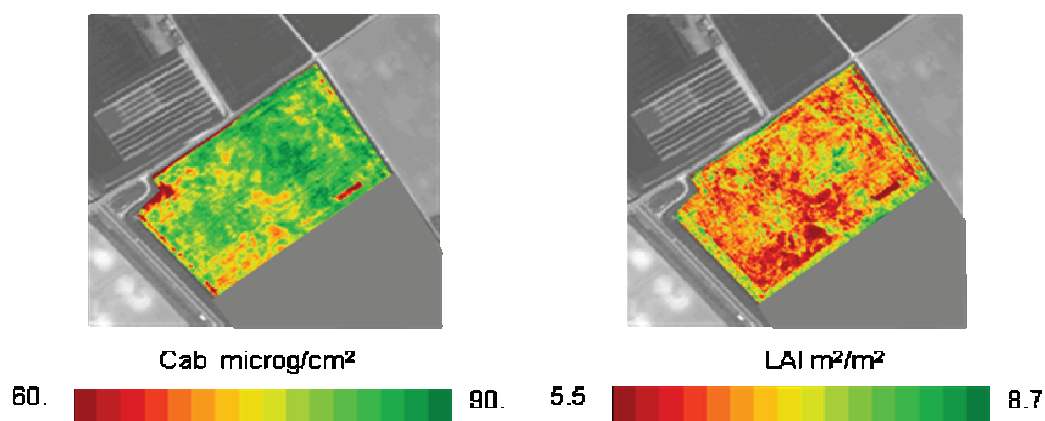
**Figure 2-7.** Carte pédologique au 1/5 000<sup>ème</sup> du bassin de Bruyères et localisation des sites (Beaudoin et al., 2005).

Les mesures de caractérisation de l'horizon de surface ont été faites sur des échantillons de sols prélevés aléatoirement dans un rayon de 10 m à l'intérieur de chaque site. Les mesures des contenus en eau et en azote minéral du sol ont été faites suivant le même protocole de prélèvement sur chacun des 31 sites, trois ou quatre fois par an. En ce qui concerne les propriétés permanentes liées au comportement hydrique, elles ont été mesurées selon la même méthodologie que pour Chambry.

## 2.4.2. Les mesures du couvert végétal

### *Les parcelles de Chambry*

Les mesures sur le couvert végétal en cours de culture sont obtenues à partir d'images de télédétection aéroportées CASI (Compact Airborne Spectrographic Imager) et satellitaires SPOT-HRV ou HRVIR, dans le domaine réflectif solaire. L'inversion des données de réflectance mesurées grâce à des modèles de transfert radiatif permet d'estimer des variables biophysiques du couvert végétal telles que le *LAI* et la teneur en chlorophylle des feuilles *Cab* (Moulin et al., 2007). Les images CASI, qui ont une grande richesse spectrale (32 bandes dans la gamme 350-1050 nm), permettent d'estimer *LAI* et *Cab* alors que les images SPOT-HRV (3 ou 4 bandes dans le vert, rouge, proche infra rouge et moyen infra rouge) ne permettent d'estimer que le *LAI*. La variable biophysique *QN* est déduite de *Cab*, sur la base de relations établies à partir de mesures : *Cab* est convertie en teneur en azote surfacique  $N_{surf}$  et *QN* est ensuite calculée par la relation  $QN = N_{surf} \times LAI$ , illustrant la dépendance qui existe entre les deux variables. Il est possible de se référer à Houlès (2004; 2007) pour plus de détails sur le passage de *Cab* à *QN*. La Figure 2-8 montre un exemple de mesures, sous forme de carte, de *LAI* et de *Cab* sur la première parcelle de Chambry.

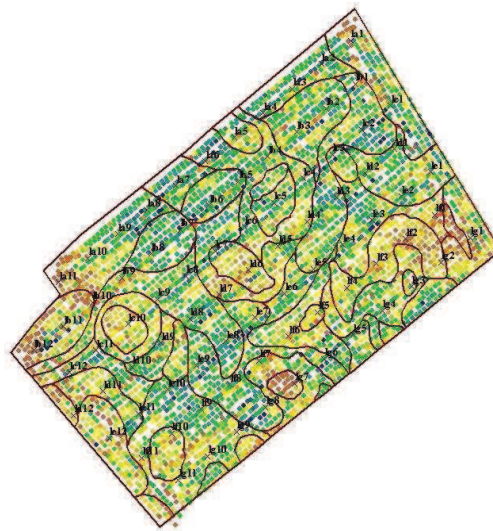


**Figure 2-8.** Exemple de carte de *LAI* et de *Cab* estimées à partir de l'image CASI du 6 Mai 2000 pour la parcelle P1 de Chambry.

Les mesures du rendement ont été faites de deux manières différentes selon le type de culture qui est mesuré. Dans le cas du blé, les mesures ont été faites à



l'aide d'un capteur de rendement embarqué sur le tracteur (Figure 2-9). On se référera à Machet (2007) pour plus de renseignements sur la cartographie du rendement du blé. En ce qui concerne la betterave, il n'existait pas de capteur adapté et les mesures ont été faites directement par prélèvement de betteraves sur les points de grille avec une surface de prélèvement de 13.5 m<sup>2</sup>. Des mesures de teneur en protéine des grains ont été réalisées pour le blé, de manière directe, par prélèvement d'échantillons et analyse au laboratoire.



**Figure 2-9.** Exemple de carte de rendement du blé pour la parcelle P1 de Chambry.

Le tableau suivant recense le nombre d'observations du couvert végétal disponibles sur Chambry, par parcelle et par année :

	Parcelle P1			
Année de récolte	2000	2001	2002	2004
Culture	Blé	Betterave	Blé	Blé
Observations	4 LAI	3 LAI	6 LAI	2 LAI
	3 QN	<i>Yld</i>	<i>Yld</i>	<i>Yld</i>
	<i>Yld</i>		<i>Prot</i>	
	<i>Prot</i>		<i>Nit</i>	

	Parcelle P2		
Année de récolte	2000	2001	2003
Culture	Betterave	Blé	Blé
Observations	1 LAI <i>Yld</i>	7 LAI 4 QN <i>Yld</i> <i>Prot</i>	3 LAI <i>Yld</i> <i>Prot</i> <i>Nit</i>

**Tableau 2-2.** Nombre de cartes d'observations du couvert végétal disponibles sur les deux parcelles de Chambry.

### *Les parcelles du bassin de Bruyères*

L'ensemble des mesures de *LAI*, de teneur en azote des feuilles, de rendement, de teneur en protéine des grains (pour le blé) et de teneur en azote du sol a été réalisé de façon directe, par prélèvement d'échantillons de plante et de sol et analyse au laboratoire. L'échantillonnage consistait en six micro-placettes de 0.3 m<sup>2</sup>, pour chacun des 31 sites. Les mesures ont été réalisées sur un nombre de dates variable selon l'année et le site. Nous avons utilisé, de ce dispositif, les données relatives aux cultures de blé et de betterave qui comportaient un nombre suffisant d'observations du couvert végétal pour mettre en œuvre le protocole d'estimation des paramètres du sol par inversion et de prédiction de variables d'intérêt (voir Chapitre 5), ce qui représente 12 sites. Pour information, le Tableau 2-3 recense le nombre total d'observations du couvert végétal et du sol pour chacun de ces 12 sites, toutes années et cultures confondues.

Site	1	2	3	4	5	6	7	8	9	10	11	12
Nombre d'observations												
<i>LAI</i>	3	6	3	4	6	3	6	3	3	3	0	3
<i>QN</i>	9	10	9	6	7	7	7	7	7	6	7	3
<i>Yld</i>	5	5	6	5	6	5	4	4	6	6	6	3
<i>Prot</i>	1	3	1	3	3	2	1	2	3	2	2	1
<i>Nit</i>	1	13	11	12	7	5	4	3	5	9	16	7

**Tableau 2-3.** Nombre total d'observations des variables d'intérêt agroenvironnemental sur les 12 sites de Bruyères retenus.

### **2.4.3. Les mesures climatiques**

Les caractéristiques climatiques locales, telles que la pluie et la température de l'air, ont été mesurées en continu sur les sites de Chambry et de Bruyères. Quant à l'évapotranspiration potentielle (calculée selon la formule de Penman), elle a été fournie par la station météorologique de Saint Quentin (Roupy). Pour information, la température annuelle moyenne est de 9.7°C, le cumul annuel des pluies atteint en moyenne 701 mm, répartis uniformément dans l'année, et le cumul annuel de l'évapotranspiration potentielle moyenne est de 664 mm.

## Chapitre 3. Sélection des paramètres du sol à estimer

### ***Article 1 : “Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined for simulating agro-environmental variables”***

Nous rappelons ici l’objectif poursuivi et la démarche mise en œuvre dans cet article, ainsi que les principaux résultats et leur intérêt dans notre démarche globale.

#### **3.1. Objectif**

Nous avons vu dans le Chapitre 1 les problèmes posés par l’estimation d’un grand nombre de paramètres du modèle STICS ainsi que les voies possibles pour réduire le nombre de paramètres à estimer. L’enjeu de ce chapitre est de sélectionner un sous-groupe de paramètres du sol, parmi les 75 existants, qui puisse être estimé à partir d’observations du couvert végétal. L’article qui en constitue le corps cherche à répondre à cette question dans un double contexte :

- celui d’un utilisateur qui souhaite savoir quels sont les paramètres du sol qui ont le plus d’influence sur les variables de sortie d’intérêt et sur lesquels il devra investir en priorité pour les mesurer,
- celui d’un utilisateur qui se pose les mêmes questions mais qui dispose d’un ensemble d’observations lui permettant d’estimer, au lieu de mesurer, les paramètres en utilisant le modèle en mode inverse.

#### **3.2. Méthodes**

Nous proposons d’atteindre cet objectif en différentes étapes successives :

1. choix d’un modèle simplifié,
2. réduction des problèmes d’identifiabilité par fixation de valeurs de paramètres,
3. sélection des paramètres importants par analyse de sensibilité.

Les étapes (2) et (3) ont été présentées au Chapitre 1.2. Nous ferons toutefois un rapide rappel ici de ces trois étapes, avant de présenter l'étape (3) qui sera développée dans l'article.

### **3.2.1. Choix d'un modèle simplifié**

La simplification de la représentation du sol et de son fonctionnement que nous avons adopté concerne :

- la prise en compte des processus simulés : nous avons choisi des options de simulation qui permettent de ne pas prendre en compte des processus relativement complexes tels que la circulation de l'eau et des solutés dans la macroporosité du sol, les fissures ou les cailloux ou bien le transfert ascendant de l'eau par capillarité, ou encore les processus de nitrification. Ce choix permet évidemment de ne pas considérer l'ensemble des paramètres sol impliqués dans la description de ces processus,
- une description simplifiée du sol : nous considérons un sol à deux couches, dont la couche de surface est d'une épaisseur fixée à 30 cm, qui est une valeur courante dans les sols agricoles classiquement travaillés.

Ces choix définissent le domaine de validité du modèle et des résultats trouvés en terme d'estimation des paramètres du sol. Toutefois, ils sont cohérents avec les sols des sites de Chambry et de Bruyères, sites sur lesquels est appliquée la méthodologie d'estimation des paramètres du sol. La non prise en compte des cailloux – qui existent dans ces sols – et de leurs propriétés, comme celle des différents horizons pédologiques réellement présents sur les sites, revient à estimer les caractéristiques d'un sol « équivalent » dont on s'est assuré qu'il permettait de décrire correctement les transferts d'eau et la minéralisation de la matière organique sur les sols considérés. Notons que des situations où les processus négligés ici seraient prépondérants obligeraient à reconsidérer ces processus et les paramètres correspondants, ou à accepter l'erreur commise sur l'estimation des paramètres.

Cette première étape permet de réduire le nombre de paramètres du sol de 75 à 18. Ces paramètres concernent ceux déjà présentés dans le Chapitre 2.4.1 : *argi*,

*Norg*, *calc*, *albedo*, *q0*, *epc(2)*, *DA(1)*, *DA(2)*, *HCC(1)*, *HCC(2)*, *HMIN(1)*, *HMIN(2)*, *Hinit(1)*, *Hinit(2)*, *NO3init(1)* et *NO3init(2)*, où le chiffre entre parenthèse représente la couche de sol concernée. Deux autres paramètres sont également considérés à ce stade : *ruisolnu* (fraction d'eau qui ruisselle sur sol nu) et *profhum* (profondeur de sol concernée par les processus d'humification).

Par ailleurs, deux relations entre les conditions initiales (*Hinit* et *NO3init*) des deux couches ont été ajoutées au modèle afin de diminuer le nombre de paramètres. Ces relations ont été déterminées sur la base de données ayant servi à la calibration des paramètres liés aux caractéristiques du blé (voir Annexe) et sont les suivantes :

$$\begin{cases} Hinit(1) = Hinit(2) \\ NO3init(1) = \frac{2}{3} NO3init(2) \end{cases} \quad (3-1)$$

A ce niveau de la sélection, le nombre de paramètres atteint le nombre de 16.

### 3.2.2. Réduction des problèmes d'identifiabilité

La méthode proposée par Makowski (2006a), visant à analyser les équations du modèle STICS afin d'éviter les problèmes d'identifiabilité en fixant certains paramètres (voir Chapitre 1.2.2), a ensuite été appliquée aux 16 paramètres issus de l'étape précédente. L'analyse des équations du modèle faisant intervenir les paramètres sol nous a permis de cibler deux paramètres: *profhum*, la profondeur de sol concernée par les processus d'humification, et *HMIN*, l'humidité du sol au point de flétrissement. En effet, le paramètre *profhum* intervient toujours dans les équations du modèle à travers un produit avec le paramètre *Norg* (voir l'Equation (1-1) du Chapitre 1.1.2). Nous proposons alors de fixer la valeur de *profhum* à une valeur nominale, à savoir 30 cm, qui correspond à une profondeur classique de sol labouré. Seule la variation de *Norg* permettra de rendre compte de la variation de la quantité de matière organique active *NHUM*. Pour le paramètre *HMIN* des deux couches de sol, nous avons vu dans l'Equation (1-2) du Chapitre 1.1.2 qu'il peut agir par différence avec le paramètre *HCC*, pour exprimer la réserve d'eau utile du sol. Plus généralement, il arrive fréquemment que ce paramètre agisse en différence

avec *HCC* dans le modèle STICS. Nous avons donc décidé de ne pas estimer *HMIN* pour les deux couches, en le fixant à une valeur nominale, propre à la parcelle considérée. Cette deuxième étape permet, en fixant *profhum* et *HMIN* des deux couches à des valeurs nominales, de diminuer le nombre de paramètres du sol à estimer de 16 à 13.

### 3.2.3. Sélection des principaux paramètres par analyse de sensibilité

La dernière étape consiste à effectuer une analyse de sensibilité globale des variables observables de STICS à ces 13 paramètres du sol, afin de distinguer les paramètres qui n'ont aucun effet sur les variables observables (à fixer à une valeur nominale) des paramètres qui ont un effet significatif (à estimer par inversion). L'issue de cette étape permet alors de déterminer le sous-groupe de paramètres du sol à estimer. La méthode d'analyse de sensibilité globale Extended FAST a été appliquée aux 13 paramètres du sol sur un ensemble de variables de sortie d'intérêt agroenvironnemental (indice foliaire, azote absorbé par la plante à des stades clés, rendement et qualité de la récolte et azote minéral du sol à la récolte).

Pour ce faire, nous considérons huit différentes configurations agropédoclimatiques sur lesquelles nous analyserons la sensibilité des variables aux paramètres. Ces configurations sont composées de :

- deux cultures annuelles différentes (blé d'hiver et betterave à sucre),
- deux climats contrastés caractérisés comme sec et humide (parmi ceux présentés au Chapitre 2.4.3),
- deux gammes de profondeurs de sol (de 30 à 100 cm pour les sols peu profonds et de 80 à 160 cm pour les sols profonds).

La méthode permet de calculer non seulement les indices de sensibilité globaux qui permettent de mesurer l'importance du paramètre sur la variation de la variable de sortie, mais aussi le coefficient de variation qui permet de déterminer l'importance relative de la variation des paramètres du sol sur celle de la variable de sortie. Ces métriques sont ensuite utilisées pour décider, selon deux optiques différentes, du sous-groupe de paramètres à considérer. La première optique est

celle de la minimisation du coût de mesure des paramètres du sol et la seconde est celle de l'estimation des paramètres par inversion du modèle.

Lorsque l'on se place dans le contexte de minimisation du coût de mesure des paramètres du sol, il est nécessaire de considérer à la fois les indices de sensibilité des paramètres et le coefficient de variation de la variable. Les variables peuvent alors être classées en deux groupes :

1. les variables ayant un petit coefficient de variation : dans ce cas les paramètres ne requièrent pas une mesure précise car ils n'ont pas un grand effet sur la simulation des variables de sortie,
2. les variables ayant un grand coefficient de variation : dans ce cas seuls les paramètres ayant un indice de sensibilité total significatif pour la variable de sortie doivent être mesurés avec précision.

Lorsqu'on se place dans le contexte d'estimation de paramètres par inversion du modèle, il n'est pas nécessaire de considérer le coefficient de variation des variables de sortie. En effet, même si une variable de sortie donnée a un faible coefficient de variation, l'estimation des paramètres du sol peut être possible à condition que l'indice de sensibilité total ne soit pas négligeable. Dans ce cas, le coût d'estimation n'en est pas accru, contrairement à l'optique mesure.

### **3.3. Résultats**

Dans l'optique de la mesure des paramètres du sol, les résultats présentés dans l'article montrent que le nombre de variables nécessitant un effort de mesure des paramètres du sol pour être simulées correctement dépend fortement de la configuration agropédoclimatique dans laquelle elles doivent être simulées. Par exemple, lorsque la betterave est simulée, le nombre de variable nécessitant un effort de mesure est plus important que lorsque le blé est simulé. Il en est de même concernant le climat : lorsque le climat est sec, ce nombre est plus important que lorsque le climat est humide. Pour les variables nécessitant un effort de mesure des paramètres du sol, le nombre de paramètres à mesurer varie entre 1 et 8 selon la



variable et la configuration. Dans le cas contraire où les variables ne nécessitent pas un effort de mesure, le nombre de paramètres à mesurer est nul.

Pour notre objectif d'estimation par inversion de STICS, la sélection du sous-groupe de paramètres à estimer avec des observations du couvert végétal sera donc faite à partir des indices de sensibilité totaux concernant les variables observables *LAI* et *QN* aux stades *AMF*, *LAX* et *FLO* pour le blé, *AMF* et *Summer* pour la betterave, ainsi que *Yld* à la récolte. La règle de décision permettant de retenir ou non un paramètre pour son estimation est la suivante :

$$\begin{cases} \exists(k,c) / ST(y_{obs}^{k,c}, \theta_i) > 0.1 \Rightarrow \theta_i = \hat{\theta}_i \\ ST(y_{obs}^{k,c}, \theta_i) < 0.1, \forall(k,c) \Rightarrow \theta_i = \theta_i^{nom} \end{cases} \quad (3-2)$$

où  $y_{obs}^{k,c}$  est la variable observable  $k$ ,  $k = 1, \dots, 7$  pour le blé et  $k = 1, \dots, 5$  pour la betterave, simulée dans la configuration agropédoclimatique  $c$ ,  $c = 1, \dots, 8$ ,  $\theta_i$ ,  $i = 1, \dots, 13$ , est un paramètre du sol proposé,  $ST(y_{obs}^{k,c}, \theta_i)$  est l'indice de sensibilité total de la variable  $y_{obs}^{k,c}$  au paramètre  $\theta_i$ ,  $\theta_i^{nom}$  est la valeur nominale de  $\theta_i$  et  $\hat{\theta}_i$  est l'estimation de  $\theta_i$ . Ainsi, si l'indice de sensibilité total est supérieur au seuil de 10% pour au moins une variable observable  $k$  simulée dans une configuration  $c$  donnée, alors le paramètre en question sera retenu pour être estimé par inversion ; sinon il sera fixé à une valeur nominale.

Le Tableau 3-1 dresse la liste des 7 paramètres du sol ainsi retenus pour l'estimation : *argi*, *Norg*, *epc(2)*, *HCC(1)*, *HCC(2)*, *Hinit* et *NO3init*. Il existe une forte variabilité entre les valeurs des indices de sensibilité totaux, ce qui signifie que l'observation de chaque variable apporte une quantité d'information très différente selon le paramètre considéré (voir Chapitre 1-3). Dans le chapitre suivant, il sera question de montrer la relation qui existe entre les indices de sensibilité globaux et la quantité d'information fournie par les observations sur les paramètres.

a)	Blé						
	$LAI_{AMF}$	$QN_{AMF}$	$LAI_{LAX}$	$QN_{LAX}$	$LAI_{FLO}$	$QN_{FLO}$	$Yld$
<i>argi</i>	0.09	<b>0.10</b>	0.07	<b>0.12</b>	0.03	<b>0.10</b>	0.02
<i>Norg</i>	0.09	<b>0.20</b>	<b>0.11</b>	<b>0.21</b>	0.04	<b>0.17</b>	0.02
<i>epc(2)</i>	<b>0.11</b>	<b>0.36</b>	<b>0.68</b>	<b>0.57</b>	<b>0.79</b>	<b>0.66</b>	<b>0.55</b>
<i>HCC(1)</i>	<b>0.22</b>	<b>0.20</b>	<b>0.37</b>	<b>0.32</b>	<b>0.45</b>	<b>0.24</b>	<b>0.49</b>
<i>HCC(2)</i>	0.07	0.03	<b>0.13</b>	0.06	<b>0.20</b>	<b>0.10</b>	<b>0.27</b>
<i>Hinit</i>	<b>0.93</b>	<b>0.90</b>	0.04	0.04	0.04	0.04	<b>0.20</b>
<i>NO3init</i>	0.09	<b>0.12</b>	0.06	<b>0.11</b>	0.03	0.08	0.01

b)	Betterave				
	$LAI_{AMF}$	$QN_{AMF}$	$LAI_{Summer}$	$QN_{Summer}$	$Yld$
<i>argi</i>	0.01	0.01	0.04	0.05	0.05
<i>Norg</i>	0.02	0.02	0.04	0.06	0.05
<i>epc(2)</i>	<b>0.18</b>	<b>0.20</b>	<b>0.57</b>	<b>0.42</b>	<b>0.49</b>
<i>HCC(1)</i>	<b>0.96</b>	<b>0.95</b>	<b>0.43</b>	<b>0.65</b>	<b>0.46</b>
<i>HCC(2)</i>	0.02	0.02	<b>0.40</b>	<b>0.15</b>	<b>0.27</b>
<i>Hinit</i>	0.02	0.04	<b>0.54</b>	<b>0.35</b>	<b>0.48</b>
<i>NO3init</i>	0.02	0.03	<b>0.10</b>	0.09	0.06

**Tableau 3-1.** Valeur maximale de l'indice de sensibilité total des paramètres sélectionnés, calculée sur l'ensemble des configurations et pour chaque variable d'intérêt a) du blé et b) de la betterave. En gras les valeurs supérieures à 10%.

### 3.4. Article 1 : “Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined for simulating agro-environmental variables”

Article soumis à Agronomy Journal.

## **Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined for simulating agro-environmental variables.**

Hubert Varella<sup>\*1</sup>, Samuel Buis<sup>1</sup>, Martine Guérif<sup>1</sup> and Marie Launay<sup>2</sup>

<sup>1</sup> INRA, UMR 1114 INRA-UAPV EMMAH, Domaine Saint Paul, Site Agroparc, F84914 Avignon Cedex 9, France

<sup>2</sup> INRA, Unite Agroclim, Domaine Saint Paul, Site Agroparc, F84914 Avignon Cedex 9, France

\* Corresponding author: Hubert Varella

### **Abstract**

The use of a crop model like STICS for appropriate management decision support requires a good knowledge of all the parameters of the model. Among them, the soil parameters are difficult to know at each point of interest and different costly techniques may be used to measure them. It is therefore important to know which soil parameters need to be determined. It can be stated that those which affect significantly the output variable deserve an accurate determination while those which slightly affect the model output variable do not. This paper demonstrates how a global sensitivity analysis method based on variance decomposition can be applied on soil parameters in order to divide them in the two categories. The Extended FAST method applied to the crop model STICS and a set of 13 soil parameters first allows to calculate the part of variance explained by each soil parameter (giving global sensitivity indices of the soil parameters) and the coefficient of variation of the output variables (measuring the effect of the parameter uncertainty on each variable). These metrics are therefore used for deciding on the importance of the parameter value measurement. Different output variables (Leaf Area Index and chlorophyll content) are evaluated at different stages of interest while others (crop yield, grain protein content, soil mineral nitrogen) are evaluated at harvest. The analysis is applied on two different annual crops (wheat and sugar beet), two contrasted weather and two types of soil depth. When the uncertainty of the output generated by the soil parameters is large (coefficient of variation  $> 1/3$ ), only the parameters having a

significant global sensitivity indices (higher than 10%) are retained. The results show that the number of soil parameters which deserve an accurate determination for simulating correctly the outputs can be significantly reduced (can reach 4) by the use of this relevant method for appropriate management decision support.

Keywords: Global sensitivity analysis, uncertainty, soil parameters, crop model STICS, management decision support, agro-environmental variables.

## **1. Introduction**

Dynamic crop models are very useful to predict the behavior of crops in their environment and are widely used in a lot of agro-environmental work such as crop monitoring, yield prediction or decision making for cultural practices (Batchelor et al., 2002; Gabrielle et al., 2002; Houlès et al., 2004). These models usually have many parameters and their spatial application for agro-environmental predictions is difficult without a good knowledge of these parameters (Launay and Guérif, 2003; Tremblay and Wallach, 2004).

The crop models parameters can be divided in three groups: those related to the agricultural techniques, those related to the genotype of the crop and those related to the soil properties. Generally, agricultural techniques are quite easy to know as they are those used by the farmer. Crop parameters can be determined from literature, or estimated from experimental work or calibrated on a large database (Flenet et al., 2003; Hadria et al., 2007; Singh et al., 2008). The knowledge of the soil parameters is an important issue because the spatial variability of the crop model simulations depends for a large part on the soil parameter values (Guérif and Duke, 1998) and predictions obtained with the model are not reliable when inaccurate parameter values are used. This knowledge may be especially difficult to acquire because parameter values can greatly vary in space (Irmak et al., 2001; Ferreyra et al., 2006). The use of existing soil maps and associated pedotransfer functions can be considered where accurate soil map are available (Nemes et al., 2009); but in many cases, the spatial accuracy of the map is too limited for accurate applications such as for example precision agriculture (King et al., 1994). In those cases, these parameters should be determined in another way. Measurements can be made

directly with soil sampling analysis at different locations of the study area or indirectly by using electrical geophysical measurements (Samouelian et al., 2005; Bourennane et al., 2007). Whatever the technique of measurement used, it is submitted to practical limitations and to time and financial constraints. Another way of gathering quite accurate values on soil parameters consists in estimating them through an inverse modeling approach using a crop model and observations on the crop state variables (Braga and Jones, 2004; Ferreyra et al., 2006; Varella et al., 2009). However, the soil parameters may not have the same contribution to the performance of the crop model and do not require the same precision of determination for a given objective: some of them deserve an accurate determination while the others can be fixed at nominal values (Bouman, 1994; St'astna and Zalud, 1999). Considering this aspect, the practical limitations of soil parameter measurements, as well as time and financial constraints should be reduced by considering only a subset of the crop model soil parameters depending on the objective and configuration of the study.

The combination of uncertainty analysis and sensitivity analysis techniques should help in identifying these key parameters. The objective of sensitivity analysis is to study how the variation of selected outputs of a model can be apportioned to different sources of variation (Saltelli et al., 2000a). In particular, sensitivity analysis methods can be used to rank uncertain input factors with respect to their effects on the model output variables by calculating quantitative or qualitative indices. Nevertheless, the fact that some factors are detected as important for a given output variable on the basis of sensitivity analysis results is not sufficient to decide that the uncertainties on these factors should be reduced. Indeed, if the variation of the considered output variable induced by the uncertainties on the factors is low then the results of sensitivity analysis on this output variable should not be taken into account. The description and quantification of these variations is the objective of uncertainty analysis.

Some authors (Bouman, 1994; Aggarwal, 1995; Blasone et al., 2008; Lawless et al., 2008; Tolson and Shoemaker, 2008; van der Keur et al., 2008) used uncertainty analysis techniques to quantify the uncertainties of a selection of crop models output variables generated by uncertainties on some selections of input parameters. Others authors (Campolongo and Saltelli, 1997; Saltelli et al., 2000b; Gomez-Delgado and Tarantola, 2006; Makowski et al., 2006b; Pathak et al., 2007)

used global sensitivity analysis to evaluate the contribution of the parameters to the variance of the model output variables. Unfortunately, few studies concerning crop models consider both uncertainty analysis and sensitivity analysis, and, to our knowledge, not at all used the combination of these techniques to identify soil parameters that need particular accuracy for simulating a set of given output variables of interest in spite of the financial and practical interests of such a study.

In this study, we propose to use a variance-based sensitivity analysis method in order to rank the soil parameters relatively to their importance on some selected outputs of the crop model STICS (Brisson et al., 2002) and to select those which deserve an accurate determining by considering also the coefficient of variation of the outputs, that is the variation of the outputs compared to their magnitude. We considered 13 soil parameters and their effects on 5 dynamic output variables of the STICS crop model, at different phenological stages, which are involved in decision making for crop management. Two different crops (winter wheat and sugar beet) growing on different seasons are considered in order to illustrate the impact of soil properties on crop growth. Each crop considered is simulated under different pedological conditions and weather.

## **2. Methods**

### **2.1 The STICS model**

The STICS model (Brisson et al., 2008) is a nonlinear dynamic crop model simulating various crops. For a given crop, STICS takes into account the weather, the type of soil and the cropping techniques used, and simulates the carbon, water and nitrogen balances of the crop-soil system on a daily time-scale. In this study, winter wheat and sugar beet crops are simulated. The crop is essentially characterized by its above-ground biomass carbon and nitrogen, and leaf area index. The soil is considered as a series of layers where the transfer of water and nitrate is described by a reservoir-type analogy. The main outputs are agronomic variables (yield, grain protein content for wheat) as well as environmental variables (water and nitrate leaching).

The STICS model includes more than 200 parameters. The global sensitivity analysis described in this study only concerns the soil parameters. The values of the parameters related to the crop have been determined either from literature, from experimental works conducted on specific processes included in the model (e.g. mineralization rate, critical nitrogen dilution curve etc.) or from a calibration based on a large experimental database (Houlès, 2004 for the wheat crop; Launay et al., 2009 for the sugar beet crop). Cropping techniques and soil parameters ranges are described in Section 2.5.

## 2.2 The soil parameters

Among the available options for simulating the soil system, the simplest was chosen in this study, by considering only the transfers in the microporosity and ignoring those in the macroporosity, the cracks, pebbles, and processes like capillary rise and nitrification. We then considered the soil as a succession of two horizontal layers, the top layer having a thickness fixed at 30 cm. Based on the measurements made on the site considered for this study (see Section 2.5) for several years, we added relations between the initial contents of water  $H_{init}$  and mineral nitrogen  $NO3_{init}$  of the two soil layers in order to limit the number of parameters that should be varied:

$$\begin{aligned}
 H_{init} &= H_{init}_{first\_layer} = H_{init}_{second\_layer} \\
 NO3_{init} &= NO3_{init}_{first\_layer} = \frac{2}{3} NO3_{init}_{second\_layer}
 \end{aligned}
 \tag{1}$$

Finally, the different hypotheses made on the soil description lead to consider a set of 13 soil parameters, defined in Table 1. They refer to permanent characteristics and initial conditions. Among the permanent characteristics, clay and organic nitrogen content of the top layer are involved mainly in organic matter decomposition processes. Water content at field capacity of both layers affects the water (and nitrogen) movements and storage in the soil reservoir and the thickness of the second layer defines the volume of the reservoir. The initial conditions correspond to the water and nitrogen content,  $H_{init}$  and  $NO3_{init}$ , at the beginning of the simulation, in this case the sowing date.

Parameter	Definition	Range	Unit	Label
<i>argi</i>	Clay content of the 1 <sup>st</sup> layer	14-37	%	<i>ar</i>
<i>Norg</i>	Organic nitrogen content of the 1 <sup>st</sup> layer	0.049-0.131	%	<i>N</i>
<i>calc</i>	Limestone content of the 1 <sup>st</sup> layer	0-28	%	<i>c</i>
<i>albedo</i>	Albedo of the bare dry soil	0.13-0.31	–	<i>al</i>
<i>q<sub>0</sub></i>	Threshold of daily evapotranspiration	7.5-14.5	mm	<i>q</i>
<i>ruisolnu</i>	Fraction of drip rainfall on a bare soil	0-0.065	–	<i>r</i>
<i>epc(2)</i>	Thickness of the 2 <sup>nd</sup> layer	0-70 or 50-130*	cm	<i>e</i>
<i>DA(1)</i>	Bulk density (1 <sup>st</sup> layer)	1.22-1.42	–	<i>D1</i>
<i>DA(2)</i>	Bulk density (2 <sup>nd</sup> layer)	1.39-1.59	–	<i>D2</i>
<i>HCC(1)</i>	Water content at field capacity (1 <sup>st</sup> layer)	14-30	g g <sup>-1</sup>	<i>H1</i>
<i>HCC(2)</i>	Water content at field capacity (2 <sup>nd</sup> layer)	14-30	g g <sup>-1</sup>	<i>H2</i>
<i>Hinit</i>	Initial water content (both layers)	4-29	% of weight	<i>h</i>
<i>NO3init</i>	Initial mineral nitrogen content (1 <sup>st</sup> layer)	4-21.5 or 12-55**	kg N ha <sup>-1</sup>	<i>n</i>

\* the first range determine a shallow soil and the second determine a deep soil.

\*\* the first range concern the wheat (cultivated after beet) and the second concern the beet (cultivated after a bare soil).

**Table 1.** Definition of the 13 soil parameters and their ranges of variation.

## 2.3 Model outputs

In this study, the STICS output variables we are mostly interested in are:

- (i) the amount of nitrogen absorbed by the plant (*QN*) and the leaf area index (*LAI*) at two (for sugar beet) or three (for wheat) different key stages during the season which are important variables for making a diagnosis on crop growth,
- (ii) the yield, and the mineral nitrogen content in the soil at harvest (for both crops) plus the grain protein content (for wheat), which are of particular interest for decision making, especially for monitoring nitrogen fertilization (Houlès et al., 2004).

The different stages of interest and the corresponding variables of interest are displayed for each crop on Table 2. For the wheat, the three key stages concern the maximum leaf growth rate – beginning of stream elongation – (*AMF*), the maximum leaf area – or booting – (*LAX*) and the flowering (*FLO*). For the sugar beet, the two key stages concern the maximum leaf growth rate (*AMF*) and the maximum leaf area (*Summer*).



Crop simulated	Variable of interest	Stage of interest	Signification of the stage
Wheat	LAI and QN	AMF	Stage of maximum leaf growth rate (beginning of steam elongation)
		LAX	Stage of maximum leaf area (booting)
		FLO	Flowering
	Yld, Prot and Nit	Harvest	Harvest
Sugar beet	LAI and QN	AMF	Stage of maximum leaf growth rate
		Summer	Day where maximum leaf area is achieved in most cases
	Yld and Nit	Harvest	Harvest

**Table 2.** Definition of the variables and the stages of interest.

## 2.4 Sensitivity and uncertainty analysis

Among the available methods of sensitivity analysis, variance-based methods are well adapted for non-linear models that need less than 1 minute for a simulation (Cariboni et al., 2004). These methods are widely used in different contexts (Campolongo and Saltelli, 1997; Saltelli et al., 2000b; Gomez-Delgado and Tarantola, 2006; Makowski et al., 2006b; Pathak et al., 2007). Their principle is to evaluate the contribution of the given uncertain factors to the variance of the model output variables selected. We will describe in this section the sensitivity indices that can be computed with these methods and the EFAST variance-based method we have used in this study to compute these indices. Uncertainty analysis is performed here by computing the coefficient of variation of the output variables considered from the simulations realized for the sensitivity analysis.

### *Sensitivity indices and coefficient of variation*

We note further  $Y$  an output variable of STICS.  $Y$  will represent in turn  $LAI$  and  $QN$  computed at the different phenological stages and the variables computed at harvest. The total variance of  $Y$ ,  $V(Y)$ , is partitioned as follows (Chan et al., 2001):

$$V(Y) = \sum_{i=1}^{13} V_i + \sum_{1 \leq i < j \leq 13} V_{ij} + \dots + V_{1,2,\dots,13}, \quad (2)$$

where  $V(Y)$  is the total variance of the output variable  $Y$  induced by the 13 soil parameters  $\theta$ ,  $V_i = V[E(Y|\theta_i)]$  measures the main effect of the parameter  $\theta_i$ ,

$i = 1, \dots, 13$ , and the other terms measure the interaction effects. Decomposition (2) is used to derive two types of sensitivity indices defined by:

$$S_i = \frac{V_i}{V(Y)}, \quad (3)$$

$$ST_i = \frac{V(Y) - V_{-i}}{V(Y)}, \quad (4)$$

where  $V_{-i}$  is the sum of all the variance terms that do not include the index  $i$ .

$S_i$  is the first-order sensitivity index of the  $i$ th parameter. It computes the fraction of  $Y$  variance explained by the uncertainty of parameter  $\theta_i$  and represents the main effect of this parameter on the output variable  $Y$ .  $ST_i$  is the total sensitivity index of the  $i$ th parameter and is the sum of all effects (first and higher order) involving the parameter  $\theta_i$ .  $S_i$  and  $ST_i$  are both in the range (0, 1), low values indicating negligible effects, and values close to 1 huge effects.  $ST_i$  takes into account both  $S_i$  and the interactions between the  $i$ th parameter and the 12 other parameters, interactions which can therefore be assessed by the difference between  $ST_i$  and  $S_i$ . The interaction terms of a set of parameters represent the fraction of  $V(Y)$  induced by these parameters but that is not explained by the sum of their main effects. The two sensitivity indices  $S_i$  and  $ST_i$  are equal if the effect of the  $i$ th parameter on the model output is independent of the values of the other parameters: in this case, there is no interaction between this parameter and the others and the model is said to be additive with respect to  $\theta_i$ . Selecting the parameters that have a negligible effect and that can thus be fixed to nominal values is called factor fixing in the literature (Ratto et al., 2007). Total effects must be considered in this case. Indeed, a factor that has a small main effect but a medium to high total effect cannot be considered as negligible: its effect depends on the value of other uncertain factors and can be important in some cases.

The coefficient of variation of the output variable  $Y$  can be calculated by:

$$CV(Y) = \frac{\sqrt{V(Y)}}{\bar{Y}} = \frac{\sigma(Y)}{\bar{Y}} \quad (5)$$

where,  $\sigma(Y)$  is the standard deviation of the output variable  $Y$  and  $\bar{Y}$  is the mean of  $Y$  induced by the 13 soil parameters  $\theta$ .

### *Extended FAST*

Sobol's method and Fourier Amplitude Sensitivity Test (FAST) are two of the most widely used methods to compute  $S_i$  and  $ST_i$  (Chan et al., 2000). We have chosen here to use the extended FAST (EFAST) method, which has been proved, in several studies (Saltelli and Bolado, 1998; Saltelli et al., 1999; Makowski et al., 2006b), to be more efficient in terms of number of model evaluations than Sobol's method. The main difficulty in evaluating the first-order and total sensitivity indices is that they require the computation of high dimensional integrals. The EFAST algorithm performs a judicious deterministic sampling to explore the parameter space which makes it possible to reduce these integrals to one-dimensional ones using Fourier decompositions. The reader interested in a detailed description of EFAST can refer to (Saltelli et al., 1999).

We have implemented the EFAST method in the Matlab® software, as well as a specific tool for computing and easily handling numerous STICS simulations. The uncertainties considered for the soil parameters are described in the next section. A preliminary study of the convergence of the sensitivity indices allowed us to set the number of simulations per parameter to 5000, leading to a total number of model runs of  $13 \times 5000 = 65000$  to compute the main and total effects for all output variables and parameters considered here. One run of the STICS model takes about 1s with a Pentium 4, 2.9 GHz processor.

## **2.5 Data**

In this study, we have considered two crops: winter wheat and sugar beet. This allows us illustrating the difference of sensitivities of different crops to the soil properties. For the same reason, each crop is simulated for two different weathers and two different types of soil depth. The weather data were obtained from the meteorological station of Roupy (49.48°N, 3.11°E). The first set of data is chosen to characterize a dry weather (1975-1976) and the second set is chosen to characterize a wet weather (1990-1991). Table 3 shows the amount of rainfall calculated for each season and weather data set. The wheat crop simulated in this study is sown on October 30<sup>th</sup> while the sugar beet crop is sown on March 30<sup>th</sup>. The amount of fertilizer

provided on wheat varies between 200 kg (shallow soil) and 240 kg (deep soil), while the amount provided on sugar beet varies between 150 kg (shallow soil) and 200 kg (deep soil).

	Spring	Summer	Autumn	Winter
Dry weather	343.4	167.8	222.4	218.8
Wet weather	361.4	247.9	239.4	316.4

**Table 3.** Amount of rainfall (in mm) calculated for each season and weather.

The range of parameter values considered in this study correspond to the soil description of the precision agriculture experimental site in northern France near Laon, Picardie (Chambry 49.35°N, 3.37°E) (Guérif *et al.*, 2001). In this study, the uncertainties of these 13 soil parameters are observed in the literature (for parameters related to albedo, evapotranspiration or drip rainfall) or measured in the experimental site (for the other parameters), and their ranges of variation are displayed on Tab. 1. Concerning the parameter *NO3init* two ranges of variation are considered, depending on the crop cultivated just before the one considered: in this study, the wheat is cultivated after sugar beet and the sugar beet is cultivated after a bare soil. The different previous crops used determine the quantity of nitrogen *NO3init* at the beginning of the corresponding crop season. The two different types of soil depth are defined by their ranges of variation (Tab. 1) and correspond to a shallow soil and a deep soil. The uncertainties considered in the global sensitivity analysis for the soil parameters are assumed independent and follow uniform distributions. The ranges of variation of the distributions are given in Tab. 1.

### 3. Results and discussion

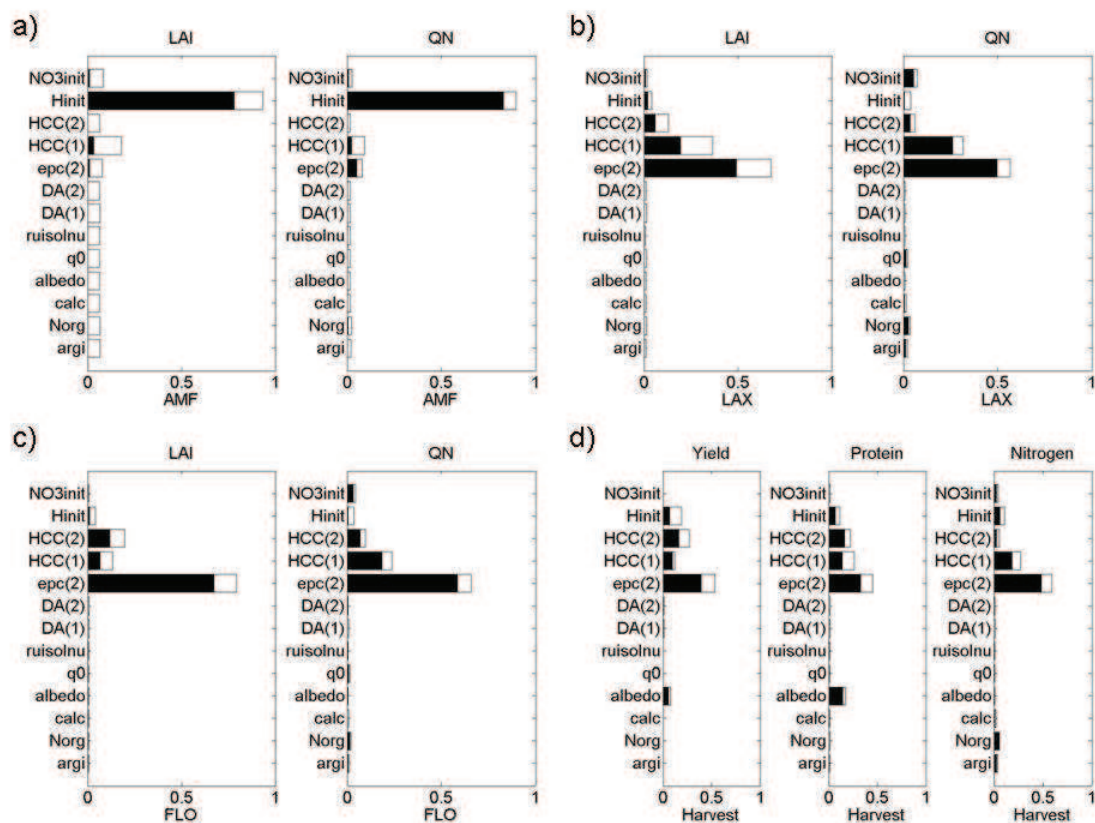
Only the main results of the study are presented here for the sake of clarity. These results concern: (i) wheat crop simulated with dry, then wet weather and a shallow soil and (ii) sugar beet crop simulated with dry, then wet weather and deep soil.

### 3.1 Global sensitivity analysis

Figure 1 shows the sensitivity indices calculated for the 13 soil parameters and for each output variable of the wheat crop simulated with a dry weather and a shallow soil. For the early stage the initial water content is dominant because in the considered weather, the rainfall is light in autumn when the wheat is sown (see Tab. 3): at the stage *AMF* (Fig. 1a), *Hinit* is the only parameter contributing (for more than 90%) to the output variance for both variables *LAI* and *QN*. In the later stages, the effects of parameters evolve with the soil volume explored by the roots (first layer, then second one) up to the flowering stage where the root development is maximum: at the stage *LAX* (Fig. 1b) and *FLO* (Fig. 1c), the effect of *Hinit* disappears and that of *HCC(1)* and *epc(2)* increase, with a dominant position of *epc(2)*. At harvest (Fig. 1d), the variables are much sensitive to *epc(2)* followed by *HCC(1)*, *HCC(2)* and *Hinit* for the three variables and *albedo* for the variable protein of the grain. In those conditions of dry weather and shallow soil, the parameters related to water availability (*epc(2)*, *HCC(1)*, *HCC(2)* and *Hinit*) are the main parameters contributing to the variance of the outputs. Those concerned by the turnover of organic nitrogen in the soil are not concerned, because the water stress is the dominant limiting factor and also because the mineralization processes are reduced by dry conditions.

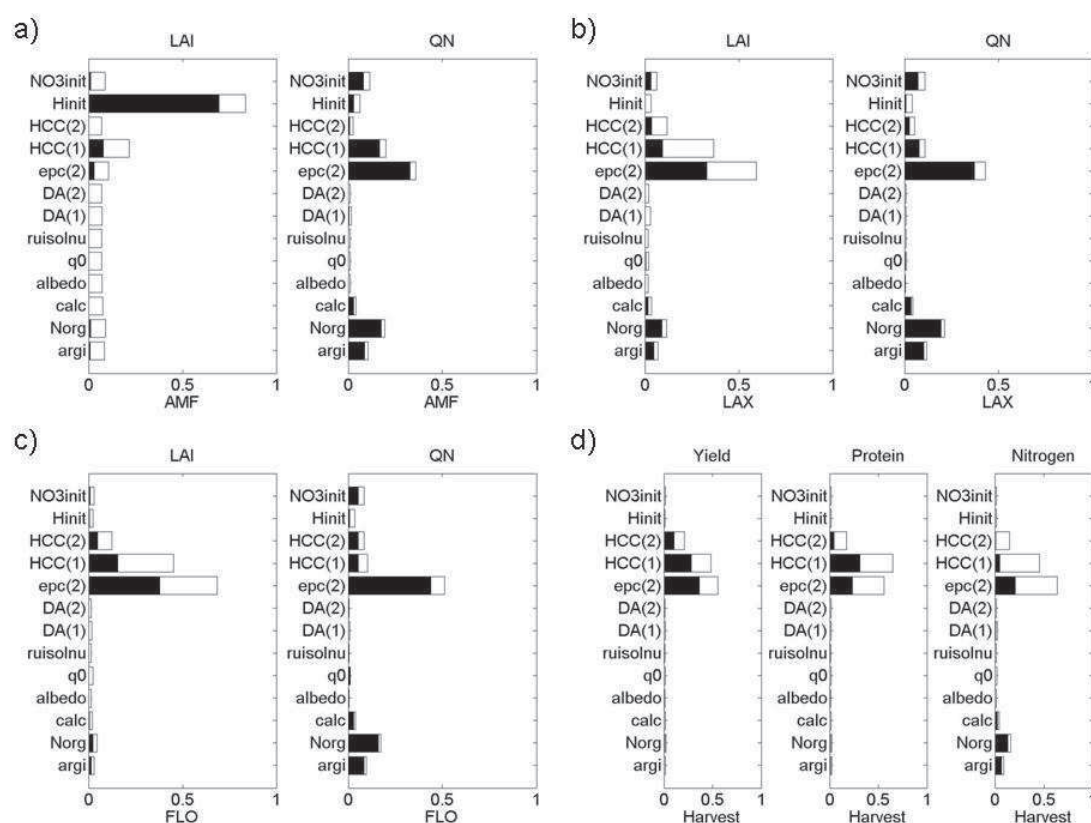
When considering wet conditions (Figure 2), the water stress is not so much a limiting factor: maximum *LAI* is equal to 3.61 in average, whereas it is equal to 2.57 in dry conditions (see Table 4). The roots grow more rapidly at the beginning of the season and the size of the soil reservoir (via the parameter *epc(2)*) is important since the *AMF* stage: the depth of roots is equal to 55.84 in average (3 months after the sowing), whereas it is equal to 45.62 in dry conditions (see Tab. 4). Moreover, in these conditions, the mineralization of the soil organic matter is increased and the effects of the concerned parameters *argi* and *Norg* do so: the cumulative mineral nitrogen arising from humus is equal to 23.95 in average (at stage *LAX*), whereas it is equal to 18.09 in dry conditions (see Tab. 4). This does not seem to influence the effects of the different parameters on *LAI* at stage *AMF* since they are very similar to these obtained with the dry weather. On the contrary, the sensitivities of the variable *QN* to the different parameters are very different from the ones obtained with a dry weather: there is no contribution of *Hinit* but *epc(2)*, *HCC(1)* and parameters involved

in the mineralization process (*argi*, *Norg* and *NO3init*) significantly contributes to the variance of this variable. This is also the case for both *LAI* and *QN* on later stages, with an increasing dominance of *epc(2)*. At harvest (Fig 2d), the variables are sensitive to the parameters *epc(2)*, *HCC(1)* and *HCC(2)* with still a slight sensitivity to *argi* and *Norg* for the soil mineral nitrogen content. The main difference between these results and those presented in Fig.1 lies in the sensitivity to parameters involved in the mineralization process (especially *argi* and *Norg*).



**Figure 1.** Sensitivity indices of the 13 soil parameters for each model output of the wheat crop simulated with a dry weather and a shallow soil. The outputs a) correspond to *LAI* and *QN* at stage *AMF*, b) correspond to *LAI* and *QN* at stage *LAX*, c) correspond to *LAI* and *QN* at stage *FLO* and d) correspond to *Yld*, *Prot* and *Nit* at *Harvest*. First-order indices are in black and interactions in white.





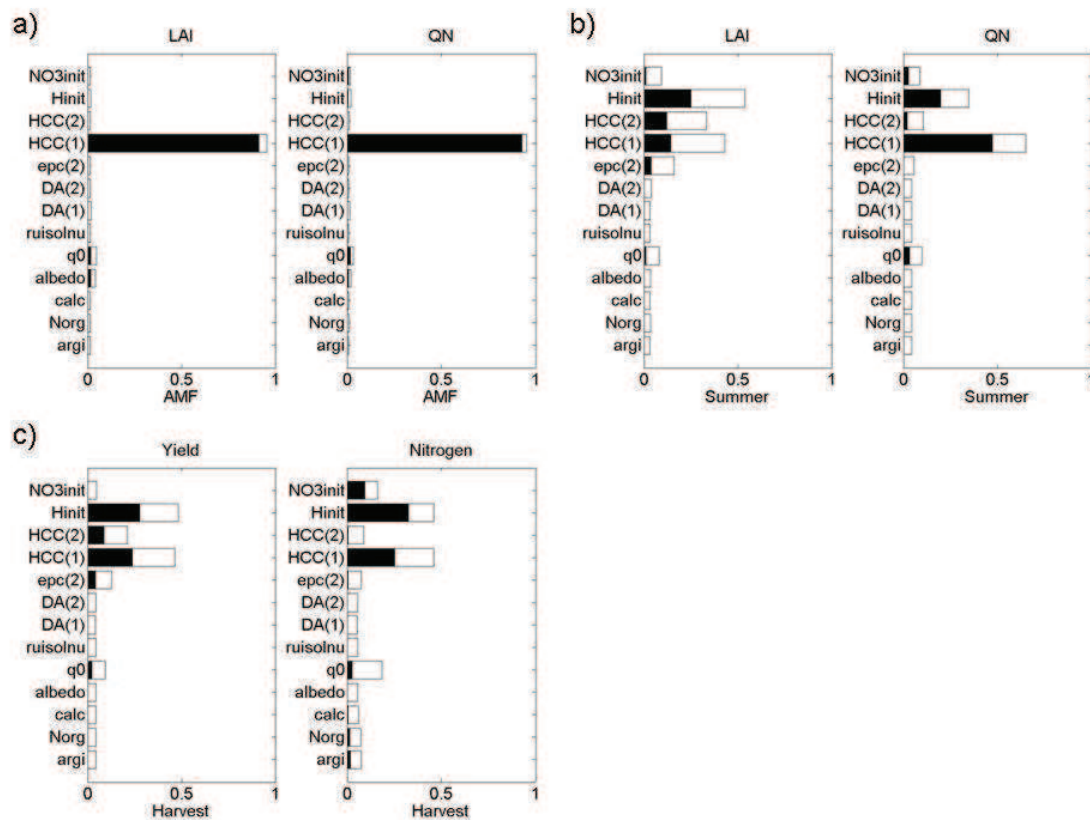
**Figure 2.** Sensitivity indices of the 13 soil parameters for each model output of the wheat crop simulated with a wet weather and a shallow soil. The outputs a) correspond to *LAI* and *QN* at stage *AMF*, b) correspond to *LAI* and *QN* at stage *LAX*, c) correspond to *LAI* and *QN* at stage *FLO* and d) correspond to *Yld*, *Prot* and *Nit* at *Harvest*. First-order indices are in black and interactions in white.

Configuration of simulation	Maximum <i>LAI</i>			<i>Qminh</i>			<i>Zrac</i>		
	min	mean	max	min	mean	max	min	mean	max
<i>W C1 -</i>	0.78	2.57	3.73	7.19	18.09	40.51	30.1	45.62	56.52
<i>W C2 -</i>	2.51	3.61	5.08	9.8	23.95	48.85	30.1	55.84	69.91
<i>SB C1 +</i>	0	1.42	4.61	0	24.91	83.38	12.06	77.58	129.61
<i>SB C2 +</i>	0.19	4	6.06	19.19	50.19	121.45	71.08	85.55	102.58

**Table 4.** Ranges of some output variables uncertainties generated by the uncertainties on the soil parameters. The output concerns the value of maximum *LAI*, the cumulative mineral nitrogen arising from humus *Qminh* (calculated at the stage *LAX* or *Summer*) and the depth of roots *Zrac* (calculated 3 months after the sowing date). The results are calculated for the four configurations of simulation: wheat crop, dry weather and shallow soil (*W C1 -*), wheat crop, wet weather and shallow soil (*W C2 -*), sugar beet crop, dry weather and deep soil (*SB C1 +*) and sugar beet crop, wet weather and deep soil (*SB C2 +*).

Figure 3 shows the sensitivity indices calculated for the 13 soil parameters and for each output variable of the sugar beet crop simulated with a deep soil and a dry weather. In this case, the crop grows mainly in summer where it experiences a severe water stress, leading to a value of maximum *LAI* equal to 1.42 in average (see Tab. 4). The depth of the second layer (parameter *epc(2)*) does not have any importance here. This is also the case for wheat crop with a deep soil (results not shown here). Indeed, as the root growth is no longer limited by the thickness of soil (the depth of roots is equal to 77.58 in average), the output variables are no longer sensitive to the parameter *epc(2)* when the soil is deep. Moreover, the outputs are not at all sensitive to the initial water content *Hinit* because the amount of rainfall is quite important in spring, when the sugar beet is sown (see Tab. 3). The soil water reserve is therefore the main limiting factor and it depends only on *HCC(1)* for the early stage *AMF*: it contributes for 95 % of the total output variance of *LAI* and *QN*. For the *Summer* stage (Fig. 3b), which correspond to the maximum of water stress index, *LAI* is mainly sensitive to parameters linked to water availability of both soil layers (*HCC(1)* and *HCC(2)*) with an increase of the sensitivity to *Hinit*. *QN* is more sensitive to characteristics of the top layer (*HCC(1)* and *Hinit*) where is concentrated the organic nitrogen, as it influences the fate of available nitrogen coming from mineralization. The same tendencies are noticed for the output variables at harvest, *Yield* being more linked to *LAI* and soil mineral nitrogen to *QN*. Many interactions are visible between all these parameters. It is also noticeable that, as in the case of wheat, the output variables have very low sensitivity to the parameters concerned with nitrogen turnover in the soil, due to the dry weather and limited mineralization. The main differences of these results with respect to those presented for the wheat crop (Fig. 1 and 2) is: (i) that *HCC(1)* contributes a lot to the variance of the output variables during all the crop season, (ii) that *Hinit* as no contribution to the variance of the output variables at the beginning of the sugar beet season and (iii) that *epc(2)* does not affect the output variables when the soil is deep.

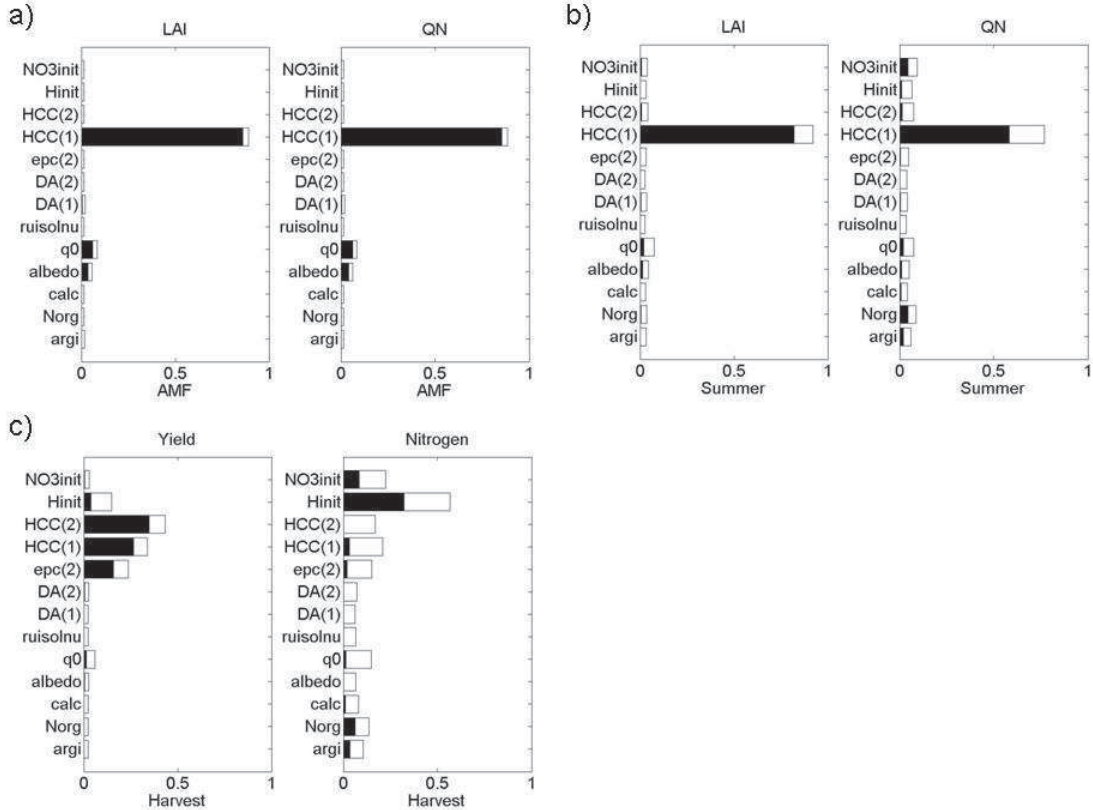




**Figure 3.** Sensitivity indices of the 13 soil parameters for each model output of the sugar beet crop simulated with a dry weather and a deep soil. The outputs a) correspond to *LAI* and *QN* at stage *AMF*, b) correspond to *LAI* and *QN* at *Summer* and c) correspond to *Yld* and *Nit* at *Harvest*. First-order indices are in black and interactions in white.

When considering wet conditions (Figure 4), the sugar beet crop growth is less affected by the water stress: maximum *LAI* is equal to 4 in average, whereas is equal to 1.42 in dry conditions (see Tab. 4). The soil water reserve of the second layer is not a limiting factor in deep soil and wet conditions for both stages *AMF* and *Summer* because the soil reservoir has a large size and the water stress is low. Thus, *LAI* and *QN* are only sensitive to the soil water reserve of the first layer which only depends on *HCC(1)* (it does not depend on *Hinit* because of the high amount of rainfall in spring). Nevertheless, the soil water reserve of the second layer becomes a limiting factor at the end of the sugar beet crop season, when the roots are deep, involving a significant sensitivity of the output *Yld* to the parameters *HCC(1)*, *HCC(2)* and *epc(2)*. Moreover, the mineralization of the soil organic matter slightly increases in wet conditions and so do the effects of the concerned parameters on *QN* at *Summer* and *Nit* at *yield*: the cumulative mineral nitrogen arising from humus is equal to 50.19 in

average, whereas is equal to 24.91 in dry conditions (see Tab. 4). The main difference between these results and those presented in Fig. 3, lies in the lower sensitivity of the soil water reserve parameters of the second layer at the two first stages of interest.



**Figure 4.** Sensitivity indices of the 13 soil parameters for each model output of the sugar beet crop simulated with a wet weather and a deep soil. The outputs a) correspond to *LAI* and *QN* at stage *AMF*, b) correspond to *LAI* and *QN* at *Summer* and c) correspond to *Yld* and *Nit* at *Harvest*. First-order indices are in black and interactions in white.

### 3.2 Total effect and coefficient of variation

For each configuration of simulation presented above, Figure 5 shows the coefficient of variation *CV* of each output variable and the corresponding total effect *ST* of each parameter. The horizontal dashed line is situated at an arbitrary minimum value *ST*=10% and the vertical dashed line is situated at another arbitrary minimum value *CV*=1/3. The threshold of 10% for *ST* has been proposed by Makowski (2006b) for screening the significant sensitivity values. When wheat crop is simulated with a

dry weather and a shallow soil (see Fig. 5a), three output variables have a coefficient of variation higher than 1/3: *Prot* ( $CV=0.37$ ), *Yld* ( $CV=0.54$ ) and *LAI* at the stage *FLO* ( $CV=0.62$ ). For these outputs, only 5 soil parameters have a *ST* higher than 10%: *epc(2)*, *HCC(1)*, *HCC(2)*, *Hinit* and *albedo*. This means that for simulating correctly these output variables of the wheat crop when the weather is dry and the soil depth is shallow, only *epc(2)*, *HCC(1)*, *HCC(2)*, *Hinit* and *albedo* have to be determined accurately (assuming the arbitrary threshold  $ST=10\%$  and  $CV=1/3$ ) and the other parameters can be fixed at a nominal value. When wheat crop is simulated with a wet weather and a shallow soil (Fig. 5b), only the variable *Nit* has a coefficient of variation slightly higher than 1/3 ( $CV=0.38$ ). The corresponding parameters having a *ST* higher than 10% are *epc(2)*, *HCC(1)*, *HCC(2)* and *Norg*, meaning that these parameters are important to be determined accurately for simulating correctly the wheat crop in this case. The first main difference between the results presented in Fig. 5a and Fig. 5b is that only one output variable has a *CV* higher than 1/3 when the weather is wet, instead of three when the weather is dry. The second main difference is that the parameters *albedo* and *Hinit*, which contribute for a significant part of the output variance when the weather is dry, are replaced by the parameter *Norg*, which is involved in the mineralization process and contribute for a significant part of the variance of the outputs when the weather is wet.



variance of all the output variables during all the wheat crop season (see Section 3.1), proves to be the most important parameter to be determined accurately for simulating the wheat crop when the type of soil depth is shallow.

The Fig. 5c shows the results when the sugar beet crop is simulated with a dry weather and a deep soil. It reveals that all the output variables have a coefficient of variation higher than 1/3 meaning that the uncertainties on the soil parameters generate a large uncertainty on the considered variables. Among those parameters, five need to be measured accurately: *epc(2)*, *HCC(1)*, *HCC(2)*, *Hinit* and *q0*. The main difference between the results presented in Fig. 5a and Fig. 5c is that the entire output variables are strongly affected by the measurement of the soil parameters when the sugar beet is simulated. When sugar beet is simulated with a deep soil and a wet weather (Fig. 5d), only the output variables *LAI* and *QN* at the stage *AMF* and *Nit* have a CV higher than 1/3 (resp. CV=0.54, 0.53 and 0.68). For *LAI* and *QN* at the stage *AMF*, the only parameter having a *ST* higher than 10% is *HCC(1)*. For *Nit*, a lot of parameters exceeds this threshold: *argi*, *Norg*, *q0*, *epc(2)*, *HCC(1)*, *HCC(2)*, *Hinit* and *NO3init*. It is thus necessary to determine accurately a lot of parameters for simulating correctly the output *Nit*, while it is necessary to determine only one parameter for simulating correctly *LAI* and *QN* at the stage *AMF*. The main difference between these results and those presented in Fig. 5c is that, excepted for the output *Nit*, at most one parameter has to be accurately known for simulating correctly the sugar beet crop in deep soil and wet conditions. The parameter *HCC(1)*, which contributes for a large part to the variance of all the output variables during all the sugar beet crop season (see Section 3.1), proves to be the most important parameter to be measured accurately for simulating the sugar beet crop when the soil is deep.

## 4. Conclusion

Global sensitivity analysis is an interesting tool for ranking parameters with respect to their contribution to the variance of the output variables of a model. However, the only use of sensitivity indices proves to be unsatisfactory for deciding which parameters should be accurately measured in a given configuration. Only the combination of uncertainty and sensitivity analysis is relevant to reach this goal. Unfortunately too few studies consider simultaneously these two aspects. We

propose in this study a simple and easy to use method that combines these two methods in order to select the parameters that needs particular accuracy for simulating a set of variables of interest with an acceptable precision. The method has three steps: (i) compute the global sensitivity indices for each uncertain parameter (ii) compute the coefficient of variation of the outputs of interest from the set of simulations performed at step (i) and (iii) select the parameters to be accurately measured for simulating correctly these outputs by setting thresholds on sensitivity indices and coefficients of variation. Of course the results of this method are strongly linked to the uncertainties hypothesized for the parameters and special attention must be paid to this aspect. Coefficients of variation and sensitivity indices thresholds should be adapted to each case depending on the level of measurements constraints and of the accuracy wishes for model output simulations.

We apply this method to the crop model STICS for selecting soil parameters that need to be measured at a field scale. Practically this needs the knowledge of the conditions under which the crop grows (weather, type of soil depth, agricultural techniques ...) and it has been shown here that the results depend on these conditions. Concerning non-permanent soil parameters such as initial conditions, the application of the method needs thus to be based on future scenarios. However, this application shows that the number of STICS soil parameters to be measured accurately for simulating correctly the output variables considered here for wheat and sugar beet crops (given the parameters uncertainties used and in the configurations studied) can be significantly reduced by the use of this method. This is of particular interest given the time and financial cost of soil measurements.

**Acknowledgements:** The financial support by the CNES and Arvalis-Institut du Végétal is gratefully acknowledged.

## References

Aggarwal, P.K., 1995. Uncertainties in Crop, Soil and Weather Inputs Used in Growth-Models - Implications for Simulated Outputs and Their Applications. *Agricultural Systems* 48 361-384.



- Batchelor, W.D., Basso, B., Paz, J.O., 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy* 18 141-158.
- Blasone, R.S., Madsen, H., Rosbjerg, D., 2008. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* 353 18-32.
- Bouman, B.A.M., 1994. A Framework to Deal with Uncertainty in Soil and Management Parameters in Crop Yield Simulation - a Case-Study for Rice. *Agricultural Systems* 46 1-17.
- Bourennane, H., King, D., Couturier, A., Nicoullaud, B., Mary, B., Richard, G., 2007. Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. *Ecological Modelling* 205 323-335.
- 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., Launay, M., Mary, B., Beaudoin, N., 2008. Conceptual basis, formalisations and parameterization of the STICS crop model. Quae: Versailles.
- Brisson, N., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M.H., Bouthier, A., Ripoche, D., Mary, B., Juste, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie* 22 69-92.
- Campolongo, F., Saltelli, A., 1997. Sensitivity analysis of an environmental model an application of different analysis methods. *Reliability Engineering & System Safety* 57 49-69.
- Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2004. The role of sensitivity analysis in ecological modelling. 4th Conference of the International-Society-for-Ecological-Informatics. Busan, SOUTH KOREA.
- Chan, K., Tarantola, S., Saltelli, A., Sobol, I.M., 2001. Variance-based methods. In: Saltelli, A., Chan, K., Scott, E.M., (Eds.), *Sensitivity analysis*. Wiley: New York.

Ferreira, R.A., Jones, J.W.,Graham, W.D., 2006. Parameterizing Spatial Crop Models with Inverse Modeling: Sources of Error and Unexpected Results. Transactions of the ASABE 49 1547-1561.

Flenet, F., Villon, P.,Ruget, F.O., 2003. Methodology of adaptation of the STICS model to a new crop: spring linseed (*Linum usitatissimum*, L.). STICS Workshop. Camargue, FRANCE.

Gabrielle, B., Roche, R., Angas, P., Cantero-Martinez, C., Cosentino, L., Mantineo, M., Langensiepen, M., Henault, C., Laville, P., Nicoullaud, B.,Gosse, G., 2002. A priori parameterisation of the CERES soil-crop models and tests against several European data sets. *Agronomie* 22 119-132.

Gomez-Delgado, M.,Tarantola, S., 2006. GLOBAL sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science* 20 449-466.

Guérif, M., Beaudoin, N., Durr, C., Machet, J.M., Mary, B., Michot, D., Moulin, D., Nicoullaud, B.,Richard, G., 2001. Designing a field experiment for assessing soil and crop spatial variability and defining site specific management strategies. Proceedings 3rd European Conference on Precision Agriculture. Montpellier, France.

Guérif, M.,Duke, C., 1998. Calibration of the SUCROS emergence and early growth module for sugar beet using optical remote sensing data assimilation. *European Journal of Agronomy* 9 127-136.

Hadria, R., Khabba, S., Lahrouni, A., Duchemin, B., Chehbouni, A., Carriou, J.,Ouzine, L., 2007. Calibration and validation of the STICS crop model for managing wheat irrigation in the semi-arid Marrakech/Al Haouzi plain. *Arabian Journal for Science and Engineering* 32 87-101.

Houlès, V., 2004. Mise au point d'un outil de modulation intra-parcellaire de la fertilisation azoté du blé d'hiver basé sur la télédétection et un modèle de culture. Thèse, Institut National Agronomique Paris-Grignon.

Houlès, V., Mary, B., Guérif, M., Makowski, D.,Juste, E., 2004. Evaluation of the crop model STICS to recommend nitrogen fertilization rates according to agro-environmental criteria. *Agronomie* 24 1-9.

Irmak, A., Jones, J.W., Batchelor, W.D.,Paz, J.O., 2001. Estimating Spatially Variable Soil Properties for Application of Crop Models in Precision Farming. Transactions of the ASAE 44 1343-1353.



King, D., Daroussin, J., Tavernier, R., 1994. Development of a Soil Geographic Database from the Soil Map of the European Communities. *Catena* 21 37-56.

Launay, M., Graux, A.-I., Brisson, N., Guéris, M., 2009. Carbohydrate remobilization from storage root to leaves after a stress release in sugar beet (*Beta vulgaris* L.): experimental and modelling approaches. *The Journal of Agricultural Science*, accepted for publication.

Launay, M., Guéris, M., 2003. Ability for a model to predict crop production variability at the regional scale: an evaluation for sugar beet. *Agronomie* 23 135-146.

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

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

Nemes, A., Timlin, D.J., Pachepsky, Y.A., Rawls, W.J., 2009. Evaluation of the Rawls et al. (1982) Pedotransfer Functions for their Applicability at the US National Scale. *Soil Science Society of America Journal* 73 1638-1645.

Pathak, T.B., Fraisse, C.W., Jones, J.W., Messina, C.D., Hoogenboom, G., 2007. Use of Global Sensitivity Analysis for CROPGRO Cotton Model Development. *Transactions of the ASABE* 50 2295-2302.

Ratto, M., Young, P.C., Romanowicz, R., Pappenberger, F., Saltelli, A., Pagano, A., 2007. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences* 11 1249-1266.

Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis* 26 445-460.

Saltelli, A., Chan, K., Scott, E.M., 2000a. *Sensitivity Analysis*. John Wiley and Sons.

Saltelli, A., Tarantola, S., Campolongo, F., 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15 377-395.

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41 39-56.

Samouelian, A., Cousin, I., Tabbagh, A., Bruand, A., Richard, G., 2005. Electrical resistivity survey in soil science: a review. *Soil & Tillage Research* 83 173-193.

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

St'astna, M., Zalud, Z., 1999. Sensitivity analysis of soil hydrologic parameters for two crop growth simulation models. *Soil & Tillage Research* 50 305-318.

Tolson, B.A., Shoemaker, C.A., 2008. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research* 44.

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

van der Keur, P., Hansen, J.R., Hansen, S., Refsgaard, J.C., 2008. Uncertainty in simulation of nitrate leaching at field and catchment scale within the Odense river basin. *Vadose Zone Journal* 7 10-21.

Varella, H., Guérif, M., Buis, S., 2009. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling & Software*. In press. doi:10.1016/j.envsoft.2009.09.012.

## **Chapitre 4. L'analyse de sensibilité globale mesure la qualité d'estimation des paramètres**

***Article 2 : "Global sensitivity analysis measures the quality of parameter estimation. The case of soil parameters and a crop model". Accepté pour publication dans Environmental Modelling & Software.***

### **4.1. Objectif**

Le sous-groupe de paramètres du sol sélectionné au Chapitre 3 est à estimer par inversion du modèle STICS avec des observations du couvert végétal. Cependant, les paramètres n'auront pas tous la même qualité d'estimation car le jeu d'observations utilisé pour faire l'inversion ne fournit par la même quantité d'information sur chacun d'entre eux. La qualité d'estimation des paramètres est définie ici comme étant l'amélioration de l'estimation par inversion, avec un jeu d'observations donné, par rapport à l'estimation issue de l'information a priori (sa valeur moyenne). La qualité d'estimation est étroitement liée à la quantité d'information apportée par le jeu d'observations sur les paramètres, dans le sens où plus cette quantité est grande meilleure est la qualité. Nous avons vu que dans le cas du modèle linéaire (voir Chapitre 1.3.1), la matrice de Fisher permet de déterminer cette quantité d'information pour un jeu d'observations donné. Il n'existe pas de correspondance dans le cas de modèle non linéaire comme STICS, mais nous avons vu (dans le Chapitre 1.3.1) que dans ce cas, la quantité d'information peut être déterminée en fonction des résultats de l'analyse de sensibilité globale des variables observables du modèle aux paramètres à estimer. C'est l'objectif de l'article qui suit.

## 4.2. Méthodes

Dans l'article qui suit, nous proposons de déterminer la quantité d'information contenue dans les observations du couvert végétal à partir des indices de sensibilité calculés par la méthode Extended FAST. Nous proposons alors deux paires de critères, une au niveau du paramètre et une au niveau du sous-groupe de paramètres. Chaque paire est composée :

- d'un critère basé sur les indices principaux et totaux, qui permet de synthétiser les résultats de l'analyse de sensibilité afin de déterminer la quantité d'information fournie par le jeu d'observations sur le ou les paramètres du sol,
- d'un critère basé sur la *RMSE* (erreur moyenne d'estimation), qui définit la qualité d'estimation du ou des paramètres comme étant l'amélioration de l'estimation par inversion relativement à l'estimation issue de l'information a priori (sa valeur moyenne).

L'information a priori, utilisée par la méthode d'estimation de paramètres, ainsi que l'incertitude sur les paramètres, utilisée par la méthode d'analyse de sensibilité, sont déterminées à partir des gammes de variation des valeurs mesurées des paramètres du sol établies sur les deux parcelles de Chambry (voir Chapitre 2.4). L'estimation des paramètres du sol par inversion est réalisée en utilisant la méthode Importance Sampling (voir Chapitre 2.3) et des observations synthétiques du couvert végétal de blé d'hiver, dans différentes configurations composée de :

- quatre climats contrastés caractérisés comme sec, moyen sec, moyen humide et humide (parmi ceux présentés au Chapitre 2.4.3),
- deux gammes de profondeurs de sol (de 30 à 100 cm pour les sols peu profonds et de 80 à 160 cm pour les sols profonds),
- deux types de précédents culturaux (betterave et pois, ce qui affecte la gamme de variation de *NO3init*),
- trois types/tailles de jeux d'observations (*LAI*, *LAI+QN*, *LAI+QN+rendement*).

L'utilité des observations synthétiques réside dans le fait qu'il est possible d'analyser la pertinence des critères proposés, indépendamment d'éventuels biais dans le modèle STICS ou dans les observations.

### 4.3. Résultats

#### *Précision d'estimation des paramètres*

Tout d'abord, les précisions d'estimation des paramètres du sol seront présentées, à savoir non relativement à l'estimation issue des valeurs moyennes de l'information a priori (résultats non présentés dans l'article), même si la précision d'estimation reste liée à l'information a priori. Le Tableau 4-1 montre la précision moyenne d'estimation des paramètres du sol : seuls les paramètres *HCC(1)*, *HCC(2)* et *Hinit* (en conditions climatiques sèches) présentent une erreur *RRMSE* en deçà de 20% (entre 16.9 et 17.9% pour *HCC(1)*, entre 16.2 et 19.6% pour *HCC(2)* et 18.1% pour *Hinit* en conditions climatiques sèches). Ces trois paramètres sont donc ceux qu'il est possible d'estimer avec le plus de précision, avec des observations du couvert végétal de blé et compte tenu de l'information a priori utilisée. Notons que le paramètre *epc(2)* devient lui aussi estimable avec une précision du même ordre lorsque le type de profondeur de sol est profond.

	Condition	<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>
<i>RRMSE (%)</i>	-	24.9	25.9	40.6	<b>17.9</b>	<b>18.6</b>	25.3	35.9
	+	24.2	26.3	23.4	<b>16.9</b>	<b>17.2</b>	23.9	36.3
	sec	25.8	26.8	31.4	<b>17</b>	<b>16.2</b>	<b>18.1</b>	33.7
	humide	23.4	25.4	32.6	<b>17.8</b>	<b>19.6</b>	31.2	38.5

**Tableau 4-1.** Précision moyenne d'estimation des paramètres du sol (*RRMSE*) avec des observations synthétiques de blé, sachant la condition agropédoclimatique : faible profondeur de sol (-), forte profondeur de sol (+), climat sec (sec) ou climat humide (humide). En gras les *RRMSE* inférieurs à 20%.

#### *Amélioration de l'estimation des paramètres*

A présent, nous allons présenter l'amélioration de l'estimation des paramètres du sol par inversion du modèle, calculée relativement à l'estimation issue des valeurs moyennes de l'information a priori (résultats non présentés dans l'article). Le Tableau 4-2 montre ces améliorations calculées par le critère *RE* : plus le critère est faible, plus l'amélioration est forte. Globalement, les améliorations les plus fortes concernent *Hinit*, *HCC(1)*, *epc(2)* et *HCC(2)*. Ces paramètres sont également ceux

qui ont la meilleure précision d'estimation. Notons cependant qu'il est possible que l'estimation d'un paramètre soit améliorée même si sa performance d'estimation est faible. Par exemple, le paramètre  $epc(2)$  a un critère  $RE$  de 0.73, lorsque la profondeur de sol est faible, alors que qu'il a un critère  $RRMSE$  de 40.6%. Si une faible précision d'estimation est liée à une forte amélioration de l'estimation (cas de  $epc(2)$ ), cela signifie que l'amélioration s'est opérée sur une large gamme de variation, rendant la précision peu significative. L'amélioration traduit l'apport de la procédure d'estimation des paramètres par inversion sur ce que l'on sait a priori des paramètres.

	Condition	$argi$	$Norg$	$epc(2)$	$HCC(1)$	$HCC(2)$	$Hinit$	$NO3init$
$RE$	–	0.96	0.98	<b>0.73</b>	<b>0.86</b>	<b>0.89</b>	<b>0.58</b>	0.91
	+	0.93	1	0.91	<b>0.81</b>	<b>0.83</b>	<b>0.55</b>	0.92
	sec	0.99	1.02	<b>0.80</b>	<b>0.82</b>	<b>0.78</b>	<b>0.41</b>	<b>0.85</b>
	humide	0.90	0.96	<b>0.84</b>	<b>0.85</b>	0.94	<b>0.71</b>	0.98

**Tableau 4-2.** Amélioration moyenne de l'estimation des paramètres du sol (critère  $RE$ ) avec des observations synthétiques de blé, sachant la condition agropédoclimatique : faible profondeur de sol (–), forte profondeur de sol (+), climat sec (sec) ou climat humide (humide). En gras les  $RE$  inférieurs à 0.90.

La configuration d'observation a un effet significatif sur l'amélioration de l'estimation. En effet, les paramètres liés à la capacité de rétention en eau du sol ( $epc(2)$ ,  $HCC(1)$ ,  $HCC(2)$  et  $Hinit$ ) ont une amélioration plus importante lorsque le climat est sec. Aussi, l'amélioration du paramètre  $epc(2)$  est plus forte en faible profondeur de sol. Ces résultats peuvent être expliqués par la sensibilité aux paramètres estimés. En climat sec, le modèle est plus sensible aux paramètres liés à l'eau du sol car ils s'expriment plus qu'en climat humide : ces paramètres sont donc plus facilement estimables en climat sec. En faible profondeur, le modèle est plus sensible à l'épaisseur du sol car elle peut limiter l'avancée du front racinaire lorsqu'elle est faible :  $epc(2)$  est donc plus facilement estimable en faible profondeur. Dans la partie suivante, nous allons présenter le lien qui existe entre analyse de sensibilité et amélioration de l'estimation des paramètres.

### *Lien entre analyse de sensibilité globale et amélioration de l'estimation*

En ce qui concerne l'étude du lien entre analyse de sensibilité et estimation des paramètres, les critères basés sur les indices de sensibilité globaux se sont révélés efficaces pour mesurer la quantité d'information contenue dans les observations et son lien avec l'amélioration de l'estimation des paramètres, appelée qualité d'estimation dans l'article. Le critère  $GMS_i$ , qui concerne le paramètre, est capable de distinguer les paramètres qui peuvent être améliorés de ceux qui ne le peuvent pas (et qui peuvent être fixés à une valeur nominale). Plus précisément, il permet de classer les paramètres entre eux en fonction de leur qualité d'estimation, notée  $RE_i$ . Quant au critère  $TGMS$ , qui concerne l'ensemble des paramètres, il permet de classer les climats et les jeux d'observations entre eux en fonction de leur capacité à estimer l'ensemble des paramètres par inversion. Il est donc lié à la qualité d'estimation de l'ensemble des paramètres, notée  $TRE$ .

L'article montre que les critères proposés permettent de diagnostiquer et de pronostiquer la qualité d'estimation des paramètres en fonction du jeu d'observations utilisé, ainsi que de classer les paramètres, les climats et les jeux d'observations entre eux en fonction de la quantité d'information contenue dans les observations et de la qualité d'estimation des paramètres du sol. Ces résultats peuvent être utilisés à des fins de compréhension des performances d'estimation des paramètres. Par exemple, lorsque le critère (celui défini au niveau du paramètre) est négatif pour un paramètre donné, la performance d'estimation est mauvaise car il agit sur les variables observées essentiellement à travers ses interactions. Lorsque le critère est nul, la performance est également mauvaise mais parce qu'il agit de manière globalement insignifiante sur les variables observées. Ces résultats peuvent aussi être utilisés à des fins d'optimisation du jeu d'observations pour maximiser les performances d'estimation des paramètres et minimiser le coût des observations. Par exemple, si l'on désire estimer un paramètre en particulier, il est possible de trouver un jeu d'observations hypothétique qui maximiserait le critère associé à ce paramètre, avant de procéder à un véritable recueil de ces observations.

L'estimation des paramètres peut être une fin en soi, à savoir qu'il est possible de n'utiliser STICS que par inversion, ou bien être le moyen par lequel les prédictions

du modèle peuvent être améliorées. Pour cela, il est nécessaire de comprendre comment les performances de prédiction du modèle réagissent en fonction des performances d'estimation des paramètres du sol. Cette démarche de compréhension sera traitée dans l'article présenté au chapitre suivant.

#### **4.4. Article 2 : “Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model”**

Article accepté pour publication dans *Environmental Modelling & Software* (Varella et al., 2009) :

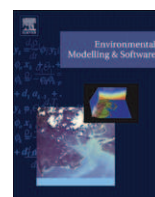
Varella, H., Guérif, M., Buis, S. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling & Software*. In press. doi:10.1016/j.envsoft.2009.09.012.





Contents lists available at ScienceDirect

# Environmental Modelling & Software

journal homepage: [www.elsevier.com/locate/envsoft](http://www.elsevier.com/locate/envsoft)

## Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model

Hubert Varella\*, Martine Guérif, Samuel Buis

INRA, UMR 1114 INRA-UAPV EMMAH, Domaine Saint Paul, Site Agroparc, F84914 Avignon Cedex 9, France

### ARTICLE INFO

#### Article history:

Received 17 March 2009

Received in revised form

29 September 2009

Accepted 30 September 2009

Available online 17 October 2009

#### Keywords:

Global sensitivity analysis

Extended FAST

Bayesian parameter estimation

GLUE

Crop model STICS

Soil parameters

### ABSTRACT

One common limitation of the use of crop models for decision making in precise crop management is the need for accurate values of soil parameters for a whole field. Estimating these parameters from data observed on the crop, using a crop model, is an interesting possibility. Nevertheless, the quality of the estimation depends on the sensitivity of model output variables to the parameters. The goal of this study is to explain the results for the quality of parameter estimation based on global sensitivity analysis (GSA). The case study consists of estimating the soil parameters by using the STICS-wheat crop model and various synthetic observations on wheat crops (*LAI*, absorbed nitrogen and grain yield). Suitable criteria summarizing the sensitivity indices of the observed variables were created in order to link GSA indices with the quality of parameter estimation. We illustrate this link on 16 different configurations of different soil, climatic and crop conditions. The GSA indices were computed by the Extended FAST method and a function of RMSE was computed with an importance sampling method based on Bayes theory (GLUE). The proposed GSA-based criteria are able to rank the parameters with respect to their quality of estimation and the different configurations (especially climate and observation set) with respect to their ability to estimate the whole parameter set. They may be used as a tool for predicting the performance of different observation datasets with regard to parameter estimation.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Dynamic crop models are very useful to predict the behavior of crops in their environment and are widely used in a lot of agro-environmental work such as crop monitoring, yield prediction or decision making for cultural practices (Batchelor et al., 2002; Gabrielle et al., 2002; Houlès et al., 2004). These models usually have many parameters and their estimation is a major problem for agro-environmental prediction (Tremblay and Wallach, 2004; Makowski et al., 2006). For spatial application, the knowledge of soil parameters is crucial since they are responsible for a major part of the variability of the crop model output variables of interest (Launay and Guérif, 2003; Irmak et al., 2001; Ferreyra et al., 2006). However, knowledge on soil properties and therefore model soil parameters is scarcely available at an appropriate scale. Direct

measurements would be highly costly and time consuming. Detailed soil maps adapted to the field scale are scarcely available, while the use of more automated techniques like remote sensing or electrical resistivity is still hampered by a lack of robust interpretation of the signal (Samouelian et al., 2005; Lagacherie et al., 2008). Moreover, these techniques do not permit to access the values of all the soil parameters required to apply a complex crop model. Fortunately, observations on crops provided by remote sensing (Weiss and Baret, 1999; Houborg and Boegh, 2008) or yield monitoring (Blackmore and Moore, 1999; Pierce et al., 1999) are widely available and allow soil parameters being estimated through the inversion of crop models. Thus Timlin et al. (2001) explored the possibility of finding the Water Holding Capacity of three soil layers by using maize yield maps and a very simple model; Braga and Jones (2004) compared the performances of two kinds of data (maize yield and soil water content) for estimating five soil parameters in nine soil layers, using the CERES crop model; Guérif et al. (2006) estimated 12 soil parameters from the assimilation of leaf area indices and absorbed nitrogen estimated from remote sensing.

\* Corresponding author. Tel.: +33 432 72 24 07; fax: +33 432 72 23 62.  
E-mail address: hvvarella@avignon.inra.fr (H. Varella).

Although the estimation process is quite efficient when a small number of parameters have a clearly identifiable influence on the observed output variables, estimating parameters of complex models such as crop models may be not so easy (Tremblay and Wallach, 2004; Launay and Guérif, 2005). It is clearly a difficult problem, as a lot of parameters have similar effects and observations are generally scarce. The use of Bayesian methods, including prior information on parameters, is a way of improving the estimation process, but the results can be deceptive (Guérif et al., 2006). One of the reasons for the difficulties encountered may be a lack of sensitivity of the observed variables to the parameters, making the estimation process inefficient. Another reason may be that the influence of the parameters on the observed variables takes place mainly through interactions, making it difficult to identify the relevant factors (Saltelli et al., 2000b). The problem of parameter identifiability is commonly studied using local sensitivity analysis (Brun et al., 2002; Rodriguez-Fernandez et al., 2006). However, these techniques are only well suited for linear or near-linear models. For complex non-linear models, only global sensitivity analysis (GSA) methods are able to give relevant information on the sensitivity of model outputs to the whole range of variation of model inputs. In recent years, many studies have focused on this subject, namely, how to choose the main parameters to be estimated (Campolongo and Saltelli, 1997; Gomez-Delgado and Tarantola, 2006; Manache and Melching, 2008; Post et al., 2008). Some of them have used crop models (Ruguet et al., 2002; Makowski et al., 2006; Jongschaap, 2007; Pathak et al., 2007) and ranked the importance of the parameters by calculating global sensitivity indices: first-order indices (the main effect of the parameter on the output) and total indices (sum of all effects involving the parameter, including the interactions with other parameters). The common practice is consistent with the principles expressed by Ratto et al. (2007). Small total sensitivity indices indicate a negligible effect of the parameter on the model output concerned. These parameters can be fixed at a nominal value ("Factor Fixing setting"). High first-order indices reveal a clearly identifiable influence of the parameter on the model output concerned, and therefore the parameters need to be determined accurately ("Factor Prioritization setting"). Small first-order indices combined with large interaction indices result in a lack of identification. In practice, the two first rules are commonly used to select the set of parameters to be estimated in a calibration problem. GSA can also be used to evaluate the quantity of information contained in a given set of observations for estimating parameters and thus to determine which is the best observation set for estimating the parameters (Kontoravdi et al., 2005). Although the results of GSA are often used to design the estimation process, the link between GSA indices and the quality of parameter estimation has never been quantified.

Our objectives in this study are twofold. Firstly, to evaluate the feasibility of characterizing soil properties of agricultural fields by the inversion of a dynamic crop model, using the many observations collected on those fields either from remote sensing or by yield monitoring. In particular, we will investigate which parameters are the most accessible in the configurations of observations available. Secondly, to use GSA results in order to measure the quantity of information contained in different sets of observations and to illustrate the link between this measurement and the quality of parameter estimates. As the performance of the estimation process is supposed to depend on several conditions such as soil type, cropping conditions (preceding crop and climate) or available observations, we chose to conduct the study on synthetic data in order to be able to generate variability in parameter retrieval performance as well as in sensitivity structure of the observed model outputs to soil parameters. This choice also allows eliminating the impact of model errors, which may complicate the interpretation of the results.

## 2. Methods

### 2.1. The crop model STICS, its parameters and output variables

#### 2.1.1. The STICS model

The STICS model (Brisson et al., 2002) is a non-linear dynamic crop model simulating the growth of various crops. For a given crop, STICS takes into account the climate, type of soil and cropping techniques to simulate the carbon, water and nitrogen balances of the crop-soil system on a daily time scale. In this study, a wheat crop is simulated. The crop is essentially characterized by its above-ground biomass carbon and nitrogen, and leaf area index. The main outputs are agronomic variables (yield, grain protein content) as well as environmental variables (water and nitrate leaching). Yield, grain protein content and nitrogen balance in the soil at harvest are of particular interest for decision making, especially for monitoring nitrogen fertilization (Houlès et al., 2004). Nitrogen absorbed by the plant and leaf area index are also important to analyze the health and growth of the plant during the crop's growing season. The STICS model includes more than 200 parameters arranged in three main groups: those related to the soil, those related to the characteristics of the plant or to the genotype, and those describing the cropping techniques. The values of the last group of parameters are usually known as they correspond to the farmer's decisions. The parameters related to the plant are generally determined either from literature, from experiments conducted on specific processes included in the model (e.g. mineralization rate, critical nitrogen dilution curve etc.) or from calibrations based on large experimental database, as is the case for the STICS model (Flenet et al., 2003). The soil parameters are difficult to determine at each point of interest and are responsible for a large part of the spatial variability of the output variable. That is why the sensitivity analysis and parameter estimation processes described in this study only concern soil parameters.

#### 2.1.2. The soil parameters

The STICS model contains about 60 soil parameters. In our case, in order to limit the problems of identifiability, the number of soil parameters to be estimated has been reduced. First, among the available options for simulating the soil system, the simplest was chosen, after checking that the model was valid for the conditions explored in our application. We then considered the soil as a succession of two horizontal layers, each characterized by a specific thickness parameter. From the observation of the tillage practices in the region around our experimental site, the thickness of the first layer was set at 0.30 m. Based on the measurements made on this precision agriculture experimental site in Chambry (49.35°N, 3.37°E) (Guérif et al., 2001), we added relations, specific to our conditions, linking the initial contents of water *Hinit* and mineral nitrogen *NO3init* of the two soil layers. We performed a first sensitivity analysis on the 13 resulting soil parameters. This allowed us to fix those whose effects on the observed variables were negligible: for each parameter we computed the values of its effects on all the observed variables considered for a lot of soil, climate and agronomic conditions, and dropped the parameters for which all these values were less than 10% of the total effects generated by the 13 parameters. We thus restricted the study to seven parameters.

The seven soil parameters considered (Table 1) characterize both water and nitrogen processes. They refer to permanent characteristics and initial conditions. Among the permanent characteristics, clay and organic nitrogen content of the top layer are involved mainly in organic matter decomposition processes and nitrogen cycle in the soil. Water content at field capacity of both layers affects the water (and nitrogen) movements and storage in the soil reservoir. Finally, the thickness of the second layer defines the volume of the reservoir. The initial conditions correspond to the water and nitrogen content, *Hinit* and *NO3init*, at the beginning of the simulation, in this case the sowing date.

**Table 1**

The 7 soil parameters and their ranges of variation.

Parameter	Definition	Range	Unit
<i>argi</i>	Clay content of the 1st layer	14–37	%
<i>Norg</i>	Organic nitrogen content of the 1st layer	0.049–0.131	%
<i>epc(2)</i>	Thickness of the 2nd layer	0–70 or 50–130 <sup>a</sup>	cm
<i>HCC(1)</i>	Water content at field capacity (1st layer)	14–30	g g <sup>-1</sup>
<i>HCC(2)</i>	Water content at field capacity (2nd layer)	14–30	g g <sup>-1</sup>
<i>Hinit</i>	Initial water content (both layers)	4–29	% of weight
<i>NO3init</i>	Initial mineral nitrogen content (1st layer)	4–21.5 or 25–86 <sup>b</sup>	kgN ha <sup>-1</sup>

<sup>a</sup> The first range is for a shallow soil and the second for a deep soil.

<sup>b</sup> The first range is for a wheat cultivated after sugar beet and the second for a wheat cultivated after pea.

### 2.1.3. Observable output variables

In this study, we focus on output variables of the STICS-wheat model which correspond to possible observations on the canopy. They consist of:

- the leaf area index ( $LAI_t$ ) and the nitrogen absorbed by the plant ( $QN_t$ ) at various dates  $t, t = 1, \dots, T$ , during the crop season – possibly derived from a set of remote sensing images (Weiss and Baret, 1999; Houborg and Boegh, 2008),
- the yield at harvest ( $Yld$ ), possibly provided by yield monitoring systems.

## 2.2. Global sensitivity analysis

Methods of global sensitivity analysis are particularly interesting because they allow the full uncertainty range of the parameters to be explored and analyzed through a complex model, and their interactions to be estimated by varying them concurrently (Saltelli et al., 2000a). Among these methods, variance-based methods use model output variance as an indicator of importance for parameters. When the number of parameters is not too large, as it is the case in this study, a variance-based method can be applied directly on the complex model. When this number is very large, it is possible to apply this method on a metamodel (polynomial approximation of the model) to calculate sensitivity indices very efficiently (Ziehn and Tomlin, 2009). A variance-based method generally needs a lot of model simulations but its application does not rely on special assumptions about the behavior of the model (such as linearity, monotonicity and additivity of the relationship between input factors and model output). Because of the non-linear structure of the STICS model and the acceptable computational cost of one run of the model (about 1s), we chose this type of method.

### 2.2.1. Variance decomposition method and sensitivity indices

We denote a given output variable of the STICS model as  $Y$ . The total variance of  $Y$ ,  $V(Y)$ , caused by variation in the 7 selected soil parameters  $\theta$ , can be partitioned as follows (Chan et al., 2000):

$$V(Y) = \sum_{i=1}^7 V_i + \sum_{1 \leq i < j \leq 7} V_{ij} + \dots + V_{1,2,\dots,7}, \quad (1)$$

where  $V_i = V[E(Y|\theta_i)]$  measures the main effect of the parameter  $\theta_i, i = 1, \dots, 7$ , and the other terms measure the interaction effects. Decomposition (2) is used to derive two types of sensitivity indices defined by:

$$S_i = \frac{V_i}{V(Y)}, \quad (2)$$

$$ST_i = \frac{V(Y) - V_{-i}}{V(Y)}, \quad (3)$$

where  $V_{-i}$  is the sum of all the variance terms that do not include the index  $i$ .

$S_i$  is the first-order (or main) sensitivity index for the  $i$ th parameter. It computes the fraction of  $Y$  variance explained by the uncertainty of parameter  $\theta_i$  and represents the main effect of this parameter on the output variable  $Y$ .

$ST_i$  is the total sensitivity index for the  $i$ th parameter and is the sum of all effects (first and higher order) involving the parameter  $\theta_i$ .

$S_i$  and  $ST_i$  are both in the range  $(0, 1)$ , low values indicating negligible effects, and values close to 1 huge effects.  $ST_i$  takes into account both  $S_i$  and the interactions between the  $i$ th parameter and the 6 other parameters, interactions which can therefore be assessed by the difference between  $ST_i$  and  $S_i$ . The interaction terms of a set of parameters represent the fraction of  $Y$  variance induced by the variance of these parameters but that cannot be explained by the sum of their main effects. The two sensitivity indices  $S_i$  and  $ST_i$  are equal if the effect of the  $i$ th parameter on the model output is independent of the values of the other parameters: in this case, there is no interaction between this parameter and the others and the model is said to be additive with respect to  $\theta_i$ .

### 2.2.2. Extended FAST

We have chosen here to use the extended FAST (EFAST) method, which has been proved, in several studies (Saltelli and Bolado, 1998; Saltelli et al., 1999; Makowski et al., 2006), to be more efficient in terms of number of model evaluations. The main difficulty in evaluating the first-order and total sensitivity indices is that they require the computation of high dimensional integrals. The EFAST algorithm performs a judicious deterministic sampling to explore the parameter space which makes it possible to reduce these integrals to one-dimensional ones using Fourier decompositions. The reader interested in a detailed description of EFAST can refer to Saltelli et al. (1999). We have implemented the EFAST method in the Matlab® software. The uncertainties considered for the soil parameters are assumed independent and follow uniform distributions. These distributions are given in Section 2.5. A preliminary study of the convergence of the sensitivity indices allowed us to set the number of simulations per parameter to 2000, leading to a total number of model runs of  $7 \times 2000 = 14000$  to compute the main and total effects for all output variables and parameters considered here. One run of the STICS model taking about 1s with a Pentium 4, 2.9 GHz processor, the overall simulation process takes about 4 h.

### 2.2.3. Criteria based on GSA indices

GSA provides main and total indices per parameter for each output variable considered. In order to summarize this information, we propose to create different criteria.

- (i) The first one is a global measure of the information contained in a set of observations to estimate each parameter:

The Global Mean Sensitivity ( $GMS_i$ ) computes the mean of the main effect of parameter  $\theta_i$  minus its interactions with the other parameters for all observed variables, each component being weighted by the degree of dependence of the corresponding output variable with the other ones in order to account for the redundancy between observable output variables:

$$GMS_i = \frac{1}{K} \sum_{k=1}^K (1 - \alpha_k) (S_i^k - R_i^k) \quad (4)$$

where  $k$  is a given observed output variable in a subset composed of  $K$  variables among  $\{LAI_t, QN_t, t = 1, \dots, T, Yld\}$ ,  $R_i^k = ST_i^k - S_i^k$  is the sum of all interaction terms including parameter  $\theta_i$  for the observed variable  $k$ .  $0 \leq \alpha_k \leq 1$  is the mean of the absolute values of the correlation coefficients  $|r_{kk'}|$  between the variable  $k$  and the other variables  $k'$  (calculated on the model simulations required for GSA):  $\alpha_k = \frac{1}{K-1} \sum_{k' \neq k} |r_{kk'}|, K > 1$ .

The  $GMS_i$  criterion is based on the following rules:

- if  $ST_i^k$  is low (and thus  $S_i^k$ ), observation  $k$  is assumed not to contain enough information to estimate parameter  $\theta_i (S_i^k - R_i^k)$  should be low,
- if  $S_i^k$  is high (and thus  $ST_i^k$ ), observation  $k$  is assumed to contain sufficient information to estimate parameter  $\theta_i (S_i^k - R_i^k)$  should be high,
- if  $S_i^k$  is low and  $ST_i^k$  is high, then the model is over-parameterized and difficulties in identifying parameter  $\theta_i$  are expected (Ratto et al., 2007):  $(S_i^k - R_i^k)$  should be low,
- high correlation between output variables indicates that the information content of these variables is redundant:  $(1 - \alpha_k)$  should be reduced.

$GMS_i$  varies within the range  $[-1, 1]$ . It tends to 1 when  $S_i^k$  is close to 1 for all observed variables and when all the observed variables are perfectly uncorrelated: in this case the model has an additive structure for the parameter  $\theta_i$  and this parameter has a clearly identifiable influence on the  $K$  observed variables.  $GMS_i$  tends to  $-1$  when  $S_i^k$  and  $R_i^k$  are close to 0 and 1 respectively for all observed variables and when all the observed variables are perfectly uncorrelated: in this case problems of identification of the parameter  $\theta_i$  are expected.

- (ii) The second criterion is calculated at the whole parameter set level:

The Total Global Mean Sensitivity ( $TGMS$ ) is the sum of the  $GMS_i$  for all parameters:

$$TGMS = \sum_{i=1}^7 GMS_i = \sum_{i=1}^7 \frac{1}{K} \sum_{k=1}^K (1 - \alpha_k) (S_i^k - R_i^k) \quad (5)$$

It measures the information contained in a set of observations to estimate the all set of parameters considered. The  $TGMS$  criterion varies within the range  $[-7, 7]$ . It tends to 1 when  $R_i^k$  is close to 0 for all parameters and all observed variables and when all the observed variables are perfectly uncorrelated: in this case the model is additive.  $TGMS$  tends to  $-7$  when  $R_i^k$  is close to 1 for all parameters and all observed variables and when all the observed variables are perfectly uncorrelated: in this case the model is expected to be unidentifiable.

## 2.3. Parameter estimation

We chose a Bayesian method which allows existing information on the parameters to be taken into account in the estimation process and an estimate of the posterior probability distribution of parameter values to be computed.

### 2.3.1. The Bayesian approach and the GLUE method

The posterior parameter distribution is given by Bayes' theorem:

$$\pi(\theta/Z) = \frac{\pi(Z/\theta)\pi(\theta)}{\pi(Z)} \quad (6)$$

where  $Z$  is the vector of total observations of size  $K$ ,  $\pi(\theta/Z)$  is the posterior parameter distribution,  $\pi(\theta)$  is the prior parameter distribution,  $\pi(Z)$  is a constant of proportionality determined by the requirement that the integral of  $\pi(\theta/Z)$  over the parameter space equals 1, and  $\pi(Z/\theta)$  is the likelihood function of the data  $Z$  given the parameters  $\theta$ . Its value is determined from the probability distribution of the errors of modeled and observed data. It is readily seen that both the prior distribution and the new data affect the posterior parameter distribution.

We chose an Importance Sampling method, GLUE (Beven and Binley, 1992; Beven and Freer, 2001; Makowski et al., 2002) which principle is to approximate the posterior parameter distribution  $\pi(\theta|Z)$  given in (6) by a discrete probability distribution  $(\theta_n, p_n)$ ,  $n = 1, \dots, N$ ,  $\sum_{n=1}^N p_n = 1$ , where  $p_n$  is the probability associated with the parameter vector  $\theta_n$ . In our case, the method proceeds as follows:

- (1) Randomly generate  $N$  vectors  $\theta_n$ ,  $n = 1, \dots, N$ , from the prior parameter distribution  $\pi(\theta)$ .
- (2) Calculate the likelihood values  $\pi(Z|\theta_n)$  and the prior density  $\pi(\theta_n)$ ,  $n = 1, \dots, N$ , associated with the different generated parameter vectors.
- (3) Calculate  $p_n = \frac{\pi(Z|\theta_n)}{\sum_{m=1}^N \pi(Z|\theta_m)}$ .

The pairs  $(\theta_n, p_n)$ ,  $n = 1, \dots, N$ , can be used to determine various characteristics of the posterior distribution, including:

- the mean of the posterior joint distribution of  $\theta$ ,  $\bar{\theta}^{post} = \sum_{n=1}^N \theta_n p_n$
- the correlation between two parameters  $\theta_i$  and  $\theta_j$ ,  $\frac{\sum_{n=1}^N (\theta_{i,n} - \bar{\theta}_i^{post})(\theta_{j,n} - \bar{\theta}_j^{post}) p_n}{\delta_i \delta_j}$ , where  $\delta_i^2 = \sum_{n=1}^N \theta_{i,n}^2 p_n - (\bar{\theta}_i^{post})^2$  is the variance of the parameter  $\theta_i$ .

This correlation coefficient permits to evaluate the pair-wise interaction structure between parameters. In particular, when a positive coefficient is detected, the pair of parameters acts in the model as a quotient or a difference, and when it is negative they act as a product or a sum (Ratto et al., 2001; Kanso et al., 2004). A high level of correlation between two parameters indicates that they may compensate each other to give the same values of model outputs. In this case, bringing new information on one of them (by incorporating new measurements or improving its prior information) may also reduce the uncertainty on the other one.

We assume that the sums of the errors of modeled and observed data are independent between dates and variables and follow normal distributions of zero mean and standard deviation  $\sigma_k$ . Thus, we use the following likelihood function:

$$\pi(Z/\theta) = \prod_{k=1}^K \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2} [y_k - f_k(\theta^{true}, x)]^2\right\} \quad (7)$$

A prior assumption is that parameters are independent and their distribution uniform. They correspond to the uncertainties used for GSA and are given in Section 2.5. We have implemented the GLUE method in the Matlab® software. A preliminary study of the convergence of the estimates allowed us to set the total number of generated parameter vectors  $N$  at 100 000.

### 2.3.2. Criteria expressing the quality of parameter estimation

As in the previous section, we created criteria related to the quality of parameter estimation both for a given parameter and for the whole set of parameters. These criteria express the error of the estimate obtained by the mean of the posterior distribution of the parameter  $i$ ,  $i = 1, \dots, 7$ :  $\bar{\theta}_i^{post}$ , relative to that obtained from the mean of the prior distribution:  $\bar{\theta}_i^{prior}$ . This relative error can be considered as an indicator of the quality of the estimation process.

For one single parameter  $\theta_i$ , the Relative Error  $RE_i$  is defined as follows:

$$RE_i = \frac{RMSE(\bar{\theta}_i^{post})}{RMSE(\bar{\theta}_i^{prior})} \quad (8)$$

where  $RMSE(\bar{\theta}_i^{post}) = \sqrt{\frac{1}{P} \sum_{p=1}^P (\theta_{i,p}^{true} - \bar{\theta}_i^{post})^2}$  and similarly for  $RMSE(\bar{\theta}_i^{prior})$ , where  $\theta_{i,p}^{true}$  is one of the true  $P$  values of soil parameter  $\theta_i$  defined to generate the synthetic observations (see Section 2.5), and  $\bar{\theta}_i^{post}$  is the corresponding estimation given by the GLUE method.

For all parameters, the second criterion, called Total Relative Error ( $TRE$ ), is defined by the mean of the seven values of  $RE_i$ :

$$TRE = \frac{1}{7} \sum_{i=1}^7 RE_i = \frac{1}{7} \sum_{i=1}^7 \frac{RMSE(\bar{\theta}_i^{post})}{RMSE(\bar{\theta}_i^{prior})} \quad (9)$$

### 2.4. Quality of the link between GSA and parameter estimation

An important issue in this work concerns the quality of the relationship between the criteria based on GSA indices and the criteria expressing the quality of parameter estimation. If the link is strong, GSA can be used to predict which parameters may be estimated with some accuracy. We use Spearman's rank correlation analysis (Spearman, 1904) to assess the quality of the link. It allows quantifying the correlation between the parameter ranking list given by the criteria based on GSA indices and the one given by the criteria expressing the quality of parameter estimation. The Spearman's rank correlation coefficient  $\rho$ , between two lists  $x_1$  and  $x_2$ , is calculated as:

$$\rho = 1 - \frac{6 \sum_{s=1}^S (x_1^s - x_2^s)^2}{S(S^2 - 1)} \quad (10)$$

where  $S$  is the size of the vectors  $x_1$  and  $x_2$ .

## 2.5. Numerical experimentation

### 2.5.1. Configurations

The STICS model output variables depend on the soil, climate and agronomic conditions for which the crop is simulated. In view of this, we use different configurations in our study, as presented in Table 2: four contrasting climates, two different soil depths (shallow and deep), and two agronomic conditions (preceding crop sugar beet and peas). The climatic data used were obtained from the meteorological station of Roupy (49.48°N, 3.11°E). Four different sets of data were chosen to characterize a dry climate (1975–1976), a wet climate (1990–1991), a medium-dry climate (1979–1980) and a medium-wet climate (1972–1973). The bounds of soil parameters' distributions were deduced from experimental data acquired on a precision agriculture site in northern France (Chambry, 49.35°N, 3.37°E) (Guérif et al., 2001). They are given in Table 1. In this application, we assume that the type of soil depth (shallow or deep) and the preceding crop (sugar beet or pea) are known. Therefore two different ranges were considered for the depth of soil  $epc(2)$  and for the mineral nitrogen content at the beginning of the wheat crop simulation  $NO3ini$ .

### 2.5.2. Generation of observations

We consider observations on wheat crops obtained for the different configurations described before. They consist of  $LAI_t$  and  $QN_t$  available at ten dates  $t$ , distributed through the wheat growing season (November 15, December 12, January 15, February 16, March 15, April 05, April 19, May 03, May 17 and June 07) and  $Yld$  at harvest. Three possible sets of observations (see Table 3) were considered for the parameter estimation experiments. In order to compute the synthetic observations, 50 vectors of true values  $\theta^{true}$ , corresponding to 50 soils, were randomly generated from the distributions defined above. Corresponding values of STICS-wheat model output variables were simulated for each configuration leading to  $50 \times 16$  simulations. Observations  $y_{q,t}$  were then computed by adding a random error term to the simulated values of the variables and dates defined above:

$$y_{q,t} = f_{q,t}(\theta^{true}, x) + \epsilon_{q,t} \quad (11)$$

where  $f_{q,t}$  is the STICS model output  $q$  ( $Yld$ ,  $LAI_t$  or  $QN_t$ ) calculated on date  $t$  (harvest for  $Yld$  or  $t = 1, \dots, T$  for  $LAI$  and  $QN$ ),  $x$  is the vector of explanatory variables and  $\epsilon_{q,t}$  is the observation error term. Following the assumptions made in Section 2.3 to compute the likelihood function of the GLUE method, the vector of observation error is given by:  $\epsilon_{q,t} \sim N(0, \sigma_{q,t}^2)$  where  $\sigma_{q,t}^2 = \sigma_{q,t}^0 f_{q,t}(\theta^{true}, x)$ ,  $\sigma_{Yld}^0 = 9\%$ ,  $\sigma_{LAI}^0 = 17\%$  and  $\sigma_{QN}^0 = 30\%$  according to results obtained in field measurements on wheat crops (Machet et al., 2007; Moulin et al., 2007).

### 2.5.3. Total number of experiments

To sum up, GSA was applied to 16 configurations involving 21 output variables, GSA-based criteria were computed for 16 configurations  $\times$  3 sets of observations, GLUE was applied and  $RE_i$  and  $TRE$  criteria were computed for 16 configurations  $\times$  3 sets of observations  $\times$  50 soils defined by 50 different  $\theta^{true}$ .

## 3. Results

### 3.1. Global sensitivity analysis

Fig. 1a shows the results for one configuration (dry climatic conditions, shallow soil, cultivated after sugar beet) labeled *dry-beet*. The sensitivity indices for all the output variables and each

**Table 2**  
16 configurations based on soil, climatic and agronomic conditions.

Climatic conditions	Soil depth	Preceding crop	Configuration label
Dry	Shallow	Sugar beet	<i>dry-beet</i>
Medium-dry	Shallow	Sugar beet	<i>mdry-beet</i>
Medium-wet	Shallow	Sugar beet	<i>mwet-beet</i>
Wet	Shallow	Sugar beet	<i>wet-beet</i>
Dry	Deep	Sugar beet	<i>dry+beet</i>
Medium-dry	Deep	Sugar beet	<i>mdry+beet</i>
Medium-wet	Deep	Sugar beet	<i>mwet+beet</i>
Wet	Deep	Sugar beet	<i>wet+beet</i>
Dry	Shallow	Pea	<i>dry-pea</i>
Medium-dry	Shallow	Pea	<i>mdry-pea</i>
Medium-wet	Shallow	Pea	<i>mwet-pea</i>
Wet	Shallow	Pea	<i>wet-pea</i>
Dry	Deep	Pea	<i>dry+pea</i>
Medium-dry	Deep	Pea	<i>mdry+pea</i>
Medium-wet	Deep	Pea	<i>mwet+pea</i>
Wet	Deep	Pea	<i>wet+pea</i>



**Table 3**  
Description of the 3 sets of observations.

Set number	Variables used	Size $K$
1	$LAI_t$ on dates $t = 1, \dots, 10^a$	$K = 10$
2	$LAI_t$ and $QN_t$ on dates $t = 1, \dots, 10$	$K = 20$
3	$LAI_t$ and $QN_t$ on dates $t = 1, \dots, 10$ , and $Yld$	$K = 21$

<sup>a</sup> See dates in Section 2.5.

parameter are presented both for first-order indices  $S_i$  (area in black) and for interaction indices  $R_i = ST_i - S_i$  (area in gray). The corresponding  $GMS_i$  criterion computed for the 3 sets of output variables is given in Table 4. Only two parameters ( $Hinit$  and  $epc(2)$ ) have strong first-order indices and little interaction with the other parameters. The others have moderate ( $HCC(1)$  and  $HCC(2)$ ) to low ( $argi$ ,  $Norg$  and  $NO3init$ ) first-order and interaction indices. The parameters  $Hinit$  and  $epc(2)$  are involved in soil water content and are determinant for the crop in dry conditions. Parameters  $argi$  and  $Norg$  which are mainly involved in organic matter mineralization induce little sensitivity because this process is blocked in dry conditions. Accordingly, the  $GMS_i$  values computed for the set #3 are high for  $Hinit$  and  $epc(2)$  (0.145 and 0.073), and close to 0 or even negative for the other parameters. These nil values correspond to two situations: either both first-order indices and interaction are low ( $argi$  and  $Norg$ ), or both first-order indices and interaction are moderate but equal ( $HCC(1)$ ,  $HCC(2)$  and  $NO3init$ ). For parameter  $epc(2)$ ,  $LAI$  and  $QN$  have similar sensitivity profiles and their indices for the last date are close to those of  $Yld$ . Thus, the information

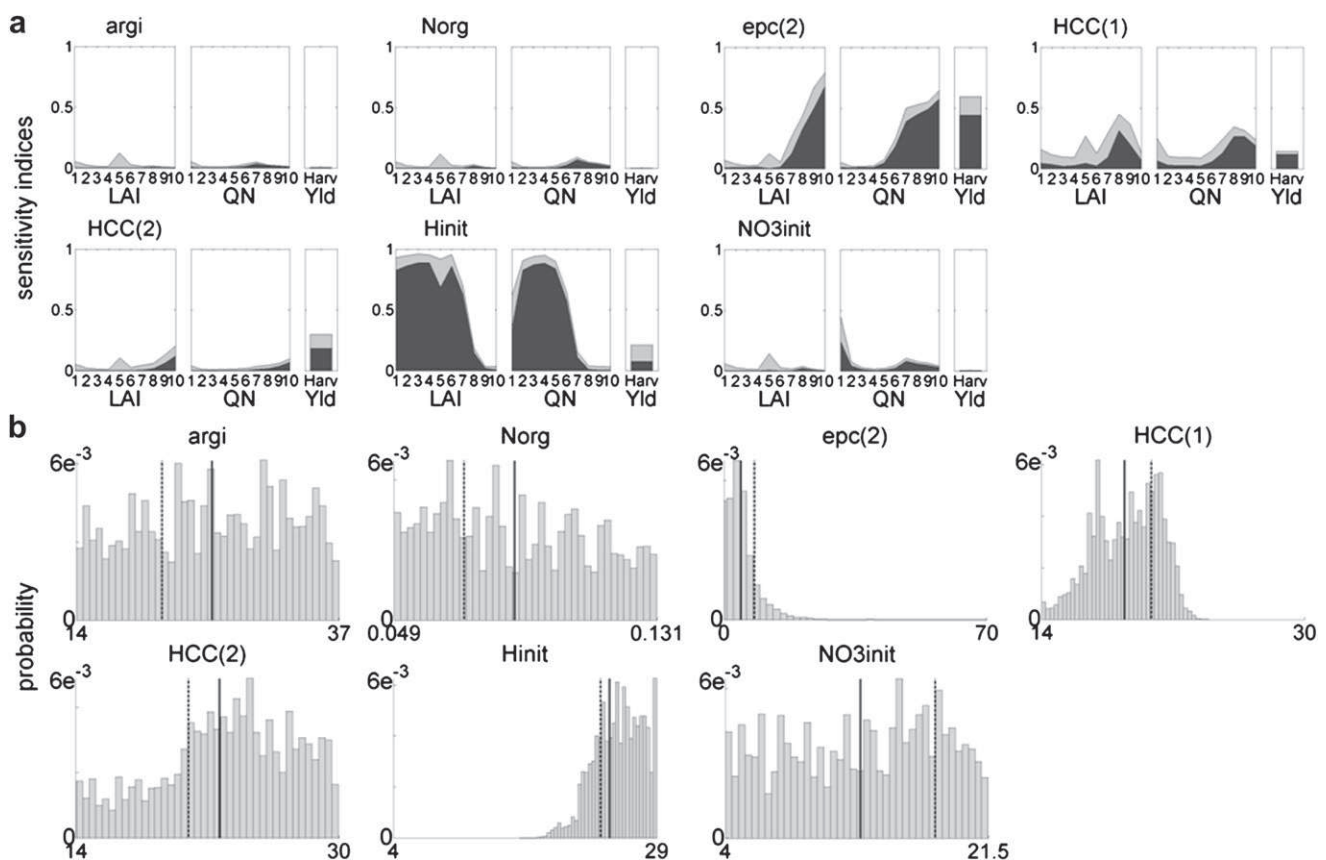
**Table 4**  
Criteria  $GMS_i$  and  $TGMS$  calculated for the *dry – beet* configuration and for the 3 sets of observations.

	$GMS_i$						$TGMS$	
	$argi$	$Norg$	$epc(2)$	$HCC(1)$	$HCC(2)$	$Hinit$	$NO3init$	Parameter set
Set 1	-0.01	-0.009	0.063	-0.008	-0.008	0.155	-0.009	0.174
Set 2	-0.006	-0.003	0.071	0.004	-0.005	0.147	-0.001	0.207
Set 3	-0.006	-0.003	0.073	0.005	-0.004	0.145	-0.001	0.209

contained in these output variables may be quite redundant for this parameter. It is confirmed by the fact that  $TGMS$  increases only by 18% from set #1 to set #2 and that the average correlation coefficient between  $LAI$  and  $QN$  is about 59% for this configuration.

Table 5 summarizes the results of the two criteria, for the 8 configurations with sugar beet as preceding crop and for the set #3 of output variables. Concerning these results, it can be said that:

- Only the parameters  $Hinit$  and  $epc(2)$  have significant values of  $GMS_i$ , as in the configuration examined previously.
- The effect of climate is significant for  $Hinit$ : the drier the climate the higher the criterion for  $Hinit$  ( $GMS_i$  decreases 0.145–0.052 from dry to wet climates for shallow soils and 0.136–0.063 for deep soils). It is also the case for  $epc(2)$  for shallow soils.
- The type of soil depth has a big effect on  $GMS_i$  for parameter  $epc(2)$ . This parameter is not linearly related to water stress on the entire range of values (the relationship is more of



**Fig. 1.** GSA results and parameter estimation results for the configuration *dry – beet*. a) Sensitivity profiles (first order: area in black and interaction: area in gray) of the output variables to the 7 parameters. On the x-axis, numbers 1–10 correspond to the 10 dates for  $LAI_t$  and  $QN_t$ , and  $Harv$  corresponds to harvest. b) Posterior distributions calculated for the 7 parameters with the set #3 of observations: one of the 50 replicates is represented. The dashed line corresponds to the true value of the parameter and the solid line corresponds to the mean of the distribution.

**Table 5**

Criteria  $GMS_i$  and  $TGMS$  for the 8 configurations having sugar beet as preceding crop and for the set #3 of output variables.  $GMS_i$  values higher than 0.03 are in bold.

	Climate	Soil depth	$GMS_i$							$TGMS$
			<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>	Parameter set
(1)	<i>dry</i>	–	–0.006	–0.003	<b>0.073</b>	0.005	–0.004	<b>0.145</b>	–0.001	0.209
(2)	<i>mdry</i>	–	0.002	0.011	<b>0.066</b>	–0.012	–0.003	<b>0.148</b>	–0.004	0.208
(3)	<i>mwet</i>	–	–0.001	0.012	<b>0.037</b>	–0.016	–0.012	<b>0.118</b>	–0.001	0.138
(4)	<i>wet</i>	–	–0.021	–0.005	<b>0.033</b>	–0.048	–0.029	<b>0.052</b>	–0.035	–0.054
(5)	<i>dry</i>	+	–0.005	–0.001	–0.008	0.019	–0.001	<b>0.136</b>	0.007	0.148
(6)	<i>mdry</i>	+	0.011	0.028	–0.004	–0.010	0.011	<b>0.128</b>	0.002	0.166
(7)	<i>mwet</i>	+	0.008	0.027	–0.015	0.017	–0.002	<b>0.118</b>	0.022	0.175
(8)	<i>wet</i>	+	–0.011	0.018	–0.034	–0.023	–0.035	<b>0.063</b>	–0.009	–0.030

a linear-plus-plateau type):  $LAI$ ,  $QN$  and  $Yld$  are sensitive to  $epc(2)$  for shallow soil (involving  $GMS_i$  between 0.033 and 0.073) whereas they are not for deeper soils (involving  $GMS_i$  close to 0 or negative).

- $TGMS$  mainly exhibits the same behavior as  $Hinit$  which is the most sensitive parameter. It increases from wet to dry climates for shallow soils: from –0.054 to 0.209; and for deeper soils: from –0.03 to 0.148.

### 3.2. Parameter estimation (GLUE)

Fig. 1b illustrates the results of the parameter estimation process described in Section 2.3 for the seven soil parameters, for the *dry – beet* configuration and the third observation set, applied to one of the 50 true soils defined in the synthetic experiments (see Section 2.5). The posterior distributions of the parameters exhibit different patterns: some are similar to the uniform prior distribution (*argi*, *Norg*, *NO3init*), some have a profoundly modified distribution with a marked mode (*Hinit* and *epc(2)*) and some have an intermediate pattern (*HCC(1)* and *HCC(2)*). These patterns will be referred to in the following as: “unmodified”, “modified” and “intermediate”. The parameter estimates (given by the mean of the posterior distributions) are close to the true values for the “modified” patterns, less close for the “intermediate” and quite dissimilar for the “unmodified”. The “modified” distributions indicate that the uncertainty of the corresponding parameters has been reduced with respect to the uncertainty associated with the prior information: observations have brought enough information to provide certainty on the estimates. On the contrary, the uncertainty of the “unmodified” parameters has only been slightly reduced. It is striking that these patterns and goodness of parameter estimates are closely related to the patterns of the GSA presented above (Fig. 1a): the higher the  $GMS_i$ , the more informative is the distribution (“modified” pattern) and the closer is the estimate to the true value. The “modified” patterns correspond to parameters whose variations induce large variance in the output variables, allowing a good maximization of the likelihood in the GLUE process, unlike the “unmodified” ones.

The results of the parameter estimation process in terms of  $RE_i$  and  $TRE$  are reported in Table 6 for the set #3 and the 8 configurations considered in the previous section (Table 5):

- $Hinit$  has the lowest value of the  $RE_i$  criteria for each configuration, meaning a large reduction of the error of estimate when using observations as compared to the error of the prior information.
- Conversely, *argi*, *Norg* and *NO3init* have generally high values of  $RE_i$ , which correspond to no reduction or even an increase of the error of estimate. The quality of estimation of *epc(2)*, *HCC(1)* and *HCC(2)* varies from medium to poor, depending on the configurations.

- The drier the climate, the better the estimation of  $Hinit$ . For example, the *dry – beet* configuration allows a greater reduction in the error on  $Hinit$  ( $RE_i = 0.428$ ) than the *wet – beet* configuration ( $RE_i = 0.727$ ). This is also the case for deep soils. Conversely, the estimation of the parameters involved in organic matter mineralization (*argi* and *Norg*) is poor but slightly improved in wet climates as compared to dry climates. However, the type of climate has little effect on the reduction of *epc(2)* estimate error.
- The type of soil depth has a big effect on the estimation of *epc(2)*: its estimate is better for shallow soils ( $RE_i$  from 0.749 to 0.694) than for deep soils ( $RE_i$  from 0.966 to 0.871). This is still in accordance with the sensitivity analysis results that show higher indices for shallow soils than for deep ones.
- The  $TRE$  criteria exhibit similar, but less pronounced, behavior to the  $RE_i$  of the best estimated parameter  $Hinit$ : they decrease slightly from wet to dry climates. The effect of soil depth on  $TRE$  is not significant.

Fig. 2 shows the correlations calculated on the posterior parameter values for each pair of parameters, for the *dry – beet* configuration and the observation set #3. High correlations are observed only between *HCC(1)* and  $Hinit$  (about 0.331 in average), between *HCC(2)* and *epc(2)* (about –0.516 in average) and between *HCC(1)* and *epc(2)* (about –0.368 in average). These results reveal that the pair (*HCC(1)*,  $Hinit$ ) acts in the STICS model as a quotient or a difference, whereas (*HCC(2)*, *epc(2)*) as (*HCC(1)*, *epc(2)*) acts as a product or a sum (see Section 2.3). Such significant correlations have been noted for configurations having a climate rather dry and only concern pairs of parameters which have good estimates: *epc(2)*, *HCC(1)*, *HCC(2)* and  $Hinit$ . These results suggest that these parameters can compensate to produce similar good simulation results with respect to observations. This can be due to model structure as well as to model error, data error and sparsity (Rodriguez-Fernandez et al., 2006). As a consequence, part of the uncertainty of the estimated parameters is due to these interactions.

The effect of the observation set on the  $TRE$  criterion is finally examined. Considering all the configurations, set #2 allows decreasing the value of  $TRE$  only about 3% compared to set #1 (0.898–0.873) while set #3 allows decreasing  $TRE$  about the same amount compared to set #2 (0.873–0.844). This is to relate with the average correlation coefficient between set #1 and set #2 which is about 61% while it is about 37% between set #2 and set #3.

### 3.3. The link between GSA and quality of parameter estimation

#### 3.3.1. At a single parameter level

In the previous sections we highlighted the analogy between the behavior of criteria based on GSA indices and the criteria expressing the quality of parameter estimation for some configurations and some sets of output variables or observations. We

**Table 6**Criteria  $RE_i$  and  $TRE$  for the 8 configurations having sugar beet as preceding crop and for the set #3 of output variables.  $RE_i$  values lower than 0.7 are in bold.

	Climate	Soil depth	$RE_i$						$TRE$	
			<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>	Parameter set
(1)	<i>dry</i>	–	1.003	1.018	0.724	0.854	0.844	<b>0.428</b>	0.85	0.817
(2)	<i>mdry</i>	–	1.007	0.887	0.749	0.824	0.79	<b>0.49</b>	0.946	0.812
(3)	<i>mwet</i>	–	1.003	0.921	<b>0.694</b>	0.743	0.922	<b>0.694</b>	0.876	0.836
(4)	<i>wet</i>	–	0.916	0.949	0.728	0.866	0.929	0.727	0.967	0.869
(5)	<i>dry</i>	+	0.979	1.014	0.871	0.779	0.708	<b>0.399</b>	0.856	0.801
(6)	<i>mdry</i>	+	0.958	0.917	0.966	0.906	0.913	<b>0.486</b>	0.982	0.875
(7)	<i>mwet</i>	+	0.931	0.915	0.962	0.811	0.886	<b>0.593</b>	0.899	0.857
(8)	<i>wet</i>	+	0.882	0.980	0.960	0.839	0.943	<b>0.697</b>	0.986	0.898

present here the results for the three sets of observations and the 16 soil, climatic and agronomic conditions.

Fig. 3 shows that a good link exists between  $GMS_i$  and  $RE_i$ : the relationship seems to be linear, the higher the  $GMS_i$  criterion, the lower the  $RE_i$  and the better the quality of estimation of the  $i$ th parameter. The results show clusters of parameters:  $h$  ( $Hinit$ ) at high  $GMS_i$  and low  $RE_i$  values,  $e(epc(2))$  at intermediate  $GMS_i$  and  $RE_i$  values, and the other parameters all grouped in the same cluster at low  $GMS_i$  and high  $RE_i$  values. Within the scattering around the relationship, the position of the parameter depends on the configuration and especially the soil depth.

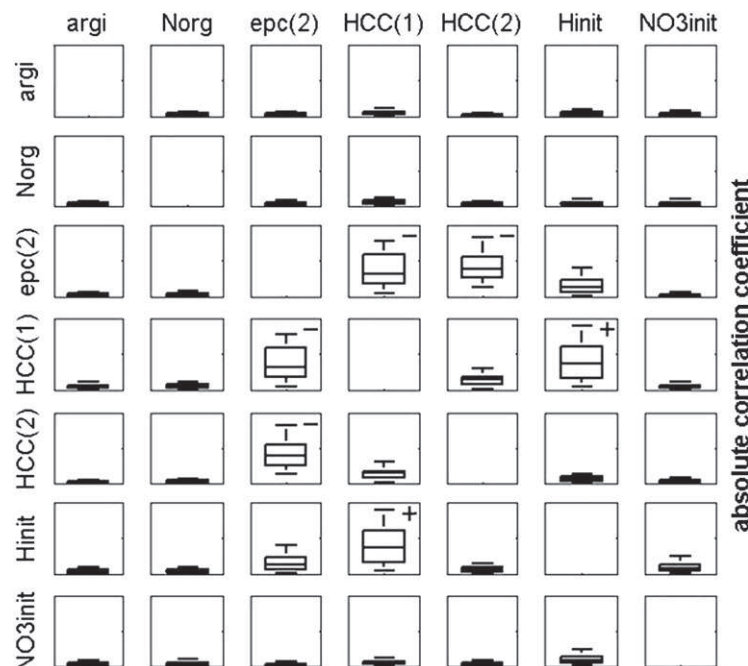
The case of parameters for which the observations contain enough information to estimate them precisely can be illustrated by the parameter  $Hinit$ . For example, for the *dry – beet* configuration and the observation set #1 (see Fig. 1a for LAI),  $Hinit$  has a big main effect ( $S_i = 0.582$  on average), leading to a high value of  $GMS_i = 0.155$ . In this case, the parameter  $Hinit$  has a low value of  $RE_i$  ( $RE_i = 0.499$ ) meaning a considerable improvement of its uncertainty through the parameter estimation process. In general, the results show that for high values of  $GMS_i$ , the reduction of the estimation error is large: a high  $GMS_i$  indicates a large improvement in parameter estimation. Small first-order indices combined with a large interaction are thought to induce problems of

identification. In our case, this applies to most of the parameters. For example, for the *wet – beet* configuration and the first observation set,  $HCC(1)$  has a small main effect ( $S_i = 0.089$  on average) and a large interaction (equal to 0.251 on average), leading to a negative value of  $GMS_i$  ( $GMS_i = -0.099$ ). In this case, its  $RE_i$  value is high ( $RE_i = 0.921$ ), meaning a poor improvement in parameter estimation. In general, the results show that for negative values of  $GMS_i$ , the reduction of the estimation error is small: a negative value of  $GMS_i$  reveals a bad quality of the parameter estimation.

Spearman's rank correlation analysis between  $GMS_i$  and  $RE_i$  allows us to quantify the quality of the relationship illustrated in Fig. 3. It is performed after discarding the parameters having a negative  $GMS_i$  (from two to five) which have always a poor quality of estimation and whose rank would still be high. The averaged Spearman's correlation coefficient is about 75.4%, which is a satisfactory value, means that the  $GMS_i$  criterion is effective for ranking the accessible parameters (for which the criterion is positive) with respect to their quality of estimates.

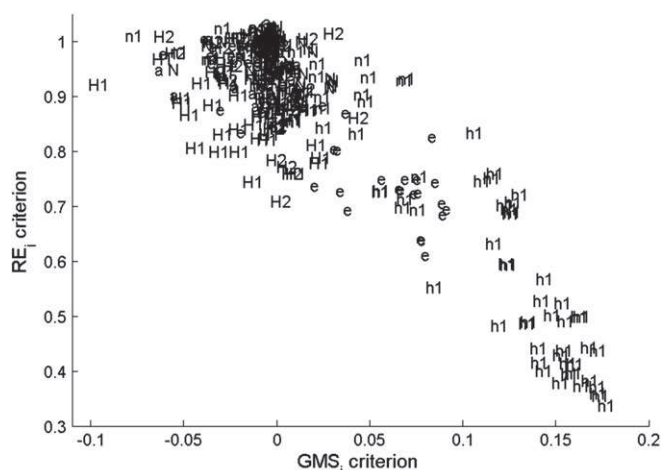
### 3.3.2. At the whole parameter set level

When considering the whole parameter set (Fig. 4a), there is a slight relationship between the  $TGMS$  criterion created with the GSA indices of the parameter set and the  $TRE$  criterion computed



**Fig. 2.** Boxplot of the absolute values of the correlation coefficients calculated on the posterior distributions of the 50 replicates, for the *dry – beet* configuration and the set #3. The sign of the 3 most correlated pairs of parameters is given in the symmetrical corresponding square.





**Fig. 3.** Scatter diagram of the criteria  $RE_i$  and  $GMS_i$  of the 7 soil parameters for the 3 sets of observations and the 16 configurations. Label *a* corresponds to *argi*, *N* to *Norg*, *e* to *epc(2)*, *H1* to *HCC(1)*, *H2* to *HCC(2)*, *h* to *Hinit* and *n* to *NO3init*.

from errors of estimate for this parameter set. The *TRE* criterion never reaches low values (the minimum value is about 0.8) even for high *TGMS* values (about 0.21), due to the relatively large number of parameters which are not easily retrievable. The global relationship is no longer linear and two clusters of configurations appear: one corresponds to shallow soils (in white symbols) and still exhibits a quite linear relationship between *TGMS* and *TRE*, the other corresponds to deep soils (in black symbols), for which the relationship is non-linear. According to the results presented Section 3.2, the first cluster leads to a slightly better estimate of the whole parameter set than the second cluster, because the estimation of parameter *epc(2)* is better for shallow soils than for deeper ones.

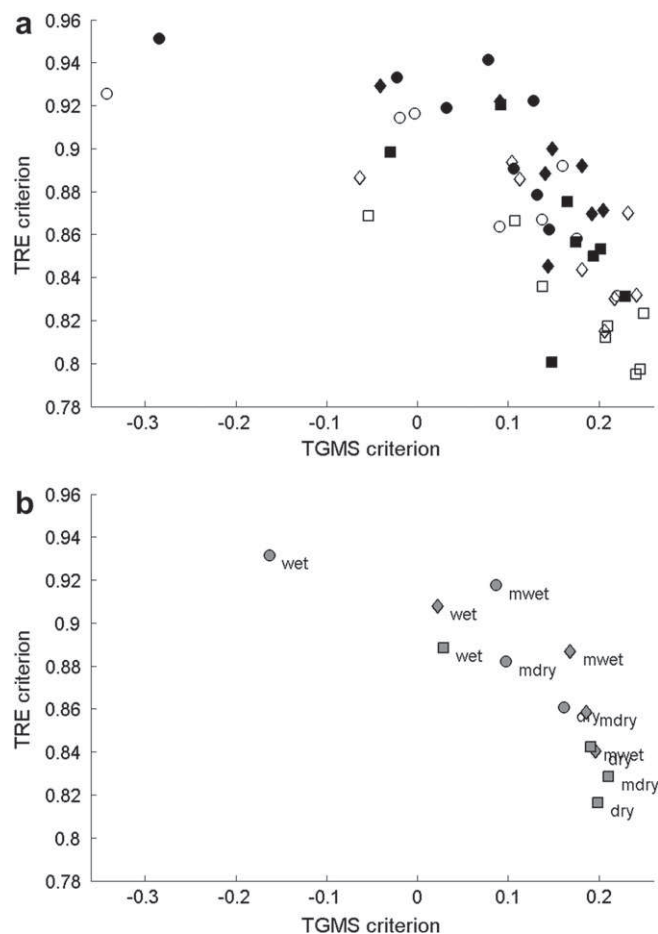
In Fig. 4b the values of *TRE* and *TGMS* have been averaged over the two soil depths and the two preceding crops. The relationship between *TRE* and *TGMS* appears stronger. The effect of climate is striking, according to results presented in the Sections 3.1 and 3.2. Configurations with a dry climate have the higher values of *TGMS* (between 0.16 and 0.2) and they correspond to the best quality of estimation of the parameter set (*TRE* between 0.81 and 0.86), unlike configurations with a wet climate (*TGMS* below 0.03 and *TRE* above 0.89). As it was seen before, the greater the number of observations considered in the estimation process, the lower is the *TRE*. As expected and as it was seen in Fig. 4b, *TGMS* often decreases when the number of observations increases. Although some of the observed variables are mutually correlated, they each improve the quality of the parameter set estimation.

Finally, the Spearman's correlation coefficients between *TGMS* and *TRE* were computed for each type of soil depth, preceding crop and observation set, in order to quantify the ranking of the four climates given by both *TGMS* and *TRE*. The averaged value is about 72%. The Spearman correlations were therefore computed for each soil depth, preceding crop and climate, in order to quantify the ranking of the three observation sets given by both *TGMS* and *TRE*. The averaged Spearman's correlation between *TGMS* and *TRE* is about 91%. Both values are very satisfactory and indicate that *TGMS* is effective to rank the climates and the observation sets with respect to their ability to estimate the soil parameters.

## 4. Discussion

### 4.1. Retrieval of soil parameters from observations on crops

The results of parameter estimation vary greatly according to the soil/climate/preceding crop configuration, the parameters, and



**Fig. 4.** a) Scatter diagram of the criteria *TRE* and *TGMS* at the whole parameter set level. a) For the 3 sets of observations and the 16 configurations. The symbol "○" corresponds to the set #1, "◆" to set #2 and "□" to set #3. A white symbol corresponds to a shallow soil and a black one to a deeper soil. b) For the 3 sets of observations and 4 types of climate (each of the 12 points is averaged over the 2 soil depth and 2 preceding crop configurations). The labels of the climate types are described in Table 2.

the set of observed variables. However, some general trends can be drawn. *Hinit* is the best estimated parameter because the observed variables are in general sensitive to this parameter (high first-order indices and poor interaction) and the drier the climate of the configuration the better is the estimate of *Hinit*. The parameters *argi* and *Norg* are the worst estimated parameters because the observed variables are in general less sensitive to those parameters (poor first-order indices and poor interaction) than to the others. The quality of *epc(2)* estimation depends on soil depth: for shallow soils its estimation is good because the observed variables are in general quite sensitive to this parameter, whereas for deep soils the estimate is poor because the observed variables in this case are not very sensitive to it. Considering the quality of estimation of the whole parameter set, the drier the climate the better is the quality of estimation, because the first-order indices are generally higher for dry climates than for wet ones. It was also observed that the greater the number of observations the better is the quality of parameter estimation. This improvement is quite small due to the correlations existing between the observed output variables *LAI*, *QN* and *Yld*. However, it must be noted that a single observation of *Yld* provides as much information for estimating the parameters as ten observations of *QN*. This can be explained by lower correlations between *Yld* and *LAI + QN* than between *LAI* and *QN*.

Concerning the parameters which are to be estimated, the question of interaction between them is essential. GSA and GLUE are complementary for explaining this point. Although GSA supplies information on the basic features of interaction structure, it does not allow a complete representation of such a structure (Ratto et al., 2001): the EFAST method does not allow computing each interaction term and computing these terms with the Sobol' method would require an enormous amount of simulations. The GLUE method completes the knowledge of the general interaction structure, allowing pair-wise interactions to be quantified through the computation of correlation coefficients between parameters. In particular, a high level of correlation between two parameters indicates possible compensation between them. This means that bringing new information on one of the two parameters may reduce the posterior uncertainty of the other. This may be done by measuring it and fixing it at the measured value, or improving its prior information, or adding new observations related to this parameter in the estimation process (observations on soil water content for example, as did Braga and Jones (2004)). In our case, if the parameter  $HCC(1)$  could be measured, the parameters  $Hinit$  and  $epc(2)$  would be better estimated than when estimated together with  $HCC(1)$ . Thus, interactions between these parameters would no longer exist and this may improve the use of the STICS crop model for appropriate management decision support. Some promising possibilities are offered by using electrical geophysical measurements (Samouelian et al., 2005).

#### 4.2. Link between GSA and quality of parameter estimates

In our results the link between the quality of parameter estimation and GSA results was illustrated through three types of behavior: high first-order indices are associated with good quality of estimation, low total indices are associated with bad quality of estimation, and high total indices combined with low first-order indices are associated with poor estimates because of interactions between parameters. Given the large number of output variables and dates considered in this application, the GSA indices had to be summarized to study the link between GSA and parameter estimation results. We proposed the  $GMS_i$  criterion accounting for both the sensitivity indices and the correlation between variables, and showed its good relationship with the criterion  $RE_i$  which measures the quality of estimation of parameter  $i$ . The criterion  $GMS_i$  proved to be effective for ranking the accessible parameters with respect to their quality of estimation. Particularly,  $GMS_i$  is able to provide information on which parameters can be estimated and which can be fixed as they do not deserve an accurate determination. The total criterion  $TGMS$  can be used to predict the ranking of the configurations with respect to their ability to retrieve the whole set of parameters, and in particular the ranking of the climates and the observation sets: it may be possible to predict which type of climate and observation set will lead to the better estimation of the whole parameter set.

These results are particularly interesting for screening the possibility of estimating parameters from a given set of available observations in a given agro-environmental context, and, following Kontoravdi et al. (2005), promote GSA as an excellent precursor to optimal experimental design.

## 5. Conclusion

This study shows on various synthetic experiments that a few soil parameters are accessible by inversion of the STICS crop model from observations of yield at harvest, leaf area index and nitrogen absorbed by the plant at various dates. However, the quality of their estimation largely depends on several factors, in particular the

climate of the observed year and the type of soil depth. In view of this, it would be useful to explore the potential of accumulating several years of observations, possibly on different crops, to maximize the effectiveness of soil parameters retrieval. Using additional observations on soil variables should also help in increasing the quality of parameter estimates that sometimes shows strong correlations between them or even retrieving parameters that are not accessible from the observations used in this study. One should also think about using other models giving access to the same or to other soil parameters from the same or from new observed variables. Nevertheless, if the value of estimating some key soil parameters lies in the information produced *per se*, it also lies in the consequent reduction in the uncertainty of the predictions by the STICS model of agro-environmental variables which are used in decision making. Further studies should focus on this aspect.

This study also shows that the quality of parameter estimation can be explained by the results of global sensitivity analysis (GSA). Suitable empirical criteria have been proposed to summarize the results of GSA which allow ranking (i) the parameters with respect to their quality of estimate and (ii) the configurations (particularly the climate and the observation set) with respect to the quality of estimation of the whole parameter set. These criteria are thus shown in our case to be useful tools for estimating the potential of given configurations of observations for retrieving soil parameter values. They may be used also for optimizing the type of observations to be acquired and the dates of acquisition. Other criteria could be proposed and other applicative studies could be useful to explore the link between GSA and parameter estimation in other cases. Finally, it would be helpful to conduct such a study on real data to assess the impact of model errors on both soil parameter retrieval and link between the proposed criteria.

## Acknowledgements

The financial support provided by CNES and Arvalis-Institut du Végétal is gratefully acknowledged. We also thank INRA and the anonymous reviewers for the helpful comments.

## References

- Batchelor, W.D., Basso, B., Paz, J.O., 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy* 18, 141–158.
- Beven, K., Binley, A., 1992. The future of distributed models – model calibration and uncertainty prediction. *Hydrological Processes* 6, 279–298.
- Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249, 11–29.
- Blackmore, S., Moore, M., 1999. Remedial correction of yield map data. *Precision Agriculture* 1, 53–56.
- 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., Ruget, F., Gate, P., Lorgeou, J., Nicoulaud, B., Tayot, X., Plenet, D., Jeuffroy, M.H., Bouthier, A., Ripoche, D., Mary, B., Juste, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie* 22, 69–92.
- Brun, R., Kuhni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter subsets. *Water Research* 36, 4113–4127.
- Campolongo, F., Saltelli, A., 1997. Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety* 57, 49–69.
- Chan, K., Tarantola, S., Saltelli, A., Sobol, I.M., 2000. Variance-based methods. In: Saltelli, A., Chan, K., Scott, E.M. (Eds.), *Sensitivity Analysis*. Wiley, New York.
- Ferreyra, R.A., Jones, J.W., Graham, W.D., 2006. Parameterizing spatial crop models with inverse modeling: sources of error and unexpected results. *Transactions of the ASABE* 49, 1547–1561.
- Flenet, F., Villon, P., Ruget, F.O., 2003. Methodology of adaptation of the STICS model to a new crop: spring linseed (*Linum usitatissimum*, L.). In: *STICS Workshop*, Camargue, France.
- Gabrielle, B., Roche, R., Angas, P., Cantero-Martinez, C., Cosentino, L., Mantione, M., Langensiepen, M., Henault, C., Lavielle, P., Nicoulaud, B., Gosse, G., 2002. A priori

- parameterisation of the CERES soil-crop models and tests against several European data sets. *Agronomie* 22, 119–132.
- Gomez-Delgado, M., Tarantola, S., 2006. Global sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science* 20, 449–466.
- Guérif, M., Beaudoin, N., Durr, C., Machet, J.M., Mary, B., Michot, D., Moulin, D., Nicoulaud, B., Richard, G., 2001. Designing a field experiment for assessing soil and crop spatial variability and defining site specific management strategies. In: *Proceedings of the Third European Conference on Precision Agriculture*, Montpellier, France.
- Guérif, M., Houllès, V., Makowski, D., Lauvernet, C., 2006. Data assimilation and parameter estimation for precision agriculture using the crop model STICS. In: Wallach, D., Makowski, D., Jones, J.W. (Eds.), *Working with Dynamic Crop Models*. Elsevier.
- Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment* 112, 186–202.
- Houllès, V., Mary, B., Guérif, M., Makowski, D., Juste, E., 2004. Evaluation of the crop model STICS to recommend nitrogen fertilization rates according to agro-environmental criteria. *Agronomie* 24, 1–9.
- Irmak, A., Jones, J.W., Batchelor, W.D., Paz, J.O., 2001. Estimating spatially variable soil properties for application of crop models in precision farming. *Transactions of the ASAE* 44, 1343–1353.
- Jongschaap, R.E.E., 2007. Sensitivity of a crop growth simulation model to variation in LAI and canopy nitrogen used for run-time calibration. *Ecological Modelling* 200, 89–98.
- Kanso, A., Chebbo, G., Tassin, B., 2004. Application of MCMC-GSA model calibration method to urban runoff quality modeling. In: *Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004)*, Santa Fe, NM.
- Kontoravdi, C., Asprey, S.P., Pistikopoulos, E.N., Mantalaris, A., 2005. Application of global sensitivity analysis to determine goals for design of experiments: an example study on antibody-producing cell cultures. *Biotechnology Progress* 21, 1128–1135.
- Lagacherie, P., Baret, F., Feret, J.B., Netto, J.M., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment* 112, 825–835.
- Launay, M., Guérif, M., 2003. Ability for a model to predict crop production variability at the regional scale: an evaluation for sugar beet. *Agronomie* 23, 135–146.
- Launay, M., Guérif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, Ecosystems & Environment* 111, 321–339.
- Machet, J.M., Couturier, A., Beaudoin, N., 2007. Cartographie du rendement du blé et des caractéristiques qualitatives des grains. In: Guérif, M., King, D. (Eds.), *Agriculture de Précision*. Quae, Versailles.
- Makowski, D., Naud, C., Jeuffroy, M.H., Barbottin, A., Monod, H., 2006. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. *Reliability Engineering & System Safety* 91, 1142–1147.
- Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie* 22, 191–203.
- Manache, G., Melching, C.S., 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. *Environmental Modelling & Software* 23, 549–562.
- Moulin, S., Zurita, R.M., Guérif, M., 2007. Estimation de variables biophysiques du couvert par ajustement de modèles de transfert radiatif sur des réflectances. In: Guérif, M., King, D. (Eds.), *Agriculture de Précision*. Quae, Versailles.
- Pathak, T.B., Fraisse, C.W., Jones, J.W., Messina, C.D., Hoogenboom, G., 2007. Use of global sensitivity analysis for CROPGRO cotton model development. *Transactions of the ASABE* 50, 2295–2302.
- Pierce, F.J., Nowak, P., Roberts, P.C., 1999. Aspects of precision agriculture. *Advances in Agronomy* 67, 1–85.
- Post, J., Hattermann, F.F., Krysanova, V., Suckow, F., 2008. Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. *Environmental Modelling & Software* 23, 125–138.
- Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications* 136, 212–224.
- Ratto, M., Young, P.C., Romanowicz, R., Pappenberger, F., Saltelli, A., Pagano, A., 2007. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences* 11, 1249–1266.
- Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems* 83, 248–265.
- Ruget, F., Brisson, N., Delecolle, R., Faivre, R., 2002. Sensitivity analysis of a crop simulation model, STICS, in order to choose the main parameters to be estimated. *Agronomie* 22, 133–158.
- Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis* 26, 445–460.
- Saltelli, A., Chan, K., Scott, E.M., 2000a. *Sensitivity Analysis*. John Wiley and Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15, 377–395.
- Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41, 39–56.
- Samouelian, A., Cousin, I., Tabbagh, A., Bruand, A., Richard, G., 2005. Electrical resistivity survey in soil science: a review. *Soil & Tillage Research* 83, 173–193.
- Spearman, C., 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15, 72–101.
- Timlin, D., Pachepsky, Y., Walthall, C., Loebel, S., 2001. The use of a water budget model and yield maps to characterize water availability in a landscape. *Soil & Tillage Research* 58, 219–231.
- Tremblay, M., Wallach, D., 2004. Comparison of parameter estimation methods for crop models. *Agronomie* 24, 351–365.
- Weiss, M., Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment* 70, 293–306.
- Ziehn, T., Tomlin, A.S., 2009. GUI-HDMR – a software tool for global sensitivity analysis of complex model. *Environmental Modelling & Software* 24, 775–785.

## Chapitre 5. Amélioration de la qualité de prédiction des variables d'intérêt à partir de l'estimation des paramètres du sol

**Article 3 : “The estimation of soil properties using observations on crop biophysical variables and the crop model STICS improves the predictions of agro-environmental variables”. Soumis à *European Journal of Agronomy*..**

### 5.1. Objectif

Nous avons vu dans le chapitre précédent que la quantité d'information apportée par le jeu d'observations, telle qu'on peut la mesurer par analyse de sensibilité détermine la qualité d'estimation des paramètres sol. Nous proposons ici d'étudier comment elle détermine également la qualité des prédictions des variables d'intérêt agroenvironnemental. Les paramètres considérés sont ceux étudiés au chapitre précédent. Les variables retenues, parmi celles étudiées au Chapitre 3, sont les variables agroenvironnementales déterminées à la récolte : rendement *Yld*, qualité de la production *Prot* (teneur en protéine pour le blé) et quantité d'azote minéral dans le sol *Nit*. La bonne prédiction de ces variables permet en effet de raisonner de manière efficace les choix de l'agriculteur vis à vis de son travail technique, en conciliant intérêts agronomique et environnemental. Par exemple, Houlès et al. (2004) ont montré que des cartes de préconisation de doses d'engrais azoté pouvaient être élaborées à partir de prédictions spatialisées de ces variables et de l'optimisation d'un critère agroenvironnemental. Nous étendrons dans cette étude la diversité des jeux d'observations en considérant, comme dans le Chapitre 3, des observations réalisées non seulement sur blé d'hiver mais aussi sur betterave sucrière, culture qui permet d'exprimer davantage les propriétés des sols.

## 5.2. Méthodes

Nous avons vu au Chapitre 1.3.2, à travers des résultats bibliographiques, que l'amélioration des prédictions à partir de l'estimation repose sur le fait que les variables à prédire et les variables observables ont des sensibilités similaires aux paramètres estimés. Nous essaierons donc d'expliquer l'amélioration des prédictions en fonction de leurs sensibilités aux paramètres du sol et de déterminer les jeux d'observations qui, grâce à leur quantité d'information efficace, permettent de réduire les incertitudes sur les prédictions. Nous avons défini cette réduction comme l'amélioration de la qualité des prédictions issues des valeurs estimées des paramètres relativement à celle des prédictions issues de l'information a priori (sa valeur moyenne). Comme pour le chapitre précédent, l'estimation des paramètres du sol par inversion sera effectuée par la méthode Importance Sampling et la même information a priori sur les paramètres sera considérée (déduite de mesures expérimentales sur le site de Chambry). Dans une première partie, ce travail est réalisé avec des observations synthétiques du couvert végétal, afin d'explorer toutes les configurations d'observations éventuelles. Ces différentes configurations seront ici composées de :

- deux cultures annuelles différentes (blé d'hiver et betterave à sucre),
- quatre climats contrastés caractérisés comme sec, moyen sec, moyen humide et humide (les même que ceux du Chapitre 4),
- deux gammes de profondeurs de sol (de 30 à 100 cm pour les sols peu profonds et de 80 à 160 cm pour les sols profonds),
- trois types/tailles de jeux d'observations : des observations composées de  $LAI$  seulement, de  $LAI+QN$ , et de  $LAI+QN$ +rendement.

Dans une seconde partie, de vraies observations réalisées sur le bassin versant de Bruyères (voir Chapitre 2.4) seront utilisées afin de valider, de manière réaliste, les résultats obtenus avec les observations synthétiques. Sachant que les sites de Chambry et de Bruyères sont proches et qu'on y retrouve des formations pédologiques voisines (voir Chapitre 2.4), l'information a priori considérée dans l'application aux observations de Bruyères est la même que celle utilisée pour les observations synthétiques (déduite de Chambry).



### 5.3. Résultats

#### *Précision et amélioration de l'estimation des paramètres*

Nous avons vu dans le Chapitre 4.3 les résultats de l'estimation des paramètres, en termes de précision et d'amélioration, lorsque des observations synthétiques sur couvert végétal de blé étaient considérées. Nous allons à présent présenter ces deux quantités dans le cas où des observations synthétiques sur couvert végétal de betterave sont considérées (résultats non présentés dans l'article). Que ce soit en termes de précision (voir le Tableau 5-1) ou en termes d'amélioration (voir le Tableau 5-2), nous voyons que les observations synthétiques de betterave permettent de diminuer significativement le critère, par rapport à ceux obtenus dans le Chapitre 4.3. Les observations de betterave sont donc plus efficaces pour estimer les paramètres du sol. Pour preuve, les paramètres  $HCC(1)$ ,  $HCC(2)$  et  $epc(2)$  ont un critère qui diminue énormément grâce à ces observations (sensibilité à ces paramètres plus importante). Par exemple, en forte profondeur de sol, le critère  $RE$  du paramètre  $HCC(1)$  passe de 0.81 (observations de blé) à 0.22 (observations de betterave). Nous noterons que la condition initiale  $Hinit$  est le seul paramètre moins bien estimé qu'avec des observations de blé (sensibilité à ce paramètre moins importante).

	Condition	<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>
<b>RRMSE (%)</b>	–	24.3	25	30.5	<b>7.8</b>	<b>15.1</b>	39	33.7
	+	24.7	25.2	<b>18.8</b>	<b>4.5</b>	<b>12.6</b>	35.1	32.1
	sec	24.2	26.1	25	<b>6.2</b>	<b>13.6</b>	33.3	31.4
	humide	24.8	24.1	24.3	6.1	14.1	40.8	34.4
	(humide +)	(24.9)	(24.6)	<b>(18.5)</b>	<b>(4.5)</b>	<b>(11.6)</b>	(39.2)	(32.4)

**Tableau 5-1.** Précision moyenne d'estimation des paramètres du sol ( $RRMSE$ ) avec des observations synthétiques de betterave, sachant la condition agropédoclimatique : faible profondeur de sol (–), forte profondeur de sol (+), climat sec (sec), climat humide (humide) ou climat humide en forte profondeur de sol (humide +). En gras les  $RRMSE$  inférieurs à 20%.

	Condition	<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>
RE	–	0.96	0.86	<b>0.49</b>	<b>0.31</b>	<b>0.63</b>	0.89	0.98
	+	0.90	0.86	0.86	<b>0.22</b>	<b>0.65</b>	0.83	0.96
	sec	0.96	0.97	<b>0.68</b>	<b>0.24</b>	<b>0.66</b>	<b>0.74</b>	0.94
	humide	0.90	<b>0.75</b>	<b>0.68</b>	<b>0.29</b>	<b>0.62</b>	0.98	1
	(humide +)	(0.88)	<b>(0.72)</b>	(0.85)	<b>(0.26)</b>	<b>(0.57)</b>	(0.95)	(1)

**Tableau 5-2.** Amélioration moyenne d'estimation des paramètres du sol (critère *RE*) avec des observations synthétiques de betterave, sachant la condition agropédoclimatique : faible profondeur de sol (–), forte profondeur de sol (+), climat sec (sec), climat humide (humide) ou climat humide en forte profondeur de sol (humide +). En gras les *RE* inférieurs à 80%.

La dernière ligne de chacun des deux tableaux précédents, concernant des conditions climatiques humides et une forte profondeur de sol, permet de comparer les résultats de l'estimation des paramètres issus d'observations synthétiques avec ceux issus d'observations réelles sur le bassin de Bruyères. Ces résultats, présentés dans le Tableau 5-3, montrent que les paramètres *HCC(1)* et *epc(2)* sont effectivement estimables avec une bonne précision lorsque des observations sur couvert de betterave sont considérées (*RRMSE* respectivement égal à 7.6 et 18.6%). Nous voyons également que l'estimation de *HCC(1)* est fortement améliorée par l'inversion avec des observations réelles de betterave (*RE* égal à 0.54).

	<i>argi</i>	<i>Norg</i>	<i>epc(2)</i>	<i>HCC(1)</i>	<i>HCC(2)</i>	<i>Hinit</i>	<i>NO3init</i>
RRMSE %	<b>17.9</b>	22.5	<b>18.6</b>	<b>7.6</b>	26.9	55	87.6
RE	0.8	<b>0.78</b>	<b>0.62</b>	<b>0.54</b>	0.95	1.37	1.41

**Tableau 5-3.** Précision (*RRMSE*) et amélioration (critère *RE*) moyenne d'estimation des paramètres du sol avec des observations réelles de betterave, pour la condition agropédoclimatique : climat humide en forte profondeur de sol (humide +). En gras les *RRMSE* inférieurs à 20% et les *RE* inférieurs à 80%.

Dans le Chapitre 4, nous avons vu que la configuration d'observation avait un effet significatif sur l'amélioration de l'estimation des paramètres : un climat sec et une faible profondeur de sol permettent d'obtenir les meilleures améliorations des paramètres liés à l'état hydrique du sol (*epc(2)*, *HCC(1)*, *HCC(2)* et *Hinit*), les autres étant difficilement estimables. Nous voyons à présent, avec ces nouveaux résultats,



que les observations sur couvert végétal de betterave permettent d'améliorer encore plus l'estimation de ces paramètres, mis à part pour la condition initiale *Hinit*. Dans la partie suivante, nous allons montrer comment l'estimation des paramètres, sous différentes configurations d'observations, peut améliorer les prédictions.

### *Amélioration des prédictions*

Les résultats de l'article montrent qu'il est possible d'améliorer la prédiction des variables agronomiques, telles que le rendement et la qualité du rendement, mais qu'il est malheureusement plus difficile d'améliorer celle des variables environnementales, telles que la quantité d'azote encore présent dans le sol à la récolte. Dans le cas des observations synthétiques, les prédictions issues des valeurs moyennes de l'information a priori peuvent être améliorées par l'estimation jusqu'à 61.4% pour le rendement et jusqu'à 58.9% pour la qualité, alors que la teneur en azote du sol ne peut être améliorée que jusqu'à 19.6%. Lorsqu'une amélioration est possible, les résultats montrent que cela vient principalement du fait que les variables à prédire sont sensibles aux mêmes paramètres que le sont les variables observables. De plus, l'article montre qu'il existe un certain degré dans les améliorations possibles, dans le sens où les jeux d'observations acquis dans différentes configurations contiennent des quantités d'information variables permettant d'améliorer l'estimation, et par conséquent la prédiction, de manière plus ou moins significative. Par exemple, nous avons vu que les conditions climatiques dans lesquelles les observations ont été recueillies ont un effet significatif sur l'estimation et par conséquent sur la prédiction, dans le sens où les conditions sèches sont plus efficaces que les conditions humides. Dans le cas des observations synthétiques, les conditions sèches améliorent les prédictions – relativement aux conditions humides – d'environ 25% pour le rendement et la qualité et d'environ 5% seulement pour la teneur en azote du sol. Le type de profondeur de sol a lui aussi un effet important dans le sens où les résultats d'estimation et de prédiction sont de meilleure qualité sur un sol peu profond que sur un sol profond. Toujours dans le cas des observations synthétiques, la prédiction du rendement est d'environ 0.4 fois meilleure pour le rendement, 0.5 fois meilleure pour qualité et 0.1 fois meilleur pour la teneur en azote du sol, lorsqu'un sol peu profond est considéré au lieu d'un sol profond. Pour finir, l'utilisation de jeux d'observations recueillis sur deux différentes

cultures (blé et betterave) nous a permis de mettre en évidence l'efficacité des observations sur un couvert de betterave dans le sens où la quantité d'information contenue dans les observations de betterave permet de mieux estimer et de mieux prédire que celle contenue dans les observations de blé. Dans le cas des observations synthétiques, l'observation du couvert de betterave améliore les prédictions – relativement aux observations du blé – d'environ 25% pour le rendement, 10% pour la qualité et 5% pour la teneur en azote du sol. Dans l'application aux observations réelles sur le site de Bruyères, nous voyons qu'il est en effet possible d'améliorer significativement la prédiction du rendement et de la qualité par l'estimation des paramètres du sol : jusqu'à 25% pour le blé ; alors que cela est assez difficile pour la teneur en azote du sol. Pour finir, l'effet de la culture observée révèle son réel potentiel : l'observation de la betterave permet d'améliorer d'environ 22% la prédiction du rendement et de la qualité du blé, par rapport à l'observation du blé lui-même. Les résultats de cet article permettent ainsi de faire un diagnostic assez large des différentes possibilités que l'on a, à partir d'éventuels jeux d'observations, pour estimer les paramètres du sol et pour améliorer les prédictions de variables d'intérêt. D'un point de vue pratique, les résultats de l'article peuvent par exemple permettre d'optimiser le recueil du jeu d'observations, en ne recueillant que celles qui mènent à une estimation et à une prédiction de qualité.

**5.4. Article 3 : “The estimation of soil properties using observations on crop biophysical variables and the crop model STICS improves the predictions of agro-environmental variables”**

Article soumis à European Journal of Agronomy.

## **The estimation of soil properties using observations on crop biophysical variables and the crop model STICS improves the predictions of agro-environmental variables.**

Hubert Varella\*<sup>1</sup>, Martine Guérif<sup>1</sup>, Samuel Buis<sup>1</sup> and Nicolas Beaudoin<sup>2</sup>

<sup>1</sup> INRA, UMR 1114 INRA-UAPV EMMAH, Domaine Saint Paul, Site Agroparc, F84914 Avignon Cedex 9, France

<sup>2</sup> INRA, Unité d'Agronomie LRM, Rue Fernand Christ, 02007 Laon Cedex, France

\* Corresponding author: Hubert Varella

### **Abstract**

The behavior of crops can be predicted when all the parameters of the crop model are well known. Among them, the soil parameters are especially difficult to determine at each location in the study area and they affect the quality of the predictions. Using data observed on crop status in the model is one way of estimating the soil parameters. Nevertheless, the results of parameter estimation depend on the observation set and the results of the predictions are thus also affected. The goal of this study is to assess the value of soil parameter estimation and prediction quality for various observation sets. To achieve it, several observation sets acquired in different conditions (winter wheat and sugar beet crops grown in different weather and cropping conditions) were used to estimate the values of the soil parameters which were then reused in the model to predict the variables of interest. Parameters were estimated using the Importance Sampling method (based on the Bayes theory). The quality of parameter estimation is then calculated (a function of RMSE) as well as the quality of predictions (a function of RMSEP). We worked first with synthetic data and then on real data. The results show that parameters related to soil water content are well estimated and the prediction of the variables of interest can thus be greatly improved. Moreover, parameter estimation and variable prediction are better when the soil is shallow and when the observations are made during dry weather and on sugar beet. The results in this paper can be used to assess the effect of the observation set on the quality of parameter estimation and variable prediction.

Keywords: Bayesian estimation, Importance Sampling, soil parameters, prediction of agro environmental variable, dynamic crop model STICS.

## 1. Introduction

Crop models are useful tools for simulating or predicting the behavior of crops subjected to different cultural practices. Such predictions are made either on a landscape or a field scale, and are widely used in a lot of agro-environmental work such as crop monitoring, yield prediction or decision making (Gabrielle et al., 2002; Houlès et al., 2004). Crop models can include more than 200 parameters whose values must be estimated from past experiments in order to predict crop behavior (Tremblay and Wallach, 2004; Makowski et al., 2006b). For spatial application a knowledge of soil parameters is even more crucial because they are responsible for much of the variability of the crop model output of variables of interest (Launay and Guérif, 2003). These parameters may be estimated from different techniques: either by soil analysis at different points in the study area, from a soil map and the application of soil transfer functions (Reynolds et al., 2000; Murphy et al., 2003), from remote sensing images (Lagacherie et al., 2008) or by using electrical resistivity measurements (Golovko and Pozdnyakov, 2007). The first method is difficult because of practical limitations, as well as time and financial constraints. Detailed soil maps suited to the scale of precision agriculture or even to that of a catchment are not usually available (King et al., 1994), while the use of remote sensing images or electrical resistivity is still hampered by a lack of robust interpretation of the signal (Lagacherie et al., 2008). Moreover, these techniques do not provide the values of all the soil parameters required for a complex crop model. Fortunately, techniques derived from remote sensing images (Weiss and Baret, 1999) or yield monitoring (Blackmore and Moore, 1999) can provide observations on crop state and thus make it possible to estimate soil parameters through the inversion of crop models.

Studies on the inversion of crop models for accessing soil parameters show that observations on the soil-crop system allow soil parameters to be estimated, and the estimates can reduce the uncertainties associated with the prediction of soil crop variables. Such observations can be made in one or several seasons of different

crops. For example, yield maps can be used (Ferreyra et al., 2006) for performing the inversion as well as data derived from remote sensing images (Guérif et al., 2006) or soil water content data (Calmon et al., 1999a), but these studies consider only one type of data from a single crop season. In other studies, some authors have estimated soil parameters by using data from more than one crop season. For example, yield maps made over two crop seasons can be used to estimate parameters (Irmak et al., 2001; Timlin et al., 2001), but in this case also only one type of data was used and not other information provided by observing different crops. Some studies took into account different types of data (Braga and Jones, 2004; Varella et al., 2008). Braga used soil water content data and yield maps for estimating the parameters and compared the effect of both types of data on the quality of parameter estimation and on the prediction of yield and soil water content. It has been proved in this study that estimating soil parameters from soil water content measurements led to better parameter estimates and predictions. Varella estimated the soil parameters using two types of data (derived from remote sensing and yield) and showed that the success of the parameter estimation process depends on the weather and the set of observations available. The sensitivity of the observable variables on the soil parameters is shown to be linked to the quality of the estimate. The weather is shown to be an important factor for the outputs to be sensitive to the soil parameters and especially to those related to the soil water content: the drier the weather, the more sensitive are the outputs to these parameters and the better are their estimates. Finally, too few studies compare the effect of the available observation set in different soil and weather conditions on parameter estimation and on prediction: Braga shows that observations on soil water content are better for estimating and predicting than yield maps, but they are very costly and require a lot of experimental work; Varella (2008) shows that dry weather gave better estimates but he only considered synthetic data.

What we propose in this paper is to evaluate the ability of various available observation sets (of different sizes and containing different types of data) and to compare the effect of the number of observed crop seasons (one or two), the type of the observed crop (winter wheat or sugar beet) and cropping conditions (weather and soil depth) on the accuracy of the estimated values and thus on the prediction of the output variables of interest. The crop model used is the dynamic model STICS

(Brisson et al., 2002) and the data consist of crop biophysical variables such as leaf area index (*LAI*), absorbed nitrogen and yield, which are currently available on a farm scale from remote sensing and yield monitors used in precision agriculture. The study is first conducted on a synthetic database designed for generating such variable conditions and as many observations as desired: several observation sets of varying size and several types of observed output variables are then considered on two different crops in different weather and cropping conditions. The synthetic study responds to all the objectives. Using the estimated values of the soil parameters, the prediction performance of agro-environmental variables is then analyzed and compared to that obtained by using the prior information on soil parameter values. Next, the parameter estimation and prediction performance are evaluated on a real experimental database. The results are then compared with those obtained with the synthetic database.

## **2. Materials and methods**

### **2.1 The crop model, output variables and soil parameters considered**

#### **2.1.1 The crop model STICS**

The STICS model (Brisson et al., 2002) is a one-dimensional dynamic crop model that simulates water, carbon and nitrogen dynamics in the soil crop system with a daily time step, as well as the behavior of many crops (wheat, sugar beet, corn, peas etc.). It distinguishes several compartments of the crop canopy and segments the soil profile into (at most) 5 layers. It considers the effects of water and nitrogen stress on plant growth and grain yield. The crop is essentially characterized by its above-ground biomass carbon and nitrogen, and leaf area index. The main outputs are agronomic variables (yield, grain protein content) as well as environmental variables (water and nitrate leaching). The STICS model is widely used in a lot of agro-environmental contexts (Ruget et al., 2002; Houlès et al., 2004; Beaudoin et al., 2008).

The model includes more than 200 parameters arranged in three main groups: those related to the characteristics of the plant or to the genotype, those related to

the soil and those describing the cropping techniques. In our case, we have used the wheat and sugar beet versions of the model for simulating these crops. The parameters related to the plant genotype were determined from literature, from experiments conducted on specific processes included in the model (e.g. critical nitrogen dilution curve) and from calibrations based on a large experimental database (Flenet et al., 2003; Hadria et al., 2007). The genotype parameter values of both crops used in this study were those given with the model ([http://www.avignon.inra.fr/agroclim\\_stics](http://www.avignon.inra.fr/agroclim_stics)) and some of them for the wheat crop were previously determined by using a large experimental database. The values of the soil and cultural technique parameters depend on the simulated case. The technique parameters are generally easily established as they are based on farmers' decisions. The soil parameters, in this case, will be estimated for each plot from observations of crops over one or several years.

### 2.1.2 Output variables considered

In this study, we focus on two types of STICS output variables. The first are observations that can be made on wheat and sugar beet canopies by automated methods. They consist of:

- the leaf area index ( $LAI_t$ ) and the nitrogen absorbed by the plant ( $QN_t$ ) at various dates  $t$  during the crop season - derived from remote sensing image inversion (Weiss and Baret, 1999; Houborg and Boegh, 2008),
- the yield at harvest ( $Yld$ ) as potentially provided by yield monitoring systems.

These output variables, are called "observable variables".

Secondly, a main objective of this study, apart from the estimation of soil parameter values, is the prediction of some output variables of interest, and its improvement as compared to those obtained without precise soil parameter values. They consist of:

- yield at harvest ( $Yld$ ),
- protein in the grain at harvest ( $Prot$ ) (only computed for the wheat crop),
- nitrogen contained in the soil at harvest ( $Nit$ ).

Yield, grain protein content and nitrogen balance in the soil at harvest are of particular interest for decision making, especially for monitoring nitrogen fertilization



(Houlès et al., 2004). Nitrogen absorbed by the plant is also important to analyze the health and growth of the plant during the crop's growing season (Baret et al., 2006).

### 2.1.3 The selection of soil parameters to estimate

The complete STICS model contains about 60 soil parameters. For our estimation purposes, it was essential to reduce their number.

First we chose the simplest options for simulating the soil system by ignoring the transfer of water and solutes within soil macropores, as well as capillary rise and nitrification, enabling us to omit the parameters associated with these processes. This choice defines -and limits- the domain of validity of the model considered and hence of the results. It is consistent with the soils contained in the databases we use in this study. In situations where these processes need to be accounted for, new parameters might need to be added to deal with these processes unless an increase in the errors is acceptable. Nevertheless, we have verified on our calibration database that the results found by not considering these processes are very close to those found by considering them: it shows the realism of the assumptions.

A second simplification was to consider the soil as two horizontal layers, each of a given thickness. From the observation of the tillage practices in the region around our reference site, the thickness of the first layer was set at 0.30 m. Based on the measurements made on this precision agriculture experimental site in northern France near Laon, Picardie (Chambry 49.35°N, 3.37°E ) (Guérif et al., 2001), we added relations, specific to our conditions, between the initial contents of water  $H_{init}$  and mineral nitrogen  $NO3_{init}$  of the two soil layers:

$$\begin{aligned} H_{init} &= H_{init}_{first\_layer} = H_{init}_{second\_layer} \\ NO3_{init} &= NO3_{init}_{first\_layer} = \frac{2}{3} NO3_{init}_{second\_layer} \end{aligned} \tag{1}$$

Finally, we performed a global sensitivity analysis on the 13 resulting soil parameters (Varela et al., 2008). It allowed us to fix those whose effect on the observed variables was negligible: for each parameter we computed the values of its effects on all the observed variables, and dropped the parameters for which all these values were less than 10% of the total effects generated by the 13 parameters. We thus restricted the study to 7 parameters.

As shown in Table 1, the 7 soil parameters considered characterize both water and nitrogen processes. They refer to permanent characteristics and initial conditions. Among the permanent characteristics, clay and organic nitrogen content of the top layer are involved mainly in organic matter decomposition processes and the soil nitrogen cycle. Water content at field capacity of both layers affects the water (and nitrogen) movements and storage in the soil reservoir. Finally, the thickness of the second layer defines the volume of the reservoir. The initial conditions correspond to the water and nitrogen content, *Hinit* and *NO3init*, at the beginning of the simulation, in this case the sowing date.

Parameter	Definition	Range	Unit
argi	Clay content of the 1 <sup>st</sup> layer	14-37 (16-25)	%
Norg	Organic nitrogen content of the 1 <sup>st</sup> layer	0.049-0.131 (0.1-0.13)	%
epc(2)	Thickness of the 2 <sup>nd</sup> layer	0-70 or 50-130* (90-120)	cm
HCC(1)	Water content at field capacity (1 <sup>st</sup> layer)	14-30 (22.5-29.5)	g g <sup>-1</sup>
HCC(2)	Water content at field capacity (2 <sup>nd</sup> layer)	14-30 (20-25.5)	g g <sup>-1</sup>
Hinit	Initial water content (both layers)	4-29 (6-25)	% of weight
NO3init	Initial mineral nitrogen content (1 <sup>st</sup> layer)	4-86 (6-50)	kg N ha <sup>-1</sup>

\* the first range is for a shallow soil and the second is for a deep soil

**Table 1.** The 7 soil parameters and their ranges of variation considered for synthetic and real experiments. The ranges of variation of the measured values on the Bruyères database are in brackets.

## 2.2 Data and numerical experiments

The objectives are to compare the effect of: (i) the size and type of observation set (ii) the number of observed crop seasons, (iii) the type of the observed crop: winter wheat or sugar beet and (iv) the given cropping conditions (weather and soil depth) on the performance of soil parameter estimation and prediction of variables of interest.

## 2.2.1 Synthetic data and experiments

### *Data and numerical experiments for parameter estimation*

A synthetic database was constituted by first drawing virtual soils and then calculating with the crop model virtual observations of wheat and sugar beet crops grown on these soils in different agricultural and weather conditions. The ranges of variation for parameters (see Table 1) are given from measurements performed on the real agricultural soils of Chambry. These soils are classified as loamy soils in the detailed soil map (Gras et al., 1961).

Two sets of 50 virtual plots with various soil properties were drawn within uniform distributions defined by the ranges given in Table 1: one set is characterized by shallow soils and the other by deep soils, in accordance with the two different ranges considered for the parameter  $epc(2)$ . One of these 50 virtual plots corresponds to a vector of the 7 soil parameter values, noted  $\theta^{true}$ .

For each virtual plot and each crop, 8 configurations were chosen for the simulation: 4 contrasting weather patterns and 2 different preceding crops (defining different initial conditions). The weather data were obtained from the meteorological station of Roupy (49.48°N, 3.11°E). The four different datasets, named C1, C2, C3 and C4, correspond to the 1975-1976, 1979-1980, 1972-1973 and 1990-1991 seasons respectively. These are characterized with respect to wheat as a dry season for C1, a medium-dry season for C2, a medium-wet season for C3 and a wet season for C4. The two different preceding crops were peas or sugar beet for wheat crops and a catch crop or bare soil for sugar beet. For each crop considered, the results of parameter estimation and prediction of variables will be shown averaged over the two preceding crops. The cropping techniques used are adapted to the crop considered and the type of soil depth from our knowledge of the farmers' practices in the region: for wheat, the sowing date is set to October 15, the amount of nitrogen fertilizer to 200 kg N per hectare for shallow soils and 240 kg for deep soils; for sugar beet the sowing date is set to March 30, the amount of nitrogen fertilizer to 150 kg N per hectare for shallow soils and 200 kg for deep soils.

We therefore created several sets of synthetic observations of the output variables  $LAI_t$  and  $QN_t$  at 10 dates  $t$  and  $Yld$  at harvest (i.e. 21 observations) for each virtual plot  $\theta^{true}$ , crop and weather/preceding crop configuration. The 10 dates were

distributed throughout the crop season, with a 30-day interval at the beginning of crop growth and a 15-day interval later on, i.e. November 15, December 12, January 15, February 16, March 15, April 5 and 19, May 3 and 17 and June 07 for wheat; May 29, June 12, 19 and 26, July 6 and 16, August 5 and 25, September 14 and October 04 for sugar beet. Computing the synthetic observations  $y_{v,t}$  entailed: (i) simulating with STICS the values of the observable variables at the chosen date and (ii) adding a random error term to the simulated values of the observable output variables. For each virtual plot  $\theta^{true}$  we calculated:

$$y_{v,t} = f_{v,t}(\theta^{true}, x_v) + \varepsilon_{v,t} \quad (2)$$

where  $f_{v,t}$  is the STICS prediction for the output  $v$  at date  $t$ :  $Yld$ ,  $LAI_t$  or  $QN_t$ ,  $t=1, \dots, 10$ ,  $x_t$  is the vector of explanatory variables and  $\varepsilon_{v,t}$  is the observation error term. These observation error terms are here assumed to be independent, unbiased and normally distributed. Moreover  $\varepsilon_{v,t} \sim N(0, [\sigma_v^0 f_{v,t}(\theta^{true}, x_t)]^2)$

and their magnitude is set from values obtained in field conditions in a precision farming project (Machet et al., 2007; Moulin et al., 2007):  $\sigma_{Yld}^0 = 9\%$ ,  $\sigma_{LAI}^0 = 17\%$  and  $\sigma_{QN}^0 = 30\%$ . 1600 sets of synthetic observations for each date and variable  $LAI$ ,  $QN$  and  $Yld$  (corresponding to 50 plots x 2 soil depths x 2 crops considered x 8 weather/preceding crop configurations) were thus created.

We did the parameter estimation experiments considering that observations were available for one or two crop seasons, with a possible alternation of crops over the crop seasons. Moreover, to evaluate the sensitivity of the results to the number and type of observations used, several sets of synthetic observations were computed. For one or two observed seasons, the first set (S1) is made up of 10 or 9 observations respectively of  $LAI$ , the second (S2) is made up of 20 or 19 observations respectively of  $LAI$  and  $QN$ , and the third (S3) is made up of 21 observations of  $LAI$ ,  $QN$  and  $Yld$  (see Table 2, Wheat and Wheat-Wheat). When only one crop season is considered, the 10 dates defined above for  $LAI$  and  $QN$  are used, plus  $Yld$ ; when two crop seasons are considered, alternate dates are used for  $LAI$  and  $QN$ , plus  $Yld$  for the two seasons. Thereby it is possible to compare the effect of the size and type of the observation set for one observed season and also the effect of the number of observed seasons for a given observation set.

Different configurations of parameter estimation experiments were thus conducted by using in turn one of the three observation sets from one or two crop seasons (Table 2). The ability to estimate the parameters will be examined as well as the effect of the size and type of the observation set and the effect of the soil depth and observed crop on parameter estimation. The effect of the available weather data on the soil parameter estimation will finally be studied. For each parameter estimation experiment, quality will be assessed by computing the criterion presented in Section 2.3.3.

Observed crop	Weather	Soil depth	Observation sets
Wheat	C1		
	C2	–	$S1 = \{LAI_{t=1, 2, 3, 4, 5, 6, 7, 8, 9, 10}\}$
	C3	+	$S2 = S1 + \{QN_{t=1, 2, 3, 4, 5, 6, 7, 8, 9, 10}\}$
	C4		$S3 = S2 + \{Yld_{harvest}\}$
Wheat-Wheat	C1-C4		$S1 = \{LAI_{t=3, 5, 7, 9}\} + \{LAI_{t=2, 4, 6, 8, 10}\}$
	C2-C3	–	$S2 = S1 + \{QN_{t=1, 3, 5, 7, 9}\} + \{QN_{t=2, 4, 6, 8, 10}\}$
	C1-C2		$S3 = S2 + \{Yld_{harvest}\} + \{Yld_{harvest}\}$
Sugar beet	C1		
	C2	–	$S1 = \{LAI_{t=1', 2', 3', 4', 5', 6', 7', 8', 9', 10'}\}$
	C3	+	$S2 = S1 + \{QN_{t=1', 2', 3', 4', 5', 6', 7', 8', 9', 10'}\}$
	C4		$S3 = S2 + \{Yld_{harvest}\}$
Wheat-Sugar beet	C1-C4		$S1 = \{LAI_{t=3, 5, 7, 9}\} + \{LAI_{t=1', 3', 5', 7', 9'}\}$
	C2-C3	–	$S2 = S1 + \{QN_{t=1, 3, 5, 7, 9}\} + \{QN_{t=1', 3', 5', 7', 9'}\}$
	C1-C2		$S3 = S2 + \{Yld_{harvest}\} + \{Yld_{harvest}\}$
Sugar beet-Sugar beet	C1-C4		$S1 = \{LAI_{t=2', 4', 6', 8'}\} + \{LAI_{t=1', 3', 5', 7', 9'}\}$
	C2-C3	–	$S2 = S1 + \{QN_{t=2', 4', 6', 8', 10'}\} + \{QN_{t=1', 3', 5', 7', 9'}\}$
	C1-C2		$S3 = S2 + \{Yld_{harvest}\} + \{Yld_{harvest}\}$

The weather pattern C1 corresponds to weather data measured in 1975-1976, C2 to 1990-1991, C3 to 1972-1973 and C4 to 1979-1980. – = shallow soil, + = deep soil. Each of the 3 observation sets (S1, S2 and S3) are considered among the following dates for the wheat: 1) 11/15, 2) 12/12, 3) 01/15, 4) 02/16, 5) 03/15, 6) 04/05, 7) 04/19, 8) 05/03, 9) 05/17, 10) 06/07, and for the sugar beet: 1') 05/29, 2') 06/12, 3') 06/19, 4') 06/26, 5') 07/06, 6') 07/16, 7') 08/05, 8') 08/25, 9') 09/14, 10') 10/04.

**Table 2.** Description of the different parameter estimation experiments conducted on synthetic observations.

### *Data and numerical experiments for predictions*

Variables of interest are predicted for crop seasons that are independent from those used in the previous step (the creation of observations for the parameter estimation process). For each of the 50 virtual plots defined above and each crop (wheat and sugar beet), 120 configurations for prediction were studied: 2 soil depths, 10 years' weather records, 3 sowing dates (15 days before and after the date used in the previous step), and 4 different cropping techniques (2 fertilizer rates, 20% more or less than the rate used in the previous step, x 2 two preceding crops). The weather data were obtained from the meteorological station of Roupy (49.48°N, 3.11°E) and are different from those used in the parameter estimation process.

For each virtual plot  $\theta^{true}$ , soil depth and prediction configurations, the values of synthetic observations of the predicted variables of interest (*Yld*, *Prot* and *Nit* at harvest) were simulated with STICS but with no error term added. The same values of the permanent soil properties as those used to create the synthetic observations in the estimation step were used for the simulations, while the values of the initial conditions (*Hinit* and *NO3init*) were randomly generated from the distributions defined in Section 2.1.3. For each parameter estimation experiment defined above, each estimated value of the permanent properties was used to predict the output variables of wheat and sugar beet crops using the STICS model. To make the prediction, two assumptions were made about the initial conditions for the predicted season: either they are unknown and thus fixed as the mean of their prior distribution, or they are known and are thus fixed at their measured values. In this way, it should be possible to quantify the effect of the knowledge of the initial conditions on the quality of the output variable prediction.

Different prediction experiments were done to measure the extent to which the prediction error could be reduced by using the estimated values of the permanent properties. The sensitivity of these results to the size and type of the observation set used for estimating the soil parameters, the soil depth and observed crop is therefore analyzed. The success in reducing the prediction error by using the available weather data for estimating the soil parameters is studied for a single observation season on the wheat crop. Finally, the effect of a possible alternation of crops over two crop seasons is studied on the prediction of the behavior of both crops.

## 2.2.2 Real experimental data and experiments

### *Dataset*

We used a database based on long-term experiments in the Bruyères catchment (49.52°N, 3.67°E). Soil and crop measurements were made from 1991 to 1999 on 36 unreplicated permanent sampling sites which are representative of the main crops and soil types. In the real experiment context, a site represents a plot in the synthetic study. The two main crops in the database are winter wheat and sugar beet, the others being oilseed rape, peas and barley. Weather data were measured continuously either at two locations on the site or from the nearest weather station. Full details on this database can be found in Beaudoin *et al.* (2005; 2008).

In these conditions, observations on crops were much scarcer than in our synthetic database. Of the 36 sites in the Bruyères database, 12 were used which provided enough observations of *LAI*, *QN* and *Yld* (at least 3, 3 and 1 respectively) for at least one crop season and extra observations of variables of interest for at least another season. These sites have similar loamy soils to those considered and characterized in the synthetic database. For these sites, the number of observations of *LAI*, *QN* and *Yld* usable for the parameter estimation experiments and the number of observations of *Yld*, *Prot* and *Nit* usable for the prediction of both crops are presented in Table 3. The number of available sites for parameter estimation experiments varies according to the observed crops: 4 for wheat, 10 for sugar beet and 12 for accumulations of crop seasons. The number of observations of the variable *j* which is predicted for both crops varies from 0 to 8. Unlike the synthetic experiments, the observation dates do not cover the entire crop season and no observations are available at the beginning of the crop season. All the weather patterns involved in the 8-year study of the Bruyères database are characterized as wet and the soil types of the 12 sites used are deep loamy soils.



Observed crop	Site	Observation sets	Prediction sets	
			Wheat predicted	Sugar beet predicted
Wheat	2	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 6 Nit}	{1 Yld + 7 Nit}
	5	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 5 Nit}	{1 Yld + 2 Nit}
	7	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 2 Nit}	{1 Yld + 2 Nit}
	12	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 3 Nit}	{1 Yld + 4 Nit}
Sugar beet	1	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 3 Nit}	{1 Yld}
	2	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 6 Nit}	{1 Yld + 7 Nit}
	3	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 3 Nit}	{2 Yld + 8 Nit}
	4	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 8 Nit}	{4 Nit}
	5	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 5 Nit}	{1 Yld + 2 Nit}
	6	{3 LAI + 3 QN + 1 Yld}	{2 Yld + 2 Prot + 3 Nit}	{1 Yld + 2 Nit}
	7	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 2 Nit}	{1 Yld + 2 Nit}
	8	{3 LAI + 3 QN + 1 Yld}	{2 Yld + 2 Prot + 3 Nit}	{ }
	9	{3 LAI + 3 QN + 1 Yld}	{3 Yld + 3 Prot + 5 Nit}	{1 Yld}
	10	{3 LAI + 3 QN + 1 Yld}	{2 Yld + 2 Prot + 5 Nit}	{2 Yld + 4 Nit}
Sugar beet-	1	{3 LAI + 9 QN + 3 Yld}	{1 Yld + 1 Prot + 3 Nit}	{1 Yld}
Wheat-Wheat	2	{6 LAI + 10 QN + 3 Yld}	{3 Yld + 3 Prot + 6 Nit}	{1 Yld + 7 Nit}
	3	{3 LAI + 9 QN + 3 Yld}	{1 Yld + 1 Prot + 3 Nit}	{2 Yld + 8 Nit}
Sugar beet-	4	{4 LAI + 6 QN + 2 Yld}	{3 Yld + 3 Prot + 8 Nit}	{4 Nit}
Wheat	5	{6 LAI + 7 QN + 2 Yld}	{3 Yld + 3 Prot + 5 Nit}	{1 Yld + 2 Nit}
	6	{3 LAI + 7 QN + 2 Yld}	{2 Yld + 2 Prot + 3 Nit}	{1 Yld + 2 Nit}
	7	{6 LAI + 7 QN + 2 Yld}	{1 Yld + 1 Prot + 2 Nit}	{1 Yld + 2 Nit}
	8	{3 LAI + 7 QN + 2 Yld}	{2 Yld + 2 Prot + 3 Nit}	{ }
	9	{3 LAI + 7 QN + 2 Yld}	{3 Yld + 3 Prot + 5 Nit}	{1 Yld}
	10	{3 LAI + 6 QN + 2 Yld}	{2 Yld + 2 Prot + 5 Nit}	{2 Yld + 4 Nit}
Wheat-Wheat	11	{7 QN + 2 Yld}	{2 Yld + 2 Prot + 8 Nit}	{2 Yld + 8 Nit}
Wheat	12	{3 LAI + 3 QN + 1 Yld}	{1 Yld + 1 Prot + 3 Nit}	{1 Yld + 4 Nit}

**Table 3.** Description of the different parameter estimation experiments and the predictions conducted on the selected data of Bruyères database, showing the available observations in the observation and prediction sets.

### Numerical experiments

Similarly to the study on synthetic data, the parameter estimation experiments done on this real database depended on the available observations. As shown in Table 3, parameter estimation involves alternating observations on wheat and sugar beet and accumulated observations on wheat and sugar beet over two and three crop seasons. As for the synthetic experiments, the variables of interest were

predicted by using the estimated values of the permanent soil parameters and by using the mean value of the distribution of the initial conditions or their measured values. We assume for the experiments on real data that the true values  $\theta^{true}$  of the parameters and of the observations of the predicted variables, denoted below  $f_j(\theta^{true})$ , are those measured on the sites. As shown in Tab. 1, the true values  $\theta^{true}$  vary between 16 and 25 for *argi*, 0.1 and 0.13 for *Norg*, 90 and 120 for *epc(2)*, 22.5 and 29.5 for *HCC(1)*, 20 and 25.5 for *HCC(2)*, 6 and 25 for *HCC(1)* and 6 and 50 for *NO3init*.

## 2.3 Parameter estimation

### 2.3.1 Method of parameter estimation

We chose a Bayesian method which takes account of existing information on the parameters to be estimated (this improves the quality of the estimation) and computes an estimate of the posterior probability distribution of parameter values (Makowski et al., 2002; Gaucherel et al., 2008).

The posterior parameter distribution is given by Bayes' theorem:

$$\pi(\theta/Y) = \frac{\pi(Y/\theta)\pi(\theta)}{\pi(Z)} \quad (3)$$

where  $Y$  is the vector of total observations of size  $K$ ,  $\pi(\theta/Y)$  is the posterior parameter distribution,  $\pi(\theta)$  is the prior parameter distribution,  $\pi(Y)$  is a constant of proportionality determined by the requirement that the integral of  $\pi(\theta/Y)$  over the parameter space equals 1, and  $\pi(Y/\theta)$  is the likelihood function. The likelihood is the probability of the data  $Y$  given the parameters  $\theta$ . Its value is determined from the probability distribution of the errors of modeled and observed data. It is readily seen that both the prior distribution and the new data affect the posterior parameter distribution.

The two most popular families of bayesian methods are Importance Sampling and MCMC (Gilks et al., 1995). These methods are based on Monte Carlo simulations and thus require a large number of model evaluations. The Importance

Sampling method (Beven and Freer, 2001; Makowski et al., 2002) was chosen for our study.

The principle of the Importance Sampling method (Beven and Binley, 1992; Beven and Freer, 2001) is to approximate the posterior parameter distribution  $\pi(\theta/Y)$  given in (3) by a discrete probability distribution  $(\theta_n, p_n)$ ,  $n=1, \dots, N$ ,  $\sum_{n=1}^N p_n = 1$ , where  $p_n$  is the probability associated with the parameter vector  $\theta_n$ . In our case, the method proceeds as follows:

(1) Randomly generate  $N$  vectors  $\theta_n$ ,  $n=1, \dots, N$ , from the prior parameter distribution  $\pi(\theta)$ .

(2) Calculate the likelihood values  $\pi(Y/\theta_n)$  for  $n=1, \dots, N$ , associated with the different generated parameter vectors.

(3) Calculate 
$$p_n = \frac{\pi(Y/\theta_n)}{\sum_{m=1}^N \pi(Y/\theta_m)}$$

The pairs  $(\theta_n, p_n)$ ,  $n=1, \dots, N$ , can be used to determine various characteristics of the posterior distribution, including the mean of the posterior joint distribution of  $\theta$ ,

$$\bar{\theta}^{post} = \sum_{n=1}^N p_n \theta_n .$$

In this study, we assume that the errors of simulated and observed data are independent between dates and variables and follow normal distributions of zero mean and standard deviation  $\sigma_k$ . Thus, we use the following likelihood function:

$$\pi(Y/\theta) = \prod_{k=1}^K \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{1}{2\sigma_k^2} [y_k - f_k(\theta, x)]^2\right\} \quad (4)$$

The parameters are assumed to be independent in our case. The prior distribution  $\pi(\theta)$  is thus the product of the different marginal prior distributions. We have implemented the Importance Sampling method using Matlab® software, and the application of the procedure simply requires the definition of the total number of generated parameter vectors  $N$ . A preliminary study of the convergence of the estimates allowed us to set this number at 10 000.

### 2.3.2 Prior information

As prior information on the parameter values, we used ranges of variation defined in Table 1 and assumed uniform (i.e.: non-informative) distributions. These ranges were obtained from measurements made on precision agriculture site of Chambry (see Section 2.1.3) on loamy soils. The soils represented in the Bruyères database are deep loamy soils and are thus similar to those of the experimental fields of Chambry, but deeper. Moreover, the prior distribution determined on Chambry is convenient to describe the parameter variability observed at Bruyères, as shown in Tab. 1. This prior distribution is thus used to estimate parameters on both experiments (synthetic and real).

### 2.3.3 Criterion expressing the quality of parameter estimation

We proposed a criterion, noted  $RE_i$  (for Relative Error of the parameter  $i$ ), to quantify the quality of parameter estimation. It computes, for each parameter, the ratio of the error of the estimated parameter  $\bar{\theta}_i^{post}$  to that of the prior information  $\bar{\theta}_i^{prior}$ :

$$RE_i = \frac{RMSE(\bar{\theta}_i^{post})}{RMSE(\bar{\theta}_i^{prior})} \quad (5)$$

$RMSE(\bar{\theta}_i^{post}) = \sqrt{\frac{1}{P} \sum_{p=1}^P (\theta_{i,p}^{true} - \bar{\theta}_{i,p}^{post})^2}$ ,  $\theta_{i,p}^{true}$  is the true value of soil parameter  $\theta_i$  for a given plot  $p$ , and  $\bar{\theta}_{i,p}^{post}$  is the corresponding estimate given by the bayesian method.

$RE_i$  quantifies how much the estimate given by the bayesian method improves ( $RE_i < 1$ ) or not ( $RE_i \geq 1$ ) the prior knowledge of the parameter value. For the synthetic database  $P$  is 50 and for the real database, it is 4, 10 and 12 for experiments on wheat, sugar beet and accumulated observations respectively.

### 2.3.4 Criterion expressing the quality of prediction

In a similar way we proposed a criterion, denoted  $REP_j$  (for Relative Error of Prediction) to quantify the quality of the prediction of the 3 agro-environmental

variables defined above. It computes the ratio of the error of prediction obtained from the mean of the posterior distributions of the parameters,  $\bar{\theta}^{post}$ , to that obtained from the mean of the prior distributions,  $\bar{\theta}^{prior}$  :

$$REP_j = \frac{RMSEP_j(\bar{\theta}^{post})}{RMSEP_j(\bar{\theta}^{prior})} \quad (6)$$

where  $RMSEP_j(\bar{\theta}^{post}) = \sqrt{\frac{1}{P \times Q_j} \sum_{p=1}^P \sum_{q=1}^{Q_j} (f_j^q(\theta_p^{true}) - f_j^q(\bar{\theta}_p^{post}))^2}$ ,  $\theta_p^{true}$  is the vector of the true values of soil parameters  $\theta$  for a given plot  $p$ ,  $\bar{\theta}_p^{post}$  is the corresponding estimate given by the Importance Sampling method, and  $f_j^q(\theta_p^{true})$  is assumed to be one of the  $Q_j$  observations of the predicted variable  $j$ , for the  $p^{th}$  vector of true values of soil parameters. For the synthetic database,  $Q_j$  is equal to 120 and for the real database its value depends on the number of available observations for the prediction as shown in Table 3.

### 3. Results and discussion

#### 3.1 Synthetic experiments

##### 3.1.1 Parameter estimates

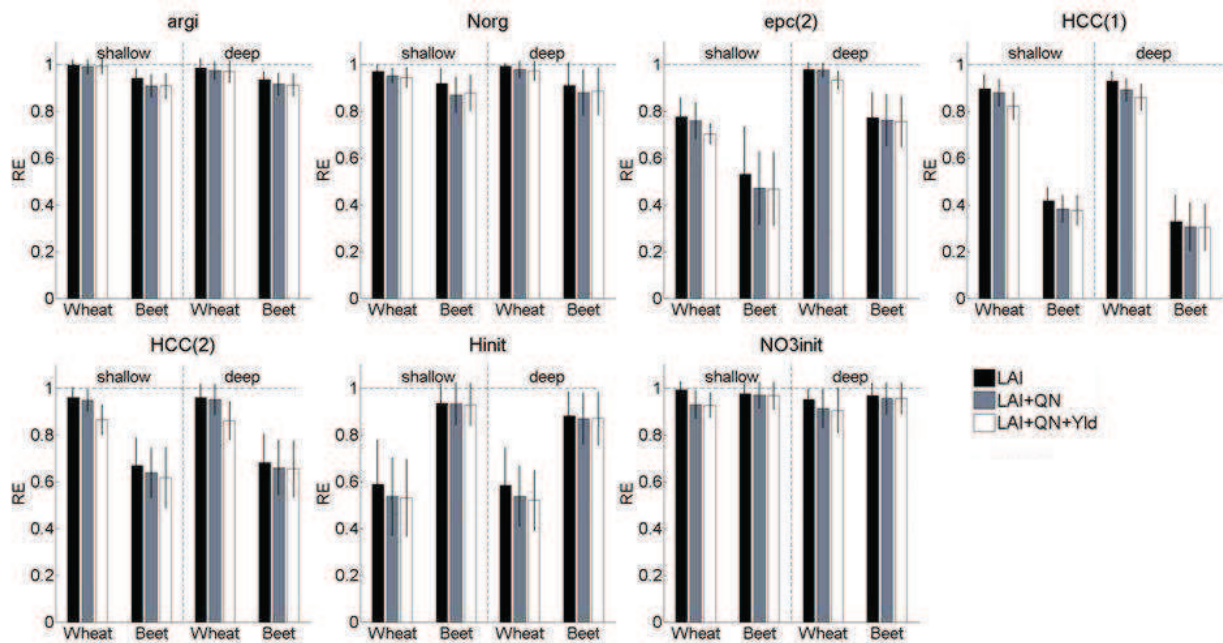
The results of parameter estimation are shown in Figure 1 as the mean of the 4 weather patterns for one season of observation on wheat and sugar beet, 2 soil depths and 3 sets of observations. The results show that the observations provide substantial information for only 4 of the 7 soil parameters: *epc(2)*, *HCC(1)*, *HCC(2)* and *Hinit*. For these parameters the maximal gain from the prior information is between 38 and 70%. The other 3 parameters are difficult to estimate with these observations: for *argi*, *Norg* and *NO3init* the maximal gain is between 5 and 12%. These results can be explained by the different levels of sensitivities of the observable variables to these parameters as shown in (Varella et al., 2008).

The crop observed plays an important role in the quality of the parameter estimation process. Observations on sugar beet always provide better estimates of

the parameters related to permanent properties ( $epc(2)$ ,  $HCC(1)$  and  $HCC(2)$ ) and to a lesser extent  $argi$  and  $Norg$ ), while observations on wheat always provide better estimates of the initial conditions ( $Hinit$  and to a lesser extent  $NO3init$ ). These results can also be explained by the sensitivities of the observed variables to the parameters. As sugar beet is a summer crop (subjected to water stress), the soil water retention properties are more important for its growth than for winter wheat (usually not subjected to water stress in this region). The observed variables of sugar beet are thus more sensitive to the soil's permanent properties, and the corresponding observations contain more information to estimate the related parameters than those related to initial conditions. The lower sensitivity of sugar beet to initial conditions as compared to wheat is because the rainfall is often heavy in spring after sugar beet sowing, reducing the importance of initial conditions compared with wheat, which is sown in autumn when the weather is drier.

The soil depth only affects the parameter  $epc(2)$  whose quality of estimation is clearly better when the soil is shallow: for observations on wheat, the gain with respect to the prior information is around 25% for shallow soil and 5% for deep soil. The corresponding values for sugar beet are 50% and 25% respectively. For the other parameters, the criterion  $ER_i$  calculated for the two soil depths has almost the same value.

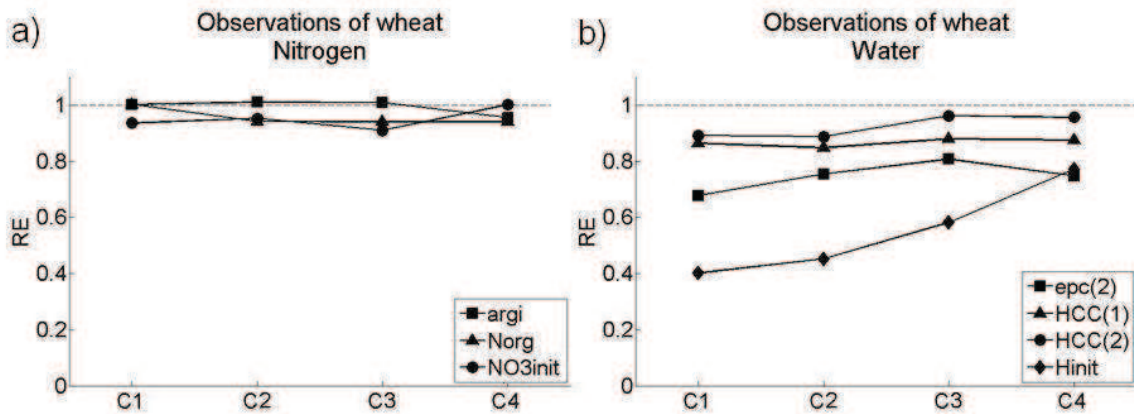
Finally, the size of the observation set only slightly affects the quality of parameter estimation even though, as expected, the bigger the observation set the better is the parameter estimation. The gain is always limited to a few percent.



**Figure 1.** Results of the parameter estimation averaged over the 4 weather patterns for one season of observation on wheat (Wheat) and sugar beet (Beet), the 2 soil depths and the 3 sets of observations.

The effect of weather on the quality of parameter estimation is then analyzed for the mean of the 3 observation sets for one season of observation on wheat on a shallow soil, as shown in Figure 2. Whatever the weather pattern, the parameter set related to soil mineral nitrogen content is difficult to estimate (Fig. 2a): the lowest value of  $RE_i$  is over 0.9. The weather only affects the quality of estimation of the parameters related to soil water content: the drier the season, the better is the estimate of these parameters. The effect of weather on *Hinit* is striking:  $RE_i$  is about 0.77 on average for the wet weather pattern and about 0.4 for the dry weather pattern (which means a gain of 23% in the first case and 60% in the second). As the sensitivity of the observable variables to the parameters related to the water content is higher in dry conditions than in wet ones, the quality of estimates of these parameters is better in the first case.





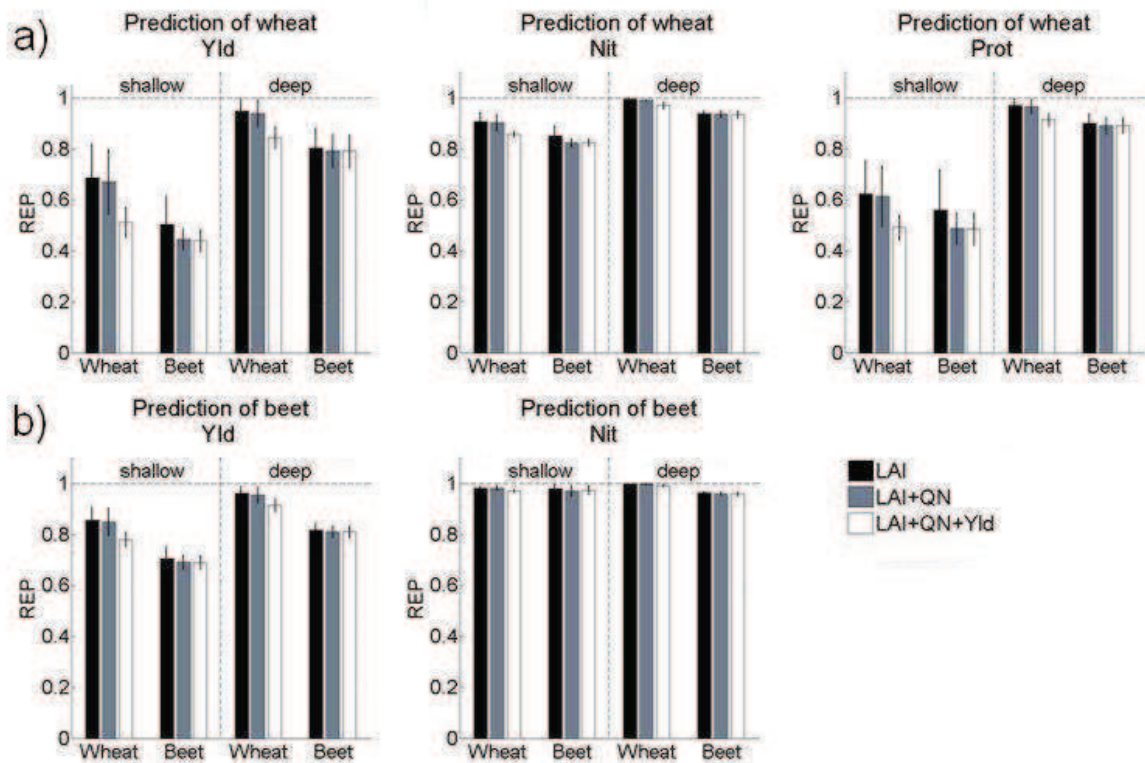
**Figure 2.** Effect of the weather on the parameter estimation averaged over the 3 observation sets, for one season of observation on wheat and a shallow soil. The parameters involved in the soil nitrogen content are in a) and those involved in the soil water content are in b). For the wheat, C1 characterizes a dry season, C2 a medium-dry season, C3 a medium-wet season and C4 a wet season.

### 3.1.2 Prediction of variables of interest

The quality of prediction will now be analyzed. Figure 3 shows the results in terms of  $REP_j$  for the prediction of the variables of interest for both crops, by using the estimated values of the permanent soil parameters (*argi*, *Norg*, *epc(2)*, *HCC(1)* and *HCC(2)*) instead of their prior values. The initial conditions (*Hinit* and *NO3init*), considered as unknown, are fixed at the mean of their distributions. As in Fig. 1,  $REP_j$  is calculated for one season of observations on wheat and sugar beet, 2 soil depths and 3 sets of observations.  $REP_j$  is then averaged over the 4 weather patterns. For both crops it can be seen that 2 of the 3 variables of interest (*Yld* and *Prot*) have a significantly lower  $REP_j$  and therefore a greatly improved quality of prediction when using the estimated values, as compared to when using prior information on the parameters. Note that the prediction quality of *Yld* is better when the wheat is predicted than when the sugar beet is predicted because the sugar beet crop is more sensitive to the initial conditions which are considered to be unknown in these results. The output variable *Nit* is only slightly affected, if at all, by the estimation of the soil parameters because it is sensitive to the initial conditions, which are not estimated, and not to the permanent parameters.

Through the estimation of the permanent soil properties, the size of the observation set slightly influences the quality of prediction. The biggest improvement between two observation sets is for parameters estimated from observations on

wheat when the wheat growth is predicted for a shallow soil and the output variable is *Yld*:  $REP_j$  is about 0.66 for set #2 and about 0.52 for set #3.



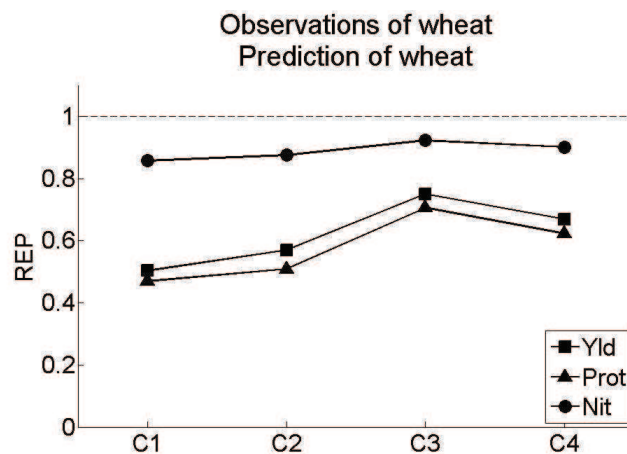
**Figure 3.** Results of the prediction of the variables of interest for a) wheat and b) sugar beet. The predictions are simulated by using the estimated values of the permanent soil parameters and by fixing the initial conditions at the mean of their distributions. The results are averaged over the 4 weather patterns for one season of observation on wheat (Wheat) and sugar beet (Beet), the 2 soil depths and the 3 sets of observations.

The crop on which observations are made for parameter estimation also plays an important role in the quality of prediction. Because observations on sugar beet improve the quality of permanent parameter estimates, the results show that the predictions are better when the soil parameters are estimated with observations made on sugar beet. This effect can be seen on the 3 predicted variables but especially on *Yld*. It is particularly striking for the prediction of *Yld* for wheat for observation set #2: the gain from prior information varies from 33 to 55% when using observations on sugar beet rather than wheat. It is striking that the improvement in the wheat prediction is greater by using observations on sugar beet than on wheat. As the sensitivity of the observed variables on sugar beet to the permanent

parameters is higher than the observed variables on wheat, the estimation of these parameters is thus better, implying a smaller uncertainty on the prediction of output variables in this case.

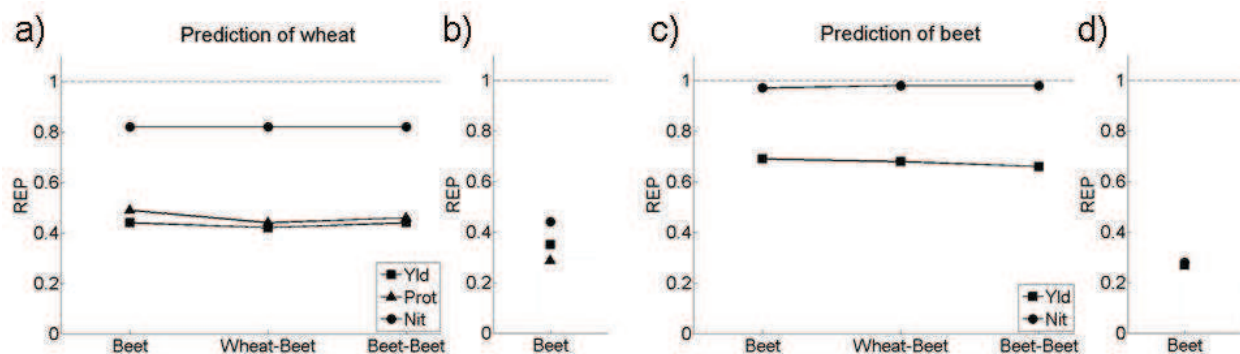
The soil depth affects the quality of the prediction, especially for the output variables *Yld* and *Prot*, which have a lower  $REP_j$  when the soil is shallow. It is not surprising, accordingly to the results of parameter estimation, because the parameter  $epc(2)$  gives better estimates in this case and because these outputs are quite sensitive to this parameter. The output variable *Nit* is not affected by the soil depth because of its lack of sensitivity to  $epc(2)$ .

The effect of weather on the quality of prediction of the variables of interest is illustrated for wheat in Figure 4 for a shallow soil and one year of observations: the results are the means of the 3 observation sets. The drier the weather (from wet C4 to dry C1) the better is the prediction, which follows from the previous results showing that the quality of estimate of permanent parameters related to the water content is affected by the weather. The  $REP_j$  value of the output variables *Yld* and *Prot* decreases significantly from 0.71 to 0.5 and 0.65 to 0.47 respectively. As before, the output variable *Nit* is not accurately predicted and is only slightly affected by the weather:  $REP_j$  decreases from 0.92 to 0.86.



**Figure 4.** Effect of the weather on prediction of the variables of interest of the wheat crop when the initial conditions are unknown. The results are averaged over the 3 observation sets for one observed season on wheat crop and a shallow soil. For the wheat, C1 characterizes a dry season, C2 a medium-dry season, C3 a medium-wet season and C4 a wet season.

Given that one season of observations on sugar beet allows better prediction (see Fig. 3), we investigated the marginal improvement due to adding another observation season on sugar beet or wheat. Figure 5a and 5c shows the results in terms of  $REP_j$  for the prediction of the variables of interest for both crops for a shallow soil and the observation set #3 averaged over the different weather patterns. In this case, as in the previous one, the initial conditions are assumed to be unknown. The results show that the addition of another observation season either on sugar beet or on wheat does not significantly improve the prediction obtained using only one observation season on sugar beet. They confirm that the better improvement in prediction is made for the wheat crop and that the output variable *Nit* is difficult to predict correctly.



**Figure 5.** Results of the prediction of the variables of interest for the wheat crop when the initial conditions are a) unknown b) known, and for the sugar beet crop when the initial conditions are c) unknown d) known. One season of observations on sugar beet (Beet) and two pairs of combined observation seasons are considered: wheat for the first season and sugar beet for the second season (Wheat-Beet), and sugar beet for both seasons (Beet-Beet). The results are averaged over the various weather patterns and shown for a shallow soil and the third set of observations.

For crop monitoring applications, the initial conditions may be considered as known (from measurements). Fig. 5b and 5d show the results obtained for the same situation as before, with the initial conditions set to their true values. In this case, the reduction of the error of prediction is striking for the sugar beet crop, which is more sensitive to the initial conditions than wheat. All the output variables are affected but the best improvement is for *Nit* which is the most sensitive to the initial conditions:  $REP_j$  decreases from 0.82 to 0.44 when wheat is predicted and from 0.97 to 0.28 when sugar beet is predicted.

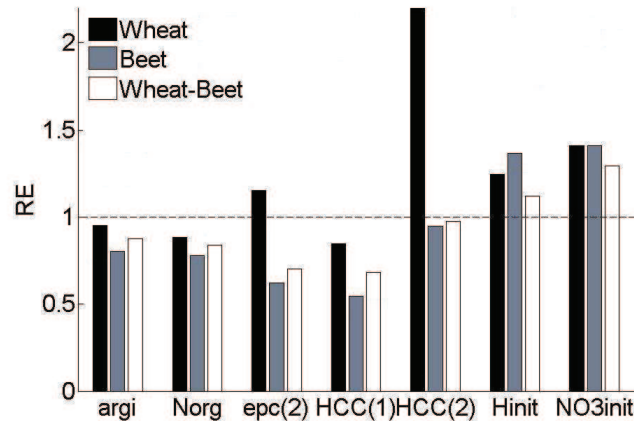
## 3.2 Real experiments

### 3.2.1 Parameter estimates

The results of parameter estimation on the real experiments at Bruyères are presented in Figure 6 for different observation sets. In every case  $HCC(1)$  is the best estimated parameter ( $RE_i$  between 0.85 and 0.54). The parameters involved in the soil mineral nitrogen content,  $argi$  and  $Norg$ , can be slightly estimated (gain between 2 and 22%), and the parameter  $epc(2)$  can be estimated only when observations on sugar beet crop are used ( $RE_i$  between 0.7 and 0.62 instead of 1.15 when observations on wheat are used). The other parameters,  $HCC(2)$ ,  $Hinit$  and  $NO3init$ , can not be estimated and have a  $RE_i$  value of 1 or slightly more. This is mainly due to a lack of observations early in wheat growth, which is when the observed variables are sensitive to  $Hinit$  and  $NO3init$ . As compared to the results obtained on synthetic data, we would argue that several types of bias have to be considered in addition on these results : the bias due to the STICS model error on the output simulation, the omission bias (Miller, 2002) and the bias on the observations. The addition of such biases increases the error of the estimated values of the parameters. This is clearly visible on the parameters with an  $RE_i$  greater than 1 ( $epc(2)$  and  $HCC(2)$  for the wheat observation set and the initial conditions  $Hinit$  and  $NO3init$ ): in these cases these biases do not allow to have a  $RMSE(\bar{\theta}_i^{post})$  lower than  $RMSE(\bar{\theta}_i^{prior})$ .

Observations on sugar beet always lead to better permanent parameter estimates than observations on wheat. The effect of the crop observed varies greatly according to the parameter:  $RE_i$  from 0.88 to 0.78 for  $Norg$ , from 1.15 to 0.62 for  $epc(2)$ , from 2.2 to 0.95 for  $HCC(2)$ . For the initial conditions, the lack of observations in the early period of the growing season prevents us from assessing the effect of the crop observed. Accumulating observations on wheat and sugar beet crops is always preferable to using only wheat observations, and the gain obtained varies according to the parameters. Nevertheless, observations on a single sugar beet crop season provide more information for estimating the permanent parameters than the accumulation of observations on wheat and sugar beet crops. The gain from using

prior information is 5% for  $HCC(2)$ , 20% for  $argi$ , 22% for  $Norg$ , 38% for  $epc(2)$  and 46% for  $HCC(1)$ .



**Figure 6.** Results of parameter estimation on the real experiments of Bruyères, for one season of observation on wheat (Wheat) or sugar beet (Beet) and for accumulated observations on wheat and sugar beet crops (Wheat-Beet).

Some of these results are quite consistent with those obtained from experiments on the same configurations of synthetic data, i.e. wet weather patterns and deep soils, as in the real data. Firstly, for a given observation set,  $HCC(1)$  is the best estimated permanent soil parameter (see Fig. 1 and 6) and can thus be estimated with great confidence. Secondly, observations on sugar beet provide very informative data for estimating the permanent soil parameters and allow better estimates of those parameters than observations on wheat. For both synthetic and real data,  $argi$  and  $Norg$  can be slightly estimated; the observed crop has little influence on their estimates. Except for observations on wheat, the results of the estimation of  $epc(2)$  and  $HCC(2)$  are consistent with those obtained on the synthetic data:  $epc(2)$  is well estimated while  $HCC(2)$  is quite difficult to predict.

Nevertheless, there are some differences from the results obtained on the synthetic data. For observations on wheat,  $epc(2)$  and  $HCC(2)$  have a very high value (above 1) of  $ER_i$  for real data while they are reasonably well estimated (about 0.75 for  $epc(2)$  and 0.9 for  $HCC(2)$ ) for synthetic data (see Fig. 6). This could be explained by the presence of increased bias (model error bias, omission bias or observation bias) in estimating these parameters. The last difference concerns the initial conditions  $Hinit$  and  $NO3init$  and the impossibility of estimate these parameters for real data

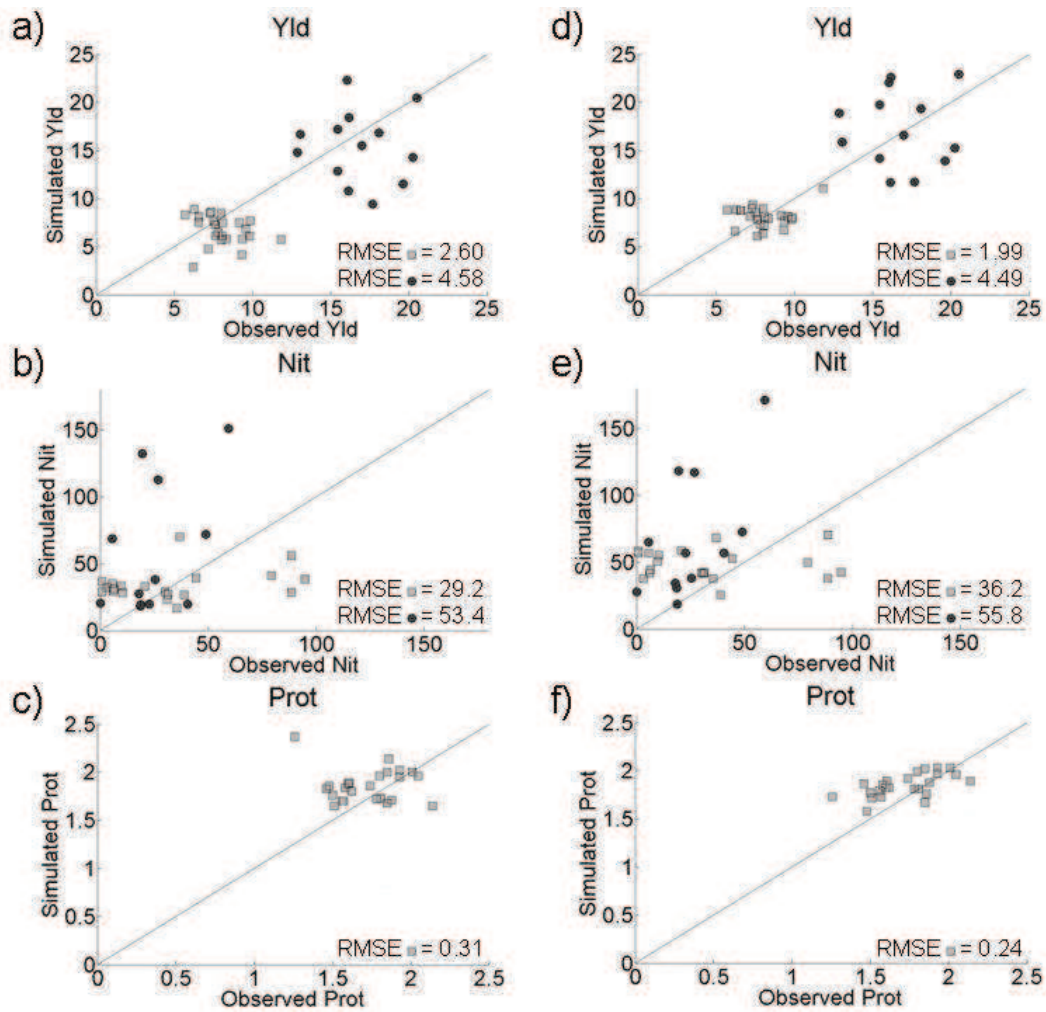


( $ER_i$  above 1) due to the lack of early observations, unlike the case of the synthetic data.

### 3.2.2 Prediction of variables of interest

Figure 7 compares observed variables with those simulated by STICS using prior information (Fig. 7a, 7b, 7c) and estimated values (Fig. 7d, 7e, 7f) of the soil parameters. The initial conditions are assumed to be unknown. For the simulations using prior information, the results are quite good for wheat *Yld* and *Prot* (Fig. 7a, 7c) although *Yld* is slightly under-estimated. The simulations are improved by the parameter estimation process (Fig. 7d, 7f) and *Yld* is no longer under-estimated: the model error is low and the simulation error is mainly due to inaccurate soil parameter values, even though the prior information is fairly suitable. Conversely, sugar beet *Yld* is less accurately simulated by the model and is not improved by parameter estimation because the model error is higher than for wheat *Yld* and cannot be reduced by estimating the soil parameters. These results are partly due to a better previous calibration of the plant parameters for wheat than for sugar beet (see Section 2.1.1), involving a smaller model error for wheat crop simulation. *Nit* (Fig 7b) is poorly predicted by the model for both wheat and sugar beet and is not improved by parameter estimation (Fig 7e): this confirms the inability of STICS to predict this variable satisfactorily (Houlès et al., 2004).

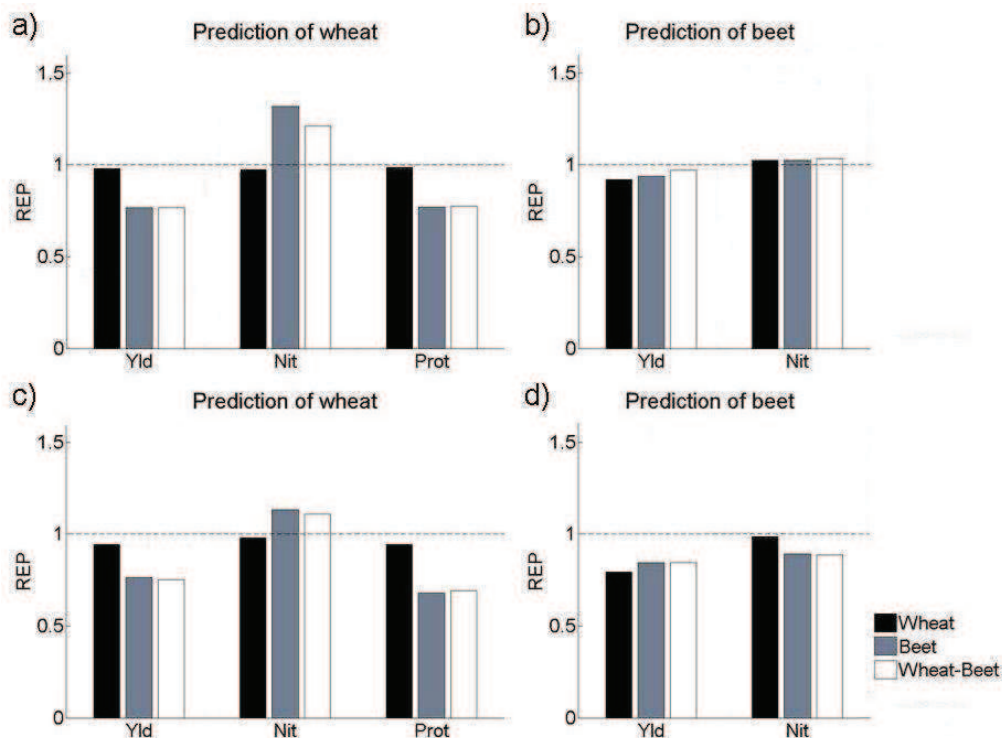




**Figure 7.** Comparison of observed and simulated variables *Yld*, *Nit* and *Prot*, obtained with the STICS model on the Bruyères database. a), b) and c) are for simulations using the prior values of the soil parameters and d), e) and f) are for those using the estimated values (those estimated with observations accumulated on wheat and sugar beet crop). In both cases, the initial conditions are unknown. □ = wheat, ○ = sugar beet.

An averaged representation of these results is given in Figure 8 where the prediction of the variables of interest for both crops after estimating the soil permanent parameters (*argi*, *Norg*, *epc(2)*, *HCC(1)* and *HCC(2)*) is shown. In a first attempt, the initial conditions (*Hinit* and *NO3init*) are assumed to be unknown and set to their prior values (Fig. 8a and 8b). The prediction of wheat variables *Yld* and *Prot* is generally improved (by about 23%) by estimating the permanent soil parameters as compared to the results obtained with prior information: this is due to their sensitivity to these parameters. For sugar beet the improvement is not significant for the output variable *Yld* (the lower  $REP_j$  value is about 0.92). However the prediction of the output *Nit* can never be improved because of a lack of sensitivity to the

permanent soil parameters but also because of a large STICS model error involved in its simulation, as stated earlier.



**Figure 8.** Results of the prediction of the variables *Yld*, *Nit* and *Prot*, on the Bruyères database, for wheat (a and c) and sugar beet (b and d), when the initial conditions are unknown (a and b) or known (c and d). The results are shown for one season of observation on wheat (Wheat) or sugar beet (Beet) and for accumulations of observations on wheat and sugar beet crops (Wheat-Beet).

The effect of the crop observed on the prediction of wheat growth is striking, but not on that of sugar beet. The error of the prediction of wheat *Yld* and *Prot* decreases considerably when using observations on sugar beet rather than on wheat: it falls from 0.98 to 0.77 for *Yld* and from 0.99 to 0.77 for *Prot*. However the observed crop does not affect the prediction of sugar beet growth or its error. The accumulation of observations on wheat and sugar beet crops improves the prediction of wheat *Yld* and *Prot* as compared to the use of a single observation season on wheat, while it has no effect on the prediction of sugar beet growth. Finally, accumulating observations on different crops does not improve the results obtained with sugar beet observations alone.

In a second step the initial conditions were known and fixed at their actual measured values (Fig. 8c and 8d). This led to an all-round improvement. The output

*Nit* for both crops is the most affected variable because of its great sensitivity to the initial conditions. The prediction error of sugar beet *Nit* can be significantly reduced when observations on sugar beet are considered and when the initial conditions are known ( $REP_j$  is about 0.78). Moreover, the *Yld* output is less improved for wheat than for sugar beet for which the error is reduced from an  $REP_j$  of about 0.92 when the initial conditions are unknown to about 0.79-0.85 when they are known. As stated before, sugar beet *Yld* is more sensitive to the initial conditions

Some of these results are quite consistent with those obtained on the synthetic data for wet conditions and deep soil. First, both experiments show that observations on sugar beet give a better prediction of wheat *Yld* and *Prot* (see Fig. 3 and 7) than those on wheat. Moreover, both studies show that combining or accumulating observations over several wheat and sugar beet crop seasons does not improve predictions any more than considering observations from a single sugar beet crop season, revealing the importance of this kind of observation for the parameter estimation and the consequent prediction. The output variable *Nit* proves to be very difficult to predict using these kinds of observations and the STICS model. The last main consistent result concerns the effect of the knowledge of the initial conditions on the reduction of the prediction error. This knowledge can reduce the prediction error of all the output variables but particularly for sugar beet because they are very sensitive to the initial conditions for this crop. The main difference between the results obtained on the real and synthetic data concerns the prediction of sugar beet growth: there is no improvement in the real case for *Yld* with sugar beet observations. In this case, a high STICS model error for the simulation of sugar beet *Yld* (see results presented above for Fig. 7) can bias these results, which is not the case for the synthetic experiments.

## 4. Conclusions

The prediction of agro-environmental variables can be improved by estimating the soil parameters from observations on crops and the STICS model. The main results found by using the synthetic observations are confirmed by the results found by using real observations, revealing the value and complementary nature of the two studies. The synthetic study enables the effects of a large number of possible

observation sets to be explored, while the real study can validate the effects of a restricted number of observation sets, taking account of the model error. The estimation of the parameters related to the soil water content can be greatly improved and the prediction of the variables of interest is thus also affected because of their sensitivity to these parameters, particularly for yield and grain protein. As Braga (2004) and Varella (2008) showed in their studies, the quantity of information available on the parameters is different according to the type of the observation set. In this study, we showed that observation sets obtained on sugar beet, in a dry season or on a shallow soil, allow better parameter estimation and thus better predictions. For even drier weather conditions or shallower soils, better estimation and prediction could be expected.

The results of parameter estimation and variable prediction are closely linked to some hypotheses made in this paper. The crucial hypothesis concerns the knowledge of the initial conditions (*Hinit* and *NO3init*) and affects only the prediction quality. If the initial conditions are assumed to be unknown, the prediction error takes into account the error involved by fixing them at a nominal value (Sobol et al., 2007). The error can be thus reduced if the initial conditions values are determined by the measurements of water and mineral nitrogen content at the beginning of the crop simulation. When a given output variable is sensitive to the initial conditions, such as *Nit* or sugar beet *Yld*, its prediction error is even more reduced by making use of this information. For crop monitoring applications it is thus important to deal with practical and financial constraints for measuring the initial conditions.

The results of parameter estimation and variable prediction are also closely linked to the prior information used to estimate parameters. In this study, this information came from examining the distribution of the soil measurements made on a soil similar to that of the real database considered here. However, if time and financial constraints allow it, the prior information could also be provided by making a partial soil analysis at certain locations within the study area of the real database and using it to provide a mean and variance for soil parameter values which might lead to better parameter estimates. The results of the parameter estimation and the prediction could be even more improved if new information could be acquired on some parameters correlated with others. This may be done by measuring them and fixing them at the measured values, improving the prior information, or adding new

observations related to these parameters in the estimation process (Varella et al., 2008). The uncertainty of parameter estimation should be thus reduced and the performance of the prediction should be greatly improved.

From a practical point of view, the study shows how the observation set chosen can influence parameter estimation and prediction. If the user is able to characterize the observation set in terms of the crop observed, the weather pattern and soil depth, it should be possible to determine which parameters can be estimated and which variables of interest can be predicted with a given accuracy. Labor and costs for collecting the data could be minimized and the performance of the methodology could be maximized: the user should seek to collect the data when the soil and weather conditions are optimal, such as when a sugar beet crop is grown in a dry season.

**Acknowledgements:** The financial support provided by CNES and Arvalis-Institut du Végétal is gratefully acknowledged.

## References

- Baret, F., Houles, V., Guerif, M., 2006. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. Symposium on Imaging Techniques for Understanding Plant Responses to Stress held at the Society-for-Experimental-Biology Meeting, Canterbury, ENGLAND, pp. 869-880.
- 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 Bruyeres catchment. *European Journal of Agronomy* 29, 46-57.
- Beaudoin, N., Saad, J.K., Van Laethem, C., Machet, J.M., Maucorps, J., Mary, B., 2005. Nitrate leaching in intensive agriculture in Northern France: Effect of farming practices, soils and crop rotations. *Agriculture, Ecosystems & Environment* 111, 292-310.
- Beven, K., Binley, A., 1992. The Future of Distributed Models - Model Calibration and Uncertainty Prediction. *Hydrological Processes* 6, 279-298.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249, 11-29.

Blackmore, S., Moore, M., 1999. Remedial correction of yield map data. *Precision Agriculture* 1, 53-56.

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., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M.H., Bouthier, A., Ripoche, D., Mary, B., Juste, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie* 22, 69-92.

Calmon, M.A., Jones, J.W., Shinde, D., Specht, J.E., 1999. Estimating parameters for soil water balance model using adaptive simulated annealing. *Applied Engineering in Agriculture* 15, 703-7013.

Ferreira, R.A., Jones, J.W., Graham, W.D., 2006. Parameterizing Spatial Crop Models with Inverse Modeling: Sources of Error and Unexpected Results. *Transactions of the ASABE* 49, 1547-1561.

Flenet, F., Villon, P., Ruget, F.O., 2003. Methodology of adaptation of the STICS model to a new crop: spring linseed (*Linum usitatissimum*, L.). *STICS Workshop, Camargue, FRANCE*, pp. 367-381.

Gabrielle, B., Roche, R., Angas, P., Cantero-Martinez, C., Cosentino, L., Mantineo, M., Langensiepen, M., Henault, C., Laville, P., Nicoullaud, B., Gosse, G., 2002. A priori parameterisation of the CERES soil-crop models and tests against several European data sets. *Agronomie* 22, 119-132.

Gaucherel, C., Campillo, F., Misson, L., Guiot, J., Boreux, J.-J., 2008. Parameterization of a process-based tree-growth model: Comparison of optimization, MCMC and Particle Filtering algorithms. *Environmental Modelling & Software* 23, 1280-1288.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1995. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Golovko, L., Pozdnyakov, A.I., 2007. Electrical geophysical methods in agriculture. *Progress of Information Technology in Agriculture*, 457-471.



Gras, R., Erhwein, J., Hermant, G., Hebert, J., 1961. Carte des sols du département de l'Aisne. Laon n 5-6. Chambre d'agriculture de l'Aine.

Guérif, M., Beaudoin, N., Durr, C., Machet, J.M., Mary, B., Michot, D., Moulin, D., Nicoullaud, B., Richard, G., 2001. Designing a field experiment for assessing soil and crop spatial variability and defining site specific management strategies. Proceedings 3rd European Conference on Precision Agriculture, Montpellier, France, pp. 677-682.

Guérif, M., Houlès, V., Makowski, D., Lauvernet, C., 2006. Data assimilation and parameter estimation for precision agriculture using the crop model STICS. In: Wallach, D., Makowski, D., Jones, J.W. (Eds.), Working with Dynamic Crop Models. Elsevier, pp. 395-401.

Hadria, R., Khabba, S., Lahrouni, A., Duchemin, B., Chehbouni, A., Carriou, J., Ouzine, L., 2007. Calibration and validation of the STICS crop model for managing wheat irrigation in the semi-arid Marrakech/Al Haouzi plain. Arabian Journal for Science and Engineering 32, 87-101.

Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. Remote Sensing of Environment 112, 186-202.

Houlès, V., Mary, B., Guérif, M., Makowski, D., Juste, E., 2004. Evaluation of the crop model STICS to recommend nitrogen fertilization rates according to agro-environmental criteria. Agronomie 24, 1-9.

Irmak, A., Jones, J.W., Batchelor, W.D., Paz, J.O., 2001. Estimating Spatially Variable Soil Properties for Application of Crop Models in Precision Farming. Transactions of the ASAE 44, 1343-1353.

King, D., Daroussin, J., Tavernier, R., 1994. Development of a Soil Geographic Database from the Soil Map of the European Communities. Catena 21, 37-56.

Lagacherie, P., Baret, F., Feret, J.B., Netto, J.M., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment 112, 825-835.

Launay, M., Guérif, M., 2003. Ability for a model to predict crop production variability at the regional scale: an evaluation for sugar beet. Agronomie 23, 135-146.

Machet, J.M., Couturier, A., Beaudoin, N., 2007. Cartographie du rendement du blé et des caractéristiques qualitatives des grains. In: Guérif, M., King, D. (Eds.), Agriculture de précision. Quae, Versailles, pp. 131-144.



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

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

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

Moulin, S., Zurita, R.M., Guérif, M., 2007. Estimation de variables biophysiques du couvert par ajustement de modèles de transfert radiatif sur des réflectances. In: Guérif, M., King, D. (Eds.), *Agriculture de précision*. Quae, Versailles, pp. 97-113.

Murphy, B., Geeves, G., Miller, M., Summerell, G., Southwell, P., Rankin, M., 2003. The application of pedotransfer functions with existing soil maps to predict soil hydraulic properties for catchment-scale hydrologic and salinity modelling. *International Congress on Modelling and Simulation*, Townsville, AUSTRALIA, pp. 502-507.

Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research* 36, 3653-3662.

Ruget, F., Brisson, N., Delecolle, R., Faivre, R., 2002. Sensitivity analysis of a crop simulation model, STICS, in order to choose the main parameters to be estimated. *Agronomie* 22, 133-158.

Sobol, I.M., Tarantola, S., Gatelli, D., Kucherenko, S.S., Mauntz, W., 2007. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety* 92, 957-960.

Timlin, D., Pachepsky, Y., Walthall, C., Loechel, S., 2001. The use of a water budget model and yield maps to characterize water availability in a landscape. *Soil & Tillage Research* 58, 219-231.

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

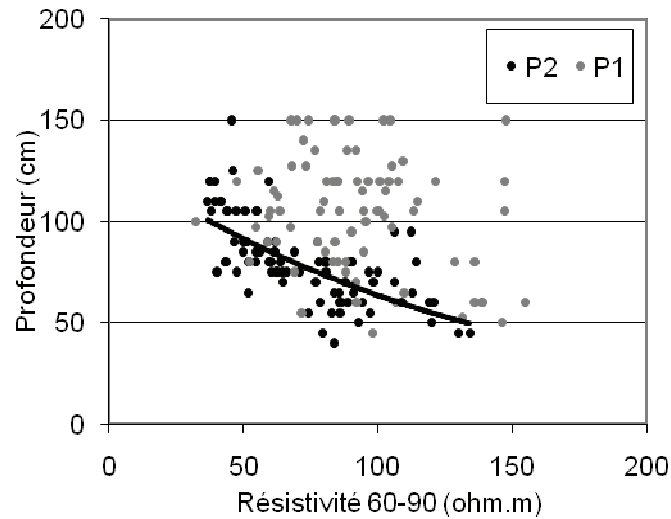
Varella, H., Guérif, M., Buis, S., 2008. Global Sensitivity Analysis (GSA) measures the quality of parameter estimation. Case of soil parameter estimation with a crop model., EGU, Vienna, Austria.

Weiss, M., Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. Remote Sensing of Environment 70, 293-306.

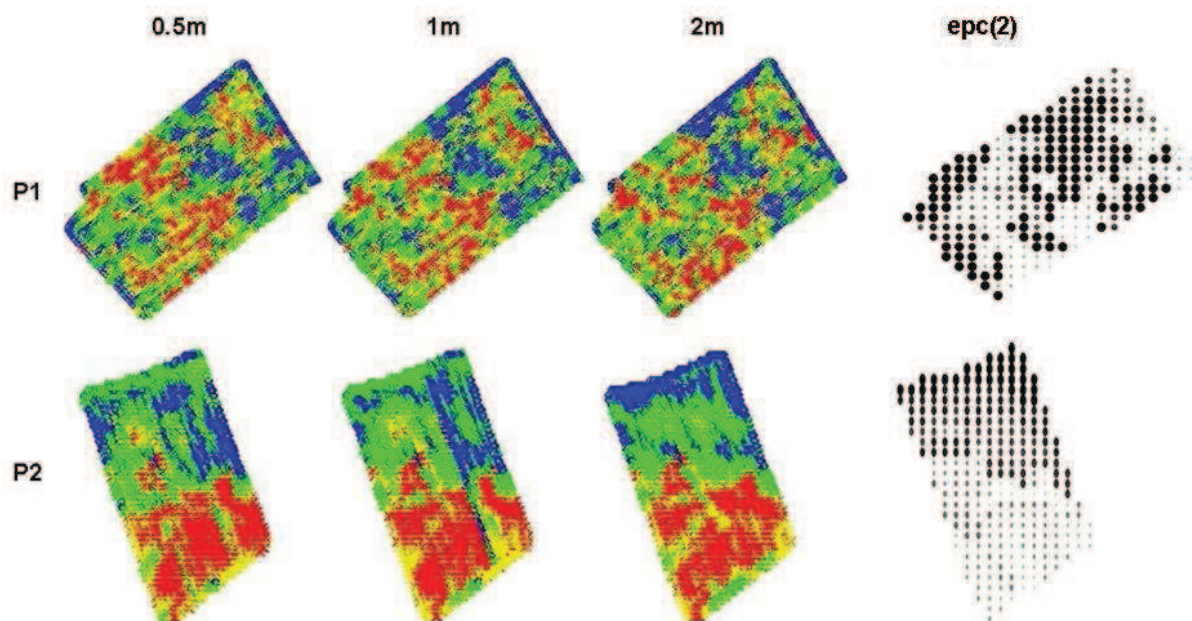
## Chapitre 6. Prise en compte d'une structure spatiale dans l'estimation des paramètres du sol

### 6.1. Position du problème et objectif

Jusqu'à présent, l'estimation des paramètres du sol a été réalisée en considérant chaque point comme indépendant de ses voisins. Or, nous avons vu au Chapitre 1.3.3 que les valeurs des paramètres sont structurées dans la parcelle, de manière plus ou moins forte. Nous avons également vu que la structure spatiale des propriétés des sols peut être étroitement liée à celle d'une mesure indirecte telle que la résistivité électrique. Par exemple, la carte de teneur en eau du sol (proche de *HCC*, en sortie d'hiver) est liée à la carte de résistivité électrique (Samouelian et al., 2005; Bourennane et al., 2007). Mertens et al. (2008) trouvent des relations entre conductivité électrique et profondeur du sol ; ces relations sont variables allant jusqu'à être de signe opposé selon les sols, en fonction de la nature du substratum. Pour des raisons peut-être analogues, dans le cas des sols des deux parcelles de Chambry (Figure 6-1), on constate que la distribution conjointe des points des mesures spatialisées de résistivité électrique de la couche de sol entre 60 et 90 cm en fonction de celles de la profondeur d'enracinement (proche du paramètre *epc(2)*), ne dégage aucune relation visible entre ces deux grandeurs. Cependant, lorsque l'on identifie les points correspondants à chaque parcelle, on note qu'il existe une relation significative pour la parcelle P2 (coefficient de corrélation égal à 0.48).



**Figure 6-1.** Relation entre la profondeur de sol et la résistivité électrique de la couche de sol entre 60 et 90 cm, pour les deux parcelles de Chambry.



**Figure 6-2.** Cartes de résistivité électrique des deux parcelles P1 et P2 pour les trois profondeurs de sol (**Figure 2-6**) et cartes de l'épaisseur de la seconde couche de sol, issues de l'étude pédologique des deux parcelles de Chambry et représentée par une résolution de 20m.

L'examen des cartes des valeurs de la résistivité électrique et du paramètre *epc(2)* (voir Figure 6-2), montre en effet qu'il est difficile de discerner une structure spatiale commune entre ces deux types de cartes pour la parcelle P1, alors que cela est plus frappant pour la parcelle P2 : pour cette parcelle, les hautes valeurs de

résistivité électrique (en rouge) correspondent à de faibles épaisseurs de sol, et vice-versa.

Si la résistivité électrique ne permet pas d'accéder directement aux propriétés des sols, elle est un bon indicateur de la structure spatiale des paramètres et doit pouvoir être prise en compte dans les méthodes d'estimation des paramètres. Ainsi Bourennane (2007) montre que la prise en compte de la structure spatiale de la résistivité électrique comme information secondaire permet d'améliorer l'interpolation spatiale par krigeage des valeurs mesurées de teneur en eau du sol. Il paraît donc judicieux dans notre démarche d'utiliser cette information sur la structure spatiale dans le processus d'estimation. Différentes approches ont été proposées pour utiliser une information additionnelle sur la structure spatiale dans le processus d'estimation, comme l'établissement de contraintes spatiales sur les valeurs estimées des paramètres (Gotzinger and Bardossy, 2007). Dans leurs travaux, Gotzinger et Bardossy utilisent des contraintes de Lipshitz et de monotonie liant, pour chaque couple de points, les valeurs de la variable aux valeurs des paramètres à estimer. Dans notre cas, le problème réside dans la définition de ces contraintes, qui sont difficiles à définir clairement et qui peuvent biaiser l'estimation spatialisée des paramètres du sol. Nous pouvons supposer, dans le processus d'estimation, que la carte des paramètres du sol est la réalisation d'un champ aléatoire (Cressie, 1993; Chilès and Delfiner, 1999), avec un bruit et une matrice de corrélation supposés connus, lui-même issu d'un champ de caractéristiques connues a priori. La prise en compte de la structure spatiale fournie par la résistivité électrique dans l'estimation des paramètres du sol est alors envisageable via l'identification d'une matrice de corrélation. Dans notre cas, il est donc nécessaire de considérer ce champ aléatoire conditionnel dans le calcul de la vraisemblance des observations afin d'estimer la carte des paramètres du sol.

Dans ce chapitre, nous présenterons une première approche méthodologique d'estimation spatialisée des paramètres du sol fondée sur ce principe, ainsi que son implémentation et nous la confronterons à une estimation où les points sont considérés comme spatialement indépendants les uns des autres.

## 6.2. Méthodes et données

Nous avons mis en œuvre deux types d'estimations :

- La première, considère les points spatialement indépendants les uns des autres et consiste en l'application de la méthode d'estimation utilisée dans les chapitres précédents à un ensemble de jeux de données observées sur les deux parcelles du dispositif de Chambry ;
- La deuxième, prend en compte l'information de la structure spatiale fournie par la résistivité électrique et consiste en une évolution de la méthode d'estimation précédente. Compte tenu du fort impact des erreurs d'observation sur le résultat de l'estimation précédente, nous avons procédé pour cette partie à des expériences sur des données synthétiques.

### 6.2.1. Estimation et prédiction sans prise en compte d'une structure spatiale sur les paramètres (observations réelles)

#### *Les observations du sol sur les deux parcelles de Chambry*

Les observations du sol, considérées ici pour valider les estimations des paramètres du sol et les prédictions des variables d'intérêt relatives au sol, sont celles faites sur les deux parcelles P1 et P2 de Chambry et qui ont été présentées au Chapitre 2.4.1. Ce sont des mesures des propriétés du sol effectuées sur un grand nombre de points de l'espace des deux parcelles, définis par les nœuds d'une grille espacés de 36 m.

#### *Les observations du couvert végétal sur les deux parcelles de Chambry*

Le Chapitre 2.4.2 recense le nombre d'observations du couvert végétal qui ont été faites sur les deux parcelles de Chambry sur différents types de support spatial. Nous utiliserons certaines d'entre elles pour estimer les paramètres du sol par inversion et d'autres pour valider la prédiction des variables d'intérêt à partir des valeurs de paramètres estimées. Le Tableau 6-1 montre quelles sont les observations que nous avons choisi d'utiliser pour l'estimation et celles choisies pour la prédiction, sur chacune des deux parcelles.

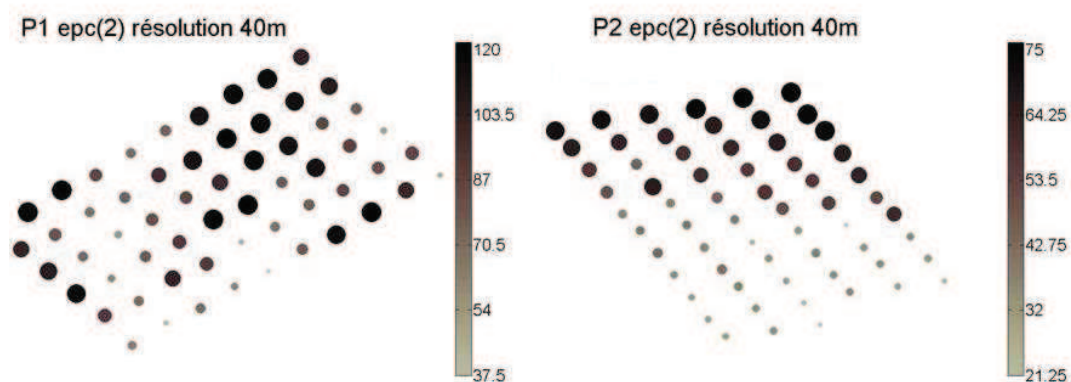
Parcelle P1				
Année de récolte	2000	2001	2002	2004
Culture	Blé	Betterave	Blé	Blé
Observations utilisées pour l'estimation	4 LAI 3 QN <i>Yld</i>	3 LAI <i>Yld</i>	Observations utilisées pour la prédiction <i>Yld</i> <i>Prot</i> <i>Nit</i>	<i>Yld</i>

Parcelle P2				
Année de récolte	2000	2001	2003	
Culture	Betterave	Blé	Blé	
Observations utilisées pour l'estimation	1 LAI <i>Yld</i>	7 LAI 4 QN <i>Yld</i>	Observations utilisées pour la prédiction	<i>Yld</i> <i>Prot</i> <i>Nit</i>

**Tableau 6-1.** Observations utilisées pour la procédure d'estimation des paramètres du sol et observations utilisées pour la prédiction des variables agroenvironnementales, pour les deux parcelles de Chambry.

#### Résolution des cartes d'observations et des valeurs estimées

Afin de réduire le coût en temps de calculs associé à l'estimation, nous avons choisi de travailler avec une résolution spatiale de 40 m. Pour cela, nous avons considéré pour l'ensemble des mesures disponibles, leurs valeurs agrégées spatialement au sein de mailles carrées de 40 m, comme présenté à la Figure 6-3.



**Figure 6-3.** Cartes de l'épaisseur de la seconde couche de sol représentées avec une résolution de 40 m.



### *Les paramètres du sol à estimer et information a priori*

Afin de simplifier davantage le problème d'estimation, nous avons décidé de réduire encore le nombre de paramètres à estimer, en utilisant les résultats obtenus dans les chapitres précédents. En effet, les paramètres permanents *argi* et *Norg* ainsi que la condition initiale *NO3init* étant difficiles à estimer à cause de la faible sensibilité des variables observables à ces paramètres, ils peuvent être fixés à une valeur nominale. Pour *Hinit*, nous le supposons connu sur toute la parcelle car sa faible variabilité intra-annuelle et intra-parcellaire, observée sur P1 et P2, permet de déterminer sa valeur sans trop d'erreur sur l'ensemble de la parcelle avec une seule mesure. Le nombre de paramètres passe de 7 à 3 paramètres : *epc(2)*, *HCC(1)* et *HCC(2)*. Les gammes de variations des paramètres qui définissent l'information a priori sont celles considérées jusqu'à présent : [1, 130] pour *epc(2)*, [14, 30] pour *HCC(1)* et [14, 30] pour *HCC(2)*.

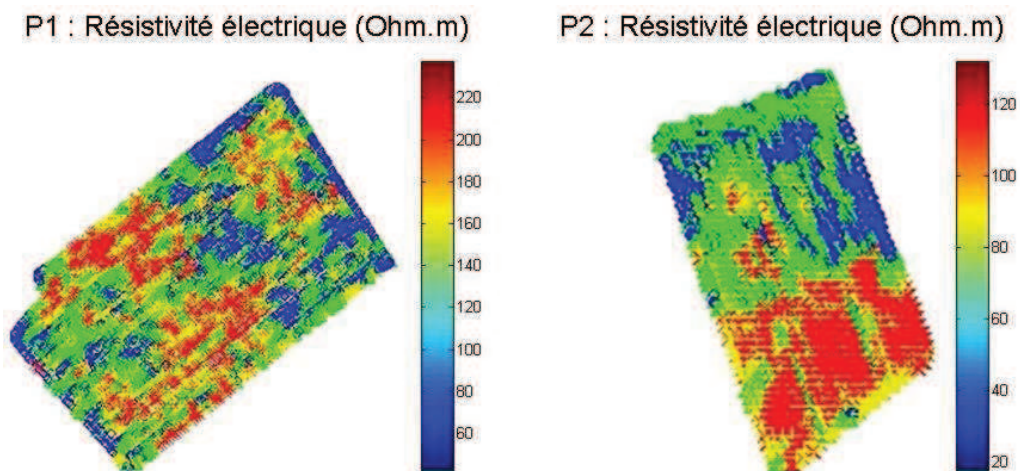
#### **6.2.2. Estimation avec prise en compte d'une structure spatiale sur les paramètres (observations synthétiques)**

Dans cette approche, nous utilisons la proximité entre les structures spatiales de la résistivité électrique et des propriétés des sols. Les observations des propriétés du sol sont les mêmes que dans le cas où aucune structure spatiale n'est prise en compte dans l'estimation. Mais ici, les observations des couverts végétaux considérées sont des observations synthétiques, obtenues à partir de simulations du modèle STICS, à la résolution spatiale de 40m. L'intérêt de ces observations synthétiques est de pouvoir tester seulement la prise en compte d'une structure spatiale dans la méthode d'estimation, tout en contrôlant l'erreur du modèle STICS (supposée nulle) et l'erreur d'observation (non biaisée et de variance connue).

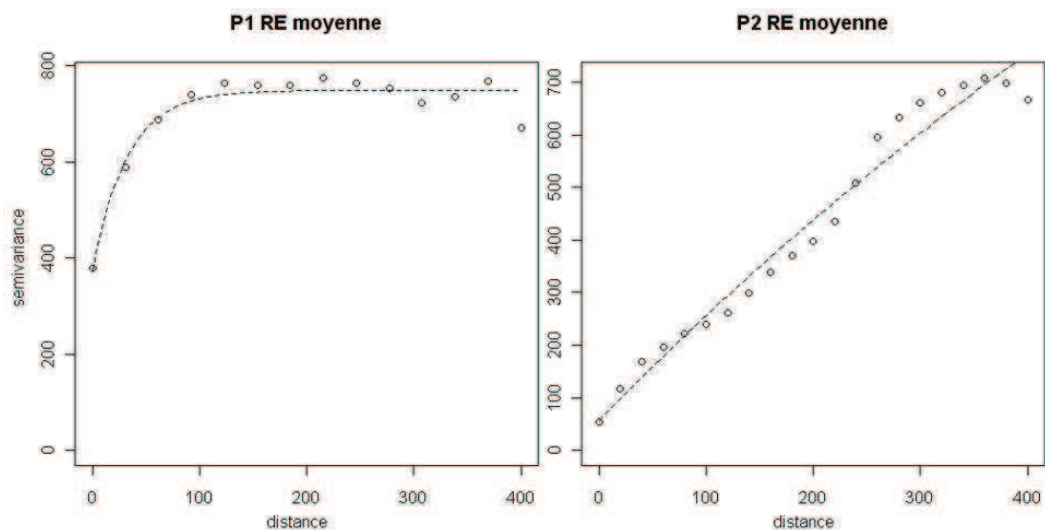
#### *Caractérisation de la structure spatiale de la résistivité électrique*

La Figure 6-2 représente les valeurs de résistivité électrique pour les trois profondeurs de sol (à savoir 0.5, 1 et 2 m) et pour les deux parcelles de Chambry. Pour chacune des parcelles, les structures spatiales de la résistivité électrique des trois profondeurs de sol sont assez similaires. Il est donc acceptable de moyenniser ces valeurs en une seule carte par parcelle dont on considère la structure spatiale

dans la procédure d'estimation des paramètres du sol. Pour chacune des deux parcelles on a calculé les valeurs moyennes de résistivité électrique (Figure 6-4) et le semi-variogramme de ces valeurs (Figure 6-5). Un variogramme (ou semi-variogramme) est une fonction utilisée en géostatistique qui permet de représenter la structure spatiale d'un champ de valeurs en estimant la variance (ou semi-variance) entre chaque couple de points de l'espace, en fonction de la distance qui sépare ces points (Cressie, 1993; Chilès and Delfiner, 1999).



**Figure 6-4.** Cartes des valeurs moyennes de résistivité électrique établies sur les deux parcelles P1 et P2.



**Figure 6-5.** Semi-variogrammes empirique (o) et ajusté (--) calculés sur les valeurs moyennes de résistivité électrique pour les deux parcelles P1 et P2.

Pour la parcelle P1, nous voyons que la variance se stabilise à partir d'une distance située en deçà de 100 m alors qu'elle ne semble pas se stabiliser pour la parcelle P2. Cela signifie que les valeurs de résistivité électrique présentent une structure spatiale beaucoup plus forte sur P2 que sur P1, avec un gradient, que l'on identifie en examinant les cartes comme étant globalement parallèle au bord long de la parcelle.

Afin de déterminer la matrice de corrélation spatiale qui pourra ensuite être prise en compte dans la méthodologie d'estimation des paramètres, un semi-variogramme exponentiel a été ajusté aux mesures de résistivité électrique (voir les courbes (--) de la Figure 6-5). Les expressions de ces fonctions théoriques ajustées sont alors les suivantes :

$$\begin{cases} \Gamma_{P1}(h) = 481[1 - \exp(-h/32)] + 285 \\ \Gamma_{P2}(h) = 2226[1 - \exp(-h/1065)] + 57 \end{cases}$$

où  $\Gamma$  est la semi-variance de la résistivité électrique pour la distance  $h$ . Parmi les valeurs ajustées, la portée  $\gamma$  indique la distance à partir de laquelle il n'y a – théoriquement – plus de variation de la variance. Pour la parcelle P1,  $\gamma$  est égale à 32 m alors qu'elle est égale à 1 065 m pour P2 (au delà de la plus grande dimension de la parcelle). A partir de ces semi-variogrammes, la matrice de corrélation spatiale  $C$  a pu être déterminée pour chacune des deux parcelles :

$$\begin{cases} C_{P1}(h) = \exp(-h/32) \\ C_{P2}(h) = \exp(-h/1065) \end{cases}$$

Ces expressions montrent bien que pour P1 la corrélation entre les valeurs de résistivité électrique décroît rapidement, en fonction de  $h$ , alors qu'elle décroît beaucoup plus lentement pour P2. La résolution spatiale des cartes (d'observations et des valeurs estimées) étant de 40 m, nous remarquons qu'elle est supérieure à la portée estimée sur P1, à savoir  $\gamma=32$  m. La prise en compte d'une structure spatiale dans l'estimation ne peut donc pas avoir d'effet sur cette parcelle. Nous choisissons alors de ne considérer que la parcelle P2 pour l'application de la méthodologie.

### Caractérisation de la structure spatiale des paramètres

L'amélioration de l'estimation des paramètres du sol grâce à l'utilisation de la structure spatiale de la résistivité électrique ne peut avoir lieu que sous l'hypothèse où la structure spatiale des paramètres du sol est similaire à celle de la résistivité électrique. Si ce n'était pas le cas, l'information apportée par la structure spatiale ne ferait que biaiser l'estimation des paramètres. Nous avons donc vérifié si, dans le cas de la parcelle P2, cette hypothèse est acceptable pour les paramètres  $epc(2)$ ,  $HCC(1)$  et  $HCC(2)$ . La Figure 6-6 montre les cartes de valeurs de ces paramètres et la Figure 6-7 les semi-variogrammes associés. Le semi-variogramme du paramètre  $epc(2)$  est strictement croissant et il est de ce fait assez similaire au semi-variogramme de la résistivité électrique ; celui de  $HCC(1)$  est lui aussi croissant même s'il admet un plateau situé entre de 150 et 200 m ; celui de  $HCC(2)$  est croissant jusqu'à une distance de 100 m puis ensuite décroissant. Dans ce cas, la structure spatiale de la résistivité électrique est étroitement liée à celle de  $epc(2)$  et dans une moindre mesure à  $HCC(1)$ .

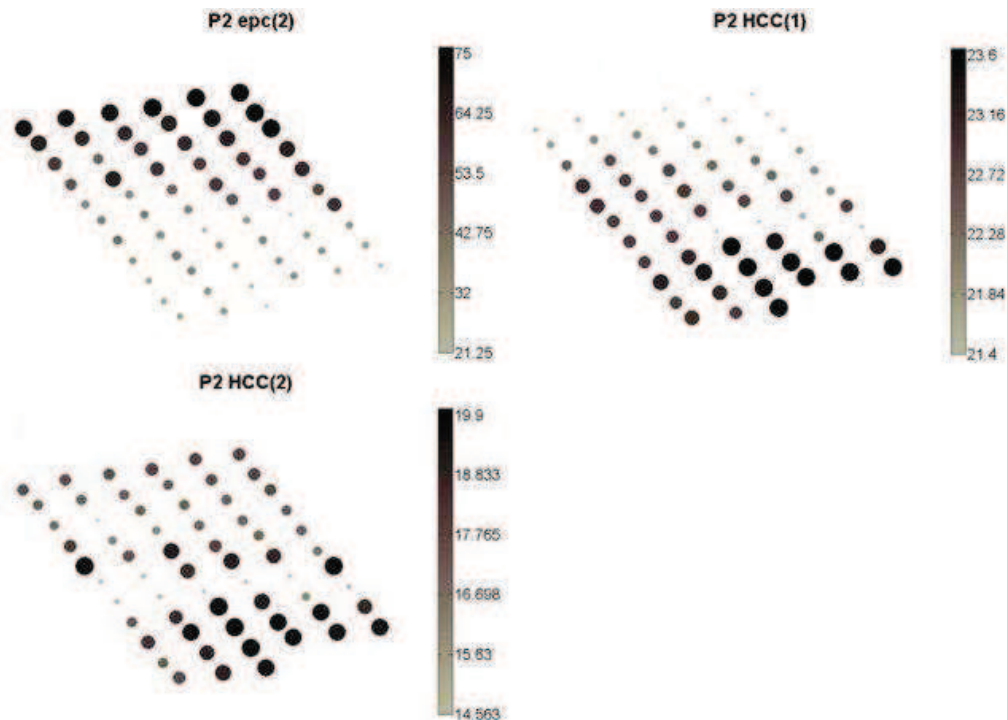
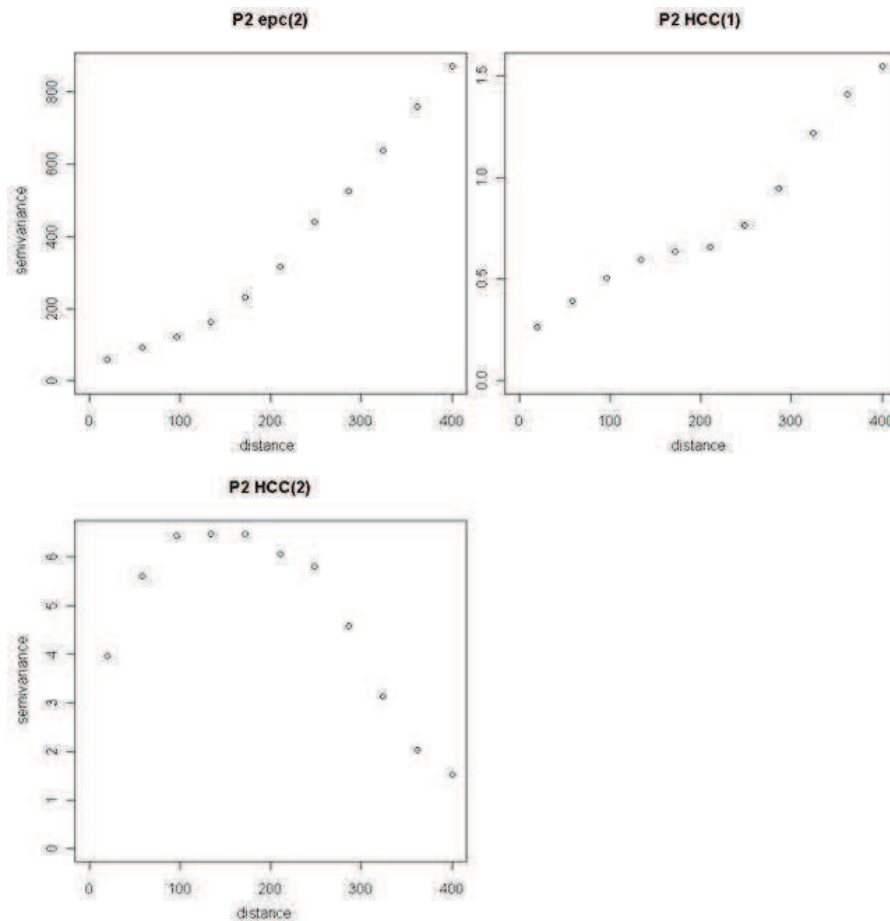


Figure 6-6. Cartes de valeurs mesurées des 3 paramètres du sol pour la parcelle P2.

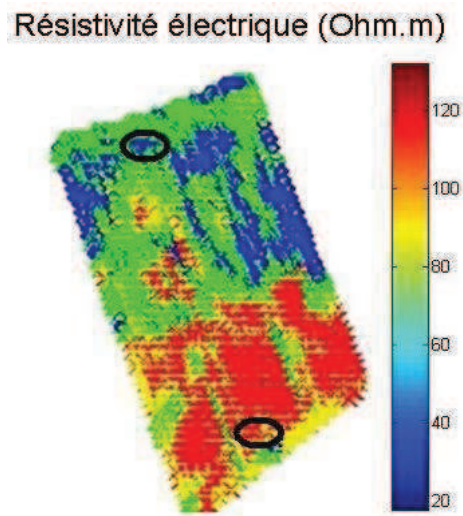


**Figure 6-7.** Semi-variogrammes empiriques des valeurs des 3 paramètres du sol pour la parcelle P2.

*Conséquences sur les paramètres du sol à estimer et information a priori*

La structure spatiale du paramètre *HCC(2)* étant plutôt différente de celle de la résistivité électrique, la prise en compte de la structure spatiale fournie par cette dernière pourrait biaiser son estimation. Il semble donc raisonnable de tester la méthode uniquement sur l'estimation des paramètres *epc(2)* et *HCC(1)*. Comme il existe une forte structure spatiale des valeurs de la résistivité électrique sur la parcelle P2, ordonnées selon un gradient (voir Figure 6-4), nous avons supposé que les valeurs des paramètres étaient elles aussi ordonnées selon ce même gradient. Cette hypothèse n'est pas contredite par la Figure 6-1 qui montre que la corrélation entre la résistivité électrique et le paramètre *epc(2)* est significative. Le fait que les valeurs soient ordonnées sur P2 selon un gradient nous permet d'améliorer de manière significative la précision de l'information a priori sur les paramètres. Grâce à ce gradient, nous considérerons qu'il est possible de mesurer les valeurs des paramètres *epc(2)* et *HCC(1)* en deux points seulement, les plus éloignés possible et le long d'un tracé le long de la parcelle. Nous pouvons fournir ainsi des bornes plus

précises des gammes de variation des paramètres. Dans le cas illustré à la Figure 6-8 où deux points sont figurés pour exemple, la gamme de variation valeurs des paramètres passerait de [1, 130] à [21.3, 70] pour  $epc(2)$  et de [14, 30] à [21.7, 23.6] pour  $HCC(1)$ . Au sein de l'algorithme d'estimation avec prise en compte d'une structure spatiale, le gradient permettra pour cette parcelle P2 de contraindre le tirage des cartes de paramètres candidates et de diminuer ainsi le temps de calcul de l'estimation.



**Figure 6-8.** Carte de résistivité électrique moyenne sur P2 avec deux points de mesures possibles pour les paramètres du sol.

#### *Les observations du couvert végétal sur la parcelle P2*

Conformément aux résultats obtenus dans le Chapitre 5, qui montrent que les meilleurs résultats en termes d'estimation des paramètres sont obtenus pour des conditions climatiques sèches et un couvert végétal de betterave, nous avons constitué un jeu d'observations synthétiques composé de 10 dates de  $LAI$  et du rendement  $Yld$  pour une culture de betterave en climat sec (celui considéré dans les Chapitre 3, 4 et 5). La création d'observations synthétiques, comme pour les Chapitre 4 et 5, est faite de la manière suivante :

$$y_{q,t} = f_{q,t}(\beta^{true}, x) + \varepsilon_{q,t} \quad (6-1)$$

où  $f_{q,t}$  est la variable de sortie de STICS  $q$  ( $LAI$  ou  $Yld$ ) calculée au temps  $t$  ( $t=1, \dots, 10$  pour  $LAI$  ou  $t$ =récolte pour  $Yld$ ),  $x$  est le vecteur des variables explicatives,  $\beta^{true}$  est le vecteur des valeurs mesurées des paramètres du sol et  $\varepsilon_{q,t} \sim N(0, \sigma_{q,t}^2)$  est le terme d'erreur d'observation dont l'écart-type respecte  $\sigma_{q,t} = \sigma_q^0 f_{q,t}(\beta^{true}, x)$ . Conformément à ce qui a été considéré aux Chapitre 4 et 5,  $\sigma_{LAI}^0 = 17\%$  et  $\sigma_{Yld}^0 = 9\%$ .

Dans la création des observations synthétiques, il aurait été plus correct de considérer une carte  $\beta^{true}$  qui aurait été la réalisation d'un champ aléatoire avec un bruit de covariance souhaitée (celle de la résistivité électrique dans notre cas). Cela aurait été plus fidèle à l'hypothèse sur la similitude des structures spatiales entre paramètres et résistivité électrique et aurait ainsi permis à la méthode d'estimation présentée ci-après d'être plus adaptée et donc plus performante. Dans ce travail, nous testerons toutefois les performances de la méthode en considérant une carte  $\beta^{true}$  non issue du champ aléatoire.

#### *Méthode d'estimation des paramètres du sol*

La prise en compte de la matrice de corrélation spatiale  $C$  dans la procédure d'estimation des paramètres se fait à travers l'écriture de la vraisemblance que nous détaillons ici.

Soient  $y_{q,t}(s)$  les observations de la variable au temps  $t$ ,  $t=1, \dots, T$  et au point de l'espace (ou site)  $s$ ,  $s=1, \dots, S$ . Les observations  $y_{q,t}(s)$  sont supposées indépendantes entre elles conditionnellement à la connaissance d'un champ aléatoire  $\beta(s)$  et des variables explicatives  $x$ . On suppose qu'elles ont une densité de probabilité conditionnelle gaussienne telle que :

$$P(y_{q,t}(s) / \beta(s)) = \prod_{t=1}^T \prod_{q=1}^Q \prod_s \frac{1}{\sqrt{2\pi\sigma_{q,t}}} \exp\left[-\frac{(y_{q,t}(s) - f_{q,t}(\beta(s), x))^2}{2\sigma_{q,t}^2}\right] \quad (6-2)$$



où  $f_{q,t}$  est la variable de sortie  $q$  du modèle STICS calculée au temps  $t$ , avec un écart-type  $\sigma_{q,t}$ .

$\beta(s)$  comprend les composantes  $\beta^1(s), \dots, \beta^P(s)$ ,  $P$  étant le nombre de paramètres, aléatoires positives et indépendantes entre elles. Pour cette raison de positivité, on prendra  $\beta(s)$  selon des lois log-normales, c'est-à-dire  $\beta^p(s) = \exp(\delta^p(s))$  où  $\delta^p(s)$  suit une loi gaussienne :

$$\delta^p(s) \sim N(\mu_{\delta^p}(s), \sigma_{\delta^p}^2 C) \quad (6-3)$$

$C$  est la matrice de corrélation spatiale (décrivant la structure spatiale de  $\delta$  et donc de  $\beta$ ) et  $\sigma_{\delta^p}^2$  est la variance commune de  $\delta^p(s)$ ,  $s = 1, \dots, S$ .

On a alors les relations suivantes entre les statistiques de  $\delta^p$  et celle de  $\beta^p$  :

$$\begin{cases} \mu_{\delta^p}(s) = \log(\mu_{\beta^p}(s)) + \frac{1}{2} \log\left(1 + \frac{\sigma_{\beta^p}^2}{\mu_{\beta^p}^2(s)}\right) \\ \sigma_{\delta^p}^2 = \log\left(1 + \frac{\sigma_{\beta^p}^2}{\mu_{\beta^p}^2(s)}\right) \end{cases} \quad (6-4)$$

où  $\mu_{\beta^p}(s)$  est l'espérance de  $\beta^p(s)$  et  $\sigma_{\beta^p}^2(s)$  est la variance de  $\beta^p(s)$ .

On suppose que la structure spatiale de  $\beta$  (et donc celle de  $\delta$ ) est liée à celle de la résistivité électrique par le biais de la corrélation  $C$ . Le but est d'estimer  $\beta^p(s)$  en considérant des informations a priori sur sa distribution spatiale. Pour cela, on part d'une loi a priori sur  $\delta^p(s)$  concernant  $\mu_{\delta^p}(s)$  et  $\sigma_{\delta^p}^2$  (et donc  $\mu_{\beta^p}(s)$  et  $\sigma_{\beta^p}^2$  par inversion du système (6-4)),  $C$  étant supposée connue. Dans le contexte de loi log-normale pour les  $\beta^p$ , il est raisonnable de supposer une loi a priori log-normale pour  $\mu_{\beta^p}(s)$  et donc une loi normale pour  $\mu_{\delta^p}(s)$ . Ces lois a priori sont supposées

indépendantes spatialement, c'est-à-dire que  $\mu_{\delta^p}(s) \sim N(m_{\delta^p}, v_{\delta^p}^2)$ , où  $m_{\delta^p}$  et  $v_{\delta^p}^2$  sont indépendants des sites. De plus, il semble également raisonnable de prendre  $\sigma_{\delta^p}^2 = v_{\delta^p}^2$  (déterministe). A ce stade, il suffit donc de connaître le champ  $\mu_{\delta^p}(s)$ , de caractéristiques connues a priori, pour déterminer  $\mu_{\beta^p}(s)$  et donc  $\beta^p(s)$ .

On souhaite estimer  $\mu_{\delta^p}(s)$  en maximisant sa loi a posteriori donnée par la formule de Bayes suivante :

$$P(\mu_{\delta^p}(s) / y_{q,t}(s)) = \frac{P(y_{q,t}(s) / \mu_{\delta^p}(s)) P(\mu_{\delta^p}(s))}{P(y_{q,t}(s))}$$

où  $P(\mu_{\delta^p}(s))$  est la loi a priori de  $\mu_{\delta^p}(s)$ ,  $P(y_{q,t}(s))$  est la loi marginale des observations et  $P(y_{q,t}(s) / \mu_{\delta^p}(s))$  peut écrire :

$$P(y_{q,t}(s) / \mu_{\delta^p}(s)) = \int_{\beta(1)} \dots \int_{\beta(s)} P(y_{q,t}(s) / \beta(s)) P(\beta(s) / \mu_{\delta^p}(s)) d\beta(1) \dots d\beta(s) \quad (6-5)$$

où  $P(y_{q,t}(s) / \beta(s))$  est la probabilité décrite à l'Equation (6-2) et  $P(\beta(s) / \mu_{\delta^p}(s))$  est déterminée à partir du système (6-4).

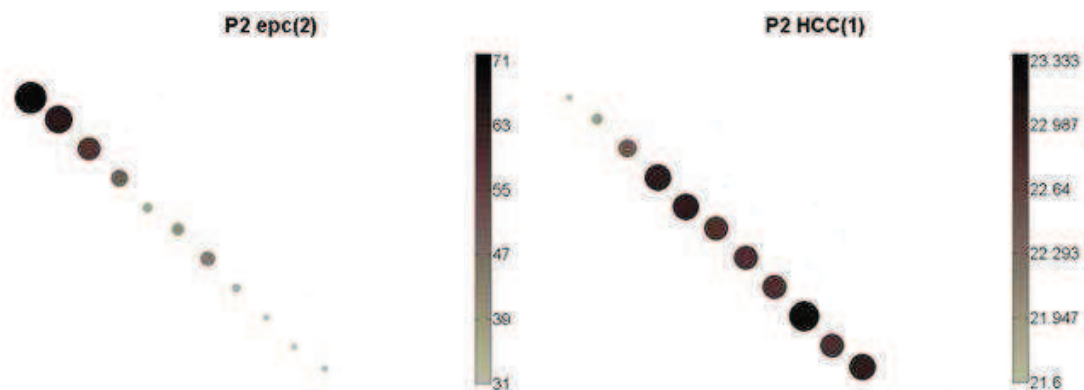
L'algorithme d'estimation de  $\mu_{\delta^p}(s)$  proposé est le suivant :

- (i) on tire  $N$  cartes de valeurs  $\mu_{\delta^p}(s)$  suivant la loi a priori  $P(\mu_{\delta^p}(s))$  telle que  $\mu_{\delta^p}(s) \sim N(m_{\delta^p}, v_{\delta^p}^2)$ , indépendamment du site,
- (ii) on tire  $M$  cartes (ou champs)  $\delta^p(s)$  suivant la loi  $\delta^p(s) \sim N(\mu_{\delta^p}(s), \sigma_{\delta^p}^2 C)$ , pour calculer l'intégrale (6-5),
- (iii) on calcule  $P(\beta(s) / \mu_{\delta^p}(s))$  à partir de la relation  $\beta^p(s) = \exp(\delta^p(s))$ , puis  $P(y_{q,t}(s) / \beta(s))$ ,
- (iv) on approche l'intégrale (6-5) par  $Q_n = \frac{1}{M} \sum_{m=1}^M P(y_{q,t}(s) / \beta(s)) P(\beta(s) / \mu_{\delta^p}(s))$
- (v) on estime  $\hat{\mu}_{\delta^p}(s)$  par le  $\mu_{\delta^p}(s)$  qui maximise  $Q_n$ .

Dans le cas de la parcelle P2, il a été plus simple de supposer que  $\mu_{\delta^p}(s)$  suit une loi a priori  $P(\mu_{\delta^p}(s))$  uniforme (au lieu de normale) telle que  $\mu_{\delta^p}(s) \sim U[a^p, b^p]$ , où  $a^p$  et  $b^p$  sont les bornes de la gamme de variation du paramètre  $p$ . De plus, nous avons supposé que les valeurs des paramètres étaient ordonnées dans la parcelle P2 selon un gradient le long de P2. Cela nous fait ajouter une étape intermédiaire au sein de l'algorithme, visant à contraindre les valeurs  $\mu_{\delta^p}(s)$ , tirées aléatoirement, selon le gradient. Cette étape se glisse entre les étapes (i) et (ii) et est telle que :

(i') on classe les valeurs  $\mu_{\delta^p}(s)$  tirées en (i) selon le gradient de la parcelle P2.

Le coût de la méthode d'estimation proposée ici est égal à  $M \times N \times S$  réalisations du modèle STICS et le temps d'une réalisation est compris entre 0.5 et 1 s (dépend de la puissance de l'ordinateur). Il n'y a pas de règle stricte concernant l'établissement des grandeurs  $M$  et  $N$ . Cependant, nous avons pu observer dans notre cas d'étude que  $M$  doit être au moins égal 10 000 et  $N$  à 20. Sachant que le nombre de points  $S$  de la parcelle est de 63 (voir Figure 6-6), le temps de calcul est au minimum égal à 73 jours. Afin de ne pas atteindre des temps de calculs si importants, nous proposons de réaliser l'estimation des paramètres sur un nombre restreint de pixels de la parcelle (constitué d'une seule rangée), comme le montre la Figure 6-9. Le nombre de pixels alors considéré est de  $S=11$ .



**Figure 6-9.** Cartes de valeurs mesurées des paramètres epc(2) et HCC(1) pour une seule rangée de points de la parcelle P2.

## Expériences réalisées

Diverses expériences ont été réalisées concernant :

1. La quantification de l'amélioration de l'estimation du couple de paramètres ( $epc(2)$ ,  $HCC(1)$ ) sans prise en compte de la structure spatiale lorsqu'une information a priori plus précise sur les paramètres  $epc(2)$  et  $HCC(1)$  est considérée. Cette information plus précise est déterminée grâce à l'observation de la structure spatiale de la résistivité électrique et à la réalisation de deux uniques mesures des propriétés des sols ;
  2. La comparaison des performances d'estimation du paramètre  $epc(2)$ , puis du couple de paramètres ( $epc(2)$ ,  $HCC(1)$ ), selon que la structure spatiale de la résistivité électrique est considérée dans l'estimation ou ne l'est pas. Cette comparaison est réalisée pour différents niveaux d'erreur sur les observations de  $LAI$  et de  $Yld$  ;
- L'analyse de l'apport relatif dans le processus d'estimation de l'information contenue dans le fait de ranger les valeurs, tirées aléatoirement, selon le gradient. Cette analyse sera entreprise dans le cas de l'estimation de  $epc(2)$  seul.

## 6.3. Résultats et discussion

### 6.3.1. Estimation et prédiction sans prise en compte d'une structure spatiale sur les paramètres (données réelles)

Rappelons que dans cette partie, l'estimation des paramètres du sol et la prédiction de variables agroenvironnementales sont entreprises avec des observations réelles de couverts végétaux et sans considérer de structure spatiale pour les paramètres. La Figure 6-10 montre les résultats de l'estimation des paramètres du sol pour la parcelle P1 et la Figure 6-11 montre les résultats de la prédiction des variables d'intérêt. Nous observons pour cette parcelle une sous-estimation de  $epc(2)$  accompagnée d'une surestimation de  $HCC(2)$  probablement due à des compensations entre ces deux paramètres. L'estimation de  $HCC(1)$  est largement surestimée et présente une forte variabilité, laquelle peut être expliquée par une information a priori trop large. En outre, la configuration d'estimation étant

constituée d'un sol plutôt profond et d'un climat globalement humide sur l'ensemble des années observées, la quantité d'information fournie par les observations sur les paramètres n'est pas la plus efficace pour estimer les paramètres (voir Chapitre 4 et 5). Même si nous disposons ici d'observations de betterave, plus efficaces pour estimer les paramètres, elles sont trop peu nombreuses pour permettre de bien estimer les paramètres. Cela explique en partie la mauvaise performance d'estimation. Même si  $epc(2)$  est sous-estimée, il semble que la surestimation de  $HCC(1)$  et de  $HCC(2)$  entraîne une surestimation de la prédiction du rendement. Pour la prédiction de la teneur en protéine (et du rendement dans une moindre mesure), nous retrouvons l'effet des trois différentes doses d'azote prescrites sur la parcelle P1 en 2002 (140, 200 et 260 kg d'azote), qui a conduit à des niveaux très contrastés de teneur en protéines (et de rendement dans une moindre mesure). La forte erreur du modèle concernant la simulation de la teneur en azote du sol (Houlès, 2004), accompagnée d'une faible sensibilité aux paramètres estimés (voir Chapitre 5), explique la mauvaise performance de prédiction pour cette variable. La Figure 6-12 montre les résultats de l'estimation des paramètres du sol pour la parcelle P2 et la Figure 6-13 montre les résultats de la prédiction des variables d'intérêt. Nous observons que l'estimation de tous les paramètres est surestimée. Comme pour P1, cela pourrait être expliqué par le fait que les informations a priori sont trop larges mais aussi par des observations admettant un fort biais positif, induisant la surestimation des paramètres. La mauvaise performance d'estimation peut également être expliquée par la configuration d'estimation, similaire à celle présente sur P1. Pour finir, la surestimation de tous les paramètres entraîne une surestimation du rendement.

Globalement, la mauvaise performance globale des prédictions sur P1 et P2 est étroitement liée à la mauvaise performance d'estimation des paramètres. Pour expliquer cela, un ensemble de causes peuvent être évoquées :

- La variance de l'erreur d'observation est sous-estimée et/ou la moyenne de l'erreur n'est pas nulle (observations biaisées). En effet, l'inversion des images de télédétection, visant à déterminer le  $LAI$  et le  $QN$ , peut donner des observations de mauvaise qualité notamment lorsque la valeur de l'épaisseur optique des aérosols

dans l'atmosphère, difficile à estimer, est mauvaise (Launay et al., 2000) ces erreurs peuvent dépasser largement les niveaux considérés ici ;

- Un biais du modèle STICS pour la simulation des variables observables, entraînant ainsi un biais dans l'inversion. Par exemple, parmi les processus liés au sol, nous avons choisi d'ignorer les remontées d'eau dans le sol par capillarité. Or ce processus peut avoir lieu dans certaines conditions sur les substrats crayeux. Dans ce cas, la simulation de la quantité d'eau du sol, traduite par l'état de croissance de la plante, est supérieure à celle réellement contenue dans l'épaisseur du sol mesurée. L'estimation de  $epc(2)$  à travers les observations de la croissance de la plante est alors surestimée (cas de P2) ;

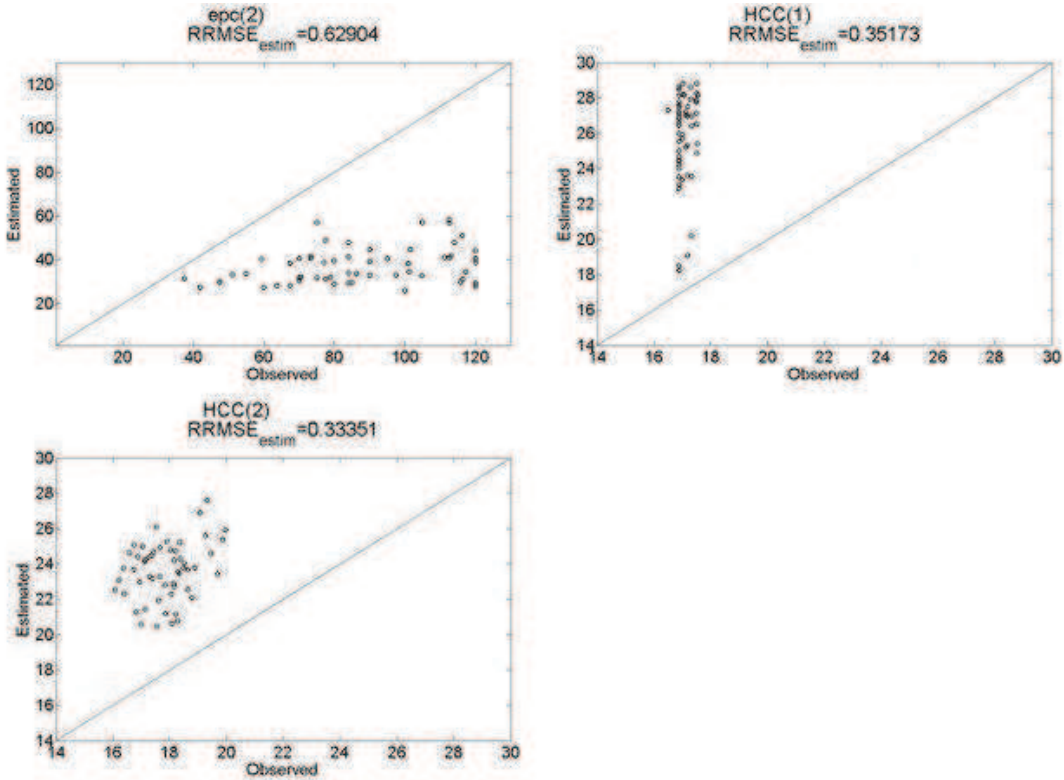
- L'erreur de mesure des valeurs des paramètres du sol qui peut être scindée en deux composantes. (i) Une erreur de mesure locale des paramètres, certainement le cas de  $epc(2)$  dont l'estimation est faite par le pédologue à partir de la profondeur à laquelle les racines lui semblent arrêter leur progression. Dans ce cas, la mesure est entachée d'une erreur pouvant être grande (la prise en compte de la racine la plus profonde peut conduire à une surestimation de  $epc(2)$ ) et dépendante du type de matériaux du sol (racines plus ou moins difficiles à observer selon le matériau). (ii) Une erreur liée à l'interpolation spatiale, qu'elle soit faite à partir de mesures sur des points de grille ou bien par zonage de la parcelle en unités de sol (définies par le pédologue) affectées à des valeurs issues de fonctions de pédotransfert locales;

- L'information a priori considérée sur les paramètres est trop imprécise. Cette information, qui est issue de l'union des gammes de mesures faites sur P1 et P2, est en effet trop large et mal centrée lorsque l'on considère chacune des deux parcelles séparément : cela est visible sur les Figure 6-10 et 6-12 où l'on voit par exemple que les observations de  $HCC(1)$  occupent la première moitié de la gamme pour P1 alors qu'elles occupent la seconde moitié pour P2 ;

- Les compensations entre les paramètres, notamment ceux qui concernent une même couche de sol, comme  $epc(2)$  et  $HCC(2)$  qui de plus jouent souvent dans les équations du modèle sous la forme de produit (permettant de calculer la capacité de rétention d'eau), pouvant conduire à un problème d'identifiabilité ;

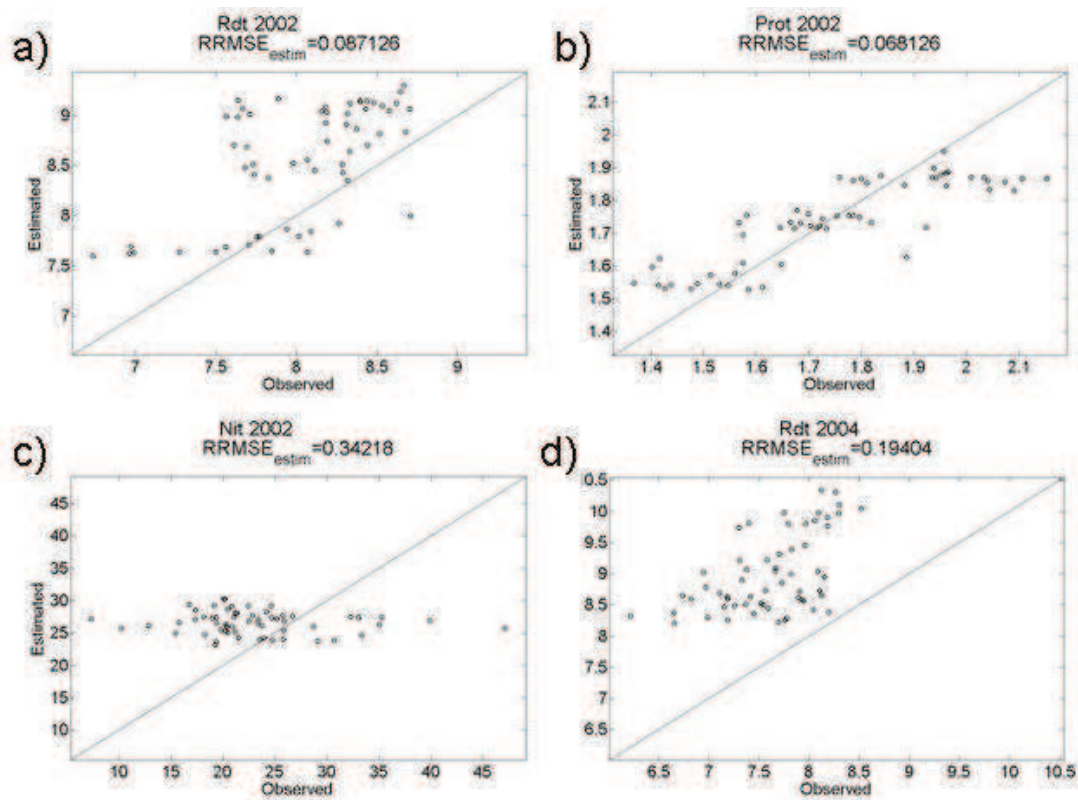
- Le biais d'omission des paramètres non estimés (voir Chapitre 1.1.2) ;

- Enfin la configuration d'estimation dans laquelle nous sommes placés (forte profondeur de sol et climat humide), dont nous avons vu dans les chapitres précédents qu'elle fournit peu d'information pour estimer les paramètres.

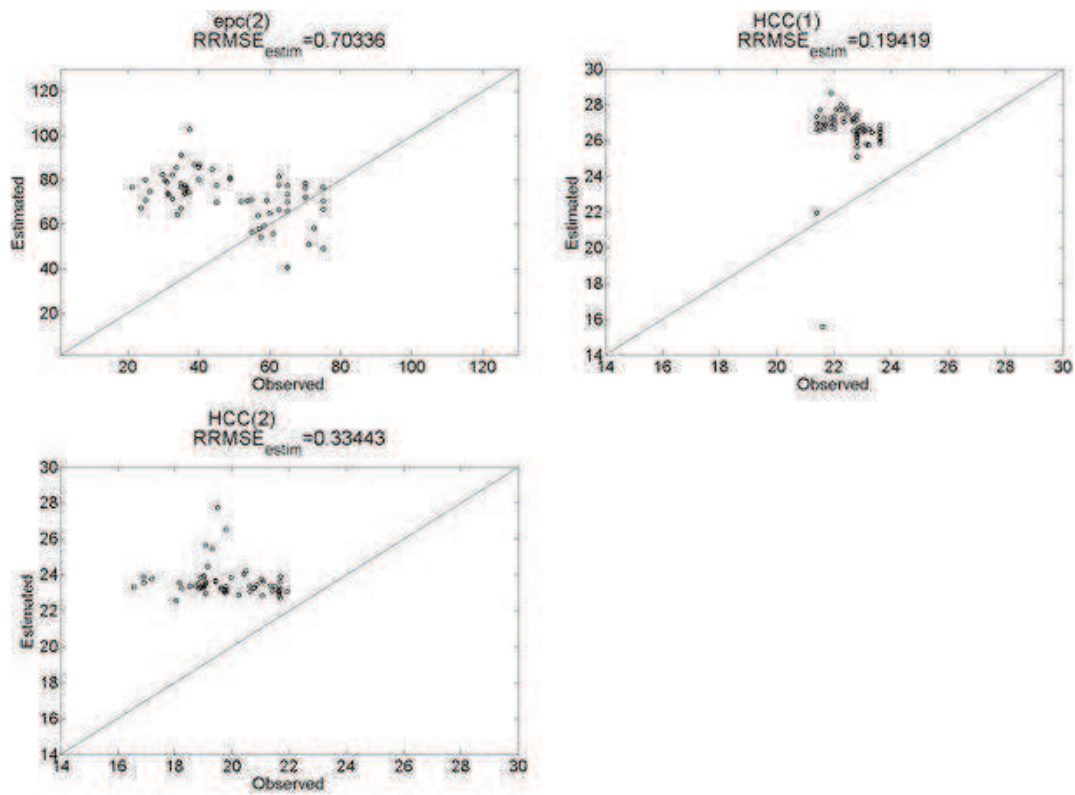


**Figure 6-10.** Comparaison entre les valeurs observées et les valeurs estimées des paramètres, pour la parcelle P1.

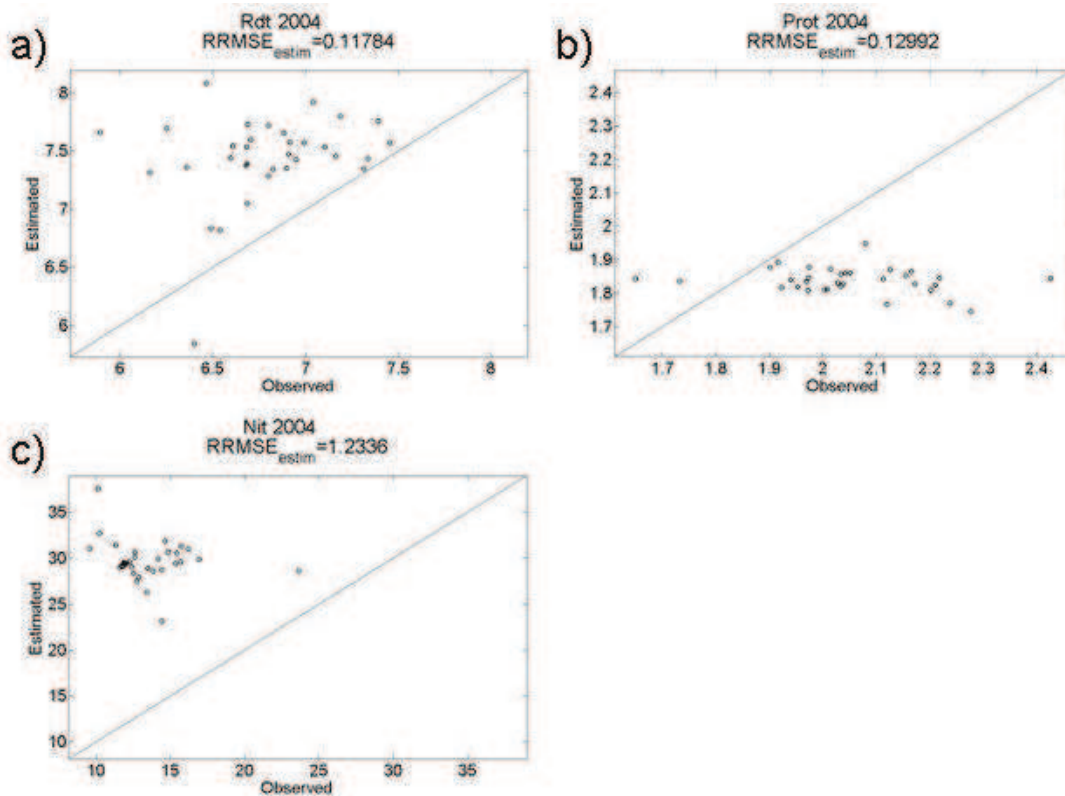




**Figure 6-11.** Comparaison entre les valeurs observées et prédites des variables : a) *Yld* du blé 2002, b) *Prot* du blé 2002, c) *Nit* du blé 2002 et d) *Yld* du blé 2004, pour la parcelle P1.



**Figure 6-12.** Comparaison entre les valeurs observées et les valeurs estimées des paramètres, pour la parcelle P2.



**Figure 6-13.** Comparaison entre les valeurs observées et prédites des variables: a) *Yld* du blé 2003, b) *Prot* du blé 2003 et c) *Nit* du blé 2003, pour la parcelle P2.

Il est important, pour bien prédire, d'avoir une estimation des paramètres du sol de qualité. Dans les configurations d'observation considérées et avec la somme des erreurs existantes (notamment liées aux observations et au modèle), les résultats obtenus montrent qu'il est difficile d'estimer les paramètres du sol. Nous allons essayer dans la partie suivante d'améliorer cette estimation en ajoutant une information supplémentaire sur la structure spatiale des paramètres, déduite de la carte des mesures de résistivité électrique.

### 6.3.2. Estimation avec prise en compte d'une structure spatiale sur les paramètres (données synthétiques)

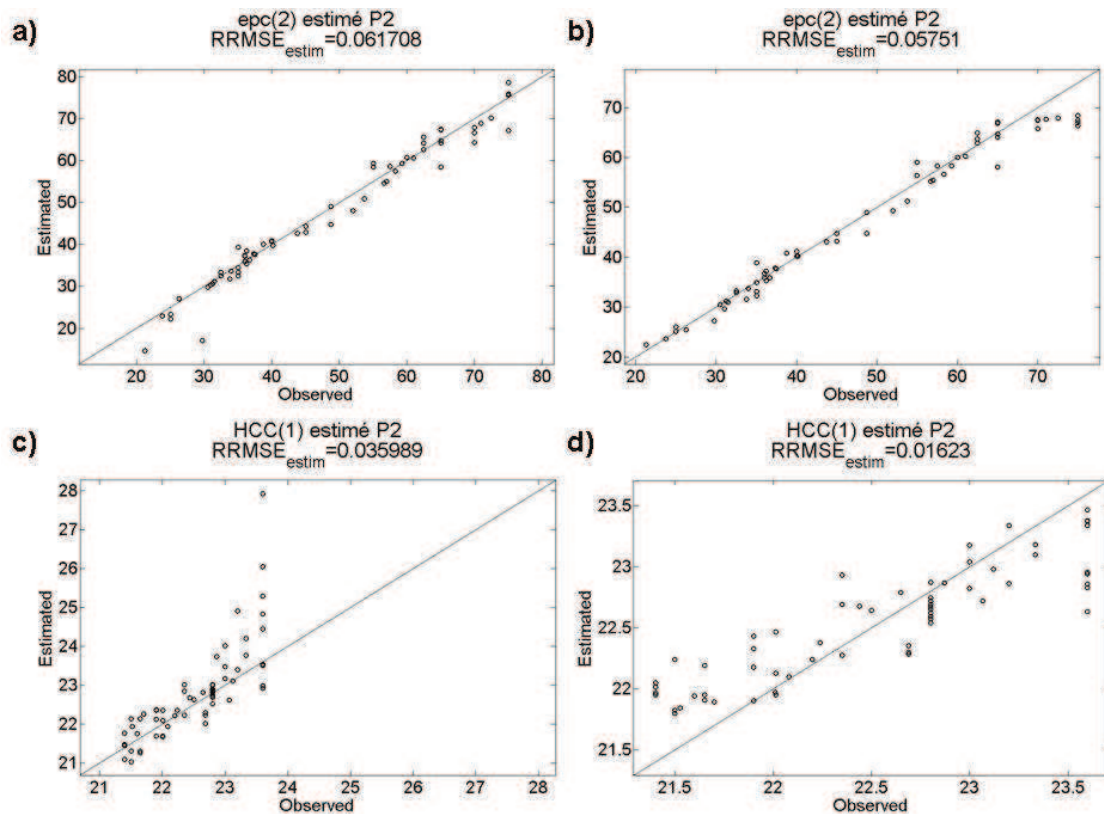
Pour éviter les problèmes de biais évoqués précédemment, nous travaillerons sur des observations synthétiques, en nous mettant dans les meilleures conditions en termes de quantité d'information disponible pour l'estimation des paramètres (observations sur culture de betterave en climat sec). Compte tenu de l'ampleur des coûts calculatoires, nous estimerons seulement les paramètres ayant une structure

spatiale liée à celle de la résistivité électrique :  $epc(2)$  et  $HCC(1)$ .- Les autres paramètres sont fixés à leurs valeurs mesurées et la parcelle P2 est réduite à une seule rangée de pixels carrés de résolution 40 m.

*Estimation sans structure spatiale (noSS) mais avec une information a priori plus précise : cas des paramètres  $epc(2)$  et  $HCC(1)$*

Dans un premier temps, nous allons quantifier l'amélioration de l'estimation des paramètres du sol grâce à l'établissement d'une information a priori plus précise sur les paramètres. Cette première étape ne considère pas la structure spatiale au sein du processus d'estimation. L'information a priori, nous le rappelons, est bâtie sur des gammes de variations dont les bornes ont été mesurées sur des points particuliers de la parcelle P2, eux-mêmes déterminés grâce à la structure spatiale de la résistivité électrique (voir la Figure 6-8). Nous comparerons cette information a priori, en terme de performance d'estimation, à celle utilisée jusqu'à présent et qui englobe les valeurs rencontrées sur les deux parcelles. La Figure 6-14 montre les performances d'estimation des paramètres  $epc(2)$  et  $HCC(1)$  en fonction des deux types d'information a priori. Contrairement aux cas d'étude précédents, l'estimation de ces paramètres apparaît très bonne même lorsque l'information a priori moins précise est utilisée. La raison principale est que, mis à part l'erreur d'observation, plus aucune autre source d'erreur n'est prise en compte dans cette expérience (notamment le biais des observations et l'erreur modèle). Ainsi, le fait que chaque paramètre soit associé à une couche de sol différente, chacune sollicitée à des moments différents durant la croissance de la plante, implique une sensibilité significative à ces paramètres – sans compensations possibles – permettant de bien les estimer. De ce fait, le rôle de l'information a priori utilisée semble minime et l'information utilisée jusqu'à présent semble être adaptée à l'échelle de la parcelle agricole pour estimer  $epc(2)$  et  $HCC(1)$  avec des observations synthétiques. Toutefois, l'utilisation de l'information a priori plus précise permet d'améliorer marginalement l'estimation : le  $RRMSE$  de  $epc(2)$  est amélioré de 6.8% et celui de  $HCC(1)$  de 54.9%. Le fait que l'amélioration soit plus grande pour  $HCC(1)$  peut être expliqué par une diminution plus conséquente de l'amplitude de la gamme de variation de ce paramètre : elle est diminuée de 87.8% alors que celle de  $epc(2)$  n'est diminuée que de 62.2%. Pour finir, nous dirons que, dans ces conditions d'estimation

très optimistes, l'établissement d'une information a priori plus précise, à partir de mesures sur un couple de points particulier de la parcelle déterminé grâce à la structure spatiale de la résistivité électrique, permet de gagner en performance d'estimation des paramètres du sol.

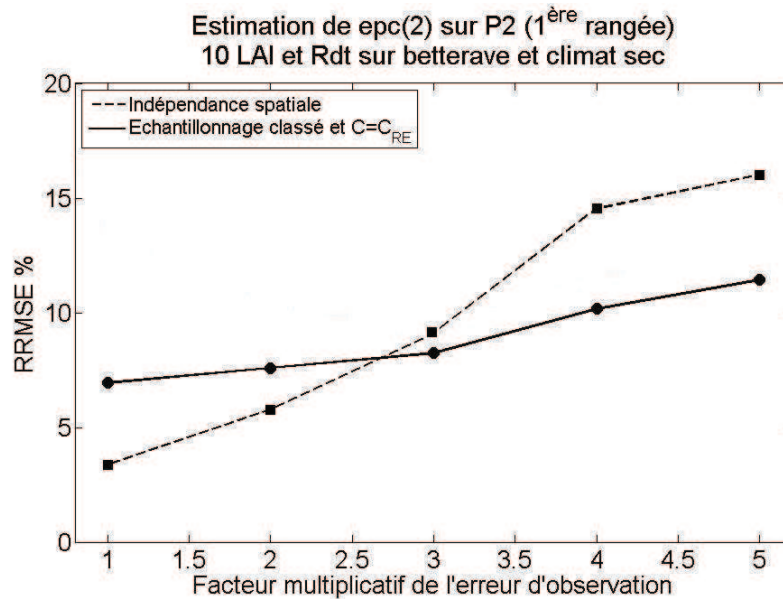


**Figure 6-14.** Comparaison entre les valeurs observées et les valeurs estimées des paramètres en fonction de l'information a priori, pour la parcelle P2. a) et c) concerne l'information a priori utilisée jusqu'à présent ; b) et d) concerne une information a priori plus précise.

#### *Estimation avec structure spatiale (SS) : cas du paramètre $epc(2)$ seul*

Nous avons estimé le paramètre  $epc(2)$  en prenant en compte la structure spatiale de la résistivité électrique dans le processus d'estimation (via sa matrice de corrélation  $C$ ), comme le propose l'algorithme présenté plus haut. Pour cela, nous avons fixé  $N = 10\,000$  et  $M = 20$ , conduisant à un nombre de réalisations du modèle STICS égal à  $10\,000 \times 20 \times 11 = 2\,200\,000$ . Nous comparerons les résultats obtenus avec ceux obtenus sans prise en compte d'une structure spatiale : dans ce cas, le nombre de tirages aléatoires effectués est de 2 000, nombre considéré comme largement suffisant pour n'estimer qu'un seul paramètre (rappelons que dans les chapitres précédents nous effectuions 10 000 tirages pour estimer 7 paramètres).

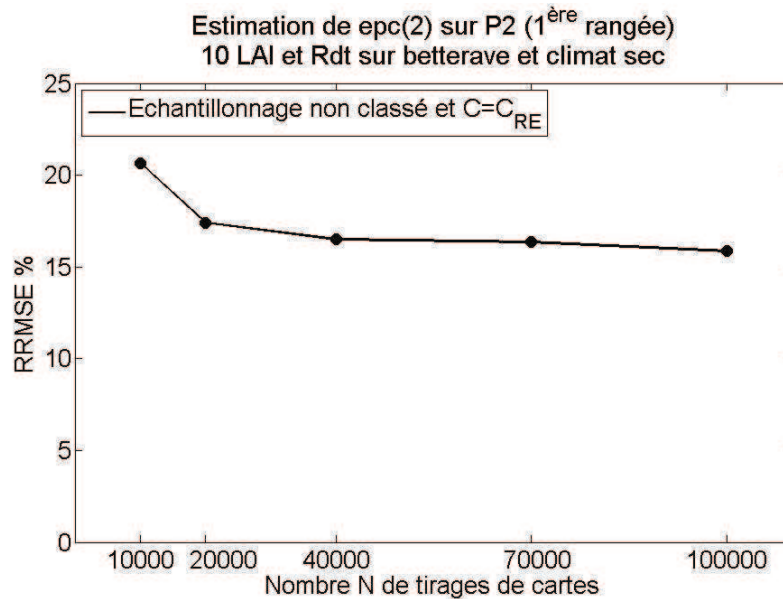
L'une des raisons évoquées pour expliquer les mauvaises performances réalisées lors de l'estimation des paramètres avec des observations réelles concernait le fait que la variance de l'erreur d'observations pouvait être plus grande que prévue. Nous avons donc testé l'algorithme avec des observations synthétiques admettant une variance d'erreur croissante. Pour cela, nous avons multiplié  $\sigma_q^0$  ( $\sigma_{LAI}^0 = 17\%$  et  $\sigma_{Yld}^0 = 9\%$ ) par un coefficient multiplicateur compris entre 1 et 5. La Figure 6-15 montre l'évolution du *RRMSE* en fonction de cette erreur d'observation. Lorsque l'estimation de *epc(2)* est faite sans prise en compte d'une structure spatiale, le *RRMSE* croît plus fortement en fonction de l'erreur d'observation (le *RRMSE* passe de 3.4% à 16%) que lorsque l'estimation prend en compte la structure spatiale de la résistivité électrique (le *RRMSE* passe alors de 6.9% à 11.4%). Les résultats montrent que pour un coefficient multiplicateur de l'erreur compris entre 1 et 2, l'estimation (SS) est moins bonne que l'estimation (*noSS*). Dans ce cas, le fait qu'il y ait peu de bruit dans les observations considérées implique comme on l'a vu une très bonne performance de l'estimation (*noSS*) et la prise en compte de la structure est moins performante. Il y a tout lieu de penser que si l'on avait pu considérer dans la création des observations une carte issue d'un champ, avec une covariance estimée à partir de la résistivité électrique, cela aurait introduit un bruit supplémentaire qui aurait probablement dégradé l'estimation (*noSS*) et permis à l'estimation avec prise en compte de la structure spatiale d'être plus performante. Pour un coefficient multiplicateur de l'erreur supérieur à 3, nous voyons que la prise en compte de la structure spatiale permet d'améliorer l'estimation. Dans ce cas, le fait que les observations soient de mauvaise qualité entraîne des valeurs aberrantes estimées par la méthode (*noSS*). Cela n'est pas le cas de l'estimation avec la méthode (SS) qui permet de lisser les valeurs estimées et de réduire ainsi l'erreur d'estimation.



**Figure 6-15.** Evolution du  $RRMSE$  en fonction d'un facteur multiplicatif de l'erreur d'observation. La courbe (--) concerne l'estimation sans prise en compte d'une structure spatiale (*noSS*), la courbe (–) concerne l'estimation avec un échantillonnage ordonné et une matrice de corrélation issue de la résistivité électrique (*SS*).

L'information contenue dans le fait de classer les valeurs tirées aléatoirement a été analysée afin de déterminer l'effet de ce classement dans le processus d'estimation : l'efficacité de l'étape (i) de l'algorithme sera ainsi testée. Pour ce faire, nous proposons de regarder les performances d'estimation de  $epc(2)$  obtenues sans ordonner les valeurs, en fonction du nombre  $N$  de tirages de cartes  $\theta(s)$ . Même pour un nombre de cartes tirées très élevé (ici jusqu'à  $N=100\ 000$ ), la Figure 6-16 montre qu'il est difficile d'estimer  $epc(2)$  correctement lorsque les valeurs tirées ne sont pas classées : le  $RRMSE$  est égal à 15.86% au minimum. Pour comparer, le fait de classer les valeurs permet, avec  $N=10\ 000$ , d'obtenir un  $RRMSE$  de 6.94%. Dans le cas de la parcelle P2, l'étape (i) de l'algorithme proposé s'avère être d'une grande efficacité pour réduire à la fois le temps de calcul d'estimation du paramètre  $epc(2)$  et l'erreur d'estimation.





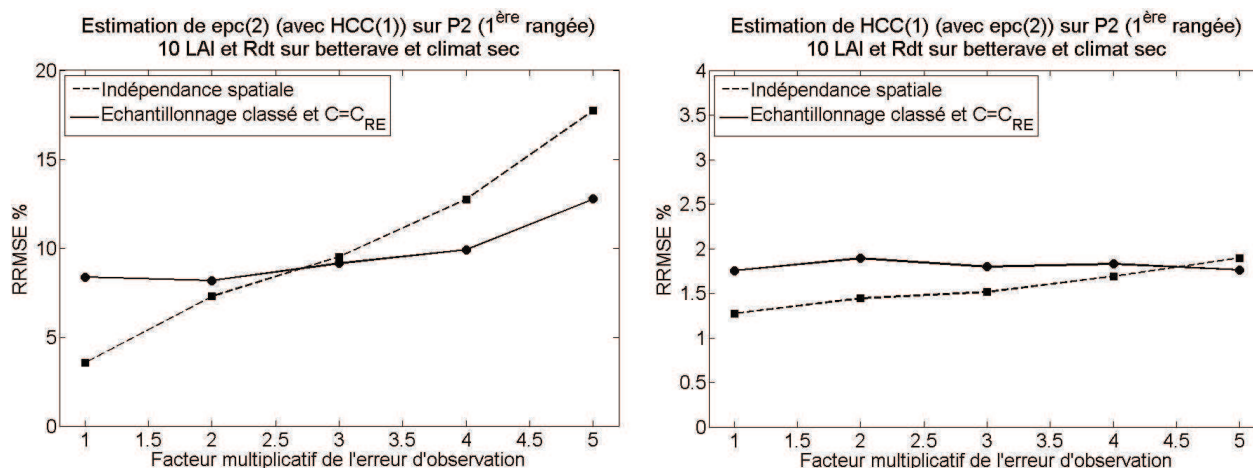
**Figure 6-16.** Evolution du  $RRMSE$  en fonction du nombre  $N$  de tirages de cartes, lorsque les valeurs tirées ne sont pas classées et pour une erreur d'observation standard.

#### *Estimation avec structure spatiale : cas des paramètres $epc(2)$ et $HCC(1)$*

A présent, nous proposons d'étendre l'exercice de comparaison des méthodes à l'estimation conjointe de  $epc(2)$  et  $HCC(1)$ . Le nombre de tirages aléatoires effectués pour estimer les paramètres sans structure spatiale est toujours de 2 000. Pour l'estimation avec structure spatiale, nous avons fixé  $N=20\ 000$  et  $M=20$ , menant à un nombre de réalisations de STICS égal à  $20\ 000 \times 20 \times 11 = 4\ 400\ 000$ . En ce qui concerne l'estimation de  $epc(2)$ , nous retrouvons globalement les mêmes conclusions que lorsqu'il était estimé seul, avec cependant un niveau de  $RRMSE$  légèrement plus élevé pour les deux méthodes dû à l'augmentation du nombre de paramètres à estimer qui passe de 1 à 2. Concernant  $HCC(1)$ , nous voyons que la prise en compte de la structure spatiale n'améliore pas l'estimation du paramètre avant que le coefficient multiplicateur atteigne à 5. Nous avons vu précédemment (Figure 6-7) que pour  $HCC(1)$  l'hypothèse sur la similarité de sa structure spatiale avec celle de la résistivité électrique n'est pas aussi évidente que pour le paramètre  $epc(2)$  : le plateau présenté par le variogramme de  $HCC(1)$  entre 150 et 200 m fait preuve d'une légère différence structurale, remettant en cause l'hypothèse. Aussi, le fait de classer les valeurs de  $HCC(1)$ , tirées aléatoirement, selon le gradient de la parcelle n'est pas totalement en adéquation avec les valeurs mesurées du paramètre (voir Figure 6-9), ce qui est probablement perturbateur pour l'estimation du



paramètre. Ces deux derniers points, soulevés pour  $HCC(1)$ , pourraient être des causes supplémentaires à la moins bonne performance de l'estimation avec prise en compte de la structure spatiale.



**Figure 6-17.** Evolution du  $RRMSE$  en fonction d'un multiple de l'erreur d'observation. La courbe -- concerne l'estimation sans prise en compte d'une structure spatiale et la courbe – concerne l'estimation avec un échantillonnage ordonné et une matrice de corrélation issue de la résistivité électrique.

## 6.4. Conclusion

Nous avons vu à quel point l'estimation des paramètres du sol avec des observations réelles s'est révélée être difficile sur les deux parcelles de Chambry et combien les sources d'erreur et de biais pouvaient être nombreuses, notamment celles qui sont liées aux observations et au modèle. Dans ces conditions de mauvaises performances, il aurait été intéressant de tester l'apport de la prise en compte d'une structure spatiale dans l'estimation des paramètres du sol. Mais nous avons estimé que trop de conditions étaient défavorables, en particulier au niveau de la quantité d'information disponible dans les observations (trop peu d'observations de betterave, climat humide). Nous avons alors préféré l'utilisation d'observations synthétiques.

Les résultats de ce chapitre montrent que la prise en compte d'une structure spatiale dans l'estimation des paramètres du sol permet d'améliorer la performance d'estimation lorsque les observations sont de mauvaise qualité. A contrario,

l'estimation avec prise en compte de la structure (SS) est moins performante que l'estimation sans prise en compte de la structure (noSS) lorsque les observations sont de qualité raisonnable et que l'on n'estime qu'un seul paramètre.

L'amélioration des performances d'estimation spatialisée des paramètres semble être liée à la véracité de l'hypothèse sur la similitude des structures spatiales des paramètres et de la résistivité électrique. L'extension à l'estimation de plusieurs paramètres pose de façon accrue le problème de la validité de cette hypothèse. En effet, nous avons vu qu'elle est vraie pour certains paramètres (cas de *epc(2)*), moins pour d'autres (cas de *HCC(1)*) et fautive pour d'autres encore (cas de *HCC(2)*) ; dans les cas où elle n'est pas valide, la méthode (SS) peut difficilement améliorer l'estimation du paramètre. Nous pouvons alors imaginer une procédure d'estimation des paramètres du sol en deux étapes. La première consisterait à estimer, sans prendre en compte de structure spatiale, les paramètres dont la structure n'est pas liée à celle de la résistivité électrique, et la seconde étape consisterait à estimer, en prenant en compte la structure spatiale, les paramètres dont la structure est liée à celle de la résistivité électrique. On pourrait aussi envisager d'incorporer plusieurs informations de structures spatiales, pour chaque type de paramètre, en supposant qu'elles soient accessibles : on peut par exemple penser à des images de télédétection sur sol nu pour caractériser la structure spatiale de l'horizon de surface.

Afin de tester plus rigoureusement l'apport de la prise en compte de la structure spatiale dans l'estimation, il faudrait être plus fidèle à l'hypothèse sur la similitude des structures spatiales entre paramètres et résistivité électrique. Pour cela, nous devrions considérer des observations synthétiques construites à partir de la carte des paramètres du sol issues de la réalisation d'un champ aléatoire de covariance connue (celle de la résistivité électrique dans notre cas). Cette considération pourrait permettre à la méthode de révéler son réel potentiel et d'être plus performante que l'estimation sans prise en compte de la structure, même avec des observations de qualité raisonnable.

Dans notre cas d'application à la parcelle P2 nous nous sommes aidés d'une contrainte qui consistait à ordonner le résultat du tirage des valeurs selon le gradient

observé de la parcelle et qui nous a permis d'augmenter grandement la performance d'estimation tout en considérant un nombre acceptable de réalisations du modèle STICS. Toutefois, l'utilisation de cette information supplémentaire est propre à la parcelle P2. Pour supprimer cette contrainte et rendre notre algorithme généralisable à tout type de parcelle, il faudrait tester, sur un modèle plus économique que STICS, l'effet du nombre de réalisations du modèle sur la performance d'estimation lorsque les valeurs ne sont pas contraintes. L'utilisation d'un modèle plus économique permettrait également de prendre en compte l'ensemble des pixels de la parcelle, ensemble sur lequel la structure spatiale issues de la résistivité électrique a été établie : les paramètres du sol pourrait alors être estimés sur le même support que celui de la structure.

Même si la méthodologie proposée dans ce chapitre n'est pas généralisable à tout type de structure spatiale et demande à être confortée, nous avons vu avec l'exemple de la parcelle P2 qu'elle est utile pour améliorer les estimations des paramètres du sol avec des observations peu précises et lorsque la parcelle présente un gradient des valeurs de paramètres. Les observations peu précises concernent souvent le cas des observations issues des images de télédétection : elles sont effectivement de mauvaise qualité lorsque les images sont inversées en *LAI* avec une valeur d'épaisseur optique imprécise. En ce sens, l'utilisation d'une information extérieure, telle que la résistivité électrique, semble porteuse d'une grande utilité dans le processus d'estimation des paramètres du sol par inversion.

Pour finir, nous avons vu qu'un certain nombre de pistes étaient tracées pour poursuivre ce travail et proposer ainsi une méthode robuste d'estimation de paramètres avec prise en compte d'une structure spatiale.

## Conclusion générale

### *Précision et l'amélioration de l'estimation des paramètres du sol*

L'objectif principal de la thèse consistait à savoir dans quelle mesure et dans quelles conditions il était possible d'estimer les paramètres du sol par inversion du modèle STICS avec des observations du couvert végétal. Avant d'obtenir une réponse à cette question, nous avons vu qu'il était primordial de commencer par sélectionner les paramètres qu'il était possible d'estimer avec ces observations, afin d'éviter les problèmes sous-jacents à l'estimation d'un grand nombre de paramètres (voir le Chapitre 1). Ce premier travail nous a permis d'aboutir à un sous-ensemble de paramètres du sol à estimer composé de paramètres impliqués dans le bilan hydrique du sol et les processus de minéralisation de la matière organique. Les différentes expériences d'estimation entreprises nous ont montré que certains paramètres étaient mieux estimables que d'autres et que cette performance dépendait des différents jeux d'observations disponibles (et de leur qualité). Les observations synthétiques sur le blé d'hiver montrent principalement que les paramètres liés à la capacité de rétention de l'eau  $HCC(1)$  et  $HCC(2)$  ont une bonne précision d'estimation avec un  $RRMSE$  inférieure à 20% (voir Chapitre 4.3). Nous avons ensuite vu, toujours sur données synthétiques, que lorsqu'on dispose d'observations sur des cultures d'été, comme la betterave, il était possible d'obtenir une précision encore meilleure des paramètres avec notamment  $HCC(1)$  et  $HCC(2)$  qui passent respectivement en dessous de 8% et de 16% d'erreur. Cela a été confirmé avec les observations réelles de betterave sur le bassin de Bruyères, où nous avons pu effectivement estimer  $HCC(1)$  avec une précision d'erreur inférieure à 8% (voir Chapitre 5.3). Quant aux observations à l'échelle intra-parcellaire sur Chambry, nous avons vu qu'il n'était pas possible d'estimer les paramètres avec une précision acceptable (voir Chapitre 6.3.1). Dans ce dernier cas de figure, il serait notamment possible qu'une mauvaise qualité des observations recueillies expliquerait en grande partie la mauvaise performance d'estimation. Néanmoins, lorsque les observations sont de qualité satisfaisante, nous pouvons dire que ce sont principalement les paramètres liés à la capacité de rétention en eau du sol, avec notamment  $HCC(1)$  et  $HCC(2)$ , qui sont les plus précisément estimés.

Nous avons montré également que l'amélioration de l'estimation que l'on peut obtenir, par rapport à l'estimation issue de l'information a priori dont on dispose, dépend fortement des configurations dans lesquelles les observations du couvert végétal ont été recueillies, et particulièrement du type de culture observé, de profondeur de sol et de climat. En effet, nous avons vu au Chapitre 5.3 que l'estimation des paramètres  $HCC(1)$  et  $HCC(2)$  était grandement améliorée par le fait de considérer des observations de betterave au lieu de blé : le critère  $RE$  passe d'environ 0.8 pour  $HCC(1)$  et  $HCC(2)$  à 0.3 pour  $HCC(1)$  et 0.65 pour  $HCC(2)$ . Le type de profondeur de sol permettait une meilleure amélioration du paramètre  $epc(2)$  : son critère  $RE$  passe de 0.86 (sol profond) à un  $RE$  de 0.49 (sol peu profond), dans l'exemple d'observations de betterave. Quant climat, il a un effet important notamment sur l'amélioration de  $Hinit$ , qui voit passer son  $RE$  de 0.71 (climat humide) à 0.41 (climat sec) lorsque des observations de blé sont considérées. La dépendance du niveau d'amélioration à la configuration s'explique par le fait que les différents jeux d'observations fournissent des quantités d'information différentes sur les paramètres à estimer, s'exprimant par des sensibilités différentes des variables observées aux paramètres du sol, et conduisant à des améliorations d'estimation différentes (voir Chapitre 4) : nous avons proposé un critère basé sur les résultats de l'analyse de sensibilité qui permettait d'expliquer cette amélioration. Finalement, nous avons constaté que les observations de betterave étaient celles qui permettaient la meilleure amélioration d'estimation, suivies des observations en climat sec et des observations sur sol peu profond. Nous constatons que le nombre de variables et le nombre d'années observées ont un rôle secondaire dans l'amélioration des estimations.

Que ce soit en termes de précision d'estimation ou en termes d'amélioration, les résultats obtenus sont liés à l'information a priori considérée dans la méthode d'estimation. En particulier, il semblerait que ces résultats soient différents selon que l'on considère une information a priori plus ou moins précise sur les paramètres, précision définie par la largeur et le positionnement de l'information autour de la valeur à estimer. Dans notre cas, nous avons considéré une distribution uniforme dont les valeurs extrêmes étaient celles observées sur l'ensemble des mesures

réalisées sur les deux parcelles. Sans possibilité de mesure sur le terrain, on pourrait imaginer une information a priori moins précise, comme celle fournie par une la carte pédologique au 1/250 000 du programme national IGCS. Mais l'on pourrait également, dans le cadre de l'agriculture de précision, imaginer disposer d'une information plus précise, comme nous l'avons tenté au Chapitre 6, en utilisant par exemple les cartes de résistivité électrique pour réaliser un zonage et pratiquer des mesures raisonnées selon ce zonage. Nous avons vu dans ce cas particulier, où nous n'estimions que deux paramètres avec une erreur d'observation et une erreur modèle parfaitement contrôlée, que l'amélioration était marginale. Toutefois, une analyse plus complète de l'effet de différentes précisions d'information a priori sur l'estimation des paramètres mériterait d'être entreprise.

#### *L'amélioration des variables agro-environnementales d'intérêt*

La prédiction des variables d'intérêt peut être améliorée lorsque ces variables sont sensibles aux paramètres dont l'estimation est améliorable à partir de l'inversion du modèle STICS et des observations du couvert végétal. Nous avons donc observé que les configurations d'observation qui sont les plus favorables à l'amélioration de l'estimation des paramètres (observations sur betterave, en climat sec et pour un sol peu profond) permettent également d'améliorer les prédictions. Par exemple, la prédiction du rendement et de la qualité du grain peut être significativement améliorée car les paramètres auxquels ces variables sont sensibles (ceux liés à la capacité de rétention en eau du sol) sont précisément ceux qu'il est possible d'améliorer par inversion. Nous avons observé pour ces variables qu'elles pouvaient être améliorées de 25% lorsque des observations de betterave étaient prises en compte (voir Chapitre 5.3). A l'opposé, la prédiction de la teneur en azote du sol est difficilement améliorable car elle n'est pas sensible aux paramètres améliorés par l'inversion, cela quelle que soit la configuration d'observation.

Les valeurs des paramètres du sol estimées par inversion de STICS peuvent également être destinées à des fins de prédiction avec un autre modèle. Dans ce cas, il semble pertinent de n'utiliser seulement les valeurs des paramètres qui peuvent effectivement être améliorées par l'inversion de STICS, à savoir ceux qui sont liés au bilan hydrique du sol, et de fixer les autres paramètres à des valeurs

déterminées par d'autres moyens. On peut bien sûr envisager également de refaire l'étude avec le modèle candidat.

#### *L'utilisation d'une information spatiale extérieure pour améliorer l'estimation*

Dans le cas spécifique de l'agriculture de précision à l'échelle intra-parcellaire, nous avons vu que de multiples sources d'erreurs, en particulier sur les observations tant du sol que du couvert végétal, étaient en grande partie responsables du mauvais résultat de l'estimation des paramètres. De plus, la haute résolution spatiale des observations et la mise en relation d'observations mesurées sur des supports de taille différente (des points, des zones définies par le pédologue, des pixels vus par un capteur de télédétection,...) conduit à des erreurs qui s'ajoutent à l'erreur propre d'observation. A cette échelle, la précision du modèle de culture devient questionnable, car elle est probablement faible par rapport à celle que l'on peut obtenir à des échelles plus grandes. Pour toutes ces raisons, nous avons cherché un moyen d'apporter une information supplémentaire dans le processus d'estimation des paramètres, liée à leur structure spatiale dans la parcelle. Nous avons proposé pour cela de valoriser les cartes de résistivité électrique du sol de la parcelle (voir Chapitre 6). Avec des observations synthétiques, nous avons montré que la prise en compte de la structure spatiale issue de la résistivité électrique, pouvait permettre d'améliorer l'estimation des paramètres lorsque l'erreur d'observation était grande. L'approche que nous avons développée est une première tentative, qui demande à être confortée, mais ce résultat est de bon augure pour l'application de la méthode sur données réelles où les erreurs d'observations sont en général fortes. Enfin, cette étude a été entreprise dans le cas d'une parcelle ayant une structure spatiale particulière, avec un gradient, et il serait bienvenu de poursuivre l'exploration des performances de la méthode en essayant de la rendre applicable à des structures spatiales quelconques.

#### *Vers un transfert opérationnel de la méthodologie*

Les cartes d'observations du couvert végétal et de résistivité électrique du sol se révèlent être des données pertinentes pour estimer les paramètres du sol, grâce à l'inversion d'un modèle de culture, et améliorer ainsi les prédictions des variables agroenvironnementales sur lesquelles reposent les règles d'élaboration des choix de



gestion technique des cultures. Ces ensembles de cartes sont généralement disponibles chez les agriculteurs engagés dans la démarche d'agriculture de précision pour lesquels l'application de ces méthodes constitue un enjeu important. D'un point de vue opérationnel, la valorisation de ces cartes d'observations à travers un modèle de culture (STICS ou un autre) doit être entreprise en deux principales étapes. La première consiste en la sélection d'un sous-groupe de paramètres qui permet d'éviter tout problème causé par l'estimation d'un trop grand nombre de paramètres. Durant cette étape, l'utilisateur sera peut-être amené à déterminer les gammes de variations des paramètres, notamment s'il utilise l'analyse de sensibilité globale pour sélectionner les paramètres. Pour cela, il devra se concentrer sur la source d'information (cartes pédologiques, mesures ponctuelles, bibliographie,...) qui lui permettra de définir la gamme de variation a priori la plus précise sur les valeurs des paramètres. La seconde étape consiste à estimer les paramètres du sous-groupe sélectionné, avec toutes les cartes d'observations disponibles. Pour cela, nous avons vu que l'utilisation d'une méthode Bayésienne s'est révélée être utile lorsque l'on dispose d'une information a priori, basée sur la gamme de variation que l'utilisateur a pu éventuellement définir lors de l'étape précédente. Les valeurs estimées des paramètres peuvent ainsi être utilisées dans le modèle afin d'améliorer les prédictions des variables d'intérêt. Un résultat central de cette thèse permet d'ajouter une plus-value à ce travail d'inversion. En repérant dans un contexte pédoclimatique donné quelles sont les configurations d'observation les plus informatives pour estimer au mieux les paramètres du sol, il est possible d'orienter l'acquisition de données sur ces configurations optimales : nous avons vu par exemple, que disposer d'observations sur betterave (et en climat sec si possible) est la configuration la plus favorable pour estimer les paramètres du sol et maximiser l'amélioration de la prédiction des variables d'intérêt, que ce soit pour le blé ou la betterave. On pourrait imaginer de proposer aux opérateurs dans ce domaine, comme Farmstar, d'étendre les produits qu'ils proposent à la betterave afin de fournir une cartographie des propriétés des sols et améliorer ainsi les prédictions de besoins en azote des cultures de blé, colza ou orge.

### *Quelques perspectives*

Avec l'évolution prévue de la technologie, concernant les images de télédétection (future disponibilité d'images à haute résolution spatiale et temporelle avec le capteur Sentinel 2 du système européen GMES) ou les capteurs embarqués sur les tracteurs, un grand nombre de cartes d'observations de qualité pourra être automatiquement acquis sur les cultures. L'utilisation des modèles de culture en sera grandement améliorée grâce au grand nombre d'observations, permettant d'une part de réduire l'erreur d'observation et d'autre part de maximiser les chances d'obtenir la configuration optimale pour estimer au mieux les paramètres du sol. Une telle acquisition d'observations pourrait également permettre de développer des méthodes efficaces quant à l'optimisation de l'utilisation des modèles de culture pour la prédiction. Par exemple, il serait envisageable d'ajouter à la phase d'estimation des paramètres du sol, une phase d'assimilation d'observations visant à filtrer les observations en cours de culture pour corriger les prédictions du modèle à la récolte. Le couplage de la phase d'estimation, entreprise sur un premier jeu d'observations, avec une phase de filtrage, d'un second jeu d'observations, pourrait alors permettre de réduire de manière efficace un certain nombre de problèmes de biais sur les prédictions.

## Annexe

# Calibration des paramètres liés aux caractéristiques génétiques du blé d'hivers

### 1. Constitution d'une base de données

Elle est constituée de 156 expérimentations provenant de divers horizons et collationnées par diverses personnes. Ces expérimentations seront appelées dans la suite USM (Unité de Simulation du Modèle). Elle concerne des essais azote ou des essais irrigation. Cette base de données est caractérisée par 22 sites-années (12 sites géographiques et 13 années climatiques entre 1988 et 2003) et 45 variables observables qui décrivent l'état de la plante et du sol. Le premier travail à d'abord été de renseigner pour chaque USM le nombre de mesures disponibles pour chaque USM et par variables observables qui concerne à la fois le sol et la plante (*LAI*, biomasse sèche, rendement et bien d'autres...). Ce type d'information est capital lorsque l'on souhaitera caler un paramètre sur une variable donnée car pour retenir une USM pour sa calibration, il est nécessaire qu'elle ait au moins une observation de la variable en question.

A partir de là, un travail de caractérisation des USM en fonction des conditions de stress azoté et hydrique de la plante a été entrepris. Pour cela, nous avons commencé par faire tourner la version V6 de STICS, avec le formalisme INN instantané et avec le fichier plante de référence (celui de 2007), sur toutes les USM afin de récupérer pour tous les jours de simulation la valeur de l'indice de stress azoté *turfac* et de stress hydrique *inns*. Ensuite, nous avons compté pour chaque USM et pour les 3 périodes : *AMF-LAX* (vitesse de croissance maximale-*LAI* maximal), *LAX-DRP* (*LAI* maximal-début du remplissage du grain) et *DRP-MAT* (début du remplissage du grain-maturité du grain), le nombre de jours où :

- $turfac \leq 0.7$
- $turfac > 0.7$
- $inns \leq 0.7$
- $inns > 0.7$

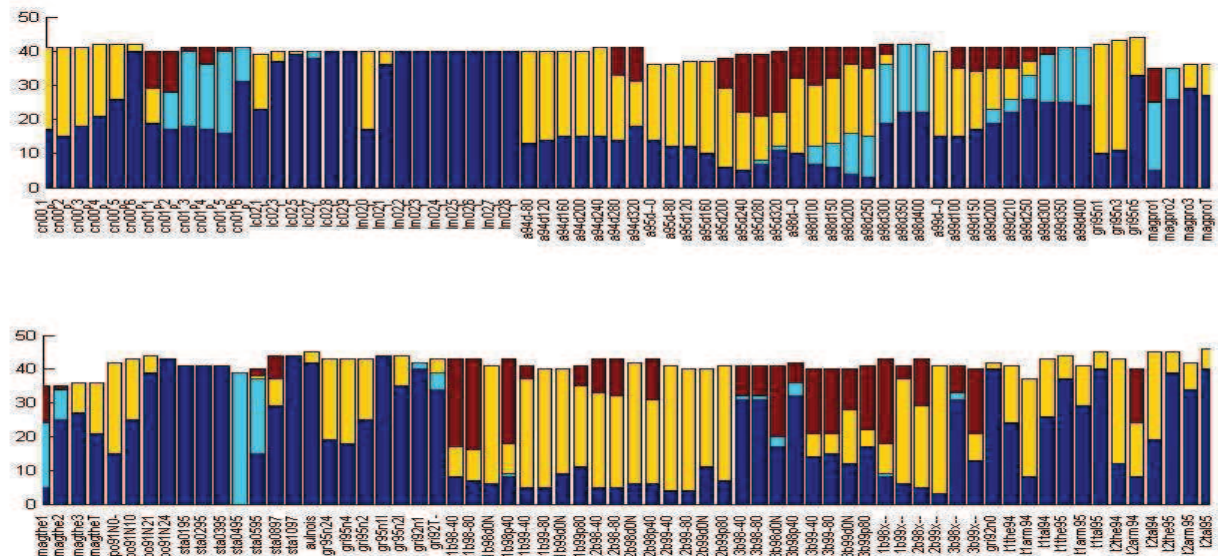
La suite de ce travail a été de renseigner les 4 groupes qui suivent et pour les 3 périodes considérées :

- aucun stress : nombre de jours où  $turfac > 0.7$  et  $inns > 0.7$
- stress hydrique : nombre de jours où  $turfac \leq 0.7$  et  $inns > 0.7$
- stress azoté : nombre de jours où  $turfac > 0.7$  et  $inns \leq 0.7$
- stress hydrique et azoté : nombre de jours où  $turfac \leq 0.7$  et  $inns \leq 0.7$

Pour chacune des 156 USM (identifiée par un code), la durée en jour de chaque période a été représentée par un histogramme à l'intérieur duquel différentes couleurs symbolisent les différentes formes de stress suivantes :

- bleu foncé : aucun stress
- bleu clair : stress hydrique (H)
- jaune : stress azoté (N)
- marron : stress hydrique et azoté (H et N)

Un exemple de résultats, présenté sur la Figure A.1, concerne la période *DRP-MAT* :



**Figure A.1.** Histogramme de la durée en jour de la période *DRP-MAT*, pour chaque USM. Le bleu foncé symbolise le nombre de jour avec aucun stress, le bleu clair avec stress hydrique, le jaune avec stress azoté et le marron avec stress hydrique et azoté.

A partir de ces résultats, le Tableau A.1 binaire (oui/non) a été renseigné pour l'ensemble des USM de la manière suivante :

Nb jour de stress H ou N <20% de la période			Nb jour de stress H seul >20% de la période			Nb jour de stress N seul >20% de la période			Nb jour de stress H et N >20% de la période		
AMF- LAX	LAX- DRP	DRP- MAT	AMF- LAX	LAX- DRP	DRP- MAT	AMF- LAX	LAX- DRP	DRP- MAT	AMF- LAX	LAX- DRP	DRP- MAT

**Tableau A.1.** Caractérisation des USM en fonction de la durée et du type de stress.

A partir de ce tableau de synthèse, nous avons ensuite décidé de classer les USM en deux sous-groupes :

- les USM dites « potentielles » : ce sont des USM ayant un nombre de jour de stress ( $H$  ou  $N$ ) < 20% pour les 3 périodes
- les USM dites « non potentielles » : ce sont des USM ayant un nombre de jour de stress ( $H$  ou  $N$ ) > 20% pour au moins 1 période

Le seuil 0.7 et le taux 20%, pour établir le tableau de synthèse ci-dessus, ont été fixés de manière à constituer les deux sous-groupes ci-dessus de façon à ce qu'ils soient les plus représentatifs en termes de nombre d'USM.

Pour finir, chacun des deux sous-groupes sera partagé en deux en fonction des USM ayant et n'ayant pas suffisamment de données de  $LAI$  afin d'être interpolées pour chaque jour et utilisées comme une données d'entrée dans le modèle. Lorsque le  $LAI$  est interpolé, il est forcé en entrée du modèle, permettant de caler les paramètres liés aux autres processus que le  $LAI$  sous l'hypothèse d'un  $LAI$  connu parfaitement. Nous disposons ainsi du Tableau A.2 qui comptabilise le nombre d'USM dans chaque cas :

	$LAI$ interpolé	$LAI$ non interpolé
Conditions potentielles	6	40
Conditions non potentielles	14	96

**Tableau A.2.** Classement des USM en conditions potentielles ou non et en présence ou non du  $LAI$  interpolé.

La méthode de calibration utilisée dans cette étude est la méthode du simplexe (Lagarias et al., 1998). Les variables observables étudiées pour appliquer

cette méthode sont celles ayant le plus d'intérêt agroenvironnemental pour la suite de notre étude et ayant assez de mesures pour mettre en œuvre la méthode, à savoir : *LAI*, biomasse sèche *masec*, rendement *mafruit* (également appelé *Yld*) et quantité d'azote contenu dans le grain *CNgrain* (également appelé *Prot*).

## **2. Le protocole de calibration**

La logique de ce travail de calibration est de calibrer le modèle étape par étape. D'une part, la production de biomasse est séparée de la production de *LAI* afin d'être calibrée en faisant l'hypothèse d'un *LAI* connu parfaitement (*LAI* forcé) ; l'éventuel biais issu de la simulation du *LAI* est alors supprimé. Une fois la production de biomasse calibrée, le *LAI* est simulé afin d'être à son tour calibré. D'autre part, les conditions potentielles sont séparées des conditions non potentielles, afin de prendre en compte ou non l'effet de l'indice de stress azoté dans la calibration d'un processus donné (biomasse ou *LAI*). Le calcul de l'indice instantané de stress azoté (Mary and Guérif, 2005), noté *inns*, est adopté dans cette étude (au lieu de l'indice cumulé). Le protocole proposé ici permet de caler étape par étape les paramètres des caractéristiques du blé, en conservant au fur et à mesure les valeurs des paramètres calées à l'étape antérieure. Les étapes établies pour ce protocole sont les suivantes :

### **1. *LAI* forcé**

- (i) production de biomasse en condition potentielle,
- (ii) production de biomasse en condition non potentielle.

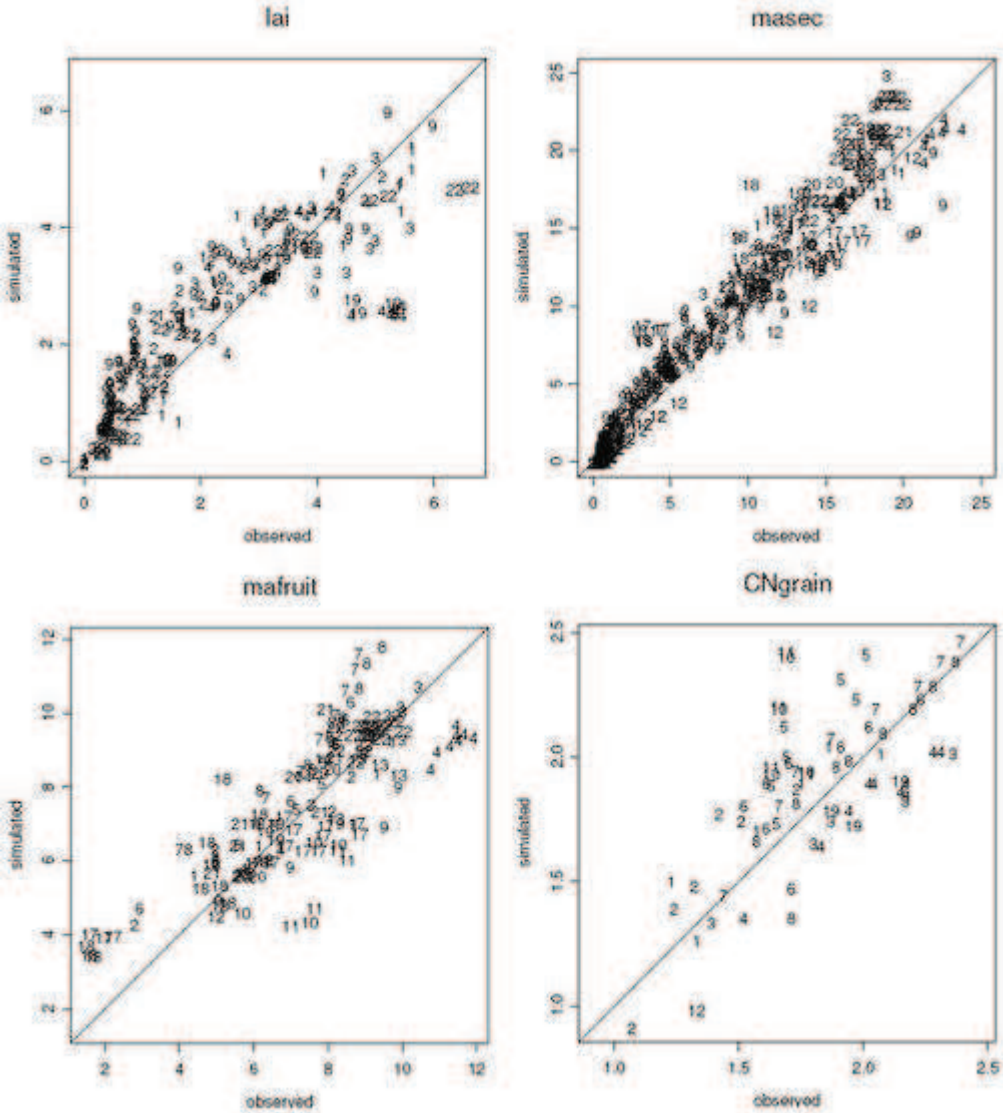
### **2. *LAI* simulé**

- (i) production de *LAI* en condition potentielle,
- (ii) production de *LAI* en condition non potentielle.

## **3. Les performances du modèle, a priori**

Le déroulement du protocole réside d'abord en un diagnostic des variables d'état mesurées, puis en un travail de recherche de paramètres à calibrer

éventuellement. Avant toute chose, la Figure A.2 montre les performances du modèle avant ce travail de calibration, toutes conditions confondues et en *LAI* simulé. La valeur du *RMSE* prise pour les quatre variables étudiées, calculée sur 156 USM, est de 0.85 pour le *LAI*, 1.92 pour *masec*, 1.35 pour *mafruit* et 0.26 pour *CNgrain*.



**Figure A.2.** Comparaison des simulations du modèle aux observations des variables observables étudiées. Chaque site-année est attaché d'un numéro.



## 4. La calibration

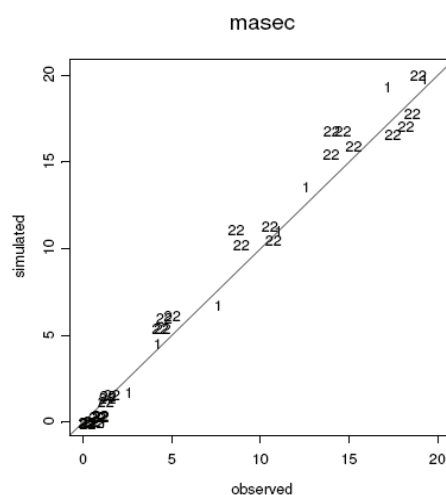
### 4.1. LAI forcé

En forçant le LAI, il est possible de ne s'intéresser qu'aux paramètres liés aux autres processus définis par variables de sortie telles que *masec*, *mafruit* ou *CNgrain*, en faisant l'hypothèse d'un LAI connu parfaitement.

#### *Production de biomasse en conditions potentielles*

- La variable *masec*

La Figure A.3 représente la performance du modèle, calculée sur 5 USM, avant le diagnostic :



**Figure A.3.** Comparaison des simulations du modèle aux observations de la variable *masec* des conditions potentielles. Chaque site-année est attaché d'un numéro.

La Figure A.3 montre que le paramétrage est très bon pour la biomasse et qu'il est difficile d'améliorer les efficacités de croissance telles que *efcroirepro* ou *efcroivej*. La valeur actuelle de ces paramètres ne sera donc pas modifiée.

- Les variables *mafruit* et *CNgrain*

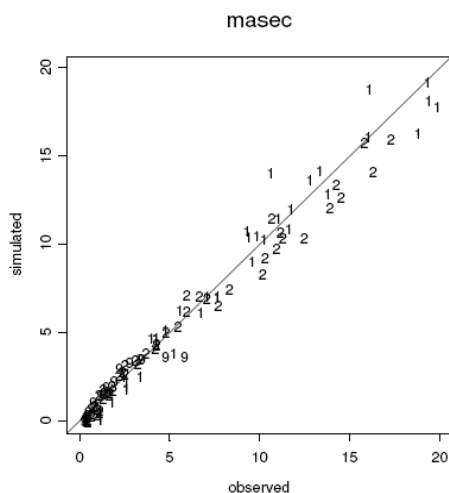
Pour le moment, aucun diagnostic sur ces deux variables ne peut être fait car on ne dispose pas suffisamment d'USM sur lesquelles on puisse établir un bilan. L'idée est de rassembler les USM potentielles et non potentielles, ayant et n'ayant

pas un *LAI* interpolé, afin d'obtenir un groupe suffisamment grand pour établir un diagnostic efficace. Ce travail sera entrepris après la dernière étape du protocole.

#### *Production de biomasse en condition non potentielle*

- La variable *masec*

La Figure A.4 représente la performance du modèle, calculée sur 12 USM, avant le diagnostic :



**Figure A.4.** Comparaison des simulations du modèle aux observations de la variable *masec* des conditions non potentielles. Chaque site-année est attaché d'un numéro.

Avec un *RMSE* de 0.93, la répartition du nuage de points concernant ce sous groupe d'USM semble satisfaisante. Une légère sous estimation est toutefois visible et à ce niveau du protocole. Il serait alors intéressant de caler le paramètre lié à l'indice de stress azoté concernant la variable *masec* : *innmin*. L'issue de cette étape permet de passer d'une valeur du paramètre *innmin* de 0.3 à 0.36 afin d'atteindre un *RMSE* de 0.89 pour *masec*.

#### **4.2. LAI simulé**

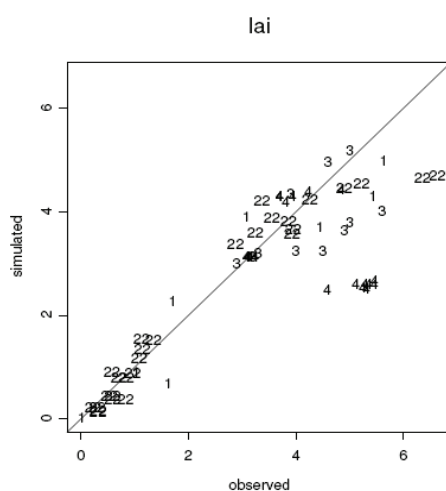
En simulant à présent le *LAI*, il est possible de ne s'intéresser qu'aux paramètres qui y sont liés, en faisant l'hypothèse que ceux liés aux autres variables soient correctement calés. Les paramètres liés au *LAI* concernent le retardement ou l'avancée des dates de stades phénologiques mais concernent aussi la vitesse de croissance de la plante. Il faut savoir que certains paramètres, comme pour les dates

de stades (mais aussi pour la durée de vie des feuilles), sont considérés comme propre à chaque variété de plante, alors que d'autres, comme la vitesse de croissance, ne le sont pas.

### *Production de LAI en condition potentielle*

#### - La variable *LAI*

La Figure A.5 représente la performance du modèle, calculée sur 18 USM, avant le diagnostic :



**Figure A.5.** Comparaison des simulations du modèle aux observations de la variable *LAI* des conditions potentielles. Chaque site-année est attaché d'un numéro.

Le *RMSE* calculé sur le nuage de points du *LAI* est de 1.05. En fin de cycle, le modèle sous-estime largement les valeurs de *LAI*. Il faut donc ajuster ce nuage de points en essayant de caler d'une part la vitesse de croissance du *LAI*, *dlaimaxbrut*, et d'autre part les paramètres qui règle le positionnement des deux stades situés dans la phase de croissance rapide, à savoir le stade *AMF* et le stade *LAX*. En effet, la ré-estimation des paramètres liés à ces stades (notés *stlevamf* et *stamflax*) a pour but de régler en partie les problèmes de développement qui se manifestent par un déphasage (en terme de positionnement des stades *amf* et *lax*) de la courbe de *LAI* simulée par le modèle avec les observations de cette variable. De plus, le paramètre de durée de vie des feuilles *durvieF* sera également ajusté.

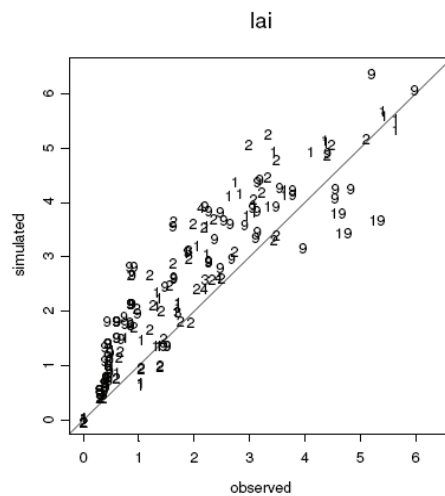
Seulement deux variétés disposent de données de *LAI* : Shango et Talent. Concernant la variété Talent, elle ne dispose pas de suffisamment d'USM sur

lesquels il est possible de caler les paramètres *stlevamf*, *stamflax* et *durvieF* d'une manière satisfaisante. Cette variété sera toutefois utilisée avec Shango pour caler *dlaimaxbrut*. Dans cette étape, nous proposons de caler successivement *dlaimaxbrut*, puis *stlevamf* et *stamflax*, et enfin *durvieF*. Il faut noter que ne disposant d'aucune date de stade observée, il n'est donc pas possible d'utiliser le module de forçage de stade proposé par le modèle utile pour caler les paramètres liés au *LAI* indépendamment des valeurs de *stlevamf* et *stamflax*. L'issue de cette étape permet de passer d'une valeur des paramètres (*dlaimaxbrut*, *stlevamf*, *stamflax*, *durvieF*) de  $(4.4 \cdot 10^{-4}, 235, 340, 200)$  à  $(4.7 \cdot 10^{-4}, 278, 352, 205)$  afin d'atteindre un *RMSE* de 0.75 pour *LAI*.

#### Production de *LAI* en condition non potentielle

##### - La variable *LAI*

La Figure A.6 représente la performance du modèle, calculée sur 24 USM, avant le diagnostic :



**Figure A.6.** Comparaison des simulations du modèle aux observations de la variable *LAI* des conditions non potentielles. Chaque site-année est attaché d'un numéro.

Le *RMSE* calculé sur le nuage de points du *LAI* est de 0.85. Il y a globalement un problème de sur-estimation du *LAI* pour les conditions non potentielles. Ce diagnostic permet de cibler directement les paramètres qui sont ici à remettre en cause : ce sont les paramètres des courbes de réponse (appelées *innsenes* et *innlai*) à l'indice de stress azoté *inns* : *innsen* et *innturgmin*. L'issue de cette étape permet

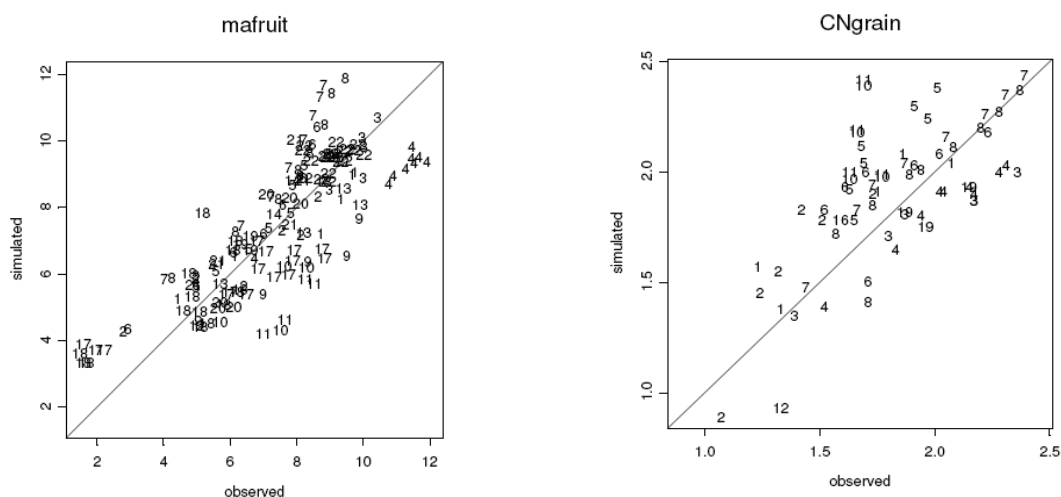
de passer d'une valeur des paramètres (*innsen*, *innturgmin*) de (0.35, -0.2) à (0.17, -0.65) afin d'atteindre un *RMSE* de 0.68 pour *LAI*.

### Conditions potentielles et non potentielles

#### - Les variables *mafruit* et *CNgrain*

A présent, il est possible d'établir un diagnostic pour ces deux variables, car le fait d'avoir calé les paramètres liés au stress azoté *inns* permet de disposer de suffisamment d'USM sur lesquelles il est possible de s'appuyer de manière satisfaisante.

La Figure A.7 représente la performance du modèle avant le diagnostic :

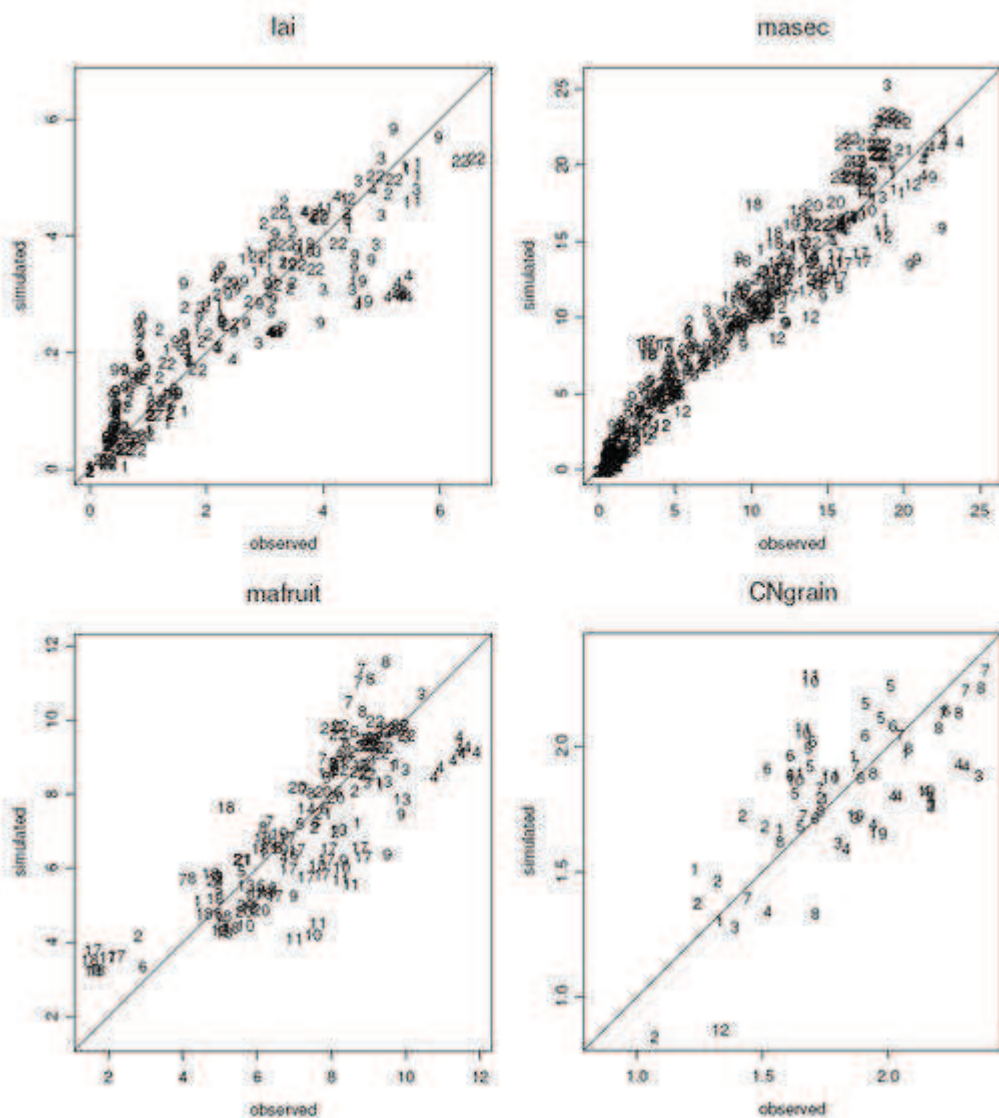


**Figure A.7.** Comparaison des simulations du modèle aux observations des variables *mafruit* et *CNgrain* des conditions potentielles et non potentielles. Chaque site-année est attaché d'un numéro.

D'une manière générale, ces deux variables présentent de la dispersion. Le *RMSE* est de 1.35 pour *mafruit* et de 0.26 pour *CNgrain*. La variable *mafruit* a de plus une légère tendance à être sous-estimée et *CNgrain* à être sur-estimée. Le lien existant entre ces deux variables concerne l'indice de récolte et les paramètres liés à cet indice, *vitirazo* et *vitircarb*, vont faire l'objet d'une calibration. Sur ces deux variables, nous proposons de caler d'abord *vitirazo* puis ensuite *vitircarb*. L'issue de cette étape permet de passer d'une valeur des paramètres (*vitirazo*, *vitircarb*) de (0.01757, 0.011) à (0.0161, 0.0107) afin d'atteindre un *RMSE* de 1.32 pour *mafruit* et de 0.24 pour *CNgrain*.

## 5. Les performances du modèle, a posteriori

La Figure A.8 montre les performances du modèle obtenues après ce travail de calibration, toutes conditions confondues et en *LAI* simulé. La valeur du *RMSE* prise pour les quatre variables étudiées est de 0.75 pour le *LAI*, 1.81 pour *masec*, 1.32 pour *mafruit* et 0.24 pour *CNgrain*.



**Figure A.8.** Comparaison des simulations du modèle aux observations des variables observables étudiées. Chaque site-année est attaché d'un numéro.

Notons que le *LAI* est au cœur du protocole proposé ici. Ceci est dû à la fois au fait que cette variable joue un rôle central au sein du modèle STICS et au fait qu'elle doit être une des variables les plus délicates à bien simuler. Notons également que l'on ne parle pas des paramètres liés au stress hydrique. En effet, notre démarche de classement des USM ne nous a pas permis de découvrir d'USM qui soit suffisamment stressée en eau afin de d'ajuster les paramètres sur le stress hydrique. Nous supposons toutefois que les paramètres par défaut sont corrects. Globalement, les améliorations apportées par le nouveau paramétrage des caractéristiques du blé pour la simulation des variables de sortie sont les suivantes :

- Le *LAI* est la variable qui est la plus affectée par le nouveau paramétrage : la dispersion et la sur-estimation ont assez bien été corrigées,
- La variable *masec* est globalement mieux simulée,
- Le biais des variables *mafruit* et *CNgrain* a légèrement été corrigé.



## Glossaire (paramètres et variables)

albedo	Albédo du sol sec nu
AMF	Stade de vitesse de croissance maximale
AR	Teneur en argile
argi	Teneur en argile de la couche de surface
Cab	Teneur en chlorophylle des feuilles
CA	Teneur en calcaire
calc	Teneur en calcaire de la couche de surface
CNgrain	Teneur en azote contenu dans le grain
DA(h)	Densité volumique de la couche de sol h
dlaimaxbrut	Vitesse de croissance du LAI
DRP	Stade du début du remplissage du grain
durvieF	Durée de vie des feuilles
epc(h)	Épaisseur de la couche de sol h
FINERT	Proportion d'azote inactif
FLO	Stade de floraison
HCC(h)	Humidité à la capacité au champ de la couche de sol h
HUMIRAC	Indice de stress hydrique à la levée
Hinit(h)	Teneur en eau dans la couche de sol h au début de la simulation
HMIN(h)	Humidité au point de flétrissement de la couche de sol h
HUMSOL	Contenu en eau du sol
innlai	Indice de stress azoté pendant la croissance du LAI, défini par les points [1, 1] et [innmin, innturgmin]
inns	Indice de stress azoté
innsenes	Indice de stress azoté pendant la sénescence, défini par les points [1, 1] et [innmin, innsen]
LAI	Indice foliaire
LAX	Stade de LAI maximal
mafruit	Rendement
masec	Biomasse sèche
MAT	Stade de maturité du grain

NHUM	Teneur en azote organique actif dans le sol
Nit	Teneur en azote minéral du sol à la récolte
NO3init(h)	Azote présent dans la couche de sol h au début de la simulation
Norg	Taux d'azote organique de la couche de surface
N <sub>surf</sub>	Teneur en azote surfacique
profhum	Profondeur de minéralisation
Prot	Teneur en protéine du grain à la récolte
q0	Seuil de cumul d'évaporation journalière
QN	Teneur en azote de la plante
ruisolnu	Fraction d'eau qui ruisselle sur sol nu
SENSRSEC	Sensibilité de la plante à la sécheresse
SF	Teneur en sable
stamflax	Positionnement du stade LAX
stlevamf	Positionnement du stade AMF
Summer	Date de LAI maximal pour la betterave
turfac	Indice de stress hydrique
vitirazo	Indice de récolte azoté
vitircarb	Indice de récolte carboné
Yld	Rendement à la récolte

## Références générales

Aggarwal, P.K., 1995. Uncertainties in Crop, Soil and Weather Inputs Used in Growth-Models - Implications for Simulated Outputs and Their Applications. *Agricultural Systems* 48 361-384.

Al Majou, H., Bruand, A., Duval, O., Cousin, I., 2007. Comparaison de fonction de pédotransfert nationales et européennes pour prédire les propriétés de rétention en eau des sols. *Etude et Gestion des Sols* 14 103-116.

Aldrich, J., 1997. R. A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science* 12 162-176.

Baret, F., Houles, V., Guerif, M., 2006. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. Symposium on Imaging Techniques for Understanding Plant Responses to Stress held at the Society-for-Experimental-Biology Meeting. Canterbury, ENGLAND.

Bastet, G., Bruand, A., Quélin, P., Cousin, I., 1998. Estimation des propriétés de rétention en eau des sols à l'aide de fonctions de pédotransfert (FTP) : Une analyse bibliographique. *Etude et Gestion des Sols* 5 7-28.

Batchelor, W.D., Basso, B., Paz, J.O., 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy* 18 141-158.

Beaudoin, N., 2006. Caractérisation expérimentale et modélisation des effets des pratiques culturales sur la pollution nitrique d'un aquifère en zone de grande culture. Thèse, Institut National Agronomique Paris-Grignon.

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 Bruyeres catchment. *European Journal of Agronomy* 29 46-57.

Beaudoin, N., Nicoullaud, 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., King, D., (Eds.), *Agriculture de précision*. Quae: Versailles.

Beaudoin, N., Saad, J.K., Van Laethem, C., Machel, J.M., Maucorps, J., Mary, B., 2005. Nitrate leaching in intensive agriculture in Northern France: Effect of farming

practices, soils and crop rotations. *Agriculture, Ecosystems & Environment* 111 292-310.

Beven, K., Binley, A., 1992. The Future of Distributed Models - Model Calibration and Uncertainty Prediction. *Hydrological Processes* 6 279-298.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249 11-29.

Blackmore, S., Moore, M., 1999. Remedial correction of yield map data. *Precision Agriculture* 1 53-56.

Blasone, R.S., Madsen, H., Rosbjerg, D., 2008. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* 353 18-32.

Bonesmo, H., Belanger, G., 2002. Timothy yield and nutritive value by the CATIMO model: I. Growth and nitrogen. *Agronomy Journal* 94 337-345.

Booltink, H.W.G., van Alphen, B.J., Batchelor, W.D., Paz, J.O., Stoorvogel, J.J., Vargas, R., 2001. Tools for optimizing management of spatially-variable fields. *Agricultural Systems* 70 445-476.

Boote, K.J., Jones, J.W., Hoogenboom, G., 1998. Simulation of crop growth: CROPGRO Model, Chapter 18. In: Peart, R.M., Curry, R.B., (Eds.), *Agricultural Systems Modeling and Simulation*. Marcel Dekker, Inc: New York.

Bouman, B.A.M., 1994. A Framework to Deal with Uncertainty in Soil and Management Parameters in Crop Yield Simulation - a Case-Study for Rice. *Agricultural Systems* 46 1-17.

Bourennane, H., King, D., Couturier, A., Nicoullaud, B., Mary, B., Richard, G., 2007. Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. *Ecological Modelling* 205 323-335.

Bouttier, F., Courtier, P., 1999. Data assimilation concepts and methods. ECMWF Meteorological Training Course Lecture Series. Reading: England.

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., Launay, M., Mary, B., Beaudoin, N., 2008. Conceptual basis, formalisations and parameterization of the STICS crop model. Quae: Versailles.

Brisson, N., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M.H., Bouthier, A., Ripoche, D., Mary, B., Juste, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie* 22 69-92.

Brisson, N., Wery, J., Boote, K., 2006. Fundamental concepts of crop models illustrated by comparative approach. In: Wallach, D., Makowski, D., Jones, J.W., (Eds.), *Working with Dynamic Crop Models*. Elsevier.

Bruand, A., Fernandez, P.N., Duval, O., 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. *Soil Use and Management* 19 232-242.

Brun, R., Kuhni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets. *Water Research* 36 4113-4127.

Calmon, M.A., Jones, J.W., Shinde, D., Specht, J.E., 1999a. Estimating parameters for soil water balance model using adaptive simulated annealing. *Applied Engineering in Agriculture* 15 703-7013.

Calmon, M.A., Jones, J.W., Shinde, D., Specht, J.E., 1999b. Estimating parameters for soil water balance models using adaptive simulated annealing. *Applied Engineering in Agriculture* 15 703-713.

Campolongo, F., Cariboni, J., Saltelli, A., 2006. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software* in press.

Campolongo, F., Saltelli, A., 1997. Sensitivity analysis of an environmental model an application of different analysis methods. *Reliability Engineering & System Safety* 57 49-69.

Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2004. The role of sensitivity analysis in ecological modelling. 4th Conference of the International-Society-for-Ecological-Informatics. Busan, SOUTH KOREA.

Chan, K., Tarantola, S., Saltelli, A., Sobol, I.M., 2000. Variance-based methods. In: Saltelli, A., Chan, K., Scott, E.M., (Eds.), *Sensitivity analysis*. Wiley: New York.

Chan, K., Tarantola, S., Saltelli, A., Sobol, I.M., 2001. Variance-based methods. In: Saltelli, A., Chan, K., Scott, E.M., (Eds.), *Sensitivity analysis*. Wiley: New York.

- Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modelling Spatial Uncertainty. Wiley Series in Probability and Statistics.
- Cressie, N., 1993. Statistics for Spatial Data: Revised Edition. John Wiley and Sons: New York, USA.
- Cukier, R.I., Fortuin, C., Shuler, K.E., Petshek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients, I, Theory. The Journal of Chemical Physics 59 3873-3878.
- Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Non-Linear Sensitivity Analysis of Multi-Parameter Model Systems. Journal of Computational Physics 26 1-42.
- Cukier, R.I., Schaibly, J.H., Shuler, K.E., 1975. Study of Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients .3. Analysis of Approximations. Journal of Chemical Physics 63 1140-1149.
- De Wit, C.T., 1978. Simulation of assimilation, respiration and transpiration of crops. Monographs, Pudoc, Wageningen.
- Dimokas, G., Tchamitchian, M., Kittas, C., 2009. Calibration and validation of a biological model to simulate the development and production of tomatoes in Mediterranean greenhouses during winter period. Biosystems Engineering 103 217-227.
- Engel, T., Hoogenboom, G., Jones, J.W., Wilkens, P.W., 1997. AEGIS/WIN: A computer program for the application of crop simulation models across geographic areas. Agronomy Journal 89 919-928.
- Ewert, F., van Oijen, M., Porter, J.R., 1999. Simulation of growth and development processes of spring wheat in response to CO<sub>2</sub> and ozone for different sites and years in Europe using mechanistic crop simulation models. European Journal of Agronomy 10 231-247.
- Ferreira, R.A., Jones, J.W., Graham, W.D., 2006. Parameterizing Spatial Crop Models with Inverse Modeling: Sources of Error and Unexpected Results. Transactions of the ASABE 49 1547-1561.
- Flenet, F., Villon, P., Ruget, F.O., 2003. Methodology of adaptation of the STICS model to a new crop: spring linseed (*Linum usitatissimum*, L.). STICS Workshop. Camargue, FRANCE.
- Gabrielle, B., Roche, R., Angas, P., Cantero-Martinez, C., Cosentino, L., Mantineo, M., Langensiepen, M., Henault, C., Laville, P., Nicoulaud, B., Gosse, G., 2002. A

priori parameterisation of the CERES soil-crop models and tests against several European data sets. *Agronomie* 22 119-132.

Garcia De Cortazar Atauri, I., 2006. Adaptation du modèle STICS à la vigne (*vitis vinifera* L.). Utilisation dans le cadre d'une étude d'impact du changement climatique à l'échelle de la France. These, Ecole Nationale Supérieure Agronomique de Montpellier.

Gauchere, C., Campillo, F., Misson, L., Guiot, J., Boreux, J.-J., 2008. Parameterization of a process-based tree-growth model: Comparison of optimization, MCMC and Particle Filtering algorithms. *Environmental Modelling & Software* 23 1280-1288.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1995. Markov Chain Monte Carlo in Practice. Chapman & Hall: London.

Godard, C., Roger-Estrade, J., Jayet, P.A., Brisson, N., Le Bas, C., 2008. Use of available information at a European level to construct crop nitrogen response curves for the regions of the EU. *Agricultural Systems* 97 68-82.

Golovko, L., Pozdnyakov, A.I., 2007. Electrical geophysical methods in agriculture. *Progress of Information Technology in Agriculture* 457-471.

Gomez-Delgado, M., Tarantola, S., 2006. GLOBAL sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of a hazardous waste disposal site in Spain. *International Journal of Geographical Information Science* 20 449-466.

Gotzinger, J., Bardossy, A., 2007. Comparison of four regionalisation methods for a distributed hydrological model. *Journal of Hydrology* 333 374-384.

Gras, R., Erhwein, J., Hermant, G., Hebert, J., 1961. Carte des sols du département de l'Aisne. Laon n 5-6. Chambre d'agriculture de l'Aine.

Grievank, A., 2000. Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation. SIAM Publisher, Philadelphia.

Guérif, M., Beaudoin, N., Durr, C., Machet, J.M., Mary, B., Michot, D., Moulin, D., Nicoulaud, B., Richard, G., 2001. Designing a field experiment for assessing soil and crop spatial variability and defining site specific management strategies. Proceedings 3rd European Conference on Precision Agriculture. Montpellier, France.

Guérif, M., Duke, C., 1998. Calibration of the SUCROS emergence and early growth module for sugar beet using optical remote sensing data assimilation. *European Journal of Agronomy* 9 127-136.



Guérif, M., Houlès, V., Makowski, D.,Lauvernet, C., 2006. Data assimilation and parameter estimation for precision agriculture using the crop model STICS. In: Wallach, D., Makowski, D.,Jones, J.W., (Eds.), Working with Dynamic Crop Models. Elsevier.

Hadria, R., Khabba, S., Lahrouni, A., Duchemin, B., Chehbouni, A., Carriou, J.,Ouzine, L., 2007. Calibration and validation of the STICS crop model for managing wheat irrigation in the semi-arid Marrakech/Al Haouzi plain. *Arabian Journal for Science and Engineering* 32 87-101.

Hald, A., 1999. On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science* 14 214-222.

Hansen, S., Jensen, H.E., Nielsen, N.E.,Swenden, H., 1990. DAISY - soil plant atmosphere system model NP0 research in the NAEP report, Nr A10. The Royal Veterinary and Agricultural University.

Hastings, W.K., 1970. Monte Carlo sampling method using Markov chainand their applications. *Biometrika* 57 97-109.

Heidmann, T., Tofteng, C., Abrahamsen, P., Plauborg, F., Hansen, S., Battilani, A., Coutinho, J., Dolezal, F., Mazurczyk, W., Ruiz, J.D.R., Takac, J.,Vacek, J., 2008. Calibration procedure for a potato crop growth model using information from across Europe. *Ecological Modelling* 211 209-223.

Hilgert, N., Rossi, V., Vila, J.P.,Wagner, V., 2005. Identification, estimation, and control of uncertain dynamic systems: A nonparametric approach. 11th International Symposium on Applied Stochastic Models and Data Analysis. Brest, FRANCE.

Houborg, R.,Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment* 112 186-202.

Houlès, V., 2004. Mise au point d'un outil de modulation intra-parcellaire de la fertilisation azoté du blé d'hiver basé sur la télédétection et un modèle de culture. Thèse, Institut National Agronomique Paris-Grignon.

Houlès, V., Guérif, M.,Mary, B., 2007. Elaboration of a nitrogen nutrition indicator for winter wheat based on leaf area index and chlorophyll content for making nitrogen recommendations. *European Journal of Agronomy* 27 1-11.

Houlès, V., Mary, B., Guérif, M., Makowski, D., Juste, E., 2004. Evaluation of the crop model STICS to recommend nitrogen fertilization rates according to agro-environmental criteria. *Agronomie* 24 1-9.

Irmak, A., Jones, J.W., Batchelor, W.D., Paz, J.O., 2001. Estimating Spatially Variable Soil Properties for Application of Crop Models in Precision Farming. *Transactions of the ASAE* 44 1343-1353.

Kanso, A., Chebbo, G., Tassin, B., 2004. Application of MCMC-GSA model calibration method to urban runoff quality modeling. 4th International Conference on Sensitivity Analysis of Model Output (SAMO 2004). Santa Fe, NM.

Kauffmann, P., 1994. *Information Estimation Tests*. Dunod: Paris.

King, D., Daroussin, J., Tavernier, R., 1994. Development of a Soil Geographic Database from the Soil Map of the European Communities. *Catena* 21 37-56.

Koda, M., McRae, G.J., J.H., S., 1979. Automatic sensitivity analysis of kinetic mechanisms. *Int. J. Chem. Kinet.*, 11 427-444.

Lagacherie, P., Baret, F., Feret, J.B., Netto, J.M., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment* 112 825-835.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *Siam Journal on Optimization* 9 112-147.

Launay, M., Graux, A.-I., Brisson, N., Guérif, M., 2009. Carbohydrate remobilization from storage root to leaves after a stress release in sugar beet (*Beta vulgaris* L.): experimental and modelling approaches. *The Journal of Agricultural Science*, accepted for publication.

Launay, M., Guérif, M., 2003. Ability for a model to predict crop production variability at the regional scale: an evaluation for sugar beet. *Agronomie* 23 135-146.

Launay, M., Guérif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture Ecosystems & Environment* 111 321-339.

Launay, M., Guérif, M., Dedieu, G., 2000. Utilisation d'un modèle de correction atmosphérique (SMAC) pour le calcul de réflectances au sol à partir d'images SPOT. *Photo-Interprétation* n°1999.

Lauvernet, C., 2005. Assimilation variationnelle d'observations de télédétection dans les modèles de fonctionnement de la végétation : utilisation du modèle adjoint et prise en compte de contraintes spatiales. These, Université Joseph Fourier Grenoble I.

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

Leenhardt, D., Wallach, D., Le Moigne, P., Guérif, M., Bruand, A., Casterad, M.A., 2006. Using crop models for multiple fields. In: Wallach, D., Makowski, D., Jones, J.W., (Eds.), *Working with Dynamic Crop Models*. Elsevier.

Machet, J.M., Couturier, A., Beaudoin, N., 2007. Cartographie du rendement du blé et des caractéristiques qualitatives des grains. In: Guérif, M., King, D., (Eds.), *Agriculture de précision*. Quae: Versailles.

Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M.H., 2006a. Parameter estimation for crop models. In: Wallach, D., Makowski, D., Jones, J.W., (Eds.), *Working with Dynamic Crop Models*. Elsevier.

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

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

Manache, G., Melching, C.S., 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. *Environmental Modelling & Software* 23 549-562.

Mary, B., Guérif, M., 2005. Effet du stress azoté sur la plante: définition d'un indice de nutrition instantané. Séminaire Stics. Carry-le-Rouet, FRANCE.

Mertens, F.M., Paetzold, S., Welp, G., 2008. Spatial heterogeneity of soil properties and its mapping with apparent electrical conductivity. *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde* 171 146-154.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21 1087-1091.

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

Mitchell, M., 1998. An introduction to genetic algorithms. MIT press.

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. Elsevier.

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* 33 161-174.

Moulin, S., Zurita, R.M., Guérif, M., 2007. Estimation de variables biophysiques du couvert par ajustement de modèles de transfert radiatif sur des réflectances. In: Guérif, M., King, D., (Eds.), Agriculture de précision. Quae: Versailles.

Murphy, B., Geeves, G., Miller, M., Summerell, G., Southwell, P., Rankin, M., 2003. The application of pedotransfer functions with existing soil maps to predict soil hydraulic properties for catchment-scale hydrologic and salinity modelling. International Congress on Modelling and Simulation. Townsville, AUSTRALIA.

Nemes, A., Timlin, D.J., Pachepsky, Y.A., Rawls, W.J., 2009. Evaluation of the Rawls et al. (1982) Pedotransfer Functions for their Applicability at the US National Scale. *Soil Science Society of America Journal* 73 1638-1645.

Nicoullaud, B., Beaudoin, N., Roque, J., Couturier, A., Maucorps, J., King, D., 2007. Cartographie des sols et agriculture de précision. In: Guérif, M., King, D., (Eds.), Agriculture de précision. Quae: Versailles.

Niu, S.S., Fisher, D.C., 1997. Detecting parameter identifiability problems in system identification. *International Journal of Adaptive Control and Signal Processing* 11 603-619.

Pang, X.P., Gupta, S.C., Moncrief, J.F., Rosen, C.J., Cheng, H.H., 1998. Evaluation of nitrate leaching potential in Minnesota glacial outwash soils using the CERES-Maize model. *Journal of Environmental Quality* 27 75-85.

Pathak, T.B., Fraise, C.W., Jones, J.W., Messina, C.D., Hoogenboom, G., 2007. Use of Global Sensitivity Analysis for CROPGRO Cotton Model Development. *Transactions of the ASABE* 50 2295-2302.

Paz, J.O., Batchelor, W.D., Babcock, B.A., Colvin, T.S., Logsdon, S.D., Kaspar, T.C., Karlen, D.L., 1999. Model-based technique to determine variable rate nitrogen for corn. *Agricultural Systems* 61 69-75.

Pronzato, L., 2008. Optimal experimental design and some related control problems. *Automatica* 44 303-325.

Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications* 136 212-224.

Ratto, M., Young, P.C., Romanowicz, R., Pappenberger, F., Saltelli, A., Pagano, A., 2007. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Hydrology and Earth System Sciences* 11 1249-1266.

Reidsma, P., Ewert, F., Boogaard, H., van Diepen, K., 2009. Regional crop modelling in Europe: The impact of climatic conditions and farm characteristics on maize yields. *Agricultural Systems* 100 51-60.

Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research* 36 3653-3662.

Ritchie, J.T., Otter, S., 1984. Ceres-Wheat: a user-oriented wheat yield model. Preliminary documentation, Agristars Publication No YM-U3-04442-JSC-18892.

Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems* 83 248-265.

Rossi, V., Vila, J.P., 2005. Filtering discrete time nonlinear system with unknown parameters: a non parametric approach. *Comptes Rendus Mathematique* 340 759-764.

Ruget, F., Brisson, N., Delecolle, R., Faivre, R., 2002. Sensitivity analysis of a crop simulation model, STICS, in order to choose the main parameters to be estimated. *Agronomie* 22 133-158.

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.

Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis* 26 445-460.

Saltelli, A., Chan, K., Scott, E.M., 2000a. *Sensitivity Analysis*. John Wiley and Sons.

Saltelli, A., Tarantola, S., Campolongo, F., 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15 377-395.

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41 39-56.

Samouelian, A., Cousin, I., Tabbagh, A., Bruand, A., Richard, G., 2005. Electrical resistivity survey in soil science: a review. *Soil & Tillage Research* 83 173-193.

Seber, G.A.F., Wild, C.J., 2003. *Nonlinear Regression*. Wiley-Interscience: Hoboken, NJ.

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

Sobol, I.M., 1993. Sensitivity analysis for non-linear mathematical models. *Math. Modelling & Comp. Exp.* 1 407-414.

Sobol, I.M., Tarantola, S., Gatelli, D., Kucherenko, S.S., Mauntz, W., 2007. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety* 92 957-960.

St'astna, M., Zalud, Z., 1999. Sensitivity analysis of soil hydrologic parameters for two crop growth simulation models. *Soil & Tillage Research* 50 305-318.

Supit, L., Hooijer, A.A., van Diepen, C.A., 1994. *System Description of the WOFOST 6.0 Crop Simulation Model Implemented in CGMS*, Office for Official Publications of the EC, Luxembourg.

Timlin, D., Pachepsky, Y., Walthall, C., Loechel, S., 2001. The use of a water budget model and yield maps to characterize water availability in a landscape. *Soil & Tillage Research* 58 219-231.

Tolson, B.A., Shoemaker, C.A., 2008. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research* 44.

Tonitto, C., David, M.B., Drinkwater, L.E., Li, C.S., 2007. Application of the DNDC model to tile-drained Illinois agroecosystems: model calibration, validation, and uncertainty analysis. *Nutrient Cycling in Agroecosystems* 78 51-63.

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

van der Keur, P., Hansen, J.R., Hansen, S., Refsgaard, J.C., 2008. Uncertainty in simulation of nitrate leaching at field and catchment scale within the Odense river basin. *Vadose Zone Journal* 7 10-21.

Varella, H., Guérif, M., Buis, S., 2008. Global Sensitivity Analysis (GSA) measures the quality of parameter estimation. Case of soil parameter estimation with a crop model., EGU. Vienna, Austria.

Varella, H., Guérif, M., Buis, S., 2009. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling & Software*. In press. doi:10.1016/j.envsoft.2009.09.012.

Varma, A., Morbidelli, M., Wu, H., 1999. *Parametric Sensitivity in Chemical Systems*. Cambridge University Press.

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. Camb* 102 371-382.

Weiss, M., Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment* 70 293-306.

Williams, J.R., Jones, C.A., Dyke, P.T., 1984. A modeling approach for determining the relationship between erosion and soil productivity. *Transactions of the ASAE* 27 129-144.