

Le modèle de culture STICS

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.

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.

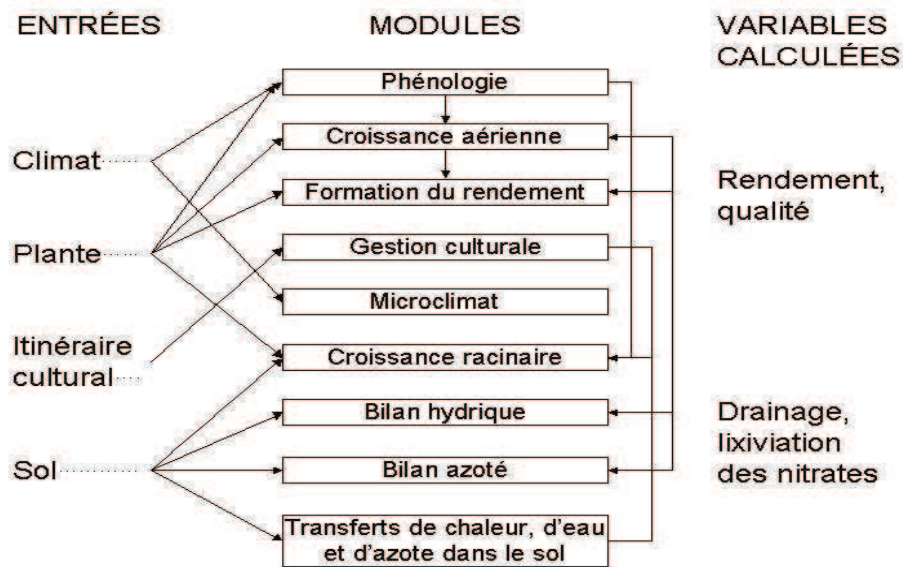


Figure 2-1. Les principaux modules du modèle de culture STICS, avec les groupes de paramètres associés, avec les entrées (variables et paramètres) et les variables de sortie. Les flèches de gauche traduisent les principales influences entre entrées et modules. Les flèches centrales symbolisent la croissance 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.

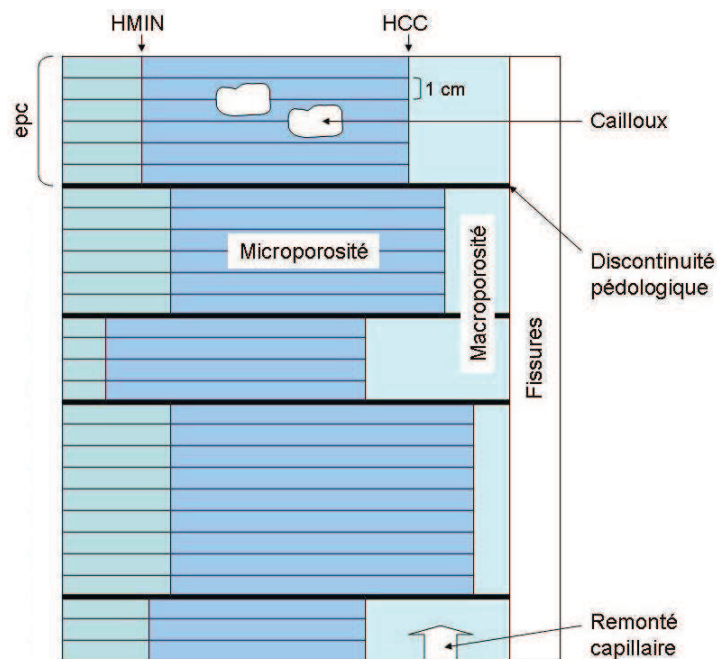


Figure 2-2. Description du profil de sol en compartiments (d'après Brisson et al., 2008)

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 1.5×10^4 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 θ_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 , θ appartenant au sous-espace Θ 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 μ 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 θ_i et θ_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 θ_i sur la variance de Y ,

et l'indice total ST_i , représentant la somme de tous les effets (principal et interactions) impliquant θ_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 θ_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 $s \in [-\pi, \pi]$, ω_i est un jeu de fréquences angulaires entières dont le choix sera discuté plus loin, φ_i est un décalage de phase tiré aléatoirement dans $[0, 2\pi[$ et G_i est appelée fonction de transformation. Le décalage φ_i est optionnel et permet d'introduire de l'aléatoire dans le parcours déterministe de Θ 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 Θ appelée "search curve" pouvant passer arbitrairement près de tout point de Θ . 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 $\omega_1 = 3$, $\omega_2 = 7$, $\omega_3 = 11$, $\varphi_1 = \varphi_2 = \varphi_3 = 0$.

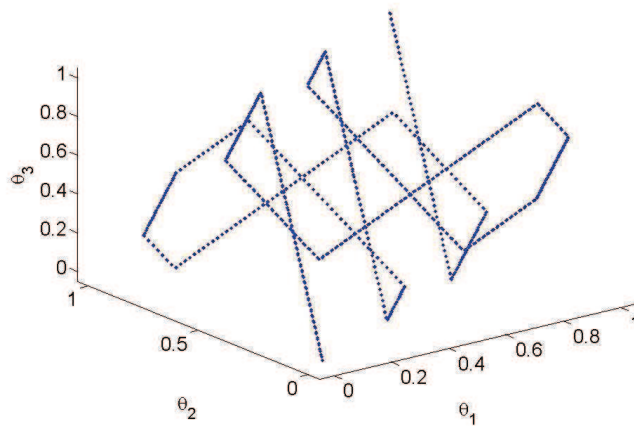


Figure 2-3. Exemple de courbe définie par l'Equation (2-4) pour 3 paramètres suivant une loi uniforme.

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 θ_i , $i=1, \dots, P$, à différentes fréquences ω_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 Λ_r correspondant à la fréquence ω_i de θ_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 θ_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 θ_i et $\theta_{\sim i}$ est ainsi concentrée autour des pics de fréquences ω_i et il considère que l'information sur tous les effets d'interaction impliquant $\theta_{\sim i}$ et non θ_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 θ_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 ω_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 θ_i :

- (i) on choisit une valeur ω_i telle que sa valeur maximale possible ω_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 θ_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[®] 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 θ . 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 :

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

- (i) génération aléatoire de N vecteurs θ_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 θ_n générés,
- (iii) calcul de la probabilité a posteriori associée au vecteur θ_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 θ_i , $\bar{\theta}_i^{post} = \sum_{n=1}^N \theta_{i,n} p_n$

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

- la corrélation entre deux paramètres θ_i et θ_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 θ_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 (θ_i, θ_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 ε_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 ε_k sont indépendants entre les différentes variables et qu'ils suivent une loi de distribution normale de moyenne 0 et d'écart-type σ_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 θ 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[®], 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 ³)	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

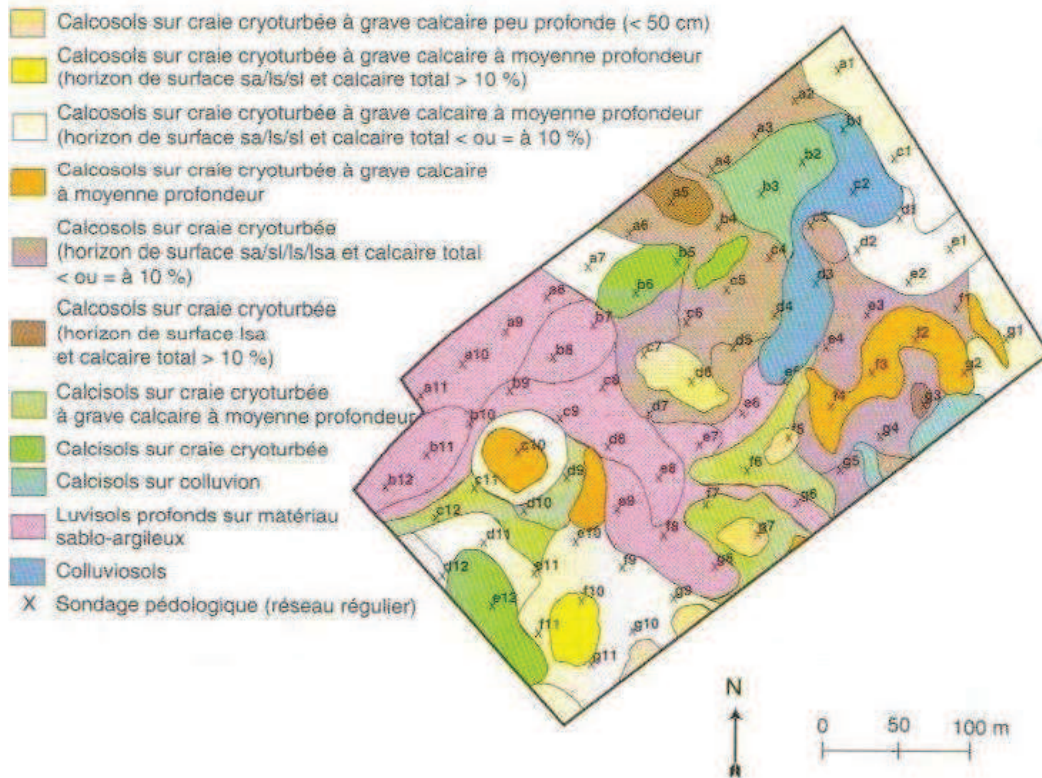
sols (dans notre cas, *HCC* et *HMIN*). Nous avons pour notre part utilisé des FPT établies localement par rapprochement entre unités de sol et propriétés mesurées au laboratoire (*HMIN*) ou in situ (*HCC*, *DA*) (Beaudoin et al., 2007). D'autres fonctions de pédotransfert ont été utilisées pour le calcul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

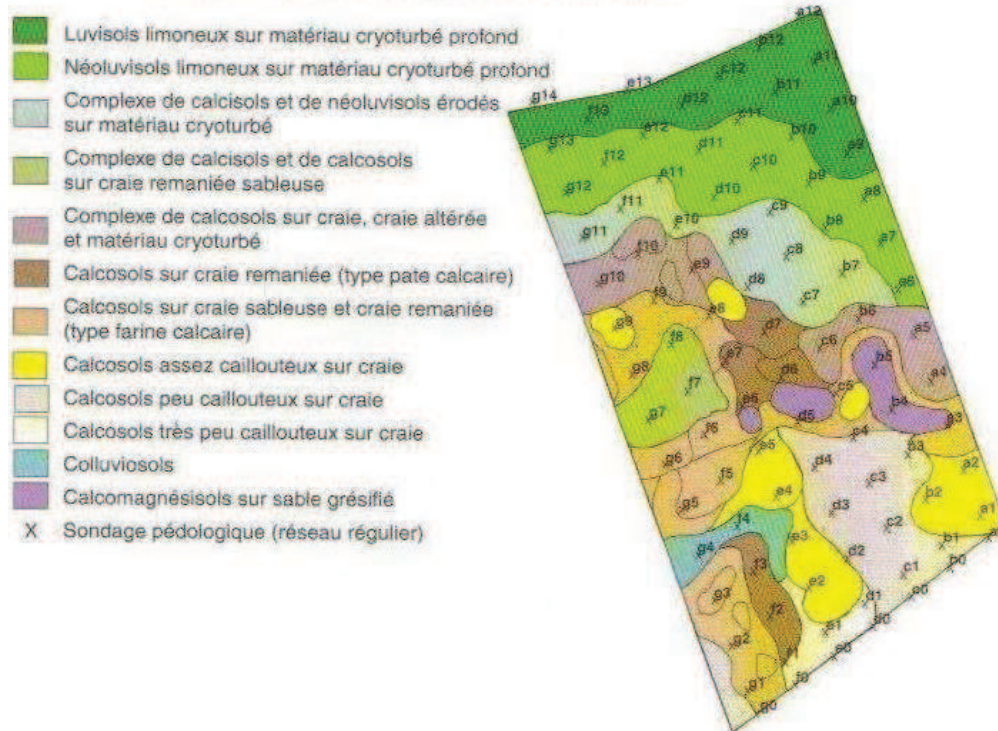


Figure 2-4. Cartes pédologiques à l'échelle du 1/3 000^{ème} des deux parcelles de Chambry. Les emplacements des nœuds de la gille (sondages pédologiques) y sont renseignés (Machet et al., 2007).

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.

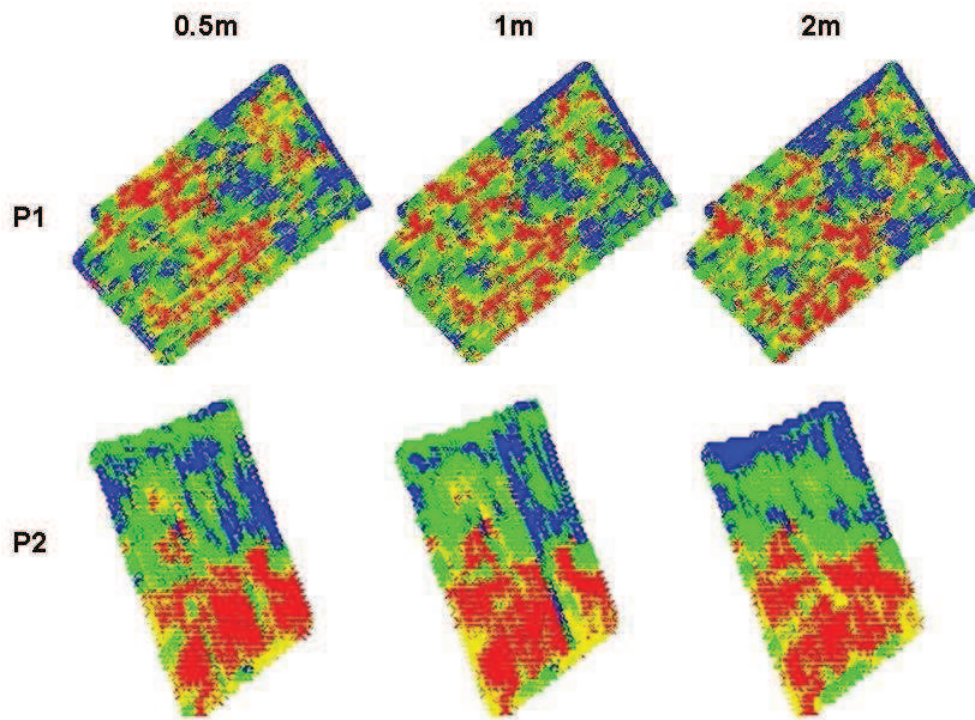


Figure 2-6. Cartes de résistivité électrique des deux parcelles P1 et P2 pour les trois profondeurs de sol.

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^{ème} (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.

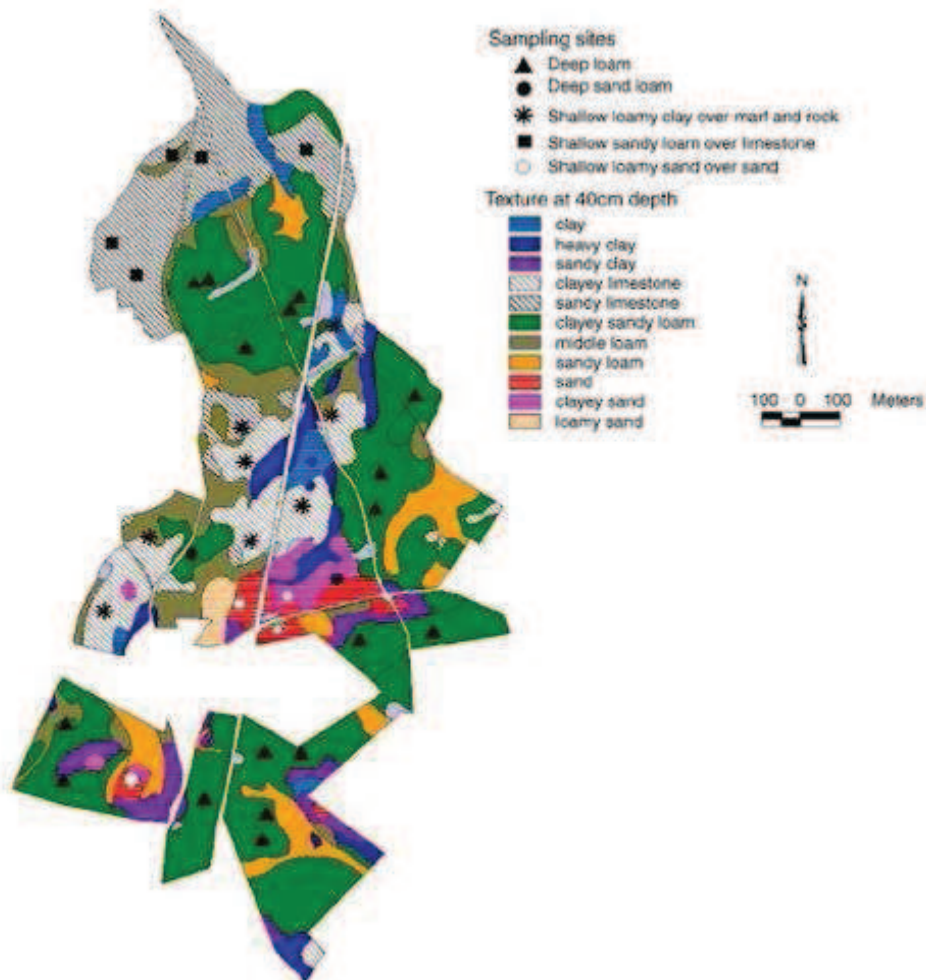


Figure 2-7. Carte pédologique au 1/5 000^{ème} du bassin de Bruyères et localisation des sites (Beaudoin et al., 2005).

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.

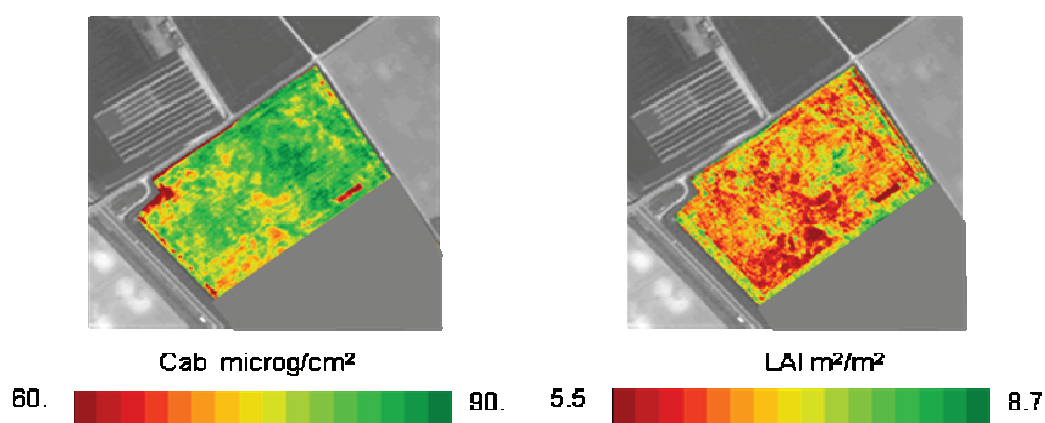


Figure 2-8. Exemple de carte de *LAI* et de *Cab* estimées à partir de l'image CASI du 6 Mai 2000 pour la parcelle P1 de Chambry.

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 m². 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.

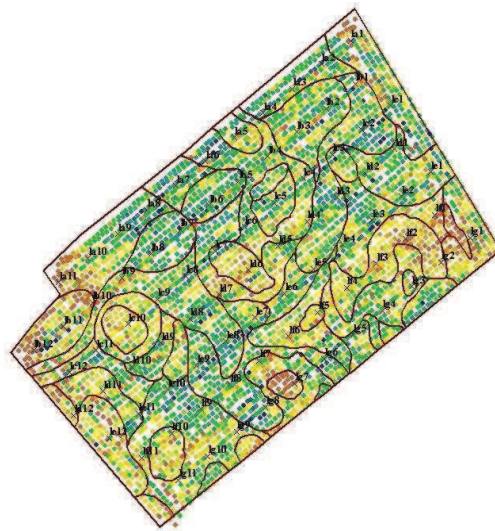


Figure 2-9. Exemple de carte de rendement du blé pour la parcelle P1 de Chambry.

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 m², 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.