



HAL
open science

Anticiper l'assolement pour mieux gérer les ressources en eau : comment valoriser les données d'occupation du sol ?

Delphine Leenhardt, Flavie Cernesson, Jean-François Mari, Delphine Mesmin

► To cite this version:

Delphine Leenhardt, Flavie Cernesson, Jean-François Mari, Delphine Mesmin. Anticiper l'assolement pour mieux gérer les ressources en eau : comment valoriser les données d'occupation du sol?. In-génieries eau-agriculture-territoires, 2005, 42, pp.13 - 22. inria-00000197v1

HAL Id: inria-00000197

<https://inria.hal.science/inria-00000197v1>

Submitted on 7 Sep 2005 (v1), last revised 28 Apr 2010 (v2)

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.

Anticiper l'assolement pour mieux gérer les ressources en eau : comment valoriser les données d'occupation du sol ?^[DBL1]

Delphine Leenhardt *, Flavie Cernesson **, Jean-François Mari *** et Delphine Mesmin */**

*Institut National de la Recherche Agronomique, UMR 1248 AgRosystèmes Cultivés et Herbagers, BP 27, 31326 Castanet Tolosan Cédex

**Ecole Nationale du Génie Rural et des Eaux et Forêts, UMR TETIS Territoires, Environnement, Télédétection et Information Spatiale Cemagref-CIRAD-Engref, Maison de la Télédétection, 500 rue J-F Breton, 34093 Montpellier Cédex 5

*** Université Nancy 2, Laboratoire Lorrain de Recherche en Informatique et ses Applications, Campus scientifique, BP 239, 54506 Vandoeuvre Cedex

RÉSUMÉ

La connaissance de l'assolement d'une région constitue un enjeu important pour les gestionnaires de la ressource en eau car cette information est un facteur clé de l'estimation des prélèvements en eau d'irrigation. Cet article restitue une expérience exploratoire ^[DBL2] basée sur le croisement de données d'occupation du sol échantillonnées dans le cadre de l'enquête Ter-Uti et de données d'occupation du sol acquises de manière exhaustive par enquête de terrain et sur une modélisation stochastique des règles de successions culturales. La discussion qui s'ensuit évalue la portée de cette étude exploratoire en fonction de nos premiers résultats et présente sur les possibilités qu'offre la télédétection.

SUMMARY

Anticipating regional land use : how taking advantage of land use data ?

Knowing the land use of a region is an important issue for water managers since it is a key factor for estimating irrigation withdrawals. This paper presents an exploratory work which involves sampled land use data from the Ter-Uti survey, exhaustive land use data from land survey and stochastic modelling of crop rotation rules. The limits of the present study, due to few case studies, are discussed, as well as the potentialities of the approach for managers and the opportunity offered by remote-sensing for improving the proposed methodology.

I. Introduction

Les gestionnaires de la ressource en eau ont besoin d'information précoce sur l'assolement du territoire. Dès le mois de juin (début de la saison d'irrigation) ils cherchent à prévoir les prélèvements en eau d'irrigation dont l'une des entrées clé est l'assolement. En effet, il est nécessaire de savoir quelles sont les cultures en place, si elles sont irriguées et quelle est leur extension en surface pour évaluer les besoins en eau d'irrigation. Les gestionnaires de la ressource en eau ont donc besoin d'une carte d'occupation du sol (COS) dès le début de la saison d'irrigation.

Une méthode classique de détermination des cartes d'occupation du sol est basée sur l'utilisation d'images satellites de dates différentes. Une image tardive d'été est nécessaire pour distinguer les cultures d'été dont les besoins d'irrigation sont très différents (par exemple maïs et tournesol). Cette méthode ne permet donc pas de satisfaire le besoin d'information précoce des gestionnaires de la ressource en eau. On recherche donc d'autres sources d'information, disponibles avant juin.

Parmi les données potentiellement utilisables, on retient les relevés annuels systématiques d'occupation du sol selon un échantillonnage statistique comme les données Ter-Utili (voir encadré n°1). En effet, l'analyse spatio-temporelle de ces données permet de définir les règles de successions culturales et de répartition spatiale de ces successions (Benoit et al., 2001). Pour satisfaire le besoin en COS précoce des gestionnaires de la ressource en eau, on veut étudier la possibilité de coupler ces règles avec une carte d'occupation du sol de l'année précédente

Cet article présente une première étude, au caractère exploratoire, visant à coupler ces différentes sources d'informations à l'aide d'outils mathématiques récemment développés en vue de fournir aux gestionnaires de l'eau des outils de prédiction précoce des assolements. Dans [DBL3] une première section le principe d'une démarche de couplage entre une carte d'occupation des sols de l'année N et des probabilités de transition entre cultures issues de modèles markoviens appliqués aux données Ter-Utili pour prédire la carte d'occupation du sol de l'année à venir (N+1). Dans la section suivante, nous présentons un premier cas d'application. Il concerne la prédiction de l'occupation du sol du bassin versant du Sousson (Gers) pour 1996. Le choix de ce cas d'étude a été dirigé par l'existence de cartes d'occupation du sol obtenues par relevés de terrain exhaustifs. Il n'a donc pas mobilisé l'utilisation d'images satellites mais nous permet de tester la méthode en toute connaissance de cause. Dans la troisième section une discussion sur l'utilisation d'images satellites s'insère dans une discussion plus générale sur les enseignements tirés de ce premier cas d'étude et sur l'utilisation de la méthode présentée dans une perspective opérationnelle de gestion de l'eau.

II. Principe de la démarche

La méthode envisagée consiste à déterminer la carte d'occupation du sol de l'année N+1 à partir de différentes sources d'information : la carte d'occupation du sol de l'année N, et des probabilités de passage d'une culture à l'autre.

1. Collecte des données d'occupation des sols

Les cartes d'occupation du sol peuvent être obtenues par relevés de terrain ou par traitement d'images satellites. Les cartographies de terrain sont rares car le travail est long et fastidieux. Les COS issues du traitement d'images offrent un potentiel intéressant car l'identification de l'occupation des sols est plus facilement automatisable, et l'on peut traiter des surfaces plus grandes. Les méthodes de détermination de l'occupation du sol, à partir d'images satellites sont en constante évolution, du fait du développement continu de nouveaux capteurs. On trouvera dans Girard et Girard (1999) une description des méthodes actuellement utilisées.

2. Détermination des probabilités de transition entre cultures

Les probabilités de transition d'une culture à l'autre résultent des modèles de Markov cachés (HMM – comme “ Hidden Markov Model ”) appliqués aux données Ter-Uti.

Les données Ter-Uti rassemblent des informations au sujet de la nature de l'occupation du sol en des points précisément identifiés et localisés (Slak, 1997). Ces informations regroupent les notions de spatialité, du fait de la localisation relative des points les uns par rapport aux autres, et de temporalité du fait du renouvellement annuel des relevés. (Encadré 1_[DBL4])

Les modèles de Markov cachés sont des modèles stochastiques adaptés à l'analyse de séquences. Ils appartiennent à la famille des techniques de “ fouille de données temporelles ” (Remaki et Meunier, 2000). Ils ont été déjà appliqués à la détection de successions culturelles en valorisant la base de données Ter-Uti (Benoît et al, 2001, Mari et Le Ber, 2004). On peut ainsi représenter les observations temporelles et spatiales d'occupation du sol comme des successions d'états où les transitions entre états dépendent, suivant l'ordre du modèle, de l'état courant et des n états précédents (Encadré 2_[DBL5]). Ainsi, si la base de donnée Ter-Uti porte sur les années $N-20$ à N , on pourra déterminer les probabilités de transition P_{ij}^n d'une culture i l'année $n-1$ vers une culture j l'année n , avec n variant de $N-19$ à N .

3. Prédiction de l'occupation du sol de l'année à venir

L'occupation du sol pour l'année $N+1$ est estimée à partir de la carte d'occupation du sol de l'année N en s'appuyant sur les probabilités de transition P_{ij}^N précédemment définies pour chacune des cultures i . Ainsi, pour chaque entité spatiale s (pixel ou parcelle), connaissant la culture i en place l'année N , on simule la distribution des cultures j possibles pour l'année $N+1$ par tirages successifs avec remise dans la loi de distribution empirique des transitions (loi multinomiale – cf encadré 3). Nous avons choisi d'effectuer 100 tirages et donc de produire, pour chaque entité spatiale s , 100 estimations d'occupation du sol pour l'année $N+1$, afin de constituer un échantillon consistant des occupations du sol possibles. Simuler 100 estimations d'occupation du sol nous permet d'exploiter les transitions possibles en introduisant du bruit, ce qui ne serait pas le cas si on prenait uniquement la culture dont la probabilité d'apparition est la plus forte. Les COS ainsi obtenues seraient directement exploitables pour simuler les volumes d'eau nécessaires pour l'irrigation. Mais, pour évaluer la

performance de la méthode, nous proposons, dans cette phase exploratoire, de « résumer » ces 100 COS en retenant, pour chaque entité s , la culture j apparaissant le plus souvent au cours des 100 tirages, autrement dit le mode ou encore la transition (i,j) de plus forte probabilité dans le diagramme de Markov résumant l'apprentissage (Figure 3).

Notons que la méthode décrite ici ne prend en compte que l'information temporelle produite par le modèle HMM, et non l'information spatiale, c'est à dire les dépendances spatiales qui pourraient être mise en évidence en étudiant par exemple les relations de voisinage, ou les interactions avec des caractéristiques physiques comme le relief ou la pédologie.

4. Évaluation de la prédiction

La méthode décrite jusqu'ici conduit donc à estimer l'occupation du sol d'une région pour l'année $N+1$ sur la base de l'occupation observée l'année N et des informations contenues dans la base de données Ter-Uti. L'évaluation de cette méthode d'estimation consiste dans un premier temps à vérifier que l'estimation produite est proche de la réalité observée. Une telle évaluation ne peut se faire qu'une fois cette année $N+1$ passée et que si l'on dispose de l'occupation du sol de l'année $N+1$, observée de manière indépendante (par relevés de terrain ou traitement d'images). Dans un deuxième temps, l'évaluation de la méthode peut consister à vérifier que celle-ci présente un avantage par rapport à d'autres méthodes d'estimation (déjà développées et/ou plus simples). Ici on s'attachera à comparer notre méthode (modèle HMM + COS de l'année N de la zone d'étude) à une estimation déduite simplement de la carte de l'occupation du sol de l'année N (ce qui revient à considérer que chaque parcelle est reconduite avec la même culture).[DBL6]

Dans ce papier, l'évaluation de la méthode a porté sur l'ensemble de la zone d'étude. On cherche à savoir (i) si l'assolement global de la zone est respecté et (ii) si, pour l'ensemble de cette zone, l'occupation du sol de chaque parcelle est correctement estimée. Pour cela deux critères d'évaluation C_1 et C_2 ont été définis. Ils sont calculés grâce aux données disponibles pour chaque entité spatiale s : deux valeurs observées d'occupation du sol (pour les années N et $N+1$) et 100 valeurs d'occupation du sol estimées pour l'année $N+1$ à partir desquelles est déterminée le mode.

Le premier critère C_1 est défini comme la racine carrée de l'écart quadratique moyen de nombre de parcelles correspondant à chaque occupation du sol :

$$C_1 = \frac{1}{S} \sqrt{\sum_{k=1}^K n(k) - \hat{n}(k)^2}$$

avec S , le nombre d'entités spatiales, k l'une des K cultures, $n(k)$ et $\hat{n}(k)$, les nombres d'entités spatiales portant la culture k respectivement observé et estimé. On veut ainsi évaluer si la COS estimée possède la même distribution de culture que la COS observée, sans se préoccuper du positionnement des cultures au sein de la zone d'étude.

Le deuxième critère C_2 correspond à la proportion de parcelles bien prédites par la méthode

d'estimation. Il vient donc en complément du critère C1 en intégrant le positionnement spatial.

Pour le calcul de ces deux critères, on procède soit en utilisant directement les 100 estimations, soit en utilisant leur mode. Dans le premier cas on dispose de 100 valeurs des critères C1 et C2 qu'on résume par leur moyenne respective.

Pour vérifier la pertinence de la méthode, on travaille tout d'abord sur l'échantillon de données Ter-Uti qui a servi à déterminer les probabilités de transition, ce qui nous permet de nous abstraire des erreurs qui pourraient être liées à des problèmes de représentativité de la zone d'étude par l'échantillonnage Ter-Uti. Dans un second temps, on s'intéresse à la zone d'étude considérée.

Enfin, on peut calculer les critères C1 et C2 non plus en se basant sur le nombre de parcelles mais sur leur surface, ce qui permet une évaluation plus réaliste pour une utilisation pratique.

III. Une première application au BV du Sousson

1. Cartes d'occupation du sol 1995 et 1996

Sur le bassin versant du Sousson (118 km²), nous avons la chance d'avoir les COS issues de relevés de terrain exhaustifs pour les années 1995 et 1996 réalisés par la CACG, comptant 3710 parcelles agricoles. (Puech et al. 1999)

Le bassin versant du Sousson peut être considéré comme exclusivement agricole (Fig. 2 a et b). Aucune industrie ou commune de plus de 200 habitants n'y est établie, 4% de la superficie totale est occupée par des habitations ou des bâtiments. Le bois occupe une part importante du bassin versant (22% en 96), particulièrement la rive droite. Près de 74% de la superficie est consacrée aux productions agricoles ; les cultures principales étant le blé d'hiver et le maïs (respectivement 15% et 17% de la surface totale), les prairies temporaires représentent 27% de la surface.

2. Description des rotations

Le modèle HMM a été appliqué à un échantillon Ter-Uti comportant 755 points et correspondant à une zone plus grande que le seul bassin versant du Sousson. La figure 1 représente les successions culturales mises en évidence par le modèle HMM. Ce graphique traduit l'existence d'un ensemble de couvert pérenne composé des bois et des roches qui n'échangent aucune surface avec les autres couverts sur toute la période. De plus, apparaissent quatre types de monoculture : maïs, prairies permanentes productives, jachères et le couple blé+orge. Le tournesol alterne uniquement avec le couple blé+orge alors que celui-ci précède et suit de façon très régulière le maïs et de façon discontinue le colza ainsi que les cultures non individualisées. Le maïs est une culture majoritairement reconduite chaque année, mais peut faire quelques échanges avec le couple blé+orge, avec les jachères en début de période et irrégulièrement avec les cultures non individualisées. Les oléagineux interviennent ponctuellement en rotation avec le maïs. Les cultures non individualisées comprennent majoritairement des prairies temporaires mais aussi des occupations du sol traduisant une activité

agricole et des occupations du sol non agricoles comme les autres sols ni altérés ni bâtis, sols artificialisés non bâtis ou les eaux permanentes et zones humides. On constate que dans la majorité des cas, la reconduction de la culture est l'évolution la plus probable, excepté pour le tournesol et le colza.

3. Estimation de l'assolement pour 1996

Les probabilités de transition retenues sont celles de l'année 1995, (transition année 1994 et année 1995 cf. diagramme de Markov Fig. 1).

L'application du programme d'estimation produit 100 cartes d'occupation du sol ; nous choisissons de résumer les différentes estimations, pour chaque parcelle, par la culture qui apparaît le plus fréquemment (le mode). On recense ainsi 1097 parcelles de non agricole (bois essentiellement), 56 parcelles de soja, 64 de tournesol, 714 de maïs, 423 de blé et orge, 1148 de prairies temporaires et 62 parcelles en autres cultures ; le colza n'apparaît pas comme une culture majoritaire, il n'est donc pas représenté sur la carte (Fig. 2c). En terme de surface, on obtient les chiffres suivants : les parcelles non agricoles couvrent 3109 ha, le maïs et les prairies temporaires dominent avec respectivement 2565 et 3902 ha, puis viennent le groupe blé-orge (1363 ha), le tournesol (208 ha) et le soja (145 ha).

4. Evaluation

Pour vérifier la pertinence de la méthode étudiée (HMM+ COS 1995), nous avons calculé les critères C1 et C2 sur les données Ter-Uti (tableau 1). On constate que si l'on retrouve bien globalement les proportions de culture observées (seulement 2% d'erreur sur la carte « moyenne » et 1.3% d'erreur sur la carte « modale »). Par contre, la probabilité que l'entité spatiale soit bien classée est améliorée dans le cas de la carte « modale » mais dégradée pour la carte « moyenne ». On met ainsi en évidence l'importance du choix du descripteur : la moyenne implique une normalité apparemment non vérifiée dans ce cas alors que le mode n'exige pas cette condition et permet de valoriser au mieux l'information issue des 100 tirages.

Sur le Sousson, on constate des valeurs moins performantes pour les deux critères, puisque l'erreur d'assolement est de 8.6%, avec seulement 60% de parcelles bien estimées.

On peut proposer comme premières pistes d'explication :

- la nature de l'échantillonnage Ter-Uti qui a une granularité importante (1 point = 100 Hectares) dans l'espace. La structure de l'échantillonnage Ter-Uti et la structure du paysage peuvent générer une sous-représentation de certaines zones ;
- les probabilités ont été calculées sur une zone plus grande que le bassin du Sousson et dont l'assolement peut être différent ;
- la non-prise en compte de contraintes spatiales qui génère des erreurs de positionnement (cf. critère C2).

Tableau 1 : Critères d'évaluation C1 et C2, calculés sur les échantillons Ter-Uti et Sousson, pour différentes méthodes d'estimation de l'occupation du sol pour l'année 1996 : (i) par simple reconduction de l'occupation du sol de 1995, (ii) par application de la méthode HMM + COS 1995 et en résumant le résultat des 100 estimations obtenues par (a) la moyenne des 100 critères (moyenne) ou (b) le critère correspondant à la carte des cultures les plus fréquentes (mode). Les valeurs entre crochets correspondent aux calculs de C1 et C2 rapportés aux surfaces des parcelles (valeurs disponibles seulement pour l'échantillon Sousson).

Echantillon	Méthode d'estimation	Critère C_1 % erreur	Critère C_2 % entités bien classées
Ter-Uti (755 points)	Reconduction de l'occupation du sol de 1995	2.4	67
Ter-Uti (755 points)	HMM + COS 1995 (moyenne)	2.0	59
Ter-Uti (755 points)	HMM + COS 1995 (mode)	1.3	69
Sousson (3710 parcelles)	Reconduction de l'occupation du sol de 1995	4.6 [4.3]	57 [61]
Sousson (3710 parcelles)	HMM + COS 1995 (moyenne)	7.4	49
Sousson (3710 parcelles)	HMM + COS 1995 (mode)	8.6 [8.7]	59 [64.5]

Considérons maintenant l'apport de la méthode HMM+COS 1995 par rapport à la reconduction simple de la COS 1995. S'il apparaît de manière globale sur l'échantillon Ter-Uti (C_1 égal à 1.3% au lieu de 2.4%), il semble masqué sur le Sousson par le problème d'échantillonnage évoqué plus haut (C_1 égal à 8.6% au lieu de 4.6%). Si l'on s'intéresse à l'occupation du sol de chaque entité spatiale (critère C_2), la méthode HMM+COS 1995 conduit à une amélioration de l'estimation (3%) à condition d'utiliser comme descripteur le mode. Cette amélioration est constatée sur les deux échantillons.

On constate une grande stabilité de l'assolement : respectivement 67% et 57% des parcelles ont le même usage en 1995 et 1996 pour l'échantillon Ter-Uti et le bassin du Sousson. Ces chiffres passent respectivement à 46% et 40% si l'on ne considère que les entités agricoles.

En terme de surface, le critère C_1 est du même ordre de grandeur que celui estimé à partir du nombre de parcelles. En ce qui concerne le critère C_2 , l'amélioration est du même ordre (3.5%) puisque respectivement 61% et 64.5% de la superficie totale du Sousson sont correctement prédites par la reconduction simple de la COS 1995 et par la méthode HMM+COS 1995. La méthode HMM+COS 1995 permet de reconstituer 52% de la surface agricole, présentant une augmentation de 4% par rapport à la reconduction de la COS 1995.

IV. Discussion-Conclusion

Le présent article présente une méthode d'estimation précoce de l'assolement. Celle-ci a été testée sur un unique cas d'étude (le bassin versant du Sousson pour la période 1995-1996) afin de fournir des premiers éléments d'appréciation. La portée de ce test a de fait été limitée par la disponibilité de données fiables d'occupation du sol sur deux années consécutives. Le test de la méthode mériterait d'être étendu à d'autres cas d'étude, en s'appuyant par exemple sur des cartes d'occupation du sol obtenues par télédétection, mais en veillant, dans ce cas, à ce que l'erreur introduite par cette technique d'observation de l'occupation du sol soit bien identifiée en tant que telle.[DBL7]

Si la méthode considérée dans cet article présente finalement un assez faible gain par rapport à la reconduction de l'assolement de l'année précédente, on peut néanmoins conclure de ce premier travail que la méthode décrite (HMM + COS N) permet d'estimer l'assolement global de la zone pour l'année N+1 assez correctement (moins de 10% d'erreur en utilisant comme descripteur la valeur modale). Toutefois cette méthode d'estimation ne permet de reconstituer une localisation correcte de l'assolement que pour 60% de la surface. Ce résultat peut être lié à différents facteurs :

- l'échantillonnage retenu : le Sousson est une zone relativement restreinte par rapport à la zone qui a servi à établir les probabilités de transition ;

- le choix des probabilités de transition : on a travaillé uniquement à partir de la transition 94-95, il pourrait être intéressant de travailler à partir de transitions sur 3, voire 4 années ;

- toutes les transitions envisagées sont temporelles : du fait qu'aucune contrainte spatiale n'a été prise en compte, il est normal que le positionnement des parcelles ne soit pas spécialement bon ainsi que le montre le critère C2 ; la prise en compte des contraintes spatiales permettrait un meilleur positionnement des parcelles et justifierait pleinement l'utilisation du critère C2 ;

- de plus, la méthode ne permet pas de détecter des changements liés à des facteurs exogènes comme une prime pour une culture accordée pour une année donnée, la volonté personnelle d'un agriculteur... [DBL8]

Pour prendre en compte des changements soudains d'occupation du sol dus à des facteurs exogènes, plusieurs solutions seraient possibles. La première, qui consisterait à intégrer ces facteurs comme nouvelles variables cachées dans le modèle HMM, semble encore difficile, à la fois pour des raisons de disponibilité de données qu'en raison des modélisations nécessaires. La seconde, qui semble plus directement opérationnelle, serait d'utiliser une image satellite en début de saison (printemps N+1). A une telle date, il est possible de savoir si le sol est nu ou non. Cette information est importante car elle permet de réduire le nombre de cultures possibles sur la parcelle étudiée: si la parcelle est détectée en sol nu, on exclura les cultures d'hiver; inversement si la parcelle est en végétation, on exclura les cultures d'été. Ce type d'information pourrait soit être intégré dans le schéma de modélisation proposé en modifiant les probabilités d'occurrence des différentes cultures, soit, plus

simplement, être utilisé pour supprimer, parmi les 100 cartes simulées, celles qui sont incompatibles avec la photographie satellitaire de printemps.[DBL9]

D'une manière générale, le recours à l'imagerie satellite semble incontournable : il n'est pas imaginable de décrire plusieurs années l'occupation du sol de zones étendues à partir d'enquêtes de terrain. L'information satellite est en effet le seul type de données qui permet une classification de l'occupation du sol sur de grands espaces avec des méthodes automatiques ou semi-automatiques. La qualité des COS ainsi obtenues dépend de nombreux facteurs : la quantité des scènes disponibles et les dates de prises de vue, le relief de la zone d'étude ainsi que les caractéristiques de son paysage, les caractéristiques du capteur, les méthodes de classification et le choix d'une nomenclature.

Pour raisonner l'utilisation de la méthode décrite dans ce papier, avec ou sans imagerie satellite, il est nécessaire de revenir aux besoins des gestionnaires de la ressource en eau, à savoir la prévision d'une quantité globale d'eau à fournir par zone de gestion pendant la campagne d'irrigation.

Ainsi se pose d'abord la question de l'unité d'espace sur laquelle faire les estimations d'occupation du sol. Dans le cas d'un gestionnaire qui doit savoir combien d'eau déstocker d'un barrage, l'unité spatiale d'intérêt est le périmètre irrigué à partir de ce barrage. Il est donc intéressé par une bonne estimation de l'occupation du sol sur ce périmètre (C1 faible). Comme il cherche à évaluer un besoin en eau global sur le périmètre, qui est une valeur agrégée sur l'ensemble des parcelles irriguées, il n'est pas spécialement intéressé à ce que l'estimation de l'assolement soit localement correct (peu importe donc la valeur de C2). Ce type de raisonnement doit permettre de déterminer le type de capteur satellite utiliser. Ainsi, dans le cas présenté, on pourrait s'appuyer sur des COS établies à partir d'images de très basse résolution spatiale (type programme MERIS) qui permettent de dégager des zones d'occupation du sol homogène. Les COS obtenues à partir d'image de basse résolution (type SPOT) ne semblent pas nécessaire : elles permettent d'obtenir plus de précision sur la répartition spatiale des cultures mais cela ne serait pas valorisé par le gestionnaire.

La méthode proposée a consisté à générer 100 cartes d'occupation du sol. Une alternative aurait pu être de ne générer que la carte des cultures ayant la plus forte probabilité d'apparaître connaissant la culture précédente. Notre approche (simulation de 100 cartes) présente une difficulté d'évaluation: il est nécessaire de disposer d'une synthèse des 100 cartes pour valider l'approche, par comparaison à une carte observée. Par contre, elle présente l'avantage de valoriser toutes les transitions entre cultures possibles avec leurs probabilités associées. A partir des 100 cartes, on dispose donc, pour chaque entité spatiale, d'une distribution de cultures possibles, et par conséquent, d'une distribution des estimations des prélèvements en eau d'irrigation. Cette approche permet donc au gestionnaire de la ressource en eau de disposer d'une estimation de l'erreur associée à la prédiction de volumes en eau, qui peut être utile dans ses prises de décision.

Remerciements

Les auteurs remercient vivement de Messieurs J-C. Teurlay et J. Lataste du Service Régional de Statistiques Agricoles de la Direction Régionale de l'Agriculture et de la Forêt de Midi-Pyrénées sans la coopération desquels cette étude n'aurait pu être réalisée.

Références

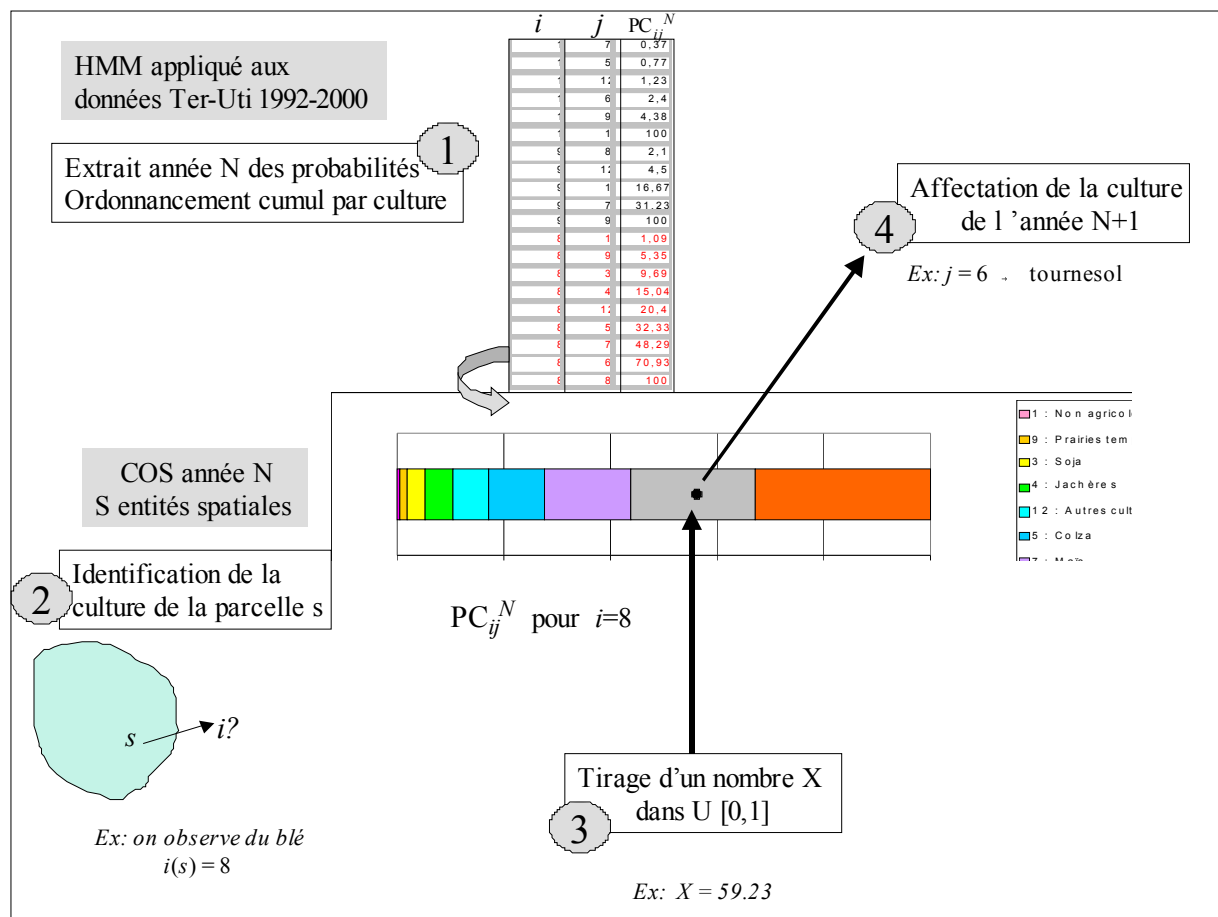
- Benoît, M., Le Ber, F., Mari, J.-F., 2001. Recherche des successions de cultures et de leurs évolutions: analyse des données Ter-Uti en Lorraine. *Agreste vision - La statistique agricole*, 31: 23-30.
- Benoît, M., Le Ber, F., Mari, J.-F., Mignolet, C., Schott, C., 2003. Carrotage, un logiciel pour la fouille de données agricoles. In :Colloque STIC et Environnement SE'2003, Rouen, France. INSA Rouen.
- Girard, M. C. and Girard, C., 1999. Traitement des données de télédétection. DUNOD, Paris. 529 p.
- Mari J.-F., Le Ber, F., sous presse. Temporal and spatial data mining with second-order hidden markov model. *Discrete Mathematics* 16p.
- Remaki L., Meunier J-G., 2000. Un modèle HMM pour la détection des mots composés dans un corpus textuel. *JADT 2000, 5èmes Journées Internationales d'Analyse Statistique des données Textuelles*.
- Puech C., Cernesson F., Balas B. 1999. Approche spatiale de la pollution par les nitrates. Bois et forêts des agriculteurs : actes du colloque de restitution de l'AIP AGRIFOR, Clermont-Ferrand, 20-21 oct. 1999, pp.191-213.
- Slak, M.F. 1997. L'évolution des paysages girondins vue par Ter-Uti. *Agreste-Les Cahiers* 21:23-33.

Figure de l'encadré 3 : Description de la méthode de prédiction de l'occupation du sol de l'année N+1 à partir d'information TerUti et d'une carte d'occupation du sol de l'année N.

Fig. 1 : Probabilités de passage d'une culture (ou groupe de cultures) à l'autre. Diagramme issu du programme HMM appliqué aux données Ter-Uti de la zone Sousson-Baïse (755 points) sur la période 1992-2000.

Fig. 2 : Cartes d'occupation du sol de la zone du Sousson : (a) observée en 1995 –COS 1995, (b) observée en 1996 –COS 1996, (c) estimée pour 1996 en utilisant la valeur modale des 100 estimations résultant de la méthode HMM+COS95.

Figure de l'encadré 3



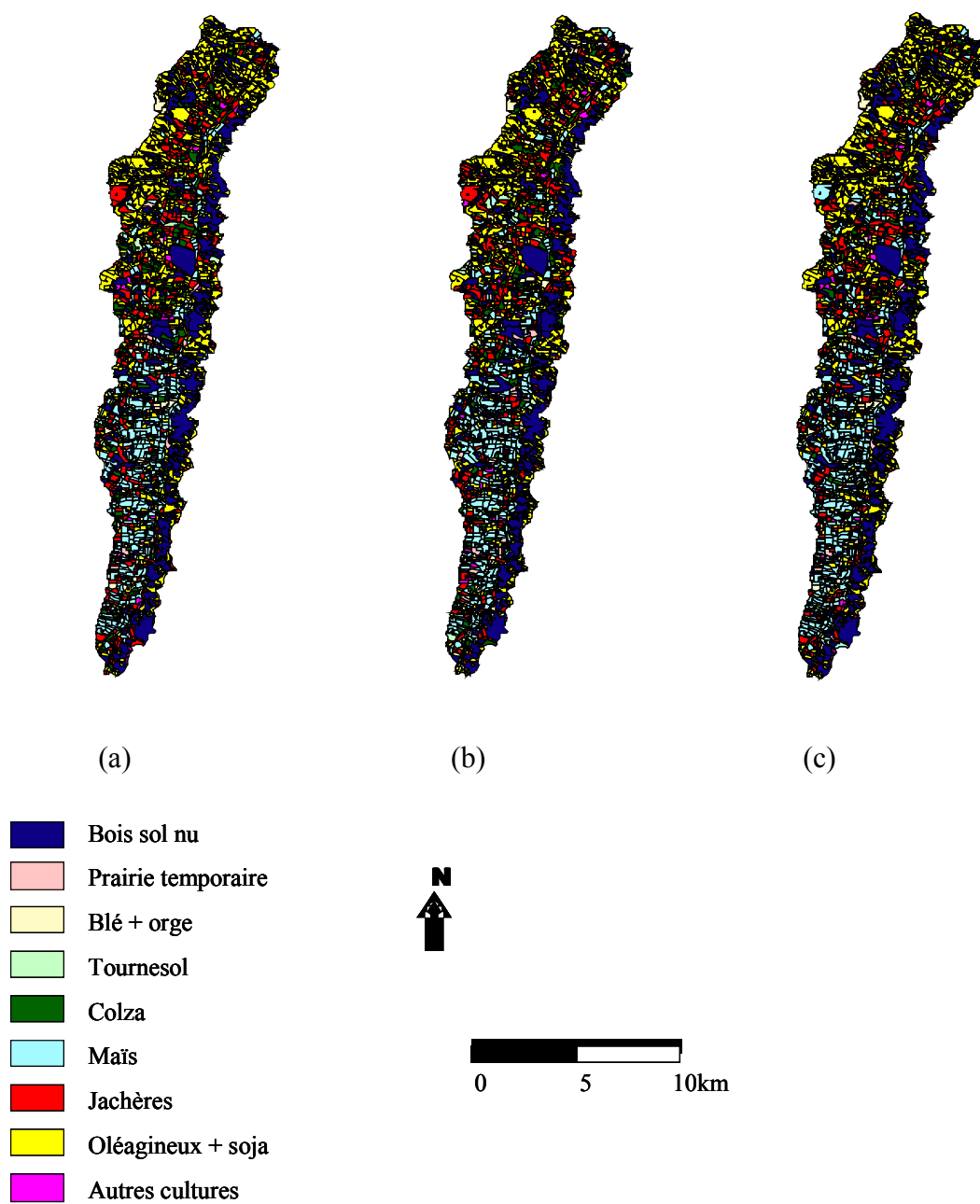
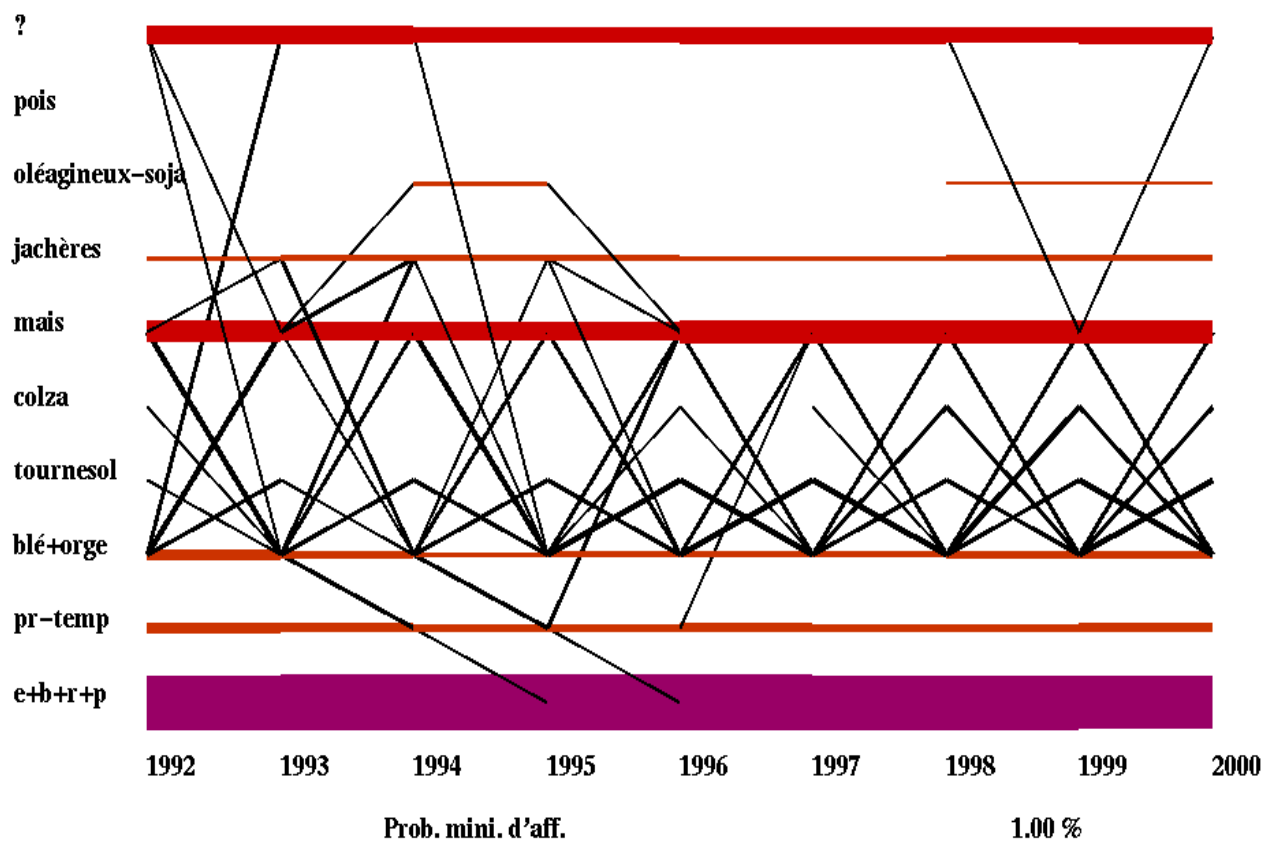


Fig 2.

Fig. 1



Encadré 1: les données Ter-Uti

Les données Ter-Uti sont collectées par le ministère de l'agriculture sur l'ensemble du territoire métropolitain. Elles représentent l'occupation du territoire chaque année. Un premier échantillonnage, effectué par l'IGN consiste à sélectionner des photos aériennes afin de couvrir la région étudiée. Deux niveaux de résolution sont obtenus. Chacune des 3820 mailles couvrant le territoire national (Fig. a) contiennent 4 photos aériennes (Fig. b). Chaque photo aérienne couvre un carré de 2 Km de côté (Fig. c). A l'intérieur, est appliquée une grille de 36 points (Fig. c) qui sont visités chaque année par un enquêteur qui relève l'occupation du point. Celle-ci (blé, tournesol, bois ou encore, zone humide, zone bâtie) est enregistrée dans un tableau où les lignes représentent les sites visités et les colonnes les années de visite. Dans notre étude nous avons 49 modalités. Un point Ter-Uti représente en moyenne 100 hectares.

Encadré 2 : Modéliser l'évolution des assolements

La modélisation stochastique est un moyen pour prendre en compte la variabilité temporelle et spatiale d'un phénomène complexe tel que l'évolution des assolements dans une région. Un modèle stochastique particulier - appelé modèle de Markov caché (ou HMM comme Hidden Markov Model) - décrit cette évolution à l'aide un automate possédant un ensemble fini d'états. Les transitions entre états sont probabilisées pour rendre compte de l'incertitude sur l'enchaînement des états. Ces probabilités sont constantes, ne dépendent pas du temps. En revanche, elles dépendent de l'état ou des n états visités auparavant. Cet automate représente une chaîne de Markov d'ordre 1 ou n . Les états sont de deux types :

- ils sont associés à une culture ou à une petit groupe d'occupations. On les appelle les états de Dirac ;
- ils sont associés à une distribution de cultures (par exemple la distribution des 49 occupations retenues dans les données Ter-Uti). On l'appelle l'état de réserve.

Un processus de simulation s'appuyant sur un HMM fonctionne de la façon suivante. A chaque pas de temps (chaque année dans notre étude) le processus change d'état selon les transitions possibles. Le choix de l'état d'arrivée à partir d'un état de départ peut se faire par un tirage selon une loi multinomiale. En effet la somme des probabilités de sortie d'un état fait toujours 1. Si l'état atteint est un état de Dirac, le processus émet la culture ou le groupe de cultures caractérisant cet état. Si l'état atteint est une distribution de cultures, un deuxième tirage (à l'aide d'une loi multinomiale) émet à son tour une des 49 occupations. On simule ainsi les occupations successives d'une parcelle.

Un gros avantage des HMM est qu'ils possèdent un algorithme d'estimation des valeurs des probabilités de transition entre états et des distributions associées aux états. (On note P_{ij}^n la probabilité de transition d'une culture i l'année $n-1$ vers une culture j l'année n).

Le résultat de l'apprentissage est représenté Figure 1. Le programme d'apprentissage aligne, pour chaque point Ter-Uti, les différentes occupations annuelles dans les états du modèle et effectue un comptage de l'utilisation des transitions année par année. Ces comptes, normalisés en fréquences, sont représentés par des épaisseurs de traits différentes. Ces fréquences sont les estimations pour les valeurs P_{ij}^n . Chaque état correspond à une ligne, qui représente une culture individualisée ou l'état de réserve, noté "?", qui en rassemble plusieurs. En abscisse, un axe chronologique figure l'évolution temporelle. Un trait représente le passage d'une culture à une autre (c'est à dire une transition) lorsque sa probabilité est supérieure à une probabilité minimale de 1 %. L'épaisseur d'un trait est proportionnelle à la probabilité. Les traits noirs indiquent un changement de culture alors que les traits rouges indiquent la reconduction de la culture ; lorsque le trait rouge est continu et n'a aucun lien avec les autres cultures, il s'agit de parcelles monoculturelles.

On peut se référer à Benoît et al. (2003) pour une description plus complète du système CarrotAge dans lequel sont implantés ces algorithme d'estimation et de visualisation.

Encadré 3 : Prédiction de l'occupation du sol de l'année à venir

L'occupation du sol pour l'année N+1 est estimée à partir de la carte d'occupation du sol de l'année N et du HMM modélisant les successions culturales, selon les étapes suivantes (cf. Figure de l'encadré).

1. Pour toutes les cultures i , on ordonne les P_{ij}^N issues du modèle HMM de manière à pouvoir les cumuler et définir les probabilités cumulées PC_{ij}^N .

Puis, pour chaque entité spatiale s (pixel ou parcelle),

2. on observe la culture i en place l'année N associée à un état de Dirac du HMM,
3. un nombre aléatoire X est tiré dans une loi uniforme dans le domaine $[0,1]$,
4. on affecte à s la culture j dont la probabilité cumulée PC_{ij}^N est égale à X ,
5. les étapes 3 et 4 sont répétées 100 fois.

Cette procédure correspond à 100 tirages d'une loi multinomiale paramétrée par les probabilités de transition au départ de l'état i du HMM l'année N. (cf. Figure 1)