
Essais de cartographie numérique des pédopaysages à 1/250 000 dans les départements de l'Aube et de la Marne

Enseignements, limites et perspectives

B. Laroche⁽¹⁾, D. Arrouays⁽¹⁾, D. Olivier⁽²⁾, et N. Lecerf⁽³⁾

- 1) INRA - Infosol Orléans - 2163, Avenue de la Pomme de Pin - CS 40001 ARDON - 45075 Orléans Cedex 2, France
Bertrand.Laroche@orleans.inra.fr
- 2) Chambre Départementale d'Agriculture de l'Aube - 2 bis rue Jeanne d'Arc - 10018 Troyes, France
- 3) Chambre Départementale d'Agriculture de la Marne - Complexe Agricole du Mont Bernard, Route de Suiippe
51009 Châlons-en-Champagne, France

RÉSUMÉ

La cartographie numérique pourrait permettre une avancée significative dans la constitution des synthèses cartographiques à 1/250 000. L'objectif de ce travail est de proposer différentes méthodes en fonction de la disponibilité des données pédologiques. Deux départements de Champagne-Ardenne ont été choisis en fonction de ce critère. Ils disposent d'une information pédologique à grande échelle, sur la totalité du territoire pour l'Aube, et de manière partielle sur la Marne. Suivant ces configurations, les méthodes de cartographie numérique employées ont été différentes. Dans le premier cas, nous avons utilisé une méthode non supervisée pour faciliter le regroupement de grands ensembles, dans le deuxième cas, nous avons appliqué une méthode supervisée basée sur des secteurs d'apprentissages et l'utilisation d'arbres de classification.

L'application de telles procédures pose la question de la crédibilité à accorder aux résultats. Par l'analyse de ces prédictions et des tests de validation, cet article apporte quelques éléments de réponse en donnant de premiers enseignements sur le potentiel, mais aussi les limites, de ces techniques de cartographie pour une utilisation en routine.

Mots clés

Cartographie numérique, donnée pédologique, synthèse cartographique, pédopaysage, transfert d'échelle.

SUMMARY**ANALYSIS OF DIGITAL SOIL MAPPING TRIALS AT 1,250 000 SCALE ON TWO FRENCH DEPARTMENTS.****Lessons, limits and prospects**

One of the alternatives for a significant progress of soil mapping programmes could be the digital soil mapping. The objective of this work is to propose various methods according to the availability of the pedological data. Two departments (Aube, Marne) of the Champagne-Ardennes region were chosen according to this criterion. They have a pedological information, on the whole territory for the Aube, and in partial areas for the Marne. According to these configurations, the methods of digital soil mapping employed were different. In the first case, we used an unsupervised method to facilitate the grouping of complex units, in the second case, we applied a supervised method based on training sectors and the use of classification trees.

The application of such procedures raises the question of the validity of the results. By analysing these predictions and their validation tests, this article brings first insights onto the potential, but also the limits, of these techniques of digital soil mapping for a use in routine.

Key-words

Digital soil mapping, soil data, synthetic map, soil-scapes, upscaling.

RESUMEN**ENSAYOS DE CARTOGRAFÍA NUMÉRICA DE LOS PEDOPAISAJES A 1/250 000 EN LOS DEPARTAMENTOS DE AUBE Y DE MARNE. Enseñanzas, límites y perspectivas.**

La cartografía numérica podría permitir un adelanto significativo en la constitución de las síntesis cartográficas a 1/250 000. El objetivo de este trabajo está proponer diferentes métodos en función de la disponibilidad de datos pedológicos. Se eligieron dos departamentos de la región de Champaña - Ardenas en función de este criterio. Disponen de una información pedológica a grande escala, sobre la totalidad del territorio para Aube, y de manera parcial para Marne. Siguiendo estas configuraciones, los métodos usados de cartografía numérica estuvieron diferentes. En el primer caso, usamos un método no supervisado para facilitar el agrupamiento de grandes conjuntos, en el segundo caso, usamos un método supervisado basado sobre sectores de aprendizajes y el uso de arboles de clasificación.

La aplicación de tales procedimientos plantea la cuestión de la credibilidad a acordar a los resultados. Por el análisis de estas predicciones y de las pruebas de validación, este artículo propone algunos elementos de respuesta que dan primeras enseñanzas sobre el potencial, pero también los límites, de estas cartografías para un uso de rutina.

Palabras clave

Cartografía numérica, datos pedológicos, síntesis cartográfica, pedopaisaje, transferencia de escala.

Malgré une accélération notable des levés cartographiques depuis la création du Groupement d'Intérêt Scientifique Sol en 2001 (Arrouays *et al.*, 2004), la France ne dispose pas encore d'une couverture cartographique complète de son territoire à une échelle supérieure à 1/1 000 000. La réalisation d'une base exhaustive à 1/250 000, dénommée Référentiel Régional Pédologique (RRP), constitue actuellement la priorité du programme national Inventaire Gestion et Conservation des Sols (IGCS). A cette échelle, les sols sont représentés sous la forme de pédopaysages qui sont des ensembles litho-géomorpho-pédologiques cohérents. Ces entités géographiques, ainsi définies, sont la représentation synthétique de l'ensemble des éléments qui constituent le paysage (roche mère, topographie, occupation du sol) et des caractéristiques des sols (types de sol, profils et horizons pédologiques).

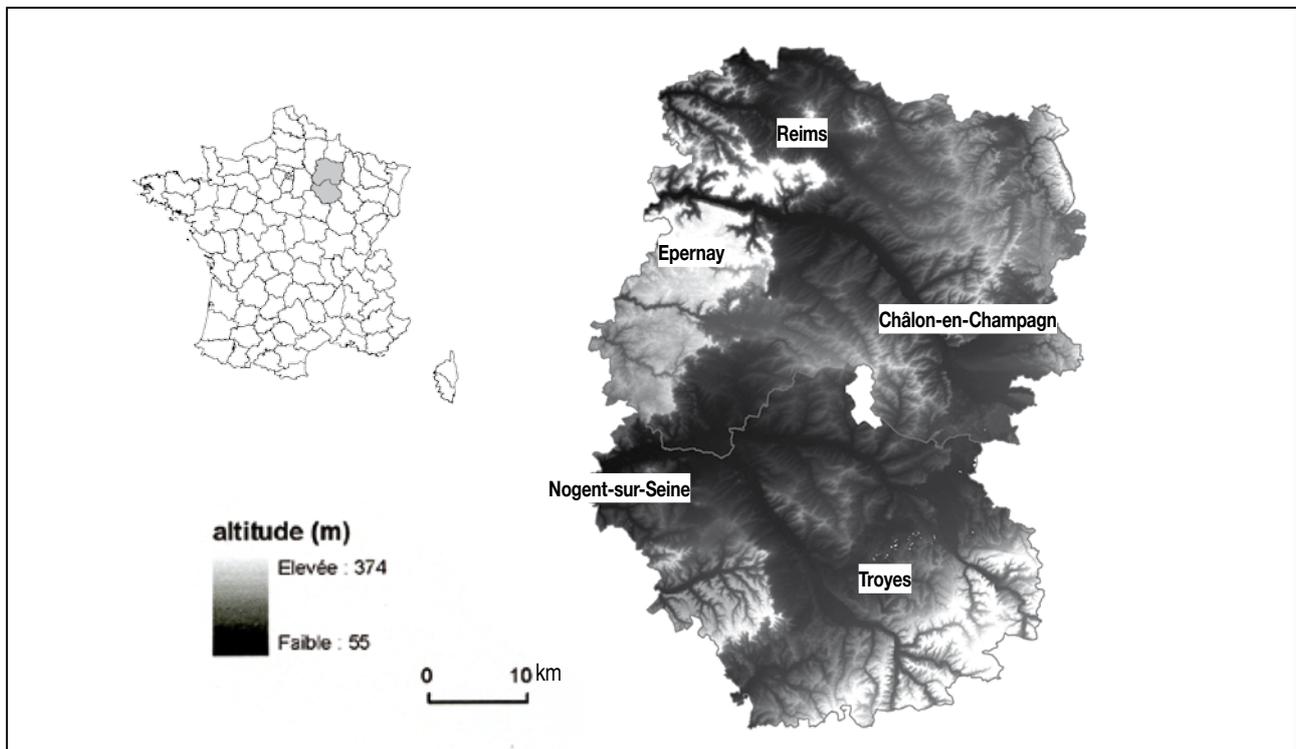
L'achèvement de cette base apparaît d'autant plus urgent qu'elle devrait constituer l'un des supports essentiels à la délimitation des zones défavorisées simples, telles que définies dans le projet de directive de la Commission Européenne (CCE, 2006). Paradoxalement, plusieurs départements français disposent de cartographies réalisées à des échelles beaucoup plus détaillées sur tout ou partie de leur territoire mais n'ont pas encore procédé à une synthèse sous forme d'un RRP validé par le Conseil Scientifique du programme IGCS (par exemple, Vienne, Vaucluse, Aisne, Tarn-et-Garonne, Aube, Cher). Ce transfert

d'échelle vers une synthèse à 1/250 000 pourrait reposer sur les méthodes de cartographie numérique ou Digital Soil Mapping (DSM) (Lagacherie, 1992; McBratney *et al.*, 2003). Les premiers exemples de délimitation automatique des pédopaysages ont été reportés dès 1996 (Legros, 1996). Depuis de nombreux modèles ont été testés dans des environnements différents et à des échelles variées, que ce soit pour estimer des propriétés de sol, des types de sols, ou réaliser des zonages agronomiques (Irvin *et al.*, 1997; Arrouays *et al.*, 1998; Bourenanne *et al.*, 2000; Dobos *et al.*, 2000; Bishop et McBratney, 2001; Gaddas, 2001; Moran et Bui, 2002; Lehmann, 2005; Cole et Boettinger, 2006; Williams, 2007). Dans certains cas, les essais ont mis en évidence des difficultés méthodologiques (Gaddas, 2001) et n'ont par conséquent pas toujours abouti à une synthèse conforme au cahier des charges IGCS (Carré *et al.*, 2008; InfoSol, 2005). Ces méthodes reposent sur des algorithmes statistiques permettant de décoder les lois de distribution de sols utilisées implicitement par le pédologue cartographe à partir des données disponibles et de les quantifier à partir des facteurs de différenciation des sols représentés sous forme numérique. L'extrapolation du modèle aux zones non cartographiées permet de prédire la distribution des sols.

L'objectif de cet article est de réaliser de manière automatique une esquisse cartographique des pédopaysages sur deux départements de la région Champagne-Ardenne (l'Aube et la

Figure 1 - Localisation de la zone d'étude.

Figure 1 - Location of the study area.



Marne) où les connaissances pédologiques sont très différentes : partielle dans le cas de la Marne et exhaustive dans le cas de l'Aube. Les méthodes de cartographie numérique mises en œuvre seront différentes dans ces deux départements tests en fonction de la disponibilité cartographique de l'information pédologique. Dans le cas de la Marne, nous utiliserons une méthode non supervisée, qui permet de définir, à partir des variables environnementales, des ensembles homogènes sans faire intervenir de connaissance *a priori*. Dans l'autre cas, le département de l'Aube sera traité par une méthode supervisée. Il s'agit alors d'extraire les règles de distribution de la variable sol à l'aide de variables environnementales (ex. topographie, géologie) sur des ensembles de points où la variable sol est connue, puis d'appliquer ces règles en dehors du secteur d'apprentissage, renseigné uniquement par les variables environnementales (Lagacherie, 1992). L'élaboration de ces RRP est fondée sur un transfert d'échelle basé sur des cartes existantes au 1 : 25 000 ou au 1 : 50 000. Le résultat attendu doit être conforme aux recommandations du Cahier des charges IGCS (InfoSol, 2005) et en continuité des RRP déjà produits dans les départements limitrophes.

MATÉRIELS ET MÉTHODES

La zone d'étude se situe sur deux départements de la région Champagne-Ardenne que sont l'Aube et la Marne (*figure 1*). La partie centrale de ces deux territoires est caractérisée par des collines ou plaines crayeuses. Elles sont bordées à l'ouest par les plateaux de la Brie et du Tardenois qui forment la cuesta de l'Île de France, et au sud-est par une zone dépressive appelée la Champagne humide qui se prolonge vers le sud-est par les plateaux calcaires du Barrois.

Les deux travaux conduits dans cette région font appel aux données pédologiques existantes.

Outre les données « sols », d'autres informations sont utilisées pour réaliser ces études :

- Un Modèle Numérique de Terrain (MNT) au pas de 90 m (SRTM, Shuttle Radar Topography Mission, (CGIAR-CSI, 2006), à partir duquel différents indices topographiques sont dérivés ; L'occupation du sol obtenue par la base de données Corine Land Cover (UE-SOeS, 2006) (2^{ème} niveau) ;
- La couche géologique (lithologie et formations superficielles lorsqu'elles sont représentées) est une information indispensable mais ces données sont parfois difficiles à acquérir du fait de l'indisponibilité de l'information souhaitée, de son hétérogénéité départementale et de son coût. Pour ce travail, nous ne disposons pas de cette information sur les 2 départements.
- La philosophie des RRP est de s'appuyer d'abord sur une première segmentation du paysage que sont les régions naturelles, même si sa précision est parfois relative. L'Aube dispose d'une esquisse pédologique au 1/250 000, qui découpe le territoire en 21 grands ensembles pédologiques (Benoit-Janin, 1971). Une

carte de 9 petites régions agricoles existe sur la Marne mais son exploitation est difficile du fait que le découpage s'appuie sur des limites administratives.

Dans les deux situations, les données d'entrée du modèle doivent avoir des caractéristiques communes : un système de projection unique, une emprise identique, et la même résolution (pixels de 100 m de côté dans notre cas).

Dans l'optique d'une représentation à 1/250 000, par convention, la plus petite surface représentable et repérable sur une carte est de $\frac{1}{4}$ cm² (carré de 5 mm de côté et/ou un cercle de 2,8 mm de rayon (Boulaine, 1980). En outre, deux arcs parallèles ne peuvent avoir un écartement inférieur à 250 m sur le terrain soit 1 mm sur la carte. Pour une échelle à 1/250 000 et conformément à la règle énoncée précédemment, les plaines alluviales présentant une largeur inférieure à 250 m ne sont pas représentées, elles sont agrégées dans le polygone qui les englobe.

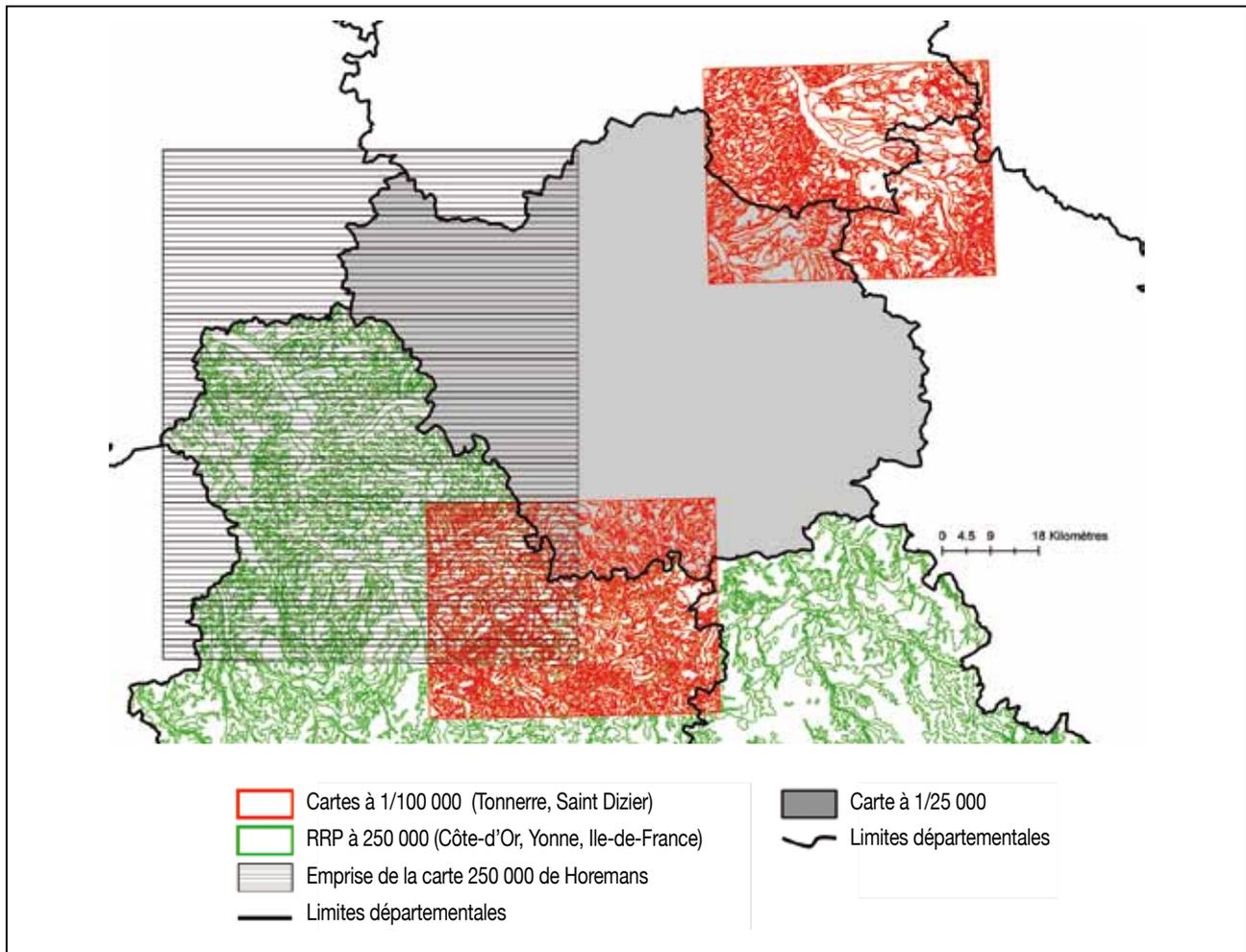
Une synthèse statistique réalisée sur 15 RRP déjà finalisés permet de donner quelques indications sur les caractéristiques des tracés des unités cartographiques de sol (Laroche et Doux, 2008). Des ordres de grandeurs du nombre de polygones, de leur taille moyenne, du nombre d'UCS rencontrés pour un RRP moyen ont pu être établis à partir de cette étude. Le nombre moyen d'UCS est de 88, leur surface moyenne est de 7720 ha et la surface moyenne d'un polygone est de 340 ha.

Définition des pédopaysages à partir d'une information à grande échelle : cas de l'Aube

Le transfert d'échelle tel qu'il a été conduit sur ce département s'est appuyé sur les différentes cartes des sols existantes (*figure 2*) :

- La carte des sols à 1/25 000, synthèse effectuée, au niveau départemental, par Kossura (1999) pour le compte de la Chambre d'Agriculture de l'Aube. Elle est issue d'un travail de synthèse des cartes existantes à grandes échelles, d'un complément par levés de terrain et d'une expertise départementale. Cette synthèse est numérisée et, à chaque polygone, est associé un type de sol décrit par une fiche synthétique (Ouisse *et al.*, 2006). Elle comprend environ 20 000 polygones et est caractérisée par environ une centaine de types de sol.
- Deux cartes pédologiques à 1/100 000 dont l'emprise chevauche en partie le département de l'Aube : Saint-Dizier (Dutil, 1992) et Tonnerre (Baize, 1976).
- Des cartes des sols extérieures au département que sont les Référentiels Régionaux Pédologiques de Côte-d'Or (Chrétien, 1996), de l'Yonne (Baize, 1994) et de l'Île-de-France (Roque, 2003). Il existe aussi une carte à 1/250 000 sur la région parisienne élargie (Horemans, 1984).

L'Aube dispose donc d'une cartographie exhaustive à 1/25 000. La technique de cartographie numérique que nous testons ici est une méthode non supervisée. Nous utilisons la combinaison de deux fonctions ISOCLUSTER et MLCLASSIFY

Figure 2 - Cartes pédologiques disponibles sur le département de l'Aube.**Figure 2** - Soil maps on the Aube department.

de l'extension spatial analyst d'Arc GIS (ESRI, 1999) afin de segmenter le territoire en entités issues de la combinaison des données prédictives que sont le substrat géologique, les dérivées du MNT et l'occupation du sol (*figure 3*). La première fonction permet de définir des classes à partir de l'analyse des valeurs des pixels (méthode des points mobiles), ces classes étant ensuite utilisées par la deuxième commande pour y classer tous les pixels de l'image.

Généralement, un découpage plus ou moins grossier en régions naturelles existe dans la littérature et peut être utilisé comme base de segmentation du territoire. Ceci peut se faire de deux manières, soit en digitalisant les régions naturelles et en les intégrant au SIG comme donnée prédictive, soit en utilisant le nombre de régions comme nombre de classes à créer lors de la première segmentation non supervisée. Dans le cadre de notre travail, la deuxième solution a été choisie. Le nombre de grandes régions naturelles est de 6 et le nombre de petites régions na-

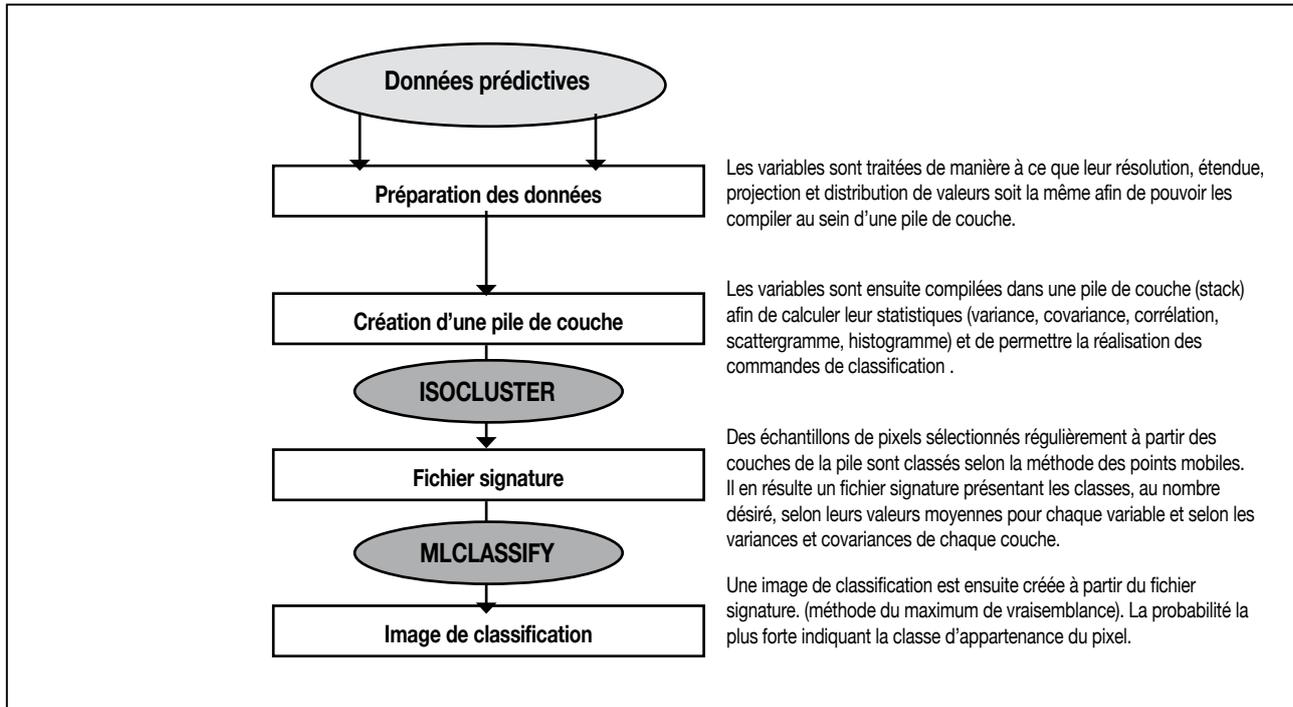
turelles de 20 (Benoit-Janin, 1971). Ces valeurs ont été utilisées pour paramétrer la fonction Isocuster lors du paramétrage du nombre de classes.

Segmentation départementale

La segmentation première en 6 classes (*figure 4A*) permet de définir des grands ensembles. Au Sud-Est on identifie les plateaux Bourguignons. La partie centrale (en bleu) caractérise la Champagne Humide. A l'ouest, dans la partie centrale, se dégage la terminaison du pays d'Othe et en vert la Champagne crayeuse. L'ensemble est découpé par les plaines alluviales de la Seine et l'Aube essentiellement. Ces dernières sont facilement identifiables, elles se distinguent nettement dans le paysage par leur particularité morphologique (pente nulle, occupation du sol différente des versants). Ce découpage en 6 classes fournit les grands éléments structurants de la future carte, en permettant de positionner des limites d'unités cartographiques particulièrement

Figure 3 - Démarche générale de la méthode non supervisée utilisée.

Figure 3 - General approach for the unsupervised method.



nettes. Pour poursuivre le travail de segmentation, il est possible de travailler sur l'ensemble du territoire, en se basant cette fois sur un nombre de classes de 20 (*figure 4B*), ou de travailler par petite région naturelle obtenue par la classification supervisée.

Segmentation par région naturelle et tracé des unités cartographiques

Les régions naturelles identifiées lors de la segmentation départementale présentent des caractéristiques du milieu naturel qui leur sont propres. Une seconde classification non supervisée est ensuite réalisée de manière indépendante et plus détaillée sur chaque région naturelle. Un exemple est présenté sur la région des plateaux calcaires bourguignons (*figure 5*) où 10 classes ont été extraites.

Le tracé des pédopaysages se fera sur la base de cette classification, tout en respectant les contraintes cartographiques d'homogénéité et de continuité des tracés des Unités Cartographiques de Sols sur l'ensemble du département.

Dans cette région naturelle, certaines morphologies bien caractéristiques telles que les plateaux et les versants peuvent être facilement distinguées par la combinaison des facteurs de la pédogenèse. Elles sont alors facilement identifiables à partir du résultat de la classification. Ces grands ensembles homogènes sont alors extraits et identifiés comme des unités cartographiques. Il est parfois pertinent de segmenter de nouveau ces

ensembles en fonction des données environnementales disponibles. L'information sur les sols peut alors intervenir dans le découpage par l'intermédiaire des données sur l'organisation des sols fournies habituellement par les sondages, ou sous la forme des plages cartographiques empruntées à la carte à 1/25 000 dans le cas du département de l'Aube.

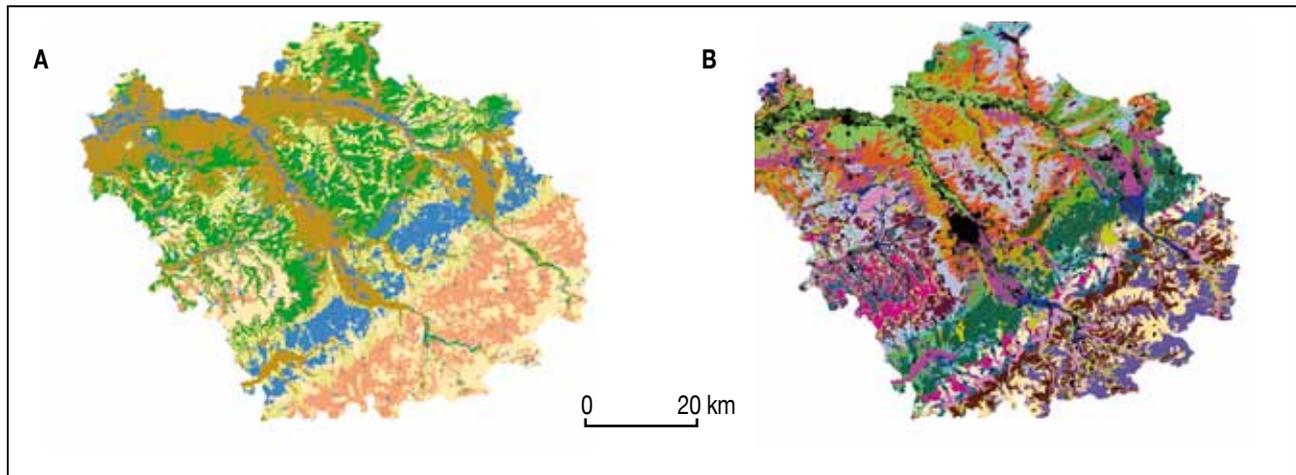
Renseignement des unités cartographiques par l'information pédologique disponible

Ces travaux de segmentation automatique constituent la base du futur découpage et sont utilisés comme un masque pour l'agrégation des polygones de la carte pédologique à 1/25 000 du département de l'Aube. On regroupe ainsi géographiquement plusieurs polygones de la carte à 1/25 000 au sein d'un polygone unique du 250 000 du fait de leur ressemblance en terme de topographie, de géologie et d'occupation du sol, en s'appuyant sur la carte des prédictions.

Avant cette phase d'agrégation, il est nécessaire de préparer la couche géométrique pédologique pour que tous les polygones aient une représentativité significative à 1/250 000. Les polygones trop petits ou dont les arcs sont pratiquement parallèles et distants de moins de 250 m sur le terrain (soit 1 mm sur la carte) sont agrégés au sein des polygones les englobant. Les unités concernées par le deuxième cas sont toutes les unités effilées (plaines alluviales, thalwegs, colluvions). Par exemple, l'unité de

Figure 4 - Segmentation du département de l'Aube (a) en 6 classes et (b) en 20 classes par une classification non supervisée.

Figure 4 - Results of the segmentation of the Aube department. Segmentation départementale.



sol (1/25 000) 31.1 (Sols peu évolués colluvionnaires hétérogènes de pente) caractérisée par 1543 polygones a été entièrement intégrée aux polygones contigus. Il est possible de réaliser cette action de manière automatique en paramétrant ArcGis, en définissant une surface minimum de polygones à conserver. Tous les polygones inférieurs à la surface définie seront fusionnés au polygone contigu de surface suffisante. Bien qu'ils n'aient plus de représentation spatiale, les sols concernés apparaîtront dans la base sémantique sous la forme d'une Unité Typologique de Sol (UTS). La richesse de l'information pédologique sera donc en partie conservée.

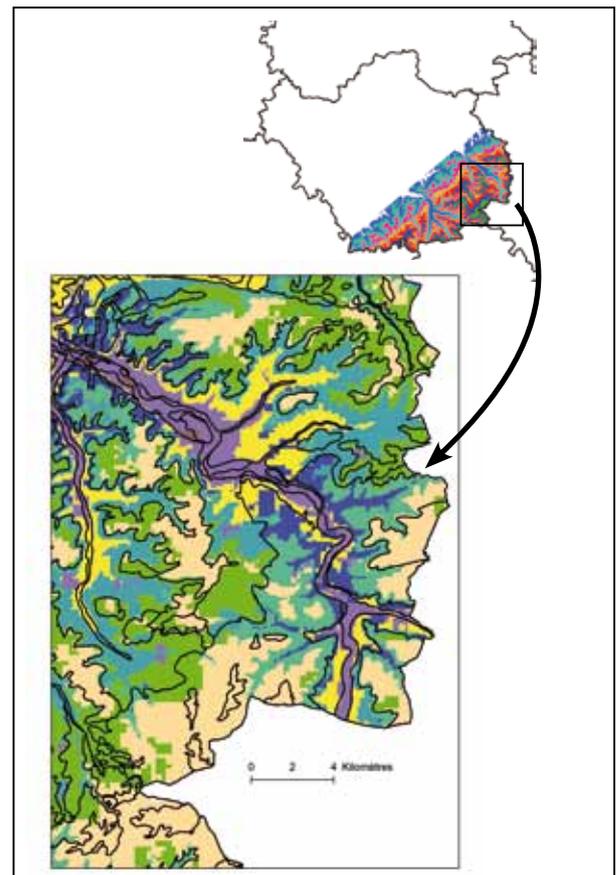
Validation à partir des connaissances existantes

Les résultats intermédiaires obtenus aux différentes étapes du tracé des pédopaysages ont été soumis à la validation par des experts locaux. Ces échanges ont permis d'affiner le découpage dans certaines zones, et de faire ressortir les particularités départementales que la classification automatique n'avait pas distinguées. Par exemple, l'unité cartographique des sols viticoles de la commune de Montgueux a été individualisée du fait de sa particularité dans le paysage, et de par sa localisation atypique et unique, ayant fait depuis longtemps l'objet d'une délimitation d'Appellation d'Origine Contrôlée.

Il est aussi indispensable de se caler sur les découpages réalisés sur les départements voisins lorsqu'ils existent, les limites des pédopaysages ne s'arrêtant pas aux limites administratives. L'homogénéité des tracés doit être continue d'un territoire à un autre puisque l'objectif du programme IGCS RRP est une cartographie harmonisée et homogène sur l'ensemble du territoire national. Sur l'Aube, les points d'ancrage étaient les deux cartes à 1/100 000 ainsi que les Référentiels Régionaux Pédologiques limitrophes. Indépendamment, ces données

Figure 5 - Segmentation en 10 classes sur la région naturelle des plateaux calcaires bourguignons par une classification non supervisée.

Figure 5 - Results of segmentation in 10 classes on a natural calcareous region.



pourraient être aussi utilisées comme zones d'apprentissage pour calibrer un modèle destiné à prédire les unités cartographiques. Cette technique est aussi utilisée en cartographie numérique sous le terme de méthode supervisée, présentée dans la deuxième partie de l'article.

- La réalisation des prédictions s'appuie donc sur une méthode descendante segmentant le territoire de manière de plus en plus fine par le découpage des entités. Le tracé final des polygones se fait par expertise en s'appuyant sur les limites fournies par la prédiction et sur l'agrégation des polygones de la carte pédologique disponible à grande échelle. A l'issue de ce travail, 99 pédopaysages ont été définis sur le département de l'Aube représentés par 1511 polygones. Un extrait est représenté en *figure 6*. où est représentée en A les polygones définis à partir de la prédiction et en B les unités cartographiques définitives.
- Une opération *a posteriori* peut être effectuée pour améliorer le rendu à 1/250 000 en lissant le contour des polygones. Cette opération permet d'éviter l'aspect dentelé des contours originels des UCS à 1/25 000. Elle peut être effectuée grâce à l'outil lisser (ESRI, 1999) qui permet d'optimiser la géométrie en fonction de l'échelle et de la résolution. Cette opération de généralisation a

pour effet de simplifier une ligne ou une limite de polygone en supprimant les petites fluctuations ou les courbures superflues de celles-ci tout en conservant sa forme globale.

Esquisse cartographique des pédopaysages en présence d'information pédologique partielle : cas de la Marne

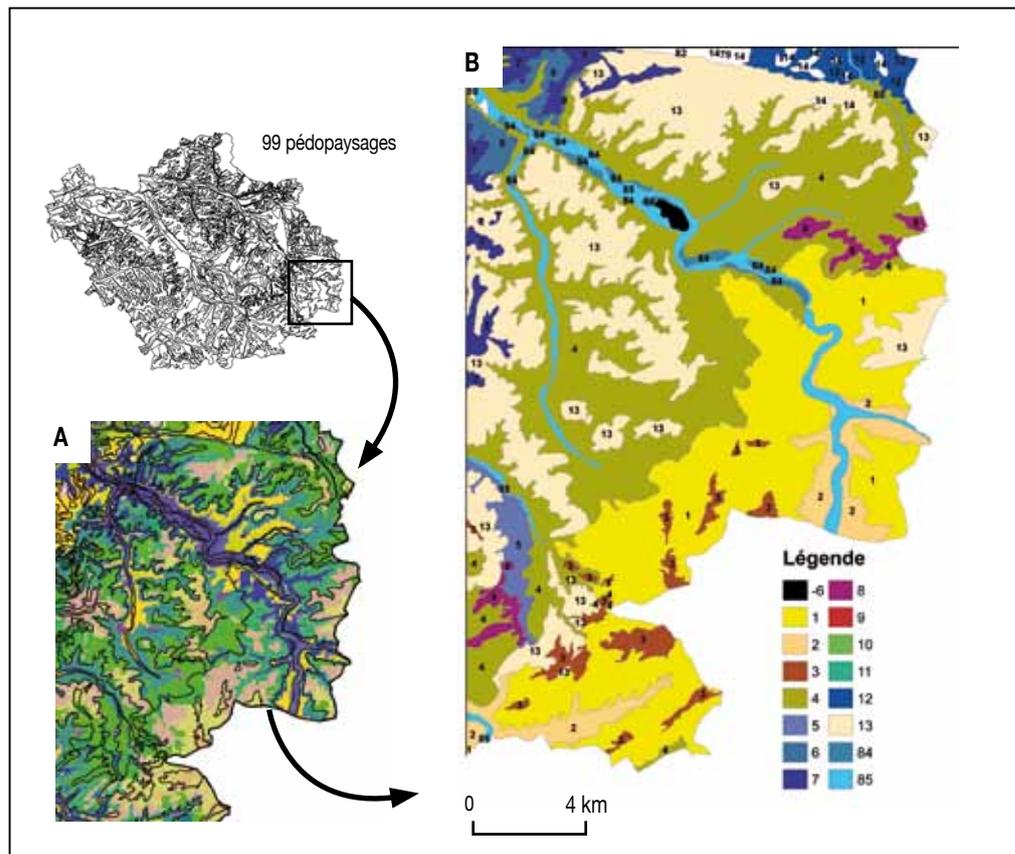
Présentation de la démarche

Les sols du département de la Marne sont partiellement cartographiés, à des échelles différentes, allant de 1/10 000 à 1/25 000. Les nombreuses études pédologiques du département (drainage, plan d'épandage...), dont l'emprise est représentée *figure 7*, représentent environ 183 000 ha. Une homogénéisation a été faite par Kossura (1999) qui a décrit les 79 types de sols répertoriés sur le département.

L'objectif est d'obtenir une cartographie des sols à 1/250 000 sur l'ensemble du territoire à partir des zones connues (*en rouge figure 7*) qui servent de secteurs d'apprentissage pour la mise au point de la méthode d'extrapolation. Nous tentons ici de prédire la répartition des sols sur les secteurs inconnus par les techniques de classification supervisée. Notre approche com-

Figure 6 - Extrait de la carte 250 000 de l'Aube. A. tracé superposé au résultat de la classification non supervisée ; B. Unités cartographiques définitives.

Figure 6 - Extract of the regional soil map.



bine une interprétation statistique basée sur les arbres de classification avec une représentation graphique sous SIG, ArcGis dans notre cas (Grinand *et al.*, 2008).

Les arbres de classification permettent la prise en compte à la fois des variables quantitatives et qualitatives. Les modèles d'arbre sont des méthodes de classification automatique par induction de règles (par exemple pente supérieure ou non à 10 %, forêt ou non...). Ces modèles sont adaptés pour décoder la structure de jeux de données complexes à partir de différentes variables explicatives et en conformité avec la valeur de la variable réponse (Friedman et Meulman, 2003). Le principe est de créer une segmentation des données par des choix binaires de manière successive afin de créer des groupes d'individus de plus en plus petits et homogènes.

Initialement développé par Breiman *et al.* (1984), le Classification And Regression Tree (CART) permet de créer un arbre selon les principes précédemment énoncés. L'analyse se fait avec CART qui est implémenté sous le logiciel R (version 2.10.0).

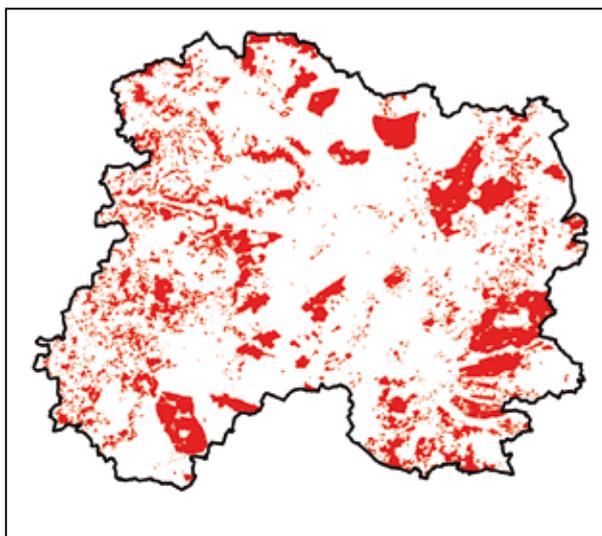
Les données d'entrée dans le modèle sont l'occupation du sol (Corine Land Cover, (UE-SOeS, 2006)), l'altitude, la pente, l'exposition, la distance hydrologique au plus proche réseau de drainage, la différence d'altitude avec le plus proche réseau de drainage, les indices de courbure, l'indice de Beven (ou Compound Topographique Index) (Beven *et al.*, 1979). Le travail de Moran et Bui (2002) a montré que l'ajout de variables lissées (par intégration des valeurs des pixels voisins par l'application d'une fenêtre mobile de convolution ayant la forme d'un cercle de 3 pixels de rayon), en plus des variables brutes dans le modèle, permet d'obtenir de meilleures classifications des sols. Cette technique permettant d'intégrer le voisinage du pixel a été appliquée.

Ces données prédictives sont introduites dans le modèle afin de le calibrer puis une extrapolation est réalisée à partir des secteurs connus sur l'ensemble du département. La méthode a été décrite par Grinand (2006). Elle est synthétisée en *figure 8*. Seules les classes de sols présentes dans les secteurs d'apprentissage peuvent être identifiées dans la mesure où « l'outil ne reconnaît que ce qu'il connaît ». En conséquence, il est important d'exploiter de nombreux secteurs afin d'avoir une bonne exhaustivité des types de sols rencontrés sur le département. La Marne répondait à cette exigence par la multitude de zones cartographiées, réparties de manière relativement homogène sur l'ensemble du département.

En sortie du modèle, une classe de sol et la probabilité d'appartenir réellement à cette classe sont affectées à chaque pixel. La classe affectée au pixel correspond à celle présentant la plus forte probabilité. Les ensembles homogènes de pixels ou leurs mosaïques organisées peuvent être regroupés dans une même unité cartographique. Le pédologue-cartographe pourra alors compléter cette esquisse en validant ou en modifiant les tracés en fonction de la réalité de la répartition des sols observée *in situ*.

Figure 7 - Portions du territoire disposant d'une information pédologique (en rouge).

Figure 7 - Existing soil maps on the department.



Résultats

Le résultat obtenu constitue une esquisse de la répartition des sols, mais les « règles » définissant l'organisation des sols établies par le modèle restent à vérifier sur le terrain. En *figure 9*, nous présentons un exemple de prédiction sur l'ensemble de la Marne en 79 classes.

La carte de probabilité donne une première information sur la validité des prédictions : évidemment, plus la probabilité présente une valeur élevée, plus la classe attribuée a de chance d'être la bonne. Cette carte permet aussi d'orienter la phase d'échantillonnage, en accentuant l'effort de terrain sur les zones à faibles probabilités ou de créer de nouveaux secteurs d'apprentissage dans ces mêmes zones.

Pour améliorer les prédictions, nous avons choisi de découper le département en 3 grands secteurs, l'ouest, la partie centrale sur craie et l'est. La prédiction ainsi obtenue permet de tracer une esquisse des pédopaysages en regroupant au sein d'un même polygone les pixels contigus d'une même classe. Une analyse plus fine permet de regrouper localement au sein d'un même polygone non plus une seule classe mais une mosaïque organisée de classes qui constitueront des unités cartographiques de sol complexes. Nous présentons un seul de ces secteurs en *figure 10 B* (la partie ouest du département : régions naturelles de la Brie Champenoise et du Tardenois). La Brie Champenoise est caractérisée par un gradient de types de sol allant du BRUNISOL, BRUNISOL rédoxique jusqu'au LUVISOL DEGRADE sur des matériaux sableux à argileux.

Le vignoble champenois est périphérique dans sa partie sud à la région du Tardenois (partie nord de notre zone). Il est caractérisé par un système de plateaux-versants, avec pour les

Figure 8 - Démarche simplifiée de la classification supervisée (d'après Grinand, 2006).

Figure 8 - Processing flowchart to supervised model.

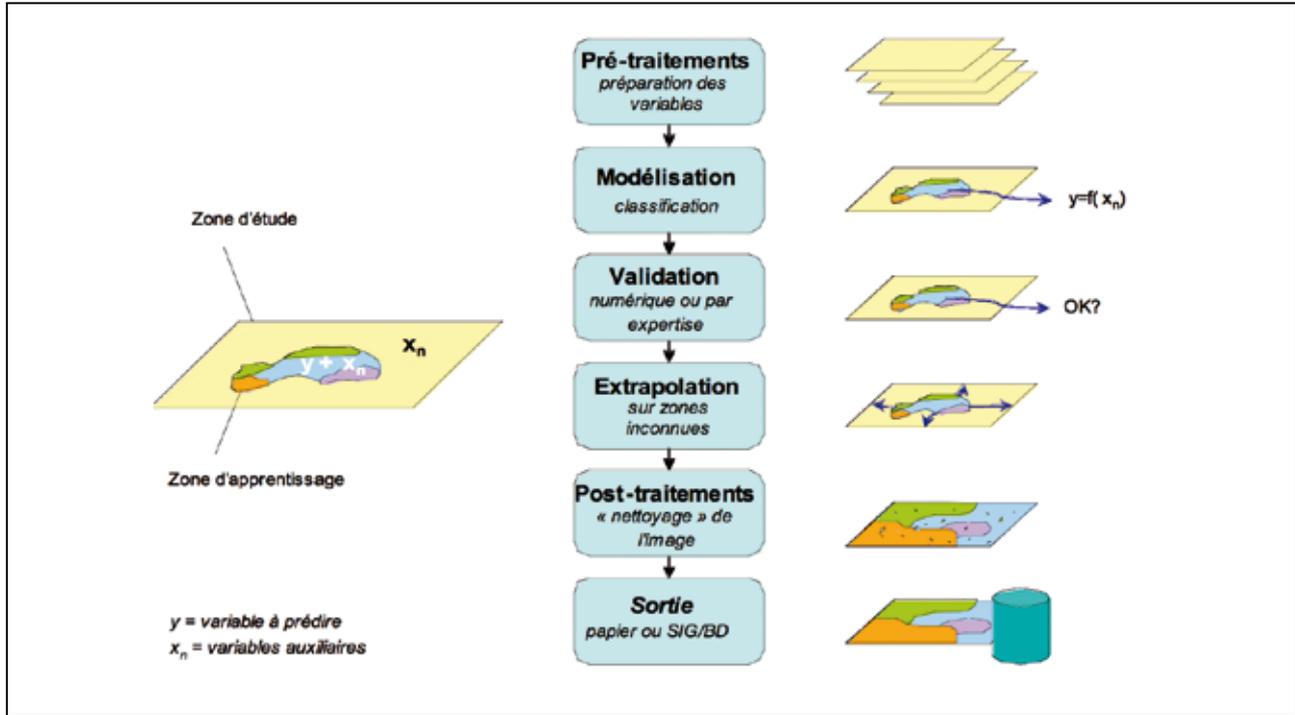
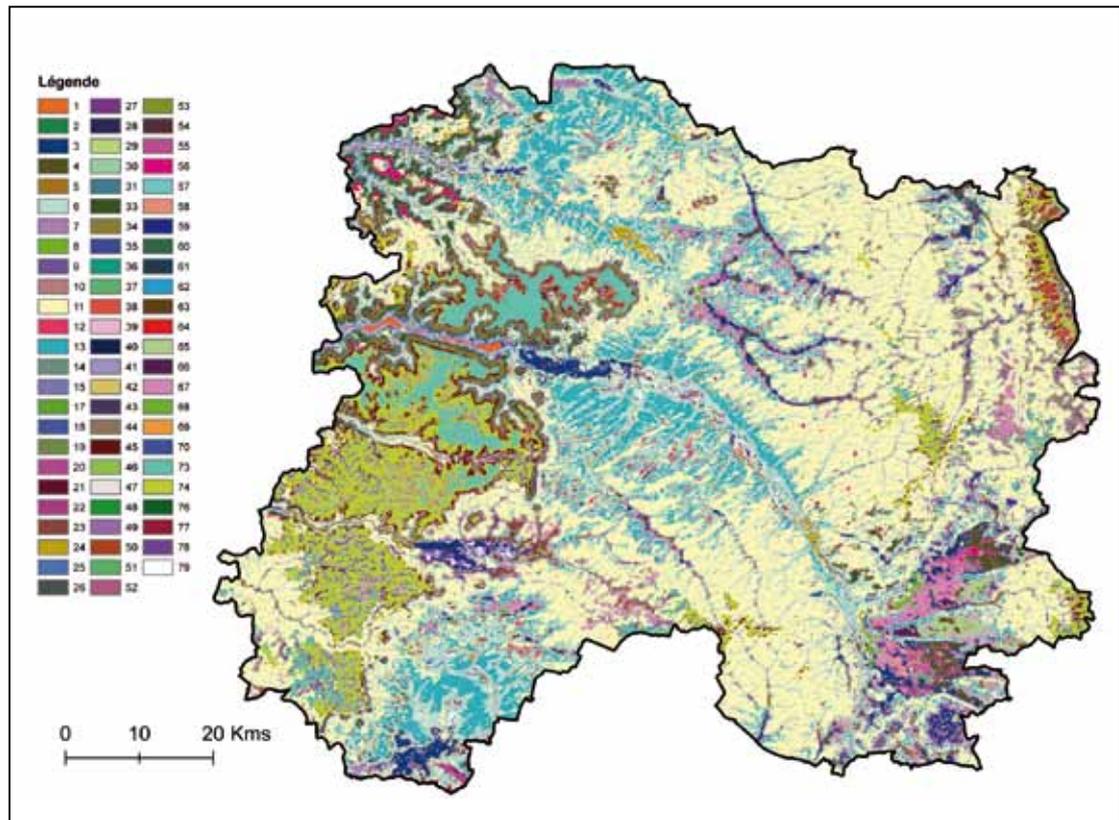


Figure 9 - Prédiction de la répartition des sols sur le département de la Marne par une classification supervisée basée sur des arbres de classification.

Figure 9 - Extrapolating regional soil distribution using supervised classification.



replats sommitaux des LUVISOLS-REDOXISOLS. Sur les versants, ce sont essentiellement des sols calcaires sur des alternances de marnes et de calcaires, qui dans la partie est sont remplacés par des sols typiques sur craie.

Validation de la méthode supervisée

Ces prédictions restent des hypothèses de répartition des sols basées sur des lois statistiques. Quelle crédibilité peut-on leur accorder? Différents outils sont à notre disposition permettant de donner une première estimation de la qualité du résultat. Les premières sources de validation sont les indices statistiques fournis par le logiciel:

- Un indice global (rapport des pixels bien classés sur l'ensemble des pixels);
- L'indice Kappa variant de 0 à 1, il exprime la réduction proportionnelle de l'erreur obtenue par une classification par rapport à celle obtenue par une procédure travaillant au hasard. Un kappa de 0,40 signifie que la classification évite 40 % des erreurs par rapport à une procédure travaillant au hasard (Girard et Girard, 1999).

L'objectif sera donc de maximiser ces deux indices.

Deux stratégies de séparation des données d'ajustement et de validation ont été établies par Grinand *et al.* (2008) afin d'évaluer la qualité de l'information produite, en jouant sur la répartition des secteurs d'apprentissage (*figure 11*).

La stratégie 1 de la *figure 11* consiste à utiliser une partie des données, choisie de manière aléatoire sur l'ensemble de la zone pour ajuster le modèle, le reste étant utilisé pour la validation.

La stratégie 2 implique que les données d'ajustement et les données de validation sont issues de deux zones différentes. La méthode est appliquée sur la zone d'apprentissage, le reste de la zone d'étude est traitée comme si elle était non couverte au plan pédologique. Elle va nous permettre de valider l'extrapolation par comparaison de la classe prédite et de la classe réelle. Pour ce deuxième cas, le choix aléatoire d'un polygone sur deux a été fait de manière à obtenir deux populations similaires mais indépendantes, avec un nombre suffisant d'individus pour que le modèle fonctionne.

Sur l'ensemble du département, la première stratégie (1) donne une prédiction avec un indice global de 0.52 et un Kappa de 0,33. La deuxième (2) donne un indice global de 0.49 et un kappa de 0.37, avec les mêmes caractéristiques d'entrée dans le modèle (échantillonnage aléatoire stratifié, 40 % des pixels sont utilisés pour la modélisation (*figure 6*)) et avec un nombre d'itérations identique (50). Une comparaison entre les prédictions 1 et 2 montre que 54 % des pixels sont affectés à la même classe lors des 2 prédictions.

En restreignant aux zones connues au niveau pédologique, il est possible de comparer la classe prédite, obtenue par le modèle, avec la classe réelle. Nous obtenons avec la stratégie 1, 50 % de pixels bien classés et 44 % avec la stratégie 2.

Spécifiquement pour la deuxième stratégie, nous pouvons affiner notre analyse, en comparant nos deux populations. Sur la zone d'ajustement, 6638 pixels sur 11080 sont affectés à la même classe que la couche de référence soit 59 %. Par contre, sur la zone d'extrapolation, seulement 2760 pixels sont affectés à la même classe que la couche de référence sur 10241 soit 27 %. C'est à dire que dans une zone non couverte par de l'information pédologique, notre modèle n'est capable de bien classer qu'environ 1 sol sur 3.

Une autre forme de validation du modèle est de s'appuyer sur des références que sont les données ponctuelles. La Chambre d'Agriculture de l'Aube dispose sur le département d'environ 590 profils. Pour notre test, nous nous sommes restreints sur la partie nord-ouest du département avec 73 profils disponibles pour la validation.

La comparaison des prédictions à ces profils donne:

La classe de référence correspond à la classe prédite dans 29 profils sur 73 (soit 40 %) avec la stratégie d'échantillonnage 1, et dans 28 cas sur 73 (37 %) avec la stratégie 2.

En revanche, si on intègre une incertitude de 100 m en prenant l'environnement du pixel en compte (la classe de référence prédite est identifiée sur l'un des pixels contigus), la qualité du classement des pixels évolue favorablement, avec pour la stratégie 1, 64 % (47 sur 73) et pour la stratégie 2, 55 % (40 sur 73) de pixels correctement prédits.

Les données ponctuelles ont un réel intérêt dans la validation des prédictions. Ce sont des données issues d'une observation directe sur le terrain, avec la seule restriction qu'elles n'intègrent pas leur environnement spatial.

Le *tableau 1* résume ces résultats en fonction des deux stratégies de séparation des données pour l'ajustement et la validation

DISCUSSION - CONCLUSION

Cet article visait à évaluer l'apport de la cartographie numérique pour accélérer la réalisation des Référentiels Régionaux Pédologiques et améliorer leur qualité dans le cadre du programme Inventaire, Gestion et Conservation des Sols. L'objectif principal est d'appréhender la répartition des sols sur un territoire donné par des méthodes de cartographie numérique, en fonction de données prédictives et de la disponibilité locale de l'information pédologique. Ces méthodes visent aussi à formaliser la réflexion que peut avoir un pédologue-cartographe lorsqu'il commence une synthèse départementale. Son approche progressive du milieu naturel versus nos données prédictives, l'intégration de l'information pédologique versus nos secteurs d'apprentissage, lui permettent d'établir des premières lois de répartition des sols. Ainsi, produit-il les premiers tracés sur la carte minute, puis il définit son plan d'échantillonnage et ainsi peut valider ses hy-

Figure 10 - Brie Champenoise et Tardenois

A : surface couverte par de l'information pédologique, B : prédiction des sols et tracé des UCS, C : carte de probabilité.

Figure 10 - Results of natural regions : Brie Champenoise and Tardenois.

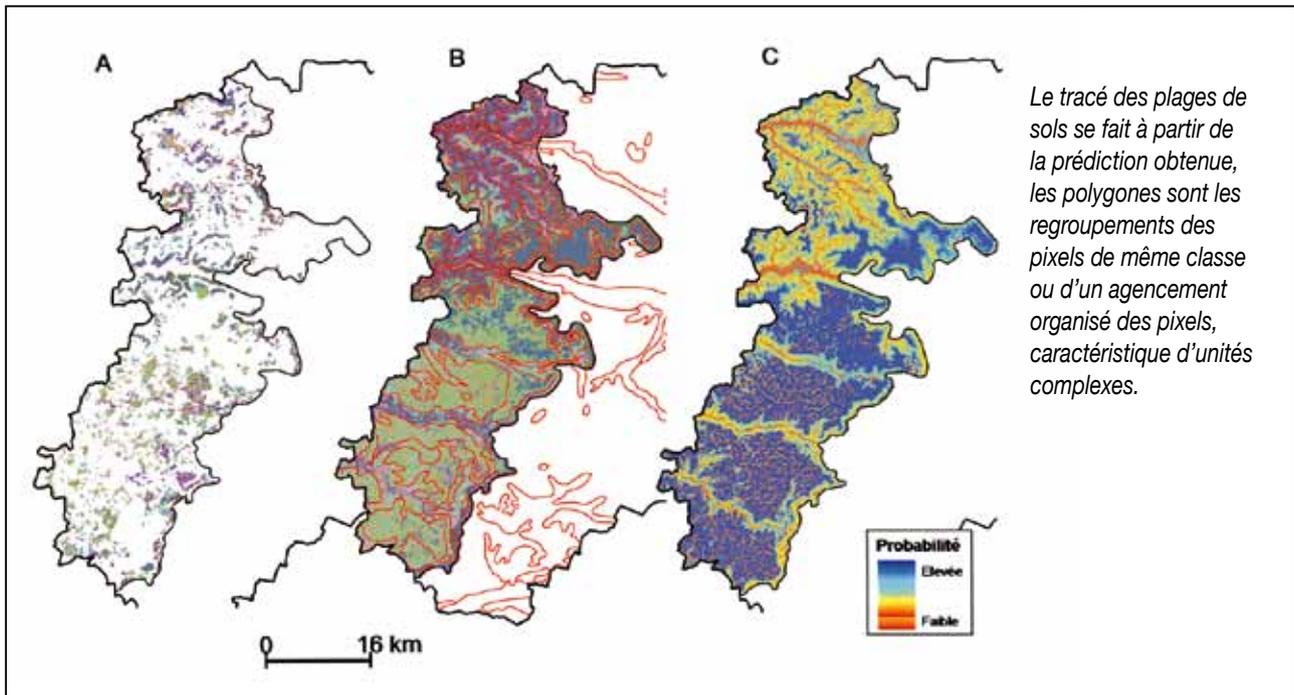
**Figure 11** - Procédures de validation (Grinand *et al.*, 2008).

Figure 11 - Validation procedures.

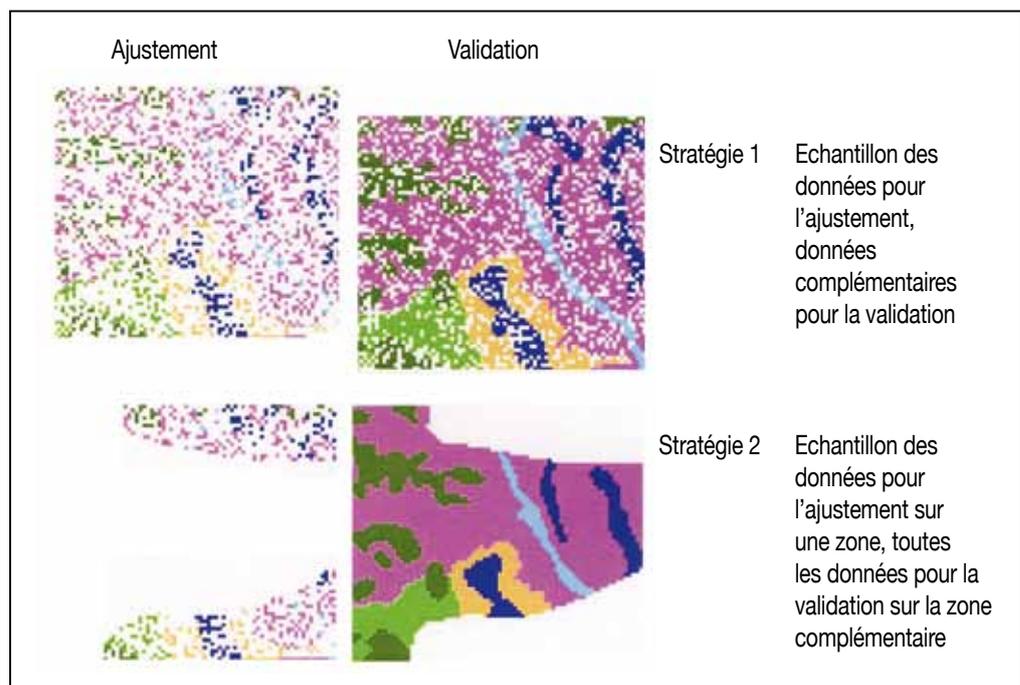


Figure 1 - Principaux résultats.**Table 1** - Mains results.

	Stratégie 1	Stratégie 2
Précision globale	0,52	0,49
Indice Kappa	0,33	0,37
Profils (pixels à pixels)	29/73 soit 40 %	28/73 soit 37 %
Profils avec intégration du voisinage	47/73 soit 64,4 %	40/73 soit 55 %

pothèses sur le terrain. Nous avons essayé de formaliser cette première phase du travail de cartographie en fournissant un découpage automatique du territoire sans *a priori*, de manière reproductible et objective. Les résultats de ces travaux restent toutefois dépendants d'une approche terrain, qu'elle soit mise en œuvre en amont (Aube), ou en aval (Marne).

Les méthodes de cartographie numérique sont caractérisées par leur objectivité et leur reproductibilité. Elles permettent de définir et de représenter graphiquement des hypothèses de répartition des sols à partir des données environnementales et pédologiques. Elles s'appuient sur des informations relativement accessibles et manipulables, ce qui permet de généraliser ces approches. Bien sûr, plus les données d'entrée seront précises et les données pédologiques importantes, et plus les résultats seront pertinents. Malgré leurs potentiels, elles restent des outils d'aide à la cartographie dont la validité des résultats reste à discuter dans tous les cas.

Ce travail montre la difficulté de traiter un département dans son ensemble, tant la diversité du milieu naturel y est importante. La quantité importante de données à manipuler peut s'avérer aussi un frein technique important. Segmenter le travail par régions naturelles semble une approche appropriée car elle est conforme à la philosophie des R.R.P. et le traitement se fait sur des ensembles moins diversifiés. Les possibilités de confusions entre des classes sont alors réduites. Dans le cas où l'information sur les régions naturelles n'existe pas au préalable, elle peut être obtenue par un premier découpage en méthode non supervisée telle qu'elle a été présentée sur le département de l'Aube.

Dans le cas où l'ensemble du territoire est cartographié à grande échelle, une méthode proposée est une classification non supervisée basée sur deux commandes disponibles dans ArcGIS et Arc/Info: ISOCUSTER et MLCLASSIFY. Les avantages de cette méthode sont la rapidité et la production d'information facilement manipulable. Par contre, elle ne fournit pas de contenu pédologique aux enveloppes ainsi définies. Elle nécessite de les croiser avec la carte des sols préexistante pour définir les types présents dans ces polygones. Elle reste soumise

à une part d'expertise pour valider et corriger les limites. Sans *a priori*, il est aussi difficile de déterminer le nombre de classes à construire pour optimiser le découpage et rester cohérent en matière de représentation. La seule possibilité est de fonctionner par tâtonnement, en essayant 5, 10, 15, 25... classes. L'ensemble de ces segmentations vont permettre d'aider au tracé des pédopaysages.

Contrairement à la méthode non supervisée, la méthode supervisée, telle qu'elle a été appliquée sur la Marne, attribue à chaque pixel une classe de sol et une probabilité d'appartenance à cette classe. Ces prédictions restent sensibles à la qualité et à la résolution des données prédictives. Elles sont aussi tributaires des secteurs d'apprentissage qui doivent être nombreux et présenter une répartition relativement homogène sur le territoire, ainsi qu'une représentativité exhaustive des types de sols du secteur.

Bien que le modèle retrouve environ 1 type de sol sur 2, les essais de validations ont montré des résultats contrastés. Le modèle fonctionne bien sur des classes de sols bien identifiées par leur morphologie, la confusion augmente sur des classes de sols proches où seul le type de sol est un élément discriminant.

Plusieurs causes peuvent être identifiées: la qualité des données prédictives, et l'absence de données géologiques sur le département de la Marne. L'intégration d'une information sur le matériau parental serait une source d'amélioration importante pour la qualité des résultats obtenus.

Cette absence est préjudiciable pour des classes de sols dont la différenciation est uniquement due au matériau: sols développés sur marne, marne crayeuse, craie marneuse, marne argileuse, calcaire dur, marne blanche dans toute la partie centrale et nord du département de la Marne. Cela induit énormément de confusion entre les différentes classes.

La cause principale des imprécisions reste certainement la complexité de l'objet recherché: le pédopaysage. Il est par définition intégrateur des éléments du milieu naturel dont le sol est un des composants. Surtout, il paraît difficile de prédire des pédopaysages à partir d'une information limitée au type de sol. Donc la prédiction telle qu'elle a été obtenue dans ce travail apparaît seulement comme une première étape. Un traitement complémentaire doit être entrepris dans le but de définir des associations ou des combinaisons de classes cohérentes afin d'aboutir à une représentation finale en pédopaysages. Cependant, la validation par les profils reste toutefois encourageante puisque dans ce cas, la comparaison se fait sur des objets similaires: type de sol à type de sol.

Les méthodes de classification automatiques basée sur la formalisation de relations sol-paysage ne permettent pas de dissocier des classes dont le facteur de différenciation dominant est de nature pédologique. Il est par exemple impossible de dissocier des sols peu épais des sols épais sur craie à partir

des variables environnementales, même en tenant compte de l'intensité de la pente. La variabilité spatiale de ces 2 types de sols n'est observable qu'à la tarière. En d'autres termes, à partir de nos données prédictives, aucune loi statistique ne peut permettre de décoder leur organisation. Une des solutions serait d'intégrer les données de sondages à notre analyse en utilisant des méthodes de type géostatistique tout en sachant que la densité minimum des observations à utiliser est fonction de l'échelle de restitution.

Enfin, très probablement, les types de sols n'étaient pas représentés exhaustivement dans les secteurs échantillonnés malgré la dispersion de la répartition des plages cartographiées sur le territoire. L'intégration de nouveaux secteurs d'apprentissages pourrait ainsi améliorer les résultats.

La carte de probabilité, associée à la carte de prédiction est un outil important à plusieurs titres : elle donne une information sur la probabilité du pixel d'appartenir réellement à la classe qui lui a été attribuée, et elle permet d'orienter le plan d'échantillonnage dans les zones où les probabilités sont les plus faibles ou les plus hétérogènes.

L'ensemble de ces travaux a été effectué sur des formations sédimentaires, des tests en Limousin ont été faits par Grinand (2006) sur les matériaux granitiques et métamorphiques. Il a montré que l'indice global de qualité était de 10 % inférieur aux valeurs présentées dans cet article. En effet, la morphologie ou la géologie ne sont pas toujours des critères discriminants de la répartition de sols sur ce type de formation et d'autres phénomènes interviennent (tectoniques...). L'interprétation des résultats de ces méthodes de classification supervisée et non supervisée ne peut donc pas se faire indépendamment de solides connaissances sur les sols de la région étudiée et de leurs facteurs de différenciation.

Une perspective intéressante pourrait être de coupler ces différentes techniques en intégrant dans la méthode supervisée, les résultats des méthodes non supervisées (première segmentation en grands ensembles homogènes). D'autres outils, intégrant des analyses de « motifs » ou de voisinage (par exemple CLAPAS (Robbez-Masson, 1994)), pourraient être combinés aux approches que nous avons présentées.

REMERCIEMENTS

Nous tenons à remercier les relecteurs, Blandine Lemerrier et Jean-Paul Legros, pour leurs relectures et leurs compléments pertinents pour la révision de cet article.

BIBLIOGRAPHIE

- Arrouays D., Daroussin J., Kicin J. C., Hassika P., 1998 - Improving topsoil carbon storage prediction using a digital elevation model in temperate forest soils of France - *Soil Science*, 163, pp. 103 - 108
- Arrouays D., Hardy R., Schnebelen N., Le Bas C., Eimberck M., Roque J., Grolleau E., Pelletier A., Doux J., Lehmann S., Saby N., King D., Jamagne M., Rat D. et P. Stengel P., 2004 - Le programme Inventaire Gestion et Conservation des Sols de France - Vol.11(3) - pp. 187-198
- Baize D., 1976 - Carte pédologique de France à 1/100 000^e, feuille de Tonnerre. INRA. 244p. + carte
- Baize D., 1994 - Référentiel régional pédologique de l'Yonne à 1/200 000^e. Régions naturelles, pédopaysages et sols. INRA 191 p. + annexes et carte
- Benoit-Janin P., 1971 - Les sols du département de l'Aube: esquisse au 1/250 000. Chaumont: SAFE, 15 p.
- Beven K.J., M.J. Kirkby, 1979 - « A physically-based, variable contributing area model of basin hydrology », *Hydrological Sciences Bulletin*, 24: 43-69.
- Bishop T.F.A & McBratney, A. B, 2001 - A comparison of prediction methods for the creation of field-extent soil property maps - *Geoderma*, 103, pp. 149-160
- Boulaine, J., 1980 - Pédologie appliquée, Coll. Sciences Agronomiques. Masson, Paris
- Bourennanne H., King D., Couturier A., 2000 - A comparison of Kriging with external Drift and simple linear regression for predicting horizon thickness with different sampling densities - *Geoderma*, 97, pp. 255 - 271
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984 - Classification and Regression trees - Wadsworth & Brooks Wadsworth statistics/probability series.
- Carré, F., Reuter, H.I., Daroussin, J., Scheurer, O., 2008 - From a Large to a Small Scale Soil Map: Top-Down Against Bottom-Up Approaches. In : *Digital Soil Mapping with Limited Data*, edited by Harterink, A. McBratney, A., Mendonca-Santos, M., Publisher: Springer Science pp. 203-212.
- CCE, 2006 - Proposition de directive européenne et du conseil définissant un cadre pour la protection des sols et modifiant la directive 2004/35/CE. COM, 232 final, 31 p.
- CGIAR-CSI, 2006 - Void-filled seamless SRTM data V1 - 2004, International Centre for Tropical Agriculture (CIAT). Available at the CGIAR-CSI 90m Database : <http://srtm.csi.cgiar.org>
- Chrétien J., 2000 - Référentiel régional pédologique de Côte d'Or à 1/250 000^e. Régions naturelles, pédopaysages et sols. INRA. 194 p. + annexes et carte
- Cole N.J., Boettinger J.L., 2006 - Pedogenic Understanding Raster-Based Classification Methodology for Mapping Soils. In : *Digital Soil Mapping*, Lagacherie, P., McBratney A.B. and Voltz, M. (ed.) Elsevier Science Publication.
- Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., Helt, T., 2000 - Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma* 97, pp. 367-391.
- Dutil P., 1992 - Carte pédologique de France à 1/100 000^e, feuille de St Dizier. INRA. 220 p. + carte
- ESRI, 1999 - (Environmental Systems Research Institute), Arc/Info: User's Guide (Redlands: ESRI).
- Friedman J.F. & Meulman J.J., 2003 - Prediction with multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22, pp. 1365-1381

- Gaddas F., 2001 - Proposition d'une méthode de cartographie des pédopaysages, application à la « moyenne vallée du Rhône » - Mémoire de Thèse, INAPG, 218 p
- Girard M. C et Girard C.M., 1999 - Traitement des données de télédétection. Dunod, 527 p.
- Grinand, C., 2006 - Applications des techniques de cartographie numérique à la délimitation de pédopaysages en Limousin. Mémoire. Orléans (FRA): Université d'Orléans. 50 p.
- Grinand C., Arrouays D., Laroche B., Martin, M.P., 2008 - Extrapolating regional landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143:180-190.
- Horemans P., 1984 - Carte pédologique de France à 1/250 000, feuille de Paris. INRA, 202 p. + carte
- Infosol, 2005 - Référentiel Régional Pédologique: Cahier des Clauses Techniques Général, 21 p.
- Irvin B.J., Ventura S.J., Slater K.B., 1997 - Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin - *Geoderma*, 77, pp. 137 - 154 p.
- Kossura C., 1999 - Synthèse des sols de l'Aube à 1/25 000^e, non publiée.
- Lagacherie P., 1992 - Formalisation des lois de distribution des sols pour automatiser la cartographie pédologique à partir d'un secteur pris comme référence - Mémoire de Thèse, Université de Montpellier, Institut National de la Recherche Agronomique, 175 p.
- Lagacherie P., 2004 - Cartographie numérique des sols: état de l'art, données et outils. Actes du colloque Digital Soil Mapping Montpellier.
- Laroche B., Doux J., 2007 - Vérifications de la qualité des Référentiels Régionaux Pédologiques en France, *Etude et Gestion des Sols*, 14(3), pp. 219-236.
- Legros JP., 1996 - Cartographie des sols: de l'analyse spatiale à la gestion des territoires - Presses polytechniques et universitaires romandes- 321 p.
- Lehmann S., 2005 - Essai de cartogénèse numérique, emploi de l'outil Clapas pour l'élaboration de la carte des sols du Gers (secteur de Mirande) - mémoire de master professionnel, Geo2Env, INRA, 48 p.
- McBratney A.B., Mendonça Santos M.L., Minasny B., 2003 - On digital Soil mapping. *Geoderma*, 117: pp. 3-52.
- Moran J.M. & Bui E.N. , 2002 - Spatial data mining for enhanced soil map modelling - *Int. J. Geographical Information Science*, 16, n°6, pp. 533-549.
- Ouisse M., Aubepart A., Kossura C., 2006 - Constitution du système géographique des sols de l'Aube.
- Robbez-Masson J.M., 1994 - Reconnaissance et délimitation de motifs d'organisation spatiale. Application à la cartographie des pédopaysages. Th. Doct. ENSA. M, 161 p.
- Roque J., 2003 - Référentiel régional pédologique de l'Île de France à 1/250 000^e. Régions naturelles, pédopaysages et sols. INRA. 224 p. + carte
- UE-SOeS, 2006 - CORINE Land Cover
- Williams C., Hargrove W., Liebman M. et James D., 2007 - Agro-ecoregionalization of Iowa using multivariate geographical clustering, *Agriculture, Ecosystems & environnement*, Volume 123, Issues 1-3, pp. 161-174.

