

Utilisation du logiciel CLAPAS pour l'aide à la délimitation de pédopaysages

Un test sur la carte des sols de Mirande (Gers, France)

S. Lehmann⁽¹⁾, J.C. Bégon⁽²⁾, M. Eimberck⁽¹⁾, J. Daroussin⁽²⁾, R. Wynns⁽³⁾ et D. Arrouays⁽¹⁾

(1) INRA Orléans, Unité InfoSol, BP 20619, F-45166 Olivet cedex 06

(2) INRA Orléans, Unité de Science du Sol, BP 20619, F-45166 Olivet cedex 06

(3) BRGM, BP 6009, F-45060 ORLEANS Cedex 2

RÉSUMÉ

L'objectif de ce travail est de tester l'apport du logiciel Clapas (Robbez-Masson, 1994) pour l'aide à la cartographie des sols dans le cadre du programme IGCS. Notre secteur d'étude couvre une surface d'environ 50 000 ha localisée dans la région des coteaux de Gascogne (Gers) (*figure 1*). Nous rappelons les principes de Clapas (*figure 2*). A partir de deux secteurs d'apprentissage (*figure 8*), nous définissons les combinaisons de données numériques (issues de la carte géologique numérisée (*figure 3*) et des indices dérivés d'un MNT au pas de 50 m) les plus pertinentes pour discriminer des Unités Cartographiques (*figures 5 et 6*). Nous procédons d'abord à une validation interne de la méthode Clapas au sein des secteurs d'apprentissage. Nous analysons la qualité des résultats à l'aide de cartes de qualité des classements (*figure 10*), et d'une matrice de confusion avec les unités de la carte de validation (*tableaux 3 et 4*). Nous procédons à plusieurs itérations en regroupant successivement des unités sur la base de leur confusion (*figure 4*). Lorsqu'un résultat satisfaisant est atteint (fixé à 75 % de bonne prédiction), nous généralisons la classification à l'ensemble de la zone d'étude. Nous la validons en la comparant à une carte des sols du même secteur au moyen d'une nouvelle matrice de confusion (validation externe). Nous montrons que cette méthode apporte une aide pratique à la délimitation d'unités pédopaysagères à l'échelle du 1:250 000 (*figure 12*). Nous proposons ensuite quelques voies d'améliorations possibles de la méthode.

Mots clés

CLAPAS, cartographie numérique, pédopaysage, modèle numérique de terrain, classification supervisée, matrice de confusion, segmentation d'image.

SUMMARY**USING CLAPAS TO HELP PREDELINEATION OF «SOILSCAPES» : Test on the soilmap of Mirande (Gers, France)**

The aim of this work is to determine the usefulness of Clapas software for soil mapping as part of the French soil surveying program : our study area covers about 50,000 ha in French Gascony (Gers department) (figure 1).

First, we recall the principles of Clapas (figure 2). Then we use two reference areas (figure 8) to define the best combinations of digital data (from the geological map (figure 3) and 50 m DEM) to discriminate mapping units (figures 5 and 6). We first operate an internal validation by applying the Clapas methodology to interpolate the classification process into the reference area. The quality of results is tested by a classification accuracy map (figure 10) and by a confusion matrix with existing maps (tables 3 and 4). Several runs are processed, by progressively grouping units which have been confounded until the result of 75 % for the classification index (figure 4).

Then we operate an external validation by generalizing the classification to the whole study area and we compare the result of the classification with a reference map using a new confusion matrix.

We show that this method is a useful tool for pre-delineation of "soilscape" units on the scale of 1:250,000 (figure 12). We finally suggest some possible ways of improving the method.

Key-words

CLAPAS, digital soil mapping, soilscape, digital elevation model, supervised classification, confusing matrix, segmentation.

RESUMEN**USO DEL PROGRAMA CLAPAS PARA LA AYUDA A LA DELIMITACIÓN DE LOS PEDOPAISAJES.****Un ensayo sobre el mapa de suelos de Mirande (Gers, Francia)**

El objetivo de este trabajo es probar el aporte del programa CLAPAS para la ayuda a la cartografía de los suelos en el cuadro del programa IGCS. El sector de estudio cubre una superficie de aproximadamente 50 000ha localizada en la región de las laderas de Gascoña (Gers) (fig 1). Se recordé los principios de CLAPAS (fig 2). A partir de dos sectores de aprendizaje (fig 8) se definieron las combinaciones de datos numéricos (que provienen del mapa geológico numerizado (fig 3) y de los índices derivados de un MNT (al paso de 50m) los mas pertinentes para discriminar unidades cartográficas (fig 5 y 6). Luego se calibro el método CLAPAS al seno de los sectores de aprendizaje. La calidad de los resultados es analizada a continuación con ayuda de mapas de calidad de clasificaciones (fig 10) y de un matriz de confusión con las unidades del mapa de validación (tabla 3 y 4). Se realizaron varias iteraciones agrupando sucesivamente unidades sobre la base de su confusión (fig 4). Cuando se logro un resultado satisfactorio (75% de buena predicción), la clasificación se generaliza al conjunto de la zona estudiada, luego validada con una comparación a un mapa de suelos del mismo sector con medio de una nueva matriz de confusión (validación externa). Se mostró que este método constituye una herramienta interesante para una predelimitación de unidades de pedopaisajes a escala del 1/250 000 (fig 12). Se proponen algunas vías de mejoramientos posibles del método.

Palabras clave

CLAPAS, cartografía numérica, pedopaisaje, modelo numérico de terreno, clasificación supervisada, matriz de confusión, segmentación de imagen

Depuis plus d'une vingtaine d'années, des travaux de recherche en matière d'aide à la cartographie automatisée des sols sont en développement constant. En France, Girard et King (Girard, 1983 ; King, 1986 ; Girard et King, 1988 ; King et Girard, 1988), figurent parmi les premiers pédologues à tester l'apport de classifications supervisées¹. Avec l'apparition des systèmes d'informations géographiques et la multiplication des sources de données externes et des outils de traitement, les recherches en cartographie numérique se sont intensifiées. Des chercheurs se sont orientés vers une classification sémantique non spatiale. C'est le cas en France de King *et al.* (1989) et Girard *et al.* (1989). D'autres ont exploré des approches fondées sur des approches géostatistiques, comme Walter (1989), Bourennane *et al.* (1996), ou des modèles linéaires de corégionalisation comme Voltz *et al.* (1997) ; Goulard et Voltz (1992). Lagacherie (1992, 2005), Lagacherie et Holmes (1997), et Lagacherie *et al.* (2001), se sont intéressés aux arbres de régression pour prédire des types de sols à partir de données environnementales. La modélisation par arbre de décision consiste en un partitionnement récursif et progressif d'un jeu de donnée en sous-classes homogènes (Breitman *et al.*, 1984). Une fois que l'arbre est construit, chaque nœud de l'arbre correspond à une règle de décision qui est utilisée pour prédire au mieux un type de sol (Scull *et al.*, 2005). Robbez-Masson (1994) est un des précurseurs des modèles de segmentation d'images intégrant la notion de voisinage spatial. Avec l'outil Clapas (Robbez-Masson, 1994), il a élaboré un modèle de classification original fondé sur la comparaison de motifs d'organisation spatiale via une distance mathématique qui rend compte de la composition spatiale d'une zone centrée sur un pixel donné. Enfin de nombreuses publications, traitent de prédictions de paramètres du sol obtenues à partir de dérivées du MNT² (Beven & Kirkby, 1979 ; Moore *et al.*, 1991, 1993 ; Cialella *et al.*, 1997 ; Chaplot *et al.*, 2000 ; Carré et Mac Bratney, 2005). Les méthodes d'arbres de régression ont été comparées avec le krigeage (Bishop et McBratney 2001, 2005). Dobos *et al.* (2000) couplent la radiométrie à haute résolution à des dérivées d'un MNT pour prédire des propriétés physiques du sol. En réalisant une étude bibliométrique des articles parus dans les 100 premiers numéros de la revue *Geoderma*, Hartemink *et al.* (2001), montrent une très forte progression des publications scientifiques dans ce domaine. McBratney *et al.* (2003) en réalisent une revue très complète d'où il ressort que d'une manière générale, la plupart des modèles prédictifs de la spatialisation des sols, intègrent, d'une manière ou d'une autre, l'équation de Jenny (1941) dans laquelle $S = f(c, o, r, p, t)^3$. Le modèle suivi par Clapas ne déroge pas à la règle.

Pour prédire des types de sols avec cet outil, il convient d'utiliser des variables qui rendent compte le mieux possible des facteurs de formation des sols.

Ces avancées scientifiques se produisent à une époque où les besoins en information sur les sols, servant à la surveillance et à la modélisation des impacts environnementaux, constituent un enjeu majeur. C'est particulièrement le cas en France, où nous ne disposons pas encore de la couverture cartographique exhaustive du territoire à des échelles plus précises que le 1/1 000 000 (Arrouays *et al.*, 2004), certaines régions pouvant même être familièrement qualifiées de « déserts pédologiques ».

Le développement des outils de cartographie numérique à des fins réellement opérationnelles sur des surfaces importantes, d'ordre régional, demeure marginal en France, à quelques exceptions notables (par exemple, dans le cadre du programme IGCS⁴, Gourmelon, 2003 ; et en Europe à des échelles plus petites, Daroussin *et al.*, sous presse). A ces échelles, de nombreux travaux ont pourtant été développés dans d'autres pays [par exemple, en Europe de l'Est (Dobos *et al.*, 2000), en Australie (Moran et Bui, 2002 ; Bui et Moran, 2003 ; Bui *et al.*, 2006), en Californie (Scull *et al.*, 2005)].

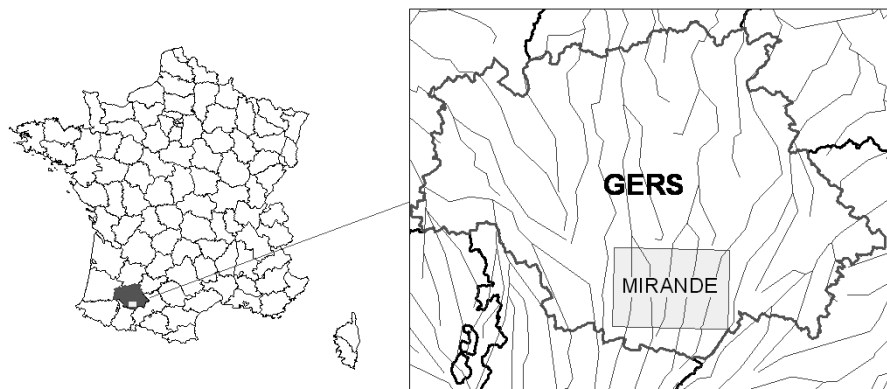
On constate donc actuellement en France un hiatus entre le développement de la recherche sur ce thème et ses applications opérationnelles. Dans le but final d'envisager le logiciel CLAPAS (Robbez-Masson, 1994) comme un outil pour l'aide à la cartographie des sols dans le cadre du programme IGCS, notre objectif est ici de tester une application précise de cet outil sur des secteurs d'apprentissages et de validation connus du pédologue. Ce logiciel a été conçu pour segmenter une image en zones homogènes au regard de sa composition texturale. Nous l'employons ici au-delà de son champ d'application comme un véritable outil de classification supervisée. Nous utilisons CLAPAS pour caractériser la composition de secteurs d'apprentissage, puis comme outil d'extrapolation que nous validons avec une carte existante afin de prédire l'apparition de certains types de sols au-delà des secteurs d'apprentissage. Ce logiciel se présente comme un logiciel libre et francophone. Il est relativement simple d'utilisation et emploie des données assez facilement accessibles. Ce point est important dans l'optique d'un transfert de technologie vers l'ensemble des partenaires du programme IGCS.

1 : Classification supervisée : Méthode déterministe destinée à obtenir des classes de pixels ou des catégories thématiques par un traitement interactif des données conditionné par la réalité de terrain (RNC, 2005).

2 : MNT : Modèle Numérique de Terrain

3 : $S = f(c, o, r, p, t)$: S représente le sol ; c, le climat ; o, les organismes vivants ; r, le relief ; p, le matériau parental ; t, le temps.

4 : IGCS : Inventaire, Gestion et Conservation des Sols. Ce programme d'inventaire multi-échelle vise à constituer des bases de données décrivant les sols et leur répartition géographique, gérées au sein d'une base à structure unique «Donesol». La principale priorité est l'achèvement et l'harmonisation de la couverture cartographique nationale à l'échelle du 1 : 250 000.

Figure 1 - Localisation de l'étude.**Figure 1** - Study area.

MATÉRIEL ET MÉTHODES

Site et sols étudiés

Notre test porte sur la carte des sols de **Mirande** (Gers), qui correspond à un extrait de l'esquisse pédologique au 1/100 000 de Auch, réalisée en 1975 par la CACG (Compagnie d'Aménagement des Coteaux de Gascogne). Les levés ont été réalisés à l'échelle du 1/20 000 par trois équipes de cartographes, et synthétisés ensuite au 1/100 000.

La coupure de Mirande recouvre la partie nord de la région de l'Astarac, sur le département du Gers. Elle est traversée du sud au nord, par un faisceau de vallées quasiment parallèles, relativement resserrées et au profil dissymétrique. Entre ces vallées, des coteaux au relief adouci sont constitués par l'empilement de couches correspondant à des cycles sédimentaires. La série débute par des lits de calcaires du burdigalien surmontés par une épaisse masse marneuse du début du miocène, suivie par de la molasse de la même époque, puis par des poudingues ou des conglomérats du quaternaire (Crouzel, 1970).

L'esquisse pédologique de Auch au 1/100 000 (Camélio, 1975) distingue quatre grandes familles de sols. Nous avons ici « traduit » dans les termes du Référentiel Pédologique (AFES, 1995), la nomenclature des types de sols détaillée dans la notice publiée en 1980.

- 1- Les FLUVIOSOLS
- 2- Les COLLUVIOSOLS
- 3- Les RENDOSOLS et les CALCOSOLS
- 4- Les CALCISOLS
- 5- Les BRUNISOLS

6- Les NÉOLUVISOLS

7- Les LUVISOLS TYPIQUES rédoxiques (appelés localement « boubènes colorées »)

8- Les LUVISOLS DÉGRADÉS (appelés aussi « boubènes blanches »)

Données utilisées

Les sources de données utilisées dans ce travail sont les suivantes :

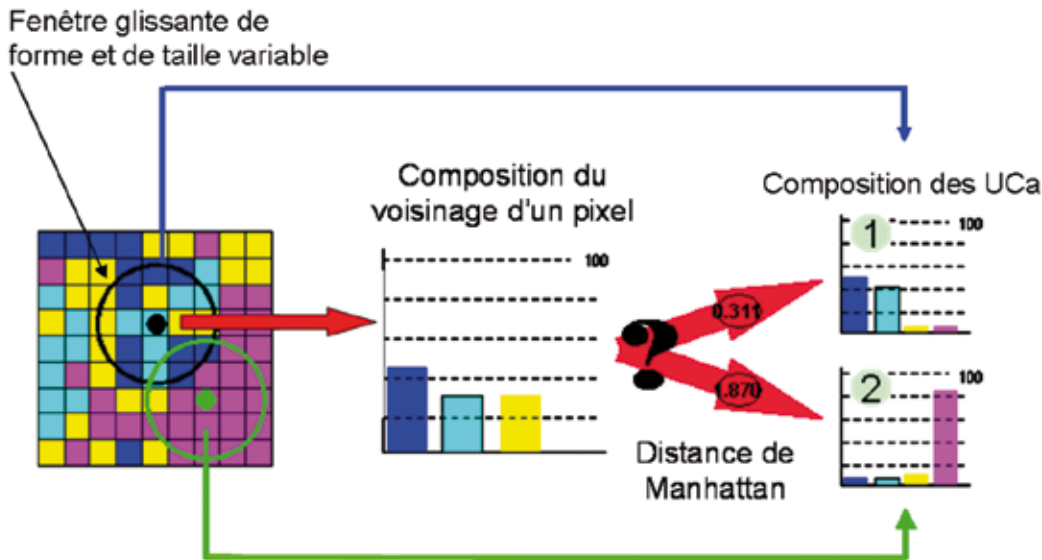
- Des cartes de sols :
 - deux minutes de terrain au 1/20 000 (numérisées) de la feuille de Mirande dont l'emprise correspondra aux unités cartographiques d'apprentissage (UCa) ;
 - la carte des sols du secteur de Mirande extraite de l'esquisse pédologique de Auch au 1/100 000, (Camélio, 1975) au format numérique, et utilisée comme carte de validation ;
- la carte géologique de la feuille de Mirande au 1/50 000 (Crouzel, 1970) au format papier et numérique ;
- Le modèle numérique d'altitude (MNA) du secteur au pas de 50 m (IGN, 2003).

Les outils de traitement de l'information

CLAPAS signifie « CLAssement de PAYSages et Segmentation ». Le logiciel CLAPAS est un outil de segmentation d'une image numérique en paysages (Robbez-Masson, 1994). La classification est fondée sur la comparaison de la composition du voisinage de chaque pixel avec celle d'unités cartographiques d'apprentissage (UCa), établies sur des secteurs de référence (Robbez-Masson, 1994). Une fenêtre glissante, de taille et de forme

Figure 2 - Principe de CLAPAS : calculer et classer des vecteurs de composition (Robbez-Masson, 1994).

Figure 2 - Principles of CLAPAS model, to calculate and classify composition vectors (Robbez-Masson, 1994).



variables (circulaire ou elliptique plus ou moins allongée) balaie l'image horizontalement et détermine la composition du voisinage du pixel situé au centre de cette fenêtre. Clapas affecte ensuite chaque pixel à une UCa selon la proximité de la composition de son voisinage avec l'UCa la plus proche au sens de la métrique de Manhattan (figure 2).

Sur la figure 2, chaque pixel exprime une combinaison de différents paramètres qui apparaît avec une couleur donnée. L'image avant classement synthétise les différentes combinaisons de classes de données d'entrée telles que des paramètres géomorphologiques, l'occupation du sol ou la lithologie. Elle est ainsi formée de combinaisons multicritères, codées sur des valeurs uniques au niveau de chaque pixel.

Après classification, l'image fournie par Clapas offre ainsi une segmentation de la zone d'étude en polygones homogènes au regard de leur composition. Elle est accompagnée d'une image des distances qui témoigne de la qualité statistique du classement.

Préparation des données de base

a) Rasterisation de la carte géologique

La carte géologique de Mirande a été simplifiée au niveau des étages géologiques et rasterisée à partir de la carte livrée par le BRGM au format vecteur d'ArcInfo® (figure 3). Dans le but de limiter le nombre des combinaisons dans l'image d'entrée de CLAPAS (255 au maximum), nous avons réduit, de quatorze à huit, les types de matériaux en regroupant, à dire d'expert, certaines

formations sur des critères lithologiques ayant une influence sur la pédogénèse.

b) Calcul des indices dérivés du MNT

Le tableau 1 recense les différents indices dérivés du MNT et calculés avec ArcGis9®.

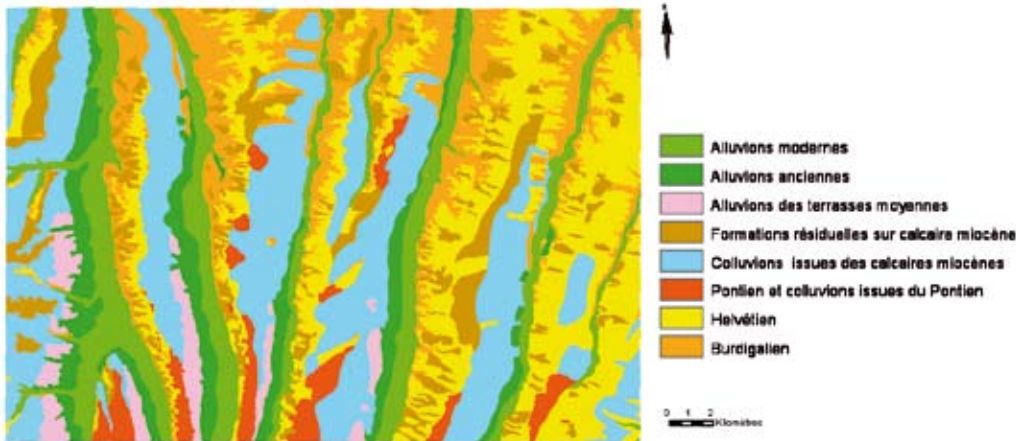
Méthode de choix des indices

Sur notre secteur d'étude, nous nommons UCa les unités cartographiques de sols qui constituent les deux secteurs d'apprentissage (Cf. Données utilisées). Pour choisir les meilleurs indices, nous nous sommes appuyés sur deux outils statistiques :

- une analyse en composantes principales des unités cartographiques d'apprentissage (UCa) en fonction des indices afin d'évaluer les corrélations entre les indices et de déterminer ceux qui expliquent le maximum de variabilité ;
- une analyse visuelle de la distribution statistique de chaque indice en fonction des UCa.

Paramètres d'entrée utilisés dans CLAPAS

Les indices retenus sont découpés en classes sur la base de leur séparabilité au vu de leur distribution statistique, puis combinés pour former une image. La valeur de chaque pixel représente une des combinaisons possibles des classes d'indices retenus. Cette image est traitée à l'aide du logiciel CLAPAS avec les paramètres décrits ci-après :

Figure 3 - Carte géologique simplifiée de Mirande (Crouzel, 1970) au format raster.**Figure 3** - Simplified geological map of Mirande (Crouzel, 1970) in raster format.**Tableau 1** - Tableau comparatif des indices dérivés du MNT.**Table 1** - Comparative table of the indexes derived from the DEM.

Indice	Nom complet	Formule	Fonction
Beven	Indice de Beven	$I_b = \text{Log} (S_{\text{surface drainée}} / \text{Tangente de la pente})$	Indice de saturation potentielle en eau des sols (Beven et Kirkby, 1979).
MRVBF	Multi Resolution Valley Bottom Flatness Index	Elle se décompose en plusieurs étapes trop complexes pour être détaillées ici. Elle utilise 2 paramètres : la pente et le "elevation percentile" (somme des pixels de faible altitude dans un rayon donné / somme des pixels de la zone concernée)	Il identifie à différentes échelles, des surfaces relativement planes et basses dans le paysage, ce qui est généralement interprété comme une cartographie des fonds des vallées (Gallant et Dowling, 2003).
MRRTF	Multi Resolution Ridge Top Flatness Index	Elle se calcule d'une façon très proche du MRVBF en modifiant le "elevation percentile" (somme des pixels de haute altitude dans un rayon donné / somme des pixels de la zone concernée)	Il identifie les sommets des interfluves et les plateaux, dominés par les processus d'érosion (Gallant et Dowling, 2003).
CTI	Compound Topographic Index	$\text{CTI} = (\text{surface drainée} + 1) * \text{pente} * \text{convexité horizontale}$	Il est généralement considéré comme l'indice d'humidité du sol.
RI	Relief Index	$\text{RI} = \text{alt min} - \text{alt max}$ (de chaque pixel dans un rayon donné)	Il constitue un indicateur de la rugosité du modelé.
Exp	Exposition	Exposition = orientation de la pente en un point donné par rapport au nord (exprimé en degrés de 0 à 360)	Il identifie les zones les plus exposées au soleil.

Taille du voisinage : 7 pixels (valeur par défaut de CLAPAS), soit, compte tenu de la résolution de notre base de données raster, 350 m.

Forme de voisinage : fenêtre de forme circulaire (forme par défaut).

Descripteur utilisé : histogramme.

Métrique statistique : Manhattan.

Ces paramètres sont conservés pour chaque itération.

Validation interne de la classification obtenue avec CLAPAS

Les UCa sont discriminées par la métrique de Manhattan (Dagnélie, 1975) (*tableau 2*). On procède ensuite à plusieurs itérations successives. Une itération consiste en la production d'une image résultat que l'on valide par l'analyse statistique de l'image des distances de classement (*figure 10*). L'image résultat est ensuite comparée à la carte des sols de validation sur le même secteur que les UCa via une matrice de confusion. Il s'agit donc d'un test

Tableau 2 - Comparaison des histogrammes des UCa par la métrique de Manhattan.**Table 2** - Comparison of the bar charts of the regions of interest by Manhattan's metric.

UCa	1	2	3	4	5	6	7	8
FLUVIOSOLS (1)		1.41	1.95	1.97	2	1.49	1.61	1.93
COLLUVIOSOLS (2)			1.33	1.12	1.44	1.81	1.61	1.44
RENDOSOLS (3)				1.34	1.63	1.82	1.95	1.75
CALCOSOLS (4)					0.85	1.59	1.83	1.23
BRUNISOLS (5)						1.65	1.69	0.94
NÉOLUVISOLS r. (6)							1.54	1.46
LUVISOLS d. prof. (7)								1.16
LUVISOLS d. lept. (8)								

Classe prédite (UCp)	Classe observée (UCa)								nbr pixels	erreur de commission	% de bonne classification
	1	2	3	4	5	6	7	8			
1	2260	710	8	0	0	999	884	33	4894	0.54	46
2	157	1454	203	353	0	236	6	0	2409	0.40	60
3	8	780	5065	2150	292	103	0	154	8532	0.41	59
4	4	128	253	669	390	38	0	207	1689	0.60	40
5	0	205	107	1122	2035	0	0	765	4234	0.52	48
6	50	41	0	83	0	135	179	33	521	0.90	10
7	62	95	0	2	0	383	2939	300	3781	0.22	78
8	0	610	111	1186	1456	417	1634	5077	10491	0.52	48
nbr pixels	2541	4003	5747	5565	4173	2311	5642	6569	36551	nbr total de pixels	
erreur de d'omission	0.11	0.64	0.12	0.88	0.51	0.94	0.48	0.23			

Tableau 3 - Matrice de confusion après la première itération (8 classes).**Table 3** - Confusion matrix after the first iteration (8 classes).

Indice global	Indice Kappa
0.49	0.42

de cohérence interne car l'apprentissage et la validation ont lieu strictement sur la même zone.

Emploi d'une matrice de confusion

Une matrice de confusion (*tableau 3*) est utilisée pour comparer le résultat de la classification générée par Clapas (Unités Cartographiques prédites), avec la carte des sols de référence du secteur de Mirande (Unités Cartographiques d'apprentissage) couvrant la même emprise (*figure 8*).

Elle se présente sous la forme d'un tableau affichant les statistiques de la précision de classification de l'image obtenue, notamment le nombre de pixels mal classés au sein des diverses classes. L'information des colonnes (*tableau 3*) correspond aux classes thématiques de la carte de référence alors que les lignes contiennent l'information thématique résultant de la classification de l'image obtenue avec Clapas. Les valeurs de la diagonale de la matrice représentent le nombre de pixels correctement classés (*tableau 3*). La somme des cellules de la diagonale de la matrice représente ainsi le nombre total de pixels correctement classés. La proportion du nombre total de pixels correctement classés par rapport au nombre total de pixels de la matrice donne l'indice global de la classification (**G**) exprimé en pourcentage (RNC, 2005).

Le total des valeurs dans les cellules non diagonales de n'importe quelle ligne représente le nombre de pixels qui ont été incorrectement attribués à des classes autres que celle représentée par la ligne. Cette valeur s'appelle « **erreur de commission** », elle peut être calculée pour chaque classe en divisant ce total par la somme des totaux des rangées. De façon similaire, le total des valeurs des cellules non diagonales de n'importe quelle colonne représente le nombre de pixels qui ont été incorrectement inclus dans la classe représentée par la colonne. Cette valeur s'appelle « **erreur d'omission** » (RNC, 2005).

Le **coefficient de Kappa** est un estimateur de la précision. Il varie de 0 à 1 et exprime la réduction proportionnelle de l'erreur obtenue par une classification donnée comparée à celle obtenue par une procédure d'affectation aléatoire. Un indice de Kappa de 0.40 indique que la classification est bonne à 40 % lorsque l'on a déduit du score obtenu, les résultats favorables dont on peut estimer qu'ils sont liés au hasard (Girard et Girard, 1999).

Sur la base de cette matrice, nous fusionnons les UCa les plus proches lors de l'itération suivante. On s'arrête à la n^{ième} itération (*figure 4*), lorsque le résultat obtenu est le meilleur compromis entre l'indice global de classification et la résolution cartographique souhaitée ; elle doit être au moins égale à une résolution compatible avec une échelle de représentation au 1/250 000.

Figure 4 - Schéma des itérations successives dans CLAPAS.

Figure 4 - Schematic diagram of iterations with CLAPAS.

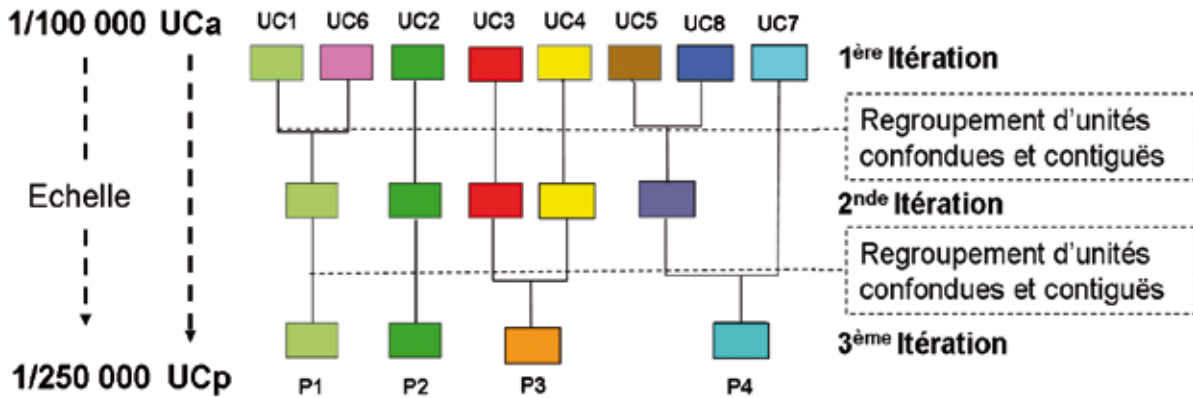


Figure 5 - Analyse en composantes principales des indices par UCa (axes F1 et F2) sur le secteur d'apprentissage.

Figure 5 - Principal Component Analysis of the main indices by soil mapping unit (axis 1 and 2) on the region of interest.

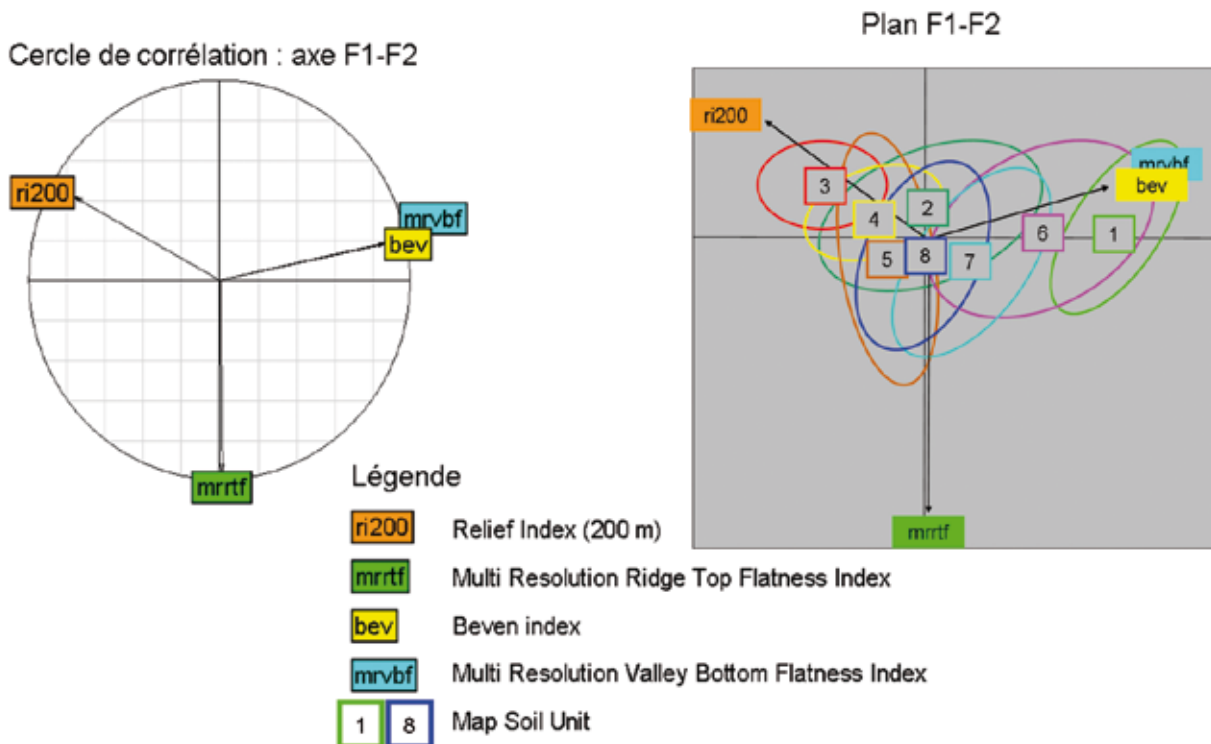


Figure 6 - Diagramme en boîte à moustaches de l'indice de Beven (a) et du relief (b) par unité cartographique de sol.

Figure 6 - Boxplots of Beven Index (a) and Relief Index (b) by soil mapping units.

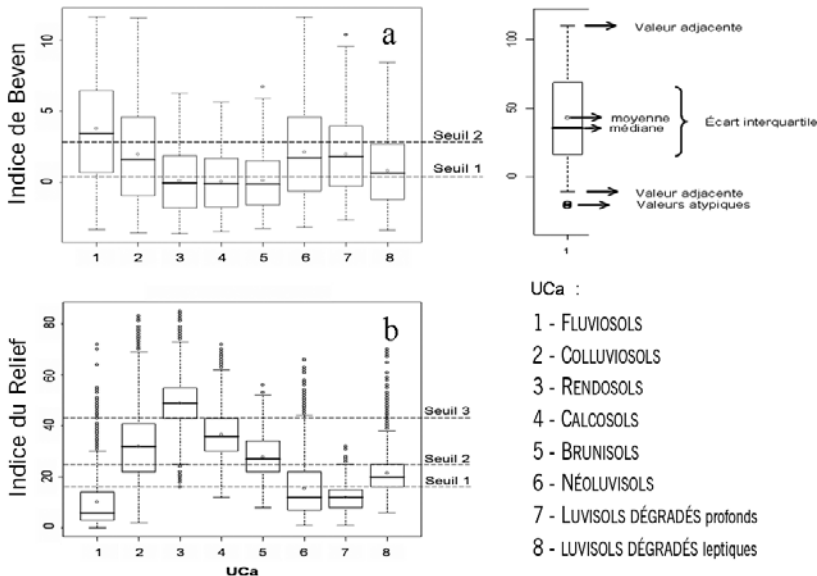
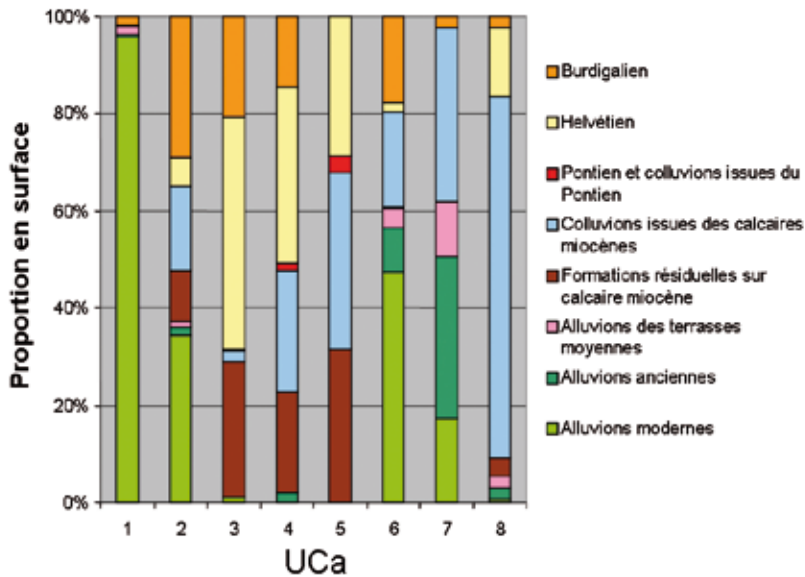


Figure 7 - Répartition des étages géologiques au sein des UCa indépendamment des autres indices.

Figure 7 - Distribution of the geological tier per soil mapping unit independently of the other indices.



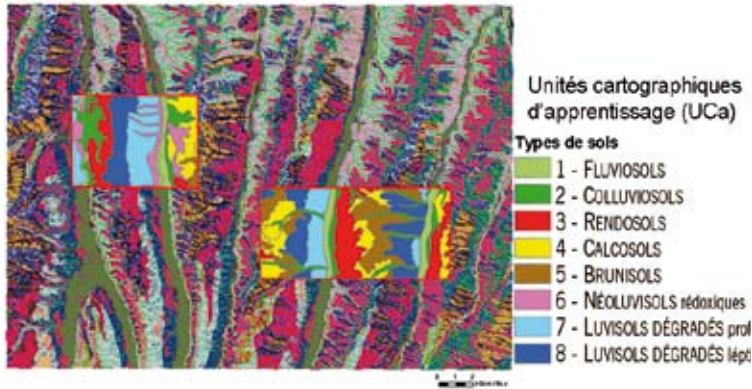


Figure 8 - Localisation des UCa (secteurs d'apprentissage) sur l'image des indices combinés.

Figure 8 - Region of interest located over the combined indices image.

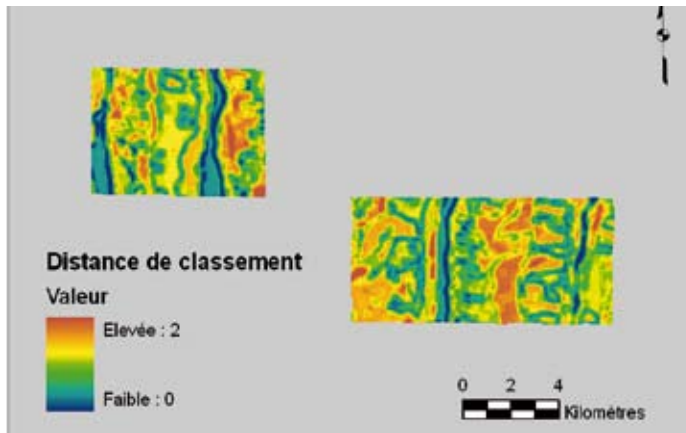


Figure 10 - Carte des distances de classement des UCp (Unités Cartographiques prédites) par rapport aux UCa (Unités Cartographiques d'apprentissage) par la métrique de Manhattan.

Figure 10 - Map of classification distances of the UCp (predicted map unit) compared to the UCa (regions of interest) by Manhattan's metric.

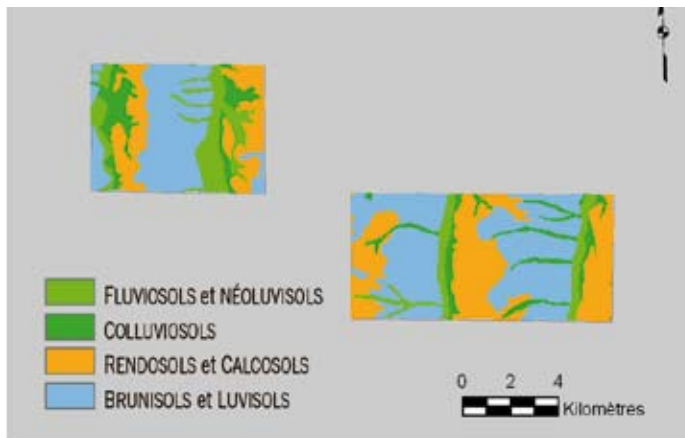


Figure 9 - Carte des unités cartographiques prédites (UCp) générée par CLAPAS lors de la première itération sur les deux secteurs de validation interne (8 classes).

Figure 9 - Soil mapping units produced by CLAPAS after the first iteration on the two regions of interest (8 classes).

Validation externe

Les classes issues de la dernière itération (UCp, Unités Cartographiques prédites) donnant le meilleur résultat (*figure 4*) sont retenues pour être généralisées à l'ensemble de la carte de Mirande avec Clapas. La carte des sols de Mirande (*Cf. Données utilisées*) sert de carte de validation. Une nouvelle matrice de confusion permet de comparer le résultat de la prédiction générée par Clapas

à la carte de validation. Il est évident que cette validation externe serait impossible en terra incognita et sert donc simplement à valider la démarche entreprise ici. Elle consiste à généraliser la classification qui a été validée sur un petit secteur de référence (validation interne), à l'ensemble d'une carte supposée présenter les mêmes caractéristiques, au regard des paramètres utilisés, sur toute son étendue.

Figure 11 - Comparaison du secteur de référence (à droite), avec celle des UCp générées par CLAPAS (à gauche) sur le même secteur (4 classes).

Figure 11 - Comparison of the reference map (right), with the one generated from CLAPAS (left) (4 classes).

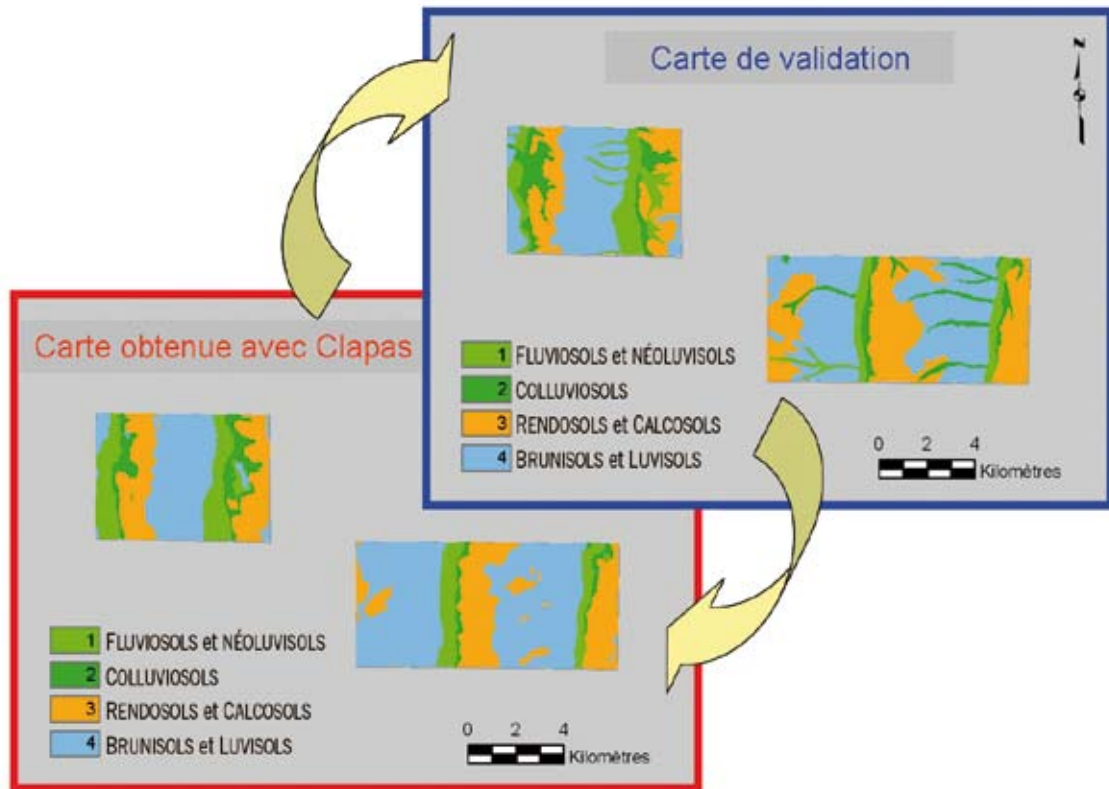


Figure 12 - Comparaison de la carte des sols simplifiée (à gauche), avec celle des UCp générées par CLAPAS (à droite) (4 classes).

Figure 12 - Comparison of the simplified soil map (left), with the one generated by CLAPAS (right) (4 classes).

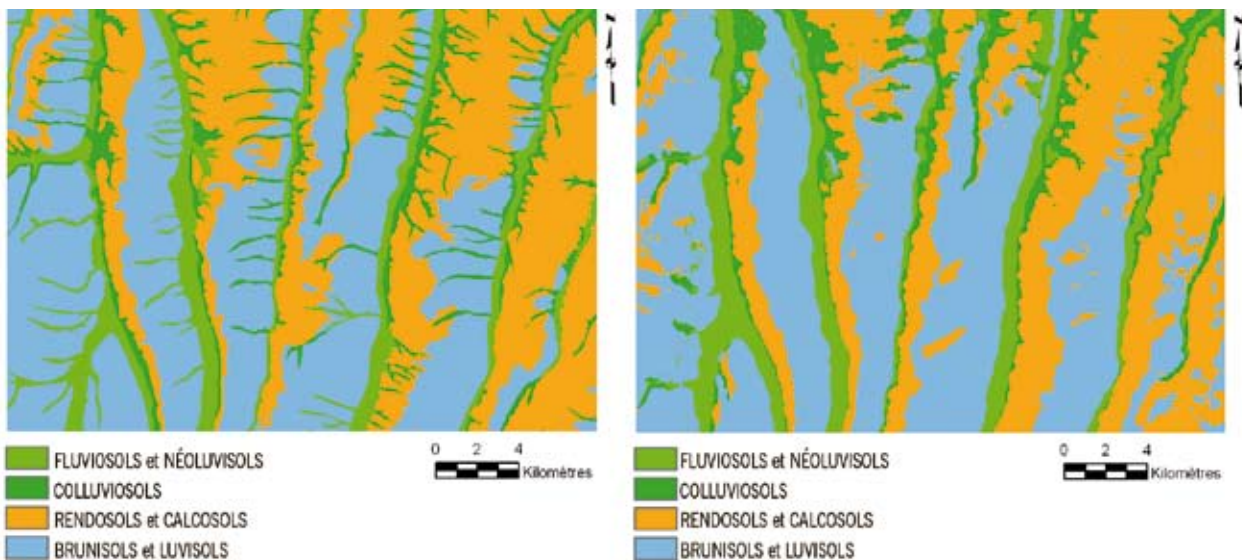


Tableau 4 - Matrice de confusion après la troisième itération (4 classes).**Table 4** - Confusion matrix after the third iteration (4 classes).

Classe prédite (Ucp)	Classe observée (Uca)				nbr pixels	erreur de commission	% de bonne classification
	1	2	3	4			
1	3444	751	91	1129	5415	0.36	64
2	393	1454	556	6	2409	0.40	60
3	153	888	8137	1043	10221	0.20	80
4	862	910	2528	14206	18506	0.23	77
nbr pixels	4852	4003	11312	16384	38551	nbr total de pixels	
erreur de d'omission	0.29	0.64	0.28	0.13			

Indice global	Indice Kappa
0.75	0.61

Tableau 5 - Matrice de confusion après la troisième itération sur l'ensemble de la carte (4 classes).**Table 5** - Confusion matrix after the third iteration on the whole map (4 classes).

Classe prédite (Ucp)	Classe observée (Uca)				nbr pixels	erreur de commission	% de bonne classification
	1	2	3	4			
1	18175	3135	662	6235	28207	0.36	64
2	4214	8181	6868	2094	21357	0.62	38
3	875	5996	60949	9877	77697	0.22	76
4	4154	3926	14957	81820	104857	0.22	76
nbr pixels	27418	21238	83436	100026	232118	nbr total de pixels	
erreur de d'omission	0.34	0.61	0.27	0.18			

Indice global	Indice Kappa
0.73	0.59

RÉSULTATS

Description statistique des UCa à fort pouvoir discriminant

Nous ne présentons ici que les indices ayant montré le meilleur potentiel de discrimination.

La *figure 5* représente l'analyse en composantes principales de notre jeu de données sur les axes F1 et F2. Ceux-ci expliquent 82 % de la variance (F1 = 53 % et F2 = 29 %).

Les ellipses colorées de cette figure correspondant aux nuages de points des UC s'étalent sur l'axe F1. Les couleurs des ellipses reprennent la légende de la *figure 8*. L'axe F1 est caractérisé, à gauche, par le vecteur « ri200 » (indice de rugosité sur une fenêtre mobile de 4 x 4 pixels) et à droite, par le vecteur « bev » (indice de Beven). Il est intéressant de constater que la répartition des ellipses de gauche à droite sur l'axe F1 correspond à l'ordre dans lequel apparaissent les sols sur une toposéquence. Cela s'explique par le fait que l'indice de Beven caractérise bien la topographie au sein du paysage. Ainsi, en remontant les versants doux, on passe progressivement des FLUVIOSOLS (UC 1), aux NÉOLUVISOLS rédoxi-ques (6), puis des LUVISOLS DÉGRADÉS profonds aux leptiques (7 et 8). Les ellipses des BRUNISOLS (5) et des CALCOSOLS (9 et 10) se recouvrent, de sorte qu'il est difficile de les discerner. En revanche, les RENDOSOLS (4) sont fortement corrélés avec le vecteur « ri200 », ce qui traduit leur position dans un relief élevé et disséqué.

Le vecteur « mrrtf » (indice MRRTF), qui contribue fortement

à l'axe 2, correspond à un étirement vertical des ellipses. Il est directement corrélé aux sommets des interfluves et des plateaux. Dans notre cas, il correspond moins à des limites d'unités cartographiques de sols qu'à des variations topographiques au sein de ces mêmes unités. Il n'apporte ainsi aucune information supplémentaire quant à leur séparabilité.

La représentation de l'indice de Beven sous forme de diagramme en boîte à moustaches (*figure 6a*) permet de distinguer trois ensembles que nous avons seuillés de façon à les séparer au mieux : 1- les FLUVIOSOLS, 2- les COLLUVIOSOLS, les NÉOLUVISOLS et les LUVISOLS DÉGRADÉS, 3 - les RENDOSOLS, les BRUNISOLS et les CALCOSOLS. La valeur négative de ce dernier groupe s'explique par la fonction logarithmique incluse dans la formule de Beven.

L'indice du relief offre d'autres opportunités pour séparer des groupes de sols (*figure 6b*). Nous avons déterminé 4 classes : 1 - les FLUVIOSOLS, les NÉOLUVISOLS rédoxi-ques et les LUVISOLS DÉGRADÉS profonds ; 2 - les LUVISOLS DÉGRADÉS leptiques ; 3 - les COLLUVIOSOLS, les BRUNISOLS et les CALCOSOLS, 4 - les RENDOSOLS.

Le choix du seuil 2 (*figure 6b*) sur ce dernier indice a été guidé par la volonté de séparer les unités 7 et 8 (LUVISOLS DÉGRADÉS épais et superficiels). On peut ainsi espérer distinguer ces plages cartographiques dans l'ensemble combinant les indices retenus.

L'indice MRVBF, fortement redondant avec l'indice de Beven (*figure 5*) n'a pas été retenu en raison d'une plus faible dispersion des moyennes par UCa au niveau des diagrammes en boîtes à moustache (non présenté ici). Les autres indices ne sont pas

présentés en raison de leur faible intérêt pour la suite des traitements.

Nous n'avons donc retenu que deux indices issus du MNA pour générer l'image d'entrée de CLAPAS : l'indice de Beven seuillé sur 3 classes et l'indice du relief seuillé sur 4 classes.

La *figure 7* représente la répartition des étages géologiques au sein des UCa. Comme attendu, les compositions de chaque UCa par rapport à ce critère, révèlent des signatures différentes. Nous intégrons donc l'information géologique à l'image d'entrée de Clapas.

Création de l'image biunivoque⁵ combinant les indices

L'image biunivoque est construite en combinant la carte géologique codée sur huit valeurs, l'indice de relief codé sur quatre valeurs et l'indice de Beven, codé sur trois valeurs. Cette image comporte 89 modalités croisées.

Validation des unités cartographiques d'apprentissage (UCa) par la métrique de Manhattan

Les 8 UCa de CLAPAS (*figure 8*) correspondent exactement à l'emprise des unités cartographiques de sols des deux secteurs de références sur lesquels nous avons réalisé l'étude statistique décrite ci-dessus

Pour être exploitables dans CLAPAS, les UCa doivent avoir des compositions de classes de pixels significativement différentes au sens de la distance de Manhattan. Ainsi, en les comparant deux à deux, une valeur proche de 2 indique une forte dissemblance de composition de leurs histogrammes, alors qu'une valeur proche de 0 indique une forte ressemblance et par conséquent l'impossibilité de les dissocier par la suite. Si la valeur de la métrique est inférieure à 1, on considère que les risques de confusions sont tels qu'ils vaut mieux fusionner les 2 classes considérées.

La boîte à outils de CLAPAS (EditPay®) permet cette comparaison.

Les résultats du *tableau 2* indiquent que les histogrammes de composition des UCa sont statistiquement différents pour toutes les UCa, à l'exception des BRUNISOLS (UC 5) que la méthode exposée ici confond *pro parte* avec les LUVISOLS DÉGRADÉS leptiques (UC 8), et surtout avec les CALCOSOLS (UC 4), leurs distances de Manhattan étant inférieures à 1.

5 : image biunivoque : image sur laquelle il n'existe qu'un seul terme pour désigner un objet et chaque terme ne doit désigner qu'un seul objet.

Résultats de la validation interne de la classification

Première itération

Les valeurs de la distance de classement correspondent à celles de la métrique de Manhattan et s'étendent de 0 à 2. Plus cette distance est élevée, proche de 2, en marron (*figure 10*) et plus Clapas a eu du mal à rapprocher une unité cartographique prédite (UCp) d'une UCa. Inversement, plus cette distance est proche de 0, en bleu, et plus l'UCp a de chance de correspondre à l'UCa. En comparant cette image à la *figure 9*, on remarque que les zones de faible distance, donc les plus faciles à classer, correspondent aux FLUVIOSOLS et aux RENDOSOLS. Inversement, les zones les plus difficiles à classer, en marron sur la *figure 10*, correspondent aux BRUNISOLS et aux COLLUVIOSOLS.

Analyse de la matrice de confusion

L'image résultat (*figure 9*) est comparée avec la carte des sols de validation (UCa du même secteur, *figure 8*) au moyen d'une matrice de confusion. Cette matrice nous permet de détecter les erreurs d'affectation et de proposer des regroupements d'UCa.

Avec un indice global de 0.49 et un kappa de 0.42, le résultat de cette première itération n'est pas très satisfaisant. Nous pouvons l'améliorer en fusionnant les classes présentant une confusion élevée (*figure 4*). La classe 6 correspondant aux NÉOLUVISOLS rédoxiques (*tableau 3*) est la moins bien prédite. La majorité des pixels de cette classe de référence (6^e colonne) a été incorrectement prédite par la classe 1 (1^{ère} ligne : 999/2 311 pixels). Nous allons donc regrouper les classes 1 et 6 lors de la 2^{ème} itération. La cinquième classe de référence (5^e colonne) correspondant aux BRUNISOLS est majoritairement prédite à tort par la classe 8 (8^e ligne : 1 456/4 173 pixels) correspondant aux LUVISOLS DÉGRADÉS leptiques. En fusionnant ces deux classes lors de la seconde itération (non détaillée ici), nous améliorons sensiblement l'indice global de classification et l'indice Kappa. Nous pouvons encore opérer quelques fusions pour atteindre l'objectif recherché d'un indice global de 75 % lors d'une troisième itération.

Troisième itération

En fusionnant les classes 7 et 8, les LUVISOLS DÉGRADÉS profonds et leptiques, (*tableau 3*) qui présentent un taux de confusion élevé (environ 1/3 de la prédiction de la classe 8 correspond en réalité à la 7^{ème} classe de référence), on améliore encore le résultat de la classification (*tableau 4*, 4^e colonne). La dernière fusion consiste à regrouper les sols calcaires (RENDOSOLS et CALCOSOLS). En effet, plus du 1/3 de la 4^{ème} classe de référence (CALCOSOLS) est faussement prédit comme appartenant à la classe 3 (RENDOSOLS), Cf. *tableau 3* : 4^e colonne (2 150/5 565 pixels). La troisième et dernière itération fournit ainsi 75 % de pixels bien classés (*figure 4*). La comparaison visuelle de la carte de référence (simplifiée à quatre UC) (*figure 11 à droite*) avec celle des unités générées par CLAPAS

(figure 11 à gauche), suffit pour juger de la forte ressemblance entre la carte prédite et la carte de validation. La matrice de confusion ci-après (tableau 4) nous le confirme.

L'indice global de bonne classification atteint 75 %. Il est conforté par un indice de kappa acceptable. Les COLLUVIOSOLS (classe 2) sont cependant surreprésentés et mal classés. On les retrouve majoritairement (un tiers de la classe) dans la classe 3. La classe 1, issue de la fusion des NÉOLUVISOLS rédoxiques avec les FLUVIOSOLS obtient un bon score (64 %), malgré un taux d'affectation erronée dans la classe 4 qui demeure important (un cinquième de la classe). La fusion des LUVISOLS avec les BRUNISOLS améliore sensiblement la pertinence de la classe 4 qui atteint ainsi 77 %. Le meilleur score, 80 % des pixels correctement affectés, est obtenu par la classe incluant les CALCOSOLS et les RENDOSOLS.

Résultats de la validation externe de la classification

Au vu des résultats obtenus par la validation interne au bout de la troisième itération, jugés satisfaisants, nous généralisons cette troisième itération à l'ensemble de la carte de Mirande. Nous comparons ensuite le résultat obtenu avec la carte des sols de Mirande sur laquelle nous avons préalablement effectué les mêmes regroupements d'unités cartographiques de sols (figure 12) à l'aide d'une nouvelle matrice de confusion (tableau 5).

Les résultats obtenus par la généralisation de la classification à l'ensemble de la carte de Mirande (tableau 5) sont très proches de ceux obtenus par la validation interne (tableau 4). L'indice global de classification reste proche de 75 % et le kappa, de 60 %. Seule la classe 2, correspondant aux COLLUVIOSOLS, subit une baisse sensible de proportion de pixels correctement affectés en passant de 60 % à 38 %. Les autres classes conservent, à un ou deux points près, les mêmes scores que lors de la validation interne.

DISCUSSION

Contraintes et atouts de la méthode CLAPAS

Les limites de CLAPAS sont de deux ordres. Elles sont d'ordre intrinsèque au logiciel d'une part, et d'ordre pratique et fonctionnel, d'autre part. Nous citerons notamment :

- Le choix des seuils des indices retenus est un paramètre d'entrée extrêmement sensible et dont la pertinence permet ou non de détecter une Unité Cartographique ; c'est donc au pédologue connaissant bien les caractéristiques des sols prédits que revient ce choix avec l'aide du géomaticien ou du statisticien.
- Les vecteurs de compositions fournis par CLAPAS définissent le contenu d'un voisinage mais n'intègrent pas la notion d'arrangement spatial (telle que la fréquence des types de contacts entre deux unités cartographiques (Robbez-Masson, 2001). Or la

prise en compte de la couverture pédologique pourrait permettre d'améliorer les prédictions à partir de lois de voisinage.

- Le nombre de valeurs de l'image sur laquelle travaille CLAPAS est actuellement limité à 255. Or ces valeurs étant issues de la combinaison des valeurs des indices utilisés, leur nombre croît très rapidement. Cela limite considérablement le nombre d'indices et nécessite un seuillage drastique pour restreindre le nombre de classes de chaque indice. En s'affranchissant de cette limite, on pourrait tester davantage d'indices et bénéficier d'une plus grande souplesse dans la classification en augmentant le nombre des classes par indices.

Du point de vue pratique et fonctionnel, la préparation des indices d'entrée est un travail fastidieux en raison de la diversité des logiciels SIG requis (ARCGIS® pour la préparation des données et une conversion au format IDRISI® pour le traitement des images par CLAPAS)

Les atouts de CLAPAS, en revanche, sont nombreux. La prise en main du logiciel est relativement aisée dès lors que les données d'entrées ont été converties au format IDRISI®. Ce logiciel présente tous les avantages d'un outil de cartogénèse numérique : objectivité, reproductibilité, généralisation à de vastes ensembles géographiques, détermination du nombre d'unités cartographiques de sols (UCS) sur des critères statistiques (Gourmelon, 2003), connaissance de l'erreur de classement par la carte des distances de Manhattan et ce qui fait son originalité, l'intégration de la notion de voisinage spatial.

La voie que nous avons exploitée ici en employant CLAPAS d'abord pour définir des secteurs de référence puis comme outil d'extrapolation de ces secteurs, met l'accent sur la dualité entre la qualité de la classification et le degré de pureté des classes recherchées. En effet, l'un se faisant au détriment de l'autre, la difficulté consiste à trouver le meilleur compromis entre un indice de classification acceptable (au moins supérieur à 70 %) et des unités cartographiques relativement pures ou tout au moins cohérentes dans les associations de sols qu'elles contiennent. C'est pourquoi nous avons pris soin de regrouper certains types de sols en grandes classes basées sur différents facteurs : la teneur en calcaire pour les RENDOSOLS et les CALCOSOLS, la structuration pour les BRUNISOLS et les LUVISOLS, et enfin le drainage naturel pour les FLUVIOSOLS et les NÉOLUVISOLS rédoxiques. Ce niveau d'agrégation est bien sûr difficilement exploitable à l'échelle du 1/100 000. En revanche, il correspond tout à fait celui que l'on cherche à obtenir au niveau des pédopaysages⁶ (Girard, 1983). Auquel cas, l'échelle de représentation du 1/250 000, d'ordre régional, trouve ici un intérêt certain. CLAPAS peut donc être utilisé par le pédologue pour aider à la prédélimitation, à l'échelle d'une région ou d'un département, des associations de sols.

6 : Pédopaysage : ensemble des horizons pédologiques et des éléments paysagiques : végétations, effet des activités humaines, géomorphologie, hydrologie, substratum, dont l'organisation spatiale permet de définir dans son ensemble une (ou une partie) d'une couverture pédologique.

Voies possibles d'amélioration

Choix des secteurs d'apprentissage

Notre approche n'utilise que deux secteurs de référence, sélectionnés à dire d'expert, qui correspondent à des minutes de cartes des sols couvrant une faible surface (64 km² au total par rapport aux 442 km² de l'ensemble de la coupure 1/50 000). Au sein d'une région donnée et inconnue, il pourrait être judicieux d'élaborer une méthode de choix des secteurs d'apprentissage de telle sorte qu'ils soient les plus représentatifs possibles des UCS à prédire. Une possibilité pourrait être de créer des histogrammes de référence sur la base de seuils déterminés par une démarche statistique. Ces histogrammes seraient ainsi non seulement représentatifs de secteurs de référence mais aussi le résultat d'une synthèse statistique propre à chaque UCS.

Choix des indices, du nombre de classes et de la géométrie des voisinages

La limite sur le nombre de valeurs pouvant être traitées par CLAPAS n'est plus aujourd'hui contrainte par le temps de calcul. Il n'est toutefois pas certain que la classification serait améliorée en multipliant le nombre d'indices et de classes qui produisent, par combinaison, ces valeurs. Un nombre trop important de classes risquerait en effet d'introduire une forme de bruit dommageable au résultat de la classification.

Bien que l'on puisse agir sur la forme et la taille de la fenêtre glissante employée pour explorer le voisinage des pixels, elle demeure fixe sur toute l'image. Il serait intéressant de pouvoir enchaîner autant d'itérations qu'il y a d'UCS en adaptant à chaque fois la forme et la taille de la fenêtre, et d'adapter les indices utilisés au type d'UCS que CLAPAS cherche à prédire. Il ne semble pas irréaliste de penser qu'une analyse du pouvoir discriminant des différentes combinaisons possibles (indice, forme et taille de la fenêtre), appliquée séparément à chaque UCS puisse permettre de délimiter pas à pas les unités recherchées. En s'appuyant sur une approche hiérarchisée par le pédologue, voire par les résultats d'une exploration systématique et automatisable, cette procédure commencerait par mettre en évidence les unités caractérisées par les « signatures CLAPAS » les plus caractéristiques. Les unités restantes seraient ensuite considérées comme des zones de transition sur lesquelles le pédologue devrait intensifier la prospection afin de mieux les définir.

Choix du descripteur de motif spatial

Le descripteur d'un secteur de référence se limite à un histogramme (vecteur de composition soumis à un classement) qui est comparé à celui du voisinage de chaque pixel de l'image traitée. Or cette approche ne tient compte ni de la disposition relative des constituants entre eux (topologie), ni des flux qu'ils échangent (Robbez-Masson, 1994). Afin d'intégrer la notion d'arrangement spatial entre les différentes classes, il serait intéressant de tester

une classification utilisant comme descripteur une matrice de co-occurrence.

CONCLUSIONS

Cette étude nous a permis de tester le logiciel CLAPAS sur la coupure IGN à 1/50 000 de Mirande. Notre approche se distingue des travaux réalisés jusqu'ici avec CLAPAS sur deux points. D'une part, nous utilisons en plus de la géologie, des indices dérivés du MNT (rugosité et Beven) dont le choix est guidé par une approche statistique intégrant la notion de zonalité. D'autre part, nous utilisons comme secteurs d'apprentissage deux cartes de sols à grande échelle couvrant environ 1/7^{ème} de la surface totale à cartographier.

Notre démarche s'appuie d'abord sur un test de cohérence interne : à chaque itération, nous comparons le résultat de la classification obtenu par Clapas avec celui de la carte de référence jusqu'à obtenir un indice global de classification de 75 % au bout de la 3^{ème} itération.

Puis on procède à une validation externe. La 3^{ème} itération est appliquée à la zone couvrant l'emprise de la carte de Mirande avant d'être confrontée à une carte des sols du même secteur. Le résultat est validé par une nouvelle matrice de confusion très semblable à celle obtenue par validation interne. On peut donc conclure que les secteurs d'apprentissage sont tout à fait représentatifs du secteur d'étude. Il aurait été intéressant d'élargir le secteur d'étude au-delà de la carte de Mirande pour connaître les limites de représentativité des secteurs d'apprentissage. Nous n'avons hélas pas pu le réaliser, faute de carte de validation.

La carte que nous obtenons ici *in fine* n'est plus tant une carte des grands types de sols qu'une carte d'un complexe de paysages pédologiques ou pédopaysages.

Ainsi, il semble que la méthode que nous avons testée, fondée sur un apprentissage à partir de secteurs de référence, est plus pertinente lorsque la généralisation est effectuée à une échelle et à des niveaux d'organisation plus synthétiques que ceux des secteurs d'apprentissage. En effet, le résultat obtenu, en fixant un seuil supérieur à 70 % de bonne prédiction, aboutit à un rendu cartographique proche des cartes pédologiques au 1/250 000 qui sectorisent non plus des types de sols mais des ensembles morphopédologiques comme les pédopaysages. Nous proposons également quelques pistes de développement des outils disponibles dans CLAPAS. Ainsi donc, CLAPAS peut constituer un outil intéressant d'aide à la cartographie à des échelles régionales en offrant au pédologue une méthode objective et reproductible lui permettant de sectoriser le paysage en entités homogènes. C'est d'ailleurs dans cet esprit que CLAPAS a été conçu d'une façon remarquablement novatrice lors de sa création par Jean-Marc Robbez-Masson.

REMERCIEMENTS

Ce travail a été initié en collaboration avec Jean Marc Robbez-Masson, auteur du logiciel CLAPAS. Il nous a accueilli et guidé avec toute la sympathie, l'enthousiasme et l'imagination qui le caractérisaient. Il nous a quitté brutalement au mois de juillet 2005, au milieu du gué, et au milieu de sa vie. Nous aurions souhaité qu'il nous accompagne plus longtemps sur ce chemin. Nous aurions été heureux et flattés qu'il puisse continuer à suivre ce travail, l'amender et qu'il accepte de le co-signer. Tous les pédologues cartographes et tous les « Digital Soil Mappers » comprendront ce que nous lui devons et ce que nous ressentons.

Nous remercions vivement Christian Walter et Jean-Paul Legros pour leur lecture critique très constructive des différentes versions de cet article.

BIBLIOGRAPHIE

- AFES, 1995 - Référentiel pédologique. Coordination : Baize D., Girard M.C., Ed. AFES/INRA. 332 p. ISBN 2-7380-0633-
- Arrouays D., Vion I., Kicin J.L., 1995 - Spatial analysis and modeling of topsoil carbon storage in temperate forest humic loamy soils of France. *Soil Science*. Vol. 159 no. 3, pp. 191-198.
- 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 Stengel P., 2004 - Le programme Inventaire gestion et Conservation des Sols. *Etude et Gestion des Sols*, 11(3), pp. 187-198.
- Beven K., Kirkby M., 1979 - A physically-based variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24, pp. 43-69.
- Bishop T.F.A., McBratney A.B. - A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, Volume 103,, Fevrier 2001, pp. 149-160.
- Bourennane H., King P., Chéry P., Bruand A., 1996 - Improving the kriging of a soil variable using slope gradient as external drift. *European Journal of Soil Science*. No. 47, pp. 473-483.
- Bourennane H., King D., Couturier A., 2000 - Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. *Geoderma*. Vol. 97, pp. 255-271.
- Bourennane H., King, D., 2003 - Using multiple external drifts to estimate a soil variable. *Geoderma*. Vol. 114 no. 1, pp. 1-18.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984 - *Classification and Regression Trees*. Wadsworth, Belmont, California, 367 p.
- Bui, E.N., Henderson, B.L., Viergever, K., 2006 - Knowledge discovery from models of soil properties. *Ecol. Model*. 191, pp. 431-446.
- Bui, E.N., Moran, C.J., 2003 - A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. *Geoderma* 111, pp. 21-24.
- Carré F., McBratney A.B., 2005 - Digital terrain mapping. *Geoderma*, Volume 128, Issues 3-4, October 2005, pp. 340-353.
- Chaplot V, Walter C, et Curmi P. 2000 - Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 97, pp. 405-422.
- Cialella, A.T., Dubayah, R., Lawrence, W., Levine, E., 1997 - Predicting soil drainage class using remotely sensed and digital elevation data. *Photogrammetric Eng. Remote Sensing* 63, pp. 171-178.
- Dagnelie P. - *Analyse statistique à plusieurs variables*. Gembloux, Presses agronomiques, 362 p., 1975
- Daroussin J., King D., Le Bas C., Vrščaj B., Dobos E., Montanarella L., *accepted* - The Soil Geographical Database of Eurasia at scale 1:1,000,000: history and perspective in digital soil mapping. *In: «Digital Soil Mapping: an introductory perspective»* - Ed. Lagacherie P., McBratney A.B., Voltz M. - Chapter 4. *Developments in Soil Science*, vol 31. Elsevier. Amsterdam.
- Dobos E., Micheli E., Baumgardner Marion F., Biehl L., Helt T. - Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*, Volume 97, November 2003, pp. 367-391.
- Dobos E., Daroussin J., Montanarella L. *A paraître* - The development of a quantitative procedure for building physiographic units for the European SOTER database. *In: Best Practice in Digital Terrain Modelling. Development and Applications in a Policy Support Environment*. Peckham R. and Jordan G. (editors), Joint Research Centre, European Commission, Ispra.
- Dobos E., Daroussin J. *A paraître* - Potential Drainage Density Index (PDD). *In: Best Practice in Digital Terrain Modelling. Development and Applications in a Policy Support Environment*. Peckham R. and Jordan G. (editors), Joint Research Centre, European Commission, Ispra.
- Gallant J.C., Dowling T.I., 2003 - A multiresolution index of valley bottom flatness for mapping deposition areas. *Water Resources Research* 39 (12) 1347 p.
- Girard M.C., 1983 - Recherche d'une modélisation en vue d'une représentation spatiale d'une couverture pédologique. Application à une région des plateaux jurassiques de Bourgogne. *Sols*, 12, 430 p.
- Girard MC, King D., 1988 - Un algorithme interactif pour la classification des horizons de la couverture pédologique. *Science du Sol*, 26, pp. 81-101.
- Girard, MC, Arousseau P, King D., Legros JP., 1989 - Apport de l'informatique à l'analyse spatiale de la couverture pédologique et à l'exploitation des cartes. *Science du Sol*, 27, pp.335-350.
- Girard M.C. & Girard M., 1999 - Traitement des données de télédétection, Editions DUNOD, ISBN 2-1000-4185-1, 529 p.
- Goulard M., et Voltz M. 1992 - Linear coregionalization model : tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24, pp. 269-286.
- Gourmelon J., 2003 - Faisabilité de l'utilisation des méthodes de segmentation sous SIG pour une stratification assistée des paysages. Application à la réalisation d'un plan d'échantillonnage pertinent dans le cadre de la constitution des bases de données géographiques « Sols et Territoires de Bourgogne ». Rapport de stage de DESS de l'Université de Bourgogne, 51p. + Annexes.
- Jenny H., 1941 - *Factors of soil formation, a system of quantitative pedology*. McGraw-Hill, New York (1941).
- King D., 1986 - Modélisation cartographique du comportement des sols basée sur l'étude de la mise en valeur du Marais de Rochefort. Thèse INA PG, 243 p.
- King D. Girard M.C., 1988 - Réflexion sur la classification des profils de la couverture pédologique. proposition d'un algorithme : VLADIMIR. *Science du Sol*, 26, pp. 239-254.
- King D., Daroussin J., Arrouays D., 1989 - Analyse cartographique et système d'information géographique en pédologie. *Science du Sol*, vol. 27 no. 1, pp. 89-92.
- Lagacherie P., 1987 - "Synthèse générale sur les études de secteur de référence drainage ONIC Ministère de l'Agriculture". Rapport CNABRL-INRA SES n°591.
- Lagacherie, P., Ledreux, C., 1992 - PROJET SAPRISTI. Systeme d'Aide à la PRédiction Intelligente des Sols par Traitement Informatique. Elaboration d'un système expert pour la reconnaissance des sols et la cartographie automatique. Séminaire INRA, Florac (FRA), 1991/10/22-24. Editions INRA, Paris, pp. 241-254.
- Lagacherie, P., Holmes, S., 1997 - Addressing geographical data errors in a classification tree for soil unit prediction. *International Journal Geographical Information Science (GBR)*. Vol. 11 no. 2, pp. 183-198.
- Lagacherie P., Robbez-Masson J.M., Nguyen-The N., Barthès, J.P., 2001 -

- Mapping of reference area representativity using a mathematical soilscape distance. *Geoderma* Vol. 101, pp. 105-118.
- Lagacherie P., 2002 - Cartographie des sols et de leurs propriétés à un niveau subrégional, Mémoire pour l'obtention du diplôme d'HDR, Université de Montpellier 2.
- McBratney A. B., Mendoça Santos M. L., Minasny B., 2003 - On digital soil mapping. *Geoderma*, Volume 117, Issues 1-2, November 2003, pp. 3-52.
- Moore I. D., Grayson R. B., Landson A. R., 1991 - Digital Terrain Modeling : a Review of Hydrological, Geomorphological, and Biological Applications. *Hydrological Processes*. Vol. 5. pp. 3-30.
- Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993 - Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57, pp. 443-452.
- Moran J.M., Bui E.N., 2002 - Spatial data mining for enhanced soil map modelling. *Int. J. Geographical Information Science* 16(6), pp. 533-549.
- 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. + annexes.
- Robbez-Masson J.M., 2000 - CLAPAS. <http://sol.ensam.inra.fr/Produits/CLAPAS/PresGen.asp> mis à jour le 15/02/2001
- Robbez-Masson J.M., Foltête J.C., Cabello, L., Flitti M., 2000 - Prise en compte du contexte spatial dans l'instrumentation de la notion de paysage. Application à une segmentation géographique assistée. *Revue Internationale de Géomatique*, 9(2), pp. 173-195.
- P. Scull, J. Franklin, O.A. Chadwick, 2005 - The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling* 181 (2005) pp. 1-15.
- Shary P.A., Sharaya L.S. and Mitusov A.V., 2002 - Fundamental quantitative methods of land surface analysis. *Geoderma*, 107(1-2) : pp. 1-32.
- Voltz M., Lagacherie P., Louchart X., 1997 - Predicting soil properties over a region using sample information from a mapped reference area. *European Journal of Soil Science (GBR)*. vol. 48, pp. 19-30.
- Vrščaj B., Daroussin J., Montanarella L. *A paraître*. SRTM as a possible source of elevation information for soil-landscape modeling. *In : Best Practice in Digital Terrain Modelling. Development and Applications in a Policy Support Environment*. Peckham R. and Jordan G. (editors), Joint Research Centre, European Commission, Ispra.
- Walter Ch. 1989 - Analyse de la variabilité spatiale de propriétés du sol au sein de trois unités cartographiques. *Science du Sol*, 27, pp. 73-76.

SITE INTERNET

- RNC, 2005 - Ressources naturelles Canada, Secteur des sciences de la Terre, Centre canadien de télédétection : http://cct.nrcan.gc.ca/glossary/index_f.php?id=3124

CARTES

- Camelio F., 1975 - Carte pédologique de Auch au 1/100 000, CACG et INRA. Notice explicative de F. Camelio, 1975, 74 p.
- Crouzel F., 1970 - Carte géologique de la France au 1/50 000, feuille de Mirande (XVIII-44) Orléans : BRGM. Notice explicative de F. Crouzel, 1970, 7 p. (document interne)
- IGN, 2003 - BDALTI au pas de 50 m. Département du Gers, Format : ESRI GRID, Projection : Lambert 2 étendu. Commandé en juin 2003 à l'IGN.

