

Utilisation d'un secteur de référence pour désagréger les unités cartographiques complexes d'un référentiel régional pédologique

Un premier essai en plaine viticole héraultaise

J. Oballos⁽¹⁾ et P. Lagacherie⁽²⁾

(1) Universidad de Los Andes, Merida, Venezuela

(2) Laboratoire sur les Interactions Sols, Agrosystèmes et Hydrosystèmes (UMR INRA-ENSAM-IRD LISAH) 2, place Viala, 24060 Montpellier cedex 1

RÉSUMÉ

Compte tenu de la complexité des couvertures pédologiques étudiées, les référentiels régionaux pédologiques réalisés en France au 1/250 000 délimitent des unités de paysage regroupant plusieurs unités typologiques de sol (UTS). Les propriétés des sols intéressant les utilisateurs étant décrites au niveau de ces UTS, leur non-délimitation des UCS est un frein important à l'utilisation effective des référentiels pédologiques régionaux. Pour surmonter cette difficulté, nous proposons de délimiter les UTS en utilisant des lois « sols-paysages » préalablement établies à partir d'une cartographie pédologique détaillée d'un secteur de référence représentatif. Les lois « sols-paysages » sont formalisées à partir des données du secteur de référence par construction interactive d'un arbre de classification utilisant une interface informatique aisément utilisable par un pédologue cartographe. Cette démarche a été testée en plaine viticole Héraultaise sur une UCS majeure du référentiel pédologique régional du Languedoc-Roussillon. L'arbre de classification obtenu utilise des variables aisément disponibles sur l'ensemble de l'UCS (distance aux UCS voisines, altitude relative et pente) et permet d'isoler avec une probabilité satisfaisante deux des quatre UTS majeures décrites dans l'UCS étudiée, les deux UTS restantes ne pouvant être localisées qu'en association avec une seconde. Les prédictions obtenues s'avèrent cohérentes avec les superficies respectives des UTS données par le référentiel régional pédologique.

Mots clés

Cartographie pédologique, lois sols-paysages, arbre de classification, changement d'échelle

SUMMARY

USING A REFERENCE AREA FOR DESAGREGATING COMPLEX SOIL MAPPING UNITS OF SMALL SCALE SOIL MAPS

Small scale (1/250 000) soil databases have been recently developed in many French regions in view to provide exhaustive soil information over vast areas with acceptable costs. However, soil mapping units (SMU) defined within these soil databases are mostly complex i.e. they include a set of distinct Soil Typological Units (STU), each of them being characterised by a set of relevant soil properties. Because

of this complexity, it is not straightforward to derive maps of soil properties from these soil databases which strongly limit their use.

To overcome this difficulty, our aim is to delineate STU within each SMU by using soil-landscape rules derived from a large scale soil survey of representative reference area. These rules produce estimates of presence of STU from easily-available maps of landscape components (e.g., in France, lithology, relief, land use). This paper describes a first evaluation of this approach in the "Plaine Viticole Centre Hérault" (figure 1 and table 1), taking the example of the desagregation of UCS 552T, a major UCS of the Languedoc-Roussillon soil database (figure 1 and table 2).

Each soil mapping units of the reference area soil map was first allocated when possible to one of the STU of the Languedoc-Roussillon on the basis of its qualitative description. This provides a simplified reference area soil map (figure 3) delineating the four major Soil Typological Units of SMU 552T, impurities - i.e. soil unit allocated to STU of neighbouring UCS)- and local soil mapping unit – i.e. soil unit not allocated to any of the STU of the regional soil database- (table 3).

Soil landscape rules were then derived by interactive construction of a classification tree using the simplified soil map of reference area as source of soil information and images of landscape components as candidate criteria of STU discrimination. The best possible classification tree was approximated following a « step by step » and a « trials and errors » approach thanks to a convivial computer interface available in software SAPRISTI (figure 2). The classification tree obtained by this approach (figure 4) allow to define 10 prediction rules – i.e. the branches of the tree - involving three types of criteria : distances to the neighbouring UCS, relative altitude and slope. Four of these rules predict major STU's of UCS 552T whereas three of them predict indistinctly impurities or local soil units. The two remaining rules define « fuzzy » zones related to uncertainty affecting relief variables. It must be noted that only two of the four predicted UTS can be delineated by the prediction rules with a satisfactory probability, the two other STU's remaining associated with one their neighbouring STU in the predictions.

The so-obtained prediction rules were then applied for desagregating UCS 552T in the study region (figure 5). A first validation of these rules were made by comparing the percentages of UTS provided by the Languedoc-Roussillon soil database with those derived from the predictions rules. Results show a good adequacy between these two informations (table 4).

This first test of desagregation of Soil Mapping Unit of regional soil database involving a large scale mapped reference area demonstrates the feasibility of such an approach. Nevertheless, several ways are to be explored to overcome the current limits :

- Defining a strategy of reference area sampling,
 - Building a quantitative approach for allocating reference area soil units to regional database STU's,
 - Using more precise maps of landscape components for defining and applying soil landscape rules (in particular for the relief),
- Dealing with fuzzy zones defined in classification trees.

Key-words

Soil survey, soil databases, downscaling, soil-landscape relations, classification tree.

RESUMEN

UTILIZACION DE UN SECTOR DE REFERENCIA PARA DESAGREGAR UNIDADES CARTOGRAFICAS COMPLEJAS DE UN REFERENCIAL PEDOLOGICO REGIONAL. Un primer ensayo en la planicie viticola del herault.

Los referenciales pedológicos regionales realizados en Francia a 1/250 000 delimitan unidades de paisaje que reagrupan varias unidades tipológicas de suelos (UTS) como consecuencia de la complejidad de las coberturas pedológicas estudiadas. Las propiedades de los suelos requeridas por los usuarios son descritas a nivel de estas UTS; la no delimitación de dichas UTS es una limitante para el aprovechamiento efectivo de la información de estos referenciales pedológicos regionales. Para superar esta dificultad, proponemos, delimitar las UTS utilizando las leyes suelo- paisaje que han sido previamente definidas a partir de una cartografía detallada de suelos de un sector de referencia representativo. Las leyes suelo – paisaje son formalizadas a partir de los datos del sector de referencia por construcción interactiva de un árbol de clasificación mediante el uso de una interfase informática fácilmente utilizable por un pedólogo – cartógrafo. Esta metodología ha sido probada en la planicie vitícola del Hérault para una de las unidades cartográficas mayores (UCS) del referencial pedológico de Languedoc – Roussillon. El árbol de clasificación obtenido utiliza variables fácilmente disponibles sobre el conjunto de las UCS (distancia a las UCS vecinas, altitud relativa y pendiente) y permite aislar con una probabilidad altamente significativa dos de las cuatro UTS mayores descritas para la UCS estudiada; las otras dos UTS, sólo pueden ser localizadas en asociación con otra UTS. Las predicciones obtenidas son coherentes con las superficies respectivas descritas en el referencial pedológico regional.

Palabras claves

Cartografía de suelos, leyes suelo-paisaje, árbol de clasificación.

Face à un besoin croissant de connaissance des sols du territoire français exprimés par de nombreux utilisateurs, le développement de référentiels pédologiques régionaux réalisés à l'échelle du 1/250 000 s'est imposé comme une solution permettant de couvrir dans des délais raisonnables l'ensemble du territoire, et ce, à des coûts acceptables (Bornand *et al.*, 1989). Cette solution est également adoptée par plusieurs pays européens (Le Bas et Jamagne, 1996) et en passe d'être généralisée à l'échelle du continent entier (Finke *et al.*, 2001). Compte tenu du niveau d'investigation choisi, l'option prise pour réaliser ces référentiels est de délimiter des unités cartographiques de sol (UCS) répondant à la notion d'unité de pédopaysage (Bornand *et al.*, 1989), portion de territoire homogène vis-à-vis d'éléments du paysage intervenant dans la différenciation des sols. Le plus souvent ces unités cartographiques regroupent en leur sein plusieurs unités typologiques de sols (UTS) qui restent non délimitées. L'inconvénient est que les propriétés des sols qui sont définies au niveau des UTS ne peuvent pas être directement spatialisées. Dès lors, la technique de spatialisation couramment pratiquée consiste à estimer ces propriétés au niveau des UCS par agrégations selon différentes techniques (de Laroche, 1994 ; Le Bas *et al.*, 1998). Cette agrégation constitue cependant

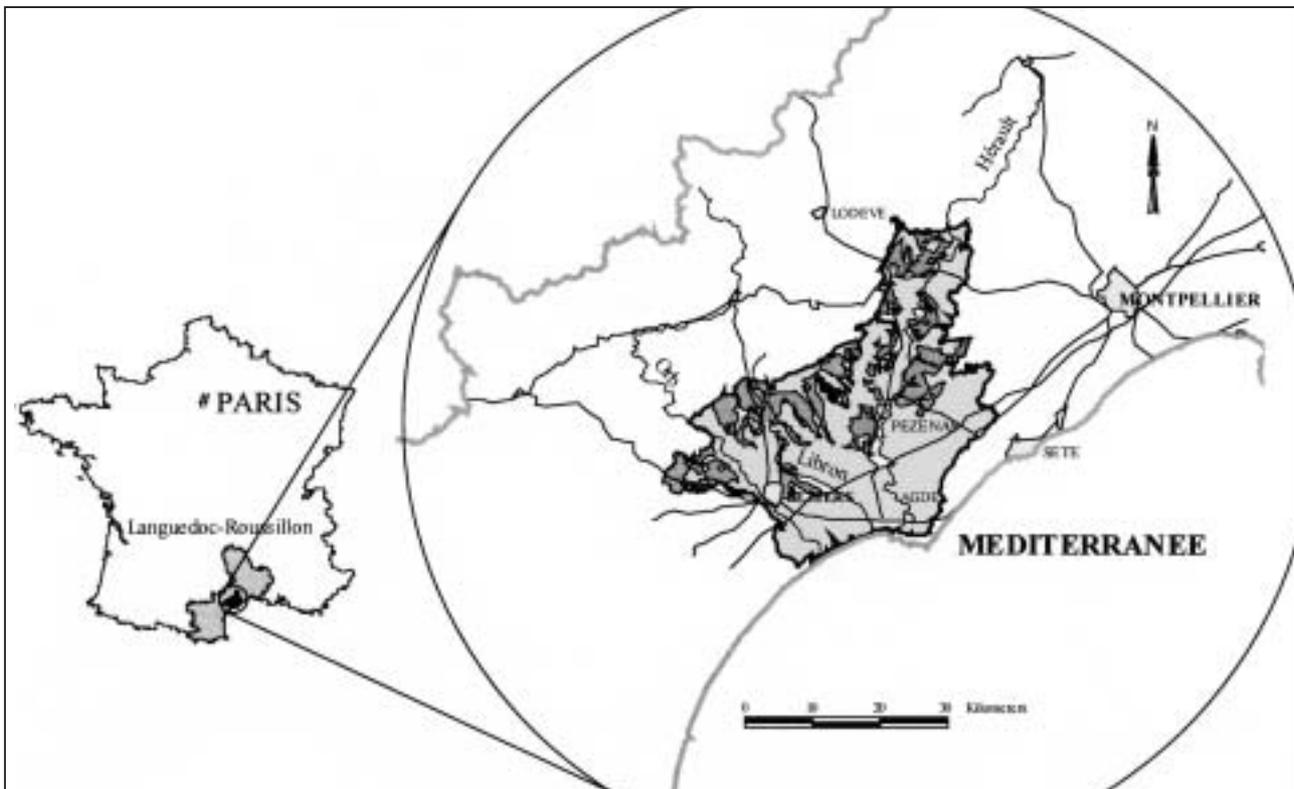
une perte de précision préjudiciable à l'utilisation effective de ces référentiels pédologiques régionaux (Le Bas *et al.*, 1998 ; Lagacherie *et al.*, 2000a).

Pour surmonter cette difficulté, des essais de désagrégation d'unités cartographiques complexes ont été menés. Ils reposent tous sur l'utilisation de lois sols-paysages (Lagacherie, 1992, Lagacherie 2002) définies comme les lois qui permettent de prédire l'occurrence des sols (et donc des UTS) à partir de cartographies d'éléments du paysage plus aisément disponibles sur de grands territoires (ex: relief, lithologie, occupation du sol). La formalisation de telles lois sols-paysages est obtenue selon deux voies différentes : soit en utilisant les informations sur l'environnement des UTS disponibles dans les légendes, notices et bases de données pédologiques (Cazemier, 1999 ; Bui et Moran, 2001), soit en utilisant des délimitations d'UTS qui peuvent exister sur des périmètres limités au sein du territoire à étudier (Bui et Moran, 2001).

Le présent papier décrit une première évaluation des possibilités et limites de cette deuxième approche dans le cas particulier où la délimitation d'UTS serait déduite d'un secteur de référence (Favrot, 1989). Cette étude constitue ainsi un premier essai d'articulation entre les deux niveaux d'investigation (référentiel 1/250 000 et secteur de référence) composant le programme national français

Figure 1 - La région d'études ■, le secteur de référence ■ et l'Unité Cartographique de Sol 552T du Référentiel Régional Pédologique du Languedoc-Roussillon ■

Figure 1 - The study region, the reference area and the Soil Mapping Unit 552T of Languedoc Roussillon soil database



« Inventaire Gestion et Conservation des sols » (Favrot *et al.*, 1994). Comme dans les précédents essais, l'approche utilise une analyse de données de type arbre de classification en y apportant cependant deux éléments nouveaux : une construction de l'arbre de classification assistée par ordinateur au moyen d'une interface graphique adaptée, ce qui permet de mobiliser les connaissances relatives au modèle d'organisation spatiale des UTS et de définir un arbre de classification floue tenant compte de l'imprécision des données géographiques utilisées. L'exemple choisi concerne la désagrégation d'une UCS majeure du référentiel pédologique du Languedoc Roussillon, l'unité 552T (collines de pentes moyennes, glacis et vallons sur molasses sablo-gréseuses du miocène), au sein de la plaine viticole « Centre Hérault » (34).

RÉGION D'ÉTUDE

La région d'étude

« plaine viticole Centre-Hérault »

Située dans la plaine côtière du Languedoc-Roussillon entre Béziers et Montpellier (*figure 1*), la région d'étude couvre environ 1 200 km². Elle est constituée par les parties aval cultivées des bassins versant de l'Hérault, de l'Orb et du Libron. C'est une plaine composée d'une succession de collines (« puech ») de glacis et de vallons où domine largement la monoculture de la vigne. Elle s'étage du bord de la Méditerranée jusqu'à 120 mètres à l'intérieur des terres.

La région d'étude présente une géologie complexe qui influence fortement la distribution des sols. Le substratum de base de la région est constitué par des sédiments marins du miocène à faciès variables : dépôts fins de marnes bleues, dépôts sableux plus grossiers avec des bancs de calcaires coquilliers, calcaire dur, molasses à dragées de quartz. Ce substratum a été partiellement recouvert au Pliocène et au quaternaire par une succession de dépôts fluviaux à pierrosité variable dont les plus récents constituent des terrasses qui s'étagent le long des cours actuels de l'Orb, de l'Hérault et de leurs affluents principaux. Une activité volcanique quaternaire (basalte et tufs volcaniques), la présence d'affleurements pré-miocènes dans l'est de la région et des phénomènes de colluvionnement toujours actifs ajoutent encore à la complexité de la distribution lithologique de la région.

L'extrême variabilité des sols de la région d'étude résultant de cette complexité est décrite dans le référentiel régional pédologique du Languedoc Roussillon (Bornand *et al.*, 1992) par 36 unités de pédopaysage (ou UCS) incluant au total 96 unités typologiques de sol. Nous présentons dans le *tableau 1* les 6 UCS principales qui couvrent au total 55 % de la région.

L'unité Cartographique de Sol étudiée

L'étude est centrée sur l'unité 552T, UCS très représentée dans la région d'étude (273 km² soit 23 % de la région). Elle regroupe 12 UTS (*tableau 2*) qui présentent toutes la particularité d'être développées sur le substratum miocène, affleurant ou colluvionné. Les deux facteurs de différenciation principaux à l'œuvre dans l'unité sont :
a) La lithologie, liée à la multiplicité des faciès reconnus dans cet

Tableau 1 - Les cinq Unités Cartographiques de Sol majeures de la région d'étude et les principaux types de sols rencontrés.

Table 1 - The five major Soil Mapping Units and main soil types of the study region

UCS	Dénomination	Nombre d' UTS	Types de sol RP (Baize et Girard, 1995)
173A	Moyenne vallée et plaine alluviale des fleuves côtiers languedociens (Orb, Hérault)	3	Fluvisols calcaire, Fluvisols neutres à calciques et Calcosols alluviaux, parfois redoxiques
309A	Moyennes terrasses caillouteuses de l'Orb et de l'Hérault (secteur médian du fleuve)	3	Fersialsols, Brunisols calciques et Calcosols
309B	Terrasses irrégulièrement caillouteuses de la Lergue et de l'Hérault en amont de Gignac	3	Brunisols calciques, Calcosols, Fersialsols calciques
309V	Hautes terrasses du Quaternaire ancien et/ou Villafranchien (Orb Hérault)	4	Fersialsols luviqes, Fersialsols, Luvisols rédoxiques
372P	Glacis pliocènes à sables et galets de Servian (Hérault)	4	Calcosols, Calcisols, Fersialsols luviqes
552T	Collines de pente moyenne, glacis et vallons sur molasses sablo-gréseuses	12	Rendisols, Calcosols, Calcisols, Calcosols redoxiques, Fersialsols

Tableau 2 - Les Unités Typologiques de Sol de l'Unité Cartographique de Sol 552T du référentiel régional pédologique du Languedoc Roussillon

Table 2 - The soil Typological Units of Soil Mapping Unit 552T of Languedoc Roussillon regional soil database

UTS	Surf % UCS	Lithologie	Relief	Types de sol RP (Baize & Girard, 1995)
974	27	Molasse, grès, calcaire gréseux miocènes	Pentes moyennes	Calcosol
505	20	Colluvions de molasse miocène	Glacis	Calcosol
506	20	Alluvio-colluvions de molasse miocène	Dépressions	Calcosol redoxique
56	10	Molasse miocène	Pente faible à moyenne	Calcosol
55	8	Molasse miocène à inclusions siliceuses	Pente faible à moyenne	Calcisol
977	4	Limon du Pliocène	Pente moyenne à forte	Calcosol
12	2	Calcaire gréseux miocène à grain de quartz	Pente faible	Rendisol
39	2	Sable issu de molasse miocène	Pente faible à moyenne	Calcosol
514	2	Calcaire lacustre massif miocène	Surfaces tabulaires	Lithosol
975	2	Calcaire blanc miocène	Replats et versants	Calcosol
11	1	Cailloutis de quartz villafranchiens	Pente nulle	Fersialsol
13	1	Sables et argiles du Pliocène	Pente faible	Fersialsol

étage (sable gréseux, grès, calcaire gréseux, calcaire gréseux à dragée de quartz) ou colluvionnés des formations voisines (sable, argile et limon du pliocène).

b) Le relief, liés aux phénomènes d'érosion et de colluvionnement récents qui isolent sur les points hauts les UTS développées sur les faciès les plus résistants à l'érosion (UTS 974, 514) et règle la profondeur des sols en fonction de la position dans le paysage, plus ou moins favorable au colluvionnement (augmentation de profondeur le long de la séquence d'UTS 974, 56, 505, 506).

Il est à noter que les UTS ne présentent pas toutes la même importance dans l'UCS 552T puisque les 4 UTS les plus représentées (974, 505, 506 et 56) couvrent 77 % de la superficie de l'UCS. La distribution de ces UTS majeures au sein de l'UCS 552T est illustrée par la coupe schématique présentée *figure 2*.

DONNÉES UTILISÉES

Les données nécessaires à l'étude envisagée sont de deux natures distinctes :

- données pédologiques permettant d'appréhender la distribution des sols;
- données sur les facteurs d'environnement permettant de prédire les sols en fonction des lois sols-paysages.

Les données pédologiques

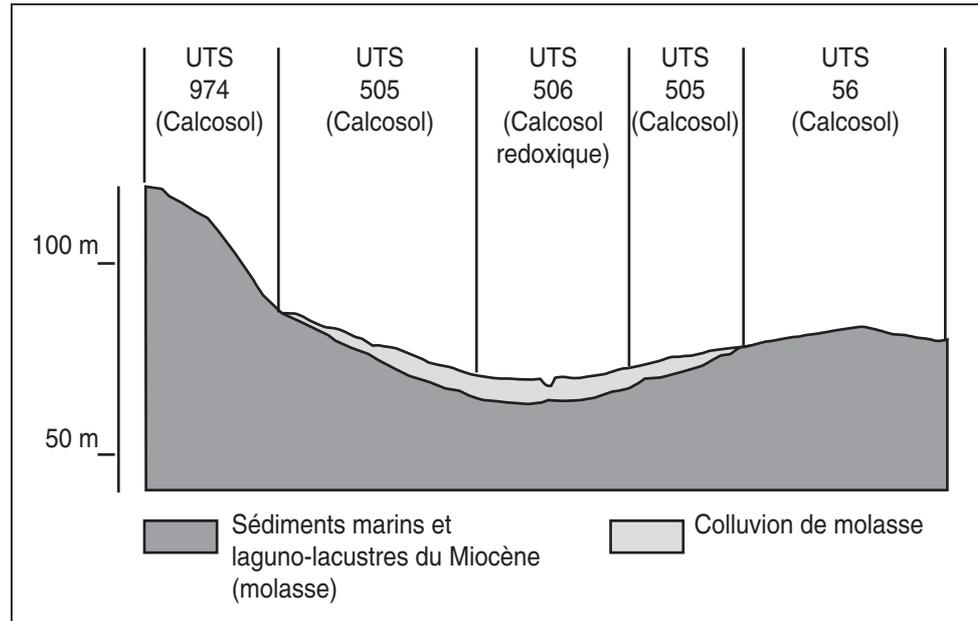
Nous situons l'étude dans le contexte du programme IGCS selon lequel le territoire français serait à terme couvert par deux types de données pédologiques spatialisées : un référentiel pédologique régional au 1/250 000 et une série de référentiels pédologiques locaux constitués par des études de type « secteur de référence » (Lagacherie *et al.*, 2000b).

Dans le cas de notre région d'étude, ces référentiels sont les suivants :

- le référentiel régional pédologique du Languedoc Roussillon (Bornand *et al.*, 1992), (Barthès *et al.*, 1999), qui, sur la région d'étude, a bénéficié d'une étude pédologique antérieure au 1/100 000 (Bonfils, 1993). Ce référentiel a été numérisé sous le SIG ARC/INFO, les données de description d'unités étant enre-

Figure 2 - Logique de distribution des UTS majeures au sein de l'UCS 552T du Référentiel Régional Pédologique du Languedoc Roussillon.

Figure 2 - Spatial organisation of the main soil typological units within the soil mapping unit 552T of Languedoc-Roussillon soil database.



gistrées sous SGBD Access (Bornand *et al.*, 1994). Il est utilisé dans la présente étude pour localiser précisément dans la région le périmètre d'étude, correspondant à la délimitation de l'UCS 552T. Il sera également utilisé en tant que donnée sur des facteurs d'environnement (cf 3.2.)

- La carte pédologique « Alignan – Roujan – Caux » située sur les trois communes citées (figure 1), d'une superficie totale de 1598 ha, est constituée par les cartes du bassin versant de Roujan (Andrieux *et al.*, 1993), échelle 1/5 000) et du secteur Alignan-Roujan-Caux « ARC » (Robbez-Masson et Barthès, 1996), échelle 1/10 000. Ces cartes ont été synthétisées par fusion des unités les plus ressemblantes au vu de leur description morphologique et constituent une carte au 1/25 000 qui a été numérisée sur ARC/INFO. Nous avons retenu comme secteur de référence la partie du secteur recoupant le périmètre de l'UCS 552T soit 552 ha et 18 unités cartographiques sur les 30 initialement délimitées sur la carte au 1/25 000. Afin d'obtenir, à partir de ce secteur de référence, une délimitation des UTS du référentiel pédologique régional, ces 18 unités ont été appariées, lorsque c'était possible, avec les UTS au vu de leurs descriptions morphologiques. Les résultats et problèmes liés à cet appariement seront présentés au début de la partie résultats.

Les données sur les facteurs d'environnement

Ces données ont été choisies selon le double critère de disponibilité sur un vaste territoire et de pertinence pour représenter un facteur d'environnement important pour la différenciation des UTS

de l'UCS 552T. Ainsi, deux types de données ont été retenus :

- données de relief,
- données de position relative par rapport aux autres UCS que l'unité 552T.

Il est à noter que nous n'avons pas utilisé dans cette étude la carte géologique au 1/50 000. En effet, il s'est avéré que ce document était en partie redondant avec le référentiel pédologique régional et qu'il n'était pas suffisamment précis cartographiquement pour délimiter les différents faciès miocènes présents au sein du périmètre de l'unité. De même, les documents cartographiques délimitant des unités de végétation ou d'occupation du sol n'ont pas été mobilisés dans cette étude, le périmètre considéré étant très majoritairement dévolu à la monoculture de la vigne.

Les données de relief

Les données de relief ont été extraites de la BD TOPO de l'Institut Géographique National qui constitue la version numérisée de la carte topographique au 1/25 000. Les courbes de niveau espacées tous les 5 mètres d'altitude contenues dans la BD TOPO ont été utilisées pour construire un modèle numérique de terrain de résolution 50 m et de précision apparente décimétrique selon la chaîne de traitement du module TIN d'ARC/INFO. Ce modèle numérique de terrain a permis de calculer deux variables de relief :

- L'altitude relative par rapport au niveau de base de la plaine viticole. Cette variable a été préférée à l'altitude simple afin de tenir compte du dénivelé général de la plaine viticole entre le piémont du massif central et la mer. Elle a été déterminée par soustraction au MNT original d'un nouveau MNT calculé en prenant, pour chaque point, l'altitude minimale dans un rayon de

3 km (taille suffisante pour intégrer un fond de vallée de l'Hérault ou de l'un de ses affluents, quelle que soit la position du site à caractériser).

- La pente, dérivée d'après la méthode de Zevenbergen et Thorne, (1987) dont l'algorithme est disponible sous ARC View.

Il est à noter que des indicateurs de relief plus élaborés utilisés dans d'autres études pour prédire les sols comme par exemple les courbures de relief (Park *et al.*, 2001) ou des indices hydrologiques (Chaplot, 1998) n'ont pas été mobilisés dans cette étude. En effet, bien qu'*a priori* plus pertinents vis-à-vis des processus de différenciation pédologique, ils se sont avérés peu efficaces dans notre contexte, étant, de par leur complexité, plus sensibles à l'imprécision du MNT que les variables décrites précédemment.

Données de position relative par rapport aux UCS voisines

Afin de pouvoir tenir compte d'une part des apports ou affleurements de matériaux présents dans les UCS voisines de 552T et, d'autre part, des incertitudes de positionnement de limites des UCS, une image de la position relative de chaque site vis-à-vis des unités voisines a été déterminée sur le périmètre de l'UCS 552T. Cette position relative a été traduite par une distance géographique à la limite la plus proche de chaque UCS voisine, distance calculée pour les trois unités majeures qui bordent sur une longueur suffisante l'unité 552T dans le secteur de référence: les plaines alluviales récentes (unité 173A), les terrasses moyennes des affluents de l'Hérault (unité 309C) et les glacis pliocènes (unité 372P).

En pratique, la distance géographique à la limite la plus proche d'une UCS a été calculée après avoir isolé dans trois thèmes ARC/VIEW distincts les limites communes entre l'UCS 552T et chacune des trois UCS voisines retenues. L'algorithme de distance euclidienne d'ARC/View a été ensuite employé pour calculer, sur chaque point de l'unité 552T, les distances géographiques recherchées.

DÉMARCHE DE FORMALISATION DES LOIS SOLS-PAYSAGES

L'objectif de la démarche est d'utiliser les données présentées précédemment pour formaliser les lois sols-paysages prédisant les UTS qui composent l'UCS 552T en tout point de cette unité. Les données utilisées consistent en un ensemble de points couvrant systématiquement (tous les 50 m) le secteur de référence, pour lesquels une identification à une UTS et les variables d'environnement sont connues (*cf section précédente*). La formalisation des lois sols-paysages résultant de la démarche est destinée à être appliquée sur l'ensemble de l'unité 552T, à l'extérieur du secteur de référence, où chaque point est uniquement renseigné par les variables d'environnement.

A la suite de travaux précédents démontrant son intérêt dans ce contexte (Lagacherie, 1992), (Bui et Moran, 2001) une démarche

par arbre de classification a été retenue pour formaliser les lois sols-paysages. Cette démarche consiste à définir une combinaison optimale de critères portant sur les variables d'environnement, de façon à segmenter un ensemble de sites en plusieurs sous-ensembles les plus homogènes possibles vis-à-vis de leur appartenance à une UTS. L'homogénéité de chaque sous-ensemble peut être évaluée quantitativement par un indicateur calculé à partir des proportions respectives des UTS dans le sous-ensemble, comme par exemple « gini index of diversity » (Breiman *et al.*, 1984) d'un sous-ensemble X ($I(X)$) :

$$I(X) = 1 - \sum_{i=1}^{i=n} p_i^2 \quad (1)$$

où p_i est la proportion de la $i^{\text{ème}}$ UTS dans le sous-ensemble X , i variant de 1 à n , le nombre d'UTS dans X . L'homogénéité maximum est atteinte lorsque $I(X) = 0$ (une seule UTS avec une proportion de 1, les autres étant à 0).

Les lois sols-paysages sformant chacune une branche de l'arbre de classification et définissant un tel sous-ensemble sont formalisées, par exemple, par la règle suivante:

si l'altitude relative au site x_0 est supérieure à 33,5 m,
et si la pente au site x_0 est inférieure à 2,5 %,
alors les probabilités d'apparition des UTS 506, 505...
sont respectivement 0,67, 0,05...

La recherche des combinaisons optimales de critères - c'est-à-dire minimisant à la fois les erreurs de prédiction et fournissant les sous ensembles les plus homogènes vis-à-vis des UTS au vu de l'indicateur présenté ci-dessus - est recherchée automatiquement par des algorithmes désormais courants dans les logiciels d'analyses de données (Breiman *et al.*, 1984), algorithmes que nous avons proposé d'améliorer pour tenir compte des erreurs affectant les données géographiques utilisées (Lagacherie et Holmes, 1997). Ce type d'algorithme s'utilise dans le cas où le nombre d'unités de sol est élevé et/ou il n'existe pas d'expertise *a priori* sur le modèle d'organisation de ces unités. Or, nous nous situons dans un contexte différent où le nombre d'UTS est limité (une seule UCS analysée à la fois) et où les lois sols-paysages sont en partie connues grâce au référentiel pédologique régional et à l'expérience tirée de l'étude d'un secteur de référence. Nous avons donc émis l'hypothèse que cette connaissance pouvait être mobilisée pour améliorer la recherche de l'arbre de classification représentant au mieux le modèle d'organisation spatial des UTS de l'UCS. Pour cela, nous avons utilisé une construction d'arbre de classification assistée par ordinateur qui constitue un des modules du logiciel SAPRISTI (Munoz *et al.*, 2002). L'approche considérée consiste, comme les algorithmes automatiques, en une construction de l'arbre de classification par étape, chaque étape visant à définir le critère optimal permettant de segmenter l'ensemble de sites traités à cette étape. L'opérateur (le pédologue) est assisté dans sa recherche du critère optimal par une interface graphique (*figure 3*). Sa démarche générale est la suivante :

- 1) Sélection par le pédologue du nœud de l'arbre à segmenter à partir d'une visualisation de l'arbre de classification courant (*planche 1a*).
- 2) Sélection par le pédologue de la variable d'environnement support à la définition du critère. SAPRISTI affiche alors l'image de la variable sélectionnée.
- 3) Recherche par le pédologue du critère optimal par approximations successives grâce à une visualisation graphique des délimitations d'UTS à isoler (*planche 1b*).
- 4) Visualisation graphique du résultat de la segmentation. Calcul par SAPRISTI du gain d'homogénéité apporté par la partition soit :

$$\Delta[I(X)] = I(X) - \left[\frac{N'}{N} I(X') + \frac{N''}{N} I(X'') \right] \quad (2)$$

$I(X)$, $I(X')$ et $I(X'')$ étant respectivement les indices de diversité du nœud père (X) et des deux nœuds fils (X' et X''), S , S' et S'' étant les superficies respectives du nœud père et des deux nœuds fils.

- 5) Validation ou rejet par le pédologue de la segmentation réalisée.
- 6) Retour à l'étape 1.

La souplesse de cette approche permet au pédologue d'utiliser des éléments d'analyse qui ne sont pas pris en considération dans les algorithmes automatiques. Ainsi, il est possible, en jouant sur l'ordre des critères et des unités à discriminer, de définir des stratégies d'identification des UTS proches de celles qui pourraient être utilisées sur le terrain. D'autre part l'imprécision des données géographiques utilisées peut être prise en compte en utilisant une segmentation floue. Ceci est réalisé en définissant des segmentations en trois sous ensembles (exemple pente : 0-2,5 %, 2,5 - 5 %, + de 5 %), le sous-ensemble intermédiaire (dans l'exemple 2,5 % et 5 %) correspondant à une zone floue, conséquence de l'imprécision liée à la variable pente.

RÉSULTATS

La démarche présentée ci-dessus a été mise en œuvre en plaine viticole centre Hérault sur l'unité cartographique de sol 552T. Nous présentons successivement les résultats concernant l'appariement des UTS avec les unités cartographiques de sol du secteur de référence, l'arbre de classification obtenu et la cartographie des UTS déduite de cet arbre sur l'ensemble de l'UCS.

Appariement des UTS avec les unités cartographiques de sol du secteur de référence

Pour chacune des 18 unités cartographiques délimitées dans le secteur de référence, un appariement avec les UTS de l'unité 552T

a été réalisé par comparaison des descriptions qualitatives figurant respectivement dans la légende de la carte et dans le référentiel régional pédologique. Ainsi, par exemple, l'unité cartographique 13 du secteur de référence décrite comme « Sols peu à moyennement profonds 30-60 cm, brun jaune, calcaires, sablo-limoneux à limono-sableux sur molasse altérée (Calcosol) » a été appariée à l'UTS 974 du référentiel Régional Pédologique décrite comme « Sol moyennement profond (40-60 cm), limono-sableux, très calcaire, brun jaune, structure polyédrique sub-anguleuse dominante (Calcosol) ». Selon les unités du secteur de référence à appairer, trois cas se sont présentés (*tableau 3*):

- *Appariement effectif avec une UTS de l'unité 552T.* Ceci concerne 5 unités du secteur de référence représentant au total 62 % en superficie de l'unité. Ces unités se répartissent dans 4 UTS majeures de l'UCS 552T dont la superficie cumulée estimée d'après les données du référentiel pédologique régional atteint 77 % (*tableau 4*), ce qui atteste de la bonne représentativité du secteur de référence vis-à-vis de l'UCS étudiée.

- *Appariement avec une UTS appartenant à une UCS autre que 552T.* Ceci concerne 10 unités du secteur de référence représentant au total 16 % en superficie de l'unité. Ces unités typologiques représentent des impuretés liées à l'imprécision de délimitation de l'UCS. Compte tenu de l'échelle du référentiel pédologique, le niveau de ces impuretés reste largement dans la norme acceptable.

- *Pas d'appariement possible avec une UTS du Référentiel Régional Pédologique.* Ceci concerne 5 unités du secteur de référence. Il s'agit d'unités résultant du voisinage entre l'UCS 552T et une UCS voisine : unité de sol située dans la zone de transition entre les deux UCS ou développée sur des colluvions issues d'un mélange entre matériaux des deux UCS. Ces unités de transition n'ont pas été décrites dans le référentiel régional vraisemblablement à cause de leur faible importance relative dans chacune des UCS qu'elles séparent.

A l'issue de cet appariement, la carte du secteur de référence a été simplifiée (*figure 3*) afin de ne conserver que les délimitations des UTS de l'unité 552T qui en sont issues. Les unités correspondant aux deux derniers cas explorés ci-dessus ont été regroupées en deux unités respectivement dénommées « impuretés » et « unités non appariées ». C'est cette carte du secteur de référence ainsi simplifiée qui est utilisée pour construire l'arbre de classification.

Construction de l'arbre de classification des UTS de l'unité 552T

La stratégie adoptée pour discriminer les 6 unités résultant de l'étape précédente a consisté, dans un premier temps, à tenter d'isoler les impuretés et les unités de transition non appariées en utilisant les images des distances aux UCS voisines, l'hypothèse sous-jacente étant que ces unités sont d'autant plus présentes que les distances aux UCS voisines sont faibles. Dans un deuxième temps, l'altitude relative a été utilisée pour distinguer d'une part les

Tableau 3 - Appariements entre Unités Typologiques de Sol de l'unité 552T et les Unités Cartographiques du secteur de référence
Table 3 - Allocation of reference area soil mapping unit to Soil Typological Unit of unit 552T

Type	Type de sol RP	UTS RRPLR	UC du SR	Superficie (% du SR)	Commentaires
UTS de l'UCS 552T	Calcosol	974	13	15,9	
	Calcosol	56	14	7,4	
	Calcosol	505	24	15,9	
	Calcosol redoxique	506	27, 28	22,8	
				Total	62,0
Impuretés	Brunisol calcique	304	7, 8	8,4	UTS décrite dans l'UCS 309C
	Calcosol	321	18, 20	4,4	UTS décrite dans l'UCS 372P
	Fluviosol calcique	101	1, 2, 3	3,0	UTS décrite dans l'UCS 173A
	Fersialsol	345	10	0.1	UTS décrite dans l'UCS 309V
				Total	15,9
UTS Non appariées	Brunisol calcique		9	11,7	Unité de transition 552T – 309C
	Calcosol		15, 17	4,8	Unités de transition 552T – 372P
	Calcosol		25	4,0	Sur colluvionnements issus de 372P
	Fluviosol calcique		29	1,5	Sur alluvions de petit cours d'eau
				Total	22,0

UTS correspondant à des affleurements du substratum (en position haute), et d'autre part les UTS développées sur des colluvionnements (en position basse). Enfin, au sein de chacun des groupes, la pente a été utilisée pour affiner la discrimination. Pour les UTS développées sur affleurement, il a été ainsi distingué l'UTS 974, la moins profonde car développée sur des positions topographiques plus favorables à l'érosion (pente forte) de l'unité 56, plus profonde car relativement protégée de l'érosion (pente faible). Pour les UTS développées sur colluvionnement, la pente a permis de distinguer l'unité 505, développée sur des glacis colluviaux de l'unité 506 occupant les dépressions à fond plat, donc de pente très faible à nulle.

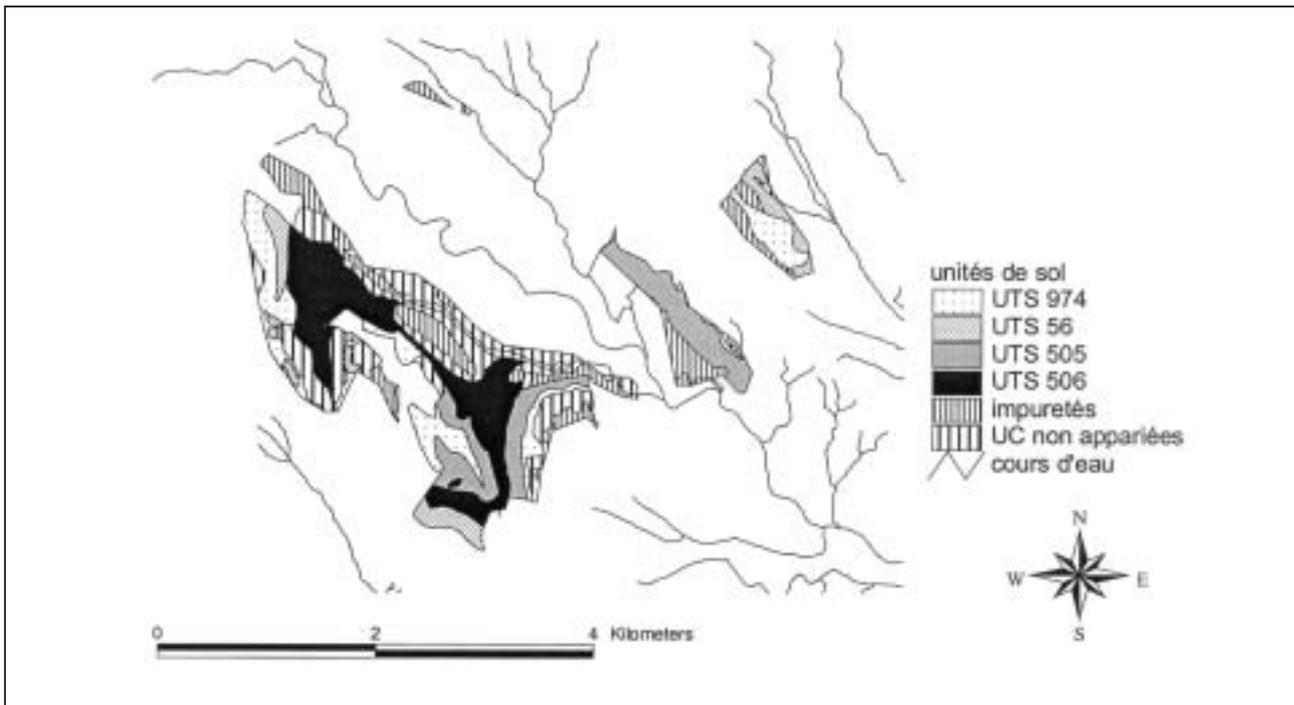
Pour chaque segmentation de l'arbre de classification utilisant la pente ou l'altitude, l'imprécision des données sur le MNT a été prise en compte en définissant, aux limites des zones discriminées, des zones de transition définies par des intervalles de confiance. Ces intervalles de confiance, définis au seuil de 80 % sur la base d'estimations d'erreur (Lagacherie, 1992) sont de 1,7 m pour l'altitude relative et de 1,25 % de la pente. Ainsi par exemple, pour une limite fixée à l'altitude relative de 35 m, la zone de transition inclura tous les points voisins de la limite dont l'altitude, estimée par le MNT, se situe entre 33,3 m et 36,7 m.

L'arbre de classification obtenu (figure 4) découpe l'UCS 552T en 10 unités de prédictions discriminées par trois types de données : les distances aux UCS voisines, l'altitude relative et la pente. Ces unités de prédictions permettent d'isoler d'une part les « impuretés »

reconnues sur le secteur de référence (unités de prédictions 1 à 3, discriminées par des distances faibles aux UCS voisines) et d'autre part les UTS appariées dans le secteur de référence (unités de prédictions 4, 6, 8, 10). Les trois unités de prédictions supplémentaires correspondent à des zones de transition, délimitées selon la méthode décrite ci-dessus. Ces trois dernières unités de prédiction ne sont pas renseignées par des probabilités d'apparition d'UTS puisqu'on peut faire l'hypothèse que ces probabilités évoluent de façon continue dans les domaines de valeur définis par les probabilités des unités de prédiction qu'elles séparent. Ainsi, par exemple, au sein de l'unité de prédiction 7 qui correspond à une zone de transition (figure 4), les probabilités d'apparition de l'UTS 505 varient entre 32 % (valeur pour l'unité de prédiction 6) et 9 % (valeur pour l'unité de prédiction 8) avec des valeurs d'autant plus proches de 9 % que l'altitude relative est proche de 36,9 m (bornes inférieures d'altitude relative de l'unité de prédiction 8).

La stratégie de découpage envisagée ne réussit à isoler avec une probabilité d'apparition satisfaisante (> 50%) que les UTS situées aux deux extrêmes de la toposéquence, respectivement l'UTS 974 (altitude élevée et pente forte, unité de prédiction 10) et l'UTS 506 (altitude faible et pente faible, unité de prédiction 4). Les deux autres UTS, 505 et 56, fortement présentes respectivement dans les unités de prédiction 6 et 8, sont confondues avec les unités qu'elles côtoient majoritairement sur la carte du secteur de référence, respectivement les UTS 974 et 506. Une présence plus importante de

Figure 3 - Carte du secteur de référence après appariement, utilisée pour formaliser les lois sols-paysages sur l'unité 552T
Figure 3 - The simplified reference area soil map used for deriving soil landscape rules



ces deux unités dans le secteur de référence serait peut-être nécessaire pour caractériser leur spécificité vis-à-vis des variables de relief utilisées.

Cartographie prédictive des UTS sur la région d'étude et validation

L'arbre de classification a été appliqué sur l'ensemble du territoire couvert par l'UCS 552T. Il en résulte une carte prédictive (*planche 2*) où sont délimitées les différentes unités de prédictions issues de l'arbre de classification, chaque unité étant caractérisée par une distribution de probabilités (*figure 4*) permettant de l'affecter à une ou deux UTS selon les cas.

Pour tenter de valider ces prédictions d'UTS, nous avons calculé les probabilités globales de chaque UTS en effectuant une moyenne des probabilités d'UTS dans chaque unité de prédiction (zone de transition exceptée), pondérée par les surfaces respectives de ces unités selon la formule suivante :

$$P_i = \frac{1}{S} \sum_j s_j P_{ij} \quad (3)$$

Avec :

P_i : probabilité globale de la $i^{\text{ème}}$ UTS de l'UCS 552T prise en compte dans les prédictions ($i = 1,4$),

S : surface totale de l'UCS 552T,

s_j : surface de la $j^{\text{ème}}$ unité de prédiction de l'arbre, j prenant les valeurs 1,2,3,4, 6,8, 10 (*voir arbre de classification figure 4*),
 p_{ij} : probabilité de la $i^{\text{ème}}$ UTS dans la $j^{\text{ème}}$ unité de prédiction.

Ces probabilités ont été comparées aux proportions des UTS estimées par les auteurs du référentiel régional pédologique (*tableau 4*). Les résultats montrent une bonne adéquation entre les deux sources de données, l'ordre d'importance des UTS majeures de l'UCS 552T étant respecté, ainsi que les ordres de grandeur des proportions respectives de ces UTS. Ainsi, les prédictions issues du secteur de référence ne semblent pas introduire de biais notable. Par contre, elles ne peuvent naturellement pas prédire les unités mineures de l'UCS 552T qui n'étaient pas représentées dans le secteur de référence.

CONCLUSION

Dans cette étude, nous avons tenté d'explorer la possibilité de désagréger les unités cartographiques de sol (UCS) d'un référentiel régional pédologique au 1/250 000 en utilisant un secteur de référence comme source de données pour formaliser les lois sols-paysages des Unités Typologiques de Sol (UTS) composant cette UCS. Les règles sols-paysages issues de cette formalisation ont été appliquées sur l'ensemble d'une UCS à partir de données géographiques disponibles sur de vastes territoires (MNT). Ceci représente, au niveau du Référentiel Régional Pédologique, une potentialité de

Figure 4 - Arbre de classification formalisant les lois sols-paysages sur l'UCS 552T. Pour chaque unité de prédiction correspondant à un nœud terminal de l'arbre, est donné le pourcentage estimé d'apparition de chaque UTS (chiffre du tableau).

Signification des abréviations : D173A ; D372P, D309C : distances aux UCS voisines 173A, 372P et 309C, AR : altitude relative, P : pente.

Figure 4 - Classification tree formalizing soil landscape rules within SMU 552T. For each prediction units which corresponds to a terminal node of the tree we provides the estimated percentage of each soil typological units.

Abbreviations meaning : D173A ; D372P, D309C : distances to neighbouring UCS 173A, 372P et 309C, AR : relative elevation, P : slope gradient.

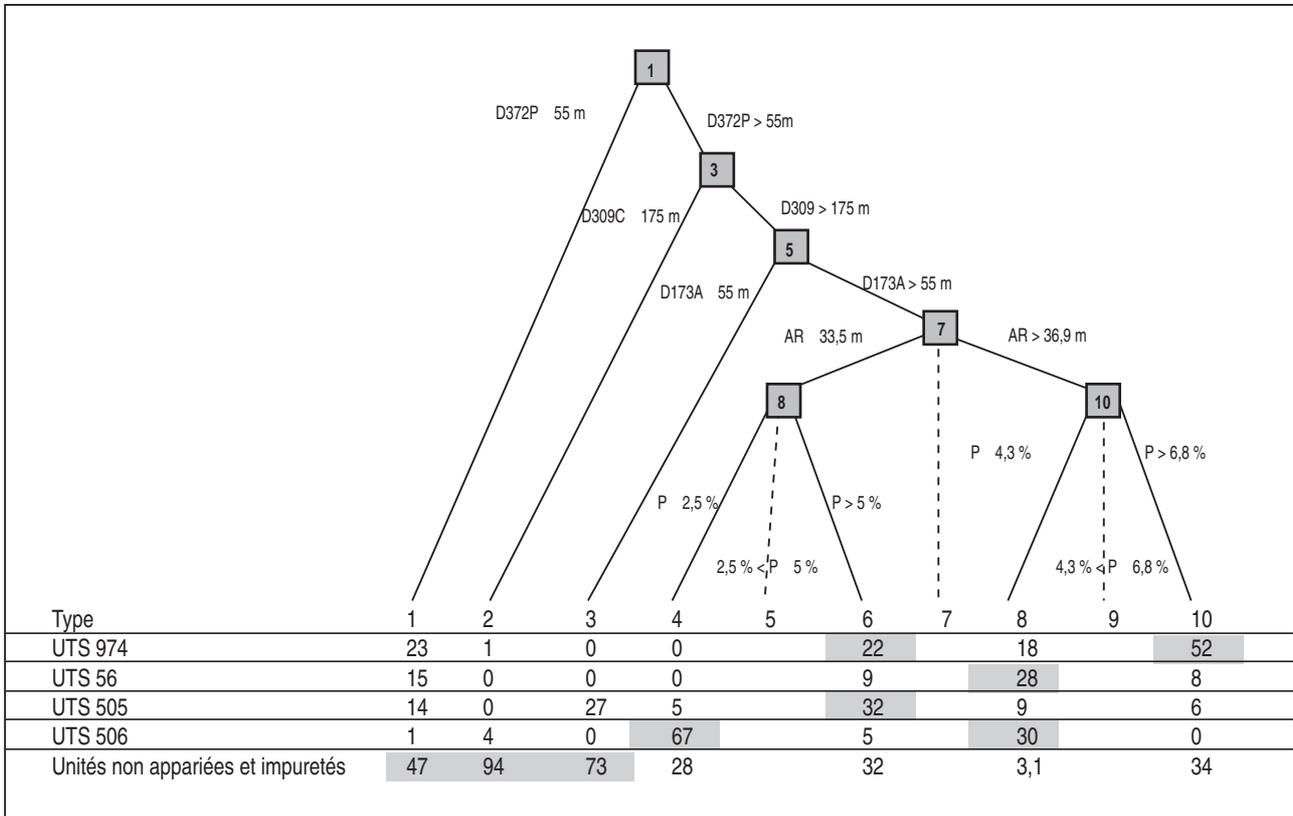


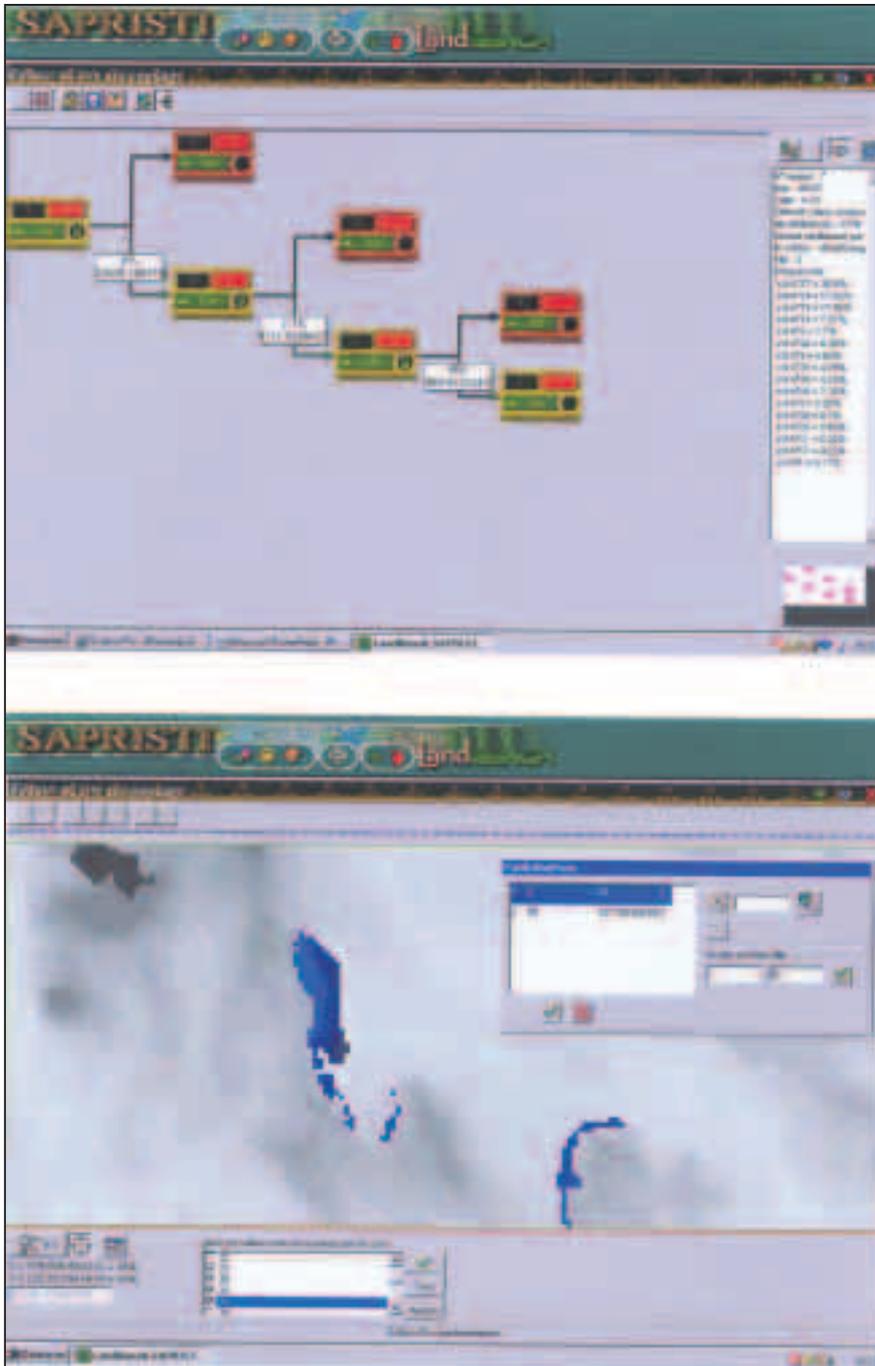
Tableau 4 - Pourcentages globaux d'apparition des Unités Typologiques de sol donnés par le Référentiel Régional Pédologique du Languedoc Roussillon et calculés à partir des prédictions

Table 4 - Global percentages of presence of STU provided by the Languedoc Roussillon Regional Soil Database and estimated from soil predictions

UTS RRPLR	Type de sol RP	Pourcentages en superficie	
		d'après RRPLR	d'après prédictions
974	Calcosol	27	23
506	Calcosol	20	18
505	Calcosol	20	18
56	Calcosol redoxique	10	12

Planche 1 - Interface graphique du logiciel de construction d'arbre de classification assisté par ordinateur : a) visualisation de l'arbre de classification courant et des caractéristiques des nœuds, b) définition interactive d'une nouvelle partition à partir d'une superposition visuelle du groupe de sols à isoler et de l'image du critère.

Planche 1 - *The Graphical interface of the interactive classification tree builder software : a) visualisation of the current classification tree and its node characteristics, b) interactive definition of a new partition using visual overlay of soil group to be isolated and of image criteria.*



gain de précision importante et illustre l'intérêt de coupler dans un même programme national de cartographie, des investigations à des échelles différentes.

L'étude exploratoire menée démontre la faisabilité et les potentialités d'une telle démarche sur l'exemple d'une UCS majeure de la plaine viticole Languedocienne. Le nombre limité d'UTS et la connaissance préalable de leur organisation spatiale nous ont amenés à utiliser une démarche de formalisation par arbre de classification qui laisse largement la place à l'intervention de l'expert pédologue. Pour cela, une interface graphique adaptée a été élaborée au sein du logiciel SAPRISTI.

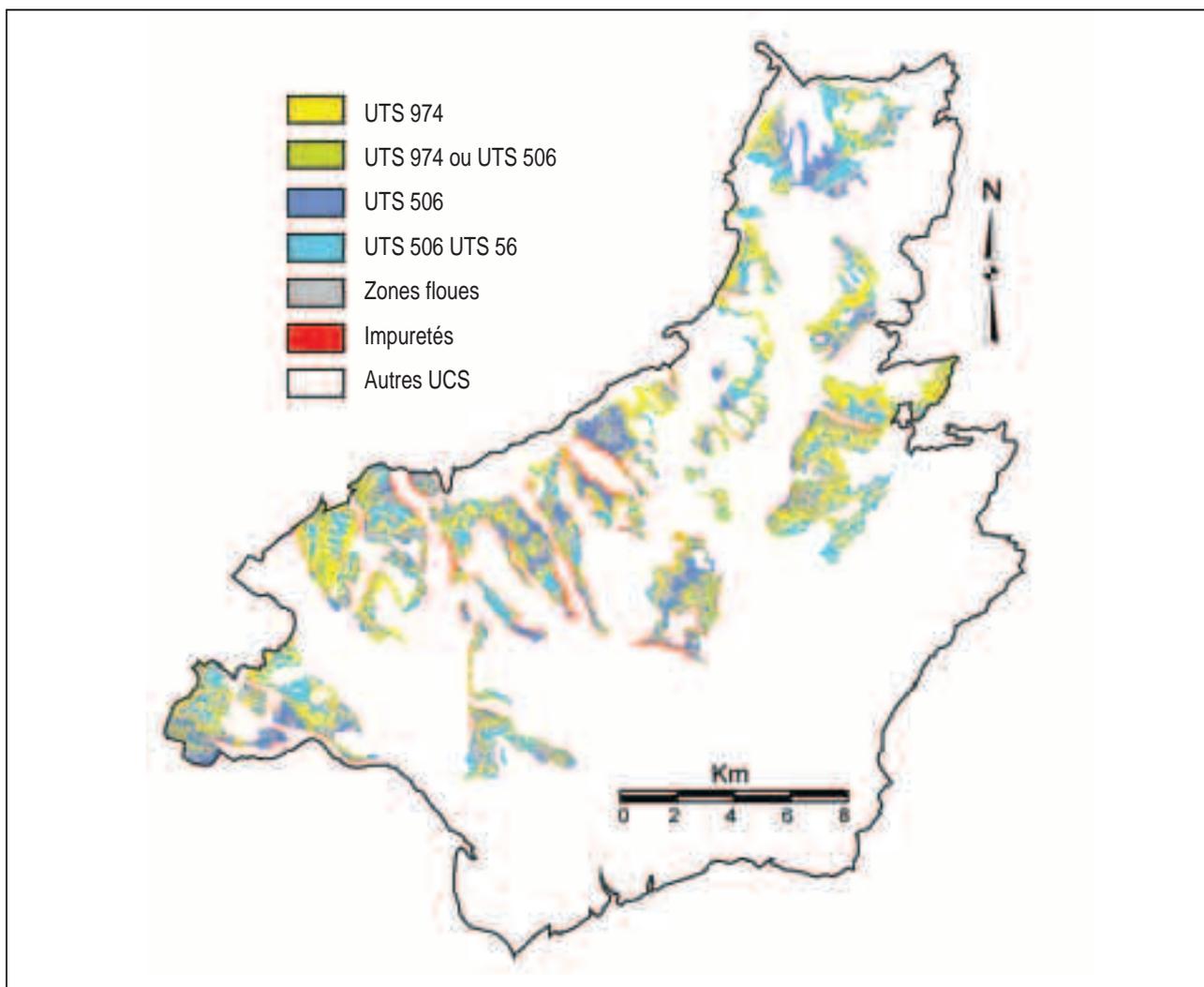
Les résultats obtenus se présentent sous la forme d'une carte prédictive couvrant l'ensemble de l'UCS, chaque unité de prédiction étant caractérisée par une probabilité d'apparition des UTS. Ces unités de prédiction isolent avec des probabilités d'apparition élevées une ou deux UTS majeures parmi les quatre présentes dans l'UCS étudiée. D'autres unités de prédiction concernent l'isolation partielle des impuretés mises en évidence par l'étude de secteur de référence. Des zones de transition entre unités de prédiction permettent de tenir compte de l'imprécision des variables d'environnement utilisée pour les prédictions. La démarche est partiellement validée par le fait que les prédictions respectent les pourcentages globaux d'occupation des UTS indiqués dans le référentiel pédologique régional.

Dans la perspective d'une utilisation effective d'une telle approche dans le cadre d'un programme de cartographie, plusieurs améliorations sont nécessaires pour surmonter les limites actuelles :

- Une stratégie d'échantillonnage du secteur de référence doit être raisonnée de façon à mieux couvrir les UTS décrites dans le référentiel régional pédologique. Dans l'étude réalisée, les unités mineures n'ont pu être prédites car elles n'existaient pas dans le secteur de référence. Un positionnement plus judicieux

Planche 2 - Carte prédictive des Unités Typologiques de Sol de l'unité 552T sur la région d'étude

Planche 2 - Prediction map of Soil Typological Unit of 552T unit over the study region



et une taille plus importante de ce secteur auraient permis, sans avoir l'ambition de caractériser les UTS très faiblement représentées, d'étendre un peu le nombre d'UTS prédites.

- L'appariement entre unités typologiques de sol définies par des auteurs différents et à des échelles différentes est une opération demandant beaucoup d'expertise et très sensible vis-à-vis des résultats. Une approche plus quantitative est nécessaire pour objectiver cette étape.

- Du fait de l'imprécision du MNT utilisé, les zones de transition couvrent une surface importante. L'estimation des probabilités d'apparition des UTS au sein de ces zones en tenant compte des unités de prédiction voisines reste à faire.

- Les prédictions d'UTS restent encore fortement limitées par la précision des données d'environnement, particulièrement, celles rela-

tives au relief. Des MNT plus précis doivent être utilisés afin de réduire l'importance des zones de transition et de permettre l'utilisation de critères de relief plus pertinents vis-à-vis des processus pédologiques gouvernant la différenciation des sols. Dans d'autres contextes pédologiques, de nouvelles données d'environnement autre que le relief (p.ex. télédétection de couverture végétale ou d'occupation du sol) doivent également être testées.

REMERCIEMENTS

Les auteurs remercient M.C. Girard et J.M. Vinatier pour leurs commentaires constructifs sur une première version de cet article.

BIBLIOGRAPHIE

- Andrieux P. *et al.*, 1993 - Le Bassin versant de Roujan. Caractéristiques générales du milieu, INRA UFR science du sol Montpellier, Montpellier. 36 pp
- Barthès J.P., Bornand M. et Falipou P., 1999 - Pédopaysages de l'Hérault. IGCS - INRA, Montpellier, 195 pp.
- Bonfils P., 1993 - Carte pédologique de la France au 1/100 000 - Feuille de Lodève (notice + carte). INRA SESCOF, 205 pp.
- Bornand M., Arrouays D., Baize D. et Jamagne M., 1989 - Cadre méthodologique d'une cartographie régionale des sols à l'échelle du 1/250 000. *Science du sol*, 27 (1) : 17-20.
- Bornand M., Barthès J.P. et Bonfils P., 1992 - Carte des pédopaysages du Languedoc Roussillon à l'échelle du 1/250,000. Laboratoire de Science du Sol, Montpellier.
- Bui E.N. et Moran C.J., 2001 - Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103 (1-2) : 79-94.
- Cazemier D., 1999 - Utilisation de l'information incertaine dérivée d'une base de données sols. Application à la cartographie des propriétés hydriques à l'échelle régionale. Thèse de Doctorat, Ecole Nationale Supérieure Agronomique, Montpellier, 169 pp.
- Chaplot V., 1998 - Organisation spatiale des sols hydromorphes de fond de vallée. Modélisation prédictive de leur distribution. Thèse de doctorat, Ecole Nationale Supérieure Agronomique de Rennes, 257 pp.
- de Laroche E., 1994 - Contribution à la phase de prototypage du Système d'Information Géographique du Languedoc Roussillon. Présentation de deux applications pilotes pour la gestion de l'espace rural, ENGREF montpellier, ENSA Montpellier, INAParis-Grignon., pp. 51pp.
- Favrot J.C., 1989 - Une stratégie d'inventaire cartographique à grande échelle: la méthode des secteurs de référence. *Science du sol*, 27 (4) : 351-368.
- Favrot J.C., Arrouays D., Bornand M., Girard M.C. et Hardy R., 1994 - Informatisation et spatialisation de la ressource sol: le programme « Inventaire, gestion et conservation des sols ». *Cahiers Agricultures*, 1994 (3) : 237-246.
- Finke P. *et al.*, 2001 - Base de données géoréférencée des sols pour l'Europe. Manuel de procédures. Bureau Européen des Sols. 174 pp
- Lagacherie P., 1992 - Formalisation des lois de distribution des sols pour automatiser la cartographie pédologique à partir d'un secteur pris comme référence. Cas de la petite région « Moyenne Vallée de l'Hérault ». thèse de doctorat, Université des Sciences et Techniques du Languedoc, Montpellier, 175 pp.
- Lagacherie P., Cazemier D.R., Martin-Clouaire R. et Wassenaar T., 2000a - A spatial approach using imprecise soil data for modelling crop yields over vast areas. *Agriculture, Ecosystems & Environment*, 81 (1) : 5-16.
- Lagacherie P. *et al.*, 2000b - Cahier des Clauses Techniques Générales (CCTG) pour la réalisation d'une étude pédologique en vue d'établir un référentiel sol local, programme Inventaire Gestion et Conservation des Sols, Ministère de l'Agriculture et de la Forêt, INRA.
- Lagacherie P. et Holmes S., 1997 - Addressing geographical data errors in a classification tree for soil unit predictions. *Int. J. Geographical Information Science*, 11 (2) : 183-198.
- Lagacherie, 2002 - Cartographie des sols et de leurs propriétés au niveau sub-régional. Mémoire d'Habilitation à Diriger des Recherches. Université de Montpellier 2, Ecole doctorale « sciences de la terre et de l'eau » 48 pages
- Le Bas C. et Jamagne M., 1996 - Soil databases to support sustainable development, SESCOF INRA - Joint Research Centre, Institute for Remote Sensing Application, Orleans.
- Le Bas C., King D., Daroussin J., Nicoullaud B. et Ngongo M., 1998 - Impact of errors in the assessment of agronomic constraints to crop production at a European scale., XVIth world congress of Soil Science, Montpellier (France).
- Munoz M., Robbez Masson J.M. et Lagacherie P., 2002 - SAPRISTI. UMR Sol & Environnement ENSAM-INRA Montpellier, Montpellier.
- Park S.J., McSweeney K. et Lowery B., 2001 - Identification of the spatial distribution of soils using a process-based terrain characterization. *Geoderma*, 103 : 249-272.
- Robbez-Masson J.M. et Barthès J.P., 1996 - Cartographie des sols des communes d'Alignan-du-Vent, Roujan et Caux. Caractérisation en vue de la maîtrise du fonctionnement hydrique. INRA-ENSAM, Montpellier.
- Zevenbergen C.G. et Thorne C., R., 1987 - Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12: 47-56.