

Vérifications de la qualité des Référentiels Régionaux Pédologiques en France :

Etat des lieux et enseignements

B. Laroche et J. Doux

INRA - Infosol Orléans - domaine de Limère, 45166 Olivet Cedex France

RÉSUMÉ

La constitution des Référentiels Régionaux Pédologiques (R.R.P.), un des volets du programme Inventaire, Gestion et Conservation des Sols, est à mi-parcours. Cette avancée nous permet de dresser un premier bilan de la qualité des bases de données produites dans ce cadre. Cette analyse bénéficie du recul du à la production et la vérification d'une quinzaine d'études (étude = un R.R.P. correspondant à un département ou à une région).

Cette vérification s'intègre dans une procédure de labellisation - ou d'attribution d'un niveau de qualité - délivrée par le ministère en charge de l'agriculture. Le protocole permet de contrôler par divers moyens (requêtes automatisées, expertise) la qualité des données graphiques (contours des unités cartographiques) et sémantiques (tables de la base de données, structurée en format Donesol). Ces étapes aboutissent à la production de recommandations de mise en conformité des R.R.P. soumis à cet exercice.

Cet article décrit l'ensemble des procédures de vérifications effectuées sur les bases de données et fait un état des lieux des R.R.P. produits. Il synthétise l'ensemble des remarques transmises aux maîtres d'ouvrages régionaux. Le non-respect de la structure de la base de données, c'est-à-dire du modèle Donesol, et la définition non conforme des unités cartographiques sont les critiques les plus souvent émises sur les R.R.P. Les autres remarques consistent principalement en des incohérences entre les champs des différentes tables.

L'objectif de cet article est de montrer d'une part les procédures de vérifications des bases de données, et d'autre part de fournir des enseignements et des recommandations pour les futurs Référentiels produits dans le cadre du programme IGCS. Plus généralement, notre travail illustre l'intérêt d'une procédure rigoureuse et systématique de vérification des bases de données géographiques sur les sols et montre les écueils les plus fréquemment rencontrés.

Mots clés

Base de données, qualité, sol, vérification.

SUMMARY**ACCURACY OF SOIL DATA BASE IN FRANCE : State of progress in 2006**

The constitution of Regional maps of soils (R.R.P.), one of the goals of the programme Inventory, Management and Conservation of Soils, is achieved for half of the French territory. This state of progress allows to realise a first assessment of the quality of databases produced within this programme. For this analysis, we use data coming from the production and the data quality checks of about fifteen studies (study = a R.R.P.).

This quality control is integrated into a procedure of labellisation - or of attribution of a quality level - which is delivered by the French ministry in charge of the agriculture. This process allows to control by various ways (systematic queries on data, expert judgement) the geometric data (boundaries of the cartographic units) and the semantic data (data includes in the tables of the database, structured in DoneSol format). These stages end in the production of recommendations for raising data quality of databases to mandatory levels.

This article describes the data control processing and reviews the quality of controlled soil databases. It synthesises all the remarks formulated to the database producers. The most important failures concern the disrespect of the structure of the data base and poor definitions of the cartographic units. Others are mainly incoherences between fields in the different tables.

The objective of this article is first to show the database control processing, and the second to supply recommendations for future mapping. More generally our work demonstrates out the interest of running rigorous and systematic data quality control, and points out the most frequent failures that were encountered.

Key-words

Soil databases, quality, soil, control

RESUMEN**VERIFICACIONES DE LA CALIDAD DE LOS REFERENCIALES REGIONALES PEDOLÓGICOS EN FRANCIA :****Estado y enseñanza**

La constitución de los Referenciales Regionales Pedológicos (R.R.P.), uno de los aspectos del programa Inventario, Gestión y Conservación de los suelos, esta a mitad del camino. El adelantado nos permite hacer un primer balance de la calidad de las bases producidas en este cuadro. Este análisis beneficia del retroceso debido a la producción y a la verificación de unas quincenas de estudios (estudio = un R.R.P. que corresponde a un departamento o a una región).

Esta verificación se integra en un procedimiento de aprobación - o atribución de un nivel de calidad - entregada por el ministerio en carga de la agricultura. El protocolo permite controlar por diversos medios (preguntas automatizadas, peritajes) la calidad de los datos gráficos (contornos de las unidades cartográficas) y semántica (tablas de la base de datos, estructura bajo formato Donesol). Estas etapas permiten la producción de recomendaciones de puestas en conformidad de los R.R.P. sometidos a este ejercicio.

Este artículo describe el conjunto de los procedimientos de verificaciones efectuadas sobre las bases de datos y hace un estado de los R.R.P. producidos. Sintetiza el conjunto de las observaciones transmitidas a las autoridades regionales contratantes. El incumplimiento de la estructura de la base de datos, es decir del modelo Donesol y la definición no conforma de las unidades cartográficas son las críticas las más a menudo emitidas sobre los R.R.P. Las otras observaciones consisten principalmente en incoherencias entre los campos de las diferentes tablas.

El objetivo de este artículo es mostrar de un lado los procedimientos de verificaciones de las bases de datos, y del otro lado proporcionar enseñanzas y recomendaciones para los futuros Referenciales producidos en el cuadro del programa IGCS. Más generalmente nuestro trabajo ilustra el interés de un procedimiento riguroso y sistemático de verificación de las bases de datos geográficos sobre los suelos y muestra los escollos los mas frecuentemente encontrados.

Palabras clave

Base de datos, calidad, suelo, verificación.

Les bases de données géographiques sur les sols sont de plus en plus fréquemment utilisées pour des applications environnementales, des évaluations de risques et des aides à la planification (Le Bas *et al.*, 2004 ; Le Bas et Schnebelen, 2006). Dans ce contexte, il est essentiel de disposer d'informations sur la qualité de ces bases de données. Ce constat a incité à faire un premier bilan des inventaires pédologiques constitués dans le cadre des Référentiels Régionaux Pédologiques (R.R.P.). Les notions de qualité ou de « précision » des cartes pédologiques, recouvrent de fait de nombreuses composantes, à la fois aux plans graphique et sémantique (Legros, 1978). Une étude alarmante a montré que 80 % des études pédologiques réalisées dans le monde seraient quasi inutilisables parce qu'elles sont incomplètes, que leur fiabilité est inconnue ou que leurs références spatiales sont erronées (Groot, 1993). De gros efforts ont été faits au niveau national, avec l'édition des normes standard de description des sols (NF X31-003) (Falipou *et al.*, 2002), ainsi qu'avec l'édition et la révision de Cahiers des Clauses Techniques Générales (Infosol, 2006). En France, Legros *et al.*, (1992) proposent de premières mesures de vérification sur les bases de données sol. Plus récemment, d'autres travaux se sont attachés à définir la qualité des bases de données géographiques sur les sols et à relier cette qualité à une échelle conseillée pour leur utilisation (Bishop *et al.*, 2001 ; Hengl et Husnjak, 2006).

Le programme Inventaire Gestion et Conservation des Sols (I.G.C.S.) a été mis en place au début des années 90, pour fédérer et harmoniser les inventaires pédologiques en France. Un des volets de ce programme est le Référentiel Régional Pédologique (R.R.P.) qui vise à produire un rendu cartographique à l'échelle 1/250 000^e, associé à une base de données structurée selon un format standard : Donesol (Gaultier *et al.*, 1993 ; Grolleau *et al.*, 2004). Afin de s'assurer de la fiabilité de la donnée produite, une procédure de vérification de la qualité a été mise en place, pour garantir aux usagers actuels ou futurs un degré de « qualité » de l'information qu'ils utilisent. Cette procédure, appliquée à une quinzaine de bases de données, permet aujourd'hui d'établir un premier bilan sur la qualité de ces bases, sur les défauts les plus fréquemment observés, et sur les préconisations à retenir pour le futur.

LE PROGRAMME IGCS

Définition

Le programme national « Inventaire, Gestion et Conservation des Sols » (I.G.C.S.), a été créé et est conduit conjointement par le Ministère chargé de l'Agriculture et l'Institut National de la Recherche Agronomique (INRA).

Une des composantes de ce programme est le **Référentiel Régional Pédologique** (R.R.P.) qui est une base de données régionale ou départementale sur les sols. Sa précision spatiale

correspond à celle d'une représentation cartographique à l'échelle 1/250 000^e. Il représente une synthèse actualisable des connaissances sur la répartition et la nature des grands types de sols d'une région ou d'un département ; il a pour objectif de déterminer et de localiser les caractéristiques principales des sols et du milieu en vue de leur gestion et de leur conservation.

La représentation cartographique consiste en un découpage en pédopaysages, qui sont définis par la combinaison des facteurs suivants : nature et organisation de la couverture pédologique (systèmes de sols), nature (lithologie) et structure géologique, forme du relief dominant ou types de relief associés (morphologie), végétation naturelle et/ou systèmes culturels existants. Sa précision sémantique est déterminée par le nombre d'observations ponctuelles (sondages et profils) et par les variables descriptives et analytiques contenues dans la base.

Ces inventaires pédologiques comprennent ainsi deux ensembles distincts reliés entre eux par le numéro d'Unité Cartographique de Sol (U.C.S.) : la base géographique et la base sémantique (King *et al.*, 1994).

- La base géographique permet la localisation spatiale des objets cartographiés. Elle est gérée au sein d'un système d'information géographique, et les entités manipulées sont des polygones (contours des plages cartographiques), et des points (emplacements des observations ponctuelles).

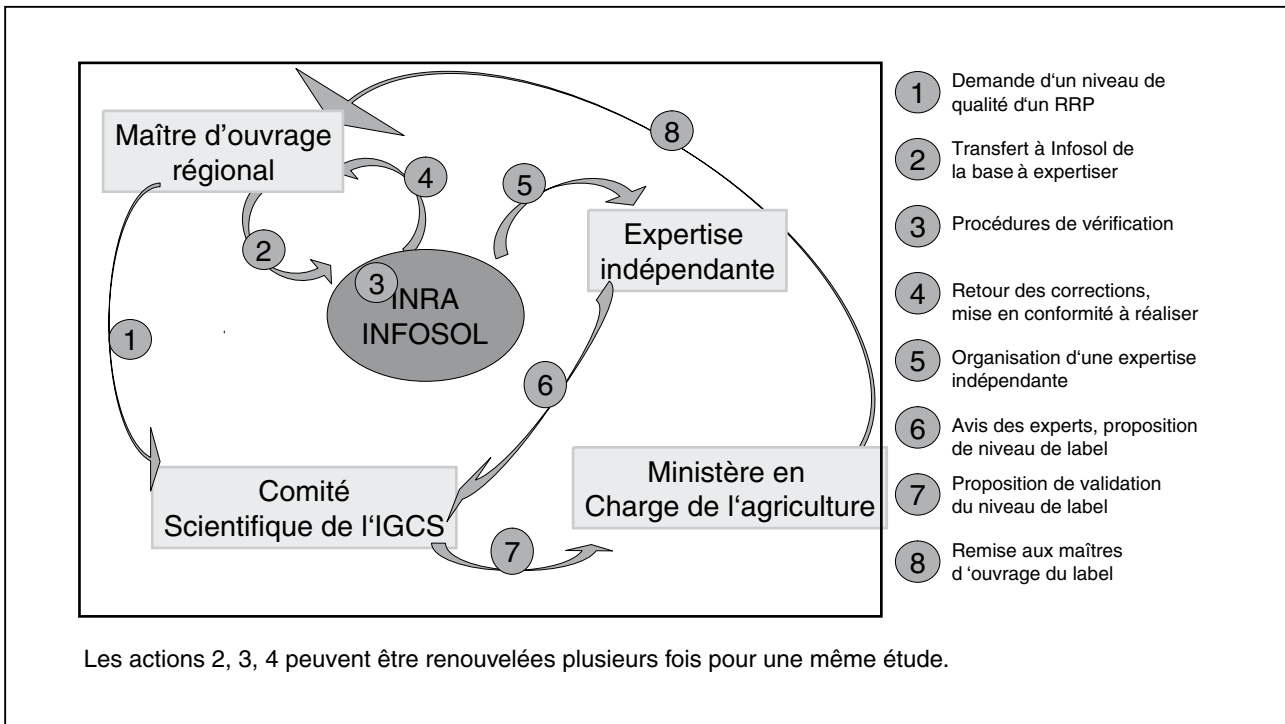
- Les objets figurés dans la base géographique sont décrits dans la base sémantique. Cette dernière est gérée au sein d'un Système de Gestion de Base de Données Relationnelle (Postgress). La spécificité de la couverture pédologique a conduit à concevoir un modèle de données original dénommé Donesol (Gaultier *et al.*, 1993 ; Grolleau *et al.*, 2004).

Cette structuration permet une manipulation aisée des informations par des logiciels d'analyses cartographiques (Système d'Information Géographique SIG). Le R.R.P. constitue ainsi l'une des bases de l'information géographique régionale. Il peut être utilisé à des fins appliquées telles que, par exemple, la planification de l'usage des terres, la protection de l'environnement, la mise en valeur agricole, les réalisations foncières ou la détermination des aires optimales pour les besoins urbains (Le Bas *et al.*, 2004 ; Le Bas et Schnebelen, 2006).

A la fin de la réalisation d'un RRP, une validation de la base de données peut être délivrée par le Ministère en charge de l'Agriculture, par l'attribution d'un niveau de qualité qui constitue une forme de certification pour les futurs utilisateurs (Infosol, 2005).

La procédure d'attribution d'un niveau de qualité

Les procédures de validation et d'attribution d'un niveau de qualité sont résumées sur la *figure 1*. Après la demande d'expertise au ministère, sa réalisation technique est confiée à l'unité Infosol ainsi qu'à deux experts indépendants. Une phase préliminaire de

Figure 1 - Les différentes étapes de la procédure de labellisation.**Figure 1** - Procedure for assessing the quality of regional maps.

vérification et de contrôle de la base de données fournie par le maître d'ouvrage est réalisée par l'unité Infosol. Cette étape permet de vérifier si le contenu est en adéquation avec le niveau de qualité visé, si le taux de remplissage des différentes tables est suffisant et si la cohérence pédologique des données est respectée.

La qualité est un terme qui peut prendre énormément de significations : dans notre cas nous considérerons qu'une base de données de qualité est « un produit dépourvu de défauts et conforme aux spécifications » (Devillers *et al.*, 2005). Dans le cas des RRP, ces spécifications sont définies par le Cahier des Clauses Techniques Générales du programme (Infosol, 2005).

Les critères d'appréciation de la qualité des données définis dans ce cahier des charges s'appuient essentiellement sur la quantité d'observations (sondages et profils, cartes préalablement existantes) effectuées pour constituer le référentiel et sur la quantité d'informations répertoriées dans les différentes tables ainsi que la cohérence des données. La détermination de chaque niveau de qualité s'apprécie précisément sur les éléments donnés ci-après :

Niveau de qualité minimum

Il correspond aux Référentiels régionaux pédologiques pour lesquels : la surface des études existantes à des échelles plus précises ($\geq 1/100\ 000$) représente moins de 10 % de la surface de la région ; les levés cartographiques sur le reste de la région ont été réalisés à un niveau de résolution minimum (c'est-à-dire un site

Une Unité Cartographique de Sol (UCS) est définie comme un morceau de la couverture pédologique. Elle a pour but de représenter l'extension géographique d'un ou de plusieurs types de sol. Elle constitue donc le regroupement d'une ou de plusieurs Unités Typologiques de Sol (U.T.S.) de façon à pouvoir en faire une représentation cartographique à une échelle donnée.

L'Unité Typologique de Sol (U.T.S.) est un volume de la couverture pédologique présentant en tous lieux de l'espace la même succession d'horizons, l'un ou l'autre de ces horizons pouvant être localement absent.

Une strate représente la variation dans l'espace d'un horizon, ou d'une couche issue d'un regroupement de plusieurs horizons. Chaque UTS est définie par la succession d'une ou plusieurs strates, et par leur organisation.

Le sondage est une description succincte du sol faite à l'aide d'une tarière

Le Profil est une coupe verticale faite dans le sol, réalisée soit à la main, soit mécaniquement. Il est décrit par une succession d'horizon caractéristique d'une UTS.

L'horizon est une couche ou un volume d'une couverture pédologique, considéré comme homogène.

décrit pour 600 ha et un site décrit et analysé pour 6 000 ha) ; la quantité et la qualité relativement faibles des données numérisées (graphiques et sémantiques) ne permettent de fournir qu'un nombre limité de cartes thématiques.

Niveau de qualité moyen

Il correspond aux RRP pour lesquels : la surface couverte par les études pédologiques préexistantes (échelles $\geq 1/100\ 000$) représente 30 à 50 % de la région ; le niveau de résolution des levés cartographiques réalisés pour le RRP est moyen (c'est-à-dire un site décrit pour 400 ha et un site décrit et analysé pour 4 000 ha) ; la caractérisation des UCS, UTS, horizons et profils est assez riche pour permettre l'obtention de cartes thématiques variées.

Niveau de qualité supérieur

Il correspond aux RRP pour lesquels : plus de 50 % de la surface de la région a déjà été cartographiée (échelle $\geq 1/100\ 000$) ; le niveau de résolution des levés réalisés sur la partie restante est très bon (c'est-à-dire un site décrit pour 200 ha et un site décrit et analysé pour 2 000 ha) ; le nombre de paramètres descriptifs cartographiques et ponctuels est très important offrant ainsi des possibilités accrues pour la production de cartes thématiques précises.

La densité des observations, dont dépendent ces niveaux de qualité, est ainsi le premier critère pris en compte puisque c'est la matière première sur laquelle s'appuie l'auteur pour renseigner les tables de données. Cependant, pour apporter une garantie de qualité aux utilisateurs, cette seule information ne suffit pas. La poursuite des contrôles se fait donc sur d'autres critères que sont, par exemple, la cohérence des données aux niveaux graphique (comparaison avec des couches références) et sémantique (comparaison des données entre elles, entre tables), et aussi leur adéquation vis-à-vis de la classification ou du référentiel utilisés.

Etat des lieux

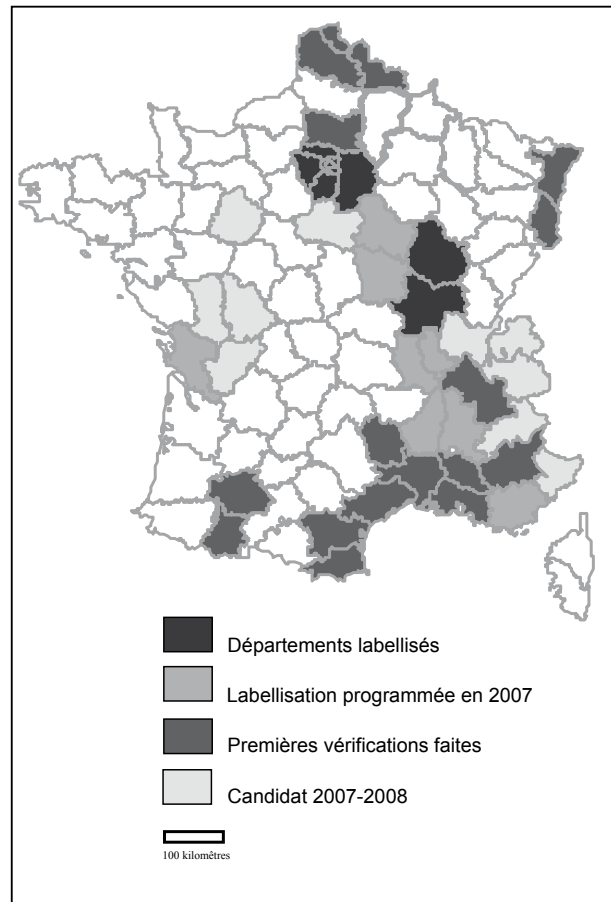
L'avancement des vérifications sur les R.R.P. est montré sur la *figure 2*.

Les régions dont le programme a été lancé le plus tôt sont essentiellement les régions Aquitaine, Bourgogne, Centre, Languedoc-Roussillon, PACA et Rhône-Alpes. Elles ont compris rapidement l'intérêt de disposer d'une information Sol capitalisée, harmonisée et disponible sous un format informatique standard.

Avec la création du GIS Sol, les régions « pionnières » ont poursuivi leurs travaux, puis de nouveaux partenaires se sont insérés dans cette dynamique de constitution des RRP. Ce cadre a d'ailleurs dépassé celui de la France métropolitaine avec la création des bases de données pour la Martinique et la Guadeloupe, en collaboration avec l'IRD. La deuxième vague concerne les régions Alsace, Bretagne, Corse, Limousin, Lorraine, Pays de Loire, Midi-Pyrénées, Nord-Pas-de-Calais, Picardie et Poitou-Charentes.

Figure 2 - Etat d'avancement des procédures de vérifications des R.R.P.

Figure 2 - State of progress of regional mapping control.



Quelques régions restent pour l'instant en marge de ce programme, ce sont l'Auvergne, la Franche-Comté, la Basse-Normandie et la Haute-Normandie. Bien qu'elles disposent de nombreuses informations sur les sols, la mise en place du projet reste souvent bloquée pour des raisons tenant sans doute à l'absence de réelle volonté politique (non implication financière des décideurs locaux). Il appartient aux responsables du programme IGCS de convaincre leurs interlocuteurs du bien fondé de la sauvegarde et de la gestion des données sols, pour le service public.

LES CONTRÔLES

L'échantillon sur lequel s'appuie cet article est constitué des R.R.P. disponibles. Ce sont essentiellement des bases issues des régions suivantes : Poitou-Charentes, Bourgogne, Rhône-Alpes, Provence Alpes Cotes d'Azur, Languedoc-Roussillon, Ile de France, Nord-Pas-de-Calais, Alsace, Midi-Pyrénées. Nous conserverons l'anonymat des études dans cet article.

La partie graphique des bases de données

Un constat sur les R.R.P.

Le premier constat est qu'il est difficile de fixer par avance un nombre moyen d'UCS adéquat pour une surface donnée (figure 3) : le nombre des UCS rencontré n'est pas proportionnel à la surface cartographiée. D'autres facteurs doivent être pris en compte pour la définition de leur nombre tels que la complexité du milieu, la perception de ce dernier par l'auteur

Les trois quarts des études présentent un nombre moyen d'UCS pour un département compris entre 60 et 110, ce qui s'inscrit

dans les préconisations de l'article 5.3.1 du Cahier des Clauses Techniques Générales (Infosol, 2005). La moyenne nationale est figurée par un trait rouge sur la figure 3 : 88 UCS par R.R.P.

La figure 4 donne une valeur moyenne de la surface de l'UCS rapportée à la surface totale. Elle est calculée par simple rapport de la surface de l'étude divisée par le nombre d'UCS. Ce graphique montre une première tendance : la surface moyenne d'une UCS croît avec la surface cartographiée lors de l'étude.

Cette corrélation s'explique en partie par le fait que plus une étude présente une surface cartographiée faible, plus les unités décrites seront découpées (par les frontières de l'étude) et maté-

Figure 3 - Nombre d'UCS pour chaque étude en fonction des surfaces cartographiées.

Figure 3 - Numbers of soil units for regional maps of soilscapes.

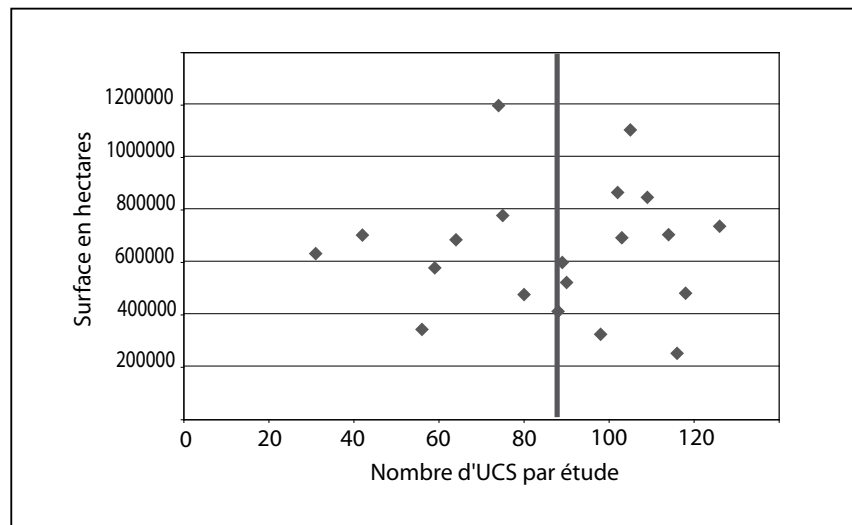


Figure 4 - Moyenne de la surface des UCS rapportée de la surface totale de l'étude.

Figure 4 - Mean of unit area reported by total surface.

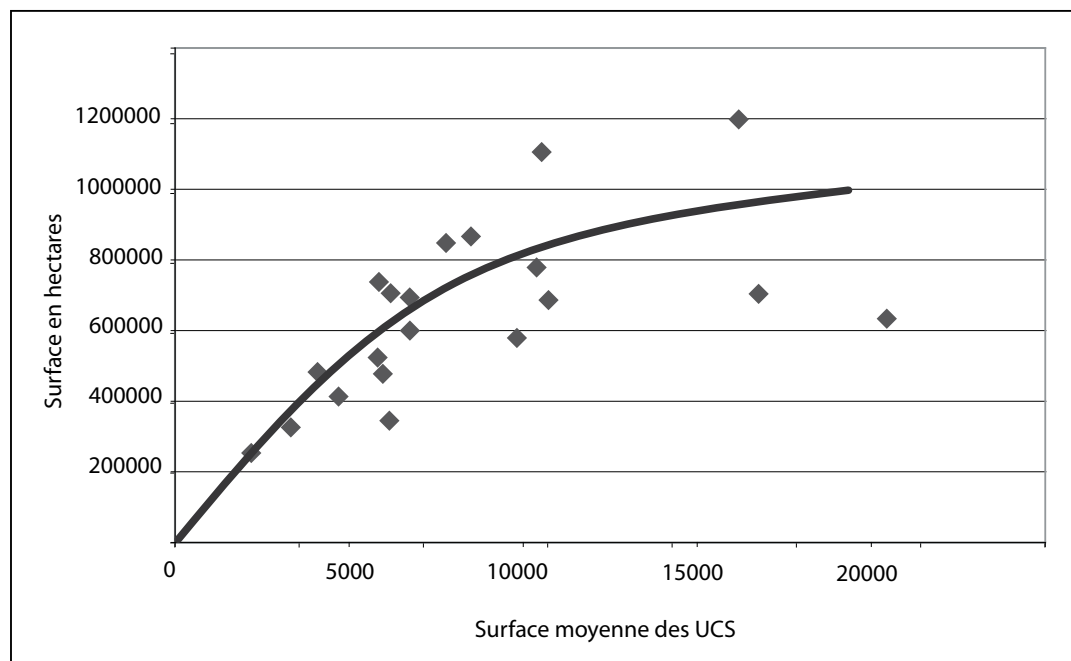
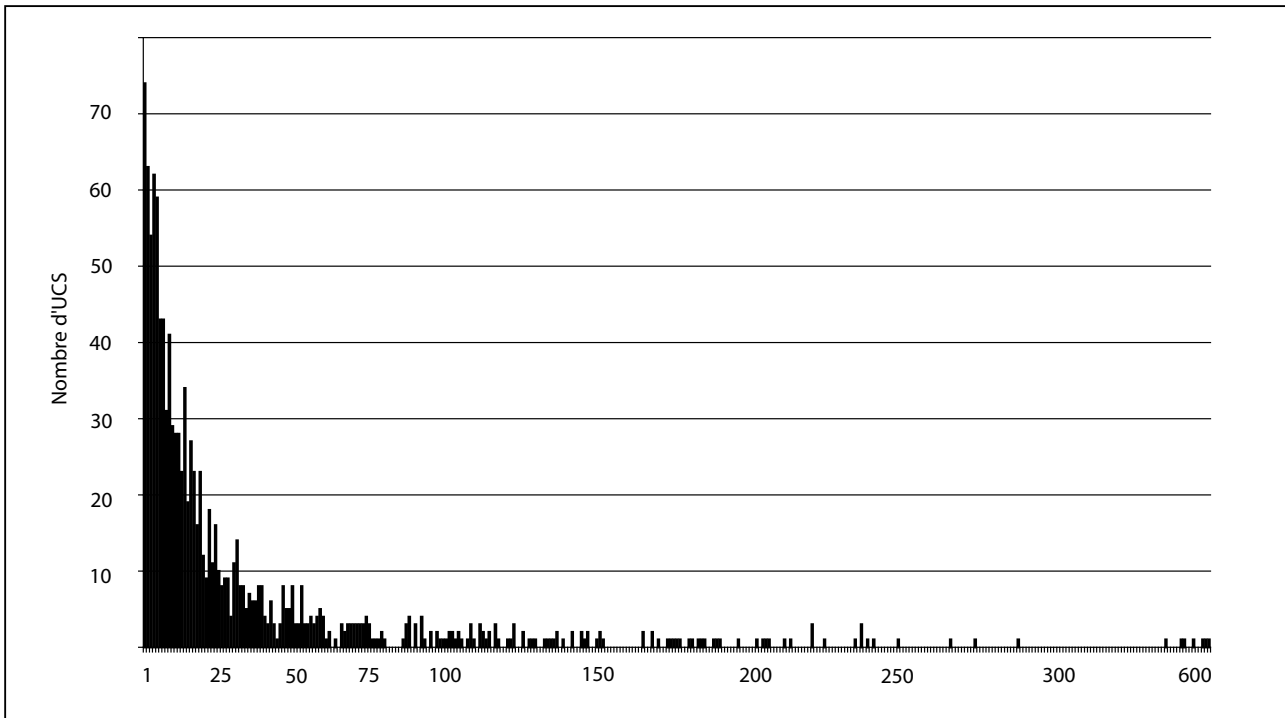


Figure 5 - Nombre de polygones par UCS.**Figure 5** - Numbers of polygons for each soil map unit.

realisent donc de faibles superficies. Les points, hors de la tendance générale, sont les études dont le nombre d'UCS est très faible par rapport à la surface cartographiée, elles présentent des unités à superficie élevée.

La surface moyenne d'une UCS pour toutes les études confondues est de 7724 ha. La plus petite, en surface, répertoriée représente 13 ha et la plus grande 178973 ha, ce qui représente une amplitude et une dispersion des valeurs très importantes.

Une réflexion importante est à mener pour normaliser la définition des pédopaysages afin qu'il y ait une certaine homogénéité dans leur détermination et représentation. Cet objectif est important puisque le but initial du programme est d'obtenir une couverture harmonisée sur l'ensemble du territoire. Le bureau européen des sols travaille sur cette problématique de mise en place d'une procédure de normalisation de la description des pédopaysages (Finke *et al.*, 2001).

Le nombre de polygones pour chaque UCS est aussi un bon indicateur sur la qualité des données (*figure 5*). Si le nombre de polygones est trop élevé, à cette échelle du 250000^e, la représentation peut apparaître trop précise par rapport au rendu souhaité, alors qu'au contraire si le nombre est faible, la représentation est peut-être trop peu précise avec un découpage peu affiné de type « patates ». D'après la *figure 5*, 50 % de l'effectif total présente un nombre de polygones par UCS inférieur à 14.

D'ailleurs sur les échantillons étudiés le nombre total de poly-

gones par étude, indépendamment de sa surface, varie de 450 à 7 500, avec une valeur moyenne qui est de 1 750 polygones.

La *figure 6* montre le nombre de polygones en fonction de la surface de l'étude. Le nombre de polygones obtenus pour les études entourées de rouge (*figure 6*) apparaît trop important par rapport à la surface couverte. Ces études présentent un découpage des unités trop précis par rapport à un rendu cartographique à 1/250 000^e. Pour le reste, le résultat est paradoxal : le nombre de polygones ne varie pas linéairement avec la surface de l'étude mais le nombre de polygone reste inférieur à 2000.

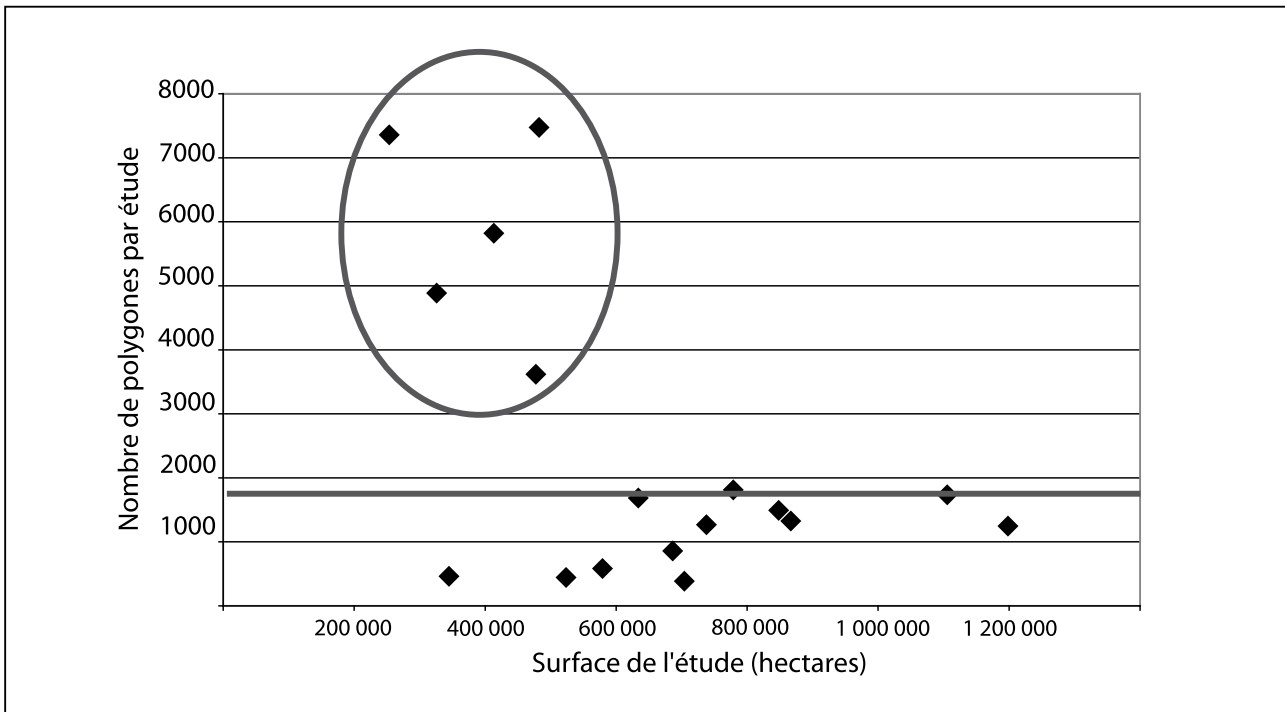
La vérification

La vérification des bases géométriques se fait en plusieurs étapes :

i) Une première appréciation visuelle du tracé est faite afin de déceler une hétérogénéité dans la représentation des pédopaysages sur toute la carte et d'examiner la compatibilité de la taille des unités avec une échelle de représentation au 1/250 000^e. La superficie et la forme des UCS doivent être compatibles avec l'échelle de représentation. On estime généralement que la plus petite surface représentable et repérable sur une carte est le 1/4 de cm² (carré de 5 mm de côté et/ou un cercle de 2,8 mm de rayon) (Boulaine, 1980). Si les contours de la plage cartographique sont pratiquement parallèles, ils doivent être au moins distants de 300 m sur le terrain. Des études ont essayé de déterminer des indices

Figure 6 - nombre de polygones en fonction de la surface de chaque étude échantillonnée (numérotées de 1 à 14).

Figure 6 - Numbers of polygons for regional maps.



mathématiques pour caractériser la forme (shape complexity index) (Legros, 1996) et la taille des polygones en fonction de l'échelle de la carte (Hengl, 2006).

L'observation statistique des valeurs des périmètres et des surfaces de tous les polygones de chaque étude est synthétisée dans la *figure 7*, elle confirme les premiers enseignements de la *figure 6*.

Cette figure montre que les études qui présentaient en *figure 6* un nombre très important de polygones, sont celles dont les caractéristiques statistiques (médiane...) présentent des valeurs les plus faibles en terme de superficie et de périmètre par rapport aux études références, déjà labellisées (5 et 13). Ce sont les études 1, 6, 9 et 11. Cette observation confirme que ces études présentent une restitution cartographique trop détaillée par rapport à l'échelle demandée. Inversement les études 3 et 14 correspondent à un découpage plutôt grossier.

Ces observations sont confrontées aux avis des auteurs qui peuvent alors justifier leurs choix lors des journées d'expertise (milieu naturel peu ou très complexe, zone de haute montagne...).

ii) Les vérifications des tracés géométriques des Unités Cartographiques de Sol se font sous SIG et par superposition à des couches références de type IGN. Les principales erreurs apparues sont marquées par des décalages par rapport à ces couches références avec une tolérance de 250 m terrain. Dans le futur il est envisagé de faire n mesures et de faire une moyenne M de ces

mesures m et d'estimer ainsi une dispersion S qui sera exprimée par la valeur de l'écart type (Devilleurs *et al.*, 2005), autrement dit la racine de la somme des carrés des écarts de chaque mesure à la mesure moyenne, divisée par le nombre de mesures réalisées. On aura alors ;

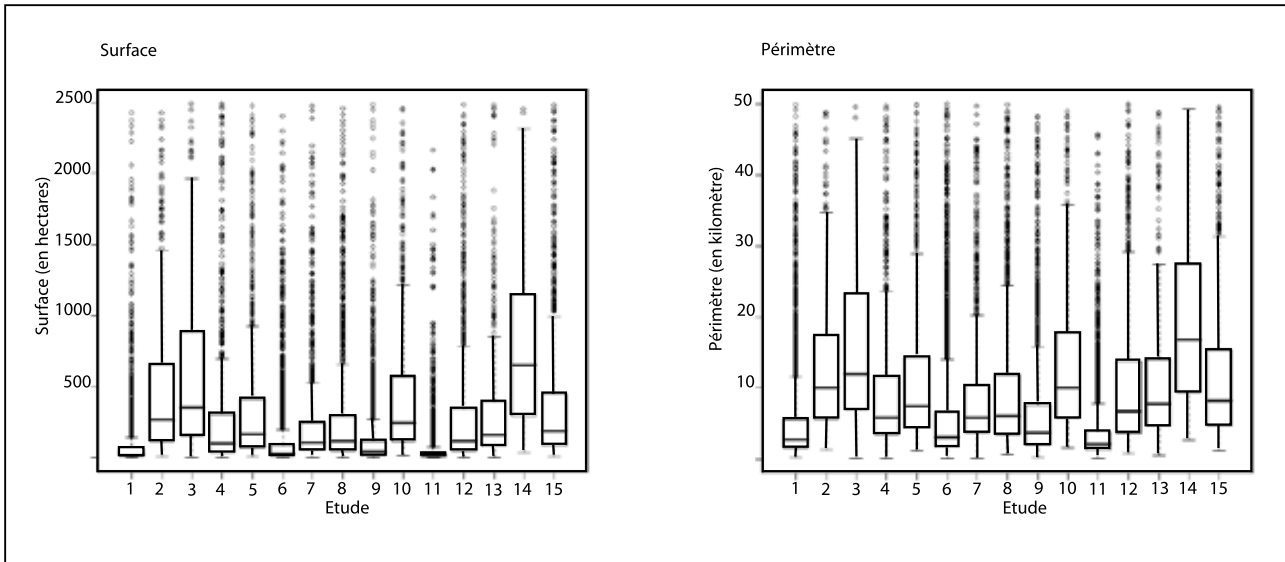
$$\text{Dispersion } S = \text{racine (somme } (M-m)^2/n)$$

Ce calcul permettra de définir statistiquement le décalage de la couche géométrique du R.R.P. et des couches références. Un seuil de dispersion maximale à ne pas dépasser pourra être alors défini.

Une autre estimation de l'écart de position entre deux lignes (la ligne produite et la ligne référence) est la distance de Hausdorff (Barillot, 2001). Une valeur d'estimation nulle de cette distance signifie que les ensembles sont superposés. Ce calcul de distance peut aussi s'appliquer sur des points tels que les profils.

Les erreurs détectées au cours des vérifications sont principalement des décalages géographiques avec par exemple des cours d'eau qui se retrouvent sur les versants au lieu de se situer dans un pédopaysage de plaine alluviale.

iii) On observe aussi de nombreuses erreurs topologiques issues pour la plupart d'une mauvaise utilisation du SIG ou tout simplement des performances limitées du logiciel utilisé pour le tracé des polygones. Une topologie défaillante peut créer des polygones non fermés, des vides entre les polygones, des micro polygones non affectés, des superpositions de polygones. Tout ceci produit une

Figure 7 - Statistiques sur le périmètre et la surface des polygones de chaque étude.**Figure 7** - Statistics of polygons properties for regional maps.

base géométrique localement fautive, difficile à utiliser, et pour le moins perturbée dans son exploitation optimale.

La base sémantique

Le constat général

Le taux de remplissage de la base de données

Le premier diagnostic de la base de données peut se faire sur le taux de remplissage de chaque table sous la forme de statistiques exprimées dans le *tableau 1*.

Globalement, les moyennes des taux de remplissage (*tableau 1*) de chaque table des Référentiels Régionaux Pédologiques sont similaires aux moyennes des autres études numérisées (100 000^e et secteurs de référence essentiellement).

Les taux de remplissage peuvent apparaître faibles pour des gestionnaires de bases de données, mais il est important d'admettre que la donnée traitée est une donnée complexe, non mathématique, dépendante du milieu naturel et de son hétérogénéité. Et n'oublions pas qu'il est prévu de stocker des informations qui, dans un cas donné, ne peuvent être fournies, par exemple, tout ce qui concerne le carbonate de calcium dans un sol acide. En d'autres termes, il est inconcevable que le taux de remplissage de la base soit de 100 %. En plus, certaines informations sont d'un accès difficile, par exemple, la valeur modale de profondeur du sol sur une portion de la couverture. Le manque de temps ou le coût trop élevé d'acquisition peuvent conduire à négliger un certains nombres de paramètres.

Les taux de remplissage assez moyen de certaines tables peuvent s'expliquer aussi par le fait que les études les plus anciennes

ont été saisies en format Donesol1 qui comportait pour certaines tables un nombre de champs moins important (U_sol, Strate).

Le faible taux de remplissage de la table ANALYSES est liée au fait qu'elle est utilisée par d'autres programmes du GIS Sol (par exemple le RMQS). Ces derniers ont nécessité l'ajout d'un grand nombre de champs dont le renseignement n'est pas obligatoire dans le cadre du RRP. Certaines analyses coûteuses sont alors absentes pour bon nombre d'horizons.

Les données surfaciques

Les données surfaciques (UCS, UTS et strates) sont les informations relatives à une portion de la couverture pédologique ; elles sont issues de la synthèse des observations ponctuelles que sont les sondages et les profils, repérables par leurs coordonnées X et Y.

Le nombre moyen d'UCS par étude est de 88. Le R.R.P. étant une représentation à 1/250 000^e, la majorité des UCS sont des unités complexes c'est-à-dire contenant plusieurs UTS. Pour l'échantillon étudié on obtient une moyenne de 2,83 UTS par UCS. La *figure 8* montre la répartition du nombre d'UTS au sein des UCS.

Il est théoriquement possible de mettre jusqu'à 99 UTS par UCS, mais le R.R.P. doit donner une vision synthétique des sols. De plus, il est difficile de concevoir l'introduction dans des modèles agronomiques ou environnementaux d'une multiplicité de sols différents, l'agrégation étant d'autant plus complexe techniquement que le nombre d'UTS est élevé. Il est conseillé dans le cahier des clauses techniques générales de limiter le nombre d'UTS à 10 par UCS. Certaines variantes pédologiques ne nécessitent pas toujours la création d'une nouvelle UTS, leurs caractéristiques peuvent parfois être intégrées dans la description des strates en élargissant les

Tableau 1 - Taux de renseignement de chaque table pour les RRP (chiffre du haut) et pour toutes les études de Donesol (chiffre du bas).
Table 1 - Statistics of soil databases for regional maps and the other studies.

Table	Nombre de champs	Moyenne (taux de remplissage)	Médiane (taux de remplissage)
U_carto	13	0,69 (0.62 Donesol)	0,73 (0.67 Donesol)
Contient	29	0,53 (0.46 Donesol)	0,67 (0.44 Donesol)
U_sol	74	0,31 (0.33 Donesol)	0,39 (0.27 Donesol)
Strate	7	0,47 (variables obligatoires)	0,4 (variables obligatoires)
Affect_strate_u_sol	14	0,74	
Profil	84	0,26 (0.33 Donesol)	0,26 (0.27 Donesol)
Horizon	113	0,13 (0.23 Donesol)	0,11 (0.17 Donesol)
Analyses	224	0,11 (0.48 obligatoire) (0.14 Donesol)	0,04 (0.47 obligatoire) (0.08 Donesol)
Affect_prof il_etude	6	0,35 (0.33 Donesol)	0,34 (0.33 Donesol)

Figure 8 - Nombre d'UTS par UCS.

Figure 8 - Numbers of Soil Typological Unit for Soil Map Unit.

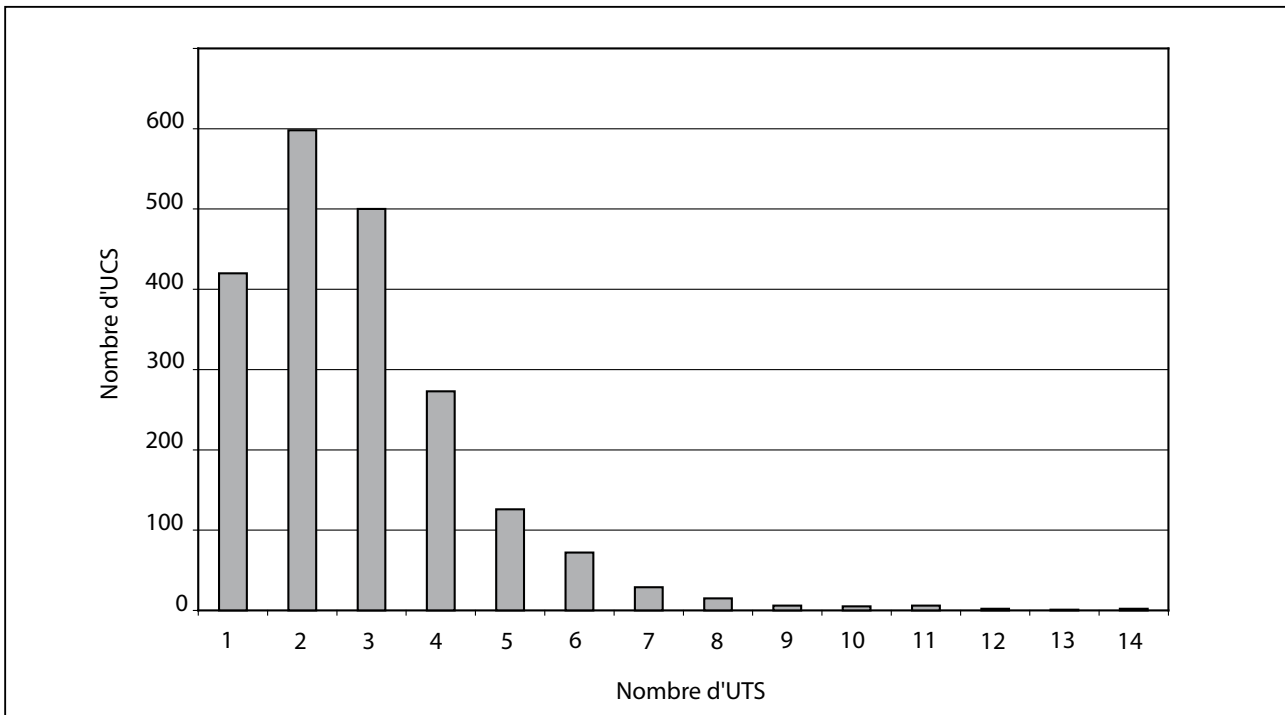


Tableau 2 - Extrait de la table AFFECT_PROFIL_ETUDE du dictionnaire de Donesol (Infosol, 2006).**Table 2** - Example of DoneSol dictionary.

Nom du champ	Intitulé	Type	Taille	Règles d'intégrité
NO_PROFIL	numéro du profil donné par l'auteur dans l'étude	texte	10	champ obligatoire
NO_UC	numéro de l'Unité Cartographique de Sol dans laquelle se situe le profil	entier	5	
NO_US	numéro de l'Unité Typologique de Sol à laquelle appartient le profil	entier	3	
REP_PROF	représentativité du profil par rapport à l'Unité Typologique de Sol	texte	1	

Tableau 3 - Exemple du renseignement du champ TYPE_REGION de la table U_carto.**Table 3** - Example of file TYPE_REGION.

Code	Signification
1	Montagnes et reliefs majeurs (dénivelées supérieures à 300 m, pentes supérieures à 30%)
2	Collines et reliefs mineurs (dénivelées inférieures à 300 m, pentes inférieures à 30%)
3	Glacis et piedmonts (pentes faibles au pied des reliefs, disséquées ou non)
4	Plateaux et hautes terres (dénivelées inférieures à 50 m, pentes de 0 à 8%, exceptionnellement jusqu'à 16%)
5	Plaines et terrasses (disséquées ou non, de toutes origines)
6	Petites vallées (replat alluvial dont la largeur est inférieure à 500 m)
7	Thalwegs (incluant les systèmes alluviaux ou colluviaux dont la largeur ne dépasse pas 200 m de large)
8	Zones dépressives (telles que marais, tourbières)

valeurs extrêmes correspondant à ces dernières. On observe que seulement 0,008 % des UCS dépassent le seuil de 10 UTS.

En moyenne, chaque UTS est décrite par **2,88 strates** avec une amplitude des valeurs moyennes allant de 2,11 à 3,9 strates/UTS/étude.

Le nombre minimum ou maximum de strates par U.T.S. n'est pas imposé. Cependant, pour chaque U.T.S., les différentes strates servant à différencier pédologiquement le type de sol doivent être présentes. Par exemple, dans le cas d'une U.T.S. de type LUVISOL, doivent être présentes au minimum les strates correspondant aux horizons A, E, et BT qui servent à différencier ce type de sol. Les strates d'une UTS ne sont pas cependant l'équivalent strict des horizons d'un profil type.

Pour la caractérisation des strates, différentes situations peuvent se présenter (Infosol, 2005) :

- Il n'y a pas de profil type associé à l'UTS considérée. Dans ce cas, les strates sont caractérisées à partir de données décrites lors de la prospection terrain par sondages à la tarière. Certaines données ne sont pas définissables par sondage ; seul un petit nombre de variables qualitatives et quantitatives, caractéristiques des strates, pourra donc être renseigné (texture, couleur, etc).

- Un seul profil type est associé à l'UTS. Dans ce cas, les données descriptives et analytiques du profil sont reprises au niveau des données des strates. Elles ont le statut de mode principal ou

de valeurs modales. Elles sont établies en utilisant les données de prospection par sondage à la tarière. Bien sûr cela ne concerne que les données accessibles directement sur le terrain.

- De nombreux profils sont associés à l'UTS. Dans ce cas, une étude statistique par horizon (ou groupe d'horizons) et par caractère peut être réalisée à partir de l'ensemble de données descriptives et analytiques issues des profils et des sondages à la tarière. Cette étude permet de renseigner les variables qualitatives et quantitatives correspondantes des strates pour rendre compte de leur variabilité au sein de l'UTS.

La qualité et la quantité du remplissage de la table strate sont des indicateurs précieux pour l'attribution du niveau de qualité, puisque une grande partie des thématiques dérivées de la base de données devra s'appuyer sur ces données synthétiques et sur leur variabilité.

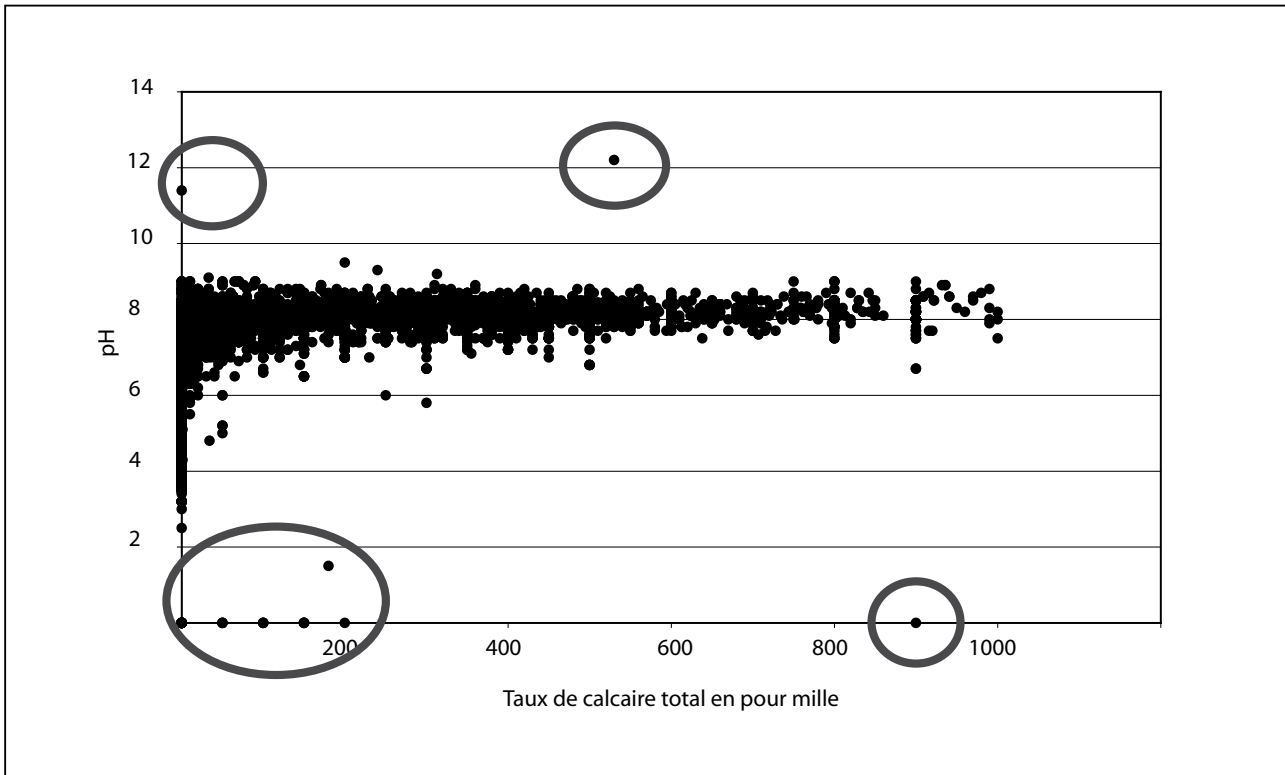
Les profils

Les sondages et les profils sont les sources principales d'informations, essentielles pour la constitution des R.R.P. C'est bien entendu un des critères de l'attribution du niveau de qualité qui s'appuie sur leur présence complète dans la base de données (table PROFIL, HORIZON et ANALYSES).

Une tendance très nette apparaît actuellement dans la constitution des R.R.P. vis à vis de la gestion des données ponctuelles : les

Figure 9 - Cohérence entre le pH et les valeurs de Calcaire total.

Figure 9 - pH and calcareous rate relationship.



sondages sont de plus en plus saisis soit directement sur le terrain soit *a posteriori* via l'interface Web. Il est important de capitaliser et de conserver ces données, car d'une part, les informations sur des supports papier disparaissent souvent lors du départ de l'auteur, et d'autre part, leur exploitation par des méthodes statistique, géostatistique, ou pour la modélisation de l'organisation des sols est en plein essor.

Principaux contrôles

Contrôle direct

Le respect de la structure de Donesol est indispensable : la base fournie par le maître d'ouvrage doit comporter l'ensemble des tables de Donesol, ainsi que tous les champs associés en conformité avec le dictionnaire. C'est-à-dire que le champ doit correspondre à la façon dont il est déclaré par rapport à son intitulé, son type, sa taille et sa spécificité (règles d'intégrité) s'il en présente une (tableau 2).

La codification des champs doit elle aussi être respectée, par son type et par les choix possibles qui sont proposés. Un exemple est présenté dans le tableau 3 pour le champ TYPE_REGION.

Ce champ TYPE_REGION (tableau 3) ne peut être intégré dans la base que sous la forme d'un code de 1 à 8. Une autre valeur sera synonyme d'erreur.

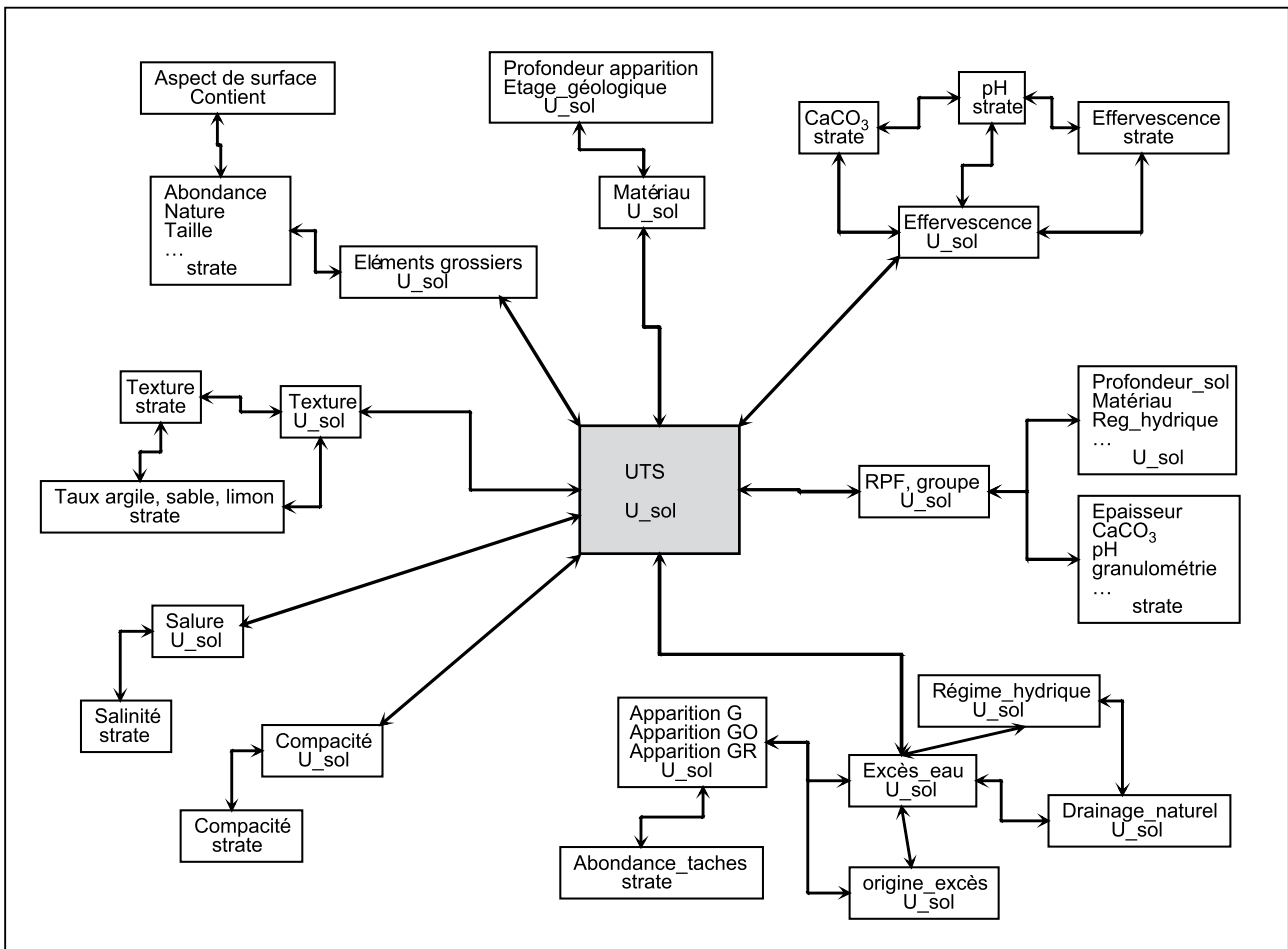
L'utilisation de l'interface Web Donesol (<http://donesol.gissol.fr/>) réduit considérablement les erreurs sur la structure et sur la codification de l'information.

De même qu'une UCS doit comporter une à plusieurs UTS qui elles-mêmes doivent être décrites par une ou plusieurs strates, un profil doit être décrit par des horizons, et des analyses (si ce n'est pas un sondage).

Contrôles indirects ou de vraisemblance

Certaines variables sont exprimées de différentes manières dans plusieurs champs de la base de données, ce qui permet de vérifier leur cohérence. Cette vérification peut se faire entre champs d'une même table ou entre deux tables. C'est le cas par exemple de la matière organique qui est renseignée quantitativement et qualitativement dans les horizons et dans les strates. Les valeurs doivent être cohérentes. De la même façon, les champs décrivant l'état hydrique du profil doivent être cohérents entre eux (présences des taches, des profondeurs d'apparition des horizons g, GO, GR, ...).

Les contrôles peuvent s'effectuer sur des variables entre lesquelles existe une forte liaison, par exemple le pH et les valeurs de calcaire (figure 9). Certaines valeurs encadrées sont manifestement fausses ($\text{pH} = 0$; $\text{pH} > 12$) ; d'autres valeurs s'écartent du nuage de points et semblent incohérentes à la vue du graphique. Leur authen-

Figure 10 - Schéma général des relations entre les champs des différentes tables.**Figure 10** - Schematic representation of the relations between items.

ticité sera vérifiée (comment ont-elle été obtenues, avec quelle méthode, est-ce une erreur de saisie, un artefact, une réalité ?) et elles seront validées, corrigées ou supprimées.

D'autres tests sont faits entre deux tables ; ainsi par exemple, une information décrite dans le profil doit être cohérente avec la même information reprise au niveau de l'horizon : si un profil est décrit comme non effervescent dans la table PROFIL, aucun des horizons ne doit présenter une effervescence différente de 0 dans la table HORIZON.

La *figure 8* permet de modéliser, pour l'UTS, les relations entre des groupes de variables sur lesquels des tests sont effectués afin de vérifier la pertinence des informations. Les variables renseignées au niveau des profils subissent les mêmes tests entre elles, et elles sont également comparées avec les données de l'UTS à laquelle le profil est rattaché.

Contrôle par rapport à la classification

La description du sol permet de le classer suivant différents référentiels (RP, CPCS, WRB). La Référence est renseignée dans la base, ce qui permet d'en vérifier la cohérence avec certains critères pédologiques présents dans la base.

Par exemple si un sol est qualifié de RENDOSOL (AFES, 1995), il doit répondre à certains critères : faible profondeur, présence ou non d'horizon diagnostique (Aca, absence de Sca), taux de calcaire total supérieur à 5 %.

Tableau 4 - Pourcentage moyen de chaque type d'erreur détectée.**Table 4** - *Pourcent of every type of error.*

Type d'erreurs détectées	Pourcentage
Problème de structure /Donesol	2 %
Erreurs sur les champs (codes erronés, incohérence par rapport au dictionnaires, ..)	13 %
Incompréhension du champ	9 %
Incohérence du renseignement des champs au sein d'une même table	11 %
Incohérence du renseignement des champs entre deux tables	46 %
Incohérence des renseignements par rapport à la classification choisie (RP, CPCS)	19 %

PREMIERS ENSEIGNEMENTS DES R.R.P. PRODUITS

Sur le plan graphique

La première question à résoudre est la définition des pédopaysages et la standardisation de leur mode de cartographie, tant les délimitations sont différentes d'une étude à l'autre. Un gros et complexe travail d'harmonisation est à mener puisque pour des départements de superficie similaire, le nombre de pédopaysages peut varier de 30 à 165. L'harmonisation inter-régionale sera difficile car la définition des pédopaysages reste dépendante de la perception du milieu par l'auteur et la délimitation de ces entités présente un caractère subjectif. Seule la standardisation de la rédaction du nom de l'UCS est actuellement imposée, ce qui sert à nommer les unités cartographiques par des termes équivalents.

Une vraie réflexion reste donc à mener en ce qui concerne l'objectivation de la délimitation des pédopaysages. Une des solutions, en amont de la prospection de terrain, pourrait être l'emploi de méthodes numériques, indépendantes de toute perception personnelle. Elles pourraient permettre de définir une première ébauche cartographique constituant un socle commun sur lequel viendrait se confirmer ou s'infirmer les hypothèses de répartition des sols. Ce premier découpage servirait aux auteurs à se caler, et ainsi de conserver une cohérence territoriale. De nombreux travaux ont été publiés sur ce thème au niveau national comme au niveau international (Dobos, 2005), (Irvin *et al.*, 1997), (McBratney *et al.*, 2003), (Robbez Masson, 1994).

L'observation des contours graphiques, au sein d'une même étude, pose le problème de l'homogénéité d'un Référentiel constitué sur un territoire où coexistent des zones à forte densité d'informations pédologiques (nombreuses études antérieures à grandes échelles) et des zones initialement plus pauvres en information. La définition et la précision du tracé des pédopaysages qui en résultent sont fréquemment différentes suivant la zone où se situent les unités. De plus, la différence n'est pas uniquement marquée dans

la précision du découpage des unités mais aussi dans le taux de renseignement de la base de données.

A l'opposé, il est aussi parfois difficile pour l'auteur de réaliser une synthèse sur des zones déjà cartographiées à plus grande échelle. Le passage à une cartographie à petite échelle peut être ressenti comme une perte d'informations. Pourtant, dans la majorité des cas, certaines UTS décrites à grande échelle ne sont que des variantes les unes par rapport aux autres ; lors de la synthèse elles sont regroupées en grande famille de sols et leur variabilité est alors exprimée par des fourchettes de valeurs plus étendues. La synthèse à petite échelle n'est donc pas une perte totale d'informations mais constitue une représentation synthétique de l'information pédologique compatible avec des utilisations à l'échelle d'un département ou d'une région.

Les problèmes liés à la topologie des couches graphiques s'estompent progressivement par la formation de plus en plus poussée des opérateurs de saisie, voire des auteurs eux-mêmes. Les performances croissantes des SIG limitent également ce type d'erreurs.

L'examen actuel des bases graphiques est exclusivement « informatique », la possibilité d'une vérification *in situ* sur le terrain serait un excellent moyen de contrôle et d'appréciation du découpage des pédopaysages fournis par l'auteur. Mais cette validation reste très marginale par le fait essentiel qu'elle est très consommatrice de temps et de crédits. Le choix d'Infosol a été de privilégier un accompagnement, dès le démarrage du projet, par des réunions techniques combinées à des sorties de terrain.

Sur le plan sémantique

Les absences, les erreurs ou les incohérences attachées aux données sémantiques sont en théorie plus faciles à détecter puisque le renseignement des champs s'appuie sur le dictionnaire Donesol, sur le cahier des charges et sur des classifications de référence (CPCS, RP).

La majorité des erreurs sont des erreurs de saisie, qui restent fréquemment difficiles à déceler sans un recours à l'expertise pédologique.

Les contraintes à la saisie sont de plus en plus intégrées dans l'interface de saisie soit lors du renseignement des champs, soit lors de la validation de l'enregistrement. C'est-à-dire qu'au niveau du champ, une vérification est opérée sur l'intitulé du champ, sur ses attributs afin qu'il soit conforme aux spécifications de Donesol. L'enregistrement de l'ensemble des champs d'une table se fait en bloc, si un de ces enregistrements est erroné, un message d'erreur apparaît.

Ce contrôle est renforcé par l'accès depuis 2006 à une interface de saisie unique et centralisée à l'unité Infosol de l'INRA d'Orléans. Cet outil remplace DonesolPC qui était déployé dans les différentes régions avec des difficultés liées à une structure figée, à des mises à jour complexes et à des modifications subies par un produit non verrouillé.

L'absence du renseignement de nombreux champs de la base de données a plusieurs causes possibles :

- l'observation n'a pas été ou n'a pas pu être effectuée, ce qui est surtout le cas lors de la reprise d'anciennes études,
- la valeur du champ est systématiquement nulle, donc par souci d'économie de temps, l'opérateur néglige de le signaler ; ce sont par exemple l'effervescence en milieu acide : pourquoi mettre une valeur puisque tous les sols sont acides et donc caractérisés par une effervescence nulle ? Il est pourtant essentiel d'intégrer le fait que les valeurs 0 et « pas d'information » n'ont pas la même signification,
- il s'agit simplement d'un oubli de saisie de la part de l'opérateur,
- certains champs ne sont parfois pas remplis délibérément : « puisque de tout façon je n'utiliserai pas cette information, je ne vais pas perdre de temps à la stocker ». Cette dernière catégorie de manque reste marginale.

L'utilisation de données anciennes est indispensable pour la capitalisation de l'information pédologique et dans un souci d'économie (ne pas refaire ce qui a déjà été fait). Mais le fait que la masse d'informations à saisir peut freiner plus d'un opérateur, ces informations restent souvent partielles, car elles ont été souvent recueillies pour des problématiques précises et non dans un souci d'inventaire général. Leur interprétation et leur exploitation restent complexes en raison des méthodes de cartographie qui n'ont pas toujours été bien caractérisées et à cause de l'évolution des méthodes analytiques qui ne sont également pas toujours parfaitement renseignées.

Le principal problème rencontré est le détournement du concept de pédopaysage et le remplissage de la base de données au profit d'une problématique particulière. Un exemple est la constitution de BD orientée dans une perspective uniquement agricole : la délimitation des unités et la typologie des sols sont fondées sur les critères de comportement agronomique, les analyses sont effectuées principalement sur les horizons labourés, etc.

Le Référentiel est - et doit rester - un inventaire des sols à vocation générique, dont la constitution est indépendante des applications thématiques envisagées. Les thématiques

d'aujourd'hui ne sont pas forcément celles de demain, et il serait alors impensable de refaire ou de refinancer une étude par le seul fait de l'absence d'informations qui paraissent à l'époque insignifiantes ou inutiles.

L'autre forme rencontrée de « détournement » est l'adaptation personnelle de la structure de la base de données par la simplification des tables, des champs, voir des codes eux-mêmes, afin de rendre son usage plus souple. Cette simplification est souvent réalisée avec l'objectif de rendre la base plus conviviale et mieux adaptée à des besoins immédiats. Le remplissage des tables est parfois partiel « puisque de tout façon je n'utiliserai pas ces champs ».

Le premier problème posé par ces systèmes parallèles est leur inadéquation au cahier des charges fixé par le programme IGCS : ils ne sont pas conformes et ne pourront ni être intégrés dans la base nationale, ni bénéficier des programmes de vérification et des procédures de labellisation. En outre, ces systèmes ne pourront bénéficier des avancées en terme de développement de thématiques et de modélisation puisque leur structure n'est pas compatible pour recevoir les modèles développés selon le standard de Donesol (détermination de la RU, des fonctions de pédotransferts...).

Les erreurs dans l'interprétation des observations pédologiques (par exemple pour attribuer un nom RP à un sol) sont difficilement détectables, à moins d'être relativement grossières. En plus, il faut souligner le fait que la constitution de ces bases est généralement confiée à des personnes compétentes, bien rodées à cet exercice. La certification du métier de pédologue pourrait être un gage supplémentaire de garantie.

En revanche, de nombreuses erreurs sont décelées par les tests de corrélation entre les champs d'une même table et entre tables (*tableau 4*). Les défauts augmentent quand la saisie a été faite table par table (saisie de toutes les UTS dans U_sol, puis toutes les strates, ...). En revanche, lorsque la saisie est réalisée séquentiellement (saisie d'une UTS, puis de ses strates, puis des profils associés,...) les erreurs de cohérences sont moins nombreuses.

Les incohérences portent principalement sur les indicateurs du régime hydrique des sols, la charge en éléments grossiers (correspondance U_sol - strates), l'adéquation des caractéristiques pédologiques au type de sol défini (CALCOSOL avec un taux de calcaire total égal à 0).

L'évaluation de la pertinence des données synthétiques ou surfaciques c'est-à-dire des strates, des UTS et des UCS, reste complexe du fait de l'absence, dans la majorité des cas, des données des sondages. Bien qu'utilisés implicitement par l'auteur pour renseigner la base et caractériser la variabilité des unités, les sondages ne sont généralement pas disponibles pour vérifier les informations, ce qui est gênant. Toutefois, de plus en plus de partenaires saisissent à présent les sondages par souci de conservation (centralisation et informatisation de la totalité de l'information acquise) et dans la prévision future d'une exploitation avec l'aide de méthodes statistiques ou géostatistiques.

La mauvaise interprétation peut aussi prendre la forme d'une incompréhension non pas de la donnée elle-même mais de la façon dont elle doit être renseignée. L'expérience a montré que la définition de certains champs de Donesol restait peu claire. Pour pallier ce problème, l'unité Infosol a mis en place des formations à Donesol (80 personnes formées à ce jour) et a réactualisé le dictionnaire de données en 2005 afin d'éclairer la manière de renseigner certains champs.

Un nouvel outil est mis en place depuis quelques mois sur le site du GIS Sol. Cette liste de discussion s'adresse à tous les utilisateurs de Donesol, elle est accessible, après abonnement, à l'adresse donesol-user@listes.inra.fr. Ce service supplémentaire permet aux usagers d'exposer leurs questions sur Donesol (demande d'explications complémentaires sur la structure, sur le mode de renseignement des champs, ...) et sur l'interface DonesolWeb (fonctionnalités actuelles, améliorations possibles, bugs rencontrés...). Elle permet aussi à l'administrateur ou responsable de programme de communiquer des informations générales par cette liste, par exemple un problème temporaire d'accès à DonesolWeb ou un changement dans la structure (exemple : le nombre de caractères du numéro d'horizon est passé de 2 à 3...). Elle sert enfin à annoncer la création de nouveaux services (possibilité d'extraction automatique d'une étude...).

Les principales demandes, archivées sur le site, intéressent la table analyses : correspondances entre unités, évolution de certaines méthodes analytiques (dénomination, protocole). Des questions fréquentes concernent aussi la transformation des coordonnées géographiques dans le système Lambert II étendu, la manière de renseigner certains champs et les problèmes de classification (exemple : humus).

CONCLUSIONS ET PERSPECTIVES

Donesol est issu d'une réflexion de plus de 30 ans qui se poursuit aujourd'hui. Sa structure multi-échelle permet de conserver, d'harmoniser et de fédérer l'information pédologique. Cette structure relativement complexe aux dires des administrateurs, a été développée dans un souci premier de stockage de l'information. De manière plus générale, la méthodologie de constitution des Référentiels a été conçue dans un souci de généricité, sans connaissance précise des exploitations futures de la base de données et sans anticipation possible des évolutions technologiques. Les utilisations actuelles donnent raison à cette démarche générique, puisque les données issues de Donesol sont aussi bien utilisées en agronomie, environnement, génie civil, pour la gestion de la faune... L'analyse des premiers Référentiels produits nous apporte quelques premiers enseignements pour la poursuite de ce programme :

La perception des pédopaysages par l'auteur reste un élément important par l'expertise qu'il apporte mais il est aussi la source d'hétérogénéité puisqu'il introduit de la subjectivité. La définition approximative des pédopaysages implique de gros efforts de

normalisation afin que les objectifs d'une couverture nationale normalisée soient atteints. Le rendu cartographique doit rester cohérent avec la réalité du pédopaysage tout en étant compatible avec l'échelle de représentation graphique (taille et forme des polygones).

Au niveau sémantique, plus de la moitié des erreurs sont des incohérences entre les champs d'une même table ou de différentes tables de la base. L'inadéquation entre le type de sol proposé et ses constituants pédologiques est une autre incohérence possible. Les atteintes à la structure de Donesol deviennent minoritaires par le verrouillage progressif du système et par l'intérêt grandissant des partenaires pour mutualiser des outils (interrogation, utilisation, diffusion) de la base, implémentés sur une structure Donesol unique et non pas sur des structures parallèles.

Cet état des lieux nous procure quelques enseignements et des références permettant de situer une étude par rapport à celles déjà réalisées.

Pour réduire les erreurs et améliorer la qualité de l'information stockée, Infosol a renforcé les procédures de suivi, de vérifications, a mis en place des formations, des listes de diffusion, une interface Web de saisie et de consultation. Cet accompagnement technique de l'unité est conforté par une proposition de norme AFNOR sur le stockage et la validation des données Sol. Elle est actuellement en cours de validation.

Comme par définition une base de donnée est évolutive, Infosol a démarré en 2007 une réflexion sur la structure de Donesol2. Le passage vers un Donesol3 se fera par une démarche à laquelle participeront tous les partenaires, producteurs de la donnée ou utilisateurs. Cette refonte intégrera des améliorations (nouveaux champs, intégration des métadonnées...) tout en tenant compte de l'acquis. Le développement de cette base doit tendre à une plus grande ouverture vis-à-vis des usagers, tout en ne perdant pas son objectif premier d'inventaire, conservation et enrichissement de la donnée pédologique.

Surface moyenne d'un département	600 000 ha
Nombre d'UCS moyen	88
Surface moyenne d'une UCS	7 724 ha
50 % des UCS ont un nombre de polygones inférieur à 14	
Nombre moyen de polygones par étude	1750
Nombre moyen d'UTS par UCS	2,83
Nombre moyen de strates par UTS	2,88
46 % des erreurs dans la base de données sont des incohérences	
19 % sont des erreurs de classification	

REMERCIEMENTS

Nous remercions Dominique Arrouays pour ses conseils et ses encouragements et pour sa relecture des premières versions de cette note. Nous remercions également Micheline Eimberck et Jean-Paul Legros pour leur relecture critique très constructive, l'ensemble des partenaires régionaux auteurs des Référentiels Régionaux Pédologiques et les nombreux organismes qui financent ces programmes.

BIBLIOGRAPHIE

- AFES, 1995 - Référentiel Pédologique, INRA édition.
- Bishop *et al.*, 2001 - Measuring the quality of digital soil maps using information criteria. *Geoderma* 103, pp. 95-111.
- Boulaine J., 1980 - Pédologie Appliquée, édition Masson, p. 220.
- Barillot X., *et al.*, 2001 - Généralisation de l'algorithme de Douglas et Peucker pour des applications cartographiques, *Bulletin du Comité Français de Cartographie*, n°169-170, pp. 42-51.
- Devillers R., Jeansoulin R., 2005 - Qualité de l'information géographique, 349 p, Hermes Science Publication Lavoisier.
- Dobos E., Daroussin J., Montanarella L., 2005 - An SRTM-based procedure to delineate SOTER Terrain Units on 1 : 1 and 1 : 5 million scales. EUR 21571 EN. Office for Official Publications of the European Communities, Luxembourg. (Document technique). 55 p.
- Falipou P., Legros J.-P., 2002 - Le système STIPA-2000 d'entrée et d'édition des données pour la base nationale de sols Donesol. *Etude et Gestion des Sols*, 9(1), pp. 55-70.
- FINKE P., *et al.*, 2001 - Bases de données géoréférencées des sols pour l'Europe. Manuel de procédures version 1.1. EUR 18092 FR, 174 p., European Communities.
- Gaultier J.P., Legros J.P., Bornand M., King D., Favrot J.C. et Hardy R., 1993 - L'organisation et la gestion des données pédologiques spatialisées : le projet Donesol. *Revue de Géomatique*, 3, pp. 235-253 - France
- Grolleau *et al.*, 2004 - Le système d'information national sur les sols : Donesol et outils associés. *Etude et Gestion des Sols*, 11(3), pp. 255-270.
- Groot, R. 1993. Making information technology work. *ITC J. 3* - pp. 228-235.
- Hengl T. *et al.*, 2006 - Evaluating Adequacy and Usability of soil maps in Croatia. *Soil-Sciences* (vol 70). pp. 920-929.
- Infosol, 2005 - Donesol version2.0 : Dictionnaire de données, 357 pages
- Infosol, 2005 - Référentiel Régional Pédologique : Cahier des Clauses Techniques Général, 21 pages.
- Irvin B.J., Ventura S.J., Slater K.B. (1997) - Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin - *Geoderma*, 77, pp. 137 - 154
- King *et al.*, 1994 - Soil space organization model and soil functioning in GIS. *Trans 15th World Soil Science Cong. (Acapulco)*, 6a - pp. 743-757.
- Le Bas C. *et al.*, 2004 - Utilisations des données Sols d'I.G.C.S. en France : un état des lieux, 11 (3), pp. 299-306.
- Le Bas C., Schnebelen N., 2006 - Utilisation des données Sols d'I.G.C.S. en France. *Etude et Gestion des Sols*, 13(3), pp. 237-246.
- Legros J.-P., 1978 - Recherche et contrôle numérique de la précision en cartographie pédologique : (1) Précision dans la délimitation des sols *Ann agron.* 29 : pp. 499-519 ; (2) Précision dans la caractérisation des unités de sols. *Ann. Agron.* 29 - pp. 583-601.
- Legros J.-P., Falipou P., Dunand-Divol F., 1992 - Vérification de la qualité de l'information dans la base de données de sol. *Science du Sol*, 30(2), pp. 117-131.
- Legros J.-P., 1996 - Cartographies des sols : de l'analyse spatiale à la gestion des territoires. Presses polytechniques et universitaires romandes. 321 p.
- McBratney A.B., Mendonça M.L., Minasny B., 2003 - On digital Soil Mapping - *Geoderma*, 117, pp. 3 -52
- Robbez-Masson J. M., 1994 - Reconnaissance et délimitation de motifs d'organisation spatiale : Application à la cartographie des pédopaysages. Thèse de doctorat, ENSA M, Montpellier, Science Agronomiques, 161 p.

PUBLICATIONS ET DOCUMENTS PUBLIÉS PAR L'AFES

Revue

Science du Sol

Revue scientifique publiée de 1952 à 1993.

Elle comporte 300 à 400 pages par an. Un index est présenté tous les ans dans le quatrième numéro.

A cessé de paraître fin 1993. Certains numéros disponibles.

La Lettre de L'Association

Publiée quatre fois par an, ce journal annonce les nouvelles de l'association, les réunions nationales et internationales; il donne des critiques d'ouvrages, de thèses, de la documentation, etc.

La Lettre est envoyée à chaque adhérent de l'association: elle accompagne l'adhésion.

Rédacteur en chef: J.P. Rossignol, ENITH, Angers.

Etude et Gestion des Sols

Revue trimestrielle, francophone traitant de la connaissance et de l'usage des sols.

Rédacteur en chef: M. Jamagne.

Secrétariat de rédaction: Micheline Eimberck et J.P. Rossignol.

Le Comité éditorial est composé de trente membres de France et de pays francophones.

Ouvrages

Le Livre Jubilaire (1984)

Point sur les acquis à cette date en matière de science du sol et de pédologie.

Fonctionnement hydrique et comportement du sol (1984)

PODZOLS et podzolisation

par D. Righi et A. Chauvel: ouvrage publié en coédition par l'AFES et l'INRA, avec le concours du CNRS, de l'ORSTOM, et de la région Poitou-Charentes (1987).

Micromorphologie des sols/Soil micromorphology

par N. Fédoroff, L.M. Bresson, Marie Agnès Courty, publié par l'AFES avec le concours du CNRS, de l'INAPG, de l'INRA, du Ministère de l'Environnement et de l'ORSTOM (1985) (épuisé).

Carte mondiale des sols et sa légende (1984)

Présentée sous forme numérique.

Le Référentiel Pédologique

Principaux sols d'Europe, deuxième édition 1995.

Ouvrage collectif publié par l'AFES et l'INRA.

Synthèse nationale des analyses de terre: Période 1990-1994

par C. Walter, C. Schwartz, B. Claudot, P.-Aurousseau et T. Bouedo, avec le concours du Ministère de l'Agriculture et de la Pêche.

Actes du XVI^e Congrès Mondial de Sciences du Sol, Montpellier - Août 1998