

# Synthèse nationale des analyses de terre réalisées entre 1990 et 1994

## I. Constitution d'une banque de données cantonale

Ch. Schwartz<sup>(1)</sup>, Ch. Walter<sup>(2)</sup>, B. Claudot<sup>(1)</sup>, Th. Bouédo<sup>(3)</sup> et P. Arousseau<sup>(3)</sup>

(1) ISA, Laboratoire Sols et Environnement - 41, rue du Port - 59046 Lille

(2) ENSA Rennes, Laboratoire de Science du Sol - 35042 Rennes

(3) ENSA Rennes, Laboratoire de Spatialisation Numérique - 35042 Rennes

### RÉSUMÉ

A partir d'informations transmises par des laboratoires d'analyses de terre, une banque de données d'analyses de terre a été constituée, concernant tout le territoire national.

Dans un premier temps, la faisabilité d'une telle banque de données est analysée. Celle-ci repose d'abord sur l'agrément des différents laboratoires par le Ministère de l'Agriculture, qui implique qu'ils utilisent les mêmes méthodes analytiques. Elle est également basée sur une validation concernant le mode d'expression des résultats et la localisation des lieux de prélèvement des échantillons, identifiée par la commune. Après élimination des données douteuses, nous avons conservé près de 300 000 analyses réalisées entre 1990 et 1994, correspondant à plus de 3 000 000 de déterminations.

La pertinence d'un traitement d'ensemble de ces informations, recueillies selon une procédure s'apparentant plus à une enquête sans stratification préalable de la population qu'à une prospection pédologique systématique, est ensuite discutée. Ces données présentent l'intérêt, par leur grand nombre, de permettre des approches statistiques, globales ou spatialisées, complémentaires des approches pédologiques habituelles.

Dans un second temps, nous montrons que le canton est l'entité spatiale élémentaire la plus pertinente pour agréger les analyses dans le cadre d'une synthèse nationale. Chaque canton est alors décrit à partir de descripteurs statistiques robustes (médianes et quartiles), conduisant à la création d'une banque de données cantonale qui est brièvement décrite.

### Mots clés

Analyse de terre, analyse spatiale, banque de données, cartographie des sols, gestion des sols

### SUMMARY

#### STATISTICAL REVIEW OF THE SOIL TESTS MADE IN FRANCE FROM 1990 TO 1994 : ESTABLISHMENT OF A SOIL ANALYSIS DATABANK.

*Using data submitted by soil testing laboratories, a soil analysis data bank has been established on a national scale.*

*As a first step, the feasibility of such a data bank is discussed. It is based on the approval of the various laboratories by the French Ministry of Agriculture, which implies that they all use the same analytical processes. It is also based on a validation procedure regarding the reporting of results and location of the sampling place, known as the "commune". After rejection of invalid data, we kept about 300 000 the analyses made between 1990 and 1994, the equivalent of more than 3 000 000 elementary results.*

The relevance of the information gained from this data set, collected using a procedure more similar to an inquiry than to an organized pedological prospection, is then discussed. It allows for interesting statistical, global or spatialised, treatments that complement usual pedological surveys.

As a second step, we show that the "canton" (which corresponds in average to 10 "communes") is the best basic unit to use for aggregate analyses in the context of a national summary survey. Each "canton" is described with robust statistical descriptors, (median or quartiles). The results are then stored in a cantonal data bank, which is briefly described.

#### **Key-words**

Soil test, spatial analysis, soil data bank, soil survey, soil management.

### **RESUMEN**

#### **SÍNTESIS NACIONAL DE LAS ANÁLISIS DE TIERRAS REALIZADAS ENTRE 1990 Y 1994**

##### **I - constitución de un banco de datos cantonales.**

A partir de las informaciones transmitidas por los laboratorios de análisis de tierras, un banco de datos de análisis de tierra, que concierne el territorio nacional fue constituido. En primer lugar se analizó la posibilidad de un tal banco de datos. Este reposa primeramente sobre el reconocimiento de los diferentes laboratorios por el Ministerio de Agricultura, lo que implica que utilizan los mismos métodos analíticos. Está también basado sobre una validación que concierne el modo de expresión de los resultados y la localización de los lugares de muestreo identificados por el municipio. Después de la eliminación de los datos dudosos, hemos conservado cerca de 300 000 análisis realizados entre 1990 y 1994, correspondiendo a más de 3 000 000 de determinaciones.

La pertinencia de un tratamiento de conjunto de estas informaciones, recogidas según un procedimiento que se aparenta más a una encuesta sin estratificación previa de la población que a una prospección pedológica sistemática, está después discutida. Estos datos presentan el interés, por su gran número, de permitir enfoques estadísticos, globales o espaciales, complementarias a los enfoques pedológicos habituales.

En segundo lugar, mostramos que el cantón es la entidad espacial elemental la más pertinente para agregar las análisis en el cuadro de una síntesis nacional. Cada cantón está entonces descrito a partir de descriptores estadísticos robustos (medianas y cuartiles), conduciendo a la creación de un banco de datos cantonal que se describe brevemente.

#### **Palabras claves**

Análisis de tierra, análisis espacial, banco de datos, cartografía de suelos, gestión de suelos.

**L**e sol est de plus en plus largement considéré comme un patrimoine menacé (INRA, 1996) qu'il importe donc de protéger. Dans ces conditions, il est paradoxal de constater que les sols français restent dans l'ensemble mal connus. En effet, la moitié du territoire national ne dispose d'aucune cartographie des sols, et les synthèses départementales ou régionales de la diversité pédologique restent trop peu nombreuses et hétérogènes (Villemain et Florentin, 1988 ; Bellivier *et al.*, 1995).

De plus, les sols cultivés (55 % du territoire national) évoluent constamment du fait des pratiques agricoles. Les propriétés concernées par cette évolution sont nombreuses : pH, densité apparente, teneurs en cations échangeables, phosphore extractible, matière organique, etc., sans oublier les éléments trace si l'on prend en compte les épandages de tous types réalisés par les agriculteurs.

Les dispositifs mis en place pour suivre l'évolution de caractéristiques des sols, tels, par exemple, l'Observatoire de la Qualité des Sols (Leprêtre et Martin, 1994 ; Bornand *et al.*, 1996) ou le programme Renecofor (ONF, 1996), ont des objectifs différents. Ils s'appuient notamment sur un nombre limité de points d'observation qui, bien qu'ils soient suivis avec précision, ne permettent pas la spatialisation sur un vaste territoire des variations mesurées.

Dans ce contexte, nous nous proposons d'analyser l'information qu'apportent les résultats des analyses de terre utilisées par les agriculteurs pour connaître et gérer les sols qu'ils cultivent. Regrouper ces résultats dans une Banque de Données d'Analyses de Terres permet de les réutiliser dans un cadre plus global, tant au plan spatial que temporel.

L'objectif de ce travail est d'obtenir, à partir de l'information ainsi recueillie, une meilleure connaissance des caractéristiques physiques et chimiques des sols cultivés du territoire français. Avec un recul plus important que les 5 années dont nous disposons pour l'instant, il est envisageable de mesurer l'évolution des paramètres de ces sols modifiés par l'activité agricole. Par ailleurs, nous envisagerons également les possibilités d'utilisation de ces résultats dans le cadre d'approches pédologique, agronomique ou environnementale. Ce travail a été conduit dans le cadre du programme IGCS (Inventaire, Gestion et Conservation des Sols) en réponse à une demande exprimée à l'AFES par la DERF (Direction de l'Espace Rural et de la Forêt au Ministère de l'Agriculture, de la Pêche) (Schvartz *et al.* 1995).

Dans cet article, nous décrivons la méthodologie de constitution de cette banque de données au plan national. Dans un second, nous en présenterons une première exploitation.

## ANALYSE PRÉALABLE DE L'INFORMATION RECUEILLIE

Les analyses utilisées ont été effectuées dans des laboratoires soumis à une procédure d'agrément de la part du Ministère de l'Agriculture et de la Pêche (arrêté du 30 décembre 1986). Cet agrément concerne autant le respect de protocoles analytiques normalisés que la fiabilité des résultats : ceux-ci sont ainsi compatibles entre eux et il est donc possible de regrouper dans une telle banque de données des résultats provenant de différents laboratoires d'analyses agréés.

### Le recueil de l'information auprès des laboratoires

#### Les réactions des laboratoires

A partir de la liste des laboratoires agréés fournie par la DERF, 48 laboratoires ont été contactés. Parmi eux, 27 ont répondu favorablement à la demande de transfert à titre gracieux de résultats d'analyses concernant la période 1990-1994.

Les raisons du refus de près de la moitié de ces laboratoires peuvent se regrouper en 2 catégories principales.

#### *Des raisons informatiques*

L'informatisation du laboratoire est en ce cas insuffisamment développée pour permettre la transmission des résultats sur support informatique. Cela peut être dû :

- soit à l'absence totale d'informatisation (c'est le cas de laboratoires ne réalisant qu'un petit nombre d'analyses),
- soit à une informatisation trop rigide où la fonction d'exportation des résultats n'a pas été prévue. Celle-ci ne peut pas être installée facilement à cause d'une absence de maîtrise de son informatique par le laboratoire (sous-traitance à une entreprise extérieure), ou d'un manque de disponibilité pour réaliser une modification qui n'est pas forcément simple, ni considérée comme importante ou urgente.

#### *Des raisons commerciales*

Elles sont exprimées par des laboratoires qui considèrent que ces informations ont un niveau de confidentialité empêchant de les communiquer, ou qu'elles présentent une valeur économique (exprimée en termes d'échange ou de vente) qu'il nous était impossible de satisfaire. Il s'agit généralement de laboratoires réalisant de grandes quantités d'analyses.

### Les données reçues

Les données reçues sont soumises à 3 étapes successives de validation : informatique, analytique et géographique.

#### *Validation informatique*

Si les données sont généralement transmises sous des formats informatiques standards tels que Base, ASCII, etc. ... et sont aisément récupérables, certaines se présentent sous des modalités de stockage plus anciennes ou moins courantes. Les

conversions nécessaires ne sont pas toujours possibles ou sont parfois incomplètes à cause de la compatibilité imparfaite des systèmes informatiques. Cela a entraîné la perte d'une partie des données reçues.

#### **Validation analytique**

Si les méthodes analytiques utilisées dans les différents laboratoires sont normalisées (AFNOR 1996), la façon d'exprimer les résultats peut varier. Outre les unités employées (pour mille ou mg/kg), les divergences peuvent porter également sur le mode d'expression de la granulométrie (100 % incluant ou non la matière organique), le mode de détermination de la matière organique (à partir du carbone organique ou de l'azote total), pour ne reprendre que les exemples les plus caractéristiques.

Cela nécessite donc une transformation des informations reçues pour obtenir une homogénéité satisfaisante des données traitées par la suite. Lorsque cette transformation n'est pas possible avec une fiabilité suffisante, les données correspondantes sont éliminées.

#### **Validation géographique**

L'information de base permettant la localisation d'un résultat est la commune de la parcelle d'où provient l'échantillon analysé.

Celle-ci peut être identifiée au départ soit par son code INSEE (forme archivée dans la banque de données finale), soit, plus généralement, par son nom. Celui-ci doit donc être converti ; s'il est erroné ou imprécis (lieux dits ou cas d'homonymie) la conversion n'est pas possible et l'analyse correspondante est éliminée.

Par ailleurs, à l'issue de cette étape, nous avons constaté des concentrations "suspectes" d'analyses provenant de communes urbaines ou parfois rurales. Ces anomalies correspondent systématiquement à des sièges ou bureaux d'organismes ayant une fonction de centralisation par rapport aux analyses de terre (coopératives, chambres d'agriculture, bureaux d'études, ...). La localisation de ces analyses n'étant pas fiable, elles ont été également éliminées.

Finalement sur les 337 173 analyses "informatiquement correctes" reçues, 296 352 ont été conservées. Le taux de rejet, proche de 12 %, nous paraît tout à fait satisfaisant, surtout dans le cadre d'une première réalisation de cette ampleur. En fait, le plus important déficit d'information provient surtout du refus de collaboration de certains laboratoires. Cependant, comme ceux-ci font partie des plus importants, il est vraisemblable que si leurs données pouvaient être disponibles, le taux de rejet ne pourrait être que diminué.

#### **Les questions posées par l'échantillonnage**

Par delà les procédures de validation évoquées ci-dessus, la fiabilité et l'intérêt de l'utilisation d'une telle banque de données sont conditionnés par la validation scientifique de son contenu. Celle-ci passe par une réflexion sur les conditions d'obtention des informations ainsi archivées.

La réalisation d'une analyse de terre comporte deux étapes principales : le prélèvement de l'échantillon, puis son analyse en laboratoire. Nous avons déjà expliqué pourquoi nous considérons la seconde phase comme satisfaisante et homogène pour toutes les données à notre disposition. Pour ce qui est de la constitution de l'échantillon, il est nécessaire d'envisager à la fois les modalités et les motivations de son prélèvement.

Par ailleurs, la réutilisation dans un cadre spatial ou temporel vaste, de données initialement conçues pour une utilisation ponctuelle, nécessite également quelques précautions.

#### **Les modalités du prélèvement**

Un échantillon représente une parcelle cultivée, ou une partie de celle-ci. Il présente donc 2 caractéristiques essentielles (AFNOR, 1996 ; ITCF, 1987) :

- il provient d'un horizon superficiel (en général 0-25 cm pour une parcelle en grande culture, 5-15 cm pour une prairie) ;
- il est constitué à partir d'un ensemble de prélèvements élémentaires mélangés entre eux pour constituer un échantillon composite intégrant la variabilité spatiale de la surface prospectée. Celle-ci correspond habituellement soit à la totalité de la (partie de) parcelle concernée, soit à un cercle de 10 à 20 mètres de diamètre considéré comme représentatif de l'ensemble par l'agriculteur et/ou le préleveur.

De plus, même si une feuille de renseignements décrivant succinctement la parcelle accompagne généralement l'échantillon au laboratoire, nous ne connaissons de son contenu que la localisation de la parcelle (obligatoire) et une indication très simplifiée de son utilisation (facultative parce que la saisie informatique de cette information n'est pas généralisée).

La connaissance qui découle d'un tel ensemble d'analyses de terre n'est donc pas assimilable à celle qui est obtenue à partir d'analyses réalisées dans le cadre d'une prospection pédologique classique ; elle ne pourra pas, en conséquence, être utilisée de la même façon.

#### **Les motivations du prélèvement**

Le prélèvement est réalisé à la demande d'un agriculteur recherchant un élément d'aide à la décision pour la gestion d'une parcelle. Ce contexte induit également une orientation particulière de la stratégie d'échantillonnage (Danneels et Schwartz, 1993) :

- seul l'espace agricole cultivé est prospecté (les forêts sont quasiment absentes),
- les parcelles utilisées en grandes cultures sont sur-représentées par rapport aux autres (principalement les prairies permanentes),
- les motivations techniques de l'agriculteur (gestion raisonnée ou diagnostic de problème) influent sur le choix des parcelles analysées et la fréquence du recours à l'analyse. La répartition des valeurs des paramètres modifiables par la pratique agricole (éléments échangeables, matière organique,

pH, ...) en est affectée, mais non celle des paramètres pérennes comme la granulométrie.

- En outre, l'hétérogénéité de la répartition régionale des informations contenues dans la banque de données est également influencée par la sensibilisation des agriculteurs à l'utilisation des analyses et donc par leur dynamisme. La spatialisation des typologies d'exploitation, en lien avec ce dynamisme, nous montre que la répartition des différents types n'est pas homogène sur le territoire (Christofini et al, 1994).

Là encore, la stratégie d'échantillonnage s'apparente plus à une enquête sans stratification préalable de la population qu'à une prospection organisée. Les informations disponibles dans une telle banque de données se distinguent donc nettement de celles qui sont obtenues dans le cadre d'une étude pédologique habituelle. Leur traitement devra donc être adapté à ce contexte.

#### **Approche spatiale : comment localiser les prélèvements ?**

Compte tenu des remarques ci-dessus, il apparaît qu'une localisation ponctuelle de ces informations par un système de coordonnées géographiques n'est pas réaliste. Le moyen le plus satisfaisant et le plus accessible (donc le plus fiable) consiste à utiliser la commune d'origine de l'échantillon.

Est-ce suffisant ? Certes, une localisation communale ne permet pas la spatialisation de la diversité des sols au sein d'une commune. Cependant cette diversité peut s'apprécier globalement à partir d'une étude statistique des résultats. Par ailleurs, le mode d'échantillonnage n'assure pas une représentation homogène de tous les types de sols existant. En effet, seules les zones agricoles sont concernées et certains sols n'y sont pas ou peu représentés ; de plus, l'importance du recours à l'analyse de terre fluctue avec le type d'utilisation de la parcelle.

Cependant, nous disposons là d'un matériau (en principe) aisément mobilisable et qui dans de nombreux secteurs de notre pays peut constituer la seule information concernant les sols qui soit utilisable à une échelle intermédiaire entre l'approche parcellaire et le 1/1 000 000.

Par ailleurs, bien que la qualité de la spatialisation de l'information ne soit pas idéale, son abondance (nombre d'analyses par commune) présente le grand intérêt de permettre une approche statistique. Celle-ci autorise, à partir de descripteurs spatialisés pertinents basés sur des représentations fréquentielles (médianes, quantiles, modes) une expression convenant bien aux besoins d'outils d'aide à la décision exigeant, pour un zonage à moyenne échelle, une approche des problèmes en termes probabilistes, ou de risque (Danneels et Schwartz, 1993). Ces données peuvent également être utilisées, seules ou associées à d'autres types d'informations, pour faire fonctionner des modèles mathématiques destinés à calculer des variables dérivées (comme l'indice de battance ou le coefficient de minéralisa-

tion de la matière organique) ou à prévoir des comportements plus globaux (risques de fuites de pesticides vers les nappes, ...).

Enfin, nous pensons que dans le cadre d'unités naturelles limitées, il est possible d'établir une relation entre des caractéristiques de l'horizon de surface et le type de sol (Schvartz et Douay, 1992). Dans ces conditions, le couplage d'une banque de donnée d'analyses avec un Système d'Information Géographique contenant des données relatives par exemple au relief, à la pédologie et/ou à l'utilisation de l'espace pourra(it) s'avérer très fructueux ...

#### **Approche dynamique : quelle pérennité de ces données ?**

Le grand nombre de données accumulées pendant plusieurs années rend accessible une étude dynamique des paramètres modifiables par l'activité agricole (Walter et al. 1995). Il est, là aussi, nécessaire d'interpréter les résultats avec prudence, autant à cause de la représentativité des parcelles échantillonnées que de la pérennité des valeurs recueillies.

Les conseils donnés habituellement par les laboratoires incitent à réaliser une nouvelle analyse sur une parcelle après un délai de 3 à 5 ans. Dans un fichier regroupant des résultats collectés sur une période supérieure à 5 ans, le risque d'en avoir deux provenant de la même parcelle n'est pas nul. De plus la seule localisation du prélèvement par la commune rend impossible une vérification rigoureuse de l'existence de tels doublons.

D'autre part, du fait de la pratique agricole, certaines valeurs évoluent. C'est d'ailleurs souvent la raison même de la demande d'analyse.

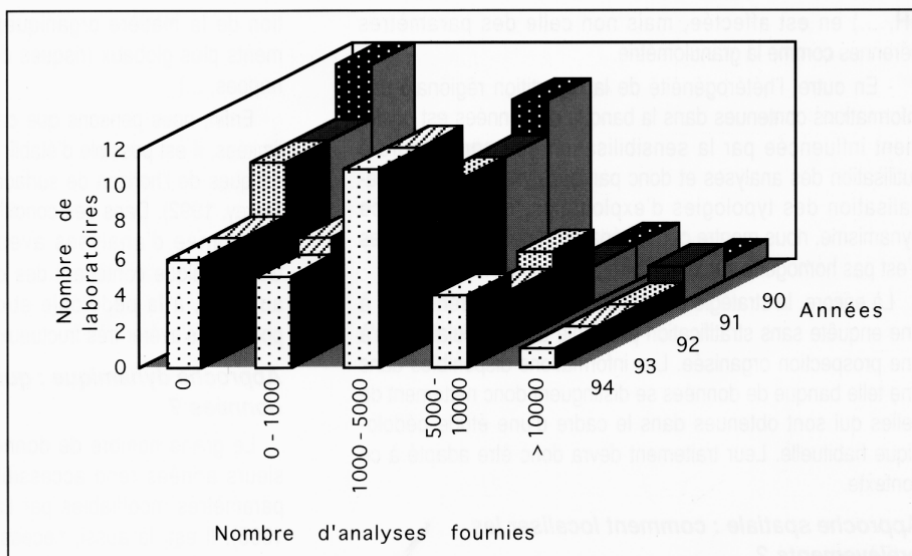
Si la validité des résultats granulométriques est pérenne, celle des autres variables dépend du comportement dans le sol des éléments correspondants. Etant donnés les mécanismes agronomiques mis en jeu, leur cinétique et la précision d'une telle approche, nous estimons, à partir des conseils habituellement donnés pour la réalisation d'une nouvelle analyse sur la même parcelle, ce délai à :

- environ 10 ans pour la matière organique et la CEC,
- 3 ans pour les déterminations classiquement appelées "chimiques",
- 1 an pour le pH.

Le statut du calcaire total est intermédiaire. En effet, seules les valeurs très faibles (quelques pour mille) sont influencées les par la pratique agricole (un apport de 5 tonnes de calcaire correspond à une variation de teneur inférieure à 2%). Au delà de quelques pour-cent, les valeurs peuvent donc être considérées comme pérennes. Pour une étude concernant ce paramètre, les seuils de valeur retenus seront donc particulièrement sensibles.

**Figure 1 :** Répartition par année du nombre d'analyses envoyées par les laboratoires.

**Figure 1 :** Annual repartition of the number of analyses sent by the laboratories.



### Les données recueillies

Pour disposer d'un nombre d'analyses suffisant tout en minimisant le risque d'avoir deux analyses provenant de la même parcelle, nous avons conservé, pour la constitution finale de la banque de données, l'ensemble des analyses reçues couvrant la période de cinq ans, allant du 1er janvier 1990 au 31 décembre 1994. Nous avons ainsi collecté environ 300 000 analyses validées (tableau 1), réparties de façon à peu près homogène sur les cinq années (figure 1).

Tableau 1 : Répartition des 27 laboratoires selon le nombre d'analyses transmises.

Table 1 : Repartition of the 27 laboratories according to the number of analyses they sent.

Nombre total d'analyses	Nombre de laboratoires concernés
moins de 5 000	11
5 000 - 10 000	7
10 000 - 15 000	3
15 000 - 20 000	2
plus de 20 000	4

La majorité des laboratoires (18) ont fourni au total moins de 10 000 analyses et seuls quatre laboratoires dépassent 20 000 analyses en cumul sur les cinq ans. Deux laboratoires

dépassent certaines années 10 000 analyses par an. Cela montre à la fois l'hétérogénéité et une concentration certaine de l'activité de l'analyse de terre en France. Cet aspect n'aurait pas été modifié si tous les laboratoires avaient accepté de participer à cette action.

Les déterminations archivées sont les suivantes :

- la granulométrie,
- le pH (eau et KCl),
- le calcaire (total et actif),
- l'azote total,
- le carbone organique,
- le phosphore extractible par les méthodes Joret-Hébert, Dyer ou Olsen<sup>1</sup>,
- le potassium, le magnésium, le calcium et le sodium échangeables,
- la capacité d'échange cationique (Metson ou Riehm)<sup>1</sup>,
- les oligo-éléments(EDTA ou solubles à l'eau)<sup>1</sup>.

Ces déterminations ne sont cependant pas systématiquement renseignées : les effectifs disponibles varient fortement de l'une à l'autre (tableau 2 et figure 2). Les plus courantes sont d'abord le pHeau, les cations échangeables (sauf le sodium), le phosphore extractible (toutes méthodes confondues) et le carbone organique, puis viennent la capacité d'échange et l'argile.

(1) : lorsqu'il existe des méthodes analytiques différentes, leur mention est conservée et associée au résultat obtenu pour permettre un traitement spécifique si nécessaire.

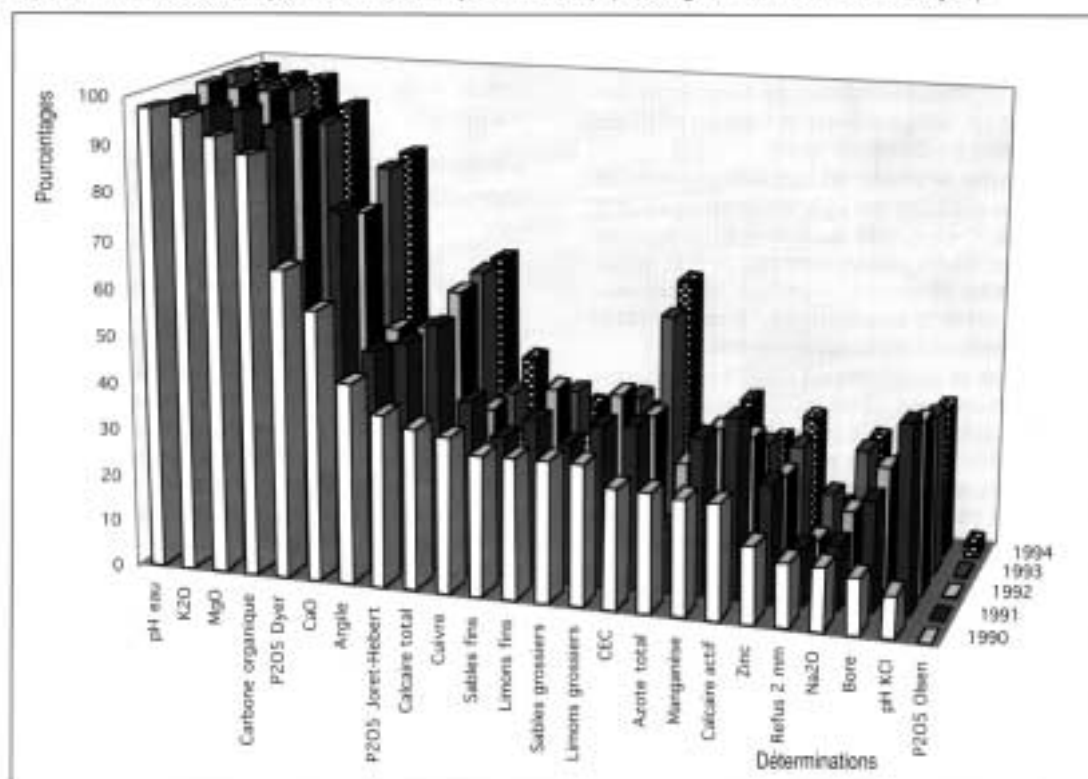
Tableau 2 : Nombre de résultats disponibles par détermination.

Table 2 : Number of available results, for each elementary determination.

déterminations	effectif total	déterminations	effectif total
Argile	122 268	P <sub>2</sub> O <sub>5</sub> Joret-Hebert	137 010
Limons fins	91 810	P <sub>2</sub> O <sub>5</sub> Dyer	144 416
Limons grossiers	92 083	P <sub>2</sub> O <sub>5</sub> Olsen	5 755
Sables fins	81 285	K <sub>2</sub> O	288 749
Sables grossiers	78 529	MgO	286 241
Refus 2 mm	28 537	CaO	233 390
pH eau	291 971	Na <sub>2</sub> O	100 569
pH KCl	75 504	CEC	147 867
Calcaire total	182 766	Bore	71 856
Calcaire actif	69 503	Cuivre	106 706
Azote total	56 780	Zinc	79 569
Carbone organique	274 014	Manganèse	90 154
		<b>Total</b>	<b>3 134 332</b>

Figure 2 : Résultats disponibles par détermination et par an (exprimés en pourcentage du nombre d'analyses annuel).

Figure 2 : Available results per year for each elementary determination (in percentage of the annual number of analyses).



Les fluctuations inter-annuelles du nombre de résultats disponibles pour chaque détermination doivent être considérées avec prudence. En effet, dans le cas de la constitution de cette première banque de données, la relative croissance du nombre de données disponibles d'une année sur l'autre (de 54 118 en 1990 à 65 577 en 1994) peut être liée à la plus grande accessibilité des analyses les plus récentes. Le trait marquant semble correspondre à la stagnation (voire à la diminution) des déterminations granulométriques accompagnée de l'accroissement du nombre de CEC réalisées. Ce fait est cohérent avec le recours de plus en plus fréquent à ce paramètre pour interpréter les résultats d'analyses chimiques et établir un conseil de fertilisation.

Le nombre moyen de déterminations par analyse est très stable au fil des années et égal à 10.

## LA CONSTITUTION D'UNE BANQUE DE DONNÉES CANTONALE

### Intérêts et limites du découpage communal

Chaque analyse est localisée par la commune de la parcelle échantillonnée. Le découpage de la France en 36 500 communes constitue donc la base d'agrégation élémentaire des résultats contenus dans la banque de données. Cependant la petite taille des entités géographiques que constituent les communes n'est pas nécessairement la plus adaptée pour toutes les exploitations possibles de ces résultats.

Les intervalles de confiance des descripteurs statistiques tels que moyenne ou médiane sont jugés satisfaisants à partir de 10 à 15 analyses (Dersigny, 1988). Seules les unités géographiques regroupant au moins 10 analyses seront donc prises en considération dans les traitements. Cependant les communes comportant plus de 10 déterminations sur la période 1990-94 sont en fait relativement peu nombreuses (tableau 3).

Si ce mode de spatialisation est adapté à des approches régionales comme celles qui ont été réalisées notamment en Bretagne et dans le Nord de la France, il ne l'est pas vraiment pour une approche que nous voulions conduire à l'échelle nationale : les restitutions cartographiques trop "pointillistes" masquent les grandes tendances puisque 15 % au maximum des communes sont concernées par une information statistiquement exploitable.

Par ailleurs, de nombreux responsables de laboratoire n'ont pas souhaité que soient diffusées des indications trop détaillées concernant leur aire d'activité. Plus vaste est le territoire sur lequel les informations sont regroupées, plus il y a de chances pour que s'y mêlent des origines différentes.

Nous avons donc convenu pour des motifs tant scientifiques que déontologiques de réaliser les traitements à l'échelle natio-

**Tableau 3 :** Variations pour quelques déterminations du nombre de communes disposant d'au moins 10 résultats (parmi les 36 500 communes françaises).

**Table 3 :** Variations according to the determinations, of the number of communes with at least 10 results (within the 36 500 communes in France).

Détermination	Nombre de communes disposant d'au moins 10 résultats
Carbone organique	5 511
pH eau	3 117
Potassium échangeable	3 104
Argile	2 423
CEC	1 442

nale en utilisant des entités spatiales plus grandes que les communes. Le fichier de données communales a cependant été conservé (mais de façon confidentielle), tant pour ne pas perdre l'information recueillie que pour permettre des croisements éventuels avec d'autres types de données, suivant des découpages différents (cartes pédologiques, géologiques, topographiques, d'utilisation du territoire...) en limitant au maximum les erreurs induites par la taille de l'unité élémentaire de spatialisation.

### L'établissement d'une base de données cantonale

#### Le choix du découpage cantonal

A partir de la commune, unité administrative élémentaire, trois bases d'agrégation sont accessibles : le canton (3 511 entités pour la France entière), la «Petite Région Agricole» (PRA, 433 entités) ou le département (95 entités).

Le découpage optimal est celui qui permet de minimiser la variabilité intra-strate des variables utilisées. Pour comparer la variabilité liée à ces trois modes de découpage de l'espace, nous avons calculé pour chaque entité spatiale les paramètres élémentaires de statistique descriptive (moyenne, écart-type, médiane, quartiles). Ce calcul a également été fait pour les communes ayant un effectif supérieur à 30 analyses et pour l'ensemble du pays afin d'avoir des points de comparaison.

Les résultats obtenus sont résumés dans le tableau 4. L'écart-type est utilisé comme indicateur de la variabilité : la comparaison de la moyenne des écarts-type intra-strates avec l'écart-type national permet d'apprécier l'intérêt des différents découpages pour décrire la variabilité spatiale d'une propriété.



**Tableau 4** : Comparaison de la variabilité intra-strate pour différents types de découpages (Commune, Canton, Petite Région Agricole, Département) à la variabilité générale observée pour l'ensemble des échantillons.**Table 4** : Comparison of the intra-strate variability for different spatial limits, within the overall variability.

	Moyenne des écarts-type intra-strates				
	Commune	Canton	PRA	Département	France entière
Argile (‰)	63	74	86	94	105
Limons totaux (‰)	94	107	129	144	186
Sables totaux (‰)	111	126	156	179	215
Carbone organique (‰)	6,3	7,0	8,1	9,0	11,2
C.E.C (meq/100g)	5,8	7,5	10,2	11,3	17,2
pH eau	0,6	0,7	0,7	0,8	1,0
P <sub>2</sub> O <sub>5</sub> Dyer (mg/kg)	189	201	212	215	247
P <sub>2</sub> O <sub>5</sub> Joret-Hébert (mg/kg)	140	164	192	222	218
Potassium (mg/kg)	119	131	143	162	177
Magnésium (mg/kg)	93	110	145	169	156

L'analyse de ce tableau conduit aux conclusions suivantes :

- la moyenne des écarts-type intra-strates croît, pour toutes les propriétés, avec la taille de l'entité géographique de référence.
- cette moyenne est dans la plupart des cas, sensiblement plus faible avec un découpage cantonal qu'avec un découpage en petites régions ; la variabilité intra-cantonale est de plus souvent peu supérieure à la variabilité intra-communale.

Par ailleurs deux types de comportements apparaissent :

- les paramètres dont la variabilité intra-cantonale est nettement inférieure à la variabilité générale : les fractions limoneuses, les sables, la CEC, le carbone organique ;
- ceux dont la variabilité intra-cantonale est plus forte (de 70 à 80 % de la variabilité générale) : l'argile, le pH, le potassium et le magnésium échangeables, le phosphore extractible.

Le premier groupe rassemble essentiellement des paramètres peu modifiés par les pratiques culturales, à la différence de ceux que l'on retrouve dans le second.

L'application de cette démarche au calcaire total présente peu d'intérêt à cause du très grand nombre de valeurs égales à zéro.

En résumé, le découpage cantonal apparaît intéressant dans la mesure où il répond aux contraintes évoquées ci-dessus, sans trop dégrader la précision de l'information traitée. C'est donc à ce niveau que nous avons choisi d'agréger les données pour en faire la synthèse à l'échelle nationale.

#### Quels descripteurs cantonaux ?

Réaliser une représentation cartographique, basée sur les unités territoriales que sont les cantons, nécessite de déterminer une valeur unique représentative de l'ensemble des analyses provenant de chaque canton. Le mode de détermination de ce

descripteur synthétique peut dépendre du type de variable. Par exemple des calculs de moyenne ou écart type ne sont pas compatibles avec des données non continues. On distingue trois types de données (tableau 5) : des données continues, correspondant aux valeurs brutes ou au résultat de calculs effectués à partir de celles-ci, et des données discrètes, provenant d'interprétation conduisant à des classes d'appréciation.

L'utilisation de la moyenne suppose une distribution gaussienne de la variable étudiée qui n'est généralement pas vérifiée ici. On préférera donc avoir recours à la médiane qui est utilisable pour toutes les variables. De plus la faible sensibilité de ce paramètre de position aux valeurs extrêmes lui confère un caractère de robustesse appréciable dans le cas de cantons disposant de peu d'analyses.

En complément de la médiane, la détermination des quartiles extrêmes permet une bonne appréciation de la dispersion d'une variable au sein d'un canton. Les quartiles nous paraissent préférables aux déciles dans la mesure où ils sont moins influencés par les distributions très étalées que nous rencontrons ; de plus, l'espace interquartiles permet de bien qualifier la partie centrale de la population.

Pour les variables non ordonnées (comme la texture), la médiane sera remplacée par l'occurrence dominante ou modale rencontrée sur chaque entité géographique. Dans ce cas, la représentation de la variabilité est moins aisée, elle peut cependant être abordée à partir de la représentativité de la classe dominante, indiquée par sa fréquence.

D'une façon plus générale, les représentations fréquentielles permettent une description quantifiée de la variabilité intra-cantonale : cela correspond à un apport important de cette banque de donnée. Ces représentations seront ensuite particulièrement

**Tableau 5** : Répartition des principales données selon leur type.**Table 5** : Repartition of the main data according to their type.

Données brutes, continues	Données calculées continues	Données d'interprétation agronomique, discrètes
argile, limons et sables	indice de battance	texture
P <sub>2</sub> O <sub>5</sub> , K <sub>2</sub> O, MgO, CEC		niveaux de P <sub>2</sub> O <sub>5</sub> , K <sub>2</sub> O, MgO
pH, CaCO <sub>3</sub>		niveaux de CaO, de pH
matière organique, Azote	coef. de minéralisation, C/N	niveaux de matière organique
oligo-éléments		niveaux d'oligo-éléments

adaptées pour une utilisation dans le cadre de modèles conduisant à des résultats évalués en termes probabilistes ou de risque.

### Les caractéristiques de la base de données obtenue

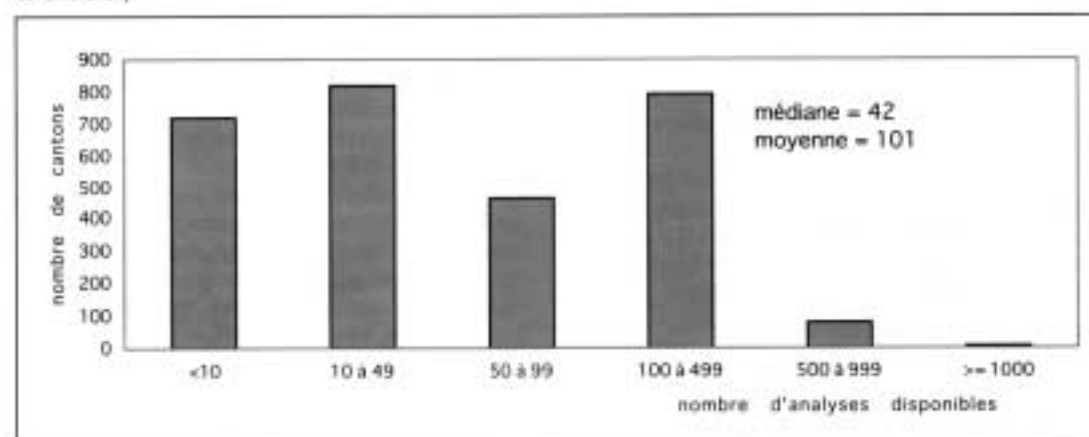
Le regroupement des analyses selon un découpage cantonal permet de disposer de résultats pour 2 881 cantons (soit 82 % de l'ensemble) et d'informations exploitables (au moins 10 analyses) pour 2 162 cantons, soit pour près de deux cantons sur 3 (figure 3).

La répartition spatiale des analyses est hétérogène (carte 1). Cette hétérogénéité est liée à trois facteurs dont les effets se superposent : l'importance de l'activité agricole, la sensibilisation des agriculteurs à l'utilisation d'analyses de terre et l'aire d'action des laboratoires. Celle-ci est d'ailleurs plus aisément perceptible

lorsque des laboratoires ayant une forte implication régionale n'ont pas souhaité transmettre de résultats.

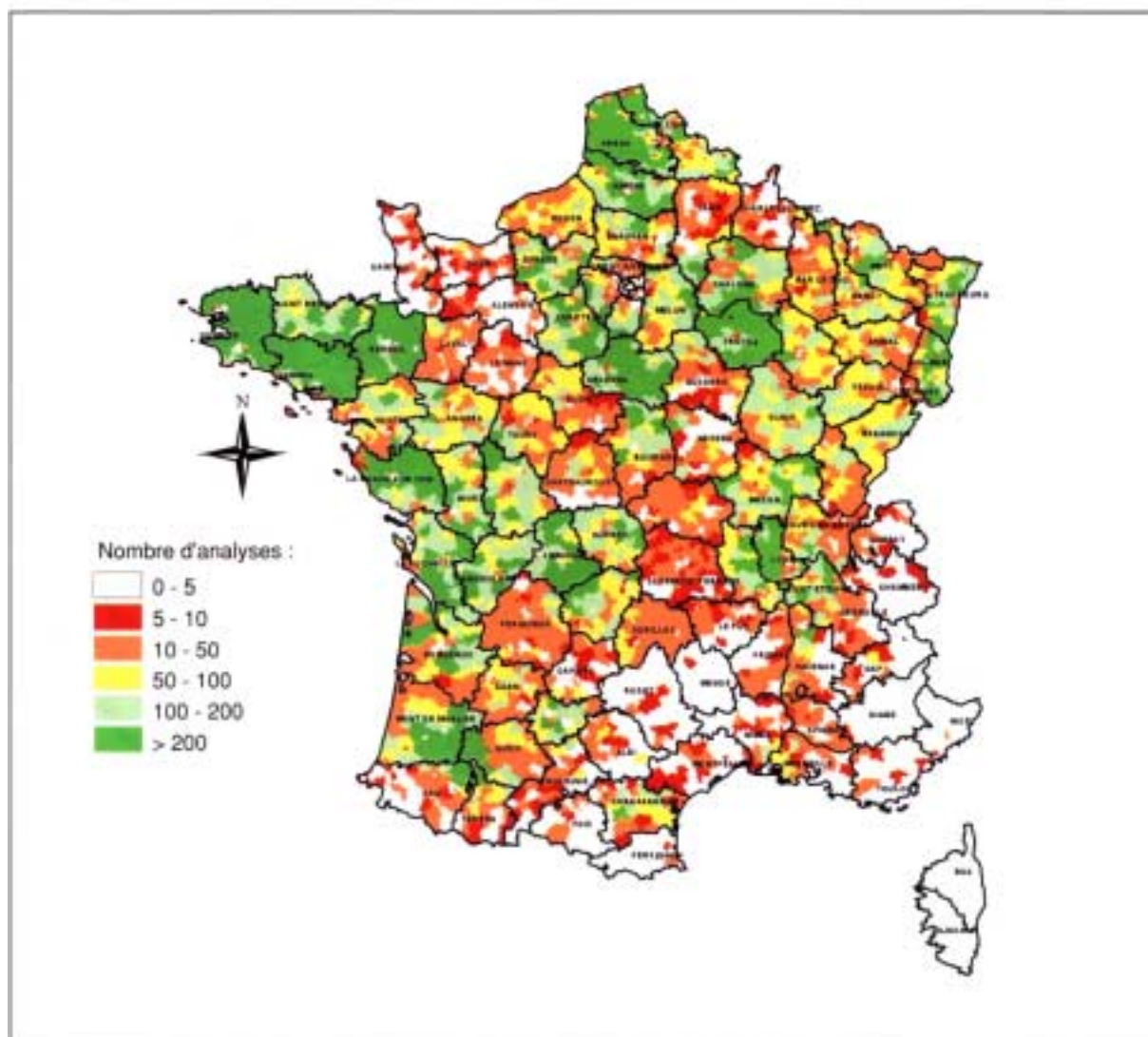
La superficie des cantons n'étant pas constante, la précision de l'information est mieux rendue par sa résolution (carte 2) qui indique la surface théorique correspondant à une analyse. La surface prise en considération est la surface totale du canton. Compte tenu de la précision assez globale de notre approche nationale cela nous paraît suffisant. Une étude plus fine, réalisée par exemple à un niveau régional, nécessiterait de considérer la surface agricole utile (SAU) du canton plutôt que sa surface totale, puisque les analyses utilisées concernent exclusivement du territoire agricole.

La résolution apparaît ici tout aussi hétérogène que la répartition numérique des analyses par canton. Elle permet

**Figure 3** : Répartition des cantons selon le nombre d'analyses disponibles (les cantons sans analyse ne sont pas pris en considération).**Figure 3** : Repartition of the cantons, according to the number of available analyses (cantons without any analysis are not taken into consideration).

**Carte 1** - Nombre d'analyses par canton.

**Map 1** - Number of analyses per «canton».

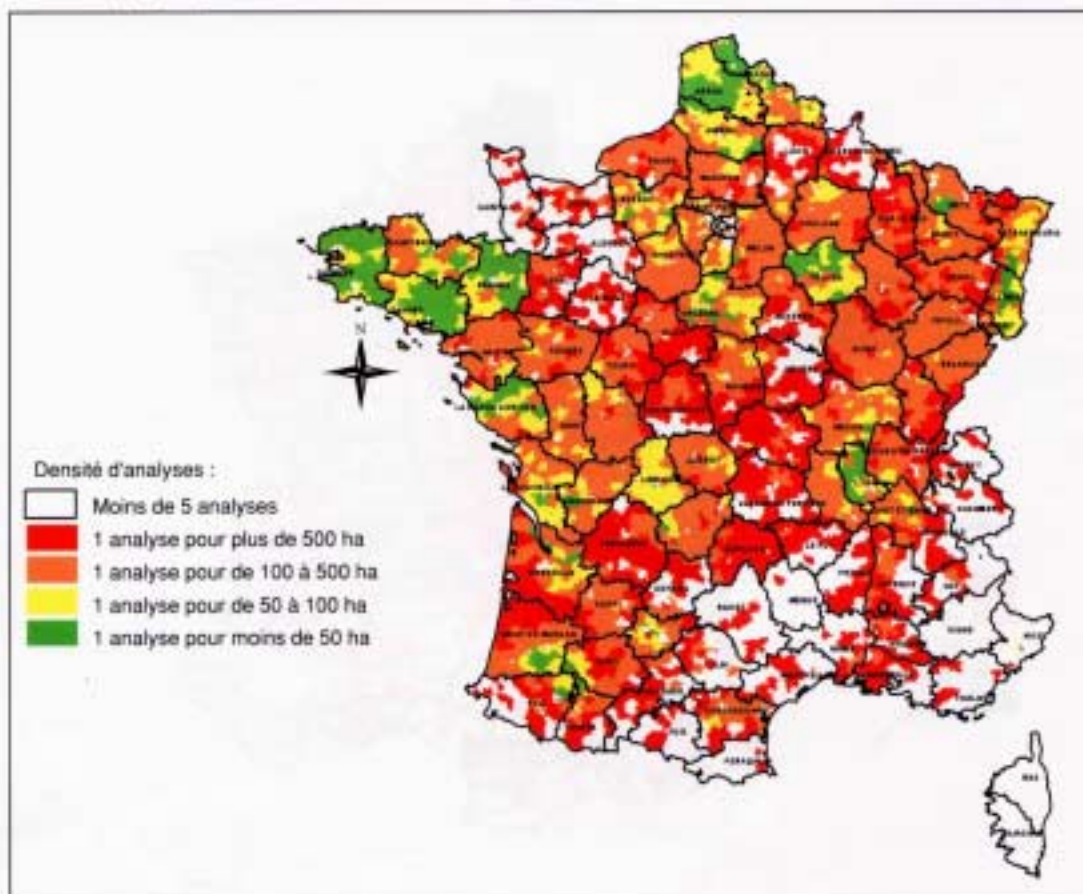


cependant de disposer d'une information supplémentaire concernant l'échelle de représentation d'une cartographie établie à partir d'une telle banque de données. La comparaison avec les indications données dans le cadre du programme IGCS (IGCS, 1991) montre qu'il n'est pas réaliste d'utiliser un document cartographique établi à partir des seules informations d'une telle banque de données à une échelle plus fine que le 1/250 000 (tableau 6).

Pour qu'elle soit réellement pertinente, la précision spatiale de l'information doit être considérée distinctement pour chaque paramètre étudié, dans la mesure où tous les échantillons de

terre ne sont pas traités suivant le même menu analytique. Nous aborderons ces aspects dans l'article suivant.

La prise en considération de la résolution est en outre un complément indispensable aux différentes validations des données de cette banque, que nous avons discutées précédemment, lorsque l'on souhaite les combiner à d'autres bases de données concernant le climat, l'utilisation de l'espace, la géomorphologie, la pédologie ou d'autres éléments. Elle est en effet nécessaire pour évaluer la précision du résultat final du traitement réalisé dans la mesure où celle-ci dépend de la confiance que l'on peut accorder à la source de données la moins précise.

**Carte 2** - Résolution de l'échantillonnage (en fonction de la surface totale des cantons).**Map 2** - Sampling density (in function of the total area of the «cantons»).**Tableau 6** : Résolution pour les cantons ayant au moins 10 analyses.**Table 6** : Resolution for the cantons with at least 10 analyses.

	moyenne	377 hectares
	médiane	192 hectares
	premier quartile	72 hectares
	dernier quartile	562 hectares
IGCS carte au 1/250 000	niveau optimal (1 observation pour 200 hectares)	51 % des cantons
	niveau minimum (1 observation pour 800 hectares)	87 % des cantons

## CONCLUSION

Cette première constitution d'une banque de données nationale d'analyses de terre a permis de regrouper les résultats d'environ 300 000 analyses de terre provenant de la plupart des régions françaises.

Ce résultat est à la fois encourageant et décevant. Encourageant car, vue la diversité et la multiplicité des partenaires à impliquer, la faisabilité d'un tel projet n'était pas acquise au départ. Encourageant également de disposer de résultats validés, exploitables, permettant une fois leur exploitation élémentaire réalisée, d'ouvrir de larges possibilités de traitements complémentaires des approches classiques de l'étude des sols et de leur cartographie (cet aspect sera développé dans l'article suivant).

Résultat un peu décevant cependant dans la mesure où il était envisageable de disposer, en cumul sur 5 années, d'une quantité d'informations beaucoup plus importante. En effet il est habituellement admis qu'environ 250 000 analyses de terre sont réalisées annuellement en France. Bien que cette évaluation soit vraisemblablement surévaluée, l'écart paraît grand entre ce qu'il était théoriquement possible de réunir et le résultat effectivement atteint. Nous en avons étudié les raisons et les plus importantes concernent des aspects déontologiques et commerciaux. En effet, les difficultés techniques sont soit faciles à surmonter, soit de faible incidence quand elles ne concernent que des quantités réduites d'informations.

Par ailleurs, il apparaît que cette approche, qui relève de l'enquête et non de la prospection systématique peut constituer un complément intéressant aux études de sols classiques. Les données sont abondantes, datées et assez facilement mobilisables pour de vastes régions ; leur localisation précise et leur représentativité par rapport à la diversité pédologique locale sont par contre plus difficiles à cerner. Sous réserve d'une rigueur scientifique suffisante, l'approche par enquête est un outil pertinent pour l'étude à moyenne échelle de questions pédologiques, agronomiques ou environnementales impliquant l'horizon de surface des sols ou le suivi dans le temps de ses caractéristiques.

## REMERCIEMENTS

Cette étude a bénéficié d'un soutien financier de la Direction de l'Espace Rural et de la Forêt du Ministère de l'Agriculture et de la Pêche, et a été réalisée sous l'égide de l'AFES. Les auteurs tiennent à remercier les différents laboratoires d'analyse de terre du GEMAS qui ont contribué à cette étude en fournissant gracieusement leurs résultats analytiques. Nos remerciements s'adressent également aux personnes du Ministère ou de l'AFES qui nous ont soutenus dans cette entreprise : Mmes M. Eimberck, G. Urbano ; MM. M. Jamagne, M.C. Girard, J.C. Rémy, J.Y. Sommier.

## BIBLIOGRAPHIE

- AFNOR, 1996 - Qualité des sols - Méthodes d'analyses. Recueil des normes françaises, 3ème édition, 534 p.
- Bellivier D., Falque M.C., Pinault N. et Favrot J.C., 1995 - REFERSOLS : REpertoire pour la France d'Etudes pédologiques avec cartographie des SOLS. Programme Inventaire, Gestion et Conservation des Sols, 91 p., annexes et 1 carte
- Bornand M., Girard M.C., King D., Martin S. et Robert M., 1996 - Besoins en données sur les sols en liaison avec les problèmes d'agriculture et d'environnement. 3èmes journées nationales de l'étude des sols, AFES. Comm. orale. Rennes 22-25 avril 1996.
- Christofini B., Roybin D., Etaix P., 1994 - Caractériser le développement régional, est-ce parler de sa qualité ? Une typologie des 128 cantons des Alpes du Nord. Etudes et recherches vol 28 Systèmes Agraires et Qualité, pp. 293-318
- Danneels V. et Schwartz Ch., 1993 - Etude méthodologique sur l'exploitation d'une banque de données régionales d'analyses de terre. Application à la région Nord-Pas de Calais. Ministère de l'Agriculture et de la Pêche-DERF, 31 p.
- Dersigny Ch., 1988 - Collection, organisation et mise à disposition de données agro-pédologiques au sein d'un référentiel pour l'Ille et Vilaine. DAA ENSAR, 54 p.
- ITCF, 1987 - Méthode d'échantillonnage sur des grandes parcelles pour une analyse physico-chimique, in Catalogue de fiches sur des méthodes d'échantillonnage, Service des Etudes Statistiques, 4 p.
- IGCS, 1991 - Cahier des clauses techniques générales pour la réalisation des référentiels régionaux pédologiques (1/250 000). Document Ministère de l'Agriculture - DERF et DGER, INRA 12 p.
- INRA, 1996 - Le sol, un patrimoine menacé ? Forum organisé dans le cadre du cinquantième anniversaire de l'INRA, Paris 24 octobre 1996
- Leprêtre A. et Martin S., 1994 - Sampling strategy of soil quality. Analysis Magazine V.22 n°3, pp. 40-43
- ONF, 1996 - Notice de présentation du Réseau National de suivi à long terme des Ecosystèmes Forestiers, document Office National des Forêts 38 p.
- Schwartz Ch. et Douay F., 1992 - Etablissement d'une clé de détermination des types de sols (exemple de la Flandre intérieure). 3èmes journées nationales de l'étude des sols, AFES. Comm. orale. Lausanne 8-10 septembre 1992.
- Schwartz Ch., Walter Ch., Claudot B., Arousseau P. et Bouédo T., 1995 - Synthèse nationale des analyses de terre. Document AFES - DERF 49 p., 4 annexes, 21 cartes
- Villemin P. et Florentin L., 1988 - Quelques caractéristiques potassiques des sols du plateau lorrain sud. Dossiers Agronomiques d'Aspach (1988),3, pp5-25.
- Walter Ch., Bouédo T., Arousseau P., 1995 - Cartographie communale des teneurs en matière organique des sols bretons et analyse de leur évolution temporelle de 1980 à 1995. Rapport final, Conseil Régional de Bretagne - Agence de l'Eau Loire Bretagne, 31 p.