

**Banque de données et contrôle de l'information pédologique.
Exemple d'application aux sols de l'Algérie du Nord.**

Djili K. , Keslani R.
Institut National Agronomique, El Harrach, Alger

Djili K. et Keslani R., 1995 – Banque de données et contrôle de l'information pédologique. Exemple d'application aux sols de l'Algérie du Nord. Ann. Agron. I.N.A., Vol.16, N° 1 et 2, pp.97 – 112.

Résumé : Les difficultés et les lenteurs d'accès à l'information rendent urgent la création de banques informatisées de données pédologiques. Cette opération qui peut nécessiter plusieurs années d'effort n'a de sens que si l'information qu'elle véhicule est complète et fiable. Dans ce sens, et à travers un exemple de création d'une banque de données pédologique relative aux études de cartographie des sols réalisées par le département de science des sols de l'INA, nous tenterons de montrer l'intérêt d'une telle opération, mais également nous essayerons de cerner les insuffisances (taux de remplissage des fichiers) et la qualité (fiabilité) de l'information pédologique disponible. Des recommandations concernant la création d'une banque de données pédologiques nationale ou régionale ainsi que la normalisation des études pédologiques seront émises.

Mots clés : banque de données, information pédologique, STIPA, profil type, Algérie du nord

Pedological databank and its information control

Abstract : Difficulties and accessibility time to get pedological information is one of many reasons to create a soil databank . This task which needs many years of hard work is worthwhile only if information is complete and reliable. From an exemple for creating a databank including soil data from cartographic surveys held by the soil departement (INA), we tried to show the interest dud the need for database creation. Also, we underlined some constraints and qualities about this available pedological information. Recommandations about national and regional databank creation and pedological studies standardization were suggested.

Keys word : databank, pedological information, STIPA, type profil, North Algeria.

INTRODUCTION

L'accès rapide à l'information et son traitement informatique et statistique s'imposent de plus en plus dans les stratégies de développement et de recherche. Pour cela, la constitution de banques de données centralisées ou régionalisées est devenue souhaitable, voire indispensable.

L'un des principaux obstacles auxquels sont confrontés les chercheurs en Algérie est l'accession à l'information scientifique. En l'absence de structures spécialisées chargées de la collecte, de la gestion et de la diffusion de l'information, les données scientifiques deviennent, dans beaucoup de cas, redondantes et inexploitable du fait de la diversité des méthodes d'étude.

Cette situation est ressentie avec acuité par la science des sols en Algérie. En effet, la diversité des organismes et des structures qui s'intéressent au sol en Algérie (universités, instituts, bureaux d'étude, agriculture...) sans coordination aucune entre eux, a fait que l'information pédologique est très disparate et difficilement accessible aux utilisateurs. Quand le pédologue réussit à l'avoir, dans beaucoup de cas elle devient inexploitable du fait de la diversité des méthodes d'étude et des méthodes d'analyse qui, souvent, sont omises d'être signalées dans les documents.

Dans ce travail, nous tenterons d'aborder deux points essentiels :

- Présenter un exemple d'une banque de données pédologiques sur les sols de certaines régions de l'Algérie du Nord.
- Analyser les données de cette banque pour tester leur fiabilité .

MATERIEL ET METHODES

1. La banque de données

Une banque de données est un ensemble de données enregistrées sur des supports accessibles à l'ordinateur. Elle correspond à un stockage et à une conservation fidèle des données et de leurs structures. Elle a pour objectifs de rassembler et de traiter sur ordinateur, avec des méthodes qui lui sont propres, des données provenant de divers inventaires. Elle nécessite donc un logiciel spécialisé et un équipement adéquat (ACCT, 1978 ; Legros, 1986).

Dans cette étude, nous avons utilisé le logiciel STIPA (Bertrand et al,1984) et comme source d'information les études cartographiques réalisées par le département de science des sols de l'INA d'Alger.

1.1. Utilisation du logiciel STIPA

STIPA, ou système de transfert de l'information pédologique et agronomique, a été conçu et réalisé en France par les chercheurs de l'INRA (Baize, 1990). Il a été mis au point pour pouvoir traiter les données localisées issues de prospections cartographiques. Le logiciel STIPA est constitué de trois parties :

- La partie description.
- La partie analyse.
- Les utilitaires.

La partie description permet de réaliser la saisie , les modifications, les contrôles , l'édition et la mise en banque des descriptions. La partie analyse permet de réaliser pratiquement les mêmes opérations mais pour les analyses. En ce qui concerne les utilitaires, ils permettent, entre autre ,l'extraction des données, la correction de la banque, la création de mini-banque...

1.2. Enregistrement des données

L'enregistrement des données a nécessité le passage par 3 phases :

- . Récolte de l'information.
- . Préparation des données.
- . Saisie ou mise en banque.

1.2.1. Récolte de l'information

Dans cette expérimentation, nous nous sommes intéressés uniquement aux travaux de cartographie des sols réalisés au département de science des sols de l'INA. Ce choix se justifie par les faits suivants :

- Accès facile à l'information : les études sont toutes centralisées au niveau de la bibliothèque de l'INA.
- Les profils dans ces études sont théoriquement les mieux décrits et les mieux analysés : il s'agit de profils types donc représentatifs d'unités de sol.
- Toutes les études de cartographie des sols à l'INA sont réalisées à de grandes échelles (1/5000 au 1/25000), par conséquent, les caractérisations morphologique et analytiques devraient être très détaillées (niveau de la série de la classification pédologique).
- L'archivage de ces études n'a été réalisé qu'après avalisation d'une commission compétente (jury de thèse).

Le Tableau I montre la répartition des études et des profils de la banque de données déjà réalisée.

Tableau I. Répartition spatiale des profils étudiés .

N° Etude	Nombre de Profils	Région	Année
1	11	Aïn El Kebira	1988
2	5	Hamadi	1986
3	7	Zemmouri	1986
4	7	Fennouane	1988
5	9	Beni Slimane	1987
6	12	Taassalat	1990
7	12	Aïn Boucif	1988
8	9	El Omaria	1989
9	18	Boumediene	1981
10	10	Souaghi	1987
11	18	Corso	1983
12	24	Thenia	1981
13	16	Ouled brahim	1980
14	10	Ksar Chellala	1979
15	9	Thenia	1984
16	9	Benchicao	1985
17	12	Région Ouest	1982
18	10	Djouab	1991
19	10	El Aïssaouia	1990
20	8	Haut Chellif	1986
21	11	Ksar Chellala	1979
22	5	Senalba	1991
23	10	Mina	1987

1. 2. 2. Préparation des données

:

Les descriptions et les analyses de tous les profils ont été transcrits sur les fiches de description et d'analyse conçues à cet effet (fiches STIPA). Ce travail, lourd et fastidieux, nécessite une bonne connaissance des méthodes d'étude utilisées dans les différents travaux. C'est une phase critique durant laquelle toute erreur de transcription pourrait se répercuter sur la fiabilité des résultats émanant de l'exploitation de la banque.

1.2.3. Saisie ou mise en banque

Cette opération consiste à relever les données de la fiche de transcription qui seront transmises et stockées dans la mémoire de l'ordinateur sous forme d'un fichier \$Bank.dat. Durant cette opération, nous avons d'abord enregistré les données de description de terrain et ensuite les données de la fiche d'analyse. Ces deux opérations auraient pu être réalisées simultanément.

RESULTATS ET DISCUSSION

Le Fichier \$Bank.dat, ou mini-banque de données, renferme 251 profils avec 923 horizons décrits et 793 analysés (Tableau II).

Tableau II. Nombre de profils et d'horizon étudiés

Nombre de profils étudiés	251
Profils décrits	251
Profils analysés	240
Profils décrits et analysés (complet)	238
Horizons décrits	923
Horizons analysés	793

L'observation du Tableau II montre que, bien qu'il s'agisse de profils types et donc de profils de caractérisation d'unités de sols, seuls 238 profils sur un total de 251 ont une caractérisation complète (décrits et analysés) ce qui représente un taux de 94,8 %. De même, il révèle que seuls 86 % des horizons sont analysés .

L'exploitation de la banque de données est une opération très recherchée par les utilisateurs. En ce qui concerne ce travail, nous nous sommes intéressés à deux objectifs :

- . Le taux de remplissage des fichiers.
- . L'analyse des données.

1. Le taux de remplissage des fichiers

Les Tableaux III et IV montrent le nombre d'observations et d'analyses ainsi que le taux de remplissage de chaque variable utilisée (923 horizons décrits et 793 horizons analysés).

Le travail du pédologue cartographe se base essentiellement sur les observations de terrain. Parmi celle-ci, une bonne description du profil type est une phase critique dans l'étape que doit parcourir le pédologue pour caractériser les unités de sol. Il est généralement admis que la description du profil doit être aussi complète que possible surtout dans les travaux de cartographie à grandes et moyenne échelles. Cependant, l'observation du Tableau III fait ressortir des disparités considérables entre les différentes variables. En effet, nous remarquons que les taux de remplissage des fiches de description sont très variables (2 à 90 %). De même, on remarque qu'aucune variable n'est renseignée à 100%.

Bien que certaines variables sont indispensables pour classer les sols au niveau de la classe (CPCS,1967 ; soil Survey Staff,1994), nous remarquons qu'elles restent mal ou très mal renseignées (effervescence HCl : 85% ; texture : 83% ; couleur : 82% ; structure : 71% ; fente de retrait: 3% ...).

L'analyse des échantillons de sol provenant de profils types est indispensable pour compléter la caractérisation morphologique et donc pour pouvoir bien classer et interpréter (pédogenèse et mise en valeur) les sols (Djili, 1994). L'observation du Tableau III fait ressortir également une grande disparité d'information entre les variables (2 à 94%). De même, nous remarquons que les variables relatives à la mise en valeur restent très mal renseignées (2 à 14%).

De ces Tableaux, il ressort qu'il existe des variables qui sont relativement bien renseignées tels que l'identification, l'effervescence, la texture, le pH, le calcaire total... avec des taux de remplissage allant de 70% à 94 % ; par contre, d'autres variables sont particulièrement délaissées telles que la description de la transition, l'abondance des pores, la densité apparente...Aucune variable n'est renseignée à 100 % . Les paramètres chimiques sont plus utilisés que les paramètres physiques (exemple sur 793 horizons analysés, la densité apparente n'a été déterminée que pour 16 horizons soit 2.02 % du total).

Tableau III : Taux de remplissage de quelques variables.

Variabes	ordre	Nombre d'observation	Taux de remplissage %
Identification	1	844	91.44
Effervescence HCl	2	793	85.91
Texture	3	774	83.85
Couleur	4	761	82.84
M.O	5	760	82.84
Humidité	6	695	75.80
Structure 1er Type	7	655	70.96
Abondance des racines	8	654	70.85
Porosité	9	647	69.00
Transition	10	567	61.43
Activité Biologique	11	448	48.54
Plasticité	12	368	39.86
Régularité transition	13	332	35.96
Compacité	14	261	28.27
Abondance éléments gros	15	255	27.62
Dimension éléments gros	16	243	26.32
Friabilité	17	227	24.59
Chenaux	18	150	16.25
Natures éléments grossiers type 1	19	149	16.14
Couleur des taches type 1	20	119	12.90
Fragilité	21	100	10.83
Structure type 2	22	94	10.18
Nature des éléments secondaires	23	89	9.64
Forme des éléments grossiers	24	74	8.02
Forme des éléments secondaires	25	62	6.71
Orientation des racines	25	62	6.71
Localisation des racines	26	50	5.41
Abondance des pores	27	42	4.54
Dimension des pores	28	29	3.14
Natures des éléments grossiers type 2	29	27	2.92
Fentes	29	27	2.92
Etat des racines	30	23	2.49
forme des taches	31	22	2.38
Pénétration des racines	32	21	2.27
Forme des racines	33	19	2.06
Aspect des faces	33	19	2.06
Contraste des taches	34	15	1.61
Couleur des taches type 2	34	15	1.61
Localisation des revêtements	35	13	1.40
Distribution des taches	35	13	1.40
Nature des éléments grossiers type 3	36	8	0.86
Couleur des taches type 3	37	3	0.32

Tableau IV. Taux de remplissage du fichier selon les variables analytiques (793 horizons).

Variabiles	Ordre	Nombre d'observations	Taux de remplissage en %
pH eau	1	748	94.32
Carbone	2	726	91.55
Azote	3	719	90.66
Sodium	4	702	88.51
C/N	5	701	88.39
Potassium	6	697	87.89
C E C	7	659	83.09
Calcaire total	8	648	81.76
Magnésium	9	642	80.95
Calcium	10	636	80.20
pH KCl	11	573	72.25
Calcaire actif	12	139	17.35
pF 4.2	13	111	14.02
pF 3	14	74	9.33
Fer total	15	72	9.08
Fer libre	15	72	9.08
PF 2.5	16	36	4.54
Densité apparente	17	16	2.02

Cependant, l'interprétation du Tableau II doit tenir compte du fait que certaines variables ne peuvent être renseignées, car inexistantes. Par exemple, l'absence des taches de type 2 ou 3 fait que leurs natures ou leurs couleurs ne peuvent être déterminées d'où un taux de remplissage qui paraît anormalement bas. Dans le même sens, la friabilité, la plasticité et la fragilité ne peuvent être renseignées à 100% car toutes ces variables dépendent du taux d'humidité initiale du sol.

Nous pensons que cette diversité des valeurs du taux de remplissage peut être expliquée par :

- L'absence de normes nationales en matière de description et d'analyse des sols.

- Le choix des variables descriptives et analytiques, ainsi que le choix des méthodes d'approche et d'analyse retenues qui dépendent souvent de l'objectif immédiat de l'étude ce qui lui permettra difficilement de s'inscrire dans un cadre national ou international d'inventaire des sols.

Ceci implique, pour l'Algérie, la nécessité de normaliser l'étude pédologique pour lui permettre de garder son caractère général et pérenne. Pour cela, la création d'un référentiel national pour l'étude et l'analyse du sol s'avère indispensable.

2. L'analyse des données

Avec 923 horizons décrits et 793 horizons analysés, cette banque renferme environ 50.000 données pédologiques (description de l'environnement, du profil et des résultats des analyses).

Souvent, à travers l'analyse des données, on tente d'examiner les relations qui existent entre deux ou plusieurs variables de la banque de données. Dans ce qui suit, seul le contrôle faisant intervenir deux variables sera présenté, il portera sur :

- Les variables quantitatives.
- Les variables qualitatives.

2.1. Relations entre les variables quantitatives

Dans ce genre de relation, un graphe de type x,y permet de visualiser la relation entre les variables prises deux à deux. L'intensité des liaisons linéaires entre les variables considérées peut être connue en calculant le coefficient de corrélation (Tableau.V) par la méthode des moindres carrés (Bouroche et Saporta, 1980 ; Vessereau, 1988).

Tableau V. Les coefficients de corrélation des différents couples.

Relation	Nombre d'observations	Coefficient de corrélation
pH – C/N	898	0.163
pH – S/T	660	0.102
pH – Calcaire total	533	0.258
Calcaire actif – Calcaire total	139	0.147
CEC – M.O	645	0.022
Fer libre – Fer total	73	0.832
pH – Calcium	645	0.289
S – C E C	660	0.77

Ce Tableau montre que les relations Fer libre – fer total et S – CEC ont des coefficients de corrélation élevés (0.91 et 0.77) . Par contre, les autres couples montrent des coefficients de corrélation variables, plutôt faibles.

Ces résultats doivent être manipulés avec prudence car dans beaucoup d'études, les bases échangeables (le Ca surtout) sont déterminés par différence (CEC- somme $Mg^{++} + K^{+} + Na^{+}$), ce qui explique la forte relation qui existe dans ce tableau entre S et la CEC.

A travers quelques exemples, nous tenterons de discuter les relations de certains paramètres à travers leurs coefficients de corrélation et leurs droites de régressions : Les relation fer libre–fer total et calcaire actif –calcaire totale seront étudiées. L'objectif vise uniquement le contrôle des résultats et exclue donc toute interprétation pédogénétique.

Pour ces deux relations, les quantités de fer libre et de calcaire actif ne sauraient dépasser les teneurs totales en fer et en calcaire (Colwell,1958 ; Segalen, 1964 ; Duchaufour et al., 1979). Si le couple fer libre – fer total est bien corrélé, l'observation des figures1 et 2 montre que la forme des nuages va dans le sens de la proportionnalité pour les deux variables de chaque relation. Effectivement, les nuages de points de chaque couple de valeurs sont effilés le long des deux droites de régression. Cependant, la forte corrélation entre le fer libre et le fer total doit tenir compte du fait que seuls les profils à sesquioxides de fer ont été analysés. En ce qui concerne la relation calcaire actif– calcaire totale, la forme de ce dernier dans le profil (dalle, croûte, encroûtement ...) n'a pas été prise en considération d'où la relative faible corrélation.

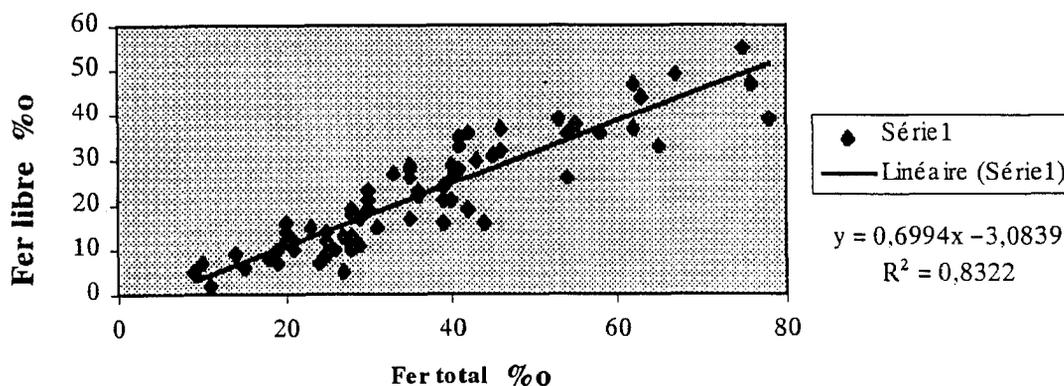


Figure 1. Relation entre la teneur en fer total et la teneur en fer libre

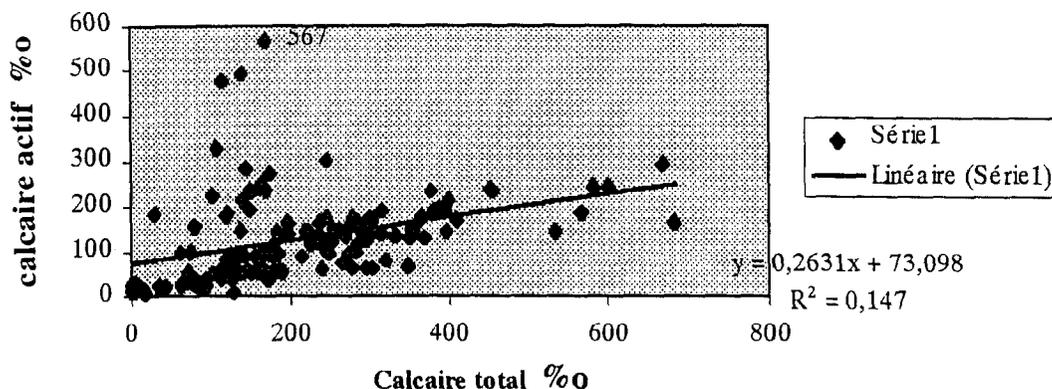


Figure 2. Relation entre la teneur en calcaire total et la teneur en calcaire actif

Dans ces relations, les erreurs ne sont observées que dans le cas du calcaire. En effet, on remarque que certaines valeurs du calcaire actif dépassent largement celles du calcaire total et ce dans 28 observations sur un total de 139, ce qui représente un taux d'erreur de 20% de couples (fig.2).

2.2. Relations entre variables qualitatives et variables quantitatives

Pour illustrer cette relation, nous avons pris l'exemple de l'effervescence qui est une variable qualitative et de la teneur en calcaire totale qui est une variable quantitative. Le test a porté sur une population de 600 couples de variables. La démarche consiste à regrouper toutes les valeurs de la variable quantitative qui appartiennent à la même classe de la variable qualitative, l'illustration étant faite par la construction d'histogramme.

Dans la classe 1 (effervescence nulle ; Fig.3), on remarque que la teneur en calcaire total peut dépasser les 11%

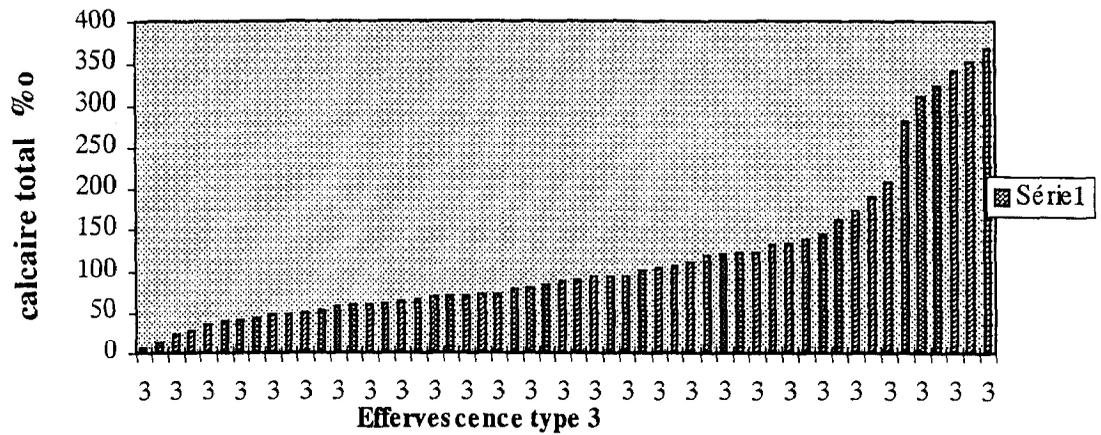


Figure 5. Relation entre la classe d'effervescence 3 et la teneur en calcaire total

La classe 4 (effervescence forte ; Fig. 6), dans 68 horizons une vive effervescence pour des échantillons dont le teneurs en calcaire varient entre 2 et 11% .

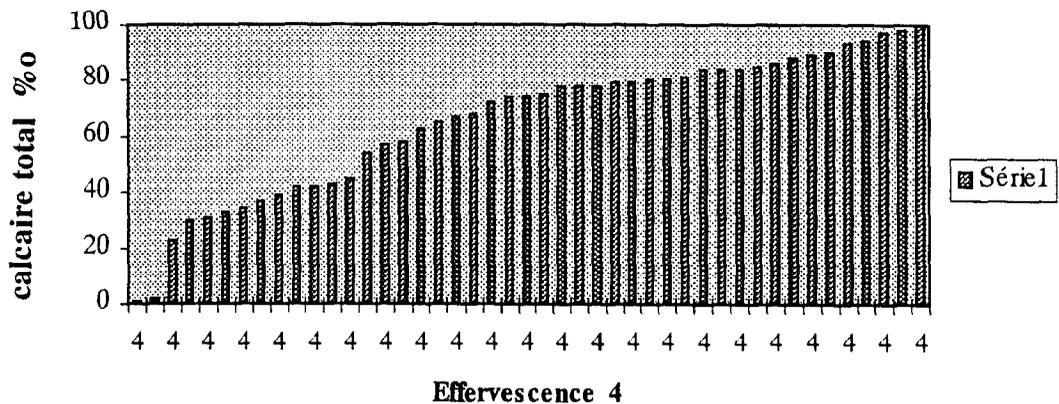


Figure 6. Relation entre la classe d'effervescence 4 et la teneur en calcaire total

Bien que la détermination du taux de calcaire total soit une opération facile à réaliser aussi bien au laboratoire qu'au niveau du terrain, nous remarquons que son appréciation demeure toujours entachée d'erreurs. En effet, il est contradictoire d'avoir

une effervescence nulle pour un taux de calcaire de l'ordre de 11%, ou une vive effervescence à l'HCl dans un horizons qui n'en contient que 2% . Cette remarque nous interpelle une fois de plus sur le contrôle et la fiabilité des résultats d'analyse ou de description.

2.3. Relations entre variables qualitatives

Le contrôle des variables qualitatives décrites sur le terrain est fait selon un découpage en un certain nombre de classes. Pour cela on construit un tableau à double entrées dans lequel on vérifie si la classe de la variable x est compatible avec la classe de la variable y. Il s'agit en fait de connaître les fréquences d'association entre les modalités de deux caractères qualitatifs. Dans notre exemple (Tableau VI), il s'agit de tester la conformité de la texture de terrain avec la plasticité.

Tableau VI. Relation entre la texture et la plasticité

Texture plasticité (P)	A	A L	A S	L	L A	L S	S	S A	S L	A L S	A S L	L A S	L S A	S L A	total	
non P	4	3	3	5	11	9	9	10	10	1	0	4	2	0	2	73
peu P	22	10	12	10	17	8	0	17	9	2	1	5	2	1	1	118
P	73	24	20	1	18	1	0	0	0	2	1	0	0	2	0	141
trés P	11	0	1	0	1	0	0	0	0	0	0	0	0	0	0	13
total	110	37	36	16	47	18	9	27	19	5	2	9	4	4	3	345

Ce choix a été dicté par le fait que la plasticité est très liée à la texture du sol. Sur le terrain, le pédologue cartographe apprécie la texture en réalisant un certain nombre de tests sur des échantillons de sol. Ces tests portent ,entre autres, sur la plasticité de l'échantillon, un sol plastique étant plus argileux qu'un sol non plastique.

Le Tableau VI, fait ressortir les fait suivants :

- 119 couples, soit 34 %, ont une concordance parfaite (exemple : texture argileuse – sol plastique).
- 128 couples, soit 37 %, ont une concordance incompatible (exemple texture limono–sableuse–sol plastique).
- Les autres couples, soit 29 %, ont une concordance approximative (exemple : texture argilo–sableuse – sol non plastique).

De ce fait, nous remarquons que sur un effectif d'échantillons qui reste relativement réduit (345), le taux d'erreur est considérable (37%). Dans ce sens,

l'étude de la variable ' texture argileuse ' fait ressortir une incompatibilité de 26 échantillons sur un total de 110, soit un taux d'erreur qui se situerait autour de 23%

Ce contrôle fait ressortir que les variables qualitatives restent mal appréciées sur le terrain.

CONCLUSION

L'analyse des données a permis de mettre en évidence l'existence d'erreurs et de contradictions entre les variables du sol. Ces insuffisances sont observées pour tous les types de variables (variables qualitatives et variables quantitatives). Par exemple, nous pouvons rencontrer des taux de calcaire actif supérieurs à ceux du calcaire total, ou une effervescence nulle alors que le taux de calcaire dans le sol est assez fort, ou encore un sol plastique pour une texture limono-sableuse.

Le contrôle des données de cette banque a fait ressortir que les erreurs sont nombreuses. L'observation de terrain, le prélèvement d'échantillons, la préparation des échantillons, les méthodes d'analyses non adaptées, la lecture des résultats, la saisie des données... sont autant de sources possibles d'erreurs. De même, il ressort que les fichiers des observations et d'analyses sont parfois très mal renseignés (< 50%). Ces remarques nous permettent de soulever le problème de la fiabilité des données pédologiques des études réalisées jusqu'à présent. A cela, deux problèmes essentiels peuvent erroner l'interprétation : les méthodes d'analyses utilisées au laboratoire, et la terminologie utilisée dans la description du terrain. En effet, l'interprétation et la comparaison des résultats ne sont possibles que s'ils se trouvent dans un même système de référence.

Cette analyse critique, rendue possible et aisée grâce à l'outil informatique, mais surtout grâce à la réalisation d'une banque de données pédologiques, mérite d'être appliquée à l'ensemble des études pédologiques déjà réalisées en Algérie. Les observations qui en découleront permettront de bien cerner les insuffisances et d'élaborer les propositions nécessaires à mettre en application pour les études futures de manière à garantir une meilleure valorisation de l'information pédologique, souvent difficile et coûteuse à obtenir.

Pour l'avenir, nous recommandons la création d'une banque de données pédologiques nationale (ou plusieurs banques régionales interchangeables), mais également la création d'un référentiel pédologique qui permettra de normaliser les travaux d'étude ou de recherche en science des sols en Algérie (méthodes d'étude, méthodes d'analyse, terminologie...).

Références

ACCT, 1978. – Normalisation et échange de données pédologiques par ordinateur, PCST, 192p.

BERTRAND R., FALIPOU P., LGROS J.P., 1984. – STIPA, Notice pour l'entrée des descriptions et analyses de sols en banque de données. Document ACCT, Paris, 136 p.

BOUROCHE J.M., SAPORTA G., 1980. – L'analyse des des données; Presse universitaire de France, Edition Que sais-je, 127 p.

COLWELL J.D., 1958. – Krasnozems in the northern New South Wales. J.Soil Sc. 9,1, 46–57.

CPCS, 1967.,– Classification des sols,E N S A Grignon, 87 p.

DJILIK.,1994. – Cartographie et prospection des sols. Cours polycopié , I.N.A. Alger , 64 p.

DUCHAUFOUR PH., SOUCHIER B., BONNEAU M., 1979.– Pédologie tome 2, constituants et propriétés du sol, édition Masson, Paris.

LEGROS J.P., 1986. – Computerized data sets dor soil maps in géographical information technology in the field of environment. Volume III. Ecole polytechnique federale de Lauzane, Suisse, 865– 213.

SEGALEN P., 1964. – Le fer dans le sol, ORSTOM, Paris,149 p.

SOIL SURVEY STAFF, 1994.– Key to soil taxonomy, USDA, Soil Consevation Service, Sixth Edition, washington , 305p.

VESSEREAU A., 1992. – Méthodes statistiques en biologie et en agronomie, Technique et documentation Lavoisier, Paris, 538p.