

CLASSOL :

Un système expert pour aider à classer les sols dans le Référentiel Pédologique

P. Falipou et J-P. Legros

ENSAM-INRA-Unité LISAH, place Viala 34060 Montpellier cedex 01

RÉSUMÉ

Un système expert, CLASSOL, a été mis au point avec pour objectif le classement assisté par ordinateur, dans le système *Référentiel Pédologique* (RP), d'un sol décrit par un utilisateur non spécialiste. La *figure 1* synthétise la démarche. Le système pose 18 questions simples à l'utilisateur. Les réponses sont choisies dans des listes pré-définies (*tableau 1*). Le logiciel examine alors exhaustivement tous les taxons du RP (*les Références*) et détermine lesquels peuvent convenir avec les réponses fournies. En général, plusieurs noms de sols sont retenus. Ils sont proposés à l'utilisateur, classés dans un ordre d'intérêt décroissant (*figure 7*), sur la base de trois critères présentés dans le texte. Il reste à l'utilisateur à consulter l'aide en ligne (*figure 3*) et l'ouvrage RP pour finir le travail et choisir entre les taxons candidats les mieux placés. Quarante tests ont été réalisés par 8 volontaires sur dix sols typiques (*tableau 3*) décrits et analysés mais dont le nom avait été caché. Les essais montrent (*tableau 4*) que le logiciel est susceptible de présenter, 3 fois sur quatre en moyenne, le bon taxon dans le trio ou quarté de tête. Toutefois, il apparaît que pour utiliser le logiciel et répondre aux questions posées, il faut un minimum de connaissances en pédologie, par exemple pour juger si un profil est différencié ou non. Dans ces conditions, le logiciel ne remplacera jamais le spécialiste. En revanche, il pourrait constituer un bon outil pédagogique pour l'enseignement, son utilisation étant assez plaisante.

Mots clés

Système expert, référentiel pédologique, classement, sol.

SUMMARY

CLASSOL: AN EXPERT SYSTEM TO ALLOCATE SOILS TO THE PEDOLOGICAL REFERENCIAL

An expert system, CLASSOL, was built in order to allocate individual soil profiles to the taxonomic units of the Pedological Referential system (RP). The fig. 1 summaries the approach. A set of 18 simple questions are asked by the system. The right answers must be chosen in lists previously made (table 1). Then the software checks all the taxa (Références) of the RP system to select these that correspond more or less to the whole lot of answers given. Several names of soils are retained. At the end of the trial, the selected taxa are presen-

ted to the user (fig. 7) ordered on the basis of three criteria defined in the publication. Then the user has to make the final choice, using the RP booklet and/or the on line assistance (fig. 3). Forty tests were made by 8 volunteers in order to identify 10 typical soils (tableau 3) characterized by their descriptions and analysis. The names of these soils were hidden. This exercise shows up (table 4) that the software is able to give the good name in the beginning of the list of names in three out of four cases. But it appears that a properly use of the software implies some knowledge of the soil science, for example to decide if the soil is "homogeneous" or with "contrasted horizons". In other words, the software can't substitute for the diagnosis of a specialist. On the other hand, this software could be a mean to learn the terminology in pleasant conditions.

Key-words

Expert system, pedological referential, soil recognition.

RESUMEN

CLASSOL: Un sistema experto, para realizar la clasificación de suelos dentro del Referentiel Pedologique

Se ha implementado un sistema experto, denominado CLASSOL, para realizar la clasificación de suelos asistida por ordenador, dentro del sistema de Referencia Edafológica (RP), a partir de la descripción de suelos realizada por un usuario no especializado. La figura 1 sintetiza el funcionamiento. El sistema plantea 18 preguntas sencillas al usuario. Las respuestas se eligen a partir de unas listas predefinidas (cuadro 1). A partir de ellas, el programa examina exhaustivamente todos los taxones del RP (las Referencias) y determina los que pueden asociarse a las respuestas recibidas. En general, se suelen proponer varias denominaciones de suelos. Estas se presentan al usuario, clasificadas según un orden decreciente de interés (figura 7) y en base a tres criterios presentados en el texto. El usuario deberá consultar la ayuda disponible en línea (figura 3) y la publicación RP para poder finalizar el trabajo y elegir los taxones más aproximados, a partir de las opciones propuestas. Se han realizado cuarenta pruebas con ocho voluntarios utilizando diez suelos típicos (cuadro 3) previamente descritos y analizados pero cuya denominación se mantuvo en secreto. Los ensayos muestran (cuadro 4) que el programa es capaz de presentar, tres veces sobre cuatro como media, la clasificación correcta entre las tres o cuatro primeras opciones ofrecidas. Sin embargo, para poder utilizar el programa y responder a las cuestiones planteadas, resulta imprescindible poseer unos mínimos conocimientos en edafología, por ejemplo, para evaluar si en un perfil se han diferenciado los horizontes. En estas situaciones, el programa nunca puede sustituir al experto. En cambio, puede considerarse un buen instrumento pedagógico en el ámbito docente por su facilidad de manejo.

Palabras claves

Sistema experto, Référentiel Pédologique, clasificación de suelos.

Les spécialistes des sols ont été obligés de constituer des classifications complexes pour bien caractériser et distinguer leurs objets d'étude. Mais, ce faisant, ils ont créé des langages spécialisés qui demandent une formation spécifique et sont difficiles d'accès au débutant. Dans ces conditions, ces classifications ou systèmes de rangement, le Référentiel Pédologique (RP) en particulier (Baize et GIRARD, 1995), n'ont sans doute pas la diffusion qu'ils mériteraient hors des milieux de spécialistes. Pourtant leur utilisation, c'est-à-dire le recours à un langage scientifique commun, est indispensable aux chercheurs comme aux techniciens. Par exemple, les chercheurs agronomes envoyant leurs publications à des revues américaines doivent donner le nom des sols sur lesquels ils travaillent, même s'ils ne sont pas spécialistes de la typologie, sous peine de voir leur manuscrit refusé. Les techniciens qui veulent utiliser les cartes de sols (ex : 1/250 000 et 1/100 000) se heurtent aussi au barrage que constitue la terminologie.

Dans ce contexte, nous avons établi un logiciel qui propose le classement automatique des sols dans le Référentiel Pédologique. Il s'agit bien de ranger des objets observés ou plutôt l'image que l'on en retient (sols-images) en les identifiant plus ou moins précisément à des types pré-établis (sols-concepts) (Baize, 1986). Les anglophones diraient que nous avons fait de *l'assignment* ou de *l'allocation*. Cela est donc différent des efforts visant à créer des catégories dans un lot de profils initialement indifférenciés, sans se référer à une classification déjà existante (GIRARD, 1983).

CHOIX D'UNE METHODE

Revue des méthodes disponibles

Plusieurs méthodes peuvent être envisagées pour bâtir un système de rangement automatique dans une classification existante (action de classement).

La première méthode consiste à constituer puis à enregistrer sur ordinateur un catalogue de sols de référence¹. Ensuite, il suffit de mesurer mathématiquement la ressemblance entre sol examiné et référence cataloguée. La ressemblance la plus nette est déterminée et le sol reçoit le nom du taxon correspondant (notion de rattachement *simple* du RP). Cependant, la ressemblance entre le sol à classer et le taxon de référence peut être approximative (notion de rattachement *imparfait* du RP). En plus, le sol à classer est susceptible de ressembler à la fois à plusieurs taxons de référence. Le rattachement est alors *double* ou *multiple* (RP, 1995). Cette approche correspond assez bien à la façon de procéder des pédologues sur le terrain, en l'absence d'outils informatiques.

Dans le cadre de stages d'étudiants, nous avons exploré il y a quelques années cette voie de la recherche de la plus grande similarité mathématique (Pavat, 1986 – Simonneaux, 1987) entre deux sols. Nous avons alors pu mettre en évidence les difficultés de l'en-

treprise (Legros, 1996). En particulier, il faut définir le poids donné à chaque variable dans le calcul de la ressemblance globale entre deux profils. Mais, ce poids ne devrait pas avoir une valeur fixe. Par exemple, le caractère « texture » est essentiel pour reconnaître un « Vertisol » mais sans importance pour définir un sol alluvial ou un sol colluvial. Dans ces conditions, tel poids ajusté à telle situation se révèle trop fort ou trop faible dans telle autre. Par ailleurs, la ressemblance, dans le cas de variables qualitatives, est une notion bien floue. Quelle est la ressemblance entre vert et jaune, entre polyédrique et cubique, etc. ? L'approche est donc difficile et teintée d'empirisme. Elle a été développée récemment dans le cadre de la *fuzzy set theory* en utilisant la méthode *k-means* (Mazaheri *et al.*, 1995). Mais les auteurs s'en servent pour bâtir un nouveau système de classification et pas pour ranger leurs sols dans la classification australienne considérée dans leur travail.

La deuxième méthode est utilisée depuis des décennies par les phytosociologues pour comparer leurs relevés de végétation. Pour l'essentiel, il s'agit de considérer comme semblables des relevés qui ont les mêmes plantes et, éventuellement, qui ont en commun le fait de ne pas avoir telle ou telle autre plante. On raisonne donc sur la base de critères « présence-absence » exploités, comme dans la méthode précédente, *via* des mesures de distance. Il est évident que cela peut être adapté à la science du sol, les plantes étant alors remplacées par les caractères (ou bien par les horizons) que le profil a ou n'a pas. La voie a déjà été explorée dans le cadre d'essais de classification automatique (Girard, 1976). Nous ne l'avons pas choisie car elle ne nous a pas semblé pertinente pour du classement. En effet, appliquée au niveau des types d'horizons du RP, elle nécessite que l'observateur sache les identifier (horizon calcaire, pétro-ferrique, sulfidique, etc.). Mais alors cet observateur est assez savant pour ne pas avoir besoin de notre logiciel pour faire la synthèse et donner un nom au sol entier ! Appliquée au niveau des caractères des horizons, la méthode implique que des experts soient capables d'établir exhaustivement la liste des caractères qu'un type d'horizon donné a le droit ou n'a pas le droit d'avoir. Une telle méthode serait mieux adaptée à la *Soil taxonomy* américaine car les horizons de diagnostic y sont décrits avec une très grande minutie.

La troisième méthode correspond en l'établissement d'un système expert organisé à partir de règles telles que : « *Si le sol a cette propriété, alors il peut être ou ne pas être de tel ou tel type...* ». Cette façon de faire, proche de celle qui sert à classer une plante dans la

¹ Le RP utilise les vocables « Références » et « Solum ». Pour éviter toute ambiguïté, le mot « référence » étant très général, nous l'employons avec une majuscule lorsqu'il s'agit d'un « taxon » du RP. Mais, il nous arrive aussi d'utiliser le mot « taxon » lorsque cela nous semble plus clair. Par ailleurs, nous préférons le mot « profil » à celui de Solum. La différence entre les deux termes est tenue en matière de classification. On nous permettra donc de rester attachés à un vocabulaire qui a fait le tour du monde, qui est compris de tous et est défendu encore à l'heure actuelle par la communauté internationale (Bockheim *et al.*, 2004).

flore de Bonnier a été testée par les américains dans le cadre de la Soil Taxonomy (Galbraith *et al.*, 1998). Par ailleurs des réflexions générales ont été menées (Dale *et al.*, 1989 ; Albretch *et al.*, 2004). C'est dans cette voie que s'inscrit notre travail. Nous n'avons pas utilisé les logiciels disponibles sur le marché (LISP, PROLOG...). Nous avons préféré bâtir notre propre système, précisément adapté à nos besoins. En particulier, nous avons exploité la notion de flou. Dans CLASSOL, les règles ne sont pas : « *Si le sol a... alors il est* » mais bien : « *Si le sol a... alors il n'est pas impossible qu'il soit* ».

Nous raisonnons à partir de caractères descriptifs qui sont en principe faciles à observer sur le terrain. Une autre tentative, en cours de mise au point, indique comment retrouver un sol du RP (Référence) lorsqu'on a identifié sur le terrain les noms RP des horizons (Baize *et al.*, 2004). Cela va constituer une aide très utile aux étudiants de la discipline. Au contraire, notre tentative est faite en direction de personnes se situant hors de notre communauté de Science du Sol.

Principe de la méthode retenue

Nous considérons d'abord que chacun des taxons du RP est *a priori* candidat à offrir son nom au sol à classer. Le logiciel pose alors une question à l'utilisateur qui dispose d'un panel de réponses toutes prêtes entre lesquelles il lui faut choisir. Par exemple, la réaction à l'acide chlorhydrique est-elle forte, faible, nulle ? Par ailleurs, nous avons bâti et inséré dans le logiciel des tables², taxon par taxon, pour examiner, en fonction de la réponse donnée par l'utilisateur, si son sol peut ou non correspondre à telle ou telle Référence du RP. C'est l'équivalent de la *base de connaissances* des systèmes experts. Par exemple, si l'utilisateur indique que l'effervescence est forte, le sol peut-il être un Calcosol, un Luvisol, un Podzosol, etc. ? Le système de codification des tables correspond à des chiffres et repose sur les bases suivantes :

- note +2 : c'est très possible (ex : une forte effervescence peut très bien correspondre à un Calcosol) ;
- note +1 : c'est possible mais ce n'est pas le cas le plus fréquent (ex : une faible effervescence peut aussi être observée dans certains types de Calcosols),
- note 0 : c'est indifférent ; par exemple, le caractère « calcaire » ou « non calcaire » ne permet ni de renforcer ni d'éliminer l'hypothèse suivant laquelle le sol à classer est un Lithosol,
- note -1 : ce n'est pas impossible mais c'est peu probable (ex : une faible effervescence peut éventuellement s'observer dans un Gypsol car ce type de sol contient parfois des particules de Ca CO₃),

² Pour le système il y a seulement une table de grande taille ; pour l'utilisateur, tout ce passe comme s'il y avait une table par Référence, table accessible via son numéro.

- note -99 : c'est strictement impossible (ex : si le sol est déclaré calcaire, fortement ou faiblement, il ne peut s'agir d'un Rankosol).

Ces notes que nous avons établies à titre d'essai, devraient être soigneusement revues par un collège d'experts. En attendant, on nous permettra de les appeler « notes des experts » pour faciliter la lecture du texte. Elles sont enregistrées dans les « *tables des experts* ».

Admettons que l'utilisateur déclare son sol « faiblement calcaire ». Le score des taxons du RP est modifié de la façon suivante : +1 pour les Calcosols, 0 pour les Lithosols, -1 pour les Gypsol, -99 pour les Rankosols.

Puis le processus recommence avec d'autres questions posées par le logiciel à l'utilisateur. Il y en a 18 au total. Ces 18 questions peuvent rapporter au mieux 2 points chacune à un taxon donné du RP. Le score maximum concevable pour un taxon est ainsi de $18 \times 2 = +36$ points. Donc, si un taxon reçoit, pour l'une des 18 questions, une seule note de -99 (Rankosol dans l'exemple ci-dessus), le score total du taxon affecté restera fortement négatif même s'il reçoit beaucoup de +2 par ailleurs. Il est de fait éliminé définitivement de la compétition.³

L'utilisateur peut sélectionner la réponse « *pas de réponse* » parce qu'il ne sait pas, hésite ou encore trouve la question hors de propos.

On calcule alors trois indicateurs :

- a) Le premier critère est le nombre de questions auxquelles une réponse a été apportée tout en ne donnant pas lieu à un score nul ; c'est en quelque sorte le nombre de questions pour lesquelles l'échange logiciel-utilisateur a été fructueux ; c'est encore le nombre de critères d'identification réellement pris en compte⁴ pour un taxon donné,
- b) Le second critère est la somme des points obtenus,
- c) Le troisième critère est la moyenne des points obtenus par question⁵. Cette moyenne est calculée pour les questions auxquelles on a attribué en retour une note différente de 0.

Les taxons sont présentés à l'utilisateur par ordre des valeurs décroissantes des critères b et c.

Au total, on n'obtient pas un seul nom de sol mais quelques noms possibles entre lesquels il faut choisir. C'est là qu'intervient la deuxième étape du travail. L'utilisateur peut, en cliquant sur les noms qui lui sont proposés, voir exactement à quels taxons cela correspond, photographies ou schémas à l'appui. En plus, il peut se reporter au RP, aux pages indiquées, ce qui lui permet d'achever l'identification

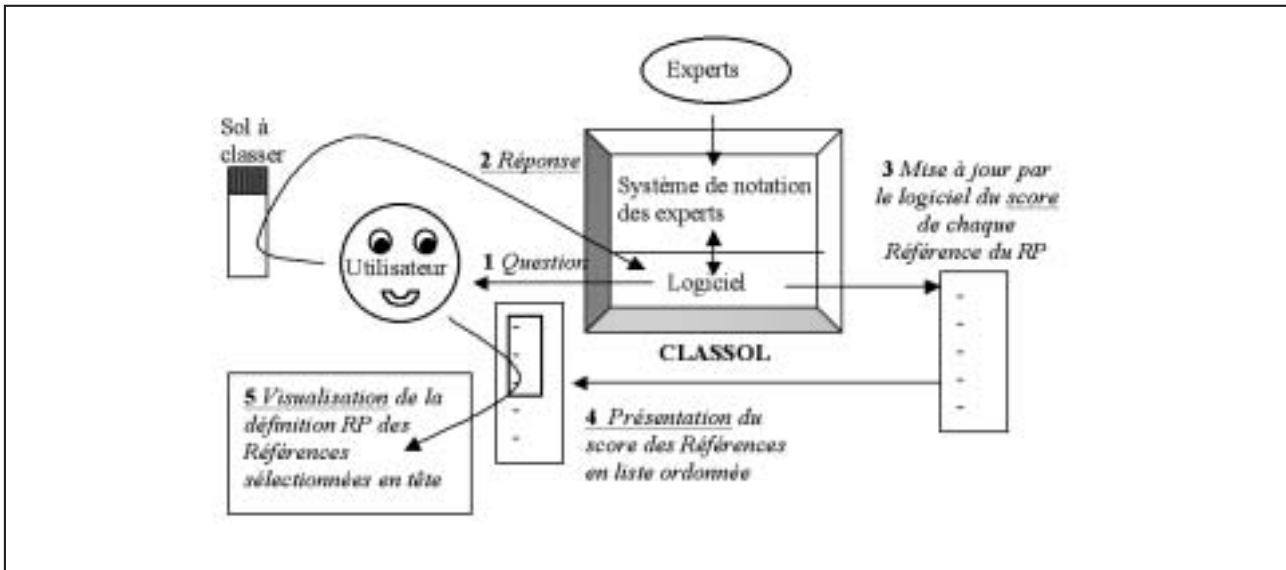
³ D'évidence une note de -35 suffirait à faire basculer le score vers les valeurs négatives puisqu'il resterait au mieux $17 \times 2 = 34$ points positifs à distribuer. Mais, avec -99, on reste en mesure d'ajouter des questions, dans le futur, sans avoir à modifier à ce niveau la volumineuse table des experts...

⁴ Mais on pourrait tout aussi bien considérer que nous confondons, pour ne pas en tenir compte, l'absence d'information et l'existence d'une information sans intérêt. S'il le faut, c'est un point sur lequel on pourra revenir facilement.

⁵ Cet Indicateur est donc le rapport des deux précédents. Le faire figurer sera agréable à tous ceux qui ne sont pas spécialistes du calcul mental avec décimales. Dans l'idéal, sa valeur atteint +2.

Figure 1 - Présentation résumée des 5 étapes de la démarche de CLASSOL

Figure 1 - Summary of the approach



Prenons un exemple : 1. le logiciel demande : « La texture du B est-elle sableuse, argileuse, etc. ? » ; 2. la réponse de l'utilisateur est, par exemple : « Sableuse » ; 3. Le logiciel consulte les tables des experts et met à jour le score des taxons (dans le cas proposé ici, les Podzols restent candidats avec +2 tandis que les Vertisols sont éliminés par un -99) ; 4. le logiciel présente alors la liste ordonnée des taxons sélectionnés à l'utilisateur ; 5. l'utilisateur peut visualiser la définition et l'image des taxons arrivés en tête. La présentation (étape 4) est obtenue dès la première question. La visualisation éventuelle (étape 5) est également accessible dès le début. Mais il est évident qu'elles prennent tout leur intérêt à la fin de la réponse aux 18 questions.

de son sol. Ainsi, Classol, ne se substitue pas à l'utilisateur, il lui permet de «dégrossir» son travail en faisant ce qui est essentiel dans une opération de classification : ouvrir le livret technique à la bonne page⁶ !

La figure 1 résume la démarche.

Liste des 18 questions posées par le logiciel

Certaines de ces 18 questions intéressent l'horizon de surface (A) qui peut être aussi l'horizon unique du profil. D'autres questions concernent l'horizon B qui ne peut être un horizon de surface (les cas exceptionnels sont donc oubliés). Plusieurs questions concernent ces deux types d'horizons (couleurs, textures, etc.). On précise s'il y a ou non un horizon blanchi (E). Mais celui-ci n'est pas décrit plus avant.

Le tableau 1 donne la liste des questions posées et les réponses possibles, tous types de sols confondus.

⁶ Signalons au passage que les taxons, décrits dans le logiciel, en utilisant les termes même du RP, n'ont pas encore tous été illustrés d'une photo. Merci à ceux qui voudraient nous aider à terminer cette partie du travail. Leur nom sera mentionné sous les clichés.

L'aide en ligne permet de savoir exactement ce qui se cache derrière les mots utilisés. Par exemple « cristalline » correspond en fait à « roches et altérites de roches cristallines et/ou métamorphique ». Les catégories retenues pour les roches sont celles proposées par D. Baize.

On pourrait discuter à l'infini du nombre des questions, du type et du nombre des réponses proposées. Tout cela est empirique ! Nous n'avons donc pas à présenter ici de trop grandes justifications qui parfois relèvent de la seule intuition. Tout se joue dans l'aptitude du système à être efficace ou non sur ces bases. Bien sûr, des améliorations pourront être apportées par la suite.

Notons qu'il faut bien choisir la philosophie présidant au remplissage des tables. Donne-t-on le droit de se tromper à l'utilisateur ? Par exemple, l'expert doit-il éliminer définitivement les Calcosols (note : -99) si la réponse est « oui » à la question : « Horizon blanchi ou appauvri ? ». En effet, cet utilisateur a pu rencontrer un encroûtement blanc de CaCO₃ ! L'expert est souvent confronté à ce type de problème. Dans l'état actuel, les tables sont construites sans prise en compte d'erreurs possibles dans les diagnostics. Le logiciel, conçu pour avoir un intérêt didactique, doit raisonner sur des bases saines. Mais, il est

Tableau 1 - Liste des questions posées à l'utilisateur par le logiciel « Classol »**Table 1** - Questions to the user with previously defined answers

N°	Questions (variables)	Réponses proposées (états possibles des variables)
1	Roche ?	Cristalline ; sédimentaire carbonatée dure ; sédimentaire carbonatée tendre ; sédimentaire non carbonatée sableuse ; sédimentaire non carbonatée argileuse ; autres roches sédimentaires ; limons éoliens ; argiles résiduelles ; alluvions fluviales, lacustres ou marines.
2	Différenciation ?	Profil peu différencié, profil très différencié
3	Horizon blanchi et ou appauvri ?	Présence, absence
4	Accumulation dans des horizons ?	Pas d'accumulation, mat.organique, Fe, Al, carbonates, sels, argile
5	Épaisseur du profil ?	0-10 cm, 10-40 cm, 40-200 cm ou plus
6	Effervescence en surface ?	Nulle, faible, forte
7	Effervescence dans l'horizon B ?	Nulle, faible, forte
8	Couleur de surface ?	Noir, gris, brun, jaune, blanc, rouge, vert, bleu, bariolé
9	Couleur horizon B ?	Noir, gris, brun, jaune, blanc, rouge, vert, bleu, bariolé
10	Texture de surface ?	Pas observable, caillouteuse, sableuse, limoneuse, argileuse, équilibrée
11	Texture horizon B ?	Pas observable, caillouteuse, sableuse, limoneuse, argileuse, équilibrée
12	Structure de surface ?	Arrondie nette, anguleuse nette, peu nette, absente.
13	Structure horizon B ?	Arrondie nette, anguleuse nette, peu nette, absente.
14	Présence d'une nappe ?	Oui, non
15	Présence d'un pseudogley ?	Oui, non
16	Présence d'un gley ?	Oui, non
17	Climat du lieu ?	Polaire/haute montagne, continental sec ou froid, tempéré humide, méditerranéen, subtropical sec, subtropical humide/équatorial
18	Microclimat ?	Conditions locales plus humides, conditions locales plus sèches

important de bien voir que l'on pourrait mettre en place une autre approche.

On note que deux questions (n°17 et 18) sont relatives au contexte dans lequel le sol est observé. Le RP, qui s'en tient à des critères objectifs et observables sur les profils, n'a pas prévu cela. Mais Classol peut partir sur d'autres pistes. Nous verrons plus loin que l'approche relève de l'enquête policière.

Nous n'avons pas utilisé directement les analyses de sols. L'appréciation tactile de la texture et la réaction à l'acide chlorhydrique peuvent pratiquement suffire pour répondre aux questions posées, dans la plupart des cas. Ceci exprime notre volonté d'aider des conseillers agricoles qui ont rarement à leur disposition des données très fouillées. Mais il est évident que l'accès à des analyses améliore la qualité des diagnostics, en particulier pour apprécier les accumulations intervenant au sein du profil (matière organique, sels, argile, fer et surtout Al qui ne se voit pas). On note que les exercices de validation présentés plus loin ont été réalisés en disposant des résultats des analyses courantes.

Difficultés méthodologiques fondamentales

Avant d'aller plus loin, il est bon d'examiner quelques questions méthodologiques liées au principe même de fonctionnement du logiciel.

Capacité de diagnostic

On peut se demander si un système de conception aussi simple est capable de retrouver le nom de sols alors que c'est un exercice qui, le plus souvent, nécessite toute la finesse de diagnostic de spécialistes expérimentés examinant de près les différents horizons du profil. En particulier, pas une des réponses aux questions posées n'a, toute seule, une valeur discriminante. Par exemple, « sol argileux » n'est pas synonyme de « Vertisol » ! Mais le logiciel travaille par recoupement. Il va apprendre en même temps que le sol argileux en question est peu différencié, bien structuré, sans horizon blanchi (élimination des Planosols), calcaire dans sa partie médiane (élimination des Pélosols), foncé, présent en climat tropical sec, etc. Au bout du compte, le diagnostic « Vertisols » sera nécessairement émis à défaut d'être prouvé. Cela ressemble à une

enquête policière où l'on sait, par exemple, que l'assassin de la rue Raymond Poincaré (RP) a une voiture noire, une grande taille et habite Montpellier. Tout est banal, mais le recoupement peut être productif.

Par ailleurs, revenons au *tableau 1* et examinons toutes les réponses susceptibles d'être fournies par l'utilisateur. Pour la première variable, il y en a 10 possibles, puis 2 pour la seconde, etc. Au total, il y a :

$10 \times 2 \times 2 \times 7 \times 3 \times 3 \times 3 \times 9 \times 6 \times 6 \times 4 \times 4 \times 2 \times 2 \times 6 \times 2$ soit près de 34 milliards de réponses différentes concevables.

L'expert, de son côté, peut caractériser un sol par 18 variables représentant tout juste 80 états au total, chacun pouvant recevoir une note prise parmi 5 différentes (2, 1, 0, -1, -99, de probable à impossible). Il a donc, à sa disposition 5^{80} possibilités pour caractériser chaque sol. C'est presque l'infini !

Cela montre que la méthode est sans doute moins grossière qu'il semble *a priori*. En tout cas, elle est évidemment suffisante pour distinguer la centaine de Références du RP !

Erreurs de diagnostic (justesse)

Lorsqu'il remplit les tables, l'expert est confronté au problème des cas exceptionnels. Par exemple, un Chernosol (tchernozem) peut-il être sableux ? C'est très peu probable même si des sols isohumiques sableux ont été décrits en Afrique sur des cordons dunaires. Dans ces conditions, l'expert qui remplit la table peut hésiter entre -1 (peu probable) et -99 (impossible). S'il choisit -1, il court ce que les statisticiens appellent le risque de première espèce (accepter une hypothèse fautive) : un sol décrit comme sableux peut être assimilé à tort à un Chernosol. En revanche, s'il choisit -99 il court le risque de deuxième espèce (refuser une hypothèse juste) : un Chernosol sableux ne pourra pas être identifié par Classol. On peut donc hésiter, dans ce cas comme dans beaucoup d'autres. En fait, il est utile de ne pas abuser des notes éliminatoires pour deux bonnes raisons. D'une part, l'erreur de première espèce (acceptation erronée) ne concerne ici qu'une variable. A l'occasion des questions suivantes, notre hypothétique « Chernosol sableux », a encore bien des raisons d'être éliminé. D'autre part, l'exercice a pour objet de présenter à l'utilisateur un panel de taxons. Le dernier tri appartient à cet utilisateur.

Exhaustivité

Il est des cas qui échappent complètement à la méthode. Ce sont ceux qui correspondent à des sols dont le classement se fait sur des critères divers qui ne sont pas pris en compte dans les 18 questions que nous posons. Par exemple, ne seront pas détectés les Anthrosols et aussi les Brunisols re-saturés, les uns et les autres étant perturbés par l'homme. Sont aussi dans la même catégorie des sols impossibles à reconnaître : les Cryosols (action de la glace), les Histosols composites, flottants ou recouverts et enfin les quasi-Luvisols. Il est évident que des questions supplémentaires permettraient de faire entrer ces Références du RP dans le champ de

diagnostic de Classol. Nous ne l'avons pas fait pour éviter de noyer l'utilisateur sous un flot de questions subsidiaires qui, la plupart du temps, ne serviraient à rien. A l'avenir, il faudra certainement complexifier le logiciel pour qu'il pose des questions complémentaires mais dans certains cas seulement. Par exemple, des questions sur la morphologie, en relation avec le rôle de la glace, ne seraient déclenchées que si le sol a été décrit sous un climat de montagne ou de haute latitude.

Ces sols impossible à reconnaître, sont cependant décrits (bouton 1 de la page d'accueil). Dans les tables des experts, leur place est également réservée. Il n'y aura donc pas de difficulté technique pour compléter le logiciel en ce qui les concerne. Dans l'état actuel, pour qu'ils ne soient pas sélectionnés par erreur et présentés à l'utilisateur comme des candidats intéressants, leurs notes ont toutes été mises à zéro. Ils ne peuvent donc pas être crédités d'un score positif.

Finesse du diagnostic (sensibilité)

Enfin, nous nous sommes demandés si la méthode était suffisamment fiable pour descendre au niveau des Références au lieu de s'en tenir à l'identification des Grands Ensembles de Références (GER). En particulier, il est difficile de distinguer les :

- Références hapliques, humiques et perhydriques dans le GER Andosols,
- Référence « saturés, méso-saturés et oligo-saturés » dans le GER Brunisols,
- Références « brunifiés, bruts et typiques » dans le GER Fluvisols,
- Références « fibriques, mésiques et sapriques » dans le GER Histosols,
- Références « dégradés, dermiques, tronqués, typiques et néo » dans le GER Luvisols,
- Références « cailloutiques et pierriques » dans le GER Peyrosols
- Références « distaux, structuraux et typiques » dans le GER Planosols.

Mais, comme on présente le panel des candidats possibles pour le classement du sol, il n'est pas mauvais, au plan pédagogique, que l'utilisateur voie apparaître, en tête de liste, plusieurs Références relatives au même GER. Cela peut lui donner l'envie d'approfondir le sujet. En plus, en procédant ainsi, il y a peu de chances que s'éliminent du classement final, à la fois la Référence intéressante et toutes les Références voisines appartenant au même GER. Bref, le risque d'erreur complète est minoré. Par exemple, si les Andosols hapliques sont éliminés à tort, Classol proposera en tête plusieurs autres Références d'Andosols.

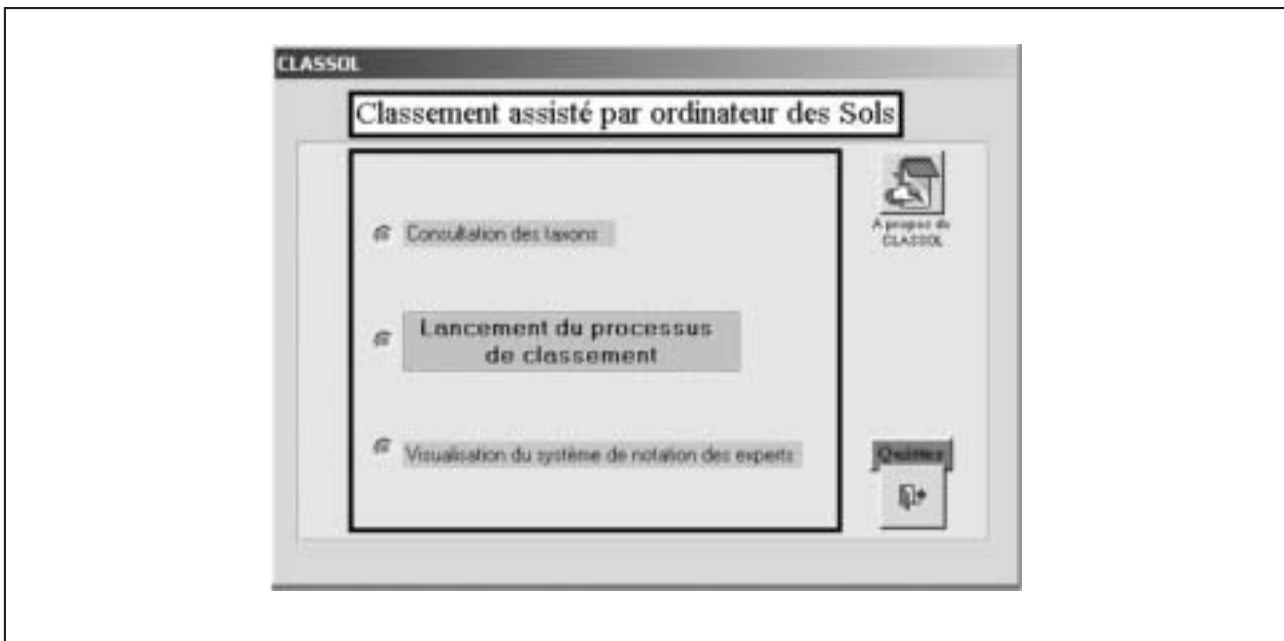
Le résumé des caractéristiques quantitatives de Classol est résumé dans le *tableau 2*.

MISE EN ŒUVRE DE CLASSOL

Les aspects techniques présentés ci-après aident à bien comprendre la partie concrète de la démarche.

Tableau 2 - Classol en quelques chiffres**Table 2** - Data inside CLASSOL

Paramètre	Nombre
Nombre de Références du RP prises en compte	102
Nombres de questions posées par le logiciel	18
Nombre de notes données par les experts pour caractériser un taxon (Référence)	80
Nombre total de notes données par les experts pour caractériser l'ensemble des Références du RP	80 x 102 = 8 160
Nombre de réponses théoriquement différentes que l'utilisateur peut donner pour caractériser un sol	34 milliards
Nombre de sols différents que pourraient caractériser les experts	5 ⁸² ≈ infini
Valeur maximale de la somme de notes	+36
Valeur maximale de la note moyenne	+ 2

Figure 2 - Lancement de l'application**Figure 2** - First screen proposed to the user

Classol a été mis au point « en local », au travers d'une petite application ACCESS.

Lorsque l'application est lancée, un premier écran apparaît (figure 2).

- Le premier bouton autorise la « Consultation des taxons » c'est-à-dire la visualisation des différentes Références du RP. On peut faire défiler les écrans à l'aide des touches du clavier « page suivante » et « page précédente ». Ces écrans peuvent servir à un apprentissage de la terminologie. Un exemple est donné dans la figure 3.

Mais il est aussi recommandé de consulter ces écrans en fin de travail lorsque, après avoir trouvé quelques Références susceptibles de correspondre au sol étudié, on veut en vérifier la pertinence. La

consultation se fait alors directement en cliquant sur les noms des sols proposés.

- Le deuxième bouton de la page d'accueil (figure 2), lance le travail de *Classement automatique*. Il fait d'abord apparaître l'écran de la figure 4. Cela correspond à la première question, qui intéresse la nature de la roche sur laquelle on trouve le profil. A ce stade, il faut sélectionner une des réponses prévues (double clic), y compris « pas de réponse » si on n'est pas en mesure de fournir une information pertinente. L'écran suivant est appelé automatiquement. L'exercice est en fait très rapide. Il faut moins de 5 minutes pour répondre aux 18 questions. Dans ces conditions, il n'a pas été prévu de procédure de correction des erreurs. Si on se trompe, il faut tout recommencer depuis le début (bouton de réinitialisation). Mais,

comme le système garde la mémoire des réponses données, il suffit alors de valider rapidement les réponses antérieures jusqu'au point où l'erreur a été commise.

Si l'utilisateur ne comprend pas bien une question, il peut en savoir plus en cliquant sur le bouton « aide ».

Au fur et à mesure de la progression du travail, une petite fenêtre rappelle à l'utilisateur les indications qu'il a lui-même fournies au système (*figure 5*).

- Le troisième bouton de la page d'accueil (*figure 6*) sert à visualiser les tables de notation des experts, celles qui servent à la reconnaissance des taxons. La *figure 6* montre comment sont organisées ces tables. Il y a une par Référence du RP. Elles sont rangées en suivant l'ordre alphabétique des Référence correspondantes mais sont identifiées par un numéro (ex : 16 pour Calcosol). Le numéro est obligatoire pour accéder directement à une table. La correspondance 'nom RP' et 'n°' est donnée dans les écrans liés au bouton 1 (*figures 2 et 3*).

Les tables des experts représentent au total 8 160 lignes et donc notes⁷.

Dans l'exemple présenté (*figure 7*) toutes sortes de types de Podzosols sont de bons candidats pour donner leur nom au sol examiné. Pour trancher, l'utilisateur utilise les méthodes indiquées plus haut : recours à l'aide en ligne, consultation du livret RP aux pages indiquées dans l'aide en ligne. En queue de liste, on laisse subsister tous les taxons candidats pourvu que la note moyenne obtenue par eux soit supérieure ou égale à 1,5. Un taxon crédité d'un tel score n'est certainement pas valable car on trouve mieux en général. Mais il est instructif d'afficher une liste de résultats pas trop courte. Par exemple, ici, les Néoluvissols n'ont pas été éliminés car eux aussi présentent un horizon blanchi.

RESULTATS ET DISCUSSION

Classol a été testé sur une version intermédiaire du logiciel qui ne comprenait ni les questions 17 et 18, ni certaines adaptations de détail dans la notation des experts. Les résultats obtenus sont donc certainement un peu inférieurs à ceux que l'on obtiendrait maintenant. Mais, nous n'avons pas jugé utile de tout recommencer, car, de toute manière, le logiciel reste en perpétuelle évolution. En fait, nous avons déjà tenu compte, dans CLASSOL, des remarques que les essais ci-dessous ont permis de dégager.

⁷ Après avoir hésité, nous n'offrons pas à l'utilisateur la possibilité de modifier le système de notation (bouton 3) ou la présentation des taxons (bouton 1) même si nous sommes persuadés de la nécessité d'améliorer encore le système. Il faut en effet éviter que circulent trop de versions disparates de Classol. L'Unité INFOSOL (INRA-Orléans) a accepté de maintenir le logiciel et d'affiner les tables en fonction des remarques qui seront faites par les utilisateurs.

Test de Classol

Pour tester le système nous avons demandé à 8 collègues, spécialisés ou non en pédologie, de classer 5 sols que nous leur avons fournis (descriptions + analyses) après avoir soigneusement masqué le nom de ces sols pour qu'il n'apparaisse pas sur leur fiche. Pour avoir des répétitions, nous nous sommes contentés de 10 profils⁸ au total si bien que chacun a été classé 4 fois. Les testeurs ont été servis au hasard mais de telle façon qu'ils aient bien 5 sols différents à identifier parmi les 10 possibles. Les sols à reconnaître sont des profils considérés comme typiques et figurant dans des notices de cartes pédologiques à 1/100 000. Anciens, ils sont identifiés dans le système CPCS si bien que nous les avons repositionnés dans le RP (*tableau 3*).

Les résultats obtenus sont présentés dans le *tableau 4*. Ils ont été synthétisés de la manière suivante. Dans l'idéal, chaque Référence à reconnaître devrait être reconnue c'est-à-dire arriver en tête dans l'essai correspondant matérialisé par l'obtention d'une liste telle que présentée plus haut (*figure 7*). Ce n'est pas toujours le cas. Dans le *tableau 4*, on donne donc le rang de classement de la Référence, dans chacun des quatre essais la concernant, ceci pour les sommes de notes et pour la moyenne. Par exemple, si on suppose qu'un Podzosol ocrique devait être identifié (cas de la *figure 7*), les rangs obtenus sont alors 4 (somme de notes - trois autres taxons font mieux) et 3 (note moyenne). Dans le *tableau 4*, on utilise le signe --- lorsque le bon nom de sol ne fait pas partie du classement final.

Les quarante essais (8 expérimentateurs x 5 sols à reconnaître) amènent les résultats suivants :

- c'est seulement dans 25 cas (somme de notes) ou 27 cas (moyenne de notes) sur 40 cas au total que le nom du sol réellement à découvrir figure dans la liste des taxons sélectionnés par le logiciel ;
- quand le nom à trouver est proposé dans la liste, il figure généralement en bonne place puisque son rang moyen se situe alors entre 2 et 3 (somme de notes) ou même vaut 2 (moyenne des notes) ; en d'autres termes, les résultats sont ou bons (dans $26/40 = 65\%$ des cas) ou bien tout à fait mauvais !
- ce qui vient d'être dit en a et b montre que, dans le cadre de nos essais, le critère « moyenne de notes » s'est révélé un peu meilleur que « somme de notes » mais il est trop tôt pour juger la différence statistiquement significative.
- certaines sols sont moins bien identifiés que d'autres, en particulier l'Organosol insaturé n'a été reconnu par personne.

Discussion

Les cas les plus intéressants à discuter sont évidemment ceux qui correspondent à un échec total.

⁸ Il s'agit bien ici de profils, décrits comme tels, et non pas de « Solums » concept introduit bien ultérieurement dans le cadre du RP.

Figure 3 - Description d'une Référence dans Classol*Figure 3* - Description of a soil type in the software help

- *L'Organosol insaturé* n'a pas été reconnu (sauf par nous-mêmes en utilisant le logiciel!). En fait, les testeurs de Classol ont été influencés par un découpage du sol, au-dessus du C, en trois horizons⁹. Ils ont donc décrit le profil comme « fortement différencié », ce qui a éliminé à la fois les *Organosols* et les *Rankosols* voi-

sins. Cela veut dire que la notion de différenciation devait être mieux expliquée dans l'aide en ligne. Ceci a été corrigé depuis l'exercice.

- Le *Calcisol*, reconnu seulement une fois sur quatre, a posé un autre problème. Il comprenait trois horizons au-dessus de la roche

Tableau 3 - Types de sols ayant servi à tester CLASSOL .*Table 3* - Group of soils that were used to test the software

N°	Type génétique décrit		Origine des données :		
	(CPCS)	Référence à reconnaître (RP)	n° de profil,	page de description,	notice
1	Sol brun calcaire	Calcisol	n° 323,	page 43,	Brive
2	Sol brun acide	Alocrisol	n° 450,	page 77,	Brive
3	Andosol vitrique	Vitrosols	n° 113,	page 171,	Privas
4	Salé hydromorphe (palus)	Thalassosols	n° 19,	page 107,	Lesparre
5	Podzol humo-ferrugineux	Podzosol humo-durique	n° 531,	p. 128,	Saint-Dié
6	Ocre-podzolique	Podzosol ocrique	n° 18,	page 64,	Saint-Dié
7	Lessivé glossique	Luvisol dégradé	n° 16,	page 59,	Saint-Dié
8	Brun calcique	Calcisol	n° 1,	page 24,	Saint-Dié
9	Ranker humifère épais	Organosol insaturé	n° 34,	page 46,	Vichy
10	Vertisol hydromorphe	Topovertisol	n° 107,	page 49,	Vichy

Sources bibliographiques : Vichy (Favrot, 1969), Brives (Bonfils, 1976), Privas (Bornand et al., 1977), Lesparre (Wilbert, 1978), Saint-Dié (Bonneau et al., 1978)

⁹ L'auteur de la description avait eu à cœur de présenter le détail du développement d'un profil destiné à illustrer une notice de la carte de France à 1/100 000...

Tableau 4 - Rangs de classement des sols à reconnaître dans la liste des sols candidats. Chaque valeur du tableau correspond à un des quarante essais (somme de notes d'un côté, moyenne de l'autre).

Table 4 - Ranks of the soils that have to be recognized as they appear in the lists of the matching taxa. Each value in the table corresponds to one of the forty trials. In the ideal, the rang is one (the good soil is proposed as the more probable).

N°	Type à reconnaître	Rang de classement des sols :							
		Selon la somme des notes				Selon la moyennes de notes			
1	Calcosol	3	3	2	2	3	1	1	2
2	Alocrisol	6	5	4	2	---	4	2	4
3	Vitrosol	1	1	---	1	1	1	4	1
4	Talassosol	5	---	---	9	1	1	4	1
5	Podzosol humo-durique	1	1	2	1	1	1	2	2
6	Podzosol ocrique	1	2	2	---	1	3	1	---
7	Luvisol dégradé	---	---	---	2	---	---	---	6
8	Calcisol	---	---	---	1	---	---	---	2
9	Organosol insaturé	---	---	---	---	---	---	---	---
10	Topovertisol	1	1	---	1	1	1	---	2

mère si bien que la notion « d'horizon médian » utilisée alors dans les questions était douteuse. Certains observateurs ont considéré comme « médian » le BC qui était calcaire. Dès lors, l'appellation Calcisol ne pouvait être retenue. C'est pourquoi nous avons mieux précisé dans Classsol quels horizons il s'agit de décrire (A, B et pour partie E).

- Le *Luvisol dégradé* a lui aussi été reconnu une seule fois sur quatre. Ce sol, très typique et accompagné d'un schéma particulièrement bien fait, aurait été immédiatement identifié par un pédologue expérimenté. Deux des testeurs n'ont pas reconnu un « horizon intermédiaire blanchi » alors que le schéma mettait en évidence un horizon qualifié de « A₂ » et caractérisé par une couleur 10YR 7/4 (le code n'était pas traduit en clair). Le troisième testeur n'a pas mentionné le « pseudogley » qui affectait le profil. Il n'y avait pas d'indication à ce sujet dans le texte de la description. Mais le schéma faisait apparaître des taches et deux des horizons avaient leur symbole accompagné d'un « g ».

- L'*Alocrisol* est mal classé. C'est à notre avis un défaut difficile à corriger dans le logiciel : tous les sols extériorisant peu de caractères très typiques (Brunisols, Calcosols, Alocrisols) sont plus difficiles à reconnaître. Mais c'est bien la même chose en l'absence d'ordinateur !

CONCLUSION

Le travail que nous avons réalisé se traduit par des résultats encourageants. Déjà, le logiciel apporte une aide sérieuse à l'identification de trois sols sur quatre. Dans le futur, d'autres essais d'utilisation feront certainement apparaître des insuffisances dans le système de notation qu'il faudra améliorer. En outre, il faudra peut-être

augmenter le nombre des questions et en revoir certaines. Chaque fois qu'une modification sera introduite, le numéro de version sera incrémenté. Les personnes se procurant le système (gracieusement) pourront donc vérifier qu'elles disposent de la dernière mouture disponible.

Mais nos essais ont aussi montré la limite fondamentale de l'approche. Au départ, nous avons cru possible de mettre le logiciel entre toutes les mains, l'expertise étant limitée à celle des concepteurs du système, tables de notation comprises. En réalité, il apparaît bien que pour répondre aux questions, il faut un minimum de culture pédologique même en disposant de l'aide en ligne. Par exemple, il faut savoir lire des résultats d'analyse c'est-à-dire examiner des textures, des couleurs, des teneurs en carbonates ou en fer pour déterminer si un profil est différencié ou non. Dans ces conditions les candidats à l'identification d'un sol, étudiants, chercheurs des autres disciplines ou techniciens des instituts spécialisés, ne pourront sans doute pas se passer de l'aide de spécialistes...

D'un autre côté, le logiciel est un bon moyen pour apprendre la terminologie, au moyen de l'aide en ligne, en l'utilisant alors comme une sorte de jeu. Nous pensons aux étudiants ayant des modules de science du sol dans leur cursus.

Enfin, notre incursion dans les systèmes experts révèle la liste des problèmes nouveaux qui se présentent dans un tel contexte : logique du système ?, complexité ? degré d'érudition supposé chez l'utilisateur ?, droit à l'erreur accordé à l'utilisateur ?, degré de sévérité idéal chez l'expert ? La réflexion amorcée pourra servir à d'autres tentatives, même d'essence différente.

En France le Référentiel Pédologique est largement utilisé. Mais au plan international, c'est la *Word Reference Base* (WRB) qui prime (WRB, 1998). Or RP et WRB ont des structures voisines

Figure 4 - Ecran « roche » (première question posée)

Figure 4 - Screen « rock » corresponding to the first question put

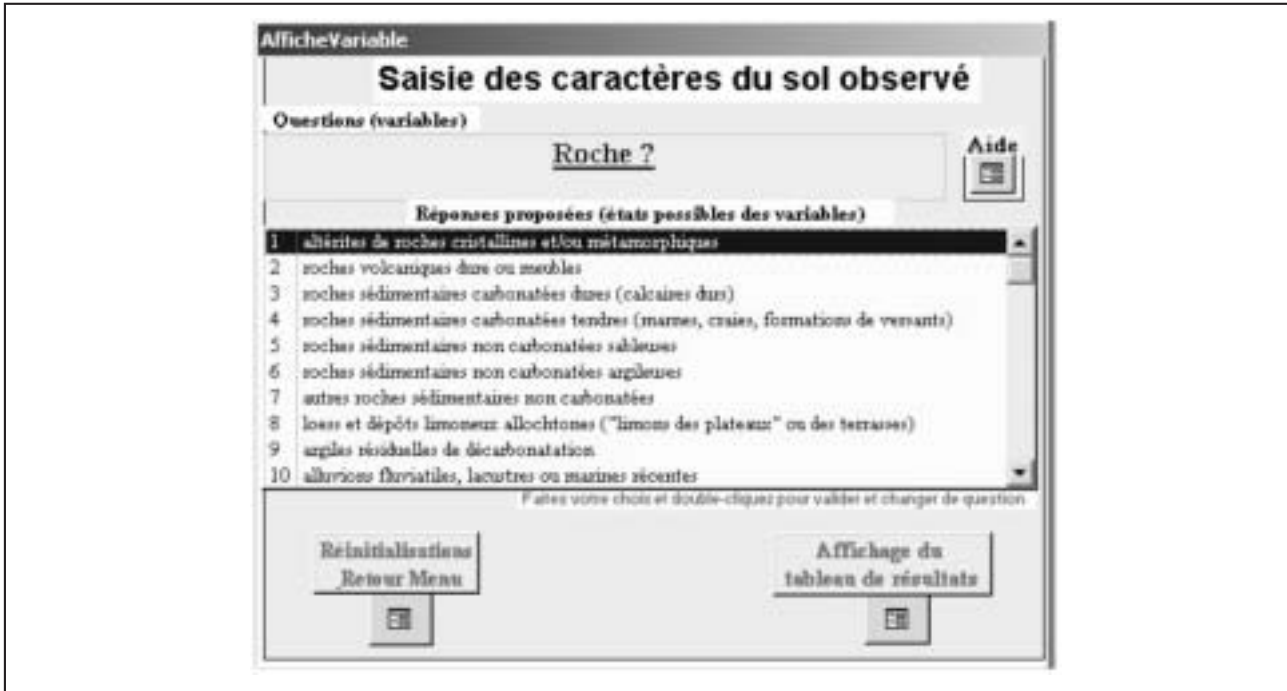


Figure 5 - Rappel des réponses déjà données par l'utilisateur

Figure 5 - List of the answer already given by the user

Variables_saisies		
1	Roche ?	1 altérites de roches cristallines et/ou métamor
2	Différenciation ?	2 profil fortement différencié
3	Horizon intermédiaire Bt	1 oui
4	Accumulation de [c] de	2 Fe
5	Épaisseur ?	3 de 40 à 200cm
6	Efficacité des horizo	1 null
7	Efficacité des horizo	1 null
8	Couleur des horizons B	1 noir
9	Couleur des horizons (B)	5 rouge
10	Texture des horizons de	2 Sableuse
11	Texture des horizons (B)	2 Sableuse
12	Structure des horizons d	1 agrégats peu sets
13	Structure des horizons (l)	1 agrégats peu sets

(Legros, 2001). Dans ce contexte, l'adaptation de Classol au WRB devrait être possible sans difficultés majeures.

Si le succès est au rendez-vous (nombreux utilisateurs) Classol sera transporté sur le net et traduit en anglais.

REMERCIEMENTS

Nous remercions ici le Ministère de l'Agriculture et de la Pêche (MAP) dont la Direction Générale de la Forêt et des Affaires Rurales (DGFAR) a débouqué les crédits nécessaires à ces travaux.

Nous remercions aussi les collègues qui ont bien voulu nous aider à tester le logiciel mis au point, par ordre alphabétique : Jean-Pierre Barthès, Guillaume Coulouma, Fabienne Déjean, François Garnier, Noëlle Guix, Olivier Hébrard, Gwen Trotoux, Jean-Marc Robbez-Masson, Patrick Zante.

BIBLIOGRAPHIE

- Albertch C., Huwe B., Jahn R., 2004 - Comparison of approaches for Automated Soil Identification. Workshop Digital Soil Mapping. Montpellier 2004, IUSS, 15 p.
- Baize D., 1986.- Couvertures pédologiques, cartographie et taxonomie. Science du Sol n° 3, 24, pp. 227-243.
- Baize D. et Girard M.C. (Coordonnateurs) et coll. 1995 - Référentiel Pédologique 1995. - INRA Éditions. 332 p.

Figure 6 - Organisation du système de notation pour un taxon (Référence)

Figure 6 - Organisation of the expert system of notation

	nomsol	nomvar	note	modalité
▶	16/Calcosols	Roche	00	altérites de roches cristallines et/ou métamorphes
	16/Calcosols	Roche	-99	roches volcaniques durs ou meulés
	16/Calcosols	Roche	2	roches sédimentaires carbonatées dures (calcaire)
	16/Calcosols	Roche	2	roches sédimentaires carbonatées tendres (marne)
	16/Calcosols	Roche	-99	roches sédimentaires non carbonatées sableuses
	16/Calcosols	Roche	-99	roches sédimentaires non carbonatées argileuses
	16/Calcosols	Roche	-99	autres roches sédimentaires non carbonatées
	16/Calcosols	Roche	1	loess et dépôts limoneux alloctones ("limons")
	16/Calcosols	Roche	-99	argiles résiduelles de décarbonatation
	16/Calcosols	Roche	1	alluvions fluviales, lacustres ou marines récentes
	16/Calcosols	Différenciation	2	profil peu différencié
	16/Calcosols	Différenciation	-1	profil fortement différencié
	16/Calcosols	lit intermédiaire blanchi	-99	oui

Figure 7 - Exemple de présentation des résultats du diagnostic de Classsol lors de la recherche d'identification d'un sol.

Figure 7 - Example of result i.e list of taxa that match more or less with the answer to the questions when one searches for a soil name

Num sol	Nom du Sol (Référence)	Nombre de notes	Points obtenus	Note moyenne	Aperçu du sol
72	Podzols Meubles	13	26	2,00	☺
74	Podzols Placiques	16	30	1,88	☺
73	Podzols Onctueux	13	23	1,77	☺
68	Podzols Durques	13	22	1,69	☺
71	Podzols Humo-Du	13	20	1,54	☺
51	Néoluvols	16	24	1,50	☺
					☺

Baize D., Jabiol B., Gobat J.M., 2004. Le Référentiel Pédologique. Premier bilan au bout de 11 années. Etude et Gestion des Sols, vol. 11, 2, pp. 149-164.

Bockheim J.G., Gennadiyev A.N., Hammer R.D., Tandarich J.P., 2004. Historical development of key concepts in pedology. Geoderma, (sous presse).

Bonfils P., 1976. Carte Pédologique de France à 1/100 000. Feuille de Brive, Notice, INRA-SESCPF, 135 p.

Bonneau M., Faivre P., Gury M., Hétier J.M., Le Tacon F., 1978. Carte Pédologique de France à 1/100 000. Feuille de Saint-Dié, Notice, INRA-SESCPF, 159 p.

Bornand M., Legros J.P., Moinereau J., 1977. Carte Pédologique de France à 1/100 000. Feuille de Privas, Notice, INRA-SESCPF, 255 p.

Dale M.B., McBratney A.B., Russel J.S., 1989. On the role of expert systems and numerical taxonomy in soil classification. Journal of Soil Science, 40, pp. 223-234.

Favrot J.C., 1969. Carte Pédologique de France à 1/100 000. Feuille de Vichy, Notice, INRA-SESCPF, 158 p.

Galbraith J.M., Bryant R.B., Ahrens R.J., 1998 - An expert system for soil taxonomy. Soil Science 163(9), pp. 749-758

Girard M.C., 1976. Recherche d'une méthodologie pédologique en matière de traitement statistique des données de sol. Application à la taxonomie et à la cartographie. SdS n° 3, pp. 177-201.

Girard M.C., 1983. Recherche d'une modélisation en vue d'une représentation spatiale de la couverture pédologique. Thèse d'Etat, Université de Paris 7, 295 p + annexes.

Legros J.P., 1996. Cartographies des sols. De l'analyse spatiale à la gestion des territoires. Presses Polytechniques et Universitaires Romandes. 321 p, 144 tableaux et figures, index.

Legros J.P., 2001. French classification systems. Encyclopedia of soil science

- ce (encyclopédie « on line »), Marcel Decker Edit (www.dekker.com).
- Mazaheri S.A., Koppi A.J., McBratney A.B., 1995. A fuzzy allocation scheme for the Australian Great Soil Groups Classification system. *European Journal of Soil Science*, December, 46, pp. 601-612.
- Pavat J.L., 1986. Contribution à l'étude de la ressemblance entre types de sols. Application aux secteurs de référence. DEA, ENSAM-USTL, 75 p + annexes.
- Simonneaux V., 1987. Mesure de la ressemblance entre des groupes de sondages à la tarière et des profils de référence. DEA INAPG-INRA, 61 p + annexes.
- Wilbert J., 1978. Carte Pédologique de France à 1/100 000. Feuille de Lesparre, Notice, INRA-SESCPF, 229 p.
- WRB, 1998. World Reference Base for Soil resources. ISSS, ISRIC, FAO. FAO report n°84, 1998, 88 p.