

Le programme Base de Données des Analyses de Terre (BDAT): Bilan de 20 ans de collecte de résultats d'analyses

N.P.A. Saby^(1*), B. Lemerrier⁽²⁾, D. Arrouays⁽¹⁾, S. Leménager⁽³⁾, B.P. Louis^(2, 4), F. Millet⁽¹⁾, J.-B. Paroissien⁽¹⁾, E. Schellenberger⁽¹⁾, H. Squidadant⁽²⁾, C. Swiderski⁽¹⁾, B. Toutain⁽¹⁾, C. Walter⁽²⁾ et M. Bardy⁽¹⁾

1) INRA, US1106 Unité Infosol, F-45075 Orléans, France.

2) AGROCAMPUS OUEST, UMR1069 SAS, 65, rue de Saint-Brieuc CS 84215 - F-35042 Rennes, France

3) DGPAAT/SSADD/SDBE/BSE Ministère en charge de l'agriculture, F-75000 Paris, France

4) INRA, UMR1069 SAS, 65 rue de Saint-Brieuc CS 84215 - F-35042 Rennes, France

*: Auteur correspondant: nicolas.saby@orleans.inra.fr

RÉSUMÉ

Le programme Base de Données des Analyses de Terre (BDAT) regroupe depuis vingt ans les résultats d'analyses d'horizons de surface de sols cultivés, effectuées sur l'ensemble du territoire national, à la demande d'agriculteurs, par des laboratoires agréés par le Ministère en charge de l'agriculture. Cet article a pour objectif de présenter les derniers résultats produits par ce programme. Il aborde successivement la présentation de l'organisation du programme, des développements méthodologiques connexes et des principaux résultats sur le statut et l'évolution des propriétés des horizons de surface des sols cultivés. A ce jour, la BDAT recense 22830147 résultats analytiques provenant de 1962238 échantillons récoltés sur la période 1990-2009. L'analyse de ces données permet par exemple de mettre en évidence à l'échelle nationale une forte hétérogénéité spatiale de la richesse en phosphore assimilable, des baisses des teneurs en carbone des sols initialement les plus pourvus et d'une hausse généralisée des pH des sols non calcaires. Cependant, les biais statistiques inhérents à la stratégie d'échantillonnage adoptée peuvent être importants et difficilement quantifiables. Des précautions doivent être prises pour interpréter les résultats issus d'analyses d'une telle base de données. Cependant, le programme BDAT constitue une source d'informations importante sur la variabilité des propriétés des horizons de surface des sols cultivés et les résultats statistiques agrégés sont aujourd'hui librement disponibles sur internet (<http://bdat.gissol.fr>).

Mots clés

Analyses de Terre, base de données, qualité des sols, surveillance, cartographie.

SUMMARY**THE FRENCH SOIL-TESTING DATABASE**

The French Soil-testing database (FSTD) gathers since twenty years the results of the soil-tests realized by commercial soil-testing laboratories approved by the Ministry of Agriculture. The objectives of this paper are three-fold: to present (i) the global gathering procedure, (ii) the inherent methodological development and (iii) the important results on the spatio-temporal monitoring of soil properties. After quality checks, analytical results from 1,962,238 samples corresponding to 22,830,147 soil-tests results are available in the FSTD. The results of the statistical analyses of these data have shown for example a large spatial heterogeneous variation of available phosphorous content, a decrease of soil carbon content or, an increase in pH of the non-calcareous soils. However, care is needed when interpreting the results from such a soil-testing database. The uncontrolled sampling strategy can lead to statistical bias difficult to assess, like heterogeneity in spatial distribution of analyses, or variability of farmers' motivations to perform soil analysis. However, the FSTB appears a relevant tool to assess the variation of the properties of the cultivated soil. Most of the statistical aggregated results are available on the web.

Key-words

Soil test, database, soil quality, monitoring, mapping.

RESUMEN**EL PROGRAMA BASE DE DATOS DE ANÁLISIS DE TIERRA (BDAT): balance de 20 años de recogida de resultados de análisis**

El programa Base de Datos de Análisis de Tierra (BDAT) reagrupa desde veinte años los resultados de análisis de horizontes de superficie de suelos cultivados, efectuados en todo el territorio nacional, a la demanda de agricultores, por laboratorios homologados por el ministerio en carga de la agricultura. Este artículo tiene como objetivo presentar los últimos resultados producidos por este programa. Aborda sucesivamente la presentación de la organización del programa, desarrollos metodológicos conexos y principales resultados sobre el estatuto y la evolución de propiedades de los horizontes de superficie de suelos cultivados. Hasta la fecha, la BDAT identifica 22830147 resultados analíticos que proviene de 1962238 muestras recogidas durante el periodo 1990-2009. El análisis de estos datos permite por ejemplo poner en evidencia a escala nacional una fuerte heterogeneidad espacial de la riqueza en fosforo asimilable, disminuciones de contenidos en carbono de suelos inicialmente los más dotados y un aumento generalizado de los pH de suelos no calcáreos. Sin embargo, los sesgos estadísticos inherentes a la estrategia de muestreo adoptada pueden ser importantes y difícilmente cuantificables. Precauciones deben ser tomadas para interpretar los resultados resultantes de análisis de una base de datos de este tipo. Sin embargo, el programa BDAT constituye una fuente importante de informaciones sobre la variabilidad de propiedades de horizontes de superficie de suelos cultivados y los resultados estadísticos agregados están hoy libremente disponibles en internet (<http://bdat.gissol.fr>).

Palabras clave

Análisis de tierra, base de datos, calidad de los suelos, vigilancia, cartografía.

L'importance du rôle des sols dans l'environnement et la nécessité de leur préservation sont aujourd'hui reconnues par tous. Les Etats ont ainsi été poussés à développer des programmes de surveillance de la qualité de leurs sols (Morvan *et al.*, 2007). Ils ont pour objectif la détermination systématique de variables du sol de façon à enregistrer leur évolution spatiale et temporelle (FAO/ECE, 1994). Même si la mise en œuvre des programmes varie entre les pays, un point commun reste que ces programmes reposent sur des dispositifs d'observations à la fois coûteux et lourds à gérer. D'autres sources d'informations sur les sols doivent donc être explorées.

L'activité agricole génère un grand nombre de résultats d'analyses de terre, utilisés principalement pour le pilotage de la fertilité des sols des parcelles. Qui plus est, ces analyses sont réalisées par des laboratoires agréés par le Ministère en charge de l'Agriculture, garantissant la compatibilité de leurs résultats et permettant leur regroupement dans une base de données unique. Historiquement, la première initiative de synthèse d'analyses de terre en France a été réalisée en Bretagne (Leleux *et al.*, 1988). Elle a permis de rassembler plus de 340000 résultats de déterminations analytiques correspondant à 71000 échantillons de sols. Cette étude a montré la faisabilité de produire des synthèses cartographiques à partir d'un grand nombre de données. L'idée a été reprise ensuite et étendue au territoire national par Walter *et al.* (1997) dans le cadre d'un partenariat avec l'Association Française d'Etude des Sols (AFES). Ce travail a permis de collecter les résultats de 330000 échantillons sur la période 1990-1995. Avec la création du Groupement d'Intérêt Scientifique (GIS) Sol en 2001, la collecte des résultats d'analyses de terre a été pérennisée et confiée à l'unité InfoSol de l'INRA d'Orléans, dans le cadre du programme financé par le Ministère en charge de l'Agriculture et appelé « Base de Données d'Analyses de Terre » (BDAT) (Saby *et al.*, 2004). Ce type d'approche se rencontre également dans d'autres pays. On trouve ainsi des exemples sur différents paramètres en Angleterre (Skinner and Todd, 1998), dans l'Est de la Caroline du Nord (Cahoon and Ensign, 2004), en Nouvelle-Zélande (Wheeler *et al.*, 2004) et aux Pays-Bas (Reijneveld *et al.*, 2009).

Reposant plus sur une démarche d'enquête que sur une prospection pédologique classique, le programme BDAT est aussi le support à des recherches méthodologiques innovantes pour l'exploitation statistique de ce type d'informations. L'objectif de cet article est de présenter les derniers résultats du programme BDAT à travers la description de l'organisation du programme, des développements méthodologiques connexes et des principaux résultats sur le statut et l'évolution des propriétés des horizons de surface des sols cultivés.

LE PROGRAMME BDAT DU GIS SOL

Le partenariat

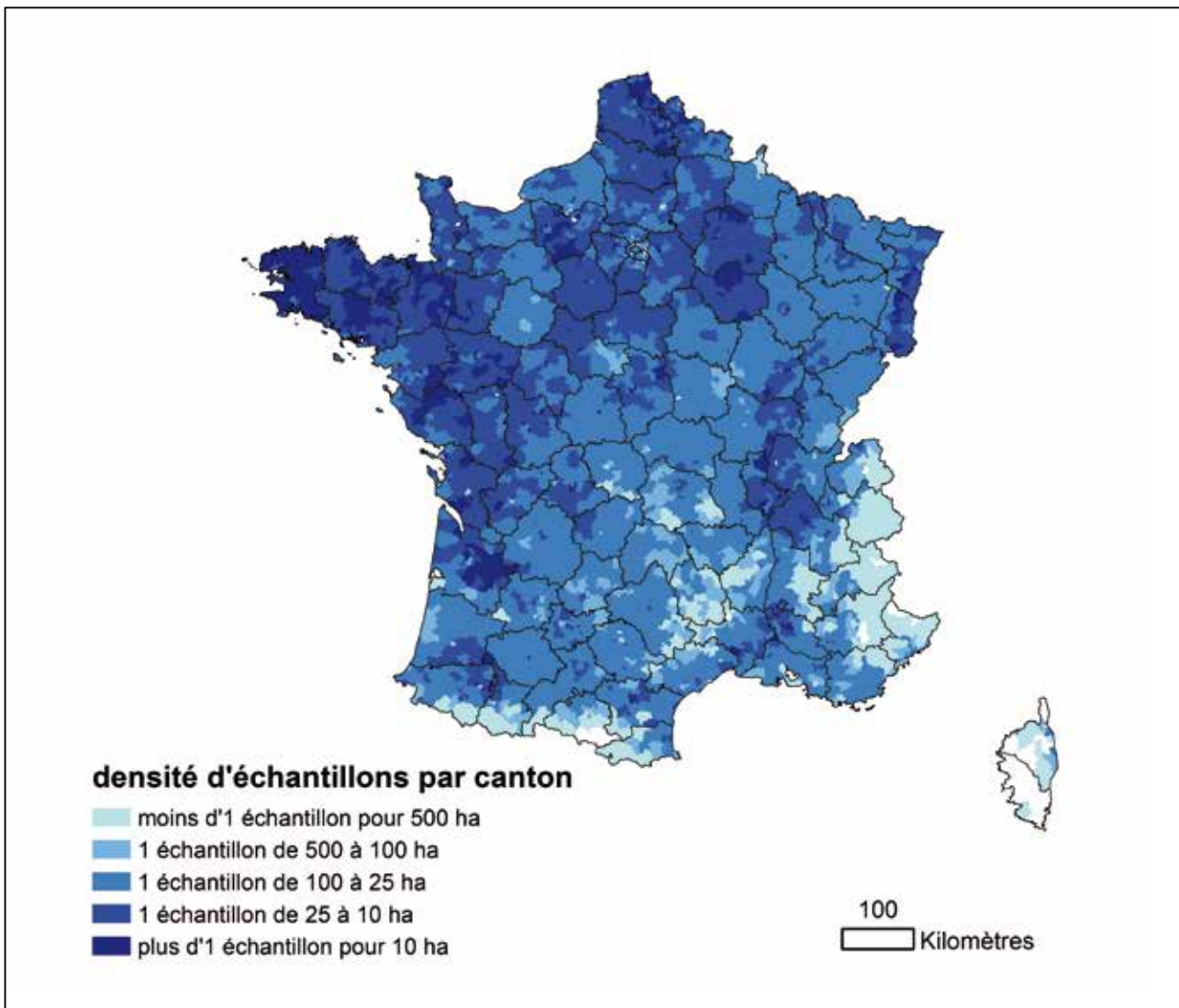
Les résultats d'analyses collectés dans le cadre de ce programme proviennent de laboratoires d'analyses de sol ayant satisfait à une procédure d'agrément proposée par le Ministère en charge de l'Agriculture et dont la liste est fournie par arrêté publié au journal officiel en fin d'année civile. En moyenne, 80 % des laboratoires collaborent, ce qui représente une trentaine de partenaires. Depuis 2001, le partenariat est régi par une convention réglant les termes du transfert des droits d'utilisation des données. Cette convention tripartite implique l'INRA Val de Loire (unité InfoSol), le Ministère en charge de l'Agriculture et le laboratoire fournisseur de données. Elle informe notamment sur les droits d'utilisation des données et stipule un certain nombre d'exigences auxquelles doivent répondre les résultats d'analyses fournis. Les principales sont les suivantes :

- la localisation: commune d'origine de l'échantillon de terre analysé,
- la présence d'au moins 5 déterminations analytiques dont le carbone organique ou l'azote total afin de pouvoir estimer le taux de matière organique,
- renseignement des métadonnées: méthode et unité d'expression de la mesure.

Une base des résultats d'analyses

Le programme BDAT a intégré les programmes du GIS Sol en 2001. Les résultats d'analyses provenant de plus de deux millions d'échantillons ont été ensuite adressés à l'unité InfoSol pour la période 1995-2009. Ces données analytiques ont subi un ensemble d'étapes de validation telles que décrites plus loin (Saby *et al.*, 2004). A l'issue de cette procédure, les résultats d'analyses issus de 1678061 échantillons se sont révélés conformes et ont pu être stockés dans la base de données. Il faut ajouter à ces effectifs les résultats de la première campagne qui a collecté des informations entre 1990 et 1995 (Walter *et al.*, 1997) et qui font grimper le nombre d'échantillons à 1962238 pour la période 1990-2009. Le taux de rejet moyen est proche de 22 %. Sa relative importance provient principalement (i) des difficultés à retrouver la commune d'origine de l'échantillon et (ii) de la présence d'échantillons de profondeur ou de sols non agricoles. La répartition spatiale de ces résultats est présentée dans la *figure 1*. On remarque que l'exhaustivité de la couverture du territoire est entachée par l'hétérogénéité de la densité des échantillons analysés. Celle-ci reste cependant suffisante pour produire des cartes des propriétés pédologiques au niveau national. La répartition temporelle est présentée dans la *figure 2* sous la forme de graphique en barres représentant les effectifs par année. La période 1995-2005 est la plus riche avec en moyenne 110000 échantillons par an. A l'hétérogénéité spa-

Figure 1 - Densité cantonale des échantillons de terre prélevés entre 1990 et 2009, dont les résultats sont intégrés au programme BDAT.
Figure 1 - Map of the density of the samples for which the soil test results achieved between 1990 and 2009 are consistent with the BDAT specifications.



tiale s'ajoute donc une hétérogénéité temporelle qu'il convient de prendre en compte dans les analyses statistiques.

Du point de vue des déterminations, les informations archivées dans la BDAT sont les suivantes:

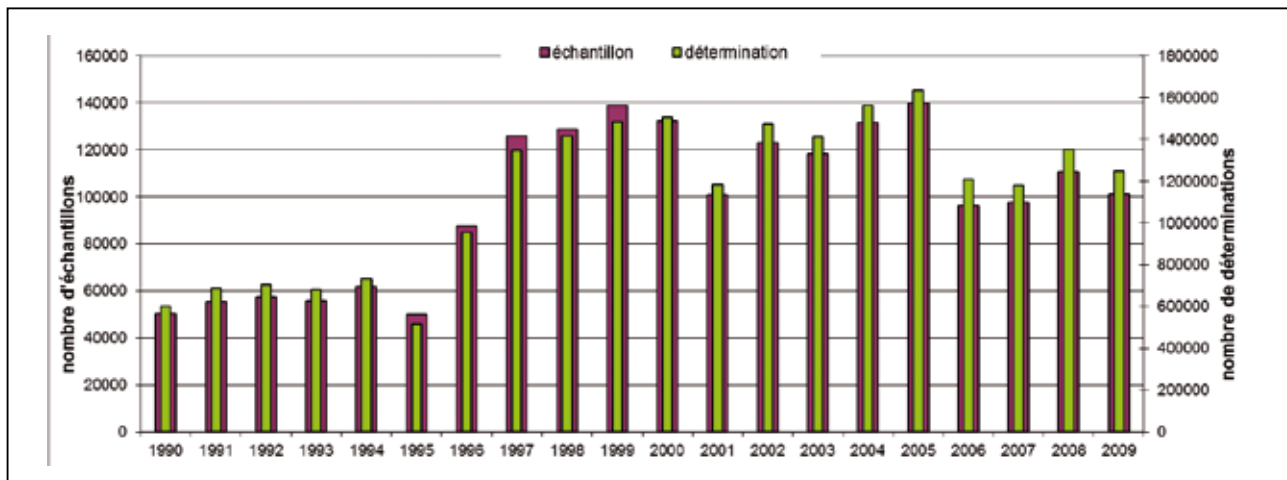
- la granulométrie (5 fractions) sans ou après décarbonatation;
- le pH (eau et KCl);
- le calcaire (total et actif);
- l'azote total;
- le carbone organique par oxydation humide ou par combustion sèche;
- le phosphore extractible par les méthodes Joret-Hébert, Dyer ou Olsen;

- le potassium, le magnésium, le calcium et le sodium échangeables;
- la Capacité d'Echange Cationique (CEC Metson ou méthode Riehm);
- le taux de saturation;
- les oligo-éléments (bore, cuivre, zinc, manganèse, fer), extraits à EDTA ou solubles à l'eau;
- une indication de l'origine géographique de la parcelle prélevée (code INSEE de la commune ou coordonnées GPS).

La somme de toutes les déterminations disponibles dans la base s'élève à 22830147 ce qui représente un taux de remplissage moyen de 36 %. Toutes les informations ne sont en effet

Figure 2 - Répartition annuelle du nombre d'échantillons (en rouge, axe de gauche) et de déterminations (en vert, axe de droite) des données de la BDAT.

Figure 2 - Annual distribution of the number of samples (in red) and analyses (in green) of the BDAT data.



pas systématiquement renseignées. Les effectifs disponibles varient fortement d'une détermination à l'autre (*figure 3*), mais aussi pour une même détermination selon la période de calcul considérée. D'une manière générale, les déterminations les plus courantes restent le carbone organique, les cations échangeables (sauf le sodium), le pH dans l'eau, la CEC Metson. La mesure de la CEC est devenue plus courante que celle du taux d'argile, ce que confirme la diminution progressive du nombre de déterminations granulométriques. Enfin, les coordonnées géographiques restent encore peu renseignées.

MÉTHODES D'ANALYSES STATISTIQUES

Les programmes de surveillance de la qualité des sols sont habituellement classés selon la stratégie d'échantillonnage sur laquelle reposent les observations (de Gruijter *et al.*, 2006). Arrouays *et al.* (2012) signalent que les réseaux d'observations construits à cet effet ont tendance soit à employer un plan d'échantillonnage systématique comme une grille (et c'est le cas du Réseau de Mesure de la Qualité des Sols (RMQS) en France), soit à utiliser un plan issu d'un tirage probabiliste. Pour les stratégies probabilistes, les sites d'échantillonnage sont choisis de manière indépendante de sorte qu'il n'est pas nécessaire de tenir compte des corrélations entre les observations. Les valeurs moyennes des propriétés du sol dans une région et les écart-types associés peuvent être estimés par des méthodes statistiques classiques dites design-based qui nécessitent peu d'hypothèses sur la distribution de la propriété. Lorsqu'on a recours à une approche de type systématique, les méthodes design-based peuvent être utilisées pour estimer la moyenne de la propriété. Cependant, le calcul de l'écart-type associé à l'estimation

nécessite le recours à une analyse dite model-based pour tenir compte de la corrélation entre les observations. Ces dernières approches peuvent également être utilisées pour produire des cartes de propriétés de sols. Les prévisions et les variances associées qui en résultent ne sont valables que si le modèle utilisé représente correctement la corrélation spatiale de la propriété.

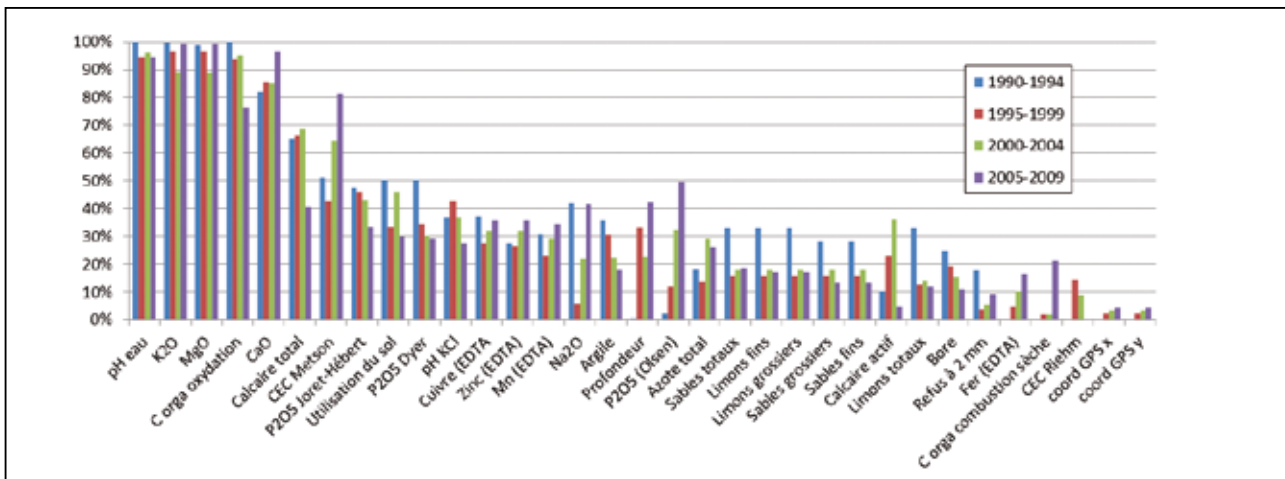
Dans le cas de la BDAT, les observations n'ont pas été collectées selon une stratégie nationale correspondant à une des deux approches statistiques présentées ci-dessus. Le risque majeur réside dans la présence de biais introduit par la sélection des sites. A titre d'exemple, il est probable que les parcelles présentant des risques de problèmes de fertilité soient préférentiellement analysées à la demande des agriculteurs. D'autre part, l'information est stockée dans la BDAT sur la base d'un support administratif (communes), dont la forme varie fortement d'une entité à l'autre, mais aussi dans le temps. Qui plus est, le découpage administratif peut être déconnecté des limites naturelles et de la distribution spatiale des grands types de sols. Un canton peut ainsi regrouper des sols très différents. Cela implique d'avoir recours à des méthodes statistiques adaptées pour la cartographie et la détection d'évolution. Le chapitre suivant présente les principaux travaux en la matière.

Production de cartographies

Une technique simple consiste à agréger les résultats d'analyses par entité géographique regroupant plusieurs communes (canton, petite région agricole...) pour en calculer et cartographier des paramètres décrivant la distribution statistique de la propriété du sol (moyenne, médiane...). Cette technique a déjà été mise en œuvre et a montré sa robustesse (Lemerrier *et al.*, 2006; Saby *et al.*, 2004; Walter *et al.*, 1997). Choisir une entité re-

Figure 3 - Effectifs disponibles par détermination, calculés sur des périodes de 5 ans entre 1990 et 2009.

Figure 3 - Number of samples available for each soil test calculated per 5 years period: 1990-1994, 1995-1999, 2000-2004, 2005-2009.



groupant plusieurs communes permet de maximiser le nombre de résultats d'analyses par entité spatiale d'agrégation, tout en garantissant le secret statistique de l'origine des données. Cependant, cette approche suppose que les localisations des échantillons ont été choisies aléatoirement. En d'autres termes, cette approche fait l'hypothèse d'ignorer la corrélation spatiale des échantillons. Pour tenir compte de cette dernière, il convient de mettre en œuvre une technique model-based. Il s'agit du modèle mixte appliqué à la géostatistique (Lark and Cullis, 2004; Saby et al., 2011). Néanmoins, dans le cadre de la BDAT, les coordonnées des échantillons sont rarement connues, rendant l'application de cette méthode délicate. L'information disponible revient à ce qu'on appellera ici les statistiques descriptives des communes (e.g. moyenne, médiane, écart-type et nombre d'échantillons). Orton et al. (2012a) proposent d'utiliser ces statistiques dans le cadre du modèle mixte grâce aux techniques de désagrégation spatiale. Ils montrent notamment comment une méthode basée sur la maximisation de la vraisemblance peut être utilisée pour estimer les paramètres du modèle de covariance spatiale. Ils démontrent également comment utiliser le meilleur prédicteur non biaisé ou krigeage pour calculer les prédictions spatiales. Cette approche permet ainsi de prendre en compte l'incertitude attachée aux données de la BDAT en (i) insérant explicitement dans le modèle statistique le nombre d'échantillons et l'écart-type et, (ii) tenant compte de la forme des polygones décrivant les cantons dans le calcul des corrélations spatiales. L'application à la cartographie du carbone organique en Franche-Comté a montré sa faisabilité (Orton et al., 2012a). Cette technique a récemment été comparée à d'autres approches à partir des données similaires récoltées aux Pays-Bas (Brus et al., 2014). Elle a alors confirmé tout son intérêt.

Mise en évidence d'évolutions temporelles

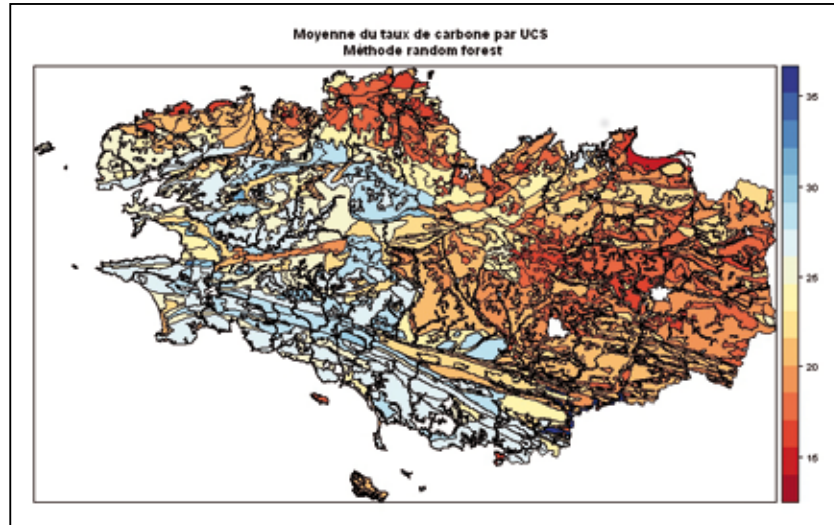
Lemercier et al. (2008) ont développé une méthode de mise en évidence d'évolution temporelle basée sur la comparaison deux à deux de populations de résultats d'analyses agrégées au niveau d'entités spatio-temporelles (cantons et périodes de 5 ans). Pour remédier aux biais issus de l'échantillonnage, les auteurs appliquent des techniques de ré-échantillonnage au sein des cantons permettant d'obtenir une résolution temporelle et spatiale constante et tenant compte de la surface agricole utile. Ils obtiennent ainsi des courbes de fréquences cumulées et des résultats de tests statistiques permettant de caractériser des évolutions. Une conséquence de cette approche est la multiplication des tests statistiques qui nécessite alors la prise en compte du risque de faux positifs (Chauveau et al., 2014). En se fondant sur une approche model-based, Orton et al. (2012b) utilisent un modèle de co-régionalisation pour cartographier et caractériser les évolutions.

Couplage avec d'autres bases de données

Les résultats analytiques stockés dans la BDAT ont récemment fait l'objet d'un couplage avec les données de la base du Référentiel Régional Pédologique (RRP) (Laroche et al., soumis). L'objectif de ce couplage est d'affecter les résultats analytiques de la BDAT non plus à des entités administratives, mais à des unités cartographiques cohérentes au niveau pédologique, et ainsi de mieux prédire certaines propriétés des sols en surface, les teneurs en carbone organique par exemple. Ce rapprochement repose sur le calcul d'une distance sémantique entre les données de la BDAT et les données des référentiels régionaux pédologiques. Plusieurs méthodes de rapprochement ont ainsi été testées correspondant à des calculs de distances par

Figure 3 - Estimation des valeurs moyennes (en g/kg) de carbone organique par UCS en Bretagne par affectation des analyses de la BDAT aux UTS et UCS du RRP par la méthode des forêts aléatoires.

Figure 3 - estimated mean values of organic carbon in g.kg⁻¹ per Soil Mapping Unit in Brittany, France. The estimation has been done using the random forest approach.



moyenne pondérée, la mise en œuvre d'ACP et l'ajustement de modèles de classification automatique de type « random forest », (Breiman, 2001). Des tests ont été mis en œuvre en Bretagne dans le cadre du programme « Sols de Bretagne » (Lemerrier *et al.*, 2010) et dans deux zones d'Alsace dans le cadre du projet ABC'Terre 'Atténuation du Bilan de gaz à effet de serre agricole et de Carbone organique des sols sur un TERRitoirE' (ABC-TERRE: Lauréat de l'appel à projets R&D REACTIF de l'ADEME en 2012). Les variables utilisées pour créer le lien sémantique devaient répondre à trois critères, i) être renseignées en nombre suffisant dans les deux bases de données et ii) ne pas avoir une évolution temporelle trop prononcée et iii) être suffisamment discriminantes pour les strates. Les variables retenues sont le pH, la capacité d'échange cationique (CEC), le taux d'argile, le taux de limons totaux, le ratio limons fins sur limons grossiers (pour la Bretagne seulement: caractérisation de l'origine autochtone ou allochtone des limons) et le calcaire total (pour l'Alsace seulement).

Bien qu'intéressante d'un point de vue méthodologie, cette méthode donne des résultats à l'heure actuelle trop mitigés pour être satisfaisants. Cependant, les estimations des valeurs de carbone organique par UCS sont cohérentes avec les connaissances a priori (*figure 4*). Les limites principales de ces méthodes de croisement de bases de données sont liées au fait qu'une grande part des données de la BDAT ne dispose pas d'information suffisante pour assurer l'affectation à une unité de sol. De plus, l'aspect temporel des données n'est pas pris en compte, or certaines variables d'affectation ou le carbone organique ont pu évoluer de façon significative au cours du temps.

UNE MINE D'OR SUR L'ÉTAT DES SOLS DE FRANCE

La BDAT est une source d'informations essentielle renseignant sur l'état des sols agricoles de France. Le rapport sur l'état des sols de France (Gis Sol, 2011) en témoigne en rassemblant ainsi un nombre important de résultats issus de la BDAT. Trois grands types de résultats présentés dans ce chapitre concernent 1) le statut des sols, 2) l'évolution temporelle des paramètres et 3) la mise en évidence de relation entre les paramètres. Trois exemples ont été retenus: le carbone organique, le phosphore extractible et le statut acido-basique.

Le carbone

Arrouays *et al.* (2006) ont montré l'impact de la distribution granulométrique sur les stocks de carbone organique présents dans l'horizon superficiel des sols cultivés français. En utilisant les valeurs des déciles supérieurs par classes texturales comme un estimateur du potentiel maximal de stockage de carbone, ils ont montré qu'il pourrait être plus efficace d'encourager des pratiques favorisant l'accumulation du carbone dans les sols ayant déjà des stocks importants, que de tenter d'augmenter les stocks dans les sols pauvres en carbone organique. Plus récemment, Angers *et al.* (2011) ont utilisé l'équation proposée par Hassink (1997) pour caractériser la variabilité spatiale du déficit de saturation des sols français en carbone organique. Ils ont ainsi démontré l'influence du climat, des propriétés du sol et de l'occupation sur le potentiel de stabilisation du carbone organique. Ils ont montré également qu'une forte proportion des sols agricoles présentait un potentiel à stocker du carbone supplémentaire. Enfin, une tendance à la baisse a été observée pour la teneur du sol en carbone dans plusieurs régions de France (e.g. Franche-Comté (Orton *et al.*, 2012b;

Saby et al., 2008); Bretagne (Lemerrier et al., 2008); Alsace (Swiderski et al., 2012)).

Le phosphore

Follain et al. (2009) ont conçu une méthodologie pour évaluer à partir des résultats de la BDAT la biodisponibilité du phosphore au niveau du territoire métropolitain. L'originalité de cette méthode tient dans l'interprétation agronomique des résultats quantitatifs issus de deux méthodes de mesures (Dyer et Joret-Hebert) à l'aide du logiciel RegiFert®. Ce dernier intègre dans son interprétation les caractéristiques du sol et les sensibilités des cultures à la disponibilité des nutriments. Cette étude a montré que, sur la période 1995-2000, 77 % des cantons présentent une majorité d'échantillons analysés situés en dessous du niveau critique à l'égard des besoins des cultures dites peu exigeantes en phosphore. En revanche, ces auteurs ont pu montrer que 23 % des cantons présenteraient une accumulation de phosphore dans le sol en quantité supérieure aux besoins des cultures. Dans ce cas, le risque de pollution est largement augmenté. En Bretagne, une région d'élevage intensif, Lemerrier et al. (2008) ont également montré une forte accumulation de phosphore dans les sols agricoles. L'excédent cantonal moyen de phosphore du sol accumulé pendant plus de 24 ans s'élève à 763 kg P/ha de surface agricole utile. Cette augmentation des stocks est un facteur de risque accru de transfert de phosphore vers les masses d'eau et donc d'eutrophisation.

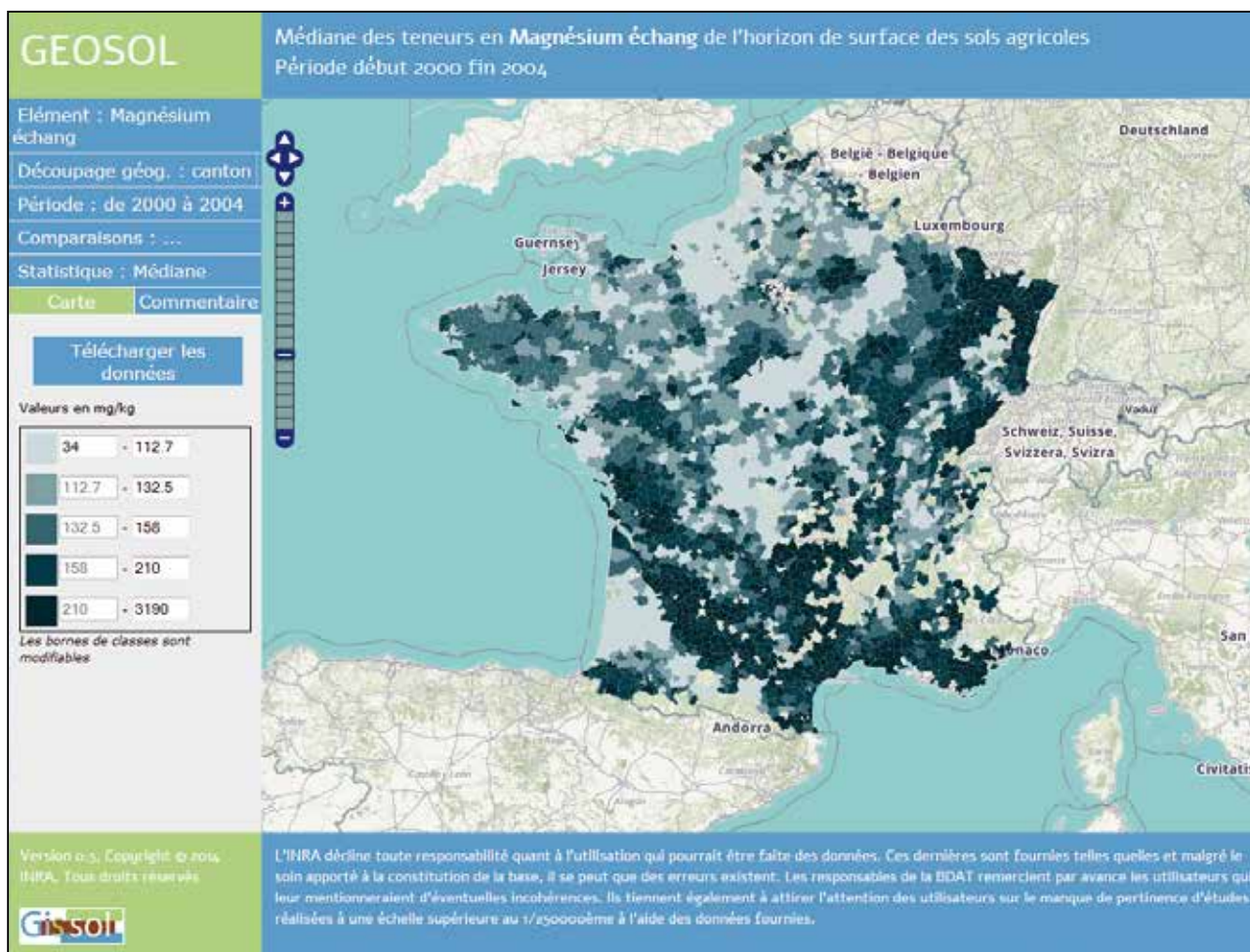
Le pH et les cations

Plus récemment, Swiderski et al. (2013) ont quantifié les évolutions temporelles observées pour le pH (mesuré dans l'eau) et le taux de saturation du sol sur la période 1996-2010. Cette quantification a consisté à comparer deux périodes (1996 à 1999 et 2006 à 2010) à l'échelle des petites régions agricoles. Il n'a pas été possible d'évaluer les tendances antérieures, car le protocole d'analyse du pH mesuré dans l'eau a subi une évolution en 1995. Les échantillons calcaires ont alors été préalablement exclus de cette analyse. Presque tous les changements importants se sont révélés être positifs, la plupart d'entre eux montrant une augmentation de pH mesuré dans l'eau de 0,2. Une tendance similaire a été observée pour le taux de saturation du sol. Cependant, pour une grande partie des cantons, aucune tendance n'a pu être mise en évidence, indiquant la nécessité de recueillir plus d'échantillons ou d'attendre plus de temps pour détecter des changements, s'ils existent. Les changements qui ont été observés pourraient être attribuables à une meilleure gestion des apports azotés ou à une diminution des dépôts atmosphériques en azote et en soufre.

UN SYSTÈME D'INFORMATION DÉCISIONNEL POUR LA DIFFUSION DES DONNÉES BDAT VIA UN SITE INTERNET ET DES SERVICES WEB

Un système d'information décisionnel (SID) rassemble des moyens, des outils et des méthodes permettant de transformer les données opérationnelles en informations pertinentes pour la prise de décision. D'après Inmon et al. (2008), le SID s'appuie sur l'entrepôt de données qui est une structure analogue à une base de données offrant une vision centralisée et la plus large possible de toutes les informations sur un sujet donné. A l'origine, cet outil a été créé pour le monde de l'entreprise afin de gérer ses données. L'entrepôt permet ainsi le stockage et l'analyse d'une grande quantité d'informations épurées, organisées, historisées (mécanisme par lequel on capture les données à chaque évolution de ces dernières) et provenant de plusieurs sources de données. Nous avons inscrit la chaîne de traitements de statistiques issues de la BDAT dans le contexte plus large de l'architecture décisionnelle du système d'information « sols ». Le premier type d'information produite correspond à la cartographie des distributions statistiques des propriétés du sol. Pour cela, les résultats d'analyses sont agrégés aux niveaux spatial et temporel en les regroupant à un grain plus grossier que la commune (régions, départements, cantons et petites régions agricoles) et selon une période temporelle de calcul qui est fixée à 5 ans pour les paramètres variables dans le temps comme le carbone organique, et 20 ans pour les paramètres plus stables comme le taux d'argile. Un ensemble de paramètres statistiques sur la distribution des propriétés du sol au sein de l'unité spatio-temporelle est calculé sur ces sous-populations: nombre de déterminations, moyenne, écart-type, déciles pour les propriétés ordinales, valeurs de la classe dominante et secondaire pour les propriétés cardinales. Le second type d'information stockée correspond à la comparaison temporelle des sous-populations des unités spatio-temporelles. L'entrepôt permet la mise en œuvre des calculs sur les résultats analytiques bruts (pH, carbone, phosphore, etc.), mais aussi sur des attributs issus d'un calcul, comme le ratio limons fins sur limons grossiers, l'interprétation agronomique du phosphore, le taux de saturation du sol ou les teneurs en carbone stabilisé par la fraction fine des sols (argile ou argile + limons fins) et calculées selon les équations d'Hassink (1997) et de Dexter (2008). L'ensemble des calculs sont ensuite stockés dans une structure, appelée « magasin de données » ou Datamart, sur laquelle repose la publication des cartes et des webservices. Ce magasin a vocation à être référencé et librement accessible en téléchargement sur un entrepôt international comme www.pangaea.de.

Plus de 2700 cartes sont ainsi générées à partir de l'entrepôt et consultables à la fois sur le site internet bdatt.gissol.fr, mais aussi à travers la publication de webservices (<http://gissol.fr/>). La figure 5 en présente les principales fonctionnalités.

Figure 5 - Organisation du site internet de la BDAT.**Figure 5** - Snapshot of the bdat site.

Le site permet de naviguer parmi les cartes et au sein d'une carte. Il offre la possibilité de télécharger la donnée de la carte sélectionnée sous la forme d'un tableur ou de l'imprimer. Un commentaire permet d'aider le lecteur dans son interprétation.

CONCLUSIONS

La collecte des résultats d'analyses de terre permet d'étudier la variabilité spatiale et temporelle des propriétés des sols fortement modifiées par l'activité humaine, ce que ne permettent pas les démarches pédologiques classiques. La démarche adoptée consiste, par voie d'enquêtes et à moindres frais, en une compilation de nombreux résultats analytiques datés et localisés, qui permettent d'obtenir une vision globale de l'ensemble d'un territoire et de dégager des tendances. Elle vient en complément de démarches expérimentales ou de programmes de cartographie

et de surveillance, mieux à même de caractériser les sols dans leur globalité et de déceler leurs évolutions.

La BDAT constitue ainsi l'un des 4 grands programmes du GIS Sol avec le Réseau de Mesure de la Qualité des sols (RMQS), l'Inventaire Gestion et Conservation des sols (IGCS) et la base des éléments traces métalliques (BDETM). Tout en capitalisant des informations précieuses sur les sols, ce programme a également permis le développement de méthodes statistiques innovantes comme les techniques de désagrégation spatiale ou de ré-échantillonnage. Cependant, il convient de limiter l'apport de cette approche à une description quantitative des variations de l'horizon de surface selon une résolution géographique qui reste assez grossière, mais qui présente l'avantage de couvrir de très larges surfaces du territoire, ce qui est un acquis très important et ouvre la voie à une meilleure prise en compte de la variabilité des sols dans des démarches d'évaluation des risques environnementaux impliquant ou concernant les sols.

REMERCIEMENTS

Le programme BDAT est financé par le Groupement d'Intérêt Scientifique Sol qui regroupe le ministère en charge de l'Agriculture, le ministère en charge de l'Environnement, l'ADEME, l'IGN, l'INRA et l'IRD. Nous remercions l'ensemble des laboratoires qui ont bien voulu participer à ce programme en fournissant les données. Nous tenons à remercier Joëlle Sauter et Stéphanie Jalabert pour leur relecture constructive de cet article.

BIBLIOGRAPHIE

- Angers, D.A., Arrouays, D., Saby, N.P.A., Walter, C., 2011 - Estimating and mapping the carbon saturation deficit of French agricultural topsoils. *Soil Use Manag* 27(4), 448-452.
- Arrouays, D., Marchant, B.P., Saby, N.P.A., Meersmans, J., Orton, T.G., Martin, M.P., Bellamy, P.H., Lark, R.M., Kibblewhite, M., 2012 - Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. *Pedosphere* 22(4), 456-469.
- Arrouays, D., Saby, N., Walter, C., Lemerrier, B., Schwartz, C., 2006 - Relationships between particle-size distribution and organic carbon in French arable topsoils. *Soil Use Manag* 22, 48-51.
- Breiman, L., 2001 - Random Forests. *Machine Learning* 45(1), 5-32.
- Brus, D.J., Orton, T.G., Walvoort, D.J.J., Reijneveld, J.A., Oenema, O., 2014 - Disaggregation of soil testing data on organic matter by the summary statistics approach to area-to-point kriging. *Geoderma* 226, 151-159.
- Cahoon, L., Ensign, S., 2004 - Spatial and temporal variability in excessive soil phosphorus levels in eastern North Carolina. *Nutr Cycl Agroecosyst* 69, 11-125.
- Chauveau, D., Saby, N.P.A., Orton, T.G., Lemerrier, B., Walter, C., Arrouays, D., 2014 - Large-scale simultaneous hypothesis testing in monitoring carbon content from French soil database - A semi-parametric mixture approach. *Geoderma* 219-220(0), 117-124.
- Dexter, A.R., Richard, G., Arrouays, D., Czyz, E.A., Jolivet, C., Duval, O., 2008 - Complexed organic matter controls soil physical properties. *Geoderma* 144(3-4), 620-627.
- FAO/ECE, 1994 - International Workshop on Harmonisation of Soil Conservation Monitoring System. In: FAO-FAO/ECE-RISSAC (Ed.), Budapest, Hungary, pp. 224.
- Follain, S., Schwartz, C., Denoroy, P., Villette, C., Saby, N.P.A., Arrouays, D., Lemerrier, B., Walter, C., 2009 - A method for assessing available phosphorus content in arable topsoils over large spatial scales. *Agron. Sustain. Dev.* 29(2), 371-379.
- Gis Sol, 2011. Synthèse sur l'état des sols de France - Groupement d'intérêt scientifique sur les sols, pp. 24.
- de Groot, J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006 - Sampling for Natural Resource Monitoring. Springer-Verlag.
- Hassink, J., 1997 - The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant Soil* 191(1), 77-87.
- Inmon, W.H., Strauss, D., Neuhoss, G., 2008 - DW 2.0: The Architecture for the Next Generation of Data Warehousing. Morgan Kaufmann Publishers Inc.
- Lark, R.M., Cullis, B.R., 2004 - Model-based analysis using REML for inference from systematically sampled data on soil. *Eur J Soil Sci* 55(4), 799-813.
- Laroche, B., Richer de Forges, A., Leménager, S., Arrouays, D., Schnebelen, N., Eimberck, M., Toutain, B., Tientcheu, E., Héliès, F., Chenu, J.P., Parot, S., Desbourdes, S., Girot, G., Voltz, M., Bardy, M., soumis - Le programme Inventaire Gestion Conservation des Sols de France: volet Référentiel Régional Pédologique. *Etude et Gestion des Sols*, 21, pp. 125-140.
- Leleux, A., Arousseau, P., Roudaut, A., 1988 - Synthèse cartographique régionale à partir de données d'analyses de terre. *Science du Sol* 26(1), 29-40.
- Lemerrier, B., Berthier, L., Buquin, G., Guichard, S., Le Bris, A.-L., Quidu, O., Walter, C., Cluzeau, D., 2010 - Programme Sols de Bretagne (2005-2010), AgroCampus Ouest, Rennes.
- Lemerrier, B., Gaudin, L., Walter, C., Arousseau, P., Arrouays, D., Schwartz, C., Saby, N.P.A., Follain, S., Abrassart, J., 2008 - Soil phosphorus monitoring at the regional level by means of a soil test database. *Soil Use Manag* 24(5), 131-138.
- Lemerrier, B., Walter, C., Schwartz, Saby, N., Arrouays, D., Follain, 2006 - Suivi des teneurs en carbone organique et en phosphore extractible dans les sols agricoles de trois régions françaises. Analyse à partir de la Base de Données des Analyses de Terre. *Etude et Gestion des Sols* 13(3), 165-179.
- Morvan, X., Richer de Forges, A., Arrouays, D., Le Bas, C., Saby, N., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Kibblewhite, M.G., Stephens, M., 2007 - Une analyse des stratégies d'échantillonnage des réseaux de surveillance de la qualité des sols en Europe. *Etude et Gestion des Sols* 14(4), 317-326.
- Orton, T.G., Saby, N.P.A., Arrouays, D., Walter, C., Lemerrier, B., Schwartz, C., Lark, R.M., 2012a - Spatial prediction of soil organic carbon from data on large and variable spatial supports. I. Inventory and mapping. *Environmetrics* 23(2), 129-147.
- Orton, T.G., Saby, N.P.A., Arrouays, D., Walter, C., Lemerrier, B., Schwartz, C., Lark, R.M., 2012b - Spatial prediction of soil organic carbon from data on large and variable spatial supports. II. Mapping temporal change. *Environmetrics* 23(2), 148-161.
- Reijneveld, A., van Wensem, J., Oenema, O., 2009 - Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. *Geoderma* 152(3-4), 231.
- Saby, N., Schwartz, C., Walter, C., Arrouays, D., Lemerrier, B., Roland, N., Squidant, H., 2004 - Base de Données des Analyses de Terre: procédure de collecte et résultats de la campagne 1995-2000. *Etude et Gestion des Sols* 11(3), 235-253.
- Saby, N.P.A., Arrouays, D., Antoni, V., Lemerrier, B., Follain, S., Walter, C., Schwartz, C., 2008 - Changes in soil organic carbon in a mountainous French region, 1990-2004. *Soil Use Manag* 24(3), 254-262.
- Saby, N.P.A., Marchant, B.P., Lark, R.M., Jolivet, C.C., Arrouays, D., 2011 - Robust geostatistical prediction of trace elements across France. *Geoderma* 162(3-4), 303-311.
- Skinner, R.J., Todd, A.D., 1998 - Twenty-five years of monitoring pH and nutrient status of soils in England and Wales. *Soil Use Manag* 14(3), 162-169.
- Swiderski, C., Saby, N.P.A., Arrouays, D., Lemerrier, B., Walter, C., Party, J.P., 2012 - Evolution des teneurs en carbone organique dans l'horizon de surface des sols cultivés en Alsace: Analyse à partir de la BDAT. *Etude et Gestion des Sols* 19(3), 179-192.
- Swiderski, C., Saby, N.P.A., Eveillard, P., Louis, B.P., Lemerrier, B., Walter, C., Arrouays, D., Bardy, M., 2013 - Évolution de la fertilité des sols d'après la Base de Données des Analyses de Terre (BDAT). In: COMIFER (Ed.), 11^e rencontre de la fertilisation raisonnée et de l'analyse, Poitiers.
- Walter, C., Schwartz, C., Claudot, B., Bouedo, T., Arousseau, P., 1997 - Synthèse nationale des analyses de terre réalisées entre 1990 et 1994 II. Descriptions statistique et cartographique de la variabilité des horizons de surface des sols cultivés. *Etude et Gestion des Sols* 4(3), 12-44.
- Wheeler, D.M., Sparling, G.P., Roberts, A.H.C., 2004 - Trends in some soil test data over a 14-year period in New Zealand. *New Zealand Journal of Agricultural Research*. 47, 155-166.