

Le programme de recherche ADEME “Bioindicateurs de l'état biologique des sols”

Ses objectifs, sa mise en œuvre et son déroulement

Guénola PERES¹, Antonio BISPO², Cécile GRAND³, Laurence GALSOMIES⁴

¹ Université de Rennes, UMR CNRS 6553 EcoBio, Station Biologique - 35380 Paimpont

² ADEME, Service Agriculture et Forêt, 20 avenue du Grésillé - BP 90406 - 49004 Angers Cedex 01-
antonio.bispo@ademe.fr

³ ADEME, Service Friches Urbaines et Sites Pollués -20 avenue du Grésillé - 49004 Angers Cedex 01
cecile.grand@ademe.fr

⁴ ADEME, Service Evaluation de la Qualité de l'Air, 27 rue Louis Vicat - 75737 Paris Cedex 15
laurence.galsomies@ademe.fr

1. Objectifs généraux du programme

Les objectifs généraux du programme étaient de fournir aux secteurs économiques et aux acteurs publics de nouveaux outils de surveillance, de caractérisation et d'évaluation des risques basés sur les propriétés biologiques du sol (Bispo et al., 2009 ; Feix et al., 2012). Plus précisément, les objectifs étaient :

- Accroître l'expertise nationale sur la bioindication et mettre en réseau les laboratoires
- Augmenter les connaissances sur la biodiversité et le fonctionnement des sols
- Développer de nouvelles méthodes de mesure de la diversité et des fonctions des organismes du sol, les valider et les normaliser dans différentes conditions de sol et d'utilisation
- Démontrer la sensibilité et la complémentarité des bioindicateurs pour caractériser différentes perturbations (ex : contamination, changement d'usage et de pratiques)
- Initier le développement de valeurs de référence (base de référence) pour interpréter les résultats actuels et à venir
- Définir des batteries de bioindicateurs ou des indices agrégés pertinents (tant en terme de sensibilité qu'en termes de fonctionnalité, transférabilité) afin de fournir aux acteurs de terrain des outils permettant de répondre à leurs besoins et attentes spécifiques en sols agricoles, forestiers et contaminés.

2. Organisation du programme de recherche

Le programme a été séquencé en 2 phases : une première phase de développement et une seconde de validation des indicateurs.

- La première phase, initiée suite à l'appel à projets lancé en 2004 en lien avec d'autres programmes pilotés par le MEEDDM, l'ANR, l'INRA et le GIS-SOL a pris fin en 2008. L'objectif spécifique de cette phase était d'étudier différents paramètres biologiques afin d'évaluer leur statut de bioindicateur et de favoriser l'émergence de nouveaux bioindicateurs. Ainsi, cette première phase réunissant 37 équipes, impliquées dans 19 projets indépendants a permis d'évaluer plus de 80 paramètres biologiques (de l'état cellulaire jusqu'aux communautés) sur des végétaux, des animaux et des microorganismes.
- La seconde phase, démarrée en 2009, avait pour principaux objectifs de tester, de valider et de calibrer un nombre réduit de bioindicateurs issus de la première phase et jugés comme étant les plus pertinents et les plus opérationnels, ceci afin de proposer aux acteurs publics de nouveaux outils de surveillance ou d'évaluation des risques basés sur des propriétés biologiques. Pour réaliser cette phase de validation, les paramètres se devaient d'être étudiés sur les mêmes sites ateliers (sites choisis selon les thématiques couvertes par l'ADEME, telles que l'épandage des déchets, sites contaminés, retombées atmosphériques, gestion de la matière organique et du stock de carbone, pollution diffuse des sols par les pratiques agricoles, n= 13) et analysés au même moment.

3. Animation et gouvernance du programme

3.1. Pilotage et animation du programme

Depuis son lancement, le programme a été piloté par les 3 ingénieurs ADEME issus des différents services (Service de l'Evaluation de la Qualité de l'air (SEQA), Service Agriculture et Forêt (SAF), Service Friches Urbaines et Sites Pollués (SFUSP)). Pour la seconde phase, compte tenu de l'ensemble des actions et des équipes à coordonner, un animateur général a été désigné à l'Université de Rennes 1 afin de suivre au quotidien l'avancement des travaux (programmer, suivre les différentes phases d'échantillonnage et d'analyse) et assurer le bon déroulement général du programme.

Une à deux fois par an, des réunions générales ont été organisées afin de discuter collectivement des protocoles, des phases d'échantillonnage et des résultats. Parallèlement, des réunions thématiques (ex : base de données, statistiques, faune du sol, microbiologie, flore) ont également été organisées par des animateurs plus sectoriels, responsables de groupes biologiques ou de traitements mathématiques.

3.2. Le comité d'orientation

Un comité d'orientation a été constitué afin de suivre les développements du programme et discuter des résultats. Celui-ci était composé :

- de scientifiques extérieurs au programme bioindicateurs, permettant ainsi un regard critique sur les résultats obtenus
- d'utilisateurs potentiels : ministères, organisations agricoles, industriels, bureau d'étude, agriculteurs, laboratoires d'analyse ...

Ce comité d'orientation a été sollicité plusieurs fois pendant les 2 phases du programme mais surtout lors de la seconde phase afin d'identifier leurs attentes vis-à-vis des bioindicateurs et les formes de restitution qui correspondaient à leurs besoins. La consultation de ce groupe de travail a été très bénéfique et a réellement permis d'affiner la transférabilité des outils biologiques proposés, approche souhaitée par l'ADEME mais aussi par les chercheurs. Pour ce faire, le comité d'orientation a été consulté via différents sondages qui ont permis d'identifier les principales attentes et les différents critères de choix des bioindicateurs, parfois divergents entre les acteurs des « Sites et Sols Pollués » et ceux des « Sites agricoles et forestiers ».

4. La première phase du programme

La première phase a été principalement consacrée au développement et à la calibration de différents bioindicateurs, dans des conditions souvent très spécifiques (ex : un seul type de sol, un seul usage). Sans entrer dans le détail des résultats acquis lors de cette première phase et qui a fait l'objet d'un numéro spécial de la revue *Etude et Gestion des Sols* (EGS, 2009 Volume 16, N°3, <http://www.afes.fr/egs.php>), les principales conclusions des recherches menées entre 2005 et 2008 sont résumées ci-après.

4.1. Les principaux acquis

Les données acquises sur les **paramètres microbiologiques** des sols (biomasse, activités enzymatiques, diversité) ont montré que ces descripteurs sont très sensibles, notamment aux conditions d'usage des sols et aux pratiques agricoles (Dequiedt *et al.*, 2009 ; El Azahri *et al.*, 2008, 2007). Leurs variabilités saisonnières et annuelles, bien que parfois importantes, permettent cependant de toujours différencier des sols en situations contrastées comme : prairies, parcelles cultivées et rotations « prairies/cultures » (Laval *et al.* ; 2009), labour/non labour (Vian *et al.*, 2009, 2008), épandages de composts (Houot *et al.*, 2009 ; Leyval *et al.*, 2009). Dans d'autres essais, s'intéressant plus particulièrement à la contamination métallique ou organique (pesticides), les activités enzymatiques ont également permis de détecter des gradients de contamination (Floch *et al.*, 2009a et 2009b).

Les résultats sur les **invertébrés du sol** (diversité de la macrofaune, des lombriciens, des nématodes et des collemboles), obtenus essentiellement dans des situations agricoles (ex : prairies, parcelles cultivées, épandage de déchets), ont permis de mettre en évidence l'effet de l'usage des sols (Cluzeau *et al.*, 2009 ; Ruiz Camacho *et al.*, 2009) mais également des pratiques agricoles (ex : fertilisation minérale ou organique, apport de pesticides, labour) sur les communautés d'invertébrés du sol (Capowiez *et al.*, 2009 ; Villenave *et al.*, 2009). Par ailleurs, des essais réalisés au laboratoire sur l'exposition de vers de terre à des sols pollués ont démontré l'existence de biomarqueurs spécifiques de certains métaux (ex : Cd) (Brulle et Vandembulcke, 2009 ; Brulle *et al.*, 2008, 2006).

Les analyses des **communautés végétales** sur d'anciens sites métallurgiques ou des sites forestiers ont mis en évidence des associations spécifiques (ex : espèces accumulatrices ou tolérantes, adaptation d'espèces) liées aux conditions de milieux et aux paramètres abiotiques des sols (ex : contaminants, pH, carbonates) (Remon *et al.*, 2009, 2005 ; Marage et Gégout, 2009 ; Dupouet *et al.*, 2009, Seynave *et al.*, 2008 ; Coudun *et al.*, 2006). En complément, la mesure de leur teneur en contaminants métalliques et

l'induction de biomarqueurs spécifiques (ex : acides gras membranaires) ont permis d'identifier les zones présentant le plus de risques de transfert (i.e. forte biodisponibilité) (Le Guédard *et al.*, 2009, 2008 ; Souguir *et al.*, 2009, Vernay *et al.*, 2009, 2008, 2007).

Les analyses de contaminants dans les **chaînes trophiques animales** (ex : invertébrés, mammifères, oiseaux) et la mesure de biomarqueurs ont permis de quantifier les risques de transfert des contaminants mais également leurs effets sur les organismes terrestres (Fritsch *et al.*, 2010 ; Gimbert *et de* Vaufleury, 2009 ; de Vaufleury *et al.*, 2009 ; Gimbert *et al.*, 2008 ; Hispard *et al.*, 2008).

4.2. Les enseignements

Lors de cette première phase, les différentes équipes impliquées ont développé et testé leurs indicateurs dans les situations qu'elles maîtrisaient le mieux (au terrain ou au laboratoire), sans avoir le temps et la possibilité de les évaluer dans d'autres conditions. De plus, les indicateurs n'ont que très rarement été croisés sur des sites communs. Il était donc impossible de les comparer, de juger de leur domaine d'utilisation et donc de définir des batteries optimales en fonction de la situation à évaluer (ex : évaluation des risques, surveillance). Dès lors, la seconde phase du programme a été mise en place en se concentrant sur un nombre plus limité d'indicateurs et sur des sites ateliers communs.

Ainsi, même si la plupart des outils présentaient un intérêt, une étape de sélection s'est avérée nécessaire afin de ne conserver que les indicateurs les plus utilisables en routine (Devillers *et al.*, 2009). Les critères retenus pour la sélection étaient à la fois scientifiques, dépendant de l'indicateur (ex : capacité à renseigner rapidement sur une modification du milieu, facilité à différencier les variations naturelles des perturbations/stress/contraintes induits par l'activité humaine) et techniques, dépendant de la méthode (ex : protocole déjà développé et possibilité de le mettre en œuvre facilement, normalisation, niveau de compétences requis, coût actuel, opérationnalité). Cette phase de sélection a permis de diminuer le nombre d'indicateurs par 2 en ne conservant que les indicateurs les plus aboutis et prometteurs. Ainsi, les indicateurs demandant encore des étapes de développement ont été éliminés (ex : puces à ADN, Sanguin *et al.*, 2006) de même que ceux basés uniquement sur des mesures de laboratoire (ex : exposition d'organismes de laboratoire à des échantillons contaminés).

5. La seconde phase du programme

Les objectifs de la seconde phase du programme étaient de mettre en œuvre tous les paramètres biologiques sélectionnés sur les mêmes sites ateliers (sites ayant des usages différents : agricole, forestier, contaminé) afin de valider leur utilisation en fonction d'une problématique. Il s'agissait d'évaluer leur pertinence en fonction de leur sensibilité (rendant compte de l'intérêt scientifique) mais également de leur « transférabilité » vers les utilisateurs, estimée notamment par rapport à des critères de coût, de facilité de mise en œuvre et facilité d'interprétation, de l'existence de structures (exemple : laboratoires) réalisant l'analyse.

5.1. Les paramètres étudiés

5.1.1. Paramètres biologiques

Les paramètres biologiques retenus suite à la première phase du programme et testés dans la phase 2, sont listés dans les tableaux ci-dessous. Ils sont détaillés pour la plupart d'entre eux à travers des « fiches outils » réalisées par les laboratoires responsables de ces paramètres.

Concernant la **microflore** des sols, ont été retenus : des paramètres de biomasse et d'abondance, des paramètres de diversité fonctionnelle (activités enzymatiques spécifiques, diversité métabolique, activité globale) et des paramètres de diversité structurelle (structure génétique bactérienne et/ou fongique basées sur l'ADN).

Types de paramètres		Descripteurs	Laboratoires impliqués
Paramètres d'effet, d'impact			
Biomasse et abondance	Biomasse totale	Biomasse moléculaire (rendement d'extraction)	INRA Dijon, UMR MSE-Genosol,
		Carbone microbien	INRA Dijon, UMR MSE
	Abondance bactérienne	Bactérie cultivable	ESITPA, Labo. Biosol
		ADNr 16S	ESITPA, Labo. Biosol
		Pseudomonas	Univ. Rouen, UMR M2C
	Abondance, biomasse fongique	Ergostérol total	ESITPA, Labo. Biosol
		Ergostérol libre	ESITPA, Labo. Biosol
PLFA (totaux, spécifiques)		INRA Versailles –UR PESSAC, ESITPA, Labo. Biosol	
	ADNr 18S	ESITPA, Labo. Biosol	
Diversité fonctionnelle	Activités enzymatiques	<u>Activité globale</u> : Déshydrogénase, Fluorescéine di acétate, Lipase, <u>Cycle du carbone</u> : Galactosidase, Laccase, Glucosidase, Cellulase, Xylanase, Arylamidase <u>Cycle du soufre</u> : Arysulfatase <u>Cycle de l'azote</u> : Uréase, N-Acétyl Glucosamidases <u>Cycle du phosphore</u> : Phosphatases acides et alcalines	INRA Versailles –UR PESSAC, ESITPA, Labo. Biosol Univ. Paul Cézanne, UMR IMEP
		Diversité métabolique bactérienne potentielle DMP	DMP (Biolog) : activité métabolique globale (AWCD), Richesse fonctionnelle (puits positifs)
	Activité globale	Minéralisation azote, Minéralisation carbone, respiration	INRA Dijon, UMR SME Univ. Paul Cézanne, UMR IMEP
Diversité structurelle génétique	Diversité génétique bactérienne/fongique	Outil : ARISA	INRA Dijon, UMR SME-Genosol,
		Outil : TTGE	CNRS Nancy, UMR LIMOS

Concernant **la faune**, des indicateurs basés sur la diversité des populations d'invertébrés du sol (macrofaune totale, lombriciens, nématodes, collemboles) ont été sélectionnés; par ailleurs, des mesures d'effet et d'accumulation ont également été réalisées sur les escargots, les vers et des micromammifères.

Types de Paramètres	Descripteurs	Equipes impliqués
Paramètres d'effet, d'impact		
Paramètre global	IBQS: indice global basé sur la macrofaune totale	IRD Bondy, UMR Bioemco
Paramètre de communauté (critères globaux, diversité fonctionnelle, diversité spécifique, indices de maturité et de structure)	Communauté nématodes (microfaune)	IRD Montpellier, UMR EcoSol – EliSol
	Communauté microarthropodes : Collemboles, Acariens (mésafaune)	INPL-ENSAIA Nancy, UMR LES
	Communauté lombricienne (macrofaune)	Univ. Rennes1, UMR EcoBio
Paramètres d'effet	Bioindicateur moléculaire lombricien (méthallothionéine à Cd)	Univ. Lille, UMR LGCgE
Paramètres d'accumulation		
Paramètres d'accumulation	Bioindicateur escargots (Indice de Bioaccumulation)	Univ. Franche-Comté, UMR «Chrono-Environnement
	Biomarqueurs micromammifères	

Concernant **la flore**, des mesures de paramètres d'impact et d'accumulation des contaminants ont été réalisées.

Types de paramètres	Descripteurs	Equipes impliqués
Paramètres d'impact	Omega3 (variation de la composition en lipides des feuilles)	Univ. Bordeaux 2, UMR Biogénèse membranaire – LEB Aquitaine Transfert
	Acides Aminés (variations de la teneur en acides aminés)	IUT Clermont-Ferrand, Labo. PBV
	Photosynthèse	IUT Clermont-Ferrand, Labo. PBV
Paramètres d'accumulation	Bioaccumulation des ETM (teneurs en métaux des feuilles)	Ecole Des Mines St Etienne, centre SPIN
	Dendrochimie (teneur en métaux des cernes de bois)	INRA Nancy, UMR EEF

Ces indicateurs (à l'exception notamment des indicateurs basés sur les micromammifères et des mesures dendrochimiques qui n'ont été réalisés que sur des sites présentant des superficies importantes) ont été mesurés sur l'ensemble de sites ateliers choisis en commun, en lien avec les thématiques prioritaires de l'ADEME.

5.1.2. Paramètres descriptifs

Afin de pouvoir expliciter au mieux les réponses biologiques et pouvoir ainsi calibrer les valeurs obtenues, un certain nombre de paramètres descriptifs complémentaires ont été mesurés ou collectés.

Métriques paysagères (IRD Bondy)

Des métriques paysagères ont été calculées à partir d'orthophotographies des différents sites de prélèvement. Les habitats ont ainsi été cartographiés et les structures paysagères décrites (ex : fragmentation, dispersion).

Données physico-chimiques

Les analyses physico-chimiques ont été réalisées par le laboratoire INRA d'Arras, à savoir :

- granulométrie, pH eau, Corg, N total, matière organique, Calcaire (CaCO₃) total, Phosphore (P₂O₅), CEC, Ca, Mg, Na, K, Fe, Mn, Al (échangeable à la cobaltihexamine),
- éléments traces totaux (Co, Cr, Cu, Ni, Zn, Pb, Cd, Tl, Mo), Arsenic et Mercure totaux
- extraction au CaCl₂ (Co, Cr, Cu, Ni, Zn, Pb, Cd, Tl, Mo, As)
- 16 HAP (ex : naphthalène, acénaphène, fluorène, phénanthrène, anthracène)
- 12 organo-chlorés,
- 10 triazines,
- 8 phényl-urées.

Méta-Données

Les données relatives aux différents sites étudiés (définissant les métadonnées) sont enregistrées, à savoir :

- Caractéristiques climatiques (pluviométrie, températures, vent)
- Historique parcellaire (usage des sols, pratiques agricoles, rotation culturales ...)

5.2. Les sites ateliers étudiés

Différents sites ateliers ont été proposés par les équipes du programme. L'ADEME a ensuite réalisé une sélection en fonction de différents critères : thématique d'étude, accès libre, confidentialité des résultats obtenus, pas de modification de l'usage pendant 3 ans, existence de données sur le site pour positionner les zones de prélèvement, antériorité de mesures de type bioindicateurs.

In fine, 12 sites ont été retenus en fonction de leur usage, de leur niveau de contamination et de leur intérêt dans le cadre des thématiques prioritaires à l'ADEME. Par ailleurs, un site proposé par l'ANDRA a été inclus dans cette liste de sites étudiés.

Ces 13 sites, répartis sur le territoire national, se distribuent selon : 4 sites forestiers, 4 sites contaminés (contaminations pluri-métallique, en arsenic, et/ou organique), 5 sites agricoles (différents usages des sols, gestion prairiale, gestion organique, mode de gestion en vergers, travail du sol). Les sites forestiers correspondent à des placettes forestières du réseau RENECOFOR¹ (gestion par l'ONF) ; ils ont été choisis en fonction de leur proximité avec les autres sites agricoles ou anthropisés étudiés. Les 13 sites étudiés définissent ainsi 47 contextes (modalités) différents.

Les principales caractéristiques des sites et des différents contextes étudiés (modalités) sont listées dans le tableau ci-dessous, les sites en eux-mêmes étant détaillés dans des « fiches sites » rédigées par les gestionnaires des sites.

¹RENECOFOR (Réseau National de suivi à long terme des ECOSystèmes FORestier) :

http://www.onf.fr/gestion_durable/sommaire/action_onf/recherche/observation/20080424-105459-848844/@@index.html?search:int=1&search_source=L3d3dy9sX29uZg==&search_group:int=237049089&search_metatype=search-type-article

Nom des sites	Gestionnaires de site	Codes des modalités	Description
GISFI - Homécourt	CNRS Nancy, UMR LIMOS	GHF GHM	Contamination en HAP faible , Friches industrielles Contamination en HAP moyenne , Friches industrielles
Metaleurop	ISA Lille, UMR LGCgE	Me98F Me117F Me117C Me103F Me103C MeTEF MeTEC Me123 MeADEME MeUsine	Contamination métallique (Pb, Zn, Cd) forte , Bois Contamination métallique (Pb, Zn, Cd) moyenne , Bois Contamination métallique (Pb, Zn, Cd) moyenne , culture Contamination métallique (Pb, Zn, Cd) faible , Bois Contamination métallique (Pb, Zn, Cd) faible , culture Contrôle, Bois Contrôle, Culture Contamination métallique (Pb, Zn, Cd) faible , Bois Bis (pour Dendrochimie) Contamination métallique (Pb, Zn, Cd) moyenne , Bois Bis (pour dendrochimie) Contamination métallique (Pb, Zn, Cd) forte , Bois Bis (pour dendrochimie)
Auzon	IUT Clermont Ferrand, Labo. PBV Collaboration BRGM	AuBHCo AuBCo AuOBCo AuFCo AuBTe AuOBTe AuPTe AuCCO AuComposite	Contamination métallique (As, Hg), Bois sur sol hydromorphe Contamination métallique (As, Hg), Bois Contamination métallique (As, Hg), Ourlet-Boisé Contamination métallique (As, Hg), Friches Contrôle Bois Contrôle, Ourlet-Boisé Contrôle Prairie Contrôle utilisé pour escargots (en remplacement du contrôle Prairie) Ensemble des sites utilisés pour les micromammifères
SHSE Crassier métallurgique stéphanois	Ecole des Mines St Etienne, SPIN	StE-A StE-I StE-R	Contamination pluri-métallique élevée, Niveau important de couvert végétal (approx. 60-70%) Contamination pluri-métallique élevée, Niveau intermédiaire de couvert végétal (approx. 30%) Contamination pluri-métallique élevée, Niveau faible de couvert végétal (approx. 5-10%)
QualiAgro-Feucherolles	INRA Grignon, UMR EGC Collaboration VERI	FeTé FeFum FeBio FeDVB FeOMR	Contrôle Epannage de fumier de bovins Epannage de compost de Biodéchets (co-compostage ordures ménagères et déchets verts) Epannage de compost de déchets verts et boue (co-compostage de déchets verts et boues épurations urbaines) Epannage de compost d'ordures ménagères résiduelles (compostage après collecte sélective des emballages « propres et secs »)

Yvetot	ESITPA Rouen, Labo. Biosol	YvSI	Prairie de Restauration (2 ans de Prairie) après 5 années de culture
		YvSII	Prairie de restauration (2 ans de Prairie) après 6 années de culture
		YvSIII	Prairie temporaire (en blé lors du prélèvement) dans une rotation culturale 2 ans maïs-blé et 4-8 ans de prairie
		YvSIV	Prairie temporaire (5 ans de prairie lors du prélèvement) dans un rotation culturale 2 ans maïs-blé et 4-8 ans de prairie
		YvPP YvGC	Prairie permanente (> 25 ans) avec couvert permanent et sans labour Grande Culture crop (10 ans) sous gestion intensive (labour, fertilisation)
BioREco- Gotheron	INRAAvignon UERI Gotheron	GOBA	Système en Agriculture Biologique ; variété "Ariane"
		GOBS	Système en Agriculture Biologique ; variété "Smoother = Golden Délicieuse"
		GOEA	Système Econome en intrants (technicité maximale pour réduire la protection chimique), variété "Ariane"
		GOES	Système Econome en intrants (technicité maximale pour réduire la protection chimique) ; variété "Smoother"
		GORA GORS	Système Raisonné (généralement protection chimique, sans prise de risque) ; variété "Ariane" Système Raisonné (généralement protection chimique, sans prise de risque) ; variété "Smoother"
Thil	ISARA Lyon, UPSP SCAB	ThTS	Travail du sol très Superficiel (moins de 5-7 cm de profondeur)
		ThLA	Labour Agronomique (labour hors-raie à 18 cm)
		ThLT	Labour Traditionnel (labour avec une charrue classique à 30 cm)
		ThTR	Travail du sol Réduit (outil à dent type chisel à 15 cm sans retournement)
RENECOFOR *	ONF Collaboration Univ. Rennes 1	F08	Forêt (Epicéa), département des Ardennes (08)
		F57	Forêt (Sapin Pectiné), département de la Moselle (57)
		F76	Forêt (Pin Sylvestre), département de Seine Maritime (76)
		F63	Forêt (Epicéa), département du Puy de Dôme (63)
OPE-ANDRA	ANDRA	ANC	Culture (Blé lors du prélèvement)
		ANF	Forêt (Hêtre)
		ANP	Prairie Permanente

5.3. Protocole et procédures de prélèvement

Compte tenu du nombre important de sites, et des contraintes de prélèvement liées aux paramètres biologiques qui font que ceux-ci ne peuvent être mesurés qu'en période d'activité favorable (printemps ou automne), il a été décidé de réaliser une première campagne de prélèvements au printemps 2009 (échantillonnage des sites de la moitié Nord de la France) et d'échantillonner au printemps 2010 les sites situés dans la partie Sud, le site de l'ANDRA ayant été échantillonné au printemps 2011.

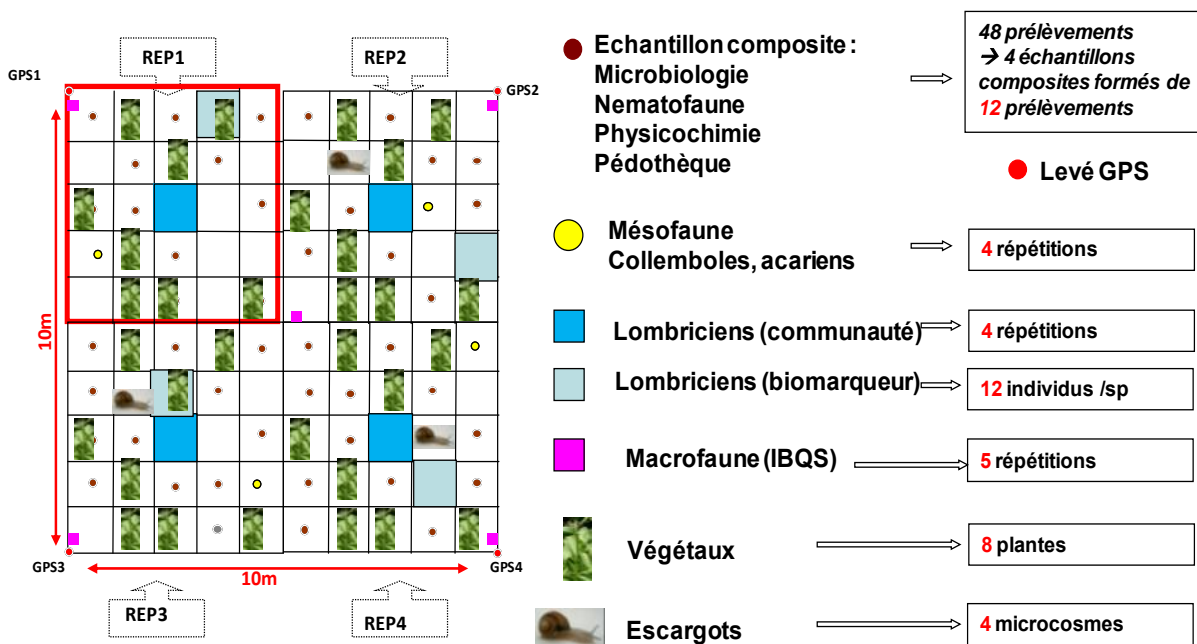
Les prélèvements de sol et de faune ont été réalisés en même temps (au printemps) alors que la flore a été échantillonnée ultérieurement, au moment le plus propice de son stade de développement.

La conduite de cette seconde phase du projet nécessitant une part importante de coordination, celle-ci a été confiée à l'Université de Rennes 1 (UMR EcoBio) qui avait précédemment réalisé le même travail dans le cadre du RMQS BioDiv (Cluzeau et al., 2012).

Pour chaque site, le gestionnaire a été chargé de définir l'emplacement exact des zones de prélèvement en fonction de stratégies propres au site (ex : gradient de contamination, d'usage, de type de sol).

Un protocole de prélèvement discuté collectivement et devant répondre aux objectifs de chaque équipe a été adopté et il a été appliqué sur chaque site par une même équipe de préleveurs (assurant la qualité des prélèvements). Le dispositif, se base sur les recommandations du groupe de travail du programme Européen ENVASSO (Bispo et al., 2007; Kibblewhite et al., 2008) ainsi que sur le dispositif appliqué dans le programme national RMQS BioDiv (Cluzeau et al., 2009 ; Cluzeau et al., 2012) et répond à la norme ISO d'échantillonnage 2311-6 : 2012 (en cours de publication).

Chaque zone de prélèvement a fait ainsi l'objet d'une stratégie d'échantillonnage identique afin d'être en mesure de comparer les données issues des différents sites. Les prélèvements ont été réalisés sur une zone d'une surface de 100 m² (10 m X 10m) sur chaque modalité, sub-divisée en 4 répétitions de 25 m² (Figure ci dessous).



Afin de pouvoir optimiser la mesure de certains paramètres biologiques, les prélèvements de sols ont été réalisés systématiquement en tout début de semaine sur les sites afin d'expédier les échantillons le plus vite aux laboratoires (livraison par Chronopost contenant des accufroids pour maintenir une température optimale). De ce fait, dans le cadre de ce programme, seul un site a pu être échantillonné par semaine. Les différentes étapes du prélèvement sont détaillées ci-après.

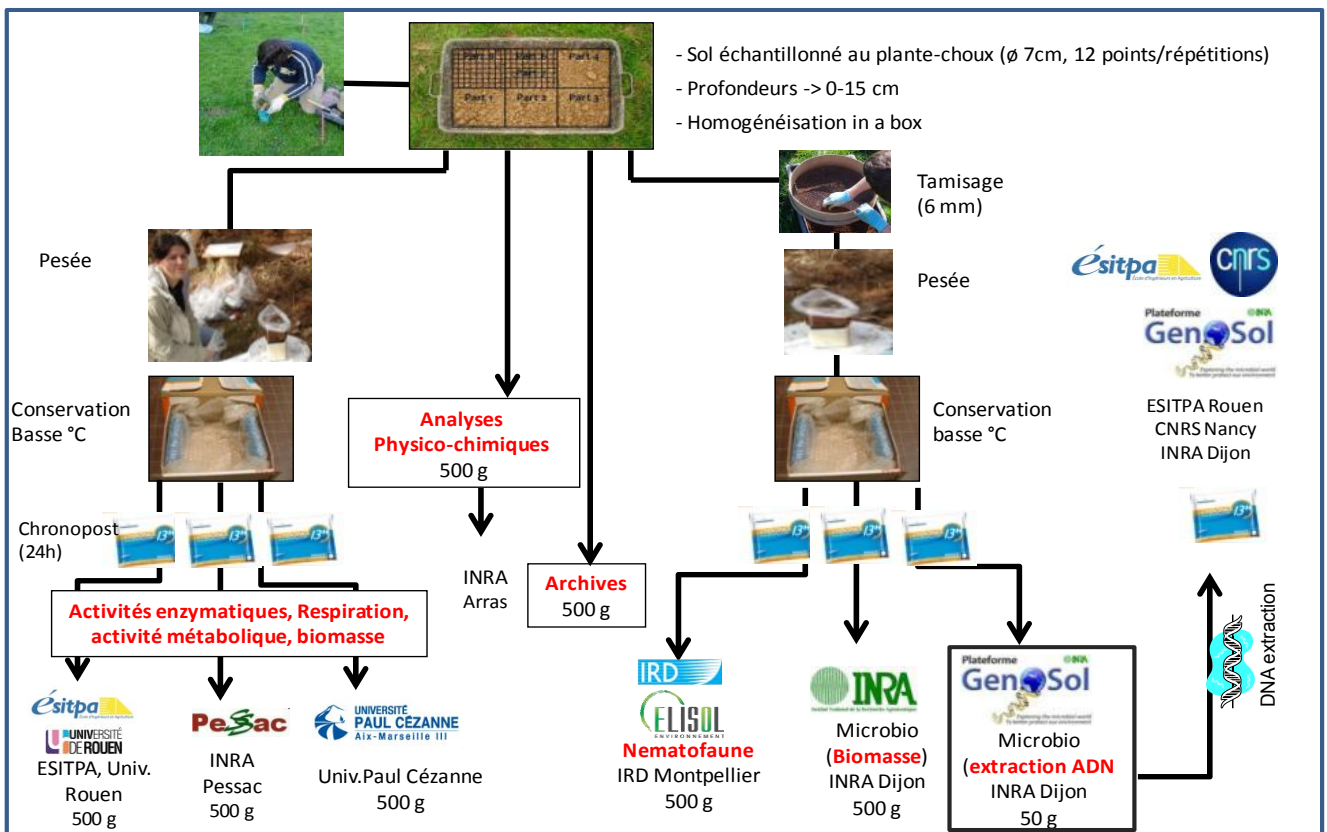
➤ **Etape 1 : positionnement de la zone de prélèvement et localisation par GPS**



➤ **Etape 2 : Prélèvement de sol** par carottier (7 cm de diamètre) sur 0-15 cm de profondeur (en système forestier, les horizons OL et OF sont enlevés, leur épaisseur respective est notée ; le prélèvement de sol se fait donc en intégrant l'horizon OH et A). Douze carottages sont réalisés par répétition de manière aléatoire, puis associés et répartis de manière homogène pour aboutir à un échantillon composite qui sera ensuite subdivisé en parts égales. Afin de répondre aux besoins de différents paramètres biologiques :

- une partie des sous-échantillons est tamisée à 6 mm, cela correspond aux paramètres : **nématodes** (cf fiche outil « nématodes »), **biomasse microbienne** et **extraction d'ADN**,
- les autres échantillons de sol, non tamisés, conservent leur structure initiale pour la mesure d'autres paramètres, à savoir : **biomasse moléculaire fongique** (cf fiches outil « biomasse moléculaire fongique estimée par ergosterol »), **abondance et biomasse microbienne** (PLFA, Pseudomonas), **activités enzymatiques** (cf fiche outil : « les activités enzymatiques »), **diversité métabolique potentielle** (cf fiche outil « diversité métabolique potentielle »), **physico-chimie** et **l'archivage**.

Ces échantillons sont pesés et expédiés aux différents laboratoires pour analyse. L'extraction d'ADN a été réalisée par un seul laboratoire (GenoSol) qui a ensuite redistribué les ADN aux différents laboratoires concernés, permettant l'étude des paramètres suivants : **diversité microbienne** (cf fiches outil « Empreinte moléculaire des communautés microbienne », « Diversité taxonomique microbienne »), **biomasse moléculaire microbienne ou fongique** (cf fiches outil « Biomasse microbienne », « Biomasse moléculaire fongique estimée par ADNr 18S »).



- **Etape 3 : Prélèvement des échantillons de sols pour la mésofaune** (collemboles, acariens – cf fiche outil « les microarthropodes »). Il se fait selon la méthode normalisée ISO 23611-2 : 2006. Les prélèvements (1 par répétition) sont réalisés à l'aide d'un plexis (6 cm de diamètre) sur une profondeur de 5 cm. Les échantillons sont envoyés directement au laboratoire qui réalise l'extraction de la mésofaune ainsi que sa détermination au niveau groupe fonctionnel.



Figure 1, 2 et 3: Étapes du prélèvement de mésofaune

- **Etape 4 : Prélèvement des lombriciens.** Ils sont prélevés selon la méthode adaptée de la norme ISO 23611-1 : 2006 qui combine une extraction chimique (formol) et un tri manuel (cf fiche outil « les vers de terre »). Un prélèvement est réalisé par répétition. Les lombriciens prélevés sont rapportés au laboratoire où ils sont déterminés finement au niveau spécifique et pesés individuellement.



Pour l'étude des **biomarqueurs lombriciens**, une solution formolée (Bouché, 1972) est appliquée à la surface du sol, les lombriciens récoltés sont rincés à grande eau. 12 individus (adultes de préférence) d'une même espèce sont envoyés vivants dans des boîtes plastiques contenant leur sol d'origine, pour être ensuite analysés au laboratoire (cf fiche outil « Expression génique de la métallothionéine (MT) chez les vers de terre (biomarqueur de contamination des sols par le Cadmium) »).

- **Etape 5 : Prélèvement de la macrofaune totale** (prélèvement réalisé par l'équipe de IRD Bondy ou par équipe de l'INRA-PESSAC). Ces prélèvements sont réalisés suivant la méthode TSBF (Tropical Soil Biology and Fertility) modifiée pour être adaptée aux milieux tempérés (Lavelle, 1988 ; Anderson et Ingram, 1993) (norme : 23611-5, 2011). Cinq répétitions sont réalisées par modalité. Cette méthode combine l'extraction au formol (1,5 L de solution à 0.2% déversé 2 fois à dix minutes d'intervalle) et un tri manuel d'un bloc de sol (25 cm X 25 CM sur 15 cm de profondeur). Les organismes récoltés sont conservés dans une solution à 4%, et rapportés au laboratoire où ils sont déterminés au niveau de l'ordre, puis si possible au niveau de la famille ou du genre.

Les étapes suivantes (6 à 8) ont été menées par les équipes porteuses du paramètre biologique, et selon des dates et des protocoles adaptés à ce paramètre.

- **Etape 6 : Etude du bioindicateur « escargot ».** La bioaccumulation a été étudiée sur des escargots de l'espèce *Cantareus aspersus* (syn. *Helix aspersa*) placés dans des microcosmes (cylindre en inox 25x25cm) sur le terrain (cf fiche outil : « Les escargots »). Les escargots sont laissés sur site pendant un temps défini (28 jours) puis ramenés au laboratoire pour être analysés (spectrométrie d'absorption, SAA ou ICPMS).



- **Etape 7 : Etude du bioindicateur « Micromammifère ».** La bioaccumulation chez les micromammifères a été étudiée uniquement sur le site d'Auzon sur 3 secteurs présentant un gradient de contamination sur des surfaces adaptées à l'aire de vie de ces vertébrés. Les micromammifères sont

échantillonnés par piégeage létal. Les micromammifères capturés sont congelés et rapportés au laboratoire pour être analysés (cf. fiche outils : « les micromammifères »).

➤ **Etape 8 : Etude des paramètres végétaux.**

La liste des sites sur lesquels les paramètres végétaux ont été étudiés est reportée dans le tableau ci-dessous.

Nom du site	Bioaccumulation des ETM	Omega3	Photosynthèse /Acides aminés	Dendrochimie
QualiAgro		x		
Thil		x		
BioREco-Gotheron		x	x	
Yvetôt	x	x		
F76	x			
F08	x			
F63	x			
F57	x			
Metaleurop (Bois)	x	x	x	x
GISFI-Homécourt	x	x	x	x
SHSE - Crassier Saint-Etienne	x	x	x	
Auzon	x	x	x	x
ANDRA		x	x	

Dans un premier temps, la végétation de chaque modalité a été prospectée afin d'identifier les principales espèces dominantes représentatives des zones étudiées. Les prélèvements des végétaux ont ensuite été réalisés en fonction de la répartition des espèces végétales présentes et en fonction de l'indicateur étudié (cf fiches outils : Les communautés végétales, Oméga3 et Photosynthèse) :

- **Bioaccumulation des ETM** : 2 stratégies de prélèvements ont été réalisées : i) la première (sur Métaeurop, Auzon) a consisté à étudier les concentrations métalliques foliaires dans des échantillons prélevés individuellement : les parties aériennes de 10 espèces dominantes ont été prélevées de façon indépendante. ii) la deuxième stratégie (sur tous les sites) a consisté à étudier les concentrations métalliques foliaires dans des échantillons composites (« pool ») : les parties aériennes de 5 espèces différentes présentes dans une même zone ont été prélevées et cet échantillonnage élémentaire a été répété 5 fois sur chaque modalité, en utilisant si possible des mélanges d'espèces différentes entre chaque réplica (cf fiche outil : Les communautés végétales).
- **Oméga 3** : 4 à 6 espèces dans la mesure du possible communes à chaque parcelle ont été identifiées. Un morceau de feuille de chaque espèce végétale est découpé à l'aide d'un ciseau et les tissus foliaires prélevés sont placés dans des tubes en verre contenant un mélange de méthanol et d'acide sulfurique, puis rapportés au laboratoire pour être analysés (cf fiche outil : Oméga3).
- **Photosynthèse** : 5 espèces végétales différentes communes à chaque modalité ont été identifiées. Les paramètres caractérisant le fonctionnement des systèmes photosynthétiques ont été mesurés directement sur le terrain : Photosynthèse nette (Pn) a été mesurée sur chaque échantillon, les autres paramètres (qP, PSII, Fv/Fm...) ont été mesurés sur des feuilles intactes (cf fiche outil : Photosynthèse).
- **Teneurs foliaires en Acides Aminés** : 6 espèces végétales différentes communes à chaque modalité ont été identifiées. Des échantillons de feuilles ont ensuite été prélevés, puis congelés (azote liquide) et mis dans de la carboglace avant d'être analysés au laboratoire par Chromatographie Liquide Haute Performance (CLHP) (Beckman system Gold).
- **Dendrochimie** : Sur le site de Métaeurop 5 arbres ont été échantillonnés (*Quercus robur* et *Acer pseudoplatanus*) sur différentes modalités et 10 arbres (*Quercus robur*) ont été échantillonnés sur l'ensemble du site d'Auzon. Pour chaque arbre, deux carottes de bois (superposées) à 1,30 m dans le tronc, ont été prélevées complétée par un prélèvement d'écorce (avec un emporte-pièce). Les mesures micro-élémentaires ont ensuite été effectuées au laboratoire par spectrométrie à dispersion d'énergie (EDS) et spectrométrie à sélection de longueur d'onde (WDS), ces techniques permettant une mesure semi-quantitative d'une large gamme d'éléments, couplées avec une interface d'imagerie (microscope électronique à balayage).

5.4 Gestion et traitement des données

Compte tenu du nombre important de données de mesures physiques, chimiques et biologiques recueillies dans le programme et du besoin avéré de traiter de manière hiérarchisée l'ensemble des résultats, il a été décidé (i) de créer une base de données commune et (ii) d'harmoniser les traitements statistiques en fonction des paramètres et des objectifs (ex : mise en évidence de doses-réponses, effet des paramètres abiotiques sur les réponses biologiques, modélisation des données, identification des complémentarités/redondances entre les outils, élaboration des valeurs de référence). Les différentes étapes de constitution de la base de données ainsi que la démarche statistique développée par le groupe de statisticiens « Biomath » et leurs principaux aboutissements sont détaillés dans Taibi et al. 2012.

6. Conclusion

Si la première phase du programme avait pour objectif de tester et de développer différents paramètres biologiques dans plusieurs situations contrastées de sol, elle n'avait pas permis de les comparer et de juger de leur complémentarité. Cette seconde phase du programme a validé ces différents bioindicateurs en définissant leur domaine d'application (ex : surveillance de la qualité des sols, diagnostics de sols pollués, requalification des sols délaissés, évaluation les modifications de pratiques agricoles ou sylvicoles). Cette mesure de l'ensemble des paramètres biologiques sélectionnés sur tous les sites communs a également permis de mieux qualifier ces outils biologiques (ex : amplitude des réponses à la présence d'un contaminant dans le sol ou à une modification du sol telle que le tassement, le changement d'usage, l'acidification, etc...) et d'évaluer leur sensibilité et leur robustesse (peut-on distinguer les réponses liées aux variations naturelles des réponses liées à un contaminant ?). Bien qu'obtenues sur un nombre limité de sites, les données de cette seconde phase du programme, complétées par celles acquises lors de la première phase, constituent une première base de références facilitant l'interprétation des données ultérieures et leur utilisation pour la surveillance et la protection des sols. Par ailleurs la consultation du comité d'orientation a conduit à la sélection de bioindicateurs, ou de batteries de bioindicateurs, ou encore d'indice agrégés en intégrant non seulement leur sensibilité mais également leur transférabilité (c'est-à-dire applicabilité) vers les utilisateurs potentiels.

L'aboutissement majeur de ce programme unique à l'échelle européenne par la multitude de paramètres testés et la variété des contextes étudiés, est la démonstration qu'il est possible de mettre en œuvre différents outils biologiques sur un site quel que soit son usage et qu'une partie de ces outils est déjà en voie d'être transférée vers les utilisateurs. D'autres vont certainement bénéficier d'énormes avancées technologiques et pourront être utilisables par les acteurs de terrain dans un futur plus ou moins proche.

Enfin il apparaît clairement que ces outils biologiques sont source d'informations complémentaires aux analyses physico-chimiques classiques, renseignant notamment des risques de transferts des contaminants vers les chaînes trophiques ou informant de l'altération ou de l'amélioration de services écosystémiques rendus par le sol. Par conséquent ces bioindicateurs peuvent constituer des outils d'aide à la décision dans la gestion d'un site pour son usage actuel (agricole, forestier) ou son usage futur (reconversion des friches, gestion sites pollués à grande échelle).

7. Bibliographie

- Arrouays D., Jolivet Cl., Boulonne L., Bodineau G., Ratié C., Saby N. et Grolleau E. 2003. Le Réseau de Mesures de la Qualité des Sols (RMQS) de France. Étude et Gestion des Sols, Volume 10, 4, 241-250
- Bispo A, Peres G, Cluzeau D, Graefe U, Römbke J, Rutgers M, Fuchs M, Sousa JP, Schulte R, Dombos M, Simon B, Gál A, Cortet J, Chaussod R, Ritz K, Creamer R, Winding A, English M, Boixadera J. Rubio JL. 2007. ENVASSO (Environmental assessment of soil for monitoring) WP 5 – Decline in soil biodiversity. EU Contract No. 022713, 22 pp.
- Brulle F et Vandenbulcke F. 2009. Développement de biomarqueurs d'exposition aux métaux. Etude et Gestion des Sols, Volume 16, 3/4, 159-173.
- Brulle F., Cocquerelle C., Wamalah A.N., Morgan A.J., Kille P., Leprêtre A. & Vandenbulcke F. (2008) cDNA cloning and expression analysis of Eisenia fetida (Annelida; Oligochaeta) phytochelatin synthase under cadmium exposure. *Ecotoxicology and Environmental Safety*, 71, 47-55
- Brulle F., Mitta G., Cocquerelle C., Vieau D., Lemièrre S., Leprêtre A. & Vandenbulcke F. (2006) Cloning and Real-Time PCR testing of 14 potential biomarkers in Eisenia fetida following cadmium exposure. *Environmental Science & Technology*, 40, 2844-2850

- Capowiez Y., Rault M., Mazzia C., Poitrenaud M., Houot S. 2009. Etude des effets des apports de produits résiduels organiques sur la macrofaune lombricienne en conditions de grandes cultures. *Etude et Gestion des Sols*, Volume 16, 3/4, 175-195.
- Cluzeau D., M. Guernion, R. Chaussod, F. Martin-Laurent, C. Villenave, J. Cortet, N. Ruiz-Camacho, C. Pernin, T. Mateille, L. Philippot, A. Bellido, L. Rougé, D. Arrouays, A. Bispo, G. Pérès. 2012. Integration of biodiversity in soil quality monitoring: Baselines for microbial and soil fauna parameters for different land-use types. *European Journal of Soil Biology*. Volume 49, 63-72
- Cluzeau D., Pérès G., Guernion M., Chaussod R., Cortet J., Fargette M., Martin-Laurent F., Mateille T., Pernin C., Ponge J-F., Ruiz-Camacho N., Villenave C., Rougé L., Mercier V., Bellido A., Cannavacciuolo M., Arrouays D., Boulonne L., Jolivet C., Lavelle P., Velasquez E., Plantard O., Walter C., Foucaud-Lemerrier B., Tico S., Giteau J-L., Bispo A. 2009. Intégration de la biodiversité des sols dans les réseaux de surveillance de la qualité des sols : exemple du programme-pilote à l'échelle régionale, le RMQS BioDiv. *Etude et Gestion des Sols*, Volume 16, 3/4, 187-201.
- Commission Européenne. 2006. Thematic Strategy for Soil Protection, Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Septembre 2006 (disponible à l'adresse: http://ec.europa.eu/environment/soil/pdf/com_2006_0231_en.pdf)
- Coudun C., Gégout J.C., Piedallu C., Rameau J.C. (2006). Soil nutritional factors improve plant species distribution models: an illustration with *Acer campestre* L. in France. *Journal of Biogeography*, 33 : 1750-1763.
- de Vaufleury A., Fritsch C., Gimbert F., Pauget B., Cœurassier M., Crini N. et Scheifler R. 2009. Utilisation et intérêts des escargots et des micromammifères pour la bioindication de la qualité des sols. *Etude et Gestion des Sols*, Volume 16, 3/4, 203-218.
- Dequiedt S., Lelièvre M., Jolivet C., Saby N., Martin M., Thioulouse J., Maron P.A., Mougé C., Chemidlin Prévost-Bouré N., Arrouays D., Lemanceau P. et Ranjard L. 2009. ECOMIC-RMQS : biogéographie microbienne à l'échelle de la France. Etat d'avancement et premiers résultats. *Etude et Gestion des Sols*, Volume 16, 3/4, 219-231.
- Devillers J. Pandard P. et Charissou. A.M. 2009. Sélection multicritère de bioindicateurs de la qualité des sols. *Etude et Gestion des Sols*, Volume 16, 3/4, 233-242.
- Dupouey J.L., Rose C., Bailly R., Behr P., Ponton S. et Weitner A. 2009. Mise au point d'indicateurs dendrochimiques des variations spatiales et temporelles de l'acidité des milieux forestiers. Rapport final du contrat DAEME n° 0562C0030. 26 p.
- El Azhari N., Chabaud S., Percept A., Bru D. And Fabrice Martin-Laurent. (2007). *pcaH*, a molecular marker for estimating the diversity of the protocatechuate-degrading bacterial community in the soil environment. *Pest Management Science*, 63:459-467.
- El Azhari N., D. Bru, A. Sarr and F. Martin-Laurent. 2008. Estimation of the density of the protocatechuate-degrading bacterial community in soil by real-time PCR. *European Journal of Soil Science*. 59: 665-673.
- Feix I., Bispo A., Grand C. Galsomies, L., 2012. Pourquoi un programme ADEME sur la Bioindication des sols ? Actes Journées Techniques Ademe Bioindicateurs&Phytotechnologies : des outils biologiques pour des sols durables, 16-17 oct 2012
- Floch C., Alarcon-Gutierrez E., Criquet S. (2009a) Metal effects on phenol oxidase activities on soils . *Ecotoxicology and Environmental Safety* 72, 108-114.
- Floch C., Capowiez Y., Criquet, S. (2009b). Enzymes activities in apple orchard agroecosystems : How are they affected by management strategy and soil properties. *Soil Biology and Biochemistry*, Volume 41, Issue 1, 61-68
- Fritsch C., Cosson R.P., Cœurassier M., Raoul F., Giraudoux P., Crini N. de Vaufleury A., Scheifler R. 2010. Responses of wild small mammals to a pollution gradient: Host factors influence metal and metallothionein levels. *Environmental Pollution*, 158 (2010) 827–840
- Gardi C., Montanarella L., Arrouays D., Bispo A., Lemanceau P., Jolivet C., Mulder C., Ranjard L., Rombke L., Rutger M., Menta C., 2009 - Soil Biodiversity monitoring in Europe: ongoing activities and challenges. *European Journal of Soil Science*. 60:807-819
- Gimbert F et de Vaufleury A. 2009. Bioindication et unités (concentrations vs quantités) : comparaison des cinétiques d'accumulation et d'élimination du Cd, Pb et Zn chez l'escargot *Helix aspersa*. *Etude et Gestion des Sols*, Volume 16, 3/4, 243-253.

- Gimbert F., Mench M., Coeurdassier C., Badot P.-M. de Vaufleury A. 2008. Kinetic and dynamic aspects of soil-plant-snail transfer of cadmium in the field. *Environmental Pollution*, 152, 736-745.
- Hispard F, Schuler D, de Vaufleury A, Scheifler R, Badot P.M. and Dallinger R. 2008. Metal distribution and metallothionein induction after cadmium exposure in the terrestrial snail *Helix aspersa* (gastropoda, pulmonata). *Environmental Toxicology and Chemistry*, 27(7), 1533-1542.
- Houot S., Cambier P., Benoit P., Deschamps M., Jaulin A., Lhoutellier C., Barriuso E. 2009. Effet d'apports de composts sur la disponibilité de micropolluants métalliques et organiques dans un sol cultivé. *Etude et Gestion des Sols*, Volume 16, 3/4, 255-274.
- ISO 23611-1. 2006. Qualité du sol -- Prélèvement des invertébrés du sol -- Partie 1: Tri manuel et extraction au formol des vers de terre.
- ISO 23611-2. 2006. Qualité du sol -- Prélèvement des invertébrés du sol -- Partie 2: Prélèvement et extraction des micro-arthropodes (Collembola et Acarina).
- ISO 23611-3. 2007. Qualité du sol -- Prélèvement des invertébrés du sol -- Partie 3: Prélèvement et extraction des enchytréides.
- ISO 23611-4. 2007. Qualité du sol -- Prélèvement des invertébrés du sol -- Partie 4: Prélèvement, extraction et identification des nématodes du sol.
- ISO 23611-5. 2011. Qualité du sol -- Prélèvement des invertébrés du sol -- Partie 5: Prélèvement et extraction des macro-invertébrés du sol
- ISO 23611-6. 2012. Qualité du sol - Prélèvement des invertébrés du sol - Partie 6 : Lignes directrices pour la conception de programmes d'échantillonnage des invertébrés du sol
- Kibblewhite, M.G., Jones, R.J.A., Baritz, R., Huber, S., Arrouays, D., Micheli, E., Stephens, M. (2008) ENVASSO final report part 1: scientific and technical activities. ENVASSO project (contract 022713) coordinated by Cranfield university, UK for scientific support to policy, European Commission 6th Framework Research Programme.
- Laval K., Mougin C., Akpa M., Barray S., Dur J-C., Gangneux C., Lebrun J., Legras M., Lepelletier P., Plassart P., Taibi S., Trinsoutrot-Gattin I. 2009. Un premier pas vers la compréhension des données biologiques. *Etude et Gestion des Sols*, Volume 16, 3/4, 275-287.
- Lavelle, P., Decaens, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, F., Mora, P., Rossi, J.P., 2006. Soil invertebrates and ecosystem services. *European Journal of Soil Biology* 42 (1), 3–15.
- Lavelle, P., Spain, A., 2006. *Soil Ecology*. Kluwer Scientific Publications, Amsterdam, The Netherlands, 651pp.
- Le Guédard M, Schraauwers B, Larrieu I, Bessoule JJ. 2008. Development of a biomarker for metal bioavailability: the lettuce Fatty Acid composition. *Environ Toxicol Chem.* 27(5):1147-51.
- Le Guédard M., Bodineau G., Houot S. et Bessoule J.J. 2009. Utilisation d'un bioindicateur lipidique végétal sur le site agricole de Feucherolles : effets de divers amendements sur les végétaux cultivés. *Etude et Gestion des Sols*, Volume 16, 3/4, 287-297.
- Leyval C., Steinberg C., Kuntz J., Beguiristain T., Edel-Hermann V, Leglize P., Gautheron N., Lebeau T. et Houot S. 2009. Impact d'amendements organiques sur la structure des communautés microbiennes des sols : choix des méthodes, validation et résultats. *Etude et Gestion des Sols*, Volume 16, 3/4, 299-312.
- Marage D. et Gégout J.-C. 2009. Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology & Biogeography*, 18: 88-97.
- Markert, B.A., Breure, A.M. and Zeichmeister H.G. 2003. *Bioindicators & biomonitors : principles, concepts, and applications*. Elsevier. Amsterdam. 997 p.
- Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M., Kibblewhite M.G. (2008) Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. *Science of the Total Environment* 391, 1-12.
- Pankhurst C., Doube B.M. and Gupta V.V.S.R. 1997. *Biological indicators of soil health*. Wallingford ; New York. CAB International. 451 p.
- Remon E., Bouchardon J-L., Joly J., Cornier B. et Faure O. 2009. Accumulation et effets des éléments métalliques sur les populations végétales spontanées de trois crassiers métallurgiques : peut-on utiliser les plantes comme bioindicateurs ? *Etude et Gestion des Sols*, Volume 16, 3/4, 313-321.

- Remon, E., Bouchardon, J-L., Cornier, B., Guy, B., Leclerc, J-C. & Faure, O. (2005) - Soil characteristics, heavy metal availability and vegetation recovery at a former metallurgical landfill: implications in risk assessment and site restoration. *Environ. Poll.* , 137 , 316-323.
- Ruiz Camacho N., Velasquez E., Pando A., Decaëns T., Dubs F. et Lavelle P. 2009. Indicateurs synthétiques de la qualité des sols. *Etude et Gestion des Sols*, Volume 16, 3/4, 323-338.
- Rutgers M., Schouten A. J., Bloem J., van Eekeren N., de Goede R. G. M., Jagersop Akkerhuis G. A. J. M., van der Wal A., Mulder C., Brussaard L., Breure A. M.. 2009 . Biological measurements in a nationwide soil monitoring network. *European Journal of Soil science*. In press
- Sanguin H., Herrera A., Oger-Desfeux C., Dechesne A., Simonet P., Navarro E., Vogel T.M., Moëgne-Loccoz Y., Nesme X. et Grundmann G. L. 2006. Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria. *Environmental Microbiology*. Vol 8 (2), 289 - 307
- Seynave, I., Gégout J.C., Hervé J.C., Dhôte J.F. 2008. Is spatial distribution of *Fagus sylvatica* limited by its potential growth? *Journal of Biogeography*, 35, 1851-1862.
- Souguir D., Goupil P., Ferjani E. et Ledoigt G. 2009. Effets génotoxiques du Cuivre chez *Vicia faba* et *Pisum sativum*. *Etude et Gestion des Sols*, Volume 16, 3/4, 339-347.
- Taïbi S., Rougé L., Thoisy-Dur J-C., Bodin J., Lepelletier P., Dantan J., Pérès G, Grand C., Bispo A. 2012. Gestion et traitement des données du programme : approche statistique et sélection d'indicateurs et de biomarqueurs dans la surveillance de la qualité des sols et l'évaluation des risques. Actes Journées Techniques Ademe Bioindicateurs&Phytotechnologies : des outils biologiques pour des sols durables, 16-17 oct 2012
- Vernay P., Austruy A., Gauthier-Moussard C et Hitmi A. 2009 Germination et fonctionnement du système photosynthétique des végétaux comme bioindicateurs de pollution des sols. *Etude et Gestion des Sols*, Volume 16, 3/4, 349-357.
- Vernay Philippe, Gauthier-Moussard Cécile, Hitmi Adnane. Interaction of bioaccumulation of heavy metal chromium with water relation, mineral nutrition and photosynthesis in developed leaves of *Lolium perenne* L. *Chemosphere* 2007; 68(8):1563-75.
- Vernay, P., Gauthier-Moussard, C., Jean, L., Bordas, F., Faure O., Ledoigt, G. & Hitmi, A. (2008) - Effect of chromium speciation on phytochemical and physiological parameters in *Datura innoxia*. *Chemosphere* , 72 , 763-771.
- Vian J.F., Peigné J., Chaussod R. et Roger-Estrade J. 2009. Effet du mode de travail du sol sur les microorganismes à l'échelle du profil cultural. *Etude et Gestion des Sols*, Volume 16, 3/4, 359-368. Vian, J.F., Peigné, J., Chaussod, R., Roger-Estrade, J., 2008. "Effects of 4 tillage systems on soil structure and soil microbial biomass in organic farming", *Soil Use Management* . Doi: 10.1111/j.1475-2743.2008.00176.x, p10.
- Villenave C., Oumar Ba A. et Rabary B. 2009. Diagnostic du fonctionnement biologique du sol par l'analyse de la nématofaune : semis direct versus labour sur les hautes terres près d'Antsirabé (Madagascar). *Etude et Gestion des Sols*, Volume 16, 3/4, 369-378.

Page de notes

Gestion et traitement des données du programme

Approche statistique de sélection d'Indicateurs et de biomarqueurs dans la surveillance de la qualité des sols et l'évaluation des risques

Salima TAÏBI¹, Laurence ROUGÉ², Jeanne-Chantal THOISY-DUR³, Jeanne BODIN¹, Patrice LEPELLETIER¹, Jérôme DANTAN¹, Guénola PÉRÈS⁴, Cécile GRAND⁵, Antonio BISPO⁵

¹ ESITPA, LAMSAD - Mont-Saint-Aignan

² Ingénierie Durable, pour l'Université de Rennes 1, CNRS "Ecobio" 6553 - Rennes

³ PESSAC UR 251 - INRA Versailles-Grignon

⁴ Université de Rennes 1, CNRS "Ecobio" - 6553 Rennes

⁵ ADEME - 20 Avenue du Grésillé - BP 90406 - 49004 Angers Cedex 01

1. Contexte et objectifs du travail

La seconde phase du programme Bioindicateurs utilise les résultats acquis par plus de 20 laboratoires sur 47 placettes de prélèvement (Pérès et al., 2012). Compte tenu du nombre très important de données (>200 000), il s'est très vite avéré nécessaire d'harmoniser et de centraliser les résultats obtenus à travers la création d'une base de données accessible de tous. Celle-ci a ainsi permis la mise à disposition des résultats pour l'ensemble des équipes en vue de croiser les données obtenues par des laboratoires différents et de faciliter la manipulation et l'analyse des données. L'objectif des traitements de données exposés ici est de hiérarchiser les bioindicateurs en fonction de leur sensibilité aux facteurs environnementaux et aux perturbations (contaminations et usage du sol) d'une part, et de proposer un indicateur agrégé de la réponse à ces facteurs d'autre part. Finalement, une démarche de sélection des bioindicateurs intégrant des critères de faisabilité pour la mise en pratique et le développement de ces outils est proposée afin de moduler les résultats précédents par la prise en compte de leur aspect technique et socio-économique. Cet article décrit brièvement la méthodologie statistique proposée par le groupe Biomath pour atteindre ces objectifs.

2. Création et fonctionnalité de la base de Données

Devant la multiplicité et l'hétérogénéité des données générées dans le cadre du programme Bioindicateurs 2 et du besoin de regrouper et d'homogénéiser l'ensemble des résultats, il a été décidé la création d'une **base de données** permettant la gestion et l'aide au traitement de l'information. Ce développement constitue une étape essentielle, car cette base est «l'outil-support» du programme, permettant de stocker en un endroit unique, sur le long terme, de manière harmonisée et structurée, l'ensemble des données recueillies. L'objectif final est de pouvoir récupérer et manipuler facilement ces données, afin d'en faciliter le traitement, en particulier multi-facteurs et inter-groupes de bioindicateurs.

2.1. Conception et développement

Un travail similaire avait été mis en œuvre dans le cadre du projet RMQS Biodiv Bretagne (première phase du programme ADEME), afin de concevoir la base de données DonECOSol (module d'EcoBioSoil, base de données de l'Université de Rennes 1). Sa structure a été pensée de manière à recueillir les données pouvant être générées dans le cadre de programmes de surveillance de la biodiversité des sols. Ce modèle de données (Cluzeau et al. 2008) a donc servi de base de réflexion et a été adapté de manière à répondre au mieux aux objectifs et aux spécificités du programme Bioindicateurs 2, identifiés au cours d'une phase essentielle d'analyse des besoins.

Concernant le développement informatique du modèle créé, le choix s'est porté sur le Système de Gestion de Bases de Données Relationnelles (SGBDR : logiciel permettant de mettre en œuvre une BDR) Microsoft Office Access®, du fait de sa simplicité d'utilisation pour des personnes non averties, leur permettant de réaliser facilement leurs propres requêtes et de personnaliser leur interface de travail (base de données personnelle) (figure 1).

Le développement de la base a débuté bien en amont des phases de prélèvement, selon un processus de co-construction avec l'ensemble des intervenants, qu'ils soient fournisseurs de données (préleveurs, gestionnaires de sites, laboratoires), ou utilisateurs potentiels (l'ensemble des équipes participantes, l'ADEME, les coordinateurs de groupe, et Biomath). En ce qui concerne les fournisseurs de données, des questionnaires ont été envoyés à chaque équipe, puis des entretiens ont été menés individuellement, afin d'identifier de manière précise l'ensemble des paramètres étudiés, ainsi que le format des données générées (types, unités...). La synthèse de ces informations a permis la construction du modèle de données spécifique, constitué de 42 tables, stockant 361 paramètres. Parallèlement, les utilisateurs ont été questionnés sur les besoins fonctionnels à intégrer : droits d'accès aux données, fonctions accessibles via l'interface... Finalement, une formation leur a été dispensée afin de se familiariser avec le fonctionnement de l'outil.

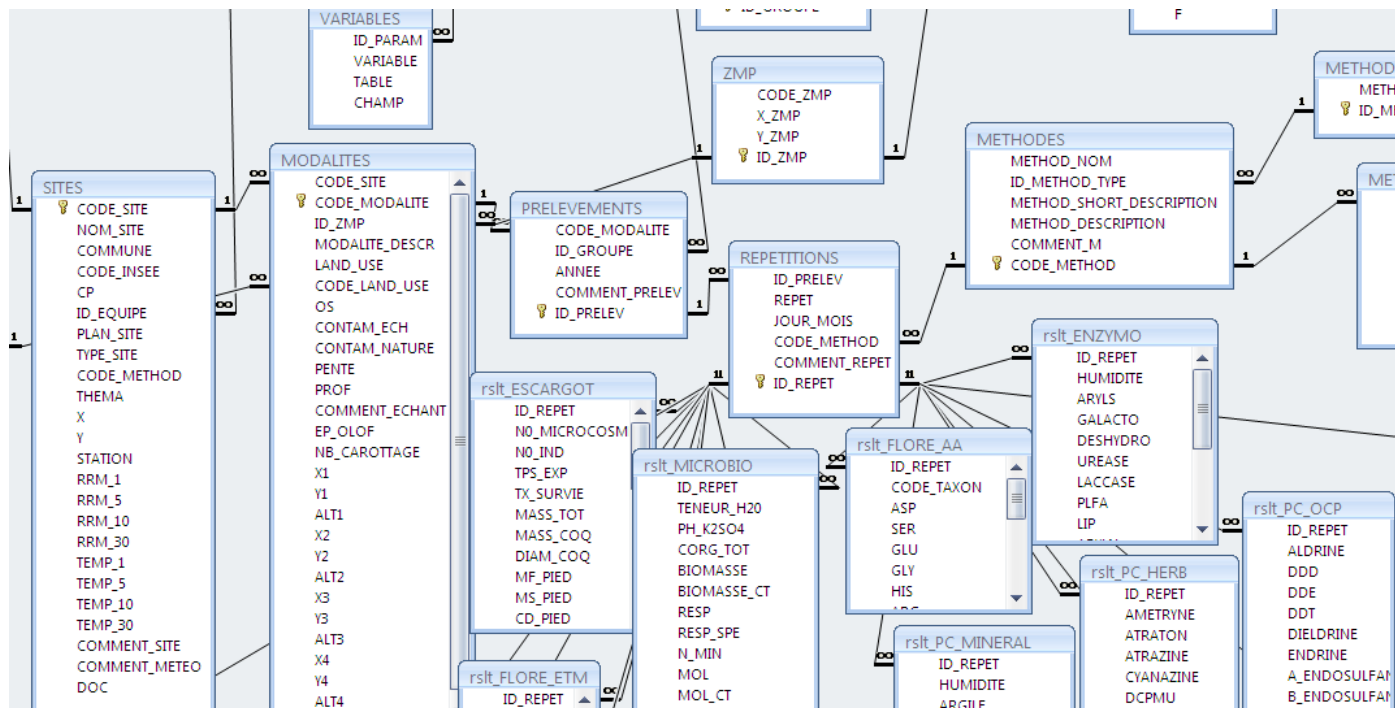


Figure 1: Extrait du modèle logique de données de la base de données BioIndicateurs 2

2.2 Saisie et validation des données : une démarche qualité

A l'issue de la phase de modélisation, des fichiers modèles de saisie en format Excel ont été créés afin de synthétiser l'ensemble des champs nécessaires à renseigner pour chaque fournisseur de données (formatage). Le transfert des données a été automatisé via la création de formulaires de mise à jour des tables afin d'éviter les erreurs de saisie. Pour améliorer la qualité de la saisie, des contrôles internes sont intégrés à certains champs (par exemple, un champ devant stocker une valeur de pH contient un contrôle interne permettant de détecter et signaler une valeur saisie non comprise entre 0 et 14), et des tests de vérification sont effectués après saisie (calculs de moyennes ou de sommes devant être comparables sur le fichier Excel et dans la base). Une codification des données manquantes (échantillon non prélevé, données aberrantes...) a été appliquée.

Au terme de la phase de saisie, une étape importante de validation des données est mise en œuvre : les différents fournisseurs sont appelés à extraire et à valider leurs données intégrées dans la base. L'ensemble des tables, des données disponibles à la consultation dans la base et leur définition sont présentées dans un document, nommé « dictionnaire des données ».

2.3. Constitution de la base de données

Les variables présentes dans la base décrivent (voir détails dans Pérès et al., 2012) :

- les caractéristiques pédologiques et pédochimiques (variables physico-chimiques),
- les contaminations en éléments métalliques et organiques,
- la bioaccumulation des éléments métalliques dans les escargots (cinétiques et statiques), la bioaccumulation des métaux dans les plantes, certains biomarqueurs, et les variables de stress des plantes pour les sites contaminés,
- les communautés des microorganismes, de la mésofaune, de la macrofaune, et les indices de diversité associés aux communautés de nématodes, de collemboles, et de lombriciens,
- la diversité microbienne,
- les activités biologiques et enzymatiques des microorganismes du sol
- les métriques paysagères et de description du paysage,

Parmi les 56 tables qui constituent la base, 42 regroupent les données mesurées in-situ par les 27 équipes partenaires. La saisie des données physico-chimiques et biologiques a révélé la diversité des variables représentant différentes formes d'indicateurs et de descripteurs du milieu environnemental. Afin de disposer de données utilisables dans le cadre d'une analyse globale, et de simplifier les démarches analytiques des partenaires, des sélections de variables ainsi que des regroupements de variables ont été décidés sur la base de l'expertise des porteurs d'indicateurs et de critères statistiques (redondances, corrélations entre variables), cela a permis la constitution d'une base de données simplifiée.

Concernant les données physico-chimiques

Les mesures réalisées décrivent la fertilité (ex : N, P, pH) mais également la contamination de chaque modalité (ex : concentrations en éléments traces métalliques, en HAPs, en pesticides organochlorés, en herbicides de famille des triazines et urées substituées). Par convention, lorsque les valeurs étaient inférieures à la limite de quantification, elles ont été remplacées par la moitié de cette valeur. L'examen des données a montré de fortes corrélations entre les HAPs, entre certains herbicides et l'aspect peu informatif de certains organochlorés. Afin de concentrer l'information, de réduire le nombre de variables et de représenter la toxicité des contaminants dans le sol, ces contaminants ont été sommés par famille chimique pour ne retenir dans la base de données simplifiée que des valeurs telles que : la somme de 4 familles d'herbicides et la somme des herbicides (triazines, urées substituées) ; la somme des HAPs légers, des HAPs lourds (plus de 4 cycles benzéniques) et la somme des HAPs pondérés par un coefficient de toxicité ; la somme des ETMs pondérés par un coefficient de toxicité.

Concernant les données biologiques

Toutes les variables biologiques présentent un format numérique à l'exception du format matriciel des variables de la diversité microbienne. Certaines des variables de mesures biologiques telles que les listes d'espèces ou de taxons ont été jugées peu pertinentes sous forme de données brutes pour l'intégration dans un traitement statistique global. Dans ce cas, le calcul d'indice ou le regroupement écologique plus intégrateur de l'information ont été retenus pour construire la base de données simplifiée. Par exemple, les communautés de faune et microfaune du sol, initialement décrites par les différentes espèces, sont in fine décrites par des abondances de groupe écologique (groupe fonctionnel) ou des indices construits sur ce principe (voir construction des indices dans les fiches outils « Les vers de terre », « Les nématodes »).

Toutes ces modifications de l'information initiale ont été réalisées par les porteurs d'indicateurs et le groupe « Biomath » afin d'éliminer les paramètres redondants ou corrélés et d'aboutir à une base de données « simplifiée » facilitant le **traitement statistique**. Ces bases de données sont accessibles à l'ensemble des partenaires du programme Bioindicateurs 2 via les interfaces spécifiquement créées.

2.4. Interfaçage

Une interface a été développée permettant la consultation conviviale, le tri et l'export simplifiés des données saisies. Les tables de la « master DataBase » ne sont pas physiquement intégrées dans l'interface mais y sont « liées ». Ce principe de fonctionnement évite de surcharger et de « polluer » la « master DataBase » et permet également à chaque utilisateur de personnaliser sa propre interface, ses propres formulaires et requêtes, sans rien modifier à la « master DataBase ».

Le fichier d'interface contient 42 requêtes et 19 formulaires permettant la consultation de la base de données complète. La base de données et son interface sont accessibles aux utilisateurs sur un site internet, sécurisées via un login et un mot de passe.

2.5. Automatisation des extractions des données

La « master Database » contient de nombreux paramètres et son modèle de données est complexe alors qu'il faut en extraire des tableaux de données directement exploitables par des logiciels de statistiques. Par ailleurs, la base ayant été alimentée en données tout au long du programme, le processus d'extraction de données a été automatisé afin de pouvoir extraire rapidement les données à chaque mise à jour de la base (Dantan et al., 2012).

Le logiciel d'automatisation développé permet d'extraire les données depuis la « master Database » puis de réaliser automatiquement des analyses statistiques préliminaires sur les données extraites (figure 2). Dans un premier temps, le logiciel lit les tables, les paramètres et les modalités dont il faut extraire les valeurs. La liste des variables à calculer est extraite d'un fichier de paramétrage externe pour éviter de recompiler le code source du logiciel. La liste des modalités est directement extraite de la table des modalités contenue dans la master Database. Enfin, pour chaque variable, les minimums, maximums, médianes, moyennes, écart-types (et autres paramètres) sont calculés puis écrits dans un tableau au format CSV.

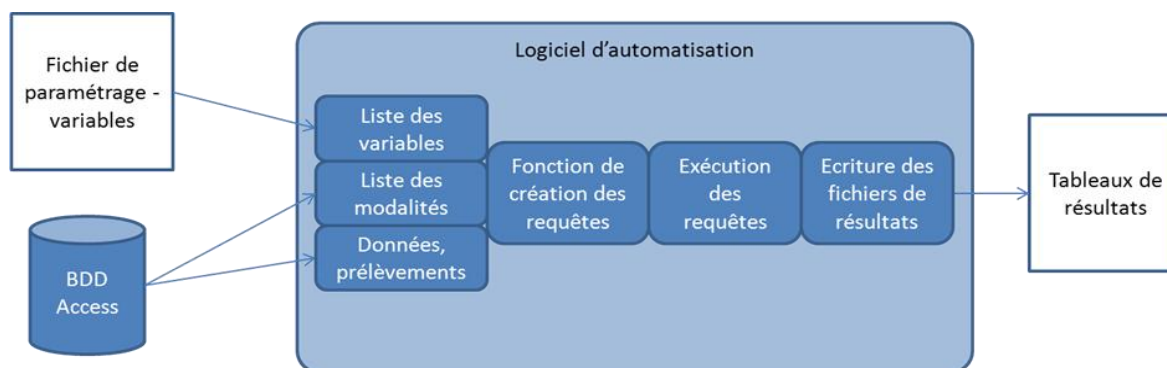


Figure 2: Architecture du logiciel d'automatisation

3. Homogénéiser le traitement des données

La qualité des données étant une question majeure et une garantie pour des résultats robustes (valeurs manquantes, aberrantes, inférieure à la limite de détection,...), il a été proposé à l'ensemble des partenaires du projet les mêmes techniques statistiques. Une première étape du traitement des données a été de proposer en 2009 un carnet de route d'analyse statistique (Taibi et al. 2009). Celui-ci est constitué de plusieurs étapes, de l'examen des données jusqu'à l'analyse multi-variée. Il a permis de poser des questions relatives à la structure de l'ensemble des données du programme et il constituait les bases nécessaires à une analyse statistique élémentaire. Un travail en appui a été aussi réalisé à la demande pour permettre le traitement des différents indicateurs.

De nombreuses questions relatives aux conditions d'application de différents types de test de comparaisons non paramétriques, au choix des paramètres de centralités et de dispersion, à l'interprétation de certains tests de l'analyse multi-variée, de l'importance des tests de corrélation ont été discutées tout au long du programme. Ces concertations ont constitué un préliminaire à la mise en place de la méthodologie appropriée pour l'analyse de l'ensemble des données. Succinctement, quelques méthodes utilisées sont présentées. Elles ont été proposées aux partenaires du projet pour des analyses propres à chaque porteur de groupe biologique :

- Les valeurs aberrantes ont été identifiées grâce à l'élaboration de boîtes à moustaches ou boîtes de Tukey et au test de Dixon. La médiane a été choisie comme caractéristique de tendance centrale. En effet celle-ci n'est pas affectée par la présence des valeurs aberrantes.
- L'intervalle interquartile renseigne sur la dispersion. Les comparaisons et tests d'influence d'un facteur ont été établis à l'aide de tests non paramétriques tels que le test de Kruskal-Wallis ou Mann-Whitney suivis de tests de classements tels que le test de Dunn.

- Pour diagnostiquer et évaluer les corrélations entre les différentes variables, on a utilisé les coefficients de corrélation des rangs de Spearman. En effet les liaisons ne sont pas toujours linéaires et ce coefficient est moins affecté par la présence de valeurs aberrantes.
- Les analyses factorielles de type ACP, AFC ou ACM ont été mises en œuvre pour les analyses multivariées (Thoisy-Dur et al., 2012).
- Pour les méthodes de classification non supervisée, la Classification ascendante hiérarchique en utilisant l'ultramétrie de Ward a été appliquée. L'analyse discriminante linéaire de Fisher a été utilisée pour classer des modalités et sélectionner les variables les plus discriminantes.

Cette méthodologie d'analyse statistique élémentaire constitue un préalable pour une analyse globale mettant en commun l'ensemble des données, afin de mettre en évidence la sensibilité de l'ensemble des variables biologiques aux facteurs environnementaux et aux perturbations définies dans le cadre de ce programme.

4. Démarche d'analyse globale

Une démarche d'analyse globale est proposée afin de déterminer la part des perturbations anthropiques (comme la contamination) devant celles des facteurs environnementaux, tel que la variabilité pédologique naturelle, les usages du sol mais aussi la structure du paysage, sur la réponse des organismes du sol. Une sélection des diverses méthodologies connues a été effectuée afin de s'adapter à la structure du jeu de données utilisé et de mettre en valeur la richesse des informations disponibles.

4.1 Définition de la base de référence

Afin de proposer un ordre de grandeur des variables biologiques mesurées en fonction de différents facteurs environnementaux, les valeurs bornées de ces variables ont été décrites dans une base de référence. Les premiers et derniers déciles (excluant les 10% des valeurs les plus basses et les 10% des valeurs les plus hautes) ont été choisis pour la définition de ces bornes afin d'exclure les valeurs extrêmes. Les valeurs en sites contaminés sont différenciées des valeurs en culture, forêts et prairies non contaminées. Une présentation (figure 3) illustrant ces gammes de valeurs a été préparée pour l'ensemble des partenaires.

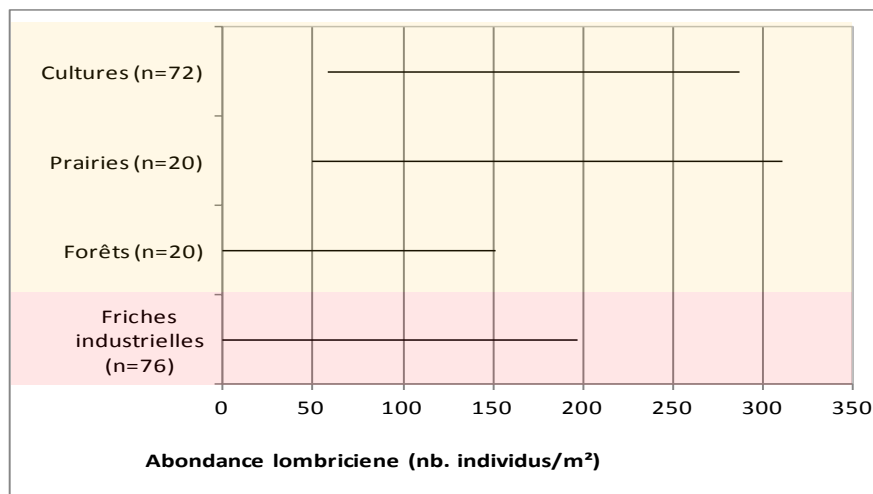


Figure 3. Valeurs de référence, pour les lombriciens dans les cultures, les prairies, les forêts et les sites influencés par des activités industrielles (actuelles ou passées) définissant des zones contaminées et regroupées sous le terme « friches industrielles ».

4.2 Intérêt des gradients du jeu de données

L'objectif étant de définir des batteries d'indicateurs sensibles à des facteurs anthropiques, un travail de définition des facteurs à tester a été mené par le groupe Biomath. Les modalités des sites ont été définies *a priori* selon des gradients de contaminations pour les sites pollués d'une part, et d'autre part selon des gradients de nutriments (carbone, azote) ou différentes pratiques agricoles (labour/non labour, Raisonné/Biologique/Economique, gestion des rotations culturales). Ces différents gradients sont parallèlement croisés avec des occupations différentes des sols définies selon Corine Land Cover et distinguant cultures (C), prairies (P), forêts (F), et bois, friches, ourlets boisés (B),

Afin de développer un traitement statistique approprié (ex : analyses discriminantes et multivariées) permettant de sélectionner des batteries d'indicateurs sensibles, à ces différents facteurs, il a été défini :

a. gradients de nutriments et des pratiques agricoles :

- Teneur en azote (utilisation de tous les sites agricoles),
- Teneur en carbone organique (utilisation de tous les sites agricoles),
- Rotation et logique de conduite sur prairies (utilisation des données du site d'Yvetot, « Yv »),
- Travail et profondeur de labour (utilisation des données du site de Thil, »Th »),
- Logique de conduite sur verger (utilisation des données du site de BioReco-Gotheron, « GO »),
- Type d'apport de matières organiques exogènes (Produits Résiduaires Organiques) (utilisation des données du site de QualiAgro-Feucherolles, « Fe »).

b. gradients de contaminations :

- Concentration en polluants organiques : HAPs pondérés par leurs degrés de toxicité, teneurs en triazines, (utilisation de tous les sites contaminés),
- Concentration en polluants métalliques : somme des ETMs pondérés ou non par leur toxicité (utilisation de tous les sites contaminés),
- Somme des excès de transfert des ETMs dans les escargots, pour tester une bioaccumulation dans les organismes indicatrice de toxicité potentielle, (utilisation de tous les sites contaminés),
- Contaminations croisées HAPs - ETMs (utilisation de tous les sites contaminés),
- Contamination croisée avec l'âge du recouvrement végétal (utilisation des données du site de St Etienne « Ste »),
- Contaminations croisées aux usages forêts et cultures (utilisation des données du site de Metaleurop « Me »),
- Contaminations à l'Arsenic croisées aux usages des ourlets boisés (utilisation des données du site d'Auzon « Au »).

Ces différents gradients sont utilisés sous forme de valeurs numériques pour former des classes. Les classes sont utilisées en l'état ou construites en s'aidant de référentiels extérieurs (définition des bornes de chacune des classes)

- selon la méthode de classification de Ward pour l'ensemble des contaminations organiques, sur tous les sites,
- en combinant les valeurs des vibrisses internes et externes des différents éléments métallique d'après les données du RMQS (Villaneau et al., 2008) et les données INRA-ASPITET pour l'arsenic (Baize, 2000) pour la contamination métallique sur tous les sites,
- en fonction de la distribution des valeurs des gradients définis ci-dessus, sur les sites agricoles et/ou sites pollués concernés par la problématique étudiée (ex : non labour/ labour moyen / fort labour).

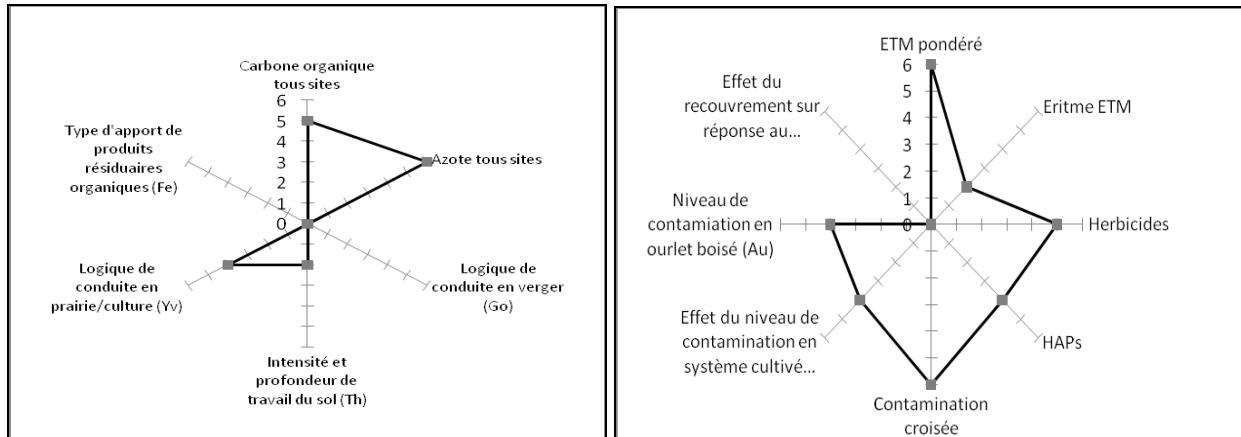
Ces classes permettent alors, dans les analyses décrites ci-après, de mettre en évidence la réponse des organismes du sol en fonction de leurs variations.

4.3. Hiérarchisation de la sensibilité de l'indicateur

Dans le but de sélectionner des indicateurs sensibles, nous avons cherché à mettre en évidence les réponses significatives de 74 variables biologiques à une perturbation liées à la contamination chimique ou une pratique agricole. Les tests sont réalisés sur le jeu de données propre à la problématique abordée, les sites pollués et les sites agricoles sont donc étudiés indépendamment. La distribution des valeurs de chaque variable biologique est testée pour chaque facteur défini ci-dessus, par un test de Kruskal-Wallis ou selon le cas par un test de Mann-Whitney. Les variables dites de bioaccumulation et d'exposition aux métaux sont testées sur sites pollués seulement.

Les variables prises indépendamment sont testées pour chaque facteur. Un score est défini selon le résultat de la p-value pour ordonner les variables biologiques : score=6 pour $p < 0,001$; score=5 pour $p < 0,01$; score=4 pour $p < 0,05$; score=2 pour $p < 0,1$ et score=0 pour $p \geq 0,1$. Les résultats de ces scores pour l'ensemble des facteurs testés sont alors représentés sur des diagrammes en radar pour chaque variable biologique (figure 4). Une vue d'ensemble de la réponse de chaque indicateur aux différents facteurs est ainsi obtenue et permet de rendre compte de l'ubiquité de certains paramètres biologiques ou à l'opposé de leur spécificité à une question/gradient (plus la zone de recouvrement du radar est importante, plus l'indicateur répond à l'ensemble des gradients étudiés).

Figure 4 : réponse de deux indicateurs (ADNR 18S à gauche et abondance d'une catégorie écologique de lombriciens à droite) aux facteurs de perturbations sur sites agricoles (à gauche) et sur sites pollués (à droite). Plus la taille de la branche du radar est grande, plus la réponse au facteur est significative.



4.4. Rôle du contexte pédologique

Les variables sélectionnées par les analyses précédentes sont potentiellement indicatrices d'une pratique ou d'une perturbation. Les organismes et la biodiversité du sol étant répartis selon les conditions environnementales telles que les facteurs pédologiques et les types de sol, ceux-ci doivent être pris en compte dans la modélisation de la réponse d'un indicateur. La part de la variabilité naturelle dans la réponse de l'indicateur à la perturbation est ainsi prise en compte. Un tel modèle de réponse de l'indicateur s'écrit donc de la manière suivante, la perturbation constituant le facteur à mettre en évidence. :

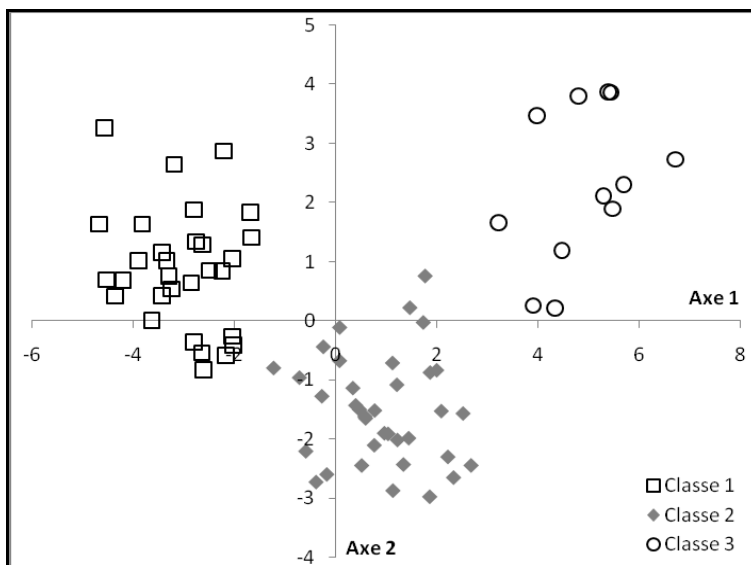
$$\text{Indicateur biologique} = \text{perturbation} + \text{facteurs pédologiques naturels} + \text{usage du sol}$$

4.5. Sélection de batteries d'indicateurs

Les analyses factorielles et les analyses discriminantes réalisées dans le cadre du programme ont pour objectif de considérer l'ensemble des variables biologiques et/ou abiotiques de la base de données. Leur principe est de déterminer une liste de variables dépendantes qui permettra d'expliquer les variations d'un facteur donné, et par extension, la liste des variables les plus impactées par ce facteur. Le grand nombre d'observations indépendantes définies dans le programme et le très grand nombre de variables biologiques à tester (une centaine) nécessite des techniques d'analyse telles que les Random Forests qui acceptent les multi-collinéarités afin d'aboutir à la sélection de batteries d'indicateurs sensibles aux différents facteurs.

A titre d'exemple, une sélection de variables sensibles aux contaminations métalliques par les ETM en inter-sites (plusieurs sites contaminés par des niveaux variables de contamination ont été pris en compte), et déterminées par la méthode des Random Forests (Laroutis et Taibi 2011) ont montré que certaines catégories écologiques de lombriciens et de nématodes mais aussi certaines activités enzymatiques, discriminent clairement le niveau de contamination en ETM (Thoisy-Dur et al., 2012). Par ailleurs, on peut également montrer les variables impactées par les teneurs en Carbone organique sur sites agricoles par une analyse discriminante (figure 5).

Figure 5 : Analyse discriminante sur les classes de Carbone organique, illustrée le long des axes 1 et 2 : les variables biologiques permettant d'expliquer ensemble la teneur en carbone organique sont : le rendement d'extraction de l'ADN du sol, la biomasse microbienne, l'ergostérol libre, la phosphatase alcaline, la phosphatase acide, la lippase, la galactosidase, la respiration spécifique, la respiration, la fda, le nombre d'unité formant colonies, l'op16S, la biomasse rapportée au carbone total, le ratio plfa saturés/monoinsaturés, les plfa gram-, l'arylamidase, la laccase, l'azote minéralisable, les plfa totaux et la xylanase.



4.6 Vers un outil d'aide à la décision

La part de chaque facteur (perturbation anthropique, variabilité pédologique naturelle, usages du sol mais aussi structure, fragmentation et hétérogénéité du paysage) dans la sélection des indicateurs sensibles sera hiérarchisée par des analyses multivariées et multifactorielles. La mesure complète des variables environnementales en complément des variables biologiques sur toutes les observations rend cette analyse effective.

Finalement, une analyse du rôle écologique des variables sélectionnées est proposée afin de définir quelles sont les composantes de l'activité biologique impactée : la biodiversité, l'implication dans un cycle biogéochimique, une fonction du sol (structure, décomposition, aération, résistance aux pathogènes...). Ce type d'analyse doit mener à des résultats intégrant non seulement les perturbations directes, mais aussi les facteurs environnementaux, et permettra d'améliorer les outils d'aide à la décision pour la gestion de sites agricoles ou pollués.

5. Outils d'opérationnalité de l'indicateur

Dans une optique de développement de l'outil, les résultats précédents de sensibilité des indicateurs biologiques nécessitent d'être replacés dans un contexte de faisabilité technique d'une part, mais aussi de compréhension par le public des utilisateurs. Il a donc été décidé d'établir des scores permettant d'apprécier l'importance de ces critères dans le choix de l'utilisation de l'outil.

Au-delà de la capacité d'un indicateur à distinguer des différences entre sites contrastées, de se comporter différemment selon des usages du sol ou encore de donner des informations sur l'état d'un sol, il est aussi important que celui-ci soit opérationnel et réponde aux critères des décideurs (Ritz et al , 2008). Une première enquête a été menée auprès des responsables d'indicateurs. Le questionnaire est structuré en trois grandes parties rassemblant des critères scientifiques, techniques et socio-économiques. Les questions posées sont de type binaire (oui ; non) suivies d'une question ouverte « texte ». Une seconde enquête a été lancée auprès de deux groupes d'utilisateurs issus du comité d'orientation du programme Bioindicateurs 2, ceux des sols agricoles (Agro) et ceux des sites et sols pollués (SSP). Les répondants devaient indiquer une note sur une échelle de 1 à 5 (1 prioritaire à 5 le moins important) et ce pour chaque critère cité. Les traitements qui en sont issus permettent de pondérer les critères en fonction des utilisateurs (Agro ou Sites et Sols Pollués).

Chaque indicateur renseigné peut être traité de la manière suivante : on somme le nombre d'occurrences positives par batterie de critères. Par exemple pour la question « Facilité d'acquisition des résultats », une réponse positive indique une note de (+1). Alors qu'une réponse négative donne une note nulle.

Le score final pour chaque indicateur résulte des formules ci-dessous. Soient C_i les critères étudiés et W_i les coefficients standardisés associés aux groupes de critères,

$$S_{\text{scien}} = \text{Somme des notes } (C_{\text{scien}}) \times W_{\text{scien}}$$

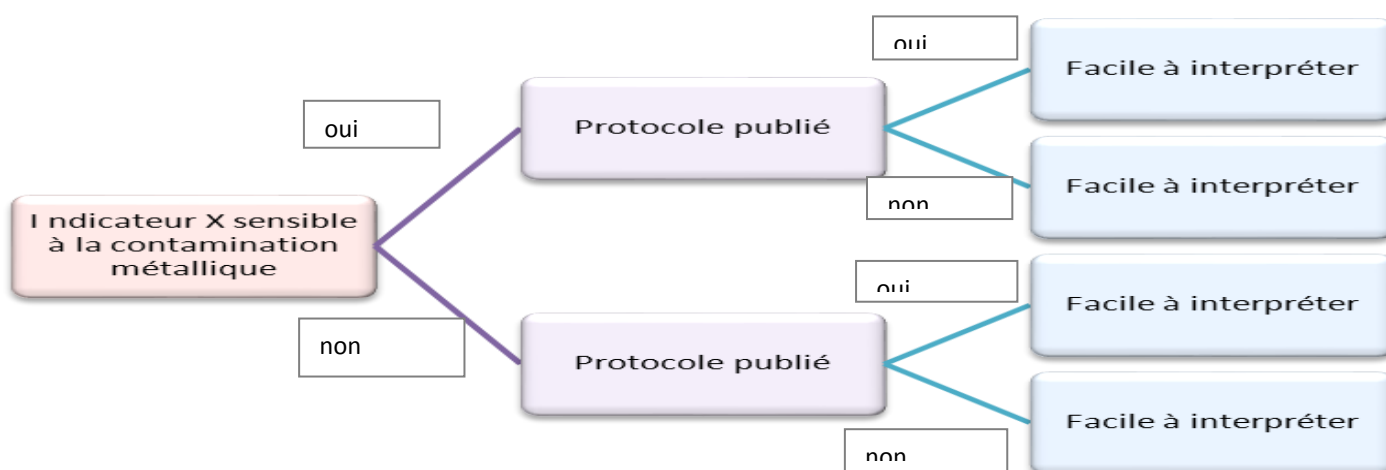
$$S_{\text{tech}} = \text{Somme des notes } (C_{\text{tech}}) \times W_{\text{tech}}$$

$$S_{\text{éco-socio}} = \text{Somme des notes } (C_{\text{éco-socio}}) \times W_{\text{éco-socio}}$$

Le dépouillement et le traitement de l'enquête utilisateurs nous fournissent les rangs moyens notés R_i et ce pour chaque groupe d'utilisateurs (Agro ou Sites et Sols Pollués). Nous déduisons donc les poids W_{tech} et $W_{\text{éco-socio}}$. Le coefficient W_{scien} découle quant à lui des calculs de sensibilité. Nous pouvons assigner par exemple une valeur résultant des tests statistiques. Il est à préciser qu'il faut attribuer un poids plus élevé à la sensibilité d'un indicateur qu'aux critères techniques ou socio-économiques.

Finalement, le score est donné par : $\text{Score (total)} = S_{\text{scien}} \text{ final} (S_{\text{tech}} \text{ final} + S_{\text{éco-socio}} \text{ final})$

L'analyse de l'arbre de décision ci-dessous permettra de confirmer les résultats obtenus lors du calcul des scores de chaque indicateur.



6 .Conclusion

Par une approche pluridisciplinaire, le groupe Biomath a établi un carnet de route et un canevas de démarches communes d'analyses destinés à l'ensemble de responsables d'indicateurs, des groupes biologiques ou des responsables de sites. La mise en œuvre par les utilisateurs s'est trouvée facilitée du fait de l'accès à une base de données commune, ainsi qu'à la matrice (réduite) des paramètres physico-chimiques. De plus tous les outils biologiques ont été mis en œuvre sur les mêmes sites et au même moment par les mêmes gestionnaires.

Les travaux de recherche « mono indicateur » permettent d'établir les valeurs de références, de mettre en évidence la sensibilité à un environnement donné (pratiques agricoles et usages/stress chimique) et de classer les indicateurs tout en prenant en compte la pédologie.

L'approche globale ainsi que la démarche pour la sélection d'une batterie d'indicateurs explicatifs et discriminants pour un stress donné à l'aide des méthodes non supervisées telle que la méthode des forêts aléatoires (Laroutis et Taïbi, 2011) donnent des résultats probants dans des contextes d'évaluation de l'état du sol à la fois complexe et dynamique.

Le sondage des utilisateurs *versus* experts apporte une plus-value pour le choix d'un indicateur dans le sens où le critère de sensibilité est certes nécessaire mais non suffisant, et ceci doit être complété par des études confirmant son opérationnalité. Finalement la complémentarité entre réponses à dire d'experts, celles des utilisateurs et les résultats fondés sur les analyses statistiques permettent de calculer un score final pour chaque indicateur.

Nous projetons de construire un indice fonctionnel d'état d'un sol à l'aide d'une démarche établie qui tienne compte de l'intégration des propriétés physicochimiques, microbiologiques, biologiques nécessaire pour établir cet indice (Taïbi et al., 2011 & 2012). Cet indice peut être défini comme le plus petit ensemble de paramètres, qui, mis en relation permet de renseigner sur la capacité d'un sol à avoir une fonction donnée.

7. Bibliographie

- Baize D., (2000). Teneurs totales en "métaux lourds" dans les sols français. Résultats généraux du programme ASPITET. *Le Courrier de l'Environnement de l'INRA*. 39, 39-54
- Bastida F., Zsolnay A., Hernandez T., Garcia C. (2008) Past, present and future of soil quality indices: A biological perspective. *Geoderma* Vol 147, Issue: 3-4, Pages: 159-171.
- Cluzeau D., Péres G., Cannavacciuolo M., Bellido A., Guernion M., Ruiz N., Cortet J., Mateille T., Martin-Laurent F., Velasquez E., Mercier V., Bispo A., Villenave C., Ranjard L., Chaussod R., Rougé L., Jolivet C., Lermecier-Foucault B., Ponge J.F., (2008). How to manage and analyse a large biodiversity data set: the case of the regional "RMQS BioDiv" experience. *International Colloquium Soil Zoology Curitiba, Brésil- 25-29 Août 2008*
- Dantan J., Polet Y., Taibi S. (2012). A KDD process to retrieve and aggregate data from relational databases. IADIS.
- Laroutis, D., Taibi, S., (2011). Discriminant analysis versus Random Forests on qualitative data: Contingent valuation method applied to the Seine estuary wetlands. *International Journal of Ecological Economics & Statistics*. 20, 1-13.
- Ritz K., Black Helaina I.J. Colin D. Campbell, Harris J.A, Wood C. (2008). Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. *Ecological Indicators*, 9, 1212-1221.
- Pérès G., Bispo A., Grand C., Galsomies L., (2012). Le programme de recherche ADEME "Bioindicateurs de l'état biologique des sols". Ses objectifs, sa mise en œuvre et son déroulement. *Actes Journées Techniques Ademe Bioindicateurs&Phytotechnologies : des outils biologiques pour des sols durables*, 16-17 octobre 2012.
- Taibi S., Lepelletier P., Dur J.-C. (2009). Carnet de route en vue d'une démarche statistique commune à tous les groupes biologiques, 9 pages. Document interne au programme ADEME BIOINDICATEURS Phase II.
- Taibi S., Lepelletier P., Pérès G., Rougé L., Dur J.-C., Bispo A.. (2011). Démarche en vue d'élaborer un indice d'état du sol. Journée de statistique de la Société Française de Statistiques, Gammarth, Tunisie. Communication orale.
- Taibi S., J.C. Thoisy-Dur, P.Lepelletier,,Rougé L, J. Dantan, A.Bispo, Grand C., Pérès G., (2012). Approche statistique de sélection d'Indicateurs et de Biomarqueurs dans la surveillance de la qualité des sols et l'évaluation des risques. Journées d'Etude des Sols, 13-23 mars 2012, Versailles. Communication orale.
- Thoisy-Dur J.C., Lepelletier P., Taibi S., Rougé L., Dantan J., Pérès G., Grand C., Bispo A. (2012). Statistical approach to select soil bioindicators for soil monitoring, risk assessment and soil characterization. Results from the French national Programme "Bioindicators". 6th SETAC World Congress, Berlin. Poster.
- Villanneau E., Perry-Giraud C., Saby N., Jolivet C., Marot F., Maton D., Floch-Barneaud A., V. Antoni et Arrouays D. (2008). Détection de valeurs anormales d'éléments traces métalliques dans les sols à l'aide du Réseau de Mesure de la Qualité des Sols. *Etude et Gestion des Sols*, 15, 183-200.