

MÉTHODOLOGIE GÉNÉRALE

2.1 Les données floristiques

Pendant le déroulement de cette thèse nous avons utilisé trois bases de données floristiques et écologiques afin de suivre et spatialiser l'évolution de l'acidité et la disponibilité en azote, la base EcoPlant (Gégout *et al.*, 2005), la base Sophy (Brisse *et al.*, 1995), et l'Inventaire Forestier National (Robert *et al.*, 2010). Elles constituent les plus grosses bases de relevés floristiques forestiers créées à l'échelle de la France.

2.1.1 Les données issues d'EcoPlant

La base de données EcoPlant regroupe des relevés floristiques forestiers accompagnés de paramètres écologiques (notamment des analyses de sol). L'objectif de cette base est d'étudier et de modéliser la distribution des espèces végétales forestières, ainsi que leur comportement et caractère indicateur vis-à-vis des facteurs climatiques et édaphiques (Gégout *et al.*, 2005). Les données sont issues de catalogues des stations forestières, des réseaux de suivi des écosystèmes forestiers européens ainsi que des travaux de thèse ou de recherche. Chaque relevé a été réalisé, dans la plupart des cas, sur des placettes d'une surface de 400 m². Pour chaque placette il existe une description floristique ainsi que des caractéristiques pédologiques, géographiques (notamment la localisation) et topographiques réalisée à la date de l'observation. Les placettes correspondent donc à des unités représentant l'espace et le temps. Parmi l'information contenue dans chaque placette de la base EcoPlant, on retrouve un relevé floristique exhaustif où soit la présence soit l'abondance/dominance des espèces est renseignée (suivant la méthode de Braun-Blanquet (1932) dans ce dernier cas). Les coordonnées géographiques des sites sont déterminées soit par localisation GPS soit par estimation à partir des cartes topographiques ou des descriptions précises des sites. La précision de localisation et l'altitude sont aussi renseignées. Un profil de sol, et des analyses physico-chimiques (plus ou moins complètes) sont réalisés pour la majorité des relevés. Les analyses chimiques du sol comprennent (entre autre) une mesure de l'acidité (pH), la teneur en Al³⁺ et CaCO₃, le taux de saturation en bases, l'azote total, le carbone organique et le phosphore assimilable. À ce jour, la base contient 6266 placettes avec des relevés floristiques.

2.1.2 Les données issues de Sophy

La base Sophy a été réalisée en recueillant les relevés phytosociologiques contenus dans de nombreuses publications scientifiques. Son objectif est de décrire les groupes socio-écologiques des végétaux (Brisse *et al.*, 1995). Elle comporte actuellement plus de 100 000 relevés obtenus sur la totalité du territoire français, soit en milieu ouvert ou forestier. Les relevés sont dans la plus part des cas réalisés sur des placettes de 400 m². Parmi l'information contenue dans la base de données Sophy, on retrouve un relevé floristique exhaustif codé en présence/absence ou en abondance/dominance. Il est également renseigné les coordonnées géographiques du relevé ainsi que la précision de leur estimation. Aucune analyse physico-chimique du sol n'est fournie.

2.1.3 Les données issues de l'inventaire forestier

L'inventaire forestier, géré par l'Institut National de l'Information Géographique et forestière (IGN), collecte des données permettant de connaître l'état et l'évolution au cours du temps de la ressource forestière française (<http://inventaire-forestier.ign.fr/spip/spip.20php?rubrique149>). La méthode d'échantillonnage de l'inventaire forestier entre 1976 et 2004 consistait en cycles d'inventaires successifs couvrant département par département, et suivant une grille systématique (1 x 1 km), le territoire forestier français métropolitain. Selon ce protocole la totalité du territoire est couverte tous les 10-12 ans (Robert *et al.*, 2010). Des relevés phytoécologiques ont été intégrés aux inventaires dendrométriques de l'inventaire forestier à partir de 1987, puis généralisés sur l'ensemble de la France à partir de 1992. En 2004, une nouvelle méthode d'échantillonnage a été adoptée, couvrant la totalité du territoire forestier français chaque année. L'avantage de cette nouvelle méthode est d'être plus souple et de s'adapter facilement à de multiples découpages spatiaux. Elle se base sur une grille systématique de 10 x 10 km, qui est déplacée de 2 kms par an, assurant ainsi une couverture kilométrique tous les 10 ans (Robert *et al.*, 2010). En d'autres termes, ce nouveau protocole procure un échantillonnage régulier dans l'espace et dans le temps. Il est fractionné en échantillons annuels présentant la même propriété de répartition uniforme dans l'espace (**Fig. 2-1**). Sur chaque placette forestière de 700 m², l'IGN réalise des prélèvements concernant le peuplement forestier, la végétation codée en présence/absence pour chacune des espèces végétales présentes et les conditions géographiques et physiques (pente, exposition, sol, etc.). Ceci s'accompagne également de mesures sur les arbres (hauteur, diamètre, etc.) et un

descriptif des peuplements. Dans ce travail de thèse nous avons utilisé la totalité des relevés échantillonnés et géoréférencés entre 1987 et 2009, soit 142 429 placettes floristiques.

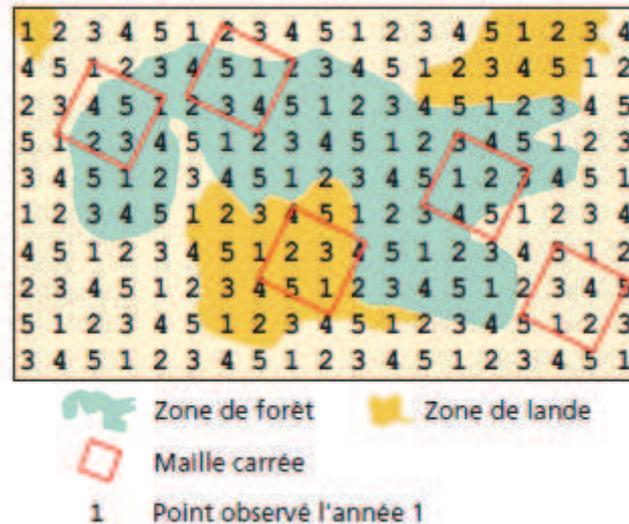


Figure 2-1 Échantillon complet systématique dans l'espace et dans le temps. Indifféremment de l'année, la maille est la même (tiré de <http://inventaire-forestier.ign.fr/spip/spip.php?20rubrique25>).

2.1.4 Les données collectées dans la forêt de Villey-Le-Sec

Un échantillonnage systématique initiale a été réalisée à la fin de juillet 1972 dans le massif forestier de Haye (y comprise la forêt de Villey-Le-Sec, située à l'ouest du massif de Haye) dans le nord-est de la France lors d'une étude précédente (Becker, 1979). L'échantillonnage se basait sur une grille de 250 x 250 m. À chaque intersection de cette grille un inventaire floristique complet a été mené sur une placette circulaire de 400 m². L'abondance/dominance des espèces a été renseignée suivant la méthode de Braun-Blanquet (1932). De toutes les placettes établies sur le massif de Haye, un total de 57 placettes était localisé dans la forêt de Villey-Le-Sec. 47 de ces placettes ont été ré-échantillonnées en mai 1991 (Thimonier *et al.*, 1994), puis au cours de cette thèse en juin/juillet 2011. Lors de cette dernière campagne, seules les 40 placettes pour lesquelles aucune perturbation majeure d'origine anthropique (e.g. coupes, sentiers de randonnées) et/ou naturelle (tempêtes Lothar et Martin de 1999) n'a été observée depuis 1991 ont été ré-inventoriées. Des marques de peinture laissées sur les arbres à l'issue de la campagne de 1972 et la cartographie des placettes inventoriées lors de la campagne de 1991 ont permis l'identification et la localisation précises des relevés durant les campagnes de 1991 et 2011, respectivement. Les relevés floristiques réalisés tant en 1991 qu'en 2011 ont suivi la méthodologie de l'étude initiale de 1972 (Becker,

1979). Des mesures de pH (H₂O) du sol ont été réalisées en complément en 1991 (Thimonier *et al.*, 1994) et 2011. En 1991, ces mesures ont été obtenues à partir d'échantillons de l'horizon supérieur de sol (Thimonier *et al.*, 1994). En 2011, le pH des placettes forestières a été mesuré à partir d'échantillons de sol prélevés en trois points différents (choisis aléatoirement) dans chaque placette de 400 m². Lors de cette dernière campagne, les relevés ont été complétés par de nombreuses mesures des propriétés physico-chimiques du sol (**Table 2-1**), la description des peuplements (renseignant notamment sur la gestion sylvicole) et des données topographiques.

Table 2-1 Présentation sommaire des données collectées lors des différentes campagnes de terrain dans la forêt de Villey-Le-Sec.

Campagne	Date	<i>n</i> placettes	Données disponibles	Auteur
1	Juillet 1972	57	- Inventaire floristique	Becker, 1979
2	Mai 1991	47	- Inventaire floristique - pH (H ₂ O)	Thimonier <i>et al.</i> , 1994
3	Juin/Juillet 2011	40	- Inventaire floristique - pH (H ₂ O) - Carbone organique - Azote total - Éléments grossiers - Graviers - Sables grossiers - Sables fins - Limons grossiers - Limons fins - Argiles - Calcaire total - Calcaire actif - Calcium échangeable - Magnésium échangeable - Potassium échangeable - Sodium échangeable - Aluminium échangeable - Manganèse échangeable - Fer échangeable - Protons échangeables - CEC pH sol - Matière organique (g/kg)	Riofrío-Dillon, 2013

2.2 Traitement des données floristiques

Étant donné la nature des bases de données, nous avons sélectionné les relevés spécifiquement forestiers en excluant ceux de milieu ouvert. Tout d'abord, nous nous sommes basés sur le territoire forestier défini par CORINE Land Cover (pour les années 1990, 2000 et 2006, à échelle 1 : 1 000 000), c.-à-d. les forêts de feuillus (code 311), les forêts de conifères (code 312) et les forêts mélangées (code 313), pour estimer la couverture spatiale des forêts françaises. Ensuite, les relevés géoréférencés et le territoire forestier (défini comme ci-dessus) furent croisés en prenant en compte la date de l'actualisation de CORINE Land Cover la plus proche de celle des relevés. Les relevés localisés dans la zone forestière ont été sélectionnés. La couverture du territoire forestier la plus ancienne dont nous disposons date de 1990, ce qui se traduit par une imprécision dans la sélection des relevés forestiers échantillonnés précédemment. Pour limiter cette incertitude nous avons conservé les relevés pré-sélectionnés sur lesquels au moins une essence forestière était inventoriée.

Finalement, nous avons extrait d'EcoPlant 6196 relevés forestiers géoréférencés compris entre 1910 et 2010, et issus de 142 études différentes, dont 3280 et 2816 placettes contiennent en outre des analyses physico-chimiques comprenant l'acidité et la disponibilité en azote du sol, respectivement, déterminées en laboratoire (**Fig. 2-2a**). De la base Sophy, nous avons extrait 32 330 relevés forestiers géoréférencés sur la période 1915-2000 (**Fig. 2-2b**). En tout, notre base de données forestière, comprenant également les placettes de l'inventaire forestier (**Fig. 2-2c**), regroupe un total de 180 955 relevés floristiques sur la période 1910-2010.

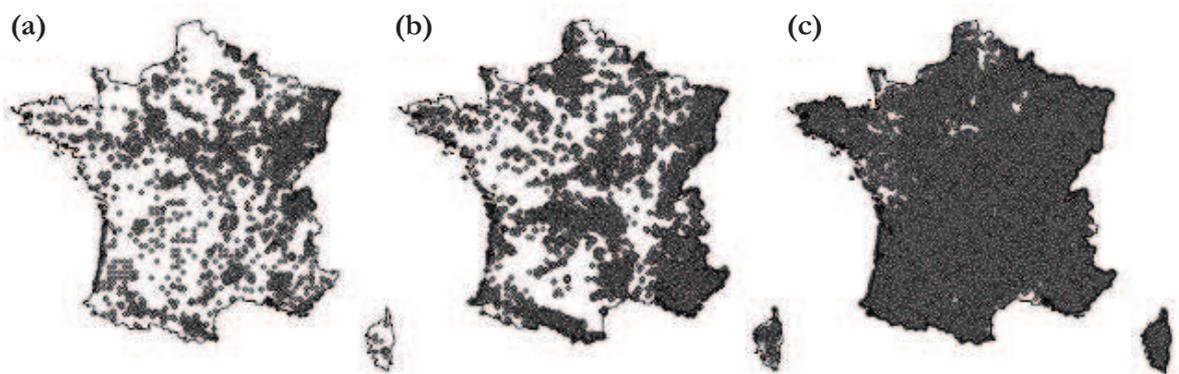


Figure 2-2 Répartition spatiale des données floristiques forestières issues des bases de données **(a)** EcoPlant ($n = 6196$ placettes), **(b)** Sophy ($n = 32\,330$ placettes), et **(c)** de l'inventaire forestier national ($n = 142\,429$ placettes).

Comme chaque base de données a été réalisée à partir d'une nomenclature taxonomique différente, une procédure d'homogénéisation des noms d'espèces a été effectuée afin de vérifier et, si nécessaire, les mettre à jour. À l'origine, nous avons compté 6887 noms d'espèces (ou sous-espèce et variété différentes) dans les trois bases de données et pour lesquels nous avons créé une nomenclature commune. Pour vérifier les synonymes des noms de plantes nous avons utilisé une base de synonymes interne au LERFoB complétée d'après la Base de Données Nomenclaturale de la Flore de France (BDNFF version 4 réalisée par Benoît Bock et Michel Kerguelen), la base Baseflor (Julve, 1998), et les trois tomes de la Flore Forestière Française (Rameau *et al.*, 1989; Rameau *et al.*, 1993; Rameau *et al.*, 2008). À la fin du traitement taxonomique, nous avons obtenu 5491 noms d'espèces (ou sous-espèce et variété différentes). Pour éviter les problèmes concernant des erreurs d'identification, nous nous sommes concentrés sur le niveau taxonomique de l'espèce, soit un pool de 4257 espèces.

2.3 Les données modélisées des dépôts atmosphériques

Il existe peu de données mesurées pour documenter les réponses de l'environnement aux dépôts atmosphériques depuis l'époque de la révolution industrielle (cf. Johnston *et al.*, 1986). Face à ce manque, des données modélisées des dépôts atmosphériques (e.g. EMEP, 2011) sont fréquemment utilisées pour estimer les effets de l'acidification et de la disponibilité en azote passés et futurs sur les écosystèmes terrestres et aquatiques (Driscoll *et al.*, 2001; Dentener *et al.*, 2006a). Dans cette thèse, les données des dépôts atmosphériques de soufre et d'azote total modélisées par l'EMEP (**Fig. 2-3**) sur la période 1910-2010 ont été utilisées pour établir la relation spatio-temporelle entre l'acidité et la disponibilité en azote bio-indiquées par la flore et les dépôts acidifiants et eutrophisants. L'EMEP est un réseau européen, qui a rassemblé et centralisé des données sur les dépôts atmosphériques à partir des réseaux nationaux. Le modèle de l'EMEP est un modèle eulérien des dépôts atmosphériques (Simpson *et al.*, 2003) intégrant des données d'émissions nationales, les vents stratosphériques et atmosphériques sur l'Europe et les données météorologiques pour déterminer les dépôts des polluants tels que ceux de soufre et d'azote sur une grille de 50 x 50 km. En France, 13 stations opérationnelles sont utilisées par l'EMEP pour mesurer des composés acides dans les dépôts secs et humides (<http://www.nilu.no/projects/cce/emepdata.html>).

Croisé *et al.* (2005) ont également modélisé les données des dépôts hors couvert de soufre, nitrate, et ammonium (entre autres ions), à partir des observations du CATAENAT (sous-réseau du réseau RENECOFOR). Basés sur une approche géostatistique, le krigeage utilisant

la structure spatiale des données de dépôts atmosphériques ainsi que des précipitations comme variable auxiliaire, les dépôts atmosphériques ont pu être estimés sur une grille de 10×10 km couvrant la totalité du territoire français. Les dépôts moyens annuels de 1993 à 1998 (**Fig. 2-4**), et de 1999 à 2004 (Croisé *et al.*, données non publiées) ont été utilisés pour la modélisation. À la différence du modèle EMEP, l'approche présentée par Croisé *et al.* (2005) est simple et n'inclut pas d'étude des mouvements des masses d'air ou d'analyses physiques/chimiques de l'atmosphère ou de polluants atmosphériques.

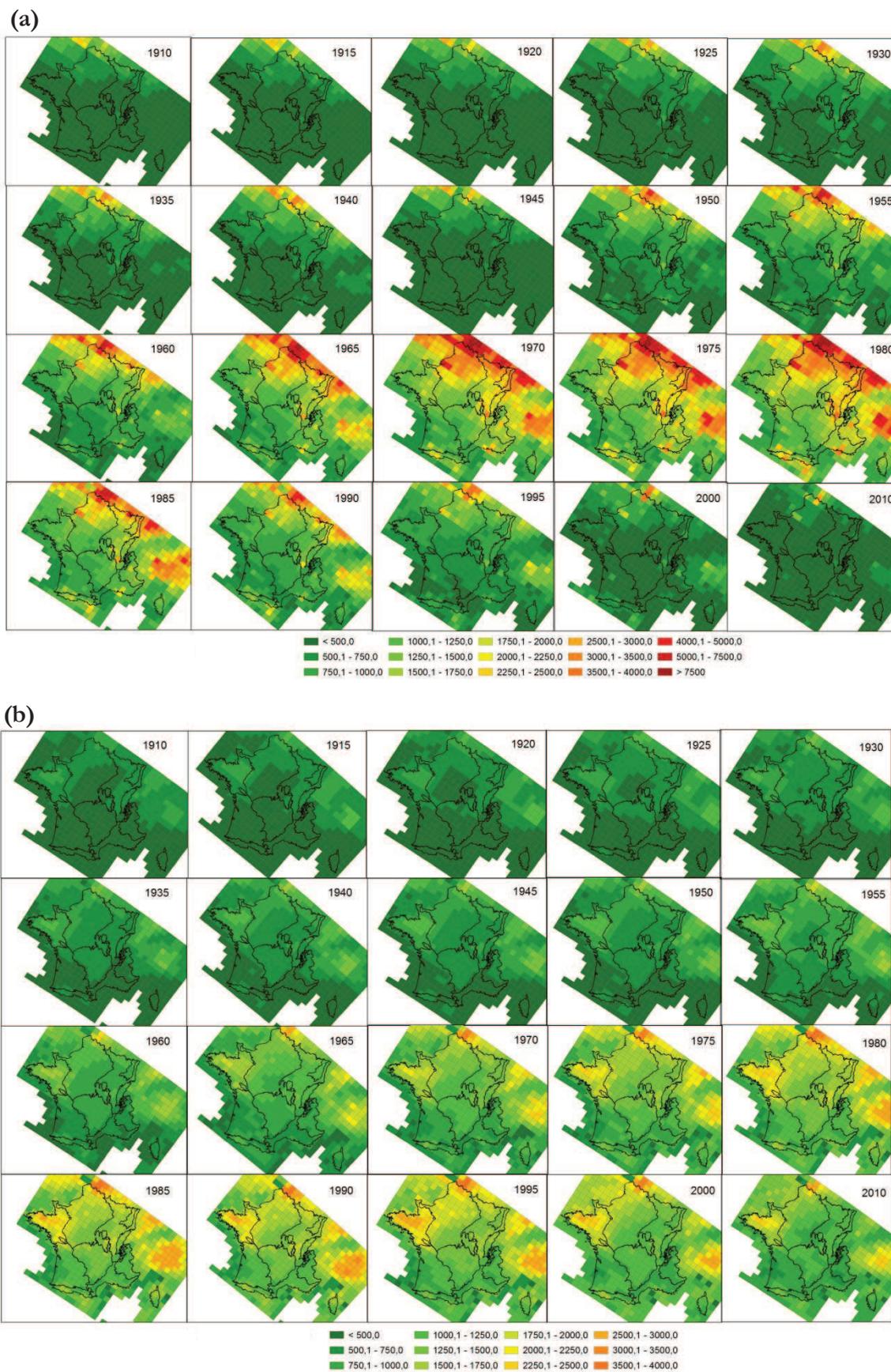


Figure 2-3 Distribution spatiale des données des dépôts atmosphériques de (a) soufre et (b) d'azote total au cours du 20^{ème} siècle modélisées par l'EMEP. Unité : mg m⁻².

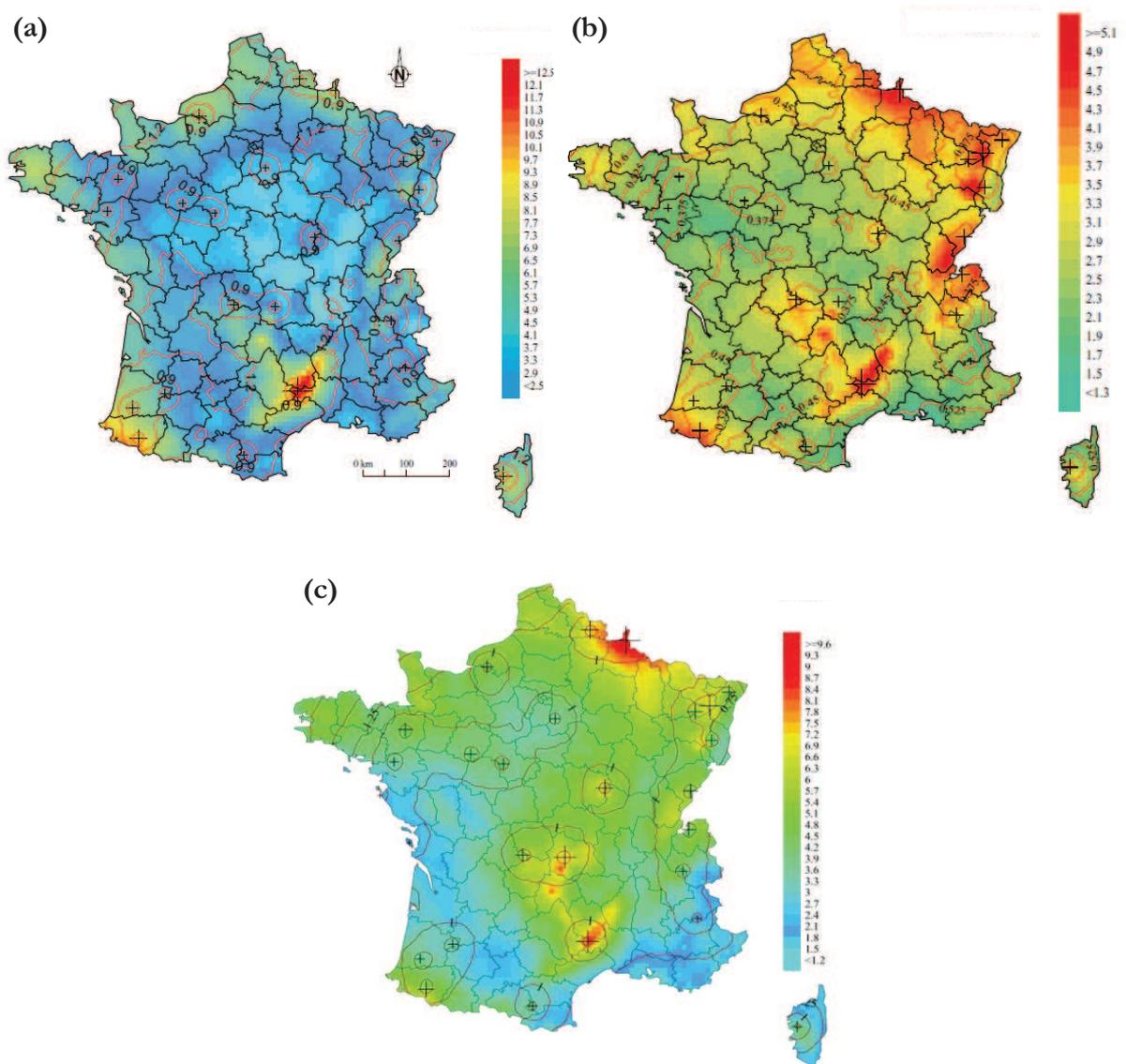


Figure 2-4 Distribution spatiale des données modélisées des dépôts hors couvert **(a)** de soufre de l'ion sulfate ($S-SO_4$) **(b)** d'azote des nitrates (NO_3) et **(c)** d'azote de l'ion ammoniacal ($N-NH_4$), basée sur les dépôts mesurés moyens annuels de la période 1993-1998 et les précipitations en tant que variable auxiliaire. Les croix représentent les placettes du sous-réseau CATAENAT ($n = 27$ placettes) utilisées pour ajuster le modèle et les traits rouges l'écart-type du krigeage. Unité : $kg\ ha^{-1}\ an^{-1}$ (tiré de Croisé *et al.*, 2005).

2.4 Les méthodes de bio-indication : comparaison de trois approches

L'utilisation des réactions spécifiques des organismes à leur environnement est la base de la bio-indication (Diekmann, 2003). En raison de la capacité des plantes à indiquer les conditions environnementales (Bertrand *et al.*, 2011b; Verheyen *et al.*, 2012; Payne *et al.*, 2013), notamment dans un contexte d'acidification et d'eutrophisation à long terme des écosystèmes (Falkengren-Grerup, 1989; Diekmann et Dupré, 1997; Hédl *et al.*, 2011), les conditions

d'acidité et de disponibilité en azote du sol ainsi que leurs changements au cours du temps sous l'influence des activités anthropiques peuvent être estimées en utilisant les données floristiques disponibles pour toute période de temps (cf. Wamelink *et al.*, 2005; Gégout, 2008).

Les méthodes de bio-indication ont évolué tant sur le plan de l'entité considérée (espèce ou communauté), que de leur performance statistique. Dans cette thèse nous avons testé trois approches de bio-indication : la moyenne des optimums, le maximum de vraisemblance et le *Weighted Averaging-Partial Least Squares* (WA-PLS). L'objectif était de les comparer dans le but de sélectionner celle caractérisée par un bon compromis entre qualité de l'ajustement et l'erreur de prédiction, mais aussi celle qui semble la plus pertinente sur le plan écologique.

La **moyenne des optimums** (OPT) est une méthode centrée sur les espèces. Elle repose sur la modélisation de la niche écologique des espèces et l'estimation des optimums d'acidité (pH) et de nutrition azotée (C:N) de la présence des espèces à partir des modèles de niche. L'optimum est la valeur d'une variable donnée (pH ou C:N dans le cas présent) pour laquelle la probabilité de présence de l'espèce est maximale. En calculant la moyenne des optimums des plantes observées sur un relevé on estime les valeurs de pH et C:N bio-indiquées du lieu (ter Braak et Barendregt, 1986) (**Fig. 2-5**).

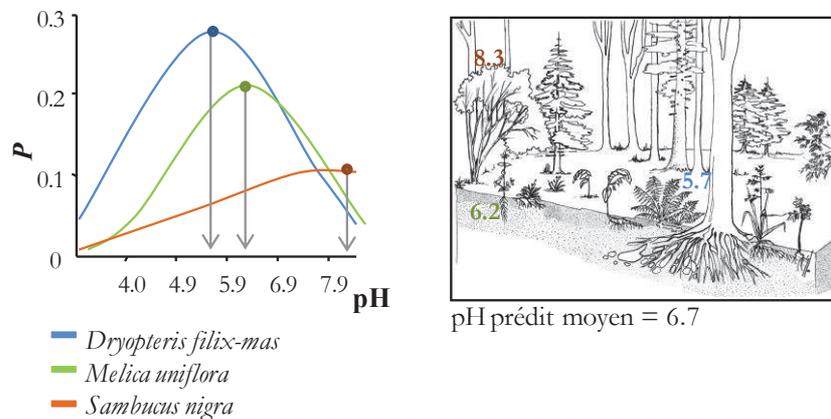


Figure 2-5 Représentation de la prédiction d'une variable écologique à partir des valeurs indicatrices des espèces.

Le **maximum de vraisemblance** (maxL) est une méthode centrée sur les espèces. Elle se base aussi sur la modélisation de la niche écologique, mais à la différence de la méthode précédente, elle examine toute la courbe de réponse pour estimer la valeur bio-indiquée. Celle-ci représente donc la valeur du gradient environnemental maximisant la probabilité de prédire (i) la **présence** des espèces dans les relevés floristiques (maxL P) ou (ii) la **présence et**

P'absence des espèces dans les relevés floristiques (maxL PA). Ce point particulier du gradient d'acidité et de disponibilité en azote dans notre cas est déterminé par maximum de vraisemblance.

La bio-indication à partir du modèle *Weighted Averaging-Partial Least Squares* (WA-PLS) est une méthode centrée sur les communautés. Celle-ci est la combinaison de la méthode *Weighted Averaging* – WA (ter Braak et Barendregt, 1986) et du modèle de régression *Partial Least Square* (PLS) (ter Braak, 1995). La WA-PLS se base sur l'ajustement d'une fonction de transfert reliant la composition spécifique des communautés aux gradients d'acidité et de disponibilité en azote (dans le cas présent). Ce modèle est considéré une approche inversée (*inverse approach*) impliquant des prédictions directes des conditions environnementales à partir de l'ajustement d'une fonction des assemblages d'espèces observés à un endroit donné avec une erreur associée (ter Braak, 1995). La WA-PLS est appropriée pour la calibration lorsque les relations espèces-environnement sont unimodales (i.e. présence d'un seul optimum dans l'espace de la niche écologique), et/ou lorsque les variables explicatives du modèle (i.e. les espèces) sont binaires (i.e. présence/absence) (ter Braak, 1995). La procédure d'ajustement du modèle est la suivante :

- (i) Le jeu de données de calibration est transformé afin de linéariser les relations entre les espèces et l'environnement. Les assemblages d'espèces sont tout d'abord pondérés par le nombre d'espèces présentes dans chaque relevé et par la fréquence des espèces sur l'ensemble des relevés (matrice x^*). La variable environnementale à prédire est aussi pondérée par le nombre d'espèces (y^*).
- (ii) Une régression PLS est conduite à partir du jeu de données de calibration transformé. Elle consiste à rechercher les combinaisons linéaires ou composantes principales ($f()$) des prédicteurs (x^*) qui maximisent la prédiction de la variable environnementale : $y^* = f(x^*) + error$. Les prédictions calculées sur la première composante principale correspondent à la moyenne pondérée des optimums des espèces composant les communautés (i.e. correspond au même résultat obtenu par la méthode WA ou à des résultats proches de la méthode OPT détaillée ci-dessus à la différence que les optimums ici sont déterminés en tenant compte de l'ensemble des espèces). Les composantes principales suivantes sont ajustées de manière à optimiser les prédictions de la variable environnementale (y^*) en utilisant les résidus de la composante précédente pour affiner l'estimation de l'optimum des

espèces (ter Braak, 1995; Brady *et al.*, 2010). Cela signifie que chaque nouvelle composante est orthogonale aux précédentes.

- (iii) Les données ayant été transformées au départ, les prédictions du modèle WA-PLS sont modifiées pour revenir à la variable environnementale originale (pour plus de détails consulter Braak et Juggins, 1993 et ter Braak, 1995). Le nombre de composantes principales maximisant l'ajustement de la WA-PLS est sélectionné à partir d'une évaluation de la performance prédictive du modèle conduite sur un jeu de données indépendant. La dernière composante sélectionnée est celle dont le biais moyen et l'erreur de prédiction sont faibles, et dont la qualité de la relation entre les prédictions et les observations (R^2) est maximale et significative.

La **Table 2-2** synthétise les résultats de l'ajustement issus de la comparaison des trois méthodes. Cette comparaison a été réalisée à partir d'un même échantillon indépendant au jeu de données de calibration. Dans ce travail de thèse nous avons choisi d'utiliser la WA-PLS comme méthode de bio-indication du fait de sa performance statistique globale (valeur de R^2 acceptable et erreur de prédiction la plus faible, **Table 2-2**) et de sa pertinence écologique (i.e. il semble plus logique d'ajuster un modèle directement à partir des communautés d'espèces qu'un modèle indépendant pour chaque espèce).

Table 2-2 Comparaison de la qualité prédictive de différentes méthodes de bio-indication des conditions d'acidité et de disponibilité en azote. La comparaison est réalisée sur un jeu de données de validation indépendant pour le pH ($n = 422$ placettes) et le C:N ($n = 400$ placettes). Les valeurs bio-indiquées ont été confrontées à celles de pH et C:N mesurées pour déterminer la qualité prédictive de chaque méthode (à partir du R^2 et la RMSE ; cette dernière quantifie l'erreur moyenne des prédictions). Pour le calcul des valeurs bio-indiquées, il y a une espèce minimum par relevé.

Méthode	pH		C:N	
	R^2	RMSE	R^2	RMSE
OPT *	0.641	0.803	0.498	4.198
maxL P *	0.586	1.053	0.332	6.203
maxL PA *	0.596	1.095	0.303	7.377
WA-PLS c5 ‡	0.78	0.626	0.283	3.192

* Méthode centrée sur les espèces. ‡ Méthode centrée sur les communautés.

OPT : moyenne des optimums de pH et C:N des espèces ; maxL P : bio-indication par maximum de vraisemblance de la probabilité de présence des espèces ; maxL PA : bio-indication par maximum de vraisemblance de la probabilité de présence et d'absence des espèces ; WA-PLS c5 : bio-indication estimée par un modèle WA-PLS ajusté à partir de 5 composantes principales.

2.5 Cartographie du pH et C:N à partir des valeurs bio-indiquées

Afin d'obtenir la répartition spatiale de l'acidité et de disponibilité en azote des sols forestiers français, nous avons interpolé par krigeage les valeurs bio-indiquées de pH et C:N prédites sur chacun des relevés (ayant au moins 3 espèces indicatrices présentes, $n_{\text{pH}} = 117\ 563$; $n_{\text{C:N}} = 110\ 923$) par le modèle WA-PLS correspondant. La **figure 2-6** schématise la procédure suivie pour cartographier les conditions d'acidité et de disponibilité en azote des sols à partir des données issues de la bio-indication. Les analyses spatiales et géostatistiques (e.g. krigeage) du pH et du C:N ont été réalisées à l'aide du logiciel ArcGIS (version 10.1; ESRI Inc., Redlands, CA, USA).

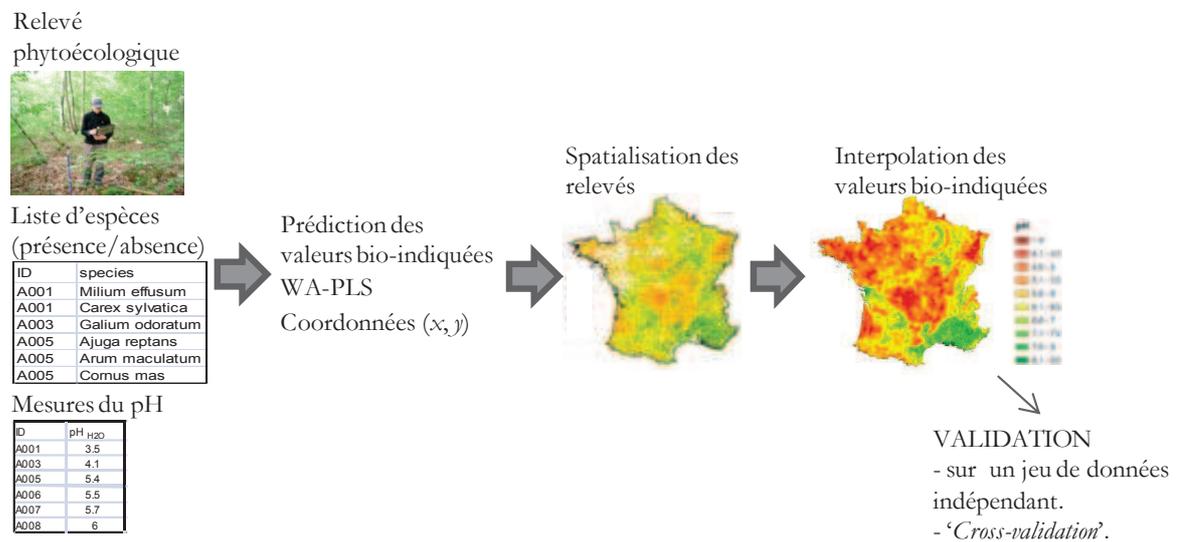


Figure 2-6 Schématisation du processus de cartographie des conditions d'acidité et de disponibilité en azote des sols forestiers français à partir des valeurs bio-indiquées.

L'ensemble des relevés de notre base de données conduit à une densité de placettes, en particulier du fait du nombre important de relevés floristiques issus de l'inventaire forestier national utilisés. Les cartes obtenues illustrent également la variabilité spatiale des valeurs de pH et C:N estimée à partir de la flore sur la période 1910-2010 (année moyenne de la représentation = 1996) (**Fig. 2-7**). On distingue notamment des régions pauvres en nutrition azotée (C:N élevé) et à caractère siliceux (pH faible) tels que le Massif Vosgien, le Massif Central, l'Aquitaine, la Bretagne. On observe également les plateaux calcaires du bassin parisien, le Jura, et la région méditerranéenne riches en azote et caractérisés par des valeurs de pH élevées. L'ensemble de ces observations démontre la pertinence générale des prédictions spatiales du modèle.

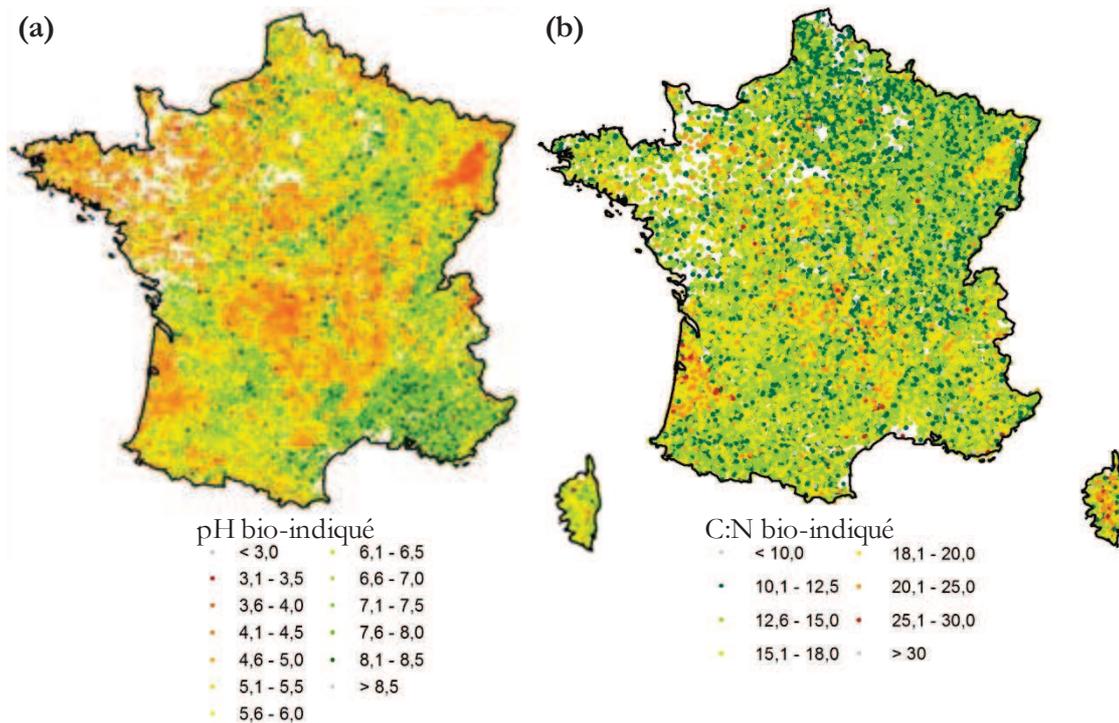


Figure 2-7 Distribution spatiale des valeurs bio-indiquées prédites sur les relevés floristiques (ayant au minimum trois espèces indicatrices par relevé) classées par intervalle de **(a)** 0.5 unités de pH ($n = 117\ 563$) et **(b)** 2.5 unités de C:N ($n = 110\ 923$).

Les paramètres du krigeage des valeurs bio-indiquées ont été optimisés par validation croisée (**Table 2-3**) dans le but, d'améliorer les prédictions et de diminuer l'erreur associée. Les cartes de répartition des valeurs de pH et du C:N issues de cette méthode sont présentées à la **figure 2-8**. Les prédictions réalisées semblent tout à fait cohérentes. On observe de vastes régions à sols acides et mauvaise nutrition azotée comme la Bretagne, les Landes, la Sologne (cohérentes avec les connaissances actuelles) ainsi que des petites zones à forte acidité comme les alluvions de la Moselle ou la forêt de Fontainebleau (également cohérente avec les connaissances actuelles de ces zones) (**Fig. 2-8**).

Table 2-3 Paramètres utilisés pour cartographier le pH et le C:N des sols forestiers français via la méthode du krigeage (*Ordinary kriging*).

		pH	C:N
Données en entrée	<i>n</i>	117 563	116 688
	Condition	<i>n</i> SpVI≥3	<i>n</i> SpVI≥3
Voisinage de la recherche	Type	Standard	Standard
	Nombre de voisins	24	24
	Géométrie du voisinage	Complet	Complet
	Angle	0	0
	Demi-axe majeur	20000	25000
	Demi-axe mineur	20000	25000
	Nombre de ' <i>lags</i> '	50	60
Variogramme	Taille du ' <i>lag</i> '	500	500
	Effet de pépite	0.357	3.905
	Erreur de mesure	0	0
	Type de modèle	Exponentiel	Exponentiel
	Portée	24982.49	29985.41
	Anisotropie	No	No
	Palier	0.500	2.427

n : nombre de placettes utilisées pour la prédiction du pH et C:N.

*n*SpVI : nombre minimal d'espèces requis pour le calcul de la valeur bio-indiquée de chaque relevé.

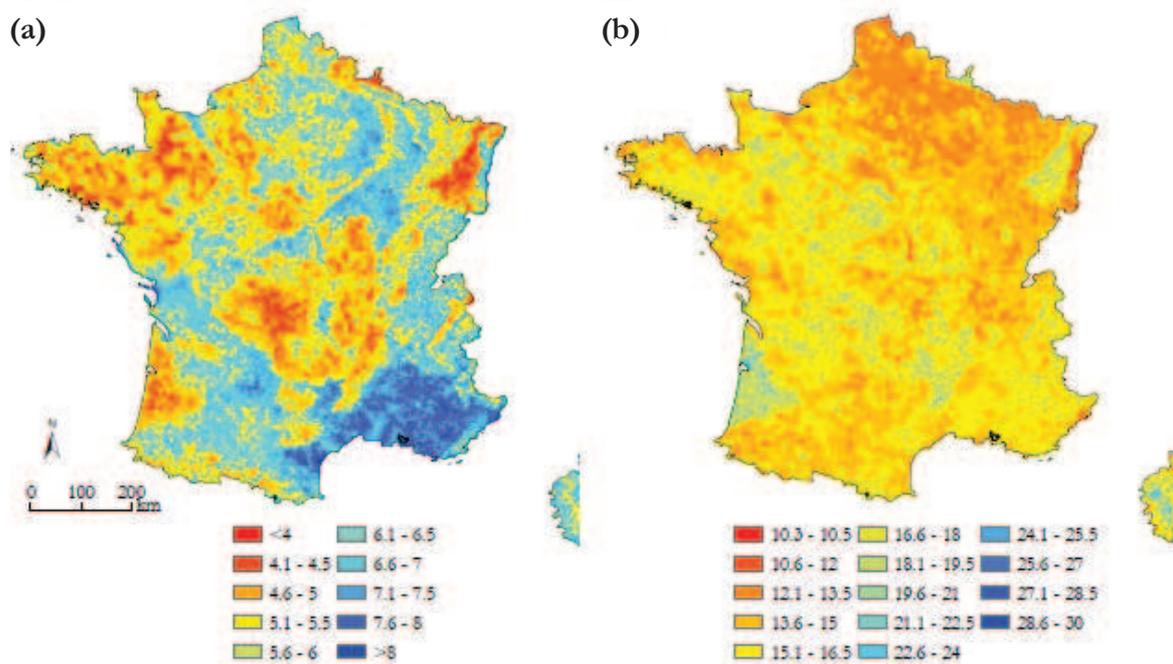


Figure 2-8 Cartes du (a) pH et (b) du C:N de l'horizon A du sol interpolées (par krigeage) à partir de leurs valeurs bio-indiquées respectives.

La précision de la prédiction d'estimation du pH et du C:N fournie par le krigeage a été calculée par validation croisée. Pour le pH, la valeur de RMSE est égale à 0.761 correspondant à un écart entre la valeur krigée et la valeur bio-indiquée supérieur au pH estimé directement sur un site à partir de son cortège floristique. Pour le C:N, la valeur de RMSE est égale à 2.253 correspondant à un écart entre la valeur krigée et la valeur bio-indiquée inférieur au C:N estimé directement sur une placette à partir de son cortège floristique.

**3 ÉVOLUTION SPATIO-TEMPORELLE DE
L'ACIDIFICATION DES SOLS FORESTIERS FRANÇAIS
AU COURS DU 20^{ème} SIÈCLE À PARTIR DU
CARACTÈRE BIO-INDICATEUR DES COMMUNAUTÉS
VÉGÉTALES**

[ARTICLE 1]

Toward a recovery time: Forest herbs insight related to anthropogenic acidification

Gabriela Riofrío-Dillon^{1,2,3}, Romain Bertrand^{1,2} & Jean-Claude Gégout^{1,2}

Global Change Biology 18 (2012), pp. 3383-3394, DOI: 10.1111/gcb.12002

3.1 Abstract

Atmospheric deposition is a global concern contributing to soil acidification and biodiversity changes in forest ecosystems. Although acidifying deposition has decreased in the last decades in Europe, few evidence of ecosystem recovery from acidification has been reported until now. The objective of this study was to reconstruct spatio-temporal changes in soil pH across the entire French forest territory over the last 100-year period through herb species assemblages. Data were collected from floristic databases resulting in a total of 120 216 plots covering French forests and spanning from 1910 to 2010. To define acidity figures, pH values were inferred from herb assemblages for each plot of the prediction dataset based on a WA-PLS model ($R^2 = 0.80$, $SD = 0.59$ for the validation dataset). Spatio-temporal trends of mean pH changes were obtained by comparing plots with respect to the period (mean year of the period = 1933, 1966, 1984, 1997, 2007) and substrate (acidic and non-acidic forest areas).

Bioindicated pH highlighted a decrease of soil pH in both acidic and non-acidic forest areas. The sharpest and most significant pH decrease occurred before 1984 in acidic areas, reaching 0.34 pH units. Subsequently, no significant changes were observed, with a tendency toward stabilization. By contrast, the pH decrease reached 0.19 pH units in non-acidic areas, only reaching significance between 1984 and 1997. Thereafter we observed a slight and significant pH increase. Spatially, pH trends revealed a regionalized character of acidification regarding the substrate, which could not be related to the extent of deposition modeled by the European Monitoring and Evaluation Programme. Both temporal and spatial trends highlight the lagged responses of non-acidic areas compared to acidic areas. Hence, floristic reconstructed pH

¹ AgroParisTech, ENGREF, UMR1092 Laboratoire d'Etude des Ressources Forêt-Bois (LERFoB), 14 rue Girardet, F-54000 Nancy, France.

² INRA, Centre de Nancy, UMR1092 Laboratoire d'Etude des Ressources Forêt-Bois (LERFoB), F-54280 Champenoux, France.

³ French Environment and Energy Management Agency, 20 avenue du Grésillé-BP 90406, 49004 Angers Cedex 01, France.

trends demonstrate a gradual cessation and recovery from acidification of French forests after a period of intense atmospheric pollution.

Keywords: acidic and non-acidic forest areas, bio-indication, France, pH, recovery, spatiotemporal trends of soil acidity, species assemblages.

3.2 Introduction

The acidity of many regions of the world has increased as a consequence of food and energy production to sustain a growing global population (Galloway, 1995). These factors have led to enhanced emissions of air pollutants which in turn lead to increased long-range transport and deposition of sulfur (S) and nitrogen (N). Both are carriers of soil acidity (Dentener *et al.*, 2006a). By approximately the mid-20th century, global S and N emissions from anthropogenic processes overtook emissions from natural processes (Galloway, 2001). The effects of atmospheric pollution are widespread and appear in a number of ways including acidification of freshwater systems (Probst *et al.*, 1999), changes in soil chemistry (De Schrijver *et al.*, 2006), damage to vegetation (Bobbink *et al.*, 2010), and changes in forest ecosystems leading to vegetation changes (Falkengren-Grerup, 1986). Most natural soils are acidic due to both organic matter production and decomposition, but human activities have contributed to making them more acidic (Galloway, 2001).

Atmospheric deposition is a global concern and a current issue. After the peak of S in the 1970s and N in 1980s, and the demonstration of its negative consequences for the environment and human health, measures were implemented to control and reduce air pollution. As a result, S and N deposition within the European area decreased significantly from 1990s to our days, by approximately 60% and 38% , respectively (EMEP, 2011). In France, S and N deposition has followed the European tendency for reduction from their levels in the 1990s (1216 mg S.m⁻² and 1530 mg N.m⁻², computed from modeled data), decreasing by approximately 70% and 21 % , respectively, in 2010 (EMEP, 2011). While S deposition has declined sharply, returning to the deposition levels of the early 20th century, N deposition has remained elevated. Although acidifying deposition has substantially decreased in the last decades, few evidence of ecosystem recovery has been reported until now (see Skjelkvåle *et al.* (2005) for freshwater evidence). A likely recovery according to Driscoll *et al.*

(2001) will be a complex, two-phase process in which chemical recovery precedes biological recovery.

The impact of increased acidifying inputs on soils and the species composition of forest ecosystems has been a major concern in North America and Europe (Aber *et al.*, 1998). Since the soils in which forests grow change slowly and forests themselves grow slowly, the effects of acidic deposition on forests may not be manifested for years to decades (Galloway, 2001). As a consequence, detection of the long-term effects on vegetation and environment is needed to better understand how they interact (Sebesta *et al.*, 2011). Similarly, the detection of such effects over a large scale area is also necessary, because potential drivers and their ecological consequences operate at national and continental scales (Smart *et al.*, 2003). However, such studies are scarce (e.g. Blake *et al.*, 1999; Emmett *et al.*, 2010; Kirk *et al.*, 2010), and the majority of existing works have reported the magnitude of changes over time but covered only small regions or specific plots (De Schrijver *et al.*, 2006; van der Heijden *et al.*, 2011).

Tracking long-term environmental changes is particularly difficult due to the limited historical data with measurements of soil parameters, which in turn is related to the low number of measured sites or monitoring programs (Dengler *et al.*, 2011). Without historical measurements, bioindicator values have become an option to determine the values of environmental parameters and to monitor their change (c.f. Braak et Dame, 1989; Birks *et al.*, 1990). Bio-indication can be defined as making use of specific reactions of organisms to their environment (Diekmann, 2003). Because of the potential ability of plants to indicate the values of environmental variables (Bertrand *et al.*, 2011b), significant insight into soil acidity changes can be identified using the available floristic data for any time period (Wamelink *et al.*, 2005). Forest inventories (mainly started up in the 1980s) and phytosociological studies providing valuable ancient floristic information (e.g. Braun, 1915) represent an important background to long-term research for detecting the impacts of long-range air pollution on vegetation. As most of the plant biodiversity in temperate forest ecosystems is represented by the herb layer, which responds sensitively to disturbances across broad spatial and temporal scales, its dynamics reflect the evolution of forest status (Thimonier *et al.*, 1992; Gilliam, 2007). Previous studies have shown significant shifts in the forest herb layer due to acidification (Diekmann et Dupré, 1997; Baeten *et al.*, 2009; Van Den Berg *et al.*, 2011). In France, while ample evidence from the northeastern region indicates that acidic deposition has deeply altered chemical soil properties, nutrient cycling and vegetation dynamics in forest ecosystems

(e.g. Thimonier *et al.*, 1994; van der Heijden *et al.*, 2011), studies from other French regions are relatively scarce (Landmann, 1995).

Here, we investigated the temporal and spatial changes in soil acidity across the entire French forest territory over the last 100-year period using the bioindicator character of forest herb assemblages. We based our study in a large number of floristic plots ($n = 120\ 216$) spread across the whole forest territory of Metropolitan France and spanning from 1910 to 2010. It allowed us to reconstruct spatio-temporal changes in forest soil pH, likely due to the effects of acidifying deposition, with respect to the substrate. We assumed that species assemblages mirrored acidity conditions. If acidification occurred, species assemblages would have reshuffled over time (Keith *et al.*, 2009), possibly in response to increasing atmospheric loads. It means we assumed that species conserve their edaphic niche, but they change their distribution in accordance with changes in soil conditions (Fehlen et Picard, 1994; Thimonier *et al.*, 1994; Hallbäcken et Zhang, 1998; Dulière *et al.*, 1999). We addressed the following questions: (i) In the context of a marked decrease of acidifying S and N deposition, do changes in soil acidity show a cessation of acidification over recent decades? (ii) Are spatio-temporal changes in soil acidity related to the nature of the geological substrate (acidic or non-acidic)? (iii) Is there a regionalized acidification related to the substrate and does it relate to the European Monitoring and Evaluation Programme (EMEP) extent of deposition changes?

3.3 Materials and methods

3.3.1 Study area

A large-scale study was realized spanning the entire French forest territory. This territory occupies a surface of 161 000 km² (29.2% of Metropolitan France as determined from the national grid of CORINE Land Cover 2006) covering a pH gradient from very acidic (3 pH units) to base-rich (>8 pH units) conditions. French forest substrate consists of bedrocks of varying base status and soils of different types and depths, which cover the whole acidity gradient. To define the substrate in which forests are located, geological information was extracted from a geological French map at scale 1:1 000 000 (Rabu et Chantraine, 2004) and then classified according to lithological composition into acidic (i.e. those with siliceous character: granite, sandstone, basalt, conglomerates) and non-acidic forest areas (i.e. those with

calcareous character: limestone, marlstone, chalk). These areas occupied 51% and 49% of total forest territory, respectively (**Fig. 3-1a**).

The French forest territory has been affected by S and N deposition over time (Landmann, 1995). A natural low deposition was registered before the industrial era, and a strong anthropogenic deposition occurred during the second half of the 20th century, ranging from 445 mg S m⁻² and 446 mg N m⁻² in 1910 to maximal values of 1983 mg S m⁻² and 1515 mg N m⁻² in the 1970s and 1980s, respectively (EMEP, 2011).

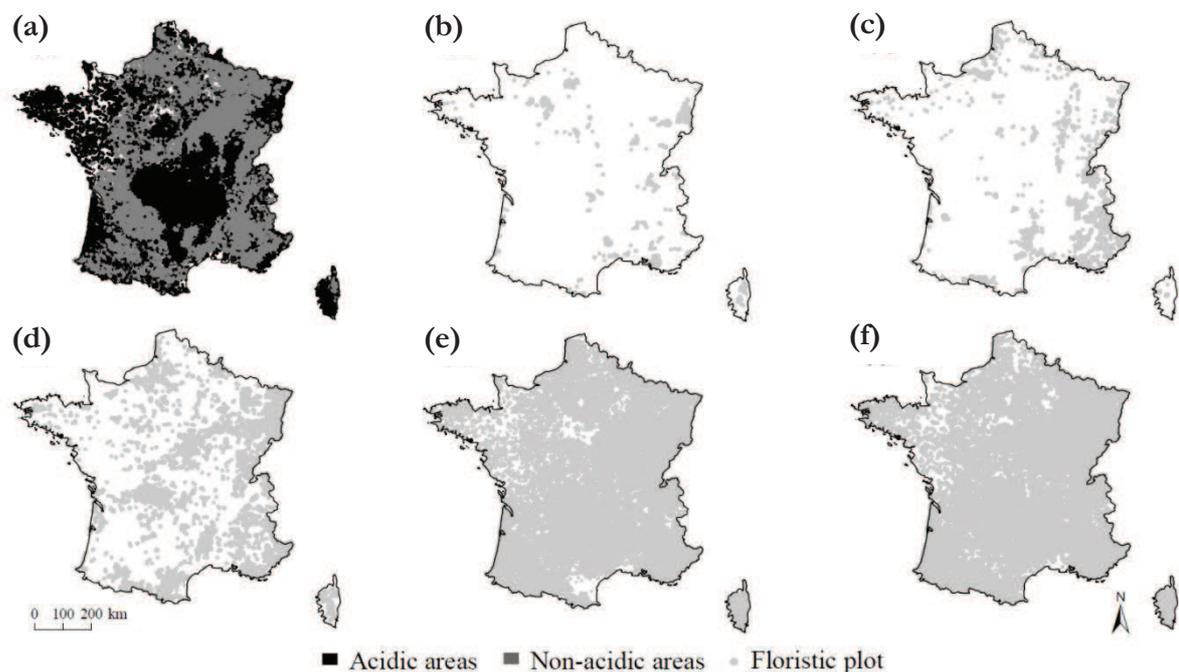


Figure 3-1 Spatial distribution of geological substrates and floristic plots across the French forest territory. **(a)** The spatial distribution of acidic and non-acidic substrates, covering 51% and 49% of forest area, respectively. The spatial distribution of floristic plots over the 20th century: **(b)** first period (1910–1949; mean year of the period= 1933; SD = 9; $n = 1202$), **(c)** second period (1950–1974; mean year of the period = 1966; SD = 6; $n = 5887$), **(d)** third period (1975–1989; mean year of the period = 1984; SD = 4; $n = 17\ 161$), **(e)** fourth period (1990–2004; mean year of the period = 1997; SD = 4; $n = 68\ 925$), and **(f)** fifth period (2005–2010; mean year of the period = 2007; SD = 1; $n = 24\ 388$). Forest cover is based on CORINE Land Cover 2006.

3.3.2 Floristic and environmental data

Three databases of floristic inventories (presence/absence data) were used: EcoPlant (Gégout *et al.*, 2005), Sophy (Brisse *et al.*, 1995), and National Forest Inventory (NFI) (Robert *et al.*, 2010). Together they provided a total of 162 100 floristic plots covering the

entire forested area of France, spanning from 1910 to 2010. They included year and location data with a degree of precision less than 1 km after being spatially georeferenced.

EcoPlant is a phytoecological database including 4913 floristic plots between 1910 and 2010; of which 2854 plots include soil pH measurements. These pH values are pH (H₂O) laboratory values measured from the upper organo-mineral A horizon of sampled soils. Sophy is a phytosociological database that includes 32 330 forest plots from 1915 to 2000. Most sampled plots from these two databases presented an area of 400 m², consistent with current phytosociological practice. The NFI database, managed by the French National Forest Inventory, includes 124 857 floristic plots spanning from 1987 to 2009. The NFI sampling method between 1976 to 2004 consisted of covering each French administrative department at a time following a systematic grid (1 x 1 km) and repeating the sampling every 10-12 years (Robert *et al.*, 2010). In 2004, a new sampling method was adopted, covering the whole French forest territory each year. It consists of a systematic grid of 10 x 10 km, which is moved 2 km each year, thereby ensuring 1 km coverage every 10 years (Robert *et al.*, 2010). For the NFI database, all plots consisted of a surface area of 700 m². Since each database contains taxonomic and nomenclatural issues, a homogenization procedure was carried out to check and, if necessary, update the names of all plant species. To avoid misidentification issues, we mostly focused on the species level. Due to their short lifespan, high population turnover rate, and likely more reactive response to contemporary global changes (Thimonier *et al.*, 1992; Falkengren-Grerup *et al.*, 2000; Bertrand *et al.*, 2011b), only forest herb species were used to reconstruct spatio-temporal changes in soil acidity.

3.3.3 Training, validation and prediction floristic datasets

The selected floristic plots were divided into 2 datasets: the training, comprising 2854 plots with measured pH values and floristic inventories from EcoPlant database, and the prediction dataset, comprising 159 246 available plots (**Table 3-1**). Floristic plots included in the training dataset were sampled between 1974 and 2010. To minimize the over sampling of some geographic regions and environmental conditions, plots had to be at least 500 m distant from each other. To ensure both a good model fit (which increases with the species frequency; ter Braak, 1995), and the use of a large dataset and pool of species to maximize the spatio-temporal representativeness of our study (which decreases with the species frequency), species with more than five occurrences were selected resulting in a total of 482 forest herb species for calibration.

To define the minimum number of species to consider for inferring the pH value of each floristic plot, an iterative sampling approach was applied over an independent dataset, the 16 x 16 Network or European Network level 1. It was created in 1989 for monitoring forest health, and includes 422 surveys with soil pH measurements and floristic inventories (Badeau et Landmann, 1996). At first, a WA-PLS model was calibrated to link the floristic assemblage of each plot with their corresponding pH measurement. Then, for the iterative approach, a random sampling of species in the plots was defined, with the number of species varying from 1 to 20 species, and iterations repeated 50 times per plot. Finally, for each of these virtual samples bioindicated pH values were predicted from the calibrated WA-PLS model, and R^2 (between predicted and measured pH values) and the root mean square error (RMSE) values were computed. The increase of R^2 and the decrease of RMSE diminish strongly when three or more species were considered, then three species was defined as the minimum number of species per plot to calibrate our predictive model. Thus, the criteria of selection to apply on the different datasets were defined as follows: species with ≥ 5 occurrences and ≥ 3 species per plot. When applying them on the training data set 2327 (81.5%) of the initial 2854 plots were selected and used to calibrate the model for inferring pH values from floristic assemblages (**Table 3-1**).

To assess the performance of the model to infer soil pH from the forest herb assemblages in an independent dataset, plots from the 16 x 16 Network were used as the validation dataset. Bioindicated pH values were computed from 326 plots (of the initial 422 plots) that met the defined criteria of selection (**Table 3-1**).

The prediction dataset was used to reconstruct pH trends between 1910 and 2010, considering the defined periods (**Fig. 3-1b, c, d, e, f; Table 3-1**). To avoid overprediction from our model, plots in the prediction dataset that met the defined criteria of selection were selected, resulting in 117 563 (73.8%) of the initial 159 246 floristic plots. Floristic plots from acidic and non-acidic forest areas were well balanced both in the training ($n = 1277$ and 1050 , respectively) and prediction dataset ($n = 54 703$ and $62 860$, respectively). The defined criteria of selection represent a good compromise between quality of the model and the available data for analyses.

Table 3-1 Description of the training, validation, and prediction datasets.

Dataset	Date of collection	Source	Available data	Application	n	$n1$
Training	1974-2010	EcoPlant	<ul style="list-style-type: none"> - Species presence/absence - Soil pH measurements 	<ul style="list-style-type: none"> - To calibrate the WA-PLS model. 	2854	2327
Validation	1989-2004	16 x 16 Network	<ul style="list-style-type: none"> - Species presence/absence - Soil pH measurements 	<ul style="list-style-type: none"> - To define the minimum number of species to consider for inferring the pH value of each floristic plot. - To validate the WA-PLS model. 	422	326
Prediction	1910-2010	EcoPlant Sophy NFI	<ul style="list-style-type: none"> - Species presence/absence 	<ul style="list-style-type: none"> - To compute bioindicated values from herb assemblages and reconstruct the pH trends between 1910 and 2010. 	159 246	117 563

n refers to the total number of plots at the origin.

$n1$ refers to n after filtering plots according the criteria of selection: species with ≥ 5 occurrences and ≥ 3 species per plot.

3.3.4 Weighted Averaging Partial Least Squares (WA-PLS) method to infer soil acidity from floristic assemblages

Among several available techniques using biotic data as a tool for reconstructing past environmental variables, WA-PLS is a powerful inverse approach, i.e. that the adjusted model predicts directly environmental variables as a transfer function from species assemblage with some error (Braak et Juggins, 1993; ter Braak *et al.*, 1993; ter Braak, 1995). WA-PLS is appropriated for calibration when the species-environment relations are unimodal (i.e. one optimum in the ecological niche space), and/or the species data are binary (presence/absence) (ter Braak, 1995). WA-PLS is a combination of both weighted averaging (ter Braak et Barendregt, 1986) and PLS methods (ter Braak, 1995). First, the training dataset is transformed to linearize species-environment relations: (i) Species assemblage dataset is weighted by both the number of species per plot and the species frequencies (hereafter x^*), and (ii) Environmental variable is weighted by the number of species per plot (hereafter y^*). Second, a PLS regression is conducted on the transformed training dataset to fit linear combinations (f) or principal components of the PLS) of the predictors (x^*) so as to maximize the prediction of the environmental variable: $y^* = f(x^*) + error$. PLS regression produces an initial component as a set of coefficients, or weighted averages of the species optima with respect to a given environmental variable. The second and further components are selected by optimizing the prediction of the environmental variable (y^*) as for the first component, and use the residual structure in the data to improve the estimates of the species optima (each new component is orthogonal to the previous one) (ter Braak, 1995; Brady *et al.*, 2010). The number of components that gives the best transfer function requires an examination of performance statistics generated by leave-one-out cross-validation (Birks, 1998; Brady *et al.*, 2010), and a confirmation on an independent dataset (ter Braak, 1995). Third, postprocessing transformation of the results of the PLS regression is required to predict values of the environmental variables (for more details read Braak et Juggins, 1993; ter Braak *et al.*, 1993; ter Braak, 1995). WA-PLS is a training procedure that has already been successfully used in pollen and species assemblages analyses to reconstruct past climatic conditions (Pla et Catalan, 2005; Bertrand *et al.*, 2011b).

Here, the WA-PLS approach was used to infer soil pH from the herb assemblages. A WA-PLS model was calibrated linking the floristic assemblage (among a pool of 482 species) of each of the 2327 plots in the training dataset with their corresponding measured pH value

(**Table 3-1**). A 5-component WA-PLS model was selected on the basis of its low standard deviation of the prediction error (SD = 0.59), low bias (mean of prediction error [bioindicated – measured pH] = 0.01 pH units) and high coefficient of determination between observed and predicted values on the validation dataset ($R^2 = 0.80$; **Fig. 3-2**). Using the prediction dataset, the pH values were inferred from the species assemblages for each floristic plot based on the calibrated 5-component WA-PLS model.

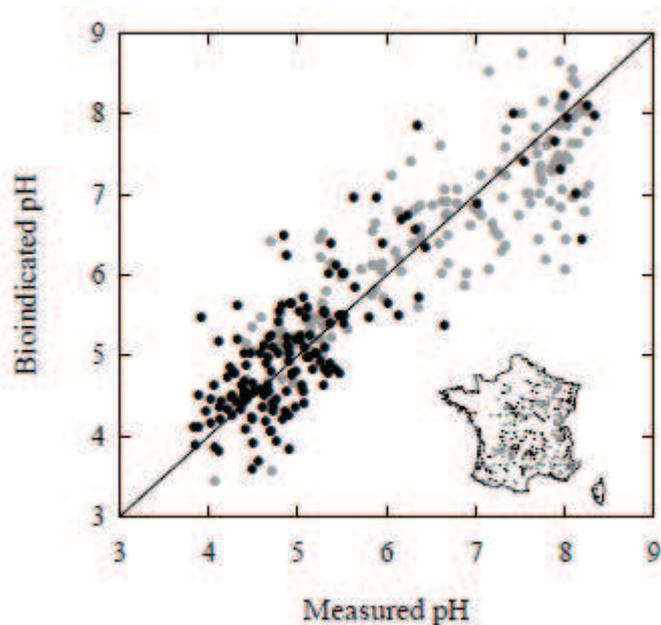


Figure 3-2 Relationship between measured pH values and bioindicated pH values (predicted from the 5-component WA-PLS model) from the validation dataset ($R^2 = 0.80$; mean difference = 0.01 pH units; SD = 0.59; $n = 326$ surveys). Black and gray points represent the spatial distribution of surveys in acidic and non-acidic areas, respectively. Solid black line represents the perfect adjustment between measured pH and bioindicated pH ($y = x$). WA-PLS = weighted averaging partial least squares.

3.3.5 Temporal sampling for analysis of acidity trends

To highlight the soil acidity changes over time, the prediction dataset was divided into five periods considering the significance of air pollution between 1910 and 2010 as well as the data availability (**Table 3-2**). The first period (1910–1949; hereafter 1933) was defined as the earliest period, with a mean acidifying deposition of approximately 570 mg S m⁻² and 562 mg N m⁻² (EMEP, 2011). Data from EcoPlant and Sophy databases were used (**Fig. 3-1b**). The second period (1950–1974; hereafter 1966) has been described as a period of high air pollution in Europe (on average 1291 mg S m⁻² and 1051 mg N m⁻² deposited) (EMEP, 2011) affecting

forest ecosystems. Data from EcoPlant and Sophy databases were used (**Fig. 3-1c**). The third period (1975–1989; hereafter 1984) was defined as a period of control of S deposition, thanks to the implementation of environmental measures (i.e., Convention on Long-range Transboundary Air Pollution, Geneva, 1979); however, atmospheric pollution continues. Data from EcoPlant, Sophy, and NFI databases were used (**Fig. 3-1d**). The fourth period (1990–2004; hereafter 1997) was defined as a period of reduction of N deposition into a lesser extent than that observed for S. Data from EcoPlant, Sophy, and NFI databases were used (**Fig. 3-1e**). The fifth period (2005–2010; hereafter 2007) was defined as a period of continued reduction of S and no change or even a trend toward increase of N deposition. Data from EcoPlant and NFI databases were used (**Fig. 3-1f**).

3.3.6 Sampling for analysis of temporal acidity trends by matching data

To reconstruct long-term changes in soil acidity, and in the absence of permanent surveys, a method that allows the comparison of floristic plots over time was used. First, the floristic plots collected most recently (i.e. 2007 period) were defined as the “reference” data which represents a large number of available floristic plots well distributed across French forest territory (**Fig. 1f**). The 93 175 remaining floristic plots spanning from 1933 to 1997 were defined as “former” data and used to compute soil pH changes regarding the “reference” plots. To control for the potential effect of spatial variability on our assessment of soil pH changes between periods (as described below), 10% of “reference” data were extracted randomly to provide “control” plots (2435 plots). Consequently, 21,953 “reference” plots were used in the matching process (**Table 3-2**).

The method used to compute temporal soil acidity changes consisted of matching each plot from the 1933, 1966, 1984, and 1997 periods (i.e. “former” data) with the nearest plot from the “reference” data, with both plots located on the same substrate. The nearest neighbor was determined by computation of the Euclidean distance (d) between floristic plots. The pH change was computed for each pair ($\Delta pH = pH_{reference} - pH_{former}$) and separate analyses using matched floristic plots were conducted for both acidic and non-acidic forest areas.

Because we aimed to minimize the ΔpH between matched plots due to geographical distance, the “reference” plots were used to explore the spatial autocorrelation between pH values. Then, a threshold of distance was defined to select plots sufficiently closed to each other to allow strong temporal analyses. The spatial autocorrelation between pH values with respect to

the substrate was analyzed using variograms (Fortin et Dale, 2005; Gribov *et al.*, 2006). Considering the variogram outputs, a threshold of distance less than 5 km between “former” and “reference” plots was selected because pH values within this radius were spatially autocorrelated. Considering this comparative distance radius, a total of 77 607 matched plots were obtained, of which 34 524 were situated in acidic areas (median d_{acidic} [1st to 3rd quartile] = 2.6 [2.0–4.5] km) and 43 083 in non-acidic areas (median $d_{non-acidic}$ [1st to 3rd quartile] = 2.4 [2.0–4.0] km) (**Table 3-2**). The statistical significance of the ΔpH between periods [ΔpH_p vs. ΔpH_{p+i}] and the statistical significance of the ΔpH of a period *per se* [ΔpH_p vs. $\Delta pH = 0$] (with p defining a period) were both tested using Wilcoxon Rank Sum test ($P < 0.05$).

To assess the validity of our method, distinguishing temporal and geographical variations, the “control” plots were matched to the nearest “reference” plot (as described above). Considering the threshold of distance less than 5 km between “control” and “reference” plots, a total of 2037 matched plots were obtained, of which 896 were situated in acidic areas and 1141 in non-acidic areas (**Table 3-2**). As each pair consisted of two floristic plots belonging to the 2007 period, no differences of pH within “control”-“reference” pairs were expected in the absence of sampling bias in our matching method. The ΔpH did not significantly differ from 0, as shown here, in either acidic (mean $\Delta pH = -0.001$ pH units [SE = 0.03], $P = 0.877$) or non-acidic areas (mean $\Delta pH = -0.011$ pH units [SE = 0.03], $P = 0.631$). Further, the median distances between “control”-“reference” pairs was 2.8 km in both acidic and non-acidic forest areas, which is comparable with distances calculated between “former” and “reference” matched plots (see above). Hence, the suitability of our method was validated, indicating no spatial bias.