Dynamique et assemblage des communautés adventices : approche par modélisation statistique

Benjamin Borgy

UNIVERSITE DE BOURGOGNE

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

THESE

Pour obtenir le grade de

DOCTEUR EN SCIENCES DE LA VIE DE L'UNIVERSITE DE BOURGOGNE

Dynamique et assemblage des communautés adventices : Approche par modélisation statistique

Présentée et soutenue publiquement par

BENJAMIN BORGY

Le 07 décembre 2011

Directeur de thèse : Xavier Reboud

JURY

B. AMIAUD, Maître de Conférences (ENSAIA-INPL, Nancy)	Rapporteur
E. PORCHER, Maître de Conférences (MNHN, Paris)	Rapporteur
F. BRETAGNOLLE, Professeur (UB, Dijon)	Président
J.N. AUBERTOT, Chargé de recherche (INRA, Toulouse)	Examinateur
M. DELOS, Ingénieur (Ministère de l'Agriculture et de la Pêche)	Invité
R. SABBADIN, Directeur de recherche (INRA, Toulouse)	Examinateur
X. REBOUD, Directeur de recherche (INRA, Dijon)	Directeur de thèse

...

Remerciements

Je tiens à adresser mes plus sincères remerciements à tout les personnes qui m'ont aidé et apporté leur soutien durant ces trois années. Je remercie particulièrement…

Xavier Reboud, de m'avoir encadré durant cette thèse, en m'accordant confiance et liberté dans mon travail de recherche. Ton soutien et ton optimisme durant les moments difficiles m'ont permit de les surmonter.

Sabrina Gaba, pour ton aide précieuse durant ces deux dernières années qui auraient été beaucoup plus difficile sans ta grande disponibilité. Travailler avec toi a été un réel plaisir.

Nathalie Peyrard et Régis Sabbadin (je me permets de vous remercier en binôme) pour votre accueil à Toulouse et votre aide indispensable en mathématique. C'est toujours aussi agréable de travailler avec vous.

Sandrine Petit, pour les multiples relectures et suggestions sur l'article WR.

Guillaume Fried, pour ton aide technique et ton expertise sur la base de données Biovigilance mais également pour les différentes relectures et suggestions que tu as pu apporter à cette thèse et aux différents articles en cours.

Marc Délos, pour ton appuie lors des démarches entreprises pour récupérer les données et qui se sont avérées fort périlleuses.

Je remercie Emmanuelle Porcher et Bernard Amiaud d'avoir bien voulu être rapporteur de cette thèse. Je remercie également Jean-Noël Aubertot, François Bretagnolle, Marc Delos et Régis Sabbadin d'avoir accepté de faire partie du Jury.

Merci aux membres du comité de pilotage : Didier Allard, Nathalie Colbach, Guillaume Fried, Nicolas Munier-Jolain, Bernard Nicolardot, Jean-Noël Aubertot et Thierry Rigaud.

De manière globale, je tiens à remercier l'ensemble du laboratoire « Agroécologie » (ex « Biologie et Gestion des Adventices »), et en particulier les différentes personnes qui ont pu m'aider dans mon travail : Bruno Chauvel, Fabrice Dessaint, Luc Biju-Duval, Claudine Chotel et toutes les personnes qui ont pu participer aux « réunions thésards ». Je remercie également Delphine Mézière et Nathalie Colbach pour leur aide précieuse lors du bouclage de ce manuscrit.

Bien entendu, je remercie mes amis « dijonnais » d'avoir rendu ces trois dernières années plus agréables qu'elles n'auraient été sans eux : Aline, Clément, Mélanie, Delphine, Mathieu, Benjamin, Emilie, Seb, Stéphane, Cécile, Solène, Coraline, Jeanne, Helmut, Thibaud, Rémi, mais aussi les Kiriens : Marion Le & La, Mathieu, Thomas, Cheb, JB et Cam.

Merci à mes parents et à ma famille pour leur soutien inconditionnel. Et surtout merci à Elodie de m'avoir supporté pendant ces deux dernières années.

Résumé

Dans un contexte de recherche de solutions pour une agriculture à la fois productive et durable, les principes, théories et méthodes de l'écologie semblent offrir des pistes à explorer pour comprendre le fonctionnement biologique des agro-écosystèmes. Notre cas d'étude concernait les informations collectées par un réseau d'observatoires de la flore adventice à l'échelon national ('Biovigilance Flore') sur lequel nous avons recherché les règles sous-jacentes à l'assemblage et la dynamique des communautés adventices au sein des parcelles de grandes cultures annuelles. Nous avons en particulier étudié les relations qui peuvent exister au sein des espèces d'une même communauté ainsi que celles affectant les relations des communautés à leur environnement. Pour tenir compte de son importance avérée dans la constitution des assemblages, nous avons conduit notre étude en tenant compte de l'abondance de la population de chaque espèce. Les principaux résultats issus des travaux réalisés au cours de cette thèse sont :

Un assemblage non aléatoire des abondances et une structuration conduisent à une homogénéisation des abondances observées à l'échelle des parcelles. Selon la densité totale d'adventices sur une parcelle, on observe des structurations opposées selon qu'une espèce domine, ou non, dans la communauté. Même si la méthodologie employée pour réaliser les relevés nous empêche d'en avoir la preuve, nous suspectons que l'abondance des adventices répond à la compétition qui s'exerce dans certaines conditions directement entre espèces adventices.

Le choix du modèle nul servant de base de comparaison dans les situations où il n'y a pas de témoin joue un rôle important dans la conduite des analyses des observations issues d'inventaires. L'échelle semi-quantitative pour rendre compte des variations d'abondance pose problème puisque aucun modèle nul n'a été développé pour ce format particulier. Un modèle adapté à notre situation, mais malgré tout générique, a été développé. Il apporte une amélioration sensible par rapport à l'existant pour étudier le degré de cooccurrence (ainsi que la dispersion de traits) dans le cas de données semi-quantitatives. Ceci a des conséquences sur notre capacité à distinguer les facteurs ayant un rôle de filtre sur la flore potentielle qui s'exprime.

La modélisation de la dynamique d'une espèce au sein d'une parcelle au fil des successions culturales nécessite de rendre compte de la dynamique du stock de graines dans le sol. Inconnu, ce stock de graines a été considéré comme une variable cachée, modélisable par l'utilisation d'un modèle de Markov caché. Cette approche s'est trouvée être remarquablement adaptée à la biologie des espèces annuelles. Nous avons ainsi pu appréhender l'effet des pratiques culturales sur la dynamique du stock de graines et de la flore levée. Cette étude a des retombées tant sur l'identification des traits d'histoire de vie que sur leur variabilité face à des stratégies de gestion qui seraient déployées pour assurer le contrôle d'une flore spécifique.

Nous avons ainsi pu appliquer avec succès quelques principes théoriques et améliorer les méthodes d'analyse afin de tenir compte des spécificités des espèces adventices, de l'amplitude des fluctuations du milieu dans lequel elles évoluent et des spécificités de nos données. Il existe une marge de progrès sur la collecte des données à réaliser sur les réseaux d'observatoires si l'on veut maîtriser la connaissance des leviers les plus à même d'aider au contrôle de la flore adventice.

Abstract

To develop solutions for a productive and sustainable agriculture, principles, theories, and methods of ecology may contribute to understand the biological processes governing the agro-ecosystem. The present case study was based on data collected by a network of observatories of weeds covering the whole of France (‘Biovigilance Flore’) and aimed at establishing for rules governing the assemblage and dynamics of weed communities in fields grown with annual crops. We particularly studied the possible relationships between species within a community, as well as the relationships between communities and their environment. Analyses were based on species abundances to take account of their effect on community assemblages. The main results of this PhD thesis are:

Abundances of weed species are not randomly distributed within a field, and this distribution as well as the community pattern result into a homogenization of observed weed abundances at the field scale. The community pattern depends on the total weed abundance in the field and on whether the species is dominant in the community. We concluded that weed abundances responded to weed-weed competition in certain conditions though the sampling methodology was not adapted to prove this assumption.

The choice of the null model for comparisons in situations without a control greatly influences the procedure for analysing assemblages. This problem arises particularly at the semi-quantitative scale since no null model has yet been developed for this particular case. Hence, a generic null model was developed for our particular situation. It considerably improved the study of co-occurrences and trait dispersal in case of semi-quantitative datasets and thus our ability to identify factors and processes determining flora composition.

Modelling the weed dynamics of a field over the crop succession needs to take into account the dynamics of the weed seed bank in the soil. As it was unknown, the seedbank was considered as a hidden variable and modelled using a Hidden Markov Model (HMM). This approach was well adapted for predicting the life cycle of annual species. We thus identified effects of cultural practices on the dynamics of the seedbank and of the emerged flora. This study allowed us to identify life-history traits and weed management strategies adapted to a specific weed flora.

We successfully applied various ecological theories and improved methods to take into account the specificities of weed species, the variations in their environment as well as the specificities of our data-set. The quality and reliability of the data collected in the observatory network must be increased if we want to correctly identify management levers likely to contribute to sustainable weed management.

Sommaire

Grille de lecture de la thèse

Nous avons fait le choix d'écrire une thèse basée sur une série d'articles soumis ou en phase finale de préparation. La partie essentielle de l'exercice scientifique académique de la thèse est contenue dans 4 articles dont la cohérence s'organise le long d'un fil directeur joignant les observations de terrain au développement des méthodologies appropriées pour leur analyse. Ce travail de thèse repose sur l'adaptation et l'application de différents concepts classiquement utilisés en écologie aux agro-écosystèmes.

1. INTRODUCTION GENERALE

Ces dernières décennies, les enjeux sociétaux et environnementaux tel que le Grenelle de l'Environnement ou le plan Ecophyto 2018 ont conduit les agriculteurs à repenser leur gestion des bio-agresseurs des cultures. Parmi les alternatives à la lutte chimique, l'intensification écologique des systèmes de cultures propose de valoriser les régulations biologiques pour maintenir une production agricole. L'intensification écologique fait appel à des concepts d'agro-écologie (voir entre autres Doré *et al.*, 2011). Cette discipline qui émergea dans les années 30 (Wezel & Soldat, 2009) est basée sur la compréhension du fonctionnement des écosystèmes cultivés afin de concevoir et d'adapter des systèmes de culture complexes, productifs et attractifs tout en limitant, si possible, le recours aux intrants (Altieri, 1995).

Cette thèse s'inscrit dans un contexte d'agroécologie où l'intérêt social et politique croissant pour la conception d'une agriculture plus durable a récemment modifié les questionnements scientifiques en ouvrant de nouveaux axes de recherche. Ceci est particulièrement vrai pour le cas adventice pour lequel la recherche c'était largement concentrée sur le conflit entre la présence/abondance en adventices et la productivité des cultures (Petit *et al.*, 2011). Néanmoins, des recherches récentes montrent que les adventices interagissent avec beaucoup d'organismes et peuvent avoir un effet sur le fonctionnement de l'agro-écosystème (par exemple Bohan *et al.*, 2011). Ainsi, il est important de comprendre les effets des interactions biotiques et abiotiques sur la constitution des communautés adventices aussi bien dans une optique de gestion que de conservation de la flore adventice.

1.1. Assemblage des communautés

L'écologie des communautés vise à étudier la cooccurrence d'un ensemble d'espèces à un instant et en un lieu donnés ; un des buts premiers est d'établir des règles générales et synthétiques pouvant expliquer les patrons d'espèces observées (MacArthur, 1972). La capacité à dégager des règles générales d'assemblage des communautés est sujet à débat, en raison de la nature complexe et parfois spécifique des processus intervenant dans les assemblages. Ainsi, les règles établies seraient trop souvent dépendantes et limitées à des domaines d'étude locaux et restreints (Simberloff, 2004 ; Roughgarden, 2009). Connor & Simberloff (1979) ont fait remarquer que les patrons de cooccurrence observés pouvaient aussi bien être générés par le seul hasard. Cela a initié une controverse sur l'importance relative du hasard et des contraintes dans la formation des communautés (Weiher & Keddy, 1995). Néanmoins, l'écologie des communautés reste une science essentielle dans un but de conservation et de gestion des espèces, en particulier dans un contexte de changement climatique et de dégradation des écosystèmes.

1.1.1. Théorie neutre de la biodiversité

La théorie neutre de la biodiversité proposée par Hubbell (2001) suggère que la stochasticité à elle seule suffit pour expliquer certains assemblages d'espèces observés, qui étaient jusqu'ici considérés comme résultants exclusivement de processus de sélection. Cette théorie s'inscrit fortement dans la continuité de la théorie de la distribution insulaire proposée par MacArthur & Wilson (1963, 1967) pour laquelle la biodiversité dépend uniquement de la taille de l'île et de sa distance au continent (considéré comme le réservoir d'espèces). Ainsi l'installation des espèces est globalement déterminée par le rapport de force entre les processus aléatoires de colonisation et d'extinction, qui sont fonction de la distance et de la taille de l'île. Cette approche, où toutes les espèces sont considérées comme équivalentes et suivent les mêmes processus de dispersion et de survie, peut être étendue à la composition des métacommunautés des différents compartiments du paysage (Loreau, *et al.*, 2003 ; Muneepeerakul *et al.*, 2008). Sans être exclusive, cette vision aléatoire de l'assemblage s'oppose malgré tout à la vision déterministe de la niche écologique où les assemblages d'espèces dépendent des caractéristiques (i) du milieu (notion de filtres) et (ii) des espèces composant le pool d'espèces.

1.1.2. *Différents types de filtre*

Les contraintes, qui opèrent comme des filtres, peuvent être classées en trois groupes (Belyea & Lancaster, 1999) (Figure 1) :

(i) les contraintes environnementales qui déterminent, par les conditions du milieu, les espèces capables de se développer sur le site (HSP) ;

(ii) les contraintes de dispersion, déterminées par la géographie, qui ressemblent les espèces capables de coloniser le site (GSP) ;

(iii) les contraintes de coexistence, déterminées par des processus d'interactions biotiques entre les espèces du pool écologique (HSP ∩ GSP = ASP) qui conduisent à la communauté observée.

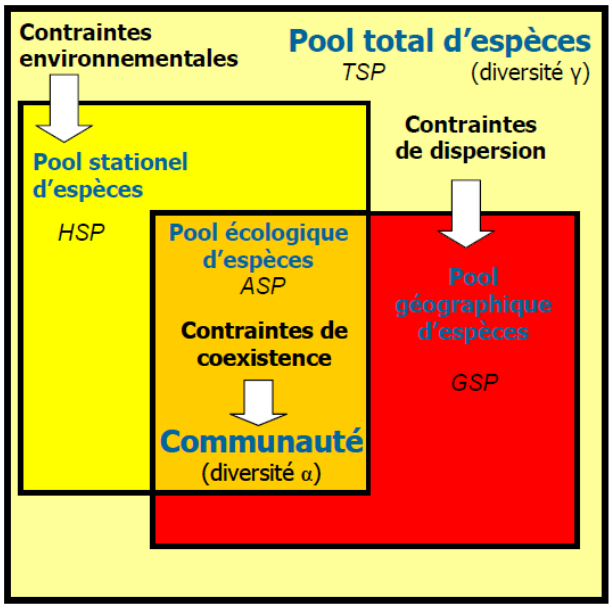


Figure 1 Les différents pools d'espèces et les contraintes correspondantes. D'après Fried (2007), à partir de Belyea & Lancaste (1999). La définition de chaque pool est donnée dans le texte.

Ainsi, la nature aléatoire de l'assemblage des communautés peut venir bruiter l'importance relative de deux des trois groupes de contraintes : les contraintes de dispersion et les contraintes de coexistence. Les contraintes environnementales ne sont pas affectées par la nature aléatoire de l'assemblage, puisqu'elles déterminent de façon purement déterministe les espèces composant le pool d'espèces présent sur un site ou une station (HSP). En revanche, les contraintes de dispersion et de coexistence sont beaucoup moins déterministes : la dispersion (souvent assimilée à de la diffusion) peut être composée d'une part de hasard relativement importante, et la compétition peut exclure une espèce de façon aléatoire entre deux espèces compétitivement équivalentes (comme pour la divergence génétique par exemple).

La structuration des communautés est ainsi définie par l'ensemble de ces contraintes mais aussi d'une part d'aléatoire, dont le poids peut être plus ou moins important suivant l'échelle

et le type de communautés étudiées. Ainsi on gardera comme définition de communauté, l'ensemble des espèces d'un même niveau trophique qui pour des raisons spatiales ou historiques se retrouvent en interaction.

1.1.3. L'approche traits fonctionnels

L'écologie fonctionnelle représente l'une des sous branches les plus explicatives de l'écologie des communautés et intéresse de plus en plus la communauté scientifique (Petchey & Gaston, 2006). Elle tend à expliquer à la fois le rôle, ou fonctions, que joue une espèce au sein d'une communauté ou d'un écosystème et à la fois le rôle, ou fonctions, associé à chaque caractère morphologique ou physiologique (ou trait fonctionnel) des différents organes composant une espèce. La structuration des communautés est décrite par les caractéristiques des espèces, aussi appelées traits fonctionnels. L'approche par traits fonctionnels permet d'expliquer la présence d'une espèce en un site (de par les conditions du milieu) et renseigne également sur les ressources pour lesquelles les espèces peuvent rentrer en compétition (de par la similarité de leur niche). On s'intéresse alors aux relations entre traits fonctionnels et conditions du milieu. La notion de fonction n'étant pas forcément mesurable directement, elle est souvent approchée par la mesure de marqueurs indicateurs du potentiel de la fonction en question (Figure 2, la capacité d'absorption [fonction] d'*Euphorbia helioscopa* est directement liée à la densité de racine [marqueur fonctionnel]).

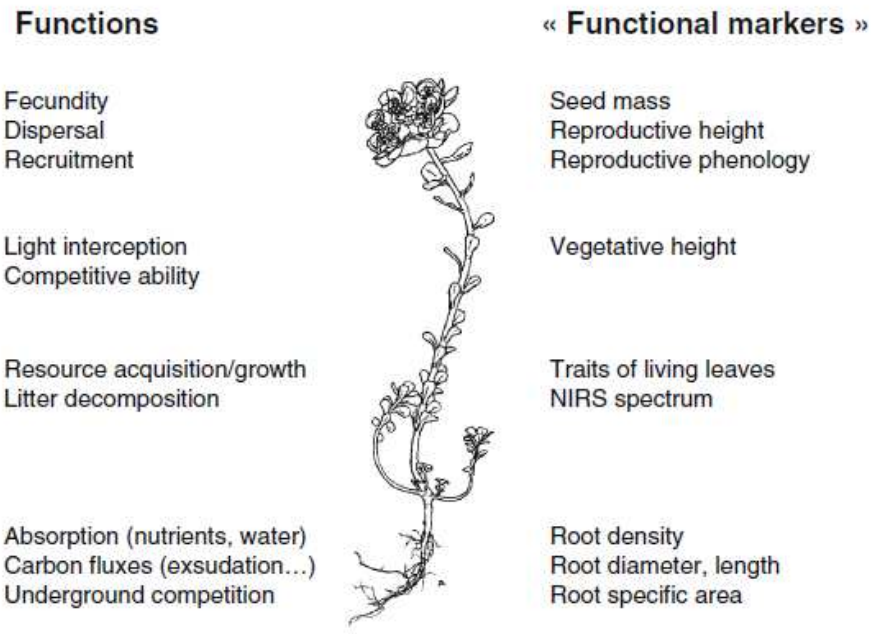


Figure 2 Exemple de relation entre fonction et marqueurs fonctionnels (traits) (Garnier & Navas, 2011). La figure montre la relation entre fonction (par exemple ceux liés à la reproduction [fécondité, dispersion , etc.]) et les marqueurs mesurables directement corrélés à cette même fonction [masse des graines, hauteur des inflorescences, etc.).

Selon cette approche fonctionnelle, le degré variable de complexité qui pouvait régner au sein de l'assemblage des espèces se trouve éclairci lorsqu'on tente d'expliquer les cooccurrences par les raisons biologiques expliquant la présence des espèces et les éventuels processus d'interaction biotique (McGill *et al.*, 2006). De plus la mise en évidence de la fonctionnalité couverte permet de qualifier la biodiversité représentée par une communauté au delà de la diversité taxonomique des espèces en présence (Mouchet *et al.*, 2010).

1.1.4. Variabilité des traits

Pour être utile à l'écologie des communautés, un trait fonctionnel est toute caractéristique mesurable, de préférence sur une échelle quantitative continue, et dont la variation inter espèces est plus importante que la variation intra espèces (McGill *et al.*, 2006 ; Violle *et al.*, 2007). En effet, les méthodes classiques d'analyse des traits fonctionnels et de leur assemblage au niveau des communautés se fait généralement par l'intermédiaire de jeux de données de contingence d'espèces (tableau espèces – site) et de jeux de données de traits moyens des espèces (tableau espèces – traits) (Dolédec *et al.*, 1996 ; Legendre *et al.*, 1997). Il parait alors évident que par l'utilisation de traits moyens d'espèces, les analyses seront peu significatives si l'on choisit des traits variant plus au sein des espèces que entre elles et où l'assemblage des traits expliquera difficilement l'assemblage des espèces (Jung *et al.*, 2010 ; Albert *et al.*, 2010). La notion de traits fonctionnels regroupe deux types de traits non exclusifs : les traits de réponse et les traits d'effet. Les traits de réponse expliquent la réponse d'une espèce à l'environnement, tandis que les traits d'effet expliquent l'effet de l'espèce sur l'environnement (Lavorel & Garnier, 2002). Dans ce travail de thèse, je me suis concentré sur les traits de réponse et leur distribution, puisque seule la réponse des espèces aux gradients environnementaux nous intéresse *a priori* lorsqu'on vise à définir des règles d'assemblage. Toutefois plusieurs études ont mis en évidence une superposition ('overlap') entre les traits de réponse et les traits d'effet suggèrent un lien entre les changements des espèces en réponse l'environnement à leur fonctionnement dans l'écosystème (voir par exemple, Garnier *et al.*, 2004, Blanco *et al.*, 2007, De Deyn *et al.*, 2008, Pakeman 2011a). Les traits de réponse peuvent cependant varier d'un environnement à un autre en fonction des conditions environnementales, et en particulier en fonction du mode de gestion des habitats. Le choix des traits fonctionnels pour l'étude de l'assemblage des communautés des agro-écosystèmes doit par conséquent être réfléchi et adapté aux systèmes étudiés.

1.1.5. <u>Dispersion des traits et processus mis en jeu</u>

Les différents processus mis en jeu dans l'assemblage des communautés n'ont pas le même impact sur l'assemblage des traits et sur la distribution de la valeur d'un trait au sein d'une communauté. En effet, les processus à « effet filtre », c'est-à-dire les contraintes environnementales et les contraintes de dispersion, vont avoir tendance à filtrer les espèces par leur caractéristiques.

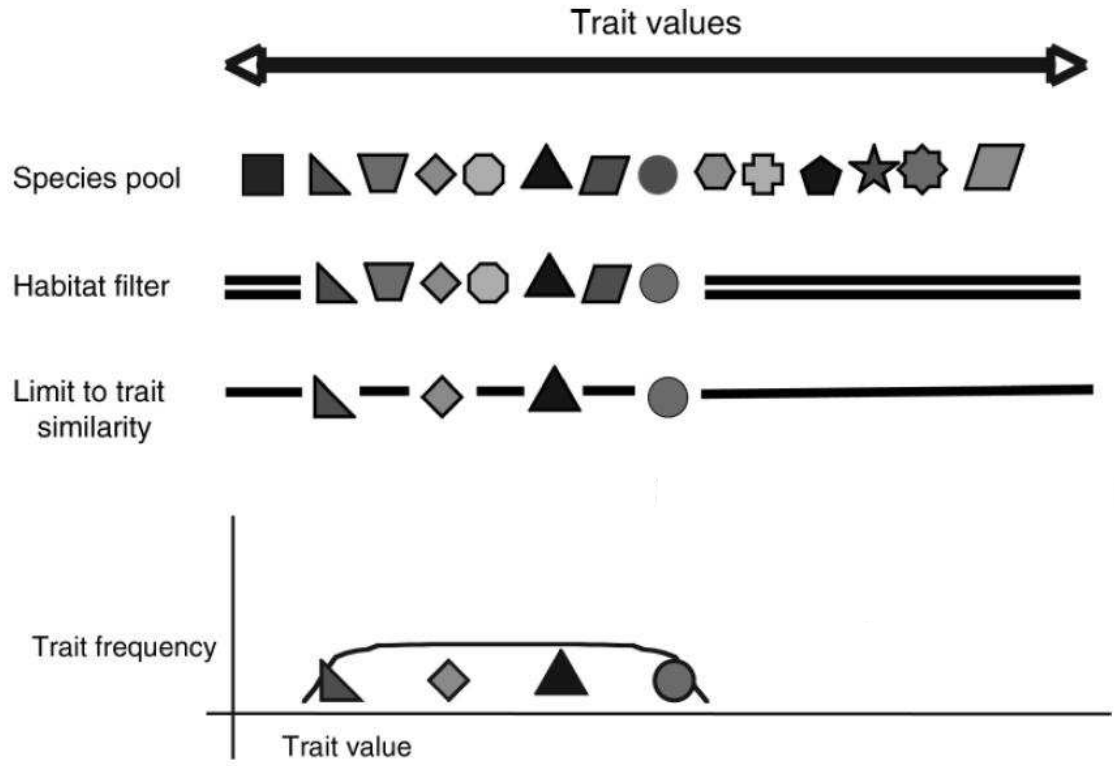


Figure 3 Effet filtre et distribution des traits. Le poids de l'effet filtre de l'habitat (Habitat filter) and la limitation de similarité des traits (Limit to trait similarity) dépendent de la combinaison des traits et des conditions abiotiques du milieu. L'effet filtre de l'habitat affecte la gamme de traits sélectionnée alors que la limitation de la similarité des traits affecte la distribution (Cornwell & Ackerly, 2009).

Les espèces occupant un habitat donné ont ainsi une probabilité plus importante d'avoir des valeurs de traits proches pour répondre aux même conditions environnementales (« niche filtering » ou « habitat filter », convergence des traits) (Figure 3). A l'opposé, les interactions biotiques au sein de la communauté vont avoir un effet différent. La compétition aura tendance à limiter la similarité des niches (« limiting similarity ») des espèces présentes en un site et ainsi à étaler l'ensemble des espèces présentes sur l'étendue des niches disponibles d'un site, ce qui limitera de fait leur zone de recouvrement donnant lieu a compétition (divergence des traits) (Cornwell & Ackerly, 2009 ; Garnier & Navas, 2011). Ainsi, la compétition tend à répartir les individus sur les gradients environnementaux, voir à créer des distributions d'exploitation des gradients très divergentes (comme observé par Grime [2006] sur des communautés d'herbacées). Définir dans quel sens sont structurées les communautés observées (convergence vs. divergence des traits) impose donc la comparaison à une

référence. Or en écologie, on accepte souvent que chaque observation soit, dans les faits, unique, pour des raisons historiques ou de spécificité locale. Le corolaire est qu'on ne dispose donc pas de référence 'toutes choses égales par ailleurs'. C'est pour résoudre cette impasse que les écologistes se sont penchés sur une solution alternative qui consiste à ce que le jeu de données devienne son propre référentiel si on le manipule avec discernement pour lui extraire une distribution attendue avec relâchement de l'hypothèse qui nous intéresse : les modèles nuls. Si les observations tombent au milieu de la distribution ainsi créée c'est qu'on ne peut rejeter l'hypothèse nulle ayant permis de construire la distribution ; à l'inverse une valeur rattachée aux observations s'écartant significativement de cette distribution trahira l'existence de règles s'éloignant de l'hypothèse nulle retenue.

Les modèles nuls permettent donc la construction de communautés de référence où toute structuration dans l'assemblage des traits a été « effacée » (l'approche modèles nuls est développée dans le chapitre 2.1).

1.2. Dynamique des communautés et approche traits d'histoire de vie

Les études sur l'assemblage des communautés s'intéressent en général à la composition d'une communauté à un moment donné. La structure et la composition des communautés varient cependant dans le temps en réponse aux variations de leur environnement biotique et abiotique (Myers & Worm, 2003 ; Knapp *et al.*, 2001 ; Garrabou *et al.*, 2002 ; Benjankar *et al.*, 2011). L'étude de la dynamique de cet assemblage impose alors de rajouter une dimension supplémentaire : le temps. Ainsi l'étude de la dynamique de l'assemblage d'une communauté se retrouve être rapidement complexe car on étudie l'évolution de l'assemblage sous un environnement changeant impliquant alors une non-stabilité des communautés (bien que ce ne soit jamais vrai, supposer la stabilité permet généralement de simplifier les processus). Néanmoins, dans le cas de communautés où les interactions biotiques sont faibles (par exemple avec peu de compétition), la dynamique globale de la communauté peut être approchée par la réunion des dynamiques individuelles des différentes espèces composant la communauté, situation grandement simplifiée (Oberdorff *et al.*, 1998 ; Hugueny *et al.*, 2000). Evidemment, le concept de communauté perd un peu de son intérêt.

1.2.1. Décomposition du cycle

Les traits d'histoire de vie permettent de décrire la dynamique d'une population en modélisant la dynamique respective des différents compartiments de l'espèce. Lotka (1925) et Volterra (1926) ont indépendamment proposé les modèles « proie-prédateur » par l'utilisation de couple d'équations différentielles pour décrire la dynamique de systèmes biologiques dans lesquels une espèce proie et une espèce prédateur interagissent. Les équations sont composées de termes représentant des taux de mortalité et de reproduction. Les modèles de Leslie sont basés sur le même principe (Caswell, 2001), mais les populations sont alors discrétisées en classes d'âges permettant une meilleure décomposition de la dynamique interne de chaque espèce et une modélisation des différents stades de vie. En effet, pour beaucoup d'espèces, les l'aptitude des individus à se maintenir dans la population (survie) et à la faire croître (reproduction) évolue au cours de leur vie. Ainsi une population donnée peut être représentée sous la forme de tables de vie. Chaque compartiment représente des stades de vie suivant par exemple des structures d'âges (juvénile, adulte). Les probabilités de transition d'un compartiment à l'autre sont des taux de survie et de reproduction attachés à chaque classe d'âge (Figure 4). Des modèles basés sur ce formalisme s'écrivent alors de manière simple, sous la forme de matrice de transition, et les différents paramètres du modèle synthétisent efficacement les différentes dynamiques internes à la population de l'espèce étudiée. L'analyse de ces matrices permet d'extraire des informations synthétiques telles que la pyramide des âges ou le taux d'accroissement intrinsèque (Caswell, 2001), voire de calculer la sensibilité de la démographie à un allongement de la durée de vie ou à l'avancement de l'âge de la première reproduction, par exemple.

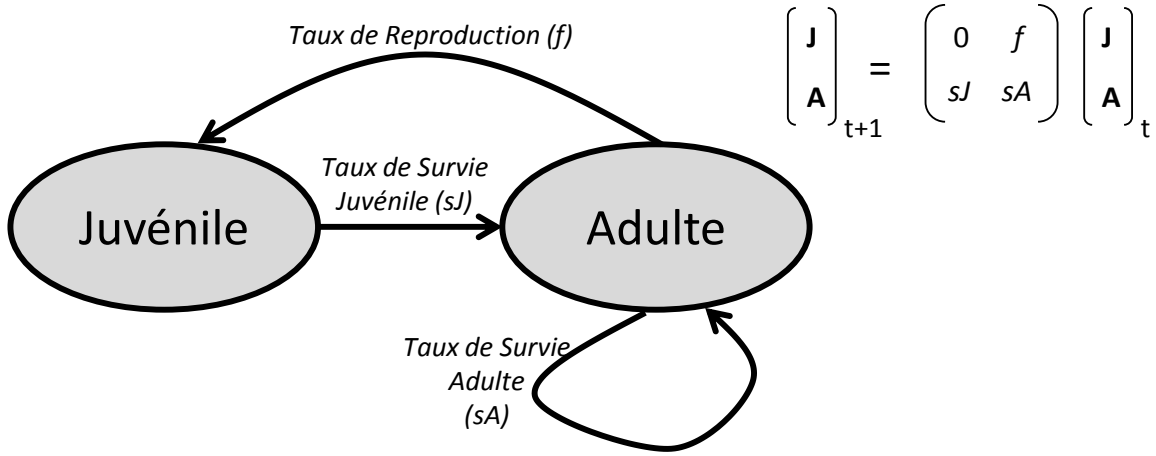


Figure 4 Cycle de vie d'une espèce. Représentation de la dynamique d'une espèce à deux classes d'âges et de trois traits d'histoire de vie contrôlant les probabilités de transition des individus et le recrutement de nouveaux individus juvéniles. La dynamique peut être ainsi réduite à une matrice de Leslie permettant de modéliser la dynamique de la population (nombre d'individus Juvéniles [J] et Adultes [A]) dans un espace de temps discret.

1.2.2. Synthèses des processus affectant la dynamique démographique

Les traits d'histoires de vie vont résumer plusieurs processus intervenant dans la dynamique d'un individu (ou d'une population mais l'on parle alors de traits démographiques) (Violle *et al.*, 2007). En effet, derrière un taux de survie vont se cacher tous les processus influençant la survie d'un individu : longévité, mortalité intrinsèque, mais aussi mortalité due à la prédation, etc. Les différences observées entre les taux de survie des différents compartiments du cycle peuvent s'expliquer par des différences biologiques et des sensibilités variables aux divers processus (Cam & Monnat, 2000). Les traits d'histoire de vie se situent donc sur un plan différent de celui des traits fonctionnels en retraçant plus spécifiquement ce qui va affecter la dynamique démographique là où les trais fonctionnels visent à préciser les conditions environnementales nécessaires à la réussite de la population ou les fonctions réalisées sur l'environnement. Ce format simple et générique permet d'avoir toutes sortes d'approches, stochastiques au grain des individus (pour les modèles individus centrés) ou déterministes (pour les modèles épidémiologiques, par exemple). Ceci peut s'avérer très utile pour la construction de modèles prédictifs adaptés aux questions posées et l'objectif visé. Les traits d'histoire de vie peuvent être estimés en conditions naturelles et permettre ainsi de positionner les espèces sur un gradient de stratégies de reproduction. Le gradient bien connu r-K en est un exemple (MacArthur, 1962 ; MacArthur & Wilson 1967). Ce gradient permet de différentier les espèces à fort taux de reproduction précoce mais avec de faibles taux de survie (r), des espèces à croissance lente mais avec une survie et une longévité importantes (K). D'après la théorie d'histoire de vie r-K de MacArthur et Wilson, les niveaux élevés de perturbations sélectionnent les espèces de courte durée de vie, à taux de croissance important et produisant beaucoup de propagules ; à l'inverse les habitats soumis à des niveaux de perturbations plus faibles sélectionnent les espèces les plus compétitives, à durée de vie

longue, produisant peu de descendants de grosse taille et dispersant peu. L'étude des traits d'histoire de vie des espèces permet ainsi d'évaluer leur valeur reproductive dans un environnement donné et leur capacité à faire face aux perturbations environnementales (Stearns, 1992). L'approche basée sur les traits d'histoire de vie permet donc de typer la stratégie évolutive des différentes espèces. Fortement dépendante des conditions du milieu, nous n'attendons pas que des traits d'histoire de vie évalués dans deux milieux différents avec des conditions environnementales différentes soient identiques (voir l'article de Nylin & Gotthard [1998] sur la plasticité des traits d'histoire de vie). La variabilité des taux d'accroissement calculés à partir des traits d'histoire de vie dans chacun de ces environnements va aussi nous renseigner sur le degré de spécialisation des espèces pour les conditions du milieux. Russo *et al.* (2005) suggèrent que les performances des espèces spécialistes, et donc leurs traits d'histoire de vie, varient significativement suivant l'habitat. Cette variabilité de succès peut ainsi servir de mesure du degré de spécialisation des espèces. De manière plus fine, on peut appréhender l'effet des conditions du milieu sur les différents compartiments de la population (classes d'âges ou stades de croissance par exemple) et sur les taux de transition entre les compartiments (exemple : Willott & Hassall, 1998). Dans un environnement perturbé où le milieu peut recouvrir différentes conditions de manière répétée, cyclique et alternative (exemple, la rotation culturale), par l'utilisation de traits d'histoire de vie relatifs à chaque état du milieu, on peut ainsi prédire la dynamique de chacun des compartiments de l'espèce au cours du temps, suivant la succession d'états du milieu. Par construction d'un modèle de dynamique des différents compartiments suivant les valeurs de différents traits d'histoire de vie, et en présence de relevés de population d'une espèce (et des différents compartiments), on peut alors inférer les traits d'histoire de vie de l'espèce compatibles avec les relevés observés suivant le choix du modèle fait. Ainsi certains traits seront plus facilement estimés via leur contribution indirecte à la dynamique démographique que par des mesures *in situ*. Par exemple, il est très difficile d'approcher par la mesure le nombre de descendants viables laissé par un individu alors que le taux d'accroissement de la population pourra en donner une estimation réaliste.

1.3. Le modèle adventice

Godhino (1984) propose de définir les plantes adventices comme« toute plante qui se développe spontanément dans les milieux modifiés par l'homme ». Selon cette définition, la flore adventice des cultures correspond à l'ensemble de la végétation spontanée présente dans un champ hormis l'espèce cultivée de l'année en cours. Ainsi, la flore adventice réunira les « mauvaises herbes » (flore sauvage spontanée) et les repousses de cultures précédentes. En France, on trouve environ 1200 espèces adventices dans les cultures (Jauzein, 2001) mais une faible fraction seulement peut être considérée comme commune. Malgré leur grande diversité, les espèces composant la flore adventice des champs cultivés partagent une faculté commune : l'adaptation aux pratiques culturales (Fried, 2007). Plusieurs caractéristiques de l'agro-écosystème cultivé en font un modèle d'étude qui pourrait être plus propice que des écosystèmes naturels pour tester des théories d'Ecologie et mesurer le poids de l'homme dans le processus d'évolution des communautés. C'est la conclusion qu'expriment Booth et Swanton (2002). Néanmoins les fortes différences entre l'agro-écosystème et le milieu naturel touchant notamment à la richesse, au degré de saturation ou encore à l'intensité des perturbations, impliquent une certaine réflexion sur l'adaptabilité des méthodes classiquement utilisées ainsi que leurs limites.

1.3.1. Spécificités des communautés adventices

Les champs cultivés constituent un milieu très artificialisé, en grande partie contrôlé par l'Homme. Ainsi, l'espèce dominante de la communauté, i.e. l'espèce cultivée, est prédéterminée. Le type et la fréquence des perturbations (i.e. les opérations de travail du sol, l'application d'herbicides) ainsi que le niveau d'éléments nutritifs apportés sont aussi entièrement contrôlés par l'Homme pour créer la situation optimale à l'expression du potentiel productif de la seule espèce cultivée. L'agro-écosystème est un système très dynamique où les contraintes (i.e. les techniques culturales) évoluent rapidement. Au cours des 50 dernières années, l'agriculture s'est considérablement intensifiée (mécanisation, recours généralisé aux fertilisants et pesticides de synthèse) entraînant des changements profonds de composition de la flore adventice (Fried, 2007). L'agriculture doit aujourd'hui répondre à de nouvelles attentes imposées par exemple, par le plan *Ecophyto 2018* qui vise à réduire de moitié la gamme et la quantité de produits phytosanitaires utilisés, ou encore à la préservation de la biodiversité des champs (conservation des espèces messicoles par exemple). Les pratiques culturales sont donc vouées à être modifiées. Cette évolution rapide des pratiques culturales continuera donc vraisemblablement à induire une forte évolution des communautés adventices aux différentes échelles (de la parcelle, de la culture, régionale et nationale). L'intensité des changements peut être approchée à travers une valeur de *turnover* défini comme la proportion d'espèces différentes entre 2 périodes. Dans les milieux agricoles ces valeurs de *turnover* sont extrêmement importantes et malheureusement souvent associées à une perte de richesse spécifique (Fried, 2007 ; Fried *et al.*, 2009a ; Fried *et al.*, 2009b).

La modification des communautés adventices en réponse à la modification des pratiques culturales, et une structuration dans la répartition des espèces à l'échelle nationale impliquent une certaine spécialisation des communautés adventices aux pratiques et à l'environnement pédoclimatique. Cette spécialisation, variable en fonction des espèces, s'exprime par une capacité à exploiter une large (espèces généralistes) ou étroite (espèces spécialistes) gamme de situations environnementales et culturales (Fried *et al*., 2010). Ainsi, les espèces, en plus de présenter différents degrés de spécialisation, sont donc aussi plus ou moins spécialistes (ou sensibles à) des conditions pédoclimatiques et/ou culturales (Fried *et al*., 2008).

Du fait des perturbations extrêmes, pratiquées à chaque campagne culturale et qui détruisent tout ou une partie de la flore levée (labour, travail du sol, traitement herbicide,…), et de la rotation culturale, une forte proportion d'espèces adventices sont des espèces thérophytes (ou annuelle), c'est-à-dire des espèces dont le cycle biologique se boucle en une saison et dont la survie pluriannuelle des populations est entièrement assurée par la production de graines (Sutcliffe & Kay, 2000 ; Sutherland, 2004). Ainsi, l'état du stock de graines dans le sol et sa dynamique jouent un rôle déterminant dans la composition des communautés de flore adventice levée. La spécialisation des espèces adventices est cependant possible grâce à la survie des graines dans le stock du sol qui permet de différer la levée des adventices jusqu'à la réunion des conditions qui leur correspondent (par exemple les conditions réunies à la date de semis d'une culture en particulier) (voir par exemple Venable & Lawlor, 1980). Cependant, moins les conditions favorables apparaissent fréquemment et plus l'espèce doit être à même de survivre dans le stock semencier : la fréquence de retour des cultures ou des opérations culturales impactent sur le mode de fonctionnement du stock semencier du sol.

Les semences des différentes espèces du stock peuvent présenter des caractéristiques (ou traits) variées : taille, poids, longévité, intensité ou hétérogénéité des dormances, etc... Chaque espèce peut ainsi être décrite comme une combinaison de traits. Par une analyse de sensibilité, Gardarin *et al*. (2010, 2011) montrent que les différentes combinaisons de traits rendent compte des différentes aptitudes des espèces à assurer leur survie en fonction de l'histoire culturale (succession culturales et itinéraires techniques) d'une parcelle. Il en résulte souvent une grande disparité de la flore quand on passe d'une parcelle à ses voisines. Le stock de semences du sol peut être très important : suivant les espèces, entre 1% et 20 % de la flore potentielle seulement s'exprime chaque année dans une parcelle cultivée (Barralis & Chadoeuf, 1988 ; Debaeke, 1988). Cela rend complexe l'étude de la végétation puisque le stock de semences peut agir comme un tampon dans la réponse des communautés aux conditions du milieu. La flore observée (composition, abondance) est donc la résultante de l'ensemble de l'historique cultural et pédoclimatique de la parcelle. Ainsi il est quasiment impossible de trouver deux situations identiques (ceci a d'ailleurs comme fâcheuse conséquence la difficulté de définir un témoin pour les études scientifiques). La dispersion des semences par les outils agricoles (Shirtliffe & Entz, 2005) ou par le vent peut conduire à la présence éphémère d'adventices dans des sites qui d'un point de vue des conditions écologiques ne leurs conviennent pas obligatoirement. Cela peut donc conduire à un « bruit de fond » supplémentaire pouvant brouiller l'analyse des relations entre flore et milieu à un instant donné (Fried, 2007).

La prise en compte du stock de graines est donc primordiale pour évaluer correctement l'impact des pratiques et de l'environnement dans l'assemblage et la dynamique des communautés adventices. Malheureusement, l'échantillonnage du stock semencier est long, fastidieux et coûteux et donc très largement hors de portée pour un grand nombre d'espèces et suivant une large gamme de conditions différentes (Barralis, 1976). Il est par conséquent rarement étudié.

1.3.2. Le réseau Biovigilance Flore

Le réseau Biovigilance Flore a été mis en place par le Ministère de l'Agriculture à partir de 2002. L'objectif de ce réseau d'observations est de détecter d'éventuels effets non-intentionnels associés aux nouvelles techniques culturales (dont l'introduction des cultures génétiquement modifiées) sur la flore et la faune des agro-écosystèmes et des milieux directement voisins (Delos *et al.*, 2006). Dans le projet initial, près de 1000 parcelles de grandes cultures réparties sur l'ensemble de la France métropolitaine étaient suivies annuellement. Sur chacune des parcelles, des relevés exhaustifs de la végétation ont été effectués et l'ensemble des pratiques culturales ainsi que quelques données majeures sur le milieu physique (pH et texture du sol) et l'environnement de la parcelle ont été recueillies (descriptif des tables de la base de données Biovigilance Flore dans les tableaux X1 et X2 de l'annexe page 152 et 153). Dans les faits, le suivi a été irrégulier, tant dans sa dimension territoriale (Figure 5) que temporelle, mais aussi entre les différentes variables mesurées.

Les relevés floristiques sont effectués sur une zone de 2000m² par un ou deux experts parcourant cette zone pendant 20 minutes et évaluant la densité de chaque espèces présentes sur une échelle semi-quantitative[1] de densité d'individus (Barralis, 1976). Une zone témoin de 100 à 200m², où aucun traitement herbicide n'a été appliqué, est également échantillonnée. Cette zone a été cultivée selon les mêmes pratiques que le reste de la parcelle cultivée échantillonnée à l'exception du traitement herbicide. Cette absence de traitement herbicide prévue par le protocole visait à permettre une expression plus importante de la flore potentiellement présente dans la zone témoin. Le technicien procédait au désherbage différé du témoin suite à sa dernière prise de mesure.

[1]L'échelle semi-quantitative est composée de 7 classes de densité : '0' : absence de l'espèce, '+' : trouvée une fois sur les 2000m²; '1' : moins de 1 individu/m²; '2' : 1-2 individus/m²; '3' : 3-20 individus/m²; '4' : 21-50 individus/m²; '5' : plus de 50 individus/m².

Figure 5 Distribution de l'échantillonnage 'Biovigilance Flore'. L'hétérogénéité de l'échantillonnage visible sur cette carte peut résulter à la fois (i) d'une réalité sous-jacente d'absence de zone de culture et donc d'adventices et (ii) de la dynamique locale de collecte et d'intérêt pour l'étude de biovigilance.

Fried a pu constater que l'ampleur de la couverture spatiale (gradients Nord-Sud et Est-Ouest d'environ 1000 km, figure 5) mettait bien en évidence des patrons d'organisation généraux au sein des communautés, dépassant les particularismes régionaux (Fried, 2007 ; Fried *et al.*, 2009a ; Fried *et al.*, 2009b ; Fried *et al.*, 2010). A l'inverse, le réseau n'a jamais visé à quantifier les processus spatiaux opérant à des échelles spatiales fines, que ce soit la dispersion des espèces entre parcelles voisines à l'échelle d'un paysage ou la distribution des espèces depuis l'intérieur du champ jusqu'aux bordures des parcelles. Pour un sous-échantillon de la base biovigilance, un suivi pluriannuel exercé sur les parcelles permet d'accéder à la trajectoire des communautés (Fried, 2007).

1.3.3. Importance de la prise en compte de l'abondance

Dans les milieux naturels ou semi-naturels, la biomasse végétale est supportée par un faible nombre d'espèces dominantes dont les traits ont ainsi un rôle important sur le fonctionnement de l'écosystème (Grime, 1998). De ce constat, l'hypothèse de « biomass ratio » de Grime (1998) suggère que l'analyse des espèces dominantes de la communauté suffirait à capturer les réponses de la communauté aux changements de l'environnement. La diversification des rotations culturales et les opérations culturales de gestion des adventices maintiennent les espèces adventices à des abondances faibles. Toutefois, comme dans les milieux (semi-) naturels, toutes les espèces adventices ne sont pas présentes à la même abondance. Ainsi, si l'on traduit les abondances des espèces observées dans un relevé en occurrence (présence/absence), une espèce en faible effectif (traduisant une non-adaptabilité aux conditions du milieu) aura le même poids qu'une espèce présente en forte abondance. Ceci pourra avoir des conséquences sur la compréhension de la réponse des espèces aux pratiques mais également sur leur effet sur l'agro-écosystème (voir Cornwell & Ackerly, 2010 ; pour un exemple en milieu non perturbé) et particulièrement la compétition avec la culture (voir

Kropff & Spitters, 1991 ; pour un exemple sur la compétition adventices-culture). Que ce soit à l'échelle de la cooccurrence des espèces, ou de la distribution des traits fonctionnels, les communautés sont ainsi décrites de manière beaucoup plus fine si on pondère la simple présence par l'abondance.

1.4. Objectifs de la thèse

Dans un contexte de conception d'une agriculture plus durable, il est primordial de mieux comprendre l'assemblage et la dynamique des communautés adventices. En effet, cela permettrait de comprendre comment mobiliser les régulations biologiques de l'agro-écosystème, de manière à mieux optimiser les stratégies de gestion des adventices pour réduire le recours aux herbicides. Les jeux de données issus du réseau 'Biovigilance Flore' donnent une photographie de la flore observée dans l'espace et dans le temps dans des parcelles de grandes cultures en France et permettent ainsi de disposer de données recouvrant ensemble large des pratiques et de territoire.

L'objectif de cette thèse est ainsi d'extraire les informations pertinentes d'un réseau d'observatoires de parcelles de grandes cultures pour identifier les mécanismes à l'origine de l'assemblage et de la dynamique des adventices. Pour cela, ce travail de thèse mobilise les principes, théories et méthodes classiquement utilisés en écologie.

Le cas adventice (et Biovigilance) pouvant être particulier, l'un des points central de cette thèse est donc d'évaluer l'adaptabilité des méthodes utilisées en écologie, et de réfléchir et mettre en place les modifications nécessaires afin d'assurer l'adéquation des méthodes à notre cas d'étude.

La composition des communautés sera abordée sous deux grandes approches :

- l'assemblage des espèces en communautés, où l'on s'intéressera aux relations qui peuvent exister entre les espèces d'une même communauté,
- la relation communautés – environnement, où l'assemblage et les dynamiques dans le temps seront uniquement appréhendés entre les espèces, prises indépendamment, et leur environnement.

Tous les travaux présentés dans ce manuscrit sont réalisés à l'échelle de l'abondance de la population de chaque espèce, afin de tenir compte de son importance dans la compréhension des assemblages.

Le Chapitre 2 traite de l'assemblage des communautés. Dans un premier temps, nous étudierons la structuration des classes d'abondances, c'est-à-dire (i) les cooccurrences entre classes d'abondances et (ii) à l'effet de l'abondance totale sur cette structuration. Dans un second temps, nous présenterons le modèle nul que nous avons développé pour l'analyse de structuration des communautés (cooccurrences entre espèce et distribution des valeurs de traits), lorsque les données d'abondance sont semi-quantitatives (i.e. classes d'abondance).

Les relations espèces – environnement sont traitées dans le Chapitre 3. Nous avons tout d'abord modélisé la niche écologique pour étudier sa capacité à expliquer l'abondance des espèces en un site donné. Pour finir, nous présenterons en quoi l'utilisation de séries temporelles de flore levée peut permettre d'identifier les traits d'histoire de vie rendant compte de la dynamique de stock de graines.

2. ASSEMBLAGE DES ESPECES ADVENTICES

Pendant cette thèse, je me suis intéressé aux règles d'assemblage des communautés adventices en étudiant dans un premier temps les patrons de cooccurrences des abondances des espèces adventices et dans un second temps la distribution des diversités fonctionnelles entre les communautés adventices. Cette deuxième partie a conduit à une réflexion sur les modèles nuls existants et sur leur pertinence pour des analyses basées sur des données dont le format de notation des abondances (échelle semi-quantitative, voir qualitative) est similaire à celui rencontré dans 'Biovigilance Flore'. Aucunes des approches existantes n'étant adaptées, j'ai alors proposé un modèle nul appelé *SwapClass* adapté des méthodes *Swap* (Gotelli, 2000). Ces travaux ont conduit à l'écriture de deux articles. Ces deux analyses se placent à deux niveaux totalement différents mais sont toutes deux basées sur la comparaison des patrons observés à des patrons théoriques suivant une hypothèse nulle. Dans ce chapitre, je présente tout d'abord l'approche par modèles nuls puis les travaux réalisés. Le premier article analyse et décrit les patrons de cooccurrences des abondances des espèces adventices.

2.1. L'approche par modèles nuls

Afin d'attester (ou non) de l'existence d'une structuration dans les inventaires biologiques et/ou les jeux de données correspondants, l'écologie des communautés a vu se développer tout un ensemble de méthodes visant à tester et observer en quoi les assemblages peuvent différer d'assemblages purement aléatoires constitués en l'absence d'un mécanisme particulier (Weiher & Keddy, 1999).

2.1.1. Philosophie de l'hypothèse nulle

Comme pour tout test statistique, l'approche par modèle nul compare les observations à celles que l'on pourrait attendre sous l'hypothèse nulle, c'est-à-dire sous l'aléa ou sous une distribution théorique particulière attendue. Ainsi, un modèle nul est défini comme un générateur de patrons, basé sur la randomisation des données observées. Cette randomisation produit une distribution des patrons attendus en l'absence d'un mécanisme écologique particulier qu'il faut donc réussir à relâcher de manière appropriée (Gotelli & Grave, 1996). Ceci revient à construire sa statistique interne. L'aléa associé au patron observé est dépendant du choix de l'hypothèse nulle et donc du modèle nul utilisé. Se pose alors le problème de choix approprié, biaisé ou non du modèle nul. En effet, derrière une hypothèse nulle peuvent se cacher plusieurs modèles nuls, différant par leur degré de restriction (ou contraintes) dans les randomisations qu'ils effectuent.

2.1.2. Contraintes des modèles nuls

Afin de générer des jeux de données randomisées qui soient « comparables » au jeu de données observées, les randomisations effectuées par le modèle nul doivent respecter un certains nombres de contraintes, sans quoi, on pourrait rejeter ou accepter à tort l'hypothèse nulle. Par exemple, lorsque l'on étudie la structuration des espèces au sein d'un jeu de données, il convient de respecter la structuration du jeu, structuration qui dépend du sujet d'étude. Il est commun en écologie de trouver des jeux de données composés de peu d'espèces fréquentes et de beaucoup d'espèces rares (Preston, 1962). Il est impératif de respecter cette structuration globale de la communauté (distribution des espèces sur un gradient rare vs. fréquente) au sein des jeux randomisés, sans quoi une différence dans les indices résumant la structuration pourrait apparaître et ceci uniquement en raison du non-respect des fréquences d'occurrences propres à chaque espèce (Gotelli, 2000). Ce serait par exemple le cas si l'on cherchait à dériver un indice de diversité basé sur le principe de Shannon $p_i*\log(p_i)$, où l'homogénéité des observations fait mathématiquement accroître l'indice. Ainsi, dans le cadre de l'étude de la structuration des espèces avec des données de présence/absence, le choix du modèle nul s'oriente généralement vers des modèles nuls assez restrictifs fixant la somme des occurrences par lignes (=nombre d'espèces par site) et par colonnes (=nombre de sites occupés par l'espèce). Ce sont les méthodes « swap » (Gotelli & Enstminger, 2001). Néanmoins, un paradoxe apparaît : on cherche à conserver une certaine

structuration qui peut être la résultante des processus de structuration que l'on souhaite justement relâcher lors des randomisations. La rareté d'une espèce est-elle uniquement due à sa façon d'exploiter le milieu ou est-elle également due aux interactions biotiques ? Ainsi les contraintes que l'on souhaite fixer sont-elles réellement indépendantes du processus à étudier ? En l'absence de présupposé, un modèle très restrictif, tant qu'il ne biaise pas les résultats, reste toutefois généralement la meilleure option (Gotelli, 2000; Ulrich & Gotelli, 2010).

2.1.3. Limite de l'approche

Comme toute méthode statistique d'analyse, l'approche par modèle nul est contrainte par certaines limites, qu'il convient de garder en mémoire, et peut être sujette à toutes sortes d'abus si certains aspects de contrôle sont négligés. Il est ainsi nécessaire de se poser un certain nombre de questions quant à la significativité et à la pertinence des randomisations que l'on effectue : toutes les espèces sont-elles équivalentes ? Est-il judicieux de mélanger des espèces provenant de milieux environnementaux très différents ou très éloignés spatialement ? Quel est le poids de l'Homme (et non de l'écologie) dans l'assemblage observé ? La méthode d'échantillonnage des espèces et des relevés est-elle standardisée ? Ainsi, malgré toutes les précautions que l'on peut apporter à la manière dont sont réalisées les randomisations et l'algorithme utilisé, il convient de prêter attention à d'autres composantes du jeu de données. Bien que le jeu de données soit utilisé comme sa propre référence, une utilisation trop naïve peut conduire à des résultats qui ont peu de sens. De plus le concept de *ceteris paribus* (« toutes choses étant égales par ailleurs ») n'est ainsi jamais vraiment validé car les conditions biogéographiques et l'influence humaine ne sont jamais égales entre deux sites (Gotelli & Grave, 1996). Par exemple, on suppose bien souvent que l'habitat préférentiel d'une espèce se trouve sur tous les sites de l'étude, ce qui est peu plausible. Bien que beaucoup de critiques aient pu être émises à l'encontre de l'approche par modèles nuls, elle reste néanmoins la méthode la plus développée en écologie des communautés et son domaine d'application va aujourd'hui beaucoup plus loin que la simple question d'origine du poids de la compétition dans l'assemblage des communautés. Initialement développés pour des questions d'écologie des communautés, il est à noter que les modèles nuls se sont aujourd'hui imposés dans d'autres domaines tels que la biologie évolutive, la paléobiologie ou encore la reconstruction phylogénétique pour ne citer que quelques exemples.

2.1.4. Interprétation

Deux visions s'opposent quant à l'interprétation des modèles nuls. Pour certains, un modèle nul n'est rien d'autre qu'une forme complexe d'un test statistique, sans considération d'un quelconque mécanisme (Simberloff, 1983). Pour d'autres, un modèle nul représente un choix de scénario explicite pour tester les effets d'interactions biotiques dans les assemblages (Colwell & Winkler, 1984). Il est certainement plus judicieux de considérer les modèles nuls comme un intermédiaire de ces deux visions: les modèles nuls permettent de décrire les assemblages des communautés, sans préciser les processus mis en jeu, et de montrer en quoi

une théorie en écologie (i.e. neutralité de l'assemblage) peut ou non générer à elle seule des patrons d'espèces similaires à ceux observés. Ils forcent généralement la communauté scientifique à réfléchir aux mécanismes alternatifs qui pourraient expliquer les patrons observés déviant de l'hypothèse nulle. De plus, le sens de la déviation par rapport à l'hypothèse nulle peut s'avérer très informatif sur l'importance relative de différents processus supposés jouer conjointement (Gotelli & Grave, 1996).

2.1.5. Modèles nuls et dispersion des traits fonctionnels

Lorsqu'on analyse la dispersion des traits fonctionnels d'une communauté et qu'on s'intéresse aux processus responsables de cette dispersion, il est nécessaire de comparer l'assemblage observé à des assemblages théoriques où tous les processus d'assemblages (cf. structuration) doivent être gommés. C'est seulement dans ces conditions que le couplage de l'approche modèle nul avec l'approche par les traits fonctionnels prend son sens. Afin de déterminer si la distribution observée d'un trait au niveau de la communauté est plutôt convergente (effet filtre environnemental et dispersion) ou divergente (compétition), on comparera nos distributions observées à des distributions théoriques sous une hypothèse nulle, où l'effet filtre des niches et la limitation de similarité des traits auront été effacés par le processus de randomisation (Figure 6).

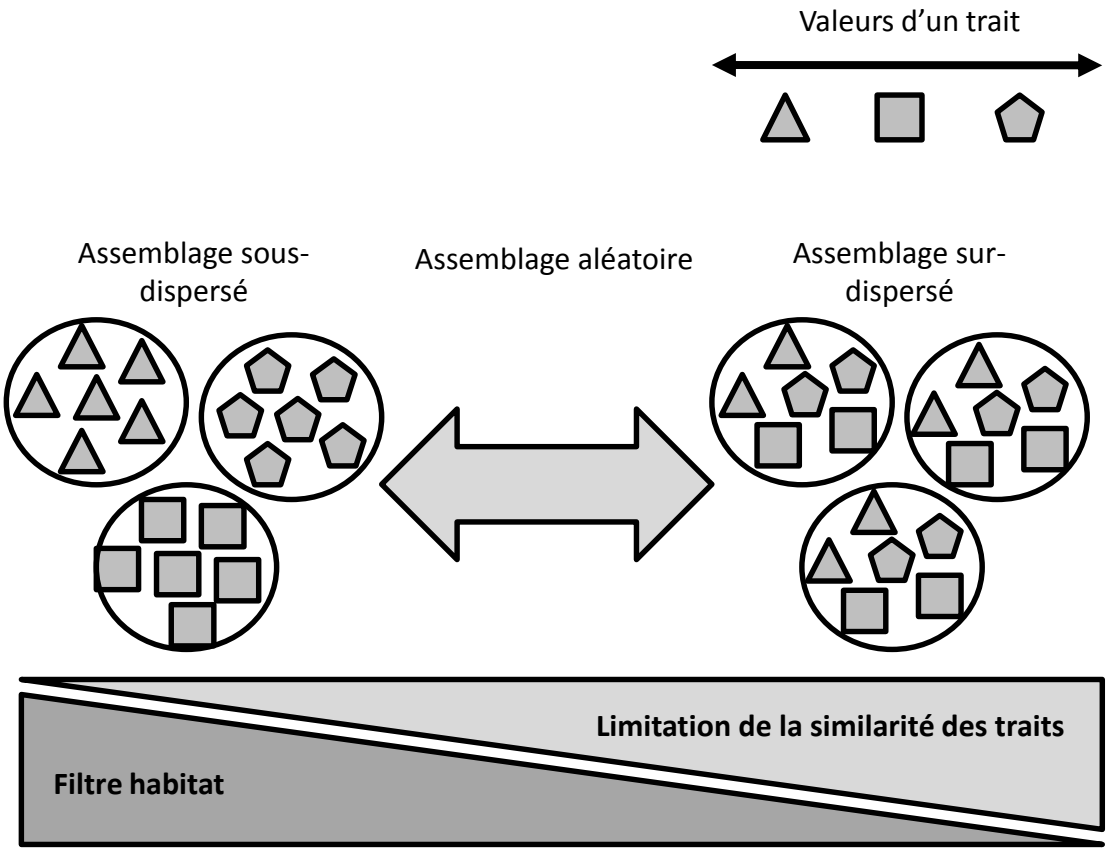


Figure 6 Distribution d'un trait et mécanisme de structuration. Les trois cercles représentent trois sites. Les deux structurations extrêmes de la distribution de la valeur du trait sont présentées : l'assemblage sous-dispersé, où l'effet filtre de l'habitat conduit les individus à présenter les mêmes valeurs de trait, et l'assemblage sur-dispersé où la compétition (induisant une limitation de la similarité des traits) conduit les espèces à minimiser la redondance des valeurs de trait. Entre ces deux assemblages extrêmes, l'assemblage peut sembler aléatoire dû à une compensation des deux mécanismes.

On peut alors tester si la distribution d'un trait est plus sous-dispersée (convergente) ou sur-dispersée (divergente) par rapport à un assemblage nul (Schamp *et al.*, 2008). Les indices de diversité fonctionnelle (voir Mason *et al.*, 2005 et Schleuter *et al.*, 2010 pour une très bonne explication des différents types d'indices de diversité fonctionnelle), permettant sa description, sont fortement dépendants des abondances relatives des espèces ainsi que de la distribution générale des différents traits (Botta-Dukát, 2005; Villéger *et al.*, 2008).

2.2. Une distribution non aléatoire des abondances des espèces adventices au sein des parcelles cultivées

La parcelle cultivée représente un environnement particulier de par sa nature extrêmement anthropisée et l'important degré de perturbations qui en découle. On sait que les interactions entre les conditions environnementales (pédoclimat et pratiques culturales) et les traits des espèces vont au moins partiellement déterminer les compositions de flore adventice observée dans une même parcelle. De plus, la dominance de la culture (de l'ordre de 80 à 400 plants/m^2) et la présence d'une forte « remise à zéro » des communautés adventices annuellement (au semis de la culture par exemple) font qu'il est courant de supposer l'inexistence de compétition inter-adventice susceptible de pousser à la sur-dispersion des traits. On attendrait donc une forte structuration des traits par effet de filtres prépondérant si le changement d'année en année des conditions ne venait pas compliquer fortement la donne. Ainsi, la très grande diversité des situations de production, associée à la diversité de composition des communautés rend difficile toute prévision. Dans un premier temps, il nous est donc paru intéressant de réaliser l'analyse de la structuration de l'abondance des espèces adventices de manière globale, sans s'intéresser aux situations (pratiques culturales et biologie des espèces) responsables de cet assemblage, et de regarder si la densité totale d'adventices influence ou non les assemblages observés.

2.2.1. Article (version post-soutenance)

Cette version est la version resoumise à *Weed Research*.

NON RANDOM DISTRIBUTION OF WEED SPECIES ABUNDANCE IN ARABLE FIELDS

B BORGY*, S GABA*, S PETIT* & X REBOUD*

*INRA, UMR 1347, Agroécologie, Dijon, F-21000, France

Running head: Distribution of abundances in weed communities

Correspondence*:* X Reboud, UMR 1347, Agroécologie, Institut National de la Recherche Agronomique, 17 rue Sully, 21065 Dijon Cedex, France. Tel: (+33) 3 80 69 31 84; Fax: (+33) 3 80 69 32 62; E-mail: reboud@dijon.inra.fr

Summary
Many decisions for pest control in agriculture are based on densities rather than species presence only. Little is known about the distribution of pest abundance and even less when several species present in a field may interfere. Following the aim of developing an ecologically based approach of weed management, we compared observed co-occurrence patterns with a null model to test whether the abundances of weed species were randomly assembled. The dataset covered the pattern of co-occurrence of abundances of weeds within 200 m^2 plots located in 1143 arable fields scattered across France.
Our results pinpointed a highly significant 'similar attraction - dissimilar repulsion' rule as species reaching similar abundance levels co-occurred significantly more often than expected under the hypothesis of a random distribution of abundance. A similar analysis applied to the 25% plots that sheltered the highest weed densities indicated that the 'similar attraction - dissimilar repulsion' rule held true for low abundance classes but reverted when at least one species was observed at densities above 20 individuals/m². Our analysis extends the analysis of community beyond the sole species richness index by accounting for their abundance. Overall, weed species interfere or respond similarly to an external biotic or abiotic factor, with resulting adjustment of densities at the scale of the agricultural field.

Keywords: null model, species assemblage, infestation, semi-quantitative scale, competition, weed, agroecology.

Introduction

Ecological theory suggests that without constraints on dispersion, the assembly of species within communities rely either on stochastic processes (Hubbell, 2001) or are driven by niche-based processes i.e. habitat filtering and biotic interactions (Cornwell & Ackerly, 2009). Elucidating community assembly rules prevailing in cultivated fields has obvious agro-ecological applications (Booth & Swanton, 2002). Indeed, a major output would be to identify conditions under which natural ecological processes may facilitate the control of crop pests without or in complement to chemical control. In the case of weed communities, a number of studies have investigated the effects of cropping systems on particular species (for example Colbach *et al.*, 2007) or on group of species (for example Legère *et al.*, 2005). These studies show that agronomic practices carried out at the field level can induce marked habitat filtering within weed communities. Weed species that do not bear the necessary life-attributes are filtered out from communities by agronomic practices such as crop type and rotation (Smith & Gross, 2007; Fried *et al.*, 2009; Gunton *et al.*, 2011), timing of tillage (Smith, 2006), levels of fertilizers and herbicide inputs (Storkey *et al.*, 2010). There is on the other hand much less available knowledge on the potential role of the biotic interactions in structuring weed communities. Weed-crop or weed-weed competition may affect the composition of the community present in a field. More generally, predation or parasitism may also contribute to filter species or group of species sharing the appropriate response traits. In cultivated fields, the implications of competition have primarily focused on crop-weed interaction and yield loss resulting from this competition, for obvious economic reasons (Berti & Zanin, 1994; Milberg & Hallgren, 2004; Harlan, 1982; Jones & Walker, 1993; Swinton *et al.,* 1994). However, some weed species are highly competitive (Storkey, 2006) and would therefore be likely to affect not only the crop but also other less-competitive weeds. Prior to the analysis of any particular mechanism accounting for particular assembly rule of species within a field each with its respective population size, one might explore whether such rules may exist and affect the way abundances co-occur between species that, otherwise, share the same agricultural, soil and climate conditions.

In this paper, we investigated patterns of co-occurrence in weed species abundance classes in 200m^2 plots using weed data collected in 1143 plots located in annual crops scattered across France. We addressed the following questions:
Are weed abundances randomly assembled within a plot, i.e. is the abundance reached by any individual species within a plot dependent on the abundance of the co-occurring species?
If such a rule exists, does it still prevail when the biotic interactions are presumably more intense i.e. for plots exhibiting high weed densities?

Materials and Methods

Data set
Weed data were obtained from 1143 arable fields scattered across France (latitudinal range of 761 km, longitudinal range 696 km) and collected between 2002 and 2007 as part of the on-going national Biovigilance project (Fried *et al.*, 2007; Fried *et al.*, 2008). The selection of surveyed fields aimed at covering the diversity of cultural practices and environmental conditions occurring across the country. Weed data were recorded after crop establishment, between 15 March and 25 July according to the crop sowing period (i.e. later for spring sown than for winter-sown crops). Within each field, the abundance of all weed species was recorded twice a year within the crop and in a paired 200 m² plot that received no herbicide so that the seedbank which is readily to germinate can express its potential. Within that untreated area, species abundance was recorded using a six-point scale (Barralis, 1976) : '+' found once in the 2000 m² area; '1' less than 1 individual/m²; '2' 1-2 individual/m²; '3' 3-20 individuals/m²; '4' 21-50 individuals/m²; '5' more than 50 individuals/m².
The entire dataset (hereafter "whole dataset") thus contained a total of 1143 sets of weed data i.e. 253 species. In this study, each field was only used once to avoid pseudo replication with 58 sampled in 2002, 215 in 2003, 383 in 2004, 207 in 2005, 227 in 2006 and 53 in 2007. We also sub-sampled the 25% of sites with the highest total weed densities i.e. the highest sums of abundance (median value of the abundance class) of all weed species occurring within the plot. This sub dataset referred to "high weed density" is composed of 286 sites and 186 species (32 sampled in 2002, 58 in 2003, 112 in 2004, 37 in 2005, 35 in 2006 and 12 in 2007). With this threshold of 25%, the total weed density always equal or exceeds 73 individuals/m^2. In this subset, there were proportionally more plots that were recorded in 2002 and 2004, and less in 2005 and 2006. 100 000 sets of 286 samples were randomly drawn in the whole data-set to generate a random distribution of the number of sites per year. The observed distribution of sites per year in the "high weed density" data-set was compared to the random distribution. Excesses and/or deficits of sites per year were identified using 2.5% and 97.5% confidence boundaries. The same procedure was applied to analyze the distribution of the type of rotations (monoculture versus rotation).

Choice of the null model

The presence-absence (abundance) matrix in which rows are species and columns are sites is a fundamental unit of analysis in community ecology. The basic principle of null model is to randomize the matrix entries i.e. species occurrence (abundance) per site in order to test deviation between the observed species community assembly and predicted ones under a random assembly hypothesis. This methodology is recognized for its robustness and lack of dependence upon assumptions regarding the kind of data and their supposed distributions. There are up to nine different null models that differ in their degree of conservative fixed constraints i.e. the occurrence frequencies among species (row totals) and species richness among sites (column totals) (Gotelli, 2000). Here, we aimed at (i) relaxing the possible association between classes of abundance of any pair of species in a site, and (ii) still accounting for a non equiprobability of the site quality i.e. sites have different species richness as well as (iii) accounting for a non-uniform distribution of species abundance in the datasets. For this purpose, species abundance per site was randomized but the global distribution of abundance classes for each species and local species diversity (with both rich and poor sites) was maintained. The most appropriate null model that fits these expectations includes two constraints (Hardy, 2008). First, in order to conserve the species richness among sites (column totals), we extended Gotelli's suggestion (Gotelli, 2000) and maintained rare/frequent and sparse/abundant characteristics of each species by retaining the number of each abundance class for each species in the null hypothesis. Fixing the totals for columns ensured that differences between observed and simulated data will only reflect the underlying structure of the dataset i.e. only the co-occurrence patterns are randomized. Second, we retained local species richness per site in order to maintain the global species co-occurrence of whole abundance classes and avoid any bias in the analysis of co-occurrence of paired abundance classes. Moreover, this constraint avoids degenerated matrices i.e. a matrix with empty lines or columns which modifies the mean abundance or richness of the subset of the non-empty lines or columns (Gotelli, 2000). Thus, species abundances were sampled independently among species and all sites are equiprobable for each of the abundance classes.

The deviation from a random assemblage of co-occurrences of species abundance was tested with these two constraints. 1000 randomized matrix were obtained by randomizing the observed matrix using the "swap" algorithm developed for presence-absence matrix and available in the library *vegan* of the statistical software R (Gotelli & Entsminger, 2001). The abundance classes of each species were then redistributed as defined by the "swap" algorithm. The procedure was applied on both datasets.

Co-occurrence index

Co-occurrence patterns in the abundance matrix were measured by computing X^{ij} the average number of species at abundance *j* in sites where at least one species has been observed at abundance *i*. X^{ij} represents the proportion of co-occurrences between abundance classes in the dataset. The measure was computed for the observed matrix (X^{ij}_{obs}) and for each of the 1000 randomised matrix (X^{ij}_{sim}). A frequency distribution was then generated with values for testing the null hypothesis that X^{ij}_{obs} was drawn at random from the distribution of X^{ij}_{sim}. Following the classical method of statistical inference, the position of the observed value in the tails of the null distribution was used to assign a probability value. The significance level (P-value) is the proportion of values that are more extreme than the observed value in the randomization distribution (Manly, 1991). Significant excess (deficit) of co-occurrences were indicated as the proportion of simulated values (X^{ij}_{sim}) higher (smaller)

than the observed value (X^{ij}_{obs}). The percentage of excess (deficit) of co-occurrences was estimated by computing a co-occurrence index:
$(X^{ij}_{obs} - \overline{X}^{ij}_{sim}) / \overline{X}^{ij}_{sim}$ where $\overline{X}$ is the average of the *X* over the 1000 randomized matrices.

Results

Characteristics of datasets

We recorded a broad range of variations in both weed species richness and total weed densities (Figure 1). The total weed abundance was positively related to species richness. From 11 species onwards, a plateau of 70 plants/m² was reached.

Table 1 presents the median local weed density i.e. median weed density per field and the median species richness in both datasets for each crop type. The subset of high weed density included sites with a total weed density ranging from 73.1 to 434.99 individuals/m². Median species richness per crop type varied for whole data-set and “high weed density” data-set between 5 to 11 and 8.5 to 11, respectively.

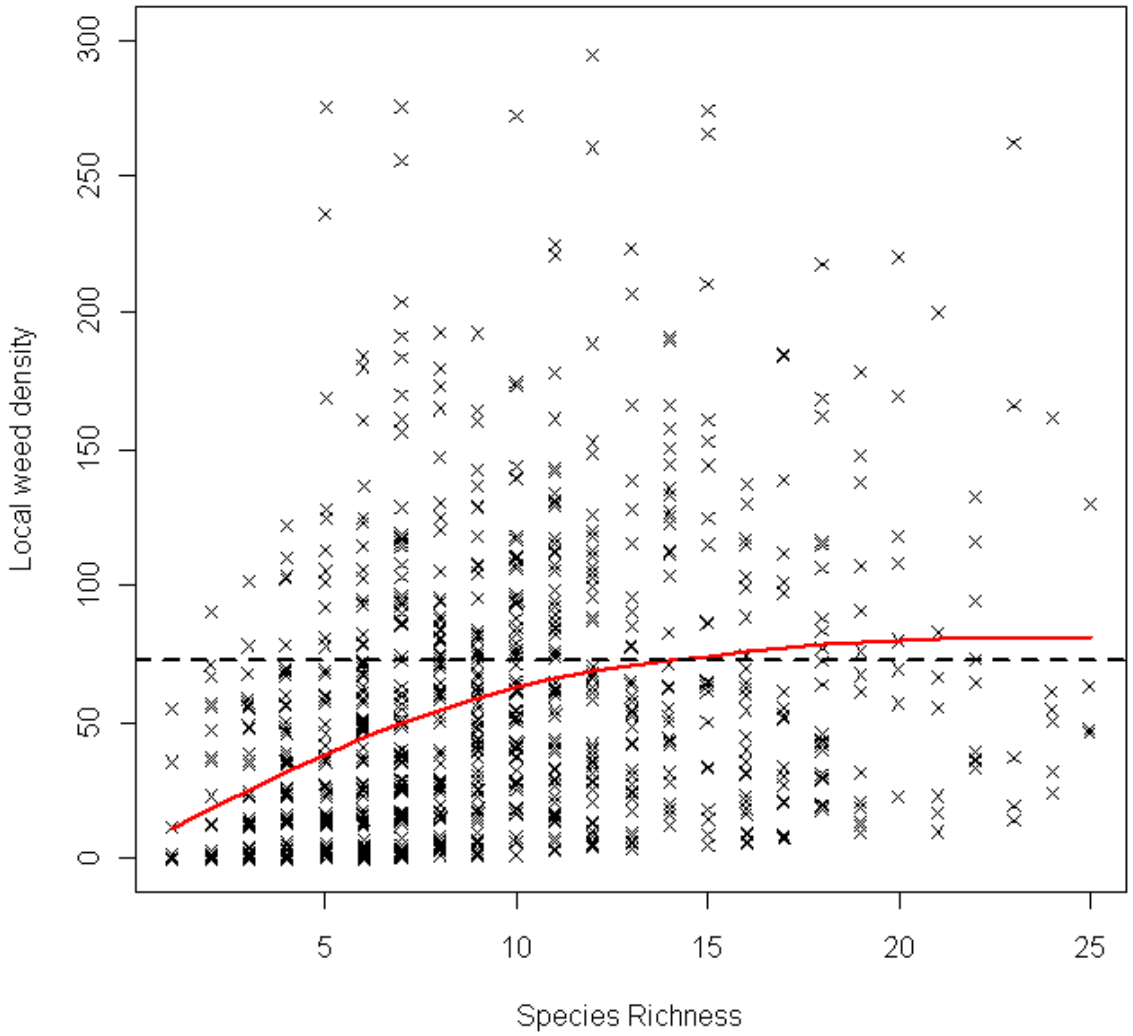


Figure 1: Relationship between species richness and local weed density in the observed whole dataset. Line represents the local polynomial regression (Loess) on observed datasets used to synthesize the relationship between weed density and species richness. Due to too small number of samples with very high species richness, Loess regression was computed only for sites with species richness lower or equal to 25 species. The parameter α which controls the degree of smoothing was arbitrary fixed to 0.75.

The distribution of crop types were slightly different in the two datasets since maize crops were in excess and winter wheat crops were in deficit in the “high weed density” dataset compared to the whole dataset. Another consequence is that the“high weed density” dataset included more fields cropped as monoculture. 24.2% of fields cropped were in monoculture in the “high weed density” while under a hypothesis of random distribution of the rotation type, an average of 8.9% was expected.

Table 1 Number of plots, median weeds density and mean species richness per crop type in the whole and in the 'high density' field datasets.

	Whole dataset			High density dataset		
Crop type	Number of plots	Median weeds density	Median species richness	Number of plots	Median weeds density	Median species richness
Beet	28 (2.4%)	28.75	8.5	6 (2.1%)	115.99	10.5
Spring barley	35 (3.1%)	34.99	8	6 (2.1%)	118.74	11
Winter barley	36 (3.1%)	40.75	5	10 (3.5%)	99.25	8.5
Maize	252 (22.0%)	62.50	9	101 (35.3%)	114.5	11
Rape	83 (7.3%)	32.49	8	12 (4.2%)	123.75	9.5
Sunflower	59 (5.2%)	20.49	11	9 (3.1%)	89.0	10
Winter wheat	384 (33.6%)	27.75	7	81 (28.3%)	110.0	10
Other/NA	242 (16.0%)	36.5	8	61 (21.3%)	110.5	10

Local weed density is the median of the sum of median values of the observed abundance classes

Distribution of abundance classes in the whole dataset

Across the 1143 plots and 253 species, 35% of the species were observed at an abundance of less than 1 individual/m² (Class1). Class+, Class2 and Class3 represented about 20% of occurrence, each. Classes of high abundances were not often recorded, respectively in 5% and 3% of observations for Class4 and Class5.

Table 2 presents the mean number of species in a site observed at abundance *j*, given that at least one species is at abundance class *i*. Regarding the co-occurrence of species abundance, we observed a negative relationship between species with low abundance and species with high abundance. The species with lower abundance classes (*i* = Class+ or Class1) tended to occur more frequently with species of low abundance classes (*j* = Class+ or Class1) than with species of high abundance classes (*j* = Class4 or Class5).

Table 2 Mean number of species for each abundance class in the whole dataset. X_{obs}^{ij} represents the mean number of species with abundance class *j* in sites (=rows) where at least one specie with abundance class *i* occurred. Values on the diagonal are written under the form *1+X* to account for the fact that, by construction, at least one abundance class *j* is already present when abundance class *i=j*. N indicates the number of concerned sites.

	Mean number of species with abundance class j in observed dataset						
i[(a)]	+	1	2	3	4	5	*N*
+	1+1.92	3.67	1.87	1.4	0.38	0.18	572
1	1.57	1+2.82	2	1.71	0.42	0.21	927
2	1.57	3.36	1+1.64	1.95	0.46	0.22	796
3	1.28	3	2.11	1+1.55	0.48	0.25	786
4	1.24	2.97	2.15	2.14	1+0.44	0.36	331
5	1.18	2.48	1.8	2	0.66	1+0.31	205

[(a)] : Sub dataset where at least one abundance of specie is equal to *i*

General pattern of abundance co-occurrence

Table 3 presents the deviations from the null model for each pair of abundances *ij* in the whole dataset. Out of the total 36 pairs of possible abundance co-occurrence, 30 significantly differed from the null model. There were excesses in the co-occurrence of similar abundance classes (the diagonal of table 3) and deficits in the co-occurrence of contrasted abundance classes (the opposite corners of Table 3). For example, in fields where at least one species was highly abundant i.e. Class5 (more than 50 individuals/m²), there were on average 33% more species of abundance Class4 (21-50 individuals/m²) in the observed dataset compared to the randomized datasets. Therefore, abundance classes observed within a plot tended to be more homogenous than what would be expected under a random assembly hypothesis.

Table 3 Proportion of number of species for each abundance class in the whole dataset. Values on the diagonal show highly significant excess of co-occurrence of species with similar abundance class, while upper right and lower left corners of the table show highly significant deficit.
Pvalue: *** <0.001; ** <0.01; * <0.05; +<0.1.

	Proportional excess or deficit of observed co-occurrence of j when i					
i [(a)]	+	1	2	3	4	5
+	0.41^{***}	0.07^{***}	-0.08^{***}	-0.29^{***}	-0.18^{***}	-0.28^{***}
1	0.03^{***}	0.11^{***}	0.043^{***}	-0.06^{***}	-0.03^{*}	-0.12^{***}
2	-0.01^{ns}	0.01^{ns}	0.12^{***}	0.03^{**}	0.03^{+}	-0.12^{***}
3	-0.2^{***}	-0.11^{***}	0.06^{***}	0.12^{***}	0.07^{*}	-0.02^{ns}
4	-0.3^{***}	-0.2^{***}	-0.01^{ns}	0.04^{ns}	0.16^{***}	0.33^{***}
5	-0.35^{***}	-0.35^{***}	-0.2^{***}	-0.05^{ns}	0.33^{***}	0.14^{***}

[(a)] : Sub dataset where at least one abundance of species is equal to *i*

Patterns of co-occurrence of abundances when local weed density is high
As previously, the analysis revealed significant excesses in the co-occurrence of species with similar abundance classes but this only held true for Class+ up to Class3. Indeed, in plots where an individual species was observed at a density higher than 20 individuals/m² (Class4 and Class5), co-occurrence of abundance classes deviated from a random co-occurrence pattern but in a totally different way (Table 4). In such plots, there were many instances of deficits of co-occurrence between Class4 or Class5 and the other abundance classes (columns 4 and 5 of Table 4). Moreover, when at least one species reached a high abundance (Class4 and Class5), species richness in the field was significantly lower than expected under the null model (Figure 2). With one species at a density above 50 individuals/m², more than two species were missing in the community. By contrast, this was compensated in sites where at least one species had a low value of abundance (Class+, Class1 and Class2), where species richness was then significantly higher in the observed dataset (Figure 1).

Table 4 Proportion of number of species for each abundance class in the "high density" subset. Excesses of similar classes (i.e., diagonal values on the table) are still observed for abundance Class+, Class1, Class2 and Class3. Deficits in cooccurrence are observed for the two highest abundance classes i.e. Class4 and Class5. P-value: *** <0.001; ** <0.01; * <0.05; +<0.1.

	Proportional excess or deficit of observed co-occurrence of j when i					
i [(a)]	+	1	2	3	4	5
+	0.55^{***}	0.15^{***}	0.05^{*}	-0.23^{***}	0.02^{ns}	-0.16^{***}
1	0.07^{***}	0.12^{***}	0.10^{***}	-0.02^{*}	0.04^{**}	-0.11^{***}
2	0.14^{***}	0.13^{***}	0.16^{***}	-0.01^{+}	0.06^{***}	-0.13^{***}
3	-0.06^{***}	-0.01^{ns}	0.03^{***}	0.06^{**}	-0.04^{**}	-0.11^{***}
4	-0.05^{+}	0.01^{ns}	-0.01^{ns}	-0.12^{***}	-0.03^{ns}	-0.25^{***}
5	-0.16^{**}	-0.17^{***}	-0.25^{***}	-0.26^{***}	-0.32^{***}	-0.05^{+}

[(a)] : Sub dataset where at least one abundance of species is equal to *i*

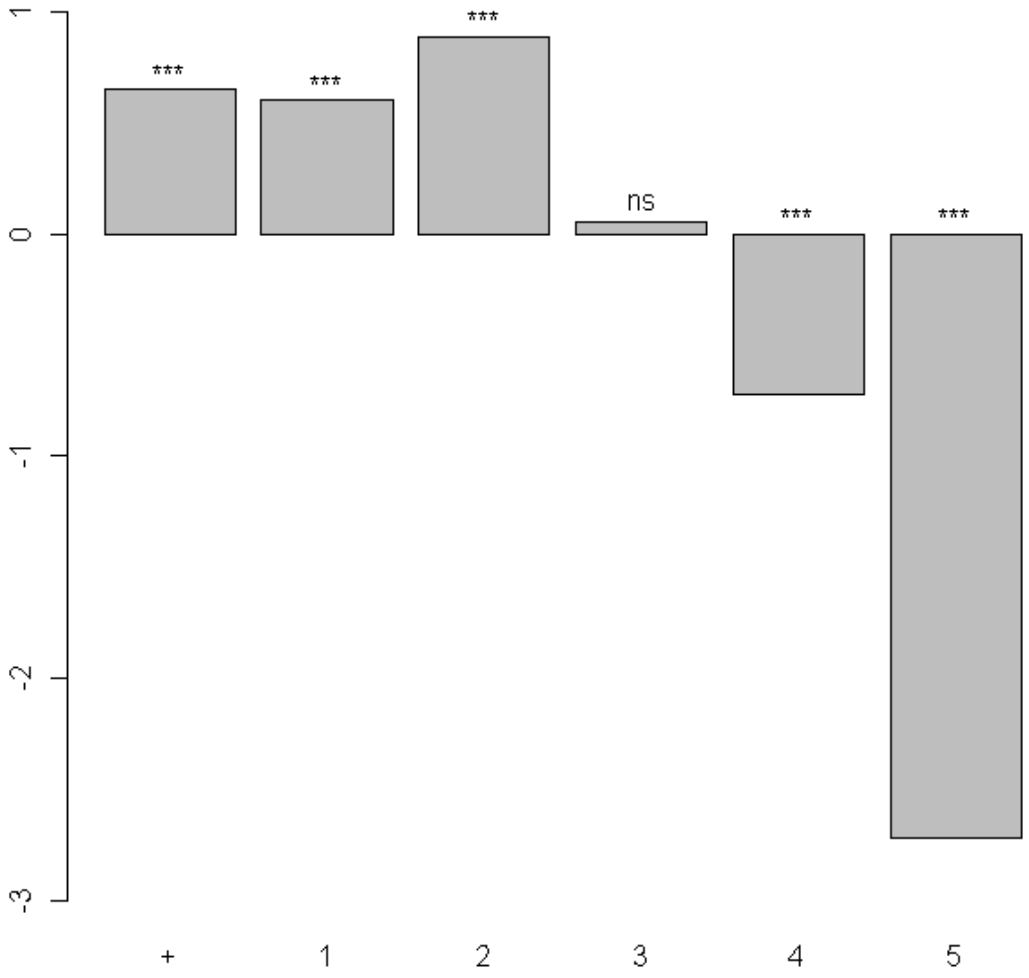


Figure 2: Differences between species richness, given their abundance class, in the observed data in the "high density" dataset and the average species richness computed over the 1000 randomized datasets obtained from the null model. Positive values represent an excess of species richness in the observed dataset while negative values represent a deficit. P-value: *** <0.001; ** <0.01; * <0.05; +<0.1.

Discussion

First, our data indicated that over a large range of crop types and management practices, weed species tend to co-occur more often at similar abundances than expected under a null hypothesis of random co-occurrence of abundances. To account for the high significance deviance from the null model, we gave it the name of 'similar attraction - dissimilar repulsion' rule. This result diverges from what is expected from the ecological concept of limiting similarity (Tilman, 1987), where an abundant species would leave only little possibility for another competing species to occupy the same niche at high abundance. Several hypotheses could explain the patterns we observed. First, conversely to natural habitats that are mostly saturated, arable fields are annually 'reinitialized'. At harvest time, almost all the vegetation is removed and therefore, during most of the growing season, plant density scarcely reaches a level of saturation at which competition (e.g. due to density-dependence process) could structure the plant community. Second, the impact of past and current weed management could strongly shape the observed co-occurrence patterns. The soil seed bank is known to be a buffered memory of past infestations and seed-movements (Barberi *et al.*, 1998). The dataset we used here encompassed a large range of weed management situations, with some fields sheltering species at low density due to intensive and effective chemical and/or mechanical control and other fields occupied by several species at medium to high densities due to a low control. Repeated gaps in the herbicide spectrum over the years may have led to increase the population size of some individual species, which therefore may have become dominant (Reddy, 2004).

This general pattern of abundance co-occurrence was however different in fields where local weed density was particularly high. In such situations, there was a tendency for repulsion between abundances classes as soon as one or several individual plants reached densities above 20 individuals/m². This could result from the fact that fields sheltering high densities of weeds are found in specific cropping systems. Indeed, in this study, the most saturated fields occurred more frequently in maize cropped as a monoculture and less in winter wheat crops, which were often part of diversified crop rotation systems (Table 1) and can lead in the dominance of a few species. Hence, a direct density effect or differences in weed community composition could lead to higher weed-weed interference. Crop diversity has an effect on both diversity and composition of the weed seed bank (Bohan *et al.*, 2011). The soil seed bank has even been suggested being a major factor that could mediate the abundance of niches and the resulting weed-crop competition relationship (Smith *et al.*, 2009). Following this logic, the co-occurrence of species with high abundance within a rotation could result from a higher number of contrasted microhabitats within the field (Smith *et al.*, 2009).

Second, as our weed data was recorded in 200 m² plot, therefore limiting the number of resource pools, it is possible that weed-weed competition could partly explain the deviations observed at high weed density. We observed here that local weed abundance reached a plateau, and that this occurred in the richest fields (Figure 1). Weed-weed competition or another biotic interaction could thus act to regulate species abundance but not necessarily its presence in a field. Moreover, the dominance of a species reaching high abundance class (class4 or class5) seems to reduce species richness in field ; beyond the positive correlation found between total abundance and species richness the saturation observe means a potential of up to two species missing when one species at least reaches a density above 50 plants per m^2 (Figure 2). Nevertheless, we cannot reject that this deficit of species could also result from a bias in the sampling method since the error of detection of species at low density may increase when other species are highly abundant. All of this suggests that some

processes either linked to species interference or to the intensity of selective pressure imposed by external factor such as cropping systems or dependence upon a common regulating predator act at high density or that their impact is only detected at high density.

More specific studies should be conducted to explore whether the signal of weed-weed interaction could be enhanced by (i) conducting samplings over smaller areas to fit the scale at which direct interference between species act, (ii) including a measure of crop density so that all the plants present are accounted for in a survey, (iii) focusing also on non-cropped habitats such as the tilled strip, where weed-crop competition is absent and (iv) conducting experiments in which weed density and/or species composition are experimentally manipulated.

Conclusions

By applying a methodology developed by ecologist to test the random assemblage of species abundances within communities, we showed that weeds do not achieve their level of infestation independently one from the other. Contrary to the expectation that a species achieving high level of infestation would reduce the possibility for other species to reach a high abundance, we found that in most agronomical situations, weed abundances are positively associated. This result may first suggest that in most situations, agro-ecosystems differ from other natural systems near biotic saturation. It also strongly suggests that weed species can access different resource pools within arable fields. Accordingly, in fields that were the most saturated with weeds, we detected the opposite trend, i.e. negative associations between abundance classes when one or few species reached high densities. Such pattern could be explained by the environmental and agronomical characteristics prevailing in the most saturated fields and/or an effect of higher biotic interactions.

Acknowledgements

Benjamin Borgy is funded through a fellowship ANR-OGM Vigiweed. Thanks to Didier Alard for fruitful discussions.

References

BARBERI P, COZZANI A, MACCHIA M & BONARI E (1998) Size and composition of weed seedbank under different management systems for continuous maize cropping. *Weed Research* 38, 319-334.

BARRALIS G (1976) Méthode d'étude des groupements adventices des cultures annuelles. In : 5ème Colloque International sur l'Ecologie et la Biologie des Mauvaises herbes, INRA, Dijon, France, 59–68.

BERTI A & ZANIN G (1994) Density equivalent - a method for forecasting yield loss caused by mixed weed populations. *Weed Research* 34, 327-332.

BOHAN DA, POWERS SJ, CHAMPION G, HAUGHTON AJ, HAWES C, SQUIRE G, CUSSANS J & MERTENS SK (2011) Modelling rotations: can crop sequences explain arable weed seedbank abundance? *Weed Research* 51, 422-432.

BOOTH BD & SWANTON CJ (2002) Assembly theory applied to weed communities. *Weed Science* 50, 2-13.

COLBACH N, CHAUVEL B, GAUVRIT C & MUNIER-JOLAIN NM (2007) Construction and evaluation of ALOMYSYS modeling the effects of cropping systems on the blackgrass life-cycle: From seedling to seed production. *Ecological Modelling* 201, 283-300.

CORNWELL WK & ACKERLY DD (2009) Community assembly and shifts in plant traits distributions across and environmental gradient in coastal California. *Ecological Monograph*, 79,109–126.
FRIED G, REBOUD X & GASQUEZ J (2007) "Le réseau "Biovigilance flore" : présentation du dispositif et synthèse des premiers résulats", *in* AFPP, ed.,'XXième Conférence du COLUMA: Journées Internationales sur la Lutte contre les Mauvaises Herbes', Dijon, France, pp. 315-325.
FRIED G, NORTON LR & REBOUD X (2008) Environmental and management factors determining weed species composition and diversity in France. *Agriculture Ecosystems & Environment* 128, 68-76.
FRIED G, CHAUVEL B & REBOUD X (2009) A functional analysis of large-scale temporal shifts from 1970 to 2000 in weed assemblages of sunflower crops in France. *Journal of Vegetation Science* 20, 49-58.
GOTELLI NJ (2000) Null model analysis of species co-occurrence patterns. *Ecology* 81, 2606-2621.
GOTELLI NJ & ENTSMINGER GL (2001) Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia* 129, 281-291.
GUNTON RM, PETIT S & GABA S (2011) Functional traits relating arable weed communities to crop characteristics. *Journal of Vegetation Science* 22, 541-550.
HARDY OJ (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology* 96, 914–926.
HARLAN JR (1982) Relationships between crops and weeds. In : Holzner W & Numata M *Biology and ecology of weeds,* Dr. W. Junk Publishers, The Hague.
HUBBELL SP (2001) The Unified Neutral Theory of Biodiversity and Biogeography, Princeton University Press, Princeton, NJ.
JONES RE & WALKER RH (1993) Effect of interspecific interference, light-intensity, and soil-moisture on soybean (glycine-max), common cocklebur (xanthium-strumarium), and sicklepod (cassia-obtusifolia) water-uptake. *Weed Science* 41, 534-540.
LEGÈRE A, STEVENSON FC & BENOIT DL (2005) Diversity and assembly of weed communities: contrasting responses across cropping systems. *Weed Research* 45, 303-315.
MANLY BFJ (1991) Randomization tests and confidence intervals. In: *Randomization and Monte Carlo methods in biology*, Chapman and Hall, United Kingdom.
MILBERG P & HALLGREN E (2004) Yield loss due to weeds in cereals and its large-scale variability in sweden. *Field Crops Research* 86, 199-209.
REDDY KN (2004) Weed control and species shift in Bromoxynil- and Glyphosate-resistant cotton (Gossypium hirsutum) rotation systems. *Weed Technology* 18, 131-139.
SMITH RG (2006) Timing of tillage is an important filter on the assembly of weed communities. *Weed Science* 54, 705-712.
SMITH RG & GROSS KL (2007) Assembly of weed communities along a crop diversity gradient. *Journal of Applied Ecology* 44, 1046-1056.
SMITH RG, MORTENSEN DA & RYAN MR (2009) A new hypothesis for the functional role of diversity in mediating resource pools and weed-crop competition in agroecosystems. *Weed Research* 50, 37-48.
STORKEY J (2006) A functional group approach to the management of UK arable weeds to support biological diversity. *Weed Research* 46, 513-522.
STORKEY J, MOSS SR & CUSSANS JW (2010) Using assembly theory to explain changes in a weed flora in response to agricultural intensification. *Weed Science* 58, 39-46.

SWINTON SM, BUHLER DD, FORCELLA F, GUNSOLUS JL & KING RP (1994) Estimation of crop yield loss due to interference by multiple weed species. *Weed Scie*nce **42**,103-109.
TILMAN D (1987) The importance of the mechanisms of interspecific competition. *The American Naturalist* 129, 769-774.

2.2.2. Répartition intra-parcellaire

Afin de pouvoir mieux évaluer une possible compétition inter-adventices, il est important de se placer à une échelle fine, compatible avec le processus d'interférence que l'on souhaite suivre et donc de disposer des répartitions spatiales à l'échelle intra-parcellaire des différentes espèces présentes. Pour cet objectif la difficulté réside dans la construction de carte de répartition spatiale alors même que l'on ne possède en général pas de relevés exhaustifs de l'ensemble de la parcelle échantillonnée. J'ai ainsi participé à la finalisation d'un article visant à développer une méthode permettant de reconstruire les cartes de répartition spatiale de l'abondance des adventices dans les parcelles à partir de ces deux types d'informations (note abondance et comptage) (Bourgeois *et al*, 2011 « Inferring weed spatial distribution from multi-type data », dans l'annexe page 162). Ce travail fait progresser la cartographie des adventices au sein des parcelles cultivées en permettant l'utilisation de données de types différentes (comptage et abondance) quantifier sur des supports de taille différente.

Cette approche permettra d'étudier la dynamique spatio-temporelle des communautés adventices dans différents systèmes de culture et les possibles interactions locales conduisant aux patrons observés. Ceci est très complémentaire des travaux conduits à des échelles plus larges de réseaux d'observatoires couvrant ainsi une grande diversité de situations.

.

2.3. *SwapClass* : un modèle nul adapté aux données de classes d'abondance

Les modèles nuls adaptés à des données de présence/absence et des données quantitatives ont largement été développés et analysés. Néanmoins, les données de terrain sont souvent collectées sous un format intermédiaire sur une échelle semi-quantitative (classes d'abondance) : l'échelle de Barralis (cf. section 1.3.2). Aucune des méthodes actuellement existantes n'étant adaptées à ce format, une majorité d'études repose sur l'acceptation d'une forte perte d'information ramenant le système à son niveau de contraste présence/absence. Comme nous pensons qu'une part importante de l'information est contenue dans la distinction entre rare et abondant, cette perte de prise en compte de l'abondance nous semble particulièrement dommageable.

Nous avons voulu remédier à cette absence de méthodes adaptées en cherchant une solution qui permette de viser les différents types d'analyses basées sur l'utilisation de modèles nuls (cooccurrences d'espèces à l'échelle semi-quantitative, analyse de dispersion des traits, etc.) et où le degré de contraintes dans les randomisations peut être d'importance relativement élevée.

Il est ainsi nécessaire de développer un modèle nul adapté au format semi-quantitatif. En se basant sur les restrictions classiquement utilisées dans le cas de modèles adaptés au format binaire (présence/absence) et/ou quantitatif, et en s'appuyant sur les méthodes classiquement utilisées (« swap »), j'ai développé une méthode de génération d'un modèle nul simple, générique et adapté au format semi-quantitatif, qui permet de conserver les règles touchant la distribution des classes d'abondances.

2.3.1. Article (version post-soutenance)

Version soumise à Ecological Modelling

SWAPCLASS: A null model adapted to abundance class data in ecology.

B. BORGY[1], X. REBOUD[1], S. GABA[1,2]

[1]INRA, UMR1347 Agroécologie, 17 rue Sully Dijon, F-21065 cedex, France

[2]Corresponding author: sabrina.gaba@dijon.inra.fr

Abstract. Null models are widely used to investigate the extent of community assembly rules. Null model algorithms are efficient for presence/absence and continuous abundance data however no null models are available for semi-quantitative data. However, many data collected from natural surveys or queries are commonly of semi-quantitative nature. Here, we present the *SwapClass* null model which accounts for such semi-quantitative data. This model is derived from the 'swap philosophy' used for presence/absence data. We tested the robustness of the model to errors of type I or type II in a trait-based analysis. The evaluation showed that the efficiency of permutations of the *SwapClass* null model and the random structure of the generated null matrices are dependent on the interaction between the size of the observed matrix, the number of abundance classes and the evenness of their distributions. The use of *SwapClass* model was no relevant for small matrices *i.e.* with a number of rows or columns inferior or equal to 10, a high number of abundance classes, a low evenness of these classes or lower number of rows (sites) compared to columns (species). However in most of the case studies, the *SwapClass* model was found to be robust to detect community assembly patterns as soon as *Nbperm* the index of mean permutation per cell of matrix exceeded 0.5. *SwapClass* is thus an efficient algorithm for investigating community assembly rules for semi-quantitative data. Moreover, the *nbPerm* function allows for testing the potential 'swapability' of a species-site matrix *i.e.* the extent of the deviation from observed communities of the null communities.

Keywords. Null model, semi-quantitative data, functional divergence, swap, community assembly.

1. Introduction

A major research focus in ecology is to understand the underlying processes that generate the patterns ofdiversity and abundance of co-occurring species. The "null model" framework provides a prominently statistical analysis for testing the likelihood of an observed pattern in absence of any particular mechanism*i.e.* resulting from stochastic processes (Gotelli and Graves, 1996). In the last decades, an increasing number of studies has investigated the extent of community assembly rules by studying the species co-occurrence (Boschilia *et al.*, 2008; Ulrich, 2004), the structure of food webs (Vázquez and Aizen, 2003) or the dispersion of functional traits (see for example Stubbs and Wilson, 2004; Cornwell and Ackerly, 2009) together with null model. The principle of a null model is to break down any supposed relationship between coexisting species by generating null communities by randomizing the observed data (or random sampling from a known or supposed distribution) (Gotelli and Graves, 1996). Though, the randomization is designed to produce a pattern that would be expected in the absence of any particular ecological processes (Gotelli and Graves, 1996). Theoretically, the significant differences between observed and simulated communities reveal non-random assembly rules (Petchey *et al.*, 2007; Weiher *et al.*, 2011). However, the specification and choice of the null model have important consequences on the results. Indeed, the test of a null hypothesis can be performed by using several null models differing in their degree of restriction on the rules of randomization of the dataset (Gotelli and Graves, 1996; Gotelli, 2000). Several null models have been developed for presence/absence data (Gotelli and Entsminger, 2001, 2003; Jonsson, 2001; Miklós and Podani, 2004; Wright *et al.*, 1998) and continuous abundance data (Hardy, 2008; Patefield, 1981; Ulrich and Gotelli, 2010). Several restriction rules can be applied either by fixing the number of sites occupied by a species (column total), the number of species in a site (row total) or both.The choice of the restriction rules used to build the null communities can lead to different outputs from the same dataset, which may cause conflicting results for the same assumption. For example, co-occurrence tests *e.g.* C-Score (Stone and Roberts, 1990) are very sensitive to the variation of species occurrence frequencies; hence species frequencies *i.e.* row total should be conserved between observed and simulated communities (Gotelli, 2000). The question of choosing the most appropriated null model has been largely studied and conceptualized by Gotelli (2000) for presence/absence data. He suggested that "swap" methods that preserve both species occurrences and species richness would be the most appropriated and conservative reference for the analysis of co-occurrence patterns. Moreover, swap methods preclude the generation of degenerate matrices (Gotelli, 2000). The situation remains more problematic in case of continuous abundance data. Indeed, in the same way that modifying species occurrences and species richness (column and row occurrence totals) can alter the index used to summarize a pattern (*e.g.* C-score), modifying species abundances and site densities (column and row abundance totals) can generate errors of type I (reject a true null hypothesis) and/or of type II (retain a false null hypothesis) through the use of indexes based on relative abundance of species. Ulrich and Gotelli (2010) suggested that the best null model for analysis of species abundance associations is an individual-based algorithm which assigns individuals randomly

to matrix cells with probabilities proportional to observed row and column abundance totals until, for each row and column, total abundances have been reached. However, species abundance is also commonly recorded by using a qualitative format such as 'none, few or many' (as you may extract from an oral query), or semi-quantitative scale with abundance (or percent of coverage) intervals expressed in a finite number of ranked categories. In this last case, a value may be associated to each class that represents the median of the corresponding interval (for example Daubenmire, 1959; Barralis, 1976; Bonham, 2004; De Bello *et al.*, 2011; Cornwell and Ackerly, 2010).

Individual-based null models are not adapted for manipulating semi-quantitative data since the decomposition of abundance values into individuals may modify the number of classes in the semi-quantitative scale. Moreover, no population-based null model which preserve the population (*sensu* Ulrich and Gotelli, 2010) abundance values and randomize their occurrences among sites or species, can fix both row and column totals (occurrence and abundance totals) at the same time. Although semi-quantitative values are commonly recorded in ecological inventories, a post transformation of the data is often realized, with a dramatic loss of information by restricting the statistical analysis to presence/absence data, for example. The pattern of abundances across replicated assemblages may potentially contain a more complex and subtle signal of community assembly rules than simple binary presence/absence matrices (Ulrich and Gotelli, 2010). In many situations such as biological conservation, control of invasive species or threshold based decision in integrated pest management, abundance was shown to be meaningful. Moreover, contrary to continuous data, many values *i.e.* abundance classes are repeated in the community matrix when using a semi-quantitative scale.

In this paper, we developed a simple null model adapted to abundance class data, the *SwapClass* model. The *SwapClass* model is derived from the 'swap philosophy' used in presence-absence data. The limited number of abundance classes and their repetitiveness allow the extension of the 'swap'method classically used in case of two classes (0 and 1). In the same way that 'swap' methods permute sub-matrices of presence/absence community matrix, semi-quantitative data can be randomized by swapping sub-matrices while row and column marginals are not modified.

After describing the model, we evaluate its efficiency for detecting random and non-random assembly patterns. To do so, we used a measure of functional trait diversity (FD) that reflects the underlying patterns in community assembly (Ackerly and Cornwell, 2007; Mouchet *et al.*, 2010). Analyses of functional trait dispersion and indexes used are known to be sensitive to relative abundances of species. Matrix characteristicsare known to have an effect on the outcome of null model analyses of species co-occurrence (Gotelli, 2000).Therefore, the model is evaluated on artificial matrices of several matrix size, number and evenness of classes, and for which the ability to be permuted is quantified.

2. Methods

2.1. '*SwapClass*' null model

In a 'swap' algorithm, randomly chosen submatrices of the form:

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \quad \text{or} \quad \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

are selected, and the cells of the submatrices are swapped. The submatrices are not necessarily formed from adjacent rows and columns; any submatrix of this form can be swapped.
Swapping creates a new matrix configuration, but does not alter row and column totals (Gotelli and Entsminger, 2001). The same method can be generalized to abundance class matrices. Submatrices of the form are randomly drawn:

$$\begin{matrix} x & y \\ y & x \end{matrix}$$

These submatrices are selected and their cells are swapped. Similarly to the presence-absence Swap, multiple classes can be swapped even though rows and columns are not adjacent or the number of several classes for each row and column are not equal. As a result, a high number of swaps break down the relationship between co-occurring abundance classes. No attention is paid on the choice of pairs of rows or columns selection i.e. for each swap rows and columns probabilities are uniform. A swap is made only if row and column totals are maintained. The algorithm swaps submatrices until the expected number of swap is reached. As in the 'swap' algorithms, performing several null matrices can be recursive (the algorithm starts from the last performed matrix) with *burn-in* (number of swaps before recording the first null communities) and *thin* parameter (number of swaps between two recordings of null communities) (Gotelli and Ulrich, 2011).

2.2. Generating artificial matrices

Artificial matrices were randomly generated to evaluate the applicability domain of the *SwapClass* algorithm *i.e.* the matrix structure, number of abundance classes and the distribution of abundances. Artificial matrices were built up using a simulation model of community assemblage and following one of three common hypotheses in ecology: null assemblage, habitat filtering and limiting trait similarity (Kraft *et al.*, 2007).
We evaluated the number of swaps that can be performed for each artificial matrix. This number was calculated by summing the number of 'swappable' sub-matrices per pair of abundance classes. The number of swappable sub-matrices reflects the extent of the deviation from observed communities of the null communities. This highly depends on the structure of

the matrix and will directly impact the output of the analysis.The smaller the number of swappable matrix, the closer to the observed matrix the null communities will be. We built an index, *nbPerm*, which precisely quantifies how matrix can be swapped. It is the sum of 'swappable' sub-matrices divided by the number of cells in the matrix (number of rows * number of columns).

2.2.1. Community assembly models

Following the methodology used by Kraft *et al.* (2007), three community assembly models simulated local assembly of species abundance classes according to the three assembly hypotheses. Species trait values and the optimum trait value per sites were randomly drawn within a uniform distribution, $U(0,1)$. We calculated the average distance between each of the optimum trait values and the total pool species trait values. For each site, we assumed that the richest sites would have a smaller mean distance between optimum traits and the poorest a higher mean distance. As a result, the optimum trait values were assigned to a site depending on its species richness and the species richness in the other sites.

Random community (H0). Species abundances per site i.e. abundance classes were randomly distributed (Fig. 1a).

Habitat filtered community (H1H). Under the habitat filtering hypothesis, we assume a trait convergence distribution (Cornwell *et al.*, 2006). To mimic this process, the number of trait values was higher around the site trait optimum. As a result, in a site, species with a trait value close to the site trait optimum have a high value of abundance class (Fib 1b).

Limiting trait similarity community (H1L). Trait divergence is assumed under limiting similarity process (Petchey *et al.*, 2007). As a result, in each site co-occurring species harbor trait values which are more different than expected for a random assembly hypothesis. To mimic this process, we first compute the distance between the trait values of each pair of species. Then we randomly remove one of the species within the pair with the smallest distance until reaching the site species richness. This process tends to increase the divergence between the remained species. Finally, abundance classes were assigned to species by assigning the highest classes to the species with the more different trait values *i.e.* higher distance.

A final step ensured that each species occurred at least once in the three community matrix H0, H1H and H1L. The community assembly models also returned the trait values of species used for H1H and H1L that will be used for trait dispersion index calculation. An example of matrices returned by the community assembly models is given in Appendix (Figure S1).

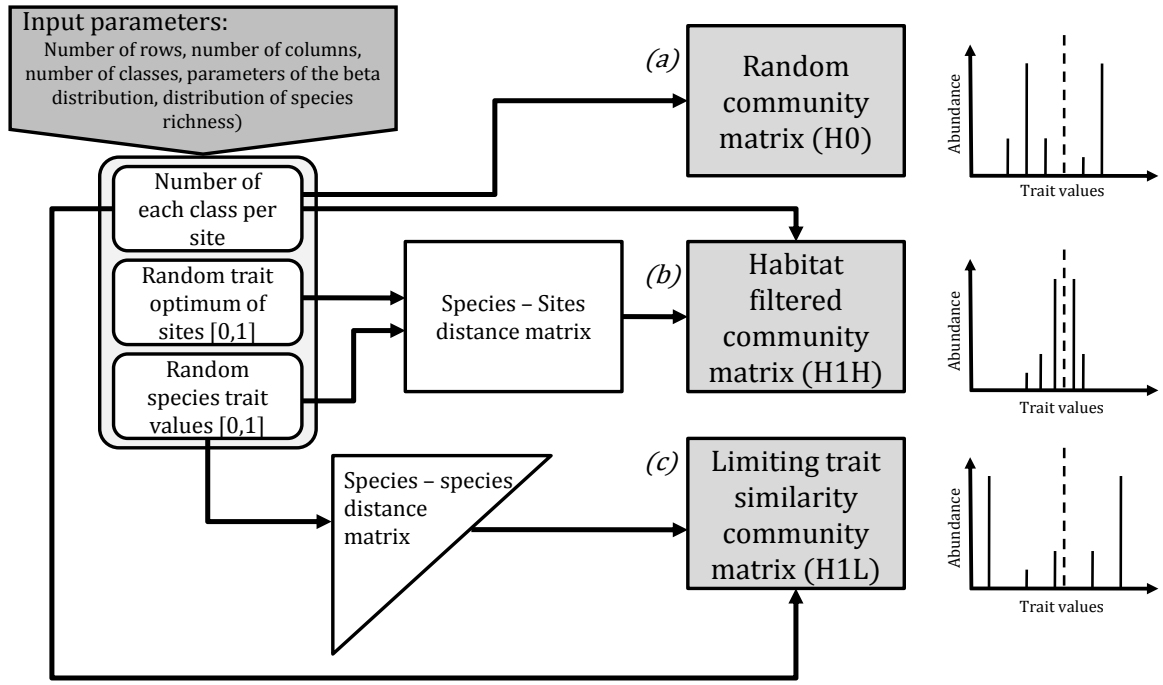


Figure 1: Procedure for the generation of community assembly models. A matrix of each assemblage hypothesis is build up from a set of input parameters. (*a*) Random community matrix (H0) is build up by randomly sampling of classes values per site. (*b*) Habitat filtered community matrix (H1H) is build up by sorting classes values per site into descending order with distance of species trait values to trait optimum of sites.(*c*) Limiting similarity community matrix (H1L) is build up by two successive steps: First, simulation in a site starts with the entire species pool. One species of the closest pair of species is randomly removed, and sequentially until the total number of species of site is reached. Then non null classes values of a site are sorted into ascending order with mean distance of each species to the other species in that site. Abscises of figures on the right side are traits values, ordinates are value of classes. Dotted lines represent the optimum trait value of a site, while continuous lines represent both trait and abundance values of species.

2.2.2. Matrix characteristics

The matrix varied according to (i) the matrix size (number of sites *i* [rows] and species *j* [columns]), (ii) the number of abundance classes *n*, (iii) the evenness of abundances classes, (iv) the distribution of species richness among the sites and (v) the assembly rule hypotheses (random community [H0], habitat filtered community [H1H] and limiting trait similarity community [H1L]) .The number of sites and species define the number of rows and columns of the three artificial matrices (H0, H1H and H1L), respectively. Values of classes are regularly distributed between 0 and 100. For example, for a 3 point-abundance scale, the values are 0, 50 and 100. Abundance classes are distributed as a Beta distribution in order to control for the evenness and the asymmetry of abundance class distribution. The species

richness per site can be either constrained or unconstrained (in this later case, only distribution of non-null classes determine species richness).
Hence, we determined the number of each class in each site for the three artificial matrices (H0, H1H, H1L). A final step ensured that each site contained at least two species (two non null classes).

2.3. Analysis of SwapClass efficiency

We evaluated the efficiency of *SwapClass* null model for detecting the assembly rules for a set of artificial matrices by conducting three different analyses:

Analysis 1 – distribution of classes. The number of rows was equal to the number of columns. Species richness distributions were unconstrained and were independently determined by the distribution of classes. We used a three factor factorial design where (i) the matrix dimension has four levels ($i = j$ {10, 30, 50, 100}), (ii) the number of classes has four levels (n {3, 5, 10, 20}) and (iii) the Beta distribution has three contrasted sets of shape parameters i.e. (1) low (α=0.1, β=1) , (2) medium (α=1, β=2) and (3) high (α=1, β=1) evenness. The parameters control for the distribution of the abundance classes which (i) follows a convex function for set 1, (ii) decreases linearly with the class values for set 2 and (iii) is uniform for set 3. This factorial design was repeated 100 times, hence a total of 4800 sets of artificial matrices H0, H1H and H1L were analyzed.

Analysis 2 – distribution of species richness. Here, we fixed the same number of rows *i* and columns *j*. Abundance classes are distributed as a Beta distribution of parameters (α=0.1, β=1) (low evenness). We used a three factor factorial design where the matrix dimension has four levels ($i = j$= e{10, 30, 50, 100}), the numbers of classes has three levels (n = e{3, 5, 10}) and three different distributions constrain species richness per site i.e. (1) uniform (from 2 to *j*), (2) normal (mean and standard deviation = *j*/2) and (3) lognormal (log-mean = log(*j*)/2 and log-standard deviation =log(*j*)/4).This design was repeated 100 times, hence a total of 3600 sets of random matrices H0, H1H and H1L were analyzed. Since the distribution of species richness was constrained by a uniform, normal or lognormal distribution, the Beta distribution was only used to define distribution of non null abundances classes.

Analysis 3 – asymmetry of the matrix. First, species richness distributions were unconstrained and were independently determined by the distribution of classes (As in Analysis 1). Abundance classes were distributed as a Beta distribution of parameters (α=0.1, β=1) (low evenness). We used a three factor factorial design where the number of rows *i* and column *j* vary between 10 and 100 by step of 10 and the number of abundance classes has three levels (n {3, 5, 10}). This design was repeated 100 times, hence a total of 30 000 sets of random matrices H0, H1H and H1L were analyzed. Second, we performed the same

analysis but with species richness constrained to a uniform distribution and abundance classes followed a uniform distribution *i.e.* (α=1, β=1).

Overall results over the three analyses. *R*esults obtained through the three analyses were merged together. We first analyzed the percentage of correct assignation of assembly pattern. Then, we analyzed for each of the three assembly patterns, the effect of *nbPerm* and of the ratio of the number of rows – number of columns on the assignation of the correct assembly pattern.

2.4. Functional Diversity criteria

Randomness, convergence or divergence of trait values in community were detected by using Rao's quadratic entropy (RaoQ) (Botta-Dukát, 2005) and the *SwapClass* null model. 200 null communities were recursively generated. The *Burn-in* parameter was set to (number of rows * number of columns * 10) while *thin* parameter was set to the maximal value between 1000 and (number of rows * number of columns). Observed mean RaoQ index was calculated from artificial community matrix and trait values of species. The distribution of the simulated mean RaoQ index was calculated from the set of simulated community matrices and trait values of species. We finally calculated LCL, the proportion of simulated values inferior or equal to observed value, and UCL, the proportion of simulated values superior or equal to observed value. A matrix was identified as random if LCL and UCL were both ≥ 0.05, as structured by habitat filtering if LCL < 0.05 or structured by limiting traits similarities if UCL < 0.05.

The codes and functions for R environment are available in supplementary files.

3. Results

3.1. Overall level of correct assignation over the three analyses

Table 1 presents the percentage of the assembly pattern predicted with the *SwapClass* model according to the assembly pattern in the artificial matrices. The three assembly patterns (H0, H1H and H1L) were correctly identified in 90.01%, 99.8% and 95.1% of the study cases, respectively. Errors in the detection of random pattern (H0) were equally distributed among H1H and H1L. Conversely, when not correctly assigned, the limiting trait similarity patterns (H1L) were in majority misclassified as random patterns (H0).

Table 1: Real against predicted pattern of artificial matrices over the three analyses (H0: random communities, H1H: habitat filtered communities and H1L: limiting traits similarity communities). This table summarizes results on the 205 200 matrices analyzed in the three analyses.

	predicted pattern				
real pattern	H0	H1H	H1L	*nbPerm**	
H0	0.9001	0.0498	0.0502	4.63 3.14)	(+/-
H1H	0.001	0.998	0.0001	3.55 2.31)	(+/-
H1L	0.0485	0.0005	0.951	0.97 0.62)	(+/-

*mean values of *nbPerm* (+/- standard deviation)

We analysed the relationship between a correct assembly pattern detection and two major components of the matrix structure *i.e.* its number of rows - number of columns ratio and its number of allowed permutations per cell (*nbPerm*). None of these components did affect the capacity to detect random patterns (H0) (Fig. 2). Habitat filtering patterns (H1H) were in general detected except when *nbPerm* was lower than 0.5. Limiting similarity patterns (H1L) were poorly detected when *nbPerm* was lower than 0.5 or when the number of rows -columns ratio was lower than 1 *i.e.* higher number of columns (species) than number of rows (sites). Our results also showed that under the methodology used to generate the three contrasted hypothesis, *nbPerm* was the highest for H0 matrices (varied between 0 and 14.2, with mean value equal to 4.63) and the lowest for H1L matrices (varied between 0 and 4.1, with mean value equal to 0.97) and intermediate for H1H matrices (*nbPerm* between 0 and 12, with mean value equal to 3.55). This result suggests that the number of 'swappable' sub-matrices was not equal between the three hypothesis tested with the highest number of potential swaps

for the random pattern. Figure S2 in appendix shows that *nbPerm* remains roughly constant through successive swaps of a matrix.

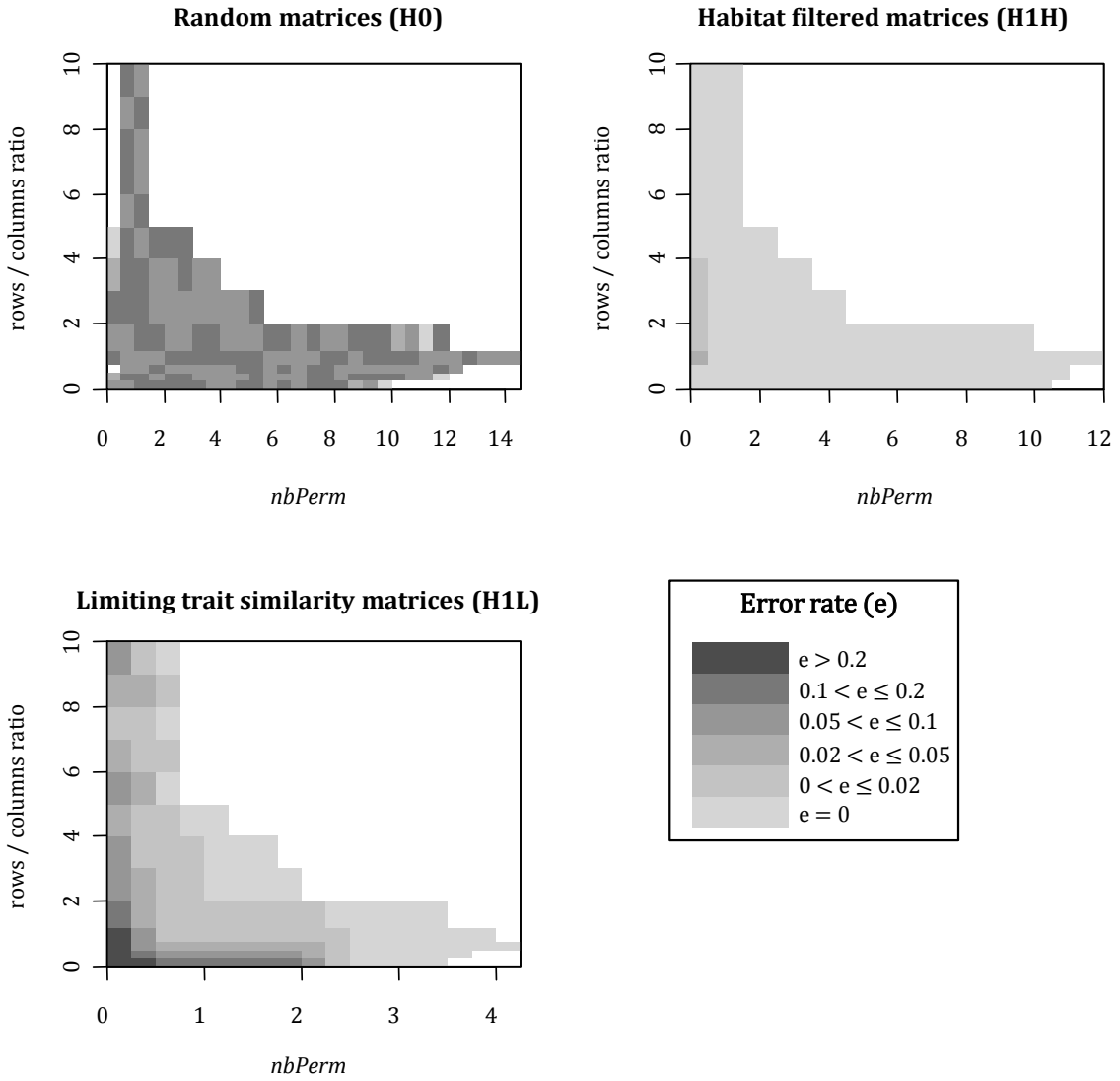


Figure 2: Error rates in the detection of H0, H1H and H1L depending on the number of swappable submatrices (*nbPerm*) and row/column ratio of the matrix. Error rate (e), the error of assembly pattern detection is presented in grey. The white area was not explored in this study. Each sub-figure was drawn from the analysis of 68 400 artificial matrices.

3.2. Analysis 1 – distribution of classes

Table 2 presents the error rates in the detection of the processes underlying the community assembly for different parameters of the Beta distribution (*i.e.* the distribution of the abundance classes), different matrix sizes and different numbers of classes. Error rates in the detection of random patterns (H0) were evenly distributed across different treatments. Error rates in the detection of habitat filtered pattern (H1H) significantly occurred for small matrices ($i = j = 10$) with 20 abundance classes (error rate between 8% and 39%). Error rates in the detection of limiting trait similarity pattern (H1L) varied between 18% and 99% for small matrices ($i = j = 10$). In the case of uniform distribution of classes, error rate was also high for $i = j = 30$ (21%). In these two latter cases (H1H and H1L), the evenness and number of classes increased the error rate of assembly pattern detection.

Table 2: Error rates in the detection of the assembly pattern for random communities (H0), habitat filtered communities (H1H) and limiting traits similarity communities (H1L), for several parameters of Beta distribution (controlling evenness), several numbers of classes (*n*) and several matrix sizes (*i* and *j* represent respectively the number of rows and columns of the matrices). Species richness was only determined by classes' distributions. Each error rate was estimated on 100 artificial matrices.

		Low evenness				Medium evenness				Uniform			
	i, j	n = 3	n = 5	n = 10	n = 20	n = 3	n = 5	n = 10	n = 20	n = 3	n = 5	n = 10	n = 20
H0	10	0.08	0.05	0.12	0.13	0.08	0.10	0.14	0.09	0.06	0.09	0.09	0.06
	30	0.11	0.16	0.09	0.08	0.11	0.09	0.14	0.10	0.06	0.11	0.15	0.15
	50	0.07	0.06	0.14	0.08	0.10	0.09	0.06	0.10	0.09	0.08	0.08	0.07
	100	0.12	0.10	0.10	0.12	0.13	0.09	0.12	0.13	0.08	0.09	0.09	0.08
H1H	10	0.01	0	0.01	0.08	0.01	0.01	0	0.18	0	0	0	0.39
	30	0	0	0	0	0	0	0	0	0	0	0	0
	50	0	0	0	0	0	0	0	0	0	0	0	0
	100	0	0	0	0	0	0	0	0	0	0	0	0
H1L	10	0.31	0.18	0.32	0.79	0.25	0.31	0.54	0.95	0.36	0.32	0.67	0.99
	30	0	0	0	0.03	0	0	0.05	0.07	0.01	0	0.04	0.21
	50	0	0	0	0	0	0	0	0.02	0	0	0	0.02
	100	0	0	0	0	0	0	0	0	0	0	0	0.01

3.3. Analysis 2 – distribution of species richness

The distribution of species richness did not affect significantly the robustness of *SwapClass* in the detection of random assembly (H0) or habitat filtering (H1H) whatever the size of the matrix and the distribution of species richness.The mean error rates in detection of random assembly (H0) for the uniform, normal and lognormal distribution of species richness were equivalent with 11%, 10.5% and 10.9% of errors, respectively. Error rates in detection of H1H were always equal to zero. Robustness in detection of limiting trait similarity pattern (H1L) was influenced by the distribution of species richness per site (Table 3). *SwapClass* was very robust in detecting H1H when species richness were distributed as a normal or lognormal distribution, except for small matrices ($i=j=10$) for which the error rates varied between 10% and 33%. We observed more contrasted results when species richness was uniformly distributed, in particularly for a small number of classes ($n= 3$) with up to 10% of errors for $i = j = 30$.

Table 3: Error rate in detection of limiting trait similarity pattern (H1L) for several constrained species richness distribution (uniform, normal and lognormal), several numbers of classes (*n*) and several matrix sizes (*i* and *j* represent respectively the number of rows and columns of the matrices). Evenness of classes was low ($\alpha=0.1$, $\beta=1$). Each error rate was estimated on 100 artificial matrices.

	uniform			normal			lognormal		
i, j	n = 3	n = 5	n = 10	n = 3	n = 5	n = 10	n = 3	n = 5	n = 10
10	0.44	0.30	0.41	0.12	0.10	0.22	0.33	0.15	0.24
30	0.10	0.02	0.03	0	0	0	0	0	0
50	0.06	0.02	0	0	0	0	0	0	0
100	0.02	0	0	0	0	0	0	0	0

3.4. Analysis 3 – asymmetry of the matrix

Neither the size of the matrix nor the number of classes significantly affected the robustness of *SwapClass* model to detect random assembly (H0) (Fig. 3). The average error rates for the three numbers of classes (n {3, 5, 10}) were equivalent with 10.5%, 10% and9.2% of errors respectively when the species richness was unconstrained, and with 9.8%, 10.1% and 9.7% of errors respectively when the species richness was constrained by a uniform distribution.
Error rates in detection of H1H were almost equal to zero (Fig. 3). Figure 3 presents detailed error rates in identification of H1L matrices. In case of unconstrained species richness distribution and low evenness distribution of abundance classes, errors in identification occurred only for very small number of rows or columns. Whatever the number of rows or classes, errors occurred frequently if the number of columns was equal to 10 (Fig. 3, top side). In case of constrained uniform distribution of species richness and uniform distribution of classes, error rates were more important. Whatever the numbers of classes, error rates were almost always greater than 5% if the number of rows were lower or equal to 30. Moreover, error rates were also relatively important if the number of columns were higher than number of rows. In that later case, increasing the number of classes seemed to decrease error rates for matrices for which the number of rows were higher than the number of columns (Fig. 3, bottom side).

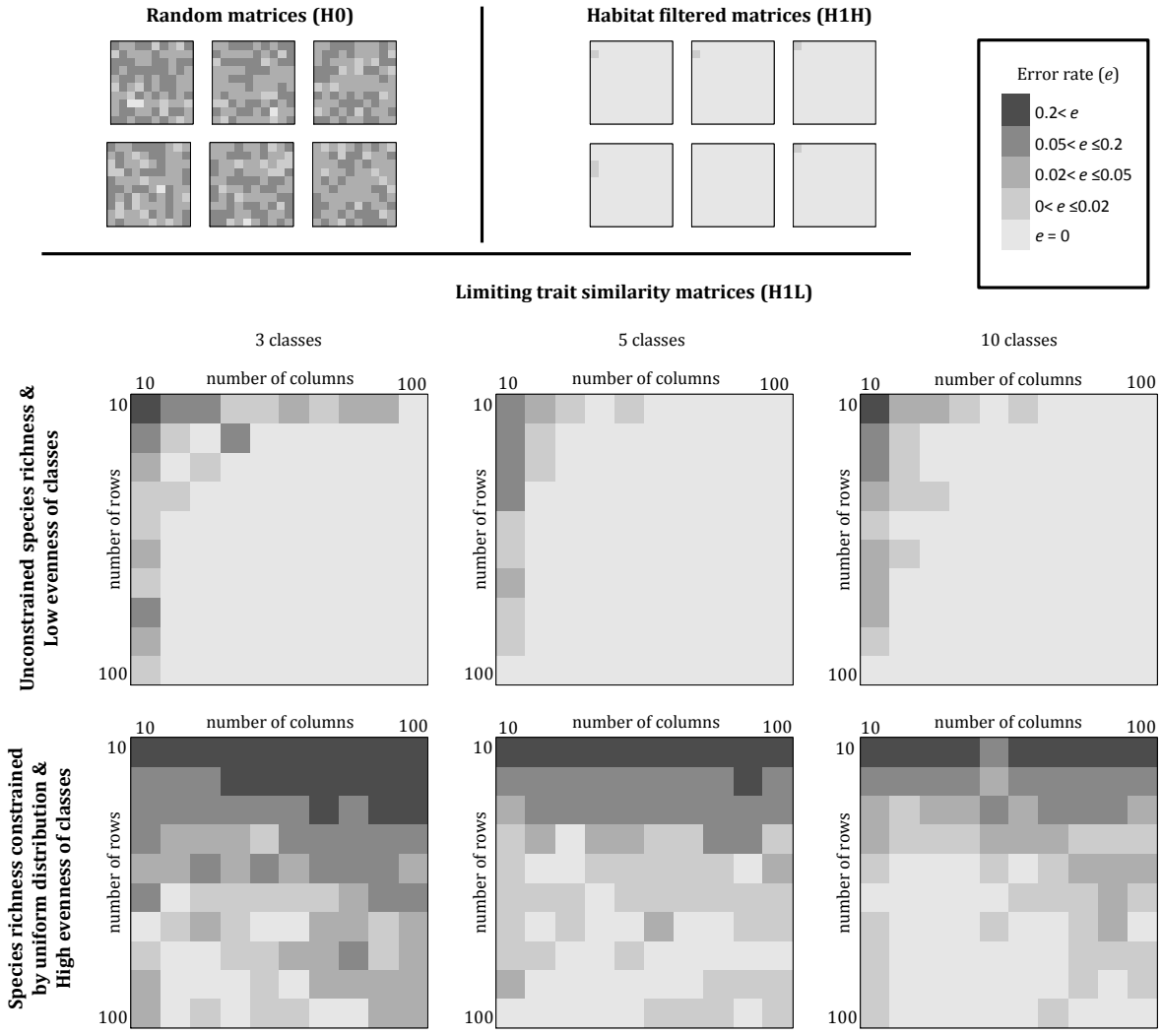


Figure 3: Error rate in detection of assembly patterns (H0: random communities, H1H: habitat filtered communities and H1L: limiting traits similarity communities), for several numbers of rows, columns and classes. The three sub-figures at the top of the figure were performed with i) unconstrained species richness and classes followed a beta distribution of parameters ($\alpha = 0.1$, $\beta = 1$), while the three sub-figures from the bottom of the figures were performed with ii) constrained species richness and abundance classes followed uniform distributions. Each error rate was estimated on 100 artificial matrices.

4. Discussion

In this paper, we present the first null model, the *SwapClass*, which is adapted to semi-quantitative (or qualitative with more than two categories) data and maintains both the distribution of each class per row and column. The evaluation of the model and the *nbPerm* index showed that its applicability domain is quite large suggesting that *SwapClass* is a robust algorithm that can be used in a wide range of ecological conditions. For most of study cases in the trait-based analysis performed here, it was not prone to Type I and II errors. However, errors in detection were observed in particular cases: for very small matrices *i.e.* with a number of rows or columns inferior or equal to 10, for matrices with more columns than rows and for matrices for which the number of allowed permutations per cell as summarized by *nbPerm* is lower than 0.5.

The error in the detection of random patterns (H0), *i.e.* Type I error, occurred at a rate of 9.99% whatever the size of the matrix, the number of abundance classes and the distribution of the abundance. At $P < 0.05$ (in either tail), if the test is robust to Type I error, around 10% of random matrices tested should have been identified as nonrandom (Gotelli, 2000). Hence, the *SwapClass* null model is not prone to Type I error meaning that the null hypothesis won't be rejected for a data set that is random.

More interestingly, our evaluation of *SwapClass* showed that the model was quite robust to detect habitat filtered or limiting similarity patterns. Over the three analyses, 97.45% of non-random patterns (H1H and H1L) were correctly assigned. The error rates were of 11.8% for matrices with less than 500 cells and of 0.2% for matriceswith at least 500 cells. However, the probability of failing to detect non-randompatterns (H1H or H1L) increases when the number of permutations that can be performed per cell *i.e. nbPerm* decreases. Within matrices with less than 500 cells, 97.2% of the identification error occurred for matrices with a value of *nbPerm* lower than 0.5.

In the extreme case of 20 classes in matrix with at least or less than 30 rows and columns, the Type II errors were important for both H1H and H1L matrices suggesting a linkage disequilibrium between cells that cannot be alleviated by the swap method. Nevertheless, this case of very high number of classes is unrealistic for ecological semi-quantitative data and, if so, one can combine adjacent classes to reduce the number of classes. As a recommendation, for high number of classes (>10), we advise to reduce the number of class by joining (rare and) adjacent classes.

Our results also showed that the distribution of abundance classes affected the probability of detecting non-random patterns. We expected that a uniform distribution of the abundance classes would increase the number of permutations that can be performed, but we observed the opposite trend. In fact, the detection of non-random patterns (and especially in H1L matrices) was relatively good when species richness was unconstrained (only the parameters of the Beta distribution determined the species richness of sites) or followed a normal or lognormal distribution, but was poor when species richness classes were uniformly distributed. We observed that the dominance of a class (typically the class 0: absence of

species) seemed to increase the number of allowed permutations.The effects of the evenness of the distribution of abundance classes were not cleared and seemed to vary with the assembly pattern tested (assembly hypothesis, matrix size and number of classes) (Table 2). Increasing evenness in the case of a small number of classes (*i.e.* 3 classes) seemed to weakly increase the error rate in detection of H1L matrices, while increasing evenness in case of high number of classes (*i.e.* 20 classes) decreased the error rate in detection of H1L matrices. Figure S3 given in appendix confirmed the idea that evenness of classes has different effects on the number of allowed permutation per cell (*nbPerm*) which can either be positive or negative depending on the number of classes. *SwapClass* is robust in detecting non-random pattern either when the distribution of abundance classes is uniform or when one or few abundance classes are dominant. This variation in the 'swapability' of any given matrix is summarized in the *nbPerm* index which is essential for indentifying applicability of the *SwapClass* null model on a particular observed matrix. Moreover, we would recommend evaluating the numbers of allowed permutations per cell also when applying the classical swap null model to presence/absence data.

Error rates in detection of limiting trait similarity pattern (H1L) occurred for matrices for which the number of rows was lower or equal to the number of columns. As every statistical test, the null model coupled to a trait-based dispersion analysis has low power at small sample size (Ulrich *et al.*, 2009) and needs more repetitions (sites) than variables (species). This allows to disentangle between the errors due to non-adapted matrix size (number of rows < number of columns) and the ones due to non-swappable patterns (*nbPerm*< 0.5). Nevertheless, the number of rows / columns ratio and the number of allowed permutations per cell (*nbPerm*) were negatively correlated for the three assembly processes (Fig. 2). Hence, a 'perfect' matrix size seemed to be constrained by a trade-off between a high ratio of number of rows / columns and a high number of allowed permutations per cells.

High differences of error rates in detecting non-random pattern between H1H and H1L may be explained by differences in the methodology used here for generating artificial matrices. We used a deterministic procedure to generate the H1H matrices and thus control the intensity of the habitat filtering process. Conversely, we used a partially stochastic procedure to generate the species assembly in the H1L matrices. Therefore, the intensity of limiting similarity process varies between the 100 artificial matrices and one may expect that in the case of a low intensity of the filter, it is more difficult for the model to significantly deviate from a random pattern and, as a corollary for the user, to detect the non-random pattern. Moreover, *nbPerm*, the number of potential permutations per cell significantly varied between the three types of artificial matrices and especially for limiting trait similarity community matrices (H1L). Indeed, in this last case, *nbPerm* was much lower than for the two other types of matrices.Therefore, in the case of the limiting similarity pattern (H1L), the matrices obtained with *SwapClass* were less different than the initial matrix *i.e.* the observed one than for the other two patterns. This would also explain the higher number of errors.

To conclude, we believe that the *SwapClass* null model algorithm is adapted for trait-based analysis i) which can be highly influenced by *e.g.* the number of species or the relative species abundances and ii) for which there was, so far, no adapted null model to common semi-quantitative dataset.

Acknowledgments

Benjamin Borgy is funded through a fellowship ANR-OGM VIGIWEED (ANR-07-POGM-003-01). The authors would like to thank Samuel Soubeyrand and Rémi Perronne for their comments.

References

Ackerly DD, Cornwell WK. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. Ecol Lett 2007;10:135-145.

Barralis G. Méthode d'étude des groupements adventices des cultures annuelles. INRA - 5ème Colloque International sur l'Ecologie et la Biologie des Mauvaises Herbes ; 1976 ; Dijon, France.

Bonham CD, Mergen DE, Montoya S. Cover estimation: a contiguous Daubermire Frame. Rangelands 2004 ;26 :17-22.

Boschilia SM, Oliveira EF, Thomaz SM. Do aquatic macrophytes co-occur randomly? An analysis of null models in a tropical floodplain. Oecologia 2008;156:203-214.

Botta-Dukát Z. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. J Veg Sci 2005;16:533-540.

Cornwell WK, Ackerly DD. Community assembly and shifts in plant trait distributions accross an environmental gradient in coastal Carlifornia. Ecol Monogr 2009;79: 109-126.

Cornwell WK, Schwilk DW, Ackerly DD. A trait-based test for habitat filtering: convex hull volume. Ecology 2006;87:1465-1471.

Daubenmire RF. Canopy coverage method of vegetation analysis. Nothwest Scientist 1959;33:43-64.

De Bello F, Doležal J, Ricotta C, Klimešová J. Plant clonal traits, coexistence and turnover in East Ladakh, Trans-Himalaya. Preslia 2011;83:315-317.

Gotelli NJ. Null model analysis of species co-occurrence patterns. Ecology 2000 ;81: 2606-2621.

Gotelli NJ, Entsminger NJ. Swap and fill algorithms in null model analysis: rethinking the knight's tour. Oecologia 2001 ;129:281-291.

Gotelli NJ, Entsminger NJ. Swap algorithms in null model analysis. Ecology 2003 ;84 :532-535.

Gotelli NJ, Graves GR. Null Models in Ecology. Washington DC :Smithsonian Institution Press; 1996.

Hardy OJ. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. J Ecol 2008 ;96 :914-926.

Jonsson BG. A null model for randomization tests of nestedness in species assemblages. Oecologia 2001 ;127:309-313.

Kraft NJB, Cornwell WK, Webb CO, Ackerly DD. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. Am Nat 2007;170:271-283.

Miklós I, Podani J. Randomization of presence-absence matrices: comments and new algorithms. Ecology 2004 ;85:86-92.

Mouchet M, Villéger S, Mason M, Mouillot D. Functional diversity measures : an overview of their redundancy and their ability to discriminate assembly rules. Funct Ecol 2010;24:867-876.

Patefield WM. Algorithm AS159. An efficient method of generating r x c tables with given row and column totals. App Stat 1981;30:91-97.

Petchey OL, Evans KL, Fishburn IS, Gaston K. Low functional diversity and no redundancy in British avian assemblages. J Anim Ecol 2007;76:977-985.

Stone L, Roberts A. The checkerboard score and species distributions. Oecologia 1990;85:74-79.

Stubbs WJ, Wilson JB. Evidence for limiting similarity in a sand dune community. J Ecol 2004;92: 557-567.

Ulrich W, Almeida-Neto M, Gotelli NJ. A consumer's guide to nestedness analysis. Oikos 2009;118:3-17.

Ulrich W, Gotelli NJ.Null model analysis of species associations using abundance data. Ecology 2010;91:3384-3397.

Ulrich W, Gotelli NJ. Over-reporting bias in null model analysis: A response to Fayle and Manica (2010). Ecol Mod 2011;222:1337-1339.

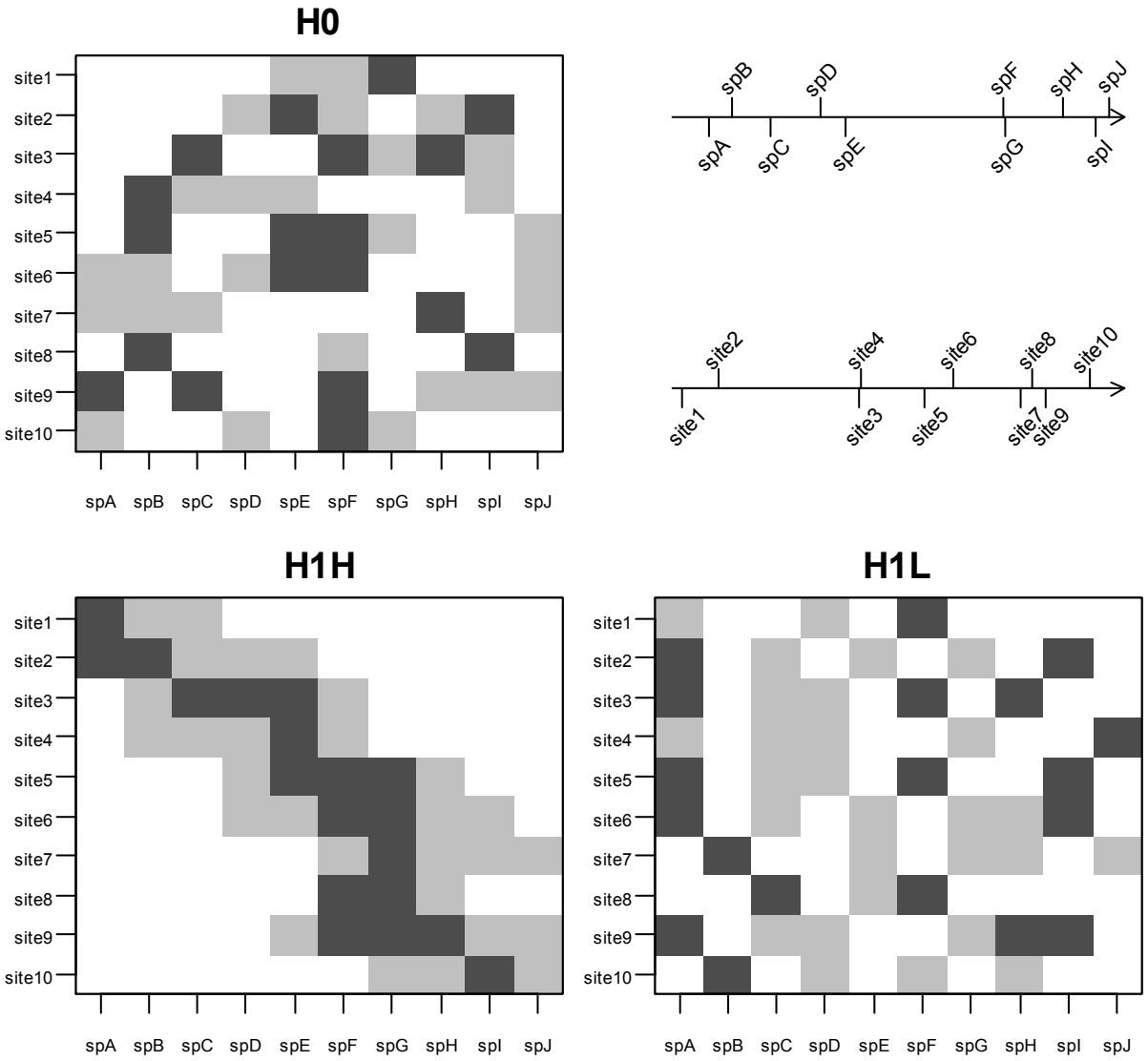
Vázquez DP, Aizen MA. Null model analyses of specialization in plant-pollinator interactions. Ecology 2003;84 : 2493–2501.

Weiher E, Freund D, Bunton T, Stefanski A, Lee T, Bentivenga S. Advances, challenges and a developing synthesis of ecological community assembly theory. Philos T Roy Soc B 2011;366:2403-2413.

Wright DH, Patterson BD, Mikkelson GM, Cutler A, Atmar W. A comparative analysis of nested subset patterns of species composition. Oecologia 1998;113 :1-20.

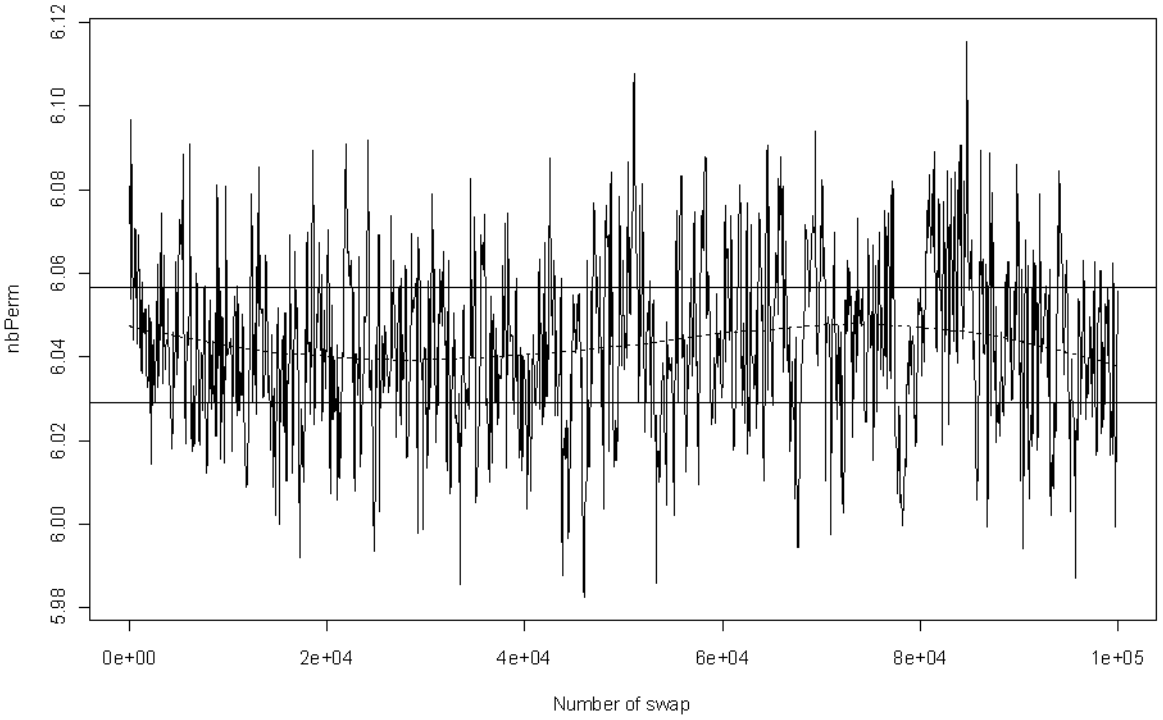
Appendices

Figure S1



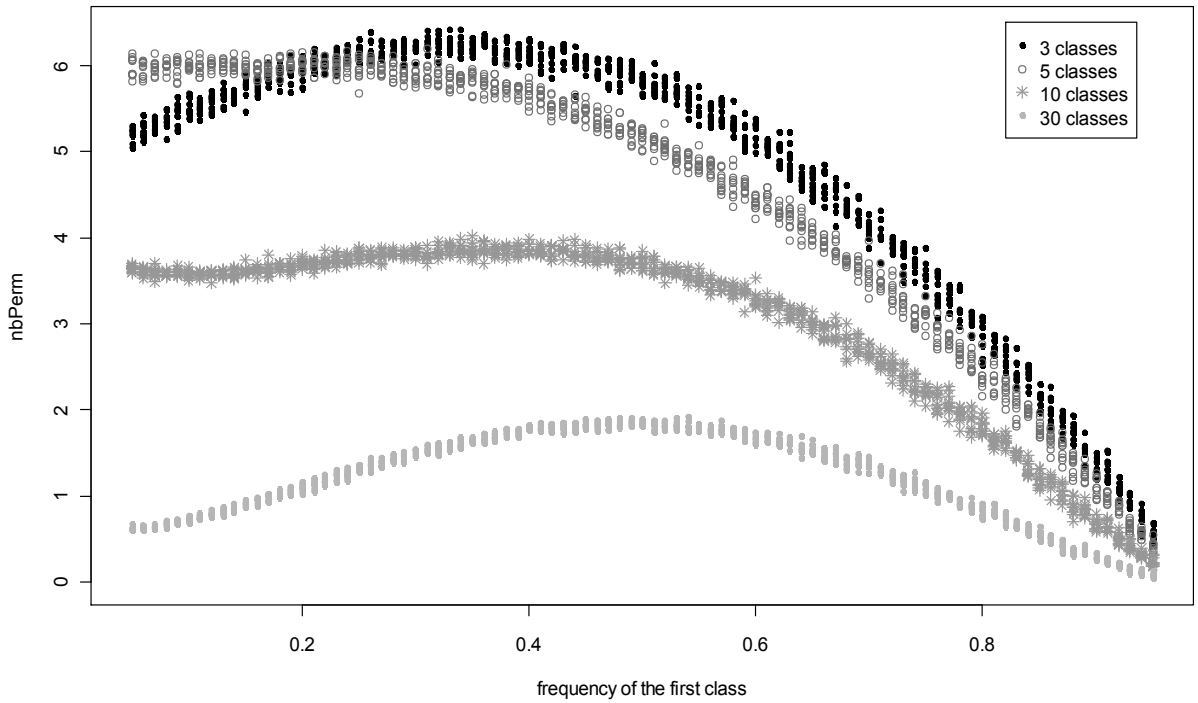
An example of matrices returned by the community assembly models in case 10 species, 10 sites and 3 abundance classes. The three community matrices (H0: random communities, H1H: habitat filtered communities and H1L: limiting traits similarity communities) have for each site (row) an equal number of several abundance classes. The two axes represent trait value of species and optimum trait value of site. The three colors represent the three abundance classes (white : 0 individual/m², grey: 50 individuals/m² and black: 100 individuals/m²).

Figure S2



Change in the number of allowed permutations per cell (*nbPerm*) with the number swaps performed. The matrix (50 rows * 50 columns) is composed by five classes uniformly sampled. The number of swaps varied between 0 and 100 000. 95% of *nbPerm* were in [5.93;6.02]. Hence, *nbPerm* remained roughly constant through swap permutations of the matrix. Dotted line represents non parametric loess regression, horizontal continuous black lines represent the first and third quantiles.

Figure S3



Variation in the number of allowed permutations per cell (*nbPerm*) with the frequency of the first abundance class. 5460 random matrices (50 rows * 50 columns) were generated over the whole analysis. The abundance scale is four points and the frequency of the first class (*p*) varies between 0.05 and 0.95. The frequency of the 3 other classes are uniformly distributed and sum at 1-*p*. In some cases, *nbPerm* was maximum for uniform distribution across all classes (in case of 3 classes, *nbPerm* was maximum for $p=1/3$). In other cases, *nbPerm* was maximum for high dominance of the first class (in case of 30 classes, *nbPerm* was maximum for p=0.5). Hence, effects of dominance of one class on the number of permutation allowed per cell (*nbPerm*) depend on the number of classes.

2.3.1. Notes sur l'importance du modèle SwapClass

Le principal apport de ce travail a donc été de fournir un modèle nul générique mais adapté à un type de notation particulier et assez fréquemment pratiqué dans les inventaires ou les enquêtes. En effet, le modèle peut être appliqué à tout type de données qualitatives composées de plus de deux modalités (ordonnées ou non) et maintient donc le nombre de chaque modalité par ligne et par colonne. De plus, nous suspectons fortement qu'il soit plus adapté que les autres méthodes utilisés pour les analyses réalisées sur la dispersion des traits. En effet, dans certaines études, afin de palier au problème de randomisation de données d'abondance, les randomisations du modèle nul permutent les traits (ou ensembles de traits) entre les espèces (exemple : Stubbs & Wilson, 2004) ou permutent les abondances des espèces par la méthode « *abuswap* » (exemple : Pakeman, 2011b). Sans vouloir faire de jeu de mot, nous pensons suite à notre étude que cela doit recouvrir un grand nombre de situations d'usage abusif ou le choix du modèle nul conduirait ainsi à modifier les dominances dans les distributions de traits sans justification. Il reste toutefois important maintenant d'évaluer les biais potentiels de notre méthode suivant la taille du jeu de données, le nombre de modalités et l'hétérogénéité dans les proportions de chaque modalité.

3. REPONSE DES ESPECES ADVENTICES A L'ENVIRONNEMENT

Les travaux visant à analyser la réponse des espèces aux filtres 'environnement' et 'pratiques culturales' ont été abordés par deux approches. Une première approche 'spatiale' (en termes de répartition entre les parcelles cultivées), exploite le concept de niche écologique des espèces adventices par l'intermédiaire d'une méthode d'analyse factorielle. La deuxième approche 'temporelle' exploite quant à elle le concept de traits d'histoire de vie pour décrire la dynamique de différentes espèces au cours des successions culturales. Ce deuxième travail, qui a donné lieu à la rédaction de l'article présenté dans la section 3.2.2, utilise les Modèles de Markov Caché pour modéliser le stock de graines qui nous est inconnu.

3.1. Délimitation d'une niche écologique à partir de relevés de flore levée

Le concept de « niche écologique » des espèces traduit (i) la position relative qu'occupe une espèce au sein d'un environnement ou écosystème, mais aussi (ii) la somme des conditions environnementales nécessaires et relations biotiques déterminant « l'enveloppe écologique » d'une espèce (Hirzel & Le Lay, 2008; Schoener, 1974). On peut identifier deux types théoriques de niche : la niche fondamentale qui représente l'espace environnemental (pédoclimatique) potentiel pouvant être occupé par l'espèce et la niche réalisée qui représente l'espace environnemental réellement occupé par l'espèce en présence de contraintes biotiques (compétition, facilitation,…) (Keddy, 1983). Evaluer la niche écologique des espèces à travers l'utilisation de jeux de données issues d'échantillonnage dans le milieu naturel et non en conditions contrôlées revient ainsi à reconstruire la niche réalisée des espèces. Une simplification des processus est donc généralement réalisée puisqu'il n'est pas rare d'analyser les liens espèces – environnement en négligeant les relations interspécifiques. On ne sait pas si cette simplification s'avère ou non audacieuse dans les conclusions qui seront tirées.

D'un point de vue conceptuel, la niche écologique est définie comme une fonction reliant la *fitness* des espèces à l'environnement (Hutchinson, 1957). Les modèles modélisant la niche (Habitat Suitability Models : HSM) permettent ainsi une meilleure prédiction par l'utilisation de gradient de plus grande cohérence dans la répartition des espèces (Hirzel & Le Lay, 2008) (Figure 7).

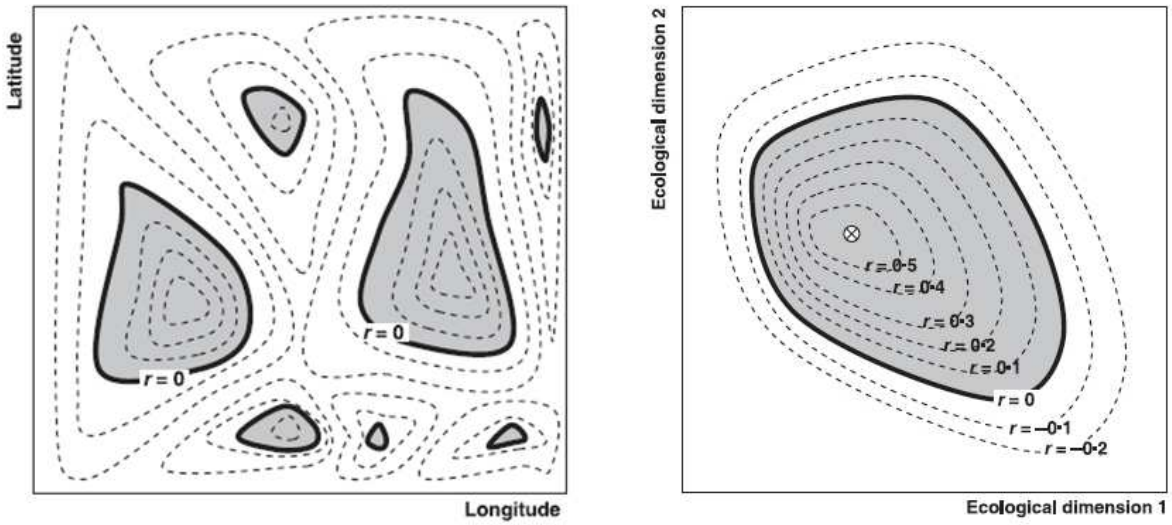


Figure 7 Représentation de la niche écologique d'une espèce dans l'espace géographique (figure de gauche) et dans l'espace écologique créé par deux variables environnementales (figure de droite). Le taux de croissance de l'espèce est représenté par les isoclines (*r*). D'après Hirzel & Le Lay, 2008. On peut distinguer des zones sources à bilan populationnel positif (r>0) et des zones puits à bilan négatif (r<0) qui ne se maintiennent que s'ils sont réalimentés.

L'espace ainsi créé à travers les conditions pédoclimatiques a ainsi un pouvoir plus explicatif. Pour le cas des espèces végétales adventices, on peut alors considérer les pratiques culturales réalisées comme faisant partie intégrante des variables environnementales. Bien que représentant une forte interaction biotique avec la communauté adventice, la culture mise en place reste un élément externe à la constitution de la communauté. Elle peut être considérée comme une variable environnementale de la parcelle au même titre que toutes autres variables pédoclimatiques et culturales. Elle résume à elle seule une forme de couvert végétal, une période culturale, un choix qualitatif et quantitatif des engrais, une gamme potentielle d'herbicides, une hauteur de coupe à la moisson, etc.…

3.1.1. Centre Abondant

La théorie du « centre abondant » qui veut que les espèces soient plus abondantes au centre de leur niche, n'est rien de plus qu'une extension du concept de niche écologique modélisant la *fitness* des espèces sur les différents gradients environnementaux (Brown, 1984) et suggérant une distribution unimodale mais pas forcément gaussienne ou symétrique des espèces le long de ces gradients (Whittaker, 1956). Une analyse récente de la littérature existante réalisée par Sagarin & Gaines (2002) montre cependant que seul un peu plus d'un tiers des études confirme cette règle. Toutefois, dans de nombreux cas, les études sont biaisées par un échantillonnage inégal, plus faible en bordure de l'aire de répartition.

Vérifier le centre abondant des niches des espèces adventices permet d'évaluer en quoi les concepts et modélisation des niches écologiques peuvent permettre la prédiction de

l'abondance des espèces adventices. Une telle prédiction constituerait une avancée significative, puisqu'au-delà de prévoir la présence des espèces, elle permettrait de cerner celles qui vont potentiellement poser des problèmes d'infestation. De par la nature du réseau Biovigilance Flore, dont l'échantillonnage se veut représentatif des cultures (et non des flores observées), on peut supposer que le biais dû à un échantillonnage plus faible en bordure de niche sera moindre. De plus, la succession culturale modifiant fréquemment les conditions du milieu et donc la position de l'espèce au sein de sa niche et l'incapacité d'une plante à se déplacer va augmenter la probabilité d'échantillonnage en bordure de niche.
De plus, le réseau biovigilance fait une notation semi-quantitative d'abondance. Les conditions semblent donc réunies pour que l'on mette à l'épreuve des faits cette théorie sur notre jeu de données.

3.1.2. L'approche par plan factoriel

Les plans factoriels permettent la constitution d'un espace environnemental prenant en compte toutes les variables mises en entrée d'analyse quel que soit leur type. Un grand nombre de méthodes a été développé pour réaliser ces plans (Chessel *et al.*, 2004 ; Dray *et al*., 2007). La première difficulté réside toutefois dans le choix de constitution judicieuse de l'espace de référence pour des espèces qui n'expriment pas les mêmes besoins ou avec des variables à disposition qui ne peuvent, dans le meilleur des cas, que rendre compte de manière indirecte de la couverture des besoins biologiques. On se retrouve alors à traiter des gradients et variables pouvant être de nature très différente, La difficulté se pose également dans le choix de la méthode et des variables à utiliser en entrée d'analyse. Si l'on suit les recommandations de Robertson *et al*. (2001) dans le cas d'un grand nombre de variables, une solution simple peut consister à bâtir des axes synthétiques d'ensembles de variables de même « nature » puis à modéliser ensuite l'espace global en travaillant sur ces nouveaux axes factoriels.

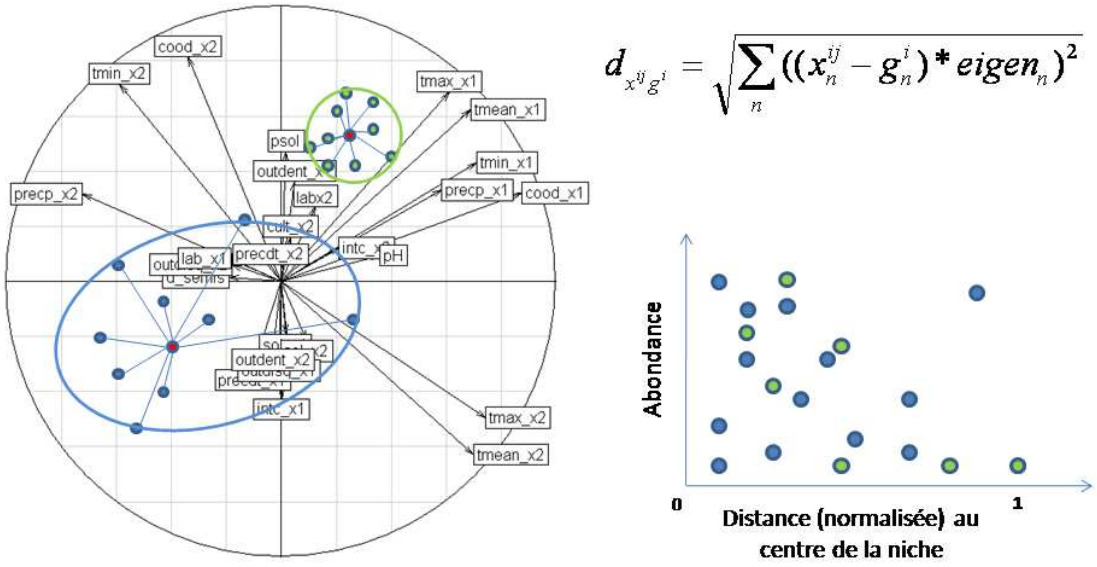


Figure 8 Représentation de l'espace environnementale et centre de gravité des niches. L'espace est basé sur les conditions pédoclimatiques et culturales. Les occurrences des espèces permettent de calculer leur centre de gravité (centre de niche) (figure de gauche). On distingue les espèces généralistes (ellipse bleue) des spécialistes (ellipse verte), et degré d'excentricité (distance au centre de l'espace). Pour chaque espèce *i*, les distances ($d_{x^{ij}g^i}$) des points *j* au centre de la niche (*g*) sont normalisées (divisions par la valeur distance maximal de l'espèce) et projetées sur une même figure en fonction de l'abondance observée (figure de droite).

L'espace ainsi créé est uniquement basé sur les variables environnementales et culturales. On peut alors projeter en variable supplémentaire l'abondance de nos espèces dans cet espace et vérifier les relations entre distance au centre de gravité de l'espèce et abondance (figure 8).

Une analyse réalisée sur les données 'Biovigilance Flore' et des données climatiques provenant du modèle WorldClim (Hijmans *et al.*, 2005) a montré qu'il existait bien une relation négative entre abondance des espèces et distance au centre de leur niche. Le tableau X3 de l'annexe (page 161) présente de manière synthétique les variables utilisées dans chacun des espaces factoriels créés par analyse en composantes principales (ACP) conduisant à la création d'axes intermédiaires synthétiques. Les variables qualitatives ont été transformées en matrices binaires suivant la méthode de Hill & Smith (1976). On réalise ensuite une ACP sur les 27 axes synthétiques créés. La connaissance du centre de gravité des occurrences de chaque espèce nous permet de calculer une distance au centre de la niche pour chaque couple « site, espèce » (distances euclidiennes sur tous les axes pondérées par les valeurs propres de chaque axe). Pour chaque espèce, on normalise les distances en les divisant par la distance maximale. La figure 9 présente les distributions et probabilités prédites par modèles linéaires des 6 classes d'abondances le long de la distance (normalisée) au centre de la niche, pour les 58 espèces les plus présentes dans notre jeu de données.

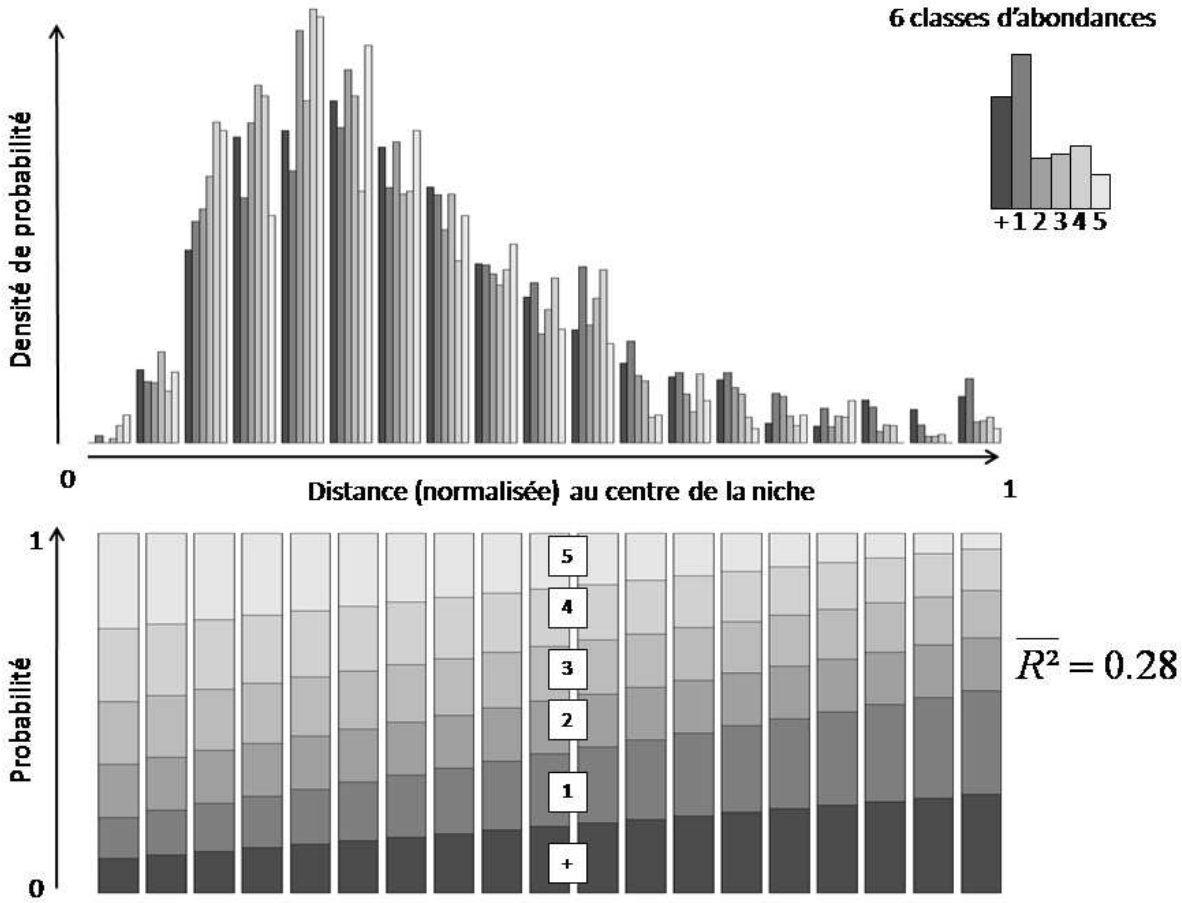


Figure 9 Distribution des six classes d'abondances le long de la distance (normalisée) au centre de la niche. La figure du haut représente la densité de probabilité de chaque classe (d'abondance croissante de + à 5), alors que la figure de droite représente les probabilités associées à chaque classe en fonction de la distance au centre, modélisées par modèle linéaire. Les modèles réalisés sur les classes + à 5 ont respectivement des R^2 égalent à 0.31, 0.80,-0.05, 0.07, 0.21 et 0.37 (moyenne = 0.28). Ainsi, on trouve une relation entre l'abondance et la distance au centre de la niche. Si elle colle a notre attendu elle n'en demeure pas moins peu forte.

La probabilité de trouver les fortes classes d'abondance (4 et 5) augmente avec la diminution de la distance au centre de la niche ; inversement la probabilité de trouver les faibles classes d'abondance (+ et 1) augmente avec la distance au centre de la niche. Bien que significative, les pentes sont peu importantes et les dispersions autour des modèles demeurent assez importantes (*cf.* R^2 moyen). Ainsi le pouvoir de nous renseigner sur l'abondance des espèces adventices en fonction des conditions du milieu et des pratiques culturales est assez faiblement prédictif même si la théorie du centre abondant semble être validée, pour nos données et suivant la méthode utilisée. Néanmoins, cette approche pourrait permettre d'appréhender la trajectoire des communautés et d'identifier les changements de pratiques culturales qui pourraient conduire à une réduction de l'abondance d'une (ou d'un ensemble d') espèce(s) problématique(s) en cherchant des conditions éloignées du centre de la niche. Une telle prévision devrait alors être mise à l'épreuve de faits expérimentaux.

3.1.3. Difficultés rencontrées et limites

Plusieurs problèmes et difficultés peuvent expliquer le manque global d'efficacité de la modélisation de la niche pour prédire l'abondance des espèces adventices et pourquoi la réflexion autour de ce travail est toujours en cours.

Manques de variables et problèmes de « faux zéros »

La modélisation de la niche écologique des espèces nécessite de décrire le mieux possible l'espace environnemental des espèces. Certaines variables identifiées comme essentielles dans l'assemblage des communautés étaient peu renseignées par le réseau biovigilance. Cela a conduit à une forte réduction du jeu de données de départ. Par exemple, alors que l'on connaît son importance (Gunton *et al.,* 2011*),* on ne dispose pas de la date de semis, sans compter qu'il faudrait maîtriser la correction à apporter selon la latitude et la longitude du point considéré. De plus, certaines variables dont l'absence de notation n'indique pas forcement une absence de la pratique (oublie de notation = faux zéros) et pour laquelle il est impossible de différencier une absence de mesure d'une absence de pratiques, viennent fortement bruiter la construction de l'espace, induisant une erreur dans les positions relatives des points.

Une méthode pas forcement adaptée

La méthode utilisée (Analyse en Composante Principales) n'est pas forcement adaptée. Tout d'abord, la construction des axes est basée sur les corrélations linéaires entre variables. Les relations entre variables, bien que supposées monotones, ne sont pas nécessairement linéaires et cela conduit à la construction d'un espace qui peut avoir peu de sens. De plus, la réduction de l'espace à quelques axes majeurs (1 ou 2) créés à partir d'un tableau de données binaires et complètement exclusif entre les colonnes (exemple : variable culture qualitative à 15 modalités transformée en un tableau binaire de 15 colonnes) revient à ne sélectionner que les premières modalités majoritaires de la variable d'origine. Cette méthode simplifiant le nombre d'axes est donc peu adaptée à notre cas. De plus, les méthodes par plan factoriel ne pouvant gérer des données manquantes autrement qu'en les remplaçant par des valeurs moyennes, la réduction du jeu de données aux sites où l'information semble de meilleure qualité peut introduire un biais important en surreprésentant ces conditions.

Une absence de prise en compte du stock de graines

C'est sans doute la critique la plus importante : l'absence de connaissance sur le stock de graines empêche d'appréhender la part du stock ayant levée suivant la position dans la niche de l'espèce. Ceci peut s'avérer problématique puisqu'il devient alors impossible de différencier un site où l'espèce est absente mais où les conditions lui sont favorables, d'un site où l'espèce est présente mais où les conditions lui sont temporairement (dû à la rotation) défavorables. De même en observant 30 levées / m² comment savoir si cela représente 80% du stock potentiel local ou seulement 1%? On tente d'appréhender l'état du stock de graines en renseignant le précédent cultural, mais cette approche conduit à une forte simplification de l'historique à une seule année et ne reflète pas nécessairement l'état du stock dont la survie peut largement excéder une année ni les variations entre espèces à dormance réduite ou

longue. Ce point pose donc un réel problème quant à l'application de telles méthodes sur ce type de jeux de données. Une réflexion sur le moyen d'y palier serait judicieuse si l'on souhaite mieux modéliser et identifier la niche écologique des espèces adventices par l'intermédiaire de jeux de données de ce type (échantillonnés en conditions naturelles). Par exemple, considérer l'ensemble de la flore ayant levée sur les *n* précédentes années permettrait d'évaluer (en termes de présence/absence) la composition en espèces potentielles du stock de graines. Cependant, une évaluation quantitative de l'état du stock par cette approche paraît peu réaliste.

Conclusion

Ainsi, modéliser la niche écologique des espèces s'avère être un réel « challenge » et nécessite de porter une attention particulière aux jeux de données, aux méthodes utilisées, ainsi qu'aux présupposés sur les besoins exprimés par les différentes espèces. Suite aux problèmes d'échantillonnage, aux forts biais introduis par une réduction de nos données (sélection des données les plus complètes uniquement) et à la faiblesse de significativités dans les résultats trouvés, il a été décidé de renvoyer à plus tard la valorisation de ce travail. Néanmoins, il reste certainement important de persister dans l'application du concept de niche écologique au cas des adventices. La spécificité du système, qui nous permet sûrement de disposer d'un échantillonnage en bordure de niche des espèces plus que pour tout autre communauté « naturelle », peut justifier à elle seule l'utilisation de ce concept afin de prédire les trajectoires des espèces adventices au sein du système cultural. De plus, d'un point de vue conceptuel, il peut être important de considérer le problème tel qu'il est en réalité : en effet, l'ensemble des pratiques étant, pour la plupart, mises en place pour empêcher une communauté adventice abondante, le lien à faire est peut-être plus dans l'établissement d'une relation entre environnement (incluant les pratiques) et espèces absentes qu'entre environnement et espèces présentes.

3.2. Estimation des traits d'histoire de vie à partir des dynamiques temporelles de flore levée

3.2.1. L'approche Modèle de Markov Caché

Un Modèle de Markov Caché (Hidden Markov Model : HMM) est un modèle statistique dans lequel le système modélisé est supposé être un processus markovien de paramètres inconnus. Même si cela n'est finalement pas compliqué, cela nécessite quelques explications.

3.2.1.1. Transition Markovienne, Chaîne et Modèle de Markov

Un processus markovien est un processus stochastique, à temps discret (et voire à espace discret) possédant la propriété de Markov : la prédiction du futur, sachant le présent, n'est pas rendue plus précise par des éléments d'information supplémentaires concernant le passé. Dans le cas d'une variable prenant plusieurs états discrets (et aléatoires) au cours du temps, les probabilités de transitions entre les différents états de la variable sont alors dites markoviennes si et seulement si les différentes probabilités ne dépendent que de l'état en cours de la variable et de l'environnement (Meyn & Tweedie, 1993). L'ensemble des probabilités déterminant les transitions entre les différents états de la variable est alors appelé Chaîne de Markov. Cette chaîne peut être définie pour un nombre fini de conditions (ou actions). Le modèle est alors communément appelé Processus de Markov (ou modèle de Markov) (Figure 10).

L'approche par modèle de Markov permet un grand nombre de développements mathématiques (étude des classes d'âge par exemple) et peut être utilisée aussi bien dans des problèmes d'inférence de paramètres que dans des problèmes de conception et d'optimisation de stratégie de gestion (Peyrard *et al.*, 2007).

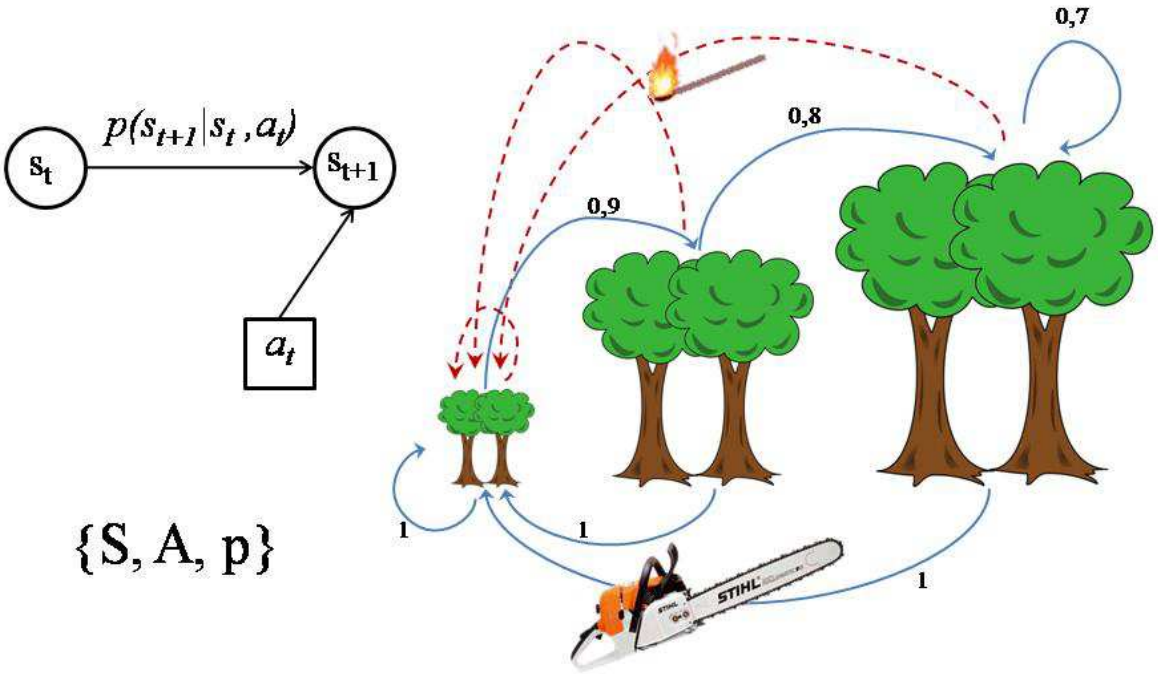


Figure 10 Exemple de modèle de Markov. L'état (s_t) d'une forêt peut être représenté sous la forme de trois classes d'âge. Une probabilité de bruler (ou de passer à la classe suivant) est associée à chaque classe et contrôle la dynamique en cas d'absence de coupe (a_t = rien). Lorsque l'on coupe la forêt (a_t = coupe), quelque soit son état, elle revient à la première classe. Dans un espace de temps discret, la probabilité de transition est donc définie par $p(s_{t+1}|s_t,a_t)$. Le système est défini par un ensemble d'états (S), un ensemble d'actions (A) et les probabilités associées à chaque transition (p).

3.2.1.2. Modèle de Markov Caché

Un Modèle de Markov Caché est un modèle de Markov composé de deux types de variables et probabilités de natures différentes. La variable d'état n'est pas directement observée, mais on observe une variable 'observable' qui ne fournit qu'une information partielle de la variable d'état. Le modèle est alors conduit à la fois par les probabilités de transition entre les états de la variable d'état et les probabilités d'observation qui déterminent les probabilités d'observer les différents états de la variable observable suivant les états de la variable d'état et l'action réalisée. La variable d'état est ainsi appelée variable cachée (Rabiner, 1989) (figure 11).

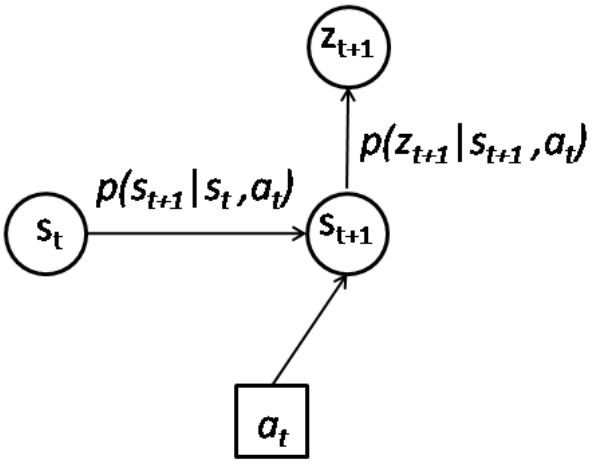


Figure 11 Représentation schématique d'un Modèle de Markov Caché (HMM). Le système est alors composé d'une variable de plus qui représente l'observation (z_t). La variable d'état (s_t) est cachée. Le système est alors défini par un deuxième type de probabilité dirigeant les observations à *t+1* suivant l'état de la variable cachée à *t+1* et l'action réalisée (a_t).

Ce type de modèle intégrant une notion d'information partiellement observable peut être très utile dans les cas où l'observation que l'on fait ne reflète pas nécessairement le véritable état de la variable étudiée. Ainsi, les modèles de Markov caché ont été utilisés dans de nombreux domaines variés où les signaux observés peuvent être bruités (par exemple en reconnaissance vocale où l'on cherche la séquence de mots la plus probable en fonction du signal vocal enregistré [Gales & Young, 2007], ou encore en écologie de la conservation, où l'on cherche à évaluer *a priori* la stratégie de gestion optimale d'une espèce cryptique [Chadès, 2008]).

3.2.1.3. Pourquoi les HMM sont-ils adaptés au cas adventices ?

En observant le cycle biologique des espèces annuelles, il apparaît évident que la flore levée d'une année n'est pas directement informative sur la flore potentielle de l'année suivante. Ainsi la modélisation du stock de graines (qui nous apparaît comme étant une variable cachée) est alors essentielle pour approcher la dynamique temporelle d'une espèce, et de sa flore levée, en un site (figure 12). De plus, le format markovien est totalement applicable au sujet d'étude : même si en apparence certains processus peuvent être fort dépendants de l'historique (i.e. la dormance des graines), l'évolution « instantanée » de l'état (qualitatif ou quantitatif) du stock dépend uniquement des conditions en cours (ainsi l'état de dormance du stock pourrait être considéré comme une variable à part entière dont l'évolution est markovienne).

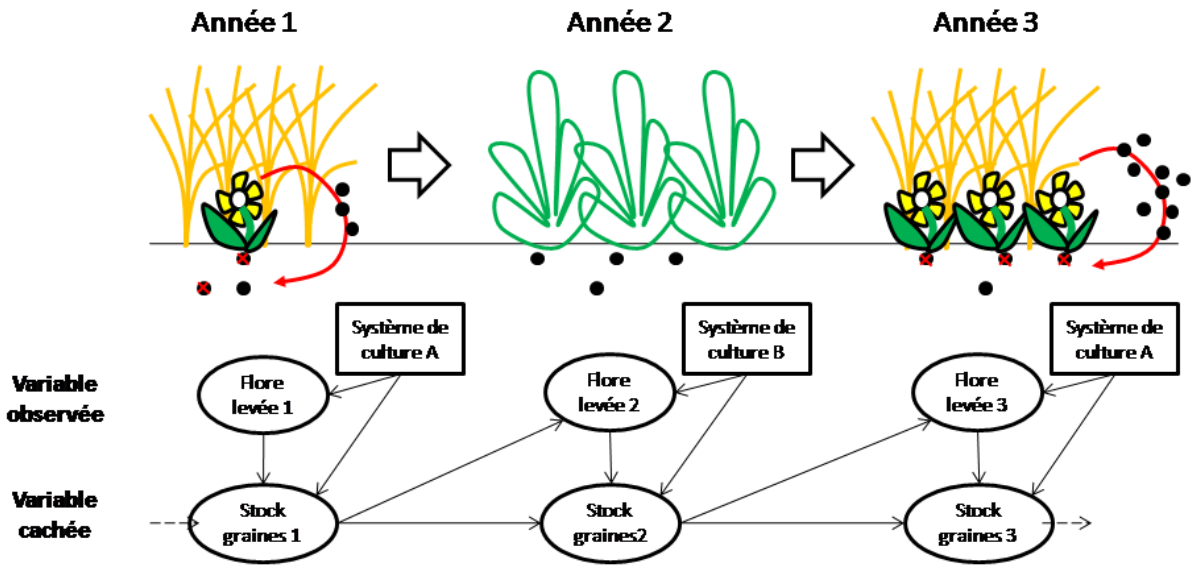


Figure 12 Dynamique simplifiée du stock de graines (cachée) et de la flore levée (observation) au fil des successions culturales. La rotation culturale est composée de deux types (A et B). La flore levée dépend du stock de graines issues des successions précédentes et du système de culture installé, le stock de graine dépend quant à lui de son état l'année précédente et de comment il a été épuisé (mort + germination) et recomposé (production de graines) en fonction du système de culture installé.

De par la rotation culturale et si on néglige l'effet climatique, le système évolue au sein de répétitions successives ou alternées de conditions similaires du milieu (système de culture), la dynamique globale peut ainsi se résumer à la dynamique réalisée au sein de chacun des systèmes de culture étudiés, simplifiant énormément la complexité du modèle.

Pour finir, l'approche probabiliste des modèles de Markov permet d'étudier uniquement certaines pratiques au sein de systèmes variés où beaucoup d'autres variables (des pratiques mais aussi des conditions pédoclimatiques) ne seraient pas renseignées au sein du modèle, avec pour seul effet de réduire les contrastes entre les probabilités des différentes actions (opérations culturales) étudiées.

3.2.2. *Estimation des traits d'histoire de vie intervenant dans la dynamique du stock de graines d'adventices à partir de séries temporelles de flore levée et par l'utilisation d'un Modèle de Markov Caché*

L'échantillonnage du stock graines au sein des parcelles cultivées est rarement réalisé de par ses difficultés : long, couteux et fastidieux. Pourtant, au sein d'un système où la flore est « remise à zéro » annuellement par les pratiques culturales (i.e. labour), il semble évident qu'il joue un rôle de poids dans l'abondance des adventices levées : dans les meilleures conditions, les levées ne pourront excéder le stock présent. Les levées sont donc contraintes par une enveloppe délimitée par le stock de graines. Dans un environnement changeant régulièrement et de façon répétitive (rotation culturale), il peut être intéressant de pouvoir appréhender l'effet des différentes pratiques sur la dynamique du stock de graines. En disposant de séries temporelles de flore levée et par l'utilisation d'un modèle basé sur le cycle biologique des plantes annuelles, on peut alors résumer l'effet des pratiques, par l'intermédiaire de traits d'histoire de vie, sur la dynamique de stock de graines et de la flore levée.

3.2.2.1. Article

Cette section est présentée sous forme d'article (version en préparation, en vue d'une soumission à la revue *Journal of Applied Ecology*).

ESTIMATION OF LIFE HISTORY TRAITS RELATED TO WEED SEEDBANK DYNAMICS FROM EMERGED FLORA TIME SERIES, USING A HIDDEN MARKOV MODEL

Benjamin Borgy[1], Sabrina Gaba[1], Nathalie Peyrard[2], Régis Sabbadin[2] & Xavier Reboud[1]*

[1] INRA, UMR 1210, Biologie et Gestion des Adventices, 17 rue Sully, 21000 Dijon, France.
[2] INRA, UR 875 Unité de Biométrie et Intelligence Artificielle, BP 27 F-31326 Castanet-Tolosan, France.

**Corresponding author :* Xavier Reboud, UMR Biologie et Gestion des Adventices, Institut National de la Recherche Agronomique, BP 86510, 21065 Dijon Cedex, France. Tel: (+33) 3 80 69 31 84; Fax: (+33) 3 80 69 32 62; E-mail: reboud@dijon.inra.fr

Summary

1. Infestation by weeds in cropping areas is an important threat to crop production. The development of sustainable cropping systems less dependent on herbicides needs to enhance knowledge on the respective efficiency of the current management actions. However, the inability to observe the weed community perfectly due to the persistent seed bank makes it difficult.

2. Using a Hidden Markov Model (HMM), we estimate from 329 arable fields three important life history traits (LHT) i.e. seed survival, seed germination and seed production, for 18 common weed species. We then test the potential efficiency of the varying management actions on weed species dynamics using a case study.

3. We found that (positive) relative growth rates were associated with combinations of life history traits. This would suggest an existence of several strategies to cope with natural and agronomical disturbances. We also found a positive correlation between germination rate estimates and an indicator value for specialization to a crop type.

4. Using our estimates, we showed that changing the proportion of the different crops in the crop sequence impacted the relative growth rate of species, with some being favoured at the expense of the others. This effect was not necessarily proportional suggesting differences in species response to management changes.

5. *Synthesis and applications*. A major challenge for designing sustainable weed management is to better understand the response of weed species to each management action of cropping systems, in particularly in the seed bank. Using a Hidden Markov Model, we estimated three major life history traits associated simply describing the seed bank dynamic and showed that weed life history strategies varied between and within crop types. We also showed that a change in the crop sequence can significantly modify the structure of weed community. This indicates that knowledge on LHT values per crop type could be a useful tool to design management policy and sustainable cropping systems.

Keywords: seedbank, hidden markov chain, population dynamics, weed management, life-history traits, survey

Introduction

Infestation by weeds in cropping areas is an important threat for crop production and management rules have been shaped to prevent their development within crop fields and the production of seeds that would provide new individual pests the next cropping seasons. In intensive agriculture, these management rules mainly rely on herbicide uses. On the other hand, the role of arable weeds for agroecosystem functioning has also been recognized (Franke et al., 2009). Weeds are food sources for animals such as pollinators that stabilize reproductive success of plant species irrespective of the population size (Backman and Tiainen, 2002;Gibbons et al., 2006) or granivorous and omnivorous arthropods such as carabid beetles (Hawes et al., 2003;Bohan et al., 2011). Thus, the development and the integration of several combined weed control measures such as integrated pest management (IPM, (Munier-Jolain et al., 2009)), have to face a set of challenging issues: they need to assure food security while reducing chemical inputs, have low cost and low environmental impacts and when possible to use biodiversity to support ecosystem services. However, deciding the best course of actions for managing weeds can be extremely difficult due to the plurispecific nature of weed community and to the complex interactions of factors such as the ecology of weed species, their heterogeneous response to management actions and the dynamic of crop sequences. The decision process is exacerbated further by our inability to observe the (potential) weed community perfectly. First, most weed seeds survive for several years in the soil and form persistent seed banks (Burnside et al., 1996;Murdoch and Ellis, 2000;Conn et al., 2006). Second, little is known about the structure and the composition of the persistent seed bank due to the tedious, expensive and time consuming work needed for counting and identifying seeds at the species level in a large number of cropping systems. Finally, a widespread estimation of weed abundance is given by semi-quantitative measure close to the farmers' perception of the weed infestation (Barralis, 1976). All together, the reasons make it difficult to assess the respective efficiency of the current management actions as well as their positive/negative interaction on the dynamics of weeds.

Life-history traits such as fecundity, seed germination, seedling survival and seed bank persistence have been investigated by weed scientists either for the understanding of ecological community structure and the coexistence of multiple competitors (Bonsall et al., 2002;Bonsall et al., 2004) or developing guidelines for the conservation of threatened species (Evans et al., 2003;Brusa et al., 2007;Moza and Bhatnagar, 2007). In crop field, the emergence stages of the weed life-cycle have been identified as being the most important in determining year-to-year changes in population numbers or in weed species occurrence (Forcella et al., 2000;Colbach et al., 2010;Gunton et al., 2011;Fried and Gaba, submitted). Weed emergence is highly dependent on the effects of cultivation techniques (e.g. tillage) in interaction with concomitant environmental variables (e.g. temperature, moisture and soil structure). Using model sensitivity analysis, Colbach et al. (2010) confirm that life-history traits related to emergence and reproduction are key parameters in the dynamics of weeds.

Moreover, the majority of weed species are therophyte species (Sutcliffe and Kay, 2000;Sutherland, 2004) which survive unfavourable season in the form of seeds, and eventually complete their life-cycle during favourable season. The dormant seeds of many species can survive for years or decades and thus delay the timing of germination until more favourable conditions occur. Seed bank can contain several hundred or thousands seeds per square meter. Values as high as 50 000 seeds are sometimes mentionned (REF), so many weed scientists share the idea that weeds are first of all seeds that sometimes give an adult visible plant. Beyond this exaggerated view, it remains clear that managing weeds require accounting for what happens underground as it is the major compartment for these therophyte species. So far, very few models have made an attempt to predict the temporal dynamic of the seed bank (e.g. Venable, 1989).

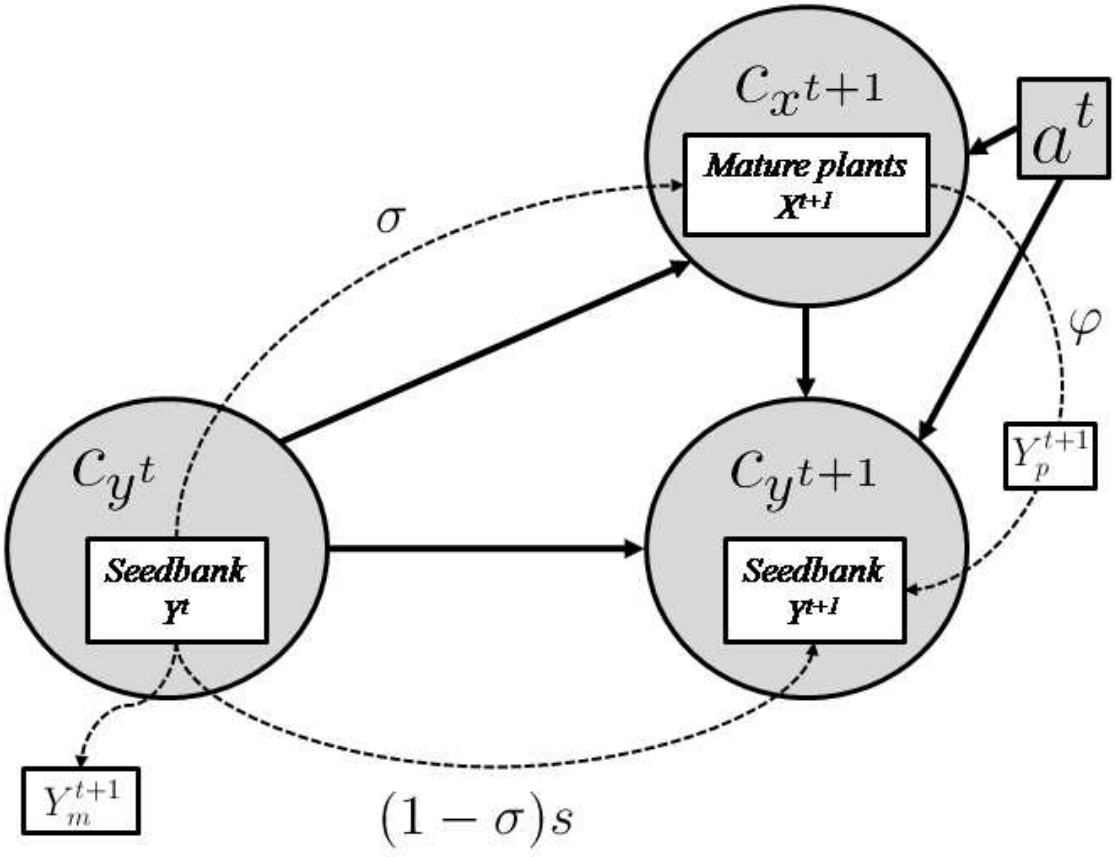
The aim of this paper is to estimate from field data three life history traits i.e. seed survival, seed germination and seed production, for at least the most frequent weed species. These estimates can be derived from time series of emerged flora data for several cultural practices. Their fitting on real situations may enable to enhance the knowledge on the reality of an interaction between weed species and management actions. To do so, weed seed dynamic will be modeled as a Markov chain describing the transitions among life cycle stages (i.e. states). With this modeling approach, we assume that knowledge of the current state is sufficient for explaining future states. In order to take into account the inability to observe the weed community perfectly, we used a special case of Markov chain model i.e. Hidden Markov Model (HMM) which enables us to deal with missing (or hidden) data such as the seed bank. In this paper, we develop a HMM on count and classes of abundance data. We estimate seed survival in the seed bank, seed germination and seed production using a large dataset covering a wide range of agricultural and environmental conditions. We then assess the potential efficiency of varying a management action on weed species in a case study where we decrease wheat occurrence by ten percent in the crop sequence. In contrast to other optimal trait measurement approach (BiolFlor (Kühn et al., 2004), LEDA (Kleyer et al., 2008)) we therefore exploit the advantage that these estimates reflect species life-history traits in “agronomic conditions”. We believe that this strategy of fitting life history trait to their real value may able to better understand the effects that current management actions have on weed communities.

Material and Methods

Hidden Markov Model

Weed life cycle is represented by a Hidden Markov Model (HMM) where the unobserved state is the seedbank state (c_y) and the visible output, depending on the hidden state and cultural conditions applied (a^t), is the emerged flora (c_x). Observation probability $P_{a^t}\left(c_{x^{t+1}} \middle| c_{y^t}\right)$ depends of cultural conditions (a^t) and seedbank state at t($c_y{}^t$), while transition probability $P_{a^t}\left(c_{y^{t+1}} \middle| c_{x^{t+1}}, c_{y^t}\right)$ depends on cultural conditions (a^t), observations ($c_x{}^{t+1}$) and seedbank state at t($c_y{}^t$) (Fig. 1). Three life-history traits (LHT) are sufficient to synthesize weed seed dynamics: germination rate (σ), seed survival rate in the seed bank (s) and seed production (φ) i.e. the number of seeds produced per plant rejoining the seed bank. The first two LHT describe the dynamics of the persistence of the seed bank while seed production illustrates seed recruitment in the seed bank. The HMM is parametrised by these three parameters which still depend and account for the management actions i.e. the cultural practices.

Figure 1: Hidden Markov Model on counts and abundance classes. White and grey symbols represent the model respectively on counts and on abundance classes.



Observation and transition probabilities of the HMM are computed from the three LHT and assumptions such as the distributions of number of survival and emerging seeds and distribution of seeds produced. Details of calculations are provided in Appendix.

Dataset and model implementation

A total of 3080 sets of crop and weed data were obtained from 385 fields across France (latitudinal range of 761km, longitudinal range 696 km) between 2002 and 2009 as part of the on-going national Biovigilance project (Fried, 2007;Fried et al., 2008). Each field was surveyed in up to eight successive years by two or more expert volunteers walking across the survey area (200m²) for a minimum of 20 min and recording the abundances of all weed species. The abundance scale proposed by Barralis (Barralis, 1976) accounts in a semi quantitative way for the number of individuals per m²: '+' denotes that only one plant was found in the 2000 m² area; '1' that less than 1 individual/m²; '2' that 1 to 2 individual were

recorded in 1m² and '3', 4 and 5 that 3 to 20, 21 to 50 and more than 50 individuals were recorded in 1m², respectively.

We considered that similar crop types present similar weed management options which have similar effects on the weed seedbank and so we restricted the various cropping systems to four management actions corresponding to four highly contrasted crop types: winter-cereals (WC), oilseed rape (OR), maize (M) and sunflower (SF). These crop types were chosen for their sufficient representativeness. For selecting the fields (or part of times series), we applied two criteria: only the 4 crop types had been sown in the field and the fields were sampled at least two consecutive years. A total of 329 fields i.e. 1191 relevés were covering these conditions. Overall, the average duration of a time series was of 3.62 years (sd= 1.19 years). Crop types were unequally represented among the relevés: half (49.6%) of the relevés where performed in WC crops, 10.2%, 29.3% and 10.8% in OR, M and SF crops, respectively. A total of 288 species were observed in the relevés. We focused on the weed species that were (i) recorded in at least in 10% of the relevés and identified as (ii) therophyte species (i.e. annual species). A total of 32 weed species was selected (Appendix, Table X3). Finally, in order to enhance the quality of the estimation, we removed the species for which the number of couple (species *i*, crop type *j*) was present in less than 10% of the total number of crop sequences. A total of 18 species was selected.

Maximum likelihood computation and estimation of life history traits values

Given the form of the HMM, the likelihood of one time series of observations, of length *T*, for one species and on one field is defined by the following formula :

$$L_n = \sum_{c_{y^0} \ldots c_{y^{t-1}}} P(c_{x^1}|c_{y^0}, a^0) P(c_{y^0}) \prod_{t=2}^{T} P(c_{x^t}|c_{y^{t-1}}, a^{t-1}) * P(c_{y^{t-1}}|c_{x^{t-1}}, c_{y^{t-2}}, a^{t-2})$$

The log-likelihood of *N* time series, corresponding to *N* fields, of observations of one species is defined by:

$$logL = -\sum_{n=1}^{N} log(L_n)$$

Since we did not want to make any assumption on the a priori probabilities of seed bank abundance class at *t=0* ($p(c_{y0})$), we considered it as an input parameter of the model, which was estimated by maximum log-likelihood, together with the values of life history traits.

We used pseudo-random search optimization of Price (Price, 1977) to estimate the parameters with a maximum log-likelihood procedure using the *pseudoOptim* function of the *R* package *FME* (Soetaert and Herman, 2009). First, predictive efficiency was evaluated by cross-validation on 4 sub-datasets of 18 species and 82 fields chosen at random (see Appendix for more details).Then, the distribution of estimated parameters (over the entire data-set) was approximated with Bayesian Monte Carlo Markov Chain (MCMC) methods

using a Gibbs sampler (Gelfand and Smith, 1990). Mean and standard deviation of parameters estimated in the four sub-datasets by cross-validation were respectively used as initial parameters and 'scale parameters' of the Gibbs sampler. 5000 iterations were performed. The distributions obtained for each species were used to calculate the mean and the standard deviation of their three life history traits.

All calculations were carried out with the R statistical software. Considering the high complexity of calculations, a part of the code was written in C langague and compiled before being used in R. The code is available upon request.

Simplification of the abundance scale

Since some abundance classes were hardly ever observed, we merged them, which resulted in a decrease of the number of abundance classes. We defined four abundance classes for the observed density of mature plants (c_x): '1' = 0 ind/m², '2' = [1:2] ind/m², '3' = [3:20] ind/m² and '4' = [21: +∞[ind/m². Since the latter abundance class is not a upper-bounded and represents a large range of variation, we must create more abundance classes of number for seeds in the bank (c_y) than the number abundance classes of observed number of mature plants (c_x). It ensures than an abundance class of number of mature plant (c_x) doesn't necessary mean that the totality of seeds in bank has emerged. Hence we created 6 classes for abundance classes of number of seed in bank (c_y): '1' = 0 ind/m², '2' = [1:2] ind/m², '3' = [3:20] ind/m², '4' = [21:60] ind/m², '5' = [61:100] ind/m² and '6' = [101:+∞ [ind/m². Hence, absolute values of estimated parameters are directly dependent from the number and range of abundance classes of seed bank (c_y). Nevertheless, whatever the discrete scale used for abundances classes of seed bank (c_y), comparison of estimated life history traits between species and between cultural practices are pertinent.

Species growth rate estimation

For each crop type, we estimated the species growth rate by using a Leslie Matrix framework incorporating the life history of weed species into a structured population model (Caswell, 2001). The asymptotic growth rate λ_c was the dominant (largest) eigenvalue of the species transition matrix A for crop type *c*. The relative growth rate of species was defined for crop type *c* as the species asymptotic growth rate divided by the mean growth rate of the 18 species. Therefore, species with a negative (positive) relative growth rate are species in relative regression (progression) against the 18 species of our data-set.

Indicator values and specialization index of weed species

We investigated the relationship between life-history traits and the index of specialization (IS) of a species for a crop type using indicator values (Indval) of weed species for four crop types. Indval were computed by the *indicspecies* R package following (Dufrêne and Legendre,

1997). IS were computed by Outlying Mean Index (OMI, function *niche* in package ade4) (Dolédec et al., 2000) using crop type as environmental variable. Indval discriminates autumn/winter and spring/summer weeds by measuring the specificity and the frequency of a species to several crop types. IS discriminates specialist and generalist weeds for crop types by measuring their sensitivity for crop types.

Management application

To underline and illustrate the management application of our model, we investigated the effect on species abundance that would result from replacing the dominant crop type i.e. WC, by any other crop type i.e. OR, SF or M at their respective frequency. We randomly replaced 10% of WC (44 relevés), by one of the three other crop types, according to their relative frequencies in the observed dataset. For each couple (field and species), the start state of the seedbank at *t=0* was the one with the highest maximum likelihood value on observed time series (the predictive efficiencies calculation are given in the Appendix). Simulated abundances of species for all time series were then computed both for the observed as well as simulated crop rotations. Simulations were run 1000 times. 10 % of the WC was randomly replaced at the beginning of each run. Species mean abundance over the entire dataset was recorded for each run. Student's tests were applied to compare the observed and simulated mean abundance for each species.

Results

Predictive efficiencies

Table 1 summarizes the predictive quality results for the 18 studied weed species. Predictive efficiencies in fields where species have been observed varied between 45.1% and 59.9% (mean=51.1%). No significant correlations were observed between predictive efficiencies and the field occurrences (proportion of fields where species has been observed) or relevés occurrences (proportion of relevés where species has been observed). However, a significant negative correlation was observed between predictive efficiencies and the mean relevés occurrence per field (relevés occurrence / fields occurrence) (Pearson's correlation test ρ= -0.59, p-value = 8.76e-3). Furthermore, predictive efficiencies were negatively correlated to the indicator value (Indval) in winter cereals (ρ = -0.49, p-value = 0.035), while there was no significant positive or negative correlation with any indicator values in oilseed rape, maize and sunflower (OR: ρ = -0.24, p-value = 0.32; M: ρ = 0.24, p-value = 0.326; SF: ρ = 0.32, p-value = 0.192). Predictive efficiencies showed a significant positive correlation with the crop type index of specialization (*IS*) (ρ = 0.52, p-value = 0.027).

Table 1: Predictive efficiency. Mean and standard deviation of predictive efficiencies for 18 species are obtained by meaning predictive efficiencies of all relevés in fields where species have been observed by cross-validation on the four validation datasets. Start values of seed bank at t=0 are chosen by maximum likelihood. "Mean predictive efficiency" represents the proportion of $X_{obs}=X_{sim}$. "Mean erroneous predictive efficiency" represents the percentage of X_{obs} in $[X_{sim}-1:X_{sim}+1]$.

BAYER Code	Latin Name	fields occurrence (percent of fields) (N=329)	relevés occurrence (percent of relevés) (N=1191)	Mean predictive efficiency (N=1000)	standard deviation
ALOMY	*Alopecurus myosuroides*	39,8 %	20,8 %	0.4512	0.1413
ANGAR	*Anagallis arvensis*	43,1 %	15,4 %	0.5574	0.1372
CHEAL	*Chenopodium album*	74,1 %	46,0 %	0.4887	0.1425
FUMOF	*Fumaria officinalis*	32,2 %	12,7 %	0.5263	0.1164
GALAP	*Galium aparine*	59,5 %	30,8 %	0.4864	0.1102
MERAN	*Mercurialis annua*	41,6 %	23,8 %	0.5207	0.1277
PAPRH	*Papaver rhoeas*	45,5 %	20,9 %	0.5403	0.1150
POAAN	*Poa annua*	38,9 %	19,9 %	0.4793	0.1504
POLAV	*Polygonum aviculare*	54,7 %	27,2 %	0.4901	0.1478
POLCO	*Fallopia convolvulus*	44,9 %	20,8 %	0.4886	0.1285
SENVU	*Senecio vulgaris*	66,5 %	35,0 %	0.4861	0.1109
SINAR	*Sinapis arvensis*	37,3 %	19,8 %	0.4570	0.1231
SOLNI	*Solanum nigrum*	56,5 %	26,7 %	0.5988	0.1352
SONAS	*Sonchus asper*	50,7 %	22,0 %	0.5895	0.1171
SONOL	*Sonchus oleraceus*	27,0 %	10,6 %	0.5622	0.1094
STEME	*Stellaria media*	48,6 %	25,6 %	0.4690	0.1429
VERHE	*Veronica hederifolia*	52,2 %	25,2 %	0.5142	0.124
VERPE	*Veronica persica*	51,6 %	24,1 %	0.4924	0.1231

Estimated life history traits and growth rates

Detailed parameter estimates obtained by Bayesian inference for the 18 weed species are presented in the Appendix (table X2a and X2b). The relative growth rates of the 18 species are shown in Figure 2a. Among the 18 species, only *Veronica persica* and *Poa annua* showed a higher relative growth rate compared to the average dynamic of the 18 weed species of the dataset in the four crop types. 6 and 7 species showed a progression in 3 and 2 crop types, respectively. The highest number of species in progression was observed in WC (13) and M (13). Conversely, more than 50% of the species (12) showed a trend toward regression in OR. Important relative growth rates were observed for 6 species. The growth rates of *Fumaria officinalis*, *Poa annua*, *Sinapis arvensis* and *Veronica persica* were 2 to 2.5 times higher than the average one in OR. Similar patterns were noticed for the growth rates of *Lamium purpureum*, *Echinochloa crus-galli* in SF and of *Veronica persica* in M. Relative growth rates were positively correlated to (i) seed survival rates in WC (ρ = 0.87, p-value = 2.3e-6), , M (ρ = 0.85, p-value = 5.35e-6) and SF (ρ = 0.63, p-value = 4.2e-3), (ii) germination rates in SF (ρ = 0.47, p-value = 0.048) and (iii) seed production in OR (ρ = 0.65, p-value = 3.2e-3) and M (ρ = 0.63, p-value = 4.4e-3).

High interspecific and intraspecific variations between life-history trait estimates were observed (Fig. 2b). None of the 18 species presented similar life-history trait values for the three estimated traits in the four crop types. A majority of the species showed the ability to produce seeds, survive in the seed bank and germinate in the four crop types, however the range of magnitude of the success highly depend on the “species x crop type” interaction. For example, *Galium aparine* could produce seeds in the four crop types but the number of seeds was much higher in SF than in the three other crops. Conversly, the germination rate of *Galium aparine* was much higher in WC than in the three other crops.

Figure 2a: Relative growth rates of seedbank for eighteen common weed species in the four major crop types. Red polygon represents growth rates of seedbank of weed species for the four studied crop types (WC = winter cereals, OR = oilseed rape, M = maize and SF = sunflower). Center of white polygons correspond to a growth rate equal to zero, borders of white polygons represent the mean growth rates for each crop type over the 18 species. These growth rates do not account for dispersion within the landscape but assume that each field is an isolated unit.

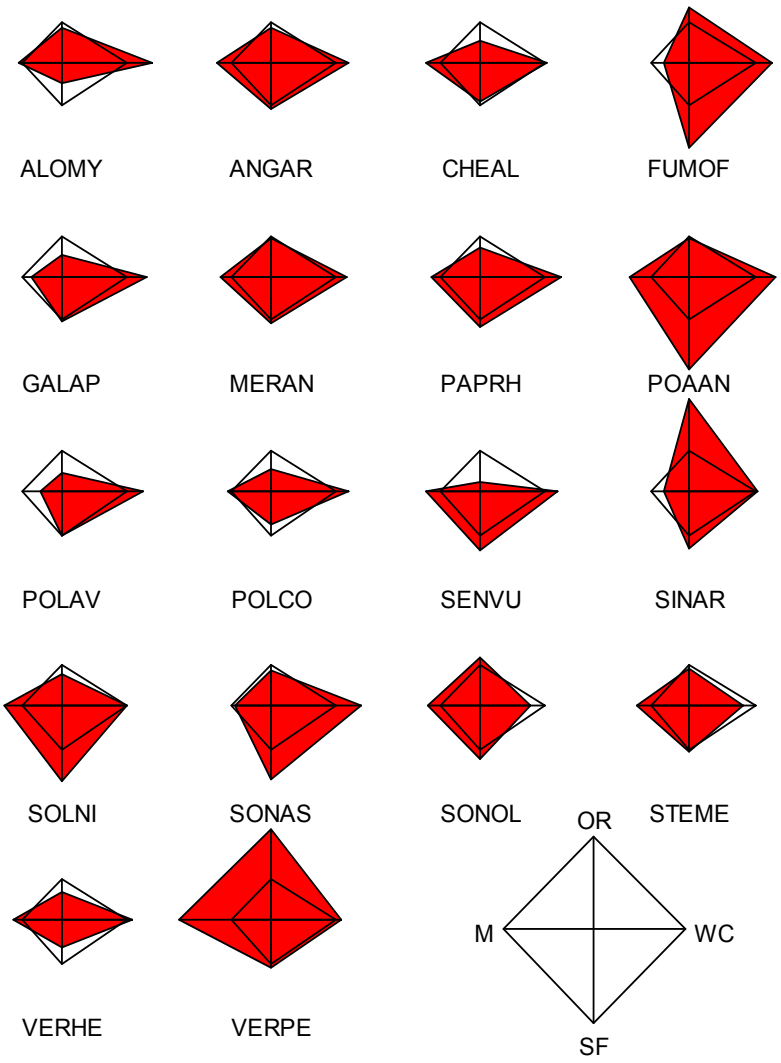
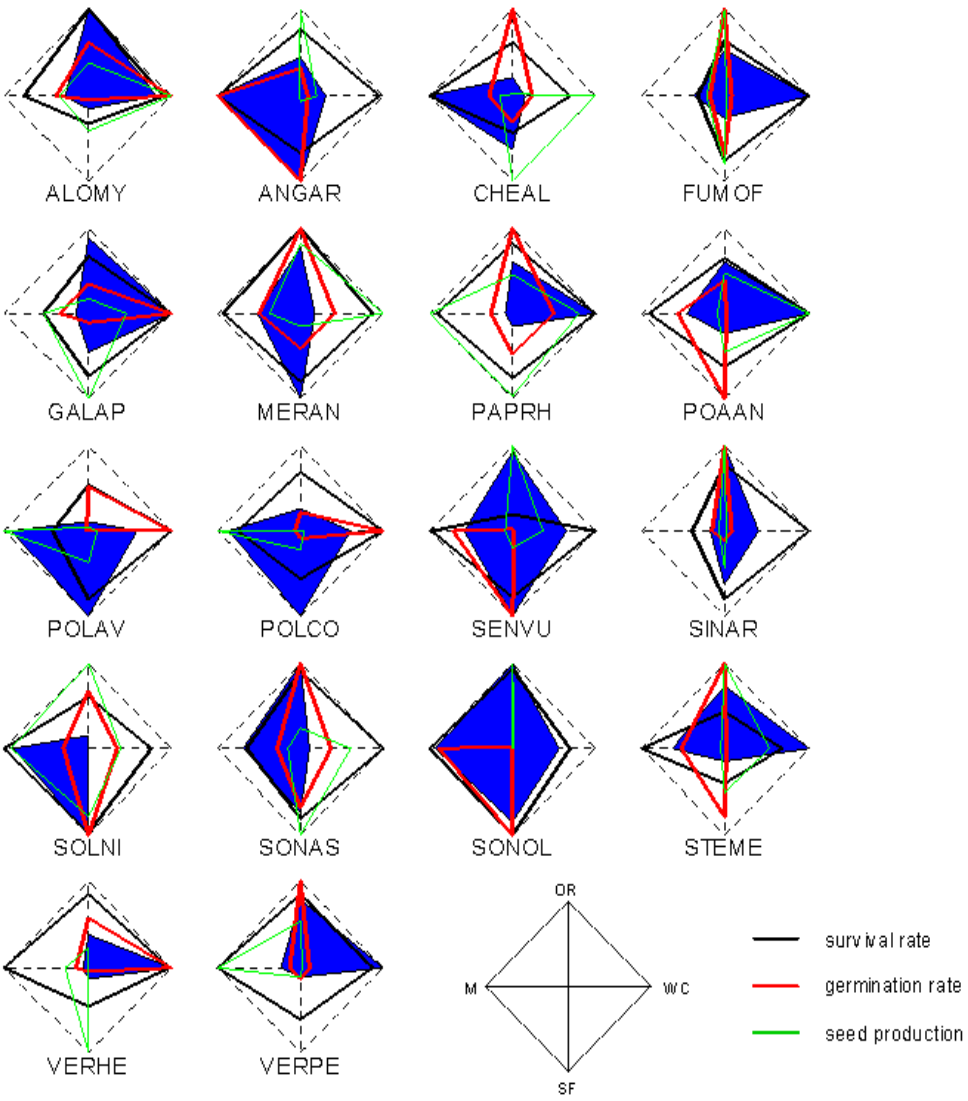


Figure 2b: Life history traits of weed species for four crop types. Species survival rates, germination rates and seed production per plant are respectively presented by black, red and green polygons for four crop types (WC = winter cereals, OR = oilseed rape, M = maize and SF = sunflower). Values are scaled by dividing each value by the maximal one in the four crop types. Hence, for each species, each scaled life history trait varies between 0 and 1 and the polygon is shifted in the direction of the crop(s) where it has its highest estimated success. Dashed white polygon represents value equal to 1 for the three life history traits, i.e. maximal values of species for each life history trait. Blue polygons represent indicator values (IndVal) of species for the four crop types. Indicator values are scaled by the maximum value of species for the four crop type. Hence, dashed lines represent the maximum of indicator values of each species.



Divergences were also observed between species (Fig. 2b) showing variation in species life-history strategies. Germination rate and seed production were globally negatively correlated to the exception of three species (*Alopecurus myosuroides*, *Fumaria officinalis* and *Mercurialis annua*) for which germination rate and seed production were positively correlated. High intraspecific variations were noticed in seed production. Only two species (*Papaver rhoeas* and *Solanum nigrum*) could produce a relative high number of seeds in the four crop types. Conversely, intraspecific variations in seed survival rate were relatively low: only three species had a survival rate lower than 50% of the maximum survival rate for at least two crop types (*Chenopodium album, Polygonum aviculare* and *Stellaria media*). The other species showed a relatively constant survival rate between crop types. Overall, nine species had survival rates superior to 50% of the maximal survival rate for all crop types.

Relationship between LHT and indicator values or Index of Specialization

For each crop type, we observed a high positive correlation between the average species relative germination rates and their Indicator values (*Indval*) (Spearman's correlation, unilateral test: WC: Rs = 0.92 p-value < 2.2e-16; OR: Rs = 0.85, p-value < 2.2e-16; M: Rs = 0.86 p-value < 2.2e-16 and SF: Rs = 0.92 p-value < 2.2e-16) (Fig. 3a). Furthermore, mean relative germination rates of species were significantly negatively correlated to mean relative seeds production in Maize (Rs = -0.44, p-value = 6.49e-3) and showed a negative correlation in winter cereals (Rs = -0.39, p-value = 0.10), Oilseed rape (Rs = -0.44, p-value = 0.06) and Sunflower (Rs = -0.39, p-value = 0.10), although not significant (Fig. 3b). No significant correlations were observed between Index of Specialization (IS) and estimated Life History Traits or growth rates.

Figure 3a: Positive relationships between relative germination rates and indicator values in several crop types of each species. Symbols indicate relative germination rate of species for each crop type (winter cereals = circle, oilseed rape = cross, maize = square, sunflower = triangle). Hence, each species is represented by one symbol of each crop type. Vertical line associated to a symbol represents quantiles 25% and 75% of relative germination rates.

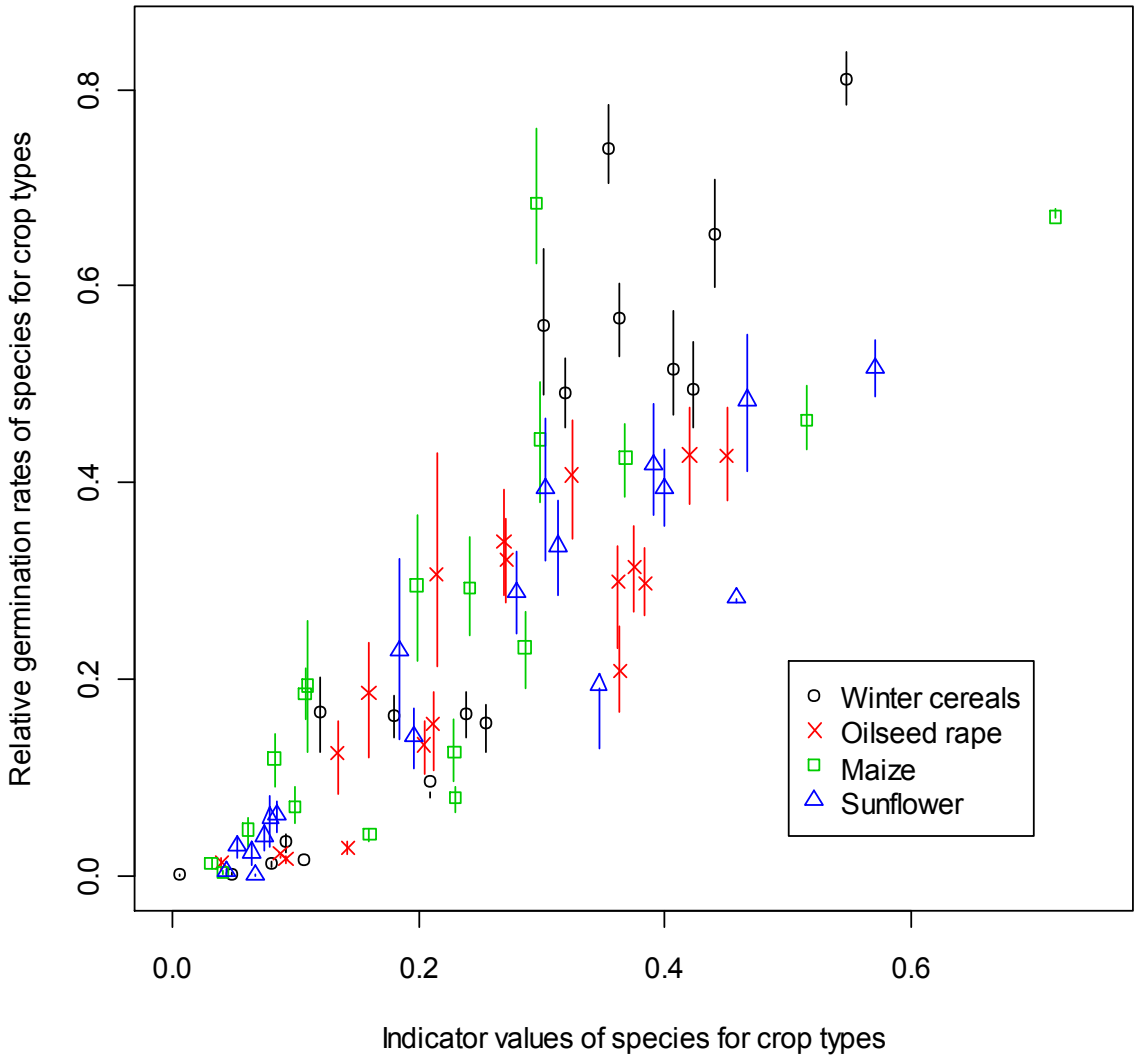
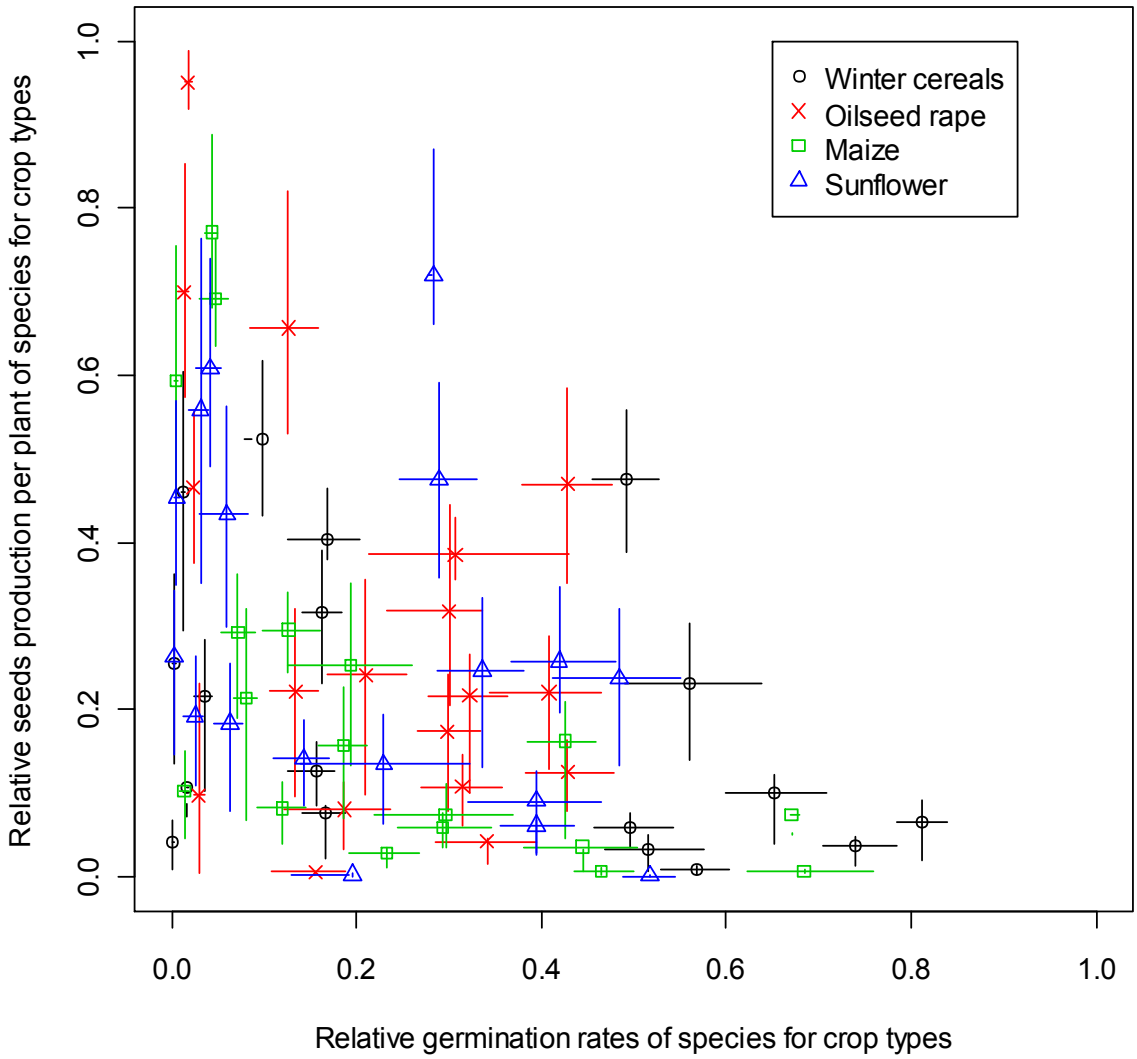


Figure 3b: **Negative relationship between relative germination rate and relative seed production per plant in the four crop types.** Symbols indicate the relative germination rate of species for each crop type (winter cereals = circle, oilseed rape = cross, maize = square, sunflower = triangle). Horizontal and vertical lines represent the quantiles 25% and 75% of relative germination rates and relative seeds productions per plant, respectively.



Management application

Table 2 summarizes the effect of a reduction of 10% of WC in the crop sequence on the mean abundance of the 18 species. A smaller occurrence of WC had a positive effect on the abundance of five of the 18 species (*Anagallis arvensis*, *Chenopodium album*, *Mercurialis annua*, *Fallopia convolvulus* and *Sonchus asper*) which had a mean abundance significantly higher in the modified crop rotations than in the observed ones. On the opposite, a reduction of 10% of the number of WC in the crop sequence significantly decreased the mean abundance of four species (*Alopecurus myosuroides*, *Fumaria officinalis*, *Poa annua* and *Stellaria media*). The other species were not affected by this particular modification of the cropping system. It is worth noting that with the exception of *Chenopodium album*, none of the 18 other species experience a variation of more than 10% when WC is reduced by 10%.

Table 2: Effect of the reduction of 10% of Winter Cereals on species mean abundances. "Mean Abundance Modification" represents the mean of the differences between the species mean abundance estimated with observed crop rotations and modified crop rotation where 10% of the WC crops were randomly replaced by one of the other crop types (according to their relative observed frequencies). Significant differences between the two distributions are indicated following: "***" = p-value<1e-3; "**" = 1e-3 ≤ p-value < 1e-2; "*" = 1e-2 ≤ p-value < 5e-2; "+" = 5e-2 ≤ p-value < 0.1 and "ns" = p-value ≥ 0.1. Numbers between brackets correspond to the percentage of modification. The reduction of WC proportion has a positive effect on species mean abundance (progressing species), no effect (no change) or a negative effect (regressing species).

BAYER Code	Mean Abundance Modification
progressing species	
ANGAR	0.0931** (2,5%)
CHEAL	0.2600*** (14,8%)
MERAN	0.1757*** (9,4%)
POLCO	0.1137*** (4,6%)
SONAS	0.0616*** (10,3%)
no change	
GALAP	-0.0198^{ns} (-2,3%)
PAPRH	-0.0067^{ns} (-1,3%)
POLAV	0.0390^{ns} (0,4%)
SENVU	-0.0126^{ns} (-0,5%)
SINAR	0.0085^{ns} (0,7%)
SOLNI	0.0402^{ns} (0,7%)
SONOL	0.0006^{ns} (0,1%)
VERHE	0.0012^{ns} (0,1%)
VERPE	-0.0144^{ns} (-2,3%)
regressing species	
ALOMY	-0.2525*** (-3,5%)
FUMOF	-0.0774*** (-5,6%)
POAAN	-0.0673* (-2,7%)
STEME	-0.1160** (-2,5%)

Discussion

To conclude, our approach opens some perspectives in weed management. Hidden Markov Model is well adapted to model weed dynamic by taking into account the dynamic of an imperfectly observed compartment i.e. the seedbank. This statistical modeling offers a complementary alternative to complex mechanistic modeling by using preexisted simple dataset. Our modeling approach and LHT estimates in the different crop types give insights into the effects of management actions on weed species dynamics which could be use to predict the changes in weed community in response to changes in agricultural practices. This knowledge is of high importance to develop of more sustainable cropping systems or to evaluate the consequences of agricultural subsidies policies.

Predictive efficiencies and quality of the estimation

Predictive efficiencies varied between 45% and 59%. They were dependent on the frequencies of occurrence of species; the lower the occurrence frequency of a species, the better the predictive efficiency. The predictive efficiencies were also sensitive to the indicator value (IndVal) in winter cereals and oilseed rape. In both crop types, the occurrence of species with low IndVal is highly dependent on the management actions i.e. agronomical practices applied and associated to these crop types. Therefore, when management actions are efficient, the probability to observe these species will be low whatever the state of the seedbank. Conversely, it is more difficult to predict the abundance of a species when agronomical practices are less efficient as the quality of the prediction then relies both on the relevance of the estimates of life history traits and on the state of seedbank. A main consequence is that the dynamics of generalist species and therefore, their response to management actions, are more difficultly determined because of their ability to grow under a wide range of agronomical conditions, decreasing the relevancy of the approximation. This seems to be confirmed by the significant positive correlation between predictive efficiency and crop type index of specialization (IS). In our study, predictions on the occurrence of weed species specialized for spring/summer crop types (maize and sunflower) with low Indval in WC and OR, were reliable. The quality of our model estimation was also confirmed by the strong positive correlation between Indval and germination rate for each crop type.

Most of weed LHT values are usually obtained in optimal conditions in experimental sites or laboratory/greenhouse conditions (Gardarin et al., 2010) or in laboratory (Bergelson and Roux, 2010). As we suspected, our LHT estimates were not correlated with values of seed mass and persistence found in the literature and in database (LEDA and BiolFlor). This contrasts with the results of prior studies (see for example for seed persistence, (Cerabolini et al., 2003), while in accordance with others (Saatkamp et al., 2009;Gardarin et al., 2010). Several hypotheses can be advanced that could explain this discrepancy. First, despite their

relative importance, few values were available for covering the range of variations for the three LHT of the 18 species studied here. This pinpoints the difficulties of measuring weed seed life-history traits. Second, our LHT values were estimated from real crop sequences and thus, strongly dependent on the set of agricultural practices, the structure of the dataset and the length of the time series (the mean length of time series per field is 3.62 years). Although we may pay attention to improve the quality of the LHT estimation, species and crop types will not be equally distributed along geographic or other ecological gradients with unknown (repercussion) on the quality of estimation for some couple "species-crop type" (see Table X3 in appendix for more details).

Growth rates and life history traits

Various combinations of life history traits were associated with species positive relative growth rates over a wide range of agronomical and ecological disturbances. However, no clear segregation between species by crop type was observed and within a crop type, no optimal response strategy to management disturbances was observed. On the contrary, species with contrasting life history strategies coexist. For example, in sunflower, *Papaver rhoeas* produces a high number of short-persistent seeds while *Senecio vulgare* produces few seeds that persist for a long time in the seed bank. Moreover, our results showed that a weed species may respond to disturbances in alternative ways since some species can have positive growth rates in two crop types either by high germination rate or high seed production (see for example, *Veronica persica*). Finally, our simulation results reveal a trade-off between germination rate and seed production in every crop type, although highly expressed in maize and sunflower. We cannot however exclude that part of this trade-off may come from an indirect effect of management actions. Farmers generally adapted their weeding strategies to cope with the more abundance and pernicious weed species. As a result, weed species with lower abundance may not be the target of treatments (mechanical or chemical) with resulting escape followed by capacity to complete their life cycle by producing seeds in some optimal conditions of low density and high access to nutriment. The strength of the trade-off between germination rate and seed production can also be exacerbated by the sampling method used in our dataset. Surveys were generally performed early in the agricultural campaign; hence, most of the individual plants sampled were at the seedling stage. Therefore, the observed abundance class of plants may not reflect the abundance of mature plants reaching bloom.

Species specialized to a crop type i.e. showing high Indval value, generally have higher germination rates. Our results are consistent with the perspective of life-history evolution and previous studies in weed research (e.g. Gunton *et al.*, 2011) pointing out the importance of the timing of germination in synchrony with the crop. Within the ephemeral environment of an arable field with a short window for growth, successful weeds normally germinate around the time the crop is sown and complete their reproductive effort before it

is harvested. Nevertheless, for several species, germination rates per crop types were not positively correlated to Indval. For example, *Mercurialis annua* has a very low estimated germination rate and a high indicator value in oilseed rape. Indicator values are computed with relevés performed on emerged flora without accounting for the seedbank dynamic. As a consequence, a high indicator value with a crop type revealed a high occurrence of the species in that crop type given the observed pattern occurrence (the dataset) without explaining the underlying processes. It is in these situations that the complementarities with other mechanistic approaches of the weed biology be welcomed. Given the crop rotation dynamic, a weed species can have a high indicator value either due to a high germination rate or to the low germination rate of an important seedbank. Thus, *Mercurialis annua* could have important Indval in oilseed rape (OR) because preceding crop was a winter cereals (WC) in which it has a high seed production which has significantly increase the number of seeds in the seed bank. Therefore, in highly disturbed habitats such as arable field, life-history traits would better reflect plant degree of specialization to agricultural conditions than Indicator values.

To summarize, the trade-off and the ability of weed species to switch their life history strategy have major consequences for management. First, decreasing the germination success, which allows an increase in the number of seed produced, would result in an increase of the population growth rate and hence, of the future infestation. Second, management actions should be designer at the crop type x species interaction level given the crop sequence, and not only at the species' internal characteristics.

Management applications and Improvement of the model and LHT estimations

Simulations based on estimated LHT may provide a useful tool to predict the changes in weed communities in response to changes in agricultural practices. Our results showed that decreasing the proportion of winter cereals can modify the weed community structure by decreasing the abundance of species recognized by weed scientist to be specialized to winter cereals crop type and conversely increasing the abundance of species benefiting from the increase of another crop types. It is noteworthy here to pinpoint that there is no proportionality between the reduction of winter wheat return in the crop sequence and the resulting change in weed abundance. For 17 out of the 18 species, the change was less than 10% so that buffering effects may prevail as either seed bank span over years or bet hedging behaviour attached to any generalist able to complete its life cycle in more than a single crop. The fact that changing crop occurrence in the cropping sequence result in varying affection of weed species is in accordance with other analyses conducted on weed communities' changes over the last decades in intensive agriculture systems. Several studies conducted in France (Fried, 2010), Northern Europe (Andreasen and Streibig, 2011;Hanzlik and Gerowitt, 2011) or Alaska (Conn et al., 2011) describe trajectories of weed species pools diversely affected but sharing the characteristics of being among the highest species turnovers.

Allowing some increased complexity of the model could improve LHT estimations. First, in order to simulate crop sequences including grassland, perennial species could be included in the model by adding the survival rate of mature plants. However, adding any new LHT would request a larger dataset. Second, seed dormancy, which is a major temporal dispersion mode, could be included by introducing a second hidden state variable of c_y representing a proportion of geminating seeds and a seed dormancy longevity rate. Finally, the error in the attribution of emerged plant abundance class could be modeled by modifying the structure of the model by introducing of (i) a second hidden state variable i.e. the "real" abundance class of mature plants (c_{xr}) and (ii) a third transition probabilities $p(c_x|c_{xr})$. The error rate of abundance class attribution could then be modeled more clearly, but these modifications of the model would imply an increase in the likelihood calculation time.

Acknowledgements

Benjamin Borgy is funded through a fellowship ANR-OGM Vigiweed. Thanks to Guillaume Fried for access to a part of the Biovigilance-Flore data-set and for fruitful discussions.

References

Andreasen, C. & Streibig, J. C. (2011) *Evaluation of changes in weed flora in arable fields of Nordic countries - based on Danish long-term surveys*. *Weed Research,* **51,** 214-226.

Backman, J. C. & Tiainen, J. (2002) *Habitat quality of field margins in a Finnish farmland area for bumblebees (Hymenoptera: Bombus and Psithyrus)*. *Agriculture Ecosystems & Environment,* **89,** 53-68.

Barralis, G. (1976) Méthode d'étude des groupements adventices des cultures annuelles. *INRA - 5ème Colloque International sur l'Ecologie et la Biologie des Mauvaises Herbes* (ed^(eds, pp. 59-68. Dijon, France.

Bergelson, J. & Roux, F. (2010) *Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nature Reviews Genetics,* **11,** 867-879.

Bohan, D. A., Powers, S. J., Champion, G., Haughton, A. J., Hawes, C., Squire, G., Cussans, J. & Mertens, S. K. (2011) *Modelling rotations: can crop sequences explain arable weed seedbank abundance? Weed Research,* **51,** 422-432.

Bonsall, M. B., Hassell, M. P. & Asefa, G. (2002) *Ecological trade-offs, resource partitioning, and coexistence in a host-parasitoid assemblage*. *Ecology,* **83,** 925-934.

Bonsall, M. B., Jansen, V. A. A. & Hassell, M. P. (2004) *Life history trade-offs assemble ecological guilds*. *Science,* **306,** 111-114.

Brusa, G., Ceriani, R. & Cerabolini, B. (2007) *Seed germination in a narrow endemic species (Telekia speciosissima, Asteraceae): implications for ex situ conservation*. *Plant Biosystems,* **141,** 56.

Burnside, O. C., Wilson, R. G., Weisberg, S. & Hubbard, K. G. (1996) *Seed longevity of 41 weed species buried 17 years in eastern and western Nebraska. Weed Science,* **44,** 74-86.

Caswell, H. (2001) *Matrix population models: Construction, analysis and interpretation. 2nd Edition.* . Sunderland, Massachusetts, USA.

Cerabolini, B., Ceriani, R. M., Caccianiga, M., De Andreis, R. & Raimondi, B. (2003) *Seed size, shape and persistence in soil: a test on Italian flora from Alps to Mediterranean coast*. *Seed Science Research*.

Colbach, N., Darmency, H. & Tricault, Y. (2010) *Identifying key life-traits for the dynamics and gene flow in a weedy crop relative: Sensitivity analysis of the GENESYS simulation model for weed beet (Beta vulgaris ssp vulgaris)*. *Ecological Modelling,* **221,** 225-237.

Conn, J. S., Beattie, K. L. & Blanchard, A. (2006) *Seed viability and dormancy of 17 weed species after 19.7 years of burial in Alaska*. *Weed Science,* **54,** 464-470.

Conn, J. S., Werdin-Pfisterer, N. R. & Beattie, K. L. (2011) *Development of the Alaska agricultural weed flora 1981-2004: a case for prevention*. *Weed Research,* **51,** 63-70.

Dolédec, S., Chessel, D. & Gimaret, C. (2000) *Niche separation in community analysis: a new method*. *Ecology,* **81,** 2914-2927.

Dufrêne, M. & Legendre, P. (1997) *Species assemblages and indicator species: The need for a flexible asymetrical approach. Ecological Monographs***,** 345-366.

Evans, E. K. M., Menges, E. S. & Gordon, D. R. (2003) *Reproductive biology of three sympatric endangered plants endemic to Florida scrub. Biological Conservation,* **111,** 235-246.

Forcella, F., Benech-Arnold, R. L., Sanchez, R. & Ghersa, C. M. (2000) *Modeling seedling emergence*. *Field Crops research,* **67,** 123-139.

Franke, A. C., Lotz, L. A., Van der Burg, W. J. & Van Overbeek, L. (2009) *The role of arable weed seeds for agroecosystem functioning*. *Weed Research,* **49,** 131-141.

Fried, G. (2007) 'Biovigilance Flore', a long-term French weed survey. *20ème Conférence du COLUMA. Journées Internationales sur la Lutte contre les Mauvaises Herbes.* (ed^(eds, pp. 315. Dijon, France.

Fried, G. (2010) *Spatial and temporal variation of weed communities of annual crops in France*. *Acta Botanica Gallica,* **157,** 183-192.

Fried, G. & Gaba, S. (submitted) *Niche filtering and limiting similarity in arable weed communities among and within fields*.

Fried, G., Norton, L. R. & Reboud, X. (2008) *Environmental and management factors determining weed species composition and diversity in France*. *Agriculture, Ecosystems & Environment,* **128,** 68-76.

Gardarin, A., Durr, C., Mannino, M. R., Busset, H. & Colbach, N. (2010) *Seed mortality in the soil is related to seed coat thickness*. *Seed Science Research,* **20,** 243-256.

Gelfand, A. E. & Smith, A. F. M. (1990) *Sampling-Based Approaches to Calculating Marginal Densities*. *Journal of the American Statistical Association,* **85,** 398-409.

Gibbons, D. W., Bohan, D. A. & Rothery, P. (2006) *Weed seed resources for birds in fields with contrasting conventional and genetically modified herbicide tolerant crops.*

Philosophical Transactions of the Royal Society London B - Biological Sciences, **273,** 1921-1928.

Gunton, R. M., Petit, S. & Gaba, S. (2011) *Functional traits relating arable weed communities to crop characteristics*. *Journal of Vegetation Science,* **22,** 541-550.

Hanzlik, K. & Gerowitt, B. (2011) *The importance of climate, site and management on weed vegetation in oilseed rape in Germany*. *Agriculture Ecosystems & Environment,* **141,** 323-331.

Hawes, C., Haughton, A. J., Osborne, J. L., Roy, D. B., Clark, S. J., Perry, J. N., Rothery, P., Bohan, D. A., Brooks, D. R., Champion, G. T., Dewar, A. M., Scott, R. J., Firbank, L. G. & Squire, G. R. (2003) *Responses of plants and intertebrate trophic groups to contrasting herbicide regimes in the Farm Scale Evaluations of genetically modified herbicide-tolerant crops. Philosophical Transactions of the Royal Society London B - Biological Sciences,* **358,** 1899-1913.

Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M., Poschlod, P., van Groenendael, J. M., Klimes, L., Klimesová, J., Klotz, S., Rusch, G. M., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Endels, P., Götzenberger, L., Hodgson, J. G., Jackel, A.-K., Dannemann, A., Kühn, I., Kunzmann, D., Ozinga, W. A., Römermann, C., Stadler, M., Schlegelmilch, J., Steendam, H. J., Tackenberg, O., Wilmann, B., Cornelissen, J. H. C., Eriksson, O., Garnier, E., Fitter, A. & Peco, B. (2008) *The LEDA Traitbase: A database of life-history traits of the Northwest European flora. Journal of Ecology,* **96,** 1266-1274.

Kühn, I., Durka, W. & Klotz, S. (2004) *BiolFlor — a new plant-trait database as a tool for plant invasion ecology*. *Diversity and Distributions,* **10,** 363-365.

Moza, M. K. & Bhatnagar, A. K. (2007) *Plant reproductive biology studies crucial for conservation. Current Biology,* **92**.

Munier-Jolain, N. M., Deytieux, V., Guillemin, J. P., Granger, S. & Gaba, S. (2009) Multi-criteria evaluation of cropping systems prototypes based on integrated weed management. *XIIIeme Colloque International sur la Biologie des Mauvaises Herbes, Dijon, France, 8-10 Septembre 2009.* (ed^(eds, pp. 268-277.

Murdoch & Ellis, C. J. (2000) Dormancy, viability and longevity. *Seeds: the ecology of regeneration in plant communities* (ed^(eds M. Fenner). CAB International, Wallingforg, UK.

Price, W. L. (1977) *A Controlled Random Search Procedure for Global Optimisation*. *The Computer Journal,* **20,** 367-370.

Saatkamp, A., Affre, L., Dutoit, T. & Poschlod, P. (2009) *The seed bank longevity index revisited: limited reliability evident from a burial experiment and database analyses*. *Annals of Botany,* **104,** 715-724.

Soetaert, K. & Herman, P. M. J. (2009) *A Practical Guide to Ecological Modelling. Using R as a Simulation Platform*. Springer-Verlag, New York, USA.

Sutcliffe, O. L. & Kay, Q. O. N. (2000) *Changes in the arable flora of central southern England since the 1960s*. *Biological Conservation,* **93,** 1-8.

Sutherland, S. (2004) *What makes a weed a weed: life history traits of native and non indigenous plants in the USA. Oecologia,* **141,** 24-39.
Venable, D. L. (1989) Modeling the evolutionary ecology of seed banks. (ed^(eds, pp. 67-87.

Appendix

HMM model on counts

For one species, under a given set of cultural practices applied between *t* and *t+1* (a^t), a very simple numerical model can be written where the number of viable seed (Y^{t+1}) at the end of the growing season is equal to this number at the beginning minus the number of seeds that died (Y_m^{t+1}), minus the number of emerging seed (X^{t+1}) plus the number of seeds produced (Y_p^{t+1}) (eq. 1):

$$Y^{t+1} = Y^t - X^{t+1} - Y_m^{t+1} + Y_p^{t+1} \quad \text{(eq. 1)}$$

Hence, the number of mature plants (X^{t+1}) depends of the number of seeds (Y^t) and the emerging rate (σ). The number of surviving seeds (Y^t- X^{t+1}-Y_m^{t+1}) depends of the number of non emerged seeds (Y^t- X^{t+1}) and of the survival rate *s*. The number of produced seeds (Y_p^{t+1}) depends of the number of mature plants (X^{t+1}) and the number of seeds produced per plant φ. Therefore, the number of seeds in bank (Y^{t+1}) depends of the number of seeds (Y^t), the non-emerged seeds survival rate *(1-σ)*s*, the number of mature plants observed (X^{t+1}) and the number of seeds produced per plant φ (fig. 1).

We assumed that the germination and death rates of one seed follow discrete Bernouilli distributions which respectively take value “emerged” and “dead” with probability σ and *1-s*. Thus the sum of emerged seeds (X^{t+1}) and the sum of dead seeds (Y_m^{t+1}) follow Binomial distributions respectively from populations of size Y^{t+1} and Y^t- X^{t+1} with probabilities equal to σ and *1-s* (eq. 2 and eq. 3).

$$X^{t+1} \sim B(Y^t, \sigma) \quad \text{(eq. 2)}$$

$$Y_m^{t+1} \sim B(Y^t - X^{t+1}, 1 - s) \quad \text{(eq. 3)}$$

The number of seeds produced per mature plant fluctuates around the expected seed production number φ. In order to take this variability into account, we assumed that the total number of seeds produced (Y_p^{t+1}) follows a Poisson distribution of mean parameter equal to the number of mature plants (X^{t+1}) multiplied by the seed production number per plant φ (eq.4).

$$Y_p^{t+1} \sim P(X^{t+1} * \varphi) \quad \text{(eq. 4)}$$

Hence, following environmental conditions and cultural practices applied between *t* and *t+1*(a^t), the dynamics of seed bank and the emergence of mature plants are defined by a set of three LHT (σ, *s* and φ).

The HMM model on counts is fully described by the two transition probabilities: $P(Y^{t+1}|Y^{t},X^{t+1})$ and from $P(X^{t+1}|Y^{t})$. These two transition probabilities are fully defined from equations (1) to (4). Their evaluation as well as the temporal structure of the HMM model is presented in more details below, in the case of the HMM model on abundance classes.

HMM model on abundance classes

In practice information about emerged flora is available in terms of abundance classes, not counts. It is also more realistic to work with abundance classes of seeds in the bank. So we extended the first HMM model based on counts to a model based on abundance classes. We defined respectively the abundance classed of seeds and of emerged plants observed as c_y and c_x. Since we did not have any information about the distribution of the counts of seeds/emerged plants within a given class, we decided to attach to each class a uniform distribution on the interval defining the class. Since the maximal abundance class is only lower bounded by *bb*, we attached a geometric distribution of parameter $p = 1/bb$ to this class. In our simple model, the abundance class of mature plants, $c_{x^{t+1}}$, observed between *t* and *t+1* only depends on the abundance class of seeds, c_{y^t}, at the end of the previous year and on the cultural practices applied between *t* and *t+1*, a^t. The abundance class of seeds, $c_{y^{t+1}}$, depends on the abundance class of seeds, c_{y^t}, the observed abundance class of plants between *t* and *t+1*, $c_{x^{t+1}}$, and the practices applied between *t* and *t+1*, a^t (fig. 1).

Hence, the probability of finding $c_{x^{t+1}}$, knowing c_{y^t} and the group of cultural practices, a^t, applied between *t* and *t+1*, is given by the formula:

$$P_{a^t}(c_{x^{t+1}}|c_{y^t}) = \sum_{y_e^{t+1} \in I_{c_{x^{t+1}}}} \sum_{y^t \in I_{c_{y^t}}} P(y^t|c_{y^t}) P(y_e^{t+1}|y^t)$$

Where $I_{c_{x^{t+1}}}$ and $I_{c_{y^t}}$ represent respectively the intervals of values of abundance classes $c_{x^{t+1}}$ and c_{y^t}.

The probability of finding $c_{y^{t+1}}$, knowing c_{y^t}, $c_{x^{t+1}}$ and the group of cultural practices, a^t, applied between *t* and *t+1*, is given by the formula:

$$P_{a^t}(c_{y^{t+1}}|c_{x^{t+1}}, c_{y^t}) = \sum_{y^{t+1} \in I_{c_{y^{t+1}}}} \sum_{y^t \in I_{c_{y^t}}} \sum_{y_e^{t+1} \in I_{C_{x^{t+1}}}} \sum_{y_m^{t+1} \in I_{c_{y^t}}} P(y_m^{t+1}|y^t, y_e^{t+1}) P(y^t|c_{y^t}) P(y_e^{t+1}|c_{x^{t+1}}) P(Y_p^{t+1} = y^{t+1} - y^t + y_e^{t+1} + y_m^{t+1}|y_e^{t+1})$$

Transition probabilities were estimated by simulation. In order to compute $P_{a^t}(c_{x^{t+1}}|c_{y^t})$ for all states c_{y^t}, *K* random values of Y^t were generated, following the distribution of c_{y^t}. For each random value of Y^t, random values of X^{t+1} were generated, following the germination rate (σ) of a^t, and attached to the corresponding classes $P_{a^t}(c_{x^{t+1}}|c_{y^t})$ is

then obtained from frequency counts. The same method was applied in order to compute $P_{a^t}\left(c_{y^{t+1}} \middle| c_{x^{t+1}}, c_{y^t}\right)$. The accuracy of this computation can be controlled by the parameter *K*. With small values of *K*, computation is faster but less accurate than with a high value of *K*.

In the abundance class model defined above, it is assumed that abundance classes are defined as interval of counts. This amounts to assuming that there is no error in counting. However, this assumption is unrealistic, especially when counts are close to borders between abundance classes. In order to take this potential counting error into account, an error rate (ε) can be included in the computation of transition probabilities. It represents the probability of an error of abundance class estimation of +/- one class (for non extreme abundance classes). This error rate (ε) was fixed to 0.2.This could seem to be very high but, due to the poor quality of our data (*cf.* sample method used), it is a plausible value.

Predictive efficiencies calculation

We evaluated the predictive efficiency and variation in parameters estimation by randomly dividing our dataset in four sub-datasets (which were our four validation data-sets). Then we created four crossed data-sets containing three of the four sub-datasets (which were our four training data-sets). For each of the four training data-sets, we computed maximum log-likelihood (using a high maximum number of iterations (50 000) and a relatively low precision criterion (1e-6) in the algorithm). The number of simulations used to evaluate transition probabilities (*K*) was fixed to 10 000.

For each couple (field,species), we fixed the start state of seedbank at *t=0* at the maximum likelihood value. Then we performed 1000 simulations for each field of the validation data-set. We defined predictive efficiency as the percentage of good attribution of observed abundance classes. As predictive efficiencies on fields where species are missing are then equal to 1, we performed predictive efficiencies only on fields where species have been observed at least once.

Table X1: Current knowledge on weed species. Ind_WC,Ind_OR, IND_M and Ind_SF represent respectively the indicator values (Indval) (Dufrêne & Legendre, 1997) of species for winter cereals (WC), oilseed rape (OR), maize (M) and sunflower (SF). IS represents Index of Specialization (Outlying Mean Index) (Dolédec et al., 2000).

Code	Latin name	Ind_WC	Ind_OR	Ind_M	Ind_SF	IS
ALOMY	*Alopecurus myosuroides*	0.3192	0.3839	0.1079	0.0648	0.4683
ANGAR	*Anagallis arvensis*	0.0922	0.1347	0.2985	0.3025	0.4545
CHEAL	*Chenopodium album*	0.1074	0.1425	0.7172	0.4575	1.3937
FUMOF	*Fumaria officinalis*	0.3017	0.1596	0.1100	0.0798	0.4557
GALAP	*Galium aparine*	0.4227	0.3754	0.0617	0.1963	0.5347
MERAN	*Mercurialis annua*	0.0810	0.3632	0.2415	0.4672	0.7674
PAPRH	*Papaver rhoeas*	0.4406	0.2693	0.0409	0.0676	0.9610
POAAN	*Poa annua*	0.3539	0.2122	0.1603	0.0849	0.3769
POLAV	*Polygonum aviculare*	0.2383	0.0404	0.3679	0.3996	0.1201
POLCO	*Fallopia convolvulus*	0.2088	0.0880	0.2959	0.3462	0.0945
SENVU	*Senecio vulgaris*	0.2554	0.3621	0.2291	0.3904	0.0791
SINAR	*Sinapis arvensis*	0.1804	0.4505	0.0830	0.2791	0.5227
SOLNI	*Solanum nigrum*	0.0059	0.0924	0.5151	0.5705	1.9965
SONAS	*Sonchus asper*	0.0489	0.4200	0.2863	0.3132	1.2111
SONOL	*Sonchus oleraceus*	0.1200	0.2150	0.1986	0.1848	0.0699
STEME	*Stellaria media*	0.3632	0.2711	0.2301	0.0534	0.2636
VERHE	*Veronica hederifolia*	0.5475	0.2048	0.0319	0.0749	1.1375
VERPE	*Veronica persica*	0.4067	0.3242	0.1000	0.0440	0.5769

Table X2a: Estimated probabilities of seedbank states at *t*=0. Table X2a gives means and standard deviations of seedbank state probabilities at t=0 obtained under the Gibb's sampler. Y1 mean values (probability of absence of species per field) are highly negatively correlated (ρ= -0.94, p-value = 4.017e-09) to the observed field occurrence of species (table 1). This strong correlation reinforces the assumption of relatively good approximation of LHT.

Code BAYER	Y1	Y2	Y3	Y4	Y5	Y6
ALOMY	0.4462+/-0.0422	0.3411+/-0.0598	0.1512+/-0.0454	0.0137+/-0.0121	0.0086+/-0.0083	0.0389+/-0.0122
ANGAR	0.3241+/-0.0424	0.5111+/-0.0851	0.1372+/-0.0689	0.0064+/-0.0038	0.0099+/-0.0082	0.0109+/-0.0047
CHEAL	0.0329+/-0.0012	0.0090+/-0.0034	0.5226+/-0.0186	0.1086+/-0.0188	0.0098+/-0.0121	0.3169+/-0.0284
FUMOF	0.4660+/-0.0823	0.2832+/-0.1458	0.2120+/-0.1127	0.0175+/-0.0107	0.0149+/-0.0103	0.0061+/-0.0046
GALAP	0.1046+/-0.0503	0.1194+/-0.0953	0.4910+/-0.1403	0.1688+/-0.0950	0.0728+/-0.0395	0.0432+/-0.0290
MERAN	0.4742+/-0.0429	0.2752+/-0.0678	0.1891+/-0.0594	0.0201+/-0.0137	0.0070+/-0.0035	0.0341+/-0.0120
PAPRH	0.3418+/-0.0684	0.1817+/-0.1206	0.2562+/-0.1235	0.1732+/-0.1132	0.0177+/-0.0129	0.0292+/-0.0195
POAAN	0.4929+/-0.0403	0.2251+/-0.0421	0.1680+/-0.0371	0.0174+/-0.0115	0.0104+/-0.0091	0.0859+/-0.0206
POLAV	0.2788+/-0.0455	0.0781+/-0.0524	0.4135+/-0.0730	0.0842+/-0.0548	0.0657+/-0.0450	0.0795+/-0.0366
POLCO	0.3303+/-0.0558	0.3072+/-0.0391	0.2404+/-0.0493	0.0171+/-0.0047	0.0137+/-0.0030	0.0910+/-0.0152
SENVU	0.0449+/-0.0418	0.1860+/-0.0921	0.6125+/-0.0881	0.0470+/-0.0183	0.0324+/-0.0238	0.0770+/-0.0328
SINAR	0.4145+/-0.0485	0.4856+/-0.0715	0.0812+/-0.0458	0.0070+/-0.0040	0.0045+/-0.0035	0.0068+/-0.0044
SOLNI	0.2181+/-0.0368	0.0555+/-0.0466	0.5803+/-0.0474	0.0125+/-0.0062	0.0041+/-0.0025	0.1291+/-0.0316
SONAS	0.2529+/-0.0581	0.0515+/-0.0330	0.2978+/-0.1369	0.2680+/-0.1482	0.0988+/-0.0790	0.0306+/-0.0184
SONOL	0.5453+/-0.0649	0.2008+/-0.0719	0.1520+/-0.0973	0.0695+/-0.0547	0.0142+/-0.0063	0.0180+/-0.0078
STEME	0.3944+/-0.0494	0.1675+/-0.0871	0.2934+/-0.0677	0.0492+/-0.0361	0.0137+/-0.0095	0.0814+/-0.0205
VERHE	0.3295+/-0.0492	0.1198+/-0.0773	0.4496+/-0.0572	0.0469+/-0.0390	0.0157+/-0.0107	0.0382+/-0.0136
VERPE	0.3009+/-0.0587	0.1097+/-0.0815	0.4458+/-0.0886	0.1185+/-0.0688	0.0168+/-0.0111	0.0081+/-0.0045

Table X2b: Estimated life history traits. Means and standard deviations of Life History Traits (LHT) distribution obtained under the Gibb's sampler. The low values of germination rates are the outcome of the high dependence of the results to the range and the number of abundance classes of number of seeds in seedbank, which were arbitrary chosen for the calculation. Change of these classes will quantitatively modify the results, but not qualitatively. Comparisons of parameter estimates between species are pertinent, since species and crop types were treated by the same way.

Code BAYER	*s* WC	σ WC	φ WC	*s* OR	σ OR	φ OR	*s* M	σ M	φ M	*s* SF	σ SF	φ SF
ALOMY	0.8303+/-0.134	0.1936+/-0.038	1.2345+/-0.378	0.9180+/-0.047	0.1189+/-0.031	0.4827+/-0.388	0.6994+/-0.144	0.0746+/-0.025	0.4198+/-0.312	0.3063+/-0.183	0.0100+/-0.007	0.5103+/-0.316
ANGAR	0.8902+/-0.083	0.0139+/-0.009	1.1475+/-0.729	0.7120+/-0.135	0.0505+/-0.033	5.5352+/-5.342	0.9320+/-0.037	0.1635+/-0.050	0.1325+/-0.112	0.6394+/-0.225	0.1626+/-0.099	0.3927+/-0.286
CHEAL	0.8815+/-0	0.0039+/-0.001	1.0469+/-0.149	0.5772+/-0.048	0.0069+/-0.001	0.8712+/-1.632	0.8840+/-0.018	0.1568+/-0.004	0.6906+/-0.012	0.6025+/-0.018	0.0662+/-0.003	9.5478+/-4.251
FUMOF	0.9416+/-0.022	0.0464+/-0.019	1.2128+/-0.685	0.6224+/-0.207	0.0159+/-0.011	0.4430+/-0.292	0.5018+/-0.163	0.0159+/-0.009	1.4260+/-1.031	0.6862+/-0.163	0.0048+/-0.003	2.6401+/-1.576
GALAP	0.8761+/-0.075	0.0333+/-0.008	1.4125+/-0.598	0.9220+/-0.056	0.0213+/-0.007	2.9708+/-1.947	0.1079+/-0.043	0.0030+/-0.001	19.754+/-9.492	0.6702+/-0.178	0.0095+/-0.003	3.5396+/-1.511
MERAN	0.9454+/-0.035	0.0098+/-0.002	6.8831+/-4.716	0.6089+/-0.211	0.1569+/-0.035	3.2670+/-2.451	0.8724+/-0.080	0.2322+/-0.093	0.7089+/-0.259	0.6048+/-0.253	0.4056+/-0.192	3.1279+/-1.737
PAPRH	0.8586+/-0.080	0.0479+/-0.022	1.5847+/-0.779	0.4687+/-0.190	0.0254+/-0.013	0.7906+/-0.611	0.3737+/-0.198	0.0003+/-0.000	14.486+/-9.390	0.6828+/-0.170	0.0001+/-9.105	5.3969+/-3.282
POAAN	0.8506+/-0.100	0.1626+/-0.024	0.2907+/-0.219	0.5939+/-0.179	0.0355+/-0.020	0.0523+/-0.037	0.6448+/-0.140	0.0093+/-0.001	7.9713+/-4.492	0.4745+/-0.215	0.0138+/-0.005	1.7188+/-1.561
POLAV	0.8140+/-0.121	0.0199+/-0.005	0.3387+/-0.260	0.5518+/-0.189	0.0016+/-0.001	4.4775+/-2.555	0.7032+/-0.160	0.0523+/-0.015	0.8073+/-0.732	0.3388+/-0.117	0.0475+/-0.011	0.2877+/-0.200
POLCO	0.8737+/-0.065	0.0116+/-0.002	18.615+/-7.266	0.3563+/-0.056	0.0030+/-0.001	15.409+/-3.752	0.9620+/-0.033	0.0844+/-0.018	0.2212+/-0.065	0.8991+/-0.024	0.0276+/-0.031	0.0882+/-0.049
SENVU	0.6817+/-0.184	0.0098+/-0.002	6.7812+/-3.734	0.5457+/-0.241	0.0190+/-0.004	17.145+/-8.910	0.8838+/-0.052	0.0080+/-0.002	15.725+/-4.133	0.8824+/-0.046	0.0281+/-0.011	13.576+/-4.716
SINAR	0.8902+/-0.079	0.0707+/-0.017	2.5662+/-1.081	0.8141+/-0.141	0.1864+/-0.043	1.0109+/-0.508	0.5712+/-0.166	0.0518+/-0.017	0.6594+/-0.400	0.7191+/-0.163	0.1265+/-0.038	4.1502+/-2.150
SOLNI	0.5761+/-0.090	0.0003+/-0.001	3.5725+/-0.457	0.7725+/-0.186	0.0044+/-0.001	181.06+/-144.9	0.8256+/-0.094	0.1173+/-0.012	0.5923+/-0.063	0.8318+/-0.121	0.1305+/-0.011	0.1324+/-0.042
SONAS	0.6301+/-0.145	0.0001+/-0.001	12.776+/-8.903	0.4006+/-0.244	0.0273+/-0.009	23.616+/-12.00	0.9102+/-0.064	0.0144+/-0.004	1.0983+/-0.644	0.3755+/-0.224	0.0215+/-0.007	11.893+/-6.830
SONOL	0.9538+/-0.044	0.0032+/-0.001	30.814+/-16.18	0.3687+/-0.185	0.0058+/-0.003	30.156+/-15.48	0.6805+/-0.154	0.0066+/-0.006	6.5381+/-5.192	0.9247+/-0.032	0.0050+/-0.004	8.0308+/-4.901
STEME	0.8227+/-0.115	0.1166+/-0.014	0.0407+/-0.024	0.7087+/-0.175	0.0676+/-0.019	0.9026+/-0.237	0.8362+/-0.104	0.0164+/-0.004	1.0896+/-0.838	0.3828+/-0.191	0.0063+/-0.003	4.0681+/-3.717
VERHE	0.8838+/-0.087	0.1084+/-0.017	0.6555+/-0.455	0.5632+/-0.200	0.0175+/-0.003	2.2369+/-1.274	0.5149+/-0.166	0.0017+/-0.001	1.1207+/-0.778	0.7211+/-0.165	0.0056+/-0.003	8.0282+/-6.276
VERPE	0.8883+/-0.043	0.0560+/-0.014	0.4799+/-0.408	0.7993+/-0.152	0.0446+/-0.014	3.8367+/-2.598	0.5724+/-0.195	0.0076+/-0.002	5.0286+/-2.598	0.5901+/-0.191	0.0005+/-0.001	8.2883+/-4.958

Table X3: Percentage of field used in calculation of life history traits values. Values correspond to the proportion of fields where crop type (WC = winter cereals, OR = oilseed rape, M = maize and SF = sunflower) has been sown at least once and species has been observed at least once in the crop sequence. These values indicate the confidence that we can have in the estimation of life history traits for each couple (species, crop type). For example, *Digitaria sanguinalis* was never observed on fields where OR was sown, hence, the estimation of LHT in OR for this species is most probably erroneous. Bold indicates species that we kept for studies.

Code BAYER	WC	OR	M	SF
ALOMY	**0.39**	**0.19**	**0.12**	**0.10**
AMARE	0.31	0.05	0.25	0.17
ANGAR	**0.36**	**0.10**	**0.23**	**0.15**
APHAR	0.20	0.12	0.06	0.05
ATXPA	0.22	0.07	0.13	0.10
CAPBP	0.33	0.13	0.21	0.09
CHEAL	**0.62**	**0.16**	**0.41**	**0.27**
DIGSA	0.10	0.00	0.14	0.04
ECHCG	0.29	0.05	0.28	0.14
EPHHE	0.24	0.11	0.09	0.12
FUMOF	**0.30**	**0.11**	**0.11**	**0.11**
GALAP	**0.58**	**0.25**	**0.16**	**0.21**
GERDI	0.23	0.13	0.08	0.07
LACSE	0.15	0.05	0.03	0.08
LAMPU	0.18	0.08	0.09	0.06
MATCH	0.28	0.13	0.11	0.07
MERAN	**0.37**	**0.13**	**0.19**	**0.18**
PAPRH	**0.44**	**0.15**	**0.13**	**0.16**
POAAN	**0.34**	**0.10**	**0.20**	**0.10**
POLAV	**0.48**	**0.13**	**0.28**	**0.20**
POLCO	**0.40**	**0.13**	**0.20**	**0.19**
POLPE	0.34	0.09	0.25	0.16
RAPRA	0.24	0.11	0.08	0.10
SENVU	**0.56**	**0.20**	**0.32**	**0.24**
SINAR	**0.35**	**0.19**	**0.12**	**0.15**
SOLNI	**0.45**	**0.12**	**0.34**	**0.23**
SONAS	**0.41**	**0.16**	**0.28**	**0.16**
SONOL	**0.21**	**0.10**	**0.14**	**0.10**
STEME	**0.41**	**0.16**	**0.26**	**0.10**
TAROF	0.16	0.06	0.11	0.07
VERHE	**0.51**	**0.20**	**0.16**	**0.19**
VERPE	**0.49**	**0.18**	**0.16**	**0.17**

3.2.3. Note sur l'apport de l'approche HMM au cas adventice

L'utilisation de modèle de Markov Caché semble bien adaptée à la modélisation de la dynamique du stock de graines. Cette approche colle à l'existence de séries temporelles de flore levée permettant de quantifier les traits d'histoire de vie relatifs à la dynamique du stock de graines, sous différences conditions culturales. Le point fort de cette approche est donc de pouvoir extraire des informations relatives à une variable complètement cachée (i.e. le stock de graines) par l'utilisation seule d'une information partielle mais récurrente (la flore levée) et la conception d'un modèle adapté à rendre compte de la dynamique de graines au cours des successions.

Dans ce papier, nous avons aussi étudié, par simulations avec notre modèle et avec les paramètres estimés, l'effet d'une modification de l'assolement comme cela pourrait découler d'une politique d'aide publique à telle ou telle culture, sur la dynamique des espèces adventices. On pourrait imaginer d'autres applications qui simuleraient, par exemple, ce qui résulterait d'une modification de la durée de vie du stock si on était en mesure d'accroitre la prédation, ou encore de simuler les effets sur le stock d'une modification de la date de récolte affectant la quantité de semences venant réalimenter le stock.

4. DISCUSSION GENERALE ET PERSPECTIVES

Les travaux que j'ai menés durant cette thèse ont montré qu'une application judicieuse des concepts traditionnellement développés en écologie sur des données issues d'un réseau de monitoring de la diversité végétale peut permettre de fournir de nouvelles perspectives et connaissances sur la constitution des communautés adventices. Le « regard » écologique utilisé pour expliquer la diversité biologique observée dans le milieu naturel convient également pour expliquer la composition et variabilité des communautés biologiques évoluant dans un milieu fortement anthropisé. Seul l'environnement change, mais pas les processus et règles d'assemblages responsables de la constitution des communautés biologiques. Les principales divergences entre les écosystèmes non ou peu perturbés et les agro-écosystèmes résident dans la hiérarchie des filtres et des processus. Par exemple, il est admis que la capacité d'accueil d'un habitat permet, par des effets densités dépendances, de réguler la taille des populations. Ce processus est un processus majeur. A l'inverse, du fait de la gestion par l'Homme, la densité dépendance a un rôle moindre dans les parcelles cultivées qui apparaissent comme des milieux non saturés. C'est donc bien cette forte différence de l'environnement de la parcelle cultivée par rapport au milieu naturel qui doit conduire à réfléchir à l'appropriation des approches méthodologiques utilisées en écologie pour l'analyse des agro-écosystèmes.

4.1. Limites et problèmes du jeu de données Biovigilance pour l'Agroécologie

Bien qu'étant certainement la meilleure base de données d'un point de vue qualitatif et quantitatif car sans réel équivalent, la base de données Biovigilance souffre de faiblesses dont il faut avoir conscience avant de mener toute analyse. Je dresse ci-dessous la liste des problèmes qui me semblent majeurs sans que cela se veuille exhaustif. Cela pourra notamment alimenter la réflexion sur les évolutions à conduire pour qu'un réseau d'observatoires puisse servir de manière élargie. A partir du moment où il est conduit dans la durée, sur un vaste territoire et coûte logiquement assez cher, une utilisation facile et étendue est donc de toute importance. On sait par ailleurs que l'exercice de se projeter dans les besoins souvent peu ou mal exprimés n'est pas chose facile. J'ai rangé les différents problèmes rencontrés en trois classes :

Données manquantes et « faux zéros »

Les importantes différences de tailles que peuvent avoir les différentes tables ne sont pas uniquement dues à des différences de format de stockage mais aussi, malheureusement, a des « efforts d'échantillonnage » très différents. En effet, il apparaît que beaucoup d'informations sont manquantes pour certaines variables et, dues à l'oubli de modalités indiquant l'absence de certaines pratiques. On se retrouve alors souvent dans l'impossibilité de différencier une absence de mesure d'une absence de pratiques (exemple : dans le cas du travail du sol non indiqué, est-on en présence d'une vraie absence de travail du sol ou d'un oubli de notation ?). L'absence trop importante de certaines pratiques (travail du sol, traitements herbicides) suggère que la base contient une part importante de « faux zéros ». De plus, certaines variables (qui ne peuvent prendre de valeur nulle) sont malgré tout composées d'une majorité de données manquantes alors que leur poids dans la détermination des communautés est évident et considéré comme important (exemple : date de semis) (Tableau 2).

absence de données	
Date de semis	4053
Texture Sol	1278
Paysage voisinage parcelle	139
Localisation	226
Culture	518
absence de notation de pratiques	
Désherbage	3213
Labour	2910
Nombre de relevés	4855

Tableau 2 Exemple de données manquantes et problèmes de « faux zéros ». Sur un jeu de données extrait de la base de données 'Biovigilance Flore' composé de 4855 relevés, on peut distinguer les variables dont on sait que la donnée est absente puisqu'essentielle, des variables où l'absence de notation peut être interprétée comme une absence de la pratique mais pour laquelle on reste incapable de différencier les « vrais » des « faux zéros ». Par exemple, 3213 relevés sur 4855 ne comportent aucune information sur le traitement herbicide réalisé, il est difficile de croire que cette absence de notation signifie, pour tous les relevés, une absence de traitement.

Pour pallier ce problème, une solution consisterait à travailler le masque informatique de saisie, en proposant des menus déroulants couvrant les situations majeures et en laissant la place pour un 'autres' complété par une ligne alimentable sans contrainte. L'enregistrement définitif vers la base pourrait être bloqué tant que la page n'est pas complète.

Absence de certaines informations

La base n'ayant pas été prévue pour faire de l'agro-écologie (ou du moins pas à une échelle très fine), certaines variables qui auraient pu s'avérer très intéressantes pour nous, manquent cruellement. Ainsi, aucune information relative à la densité de semis ou au stade phénologique de la culture n'a été décrite. La culture étant le compétiteur principal de la parcelle, ces informations auraient permis par exemple, de mieux appréhender les interactions inter-adventices. Même si la notation des espèces adventices est réalisée sous un format semi-quantitatif (classe d'abondance) ce qui donne une idée de la taille de la population d'une espèce, l'absence de stade phénologique moyen ne permet pas de préciser son état. Cette limite est importante puisqu'elle conduit à comptabiliser de la même manière une levée massive de plantules, dont la majorité mourra en quelques jours, et des plantes au stade floraison qui pourraient boucler leur cycle.

Ce problème est profond. Une solution passe par la réflexion et la co-construction en amont de ce qui est important. Ceci nécessite de réunir une expertise étendue, souvent au-delà de la sphère que l'on imagine initialement.

Une échelle d'échantillonnage très large

La présence d'adventices et leur répartition souvent en « patch » à petite échelle peut s'avérer être souvent aléatoire. Afin de capturer le maximum d'espèces et de représenter au mieux la diversité d'une parcelle, l'échelle d'échantillonnage a été choisie relativement grande, environ 200m². Ainsi, si l'on tente d'analyser la structuration des communautés et les mécanismes d'assemblages, on se retrouve à considérer un ensemble d'espèces qui peuvent dans les faits être très éloignées dans l'espace. Néanmoins, nous pensons que les processus d'assemblage locaux peuvent se ressentir à une échelle aussi large, mais il faut garder à l'esprit que l'échelle peut venir bruiter les analyses dans le cadre d'études basées sur la structuration. Une telle échelle est de toute façon peu compatible avec le principe même de communauté. D'ailleurs cela nous a souvent conduit à parler de 'flore levée' ou de 'relevé' plutôt que de communauté.

Pas de bonne solution à ce problème puisque chaque échelle a ses avantages et ses inconvénients. Plusieurs protocoles sont toutefois basés sur des inventaires emboîtés du type 10 fois 4m², par exemple qui permettent de capturer différents 'grains'.

4.2. Notion de communautés

La réflexion conduite lors de ces différents travaux nous amène donc à réfléchir sur la notion de communauté adventice. En effet, les flores levées semblant peu interagir, en raison des fortes perturbations anthropiques et à la « remise à zéro » annuelle du couvert végétal, la notion de communautés biologiques (i.e. interaction d'organismes partageant un environnement commun) ne semblent peut être pas adaptée au cas adventice. Néanmoins, les stratégies de contrôle efficaces mises en place sont très certainement seules responsables du peu de compétition inter-adventice observée et les espèces adventices restent tout de même potentiellement en concurrence pour la ressource. De plus, la communauté biologique peut être définie au sens plus large comme un ensemble de populations d'espèces différentes présentant une unité fonctionnelle commune, c'est-à-dire leur capacité à être « adventice ».

Au-delà d'être adaptées aux conditions du milieu au moment de leur levée (i.e. au moment du semis de la culture), les espèces adventices sont aussi adaptées à la rotation culturale (motif de succession de cultures répété dans le temps), permettant la spécialisation de certaines espèces pour des conditions rares (i.e. peu fréquentes dans le temps) mais qui leur sont alors très profitables. De plus, la flore levée est directement liée au stock de graines présent. Ainsi, la notion de communautés devrait-elle plutôt s'appliquer au stock de graines ? Cela complique alors la notion de communautés puisque les espèces se retrouvent en interaction par sous-ensembles levant plus ou moins en même temps. La dynamique cyclique (i.e. répétition des mêmes séquences de culture au cours du temps) impose de réfléchir à la définition précise que l'on peut ou que l'on souhaite donner à une communauté adventice.

4.3. Apport à l'agroécologie et au cas adventice

L'apport de cette thèse à l'agroécologie peut être regroupé en deux points.

4.3.1. Prise en compte de l'abondance

La prise en compte de l'abondance dans l'assemblage et la dynamique des espèces adventices est une question centrale pour mieux quantifier les structurations et les processus mis en jeux, et a ainsi été abordée comme un des axes centraux des travaux réalisés au cours de cette thèse. L'assemblage non-aléatoire observé des notes d'abondance suggère que les travaux visant à expliquer l'abondance des différentes espèces levées au sein d'une parcelle doivent être approfondis et peuvent offrir certaines perspectives dans l'identification des « leviers » de gestion permettant de réduire la pression exercée par la communauté adventice.

De plus, réaliser des analyses au niveau de l'abondance est très certainement le seul moyen de différencier une présence fortuite d'une vraie présence résultant d'une accumulation de conditions suffisamment favorables. En traquant les règles sur l'abondance on espère se mettre en position de comprendre.

4.3.2. Des modèles adaptés

L'apport de la modélisation réalisée au cours de cette thèse se résume donc en deux approches statistiques visant à mieux comprendre l'assemblage et la dynamique des flores adventices.

Le choix du modèle nul le plus adapté à la question abordée a représenté un point de toute importance et a conduit au développement d'un modèle nul adapté aux données de type semi-quantitatif. Ce modèle nul permet ainsi de mieux appréhender les assemblages, au niveau de l'abondance, des communautés adventices et semble apporter une réelle plus-value dans les analyses de dispersion des traits fonctionnels des espèces adventices (par l'utilisation de données semi-quantitatives).

La modélisation par modèle de Markov caché représente la deuxième approche mise en place et semble être particulièrement adaptée au cas adventice pour analyser et extraire un maximum d'informations des séries de flore observée. Cette approche, modélisant le stock de graine permet ainsi d'appréhender les réels effets du système de culture sur la flore amenée à lever. Bien que seules les relations espèces-environnement aient été abordées au sein d'une seule parcelle, une modélisation sur graphe permettrait théoriquement de considérer soit le voisinage parcellaire et les processus de dispersion, soit les espèces co-occurrentes et ainsi de considérer les processus d'interactions biotiques inter-adventices. Néanmoins, cette première approche simple de la dynamique offre déjà certaines perspectives sur l'identification des leviers de gestion potentiels.

4.4. Perspectives

Les travaux réalisés durant cette thèse ouvrent de nouvelles perspectives sur l'exploitation des données issues du réseau 'Biovigilance Flore', mais aussi de manière plus générique, sur l'exploitation de données semi-quantitatives et la description des dynamiques des espèces adventices au sein de rotations culturales.

Certains oublis ou négligences dans la composition de la base de données ont souvent été problématiques et nous ont conduits à devoir « tempérer » nos conclusions par manque de puissance ou d'accès à certaines informations. Ainsi les quelques suggestions énumérées dans la section 4.1 (« Limites et problèmes du jeu de données Biovigilance pour l'agroécologie ») permettraient très certainement de pouvoir plus appuyer nos conclusions, particulièrement celles relatives aux questions de compétition entre adventices. Dans l'état actuel de la base de données, l'étendue des analyses et des études de type écologique qui peuvent encore être explorées reste néanmoins importante. Je dresse ainsi quelques exemples d'études qui pourraient s'avérer intéressantes ; cette liste, loin d'être exhaustive, ne représente pas non plus nécessairement les études les plus pertinentes mais plutôt celles qui me porteraient à cœur :

- Une étude combinant les deux grandes approches étudiées (assemblage nul et effet filtre habitat) pourrait s'avérée très pertinentes. Au-delà d'une randomisation par stratification basée uniquement sur les cultures, une description de la niche des espèces en amont permettrait de calculer des probabilités d'occurrences par site. Ces probabilités nous permettraient de réaliser des assemblages nuls plus en accord avec la réalité environnementale et ainsi de plus se concentrer sur les éventuels interactions biotiques entre espèces adventices.

- Une interpolation des répartitions des différentes espèces à l'échelle du territoire français pour permettre la constitution de carte de répartitions. Cette interpolation pourrait être affinée par des variables biogéographiques, pédoclimatiques et culturales ; et pourrait permettre d'appréhender l'évolution de la flore adventice à l'échelle du territoire sous des scénarios de changements climatiques ou de modifications de l'agriculture.

- Un modèle prédictif simple, disponible sur une page internet, de flore levée en fonction d'une description sommaire des pratiques réalisées pourrait être aussi un bon retour pour le réseau d'observation. L'agriculteur ou le technicien assurant le suivi pourrait également renseigner les espèces dominantes (ou ayant posées problèmes) les *n* années précédentes, permettant ainsi d'affiner les prédictions et les risques potentiels.

- S'intéresser à la manière dont on pourrait traiter les données manquantes et le problème de « faux zéros » permettrait très certainement d'exploiter au mieux les données actuelles. Au-delà de la solution qui consisterait à contacter tous les organismes et personnes impliquées dans la récolte et la saisie des données (qui n'est certainement pas réalisable par un doctorant), certaines méthodes statistiques sont connues pour permettre le remplissage des données manquantes et l'identification des vraies absences de certaines pratiques. Je n'ai ainsi aucune solution ou méthode à proposer actuellement ; mise à part peut être l'utilisation de Chaîne de Markov Cachée où la notation représenterait une information partielle de l'état

réelle de la variable. Cette idée me vient bien évidement par affinité (et connaissance) pour la méthode, mais elle n'en reste pas moins, à mes yeux, complexe et difficile à mettre en place dans l'état actuel des connaissances de l'effet réel des pratiques sur la dynamique du stock de graines.

- Enfin, le pouvoir prédictif de la flore observée à décrire l'état du milieu n'a pas du tout été abordée au cours de cette thèse et n'en reste pas moins très intéressante. Le recouvrement des niches des différentes espèces présentes en un site peut très certainement nous renseigner précisément sur l'état réel du milieu et ceci mieux que chaque espèce prise isolément. Cette approche pourrait aussi bien apporter de nouvelles informations dans la manière de compléter les données manquantes, mais pourrait aussi servir de base dans un modèle prédictif l'abondance totale d'adventices attendue en un point.

5. REFERENCES BIBLIOGRAPHIQUES

- Albert C.H., Thuiller W., Yoccoz N.G., Soudant A., Boucher F., Saccone P. & Lavorel S. (2010) Intraspecific functional variability: extent, structure and sources of variation. *Journal of Ecology* **98 :** 604-613.
- Altieri M.A. (1995) Agroecology: The Science of Sustainable Agriculture. Westview Press. London. United-Kingdom.
- Barralis G. (1976) Méthode d'étude des groupements adventices des cultures annuelles. *INRA - 5ème Colloque International sur l'Ecologie et la Biologie des Mauvaises Herbes.* Dijon, France 59-68.
- Barralis G. & Chadoeuf R. (1988) Relation entre flore potentielle et flore réelle des champs cultivés. *VIIIème colloque International sur l'Ecologie et la Biologie des Mauvaises herbes*. Dijon, France 43-52.
- Belyea L.R. & Lancaster J. (1999) Assembly rules within a contingent ecology. *Oikos* **86** : 402-416.
- Benjankar R., Egger G., Jorde K., Goodwin P. & Glenn N.F. (2011) Dynamic floodplain vegetation model development for the Kootenai River, USA. *Journal of Environmental Management* **92**(12) :3058-3070.
- Blanco C.C., Sosinski E.E., Santos Jr. B.R.C., da Silva M.A. & Pillar V.D. (2007) On the overlap between effect and response plant functional types linked to grazing. *Community Ecology* **8**:57–65.
- Bohan D.A., Boursault A., Brooks D.R. & Petit S. (2011) National-scale regulation of the weed seedbank by carabid predators. *Journal of Applied Ecology* **48**(4): 888–898.
- Booth B.D. & Swanton C.J. (2002) Assembly theory applied to weed communities. *Weed Science* **50** : 2-13.
- Botta-Dukát Z. (2005) Rao's quadratic e*ntropy as a measure of functional d*i**ve**rsity based on multiple traits. *Journal of Vegetation Science* **16**:533-540.
- Brown J.H. (1984) On the relationship between abundance and distribution of species. *American Naturalist* **124** : 255-279.
- Cam E. & Monnat J.Y. (2000) Stratification based on reproductive state reveals contrasting patterns of age-related variation in demographic parameters in the kittiwake. *Oikos* **90** :560-574.
- Caswell H. (2001) Matrix population models: Construction, analysis and interpretation, 2nd Edition. Sinauer Associates. Massachusetts, USA.
- Chadès I., McDonald-Madden E., McCarthy M.A., Wintle B., Linkie M. and Possingham H.P. (2008) When to stop managing or surveying cryptic threatened species. *PNAS* **105**(37) :13936-13940.
- Chessel D., Dufour A.B. & Thioulouse J. (2004) The ade4 package-I- One-table methods. *R News*. 4: 5-10.
- Colwell R.K. & D.W. Winkler. (1984) A null model for null models in biogeography. pp. 34-359 in: *Ecological Communities: Conceptual Issues and the Evidence. D. R.* Strong Jr., Simberloff D., Abele L.G. and Thistle A.B..

- Connor E.F. & Simberloff D. (1979) The assembly of species communities: Chance or competition? *Ecology* **60**: 1132-1140.
- Cornwell W.K & Ackerly D.D. (2009) Community assembly and shifts in plant trait distributions accross an environmental gradient in coastal Carlifornia. *Ecological Monographs* **79**(1): 109-126.
- Cornwell W.K & Ackerly D.D. (20010) A link between plant traits and abundance: evidence from woody plants in coastal California. *Journal of Ecology* **98**: 814-821.
- De Deyn G.B., Cornelissen J.H.C., & Bardgett R.D. (2008) Plant functional traits and soil carbon sequestration. *Ecology Letters* **11**:516–531.
- Debaeke P. (1988) Dynamique de quelques dicotylédones adventices en culture de céréales. I. Relation flore levée-stock semencier. *Weed Research* 28:251-263.
- Delos M., Hervieu F., Folcher L., Micoud A. & Eychenne N. (2006) La «Biovigilance», des OGM au général. Exemple du suivi des grandes cultures en France. *Phytoma-LDV* **589**: 44-48.
- Dolédec S., Chessel D., Ter Braak C.J.F. & Champely S. (1996) Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* **3**: 143-166.
- Doré T., Makowski D., Malézieux E., Munier-Jolain N., Tchamitchian M. & Tittonell P. (2011) Facing up to the paradigm of ecological intensification in agronomy: Revisiting methods, concepts and knowledge. *European Journal of Agronomy* **34**(4) : 197-210.
- Dray S., Dufour A.B. & Chessel D. (2007) The ade4 package-II: Two-table and K-table methods. *R News*. 7(2): 47-52.
- Fried G. (2007) Variations spatiales et temporelles des communautés adventices des cultures annuelles en France. Thèse de docteur en 3ème cycle, Université de Bourgogne, Dijon, France.
- Fried G., Norton L.R. & Reboud X. (2008) Environmental and management factors determining weed species composition and diversity in France. *Agriculture, Ecosystems & Environment*, **128**(1-2) : 68-76.
- Fried G., Petit S., Dessaint F., & Reboud X. (2009a) Arable weed decline in Northern France: Crop edges as refugia for weed conservation? Biological Conservation 142(1) : 238-243.
- Fried G., Chauvel B. & Reboud, X. (2009b) A functional analysis of large-scale temporal shifts from 1970 to 2000 in weed assemblages of sunflower crops in France. *Journal of Vegetation Science*, **20**: 49–58.
- Fried G., Petit S. & Reboud X. (2010) A specialist-generalist classification of the arable flora and its response to changes in agricultural practices. *BMC Ecology* **10:**20.
- Gales M. & Young S. (2007) The Application of Hidden Markov Models in Speech Recognition. Foundations and Trends_ in Signal Processing 1(3) : 195-304.
- Gardarin A., Guillemin J.P., Munier-Jolain N.M. & Colbach N. (2010) Estimation of key parameters for weed population dynamics models : Base temperature and base water potential for germination. *European Journal of Agronomy* **32**(2) : 162-168.

- Garrabou J., Ballesteros E. & Zabala M. (2002) Structure and Dynamics of Noth-western Mediterranean Rocky Benthic Communities along a Depth Gradient. *Estuarine, Coastal and Shelf Science* **55**(3) :493-508.
- Gardarin A., Dürr C. & Colbach N. (2011) Prediction of germination rates of weed species : Relationships between germination speed parameters and species traits. *Ecological Modelling* **222**(3) : 626-636.
- Garnier E., Cortez J., Billèss G., Navas M.-L., Roumet C., Debussche M., Laurent G., Blanchard G., Aubry D., Bellmann A., Neill C. & Toussaint J.-P. (2004) Plant functional markers capture ecosystem properties during secondary succession. *Ecology* **85**:2630–2637.
- Garnier E. & Navas M.-L. (2011) A trait-based approach to comparative functional plant ecology : concept, methods and applications for agroecology. A review. *Agronomical sustainable development*.
- Godinho I. (1984) Les définitions d' "adventice" et de "mauvaise herbe". *Weed Research* **24** : 121-125.
- Gotelli N.J. & Enstminger G.L. (2001) Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia* **129** : 281-291.
- Gotelli N.J. & Graves G.R. (1996) Null Models in Ecology. Smithsonian Institution Press, Washington, DC.
- Gotelli N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology* **81**: 2606-2621.
- Grime J.P. (1998) Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology* **86**(6): 902–910.
- Grime J.P. (2006) Trait convergence and trait divergence in herbaceous plant communities: Mechanisms and consequences. *Journal of Vegetation Science*, 17: 255–260.
- Gunton R.M., Petit S. & Gaba S. (2011) Functional traits relating arable weed communities to crop characteristics. *Journal of Vegetation Science* **22**: 541–550.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965-1978.
- Hill M. O. & Smith A.J.E. (1976) Principal component analysis of taxonomic data with multistate discrete c*haracters. Taxon*, **25** : 249-255.
- Hirze**l** A.H. & Le Lay G. (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45** : 1372–1381.
- Hubbell S.P. (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, New Jersey, USA.
- Hugueny B. & Cornell H. (2000) Predicting the relationship between local and regional species richness from a patch occupancy dynamics model. *Journal of Animal Ecology* **69**(2) : 194–200.
- Hutchinson G.E. (1957) Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology*, **22 :** 415–427.

- Jauzein P. (2001) Biodiversité des champs cultivés : l'enrichissement floristique. *Dossier de l'environnement de l'INRA* **21** : 43-64.
- Jung V., Violle C., Mondy C., Hoffmann L. & Muller S. (2010) Intraspecific variability and trait-based community assembly. *Journal of Ecology* **98** : 1134-1140.
- Keddy P.A. (1983) Shoreline Vegetation in Axe Lake, Ontario: Effects of Exposure on Zonation Patterns. *Ecology* **64**(2) : 331–344.
- Knapp R.A., Matthews K.R. & Sarnelle O. (2001) Resistance and resilience of Alpine lake fauna to fish introductions. *Ecological Monographs* **71**(3) : 401-421.
- Kropff M.J. & Spitters C.J.T. (1991) A simple-model of crop loss by weed competition from early observations on relative leaf-area of the weeds. *Weed Research* **31**(2): 97-105.
- Lavorel S. & Garnier E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology* **16** : 545-556.
- Legendre P., Galzin R. &Harmelin-Vivien M.L. (1997) Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* **78**: 547-562.
- Loreau M., Mouquet N. & Gonzalez A. (2003) Biodiversity as spatial insurance in heterogeneous landscapes. *PNAS* **100**(22): 12765-12770.
- Lotka A.J. (1925) Elements of physical biology. Williams & Wilkins Co. Baltimore, USA.
- MacArthur R.H. (1962) Some generalized theorems of natural selection. *PNAS* **48**: 1893–1897.
- MacArthur R.H. & Wilson E.O. (1963) An equilibrium theory of insular zoogeography, *Evolution* **17**(4) : 373-387.
- MacArthur R.H. & Wilson E.O. (1967) *The Theory of Island Biogeography*, Princeton University Press, New Jersey, USA.
- MacArthur R.H. (1972) Geographical Ecology: Patterns in the Distribution of Species. Harper & Row, New York.
- Mason N.W.H., Mouillot D., Lee W.G. & Wilson J.B. (2005) Functional richness, functional evenness and functional divergence : the primary components of functional diversity. *Oikos* **111** :112-118.
- Meyn S.P. & Tweedie R.L. (1993) Markov chains and stochastic stability. Springer-Verlag, London.
- McGill B.J., Enquist B.J., Weiher E. & Westoby M. (2006) Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21** : 178–185.
- Mouchet M.A., Villéger S., Mason N.W.H. and Mouillot D. (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Functional Ecology, **24**: 867–876.
- Muneepeerakul R., Bertuzzo E., Lynch H.J., Fagan W.F., Rinaldo A. & Rodriguez-Iturbe I. (2008) Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. *Nature* **453:** 220-222.
- Myers R.A. & Worm B. (2003) Rapid worldwide depletion of predatory fish communities. *Nature* **423** : 280-283.

- Nylin S. & Gotthard K. (1998) Plasticity in life-history traits. *Annual Review of Entomology* **43**:63–83.
- Oberdorff T., Hugueny B., Compin A. & Belkessam D. (1998) Non-interactive fish communities in the coastal streams of North-western France. *Journal of Animal Ecology* **67**(3): 472–484.
- Pakeman R.J. (2011a) Multivariate identification of plant functional response and effect traits in an agricultural landscape. *Ecology* **92**(6) :1353–1365.
- Pakeman R.J. (2011b) Functional diversity indices reveal the impacts of land use intensification on plant community assembly. *Journal of Ecology* **99**(5): 1143–1151.
- Petchey O.L. & Gaston K.J. (2006) Functional diversity: back to basics and looking forward. *Ecology Letters* **9**: 741–758.
- Petit S., Boursault A., Le Guilloux M., Munier-Jolain N. & Reboud X. (2011) Weeds in agricultural landscape, A review. *Agronomy for sustainable development* **31**(2) : 309-317.
- Peyrard N., Sabbadin R., Lô-Pelzer E. & Aubertot J.N. (2007) A graph-based Markov decision process framework for optimising collective management of diseases in agriculture: application to blackleg on canola *Proceeding of the 17th International Congress on Modelling and Simulation,* Christchurch, New Zealand.
- Preston F.W. (1962) The canonical distribution of commonness and rarity: parts I and 2. *Ecology* **43**:185–215,410–432.
- Rabiner L.R. (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE **77**(2) : 257-286.
- Robertson M.P., Caithness N. & Villet M.H. (2001) A PCA-Based Modelling Technique for Predicting Environmental Suitability for Organisms from Presence Records. *Diversity and Distributions* **7**(1) : 15-27.
- Roughgarden J. (2009) Is there a general theory of community ecology? *Biology Philosophy* **24**(4) : 521-529.
- Russo S.E., Davies S.J., King D.A. & Tan S. (2005) Soil-related performance variation and distributions of tree species in a bornean rain forest. *Journal of ecology* **93**: 879–889.
- Sagarin R.D. & Gaines S.D. (2002) The 'abundant centre' distribution: to what extent is it a biogeographical rule? *Ecology Letters* **5** : 137-147.
- Schamp B.S., Chau J. & Aarssen L.W. (2008) Dispersion of traits related to competitive ability in an old-field plant community. *Journal of Ecology* **96**: 204-212.
- Schleuter D., Daufresne M., Massol F. & Argillier C. (2010) A user's guide to functional diversity indices. *Ecological Monographs* **80**(3) :464-484.
- Schoener T.W. (1974) Resource partitioning in ecological communities. *Science* **185**:27-39.
- Shirtliffe S.J. & Entz M.H. (2005) Chaff collection reduces seed dispersal of wild oat (Avena fatua) by a combine harvester. *Weed Science* **53**:465-470.
- Simberloff D. (1983) Competition theory, hypothesis-testing, and other community ecological buzzwords. *American Naturalist* **122**:626-635.

- Simberloff D. (2004) Community ecology: is it time to move on? *American Naturalist* **163**(6) : 787-799.
- Stearns S.C. (1992) The Evolution of Life Histories. Oxford University Press, New York, USA.
- Stubbs W.J. & Wilson J.B. (2004) Evidence for limiting similarity in a sand dune community. *Journal of Ecology* **92** :557-567.
- Sutcliffe O.L. & Kay Q.O.N. (2000) Changes in the arable flora of central southern England since the 1960s. *Biological Conservation* **93:** 1-8.
- Sutherland S. (2004) What makes a weed a weed: life history traits of native and non indigenous plants in the USA. *Oecologia*, **141:** 24-39.
- Ulrich W. & N.J. Gotelli. (2010) Null model analysis of species associations using abundance data. *Ecology* **91**:3384-3397.
- Venable D.L. & Lawlor L. (1980) Delayed germination and dispersal of desert annuals: escape in space and time. *Oecologia* **46**: 272–282.
- Villéger S., Mason N.W.H. & Mouillot D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* **89**:2290-2301.
- Violle C., Navas M.-L., Vile D., Kazakou E., Fortunel C., Hummel I. & Garnier E. (2007) Let the concept of trait be functional! *Oikos* **116** : 882-892.
- Volterra V. (1926) Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. Mem. R. Accad. Naz. dei Lincei. Ser. VI, vol. 2.
- Weiher E. & Keddy P.A. (1995) Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* **74** :159-164.
- Weiher E., & Keddy P.A. (1999) Ecological assembly rules, perspectives, advances, retreats. Cambridge University.
- Wezel A. & Soldat, V. (2009) A quantitative and qualitative historical analysis of the scientific discipline agroecology. *International Journal of Agricultural Sustainability* **7** (1): 3-18.
- Whittaker R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs* **26**:1-80.
- Willott S.J. & Hassall M. (1998) Life-history responses of British grasshoppers (Orthoptera: Acrididae) to temperature change. *Functional Ecology* **12**: 232–241.

6. ANNEXE

Tableau X1 Description et taille des 10 tables composant la base de données Biovigilance.

Table	Nb de champs	Nb de lignes	Description
dbo_APARC_MATORG	6	1094	Apport matière organique
dbo_APARC_STE	17	1380	Pédologie
dbo_v_bordures	10	2800	Description des bordures de champs
dbo_v_cult_vois	8	12559	Cultures voisines
dbo_v_engrais	10	2694	Apport engrais
dbo_v_environnement	5	1380	Surface zones échantillonnées
dbo_v_obs_flore	46	410879	Relevés floristiques + type de cultures
dbo_v_pc_mecanique	9	6399	Travail mécanique
dbo_v_sensibilite	8	21	Impact herbicide sur la culture
dbo_v_trait_met	17	7790	Traitements phytosanitaires

Tableau X2 Description des champs de chaque table de la base de données Biovigilance. Les descriptions sont principalement issues de dires d'experts et/ou par interprétation du nom du champ, expliquant l'absence d'information pour certaines champs (?).

Table	Nom de Champ	Description du Champ
dbo_APARC_MATORG	PARC_CDN	Code interne de la parcelle
	MO_ANNEE	Année de l'apport
	MO_HUMIFERE_ON	Humifère (Oui Non)
	MO_TAUXMO	Taux de matière organique
	MO_PH	Taux de ph
	MO_TAUXCALCAIRE	Taux de calcaire
dbo_APARC_STE	PARC_CDN	Code interne de la parcelle
	SOL_NOM_LOCAL	Nom local du sol
	SOL_TEXTURE	Texture du sol
	SOL_HUMIFERE_ON	Humifère (Oui Non)
	SOL_TAUX_MO	Taux de matière organique
	SOL_ANNEE_MO	Année de l'apport en matière organique
	SOL_PROFONDEUR	Code profondeur du sol
	SOL_ACIDITE	Code acidité du sol

	SOL_PH	ph du sol
	SOL_ANNEE_PH	année de mesure du ph
	SOL_DRAINAGE	Code drainage du sol
	SOL_ARGIL_PC	Pourcentage d'argile
	SOL_SABLE_PC	Pourcentage de sable
	SOL_LIMON_PC	Pourcentage de limon
	TOPO_TYPE	Code type de sol
	TOPO_CMT	Commentaire
	SOL_TAUX_CALCAIRE	Taux de calcaire
dbo_v_bordures	PARC_CDN	Code interne de la parcelle
	ENV_DATE	Date à laquelle à été noté l'environnement
	ANNEE	Année du relevé
	TEMOIN_ON	Parcelle témoin (« O ») ou non (« N »)
	BORD_NO	chaque numero correspond à côté du polygone parcelle
	BORD_TYPE	Type de bordure externe du champ
	BORD_PC	% représenté par le côté du polygone
	BORD_ENT	Mode d'entretien de la bordure

	BORD_CMT	Commentaire
	BORD_DIST_TM	Distance de la zone témoin à la bordure du champ
dbo_v_cult_vois	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	ANNEE	Année du relevé
	OBS_NO	Numéro correspondant à un des côtés du champ
	PARC_CDN	Code interne de la parcelle
	TABL_CDN	Code du tableau
	VIVSYN_CDN	?
	CULTURE_LB	Culture présente sur les parcelles voisines de la parcelle Biovigilance
dbo_v_engrais	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	TYPE_ENGRAIS	Type d'apport
	PARC_CDN	Code interne de la parcelle
	TABL_CDN	Code du tableau
	APPORT	Quantité d'apport
	STARTER_ON	Engrais starter (Oui Non)

	FORME	Forme d'apport
	LOCAL_ON	Localisé (Oui Non)
	ANNEE	Année du relevé
dbo_v_environnement	PARC_CDN	Code interne de la parcelle
	ENV_DATE	Date du relevé de l'environnement de la parcelle
	ANNEE	Année du relevé
	SURF_PARCELLE	Surface de la parcelle (m²)
	SURF_TEMOIN	Surface de la zone témoin (m²)
dbo_v_obs_flore	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	PARC_CDN	Code interne de la parcelle
	ANNEE	Année du relevé
	REGION	Code de la région
	CAMPAGNE	Année de campagne
	OBS	Code Observateur
	OBSERVATEUR	Libellé observateur
	DEPT	Code département

	DEPARTEMENT	Libellé du département
	COM	Code INSEE de la commune
	COMMUNE	Libellé de la commune
	PARCELLE	Code parcelle
	PARC_LIEU	Lieu-dit de la parcelle
	TEMOIN_ON	Parcelle témoin (« O ») ou non (« N »)
	PLACETTE	Nom de la placette (Témoin ou parcelle)
	VEG	Code interne du végétal = culture
	VEGETAL	Libellé du végétal = culture
	CODE_BAYER	Code Bayer Culture
	VAR	Code interne de la variété
	VARIETE	Libellé de la variété
	VEGI	Code Interculture
	VEGETAL_INTER	Interculture
	PAR	Code interne du parasite
	PAR_BAYER	Code Bayer du parasite (Uniquement si OEPP)
	PARASITE	Libellé du parasite

	MESURE	Type de Mesure (CP: Comptage Parasite, ECHR: Echantillon réel, ECHT: Echantillon théorique, NOTE: notations de 0 à VMAX, P/A: Existence (Prés/Abs), PCM: Pratique Culturale mécanique, PI: Piègeage, PH: Phénologie, TR: Traitement chimique, CMT: Commentaires, EL: Elevage, OBS: Observation quelconque, MET: Météo, NOTA: Notation densité adventices
	REL_NO	Numéro d'observation (101 = 1ère Observation, 102 = 2e …)
	DIFFDATE	Nombre de jour entre la date du relevé et la date du jour (=date de l'extraction de la base?)
	OBS_VAL	Valeur du relevé
	ECHANTILLON	Echantillon du relevé
	VMAX	Valeur maximum du relevé
	ORGANE	Organe
	GROUPE	Groupe de l'observation (ici, toujours : BIOVIG)
	OBS_VAL2	2e valeur de relevé (ex des échantillons théoriques)
	OBS_VAL3	3e valeur d'observation
	OBS_VAL4	4e valeur d'observation
	TABL_LB	Libellé du tableau
	TABL_GRP	Groupe du tableau (ici, toujours :BIOVIG)

	GEST	Code du gestionnaire
	GRP2	Regroupement N°2
	GRP3	Regroupement N°3
	ACTIF_ON	Observation Active ou non
	GEST_OBR_TYPO	?
	PARC_TYPO	?
	OBS_TYPO	?
dbo_v_pc_mecanique	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	OBS_NO	ce numéro est incrémenté si à la même date il y a le même TYP_OUTIL mais dans le détail pas le même OUTIL
	PARC_CDN	Code interne de la parcelle
	TABL_CDN	Code du tableau
	TYP_OUTIL	Grande catégorie d'outils (dents, disques, labour, bineuses, etc.)
	OUTIL	Description précise de l'outil (difficile, mélange entre marque et type d'outil)
	PROFONDEUR	Profondeur de travail du sol (en cm)
	ANNEE	Année du relevé

dbo_v_sensibilite	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	ANNEE_ENCOURS_ON	Impact des traitements de l'année N ou N-1 ?
	SENS_TYPE	Type de symptôme observé sur la culture (jaunissement, nanisme, etc.)
	PARC_CDN	Code interne de la parcelle
	TABL_CDN	Code du tableau
	SENS_CLASSE	?
	ANNEE	Année du relevé
dbo_v_trait_met	POBS_CDN	Code interne de l'observation
	OBS_DATE	Date du relevé
	ANNEE	Année du relevé
	OBS_NO	Numéro incrémenté si plusieurs traitements à la même date
	PARC_CDN	Code interne de la parcelle
	TABL_CDN	Code du tableau
	INT_AMM_RFN	cela doit correspondre à un code par produit phyto
	PRODUIT	nom commercial du produit phytos
	DOSE	dose

	UNIT	21=L/HA ; 23=KG/HA ; 30=L/Q ; 42=g/Q voir avec J GASQUEZ!
	NON_HOMOLOG	?
	LOCAL_ON	?
	TMIN	conditions météo lors des traitements, utilisées uniquement pour expliquer si un traitement a raté
	TMAX	?
	PLUIE	?
	HYGRO	?
	VENT	?

Tableau X3 Origine et Description des données utilisées pour générer les axes synthétiques de l'environnement pédoclimatique et culturale utilisés pour modéliser la niche écologique des espèces.

Tableaux données	Origine	Nb variables	Type	Nb axes choisis
Culture	Biovigilance	15	binaire	2
Précédent	Biovigilance	16	binaire	2
Interculture	Biovigilance	15	binaire	2
Date de semis	Biovigilance	1	numérique	1
Coordonnées	Biovigilance	2	numérique	2
Structure sol	Biovigilance	3	numérique	2
pH	Biovigilance	1	numérique	1
Précipitation mensuelle	Worldclim	12	numérique	2
Température mensuelle min	Worldclim	12	numérique	2
Tempéraure mensuelle moyenne	Worldclim	12	numérique	2
Température mensuelle max	Worldclim	12	numérique	2
Préparation sol	Biovigilance	1	binaire	1
Outil à disques	Biovigilance	3	binaire	2
Outil à dents	Biovigilance	3	binaire	2
Travail du Sol	Biovigilance	3	binaire	2
Total		**111**		**27**

6.1.1. *Article SwapClass (version pré-soutenance)*

Cette section est présentée sous forme d'article (version en préparation, en vue d'une soumission à la revue *Ecological Modeling*)

SWAPCLASS: A NULL MODEL ADAPTED TO ABUNDANCE CLASSES DATASETS IN ECOLOGY

Benjamin BORGY[1], Xavier REBOUD[1] & Sabrina GABA[1]*

[1] INRA, UMR 1210, Biologie et Gestion des Adventices, 17 rue Sully, 21000 Dijon, France.

**Corresponding author :* Sabrina Gaba, UMR Biologie et Gestion des Adventices, Institut National de la Recherche Agronomique, BP 86510, 21065 Dijon Cedex, France. Tel: (+33) 3 80 69 31 87; Fax: (+33) 3 80 69 32 62; E-mail: sabrina.gaba@dijon.inra.fr

Abstract

What can you do when the information you get from field surveys is an answer such as no, yes a few or yes many?

Although semi-quantitative values are commonly recorded in ecological inventories, post transformation of the data is often realized, with a resulting loss of information by restricting the statistical analysis to a presence/absence contrast. So the stake is to adapt statistical approach to account for the variation in abundance that often has high ecological significance.

Many distributions in ecology do not follow a random pattern.

Abundance or density has been stressed as the next step to account for beyond the presence/absence of species in a given habitat. However the Null models that have been proposed in ecology to conduct the statistical analysis of non random distributions, do not cover so far the situation where the data is extracted from a query in a semi quantitative manner such as a distinction between 'few or many'.

Here, we have extended an existing algorithm of randomization that not only keep rows and columns marginals but also allows to fix the distribution for the interval classes.

On an ad hoc dataset we show that our procedure has a substantial increase fit.

On a case study we show that the choice the null model may significantly alter the conclusion that would be derived.

Introduction

A major research focus in Ecology has been the search of community assembly rules to study of patterns in the diversity, abundance, and composition of species in communities, and the processes underlying these patterns. Null models provide a prominently statistical test for whether an observed pattern is likely in the absence of a particular mechanism (Weiher & Keddy 1999). A growing number of studies has checked for assemblage rules by studying species co-occurrence (Boschilia et al., 2008; Ulrich, 2004), structure of food webs (Vázquez *et al.*, 2003) or the dispersion of functional traits (Stubbs & Wilson, 2004 ; Ackerly & Cornwell, 2009). In all these studies, a null model can be always defined as a pattern-generating model that is based on randomization of data (or random sampling from a known or imagined distribution). The randomization is designed to produce a pattern that would be expected in the absence of a particular ecological mechanism (Gotelli & Graves, 1996).
The evaluation of non-random pattern in the observed species distribution in communities highly dependents on the null hypothesis and on the choice and assumptions of the underlining null model on which the data are fitted. The test of a null hypothesis can be performed using several null models differing in their degree of restriction on the rules of randomization of the dataset. Therefore, different results and interpretations can be drawn by testing the same null hypothesis on a dataset. For example, co-occurrence tests used for testing for patterns of species co-occurrence are very sensitive to the variation of species occurrence frequencies; so species frequencies should be conserved between observed and randomised communities (Gotelli, 2000).

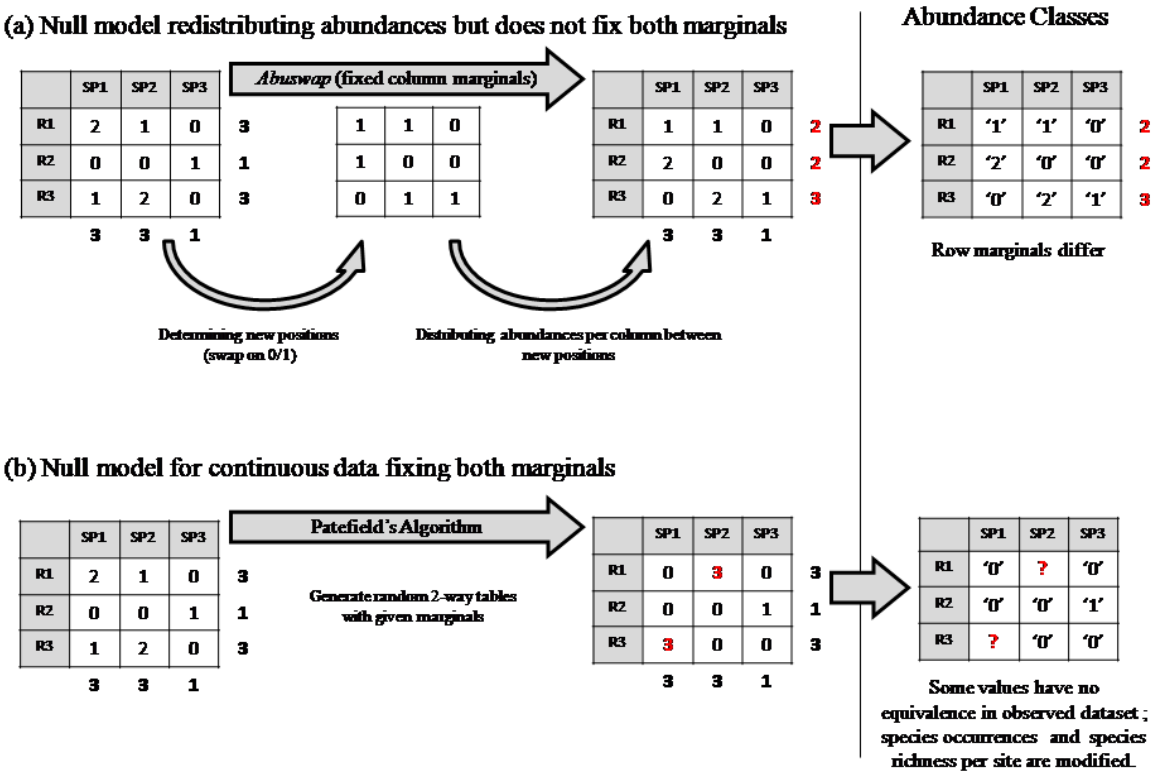
Some communities present datasets with highly structured row (sites) and column (species) marginals. For example, it is a very common pattern that ecological dataset are composed by few common species and many very rare species (Preston, 1962) and lognormal distribution seems to best fit this species abundance distribution (Ulrich et al., 2010). Moreover, for some ecological dataset, in particular in human-driven ecosystems such as arable fields, the distribution of the total abundance species in the communities is highly structured. As consequence, the analysis of species abundance co-occurrences requires to conserve the total abundance pattern in absence of any assumption on the distribution of the abundance per sites (Ulrich et al., 2010). This would be, for example, the case when the dataset cover both environmental rich and poor sites. Hence, the pattern-summarizing index can be highly

altered by unrestricted randomization leading in non comparable randomized datasets. The problem of choosing the adequate null model can become extremely tricky.

Several null models have been developed for presence/absence and continuous abundance data (Gotelli & Graves, 1996; Oksanen et al., 2011), and different restriction rules may be applied either by fixing the number of sites occupied by a species (column total), the number of species in a site (row total) or both. However, the abundance of species may also be recorded by using a qualitative format such as 'few or many' (as you may extract from an oral query), or semi-quantitative scale with abundance intervals. In the last case, the values of classes represent the median of intervals (for example Barralis, 1976). None of the null models developed so far is adapted to account for this kind of data format. Although semi-quantitative values are commonly recorded in ecological inventories, post transformation of the data is often realized, with a resulting loss of information by restricting the statistical analysis to a presence/absence contrast. So the stake is to adapt statistical approach to account for the variation in abundance that often has high ecological significance.

Among the existing null models, some can be adapted to abundance class data. The *Abuswap* model developed by Hardy (2008) can be used with abundance classes. However, the randomization process can only retain the observed row or column totals (Fig. 1a) separately. Quantitative null models (e.g. Patefield's algorithm) (Patefield, 1981) can retained both row and column totals and can be adapted by considering semi-quantitative values as quantitative i.e. abundance classes. However, the randomization process does not concern the observed abundance classes and some classes may be either created or removed by the randomization process generating numerical values with no equivalence in original matrix (Fig. 1b). The quantitative null models are thus inappropriate when working with abundance class data. Moreover, even though row and column totals are both fixed, the frequency of species occurrences and species richness per site are modified. Without any assumption, there is no reason to modify the observed patterns of species occurrences. This is particularly true in cases of analysis with indexes based on occurrence or relative abundance of species since all these modifications can highly alter simulated indexes leading in an erroneous acceptance or reject of the null hypothesis.

Figure 1 Incompatibility between abundance classes format and existing null models. (a) Abuswap null model with fixing column marginals allows dealing with abundance classes but does not preserve row marginals (sum of classes' values per row). (b) Patefield's algorithm allows both fixing of row and column marginals but can generate quantitative values with no equivalence in used class format.



To alleviate the limits of the current algorithms, we developed a simple null model adapted to abundance class data. This model is derived from the 'swap philosophy' used in presence/absence data-set. The limited number of abundance classes and their repetitiveness allow extension of this method classically used in case of two classes (0 and 1). Hence, it allows restricted randomization of dataset by fixing row and column sums and row and column sums of occurrences. The row and column totals of the observed and randomized datasets are strictly equal and only co-occurrences of species (in term of presence/absence and abundance) differ. We call this new null model '*SwapClass*'.

Analyses of functional trait dispersion and indexes used are sensitive to relative abundances of species. This is particularly true when calculating a community mean trait index (CWM).

Hence it is interesting to study how the absence of restriction on rows marginals can affect results and interpretation. There is a debate on the existence of competition within weed communities which should result in overdispersion of traits in the community to limit competition between individuals, but it is still poorly documented. Analysis of trait linked to competition dispersion could supply some answer on it.

In this paper, we first present the SwapClass model. Then we compare its efficiency to *Abuswap* null model by fixing rows or columns marginals using artificial datasets. Several simulated dataset are used in order to explore different species assembly patterns: from random to co-occurrence. Statistical tests are used to detect randomness or species co-occurrence pattern in the artificial abundance dataset. Finally we demonstrate the importance of choosing with caution a null model adapted to the structure of the observed dataset. To do so, we applied and compared the respective results given when using each of two null models (*Abuswap* with fixed columns marginals and *SwapClass*) in a dispersion trait analysis on weed communities. Since functional diversity indexes are sensitive to relative abundance of species, we expect differences between results of several null models.

Materials and Methods

'SwapClass' null model

In a 'swap' algorithm, randomly chosen submatrices of the form:

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \quad \text{or} \quad \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

are selected, and the cells of the submatrices are swap. The submatrices are not necessarily formed from adjacent rows and columns; any submatrix of this form can be swapped. Swapping creates a new matrix configuration, but does not alter row and column totals (Gotelli & Entsminger 2003). Same method can be applied on abundance class matrices, and then randomly chosen submatrices of the form:

$$\begin{matrix} x & y \\ y & x \end{matrix}$$

are selected, and the cells of the submatrices are swap in the same way. Swap on multiple classes keeps same properties than in presence/absence case (rows and columns not necessary adjacent, equal number of several classes for each row and column). Hence, high number of swaps destructs co-occurrences of abundance classes within pairs of species. Algorithm pays no attention to the choice of pairs of rows or columns selection; for each

swap, rows and columns probabilities are uniform. A swap is made only if fill does not change. The algorithm swaps submatrices until the expected number of swap is reached. As in 'swap' algorithms, performing null matrices can be recursive (the algorithm starts from the last performed matrix). The code is available upon request.

Diagnostic test on artificial random and structured matrices

The efficiency of *SwapClass* null model to detect randomness and pattern of artificial matrices is compared to those of *Abuswap* null model by preserving row and column occurrences, and column or row sums. 100 sets of couple of matrices are generated. Each couple is composed by one structured matrix and one random matrix. Cells of artificial matrices are composed by six abundance classes ('0' = 0 ind/m², '1' = [1:3] ind/m², '2' = [4:8] ind/m², '3' = [9:18] ind/m², '4' = [19:30] ind/m² and '5' = [31:+∞ [; numerical value of each class are median value of each interval; class '5' has a numeric value equal to 35 ind/m²). Pattern in non random matrix is performed by both modeling filter effect and resource partitioning between species. Species abundances distribution follows a lognormal distribution (Ulrich et al., 2010). Random matrix is performed by generating random matrix using Patefield's algorithm given marginals of the structured matrix generated (Patefield, 1981) (see Figure 1 in Appendix for more information about computation of random and structured artificial dataset). Each matrix is composed by 10 species (=columns) and 200 relevés (= rows).

To evaluate importance of preserving marginals of matrix in case structured marginals, we test efficiencies of three null models (*Abuswap* (Hardy, 2008) with fixed column marginals, *Abuswap* with fixed row marginals, and *SwapClass*) on two different patterns of row sums: (i) row sums follow a uniform distribution and (ii) row sums follow a lognormal distribution. We decide to test these two distributions of total abundance per site since the first one is the actually used distribution in preceding papers testing null model efficiencies on artificial matrices (Ulrich & Gotelli, 2010) and the second one is the actually observed distribution of total abundance in weed communities. Moreover, we think that this second distribution exacerbates impact of row sums distribution modification on studied indexes.

Efficiencies of null models to detect randomness and pattern are evaluated by using two co-occurrence metrics for abundance data: 'Mantel' and CA_{ST} indexes (see appendix) (Ulrich & Gotelli, 2010), and one occurrence metric for presence/absence data: C-score (Stone and Roberts, 1990).

For each artificial matrix, 100 null matrices are performed using the three null models. Since 'Mantel' test is based on correlation between original matrix and null matrices, and between

null matrices themselves, generating null matrices had not to be recursive; hence each null matrix is performed form the original one. We selected a high number of swaps: Number of discard communities before proper analysis is of 10 000 in *Abuswap* methods, and number of swaps performed is of 10 000 in *SwapClass* method. We check that this number is enough for both methods by plotting Bray-Curtis distances between artificial matrices trough generation by recursive way. We finally count the percentage of matrices identified as being segregated or aggregated (using the upper [UCL] and the lower [LCL] 97.5% confidence limits) in case of random matrices (H0) or structured matrices (H1).

Analysis of pattern of traits linked to competition in weed communities

We used the 'Biovigilance Flore' framework, a weed survey that has been set up in France to quantify the impact of agricultural practices and landscape organization on weed species diversity and abundance (Fried, 2007; Fried et al., 2008). The survey is carried out across a large number of fields chosen to cover the diversity of cultural practices and environmental conditions present in arable fields.

In this survey, the evaluation of abundance is a bit more sophisticated than a general perception such as 'few versus many'. A control plot of 200 m2 received the same usual techniques as the normal field with the exception of all herbicide(s) treatment(s). Each species' abundance was estimated in untreated area by two or more trained persons using six abundance classes. This method takes into account the number of individuals per m² using the following scale intervals as developed by Barralis (Barralis, 1976) : '+' found once in the 2000 m² area; '1' less than 1 individual/m²; '2' 1-2 individual/m²; '3' 3-20 individuals/m²; '4' 21-50 individuals/m²; '5' more than 50 individuals/m². The dataset is restricted to encompass 1143 sites and 253 species so that each particular crop field is only used once, thus avoiding pseudo replication (58 sampled in 2002, 215 in 2003, 383 in 2004, 207 in 2005, 227 in 2006 and 53 in 2007). Tree and cultivar species are removed from data-set, as well as relevés (sites) for which species richness is lower than three. To allow stratified randomization according to cultural type, we keep only sites where the four more frequent cultural type have been applied (Winter wheat = 303, Oilseed rape = 66, Maize = 241 and Sunflower = 53). Finally, the data-set is composed by 663 sites and 199 species.

Trait-based community analysis requires the selection of traits that are critical to the community processes of interest. In this study, we focus on the response of weeds to competition : plant height (Gaudet & Keddy, 1988 ; Goldberg, 1997), seed weight which is a good proxy of seedling competition ability (Eriksson, 1997;Turnbull et al., 1999; Storkey *et al.*, 2010) , specific leaf area (SLA, leaf area divided by dry mass) which is associated with rapid biomass accumulation and a good competitiveness for light interception at early stages and life form. Trait values were obtained in standardized databases: LEDA (Kleyer et al.,

2008), Biolflor (Kühn et al., 2004), Ecoflora (REF) and Baseflore (REF). Missing values were replaced by average (or dominant) value of the traits.

Three indices of functional diversity (FD) were computed on the combination of the four traits using the "FD package" (Laliberté & Legendre,. 2010; Laliberté & Legendre,. 2011) implemented in the R software (R Development Core Team 2010): functional richness (FRic) represents the amount of functional space filled by the community (Villégier et al., 2008), functional evenness (FEve) describes the evenness of abundance distribution in a functional trait space (Mason et al. 2005) and Rao's quadratic entropy (RaoQ) used as a measure of divergence of community by quantifying mean distance between individuals in a functional trait space (Botta-Dukát, 2005).
We evaluate difference and significance between observed and mean of null communities for FD indexes according to two null models: *Abuswap* with fixed column marginals and *SwapClass*. We decide not to test *Abuswap* with fixed row marginals since we know that species occurrence and abundance modification strongly modify mean traits values of communities.

Results

Performance of null model algorithms on artificial matrices

For the uniform and lognormal distributions of the sums of abundances per sites, the 'Mantel' index showed that the *Abuswap* method with fixed column totals confirmed the null hypothesis in 83% and 92% of random matrices, respectively (H0 in Table 1a and 1b). Results on the *Abuswap* with fixed rows were very bad, since respectively 8% and 3% of random matrices have been correctly identified. The *SwapClass* method presents the best result with 100% of random matrices identified as random for both distributions of sums of rows. Using CA_{ST}, *Abuswap* with fixed columns or fixed rows had never correctly identified any random matrices, while *SwapClass* identified respectively 89% and 94% of them.

Table 1 Results on data-sets for which sums of abundances per row follow a uniform distribution (a) or a lognormal distribution (b). Values represent the percentage of random matrices (H0) or structured matrices (H1) identified by three metrics of species covariation as being segregated or aggregated (using the upper (UCL) and lower (LCL) 97.5% confidence limits. Low values for both LCL and UCL means randomness of metric; hence, by using Mantel index, *Abuswap* with fixed column marginals detects 83% (100% - 17%) of random matrices (H0) as being random, in case uniform distribution of sums of abundances.

	(a) **Uniform**						**(b) Lognormal**					
	Mantel		**CAst**		**C-score**		**Mantel**		**CAst**		**C-score**	
	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
H0												
Abuswap fix. col. mar.	17%	0%	100%	0%	0%	32%	8%	0%	100%	0%	0%	18%
Abuswap fix. row mar.	82%	0%	100%	0%	0%	29%	97%	0%	100%	0%	1%	15%
SwapClass	0%	0%	0%	11%	1%	5%	0%	0%	2%	4%	2%	5%
H1												
Abuswap fix. col. mar.	100%	0%	0%	67%	83%	0%	100%	0%	0%	93%	20%	2%
Abuswap fix. row mar.	99%	0%	60%	19%	80%	0%	52%	0%	25%	52%	23%	1%
SwapClass	100%	0%	0%	80%	74%	0%	100%	0%	0%	95%	29%	1%

Analysis of results obtained on structured matrices (H1) showed that *Abuswap* with fixed columns and *SwapClass* have equally identified structured matrices with 'Mantel' index as 100% of structured matrices have been correctly identified for both distributions. Through the use of CA_{ST}, *SwapClass* has a better score than *Abuswap* with fixed columns marginals by correctly identifing respectively 80% and 95% of structured matrices for uniform and lognormal distributions against 67% and 93% of structured matrices correctly identified with *Abuswap* fixed columns). Results on structured matrices through use of *Abuswap* with fixed rows marginals and mantel test, were very bad since 52% of matrices have been identified as structured, in case of lognormal distribution (trough CA_{ST}, 77% of structured matrices have been identified as non random).

Analysis of pattern of traits linked to competition in weed communities: comparison between Abuswap and SwapClass

Functional diversity indexes showed significant differences between observed mean indexes obtained with the *Abuswap* null model (fixed columns marginals) and those obtained with *SwapClass* null model(Table 2a and 2b).

Table 2 Comparison of functional diversity indexes for two null models. Values indicate differences between observed values and mean of simulated values (positive= observed > mean of simulated) for three functional diversity indexes: FRic = Richness, FEve = Evenness and RaoQ = Divergence for the whole data-set. Significances are noted following: "***" = p-value < 1e-3; "**" = 1e-3 ≤ p-value < 1e-2; "*" = 1e-2 ≤ p-value < 5e-2; "+" = 5e-2 ≤ p-value < 0.1 and "ns" = p-value ≥ 0.1.

	(a) *Abuswap* (fixed column marginals)			**(b) *SwapClass***		
crop type	FRic	FEve	RaoQ	FRic	FEve	RaoQ
WW	**-0.00375****	**0.053217*****	**0.00109*****	**-0,00367****	0,00554ns	**-0,00069***
OR	**-0.03723+**	**0.085*****	0.00046ns	**-0,00366+**	**0,02022+**	**-0,0012***
M	-0.00184ns	**0.02528*****	-0.00004ns	-0,0154ns	**-0,02216****	**-0,00062***
S	**-0.00714***	-0.00524ns	-0.00259ns	**-0,00698***	**-0,03826*****	**-0,00155***

Functional Richness's (FRic) for both null models were similar, but functional evenness (FEve) and functional divergence (RaoQ) showed significant differences between the two null models. Observed FEve were significantly higher than random for WW, OR and M in case of using *Abuswap* model (Table 2a), while observed FEve was higher only in OR and were lower in M and S in case of using *SwapClass* model (Table 2b). Observed RaoQ was higher than random in WW in case of using *Abuswap* model (Table 2a) while observed RaoQ were lower in all crop types in case of using *SwapClass* model (Table 2b).

Discussion

Performance of null model algorithms on artificial matrices

Results obtained with different indexes may lead to different conclusion about the quality of several null models studied. But mantel index seems to be the better index to detect random or structured data-set, whatever the null model used. And whatever the used index, *SwapClass* seems to best discriminate random and structured matrices.

Efficiency of CA_{ST} index is still controversial. Gotteli suggests a better behavior of Mantel index. Moreover, Gotelli developed this index for quantitative data and we ignore what are the impacts of using this index on abundance class data. The power of CA_{ST} to detect random

or structured matrices was always lower than the power of Mantel index. Finally, CA_{ST} index has perfectly detected random matrices with *Abuswap* method, wathever the fixed marginals. This result is surprising since we know that not fixing columns marginals seriously impacts the dataset by modifying frequence and abundance of several species. This may warn us on the shady result obtained trough CA_{ST} index.

To conclude, these results reinforce the idea that modification of rows or columns marginals has an important impact on results and interpretation, in agreement with Gotelli's results (Ulrich & Gotelli, 2010). We think that the null model that we developed especially for abundance classes data-set can offer stronger result than other null model designed to presence/absence or quantitative data-set. The way used to randomize observed matrix could certainly generate bias in case of too small matrices or with some abundance classes present in very low frequencies. However, this null model fixes both sums of occurrence and sums of abundance of each row and columns. This double fixing (abundance and occurrence) could be too restrictive, and someone could want to fix only abundance and not occurrence. In this case, this null model is not adapted and use of Patefield's algorithm remains the good alternative.

Hence, the *SwapClass* null model seems to be the best model for abundance classes data-set, of course, this model is not necessary the best model to use on numerical or presence/absence, and we suggest refereeing to Gotelli work in these cases. It's only the high specificity of situations where abundances classes need to be accounted for that highly structured rows and columns marginals of data-set stimulates the need for adapted tools.

Analysis of pattern of traits linked to competition in weed communities: comparison between Abuswap and SwapClass

Results obtained by *Abuswap* null model suggest an over-dispersion of functional traits linked to competition, while results obtained by *SwapClass* null model would conversly suggest an under-dispersion of functional traits linked to competition. Results on functional richness are similar between the two null models (Table 2). This result isn't surprising since FRic index only use occurrences of species (and not abundances) to defined range of variations (i.e. richness) on each trait (Villéger et al., 2008). In fact, both null models have a very small and similar effect on presence/absence data. By contrast, results on Functional evenness are more contrasted. Using *Abuswap* null model, results show that observed evenness is significantly higher in WW, OR and M; these results are partially in opposition to results obtained with the *SwapClass* null model, for which evenness was higher only in OR and lower in M and S (Table 2). Similar behaviors can be observed for the divergence: by using *Abuswap*, RaoQ index was higher only in WW, while by using *SwapClass* model, RaoQ

was significantly lower in all crop types. Hence, use of a null model for which rows marginals weren't fixed can lead to opposite conclusion about the role played by competition to limit functional traits similarities between the weed species found in a same site. Since abundance co-occurrence and functional diversity indexes are highly dependent to relative abundance of species observed in a same site, a null model that doesn't fix rows marginals would logically be inappropriate for a data-set with high structure of its rows marginals.

Moreover, results obtained through *SwapClass* model are more plausible than those obtained through *Abuswap* model. In fact, it is well known that the major competitor of weed communities is represented by the dominant and more numerous species which is the crop itself ; so that few competition is expected to occur between weed species. Hence, competition with crop represents a high filter effect in weed assembly rules and lead in an under-dispersion of traits linked to competition.

To conclude, we suggest the use of *SwapClass* null model in trait dispersion analysis in case of abundance class dataset. Since this null model fixes row and column marginals of occurrences and abundances, it seems to be the most adapted.

Acknowledgement

Benjamin Borgy is funded through a fellowship ANR-OGM Vigiweed.

References

Barralis G. (1976) Méthode d'étude des groupements adventices des cultures annuelles. *INRA - 5ème Colloque International sur l'Ecologie et la Biologie des Mauvaises Herbes.* Dijon, France.

Boschilia S.M., Oliveira E.F. & Thomaz S.M. (2008) Do aquatic macrophytes co-occur randomly? An analysis of null models in a tropical floodplain. *Oecologia* **156**:203-214.

Botta-Dukát, Z. (2005) Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science* **16**:533-540.

Cornwell W.K & Ackerly D.D. (2009) Community assembly and shifts in plant trait distributions accross an environmental gradient in coastal Carlifornia. *Ecological Monographs* **79**(1): 109-126.

Eriksson O. (1997) Clonal life histories and the evolution of seed recruitment. In: de Kroon H., van Groenendael J. (eds), The Ecology and Evolution of Clonal Plants. Backhuys, Leiden, pp. 211–226

Fried G. (2007) 'Biovigilance Flore', a long-term French weed survey. *20ème Conférence du COLUMA. Journées Internationales sur la Lutte contre les Mauvaises Herbes.* Dijon, France.

Fried G., Norton L.R., Reboud X., (2008) Environmental and management factors determining weed species composition and diversity in France. *Agriculture, Ecosystems & Environment* **128**: 68-76

Gaudet C.L. and Keddy P.A. (1988) Predicting competitive ability from plant traits: a comparative approach. *Nature* **334**:242-243.

Goldberg D.E. (1997) Competitive ability: definitions, contingency and correlated traits, J. Silvertown, M. Franco, J.L. Harper, Editors, *Plant Life Histories*, Cambridge University Press, Cambridge, pp. 283–306.

Gotelli N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology* **81**: 2606-2621.

Gotelli N.J., and Graves G.R. (1996) Null Models in Ecology. Smithsonian Institution Press, Washington, DC.

Hardy, O.J. (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology* **96**, 914–926.

Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens and Helene Wagner (2011). vegan: Community Ecology Package. R package version 1.17-10. http://CRAN.R-project.org/package=vegan

Kleyer, M., Bekker, R., Knevel, I., Bakker, J., Thompson, K., Sonnenschein, M., Poschlod, P., Van Groenendael, J., Klimeš, L., Klimešová, J., Klotz, S., Rusch, G., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Dannemann, A., Endels, P., Götzenberger, L., Hodgson, J., Jackel, A.-K., Kühn, I., Kunzmann, D., Ozinga, W., Römermann, C., Stadler, M., Schlegelmilch, J., Steendam, H., Tackenberg, O., Wilmann, B., Cornelissen, J., Eriksson, O., Garnier, E. and Peco, B. (2008), The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology, 96: 1266–1274.

Kuhn I., Durka W. & Klotz S. (2004). - BiolFlor - a new plant-trait database as a tool for plant invasion ecology. Divers. Distrib., **10**, 5-6: 363-365.

Laliberté E. & Legendre P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**:299-305.

Laliberté, E. & B. Shipley. (2011). FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-11.

Mason N.W.H., Mouillot D., Lee W.G. & Wilson J.B. (2005) Functional richness, functional evenness and functional divergence : the primary components of functional diversity. *Oikos* **111** :112-118.

Patefield W.M. (1981) Algorithm AS159. An efficient method of generating r x c tables with given row and column totals. *Applied Statistics* **30**, 91–97.

Preston F.W. 1962. The canonical distribution of commonness and rarity: parts I and 2. *Ecology* **43**:185–215,410–432.

Stone L. & Roberts A. (1990) The checkerboard score and species distributions. *Oecologia* **85**:74–79.

Storkey J., Moss S.R. & Cussans J.W. (2010) Using Assembly Theory to Explain Changes in a Weed Flora in Response to Agricultural Intensification. *Weed Science*, **58**(1), p.39-46.

Stubbs W.J. & Wilson J.B. (2004) Evidence for limiting similarity in a sand dune community. *Journal of Ecology* **92** :557-567.

Turnbull L.A., Rees M. & Crawley M.J. (1999) Seed mass and the competition/colonization trade-off: a sowing experiment. Journal of Ecology, 87: 899–912.

Ulrich W. (2004) species co-occurrences and neutral models: reassessing J.M. Diamond's assembly rules. *Oikos* **107**:603-609.

Ulrich W., Ollik, M. & Ugland. K.I. (2010) A meta-analysis of species–abundance distributions. *Oikos* **119**:1149–1155.

Ulrich W. and Gotelli N.J. (2010) Null model analysis of species associations using abundance data. *Ecology* **91**:3384-3397.

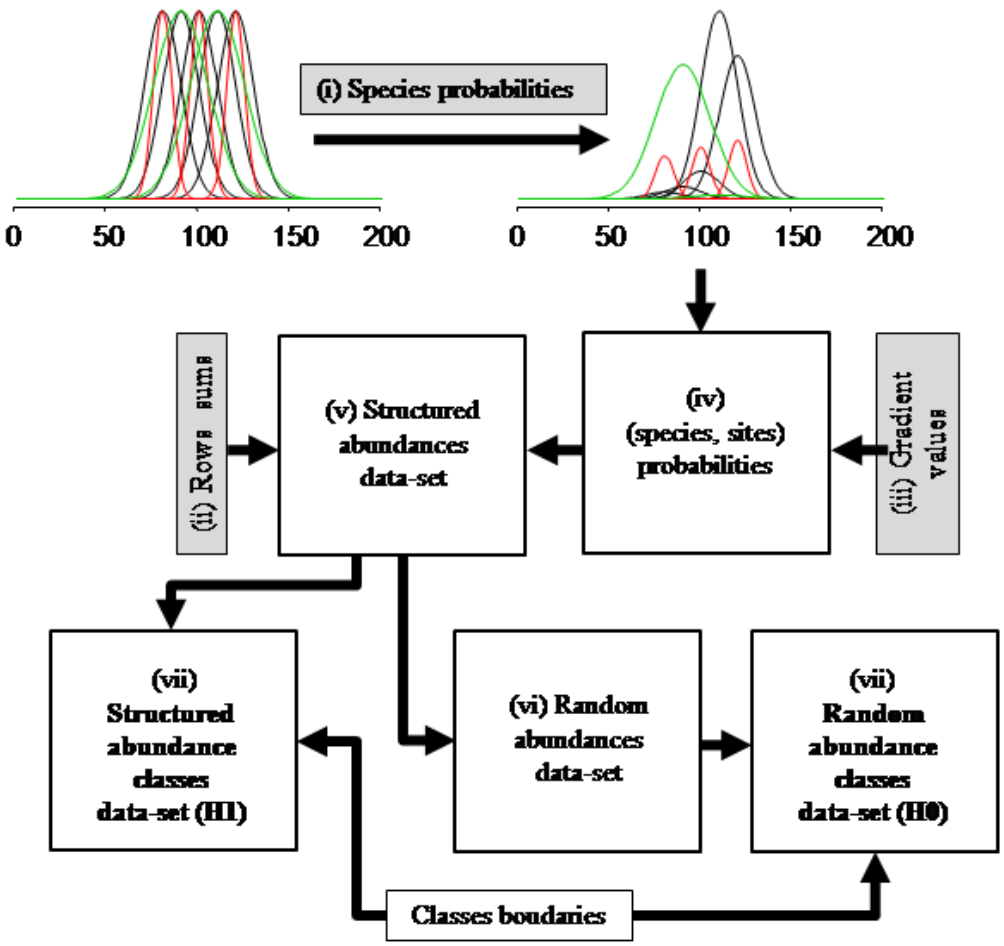
Vázquez D.P. & Aizen M.A. (2003) Null model analyses of specialization in plant-pollinator interactions. *Ecology* **84**, 2493–2501

Villéger S., Mason N.W.H. & Mouillot D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* **89**:2290-2301.

Weiher, E., & Keddy P. (1999) Ecological assembly rules, perspectives, advances, retreats. Cambridge University Press, Cambridge, UK.

Appendix

Figure X1 Computation of structured and random artificial datasets. First, species specialization and preferences for gradient are defined. For each run, row sums, gradient values, global probabilities of species were randomly choose (grey boxes). (i) Species probabilities were randomly chosen following a log normal distribution (few species were frequent, lots of species were rare). (ii) Row sums were randomly chosen following a log normal distribution. (iii) Gradient values were randomly chosen following a uniform distribution. (iv) (species, sites) probabilities were computed from gradient values and species distributions on gradient. (v) total abundance per site (row sum) were randomly distributed among species according species probabilities per site. (vi) Marginals of structured abundance dataset (rows and columns sums) were used to compute random abundance dataset using Patefield's algorithm. (vii) Structured and random abundances datasets were transformed in abundance classes dataset according to classes boundaries.



Co-occurrences indexes

Mantel index: The Mantel test identifies nonrandom correlations between two matrices (Mantel 1967). To assess whether the matrix was nonrandom, we used the mean Mantel correlation between the original matrix (using the Spearman correlation and Bray-Curtis distance on columns) and the matrices generated by different null model algorithms; and compare it to the distribution of Mantel correlation between all the matrices generated by different null model algorithms. The Mantel test used in this way can indentify nonrandomness, but it does not indicate whether an observed matrix is unusually aggregated or segregated (Ulrich & Gotelli, 2010).

CA_{ST}: We used an abundance analog of ''checkerboard'' distributions (Diamond 1975). The more checkerboard units there are in a matrix, the more segregated species are in their occurrence. We define an ''abundance checkerboard'' as a 2 x 2 submatrix of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad a > b \qquad a > c \qquad d > b \qquad d > c$$

or $a < b \qquad a < c \qquad d < b \qquad d < c$

where a, b, c, and d represent the abundances of two species in two different sites. The metric CA is a count of the total number of abundance checkerboards in the matrix. This metric can be standardized with regard to matrix size (m rows, n columns) by

$$CA_{ST} = \frac{4CA}{m(m-1)n(n-1)}$$

The standardized CA value can range from 0.0 to 1.0, with high values of CA indicating more negative covariation in abundances (Ulrich & Gotelli, 2010).

References of appendix

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220.

Ulrich. W. and N.J. Gotelli. 2010. Null model analysis of species associations using abundance data. Ecology 91:3384-3397.

6.1.2. Résultats supplémentaires découlant de l'article

Les résultats issus de cet article, sur des données artificielles, montrent que notre modèle nul (*SwapClass*) semble plus apte à détecter la présence ou non d'une structuration dans la cooccurrence des espèces, à l'échelle de l'abondance, que les deux variantes du modèle *Abuswap* (sommes des colonnes ou sommes des lignes fixes). De plus, les résultats montrent aussi que par l'utilisation d'un modèle nul adapté (*SwapClass*) à notre jeu de données (classes d'abondance) et suivant l'analyse que l'on souhaite réaliser (comparer les distributions de traits observées à des distributions « nulles »), la diversité fonctionnelle des traits liés à la compétition semble être sous-dispersée suggérant un effet filtre important des pratiques et de l'environnement (filtre habitat). Néanmoins, les deux processus pouvant conduire à un assemblage non-aléatoire des traits, le filtre habitat et la compétition, vont avoir des effets contraires entre sous- et sur-dispersion. Ces deux processus n'étant pas exclusifs, le fort effet de l'un peut venir obscurcir l'effet de l'autre. Comme pour l'analyse de l'assemblage des notes d'abondance, il peut alors être intéressant de se concentrer sur un sous-échantillon pour estimer l'effet que pourrait avoir la densité totale d'adventices sur l'assemblage de ces traits. En effet, si la compétition inter-adventice limite la similarité des traits liés à la compétition, la diversité fonctionnelle de ces mêmes traits devrait alors être plus sur-dispersée au sein d'un jeu de données composé des seuls sites où la densité totale d'adventices est la plus importante. Ainsi, nous avons alors réalisé la même analyse (modèle nul *SwapClass* + indice de diversité fonctionnelle) sur les 25% des sites ayant les densités totales d'adventices les plus importantes (le jeu de données est alors composé de 168 sites et de 150 espèces). Les différences entre valeurs observées et simulées des trois indices de diversité fonctionnelle présentent des différences importantes entre le jeu de données total et le jeu de données réduit aux « fortes densités » (Tableau 1). Les richesses et les équitabilités fonctionnelles (respectivement FRic et FEve) qui étaient presque exclusivement inférieures aux valeurs obtenues sous une hypothèse d' « aléa » dans le jeu de données total, se retrouvent maintenant être non différentes statistiquement dans le jeu de données « fortes densités ». Les divergences fonctionnelles (RaoQ) qui étaient toutes significativement inférieures à l' « aléa » ne le sont plus que pour deux types de cultures (Maïs [M] et Tournesol [S]).

	(a) Jeu de données entier			(b) Jeu de données "fortes densités"		
culture	FRic	FEve	RaoQ	FRic	FEve	RaoQ
WW	**-0,00367****	0,00554ns	**-0,00069***	-0,0014ns	-0,00086ns	-0,00027ns
OR	**-0,00366+**	**0,02022+**	**-0,0012***	-0,00032ns	**0,0981****	-0,00073ns
M	-0,0154ns	**-0,02216****	**-0,00062***	-0,00173ns	-0,00992ns	**-0,00085***
S	**-0,00698***	**-0,03826*****	**-0,00155***	0,0066ns	**-0,0509+**	**-0,00202***

Tableau 1 Indices de diversité fonctionnelle et effet de la densité totale. Les valeurs indiquent la différence entre valeurs observées et moyennes des valeurs simulées (positive= observée > moyenne des simulées) pour 3 indices de diversité fonctionnelle: FRic = Richesse, FEve = équitabilité et RaoQ = divergence pour le jeu de données entier (a) et le jeu de données réduit aux fortes densités (b). Les significativités sont notées suivant : "***" = p-value < 1e-3; "**" = 1e-3 ≤ p-value < 1e-2; "*" = 1e-2 ≤ p-value < 5e-2; "+" = 5e-2 ≤ p-value < 0.1 and "ns" = p-value ≥ 0.1. Le modèle nul *SwapClass* a été utilisé pour réaliser les distributions simulées sous l'hypothèse nulle. Les cultures blé d'hiver, colza, maïs et tournesol sont respectivement représentées par WW, OR, M et S.

Le fort effet filtre habitat que l'on observait à l'échelle du jeu de données entier, se retrouve être d'importance moindre à l'échelle du jeu de données réduit aux 25% des sites les plus denses en adventices. Ainsi, dans ce jeu de données réduit, même si nous n'observons pas pour autant de sur-dispersion des traits liés à la compétition, nous pouvons suspecter que la compétition vient réduire la sous-dispersion des traits par une importance relative plus importante (comparée à celle dans le jeu de données total). On peut aussi supposer que ce sous jeu de données regroupe des situations culturales où l'intensité des perturbations liées aux pratiques est plus faible, laissant ainsi passer plus de valeurs de trait et expliquant ainsi la perte de significativité des déficits de richesse (les différences négatives entre valeurs observées et moyenne des simulées de FRic qui étaient significatives dans le jeu de données total ne le sont plus dans le jeu de données réduit aux « fortes densités »). Néanmoins, cette deuxième hypothèse ne peut expliquer les différences d'équitabilité (FEve) et de divergence (RaoQ) entre les deux jeux de données (les indices sont censés être indépendants). Ainsi, ceci confirmerait nos attentes : en se limitant aux situations les plus à même d'être en condition de compétition, on s'éloigne du modèle « filtre ». Ces deux forces seraient d'une intensité suffisamment similaire pour s'annuler. Ces résultats appuient l'idée que la compétition inter-adventice soit en moyenne de trop faible intensité pour façonner les communautés et ce, principalement du fait que les communautés adventices atteignent rarement des degrés de saturation suffisamment importants pour induire une compétition entre elles. Néanmoins, la densité en adventices ayant un effet sur la dispersion des traits, et réduisant la sous-dispersion, on peut facilement penser que la compétition doit limiter la similarité des traits dans certaines conditions de fortes densités d'adventices.

6.2. Article: Inferring weed spatial distribution from multi-type data (paru dans Ecological Modelling 226 (2012) 92-98)

Ecological Modelling 226 (2012) 92–98

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Short communication

Inferring weed spatial distribution from multi-type data

A. Bourgeois[a], S. Gaba[b], N. Munier-Jolain[b], B. Borgy[b], P. Monestiez[a], S. Soubeyrand[a,*]

[a] INRA, UR546 Biostatistique et Processus Spatiaux, F-84914 Avignon, France
[b] INRA, UMR1210 Biologie et Gestion des Adventices, F-21065 Dijon, France

ARTICLE INFO

Article history:
Received 28 July 2011
Received in revised form 7 October 2011
Accepted 10 October 2011

Keywords:
Weed mapping
Weed patch
Bayesian hierarchical model
Spatial interpolation
Log Gaussian Cox process

ABSTRACT

An accurate weed management in a context of sustainable agriculture relies on the knowledge about spatial weed distribution within fields. To improve the representation of patchy spatial distributions of weeds, several sampling strategies are used and lead to various weed measurements (abundance, count, patch boundaries). Here, we propose a hierarchical Bayesian model which includes such multi-type data and which allows the interpolation of weed spatial distributions (using a MCMC algorithm). The weed pattern is modeled with a log Gaussian Cox process and the various weed measurements are modeled with different observation processes. The application of the method to simulated data shows the advantage of combining several types of data (instead of using only one type of data). The method is also applied to infer the weed spatial distribution for real data.

1. Introduction

Weeds in arable fields are a potential threat for valuable crops. Indeed, weeds may affect the crop yield and the harvest quality by introduction of impurities (Sen, 1998). On the other hand, weeds may provide food and shelter for invertebrates and birds (Holmes and Froud-Williams, 2005; Meiss et al., 2010). In a context of increasing the sustainability of agro-ecosystems, this dual function of weeds needs to be optimized. At the field scale, the choice of agricultural practices for accurate weed management, in a context of herbicide use reduction, requires that we enhance our knowledge about the spatial distribution and dynamics of weed populations.

One way to acquire that knowledge is to make weed maps. Most weed maps in literature were generated with kriging (Cardina et al., 1995; Heisel et al., 1996; Clay et al., 1999), even though kriging is criticized for producing weed maps with less variation in the population at short distance than is realistic (Rew et al., 2001). The map quality relies on the spatial interpolation method (Heisel et al., 1996; Dille et al., 2002; Guillot et al., 2009) and the quality of sampling (Cousens et al., 2002). Some models have been proposed to improve spatial interpolations (Brix and Møller, 2001; Brix and Chadœuf, 2002; Kruijer et al., 2007). For example, Brix and Møller (2001) proposed a space-time multitype log Gaussian Cox process which includes pairwise interaction terms allowing the modeling of aggregation at large scale and of regularity at small scale.

Although the latter models better fit the data, they are traditionally built to map weed counting data collected over fixed-size quadrats. However, some sampling strategies have been developed to deal with time consuming manual weed counting and poor representation of weed patches. For example, to evaluate the performance of integrated weed management in four cropping systems in a long term experiment, within field weed abundance was assessed by (i) counting weeds in 30–40 quadrats of size 0.36 m^2, (ii) giving abundance notes in 30–100 quadrats of size 16 m^2 and (iii) giving abundance notes in weed patches with boundaries determined by GPS (Chikowo et al., 2009; Munier-Jolain et al., 2004, 2008). Mapping these three types of data is challenging because of their different natures (counting versus classes) but also because they have been assessed at different scales.

In this short communication, we propose a hierarchical Bayesian model (Clark, 2005; Wikle, 2003) which takes into account the three types of weed data (i), (ii) and (iii); see also Gotway and Young (2002) for an overview of the multi-type data topic. Our contribution is in line with the articles by Brix and Møller (2001) and Brix and Chadœuf (2002) who proposed to interpolate weed counting data with Cox processes. The novelty of our approach lies in the use of three sub-models built for the three types of data. In Section 2, the hierarchical Bayesian model including the three sub models is detailed, and the estimation and interpolation method is presented. In Section 3, the method is applied to simulated and real data and some perspectives are discussed.

* Corresponding author.
E-mail address: Samuel.Soubeyrand@avignon.inra.fr (S. Soubeyrand).

0304-3800/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2011.10.010

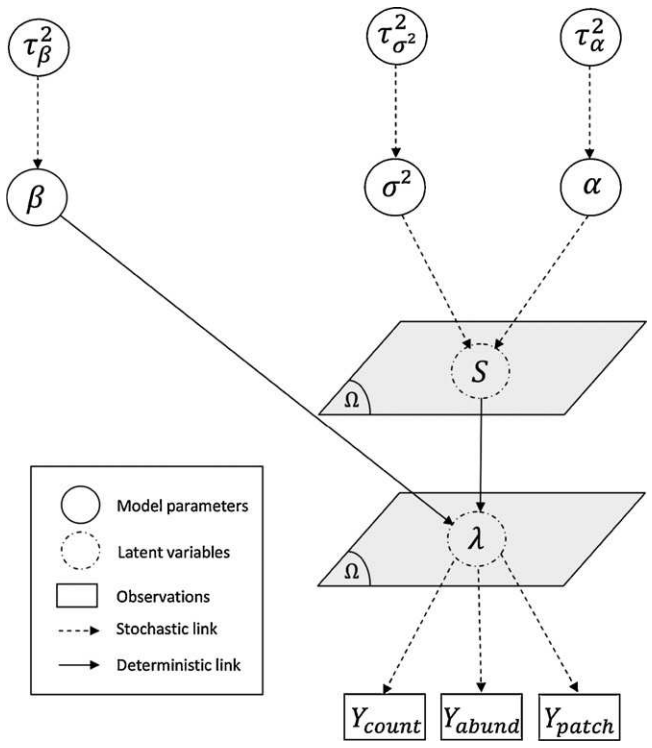


Fig. 1. Direct acyclic graph showing the structure of the hierarchical Bayesian model.

2. Method

2.1. Hierarchical Bayesian model

In what follows, we set the point pattern model used to describe weed locations in a field. Then, we model the three types of data conditionally on the point pattern model and propose prior distributions for the model parameters. The direct acyclic graph in Fig. 1 illustrates the structure of the model.

2.1.1. Weed locations

Let $\Omega \subset \mathbb{R}^2$ be the domain under study, typically a field plot. We assume that weed locations form over Ω a log Gaussian Cox process (Møller et al., 1998). The log Gaussian Cox process is a doubly stochastic point process (Diggle, 2003; Illian et al., 2008; Stoyan et al., 1995) with intensity λ modeled as a log-normal random field. In the following, the weed intensity function λ satisfies:

$$\lambda(x) = \exp(\beta + S(x))$$

where $\beta \in \mathbb{R}$, S is a Gaussian random field with stationary exponential spatial covariance function: $C(x, x') = \sigma^2 \exp(-\alpha||x - x'||)$, $||\cdot||$ is the Euclidean distance, σ and α are in $\mathbb{R}^*_+$. If spatial explanatory variables are available (it was not the case in our data set), they may be added to $\beta + S(x)$ like in Diggle et al. (1998).

2.1.2. Counting data

The first type of data is the counting of weeds over disjoint quadrats $A_1, \ldots, A_I, I \in \mathbb{N}$, included in Ω. Let Y_i be the count of weeds in quadrat A_i. Under the process for weed locations defined above, Y_i given λ follows a Poisson distribution with mean $\Lambda(A_i) = \int_{A_i} \lambda(x)dx$ and for all $i \in \{1, \ldots, I\}$

$$\mathbb{P}(Y_i = n \mid \lambda) = e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^n}{n!}, \qquad \forall n \in \mathbb{N}.$$

2.1.3. Abundance data

The second type of data is the assessment of the quantity of weeds over disjoint quadrats $A_{I+1}, \ldots, A_{I+J}$ included in $\Omega, J \in \mathbb{N}$, using a simplified version of the Barralis scale (Barralis, 1976; Munier-Jolain, 2010). For $i \in \{I+1, \ldots, I+J\}$, if the number of weeds Y_i in A_i is low, i.e. less than or equal to n_1, then this number is observed; if the number of weeds is high, i.e. greater than n_1, then Y_i is censored in the Q intervals $(n_1, n_2], \ldots, (n_{Q-1}, n_Q], (n_Q, n_{Q+1} = \infty)$. Values of $n_1, \ldots, n_{Q+1}$ for the applications are provided in Appendix A. Under the process for weed locations defined above, Y_i given λ is Poisson distributed with mean $\Lambda(A_i) = \int_{A_i} \lambda(x)dx$ (as above) and for all $i \in \{I+1, \ldots, I+J\}$

$$\mathbb{P}(Y_i = n \mid \lambda) = e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^n}{n!}, \quad \forall n \in \{0, 1, \ldots, n_1\},$$

$$\mathbb{P}(Y_i \in (n_q, n_{q+1}] \mid \lambda) = \sum_{n=n_q+1}^{n_{q+1}} e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^n}{n!}, \quad \forall q \in \{1, 2, \ldots, Q\}.$$

2.1.4. Patch data

The third type of data is the counting of weeds over patches $A_{I+J+1}, \ldots, A_{I+J+K}$ included in $\Omega, K \in \mathbb{N}$, with high weed densities with respect to the surroundings of these patches. The surroundings are denoted by $\tilde{A}_{I+J+1}, \ldots, \tilde{A}_{I+J+K}$. For $i \in \{I+J+1, \ldots, I+J+K\}$, the number of weeds per area unit in patch A_i is assumed to be γ times higher than the number of weeds per unit area in the patch surrounding $\tilde{A}_i$ ($\gamma \geq 1$). Under the process for weed locations defined above, for all $i \in \{I+J+1, \ldots, I+J+K\}$, Y_i given λ is Poisson distributed with mean $\Lambda(A_i) = \int_{A_i} \lambda(x)dx$ (as above) and $Y_i/|A_i| \geq \gamma \tilde{Y}_i/|\tilde{A}_i|$; $|A_i|$ and $|\tilde{A}_i|$ are the areas of A_i and $\tilde{A}_i$. Consequently, for all $i \in \{I+J+1, \ldots, I+J+K\}$ and $n \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{P}(Y_i = n, Y_i/|A_i| \geq \gamma \tilde{Y}_i/|\tilde{A}_i| \mid \lambda) \\
&= \mathbb{P}(\tilde{Y}_i \leq Y_i |\tilde{A}_i|/\gamma |A_i| \mid Y_i = n, \lambda) \mathbb{P}(Y_i = n \mid \lambda) \\
&= \left(\sum_{n'=0}^{\lfloor n|\tilde{A}_i|/\gamma|A_i| \rfloor} \mathbb{P}(\tilde{Y}_i = n' \mid \lambda) \right) \mathbb{P}(Y_i = n \mid \lambda) \\
&= \left(\sum_{n'=0}^{\lfloor n|\tilde{A}_i|/\gamma|A_i| \rfloor} e^{-\Lambda(\tilde{A}_i)} \frac{\Lambda(\tilde{A}_i)^{n'}}{n'!} \right) e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^n}{n!},
\end{aligned}$$

where $\lfloor n|\tilde{A}_i|/\gamma|A_i| \rfloor$ is the floor value of $n|\tilde{A}_i|/\gamma|A_i|$.

2.1.5. Prior distributions for the parameters

In this article, the parameter γ is assumed to be equal to one. This is the most conservative value for γ when no additional information is available; the specification of γ is discussed in Section 3. For β, $\log \sigma$ and $\log \alpha$, we assumed independent centered normal prior distributions with variances τ^2_β, $\tau^2_{\sigma^2}$ and τ^2_α equal to 100^2. Vague priors were used because no information was available on the parameters. Thus, hierarchical Bayesian modeling is not invoked to include expert information but to exploit MCMC which allows us to provide a posterior distribution for the weed intensity function.

2.2. Estimation and interpolation

Assuming that the contours of the sampling units (i.e. the quadrats, the patches and the patch surroundings) are known, the hierarchical model presented above may be used to write a posterior distribution allowing the interpolation of the weed intensity function.

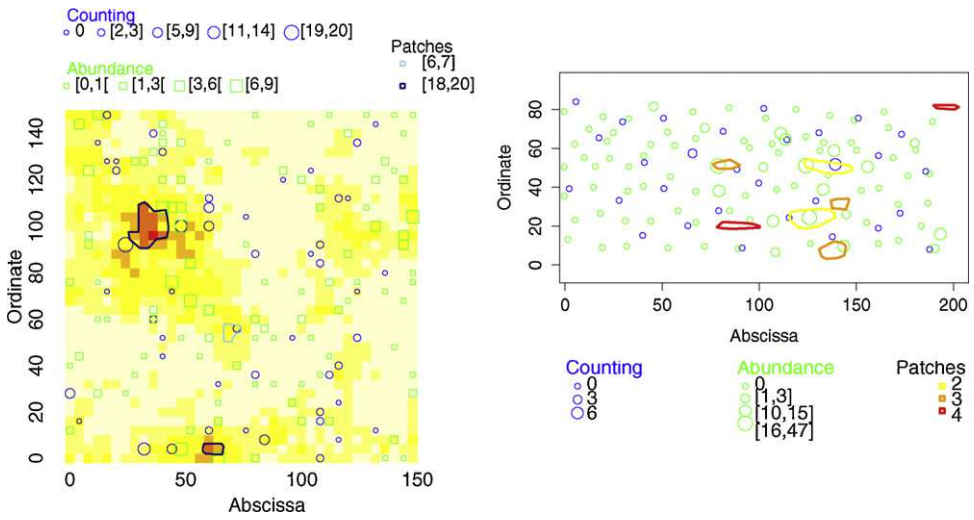


Fig. 2. Simulated data set (left) and real data set (right; measures of cleavers in a wheat field in May 2006, France). For the simulated data set, the true weed intensity function λ is shown. The intervals provided in the legends merge the counts and intervals in the data tables.

2.2.1. Posterior distribution

Let $\mathbf{Y}$ denote the set of counting, abundance and patch data. From the dependence structure provided by Fig. 1, the joint posterior distribution of the unknowns (weed intensity function and parameters) is proportional to

$$\begin{aligned} p(\lambda, \beta, \sigma, \alpha \mid \mathbf{Y}) &\propto p(\mathbf{Y}, \lambda \mid \beta, \sigma, \alpha)\pi(\beta, \sigma, \alpha) \\ &= p(\mathbf{Y} \mid \lambda)p(\lambda \mid \beta, \sigma, \alpha)\pi(\beta, \sigma, \alpha) \\ &= p(\mathbf{Y} \mid \lambda)\mathbf{1}\{\lambda \equiv \exp(\beta + S)\}p(S \mid \sigma, \alpha)\pi(\beta, \sigma, \alpha). \end{aligned} \tag{1}$$

where $p(\mathbf{Y}, \lambda \mid \beta, \sigma, \alpha)$ is the complete likelihood of the model; π is the joint prior distribution of the parameters, i.e. a product of normal densities; $p(\mathbf{Y} \mid \lambda)$ is the conditional distribution of the data given λ; $p(\lambda \mid \beta, \sigma, \alpha)$ is the distribution of λ which can be written as the product of the indicator function $\mathbf{1}\{\lambda \equiv \exp(\beta + S)\}$—which equals one if λ coincides with $\exp(\beta + S)$—and the distribution $p(S \mid \sigma, \alpha)$ of the Gaussian random field S. The indicator function appears because of the deterministic link between λ, β and S.

2.2.2. Conditional distribution $p(\mathbf{Y} \mid \lambda)$

Regarding the J abundance data, we suppose that there are J_1 observed counts, where $0 \le J_1 \le J$, and $J - J_1$ counts censored in intervals. Then, $\mathbf{Y}$ can be written as follows:

$$\mathbf{Y} = \left\{ y_1, \ldots, y_{I+J_1}, [\underline{y}_{I+J_1+1}, \overline{y}_{I+J_1+1}], \ldots, [\underline{y}_{I+J}, \overline{y}_{I+J}], y_{I+J+1}, \ldots, y_{I+J+K} \right\},$$

where the symbols y_i denote the observed counts and the symbols $\underline{y}_i$ and $\overline{y}_i$ denote the lower and upper bounds of the intervals in which some of the observed counts are censored. Given the contours of the sampling units, the distribution $p(\mathbf{Y} \mid \lambda)$ can be written, if $J_1 < J$:

$$p(\mathbf{Y} \mid \lambda) = \left(\prod_{i \in \{1, \ldots, I+J_1\} \bigcup \{I+J+1, \ldots, I+J+K\}} e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^{y_i}}{y_i!} \right) \times \prod_{i=I+J_1+1}^{I+J} \left(\sum_{y=\underline{y}_i}^{\overline{y}_i} e^{-\Lambda(A_i)} \frac{\Lambda(A_i)^{y}}{y!} \right)$$

$$\times \prod_{i=I+J+1}^{I+J+K} \left(\sum_{y=0}^{\lfloor y_i |\tilde{A}_i| / \gamma |A_i| \rfloor} e^{-\Lambda(\tilde{A}_i)} \frac{\Lambda(\tilde{A}_i)^{y}}{y!} \right). \tag{2}$$

If $J_1 = J$ (no count censored in interval), then the second product in Eq. (2) has to be deleted.

Some of the sampling units may overlap (in the real data, there are 12 overlaps for 135 sampling units). For such overlapping sampling units, the corresponding weed measures are dependent conditional on λ. This dependence is ignored in Eq. (2); an approach to take it into account is discussed in Section 3.

2.2.3. Integral approximation and distribution $p(S \mid \sigma, \alpha)$

The weed intensity function λ being the function to be estimated, the integrals $\Lambda(A_i)$, $i = 1, \ldots, I+J+K$, and $\Lambda(\tilde{A}_i)$, $i = I+J+1, \ldots, I+J+K$, are unknown. These integrals are approximated as follows: let $\lambda(x_1), \ldots, \lambda(x_M)$ denote the values of λ in a finite number of points $x_1, \ldots, x_M$ included in the sampling units; let A denote any sampling unit; the approximation of $\Lambda(A)$ is

$$\check{\Lambda}_M(A) = \frac{|A|}{\sum_{m=1}^{M} \mathbf{1}(x_m \in A)} \sum_{m=1}^{M} \lambda(x_m)\mathbf{1}(x_m \in A) \approx \Lambda(A),$$

where $\mathbf{1}(\cdot)$ is the indicator function.

In the estimation algorithm, the approximation $\check{\Lambda}_M$ replaces the function Λ in Eq. (3). It follows that the distribution $p(S \mid \sigma, \alpha)$ in Eq. (1) reduces to the distribution of the spatial Gaussian vector $S(x_1), \ldots, S(x_M)$; the expression of this distribution is given in Stein (1999, Appendix).

In the estimation algorithm, the values of $S(x_1), \ldots, S(x_M)$ are updated at each iteration. Thus, for large M the algorithm may be very time consuming (e.g. for the real data, we set $M = 152$ and it took about 40 h to run 10^5 MCMC-iterations with an up-to-date computer and the R software).

2.2.4. MCMC algorithm

Our model with various observation processes is an extension of spatial generalized linear mixed models used in model-based geostatistics (Diggle et al., 1998). Thus, we adapted the MCMC algorithm presented by Diggle et al. (1998) to (i) estimate $\lambda(x_1), \ldots, \lambda(x_M)$ and (ii) interpolate λ at the nodes of a grid covering the study domain Ω. The main adaptation deals with the expression of the

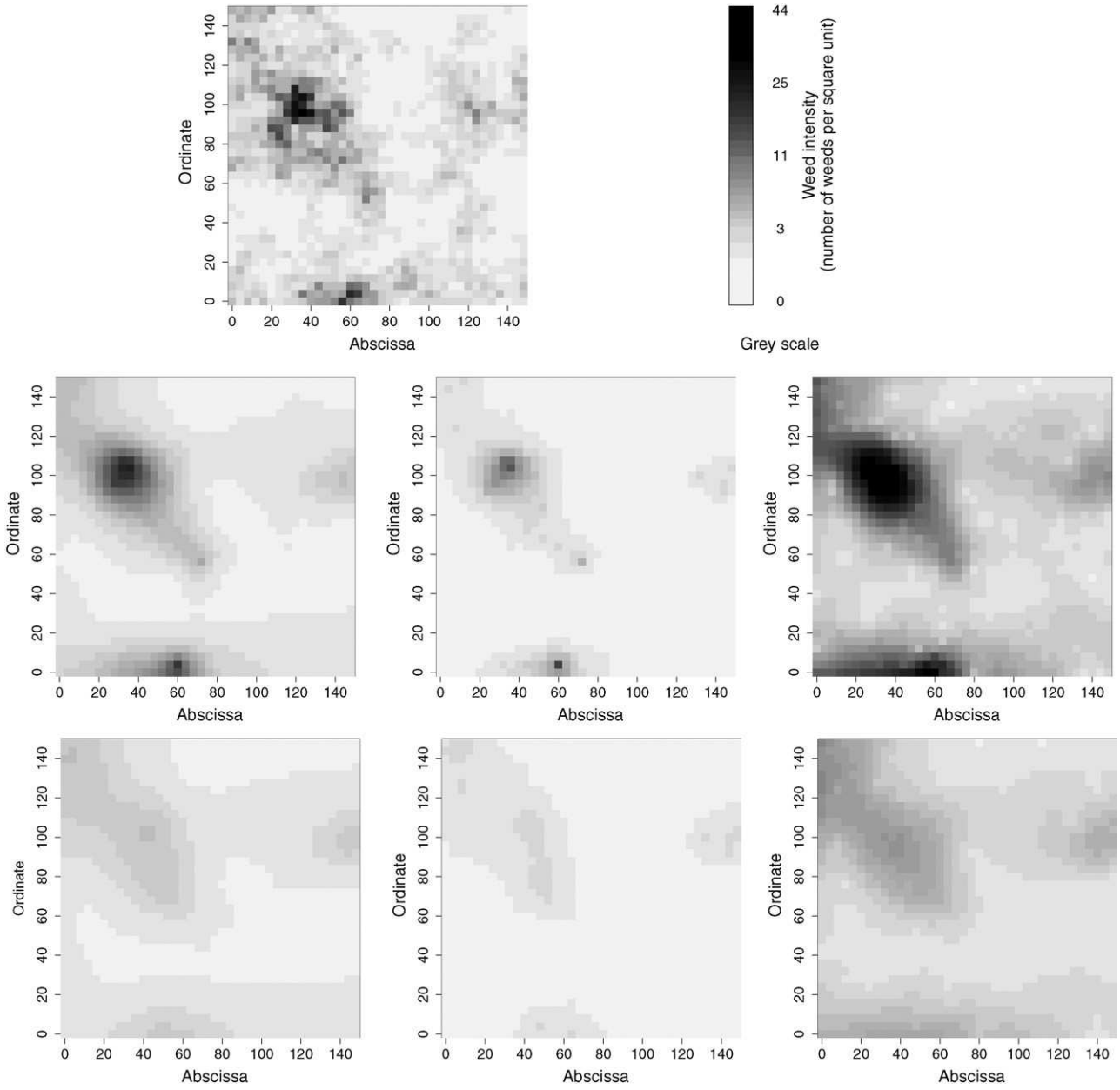


Fig. 3. Interpolation for the simulated data. Top: true weed intensity function λ and grey scale legend (for all the maps). Second line, from left to right: interpolated posterior median of λ and posterior quantiles of order 0.005 and 0.995 using all the data. Third line: same plots obtained when only the abundance data are used.

likelihood (see above). Information about the starting values and the proposal distributions are provided in Appendix A.

3. Results and discussion

We applied the method proposed above to a simulated data set and a real one (available at http://samuel.biosp.org). The corresponding weed measurements are shown on Fig. 2.

3.1. Simulated data

The simulated data set was generated under the hierarchical model of Section 2.1 with $(\beta, \sigma^2, \alpha) = (0.5, 2, 1/35)$, $\Omega = [0, 152] \times [0, 152]$. The true weed intensity function λ and the collected data are shown on the left panel of Fig. 2. There are $I + J + 2K = 50 + 100 + 2 \times 3 = 156$ sampling units (2 times K because there are K patches and K patch surroundings); there are seven pairs of overlapping sampling units. For the integral approximation in the estimation algorithm, we set $M = 158$ points distributed in the 156 sampling units. We used only one point per quadrat because each quadrat area was less than 0.07% of the total area of Ω. For the patches and patch surroundings we used numbers of points proportional to their areas. For the distribution of the number of weeds in the patch surroundings, we used $\gamma = 1$.

The interpolated posterior median of λ and the posterior quantiles of order 0.005 and 0.995 are displayed on Fig. 3 (second line). We also drawn the analogue maps obtained when only the abundance data are used (third line). Visually, the true weed intensity function λ (top left panel of Fig. 3) is correctly interpolated by its posterior distribution obtained with all the data. 85.9% of the true values of λ are in the corresponding marginal 99%-posterior intervals; see Table 1. This percentage is lower than expected but will increase with the quantity of information brought by the data. Besides, despite the partial inadequacy between the true λ and its

Table 1
Coverage statistics for the simulated data set. Coverage of the true values of λ by their marginal 99%-posterior intervals (PI); assessed for 1444 values of λ at the grid nodes. Coverage of the observations by their marginal 99%-PI obtained by simulation under the posterior distributions of the unknowns; For any observed weed count, we checked if the count was in its PI; For any count censored in an interval, we checked if this interval was intersecting the PI.

	All data	Only abundance data
Coverage of λ values	85.9%	63.9%
Coverage of observations	99.4%	99.0%

estimation, 99.4% of the observations are correctly predicted by the posterior model; see Table 1. When only the abundance data are used, the interpolation is poorer; see Fig. 3 (third line) and Table 1. This illustrates the advantage of combining the three types of data.

3.2. Real data

The real data set concerns the cleavers sampled in May 2006 in a wheat field located near Dijon, France. There are $I+J+2K=30+91+2\times 7=135$ sampling units; there are twelve pairs of overlapping sampling units. For the integral approximation in the estimation algorithm, we set $M=152$ points distributed in the 135 sampling units. For the distribution of the number of weeds in the patch surroundings, we used $\gamma=1$.

The interpolated posterior median of λ and the posterior quantiles of order 0.005 and 0.995 are displayed on Fig. 4. 100% of the observations are covered by their marginal 99%-posterior intervals (PI) obtained by simulation under the posterior distributions of the unknowns; The legend of Table 1 explains how this coverage is computed. Moreover, we carried out a posterior predictive assessment of the model fitness by applying the approach of Gelman et al. (1996). The posterior predictive p-value which equals 0.51 supports the adequacy of the fitted model to the real data; see Appendix A for details about the test.

Such interpolations could be used to study the spatio-temporal dynamics of weeds and the interaction between different weed species.

3.3. Methodological perspectives

With the method presented above we progressed in the mapping of the weed spatial distribution in a field because we are able to include in the interpolation several types of data. However, some points may be improved to obtain a more accurate inference. Indeed, it should be possible to take into account:

- the uncertainty of contours for large quadrats, patches and surrounding patches;
- the uncertainty in the intervals in which the counts are censored for abundance data and the uncertainty in the counts for patch data, because these observations are only based on visual assessment in the real data;
- the dependence of observations made for overlapping sampling units. It could be easy to take into account this dependence for quadrats included in larger sampling units (differences between two counts or a count and an interval). However, taking into account this dependence for partially overlapping sampling units is not obvious.

Improving our method in these directions could be possible by including supplementary latent variables in the model, but this solution may lead to very time-consuming algorithms.

Another improvement could be the assessment of the parameter γ which governs the distribution of the number of weeds in the patch surroundings. In theory, this parameter could be estimated with the MCMC algorithm. Counting and abundance data collected within the patch surroundings would help the MCMC to provide a posterior distribution for γ. However, in the data set analyzed in this article, the number of such data was too low. Thus, we preferred to use the conservative value $\gamma=1$. $\gamma=1$ implies that the density of weeds within the patch (DWP) is only greater than or equal to the density of weeds within the patch surrounding (DWPS). $\gamma>1$ would imply that DWP is greater than DWPS. In the interpolation of the weed intensity function, $\gamma>>1$ would make appear rings of very low intensity around the patches.

3.4. From kriging to model-based geostatistics

We mentioned in the introduction that kriging is often applied to interpolate weed spatial distributions. However, with multitype data and heterogeneous sampling units, regular kriging cannot be

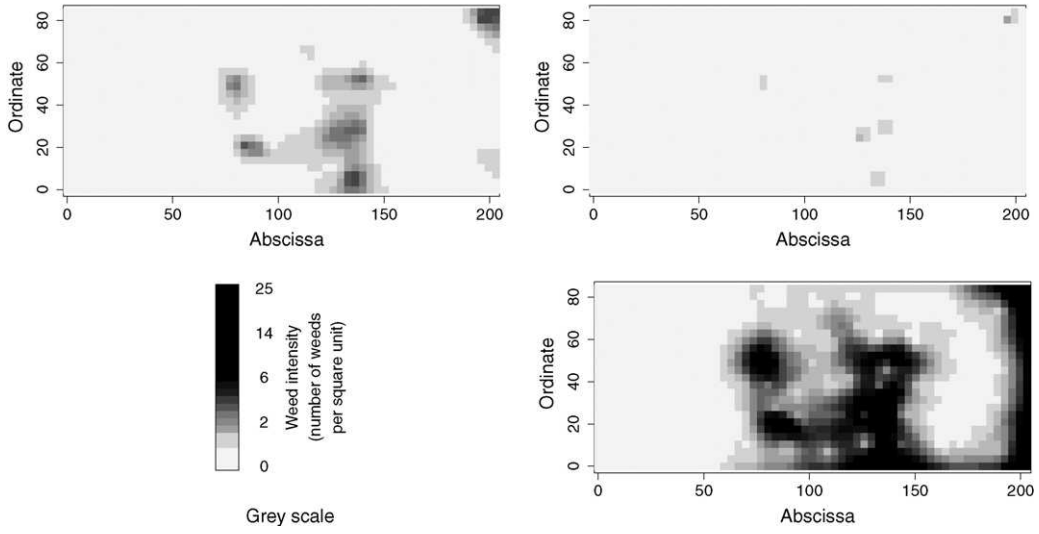


Fig. 4. Interpolation for the real data. Left: interpolated posterior median of λ and grey scale legend. Right: posterior quantiles of order 0.005 (top) and 0.995 (bottom).

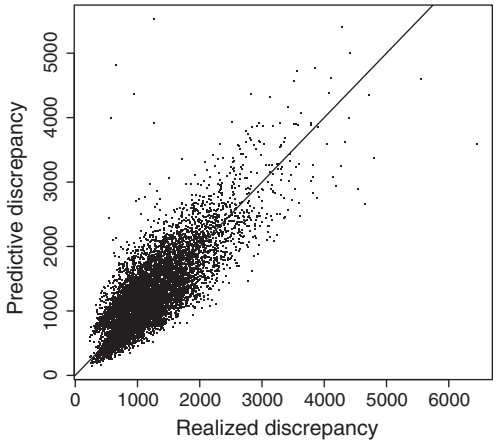


Fig. 5. Posterior predictive model check. Scatterplot of predictive versus realized discrepancies under the joint posterior distribution of λ. The *p*-value is estimated by the proportion of points above the 45° line.

directly applied. One must first transform the data and discard those which cannot be transformed correctly. For instance, we should have to homogenize supports and distributions of counting, abundance and patch data, and discard patch surrounding data to be able to apply indicator kriging, ordinary kriging (Chilès and Delfiner, 1999) or Poisson kriging (Monestiez et al., 2006). Since the article by Diggle et al. (1998), kriging (naturally associated with Gaussian assumptions, see Diggle et al., 1998; Stein, 1999) has been extended to model-based kriging which is associated with spatial generalized linear mixed modeling. Thus, the approach proposed in this communication is in this vein and is nothing else than model-based kriging with various observation models.

Acknowledgements

The authors thank an anonymous reviewer for his suggestions as well as Edith Gabriel for her comments on an early draft of the manuscript. This work was supported by the ANR grant STRA-08-02 Advherb.

Appendix A.

A.1. Interval values

Regarding the abundance data in the simulated and real applications, the intervals in which the weed counts are censored are $(n_1, n_2] = (9, 15]$, $(n_2, n_3] = (15, 47]$, $(n_3, n_4] = (47, 319]$, $(n_4, n_5] = (319, 799]$, $(n_5, n_6] = (799, 7999]$, $(n_6, \infty) = (7999, \infty)$. They correspond to a simplification of the Barralis scale (Barralis, 1976; Munier-Jolain, 2010).

A.2. MCMC tuning

The starting values of $S(x_1), \ldots, S(x_M)$ were fixed at zero and arbitrary starting values for β, σ and α were chosen so that the initial values of the log-likelihood and the log-priors were finite. Regarding the proposal distributions, new values for $S(x_1), \ldots, S(x_M)$, β, $\log\sigma$ and $\log\alpha$ were proposed with univariate normal distributions centered on the current values. Block updating was used only for $\log\sigma$ and $\log\alpha$. Besides, we ran in each case (simulated and real data) 10^5 MCMC-iterations and obtained the posterior sample of λ by sub-sampling in the chains every 10 iterations after a burnin of 10^4 iterations. The interpolation of λ was made for the 9000 iterations which were sub-sampled.

A.3. Posterior predictive assessment of the model fitness

We applied the approach of Gelman et al. (1996) by using a χ^2-like discrepancy:

$$\chi^2(\mathbf{Y};\lambda^{(z)}) = \left(\sum_{i\in\{1,\ldots,I+J_1\}\bigcup\{I+J+1,\ldots,I+J+K\}} \frac{\left\{y_i - \check{\Lambda}_M^{(z)}(A_i)\right\}^2}{\check{\Lambda}_M^{(z)}(A_i)} \right) + \left(\sum_{i=I+J_1+1}^{I+J} \frac{\left\{\frac{1}{2}(\bar{y}_i - \underline{y}_i) - \check{\Lambda}_M^{(z)}(A_i)\right\}^2}{\check{\Lambda}_M^{(z)}(A_i)} \right) + \left(\sum_{i=I+J+1}^{I+J+K} \frac{\left\{\left(\frac{1}{2}\left\lfloor \frac{y_i|\tilde{A}_i|}{\gamma|A_i|}\right\rfloor - \check{\Lambda}_M^{(z)}(A_i)\right)\right\}^2}{\check{\Lambda}_M^{(z)}(A_i)} \right). \quad (3)$$

where $\lambda^{(z)}$ is the state of λ in the z-th iteration of the MCMC algorithm and

$$\check{\Lambda}_M^{(z)}(A_i) = \frac{|A_i|}{\sum_{m=1}^M \mathbf{1}(x_m \in A_i)} \sum_{m=1}^M \lambda^{(z)}(x_m)\mathbf{1}(x_m \in A_i).$$

For the abundance data and patch surrounding data censored in intervals, we used the middles of the intervals instead of the unobserved weed counts y_i. For abundance data, the middle of the interval $[\underline{y}_i, \bar{y}_i]$ is $\frac{1}{2}(\bar{y}_i - \underline{y}_i)$; for patch surrounding data, the middle of the interval $\left[0, \left\lfloor \frac{y_i|\tilde{A}_i|}{\gamma|A_i|}\right\rfloor\right]$ is $\frac{1}{2}\left\lfloor \frac{y_i|\tilde{A}_i|}{\gamma|A_i|}\right\rfloor$.

The posterior predictive *p*-value was obtained as follows. For each $\lambda^{(z)}$ a replicated data set $\mathbf{Y}_{\text{rep}}^{(z)}$ was simulated. Then, we computed the realized discrepancy $\chi^2(\mathbf{Y};\lambda^{(z)})$ and the predictive discrepancy $\chi^2(\mathbf{Y}_{\text{rep}}^{(z)};\lambda^{(z)})$ for each iteration z and drawn the corresponding scatterplot; see Fig. 5. The *p*-value was estimated by the proportion of points above the 45° line, i.e. *p*-value=0.51.

Appendix B. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2011.10.010.

References

Barralis, G., 1976. Méthode d'étude des groupements adventices des cultures annuelles: application la Côte d'Or. Vème Colloque International d'Ecologie et de Biologie des Mauvaises Herbes, Dijon, 59–68.
Brix, A., Chadœuf, J., 2002. Spatio-temporal modelling of weeds by shot-noise G Cox processes. Biometrical Journal 44, 83–99.
Brix, A., Møller, J., 2001. Space-time multi type log Gaussian Cox processes with a view to modelling weeds. Scandinavian Journal of Statistics 28, 471–488.
Cardina, J.D., Sparrow, H., McCoy, E., 1995. Analysis of spatial distribution of common lamsquarters chenopodium album in no-till soybean. Weed Science 44, 298–308.
Chikowo, R., Faloya, V., Petit, S., Munier-Jolain, N., 2009. Integrated weed management systems allow reduced reliance on herbicides and long-term weed control. Agriculture, Ecosystems & Environment 132, 237–242.
Chilès, J., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty, vol. 344. Wiley-Interscience.
Clark, J.S., 2005. Why environmental scientists are becoming bayesians. Ecology Letters 8, 2–14.
Clay, S.A., Lems, G.J., Clay, D.E., Forcella, F., Ellsbury, M.E., Carlson, C.G., 1999. Analysis of spatial distribution of common lamsquarters chenopodium album in no-till soybean. Weed Science 47, 674–681.

Cousens, R., Brown, R., McBratney, A., Whelan, B., Moerkerk, M., 2002. Sampling strategy is important for producing weed maps: a case study using kriging. Weed Science 50, 542–546.
Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns. Oxford University Press, New York.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Journal of the Royal Statistical Society, C 47, 299–350.
Dille, J., Milner, M., Groeteke, J., Mortensen, D., Williams, M., 2002. How good is your weed map? A comparison of spatial interpolators. Weed Science 51, 44–55.
Gelman, A., Meng, X., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6, 733–759.
Gotway, C.A., Young, L.J., 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97, 632–648.
Guillot, G., Loren, N., Rudemo, M., 2009. Spatial prediction of weed intensities from exact count data and image-based estimates. Journal of the Royal Statistical Society Series C: Applied Statistics 58, 525–542.
Heisel, T., Andreasen, C., Ersboll, A.K., 1996. Annual weed distributions can be mapped with kriging. Weed Science 36, 325–333.
Holmes, R.J., Froud-Williams, R.J., 2005. Post-dispersal weed seed predation by avian and non-avian predators. Agriculture, Ecosystems & Environment 105, 23–27.
Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. Wiley.
Kruijer, W., Stein, A., Schaafsma, W., Heijting, S., 2007. Analyzing spatial count data, with an application to weed counts. Environmental and Ecological Statistics 14, 399–410.
Meiss, H., Le Lagadec, L., Munier-Jolain, N., Waldhardt, R., Petit, S., 2010. Weed seed predation increases with vegetation cover in perennial forage crops. Agriculture, Ecosystems & Environment 138, 10–16.
Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox process. Scandinavian Journal of Statistics 25, 451–482.
Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J., Guinet, C., 2006. Geostatistical modelling of spatial distribution of balaenoptera physalus in the northwestern mediterranean sea from sparse count data and heterogeneous observation efforts. Ecological Modelling 193 (3–4), 615–628.
Munier-Jolain, N., 2010. Rapid weed survey at the field scale. Technical report, INRA—Quantipest platform.
Munier-Jolain, N., Deytieux, V., Guillemin, J.P., Granger, S., Gaba, S., 2008. Conception et évaluation multicritères de prototypes de systèmes de culture dans le cadre de la protection intégrée contre la flore adventice en grandes cultures. Innovations Agronomiques 3, 75–88.
Munier-Jolain, N., Faloya, V., Davaine, J.B., Biju-Duval, L., Meunier, D., Martin, C., Charles, R., 2004. A cropping system experiment for testing the principles of integrated weed management. In: Annales AFPP, XIIème colloque international sur la lutte contre les mauvaises herbes, Dijon, pp. 147–156.
Rew, L.J., Whelan, B., McBratney, A.B., 2001. Does kriging predict weed distributions accurately enough for site-specific weed control? Weed Research 41, 245–263.
Sen, D.N., 1998. Key factors affecting weed-crop balance in agroecosystems. In: Altieri, M.A., Liebman, M. (Eds.), Weed Management in Agroecosystems: Ecological Approaches. CRC Press, New York, pp. 157–182.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag, New York.
Stoyan, D., Kendall, W.S., Mecke, J., 1995. Stochastic Geometry and its Applications, 2nd ed. Wiley, Chichester.
Wikle, C.K., 2003. Hierarchical models in environmental science. International Statistical Review 71, 181–199.