

**UNIVERSITE MONTPELLIER II  
SCIENCES ET TECHNIQUES DU LANGUEDOC**

**T H E S E**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE MONTPELLIER II**

*Discipline* : Biologie des Populations & Écologie

*Formation Doctorale* : Biologie de l'Évolution & Écologie

*Ecole Doctorale* : Biologie des Systèmes Intégrés

présentée et soutenue publiquement

par

François MUNOZ

Le 19 mai 2006

**Distribution régionale des espèces et dynamique des métapopulations :  
Modèle hiérarchique d'habitat et inférence du taux relatif extinction/colonisation**

**JURY**

M. CHEPTOU Pierre-Olivier, Chargé de Recherche CNRS, CEFE  
M. GARRAUD Luc, Botaniste, CBN Alpin  
M. GODELLE Bernard, Professeur, Univ. Montpellier II  
M. GOUYON Pierre-Henri, Professeur, MNHN Paris  
M. KJELLBERG Finn, Directeur de Recherche CNRS, CEFE  
Mme LESNE Annick, Maître de Conférences, Univ. Paris VI  
M. REBOUD Xavier, Directeur de Recherche INRA, INRA Dijon

Directeur de Thèse  
Membre invité  
Président  
Rapporteur  
Directeur de Thèse  
Rapporteur  
Examineur

## Remerciements

Voici la traditionnelle cérémonie d'attribution des remerciements, moment solennel au décorum soigné. Sans plus attendre, voici les catégories reines du palmarès.

Pour leur capacité à m'encadrer et à me supporter durant trois ans (on appréciera les différentes acceptions des verbes), j'appelle mes adorés directeurs de thèse, Pierre-Olivier Cheptou et Finn Kjellberg. Luc Garraud, ami et botaniste passionnant, est à l'origine et au cœur de ma démarche : hommage lui est donc rendu. En mémoire des discussions précieuses et passionnantes, je salue collectivement Thierry Boulonier, Jérôme Chave, Pierre Couteron, Stéphane Dray, Jean-Dominique Lebreton, Annick Lesne, Raphaël Pélissier et Marc-André Selosse. Pour son esprit critique incomparable et rafraîchissant, je propose un prix spécial à Daniel Chessel.

A tous les amis et camarades avec qui j'ai eu le bonheur de partager quelques années de labeur, j'offre humblement mon affection : Cindy Adolphe, Delphine Aigoin, Emilie Andrieu, Nadine Boulant, Paul Caplat, Antoine Dornier, Clélia Sirami, pour ne citer que les plus proches. Je souhaite associer à ces pensées d'amitiés Philippe Jarne et Patrice David, avec qui j'ai réalisé mes premiers pas dans le monde de la recherche.

Et pour finir, voici l'hommage appuyé à un ami incomparable, homme de grande âme et de cœur, dont le soutien et l'humanité furent assurément essentiels. Je le nomme donc en point d'orgue de la cérémonie : Alain Renaux.

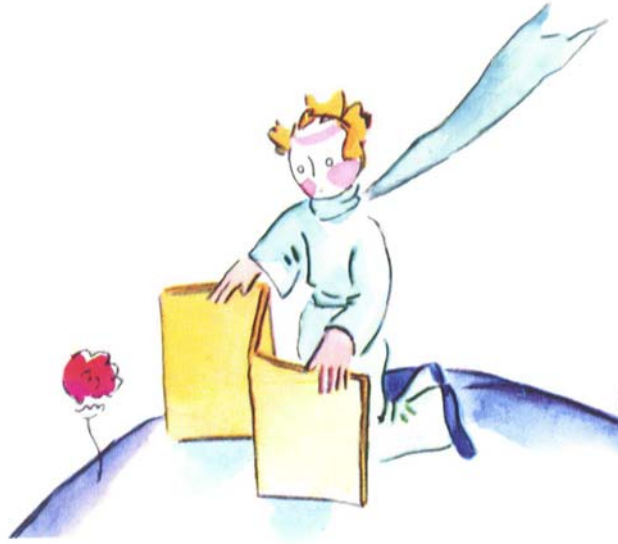
# Plan général

Remerciements .....	1
Plan général .....	2
Introduction .....	7
1. Dynamique spatiale des populations .....	8
1.1. L'Atlas de la Flore de la Drôme .....	8
1.2. Vision dynamique de la niche et métapopulations .....	11
2. Quatre paradigmes en question .....	17
2.1. Cadre de base : système, objets, états .....	17
2.2. L'entité biologique en écologie : paradigme biologiste .....	18
2.3. Déterminisme vs. stochasticité : paradigme de la niche .....	20
2.4. Paradigme de l'échelle et paradigme de l'équilibre .....	21
2.4.1. Echelle des objets .....	21
2.4.2. Equilibre du système .....	22
2.4.3. Stabilité du système .....	24
2.5. Contingence et émergence : paradigmes en question .....	25
2.5.1. Prise en compte des événements contingents .....	25
2.5.2. Signification des propriétés émergentes .....	29
3. Structure de causalité .....	29
4. Axes de travail .....	32
I. Métapopulations simulées en habitat structuré .....	35
1. Concept de métapopulation : modèle de Levins .....	35
1.1. Formulation du modèle .....	35
1.2. Usage d'une représentation discrète ou continue .....	37
2. Procédure de simulation .....	38
2.1. Paysage binaire d'habitat .....	38
2.2. Simulation de la dynamique métapopulationnelle .....	39
II. Criticalité et auto-organisation des métapopulations .....	43
1. Singularité des processus de contact .....	43
1.1. Limite de la représentation quasi-stationnaire .....	44
1.2. Notion d'auto-organisation .....	47
1.3. Seuil critique d'extinction .....	49

1.4.	Criticalités et auto-organisation .....	53
2.	Structure auto-organisée des métapopulations .....	56
2.1.	Seuil d'extinction des métapopulations et vision dynamique de l'habitat.....	58
2.2.	Propriétés émergentes .....	66
2.3.	Une question d'échelles .....	68
III.	Modélisation inférentielle : des effets aux causes .....	70
1.	Mesurer l'espace écologique : pertinence de l'analyse spectrale .....	71
1.1.	Qu'est-ce que l'analyse spectrale ?.....	72
1.2.	Stationnarité spatiale, principe d'incertitude et lissage.....	73
1.3.	Analyse de Fourier et méthode PCNM spectrale.....	77
1.4.	Découplage des dimensions spatiales et non-spatiales de la niche.....	82
1.5.	Saturation de l'habitat vs convergence structurale a priori/a posteriori.....	86
1.6.	Inférences des paramètres de l'habitat a priori et a posteriori .....	89
2.	Etude des effets de structuration spatiale par partition de variation.....	90
2.1.	L'espace n'est pas une variable comme les autres.....	90
2.2.	Cause ou effet ? Notion de variable spatiale.....	92
2.3.	Trouver un sens écologique : approche informationnelle.....	93
2.4.	Variations sur le thème de la partition .....	94
3.	Dualité vs. causalité : étude critique de la méthode PCNM .....	94
3.1.	Méthode PCNM causale .....	95
3.2.	Comparaison avec la méthode PCNM spectrale : rôle capital du lissage .....	96
3.3.	Retour sur la spatialisation de la niche : méthode de partition spectrale .....	100
3.4.	Test sur données simulées et perspectives pour l'inférence .....	104
3.4.1.	Découplage de la structure habitat et de la structure métapopulation.....	104
3.4.2.	Perspectives d'inférence .....	105
IV.	Modélisation analytique : des causes aux effets .....	108
1.	Approche relativiste et informationnelle .....	108
2.	Représentation spatialement implicite par approximations de paires.....	110
2.1.	Principe général .....	112
2.2.	Modélisation de la dynamique du système .....	113
2.3.	Etude des solutions à l'état stationnaire.....	114
2.4.	Inférence par Pair Approximation .....	115
3.	Modèles de métapopulation spatialement réalistes .....	118
3.1.	Définition du système et équation de base.....	119
3.2.	Stationnarité temporelle.....	122

3.3.	Formulation matricielle .....	124
3.4.	Notion de stationnarité spatiale – Linéarisation de l'équation de dynamique .....	125
3.5.	Prise en compte de la structure de l'habitat .....	127
3.6.	Modèle analytique spectral .....	128
3.7.	Perspectives ouvertes.....	130
4.	Inférence de la dynamique métapopulationnelle .....	130
4.1.	Méthode 1 : inférence globale à partir des probabilités locales moyennées dans le temps .....	131
4.2.	Méthode 2 : inférence locale à partir des probabilités moyennées dans le temps.....	134
4.3.	Méthode 2 : inférence locale à partir des occurrences instantanées agrégées spatialement .....	139
4.4.	Méthode 3 : préférences d'habitat et dynamique métapopulationnelle .....	141
4.5.	Bilan : quel choix pour l'estimation à partir de données réelles ? .....	144
V.	Dynamique spatiale et écologie des plantes vasculaires du département de la Drôme .....	146
1.	Données floristiques et environnementales du département de la Drôme.....	146
1.1.	Inventaire floristique.....	146
1.2.	Données environnementales .....	147
2.	Caractères généraux de la Flore drômoise.....	148
2.1.	De la garrigue méditerranéenne aux prairies subalpines .....	149
2.2.	Composante méditerranéo-montagnarde .....	151
2.3.	Les forêts tempérées médio-européennes .....	152
2.4.	Vallées et zones humides .....	154
2.5.	Le syndrome orchidées .....	155
3.	Estimation du ratio $r$ extinction / colonisation .....	158
3.1.	Agrégation spatiale des données .....	158
3.2.	Cartes d'habitat potentiel .....	159
3.2.1.	Variables d'habitat .....	160
3.2.2.	Deux représentations de l'habitat.....	161
3.3.	Validation de la procédure d'inférence.....	164
3.4.	Normalisation de l'habitat potentiel .....	166
3.5.	Caractéristiques générales des estimations pour les végétaux de la Drôme .....	167
4.	Impact des activités humaines sur la distribution spatiale des végétaux .....	172
4.1.	Vision relativiste de l'habitat et de la dynamique métapopulationnelle .....	172
4.2.	La fable de l'orchidée et de la fougère .....	174
4.2.1.	La néottie nid d'oiseau <i>Neottia nidus-avis</i> .....	174
4.2.2.	Le capillaire noir <i>Asplenium adiantum-nigrum</i> .....	176
4.3.	Syndrome de rudéralité.....	179
5.	Epilogue.....	180

Synthèse et discussion .....	181
1. Représentations de la dynamique métapopulationnelle : une question d'échelles .....	182
1.1. Etat quasi-stationnaire – auto-organisation.....	183
1.2. Démarche analytique d'inférence – principe d'incertitude.....	183
1.3. Paradigme des échelles intermédiaires .....	184
1.4. Péroration - le démon de Maxwell .....	185
2. Représentations de la niche écologique : habitats <i>a priori</i> et <i>a posteriori</i> .....	187
2.1. Découplage des composantes statiques et dynamiques de la niche .....	188
2.2. Notion d'habitat potentiel et modèle régressif.....	189
2.3. Intérêt d'une définition arbitraire de l'habitat potentiel.....	191
3. Perspectives .....	192
3.1. Pistes d'approfondissement méthodologique .....	192
3.2. Inventaire floristique de la Drôme .....	195
4. Conclusion.....	196
Bibliographie .....	201
Annexe 1 : Plate-forme de simulation <i>SimulMetapop</i> .....	212
Annexe 2 : Sélection des variables environnementales pour le département de la Drôme .....	214
1. Catégories géologiques.....	214
2. Catégories d'occupation des sols.....	215
3. Traitement des données météorologiques.....	217
Annexe 3 : Estimation des paramètres de dynamique spatiale des plantes de la Drôme.....	xxx
Annexe 4 : Spectral analysis of distribution maps provides insights into metapopulation dynamics .....	218



- J'ai aussi une fleur.
- Nous ne notons pas les fleurs, dit le géographe.
- Pourquoi ça ! C'est le plus joli !
- Parce que les fleurs sont éphémères.
- Qu'est-ce que ça signifie : « éphémère » ? (...)
- Ça signifie « qui est menacé de disparition prochaine ».
- Ma fleur est menacée de disparition prochaine ?
- Bien sûr.

Ma fleur est éphémère, se dit le petit prince, et elle n'a que quatre épines pour se défendre contre le monde ! Et je l'ai laissée toute seule chez moi !

Antoine de Saint Exupéry (1943). *Le Petit Prince*.

## Introduction

« Pourquoi une population de plantes (ou de papillons, ou d'oiseaux...) se trouve-t-elle en un endroit donné, et pas dans une multitude de sites semblables où l'espèce est absente? » C'est une question que tout naturaliste s'est posée un jour ou l'autre. Je prendrais l'exemple simple d'une communauté de laïches arctico-alpines que je recherchai en 2004 dans les alluvions morainiques des hautes montagnes de Savoie. Des heures de prospection dans un habitat remarquablement uniforme furent nécessaires avant de trouver quelques mètres carrés de ces précieuses plantes. La première idée qui me vint à l'esprit est l'existence de fortes perturbations du milieu, lors des crues des ruisseaux, engendrant de forts taux d'extinction locale. Ainsi ces plantes doivent sans cesse recoloniser les zones rasées par les flots. Mais à conditions égales de perturbation, je constatai également que certaines plantes colonisaient plus vite que d'autres les zones libérées, de petits joncs précédant en général les laïches bicolores et maritimes. Ainsi la capacité de colonisation des différentes espèces déterminait également leur présence à l'échelle de ces alluvions.

Cette capacité à occuper le milieu, en résistant aux perturbations et/ou en colonisant rapidement les sites favorables est une caractéristique adaptative importante qui détermine le devenir à long terme des taxons. L'occupation actuelle de l'espace disponible par les organismes représente un champ d'investigation majeur de l'Ecologie. Cela implique différents types de phénomènes, depuis l'échelle locale et les questions autécologiques jusqu'à l'échelle biogéographique. Dans quelle mesure la distribution spatiale des habitats et des populations est-elle porteuse d'information sur ces processus ?

Une telle question, ancienne et fondamentale en Ecologie, aiguise toujours les débats théoriques (Kareiva and Wennergren 1995), et son traitement fait sans cesse référence à de nombreux paradigmes fondamentaux, que nous évoquerons un peu plus tard (Chap. I, §2). Dans ce contexte, la mise en relation des patrons observés avec les processus écologiques demeure une réelle gageure scientifique (cf. par ex. Murrell et al. 2001). Il semble que le problème majeur soit de savoir lire et comprendre l'information disponible, et de ne pas aller au-delà de ce qu'il est possible de dire. Il faut donc partir en quête d'une « bonne » mesure de la structure spatiale et temporelle des entités biologiques (individus, populations,



communautés), c'est à dire une mesure susceptible de mettre en exergue la signature spatiale des processus écologiques dynamiques.

En définitive, à l'instar du botaniste cherchant les laïches arctico-alpines au milieu des moraines, je suis amené à me poser deux questions :

- Comment mesurer de manière écologiquement pertinente la structure spatiale des populations ?
- Comment relier cette structure aux processus écologiques sous-jacents ?

## 1. Dynamique spatiale des populations

Afin de connaître la distribution spatiale des espèces, les naturalistes recueillent en général des données d'occurrence sur un réseau de stations, appelés souvent « pointages ». Ils renseignent alors la présence d'une population en un point donné, sans fournir davantage d'information sur son état démographique.

Dans un tel contexte, la perspective centrale de mon travail repose sur la question suivante : **que peut-on inférer sur la dynamique de renouvellement des populations, dans l'espace et dans le temps, à partir de simples données d'occurrences ?**

J'ai choisi d'appréhender d'abord cette question d'un point de vue théorique, afin de clairement définir les concepts en jeu, et d'établir si la dynamique spatiale et temporelle d'extinctions et de colonisations de populations peut induire des structures particulières. L'objet de cette démarche est d'évaluer l'importance de tels phénomènes pour la compréhension de l'écologie et de l'évolution des espèces. Fort des enseignements de cette recherche théorique, j'ai pu alors construire un support analytique simple et robuste pour étudier dans un cas concret la distribution d'occurrences d'espèces de plantes.

### 1.1. *L'Atlas de la Flore de la Drôme*

Le système auquel je m'intéresse est un espace écologique régional, au sein duquel des populations de plantes sont répertoriées et localisées dans le temps et dans l'espace. Un tel inventaire a été réalisé sous la direction de Luc Garraud dans le département de la Drôme, sur une dizaine d'années. Les inventaires floristiques réalisés ont donné lieu à la publication d'un Atlas (Garraud 2004). Il y a ainsi, en ne conservant que les données fiables et validées, un ensemble de 11249 points d'échantillonnage sur tout le département (voir Figure 1), pour un

ensemble de plus de 2700 taxons recensés. Il s'agit là de l'un des jeux de données floristiques qualitativement et quantitativement les plus complets de France.

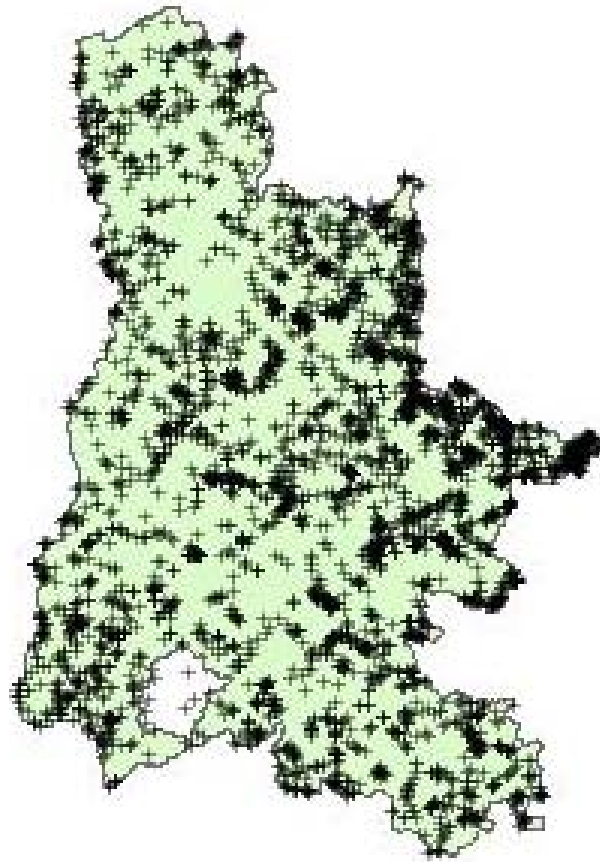


Figure 1 : Echantillonnage floristique effectué par Garraud (2004). Les 11249 points de relevé sont représentés, incluant ceux effectués dans l'enclave du Vaucluse.

Quoi qu'il en soit, nous sommes en présence d'une somme descriptive d'informations floristiques dont nous ignorons *a priori* la qualité et le sens écologique. Or à une telle échelle spatiale, conduire des expérimentations systématiques sur la dynamique des populations serait impossible. L'idée d'inférer la nature des processus structurant la distribution spatiale des espèces, à partir des données d'occurrence, constitue donc une solution alternative importante. Une telle recherche a été largement promue par les théories récentes mettant en jeu la modélisation de processus dynamiques à différentes échelles spatiales (métapopulations, Hanski 1998; métacommunautés, Hubbell 2001). Dans la limite des résultats de notre recherche théorique sur ce thème, les données floristiques de Garraud (2004) sont les plus à même, par leur qualité, d'apporter des informations substantielles sur les processus écologiques dynamiques qui nous intéressent.

Nous partons en définitive de l'idée que l'information de présence / absence des espèces dans l'espace, sur un ensemble de points d'échantillonnage, doit être porteuse d'information sur leur écologie, car elle est le résultat de la dynamique dans le temps et dans l'espace des populations de l'espèce. Contrairement à la réalisation d'une expérience, nous ne contrôlons pas ici les conditions d'évolution du système, et de nombreux mécanismes inconnus au premier abord sont impliqués, ainsi que des phénomènes contingents liés à la biogéographie et à la présence de l'homme.

Bien que le corpus théorique et les moyens techniques soient en l'occurrence issus d'une évolution récente de la science écologique, il est utile de rappeler que ce type de préoccupation a motivé dès l'origine les promoteurs de la géographie botanique. Ainsi les termes employés par le célèbre botaniste De Candolle dans son traité de géographie botanique (De Candolle 1855) ne manquent pas de modernité : « Une espèce étant donnée dans un pays quelconque, on voit aisément qu'elle est douée de moyens de diffusion très nombreux, très vivaces, mais qu'elle est contenue dans certaines limites géographiques : 1° par des obstacles matériels, comme la largeur des bras de mer et l'étendue des océans ; 2° par des conditions de climat qui l'empêchent de vivre ou de se reproduire au-delà de certaines limites. Les plantes surmontent quelquefois les obstacles matériels, grâce à leurs moyens de dissémination et aux transports accidentels, provenant de l'homme, des animaux, des vents ou des courants ; mais elles ne sauraient vaincre l'action continue d'un climat contraire, de sorte que sur la ligne où s'engage avec lui le combat, c'est toujours, ou du moins c'est, à la longue, toujours le climat qui reste victorieux. »

Nous verrons que les organismes possèdent des capacités biologiques leur permettant d'occuper certains types d'habitat. Nous verrons qu'afin d'accéder aux espaces favorables, ils sont dotés de moyens de dispersion et de colonisation. Au cours du temps, nous pouvons imaginer, à l'instar de De Candolle, l'action d'organismes qui conquièrent peu à peu l'espace favorable. Mais il y a des obstacles, des chemins difficiles et longs pour atteindre et occuper tous les espaces disponibles. Et le climat, ainsi que d'autres variables importantes de l'environnement, changent, et en changeant ces variables apportent des contraintes nouvelles à la dynamique des espèces. **La question du temps écologique sera au cœur de mon travail ; c'est ce temps qui détermine dans quelle mesure une espèce peut occuper un espace disponible, ses caractéristiques biologiques étant figées et la structure de l'habitat étant statique.** Il se distingue clairement du temps évolutif, où les exigences des espèces

elles-mêmes changent, conduisant à la conquête de nouveaux milieux. Les variables dynamiques nombreuses et enchevêtrées nous interpellent. La question des échelles de temps et d'espace sera une clé essentielle pour pouvoir comprendre les couplages et découplages possibles entre processus écologiques et évolutifs.

### *1.2. Vision dynamique de la niche et métapopulations*

Dans le temps et l'espace écologique, les objets que nous considérons sont les populations des espèces, dont on connaît la présence ou l'absence grâce aux inventaires sur le terrain. Considérons une population en un point d'échantillonnage donné, et à un instant donné. Cette population sera susceptible de disperser des propagules dans les espaces alentours, de recevoir des propagules des espaces alentours, ou bien encore de s'éteindre.

En fonction des exigences de l'espèce, les variables environnementales permettent de définir des conditions favorables au sens de la niche réalisée pour permettre le maintien démographique de la population. Si les conditions environnementales sont défavorables, la population est vouée à disparaître. Ce sont des conditions déterministes de subsistance (paradigme de la niche, §2.3).

A ces conditions s'ajoute la stochasticité des variables environnementales, qui peuvent devenir défavorables temporairement et compromettre le maintien de l'espèce, ainsi que la stochasticité démographique qui peut induire l'extinction de la population. La viabilité de la population dépendra des valeurs des variables environnementales au site considéré, et des propres traits démographiques de l'espèce. Les régimes stochastiques de perturbation affectent donc le taux d'extinction des populations. Dans le même temps, les propagules vont pouvoir établir de nouvelles populations avec une certaine probabilité.

Dans ce contexte, « les choses bougent » sans arrêt : il y a en permanence des populations qui disparaissent, et d'autres qui apparaissent. La distribution observée des populations à un instant donné est donc le produit complexe de tels processus stochastiques et déterministes, et notre propos est d'éclairer la nature de ce lien. C'est dans cette logique qu'a été introduite la notion de *métapopulation* (Encadré 1).

#### Encadré 1 : Le concept de métapopulation

Une métapopulation est un ensemble de populations en interaction, connectées par la dispersion d'individus entre populations. Au cours du temps, au gré d'événements d'extinction et de colonisation, des populations apparaissent et disparaissent. Sur la base d'un réseau de sites occupables par des populations, on s'attache alors à connaître la densité de sites occupés à un moment donné. Le comportement dynamique d'un tel réseau de populations a été étudié d'abord par Levins (1969), par le biais d'un modèle simple qui sera présenté au Chapitre I. L'utilisation du concept s'est depuis considérablement étendue, notamment comme support à des modèles d'évolution de la dispersion (voir Olivieri et al. 1995). Toutefois, tout réseau de populations ne doit pas être considéré comme une métapopulation. En fait le concept de métapopulation recouvre plusieurs hypothèses importantes :

- A l'échelle de la métapopulation, la fréquence des événements d'extinction et de colonisation est suffisamment faible pour que le temps d'arrivée des populations à leur équilibre démographique soit faible devant les temps moyens de colonisation et d'extinction. Cette hypothèse permet donc de dissocier l'étude de la dynamique démographique de l'étude de la dynamique d'extinction/colonisation.
- Les processus d'extinction et de colonisation sont aléatoires, et n'impliquent pas en particulier de préférence d'habitat. Le problème se posera en particulier chez les animaux, en cas de dispersion directionnelle. En outre, la représentation repose sur une représentation uniforme de la qualité d'habitat, pour la persistance et la reproduction. D'éventuelles hétérogénéités de qualité peuvent être prises en compte indirectement dans les modèles récents (Ovaskainen and Hanski 2001).

Nous utiliserons et discuterons ce formalisme, qui constituera notre modèle écologique de référence.

Nous situons notre propos vis à vis de trois grandes familles de processus régissant la distribution spatiale des espèces :

- Déterminisme de la niche. L'habitat correspond à la réponse déterminée d'une espèce aux conditions environnementales. Cette réponse reflète des exigences spécifiques, dépendant directement de la biologie de l'organisme considéré. C'est la niche réalisée de Hutchinson (1957). La distribution des populations est alors déterminée directement par

la configuration spatiale de l'habitat physique, par le biais des exigences intrinsèques de l'espèce. C'est le principe repris par les nombreux modèles récents d'habitat potentiel (Austin 2002). Nous soulignerons que l'utilisation sans discernement de ce principe, négligeant l'effet de tout processus dynamique, est susceptible de conduire à des conclusions erronées sur les préférences écologiques des espèces.

- Dynamique stochastique des populations. Dans un modèle de type métapopulation, la dynamique des populations dépend de deux types d'événements : colonisation et extinction. En outre, pour un même réseau de milieux favorables, la dynamique des espèces variera en fonction des caractéristiques biologiques propres. Nous supposons donc que les événements de colonisation-extinction sont une réponse intrinsèque à ces caractéristiques spécifiques. Il y a donc un contrôle biotique qui détermine le degré d'occupation de l'espace potentiellement favorable. Notre approche se situera à ce niveau de perception, puisque c'est celui que l'on peut observer (par exemple sur un inventaire comme celui de la Flore de la Drome). Pour des modèles à contrôle biotique, nous soulignerons que la structuration induite sur un temps court et une étendue limitée par les événements de colonisation-extinction présente des caractéristiques particulières. Celles-ci traduisent la réponse du système à d'incessantes perturbations stochastiques, d'amplitudes variables, dites *contingentes* (voir Encadré 2), qui entraînent l'apparition de structures éphémères. Sur un temps très long, les modèles stochastiques négligent de telles structures fugaces. Mais dans la réalité de notre système écologique, où la gamme de temps et d'espace est très restreinte, ce sont des structures qu'il est important de comprendre. On dit qu'elles *émergent* du fonctionnement local des populations. Leur comportement est résumé par le concept d'*auto-organisation* : il s'agit d'une émergence spontanée et hors équilibre de structures macroscopiques, résultant du comportement interactif d'un grand nombre d'entités élémentaires (Coveney 2003), en l'occurrence ici de populations.
- Facteurs historiques. Certains événements contingents sont très rares et d'amplitude exceptionnelle. Dès lors, ils apparaissent comme des événements ponctuels dans l'espace et dans le temps écologiques. Par leur impact important sur les écosystèmes, ces événements sont susceptibles de marquer profondément la distribution spatiale ultérieure des populations. Par ailleurs, ils n'engendrent pas clairement de structure auto-organisée identifiable, qui résulterait de la répétition d'événements comparables. On entre alors

dans une représentation *historique*, où des évènements contingents ponctuels, identifiés et exceptionnels, jouent un rôle crucial, sans être clairement récurrents. Dans le cadre de modèles d'occupation déterministes (niche) ou stochastiques (dynamique des populations), ces événements constituent un bruit non structuré dans l'espace et dans le temps, autour d'une distribution attendue « à l'équilibre » : ils réduisent l'information disponible sans en changer la nature. Nous estimerons toutefois la sensibilité de nos méthodes d'analyse sur des situations actuelles que nous savons fortement déterminées par des facteurs historiques. C'est le cas par exemple de l'extension altitudinale des pelouses de montagne, liée au pâturage, ou des phénomènes d'invasion, associés à des introductions récentes très généralement favorisées par l'Homme. On se trouve en l'occurrence souvent dans des situations transitoires (voir Encadré 3). Toutefois, même à l'arrivée à l'équilibre, la position initiale du (ou des) foyer(s) d'invasion influence la distribution spatiale à l'équilibre, en fonction par exemple des barrières écologiques rencontrées. Des phénomènes contingents à grande échelle (ouverture d'océans, surrection de montagnes) ont pu ainsi être des moteurs de diversification, et occasionner notamment des processus de vicariance. C'est un thème essentiel de la biogéographie et de la biologie évolutive.

#### Encadré 2 : Qu'est-ce qu'un évènement contingent ?

Il semble presque impossible de trouver des publications scientifiques qui définissent, quand elles en font référence, la notion de contingence. C'est en fait un concept philosophique dont la définition première peut paraître ésotérique : est contingent un évènement qui est « susceptible d'être ou de ne pas être, de se produire ou de ne pas se produire » (Trésor de la Langue Française informatisé, ATILF 2004). Cela s'oppose aux concepts de nécessité et de déterminisme.

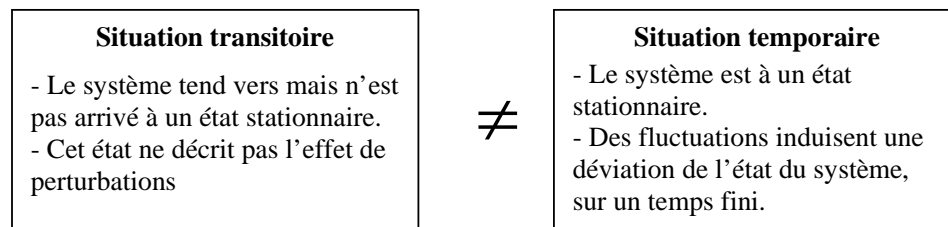
Dans *La Nausée*, J-P Sartre fait la réflexion suivante sur la contingence : « L'essentiel c'est la contingence. Je veux dire que, par définition l'existence n'est pas la nécessité. Exister c'est être là simplement, les existants apparaissent, se laissent rencontrer, mais on ne peut jamais les déduire. » C'est une bonne définition pour le cadre de notre travail, une définition *existentialiste*. La contingence est ancrée dans l'instant et le local, dans l'existence individuelle des populations ; c'est un ensemble d'événements imprévisibles, mais néanmoins inéluctables, dont l'incidence sur le devenir du système métapopulation nous intéressera tout particulièrement.

### Encadré 3 : Situations temporaires vs. transitoires

Pour la suite de mon travail, et pour éviter toute confusion, nous distinguerons strictement l'usage des termes *temporaires* et *transitoires*, de la manière suivante :

- Une **situation transitoire** représente une phase dynamique où **le système n'est pas parvenu à un état stationnaire**. Dès lors les propriétés du système ne sont pas décrites par les propriétés d'un modèle à l'équilibre.
- Une **situation temporaire** décrira en revanche l'effet de fluctuations, ou de perturbations, sur le comportement du **système à l'état stationnaire**. Leur ampleur et leur incidence pourront varier, et nous nous intéresserons plus particulièrement à leur prise en compte au cours de mon travail. Cette terminologie correspond donc à l'étude d'un équilibre stable, tel que les déviations du système soient décrites sur des périodes de temps finies.

Nous attirons particulièrement l'attention du lecteur sur cette convention, résumée par le schéma suivant :



Pour exprimer ces différentes conceptions d'habitat, j'introduis deux représentations différentes, appelées habitat *a priori* et habitat *a posteriori* (Figure 2). Cette terminologie est une convention créée et utilisée dans le cadre de mon travail. L'idée est de traduire le passage d'une échelle de compréhension des processus à une autre. Le concept statique de la niche, au sens des niches fondamentale et réalisée, traduit une perception centrée sur l'individu. Ce que représente la niche, ce sont des exigences individuelles, liées aux caractéristiques biologiques de l'organisme concerné. Cela constitue à ce titre une représentation autécologique. Dès lors, l'habitat *a priori* exprime spatialement ces exigences individuelles, et extrapole l'espace occupable par l'espèce. On passe ainsi d'une représentation centrée sur l'individu à une représentation centrée sur l'espèce. Cet habitat est l'expression *a priori* d'un espace potentiellement occupable par l'espèce, au regard des exigences biologiques des individus.



La configuration spatiale des ressources nécessaires à la survie des individus représente de fait une contrainte forte pour leur dynamique spatiale. Nous nous intéressons donc à la dynamique spatiale et temporelle des populations au sein de cet espace (concept de métapopulation). L'espace réellement occupé par les populations **dans un temps écologique fini** est un sous-ensemble de l'habitat *a priori*, appelé habitat *a posteriori*. C'est un espace réellement occupable sur un temps considéré, calculé *a posteriori* d'après un modèle dynamique.

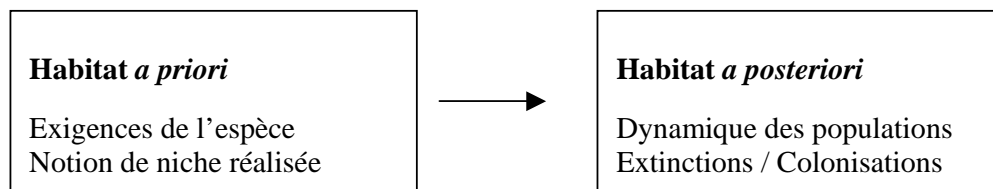


Figure 2 : Formalisation théorique de l'habitat. L'habitat *a priori* est conçu comme la représentation spatialisée de la niche réalisée de Hutchinson. Il est déterminé directement par la structure spatiale des variables environnementales. A cela se superposent des processus stochastiques de colonisation / extinction qui induisent une occupation partielle de cet habitat. L'habitat occupé au sein de l'habitat *a priori* est l'habitat *a posteriori*.

Le seul moyen direct de connaître les exigences d'une espèce (niche fondamentale et réalisée) est la planification de manipulations contrôlées, lourdes, à grande échelle et impliquant un grand nombre de facteurs d'habitat. On utilise donc en général les informations d'occurrence et de non-occurrence pour établir indirectement un lien entre l'occupation de l'espace et la variation des variables environnementales. Ce sont des modèles causaux, de type modèle linéaire en général.

L'information d'occurrence et de non-occurrence peut être ainsi mise en relation avec la variation des conditions environnementales. Mais cette information porte aussi la signature des processus dynamiques, en particulier concernant la colonisation et l'extinction des populations. Un problème crucial, tant sur le plan de la théorie écologique que sur celui des techniques statistiques, est d'être en mesure d'identifier l'effet de ces deux composantes de nature différentes (cf. Ives et al. 1998; Svenning and Skov 2002). Une espèce très localisée pourra être considérée comme possédant un habitat très restreint et une forte persistance locale, ou bien un habitat plus général et un fort taux d'extinction locale : le choix de l'une ou l'autre des interprétations risquera d'être équivoque. Des espèces très dispersantes pourront *a contrario* être trouvées dans des conditions non favorables à leur persistance mais se

maintenir néanmoins par un apport régulier de propagules (système puits-source, voir Dias 1996) : là encore, il y a un biais possible d'estimation de la niche.

Dans cette perspective, l'objet de mon travail est **d'étudier dans quelle mesure la distribution spatiale des populations peut nous renseigner simultanément sur les composantes de la niche de l'espèce, et sur les caractéristiques de sa dynamique spatiale.** Il s'agit donc d'éclaircir le risque d'équivocité de l'information spatiale.

Pour mettre en œuvre cette démarche, il convient de bien identifier la nature des hypothèses et des paradigmes impliqués. En effet la compréhension de la structure de causalité fait appel à certains concepts et à certains outils dont nous verrons que les choix sont loin d'être anodins. C'est pourquoi, il apparaît opportun de consacrer une partie de l'Introduction à la mise en perspectives des paradigmes clés auxquels est associée ma problématique de travail (§2). Il est également important de clarifier quels angles d'approche il est possible de choisir pour comprendre la structure de causalité entre processus locaux et structures spatiales résultantes (§3).

## **2. Quatre paradigmes en question**

Voici en prélude quelques notions et paradigmes clés ayant motivé et orienté la conduite de ma recherche. Il s'agit d'attirer l'attention sur la logique même de construction des modèles écologiques en général, en soulignant l'importance de quatre paradigmes. Cette logique en elle-même sera discutée tout au long de mon travail, car les choix de représentation et de modélisation de la réalité écologique auront une importance capitale dans le traitement de mes questions scientifiques.

*Paradigme* : Conception théorique dominante ayant cours à une certaine époque dans une communauté scientifique donnée, qui fonde les types d'explications envisageables, et les types de faits à découvrir dans une science donnée.  
(Trésor de la Langue Française informatisé, ATILF 2004)

### *2.1. Cadre de base : système, objets, états*

Au cœur de la démarche scientifique est la formulation de questions, qui motive et oriente la mise en place d'expériences et permet d'organiser les observations. S'agissant de traiter

une question, il faut d'abord définir un système d'étude, mettant en jeu des objets dont on mesure un ou plusieurs états. Le système est caractérisé par un ensemble de variables d'environnement  $H_i$  sur l'ensemble de son étendue. Il peut être ouvert s'il peut y avoir une incidence d'éléments extérieurs au système, il est fermé sinon. On s'intéresse alors aux processus qui régissent l'état des objets au sein du système, en général dans le temps et dans l'espace. Le système d'étude regroupe donc les objets dans un espace clairement délimité (en général deux dimensions spatiales  $x$  et  $y$ , une dimension temporelle  $t$ ), et avec des règles de fonctionnement et d'interactions définies. La manière de définir le système et les objets, ainsi que les règles d'interaction, constituent un schéma conceptuel dans lequel le chercheur se place. Considérons par exemple un système stochastique classique en démographie : le système est une population, les objets sont des individus, les états les probabilités des individus de se trouver à un certain âge, et les règles de fonctionnement permettent de définir les probabilités de transition entre classes d'âge.

Dans ce cadre conceptuel, le chercheur formule une théorie, dont il met en question les attendus à l'aide d'observations et d'expériences. On considère classiquement, dans le cadre poppérien falsificationniste, qu'une théorie n'est scientifique que si elle est falsifiable, donc si l'on peut infirmer les attendus (Popper 1959).

## *2.2. L'entité biologique en écologie : paradigme biologiste*

L'écologie est revendiquée être une science biologique, et dès lors ses systèmes d'étude sont centrés sur l'organisme biologique (objet), confronté à un environnement dans lequel il évolue et avec lequel il interagit (système). C'est sur la base de telles interactions que l'on souhaite étudier la dynamique des populations d'individus conspécifiques (écologie des populations), des communautés d'individus de différents taxons (écologie des communautés), l'existence de fonctions caractérisant les organismes (écologie fonctionnelle), ou définir une formalisation générale des flux d'énergie et de matière à travers les individus en interaction (écologie écosystémique) (Barbault 2000).

Il y a ainsi différentes approches orientées vers la compréhension de la place de l'entité biologique dans son environnement par le biais de ses fonctions et de ses interactions directes. Le cœur même de l'écologie est donc de donner un sens à l'immense diversité biologique, de formes et de fonctions, que les naturalistes ont soigneusement répertoriée.

La *macroécologie* est un domaine de l'écologie qui pose cependant des problèmes différents (Marquet 2002; Nee 2002). En effet, il se caractérise par un système d'étude très étendu, dans le temps et dans l'espace : cela tend à démarquer ses problématiques de celles d'une approche purement organisme-centrée. La macroécologie traite notamment de grandeurs plus abstraites, appelées diversité biologique, distributions d'abondance et de fréquences, qui mettent en jeu un grand nombre d'entités biologiques différentes. La théorie macroécologique observe des lois « universelles » et empiriques qui semblent régir ces grandeurs sur toute la planète. Dans un tel système d'étude, où l'organisme biologique est une particule élémentaire, les caractéristiques macroscopiques sont largement déconnectées des caractéristiques individuelles des organismes. Dans la théorie neutraliste des communautés (Hubbell 2001), elles traduisent plutôt des caractéristiques dynamiques globales des communautés.

Cette philosophie holiste demeure souvent contestée, la signification biologique et la falsifiabilité des lois à grandes échelles étant critiquées (McGill 2003). Le fait de ne pas pouvoir relier directement une loi macroscopique à des mécanismes biologiques clairement identifiables est alors interprété comme une invalidation de la loi. On peut cependant s'interroger sur le fait que les interactions individuelles entre organismes puissent être directement interpolées pour expliquer l'ensemble des patrons observés de biodiversité, à toutes les échelles. Des structures émergentes et temporaires peuvent apparaître à une échelle intermédiaire de temps et d'espace, et présenter des caractéristiques cohérentes à grande échelle, sans être le reflet direct et déterminé des processus biologiques locaux (Barabasi and Albert 1999): c'est l'auto-organisation. C'est une problématique importante en physique (Ciliberti et al. 2000). Néanmoins une telle conception en biologie tend à introduire comme contraintes dans la survie des populations, et donc des espèces, d'autres forces que la sélection naturelle individuelle et gène-centrée (Stewart 2003); cette vision élargie reste de ce fait encore peu répandue, d'autant plus qu'elle apparaît difficile à mettre en œuvre de façon rigoureuse (voir notamment Johnson and Boerlijst 2002; Ulanowicz 2002; Leoncini and Rivault 2005).

A l'ombre du néo-darwinisme, le paradigme biologiste réductionniste ne doit pas devenir un dogme qui occulterait les processus interactifs auto-organisés susceptibles de jouer un rôle important dans la structuration spatiale et temporelle des organismes. La notion de complexité écologique est au cœur de la compréhension holistique de ces phénomènes, où un

grand nombre d'éléments discrets (individus ou populations) sont en interaction et exhibent dans leur ensemble des structures particulières (Li 2000b; Levin 2002; Wu and David 2002).

A titre d'exemple, dans l'étude de l'évolution du sexe, le fait que les processus de sélection au niveau individuel ne puissent pas seuls expliquer la prédominance des organismes sexués a cristallisé un antagonisme fort entre les tenants réductionnistes de la sélection individuelle et les tenants holistes de la sélection de groupe. Mais une attitude sage serait clairement dans une prise en compte raisonnée de l'ensemble de ces échelles (Gouyon 1999), et en définitive dans la compréhension des liens dynamiques entre les niveaux individuel et spécifique, par le biais notamment de l'auto-organisation et des structures émergentes.

### *2.3. Déterminisme vs. stochasticité : paradigme de la niche*

La survie des organismes dépend d'une biochimie complexe répondant à certaines propriétés du milieu physique. Une plante qui n'est pas adaptée à la survie en milieu salé ne pourra pas perdurer dans un tel milieu et sera rapidement éliminée, car elle sera incapable de réaliser les fonctions biologiques vitales sous l'effet de la toxicité du sel et/ou de la pression osmotique du sol. C'est ainsi que l'on définit des plantes croissant en milieu salé, dites halophytes, et des plantes ne tolérant pas le sel, dites glycophytes.

Il en est ainsi concernant la relation des organismes avec de nombreuses variables physiques du milieu : disponibilité en eau, températures, nature du substrat... La diversification des stratégies de survie des organismes en fonction de conditions abiotiques variables peut constituer une source de diversité des organismes biologiques (Kisdi 2002; Via 2002). Les conditions physiques de l'environnement permettant la persistance d'une espèce à un endroit donné définissent sa niche fondamentale (Hutchinson 1957). C'est une relation déterministe entre des variables d'environnement et la présence potentielle d'une espèce.

La survie d'une espèce dans une communauté dépend aussi des interactions entre les individus. Les individus, qu'ils soient conspécifiques ou non, sont en compétition pour les ressources qu'ils partagent, eau, nutriments, lumière, espace. Il peut dans certains cas y avoir des interactions positives (facilitation si c'est dans un seul sens, mutualisme dans les deux). Ces interactions entre organismes influent sur leur capacité à survivre. C'est pourquoi on définit pour une espèce sa **niche réalisée** comme l'ensemble des conditions biotiques et

abiotiques nécessaires à sa persistance. C'est encore une vision déterministe de la relation entre l'espèce et son environnement local.

Maintenant, si l'on considère la distribution des espèces dans une région donnée, on constate en général que cette niche dite « réalisée » n'est pas complètement occupée, ou au contraire que des zones qui devraient être normalement en dehors de la niche réalisée sont occupées (systèmes source-puits, Dias 1996). Dans ces cas-là, la présence de l'organisme dans un lieu donné ne dépend pas simplement de ses exigences locales, mais aussi de phénomènes dynamiques qui ont permis à un ou à des individus de s'implanter à cet endroit. Il y a alors variation temporelle de la présence de l'espèce, même si les conditions sont en moyenne favorables. La stochasticité du milieu, de la démographie de l'espèce et des autres espèces induit des extinctions locales. A l'équilibre, ces extinctions sont compensées par l'émission de propagules susceptibles de coloniser ou de recoloniser des sites favorables. Ainsi, au-delà de la vision déterministe de la niche, basée sur le déterminisme des exigences de l'espèce, il faut donc prendre en compte l'effet des processus stochastiques dans la structuration des populations et des communautés (cf. Pulliam 2000). Ce sont des processus stochastiques de type Vie-Mort, qui peuvent intervenir pour différents types d'objets : au niveau des individus (démographie), des populations (métapopulation), des communautés (métacommunauté).

## *2.4. Paradigme de l'échelle et paradigme de l'équilibre*

### *2.4.1. Echelle des objets*

Dans le système d'étude en écologie, le temps et l'espace sont en général des grandeurs discrètes et délimitées :

- Ces grandeurs sont discrètes, car il existe une entité élémentaire dans l'espace (l'objet = l'individu, la population, la communauté) et dans le temps (unité de temps = année ou génération par exemple).
- Ces grandeurs sont délimitées, car il existe une plage de temps d'étude du système (quelques générations par exemple), et une délimitation spatiale (échelle de la station, du paysage, de la région).

Dans le modèle de métapopulation dans un espace fini, la référence spatiale des populations est discrète. On pourrait envisager une représentation différente, avec une représentation continue de l'espace, et une variation continue spatialement de la probabilité locale d'occupation. Il existe de tels modèles continus de dynamique spatiale populationnelle, s'apparentant à des processus de type réaction-diffusion (Cantrell and Cosner 2003).

Ceux-ci présentent en général l'avantage de permettre des traitements mathématiques plus simples et plus puissants (voir Law et al. 2003). Ils seront en outre pertinents dans des situations où les structures spatiales résultent essentiellement de processus au niveau individuel, dans des milieux très contraints notamment (HilleRisLambers et al. 2001). Cela ne rend toutefois pas compte de l'existence de hiérarchies écologiques d'entités discrètes, comme les concepts de population (Berryman 2002) et de métapopulation l'expriment (Levins 1969). Dès lors, on s'attend à ce que le comportement du système soit par nature très différent avec une représentation discrète ou continue de l'espace, et l'on peut s'attendre à prédire des propriétés différentes: nous reviendrons sur cette idée dans le cadre de l'auto-organisation (voir Chapitre II).

Au sein d'une représentation discrète des états, la définition de *l'échelle des objets* et du système sera en particulier primordiale pour les choix de modélisation : le devenir théorique du système dépend de l'échelle choisie de représentation (Dungan et al. 2002).

#### 2.4.2. Equilibre du système

Un aspect crucial de la modélisation d'un processus dynamique est l'étude, s'il existe, de l'équilibre du système. Pour résumer, l'état d'un objet du système est le produit de processus stochastiques et déterministes. Quand on s'intéresse aux caractéristiques biologiques des organismes, on souhaite en général s'affranchir de la variation spatiale et temporelle des distributions induites par ces processus, pour se concentrer sur des situations où les états des objets ne dépendent ni de l'espace, ni du temps. On s'intéresse alors à des états à l'équilibre, tels qu'en général on puisse s'abstraire de leur dépendance temporelle et/ou spatiale (notion de stationnarité). Ce sont des approximations, des simplifications théoriques afin de pouvoir appréhender à l'aide d'idées et d'outils simples le comportement des systèmes écologiques (O'Neill 2001). Dans les modèles d'équilibre classiques, c'est l'état des objets qui est traité à l'équilibre par rapport à leur environnement, et ce traitement dépend donc de l'étendue spatiale et temporelle des objets.

Dans tous les cas, on cherche donc à caractériser le système à l'équilibre au travers de grandeurs simples, synthétiques, et notamment constantes dans le temps et/ou l'espace. Cette démarche a des vertus prédictives, mais est une gageure pour l'inférence. En effet, si l'on observe un système sur une étendue limitée et à une date donnée (par exemple, distribution des plantes de la Drome à un instant  $t$ ), on ne connaîtra qu'une seule réalisation du (ou des) processus stochastique(s) qui nous intéresse(nt) (par exemple, dynamique des populations de plantes de la Drome). L'accès aux grandeurs synthétiques caractérisant l'état stationnaire sera problématique, ou du moins fortement teintée d'incertitude. Un test d'arrivée à l'état stationnaire sur cette seule base serait nécessairement faible. Avec une représentation trop simplifiée du système à l'état stationnaire, l'hypothèse d'arrivée à cet état sera dès lors souvent arbitraire et non falsifiable.

Si l'on décide maintenant de changer la nature des objets élémentaires en les agrégeant par exemple (en traitant de populations au lieu d'individus par exemple), on change aussi la définition des états des objets, et on change évidemment la nature des équilibres. **La notion d'équilibre est donc échelle-dépendante.** Dans les modèles continus, on se libère de cette dépendance en supprimant la notion de discrétisation des objets.

On parle en outre de systèmes « scale-free » quand les propriétés du système sont invariantes en changeant l'échelle des objets ; cela existe aussi bien en représentation discrète que continue. On entre alors dans le domaine des structures auto-similaires, fractales et multi-fractales, qui suscitent notamment un engouement important en écologie du paysage (Li 2000a; Bolliger et al. 2003). Les phénomènes d'auto-organisation sont en général invoqués pour expliquer de telles structures (Barabasi and Albert 1999). Les grandeurs auto-similaires ne présentent pas d'échelle de variation caractéristique. Elles vont présenter alors des fluctuations à toutes les échelles, et typiquement s'exprimer sous forme de lois de puissance (Barabasi and Albert 1999). La notion d'équilibre doit alors être adaptée à une telle absence d'échelle référence, et être donc basée sur des statistiques et des lois « échelles-indépendantes ». On s'intéresse notamment aux lois de puissance, nombreuses et caractéristiques des phénomènes auto-organisés (voir par ex. Barabasi and Albert 1999; Pascual et al. 2002). Il est important de souligner alors que l'échelle (ou l'absence d'échelle) des patrons observés ne définit pas nécessairement l'échelle des processus causaux. C'est ce qu'exprime l'idée d'émergence.



### 2.4.3. Stabilité du système

Sous-jacente à l'idée d'équilibre déterministe est la théorie des petites perturbations. Les petites perturbations autour de l'équilibre engendrent une réponse linéaire (approximation au premier ordre): une petite perturbation engendre un faible écart à l'équilibre, et un retour plus rapide. Se pose ainsi la question de la stabilité de l'état d'équilibre du système, ou en abrégé la question de la stabilité du système. Il est possible d'aller au-delà de la simple approximation des petites perturbations, afin d'accéder à une caractérisation plus fine de la réponse aux perturbations des systèmes à l'équilibre: les travaux de Lyapounov notamment sont en la matière fondateurs (voir par exemple Vidyasagar 2002). Ces questions se rattachent aux concepts de physique statistique et soulignent une distinction importante entre : (i) des **systèmes proches de l'équilibre**, dont le comportement reste lié à une logique de perturbation autour de cet état, et (ii) des **systèmes loin de l'équilibre mais présentant une forte stabilité**, dont les structures peuvent présenter un ordre remarquable (voir Prigogine 1977). Aussi la notion de stabilité doit-elle être formellement dissociée de celle d'équilibre.

Ces considérations demeurent aujourd'hui étrangères au monde de l'écologie, plus familier de la théorie des probabilités que des concepts de physique statistique. Dans cet esprit cependant, la modélisation stochastique permet une prise en compte approfondie et explicite des effets des perturbations autour des états à l'équilibre. De fait, des fluctuations aléatoires de plus ou moins grande ampleur sont susceptibles de conduire à la disparition de tout système, même à l'équilibre, fut-ce au terme d'un temps très long. L'étude du temps d'extinction détermine alors le degré de stabilité des états stationnaires stochastiques : un système est plus stable quand son temps d'extinction est plus élevé.

La stabilité du système écologique dépend en particulier de la complexité de sa structure, à différentes échelles spatiales et temporelles. C'est un fait particulièrement reconnu et étudié dans le cadre des chaînes trophiques (MacArthur 1955). La notion de stabilité nous conduit souvent en biologie à la notion de robustesse d'un réseau d'entités en interaction. La compréhension et la prédiction de la robustesse des réseaux constituent actuellement un thème de recherche majeur, en particulier pour les réseaux informatiques de type Internet. Elle implique une manière synthétique et intégrative de penser les systèmes, en particulier les systèmes écologiques (de Rosnay 1975; Proulx et al. 2005). Un principe clé du concept de métapopulation est qu'un réseau de populations, liées par des échanges de propagules, possède un comportement dynamique global qui lui confère un temps de persistance très supérieur à celui des populations individuelles (Levins 1969; Hanski and Gilpin 1997). Cette

hiérarchie de structures transitoires est au cœur du paradigme de la métapopulation en génétique des populations : elle détermine fortement la nature des stratégies de dispersion évolutivement stables (Olivieri 1990). Le réseau des populations est d'autant moins sensible aux perturbations locales que le couplage démographique est faible et que les interactions sont nombreuses : on retrouve la philosophie fondamentale de la théorie des métapopulations, qui est donc bien une question de robustesse de réseau. Par la suite, cette idée a été étendue en écologie au concept de réseau de communautés, ou métacommunautés (Hubbell 2001), et même aux réseaux d'écosystèmes, ou métaécosystèmes (Loreau et al. 2003). Notons que ces idées ont été développées en ingénierie des systèmes, que ce soit en automatique ou en mécanique, avec notamment les notions de degrés d'inertie et d'hyperstaticité. Ces termes suffisent à faire clairement écho à l'idée de robustesse des réseaux, que l'on cherche en l'occurrence à formaliser pour la mettre à profit.

## *2.5. Contingence et émergence : paradigmes en question*

### *2.5.1. Prise en compte des événements contingents*

Quand on parle d'état d'équilibre, on s'intéresse à une situation où le système est invariant. En l'occurrence nous nous intéressons à une invariance dans le temps et/ou l'espace. Les conditions d'existence et d'accession à cet état doivent être étudiées. En particulier, il peut y avoir plusieurs états d'équilibres, et la convergence vers un état possible d'équilibre n'est pas systématique.

On peut étudier et caractériser un état équilibre au moyen d'une formulation déterministe, avec éventuellement une connaissance requise des conditions initiales. On doit cependant adapter la démarche à la prise en compte de certaines contraintes importantes. L'hypothèse d'invariance du système constitue de fait une approximation très forte et abstraite : tout système est voué au changement (voir Encadré 4). Dans le cas de la dynamique des métapopulations, le temps d'extinction caractérise l'espérance de vie de la métapopulation. Il exprime une incidence globale des effets stochastiques sur le long terme. A plus courte échelle de temps et d'espace, les variations stochastiques peuvent également induire des régimes particuliers non modélisables par des modèles déterministes à l'équilibre, comme par exemple des régimes d'avalanches. Nous nous intéresserons spécifiquement à ceux-ci au cours du mémoire.

#### Encadré 4: Qu'est ce qu'un état quasi-stationnaire ?

Il est souvent fait référence à la notion d'état quasi-stationnaire dans l'étude des processus stochastiques possédant un état absorbant, en l'occurrence pour les métapopulations un état absorbant nul (extinction). L'utilisation du terme est dans ces contextes devenue systématique, au point qu'il est devenu très rare de trouver des publications qui prennent la peine de définir le terme. Il est toutefois important de bien clarifier sa signification pour la suite de mon travail (cf. Dickman and Vidigal 2002).

Considérons un processus stochastique qui présente un vecteur d'états  $N_t$  positif, qui possède  $N = 0$  comme état absorbant. Cet état est en fait le seul susceptible d'être absolument stationnaire. Considérons que le système admet également une description **macroscopique** par un autre état stationnaire actif (non nul et effectif)  $N_{st} > 0$  (les fluctuations sont donc négligées). L'idée est alors que la distribution de probabilité  $p_n(t)$  des états bifurque, après une situation transitoire (voir Encadré 3), en **deux composantes, une concentrée sur l'état absorbant, l'autre,  $q_n(t)$ , sur la valeur macroscopique stationnaire  $N_{st}$** . Cela s'exprime par

l'équation suivante :  $p_n(t) = \left(1 - \sum_{n \geq 1} q_n(t)\right) \delta_{n,0} + q_n(t)$ , avec  $\delta_{n,0}$  la fonction de Dirac valant 1 si  $n = 0$  et 0 sinon, et  $q_0(t) = 0$ .

L'état quasi-stationnaire est défini par :  $p_n(t) = \left(\sum_{n \geq 1} q_n(t)\right) \bar{p}_n$ , pour  $n \geq 1$ , avec  $\bar{p}_n$  indépendant du temps.

Le formalisme stochastique offre certes un degré supérieur de réalisme, mais ce réalisme est tempéré par la nécessité de simplifier les modèles. Ainsi la notion d'état quasi-stationnaire d'un processus de Markov (Encadré 4) demeure assez rigide. Elle est traitée en général par des approximations déterministes sur les principaux moments (moyenne et variance) : voir Ovaskainen et Hanski (2001) dans le cadre des métapopulations. Dans la suite de mon travail, il me faudra savoir si une distribution donnée des états des objets, à un instant et en un lieu donnés, est bien exprimée par l'état quasi-stationnaire d'un processus de contact stochastique (données floristiques issues de l'Atlas de la Flore de la Drôme, voir §1.1).

La modélisation stochastique permettra de caractériser certaines statistiques agrégées dans le temps ou dans l'espace : pour les métapopulations, il s'agira ainsi de la probabilité moyenne d'occupation de l'habitat à l'état quasi-stationnaire, le temps moyen d'extinction, le temps moyen d'arrivée à l'équilibre (Ovaskainen and Hanski 2002). Dans un modèle

purement déterministe de métapopulation (type champ moyen, Levins 1969), on mettra en regard de ces grandeurs le temps de relaxation, qui mesure le temps de retour à l'équilibre après perturbation (réduction de l'écart à l'équilibre par un facteur  $e^{-1}$ ) (Etienne and Nagelkerke 2002).

Or, quand nous nous intéressons à la réalisation de processus stochastiques, en l'occurrence dans mon travail à la dynamique spatiale et temporelle de populations de plantes, des phénomènes contingents, localisés dans le temps et/ou l'espace, sont susceptibles d'initier de fortes altérations des distributions attendues. Dans l'étude d'un système écologique, nous nous intéresserons à la caractérisation d'un objet par rapport à son environnement. Ce faisant, nous distinguons l'objet d'étude et les objets qui sont autour et qui constituent son environnement. L'existence de l'objet focal, comme celle des objets de l'environnement est émaillée d'événements contingents (Encadré 2). Quand on modélise le comportement du système, le souci de simplifier la représentation des processus amène à représenter de manière dissymétrique ces deux catégories d'objets : les objets de l'environnement ne sont en général pas identifiés précisément, mais représentés par des variables synthétiques extrinsèques, l'objet focal est caractérisé par des variables d'état. Notre dialectique des facteurs intrinsèque/extrinsèque est inspirée de la terminologie utilisée pour l'étude de la dynamique des populations. Cela conduit traditionnellement à représenter deux types d'événements contingents :

- Les contingences extrinsèques au fonctionnement de l'objet biologique. Les contingences géomorphologiques (surrection des montagnes, volcanisme, formation de mers et océans) et climatiques (réchauffements, refroidissements, accompagnés de variations des niveaux des mers) façonnent les biotas, à échelle biogéographique. L'Homme par son action joue actuellement un rôle majeur dans la modification des distributions des espèces. A une échelle de temps écologique, ces phénomènes s'écartent notamment du cadre des processus stochastiques précédemment évoqués par leur occurrence très irrégulière, alors que les distributions de probabilité de colonisation et d'extinction des espèces sont décrites indépendamment du temps et de l'espace. Les événements contingents ne peuvent donc pas être inclus dans la modélisation classique d'un système stationnaire. Ils peuvent cependant avoir des incidences fondamentales sur le devenir des systèmes biologiques (voir §1.2).

- Les contingences intrinsèques au fonctionnement de l'objet biologique. La dynamique des objets étudiés, en ce qui nous concerne la dynamique des populations, présente des variations aléatoires et imprévisibles : c'est le principe même de la stochasticité. Cependant, dans un souci déjà évoqué de simplification, les modèles stochastiques reposent souvent sur des statistiques exactes, pouvant nécessiter un nombre important de degrés de liberté d'estimation. Ils deviennent alors dans leur forme très déterministes (Ovaskainen and Hanski 2001). En particulier les états quasi-stationnaires sont caractérisés par une composante constante macroscopique (voir Encadré 4), considérée en espérance sur un temps infini. Les petits écarts ponctuels à ces états (dans le temps et dans l'espace) sont en principe négligés ou assimilés à du bruit. Pourtant il peut y avoir, sur un laps de temps et une étendue limités, l'apparition d'une certaine cohérence dans la succession de ces fluctuations dans le temps et dans l'espace (lois de puissance dans le cas de la criticalité auto-organisée, Bak 1997). Bien que les événements soient individuellement contingents et imprédictibles, la succession de leurs occurrences peut engendrer des structures déterminées (Luisi 2003) : cette idée est au cœur de la théorie de l'auto-organisation, et fera l'objet d'un traitement spécifique (Chapitre II).

Le fait que l'importance des phénomènes contingents soit négligée dans les théories écologiques classiques (niche, biogéographie insulaire, climax) n'est pas anodin, et revêt peut-être une réelle signification épistémologique. L'écologie est devenue une science en se démarquant dans les années 1960 et 1970 des démarches descriptives et historiques, telles l'histoire naturelle et la biogéographie (par ex. De Candolle 1855). L'identité de l'écologie s'est donc formée autour d'un dogme, celui d'une science basée sur des lois générales et universelles, se rapprochant ainsi clairement des sciences physiques et s'écartant résolument de ses racines historiques. Actuellement, les limites de cette démarche sont reconnues et les événements contingents retrouvent leur place légitime dans les théories écologiques (Taylor and Haila 2001). L'essor de la biologie de la conservation y a contribué (Rozdilsky et al. 2001), bien que certains débats récurrents reprennent parfois des positions dogmatiques (Blackburn and Gaston 2002; Fisher 2002; Marquet 2002; Nee 2002). Un des soucis majeurs de mon travail sera de souligner l'importance des phénomènes contingents, et de montrer que leur prise en compte peut s'inscrire dans une démarche scientifique de compréhension des structures écologiques observées, en particulier avec un espace et une durée d'observation limités.

### 2.5.2. Signification des propriétés émergentes

En dernier lieu, le paradigme d'indépendance des échelles suppose qu'il est possible de négliger l'effet sur le système étudié de phénomènes intervenant à des échelles plus grandes ou plus petites que la plage d'échelle du système. L'effet des échelles inférieures peut en général être découplé par l'usage de paramètres « résumant » cette incidence : dans un modèle de métapopulation par exemple (Levins 1969), la viabilité démographique des populations est incluse dans la définition de la probabilité d'extinction.

Cependant ce type d'approximation peut être mis en défaut par l'apparition d'une incidence renforcée des processus microscopiques sur les échelles supérieures mésoscopiques et macroscopiques. Dans ce cas, les solutions consistant à simplifier leur effet sur les niveaux supérieurs deviennent obsolètes. Dès lors l'existence de phénomènes impliquant des effets de structuration sur un grand nombre d'échelles ne doit pas être systématiquement occultée (Levin et al. 1997; Wootton 2001; Dungan et al. 2002). L'idée d'émergence devient prégnante : l'état des objets du système à une échelle donnée peut être étroitement lié à des mécanismes opérant à des échelles sous-jacentes non prises en compte (Wootton 2001; Reuter et al. 2005).

Cette notion d'émergence est au cœur de la théorie de l'auto-organisation, phénomène encore mal compris mais de plus en plus évoqué (Taylor and Haila 2001; Skar 2003). La carence de fondement théorique et d'interprétation mécanistique entrave la démonstration et la prédiction d'un tel phénomène. Dans un contexte de cloisonnement entre disciplines se focalisant sur des niveaux différents de perception, comprendre l'interaction entre des processus intervenant à différentes échelles demeure en effet difficile.

## 3. Structure de causalité

L'essentiel du travail de compréhension des phénomènes écologiques consiste en définitive à établir une structure de causalité au sein des systèmes étudiés. Nous partons de l'axiome que l'état des objets est au moins en partie le résultat de variables et de mécanismes causaux sous-jacents. On cherche alors à décrire et à tester la relation entre des variables d'effet (états des objets) et des variables causales supputées.

- Logique hypothético-déductive = test d'hypothèses

En biologie, les propriétés de l'objet et des mécanismes biologiques constituent en général le point focal de l'étude. On formule des hypothèses sur ces propriétés et on en déduit des

attendus sur les états des objets à l'intérieur du système, en général sous conditions d'équilibre. Sur la base d'observations ou d'expériences, les attendus peuvent être testés et permettent d'infirmer ou non les hypothèses sous-jacentes : c'est une démarche hypothético-déductive.

Les critères de décision des tests d'hypothèse en écologie reposent très généralement sur l'erreur statistique de type I. Cela signifie qu'on accepte une probabilité donnée de rejeter une hypothèse formulée quand elle est vraie. En faisant cela, on ne valide pas l'hypothèse, mais on considère dans quelle mesure on peut l'infirmer. L'erreur de type II en revanche prend en compte la probabilité d'accepter une hypothèse alors qu'elle est fausse. Il y a une tendance actuelle à utiliser l'erreur de type I pour calculer les intervalles de confiance et valider des hypothèses, sans se préoccuper de l'effet de l'erreur de type II : cela peut conduire à d'importantes erreurs de décision, comme dans le cas des *Contrastes Phylogénétiquement Indépendants* (Quader et al. 2004). Cette évolution traduit une déviation épistémologique implicite en écologie, depuis la logique poppérienne de falsification des hypothèses, vers une logique inférentielle de validation (Lecoutre 2005).

- Logique inférentielle = lecture du réel

De fait on a souvent recours en biologie à une logique inférentielle explicite. Il s'agit d'estimer les paramètres d'un modèle censé représenter le plus fidèlement possible la structure des données réelles : on le dit alors *ad hoc*. L'analyse causale, à grand renfort de régressions linéaires ou généralisées (GLM – *Generalized Linear Models* – linéaires, et GAM – *Generalized Additive Models* – non linéaires), paramètre de manière « optimale » la relation entre des variables dites causales et la ou les variables d'effet d'intérêt. Il n'y a pas alors d'attendu précis et testable, mais un modèle *ad hoc* dont la souplesse (évaluée souvent par ses degrés de liberté) permettra une représentation précise ou non de cette relation. Il y a alors une dimension algorithmique essentielle, propre au problème de l'ajustement. Les paramètres du modèle sont inférés à partir des données disponibles. L'idée d'analyse causale (Legendre and Legendre 1998) est ainsi fondamentalement inférentielle. On bascule du test d'hypothèses, dans une logique clairement hypothético-déductive, au test d'ajustement de modèles, fonction des données réelles et sans référence directe à une structure d'hypothèses.

On peut aller plus loin encore dans cette logique, et orienter la construction même du modèle *ad hoc* d'inférence par l'observation et la réalisation en amont de mesures sur le système réel. Il s'agit alors d'évaluer la nature et la qualité de l'information disponible pour la compréhension des phénomènes écologiques (Fortin et al. 2003), préalable à la démarche

d'inférence. Cette observation est susceptible d'orienter la modélisation du système, comme par exemple en analyse bayésienne (Ellison 1996). Cette étape préalable appartient au registre de l'observation et de la recherche d'information écologiquement pertinente : elle nous conduit vers une logique informationnelle.

- Logique informationnelle – interférence test d'hypothèse / lecture du réel

Actuellement la disjonction entre approches hypothético-déductive et inférentielle tend à s'estomper du fait du recours massif aux critères informationnels (Johnson and Omland 2004). On considère alors des modèles sur-paramétrés, mécanistiques ou non, que l'on essaie d'ajuster aux données mesurées sur le réel. Le paramétrage est réduit progressivement et séquentiellement par une sélection *ad hoc* des paramètres les plus pertinents, en fonction de la quantité de variance que la large gamme des modèles associés est en mesure de décrire. **Il s'agit ainsi de sélectionner des paramètres qui décrivent bien, et non qui expliquent bien, la structure des données mesurées.**

Bien que la logique fondamentale puisse être hypothético-déductive (formalisation et test d'attendus), les attendus peuvent être vidés de leur valeur théorique en leur conférant trop de degrés de libertés. Il y a notamment le risque d'équivocité : plusieurs jeux bien différents de valeurs des paramètres peuvent décrire de manière équivalente les données, mais avec un seul jeu, voire aucun, possédant une réelle valeur explicative. On en est alors réduit à une analyse descendante délicate de sélection de variables (George 2000), dans une approche inférentielle. Une bonne philosophie demeure de privilégier des modèles et des attendus simples.

Quelle que soit la logique adoptée, l'objectif est bien de reconstituer une structure de causalité au sein du système étudié. **La question essentielle est de savoir si l'on décrit bien et/ou si l'on explique bien ce qui se passe dans le système d'intérêt.** En définitive la confusion entre les démarches hypothético-déductive et inférentielle ne doit pas être sous-estimée et revêt une réelle signification épistémologique :

- Les ressources informatiques de plus en plus performantes permettent de traiter des quantités d'information de plus en plus importantes. On peut supposer que l'absence de tels moyens a par le passé favorisé la démarche hypothético-déductive, qui seule traite d'attendus simples pour des systèmes complexes (Levin et al. 1997). La valeur démonstrative des simulations et automates cellulaires de plus en plus massivement utilisés reste cependant sujette à caution (Molofsky and Bever 2004).



- Les progrès dans la modélisation des systèmes stochastiques a fait entrevoir l'existence de phénomènes complexes de nature chaotique, ou à la limite du chaos (criticalité, voir Chapitre II). Des propriétés émergentes existent souvent à des échelles supérieures (Reuter et al. 2005). La culture inférentielle et empiriste de la physique (Goldenfeld and Kadanoff 1999) pèse certainement actuellement sur l'évolution des idées en écologie. Toutefois les gammes d'échelles impliquées et les comportements des objets sont très différents, ce qui implique d'envisager avec prudence toute transposition théorique (Ricotta 2000).

Les questions abordées au cours de mon travail vont m'amener à revenir régulièrement sur cette dialectique de représentation de la causalité : ascendante (hypothético-déductive) vs. descendante (inférentielle). Nous verrons que le choix de la démarche, ainsi que la discussion des quatre paradigmes fondamentaux, seront primordiaux pour bien comprendre ce que révèle la structuration spatiale des populations au sein d'une métapopulation.

#### **4. Axes de travail**

La préoccupation centrale de mon travail est de donner un sens à la structure spatiale et/ou temporelle des populations d'organismes, afin d'inférer leur dynamique, et donc leur capacité d'occupation de l'espace disponible. La compréhension du lien entre processus dynamiques et structures résultantes dans l'espace et dans le temps sera donc la problématique essentielle. Cela pose des problèmes théoriques et techniques importants. Ma démarche sera dès lors d'utiliser plusieurs angles d'approches, partiels mais complémentaires, qui permettront d'évaluer la nature de l'information fournie par la structure observée des populations.

J'ai souligné dans la présente Introduction l'importance des concepts et paradigmes dominants qui, s'ils ne sont pas bien compris, obscurcissent la démarche scientifique. Mon travail n'a pas vocation à créer un socle théorique nouveau pour la modélisation de la dynamique des métapopulations, le foisonnement actuel offrant déjà une base de travail riche. En revanche, à partir d'une réflexion théorique élargie du lien entre structures spatiales émergentes et mécanismes sous-jacents (Chapitre II), je propose une analyse essentiellement informationnelle des différentes méthodes. Cela signifie que la confrontation d'approches de nature différentes, imparfaites mais complémentaires, permettra d'appréhender leur efficacité, mais aussi de comprendre *pourquoi elles ne fonctionnent pas toujours bien*.

Trop souvent les défaillances des modèles sont attribuées aux hypothèses formulées dans la construction des modèles, ou encore à des problèmes purement techniques, et ne donnent pas lieu à une réflexion sur des paradigmes utilisés de façon implicite. A plusieurs reprises au cours de mon travail (Chapitres III et IV), une relecture des atouts et faiblesses des méthodes classiques montrera à cet égard l'importance de facteurs jusqu'ici sous-estimés voire ignorés.

Le plan adopté répond à ces préoccupations, et suit une trame logique :

- Chapitre I. Pour comprendre le lien entre dynamique métapopulationnelle et structures spatiales résultantes, il faut recourir à une expérience en « conditions contrôlées », où les différents paramètres en jeu peuvent être contrôlés et modulés. Dans cet esprit, une dynamique métapopulationnelle dans un habitat simple et contrôlé a été simulée sur une large gamme de variation des paramètres. Ces simulations servent de base de référence tout au long de mon travail.
- Chapitre II : Je m'intéresse à un mécanisme important pour comprendre le lien entre dynamique locale des populations et structures émergentes : l'auto-organisation. Ce mécanisme est présenté et expliqué au regard des objectifs de mon travail, c'est-à-dire en évaluant les conséquences pratiques sur l'étude des états quasi-stationnaires.
- Chapitre III : Le problème du choix d'une mesure pertinente de la structure spatiale des populations est posé. L'analyse spectrale est supposée donner une représentation pertinente des structures résultant de processus d'extinction/colonisation. Nous évaluons la sensibilité des mesures spectrales aux paramètres clés de la dynamique métapopulationnelle. Nous montrons en particulier l'importance cruciale du lissage et du principe d'incertitude. Nous ébauchons enfin une nouvelle approche spectrale pour la partition de la variation spatiale d'une variable, à des fins d'inférence.
- Chapitre IV : Ce Chapitre forme avec le précédent un diptyque offrant deux visions très différentes. A la vision informationnelle et inférentielle du Chapitre III (approche descendante), nous substituons dans ce Chapitre une démarche analytique, basée sur la modélisation du processus métapopulationnel (approche ascendante). L'étude du modèle reste étroitement associée à la discussion de la méthode d'ajustement aux données réelles,

dans une logique inférentielle. Ce Chapitre aboutit alors à une discussion sur le choix de l'approche de modélisation, en fonction de la nature des données d'occupation disponibles dans le temps et dans l'espace.

- Chapitre V : Les résultats du Chapitre IV permettent d'étudier la dynamique métapopulationnelle des plantes du département de la Drôme, à partir de l'Atlas de Garraud (2004). L'évaluation de cette dynamique est relative à la connaissance que l'on a des données environnementales, et dès lors à la définition de l'habitat potentiel de l'espèce. Nous avons envisagé deux types d'habitat potentiel : l'un basé sur des variables purement bioclimatiques, l'autre sur des variables d'occupation des sols. Nous avons pu évaluer la dynamique des espèces au sens des deux habitats, et tester enfin l'effet de l'action de l'Homme sur la dynamique spatiale des espèces.
- *Synthèse et discussion* : La synthèse sera le pendant des idées et questions générales formulées en Introduction. Elle discutera en particulier les opportunités nouvelles que notre approche inférentielle apporte, et mettra l'accent sur les résultats théoriques forts issus des Chapitres III et IV. Enfin les résultats fructueux obtenus dans l'étude de la Flore de la Drôme souligneront l'opportunité de notre démarche, dévolue à la recherche d'une estimation écologiquement pertinente de la structure spatiale des populations. Les perspectives en terme de tests d'hypothèses écologiques seront alors esquissées.

# I. Métapopulations simulées en habitat structuré

Plusieurs approches vont être conduites et discutées dans les parties suivantes. Elles feront appel à différentes hypothèses et représentations **écologiques** des populations ; nous n'aborderons pas l'analyse évolutive des stratégies de persistance et de dispersion au sein des métapopulations.

Afin de mieux comprendre le fonctionnement et les attendus de ces approches, des simulations ont été conduites. Il s'agit de métapopulations évoluant dans un espace discrétisé, sous la forme d'une grille, où chaque cellule de la grille représente une étendue spatiale abritant potentiellement une population. Au sens de l'habitat *a priori*, les cellules peuvent être favorables ou défavorables au maintien des populations. Le temps aussi est discrétisé en générations successives de populations : on définit des probabilités de colonisation et d'extinction d'une génération à la suivante (processus de Markov en temps discret).

Ces simulations constituent une expérience théorique, permettant de mettre en évidence, en conditions contrôlées, le lien entre des processus simples et les structures spatiales obtenues. C'est donc un outil pour comprendre ces mécanismes (Peck 2004).

Nous précisons dans un premier temps le concept de métapopulation, en nous appuyant sur les modèles fondamentaux associés. Dans un second temps nous détaillons la mise en place de métapopulations simulées. Les hypothèses et les paramètres choisis seront détaillés et justifiés.

## 1. Concept de métapopulation : modèle de Levins

Le modèle classique de Levins (1969) est fondateur dans l'étude des métapopulations, et situe clairement les bases conceptuelles que nous discuterons tout au long de mon travail. Nous présentons donc la logique de ce modèle en temps continu, socle de nombreux développements spatialement plus réalistes apparus récemment.

### 1.1. Formulation du modèle

Le modèle formule, dans l'espace et dans le temps, une dynamique stochastique d'occupation par une espèce focale, soumise à des événements de colonisation et d'extinction. C'est un modèle de type *champ-moyen*, où les événements interviennent uniformément dans

l'espace : la probabilité d'extinction locale, ainsi que la probabilité de colonisation depuis une population source, ont des distributions uniformes dans l'espace et dans le temps.

On considère ainsi un ensemble de cellules occupées et inoccupées (sites de même aire). On note  $\rho$  la densité de cellules occupées. Soit  $c$  et  $e$  respectivement, par unité de temps, le taux de colonisation d'une cellule vide depuis une cellule occupée voisine et le taux d'extinction d'une cellule. Un bilan de flux d'entrée-sortie sur une cellule pendant un temps  $dt$  s'écrit alors :

$$d\rho = c \cdot dt \times \rho(t) \cdot (1 - \rho(t)) - e \cdot dt \times \rho(t)$$

$$\frac{d\rho}{dt} = c \times \rho \times (1 - \rho) - e \times \rho$$

L'équilibre  $\rho^*$  et le temps de relaxation  $\tau$  du modèle sont tels que :

$$\rho^* = \begin{cases} 0 & \frac{c}{e} \leq 1 \\ 1 - \frac{e}{c} & \frac{c}{e} > 1 \end{cases} \quad \tau = \begin{cases} \frac{1}{e - c} & \frac{c}{e} \leq 1 \\ \frac{1}{c - e} & \frac{c}{e} > 1 \end{cases}$$

Le temps de relaxation est une mesure du temps de retour à l'équilibre après perturbation : c'est le temps nécessaire pour que l'écart après perturbation par rapport à l'équilibre soit réduit d'un facteur  $e^{-1} = 0.368$  (NB : ici  $e$  est le nombre d'Euler). Plus le temps de relaxation est grand, moins l'équilibre est attractif. Quand on travaille sur une représentation déterministe d'un processus stochastique, le temps de relaxation peut être considéré comme une estimation indirecte du temps d'extinction (Etienne and Nagelkerke 2002).

L'effet de la perte d'habitat peut être introduit en considérant Lande (1987):

$$\frac{d\rho}{dt} = c \times \rho \times (h - \rho) - e \times \rho, \text{ où } h \text{ est la fraction de cellules favorables.}$$

$$\text{On a à l'équilibre } \rho^* = \begin{cases} 0 & \frac{c}{e} \leq 1 \\ h - \frac{e}{c} & \frac{c}{e} > 1 \end{cases}. \text{ La proportion de sites vides s'écrit } h - \rho^* = \frac{e}{c} \text{ pour}$$

$h > \frac{e}{c}$  et est indépendante de  $h$ .  $h = \frac{e}{c}$  est appelé le seuil d'extinction. Il en résulte la règle de Levins : “A necessary and sufficient condition for metapopulation survival is that the remaining number of habitat patches following a reduction in patch number exceeds the number of empty but suitable patches prior to patch destruction.” Cette assertion a été souvent discutée au regard des hypothèses du modèle.

Le formalisme de Levins est en effet sujet aux hypothèses suivantes :

- *Absence de structure spatiale : pas de localisation des cellules ni de corrélation locale.*
- Les cellules et populations locales sont identiques en terme de dynamique locale.
- Les dynamiques locales sont ignorées ou les dynamiques locales se produisent bien plus rapidement que les dynamiques métapopulationnelles, ce qui implique :
  - La migration a un effet négligeable sur les dynamiques locales.
- *Le choix des cellules par les colonisateurs est aléatoire (pas de préférence d'habitat).*

Ces hypothèses impliquent que l'on peut décrire simplement les cellules de manière binaire : 1 (occupés) et 0 (inoccupés). Les hypothèses en italiques seront celles qui devront être étudiées et relâchées dans la suite de mon travail: il s'agira d'étudier l'incidence de la présence d'habitat défavorable (effet de sa géométrie) et de la structure spatiale des populations (corrélations).

Etienne (2000) a étudié l'effet de l'extension du modèle de Levins en relâchant trois des hypothèses initiales : introduction de populations de tailles différentes, introduction d'un effet *rescue* (modification du taux d'extinction local grâce à l'immigration), et introduction d'une préférence de cellule. L'auteur a montré que l'on peut aménager l'équation fondamentale de Levins pour prendre en compte ces effets, sans modifier fondamentalement le comportement du système.

*Nota Bene* : Par souci d'alléger la représentation des équations, l'opération de multiplication sera en général représentée par le point (.) dans la suite du travail.

### *1.2. Usage d'une représentation discrète ou continue*

Le modèle de Levins n'offre pas de référence à une échelle spatiale et temporelle explicite des processus. L'espace et le temps sont continus. Il s'agit en fait d'une approximation, utile pour faire des prédictions générales sur les systèmes réels, en l'occurrence par le biais de la densité d'habitat occupé. Cette densité, exprimée dans un espace et temps continus, peut être évaluée aussi bien dans un espace et un temps discrets, et c'est ce qui est fait en général.

Le lecteur pourra trouver paradoxal de ne pas avoir alors de référence à l'état discret des populations dans l'espace ; la notion de métapopulation est en effet fondée sur cette représentation discrète. Il faut bien avoir à l'esprit, à ce stade, que les modèles que nous

présenterons tout au long du mémoire ont vocation à *impliciter* la représentation des processus réels, afin d'accéder à la possibilité d'un traitement analytique et à obtenir des prédictions simples. Mais celles-ci seront toujours partielles et synthétiques, s'accommodant ainsi de certaines approximations.

Cette idée sera rappelée tout au long des chapitres qui suivent. Quoi qu'il en soit, nous présenterons ci-après une logique différente, une logique de *simulation*. Dans cette logique en effet, il ne s'agit pas de formuler des prédictions synthétiques sur le comportement du système, mais d'essayer de reproduire fidèlement ce comportement. Nous utilisons certes une représentation simplifiée de certains facteurs, comme la structure de l'habitat, ou la variation spatiale et temporelle des événements de colonisation-extinction. Mais l'objectif reste d'*explicit*er la dynamique métapopulationnelle, de lui donner une référence spatiale et temporelle claire et réaliste.

Dans la suite du travail, nous aurons à confronter la démarche d'explicitation de type simulation à celle d'implication de type modélisation. Cette dualité sera au cœur de la compréhension du lien entre les processus dynamiques, implicites, et les structures émergentes, explicites. Nous verrons aussi qu'une représentation discrète du processus métapopulationnel revêt une grande importance théorique (Chapitre II). L'inférence sera en définitive le point ultime de rencontre des deux logiques de représentation de la dynamique.

## 2. Procédure de simulation

Nous abordons dans cette partie la méthodologie adoptée pour simuler la dynamique de métapopulations dans un habitat structuré. Ces métapopulations *théoriques* de référence présentées et utilisées à maintes reprises par la suite ont été simulées à l'aide du logiciel *Mathematica 5.0* (Wolfram Research 2003).

A titre d'amélioration de la procédure de simulation, nous présentons en Annexe 1 la mise en place d'une plate-forme autonome de simulation, ayant vocation à simplifier et à améliorer considérablement la rapidité des calculs.

### 2.1. Paysage binaire d'habitat

Nous choisissons un paramétrage simplifié pour représenter la structure spatiale d'un habitat binaire :  **$p$**  est la **densité** de cellules favorables au sens de l'habitat *a priori*, tandis que

**$q$  est l'agrégation de l'habitat**, c'est à dire la probabilité conditionnelle qu'une cellule voisine d'une cellule favorable soit favorable. Ces deux paramètres sont des indices spatiaux caractéristiques et semblent écologiquement pertinents pour l'étude des habitats réels (O'Neill et al. 1988; Cain et al. 1997).

Nous utilisons l'algorithme et le programme de Hiebeler (2000) pour constituer un ensemble de cartes binaires d'habitat, avec différents niveaux de  $p$  et  $q$  (voir Figure 3b). Les valeurs de ces paramètres sont définies de 0.1 à 0.9, et de 0.1 en 0.1 : il y a donc 9 modalités pour chaque paramètre.

Il existe en outre des contraintes spatiales qui limitent le nombre de paires  $(p, q)$  possibles. Une cellule favorable est représentée par 1, une cellule défavorable par 0. Soit  $P(00)$  la probabilité de tirer une paire de cellules 0-0,  $P(10)$  de tirer une paire 0-1 ou 1-0 et  $P(11)$  de tirer une paire 1-1, ces probabilités sont reliées par la relation suivante :  $P(00) = 1 - P(11) - 2.P(01) = 1 - pq - 2p(1 - q) \geq 0$ . La grandeur  $P(00)$  étant positive, cela implique donc d'avoir  $q \geq 2 - \frac{1}{p}$ . A partir de  $p = 0.6$ , certaines valeurs de  $q$  ne sont donc pas possibles. Nous avons en définitive un jeu de 58 cartes différentes.

## 2.2. Simulation de la dynamique métapopulationnelle

Pour une carte donnée, nous simulons la dynamique des populations au sein de l'habitat favorable. L'état d'une cellule d'habitat favorable peut être occupé ou inoccupé. Quand une cellule est colonisée, on considère que l'équilibre démographique est atteint en une génération. Cette hypothèse classique du modèle de métapopulation est souvent discutée, mais nous ne considérerons pas par la suite d'effet démographique.

En définitive, l'état des cellules varie dans le temps et dans l'espace au gré des événements de colonisation et d'extinction :

- La probabilité d'extinction  $e$  des cellules par génération est uniforme dans l'espace et dans le temps (i.e. en espérance indépendante de la position et de la date). Au cours de la simulation, le nombre d'événements d'extinction à chaque génération est déterminé par un tirage binomial de probabilité  $e = cste(x, y, t)$ . Ce modèle d'extinction prend en compte implicitement des événements stochastiques d'occurrence uniforme liés à la démographie, ou à des variations catastrophiques de l'habitat. Il ne prend pas en compte la possibilité d'effets successionnels, où le taux d'extinction peut dépendre de l'âge des populations.



Ces effets peuvent être importants dans une perspective évolutive (stratégies de dispersion, Olivieri et al. 1995). Cependant, dans un temps écologique, nous considérons que la prise en compte implicite des effets successionnels dépendra étroitement de la définition des variables d'habitat (Chapitre V).

- Les événements de colonisation dépendent de la distance aux populations sources. On considère une probabilité *conditionnelle* de colonisation décroissant exponentiellement

avec la distance :  $c(i) = \frac{e}{r} \cdot \frac{\text{Exp}(-i)}{\sum_{k=1, \dots, 5} 8.k.\text{Exp}(-k)}$ . C'est la probabilité de coloniser une cellule

située dans le voisinage  $i$  de la cellule d'origine, en supposant la cellule d'arrivée présenter un habitat favorable. Ce noyau de colonisation est défini sur les 5 premiers voisinages  $i$  (voir Figure 3a). Le nombre de cellules colonisées à une génération donnée, depuis une cellule donnée, et dans un voisinage  $i$ , est déterminé par un tirage binomial de paramètres  $\text{Bin}(8.i, c(i))$ .

Durant la phase quasi-stationnaire, les événements de colonisation et d'extinction s'équilibrent. Il en résulte qu'une augmentation des taux de colonisation et d'extinction dans les mêmes proportions aboutit à un état quasi-stationnaire inchangé. Pour illustrer cette idée, considérons une représentation simple de la dynamique du système en habitat uniforme :

$$\frac{d\rho(t)}{dt} = C_t \cdot \rho(t) \cdot (1 - \rho(t)) - e \cdot \rho(t), \text{ avec } \rho(t) \text{ la densité de cellules occupées à l'instant } t.$$

La densité stationnaire  $\rho_{st}$  sera déterminée par la relation suivante à l'équilibre:

$$C_t \cdot \rho_{st} \cdot (1 - \rho_{st}) = e \cdot \rho_{st}$$

Cela conduit à écrire  $\rho_{st} = 1 - \frac{e}{C_t}$ . On retrouve l'idée intuitive que la probabilité

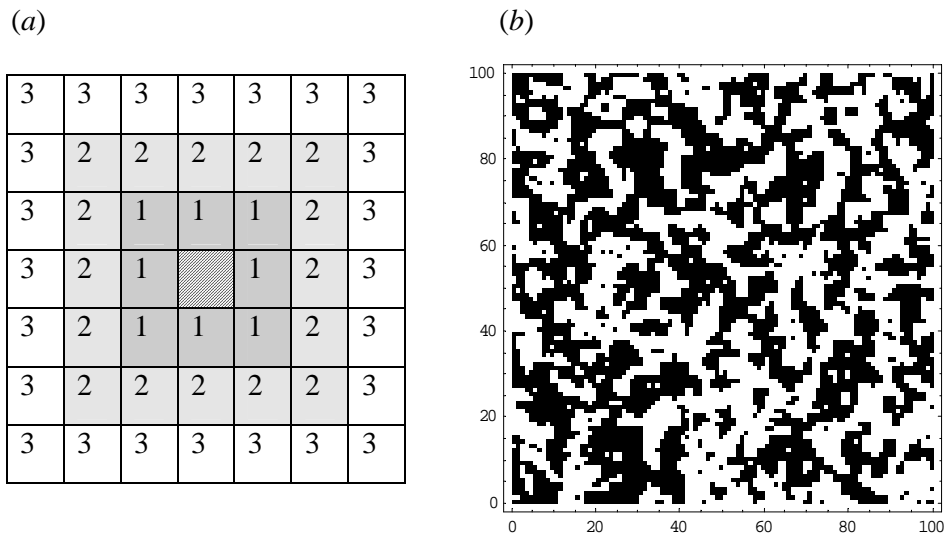
d'extinction doit demeurer inférieure à la probabilité conditionnelle de colonisation, soit

$\frac{e}{C_t} < 1$ . Ce modèle très simple est une simplification où le terme de colonisation

$C_t \cdot \rho(t) \cdot (1 - \rho(t))$  dépend : (i) de l'état d'occupation des cellules sources  $\rho(t)$ , (ii) de l'état d'occupation de la cellule d'arrivée  $(1 - \rho(t))$ . La simplification consiste à établir une probabilité d'occupation identique et constante dans l'espace pour toutes ces cellules : on parle alors d'approximation *champ-moyen*.

Dans un modèle de métapopulation spatialement réaliste (probabilité d'occupation non spatialement uniforme), l'équilibre entre événements de colonisation et d'extinction permet de décrire une distribution quasi-stationnaire non nulle des probabilités locales d'occupation, vers laquelle converge le système (Ovaskainen and Hanski 2001). Cette convergence peut éventuellement être conditionnée à un nombre initial seuil de populations (invasibilité de la métapopulation). On estime les probabilités quasi-stationnaires d'occupation par une moyenne temporelle des états des cellules (occupé/inoccupé), durant l'état quasi-stationnaire (voir Figure 3c). Par ailleurs, la description de l'état quasi-stationnaire est essentiellement déterminée par le seul paramètre  $r=c(i)=\frac{e}{\sum_{k=1,\dots,5} 8.k.c(k)}=\frac{e}{C_i}$ , où  $C_i = \sum_{k=1,\dots,5} 8.k.c(k)$  représente le taux de colonisation global depuis une population source (Ovaskainen and Hanski 2001).

Il est important de souligner que les paramètres  $p$ ,  $q$  et  $r$  ne sont pas indépendants. Nous avons vu (§2) qu'une contrainte topologique relie les valeurs accessibles de  $p$  et  $q$ . Il existe un autre facteur de dépendance, directement lié à la dynamique métapopulationnelle. Considérant les métapopulations persistant à l'état quasi-stationnaire (*i.e.* n'aboutissant pas à l'extinction de toutes les populations), la gamme de valeurs de  $r$  possibles dépendra de la configuration de l'habitat, donc directement de  $p$  et  $q$ . Cette double dépendance est représentée graphiquement à la Figure 4.



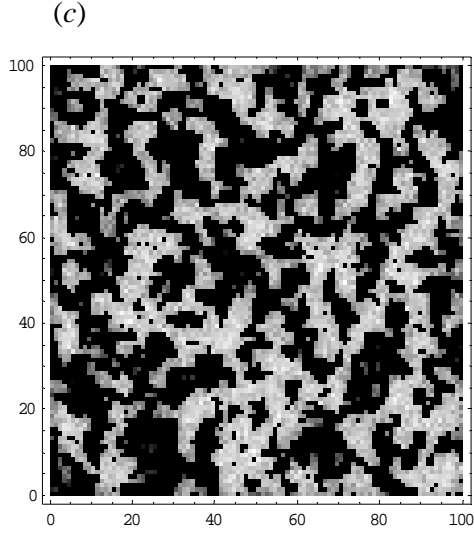


Figure 3 : Métapopulations simulées en habitat binaire. (a) La fonction de colonisation utilisée décroît avec la distance à la population source (cellule hachurée). Pour simplifier, la distance considérée est le rang du voisinage (ici numéroté de 1 à 3). (b) Paysage binaire d'habitat obtenu avec l'algorithme de Hiebeler (2000), pour les valeurs de densité  $p = 0.5$  et d'agrégation  $q = 0.8$ . Les cellules noires représentent l'habitat *a priori* défavorable. (c) Probabilités locales d'occupation de l'habitat favorable durant l'état quasi-stationnaire pour  $r = 4$ . Les probabilités sont représentées en niveaux de gris, plus clair = plus élevée, plus foncé = plus faible. Ces probabilités sont estimées au moyen d'une moyenne sur 300 générations des états des cellules (occupé/inoccupé), durant l'état quasi-stationnaire. Les cellules noires représentent l'habitat *a priori* défavorable.

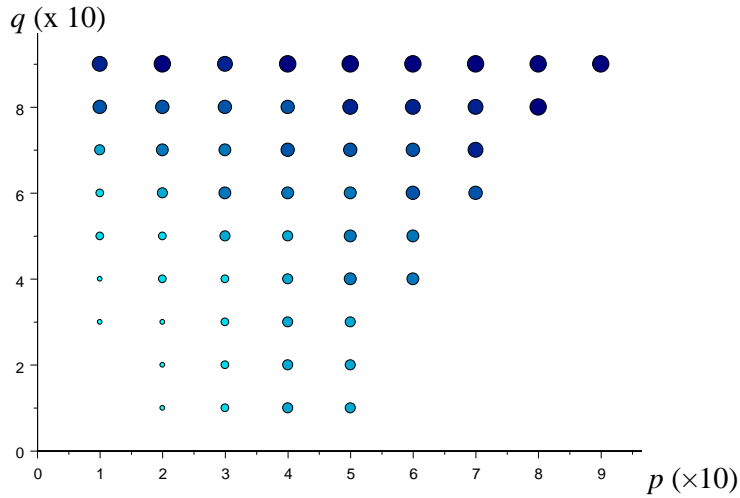


Figure 4 : Gamme des paramètres  $(p, q, r)$  associés à la persistance des métapopulations théoriques à l'état quasi-stationnaire. Il existe une limite inférieure à  $q$  considérant la variation de  $p$ , liée à des contraintes topologiques. Pour les autres combinaisons  $(p, q)$ , la gamme de valeurs de  $r$  assurant la persistance de la métapopulation augmente à  $p$  fixé quand  $q$  augmente, et augmente à  $q$  fixé quand  $p$  augmente. La valeur maximale de  $r$  pour un couple  $(p, q)$  fixé est représentée par la taille et la couleur des points : plus gros et plus foncé = valeur plus élevée.

## II. Criticalité et auto-organisation des métapopulations

Dans ce Chapitre, mon propos est d'attirer l'attention du lecteur sur des phénomènes souvent méconnus mais importants pour l'étude de la structure spatiale et temporelle des métapopulations :

- d'abord l'existence, même en habitat uniforme, d'une structuration spatiale et temporelle particulière résultant de phénomènes d'avalanches,
- ensuite l'existence d'un seuil critique séparant un régime de persistance globale de la métapopulation et un régime d'extinction globale de la métapopulation,
- enfin l'existence de structures auto-organisées au voisinage des seuils critiques, que nous illustrerons sur une typologie simple.

Mon souci sera dans un premier temps de donner un aperçu théorique assez large, résumant les idées et concepts essentiels associés à ces phénomènes. Dans un second temps, je montrerai l'existence et l'importance de l'auto-organisation dans la structure des métapopulations simulées dans un habitat structuré binaire (voir partie précédente).

Ces éléments définiront une trame théorique sur laquelle seront bâtis les deux Chapitres suivants, dévolus à l'étude du lien entre le processus dynamique métapopulationnel et les structures de populations résultantes dans l'espace et le temps.

### 1. Singularité des processus de contact

La dynamique spatiale de colonisation-extinction au sein des métapopulations correspond à une famille de processus appelés processus de contact :

- Ils sont associés à une représentation discrète des états des objets, par exemple binaire actif/inactif. Dans notre cas, les objets sont des populations et nous traitons des états binaires occupés/inoccupés, ou éventuellement ternaires habitat favorable occupé/ habitat favorable inoccupé/ habitat défavorable.
- Comme leur nom l'indique, les processus de contact mettent en jeu des événements définis au voisinage des zones activées, dans notre cas les processus de colonisations sont définis au voisinage des populations occupées.

### *1.1. Limite de la représentation quasi-stationnaire*

Nous ne développerons pas ici la formalisation de ce type de processus sous forme d'équations. Nous renvoyons l'attente du lecteur en ce sens à l'examen du Chapitre IV (§3). Nous nous concentrerons ici sur certaines caractéristiques particulières des processus de contact, plus aisément comprises par une représentation intuitive que mathématique.

La première chose à souligner, simple mais importante, est le fait que les processus de contact sont *non linéaires*. Cela est lié à la localisation spatiale des processus : en l'occurrence la colonisation d'une cellule d'habitat dépend du produit de sa propre probabilité d'occupation par la probabilité d'occupation des cellules voisines. Et nous avons ainsi une dépendance dynamique non linéaire entre les états des cellules. Cette non linéarité est fondamentale ; nous verrons qu'elle est au cœur des propriétés particulières que nous développons dans ce Chapitre.

Les processus de contact sont par ailleurs des processus stochastiques dont les règles d'évolution ne mettant en jeu que deux générations successives. La distribution des états quasi-stationnaires peut alors être décrite dans un contexte markovien (Lasserre and Pearce 2001; Dickman and Martins de Oliveira 2004), et les processus de contact auront des propriétés de convergence vers une distribution quasi-stationnaire unique. Toutefois, du fait de la non linéarité du processus, **une représentation matricielle, sous forme de chaîne de Markov, est impossible**. Cela interdit le recours aux propriétés bien connues des modèles matriciels de dynamique, utiles pour l'étude des conditions de convergence et de stabilité des états quasi-stationnaires.

Si l'espérance quasi-stationnaire des états du système est prédite par un modèle de Markov, d'autres aspects de la dynamique quasi-stationnaire sont plus délicats à définir. Le problème est la prise en compte de certaines situations particulières liées à la nature discrète des états :

- De manière stochastique, il peut y avoir un agrégat local de populations, qui engendrent un pool de colonisation plus fort, et une « avalanche » de colonisation, comme une sorte d'effet dominos.
- Inversement, il peut y avoir des avalanches d'extinction à la frontière des zones inoccupées.

Les avalanches possèdent des amplitudes variables. Plus l'avalanche sera importante, plus il faudra de temps pour que le système revienne à une situation proche de son équilibre, car il y a peu de chance qu'une avalanche inverse soit immédiatement consécutive : il y a une certaine inertie du système, qui présente ainsi une corrélation spatiale et temporelle des écarts à l'équilibre. Il est en définitive important de déterminer, par rapport aux inférences que nous souhaitons faire et à la nature des données disponibles, si la prise en compte de ces phénomènes est indispensable ou non.

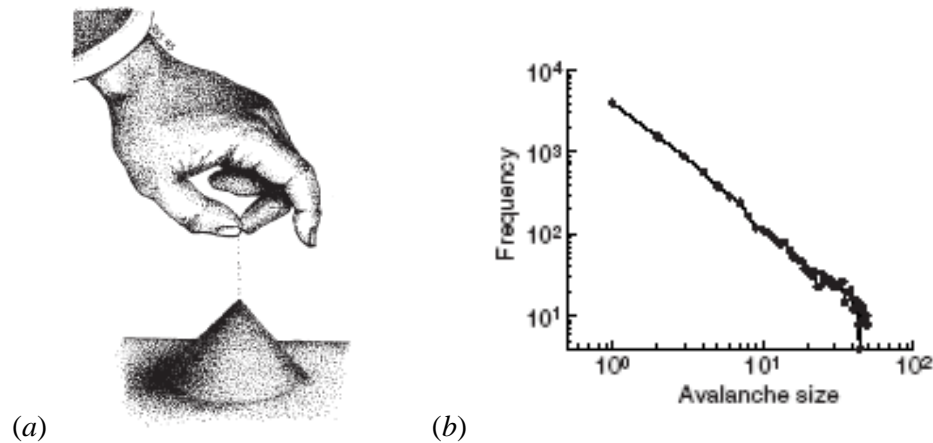
Pour un modèle stochastique, ce sont là des phénomènes dits contingents (voir Introduction, Encadré 2). On considère qu'ils engendrent une variation aléatoire et non structurée des états (bruit blanc) : c'est dans cette logique que se détermine le temps d'extinction. Dans une représentation déterministe de l'équilibre, on considère généralement que les perturbations n'ont pas d'incidence sur le produit du processus : suivant la théorie des petites perturbations, l'équilibre apparaît immuable.

Dans les deux cas, la distribution quasi-stationnaire est obtenue de manière déterministe à partir des probabilités de transition au sein du système (Ovaskainen and Hanski 2001). A ce stade, les phénomènes contingents ponctuels, engendrant des avalanches, ne sont pas pris en compte. Leur prise en compte poserait d'ailleurs des problèmes techniques et théoriques :

- Du fait de leur nature contingente, les avalanches ne seront pas modélisables par des distributions de probabilités simples. Il n'y a pas de relation univoque entre de petites perturbations et l'amplitude des avalanches induites.
- L'amplitude et la fréquence des avalanches sont caractérisées par des lois d'échelle (Solé et al. 1999). Malgré leur occurrence *a priori* ponctuelle et locale, les avalanches peuvent avoir des répercussions à grande échelle de temps et d'espace (inertie du système), et donc sur la caractérisation de l'état stationnaire du système (voir Encadré 5).

#### Encadré 5: Avalanches et lois d'échelle

Voici une illustration de la célèbre expérience du tas de sable (Figure *a*). Sur un plan horizontal, un apport régulier de grains de sable est fait. Un tas de forme conique se forme peu à peu, jusqu'à atteindre une pente critique caractéristique. A partir de là, des avalanches de grains de sable se produisent, d'amplitude et d'occurrence variables. Ces avalanches peuvent engendrer des écarts importants à la pente critique caractéristique, mais à long terme le système tend à osciller autour de cette valeur. La fréquence et l'ampleur (en nombre de grains impliqués) des avalanches sont reliées par une loi d'échelle appelée loi puissance, linéaire en coordonnées log-log (Figure *b*).



(Illustrations issues de Solé et al. 1999)

Les avalanches sont des phénomènes susceptibles de se produire dans de nombreux systèmes écologiques et évolutifs, dès lors que de nombreuses entités ponctuelles en interaction sont impliquées (Solé et al. 1999). On a suggéré en particulier que les phases d'extinctions dans les registres fossiles, d'amplitudes variables, relevaient de phénomènes d'avalanches. Cette théorie s'appuie sur un modèle intuitif simple et fameux, le modèle de Bak et Sneppen (1993).

L'étude des processus de contacts est en définitive à l'interface de la modélisation stochastique, héritée de la théorie des probabilités, et de la physique statistique, qui pose un regard davantage informationnel sur les processus aléatoires (notion centrale d'entropie) (Franc 2004). La compréhension des structures spatiales et temporelles induites par de tels processus nécessite alors d'adopter une double culture. Il y a ainsi deux idées fortes à retenir :

- En temps infini, l'effet des phénomènes contingents est négligeable et les probabilités d'occupation sont représentées par le produit quasi-stationnaire d'un processus markovien (voir Encadré 4).
- En temps fini, l'existence d'une hétérogénéité spatiale et temporelle, induite par la stochasticité et structurée par des avalanches est susceptible d'avoir une incidence sur l'estimation en espérance des états quasi-stationnaires. Ce problème se superpose au problème purement statistique de l'estimation, et en est complètement indépendant.

Les processus de contact ont été essentiellement étudiés dans le contexte d'un environnement (i.e. habitat) uniforme. Leurs propriétés structurales dans le temps et dans l'espace, liées aux avalanches, sont cependant confirmées dans le cas d'un environnement hétérogène (Szabo et al. 2002). Prédire la structuration spatiale des objets induite par un processus de contact constitue en définitive une difficulté et un enjeu majeur pour l'inférence des interactions élémentaires sous-jacentes (Snyder and Nisbet 2000).

### *1.2. Notion d'auto-organisation*

L'auto-organisation est définie par Coveney (2003) de la manière suivante : “(it) is the spontaneous emergence of non-equilibrium structural organization on a macroscopic level, due to the collective interactions between a large number of (usually simple) microscopic objects. Such structural organization may be of a spatial, temporal or spatio-temporal nature, and is thus an emergent property”.

Les processus de contact (Weitz 1997), et plus généralement les processus dissipatifs non linéaires (Coveney 2003), peuvent induire des structures auto-organisées dans le temps et l'espace. La théorie de l'auto-organisation prédit une relation entre l'amplitude et la fréquence des avalanches (Camazine et al. 2001). Ainsi une avalanche de colonisations ou d'extinctions est un phénomène auto-organisé, résultant des interactions locales entre un grand nombre de populations, qui induit l'apparition de propriétés émergentes, en terme de lois d'échelles spatiales et temporelles d'occurrences d'avalanches.

L'auto-organisation est un phénomène invoqué dans de nombreux domaines scientifiques, en particulier en biologie (Camazine et al. 2001). La formation de structures morphologiques cohérentes et régulières, comme les taches du léopard ou les motifs de nombreux coquillages résultent de processus auto-organisés lors de la croissance des organismes. Certains patrons



de végétation spectaculaires, en particulier dans les savanes et zones arides, exhibent également des structures spatiales remarquables auto-organisées (Rietkerk et al. 2004). L'auto-organisation des mosaïques de végétation peut exister également dans des contextes moins extrêmes et davantage façonnés par l'Homme (Bolliger et al. 2003) : il s'agit donc probablement d'un déterminant important et répandu de la structuration spatiale des écosystèmes.

L'auto-organisation demeure une notion assez lâche recouvrant des phénomènes très différents. Ainsi plusieurs types d'auto-organisation ont été décrits (Camazine et al. 2001), et la nomenclature souffre actuellement d'une absence de codification claire (Pascual and Guichard 2005).

Certains types d'auto-organisation ont été largement discutés dans le cadre des populations couplées démographiquement. La synchronie démographique des populations est dans cette perspective un phénomène important (Kaitala and Ranta 1998; Kaitala et al. 2001; Ranta et al. 2002). Elle peut être imputée essentiellement à la corrélation spatiale et temporelle des variables environnementales (Koenig 2002), à l'effet de relations antagonistes trophiques proies-prédateurs (Pascual et al. 2002), ou encore à des connexions dynamiques intrinsèques entre populations. Les effets Allee (seuil de densité initiale nécessaire au développement de la métapopulation) et *rescue* (modification du taux d'extinction local grâce à l'immigration) (voir Hui and Li 2003) constituent un lien entre les processus démographiques et la dynamique des métapopulations. L'auto-organisation induite par la connexion dynamique des démographies locales est souvent associée au thème de la *complexité dynamique*. Les modèles associés, en général déterministes, fournissent des diagrammes de bifurcation qui illustrent les changements de régime du système (notion de chaos déterministe, d'attracteurs étranges...).

Nous ne nous intéresserons pas à de tels modèles fondés sur des couplages démographiques, car ce n'est pas la logique de l'approche métapopulation *stricto sensu*. En tant que processus de contact, une dynamique métapopulationnelle stricte (découplage des processus démographiques) peut aussi posséder des propriétés d'auto-organisation. Nous chercherons par la suite à affiner la compréhension de ces propriétés, afin de mieux évaluer leurs conséquences.

### 1.3. Seuil critique d'extinction

Il existe de nombreux exemples de seuils séparant des régimes de comportement complètement différents et indépendants au sein d'un système. Dans le cas des métapopulations (processus de contact), il existe une transition de phase de type bifurcation (voir Encadré 6) ; d'un côté du seuil, il y a un état absorbant stable où toutes les populations sont éteintes, et de l'autre côté du seuil un état stable où un nombre non nul de populations persiste. Cette transition apparaît en faisant varier le taux de colonisation, à taux d'extinction constant (Franc 2004). On parle alors de seuil d'extinction de la métapopulation (Ovaskainen et al. 2002). Ce seuil d'extinction est déterminé par la quantité et la configuration des cellules d'habitat favorable (Ovaskainen and Hanski 2001). Ainsi, pour une espèce possédant une dynamique donnée de colonisation et d'extinction, une fragmentation accrue de l'habitat fera varier la valeur du seuil, jusqu'à ce que les caractéristiques de l'espèce franchissent le seuil, et que l'on aboutisse à l'extinction déterministe de l'ensemble de la métapopulation. Ce seuil exprime donc la capacité d'un paysage à accueillir une métapopulation, et est donc également qualifiée de capacité métapopulationnelle de l'habitat (Ovaskainen and Hanski 2001). Cette notion est importante notamment pour des problématiques de biologie de la conservation (Hanski and Ovaskainen 2000).

Il ne faut pas confondre le *seuil d'extinction global* de la métapopulation avec la notion de *temps d'extinction* de la métapopulation. Celle-ci décrit l'effet des variations stochastiques durant l'état quasi-stationnaire (voir Encadré 4, p.26) : l'absence de populations étant un état absorbant, il est inéluctable que des fluctuations aléatoires conduisent à terme à l'extinction globale de la métapopulation (métaphore du *gambler's ruin* en théorie des jeux).

Si l'on considère une représentation simplifiée de la dynamique métapopulationnelle, de type champ moyen (voir Chapitre I), le comportement du système correspond à un processus de bifurcation bien connu (voir Encadré 6). Quand on adopte une représentation spatialement plus réaliste du système, le seuil critique existe toujours, mais le comportement du système devient plus complexe à décrire.

De très nombreux exemples de transitions de phase existent dans la nature et dans les sciences humaines (Brunk 2000). Ils ont suscité un large engouement, peut-être excessif (Bak 1997; Frigg 2003). Parmi ces exemples, il y a des phénomènes aussi familier que le passage entre états solide et liquide, entre état magnétique et non magnétique pour les aimants, supra-conducteur et non supra-conducteur, cristallisé et amorphe (voir Stanley 1971).

Il peut alors se produire des basculements plus ou moins rapides et catastrophiques entre des régimes séparés par un seuil critique (pour les métapopulations, voir Labra et al. 2003). En écologie, cela concerne par exemple l'alternance entre perturbations et phases de rétablissement (incendies, successions en général), la perturbation agissant comme une dissipation de la « négentropie » (entropie négative, exprimant un ordre structurel hors équilibre, voir Schrodinger 1944) accumulée pendant la phase de rétablissement. A grandes échelles de temps et d'espace, de tels basculements peuvent correspondre à des changements phytoclimatiques spectaculaires. Au Sahara par exemple, la désertification a constitué un changement majeur et soudain de la végétation (voir Scheffer and Carpenter 2003, Fig.1). Ce basculement serait apparu au cours d'un changement climatique graduel, dépendant directement de l'insolation : en l'occurrence la perturbation n'est pas ponctuelle mais représente un seuil dans la variation continue d'une variable environnementale.

Les tremblements de terre, certaines catastrophes climatiques (cyclones, typhons) et évolutives (crises d'extinction), les feux de forêt peuvent être interprétés comme des exemples de ruptures soudaines et majeures entre deux régimes au voisinage d'un seuil critique (Solé et al. 1999; Pascual and Guichard 2005). Dans tous les cas, on passe d'un système très organisé, à grande échelle (par exemple en physique un solide, un aimant, un supra-conducteur, un cristal), à un état où le comportement aléatoire des entités élémentaires perd toute cette cohérence (liquide, métal non aimanté, solide amorphe). Pourtant les règles de fonctionnement sont inchangées : le système est le même. Seules des conditions environnementales ont changé ; très souvent en physique il s'agit de la température (voir en particulier la théorie de Landau sur les transitions de phase).

#### Encadré 6 : Point de bifurcation et modèle de Levins

L'équation champ-moyen de dynamique de métapopulation (Levins 1969) appartient, par un changement de variable adéquat, à la famille d'équations différentielles  $\dot{y} = \mu \cdot y - y^2 = f(y, \mu)$ , avec  $\mu = C_i - e$ . Les équations dynamiques de cette famille se caractérisent par un comportement de *bifurcation* dite *transcritique*. Trois propriétés caractérisent ce comportement (résumées par la Figure ci-après):

- Deux courbes de points fixes (*i.e.* telles que  $f(y, \mu) = 0$ ) passent par (0,0), une donnée par  $y = 0$ , l'autre par  $\mu = y$ .
- Les deux courbes existent des deux côtés de  $\mu = 0$ .
- La stabilité suivant les deux courbes s'inverse au passage de  $\mu = 0$ .

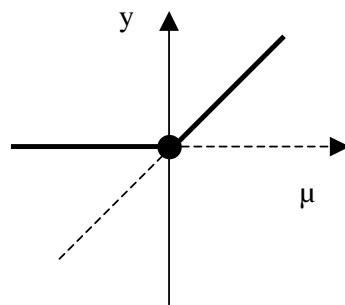


Figure : Le point de bifurcation transcritique, représenté par le point noir, représente un basculement de stabilité entre deux courbes définies par  $y = 0$  et  $\mu = y$ . Les portions grasses des courbes sont stables, les pointillées instables.

Le modèle de Levins est exprimé dans un contexte spatialement implicite, où toutes les cellules d'habitat sont équidistantes, et ne prend donc pas en compte l'existence possible d'hétérogénéités de structure spatiale de l'habitat. Il est possible d'adapter simplement le modèle pour prendre en compte la variation de la densité d'habitat (Lande 1987). On conserve une formulation du type  $\dot{y} = \mu \cdot y - y^2 = f(y, \mu)$ , avec maintenant  $\mu = h \cdot C_i - e$ . Dans ce cas, le seuil critique dépend explicitement de la densité d'habitat, tout en restant dans une logique de bifurcation.

Ce qui sera de nature à compliquer les choses est la prise en compte explicite de la structure spatiale de l'habitat, avec l'effet de la connectivité et des hétérogénéités de structure. Dans ce cas, une contrainte structurelle vient donc se superposer à l'existence de la bifurcation. D'autres problèmes apparaissent, qui sont du domaine des flux de matière et d'informations dans des réseaux : c'est le domaine des phénomènes de percolation.

La notion de seuil critique paraît en définitive assez familière et répandue dans les systèmes physiques et biologiques. Elle recouvre toutefois des phénomènes différents, et fait l'objet d'une riche littérature sur laquelle nous ne nous étendrons pas. La terminologie reste au demeurant peu explicite. Nous soulignerons cependant, dans le cadre de l'étude des métapopulations, l'importance de bien distinguer deux types de criticalité :

- *Criticalité de type bifurcation.* C'est une criticalité dynamique, dans le sens où elle exprime un changement de comportement dans un modèle purement dynamique. Nous l'avons présentée pour le modèle de Levins (Encadré 6), en habitat non structuré.
- *Criticalité de type percolation.* C'est une criticalité statique, dans le sens où elle exprimera l'effet d'une structure statique de substrat (ici l'habitat) sur des flux de matière ou d'énergie (ici des propagules).

Cette dualité statique/dynamique n'est pas consacrée dans la littérature, mais elle cadre bien avec la problématique des métapopulations. Elle cadre bien en effet avec l'hypothèse clé de découplage des dynamiques lentes et rapides. Rappelons que ce découplage apparaît à deux niveaux. D'une part la dynamique démographique est supposée être rapide devant celle de la métapopulation, d'où l'hypothèse que les équilibres démographiques sont toujours atteints. D'autre part la dynamique de l'habitat est supposée être lente devant celle de la métapopulation, d'où l'hypothèse que l'habitat est statique.

Aussi perçoit-on ici une dialectique écologique fondamentale, exprimée par le diptyque habitat *a priori*-habitat *a posteriori* (voir Introduction, p.11):

- L'habitat *a priori* impose une contrainte statique de structure, et constitue une criticalité statique, de type percolation.
- L'habitat *a posteriori* résulte d'une contrainte dynamique, basée sur la dynamique de colonisation-extinction, et exprime une contrainte de criticalité dynamique, de type bifurcation.

Il est essentiel de comprendre que les modèles spatialement réalistes de métapopulations, au sens de Ovaskainen et Hanski (2001), sont clairement fondés sur cette dialectique. Ainsi leur détermination du seuil critique d'extinction de la métapopulation est basée (voir Encadré 8, p.59) :

- sur une matrice de paysage, avec une matrice de connectivité des cellules d'habitat, qui exprime la contrainte statique de structure de l'habitat (contrainte de percolation),

- mais aussi sur la définition de la fonction de colonisation et d'extinction, qui détermine le calcul des connectivités.

Le seuil critique d'extinction qui est obtenu dépend donc à la fois des caractéristiques dynamiques intrinsèques de l'organisme considéré (capacité de colonisation et de persistance locale), et des caractéristiques structurelles de l'habitat physique.

**La distinction des deux notions d'habitat *a priori* et *a posteriori* revêt en définitive un sens véritable et profond en terme de mécanismes écologiques, dès lors que l'on introduit un découplage entre les processus lents de dynamique d'habitat, et rapides de dynamique de populations.** Cette idée sera un fondement important de tout mon travail.

#### *1.4. Criticalités et auto-organisation*

Les phénomènes de criticalité caractérisent donc souvent la dynamique non linéaire de systèmes ouverts, dissipatifs. Ils s'accompagnent de l'apparition d'une invariance d'échelle (structures *scale-free*, voir p.22), symptomatique de phénomènes d'auto-organisation. On utilise parfois la notion de **paramètre d'ordre** pour décrire la cohérence structurale du système : au voisinage du seuil critique, sa valeur bascule, et ce basculement s'accompagne d'une divergence des longueurs et temps de corrélation. Ce changement est discontinu. Le **paramètre de contrôle** est en revanche une caractéristique du système dont la variation est continue et qui possède une valeur seuil définissant l'état critique du système. Par exemple, dans le cas de la percolation, considérons un espace à deux dimensions où l'état d'un point est binaire, conducteur ou non-conducteur. Le paramètre de contrôle est la densité de points conducteurs ; le paramètre d'ordre est la probabilité d'existence d'un agrégat infini de points conducteurs. Il existe une valeur seuil du paramètre de contrôle, appelée seuil de percolation, où le paramètre d'ordre va basculer d'une probabilité non nulle à une probabilité nulle. Il est utile de raisonner à partir de ces deux paramètres et de leur interaction pour comprendre la nature des phénomènes de criticalité (Sornette 1992; Gil and Sornette 1996).

On peut schématiquement représenter trois types de criticalité, en fonction de la nature des perturbations (associées aux variations du paramètre de contrôle), de la réponse du système aux perturbations (associée aux variations du paramètre d'ordre), et du type d'auto-organisation induite (Pascual and Guichard 2005) :

- *Criticalité de type transition de phase.* L'effet des perturbations (extinctions) est uniforme dans l'espace et le temps, l'habitat est non structuré. C'est le cadre des modèles de Levins (1969) et Lande (1987), par le biais des extinctions uniformes, et plus généralement des processus de contact en habitat homogène. Le système peut être décrit par deux états, occupé et inoccupé. Ce type de fonctionnement s'apparente au modèle de Ising en physique, décrivant l'orientation des spins de particules dans un aimant (deux états). L'état des particules dépend de la configuration locale des spins, donc de l'état des voisins. C'est également le cadre des modèles épidémiologiques binaires SI (*susceptible/infective*) de dynamique d'infection, et des modèles forestiers de dynamique de régénération par volis (destruction par le vent). Dans ce type de système, il existe un seuil critique marquant une transition de phase avec un état absorbant nul stable d'un côté, et un état stable non nul de l'autre. On peut créer les conditions pour que le système se trouve dans un état critique, en intervenant sur le paramètre de contrôle, en l'occurrence l'intensité des perturbations. Dès lors le système manifestera une structure spatiale et temporelle fortement auto-organisée, avec divergence des longueurs et des temps de corrélation. Cependant, **le système n'évoluera pas spontanément vers cet état critique.** Une intervention externe sur le paramètre de contrôle est requise.
- *Criticalité auto-organisée.* Considérons à présent des régimes de perturbation plus localisés dans le temps et dans l'espace, comme l'étincelle déclenchant un feu de forêt, ou l'arrivée d'une souche infectieuse dans une population d'organismes vulnérables. Considérons par ailleurs un système à trois états : susceptible à la perturbation, détruit par la perturbation ou en phase de rétablissement. Un modèle de dynamique infectieuse SIR (*susceptible/infective/recovery*) représente ce type de système ; les individus en voie de guérison ont acquis une immunité telle qu'une certaine densité de ces individus est susceptible de limiter la propagation de toute infection future. On peut représenter de manière analogue les dynamiques forestières soumises à incendie, où les phases jeunes de repeuplement ne seront pas propices à la propagation de l'incendie. Ce type de modèle à trois états, avec des perturbations localisées, et une probabilité de rétablissement uniforme dans l'espace et le temps, induit une **évolution naturelle du système vers un état critique.** Ce comportement correspond à une rétroaction du paramètre d'ordre sur le paramètre de contrôle (Sornette 1992) ; cette rétroaction est la clé de l'évolution spontanée vers l'état critique. Le paramètre d'ordre correspond ici à la probabilité d'existence d'un ensemble de grande taille d'individus susceptibles à l'incendie ou à la maladie.

L'intensité des perturbations dépend de l'état du paramètre d'ordre. L'état critique correspond alors à une structure spatiale et temporelle fortement auto-organisée (divergence des longueurs et temps de corrélation) : on parle de *criticalité auto-organisée*. Cette notion est souvent invoquée et discutée pour comprendre la structure fractale des communautés végétales (Li 2000a; Bolliger et al. 2003). Elle fut initialement illustrée par le comportement des tas de sable (voir Encadré 7). Un ingrédient fondamental dans ce type de système est la disjonction entre l'échelle temporelle d'occurrence des perturbations (lente), et celle de rétablissement ou de relaxation (rapide) (voir Gil and Sornette 1996).

- *Criticalité robuste*. Considérons une variante du modèle SIR, ou du modèle incendie de forêt, que nous venons d'évoquer, avec une probabilité de rétablissement hétérogène, localisée, au lieu d'être uniforme. Cette localisation spatiale de la phase de rétablissement sera notamment due aux limites à la dispersion, quand par exemple il n'y pas de graines viables présentes après un incendie. Le comportement induit est proche de la *criticalité auto-organisée stricte*, à quelques nuances près. Cela réduit notamment la différence d'échelle de temps entre l'occurrence des perturbations et des rétablissements. La diffusion spatiale post-rétablissement sera plus lente, les réponses aux perturbations seront de moindre étendue, les fluctuations temporelles moins importantes. La disjonction des échelles temporelles, nécessaire à la *criticalité auto-organisée*, disparaît ; cependant le système conserve des propriétés de convergence vers un état critique, et des propriétés d'auto-organisation. C'est le sens du qualificatif de *criticalité « robuste »*. On inclura dans ce type de système des situations d'interactions trophiques, comme des modèles prédateur-proie.



#### Encadré 7 : Le paradigme du tas de sable

Halley et al. (2004) définissent la criticalité auto-organisée de la manière suivante : « A system which maintains itself in a critical state through a balance between a steady accumulation of pressure and its release through a series of discrete events. The classic paradigm is the sandpile with a steady trickle of grains to the top. The pile maintains approximately the same steepness through regular avalanches. Though the exact timing is random, pressure accumulates continuously between events so the expected magnitude of an event increases, the longer it is delayed. Other models of SOC are forest fires (with fuel accumulating) and earthquakes (shear pressure accumulation). A key signature of SOC is that the magnitudes of events and the intervals between them follow power laws. » (voir aussi Li 2002).

Dans une telle situation, le système converge vers un état critique et s'y maintient naturellement. Il existe alors des propriétés émergentes déterminées et constantes du système, par exemple la pente du tas de sable (voir Dickman et al. 2000). Les variations autour de cet état seront de nature catastrophique (avalanche de grains de sables, tremblement de terre, feu de forêt). Ainsi, alors qu'un équilibre déterministe n'admet que de petites perturbations, alors qu'un état quasi-stationnaire stochastique ne représente qu'un bruit blanc autour de l'attendu quasi-stationnaire, le système critique auto-organisé peut présenter des variations brutales et importantes, sans que cela entraîne de changement de régime. Ces variations présentent une amplitude et une fréquence inversement corrélées : on parle de bruit rouge ou de bruit  $1/f$ .

La dynamique métapopulationnelle est fondamentalement un processus de contact, avec l'existence d'un seuil critique d'extinction, accompagné d'un phénomène d'auto-organisation. Nous soulignerons dans la partie suivante l'importance d'une prise en compte de cette auto-organisation pour l'étude de la structure spatiale et temporelle des métapopulations.

## **2. Structure auto-organisée des métapopulations**

Le déterminant environnemental de la structure spatiale et temporelle des métapopulations est la distribution de l'habitat. Un changement dans la structure et l'abondance de l'habitat peut engendrer un effet de seuil, avec une extinction brutale de la métapopulation. Nous avons souligné que ce facteur constitue une contrainte statique, dans le sens où l'on néglige la dynamique de l'habitat.

Nous avons également souligné que la dynamique d’extinction-colonisation était de nature à créer un phénomène de transition de phase, et ce même dans un modèle non spatialisé (voir Encadré 6). L’existence et la détermination du *seuil critique d’extinction* dans un habitat structuré ont été largement étudiées (voir Ovaskainen and Hanski 2001; Ovaskainen et al. 2002; Bascompte 2003), et met en jeu à la fois la prise en compte du comportement dynamique des espèces (taux de colonisation et d’extinction), et la structure du réseau de cellules d’habitat.

Il est délicat de définir clairement la nature de la criticalité impliquée dans un système métapopulationnel en habitat structuré, car il ne s’intègre pas clairement dans les schémas classiques que nous avons présentés (voir §1.4). De fait, au voisinage du seuil d’extinction, il est connu qu’une métapopulation arbore une divergence des temps et des distances de corrélation :

- On sait qu’il existe des corrélations spatiales à grande échelle au voisinage du seuil d’extinction (voir par ex. Bascompte 2001).
- Un phénomène bien connu en biologie de la conservation est la dette d’extinction (*extinction debt*) : une métapopulation condamnée à disparaître, peut persister pendant très longtemps, du fait d’une corrélation temporelle à grande échelle (Hanski and Ovaskainen 2002; Oborny et al. 2005).

Cette divergence des temps et longueurs de corrélation au voisinage du seuil critique est au demeurant une caractéristique générale des processus de contact (Dickman and Martins de Oliveira 2004). Un des champs d’application les plus répandus de ce type de phénomène est l’étude de la percolation. La percolation dirigée (*directed percolation*) est le processus de contact le plus largement étudié (Hinrichsen 2000) : le phénomène de percolation traduit l’écoulement cohérent d’un corps au sein d’un système physique sur toute sa dimension spatiale. Il y a un effet de seuil séparant l’état où la percolation est possible, de celui où elle est impossible.

La criticalité du phénomène de percolation renvoie au problème de la connectivité de l’habitat (cf. Brooks 2003). Le comportement d’une espèce au sein d’un habitat spatialement structuré dépend de sa propre dynamique spatiale, tandis que la percolation est dépendante de l’échelle spatiale d’interaction entre l’espèce et son habitat (Keitt et al. 1997).

En définitive, bien que les métapopulations arborent au voisinage du seuil critique une structure auto-organisée, il n'y a pas de convergence naturelle vers un état critique, et donc *a priori* pas de criticalité auto-organisée. La conjonction d'une criticalité de type percolation et d'une criticalité de type bifurcation sera pour l'heure le principal fait à retenir. Nous ne nous attarderons pas plus avant sur la définition du type de criticalité.

Nous nous intéresserons en revanche à l'impact de l'auto-organisation spatiale et temporelle théorique des populations, notamment au voisinage du seuil critique, sur l'étude des métapopulations réelles.

En effet, cette incidence semble paradoxalement sous-estimée dans les modèles théoriques de dynamique, malgré l'engouement sur le thème de l'auto-organisation. En particulier, nous montrerons l'importance de la prise en compte de l'auto-organisation quand l'état quasi-stationnaire des métapopulations est déterminé sur un temps et un espace fini. Plus on est proche du seuil critique, plus le poids de l'auto-organisation dans la prédiction de l'état du système sera important.

### *2.1. Seuil d'extinction des métapopulations et vision dynamique de l'habitat*

Le seuil critique d'extinction est une caractéristique topologique du réseau des cellules d'habitat favorable, et dépend en particulier de leur nombre et de leur structure de connectivité (Hanski and Ovaskainen 2000; Ovaskainen and Hanski 2001). La méthode de calcul classique s'appuie dès lors sur la matrice de connectivité des cellules, qui dépend de la fonction de colonisation et de la taille des cellules (voir Encadré 8).

### Encadré 8 : Calcul du seuil critique d'extinction pour un graphe de cellules d'habitat

Le paramètre  $\lambda_M$ , que l'on peut appeler *seuil critique d'extinction* (Bascompte 2003), ou *capacité de persistance* de la métapopulation, définit pour un réseau donné de cellules d'habitat favorable une condition de seuil sur le maintien à long terme de la métapopulation. Le calcul que nous présentons ici pour ce **paramètre topologique** est proposé par Ovaskainen et Hanski (2001). Il repose sur une version spatialement réaliste du modèle de Levins (1969), déterministe et exprimée en temps continu. Le passage en temps discret est possible et étudié par Ovaskainen et Hanski (2001) dans le cadre des modèles à fonction d'incidence (*Incidence Function Model*).

Le calcul est fondé l'équation suivante de dynamique métapopulationnelle :

$$\frac{dp_i(t)}{dt} = C_i(p(t)).(1 - p_i(t)) - E_i.p_i(t), \text{ où } C_i(p(t)) \text{ est le taux de colonisation des cellules}$$

vides, et  $E_i$  les taux d'extinction des cellules occupées.

Les auteurs définissent en outre les écritures suivantes :  $E_i = e/A_i$  et  $C_i(p(t)) = c \cdot \sum_{j \neq i} A_j \cdot e^{-\alpha d_{ij}} p_j(t)$ . Ils constituent la matrice  $M$  dont les termes sont  $m_{ij} = A_i \cdot A_j \cdot e^{-\alpha d_{ij}}$ . La condition de seuil de persistance porte sur  $\delta = e / c$ , soit une persistance de la métapopulation assurée pour  $\delta < \lambda_M$ .

Les auteurs montrent que  $\lambda_M$  est la valeur propre principale de  $M$ . Nous invitons le lecteur intéressé par de plus amples détails de démonstration à se reporter au travail de Ovaskainen et Hanski (2001).

Dans le cas de nos métapopulations simulées (Chapitre I), les cellules présentent une superficie identique, que l'on peut fixer à 1. On écrit alors  $A_i = A_j = 1$ , ainsi que  $\alpha = 1$  et

$$c = \frac{Ct}{\sum_{k=1 \dots 5} 8.k.e^{-k}} \text{ pour retrouver le paramétrage utilisé pour nos métapopulations simulées.}$$

En définitive la condition de seuil sur  $r = e / Ct$  est facilement déduite de celle sur  $\delta$ , grâce à la diagonalisation de la matrice  $M$ , dite matrice de paysage. Ces calculs ont été réalisés pour l'ensemble des cartes d'habitat *a priori* et *a posteriori* associées à nos métapopulations simulées. Nous avons eu recours pour cela au logiciel *Matlab*.

L'ambiguïté de la définition de l'habitat dans ce type de représentation est à l'origine d'un débat animé sur la notion de métapopulation (Freckleton and Watkinson 2002; Ehrlén and Eriksson 2003; Freckleton and Watkinson 2003; Baguette 2004; Hanski 2004) :

- Les plus rigoristes revendiquent que la représentation métapopulationnelle ne concerne que les cellules d'habitat où des événements de colonisation sont observables, sous-entendu dans un temps écologique. Cela dépend donc de l'échelle spatiale et temporelle d'observation, et rejoint partiellement notre objectif de travail, la signification écologique de l'habitat *a posteriori*. Il faut toutefois se méfier de la différence conceptuelle entre colonisation observable et colonisation observée. *Observable* se rapporte à une probabilité, *observé* se rapporte à une réalisation d'événement (qui peut porter sur de nombreuses générations).
- D'autres incluent toutes les cellules dont l'habitat est favorable dans l'absolu d'après la connaissance que l'on a des exigences de l'espèce. Cela rejoint la vision hutchinsonienne de la niche, et bien évidemment notre vision de l'habitat *a priori*.

Nous sommes ainsi au cœur du problème de la définition de la niche d'une espèce, et en particulier de la prise en compte de sa dimension dynamique (voir Introduction §1.2). D'une part, si l'on considère l'habitat réellement accédé **dans un temps écologique**, le seuil critique d'extinction de la métapopulation sera défini à partir du graphe de l'habitat *a posteriori*. En revanche, si l'on considère l'habitat potentiellement favorable et accessible, dans la logique d'un état stationnaire stochastique exprimé **en un temps infini**, le seuil d'extinction serait défini à partir de l'ensemble des cellules de l'habitat *a priori*. Donc les topologies de l'espace accessible et de l'espace accédé sont susceptibles de différer. Le seuil d'extinction de la métapopulation n'est alors pas nécessairement le même dans une conception d'habitat *a priori* et dans une conception d'habitat *a posteriori*.

Le fait qu'une espèce n'occupe dans un temps écologique qu'une partie de l'espace *a priori* disponible nous amène à nous interroger sur la dynamique des espèces au sein de l'habitat *a posteriori*. L'habitat *a posteriori* représente la réalisation d'une occupation partielle de l'habitat *a priori* durant l'état quasi-stationnaire, et est susceptible d'être façonnée par des événements contingents (voir §1.1). Dans ces conditions, il est important que l'étude de la persistance d'une métapopulation, pour la mise en place de réserves notamment, intègre la prise en compte des structures contingentes particulières. L'existence de structures auto-organisées dans l'espace, façonnant l'habitat *a posteriori*, aurait ainsi une

importance écologique cruciale. Nous nous proposons donc d'examiner, sur la base de nos métapopulations simulées, s'il existe une différence entre le seuil d'extinction calculé pour l'habitat *a priori* et le seuil calculé pour l'habitat *a posteriori*.

Considérons donc nos métapopulations théoriques simulées en habitat binaire (voir Chapitre I). Les seuils d'extinctions sont calculés pour les 227 distributions quasi-stationnaires non nulles, pour les deux représentations de l'habitat, et suivant la méthode de Hanski (1998) (voir Encadré 8). Ces valeurs théoriques sont confrontées aux seuils observés, c'est à dire aux valeurs  $r$  limites obtenues par simulations pour chaque carte d'habitat : c'est donc un seuil observé évalué grossièrement à 0.1 près (Figure 5a).

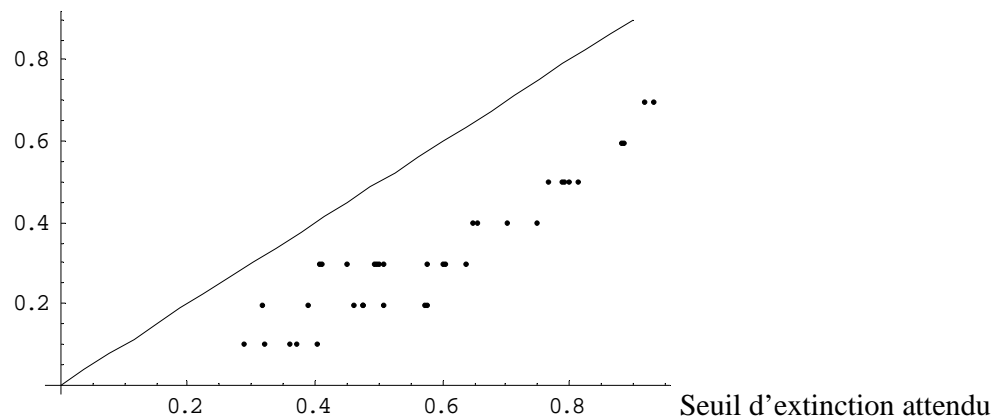
Nous montrons que la relation entre les seuils critiques d'extinction calculés pour l'habitat *a priori* et pour l'habitat *a posteriori* dépend fortement du nombre de générations disponibles pour estimer l'habitat *a posteriori* (Figure 5b). Quand il y a un nombre élevé de générations d'estimations (ici 300), les seuils critiques de l'habitat *a priori* et de l'habitat *a posteriori* coïncident. En fait la structure locale de l'habitat *a posteriori* ne diffère pas de la structure de l'habitat *a priori*. Notons que l'étendue spatiale de l'habitat *a posteriori* peut être cependant très différente de celle de l'habitat *a priori* (voir Figure 6a).

Maintenant si l'on considère l'estimation d'un habitat *a posteriori* sur 5 générations d'observation seulement, la coïncidence des seuils d'extinction des habitats *a priori* et *a posteriori* n'est plus constatée (Figure 5b). Il y a en revanche un biais systématique vers un seuil plus faible quand on estime l'habitat *a posteriori* sur un nombre faible de générations. Il y a donc alors l'apparence d'une structure spatiale des populations plus « précaire ». La coïncidence de la structure locale de l'habitat *a priori* et de l'habitat *a posteriori* est donc asymptotique, **mais apparaît sans nécessairement que les deux habitats soient identiques** (Figure 6b). En fait, il apparaît plutôt que l'habitat *a posteriori* permet une estimation correcte lorsqu'il y a **exploration complète sur une zone localisée de l'habitat *a priori***. Quand le nombre de générations d'estimation est insuffisant, cette convergence de structure ne peut pas être constatée. Transitoirement les espèces explorent une portion de l'espace qui apparaît moins favorable en terme de connectivité que ce que permettrait une distribution plus « uniforme » dans l'habitat disponible localement. Nous constatons en outre que l'écart des seuils d'extinction à 300 et 5 générations apparaît linéairement relié au paramètre  $r$  effectif des simulations (Figure 5c). De fait, lorsqu'une espèce occupe fortement son habitat local ( $r$  faible) on obtient rapidement une bonne conjonction de la structure locale des habitats *a*

*priori* et *a posteriori*. Inversement, lorsqu'une espèce occupe peu son habitat ( $r$  élevé), il faut bien plus de générations pour que la structure locale de l'habitat *a posteriori* constitue une bonne image de la structure de l'habitat *a priori*.

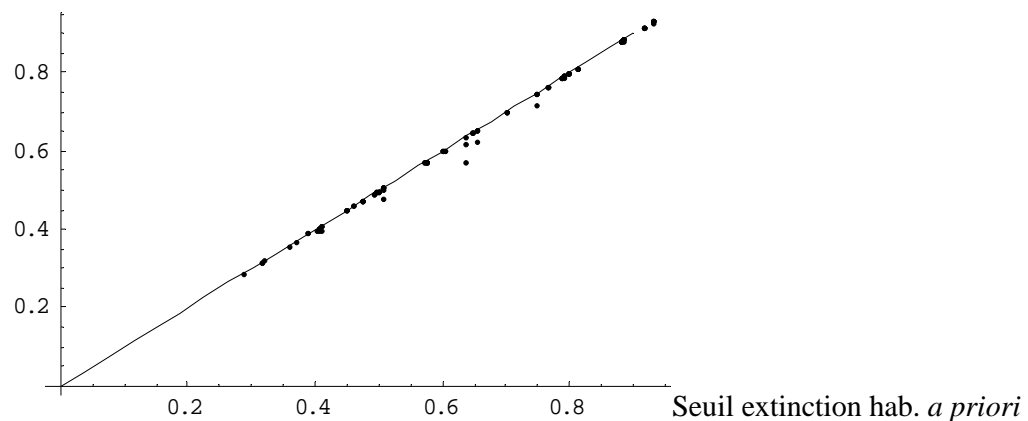
- (a) Calcul du seuil d'extinction en considérant le graphe d'habitat *a priori*.

Seuil d'extinction observé

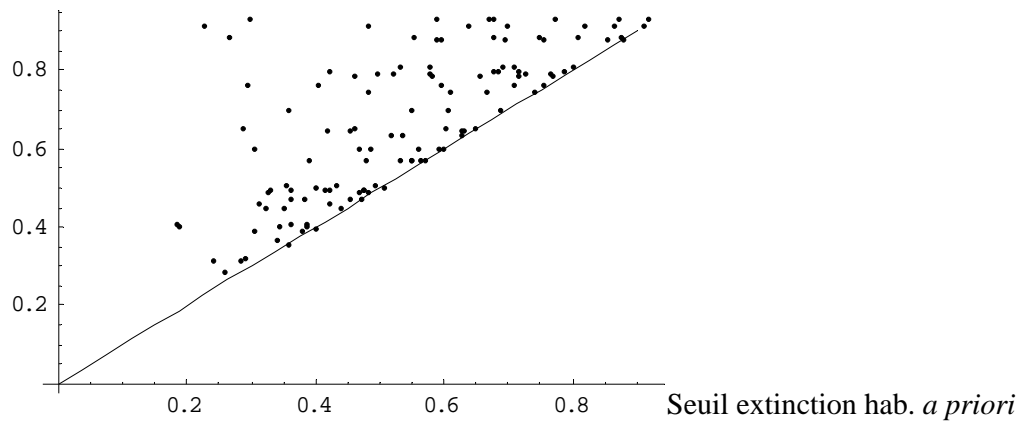


- (b) Comparaison des seuils critiques calculés pour l'habitat *a priori* et l'habitat *a posteriori*, pour deux gammes temporelles d'estimation de l'habitat *a posteriori*.

Seuil extinction hab. *a posteriori* estimé sur 300 générations



Seuil extinction hab. *a posteriori* estimé sur 5 générations



(c) Seuil extinction hab. *a posteriori* estimé sur 300 gén. – Seuil sur 5 gén.

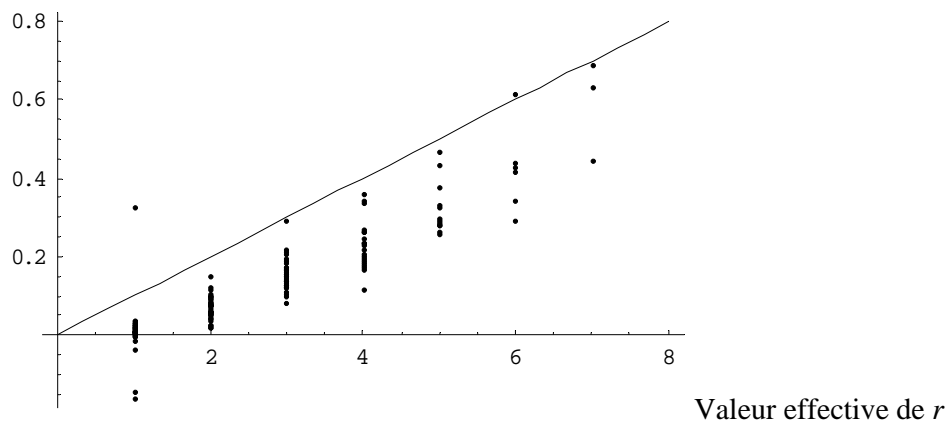


Figure 5 : Calcul des seuils critiques d'extinction des métapopulations simulées en habitat binaire, selon la méthode de Hanski and Ovaskainen (2000). Pour toutes les figures, la droite  $y = x$  est représentée en traits continus.

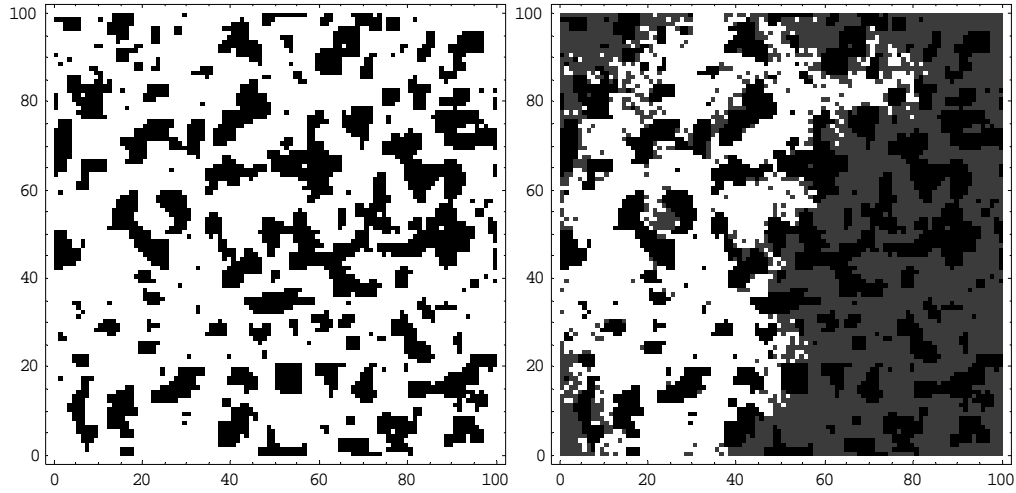
(a) Relation entre le seuil critique d'extinction théorique et le seuil d'extinction obtenu par simulations sur 8000 générations, pour chaque carte d'habitat binaire paramétrée  $(p, q)$  (1 point = 1 carte).

(b) Comparaison des seuils d'extinction théoriques calculés pour l'habitat *a priori* et l'habitat *a posteriori*. Deux différentes gammes temporelles d'estimation de l'habitat *a posteriori* sont proposées : sur 300 générations, et sur 5 générations. Dans le premier cas, le seuil critique d'extinction de l'habitat *a posteriori* coïncide avec celui de l'habitat *a priori*. Dans le second cas, il y a un biais fort vers des seuils critiques plus faibles en habitat *a posteriori*.

(c) On représente la différence entre les seuils d'extinctions de l'habitat *a posteriori* sur 300 générations et sur 5 générations, en fonction de la valeur effective du paramètre  $r$  pour chaque simulation.

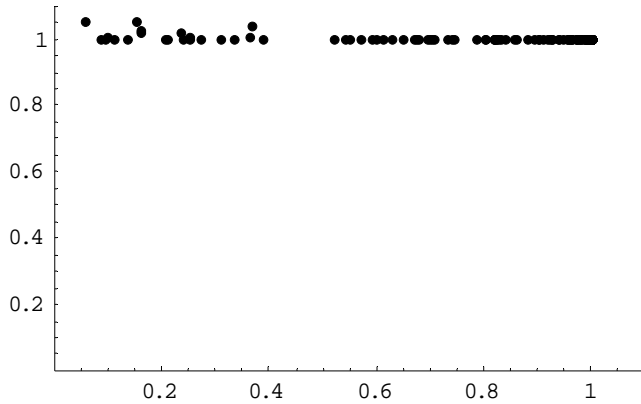


(a) Illustration de cartes d'habitat *a priori* et *a posteriori*, avec  $p = 0.7$ ,  $q = 0.9$  et  $r = 0.7$ .  
Habitat *a priori* Habitat *a posteriori* 300 gén.



(b) La convergence des seuils critiques d'extinction de ces habitats ne signifie pas une saturation de l'habitat *a priori*.

Ratio des seuils calculés pour hab. *a priori* et *a post* sur 300 gén.



Prop. d'hab. *a post* sur hab. *a priori* sur 300 gén.

Figure 6 : (a) Cartes d'habitat *a priori* (à gauche) et *a posteriori* sur 300 générations (à droite), pour  $p = 0.7$ ,  $q = 0.9$  et  $r = 0.7$ . L'habitat *a posteriori* est inclus dans l'habitat *a priori*, et l'habitat *a priori* inoccupé est alors représenté en gris. (b) Relation entre le taux de saturation habitat *a posteriori* sur habitat *a priori*, et rapport des seuils critiques d'extinctions entre les deux habitats (121 simulations représentées).

Ces figures illustrent d'une part que souvent l'habitat *a posteriori* calculé sur 300 générations ne correspond pas à une saturation de l'habitat *a priori*, et que d'autre part cela contraste avec la forte convergence structurale observée lors du calcul des seuils d'extinction sur 300 générations (voir Figure 5a).

Une autre information importante pour notre propos est l'absence de concordance entre le seuil d'extinction de l'habitat *a posteriori* sur 5 générations et la valeur du paramètre  $r$  des simulations (non montré). On aurait obtenu une telle égalité en cas de convergence de la distribution des populations vers une structure spatiale critique à courte échelle de temps. En l'absence d'une telle convergence, nous en déduisons que la criticalité de type percolation induite par la structure de l'habitat ne constitue pas pour la dynamique métapopulationnelle une criticalité auto-organisée (voir §1.4). Enfin, nous signalons que les résultats obtenus avec un habitat *a posteriori* estimé sur une seule génération conduisent aux mêmes observations générales qu'avec l'estimation sur 5 générations.

Plusieurs aspects cruciaux peuvent finalement être soulignés :

- **Les populations occupent peu à peu l'habitat *a priori*, mais à un rythme relativement lent.** Pour 300 générations d'estimation de l'habitat occupé, avec un turnover élevé d'extinctions et de colonisations (voir Chapitre I,  $e = 0.09$ ), il y a souvent une occupation très partielle de l'habitat disponible (Figure 6a).
- En revanche, **la structure de l'habitat occupé tend nettement plus vite vers une structure proche de celle de l'ensemble de l'habitat occupable** (Figure 6b).
- **Sur un temps écologique, il y a donc une occupation partielle de l'habitat disponible, indépendamment de toute contrainte biogéographique,** simplement par le jeu des processus stochastiques. La Figure 6a montre un effet de contingentement qui est purement stochastique, mais pourrait sur le terrain être faussement conçu comme un effet biogéographique. **Si l'on souhaite caractériser la structure de l'habitat *a priori* à partir des données d'occupation des populations, c'est donc la structure de connectivité (Figure 5b) et non la densité des patches occupés (Figure 6a) qui importe.** Nous retrouverons plus loin cette conclusion importante (Ch. III.1.4).
- **Enfin la structure spatiale d'occupation est contrainte par la structure de l'habitat, par la dynamique métapopulationnelle, mais ne correspond pas à une situation de criticalité auto-organisée.**

Cela souligne en particulier que le débat autour d'une représentation exhaustive de l'habitat accessible au sein d'une métapopulation n'est pas nécessairement pertinent. C'est la structure d'occupation sur un temps écologique, en particulier en terme de connectivité générale, qui exprime l'empreinte de l'habitat *a priori*. En revanche, la dynamique

métapopulationnelle contraint notablement non seulement le degré de saturation instantanée de l'habitat *a priori*, mais aussi le rythme d'exploration spatiale de cet habitat.

Cela ouvre enfin deux perspectives majeures pour l'inférence, que nous retrouverons dans les Chapitres suivants. On constate sur la Figure 6a qu'il y a une zone de présence des populations, dans laquelle les propriétés de turn-over de populations pourront être observées et une zone non occupée pendant la période d'observation, donc non informative. Comme nous le verrons plus loin, cette hétérogénéité est de nature à entraver l'inférence, car sa nature contingente complique les tentatives de modélisation. On a alors deux grands types de solution pour s'affranchir de ce problème :

- Une solution consiste à n'effectuer des estimations de paramètres que là où il y a de l'information, c'est à dire en introduisant de façon explicite qu'on ne réalise les estimations que dans la zone active.
- Une autre solution, *a priori* plus élégante, est d'explicitier la *structure globale*, zone occupée versus zone non occupée, au moyen d'un seul paramètre. Les autres paramètres, dissociés de celui-ci, exprimeront alors la *structure locale* au sein de la zone de présence des populations. Nous verrons en particulier que c'est ce que peut permettre de réaliser une analyse spectrale (Ch. III.1), dissociant la signature spatiale de la dynamique métapopulationnelle de l'effet « grande zone inoccupée ».

## 2.2. Propriétés émergentes

L'auto-organisation des processus de contact est difficilement représentable mathématiquement, car il faut pouvoir modéliser l'incidence d'événements contingents. Une petite perturbation peut engendrer une avalanche, comme elle peut engendrer un petit écart à l'équilibre sans avalanche. Il s'agit avant tout d'un problème d'échelle. Pour un modèle stochastique comme pour un modèle déterministe, l'état **asymptotique** du système dépend des conditions locales d'équilibre des objets avec leur environnement. Quand il y a des avalanches, les objets ne sont **temporairement** plus en équilibre avec leur environnement immédiat (voir Encadré 3), et c'est au niveau des agrégats qu'il faut travailler : il y a un changement d'échelle de temps et d'espace. Dans ce contexte, **un agrégat de cellules occupées devient une entité fondamentale** avec un taux donné de colonisation des zones voisines. L'échelon local ne permet plus de décrire correctement les conditions d'équilibre : **il faut passer à une échelle supérieure, à un échelon intermédiaire dans la représentation spatiale et temporelle du système.**

Cette propriété de changement d'échelle n'est pas modélisable par les techniques habituelles de représentation des processus dynamiques, qu'elles soient discrètes ou continues. Il y a cependant des solutions pour représenter l'invariance d'échelle des processus auto-organisés et en déduire certaines de leurs caractéristiques (Weitz 1997; Ciliberti et al. 2000). Cela permet de caractériser certaines lois d'échelle dans le temps et dans l'espace (Barabasi and Albert 1999). Ces structures ne sont pas modélisables par un modèle d'équilibre local, et sont qualifiées de structures émergentes. Les structures spatiales résultantes sont riches et diversifiées (cf. par exemple Wakano et al. 2003).

Pour illustrer l'idée d'émergence, nous étudions le spectre d'une séquence temporelle de densités de populations au sein d'une métapopulation, au moyen d'une analyse de Fourier. Cette représentation décrit la structure d'un signal temporel sur une échelle de fréquences : s'il existe une périodicité forte du signal, le spectre possèdera une valeur élevée pour la fréquence associée. Un spectre lissé est un spectre corrigé afin de limiter l'erreur d'estimation (voir Chapitre III). Pour un processus de contact tel que notre modèle de métapopulation, les petites fluctuations seront fréquentes et les grandes plus rares, ce qui induira ce que l'on appelle une coloration du spectre, ou encore une loi de puissance : il y aura en coordonnées log-log une relation linéaire négative entre les fréquences et les coefficients du spectre. Cela est en théorie d'autant plus marqué que l'on se situe à proximité du seuil critique d'extinction (Weitz 1997; Dickman and Martins de Oliveira 2004).

La Figure 7 représente une telle coloration linéaire en log-log. Le spectre représenté est une approximation de la structure obtenue sur 8 simulations de métapopulations en habitat uniforme ( $r = [0.1, \dots, 0.8]$ ). Nous réalisons une Analyse en Composantes Principales sur les spectres lissés des simulations et mettons en exergue cette structure commune et prépondérante sur les spectres (91% de la variance sur les 8 spectres décrit par la première composante CP1). Nous avons ainsi l'exemple d'un processus sans mémoire, exclusivement décrit par des événements d'une génération à la suivante, mais qui induit dans le temps une structure spectrale singulière et corrélée sur toute la gamme des fréquences. Il s'agit en l'occurrence d'une propriété émergente.

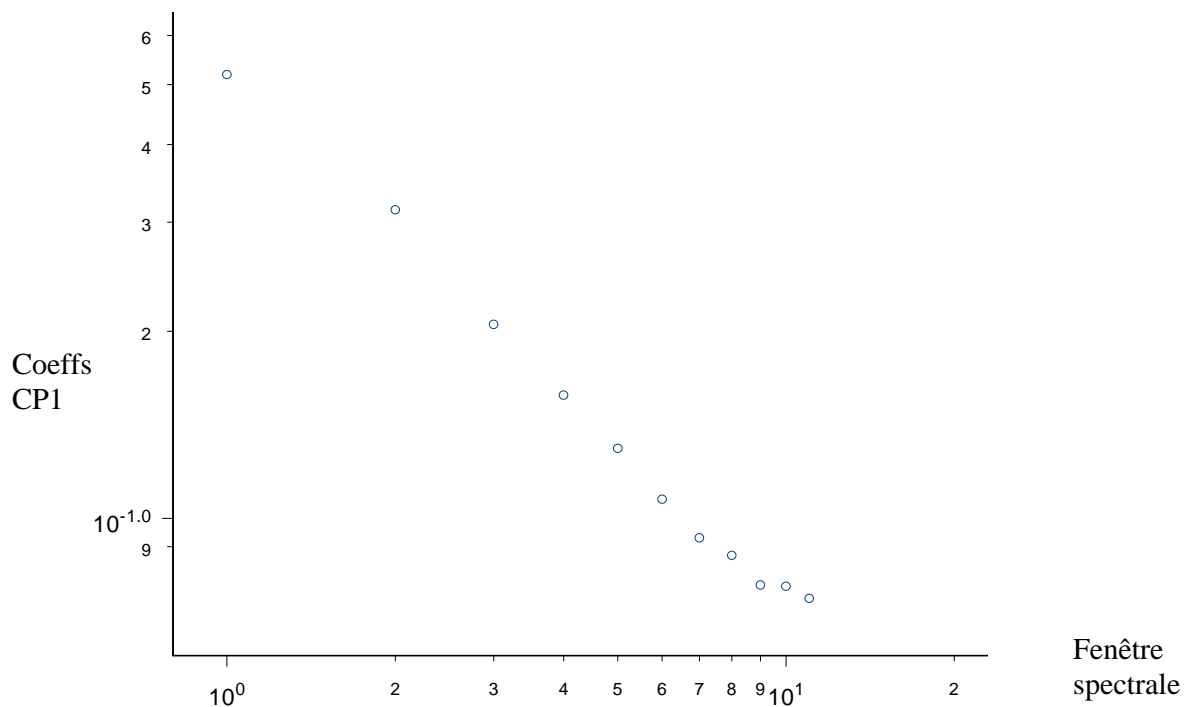


Figure 7 : Représentation de la première composante principale de l'ACP (91% de variation expliquée) réalisée sur les spectres temporels lissés des effectifs quasi-stationnaires pour les 227 métapopulations simulées en habitat structuré (300 générations utilisées). La représentation en log-log permet de mettre en évidence une relation linéaire négative caractéristique représentant la coloration du spectre.

Il est difficile de caractériser et de tester l'existence de telles lois d'échelle, comme il est difficile de trouver des conditions suffisantes pour montrer l'existence d'auto-organisation. Cela est à l'origine d'un débat passionné entre les tenants de l'auto-organisation et leurs détracteurs (Halley et al. 2004). En particulier d'autres processus peuvent induire des lois de puissance : on doit se contenter de conditions nécessaires et non suffisantes (voir la démarche de Bak 1997).

Mon propos n'est pas de rentrer dans un tel débat mais de souligner l'existence possible d'événements contingents induisant des avalanches de populations. Nous verrons les limites des modèles dynamiques usuels dans les situations où l'auto-organisation est la plus probable, ce qui servira d'appui sinon de démonstration en faveur de cette théorie.

### 2.3. Une question d'échelles

Nous soulignons en définitive l'opportunité de dissocier l'échelle de représentation du processus de l'échelle de représentation des structures résultantes :

- Nous avons parlé d'une inertie des avalanches d'événements de colonisation et d'extinction, car plus une avalanche est importante, plus le retour à l'équilibre est long. Il s'agit d'une *inertie locale*, fruit des règles locales d'interaction par colonisation.
- Si l'on caractérise le système par des grandeurs globales, sur une métapopulation entière, la perception de ces phénomènes locaux n'est pas directe et est susceptible d'être altérée par les effets d'émergence. La variation d'occupation des populations dans le temps et dans l'espace représente alors une *inertie globale* du système, liée à la variation du nombre total de populations.
- Bien que ces inerties soient reliées, leur échelle de représentation diffère et elles n'ont donc pas la même valeur pour la compréhension et l'inférence des processus. L'inertie globale exprimera des propriétés émergentes sur le nombre total de populations. L'inertie locale exprimera plus directement l'auto-organisation au travers des avalanches locales. Les processus mis en évidence et les structures résultantes sont susceptibles de différer.

**L'idée est que l'on ne voit pas la même chose quand on étudie le système métapopulation à partir de statistiques locales et à partir de statistiques globales. Nous montrerons ainsi, pour les questions que nous posons, que l'échelle pertinente se situe entre les deux.** En physique, cette problématique importante est traduite par le triptyque microscopique – mésoscopique - macroscopique. L'échelle microscopique est le lieu des phénomènes locaux stochastiques, l'échelle mésoscopique le lieu des phénomènes auto-organisés, et l'échelle macroscopique le lieu des propriétés émergentes déterministes. Suivant le choix de l'échelle, on ne recourra pas aux mêmes formalismes, et surtout on n'accédera pas aux mêmes propriétés.

Cela aura des incidences sur le choix de la mesure, de la méthode de modélisation et enfin de la méthode de validation des attendus théoriques sur la structure des métapopulations. Cette idée essentielle sera au cœur des parties suivantes. Nous verrons qu'un modèle centré sur une représentation locale de l'équilibre est systématiquement entaché d'erreur quand les effets d'auto-organisation sont forts. **En considérant le comportement du système à une échelle supérieure, en fait à un échelon intermédiaire entre le niveau local et le niveau global, nous verrons qu'il est possible de surmonter les effets confondants de l'auto-organisation, sans qu'il soit nécessaire de modéliser explicitement les structures émergentes auto-organisées.**

### III. Modélisation inférentielle : des effets aux causes

Quand on s'intéresse aux mécanismes et aux conséquences d'un processus écologique, nous avons souligné en Introduction l'importance de la définition précise du système et des objets étudiés. Dans mon travail, le processus étudié est la dynamique métapopulationnelle en habitat structuré. Le système est une étendue spatiale de cellules discrètes où les conditions environnementales varient, et où des populations (les objets) apparaissent et disparaissent au gré des événements de colonisation et d'extinction.

A partir de ce point, nous nous interrogeons sur les conséquences du processus dynamique métapopulationnel, en terme de structurations spatiale et temporelle des populations. Il nous faut donc d'abord décider comment mesurer ces structures. Dans ce Chapitre, nous considérons l'importance du choix d'une *mesure pertinente*. Nous ne modélisons donc pas le mécanisme instaurant une relation entre des caractéristiques locales et des structures globales émergentes ; ce sera l'objet du Chapitre IV. Nous nous intéressons en revanche à une manière informative de mesurer les structures résultantes. Cette démarche présentera deux étapes :

- Une manière simple d'estimer l'information exprimée sera d'abord de comparer les mesures obtenues pour un ensemble de métapopulations simulées sur une grande gamme de conditions d'habitat et de dynamiques de colonisation-extinction. Si les résultats de cette comparaison permettent de décrire une forte discrimination des métapopulations, cela signifie que la mesure utilisée capture une information importante de notre système. Si en outre la comparaison met à jour un nombre de degrés de libertés similaire à celui introduit pour décrire initialement le système, notre description apparaîtra pertinente et permettra de décrire effectivement la structure de fonctionnement des métapopulations. Bien évidemment, nous chercherons en particulier une mesure qui pourra exprimer des aspects importants *de nature écologique*.
- Sur ce principe, **la recherche d'une mesure pertinente est un préalable à l'écriture de statistiques d'inférence sur des paramètres écologiques**. En l'occurrence, nous rechercherons la possibilité d'inférer à partir d'une bonne mesure spatiale les paramètres utilisés pour la simulation des métapopulations (Chapitre I) : la densité d'habitat  $p$ , l'agrégation d'habitat  $q$ , et le ratio extinction/colonisation  $r$ . Il faut bien noter en

définitive que l'échec éventuel de la démarche d'inférence traduirait soit l'inadéquation de la mesure spatiale à répondre à des questions écologiques, soit le fait que les paramètres  $p$ ,  $q$  et  $r$  sont de mauvais index écologiques, soit enfin une combinaison des deux causes. Nous faisons donc l'hypothèse *a priori* que  $p$ ,  $q$  et  $r$  sont des paramètres pertinents, et nous concentrons sur la recherche d'une mesure pertinente de la structure spatiale.

Une telle recherche n'est ni arbitraire, ni purement descriptive. Le choix de la mesure sera orienté par des hypothèses sur la nature de la relation entre processus métapopulationnel et structures résultantes. Sous cet angle, la recherche d'une mesure pertinente est un moyen de confirmer ou d'infirmer ces hypothèses, et est donc étroitement liée à la démarche hypothético-déductive.

## **1. Mesurer l'espace écologique : pertinence de l'analyse spectrale**

De nombreuses méthodes d'analyse spatiale existent, et sont associées également à de nombreuses mesures différentes de la structure spatiale (Liebhold and Gurevitch 2002). Le sens écologique et le lien entre ces différentes approches semblent être aujourd'hui très mal maîtrisé (Dale et al. 2002).

Dans un habitat uniformément favorable, le modèle le plus simple de dynamique métapopulationnelle implique des extinctions uniformément réparties et des mouvements de colonisation de proche en proche ; elle s'apparente ainsi à un processus de réaction-diffusion. Notre hypothèse est alors qu'une telle dynamique est susceptible d'engendrer des structures d'occupation dans l'espace et dans le temps *autocorrélées* et *périodiques* simples. Par ailleurs, cette dynamique va en fait la plupart du temps se produire au sein d'un habitat hétérogène, structuré en secteurs favorables ou non à la persistance de population(s). Cette structure d'habitat sera souvent elle-même caractérisée par une autocorrélation et une périodicité spatiale plus ou moins importantes (O'Neill et al. 1988).

Nous nous sommes dès lors intéressés aux techniques d'analyse spectrale, qui permettent de mesurer et d'étudier l'amplitude de telles structures auto-corrélées et périodiques (Platt and Denman 1975; Chung 1994). Il s'agit en particulier de démêler la structuration engendrée par la dynamique de l'organisme considéré (extinctions et colonisations), et la structure de son habitat favorable. Nous supposons donc que l'analyse spectrale fournira une mesure



pertinente de telles structures et permettra d'identifier l'effet des variations de dynamique métapopulationnelle. Nous testons et discutons dans la présente partie cette hypothèse.

### *1.1. Qu'est-ce que l'analyse spectrale ?*

Au sein du système écologique étudié, on définit un système de coordonnées pour le positionnement spatial des objets, très généralement la position géographique. On définit en outre une mesure de distance pour le positionnement des objets entre eux. On utilise en général les distances géographiques, mais on peut aussi définir des métriques dites écologiques prenant en compte la dissimilarité écologique des sites et des communautés, en fonction des processus de colonisation, de la présence de barrière physique à la dispersion...

Une fois que l'on a choisi une mesure de position et une mesure de distance, on travaille sur la relation spatiale entre les états des objets. L'état des objets est à la base une fonction dépendante de la position et les processus écologiques régissent la variation des états dans l'espace. L'analyse spectrale permet de représenter l'état des objets dans un autre référentiel associé à la mesure de distance. Une nouvelle fonction d'état est alors exprimée par un ensemble de composantes spectrales associées à différents niveaux de corrélation ou de périodicité spatiale : cette nouvelle fonction est le spectre des états.

Nous souhaitons nous éloigner d'une analyse précise « cartographique » des structures spatiales, qui mettrait en exergue la position des objets, et nous préférons étudier spécifiquement les patrons de corrélation et de périodicité résultant 1) de la structure de l'habitat et 2) des caractéristiques de l'espèce en terme d'extinction et colonisation. L'objectif est donc de passer d'une représentation explicite de l'espace à une représentation implicite, exprimant les structures corrélées et périodiques.

L'analyse de spectre de puissance (*power spectrum analysis*) est une forme d'analyse de variance, où la variance spatiale du signal centré (moyenne nulle) est décomposée suivant des classes de fréquences (ou de corrélations) spatiales (Platt and Denman 1975). Le terme de spectre est utilisé dans de nombreux contextes en mathématiques et en physique. Newton semble être le premier à l'avoir présenté dans un contexte scientifique, pour traduire la décomposition de la lumière blanche en un faisceau de rayons colorés, grâce à un prisme de verre. L'usage de ce terme traduit l'idée de décomposition d'un signal en une multitude de composantes représentant un gradient de longueurs d'onde ou de fréquences (dans le cadre de Newton, il s'agira des longueurs d'ondes électromagnétiques).

Ce que nous proposons de faire, c'est donc de décomposer la variance spatiale d'une structure de populations en un ensemble de composantes spectrales, associées à différentes périodes ou degrés de corrélation spatiale. Ainsi pourra-t-on caractériser l'importance au sein d'une structure de certaines gammes de fréquences, parce qu'elles décrivent une partie importante de la variance totale.

Mathématiquement, l'analyse spectrale est fondée sur un changement de référentiel, qui permet de passer de l'espace des positions à l'espace des fréquences (ou des corrélations). Cette opération présente la propriété importante de conserver intégralement l'information dans les deux espaces. Il y a en particulier une réversibilité du processus : on peut passer de la fonction d'état au spectre des états et réciproquement par une simple transformation mathématique. On dit qu'il y a **dualité entre l'espace réel et l'espace spectral**.

La lecture intuitive d'un spectre sera plus délicate que celle d'une carte, car plus abstraite. Cependant, nous ciblerons ainsi spécifiquement une fraction de l'information spatiale, supposée pertinente pour traiter nos questions écologiques. Il demeure qu'il faudra veiller à adopter une lecture prudente et robuste des informations spectrales, afin de se prémunir d'éventuels artéfacts.

#### Note importante

Les propriétés de l'analyse spectrale qui nous intéressent ici ont trait à la possibilité de décomposer la variance spatiale sur une échelle de fréquences ou de corrélations. Ces propriétés apparaissent partagées par des objets mathématiques de natures fondamentalement différentes, bien que la littérature souffre d'une absence de synthèse claire sur le sujet. Ainsi, malgré leur différence de nature mathématique, les deux méthodes que nous présenterons par la suite auront des caractéristiques et propriétés communes, au point de rencontre des idées de représentations fréquentielle et corrélative.

Les différences plus subtiles et mathématiques seront en dehors du cadre de mon travail. Nous nous contenterons de souligner que malgré des natures fondamentalement différentes, les deux méthodes choisies sont de nature à exprimer les mêmes propriétés écologiques.

### *1.2. Stationnarité spatiale, principe d'incertitude et lissage*

L'hypothèse de l'analyse spectrale la plus souvent discutée est la stationnarité spatiale. La décomposition fréquentielle *unique* d'un signal est peu informative sur les variations dans

l'espace et/ou le temps de la structure du signal : il peut y avoir par exemple localement des variations à haute fréquence et ailleurs des variations à basse fréquence, qui ne seront pas exprimées par l'analyse spectrale. L'analyse spectrale ne sera pas pertinente pour représenter de telles variations de la structure fréquentielle.

La dynamique métapopulationnelle étant reliée au comportement intrinsèque de l'espèce, nous conjecturons que les caractéristiques fréquentielles associées ne dépendent pas de la position, donc respectent la condition de stationnarité spatiale. Actuellement, une démarche alternative en vogue qui prend en compte les variations spatiales ou temporelles de structure fréquentielle est l'analyse en ondelettes (voir par exemple Keitt 2000). La contrepartie à une telle levée de l'hypothèse de stationnarité spatiale est une inflation de degrés de liberté, rendant nécessaire le recours à des jeux de données de grandes tailles. De fait l'analyse en ondelettes est actuellement dévolue à l'étude de jeux de données de grandes tailles, par exemple en météorologie (Torrence and Compo 1998), et en analyse d'images de télédétection (Csillag and Kabos 2002) .

Il existe une relation directe entre cette hypothèse et l'hypothèse de stationnarité spatiale couramment utilisée en écologie (voir Legendre and Legendre 1998), par le biais des covariogrammes (Platt and Denman 1975). Toutefois, elle est également reliée à une autre propriété propre à l'analyse spectrale qui demeure trop souvent occultée : le principe d'incertitude.

Ce principe stipule que l'on ne peut pas caractériser précisément et simultanément la valeur d'une variable en un point et la valeur des composantes spectrales de cette variable. Cela implique l'existence d'un biais d'estimation des composantes spectrales quand on passe de l'espace réel à l'espace spectral. On recourt en général au lissage, une technique qui permet de moyenner l'estimation des composantes spectrales sur une gamme de valeurs de fréquence (voir par exemple Platt and Denman 1975). Pour un entier  $s$  diviseur du nombre total de composantes fréquentielles du spectre initial, il s'agit de regrouper les composantes fréquentielles initiales  $(u, v)$  en intervalles  $I_{i,j}$ :

$$I_{i,j} = \left[ \frac{s(i-1)}{n}, \frac{s.i-1}{n} \right] \times \left[ \frac{s.(j-1)}{n}, \frac{s.j-1}{n} \right], \text{ pour } i, j = 1, \dots, \text{Int}(n/s)$$

Exemple de procédure de lissage: 
$$\text{smooth}S(X)(I_{i,j}) = \frac{\sum_{(u,v) \in I_{i,j}} S(X)(u,v)}{k}$$

$\text{Int}$  est la fonction partie entière (arrondi au plus petit entier en valeur absolue).

$S(X)$  est le spectre initial d'une variable  $X$ ,  $smoothS(X)$  est le spectre lissé associé.

Ainsi on ne connaît pas précisément la fréquence associée aux composantes lissées. **Il y a donc une incertitude sur la détermination des composantes fréquentielles.** Cela est étroitement lié au *principe d'incertitude d'Heisenberg*, qui impose une incertitude sur la connaissance simultanée de la position et de la vitesse des particules en mécanique quantique. Les fondements du principe sont statistiques et ne dépendent pas de la nature des objets : il est aussi bien valable en biologie qu'en mécanique quantique. Il ne faut enfin pas confondre ce principe avec les artéfacts purement calculatoires de l'analyse spectrale, comme le repliement de spectre (voir Encadré 9).

**La perte de précision va s'accompagner d'un gain de robustesse, et plus on perdra en précision, plus on gagnera en robustesse d'estimation.** Or nous avons plus besoin de robustesse que de précision dans nos estimations. En reconnaissant l'existence de ce compromis entre robustesse et précision, nous sommes en mesure d'effectuer un choix en accord avec nos objectifs d'inférence.

Le lissage suppose en définitive une rupture du principe de dualité : le passage d'un espace à l'autre peut se faire en théorie dans les deux sens, mais le lissage nécessaire n'est pas une opération inversible. On peut donc calculer un spectre à partir des positions, mais on ne peut pas retrouver l'information de position à partir d'un spectre lissé. Il est important de noter que l'hypothèse de stationnarité spatiale peut donc ne pas être pertinente pour décrire un processus dans l'espace réel, mais peut être appropriée pour décrire les structures résultantes dans l'espace spectral. **Le lissage permet en effet de négliger une partie des hétérogénéités spatiales, et de souligner les échelles importantes de structuration du signal : cela ne nécessite donc pas une représentation précise de l'ensemble des composantes fréquentielles.** Cette idée apparaît fondamentalement en adéquation avec l'idée d'échelle en écologie, en écho à une vision hiérarchisée des processus et des objets (Levin 1992).

Concernant l'étude de la dynamique des métapopulations, le lissage aura une implication essentielle : l'analyse spectrale sera en effet moins sensible aux caractéristiques contingentes des structures d'habitat et d'occupation. Nous démontrerons l'importance écologique de cette propriété dans la partie IV.3.

#### Encadré 9 : Ne pas confondre principe d'incertitude et artéfacts spectraux

L'utilisation de l'analyse spectrale doit être accompagnée de certaines mises en garde, afin de ne pas interpréter faussement des éléments artéfactuels au sein des signatures fréquentielles.

L'artéfact le plus important est le repliement du spectre (*aliasing*). Si l'on considère un signal échantillonné sur un plan régulier d'intervalle  $\Delta$ , il est intuitif que l'on ne pourra pas exprimer dans l'analyse spectrale des fréquences représentant des fluctuations plus courtes que  $\Delta$  (hautes fréquences). La plupart des méthodes d'analyse spectrale pourront fournir toutefois une estimation de ces composantes, qui seront dès lors artéfactuelles et n'auront pas de valeur explicative (voir Figure). On se retrouvera avec des composantes à basse fréquence qui seront des artéfact du signal réel à haute fréquence. On parle de repliement du spectre parce que ces composantes artéfactuelles induisent une symétrie au sein du spectre obtenu.

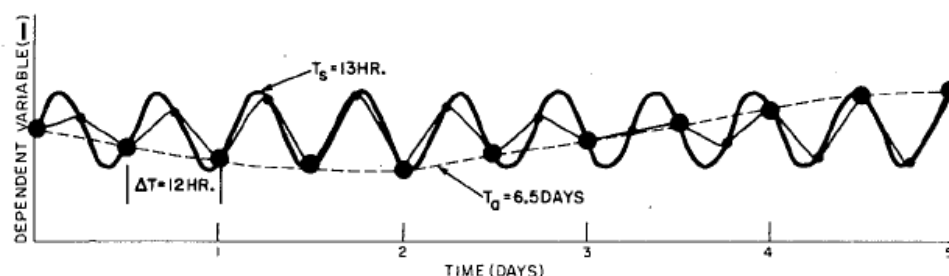


Figure : Illustration du phénomène de repliement du spectre (Platt and Denman 1975). La ligne continue grasse représente le signal à échantillonner, périodique de période  $T_s = 13$  heures. Les gros points représentent un échantillonnage discret toutes les douze heures  $\Delta T = 12 \text{ h}$ , et la ligne pointillée représente la périodicité apparente et artéfactuelle à 6.5 jours. Les petits points représentent un échantillonnage toutes les 6 h, de nature à exprimer la réelle périodicité de 13 h.

En résumé, on peut montrer que les fréquences inférieures à  $1/2\Delta$  dans le spectre du signal échantillonné seront artéfactuelles. Le *critère de Nyquist* consiste à tronquer le spectre obtenu pour supprimer les artéfacts, en ne conservant que les fréquences supérieures à  $1/2\Delta$ .

Il s'agit là d'un artéfact strictement lié à l'échantillonnage. C'est un problème de nature bien différente de la question de l'incertitude que nous abordons par ailleurs. L'incertitude est un problème statistique, liée au fait que la variance d'estimation des composantes spectrales d'un signal échantillonné ne tend pas vers zéro quand la densité d'échantillonnage augmente (Platt and Denman 1975). L'erreur standard d'estimation reste au contraire supérieure à 100 % de la valeur des composantes spectrales à estimer, et constitue donc une source de bruit majeure. La procédure de lissage permet de réduire cette erreur, au prix d'une perte de précision sur le découpage des classes de fréquences.

### 1.3. Analyse de Fourier et méthode PCNM spectrale

Dans le cadre de mon travail, j'utiliserai en parallèle deux méthodes d'analyse spectrale : l'analyse de Fourier, et l'analyse PCNM (*Principal Coordinates of Neighbour Matrices*) (Chapitre III). Ces deux méthodes ont été introduites dans des domaines bien distincts :

- L'analyse de Fourier décrit à la base un signal continu dans un espace vectoriel comme une combinaison linéaire complexe de composantes périodiques sinusoïdales (Figure 8a) (voir Platt and Denman 1975). La notation exponentielle complexe est un outil mathématique simple qui a rendu cette méthode très populaire pour la description de processus déterministes linéaires. Elle peut être utilisée également pour décrire des signaux dans des bases discrètes, ce qui est notre cas puisque nous travaillerons en temps et en espace discret. En tout état de cause, la méthode de Fourier est utilisable pour décrire des structures exprimées sur un échantillonnage discret. C'est aussi une méthode sensible aux effets de bord quand l'espace d'étude est borné. De tels biais sont susceptibles d'apparaître dans l'analyse des métapopulations simulées, ainsi que dans l'analyse des données floristiques de la Drome. C'est pourquoi nous nous sommes aussi intéressés à une autre méthode d'analyse spectrale.
- L'analyse PCNM (*Principal Coordinates of Neighbour Matrices*) est issue de la théorie spectrale des graphes (Chung 1994; Borcard et al. 2003), et a été introduite en écologie dans un contexte d'analyse spatiale plutôt que spectrale (Borcard and Legendre 2002). Elle repose sur la décomposition de la variance d'une variable exprimée sur les nœuds d'un graphe, dit graphe d'échantillonnage (voir aussi Méot et al. 1998). Le spectre est un ensemble de coefficients associés à des composantes spectrales échelonnées sur un gradient de corrélation spatiale (voir Figure 8b). Contrairement à l'analyse de Fourier, ces composantes ne sont pas fixes et dépendent de la structure du graphe d'échantillonnage. C'est pourquoi la méthode d'analyse PCNM peut indifféremment être utilisée avec des échantillonnages réguliers et irréguliers.

(a) Analyse spectrale de Fourier

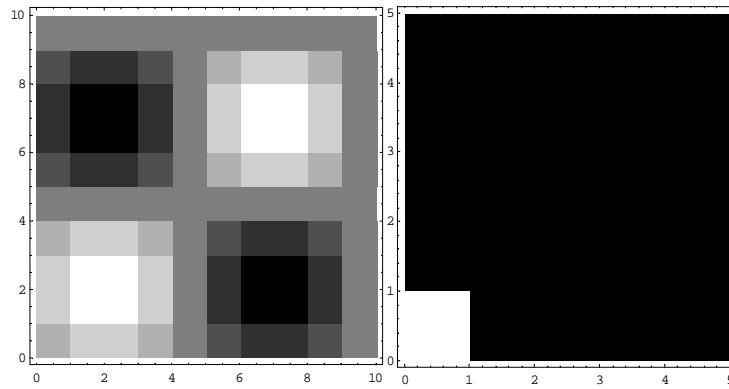
*Domaine réel*

*Domaine fréquentiel*

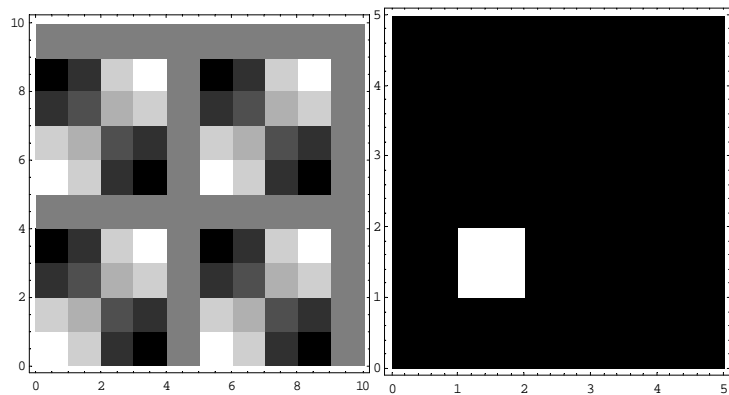
$$f(x,y)=\sin\left(\frac{2.\pi.x}{T}\right)\sin\left(\frac{2.\pi.y}{T}\right)$$

$$F(u,v)=\frac{N}{T}$$

$T = 10$



$T=5$



Coordonnées géographiques

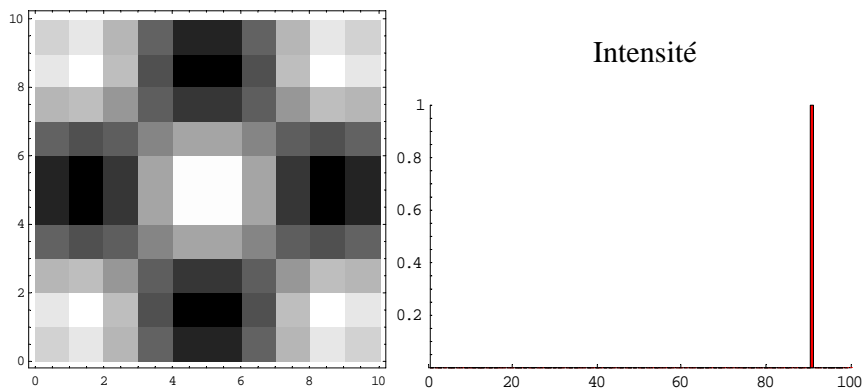
Fréquences

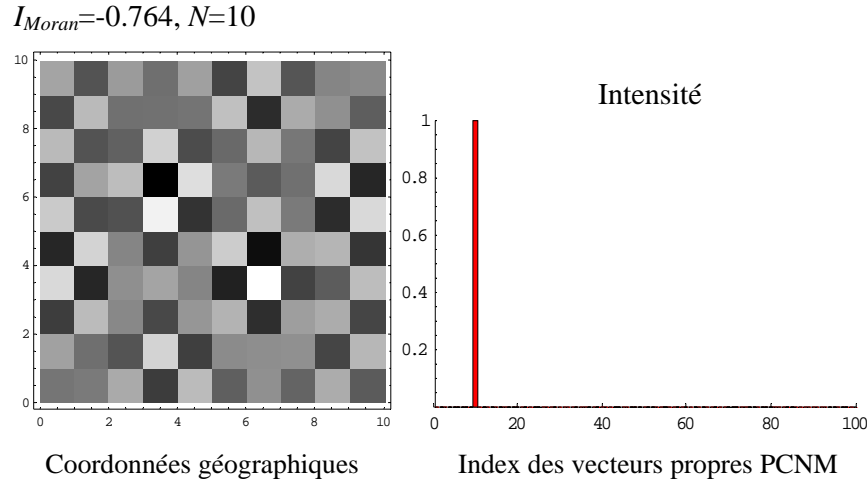
(b) Analyse spectrale PCNM

*Domaine réel*

*Domaine PCNM*

$I_{Moran}=0.73$





(c) Autocorrélation spatiale des vecteurs propres PCNM

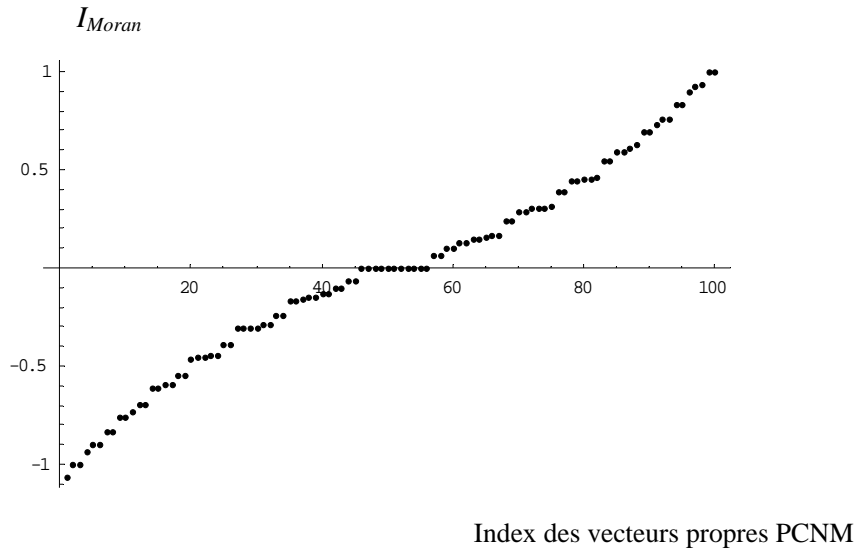


Figure 8 : Illustration des décompositions spectrales de Fourier et PCNM, pour des structures spatiales exprimées sur une grille de côté  $N = 10$ .

(a) Exemple d'analyse de Fourier pour une structure spatiale sinusoïdale pure : représentation de la structure spatiale à gauche (domaine réel), et de la structure fréquentielle associée à droite (domaine fréquentiel). Les valeurs sont représentées en niveaux de gris, du noir au blanc par ordre croissant. La fonction d'analyse de Fourier est la transformée discrète 2D proposée par *Mathematica* (fonction *Fourier*).

(b) Exemple d'analyse spectrale PCNM (sur la grille régulière de côté  $N$ ) pour deux structures spatiales, l'une fortement autocorrélée positivement, l'autre fortement autocorrélée négativement : représentation de la structure spatiale à gauche (domaine réel), et des coordonnées associées dans la base PCNM à droite (domaine PCNM). L'autocorrélation est mesurée par l'indice  $I$  de Moran, pour des corrélations au premier voisinage. Les structures représentées sont en fait deux vecteurs propres de l'analyse PCNM, ce qui explique la présence d'une seule composante non nulle dans leurs spectres. Ce sont des structures autocorrélées « pures », dans le sens où elles respectent l'hypothèse de stationnarité spatiale.



(c) Echelle des autocorrélations spatiales des vecteurs propres définissant la base PCNM ( $I$  de Moran), pour l'analyse montrée au (b). Voir aussi l'Encadré 10 pour de plus amples détails sur l'analyse PCNM.

Malgré la diversité des méthodes d'analyse spectrale que l'on peut trouver, les propriétés fondamentales (dualité, incertitude) sont par construction les mêmes. La méthode PCNM offre un formalisme intéressant et général, car il introduit explicitement la mesure de distance dans le calcul des vecteurs propres PCNM. On peut ainsi changer de base de décomposition spectrale en changeant de métrique. L'analyse spectrale d'un signal traduit alors sa structuration à différentes périodes ou fréquences pour une métrique donnée (Dray et al. 2004). C'est ainsi que la méthode PCNM peut être utilisée pour étudier la structuration de distances écologiques, par exemple entre compositions de communautés (Legendre and Anderson 1999; Legendre et al. 2005), ou encore phylogénétiques (Desdevises et al. 2003). Nous verrons cependant que les implications de la nature spectrale de la méthode PCNM sont méconnues ou mal exploitées, ce qui peut conduire facilement à des inférences erronées (voir Chap. IV §2.4).

Pour le calcul d'un spectre de Fourier, le lecteur pourra se référer à l'abondante littérature sur le sujet (voir notamment Platt and Denman 1975; Legendre and Legendre 1998). Pour le calcul de la base spectrale PCNM, on utilise une analyse en coordonnées principales de la matrice des distances (Legendre and Legendre 1998; Borcard and Legendre 2002). On cherche alors à représenter la signature d'un signal dans cette base. La méthode classique de (Borcard and Legendre 2002), que j'appellerai *méthode PCNM causale*, utilise les vecteurs propres de la base PCNM comme variables artificielles (« dummy variables ») et recherche par une analyse causale la relation entre le signal à analyser et ces variables.

Nous proposons une autre méthode de calcul du spectre PCNM d'un signal, plus conforme à la logique de l'analyse spectrale, car elle respecte en particulier le principe de dualité (Encadré 10). Nous l'appellerons donc *méthode PCNM spectrale*. Nous verrons en outre que cette méthode permet de respecter le principe d'incertitude, ce qui n'est pas le cas de la méthode PCNM causale. La comparaison des méthodes PCNM causales et spectrales sera discutée plus loin (Chap. IV § 2.4).

#### Encadré 10 : méthode d'analyse PCNM spectrale

La base PCNM est constituée des vecteurs propres  $V_i$  de l'analyse en coordonnées principales de la matrice des distances, et représente alors un gradient de corrélation spatiale (Borcard and Legendre 2002) (voir Figure 8c). Nous travaillons sur les vecteurs propres normalisés à 1, ce qui permet le calcul du spectre de puissance (décomposition de la variance). La conservation des distances dans l'espace PCNM est possible si les vecteurs propres sont normalisés à la racine carrée des vecteurs propres associés (pour démonstration, voir Legendre and Legendre 1998). Dans l'optique d'une décomposition de la variation spatiale à différentes échelles (Chung 1994; Méot et al. 1998), je travaillerai dans la suite de mon propos sur les spectres de puissance, donc avec des vecteurs propres normalisés à 1.

Les  $p$  vecteurs  $V_i$  considérés sont donc orthonormaux. A noter que le vecteur  $\mathbf{1}$  est vecteur propre associé à la valeur propre 0. Cela implique, sous la condition d'orthogonalité, que tous les autres vecteurs propres ont une moyenne nulle.

On notera  $M$  la matrice de transition, telle que :

$$M = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \begin{pmatrix} V_1 & \cdots & V_p \\ m_{11} & \cdots & m_{1p} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{np} \end{pmatrix}$$

Soit une variable  $x$  définie par un vecteur  $X$  dans l'espace des sites, et  $S_X$  dans l'espace spectral PCNM (triplet  $(x, X, S_X)$ ). En d'autres termes,  $S_X$  est le spectre PCNM de  $X$ . On peut écrire :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad S_X = \begin{pmatrix} S_{X1} \\ \vdots \\ S_{Xp} \end{pmatrix}$$

Ou encore

$$X = \sum_{i=1}^n X_i \cdot e_i = \sum_{j=1}^p S_{Xj} \cdot V_j = \sum_{j=1}^p S_{Xj} \cdot \sum_{i=1}^n m_{ij} e_i$$

Soit :

$$X_i = \sum_{j=1}^p m_{ij} \cdot S_{Xj} = (M \cdot S)_i$$

Une transformation matricielle permet donc de passer de l'espace réel à l'espace spectral

PCNM. L'analyse PCNM est donc fondamentalement une opération algébrique permettant de passer d'une base vectorielle où chaque point représente une dimension, à une base vectorielle où les vecteurs propres sont associés à des degrés de corrélation spatiale. A ce niveau, il y a adéquation stricte de l'information exprimée dans les deux espaces, car on peut passer indifféremment de l'un à l'autre par transformation matricielle. Ainsi la variance spatiale globale dans le domaine réel est exprimée par la somme des composantes du spectre de puissance.

**La terminologie utilisée pour décrire l'analyse spectrale PCNM nécessite de faire un distinguo important :**

- **Les vecteurs propres** de la base PCNM sont les *vecteurs*  $V_i$  issus de l'analyse en coordonnées principales de la matrice de distances. Ils sont ordonnés sur un gradient d'autocorrélation spatiale (Figure 8c), et ne dépendent que de la configuration spatiale de l'échantillonnage.
- **Les composantes** du spectre PCNM d'une variable  $X$  sont des *scalaires*  $S_{xi}$  qui expriment les coordonnées  $i$  de  $X$  dans la base PCNM des vecteurs  $V_i$  (Figure 8b). Les composantes dépendent de et expriment la structure spatiale de la variable  $X$ , connaissant sa valeur en chaque point de l'échantillonnage d'étude (espace réel).

#### 1.4. Découplage des dimensions spatiales et non-spatiales de la niche

La méthode d'analyse PCNM spectrale permet de respecter le principe de dualité (Encadré 10), grâce à une transformation algébrique matricielle inversible. L'analyse de Fourier respecte également ce principe : la transformation du domaine réel au domaine spectral se fait au moyen d'une transformation inversible appelée *transformée de Fourier*. Dans toutes mes analyses de Fourier, j'utiliserais une transformation discrète en deux dimensions, réalisée avec la fonction *Fourier* du logiciel *Mathematica*.

Insistons encore sur l'importance du principe d'incertitude en analyse spectrale (Chap. IV §1.2). Ce principe se traduit en pratique par la nécessité de lisser les spectres, c'est-à-dire de diminuer le nombre de composantes spectrales en calculant des moyennes sur des intervalles de fréquence ou de corrélation. Soulignons que différentes méthodes de lissage existent ; l'important est dans l'absolu de réduire le nombre de degrés de liberté des composantes spectrales, en regroupant les fréquences proches (voir Platt and Denman 1975).

Nous avons réalisé une opération de lissage simple et analogue sur les spectres de puissance de Fourier et PCNM. Nous présentons la procédure de lissage pour l'analyse de Fourier. Soit un entier  $s$  diviseur de  $n$ , nous partitionnons les fréquences initiales  $(p, q)$  en intervalles  $I_{i,j}$ :

$$I_{i,j} = \left[ \frac{s(i-1)}{n}, \frac{si-1}{n} \right] \times \left[ \frac{s(j-1)}{n}, \frac{s.j-1}{n} \right], \text{ pour } i, j = 1, \dots, \text{Int}(n/s)$$

$$(u, v) \in I_{i,j} \quad \text{si} \quad i = \text{Int}\left(\frac{u.n}{s} + 1\right) \text{ et } j = \text{Int}\left(\frac{v.n}{s} + 1\right)$$

$\text{Int}$  est la fonction partie entière (arrondi à l'entier inférieur en valeur absolue).

Le spectre de puissance de Fourier lissé est défini de la manière suivante :

$$\text{smooth}\mathfrak{Z}(\bar{X})(I_{i,j}) = \sqrt{\frac{\sum_{(p,q) \in I_{i,j}} |\mathfrak{Z}(\bar{X})(u, v)|^2}{k}}$$

A noter que la composante constante du spectre ( $u=0$  et  $v=0$ ), est retirée pour ne prendre en compte que les composantes fréquentielles réelles : la composante constante représente la moyenne spatiale du signal. La supprimer revient à effectuer un centrage préalable dans l'espace de la structure étudiée.

Pour les spectres PCNM, la démarche est semblable, et même plus simple puisque le spectre est exprimé dans un espace à une seule dimension.

Le manuscrit de l'article en Annexe 4 présente notre étude et nos résultats sur l'analyse des spectres de Fourier spatiaux et lissés des métapopulations simulées en habitat binaire, à leur état quasi-stationnaire (pour rappel sur les simulations, voir partie II). Nous montrons que l'analyse spectrale de Fourier constitue une mesure écologiquement pertinente de la structure spatiale des populations, car elle est très sensible à la variation des paramètres de dynamique métapopulationnelle,  $p$ ,  $q$  et  $r$ . La Figure 9 en est l'illustration essentielle. L'Analyse en Composantes Principales de l'ensemble des spectres de Fourier des métapopulations quasi-stationnaires est représentée. Les deux premiers axes de l'ACP expriment 90% de la variance totale des spectres (50% pour FPC1, 40% pour FPC2). Visuellement, le premier axe est fortement associé à la variation de  $r$  (Figure 9b), et le second axe à la variation de  $p$  (Figure 9a) et de  $q$ .

Cependant il est important de rappeler que les paramètres sont fortement corrélés entre eux, et de souligner que la structure particulière du nuage de point est de nature à fausser les corrélations statistiques. Nous avons donc considéré les corrélations entre les paramètres  $p$ ,  $q$ ,

$r$  d'une part, et les signatures spectrales FPC1 et FPC2 d'autre part. Les résultats sont représentés sur le Tableau 1. On constate que les paramètres  $p$ ,  $q$  et  $r$  sont tous significativement corrélés entre eux. En outre, la densité d'habitat  $p$  est corrélée significativement à FPC1 et à FPC2. En revanche, il apparaît bien que  $r$  est corrélé à FPC1 mais pas à FPC2, et que  $q$  est corrélé à FPC2 mais pas à FPC1. Le découplage spectral FPC1 vs. FPC2 exprime davantage une ségrégation entre la dynamique métapopulation, exprimée par  $r$ , et l'agrégation spatiale de l'habitat, exprimée par  $q$ .  $p$  n'apparaît pas être un indicateur pertinent pour exprimer ce découplage.

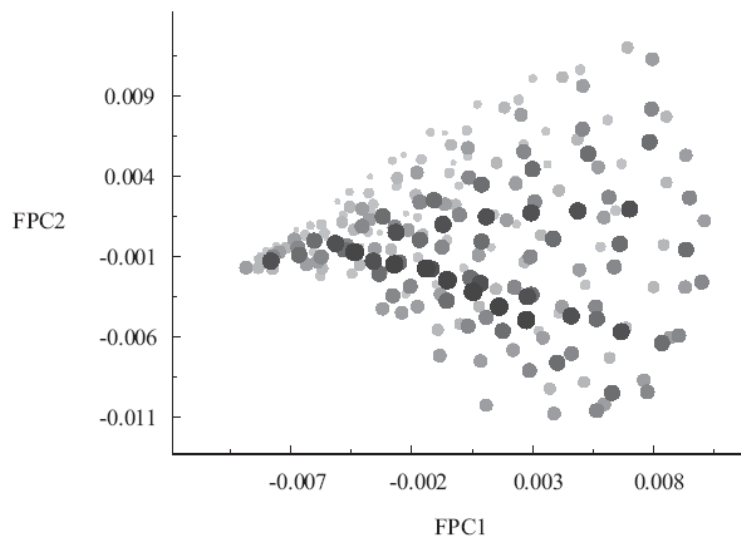
	FPC1	FPC2	p	q	r
FPC1	1	0	<b>0.2914***</b>	-0.0202	<b>-0.6985***</b>
FPC2		1	<b>0.2727***</b>	<b>-0.667***</b>	-0.069
P			1	<b>0.197**</b>	<b>0.2553***</b>
Q				1	<b>0.4320***</b>
R					1

Tableau 1 : Tableau des corrélations (coefficient de Pearson) entre les paramètres de dynamique métapopulationnelle  $p$ ,  $q$  et  $r$  d'une part, et les deux premiers axes principaux issus de l'analyse des spectres de puissance de Fourier des métapopulations simulées quasi-stationnaires. Le test associé est un test  $t$  paramétrique, l'hypothèse nulle étant celle de non-corrélation entre les vecteurs. Le rejet de l'hypothèse à 5% est représenté par une étoile \*, à 1% par deux étoiles \*\*, et à 0.1% par trois étoiles \*\*\*.

Ainsi les signatures spectrales sont très sensibles aux paramètres  $q$  et  $r$  (90% de variance totale sur les deux premiers axes, eux-mêmes liés fortement à ces paramètres), et surtout expriment un découplage entre deux dimensions : la dimension de la niche *a priori* (paramètre  $q$ ), exprimée par FPC2, et la dimension de la niche *a posteriori* (paramètre  $r$ ), exprimée par FPC1. Nous avons donc ici une mesure pertinente de la structure spatiale pour représenter l'effet de la dynamique métapopulationnelle, et un moyen de découpler les dimensions statiques et dynamiques de la niche, exprimées par les notions de niche *a priori* et *a posteriori* (cf. I.1.2). Il est intéressant de noter ici que l'effet de la structure de l'habitat s'exprime essentiellement sur une seule dimension. Bien que la question déborde du cadre de notre travail, cela suggère qu'il puisse y avoir une façon de représenter l'essentiel de la structure de l'habitat à travers un seul paramètre *pertinent*. Voir le manuscrit en Annexe 4 pour de plus amples développements.

(a)

Variation for parameter  $p$



(b)

Variation for parameter  $r$

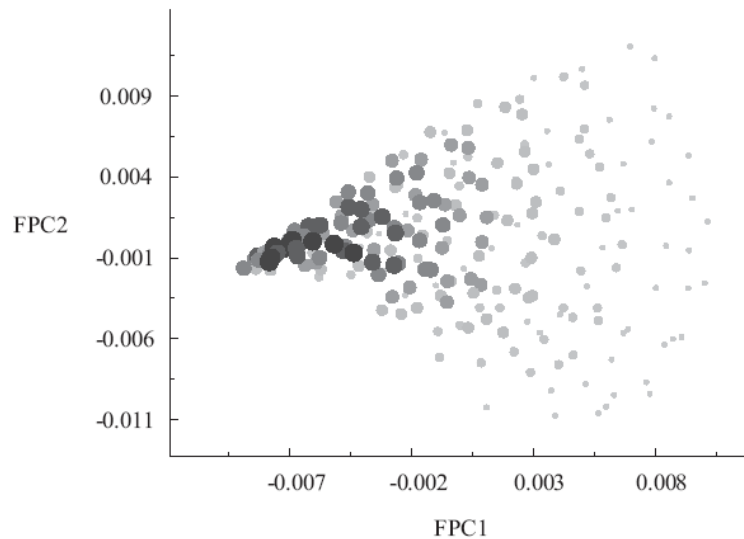


Figure 9 : Variation des paramètres  $p$  (Figure a) et  $r$  (Figure b), exprimée par la couleur et la taille des points (plus gros et plus sombre = plus élevé), dans l'espace des deux premiers axes FPC1 et FPC2 de l'Analyse en Composantes Principales de l'ensemble des spectres de Fourier des métapopulations simulées quasi-stationnaires (1 point = 1 simulation).

Des résultats très semblables sont obtenus en substituant à l'analyse de Fourier une analyse PCNM spectrale. Notre technique d'analyse spectrale PCNM apporte donc des

signatures analogues. Elle présente un atout supplémentaire : elle conserve en effet les mêmes propriétés que ce soit avec un échantillonnage régulier ou irrégulier de l'espace. Cela constitue un avantage pratique important, car il va être possible d'utiliser ce type de décomposition pour les données de la Drôme, disposées sur un échantillonnage irrégulier.

### 1.5. Saturation de l'habitat vs convergence structurale *a priori/a posteriori*

Nos résultats sont de nature à confirmer les conclusions sur l'analyse des seuils critiques d'extinction, pour les habitats *a priori* et *a posteriori* (voir Ch. II.2.1). Nous soulignons alors que la convergence de la structure de connectivité des deux types d'habitat était nettement plus rapide que la saturation de l'habitat *a priori* dans son ensemble. Le paramètre  $q$  est de nature à exprimer, au moins partiellement, la structure de connectivité de l'habitat, et est donc un bon index de la composante habitat *a priori* dans l'analyse spectrale. En revanche, le paramètre  $p$  est un index de densité de l'habitat *a priori*. Ce qui va avoir à la fois un effet sur le remplissage de cet habitat mais aussi sur la connectivité entre taches d'habitat favorables. Il va donc avoir un effet ambigu sur les deux premiers axes

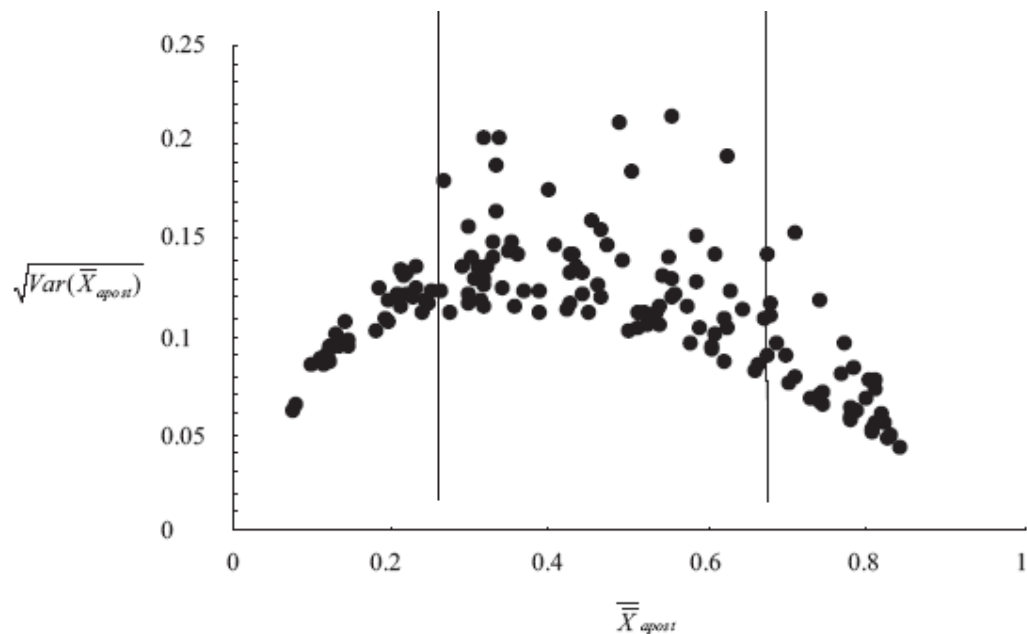
Nous pouvons produire en définitive un schéma simple :

- **La structure spatiale d'occupation moyenne au sein de l'espace occupé permet d'exprimer la dynamique métapopulationnelle.** Cela est résumé par la Figure 10. D'une part, la variation spatiale des probabilités d'occupation locales dans l'habitat *a posteriori*, notée  $Var(\bar{X}_{apost})$ , présente un patron caractéristique « courbe en cloche » en fonction de la moyenne d'occupation au sein de l'habitat *a posteriori* (Figure 10a). D'autre part, l'axe 1 de l'ACP sur les spectres de puissance de Fourier traduit la variation de  $\bar{\bar{X}}_{apost}$ , avec un centrage sur le sommet de la courbe en cloche de  $Var(\bar{X}_{apost})$  (Figure 10b). Enfin, l'occupation moyenne dans l'habitat *a posteriori* est reliée au paramètre de dynamique métapopulationnelle  $r$  (Figure 10c). L'axe 1 de l'ACP exprime donc la structure spatiale de l'occupation au sein de l'habitat *a posteriori*, et est relié fortement à la variation de  $r$ .
- **La connectivité générale du graphe des cellules occupées permet d'exprimer l'agrégation de l'habitat *a priori*.** C'est ce qu'exprime de fait l'axe 2 de l'ACP sur les spectres de puissance de Fourier. Cela résulte de la convergence structurale entre habitat *a priori* et *a posteriori*, quand l'habitat *a posteriori* est estimé avec un nombre suffisant de

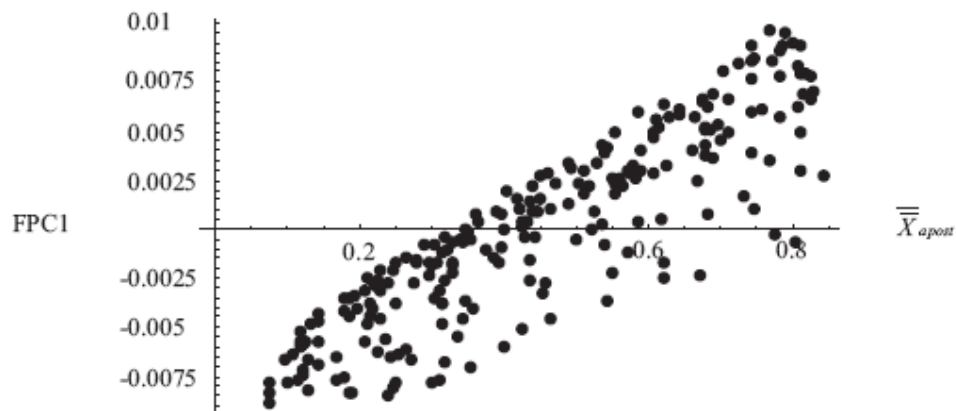
génération (Ch. II.2.1). Cette contrainte de convergence explique en outre que le découplage puisse être perdu pour un nombre trop faible de générations d'observations (résultat non montré).

Il s'agit d'un schéma cohérent avec nos analyses précédentes (Ch. II.2.1) ; ce qui est nouveau et remarquable est la possibilité de découpler les deux niveaux d'information.

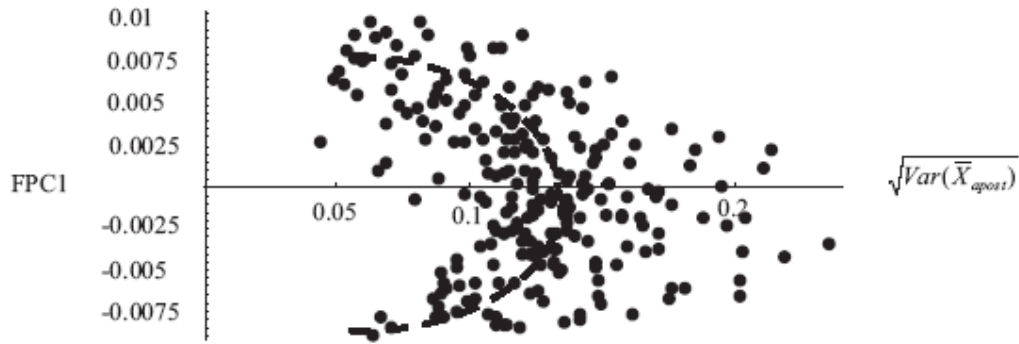
(a) Structure d'occupation de l'habitat *a posteriori*.



(b) Relation entre la structure d'occupation de l'habitat *a posteriori*, et le premier axe de l'ACP sur les spectres de puissance de Fourier.







(c) Relation entre la probabilité moyenne d'occupation locale dans l'habitat *a posteriori*, et le paramètre de dynamique métapopulationnelle  $r$ .

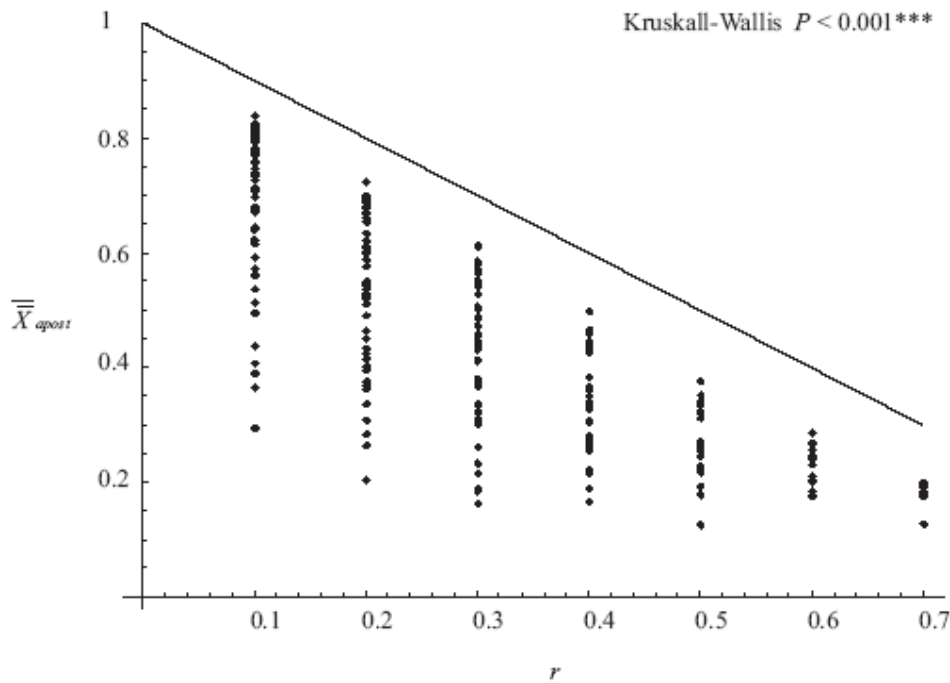


Figure 10 : La structure spatiale d'occupation de l'habitat *a posteriori* est caractérisée par la relation entre variance spatiale  $Var(\bar{X}_{apost})$  des probabilités locales d'occupation moyennées dans le temps  $\bar{X}_{apost}$ , et la moyenne  $\bar{\bar{X}}_{apost}$  de ces probabilités sur l'ensemble de l'habitat *a posteriori* (Fig. a). La relation entre le premier axe de l'ACP sur les spectres de puissance de Fourier et ces mesures montre que cet axe exprime la structure spatiale d'occupation au sein de l'habitat *a posteriori*. En outre la relation entre  $\bar{X}_{apost}$  et  $r$  montre le lien entre la structure d'occupation de l'habitat *a posteriori* et la variation du paramètre dynamique  $r$ . Cette dépendance induit une dépendance du premier axe d'ACP avec  $r$ .

### 1.6. Inférences des paramètres de l'habitat *a priori* et *a posteriori*

En introduction de ce Chapitre, nous avons proposé une démarche en deux temps : (i) rechercher une mesure spatiale pertinente, dans la mesure où elle permet de bien discriminer les structures exprimées par les métapopulations simulées, (ii) rechercher une procédure d'inférence qui permette de relier cette mesure pertinente à des paramètres écologiques pertinents bien choisis. Nous avons en particulier postulé que les paramètres  $p$ ,  $q$  et  $r$  utilisés pour la simulation des dynamiques métapopulationnelles pouvaient représenter des paramètres écologiquement pertinents.

Dans le paragraphe précédent, nous avons franchi avec succès la première étape, en montrant que l'analyse spectrale était porteuse d'information sur la structure des métapopulations, une information exprimée par deux degrés de liberté. Nous avons même fait un pas dans la deuxième étape, en montrant que les corrélations entre  $p$ ,  $q$ ,  $r$  d'une part, et les signatures représentant ces deux degrés de liberté d'autre part, permettaient d'éclairer la signification écologique de l'information exprimée. Un degré de liberté représente essentiellement le degré d'agrégation de l'habitat (paramètre  $q$ ), un autre degré de liberté exprime la dynamique métapopulationnelle intrinsèque (paramètre  $r$ ).

Ces éléments suggèrent que l'analyse spectrale constitue une bonne base pour l'inférence séparée des caractéristiques d'habitat *a priori* et des caractéristiques de dynamique métapopulationnelle. Il existe cependant un facteur confondant, la densité d'habitat  $p$ , liée significativement aux deux signatures spectrales présentées, et également corrélée fortement aux paramètres  $q$  et  $r$ . L'absence d'indépendance claire entre les paramètres putativement pertinents  $p$ ,  $q$  et  $r$  est une entrave évidente à l'inférence, car elle est de nature à créer des multiplicités d'inférences possibles, pour une même structure spatiale de métapopulation. Nos tentatives dans ce sens ont effectivement été confrontées à ce problème.

En outre, nos tentatives ont été basées sur des espérances d'occupation locale estimées sur beaucoup de générations à l'état quasi-stationnaire, c'est-à-dire après une certaine homogénéisation spatiale. Si on considérait plutôt des probabilités estimées sur peu de générations, avec une structure spatiale plus sensible aux effets d'auto-organisation, l'information spectrale pourrait être plus riche et permettre de découpler plus nettement les effets habitat *a priori* / *a posteriori* ; il y aurait une meilleure opportunité d'inférence de  $r$ .

En conclusion, nous n'aboutirons pas dans le cadre de mon travail à la formulation d'une méthode d'inférence à partir des spectres, qui reste à bien des égards prématurée. Nous

aurons néanmoins par la suite d'autres opportunités de mettre en valeur la richesse de l'information spectrale. Nous explorerons par ailleurs dans le Chapitre suivant une autre voie, à partir d'une modélisation spatiale analytique, qui permettra d'aboutir avec succès à des solutions d'inférence.

## 2. Etude des effets de structuration spatiale par partition de variation

Une autre manière d'étudier la signification écologique des structures spatiales repose sur le principe de partition de la variation, mis en avant par Borcard et al. (1992). Il s'agit de découper la variation d'une variable d'effet suivant différents groupes de variables causales, dont certaines peuvent exprimer des aspects de structuration spatiale. Cette démarche est fondamentalement de nature inférentielle : la quantification de la déviance expliquée par différents groupes de variables permet d'inférer les rôles respectifs de ces variables.

Cette méthode a donc été promue pour évaluer l'incidence de la structuration spatiale sur les relations causales que les écologistes cherchent à mettre en évidence (cf. Arenas and Cuadras 2002). Il s'agit de traiter alors l'espace comme une variable comme une autre.

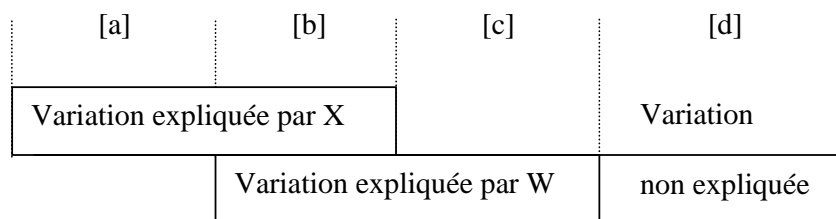


Figure 11 : Partition de la variation de la variable de réponse entre deux jeux de variables explicatives X (environnement seul) et W (structure spatiale).

### 2.1. L'espace n'est pas une variable comme les autres

Jusqu'ici, nous avons souligné que la structuration spatiale des populations est la résultante de processus dynamiques, relevant à la fois de la structuration de l'habitat et de la capacité intrinsèque de l'espèce à occuper cet habitat. Dans ces conditions, la structure spatiale des distributions observées est une variable d'effet. Toutefois la présence de corrélation spatiale a avant tout été perçue historiquement comme une source de biais

statistique dans les méthodes d'analyse causale (régressions linéaires et modèles dérivés) (cf. Keitt et al. 2002). Elle introduit en effet une rupture de l'hypothèse d'indépendance des données mesurées sur les points d'échantillonnage. Les problèmes engendrés sont de différents ordres :

- Le nombre effectif de points indépendants doit être pris en compte dans le calcul des degrés de liberté, sous peine de surestimer l'information disponible. C'est très classique dans les analyses de niche, où les phénomènes d'agrégation doivent être pris en compte.
- La non-normalité des variables de la régression, et notamment des résidus ne permet pas de réaliser correctement les tests paramétriques de significativité des variables.

Pour comprendre la signification de la structuration spatiale des variables considérées, il faut prendre en compte plusieurs schémas de causalité possibles. Pour cela nous faisons ici référence au paradigme de la niche.

- *Remplissage de niche.* La variable d'effet peut répondre à une variable causale donnée, et sa structuration spatiale être la conséquence directe de cette relation. Dans ce cas, il y a une forte relation entre la structuration spatiale de la variable d'effet et de la variable causale. Nous obtiendrons ce type de situations pour des valeurs de  $r$  faible, c'est à dire pour un faible taux relatif d'extinction et donc un fort taux d'occupation de l'habitat.
- *Composantes cachées de la niche.* Dans le même ordre d'idée, la variable d'effet peut répondre fortement à une variable causale non prise en compte dans l'analyse.
- *Dynamique spatiale.* La relation causale statique ne prend pas en compte les phénomènes dynamiques affectant la distribution des populations. Ces phénomènes peuvent induire des structures particulières, qui s'ajoutent et se combinent aux effets de structuration de l'habitat. Ils introduisent donc des degrés de liberté supplémentaires dans la structuration spatiale des populations. La dynamique spatiale des populations aura un effet d'autant plus fort que l'habitat sera peu rempli, c'est à dire pour des valeurs de  $r$  fortes (taux d'extinction relatif élevé).

Dans les deux derniers cas, une indépendance de principe existe entre la structuration spatiale des variables causales et celle de la variable d'effet. Cela dépend toutefois aussi de la nature et de la qualité de l'information spatiale prise en compte, d'où l'importance d'une mesure pertinente de la structure spatiale.

## 2.2. Cause ou effet ? Notion de variable spatiale

Ces considérations théoriques revêtent une importance pratique considérable. Les effets délétères sur l'inférence des dépendances entre variables sont considérablement amplifiés si l'on n'est pas en mesure de donner un sens à leur structure spatiale. S'inscrivant dans le cadre du paradigme de l'échelle, Borcard et Legendre (2002) ont proposé une quantification des structurations spatiales à différentes échelles par la méthode PCNM. Le fait de découper l'information de structuration spatiale à différentes échelles permettrait selon eux de dissocier l'effet de phénomènes intervenant à différentes échelles.

D'autres méthodes existent qui ne prennent pas en compte une telle hiérarchisation des échelles spatiales : elles dissocient la caractérisation des échelles larges (trend surface analysis) de celle des échelles fines (autorégression) (Lichstein et al. 2002). Dans tous les cas, elles répondent au même paradigme de l'échelle, à savoir le présupposé qu'un phénomène opérant à une certaine échelle tend à avoir une signature spatiale à cette même échelle (Legendre and Borcard 2003).

Ces méthodes introduisent en général de fausses variables causales (*dummy variables*) de nature spatiale, que ce soit des polynômes de coordonnées spatiales (trend surface analysis), des composantes spectrales (PCNM), ou des matrices d'autorégression. Au regard des effets confondants de la structuration spatiale évoqués précédemment, cette démarche n'est pas toujours pertinente et ne doit donc pas être systématique :

- Dans le cas l'habitat *a priori* serait rempli, que les composantes soient cachées ou non, les variables spatiales représenteraient une majeure partie de l'information contenue dans les variables causales. Dans ce cas un modèle causal non spatialisé permet de prendre en compte correctement l'effet des variables statiques d'habitat (niche réalisée sensu Hutchinson 1957). Créer des variables spatiales *causales* serait en revanche une source d'erreur. De fait la structure de corrélation des variables causales d'habitat n'est elle-même pas une variable causale. Introduire cette structure comme variable causale n'a alors pas de sens écologique, et induit la prise en compte d'un effet parasite « carte d'habitat » qui nous détourne de la compréhension des processus de dynamique des populations.
- Sous l'hypothèse que l'habitat *a priori* n'est pas rempli, on prendra en revanche en compte l'existence de processus dynamiques (notamment colonisation et extinction de populations), en considérant généralement un modèle d'autorégression. Dans ce cas la

structure spatiale est un effet de mécanismes écologiques à l'intérieur du système, et l'analyse causale n'apparaît pas être pertinente pour la prendre en compte. Toutefois une autorégression permet de « court-circuiter » ce problème en introduisant la variable d'effet au sein des variables causales, et de décrire la relation entre les sites voisins par exemple.

### *2.3. Trouver un sens écologique : approche informationnelle*

Nous avons vu que le problème est de mesurer correctement l'espace et de savoir démêler l'effet des variables causales d'autres processus dynamiques. Toutefois, pour une mesure spatiale donnée, considérée comme un référentiel du système, on peut évaluer la quantité d'information spatiale associée à différents paramètres du système. La méthode de partition de la variance est en définitive de cette nature. Un référentiel spatial est introduit au travers d'un groupe de variables spatiales, que ce soit des variables PCNM ou autre chose (par exemple des polynômes de coordonnées, Borcard and Legendre 1994). La partition permet en outre d'évaluer les redondances d'information entre différents groupes de variables. On ne cherche pas à décrire complètement l'effet des mécanismes sous-jacents, mais simplement de donner un sens à une information parcellaire sur les variables en jeu.

C'est très différent d'une démarche analytique ascendante, où une hypothèse précise donne lieu à des prédictions précises. Celles-ci sont évaluées dans des conditions généralement impropres à leur évaluation, du fait de l'inadéquation des nombreuses hypothèses restrictives inévitablement formulées. Il faut alors construire des hypothèses alternatives et les tester séquentiellement. Cette approche conduira en général à une avalanche d'hypothèses alternatives, du fait du grand nombre de simplifications nécessairement formulées sur les systèmes écologiques. De plus la validation d'une hypothèse ne signifiera pas qu'elle est formellement juste, suivant le principe de falsifiabilité.

Dans l'approche informationnelle en revanche, ce n'est plus le modèle mais les données mesurées par l'expérimentateur qui sont au cœur de la démarche (cf. Levin et al. 1997). Il s'agit alors d'évaluer la quantité d'information que plusieurs modèles indépendants ou non peuvent exprimer. Il n'y a donc pas de test séquentiel d'hypothèses alternatives. La gamme de quantité d'information expliquée sur l'ensemble des modèles représente l'information disponible dans les données pour la question à laquelle notre gamme de modèles cherche à répondre. Si cette gamme est suffisamment large et que un ou plusieurs modèles sont de qualité nettement supérieure, on pourra conclure à sa ou à leur pertinence (Johnson and

Omland 2004). Il ne s'agit pas d'un test formel d'hypothèse, mais de la construction *ad hoc* d'un modèle descripteur pertinent. Selon la nature des modèles utilisés, on pourra orienter la signification écologique d'une telle description. La réponse ne sera finalement pas de dire qu'un modèle est vrai ou faux, mais de dire qu'un modèle est informatif ou non. C'est donner un sens écologique à une information écologique parcellaire et non accessible directement à notre compréhension.

#### 2.4. Variations sur le thème de la partition

La partition peut être réalisée sur la base de variables de structure spatiale qui ne représentent pas une distance géographique. Il est possible de réaliser une partition de variation d'une variable écologique au sein de communautés en fonction de la similarité de ces communautés (voir Legendre et al. 2005), ou encore de réaliser une partition d'un trait morphologique ou écologique en fonction de la distance phylogénétique des espèces considérées (voir Desdevises et al. 2003).

On peut bien sûr réaliser des partitions de variation sans nécessairement utiliser de variables virtuelles (*dummy variables*), issues de la décomposition spectrale d'une matrice de distances. On peut ainsi simplement décomposer la variation d'une variable d'effet suivant plusieurs catégories de variables causales écologiques (Heikkinen et al. 2005).

Nous verrons dans la partie suivante que le principe causal de la partition, tel qu'il est proposé par Borcard et al. (1992), pose des écueils semblant incontournables, tant en terme de validité statistique que de pertinence écologique. Toutefois, nous montrerons que l'analyse spectrale, en offrant un moyen de découper l'information écologique suivant différentes échelles spatiales, peut offrir des solutions plus rigoureuses et plus robustes pour comparer la structure spatiale de différentes variables.

### 3. Dualité vs. causalité : étude critique de la méthode PCNM

En résumé, les approches développées en écologie spatiale depuis quelques années sont très ambiguës car elles mélangent souvent inférence et test d'hypothèse. Cette confusion est dérangeante, car elle traduit un certain désarroi des chercheurs face à la complexité des données qu'ils doivent traiter. On se réfugie alors derrière un certain paradigme de l'écologie spatiale, qui s'articule autour de deux axes contradictoires :

- L'utilisation d'une modélisation causale de la structure spatiale des données, à grand renfort de modèles linéaires ou de modèles multivariés canoniques. La vertu prédictive d'une telle approche a encouragé son utilisation massive en Biologie de la Conservation (ex : Araújo and Williams 2000; Thuiller 2003). Elle ne permet pas cependant une prise en compte explicite des processus écologiques (Austin 2002). L'incidence de l'espace est toujours considérée au niveau des variables causales, que ce soit en ajoutant de fausses variables spatiales (y compris auto-régressives), ou en réalisant des régressions partielles destinées à supprimer la corrélation spatiale des données (Keitt et al. 2002). Nous avons vu qu'une telle démarche inspirée de l'approche hypothético-déductive constitue une vision biaisée de la signification écologique de l'espace.
- Il faut compenser l'impossibilité de tester des hypothèses précises sur des données réelles incompatibles avec des modèles restrictifs. On utilise donc en général des modèles très souples, vidés en fait de sens écologique, et on sélectionne des paramètres jusqu'à obtenir une représentation « parcimonieuse » des données réelles. Un exemple extrême est l'utilisation des GAM (*Generalized Additive Models*) et des polynômes de variables causales, où l'on privilégie la qualité d'ajustement à la pertinence du modèle (Thuiller et al. 2003). On est donc dans une démarche inférentielle, bien que la modélisation causale soit à l'origine hypothético-déductive (voir la nature des hypothèses des tests paramétriques). Par ailleurs les mesures spatiales utilisées sont rarement discutées pour leur pertinence écologique (Dungan et al. 2002).

Afin de discuter les problèmes que peut poser un tel paradigme, nous proposons de confronter la méthode PCNM causale d'analyse spatiale proposée par Borcard et Legendre (2002) à notre méthode d'analyse spatiale PCNM de nature spectrale (voir Chap. III §1.3, et notamment Encadré 10).

### 3.1. Méthode PCNM causale

Comme jeu de données de référence, nous utilisons les métapopulations simulées en habitat structuré, déjà présentées dans la première partie. Le plan d'échantillonnage (2500 points sur un maillage carré régulier) et les composantes de l'analyse PCNM ont été également présentés. Nous utilisons à présent ces composantes au sein d'une analyse causale conforme à la méthodologie de Borcard et Legendre (2002). Cette approche est largement utilisée actuellement en écologie (ex : Felizola Diniz-Filho and Bini 2005).



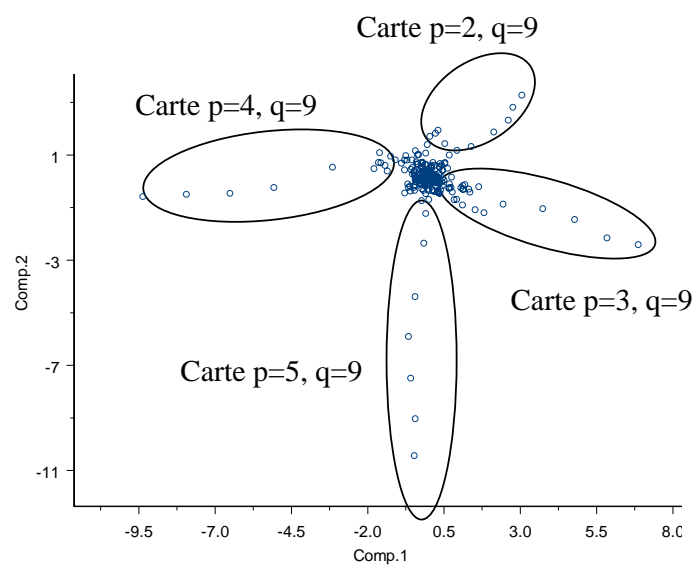
L'analyse causale PCNM est très différente dans son esprit de l'approche PCNM spectrale que nous avons déjà étudiée. Elle comporte plusieurs aspects importants :

- La variable d'effet Y est la probabilité d'occupation des points d'échantillonnage, sur 300 générations. Les composantes PCNM sont les variables causales W, et l'analyse de la relation entre Y et les variables causales W se fait grâce à un modèle linéaire généralisé gaussien (avec le logiciel SPlus).
- Par définition, il y a autant de variables PCNM que de points d'échantillonnages. (Borcard and Legendre 2002) recommandent de n'utiliser que les composantes PCNM associées à des valeurs propres positives, sans étayer cette recommandation par un argumentaire théorique. C'est donc ce que nous avons fait, bien que cette position apparaisse discutable.
- Avant de réaliser cette régression, on doit opérer une sélection préalable des variables PCNM. Même en ne sélectionnant que les composantes associées à des valeurs propres positives, nous conservons 1753 composantes sur 2500 à l'origine. Il est bien évidemment impossible de réaliser un GLM (*General Linear Model*) avec 1753 variables et 2500 points. Mais nous pouvons tirer profit de l'indépendance des composantes PCNM pour faire une étude des corrélations entre Y et les composantes PCNM. On peut classer les composantes PCNM par ordre décroissant de corrélation avec la variable d'effet Y. On décide alors de sélectionner les meilleures composantes PCNM, jusqu'à atteindre un seuil de variance expliquée correspondant au tiers de la variance totale de Y.
- Le nombre résultant de composantes PCNM sélectionnées se situe entre 6 et 129, donc dans une gamme de valeurs acceptable pour réaliser le GLM. Celui-ci est donc réalisé pour les 227 simulations présentant une occupation non nulle à l'état quasi-stationnaire. Les variables PCNM sont sélectionnées une dernière fois au cours de cette analyse suivant des critères d'AIC (*Akaike Information Criterion*) (fonction *step* de SPlus). Ce sont les modèles les plus parcimonieux qui sont donc conservés.

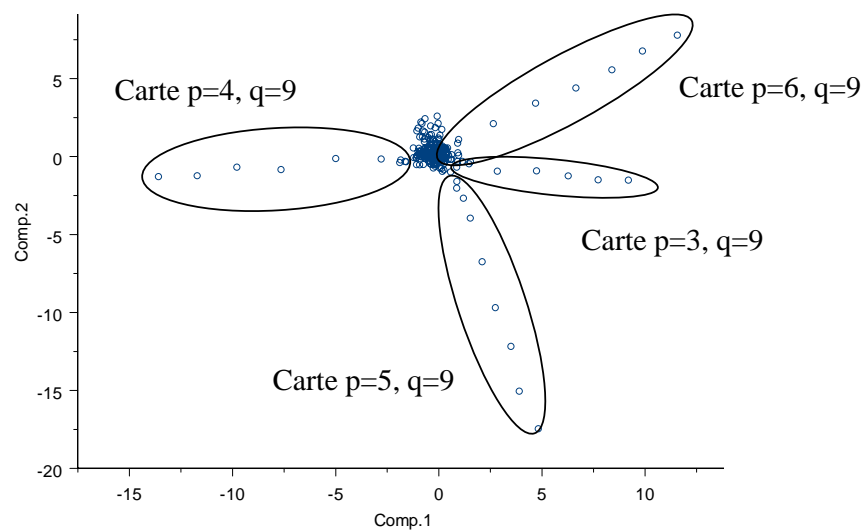
### 3.2. Comparaison avec la méthode PCNM spectrale : rôle capital du lissage

Les « spectres » issus de cette analyse causale sont comparés à ceux obtenus par le biais de l'analyse PCNM spectrale (voir Encadré 10). Pour chacune des deux méthodes, nous analysons l'information contenue dans les spectres au moyen d'une Analyse en Composantes Principales sur les 227 spectres de métapopulations simulées.

(a)



(b)



(c)

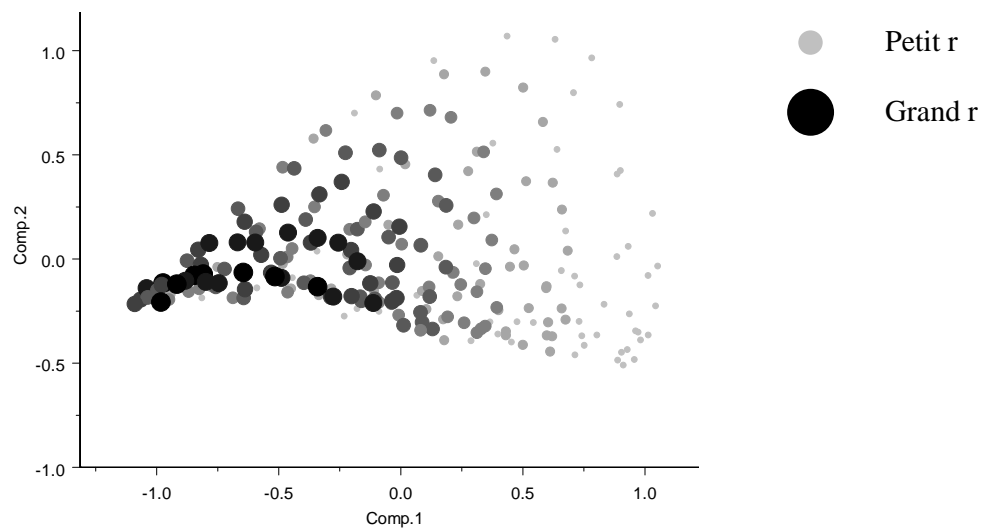


Figure 12 : Représentation des points-simulations dans l'espace défini par les deux premiers axes de l'ACP réalisée sur les spectres des 227 simulations : (a) spectres obtenus par la méthode PCNM causale ; (b) spectres bruts non lissés obtenus par la méthode PCNM spectrale ; (c) spectres lissés obtenus par la méthode PCNM spectrale. Pour les figures (a) et (b), les axes de l'ACP reflètent les effets des différentes cartes d'habitat utilisées pour les simulations (couples  $p, q$ ). Pour la figure (c), le premier axe de l'ACP représente la variation du paramètre de dynamique métapopulationnelle  $r$ , et le deuxième l'effet de  $q$ . Il n'y a pas d'effet carte pour la figure (c).

L'ACP sur les 227 spectres de l'analyse causale présentent 36 composantes qui expliquent chacune entre 1 et 5.1% de la variance. Les nuages des points dans les différents systèmes d'axes d'ACP montrent que les axes sont fortement associés à des séquences de points correspondant à une même carte d'habitat (le nuage des points sur les deux premiers axes est illustré, Figure 12a). En guise de point de repère, rappelons que nous avons utilisé pour nos simulations 58 cartes différentes d'habitat.

En revanche l'ACP sur les spectres de l'analyse PCNM spectrale présente 94% de variation expliquée par les deux premiers axes, et surtout pas d'effet carte (Figure 12c): la variation sur le premier axe est sensible à la dynamique métapopulationnelle (paramètre  $r$ ), la variation du deuxième axe est sensible à la variation de l'agrégation d'habitat (paramètre  $q$ ). Cela vient conforter les commentaires proposés au Chap. III §1.2 sur le principe d'incertitude. L'opération de lissage fait fondamentalement différer la méthode PCNM spectrale de la méthode PCNM causale. Nous avons souligné que cette opération permet de mettre en exergue les *échelles pertinentes* de structuration du signal, en gommant notamment l'effet d'hétérogénéités locales contingentes et sans réelle signification écologique.

La comparaison des deux méthodes traduit donc une différence entre les spectres quant à la nature de l'information fournie:

- L'analyse spectrale PCNM (resp. l'analyse spectrale de Fourier) s'abstrait des effets de cartes en rendant essentiellement compte de caractéristiques globales de corrélation spatiale (resp. de périodicité spatiale). Celles-ci traduisent la variation de structuration de l'habitat et de dynamique métapopulationnelle dans le cadre de nos simulations. Le fait que l'on puisse découpler ces deux composantes de variation spectrale montre qu'il n'y a pas d'effet carte. Notre usage de la méthode spectrale a été motivé initialement par le souhait de se débarrasser des effets contingents de cartes, pour souligner les structures écologiquement pertinentes (voir introduction du Chap. III, §1). Cet objectif est donc présentement atteint

- L'utilisation des composantes PCNM dans une analyse causale met en exergue des relations contingentes entre les variables spatiales PCNM et la structuration de l'habitat : l'effet carte est très fort, et ce sont les simulations ayant des cartes identiques qui déterminent les axes d'ACP. L'absence de lissage souligne les variations de fréquences ou de corrélations liées aux hétérogénéités spatiales au sein d'une carte, et en particulier liées aux aléas de sa construction. Nous obtenons donc, conformément au principe d'incertitude, une bonne estimation sur les positions et une mauvaise information sur la périodicité.

L'utilisation des composantes PCNM selon la méthode de Borcard et Legendre (2002) fournit des informations contingentes, très dépendantes du jeu de données utilisé. Elle ne permet pas de rendre compte de caractéristiques spatiales pertinentes pour la compréhension des processus dynamiques dans l'espace. Une analogie permet d'illustrer simplement cette idée (Kauffman 1993, p.13). Considérons deux collines identiques. Sous l'effet des précipitations, chaque colline développe un réseau de ruissellement et de ravinement. Ces réseaux sont propres à chacune des collines, en raison des contingences particulières de structure des roches, de direction des vents... La structure résultante est donc le fruit de l'histoire particulière de chaque colline, mais cependant la structure macroscopique des écoulements, avec des bifurcations et des convergences, présente des caractéristiques communes, que les hydrologues formalisent sous forme de lois. On peut supposer qu'une analyse de la structure spatiale des ruissellements par la méthode PCNM causale ferait ressortir des caractéristiques contingentes, locales et particulières à chaque colline, tandis que la méthode PCNM spectrale mettrait en exergue des aspects plus généraux, liés à la structure globale du système.

*A fortiori* quand une composante PCNM ressort significativement par la méthode causale, affirmer qu'il existe un processus opérant à l'échelle spatiale associée à cette composante risque fort d'être faux. Cette composante risque fort de refléter des structures contingentes. La relation « process-pattern » ne peut donc être comprise qu'au moyen de la version spectrale lissée du traitement des composantes PCNM.

Mais pourquoi les résultats obtenus suivant les deux méthodes sont-ils aussi radicalement différents ? La réponse réside dans un élément capital de l'analyse spectrale : le lissage. En réalité, les résultats de l'analyse spectrale PCNM seraient analogues à ceux de l'analyse

causale PCNM si on ne faisait pas le lissage (Figure 12*b*): l'information fournie serait contingente et fortement liée à l'effet carte.

Nous avons vu que le lissage permet d'améliorer l'estimation des composantes spectrales : sans lissage, il y a une variance d'erreur de l'ordre de 100% sur l'estimation de ces composantes. Donc, les composantes brutes avant lissage présentent une variance d'erreur de 100%, ce qui gomme l'information spectrale. Les similarités entre spectres ne dépendent alors que de similarités contingentes entre paysages, aboutissant à un résultat similaire de l'algorithme d'ajustement : c'est le cas quand on traite des métapopulations correspondant à un même paysage d'habitat. Le fait de traiter un grand nombre de variables causales accroît considérablement le risque de sélectionner des variables causales ayant une relation « par hasard » avec la variable d'effet.

En définitive, le problème crucial est qu'il n'est pas possible de réaliser un lissage avec la méthode causale, car il doit être postérieur au calcul du spectre brut. Dans l'analyse causale, on doit nécessairement sélectionner au préalable les composantes PCNM pour réduire le nombre de variables contenues dans le GLM. Cela implique que la partition de la variation des variables d'effet suivant différents groupes de variables causales ne peut pas se faire par la méthode causale jusqu'ici proposée par Borcard et Legendre (2002), mais doit se faire au moyen d'une véritable approche spectrale.

### *3.3. Retour sur la spatialisation de la niche : méthode de partition spectrale*

Nous avons montré que la méthode PCNM causale est sensible aux hétérogénéités contingentes de la structure spatiale, pour une carte donnée, et ne fournit pas une caractérisation globale et écologiquement pertinente. Nous avons vu cependant que le principe de partition de la variation suivant deux groupes de variables environnementales et spatiales PCNM offre un outil intéressant pour l'étude de la composante spatiale de la niche. Nous proposons donc une nouvelle méthodologie de partition de la variation, prenant en compte la nécessité d'une méthode spectrale de calcul des composantes PCNM.

Il s'agit de distinguer deux niveaux de compréhension de la relation causale entre données environnementales et variables d'effet (présence-absence, diversité) :

- La variable d'effet *Y* et les variables environnementales *X* peuvent être décrites aussi bien dans l'espace réel que dans l'espace spectral (principe de dualité). Le spectre de *Y* est

noté  $S_Y$ , celui de  $X$   $S_X$ . Malgré la dualité mathématique de ces espaces, nous avons vu que l'analyse spectrale de données réelles nécessite de recourir au lissage et de perdre cette réciprocité (principe d'incertitude). Dès lors la description des variables dans les deux espaces ne traduit pas les mêmes propriétés. Dans l'espace spectral, on met l'accent sur les structures périodiques régulières, indépendamment de l'information de position (principe de stationnarité spatiale). Cela explique que l'analyse spectrale permette de s'abstraire des contingences de position (paragraphe précédent).

- L'analyse causale de  $Y$  sur  $X$  n'est donc pas équivalente à celle de  $S_Y$  sur  $S_X$ . La relation entre  $Y$  sur  $X$  tirera au maximum profit de l'information disponible dans les données, mais surestimera du même coup cette quantité d'information disponible en négligeant les relations spatiales entre les mesures ponctuelles : c'est une estimation *a maxima* de la niche écologique. En revanche la recherche d'une relation entre  $S_Y$  et  $S_X$  ne conservera de l'information mesurée aux points d'échantillonnage qu'une composante de structuration spatiale, en gommant les contingences de position : c'est une estimation purement spatiale de la niche. Nous avons montré que cette estimation spectrale de la niche permettait de faire ressortir une information écologique pertinente pour la compréhension des mécanismes dynamiques dans l'espace.

Comme il ne faut pas présenter sur un même plan les données spectrales et réelles, nous proposons une nouvelle conception de partition de la variance spatiale, basée exclusivement sur le concept de spectre PCNM.

Cette démarche s'appuie sur les propriétés suivantes:

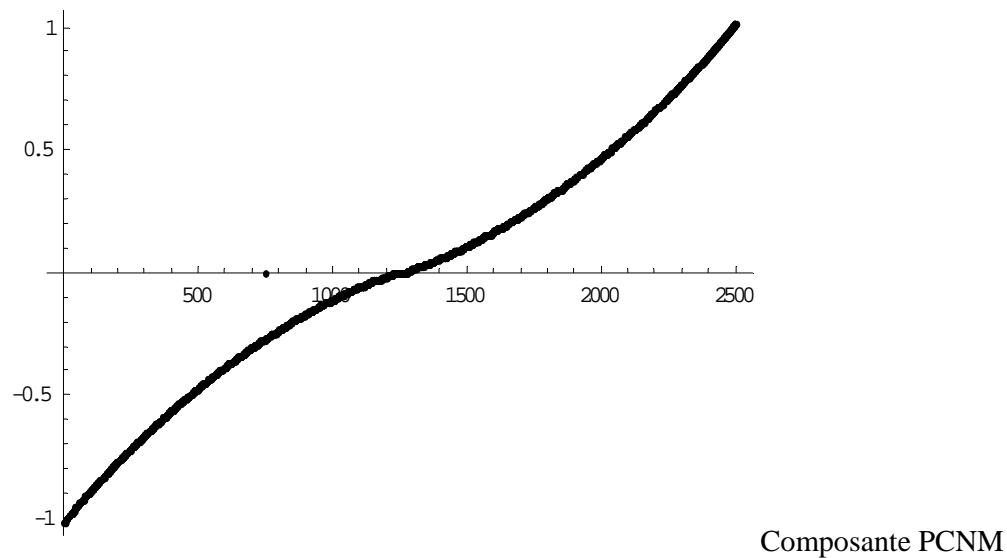
- Les vecteurs constituant la base brute de décomposition spectrale PCNM ( $n$  vecteurs pour  $n$  points d'échantillonnage) peuvent être hiérarchisés suivant un gradient de corrélation spatiale. Par convention, nous ordonnons les composantes par ordre croissant des valeurs propres associées, ce qui revient à les ordonner depuis les autocorrélations les plus négatives vers les autocorrélations les plus positives, via l'autocorrélation nulle (Dray et al. 2004). On peut donc associer à chacun des vecteurs un indice d'autocorrélation caractéristique. La Figure 13a montre le gradient de corrélation spatiale des composantes PCNM au sens de l'indice  $I$  de Moran, pour un échantillonnage de 2500 points sur une grille régulière, et pour les premiers voisins (distance unité par convention).
- L'indice de Moran est une statistique qui permet de tester dans quelle mesure une distribution spatiale donnée diffère d'une distribution complètement aléatoire, via un test de Mantel. On peut ainsi définir quelles sont les composantes PCNM dont la signature

spatiale est significativement structurée ou non. Nous avons donc défini un ensemble de composantes PCNM dont la structure ne diffère pas significativement à 5% d'une distribution aléatoire (pour la méthode, voir Legendre and Legendre 1998, p.719). Il s'agit des composantes dont l'indice est compris entre 1148 et 1351 inclus (voir Figure 13b).

- Il y a plusieurs représentations possibles du spectre PCNM. L'une d'elle, appelée spectre de puissance, réalise une décomposition de la variance spatiale totale du signal sur un gradient de corrélation spatiale exprimé par la base PCNM.
- Le cospectre PCNM de deux signaux représente de la même manière une décomposition de la covariance spatiale totale entre deux signaux, sur un gradient de corrélation spatiale exprimé par la base PCNM.

(a) Vue d'ensemble de l'autocorrélation spatiale des composantes PCNM

I de Moran



(b) Composantes PCNM non significativement structurées dans l'espace

I de Moran

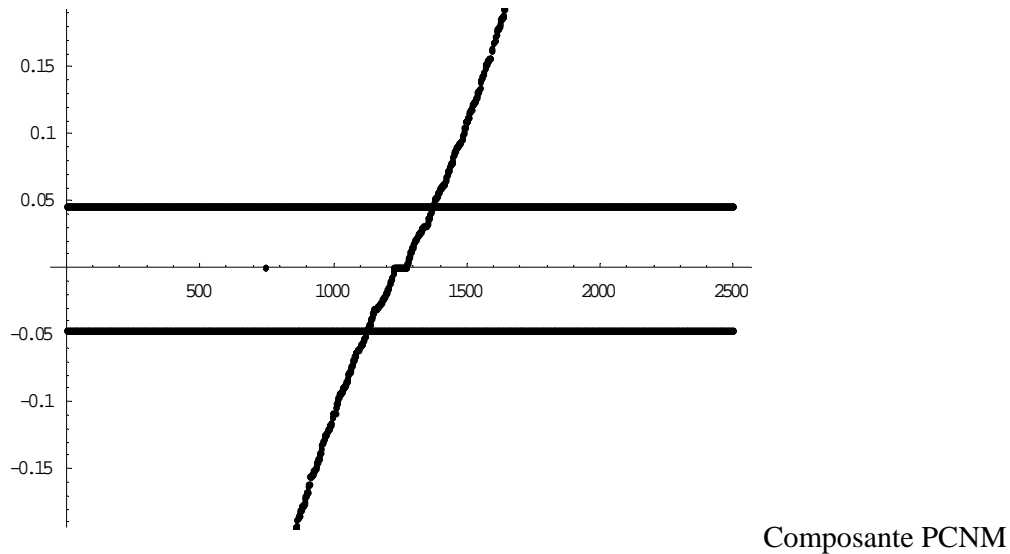


Figure 13 : Représentation de l'indice de corrélation spatiale au sens du I de Moran pour les vecteurs PCNM d'une décomposition réalisée pour un graphe d'échantillonnage de 2500 points disposés sur une grille carrée régulière. (a) Gamme des valeurs d'autocorrélation représentée par l'ensemble des 2500 composantes PCNM. (b) Zone de non significativité à 5% de la structure spatiale d'autocorrélation (hachuré), et composantes PCNM associées.

Ces quatre propriétés essentielles nous amènent à définir de nouveaux indices de partition de la variance spatiale :

- Considérant le spectre de variance PCNM d'un signal donné  $X$ , il est possible de définir une fraction de variance associée à des composantes PCNM significativement autocorrélées ( $Var_{spat}(X)$ ), et une fraction associée aux composantes 1148 à 1351 incluses non significativement autocorrélées ( $Var_{nonspat}(X)$ ).
- De même, il est possible de définir la covariance de deux signaux  $X$  et  $Y$  significativement corrélée ( $Covar_{spat}(X, Y)$ ) et non significativement corrélée ( $Covar_{nonspat}(X, Y)$ ).
- En l'occurrence, nous nous intéressons à la structuration spatiale des métapopulations dans un habitat structuré.  $X$  sera donc la distribution spatiale des populations, et  $Y$  la distribution spatiale de l'habitat associé.

Deux indices de partition spatiale de la variance nous intéresseront dès lors au premier chef :  $part(X) = Var_{spat}(X) / (Var_{spat}(X) + Var_{nonspat}(X))$  est le degré de structure spatiale de la



distribution des populations,  $part(X,Y) = Covar_{spat}(X,Y) / (Covar_{spat}(X,Y) + Covar_{nonspat}(X,Y))$  est le degré de structure spatiale de la covariance entre la distribution des populations et celle de l'habitat.

On peut aller plus loin dans cette logique en rappelant que la gamme des corrélations spatiales exprimées par les composantes PCNM incluent des valeurs négatives (autocorrélation négative, ou anti-corrélation), et des valeurs positives (autocorrélation positive). Il est possible de scinder les indices de partition pour séparer les composantes de signe positif et négatif d'autocorrélation : on aurait  $part^+(X) = Var^+_{spat}(X) / (Var^+_{spat}(X) + Var^-_{spat}(X) + Var_{nonspat}(X))$  et  $part^-(X) = Var^-_{spat}(X) / (Var^+_{spat}(X) + Var^-_{spat}(X) + Var_{nonspat}(X))$ . On procéderait de même pour la partition de la covariance.

### 3.4. Test sur données simulées et perspectives pour l'inférence

Nous étudions cette nouvelle approche de partition de la variance et de la covariance spatiale sur les métapopulations simulées en habitat binaire.  $X$  est ici la distribution moyenne dans le temps des populations à l'état quasi-stationnaire, tandis que  $Y$  est la distribution de l'habitat binaire associé. Ces variables sont centrées, c'est à dire que les composantes constantes de leurs spectres respectifs et de leur cospectre sont nulles.

#### 3.4.1. Découplage de la structure habitat et de la structure métapopulation

Soulignons en premier lieu que pour une carte  $(p, q)$  donnée, l'augmentation de  $r$  induit une augmentation de  $part^-(X)$  et une diminution de  $part^+(X)$ . Cela s'explique simplement en considérant qu'une augmentation de  $r$  correspond à une augmentation du taux d'extinction relatif. Celui-ci tend à introduire une anti-corrélation plus forte en rompant la similarité entre voisins par des événements d'extinction plus fréquents. Cette interprétation est confirmée par les résultats obtenus en considérant  $X_t$ , la distribution instantanée des populations : dans ce cas l'incidence des extinctions en terme d'anti-corrélation est exacerbée car non lissée par une moyenne dans le temps. On a donc naturellement  $part^-(X_t) > part^-(X)$ , et une augmentation plus marquée de  $part^-(X_t)$  quand  $r$  augmente. En définitive  $part(X) = part^-(X) + part^+(X)$  augmente quand  $r$  augmente, ce qui signifie que la structuration spatiale globale est plus forte : c'est de fait ce que l'on attend quand on approche du seuil critique d'extinction de la métapopulation et que le phénomène d'auto-organisation est renforcé.

*A contrario*, pour  $(p, q)$  fixé et  $r$  augmentant,  $part^-(X, Y)$  diminue,  $part^+(X, Y)$  augmente et  $part(X, Y)$  diminue. Alors que la structure spatiale se renforce à mesure que l'on approche du seuil critique d'extinction, il y a aussi une diminution du taux d'occupation de l'habitat et un découplage progressif entre la structure spatiale des populations et celle de l'habitat. Cela vient conforter les idées formulées au Chap. III §2.1 sur le découplage des structures de l'habitat *a priori* et des structures engendrées par la dynamique spatiale des populations.

Ces deux propriétés importantes, prises conjointement, constituent un argument en faveur d'une auto-organisation accrue au voisinage du seuil d'extinction: à mesure que l'on approche du seuil, la structure des populations est de moins en moins dépendante de celle de l'habitat, mais devient néanmoins de plus en plus forte, sous l'effet de la dynamique intrinsèque auto-organisée.

En définitive, il ne faudra pas appréhender de la même manière la structure spatiale d'une espèce rare ultra-spécialisée mais saturant son habitat, et la structure spatiale d'une espèce rare présentant un habitat répandu mais un taux d'extinction relatif élevé. Si l'on recherche une relation causale pour définir la niche fondamentale de ces espèces, la prise en compte de l'espace sera foncièrement différente dans les deux cas (voir Chap. III §2.1): si l'espèce sature son habitat, prendre en compte la structure spatiale de l'habitat est source de biais, si l'espèce ne sature pas son habitat, c'est ne pas prendre en compte la structure spatiale de l'habitat qui sera source de biais. Cela est susceptible d'avoir des effets dévastateurs pour comparer l'écologie des espèces rares: il n'y a pas une unité écologique dans le concept d'espèce rare.

### 3.4.2. Perspectives d'inférence

Nous étudions à présent la relation entre les indices  $part(X)$  et  $part(X, Y)$  d'une part, et d'autre part les paramètres de structure spatiale de l'habitat  $p$  et  $q$ , et de dynamique spatiale des populations  $r$ . Des régressions linéaires multiples sont pratiquées, avec comme variables d'effet ces indices, et comme variables causales:  $p$ ,  $q$ ,  $r$ ,  $p^2$ ,  $q^2$ ,  $p \times q$ ,  $r \times p$  et  $r \times q$ . Les composantes quadratiques sont introduites pour rappeler d'une part l'interaction possible entre paramètres dans la réponse des espèces, d'autre part la prise en compte d'indices supplémentaires de structuration spatiale de l'habitat, en particulier la probabilité de paires 1/1, décrite par  $p \times q$ .

Les résultats de cette analyse sont donnés dans le Tableau 2 (calculs réalisés avec SPlus). Nous pouvons ici souligner une propriété importante : les termes de la régression contenant  $r$  (i.e.  $r$ ,  $p \times r$  et  $q \times r$ ) sont significativement différents de 0 dans l'analyse de  $part(X)$ , mais tous non significativement de 0 pour  $part(X,Y)$ . Sur l'ensemble des cartes d'habitat, cela indique donc que la part de covariance structurée dans l'espace entre les métapopulations et leur habitat ne dépend pas de la dynamique intrinsèque des populations, mais uniquement de la structure de leur habitat. Nous avons cependant vu que pour une carte donnée, il y a une diminution de la part spatiale de la covariance quand  $r$  augmente. *A contrario*, la proportion de variance spatialement structurée des métapopulations dépend significativement des paramètres  $r$ ,  $p \times r$  et  $q \times r$ . Notons que la variance spatiale brute, elle, dépend de  $r$ , et donc que l'indépendance demeure une propriété de la partition relative spatiale/non spatiale de la covariance entre  $X$  et  $Y$ . Au demeurant ces résultats soulignent une fois encore l'incidence de l'auto-organisation quand  $r$  augmente, la structure spatiale des populations se renforçant, là où le couplage des structures spatiales des populations et de l'habitat s'atténue.

	<i>part(X)</i>				<i>part(X,Y)</i>			
	Value	Std. Error	<i>t</i> value	Pr(> t )	Value	Std. Error	<i>t</i> value	Pr(> t )
(Intercept)	<b>92.3188</b>	<b>0.7403</b>	<b>124.7005</b>	<b>0.0000</b>	<b>93.2105</b>	<b>1.1394</b>	<b>81.8098</b>	<b>0.0000</b>
<i>p</i>	0.0807	0.1989	0.4055	0.6855	0.5590	0.3061	1.8260	0.0692
<i>q</i>	<b>-1.1068</b>	<b>0.1771</b>	<b>-6.2483</b>	<b>0.0000</b>	<b>-1.7765</b>	<b>0.2726</b>	<b>-6.5169</b>	<b>0.0000</b>
<i>r</i>	<b>1.3262</b>	<b>0.2681</b>	<b>4.9458</b>	<b>0.0000</b>	0.4983	0.4127	1.2074	0.2286
$p^2$	<b>-0.0662</b>	<b>0.0182</b>	<b>-3.6438</b>	<b>0.0003</b>	-0.0682	0.0280	-2.4392	0.0155
$q^2$	<b>0.2302</b>	<b>0.0156</b>	<b>14.7597</b>	<b>0.0000</b>	<b>0.2778</b>	<b>0.0240</b>	<b>11.5731</b>	<b>0.0000</b>
$p \times q$	<b>-0.0715</b>	<b>0.0256</b>	<b>-2.7994</b>	<b>0.0056</b>	<b>-0.1072</b>	<b>0.0393</b>	<b>-2.7263</b>	<b>0.0069</b>
$r \times p$	<b>0.0770</b>	<b>0.0228</b>	<b>3.3773</b>	<b>0.0009</b>	-0.0279	0.0351	-0.7947	0.4277
$r \times q$	<b>-0.2013</b>	<b>0.0311</b>	<b>-6.4790</b>	<b>0.0000</b>	-0.0161	0.0478	-0.3363	0.7369

Tableau 2: Analyse de la relation entre le degré de structuration spatiale de la variance d'occupation des métapopulations ( $part(X)$ ), et de structuration spatiale de la covariance entre l'occupation et l'habitat, en fonctions des paramètres  $p$ ,  $q$  de structure de l'habitat, et  $r$  de dynamique métapopulationnelle. Les objets sont les métapopulations simulées en habitat binaire. Cette analyse repose sur une régression linéaire multiple dont les résultats sont ici indiqués.

D'un point de vue pratique,  $part(X)$  sera un bon indice pour l'estimation de  $r$ , à condition de connaître  $p$  et  $q$ .  $r$  peut donc être estimé de manière conditionnelle, en estimant au préalable une distribution supputée d'habitat, avec un choix arbitraire de variables et une méthode de détermination *ad hoc* de l'habitat (généralement par GLM). Le  $r$  estimé correspond alors à la dynamique spatiale des populations sous condition d'un habitat donné. On peut en particulier obtenir plusieurs estimations de  $r$ , suivant différents contextes d'habitat.

On retombe cependant sur un problème de base : l'habitat et donc sa structure ne peuvent être perçus (estimation de  $p$  et  $q$ ) que dans une zone active où les populations sont présentes. Lorsque la valeur de  $r$  sera élevée, il faudra ou bien faire une estimation locale, de fait biaisée, de l'habitat à un instant donné, ou bien attendre de très nombreuses générations pour avoir une image stabilisée de l'occupation d'habitat. Il est donc vraisemblable qu'il faudra faire un compromis.

## IV. Modélisation analytique : des causes aux effets

Dans le Chapitre précédent, nous avons analysé directement la relation entre différentes statistiques spatiales supposées pertinentes (issues de l'analyse spectrale ou de la partition de variation), et les paramètres de la dynamique métapopulationnelle ( $p$ ,  $q$  et  $r$ , voir Chapitre I). Nous avons ainsi pu évaluer les intérêts et les inconvénients de ces représentations. Cette évaluation était fondée sur des hypothèses *ad hoc* et intuitives concernant la relation entre patrons spatiaux et processus sous-jacents (structures corrélées, découplage des signatures spatiales).

Pour compléter cette démarche, nous nous intéressons ici à la représentation mécanistique du processus métapopulationnel, afin d'éclairer la nature du lien entre la dynamique locale et les structures émergentes. Nous nous appuyons donc sur des modèles analytiques permettant de représenter et de résoudre algébriquement les états des populations à l'état quasi-stationnaire. Nous utiliserons en particulier plusieurs supports analytiques, ainsi que plusieurs démarches d'estimation du paramètre clé d'extinction/colonisation  $r$ , afin de comprendre l'effet de différents systèmes de représentation et conclure sur leurs applicabilités respectives.

### 1. Approche relativiste et informationnelle

Il n'y a pas de manière unique de modéliser la dynamique des métapopulations. Ce qui apparaît plus important encore, c'est qu'il n'existe pas dans l'absolu de démarche de modélisation plus vraie que les autres.

Pour répondre à la question du lien entre les processus dynamiques de colonisation-extinction et les structures spatiales d'occupation résultante, j'ai été amené à considérer plusieurs manières de modéliser un même processus. Loin d'être redondante, cette démarche permet d'éclairer différents aspects de la relation patron / processus, et milite en faveur du recours systématique à plusieurs démarches de modélisation : on introduit ainsi l'effet « système de représentation ».

S'enfermer dans un formalisme unique, ce serait perdre le recul nécessaire pour évaluer l'effet des nombreux présupposés et simplifications, par rapport à l'effet des processus réels

au sein du système étudié. Si un modèle peut être utilisé pour falsifier un faisceau d'hypothèses, il ne peut pas être utilisé de manière absolue comme support de validation. En revanche l'utilisation croisée de plusieurs démarches de modélisation permet de comprendre et d'évaluer la nature de l'information contenue dans les données réelles afin d'orienter correctement l'inférence.

Comme nous l'avons précisé en Introduction, la définition du système et de ses états est l'étape cruciale de la démarche de modélisation. Avant même de décrire les processus, on définit ainsi une certaine vision, un cadre d'étude. Dans mon cas, le système est une étendue spatiale présentant des cellules identiques susceptibles de recevoir des populations. Celles-ci sont soumises à des événements de colonisation et d'extinction. Les états du système peuvent être définis de différentes manières, ce qui oriente le choix de la modélisation :

- On s'intéresse à des mesures globales d'occupation de l'espace : **densité globale** d'occupation de l'habitat, et **corrélations spatiales globales** des occurrences. Au régime quasi-stationnaire, ces états sont évalués par des statistiques agrégées dans l'espace, pour lesquelles on dispose éventuellement de répliquats temporels. C'est le domaine des modèles spatialement implicites, par exemple de type champ moyen (*Mean-Field*) pour la prédiction de la densité d'occupation (modèle de Levins, en temps continu), ou de type approximations de paires (*Pair Approximation*, en temps continu) pour la prédiction des corrélations spatiales à différentes échelles (Ellner 2001). Ces modèles cités en temps continu permettent l'obtention de solutions analytiques simples, mais au prix d'une information très condensée, et dès lors peut-être peu pertinente, sur les patrons d'occupation.
- On s'intéresse à la **probabilité locale d'occupation** des cellules, dans une représentation en temps discret. Deux cas de figure très distincts, du point de vue de la culture scientifique, peuvent être distingués :
  - Le modélisateur formule l'hypothèse d'absence de corrélation temporelle des occupations locales lors de l'état quasi-stationnaire. Cette hypothèse d'indépendance nous place dans le cadre de la théorie des probabilités, et **l'espérance des probabilités locales** quasi-stationnaire peut être estimée par la moyenne temporelle des réalisations instantanées d'occupation. Les modèles stochastiques markoviens en temps discret, de type *Incidence Function Model*, correspondent à une telle représentation.

- Le modélisateur peut s'intéresser aux structures couplées spatiales et temporelles induites par les séquences de colonisations et d'extinctions, même durant l'état quasi-stationnaire. On s'intéresse alors à la probabilité **locale et instantanée** d'occupation des cellules. Cette représentation est une gageure, car on ne dispose pas à la fois de répliquats dans l'espace et dans le temps. On doit donc se limiter à la connaissance des données de présence-absence, donc aux réalisations. C'est le domaine des processus de contact (*Contact Process*), et des phénomènes d'auto-organisation (voir Chapitre II). Les phénomènes abordés alors sortent du champ d'investigation de la théorie des probabilités pour entrer dans celui de la physique statistique.

Nous allons voir que ces différentes approches de modélisation apportent des informations complémentaires. Cet éclairage croisé nous permettra d'apporter des éléments inédits pour la compréhension du lien entre dynamique métapopulationnelle et structure spatiale résultante.

#### Note importante

Dans mon travail, je m'intéresse aux états quasi-stationnaires de processus markoviens, indépendamment des conditions initiales d'occupation de l'espace, et donc indépendamment des régimes transitoires (voir Encadré 3). Ce faisant, je ne m'intéresse pas aux conditions d'invasibilité du milieu, conditions susceptibles de varier en fonction de la nature des modèles, en particulier en temps discret vs. continu (Ovaskainen and Hanski 2001).

En particulier l'existence d'effets Allee au sein des métapopulations est certes une question écologique importante (Harding and McNamara 2002; Hui and Li 2003, 2004), mais elle se situe en dehors du champ de mon travail. Rappelons que le point de départ de ma démarche est la compréhension des processus dynamiques façonnant la distribution actuelle et observée des espèces, sur la base de données d'inventaires. Dans un tel contexte, la question de l'invasibilité n'est pas posée, et nous ne considérerons pas par la suite les cas d'espèces très rares pouvant constituer des situations transitoires ambiguës soumises aux problèmes d'invasion.

## **2. Représentation spatialement implicite par approximations de paires**

Le principe du modèle champ-moyen de Levins (voir Chap. I §1, p.35) est de réduire au maximum la prise en compte de la structuration spatiale des processus, afin de permettre un

traitement analytique simple. Toutefois, cela ne permet qu'une appréhension globale et grossière du comportement du système. Notre préoccupation étant de comprendre la signification écologique de la structure spatiale des populations, nous cherchons plutôt à exprimer le lien entre des mesures de structuration spatiale et les paramètres des processus dynamiques de colonisation-extinction.

La manière la plus simple d'améliorer le modèle champ-moyen dans ce sens est de considérer de nouveaux états du système, des états choisis pour représenter la structure spatiale des populations. C'est l'esprit des méthodes d'approximation par paires, ou *Pair Approximation*. Le modèle fondamental de Levins s'intéresse uniquement à l'état individuel des cellules d'habitat, occupé ou inoccupé, tandis que le modèle de *Pair Approximation* introduit en outre l'état des paires de cellules : ce peut être par exemple l'état de deux cellules distantes de  $n$  cases, *occupé-occupé* ou *inoccupé-occupé*, ou encore *inoccupé-inoccupé*. Des équations représentent l'évolution des différents états, avec un formalisme déterministe en temps continu dans l'esprit du modèle de Levins. On obtient alors un système d'équations différentielles couplées, dont on recherche une résolution analytique.

Les modèles de *Pair Approximation* sont très utilisés depuis longtemps en physique, et connaissent en écologie un développement récent pour l'étude des dynamiques spatiales de populations (Ellner et al. 1998; Hiebeler 2000; Iwasa 2000; Snyder and Nisbet 2000; Ellner 2001; Ovaskainen et al. 2002; Thomson and Ellner 2003). Cette approche permet d'évaluer un effet global des processus de colonisation-extinction sur la corrélation spatiale des populations, sur une ou plusieurs classes de distances. Les probabilités de paires sont donc des grandeurs déterminées et uniformes dans l'espace. On peut utiliser les prédictions du modèle comme base d'inférence sur les paramètres de colonisation-extinction, à condition de disposer de données réelles présentant une telle uniformité spatiale des probabilités de paires.

Parce que nous souhaitons éclairer le lien entre la dynamique d'extinction-colonisation et la structure spatiale des populations, nous développons ici un modèle de type *Pair Approximation* exprimant les probabilités de paires de cellules occupées sur plusieurs voisinages. Les résultats obtenus nous permettent d'envisager une solution d'inférence sur ces probabilités. Nous étudierons en dernier lieu les atouts et les limites d'une telle démarche.



## 2.1. Principe général

On considère une aire de dimension infinie divisée en mailles égales (par exemple des carrés, ou tout motif constituant un pavage). On s'intéresse à l'occurrence des populations d'une espèce  $S$ . Pour simplifier, l'habitat de cette espèce est soit favorable, soit défavorable et chaque maille se trouve à un instant donné dans l'un des deux états (discrétisation spatiale de l'habitat). On s'intéresse ici à l'abondance de l'espèce  $S$  en terme de nombre d'occurrences et non en terme d'effectifs. Dès lors une maille d'habitat favorable est soit occupée, soit inoccupée par l'espèce  $S$ , abstraction faite de tout phénomène intra-maille, en particulier lié à la démographie.

Le système est soumis à des événements qui déterminent son évolution. Une maille occupée, *i.e.* une population, est soumise à un aléa d'extinction par unité de temps noté  $e$ . En outre chaque population disperse des propagules dans les mailles voisines. Dès lors on définit pour chaque maille occupée (état 1) une probabilité  $\beta$  par unité de temps de coloniser une maille voisine inoccupée si elle est favorable (état 0). L'état 2 décrit l'habitat défavorable.

La dispersion depuis une maille occupée vers une maille vide voisine est fonction de la distance d'une part, et de la disponibilité en sites vides favorables d'autre part. Les distributions spatiales des sites occupés et de l'habitat influent donc sur la dynamique globale du système. La probabilité de changement d'état d'une maille dépend de la nature des mailles voisines. L'idée de l'approximation par paire est de restituer cet effet de la structure spatiale « au premier ordre », en prenant en compte la dynamique de doublets de mailles voisines. Par extension, on parle de *n*-approximation, quand on étudie les dynamiques des 2-,3-..., *n*-uplets.

La capacité de colonisation dans l'espace infini d'une population ponctuelle peut être décrite en fonction de la distance à la maille source, ce que l'on appellera le degré de voisinage. Les voisins adjacents à la maille source constituent son premier voisinage. Les voisins adjacents au premier voisinage constituent le second voisinage, et ainsi de suite. On note  $\beta_i$  la probabilité par unité de temps qu'une maille occupée colonise une maille favorable inoccupée appartenant à son voisinage  $i$ . Enfin, chaque voisinage  $i$  est constitué de  $z_i$  mailles. Si l'on considère des mailles carrées, et dans un schéma simplifié, les voisinages figurent des « anneaux » autour de la maille centrale, ce qui permet d'écrire simplement  $z_i = 8.i$ .

## 2.2. Modélisation de la dynamique du système

Le formalisme proposé repose sur celui décrit par Ellner (2001). Il est étendu à la prise en compte d'une matrice binaire d'habitat favorable / défavorable.

Afin de décrire les événements de colonisation, on définit les densités absolues de couples de mailles distantes de  $i$  :  $\rho_{00}^{(i)}, \rho_{11}^{(i)}, \rho_{22}^{(i)}, \rho_{10}^{(i)} = \rho_{01}^{(i)}, \rho_{20}^{(i)} = \rho_{02}^{(i)}, \rho_{12}^{(i)} = \rho_{21}^{(i)}$ .

De la même manière, on parle de probabilité conditionnelle de rencontrer une maille de type 1

à une distance  $i$  d'une maille de type 0 :  $q_{1/0}^{(i)} = \frac{\rho_{10}^{(i)}}{\rho_0}$ .

On utilise ici une représentation temporelle continue. Les mailles de type 2 sont intangibles, d'où  $\rho_{22}^{(i)} = \text{Cste}(t)$  et  $\rho_2 = \text{Cste}(t)$ . En revanche les mailles de types 0 et 1 sont soumises aux événements de colonisation et d'extinction.  $\rho_0, \rho_1, \rho_{00}^{(i)}, \rho_{11}^{(i)}, \rho_{10}^{(i)}, \rho_{20}^{(i)}, \rho_{12}^{(i)}$  sont donc des fonctions du temps. Notons qu'à tout instant on a les relations suivantes :

$$\text{Eq. 1} \quad \begin{cases} \rho_0 = 1 - \rho_2 - \rho_1 \\ \rho_{00}^{(i)} = \rho_0 - \rho_{10}^{(i)} - \rho_{20}^{(i)} \\ \rho_{20}^{(i)} = \rho_2 - \rho_{22}^{(i)} - \rho_{12}^{(i)} \\ \rho_{10}^{(i)} = \rho_1 - \rho_{12}^{(i)} - \rho_{11}^{(i)} \end{cases}$$

D'où il apparaît que la dynamique du système peut être entièrement décrite par l'étude de trois variables, par exemple  $\rho_1, \rho_{11}^{(i)}, \rho_{12}^{(i)}$ .

Considérons donc la dynamique de ces variables. Nous réalisons d'abord le bilan de flux de mailles acquérant et perdant l'état 1 entre  $t$  et  $t+dt$ . Premièrement, une maille de type 0 peut être colonisée par des mailles voisines occupées, avec la probabilité  $\sum_i z_i \beta_i q_{1/0}^{(i)}$ . Dans

le même intervalle de temps, une maille de type 1 peut passer à l'état 0 par extinction, à la probabilité  $e$ . Il en résulte l'Eq. 2:

$$\text{Eq. 2} \quad \frac{d\rho_1}{dt} = -e \cdot \rho_1 + \rho_0 \sum_i z_i \beta_i q_{1/0}^{(i)}$$

Le même type d'analyse peut être réalisé pour des paires de mailles. Considérons donc à un instant  $t$  l'ensemble des paires de mailles occupées distantes de  $i$  (notées  $II^{(i)}$ ). Durant un intervalle de temps  $dt$ , une extinction peut se produire avec la probabilité  $e$  dans chacune de ces mailles, d'où une disparition de paires  $II^{(i)}$  à la probabilité  $2e$ . En revanche durant le même intervalle de temps des mailles de type  $10^{(i)}$  et  $00^{(i)}$  peuvent passer à l'état  $11^{(i)}$  par des

événements de colonisation. On néglige ici le cas des transitions  $00^{(i)} \rightarrow 11^{(i)}$ , par une approximation au premier ordre sur les  $\beta_i$ : les doubles événements de colonisation sont négligés devant les événements de colonisation simples.

Les transitions  $10^{(i)} \rightarrow 11^{(i)}$  sont de deux types. La maille inoccupée d'une paire  $10^{(i)}$  peut être colonisée par l'autre maille occupée, suivant la probabilité  $\beta_i$ . Elle peut aussi être colonisée par les mailles occupées parmi les  $z_i-1$  autres mailles du voisinage  $i$ , suivant la probabilité  $(z_i-1)\beta_i \cdot q_{1/0}^{(i)}$ . Enfin elle peut être colonisée par les mailles occupées des autres voisinages, suivant la probabilité  $\sum_{j \neq i} z_j \beta_j q_{1/0}^{(j)}$ .

On analyse de la même manière les transitions  $01^{(i)} \rightarrow 11^{(i)}$ . Il résulte de ces bilans l'Eq. 3:

$$\text{Eq. 3} \quad \frac{1}{2} \cdot \frac{d\rho_{11}^{(i)}}{dt} = -e \cdot \rho_{11}^{(i)} + \rho_{10}^{(i)} \cdot \left( \beta_i + (z_i - 1)\beta_i q_{1/0}^{(i)} + \sum_{j \neq i} z_j \beta_j q_{1/0}^{(j)} \right)$$

Considérons enfin l'ensemble des couples de mailles occupées / habitat défavorable distantes de  $i$  (notées  $12^{(i)}$ ). Pendant l'intervalle  $dt$ , la transition  $12^{(i)} \rightarrow 02^{(i)}$  correspond à une extinction, avec une probabilité  $e$ . La transition  $02^{(i)} \rightarrow 12^{(i)}$  est possible soit par colonisation de la maille 0 par une maille occupée de son voisinage  $i$  (probabilité  $(z_i-1)\beta_i \cdot q_{1/0}^{(i)}$ ), soit par colonisation de la maille 0 par une maille occupée d'un autre voisinage.

Il vient alors l'Eq. 4:

$$\text{Eq. 4} \quad \frac{d\rho_{12}^{(i)}}{dt} = -e \cdot \rho_{12}^{(i)} + \rho_{02}^{(i)} \cdot \left( (z_i - 1)\beta_i q_{1/0}^{(i)} + \sum_{j \neq i} z_j \beta_j q_{1/0}^{(j)} \right)$$

Eq. 1 permet de réécrire ces équations différentielles en fonction des seules variables  $\rho_1$ ,  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$ .

### 2.3. Etude des solutions à l'état stationnaire

L'Eq. 2 permet d'écrire :  $\sum_i z_i \beta_i q_{1/0}^{(i)} = \frac{1}{\rho_0} \left( \frac{d\rho_1}{dt} + e \cdot \rho_1 \right)$ . On peut dès lors opérer une substitution dans les équations Eq. 3 et Eq. 4 :

$$\text{Eq. 5} \quad \frac{1}{2} \cdot \frac{d\rho_{11}^{(i)}}{dt} = -e \cdot \rho_{11}^{(i)} + \rho_{10}^{(i)} \cdot \left( \beta_i (1 - q_{1/0}^{(i)}) + \frac{1}{\rho_0} \left( \frac{d\rho_1}{dt} + e \cdot \rho_1 \right) \right)$$

$$\text{Eq. 6} \quad \frac{d\rho_{12}^{(i)}}{dt} = -e \cdot \rho_{12}^{(i)} + \rho_{02}^{(i)} \left( \frac{1}{\rho_0} \left( \frac{d\rho_1}{dt} + e \cdot \rho_1 \right) - \beta_i q_{1/0}^{(i)} \right)$$

A l'état stationnaire, les dérivées temporelles sont nulles, d'où :

$$\text{Eq. 7} \quad -e \cdot \rho_0 \cdot \rho_{11}^{(i)} + \rho_{10}^{(i)} \cdot (\beta_i (\rho_0 - \rho_{10}^{(i)}) + e \cdot \rho_1) = 0$$

$$\text{Eq. 8} \quad -e \cdot \rho_0 \cdot \rho_{12}^{(i)} + \rho_{02}^{(i)} \cdot (e \cdot \rho_1 - \beta_i \rho_{10}^{(i)}) = 0$$

Ce qui permet d'exprimer  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$  en fonction de  $\rho_1$  en travaillant à  $i$  fixé avec les deux seules équations Eq. 7 et Eq. 8. On utilise *in fine* Eq. 2 pour déterminer  $\rho_1$ .

Notons qu'à cet état stationnaire, s'il existe, les paramètres  $e$  et  $\beta_i$  ne correspondent qu'à un seul degré de liberté: le système à l'équilibre peut être décrit en fonction de  $r_i = e / \beta_i$ . Extinction et colonisation interviennent ainsi de manière opposée.

#### 2.4. Inférence par Pair Approximation

Les équations Eq. 7 et Eq. 8 permettent de prédire les probabilités de paires  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$  en fonction de  $\rho_1$  indépendamment des autres voisinages. Grâce à ce découplage, nous pouvons résoudre analytiquement ces équations avec *Mathematica 5.0*. Les expressions symboliques compliquées de  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$  ne seront cependant pas reprises ici. Nous nous intéressons à l'adéquation des prédictions du modèle analytique avec les probabilités de paires observées dans nos métapopulations simulées en habitat binaire.

Les auteurs ayant utilisés la méthode de *Pair Approximation* pour l'étude des métapopulations se sont intéressés avant tout à la prédiction de  $\rho_1$  pour des modèles ne faisant intervenir qu'un voisinage. J'adopte de mon côté une démarche différente : pour chaque métapopulation simulée, j'utilise la valeur observée de  $\rho_1$ , c'est à dire mesurée sur la métapopulation simulée, afin de résoudre le système d'équations Eq. 7 + Eq. 8, et prédire le profil des probabilités de paires occupées.

Les prédictions sur les probabilités de paires sont donc soumises à une estimation correcte de  $\rho_1$  à partir des observations. Quoi qu'il en soit, les profils calculés pour nos simulations exhibent une très bonne adéquation entre valeurs estimées et attendues (Figure 14).

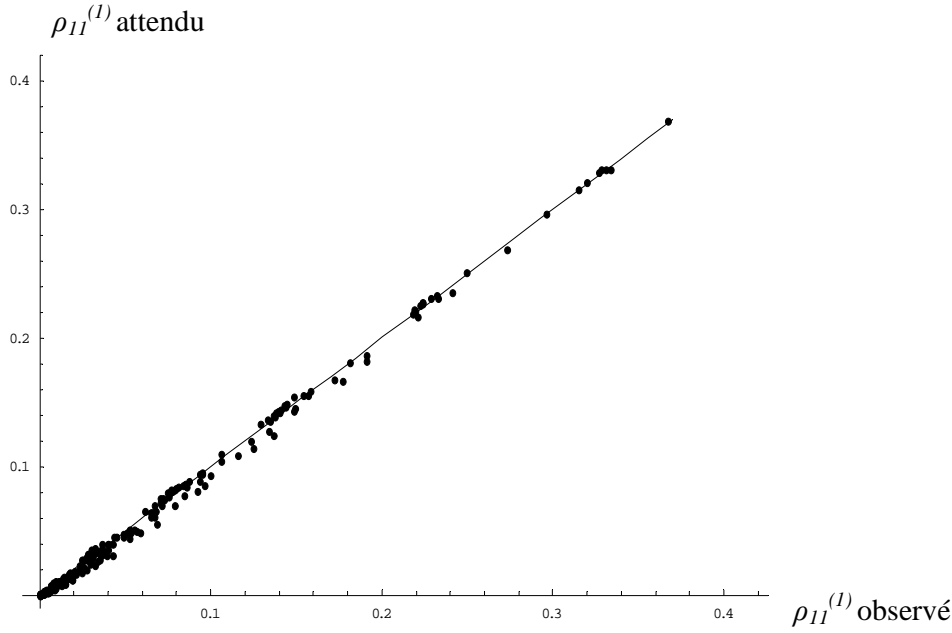


Figure 14 : Comparaison des valeurs observées de  $\rho_{11}^{(1)}$  aux valeurs attendues par *Pair Approximation*. Chaque point correspond à l'une des 227 métapopulations simulées en habitat binaire ayant une distribution quasi-stationnaire non nulle. On observe des qualités d'ajustement similaires pour  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$  sur les 5 premiers voisinages  $i$ . Les valeurs observées de  $\rho_1$  sont utilisées pour l'obtention des valeurs analytiques des probabilités de paires.

La prédiction analytique des profils de probabilité de paires peut permettre l'inférence des paramètres du modèle. La priorité est naturellement donnée à l'inférence du paramètre clé ratio d'extinction / colonisation par voisinage,  $r_i = e / \beta_i$ .

Nous avons étudié la possibilité d'inférer ce paramètre en supposant un noyau de dispersion exponentiel, tel qu'il a été utilisé pour les simulations :  $\beta_i = \frac{C_t \times \exp(-i)}{\sum_{1 \leq i \leq 5} 8 \times i \times \exp(-i)}$ . Nous

nous attachons donc à inférer le paramètre  $r = e / C_t$  à partir de l'ensemble du profil des probabilités de paires  $\rho_{11}^{(i)}$  et  $\rho_{12}^{(i)}$  sur les 5 premiers voisinages soit sur 10 degrés de liberté. Nous constatons que l'écart entre distributions prédites et observées au sens des moindres carrés est très peu sensible aux valeurs de  $r$ , et la minimisation des écarts n'aboutit pas à une estimation correcte de  $r$ . Cette estimation est même uniforme et indépendante des valeurs réelles de  $r$ .

Cela n'est pas surprenant si l'on considère que les fluctuations de  $r$  peuvent résulter d'une variation simple du taux d'extinction, pour un noyau fixé de colonisation. La probabilité

d'extinction étant uniforme dans l'espace, elle n'affecte pas le profil des probabilités de paires.

On attendrait en revanche que les profils de probabilités de paires soient davantage sensibles à la nature de la fonction de colonisation. Nous avons testé l'effet de la variation des paramètres  $\alpha$  et  $\gamma$  dans l'expression  $\beta_i = \frac{C_i \times \exp(-\alpha \times i^\gamma)}{\sum_{1 \leq i \leq 5} 8 \times i \times \exp(-\alpha \times i^\gamma)}$ . Là encore, nous

constatons une très faible sensibilité des profils de probabilités de paires aux paramètres  $\alpha$  et  $\gamma$ . Toutefois nous considérons toujours en l'occurrence des profils exponentiels de colonisation.

Intéressons nous maintenant à une famille différente de profils de probabilité de colonisation, avec une décroissance hyperbolique par exemple. Nous pouvons discriminer sur la prédiction des probabilités de paires l'effet d'un profil de colonisation « plutôt » hyperbolique ou plutôt exponentiel, en modulant le paramètre  $\alpha$  de la probabilité de

$$\text{colonisation } \beta_i = C_i \times \frac{\alpha \cdot \exp(-i) + (1 - \alpha) \cdot \frac{1}{i}}{\sum_{1 \leq i \leq 5} 8 \times i \times \left( \alpha \cdot \exp(-i) + (1 - \alpha) \cdot \frac{1}{i} \right)}. \quad \alpha \text{ variera entre 0 et 1 inclus.}$$

Nous avons recherché, au sens des moindres carrés, l'ajustement optimal du modèle PA pour cette fonction de colonisation, en fonction de  $\alpha$  et sur les profils de probabilités de paires des métapopulations simulées non nulles. Nous aboutissons pour 130 des 227 simulations à une valeur optimale de  $\alpha$  de 1 exactement. Ce sont des situations où nous sommes donc en mesure de discriminer le véritable modèle exponentiel de dispersion par rapport à un modèle hyperbolique « nul ».

En outre, quand un ajustement de  $\alpha$  différent de 1 est obtenu, nous constatons une sensibilité faible de l'ajustement (valeur absolue de la dérivée seconde de la fonction d'ajustement au point optimum), par rapport aux valeurs de sensibilité obtenues en général pour les estimations à 1 exactement (Figure 15). Cela suggère que dans certains cas l'information de structure spatiale des populations n'est pas suffisamment riche pour discriminer les modèles de dispersion exponentiel et hyperbolique. Des approfondissements de la méthode seront nécessaires pour éclairer la nature de ces limitations, et le cas échéant pour proposer des moyens opérationnels d'inférence des profils de colonisation. Une perspective majeure serait en particulier d'utiliser une formulation plus générale de la

fonction de colonisation pour discriminer un profil dit à queue grasse et un profil dit sans queue grasse. Ce serait un outil inférentiel précieux pour appréhender l'importance des événements de dispersion à longue distance à partir de l'analyse de la distribution actuelle des organismes (voir Clark 1998).

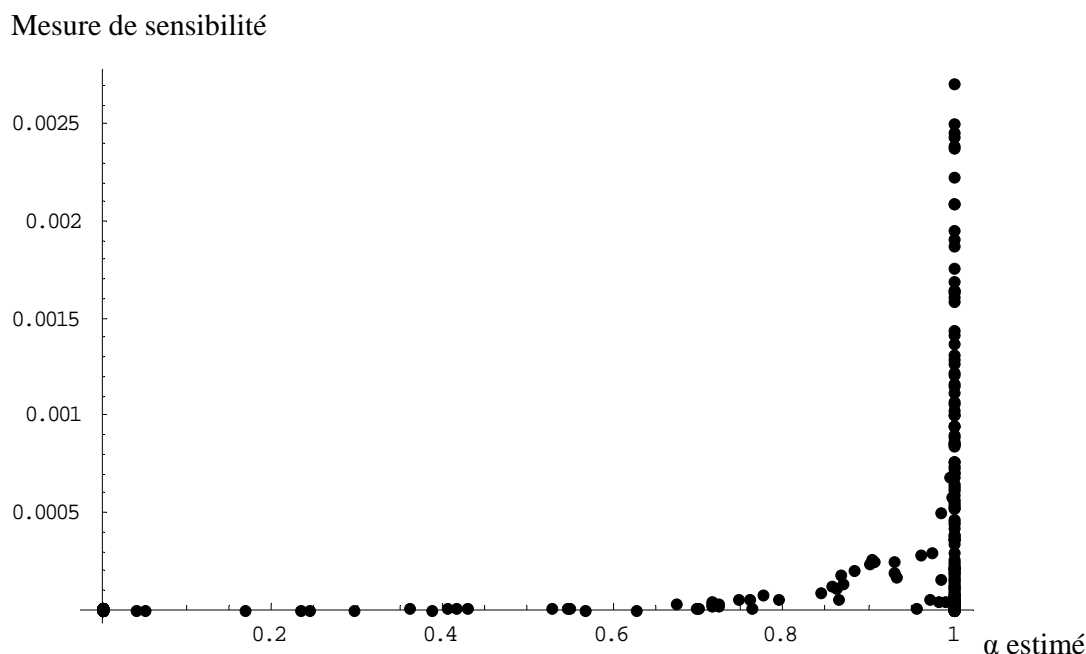


Figure 15 : Analyse de la sensibilité de l'analyse par Pair Approximation des profils de probabilités de colonisation, pour les 227 métapopulations non nulles simulées en habitat binaire (Chapitre I). En abscisses est représentée l'estimation par ajustement du paramètre  $\alpha$  qui mesure l'incidence d'un profil hyperbolique par rapport à un profil exponentiel de

probabilités de colonisation  $\left( \beta_i = C_i \times \frac{\alpha \cdot \exp(-i) + (1 - \alpha) \cdot \frac{1}{i}}{\sum_{1 \leq i \leq 5} 8 \times i \times \left( \alpha \cdot \exp(-i) + (1 - \alpha) \cdot \frac{1}{i} \right)} \right)$ . En ordonnées sont

représentées les valeurs associées de sensibilité d'ajustement. La sensibilité est ici décrite par la valeur absolue de la dérivée seconde de la fonction d'ajustement au point d'optimisation. Chaque point représente une des 227 métapopulations simulées pour lesquelles l'ajustement a été effectué.

### 3. Modèles de métapopulation spatialement réalistes

Les équations locales d'équilibre extinction/colonisation sont à la base de l'étude des états quasi-stationnaires des métapopulations, et ont été largement utilisées dans la littérature. Nous proposons ici de reprendre ces équations sous leur formulation la plus générale. Nous présenterons et discuterons deux hypothèses fondamentales souvent formulées :

- L'hypothèse de stationnarité spatiale, ou stationnarité au second ordre (voir Legendre and Legendre 1998). Cela signifie notamment que la covariance spatiale entre l'état des voisins, fonction de la distance, présente le même profil en tout point de l'espace. Dans notre formalisme métapopulationnel, cela est au mieux envisageable en habitat uniforme, mais trivialement faux en habitat structuré.
- La non-covariance dans le temps de la probabilité d'occupation de cellules voisines. Cette hypothèse impliquera que l'espérance temporelle d'occupation d'un patch n'est pas fonction de sa covariance au cours du temps avec l'état des patches voisins (voir §3.2 et §4.1). L'hypothèse semble fausse à première vue, car l'état des voisins est lié par des processus de colonisation. Mais cela a un sens quand on représente l'espérance des probabilités d'occupation en temps infini. Dans ce cas en effet, quand on estime l'espérance des probabilités d'occupation locale, les variations stochastiques associées aux événements de colonisation et d'extinction sont de nature à rompre les corrélations locales et temporaires qui peuvent apparaître entre les états des voisins. La covariance ne sera en revanche pas nulle si l'intervalle de temps considéré pour estimer cette espérance n'est pas suffisamment long. La covariance sera en particulier élevée au voisinage du seuil critique d'extinction, et sera de nature à introduire l'effet de l'auto-organisation dans l'équation quasi-stationnaire.

Ces hypothèses ont en définitive vocation à « impliquer » la représentation de l'espace et du temps dans le modèle analytique, afin d'obtenir des solutions plus simples. On rejoint ainsi plus ou moins consciemment la logique des modèles spatialement implicites du type *Pair Approximation*.

Dans la partie suivante cependant, nous montrerons que la résolution analytique complète du modèle de dynamique métapopulationnelle n'est pas nécessaire pour pouvoir faire des inférences sur les paramètres de cette dynamique. Nous nous attachons donc ici à comprendre la structure du modèle analytique sans chercher à le résoudre.

### 3.1. Définition du système et équation de base

Le système étudié est un ensemble de populations dans un espace fini discret correspondant à un espace écologique régional : chaque entité spatiale peut contenir une population et une seule. Nous nous plaçons en premier lieu dans un habitat uniforme. Les



objets considérés sont les populations, caractérisés par les coordonnées de l'entité spatiale dans laquelle ils se trouvent. Leur état est décrit par une probabilité d'occupation.

Le temps est appréhendé ici de manière discrète, l'unité de temps correspondant au temps de *génération* d'une population (temps d'arrivée à l'équilibre démographique). Un temps de génération représente un cycle de vie *arbitraire* d'une population, nécessairement plus long que le temps d'arrivée à l'équilibre démographique.

Deux échelles de temps sont dès lors considérées :

- L'échelle ponctuelle, associée à l'étude de la probabilité instantanée d'occupation des entités spatiales.
- L'échelle agrégée, associée à l'étude de la probabilité moyenne d'occupation des entités spatiales sur une plage de temps donnée.

Les flux de populations correspondant aux événements de colonisation et d'extinction sont représentés au niveau de l'entité locale par l'équation suivante :

$$p_{xy}(t+1) - p_{xy}(t) = (Taux\_colonisation_{xy})[1 - p_{xy}(t)] - (Taux\_extinction_{xy})p_{xy}(t)$$

$$p_{xy}(t+1) - p_{xy}(t) = \left[ \sum_{\substack{k=0, l=0 \\ (k,l) \neq (x,y)}}^{N-1} c(d_{xy,kl}) p_{kl}(t) \right] [1 - p_{xy}(t)] - e \cdot p_{xy}(t)$$

Eq. 9 $p_{xy}(t+1) - p_{xy}(t) = \left[ \sum_{k=0, l=0}^{N-1} c(d_{xy,kl}) p_{kl}(t) \right] [1 - p_{xy}(t)] - e \cdot p_{xy}(t) \quad (\text{si } c(0) = 0)$
---

$d_{xy,kl}$  désigne la distance euclidienne entre les points  $(x, y)$  et  $(k, l)$ .  $c(d)$  est une fonction de colonisation caractéristique de l'espèce considérée, telle que  $c(0)=0$  (convention pour simplifier l'écriture de l'équation).

Dans un espace discrétisé à deux dimensions, le produit de convolution est défini par la relation suivante :  $f \otimes g(p, q) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cdot g(p-x, q-y)$

On remarque alors la propriété suivante :

$$\sum_{k=0, l=0}^{N-1} c(d_{xy,kl}) p_{kl}(t) = \sum_{k=0, l=0}^{N-1} c(\sqrt{(x-k)^2 + (y-l)^2}) p_{kl}(t) = p \otimes c(x, y)$$

Avec  $c'(X, Y) = c(\sqrt{X^2 + Y^2}) = c \circ N(X, Y)$ ,  $N$  étant ici la norme euclidienne.

Ce produit de convolution est un élément clé de l'équation de dynamique métapopulationnelle (voir Encadré 11).

L'Eq. 9 peut alors écrire de manière plus simple :

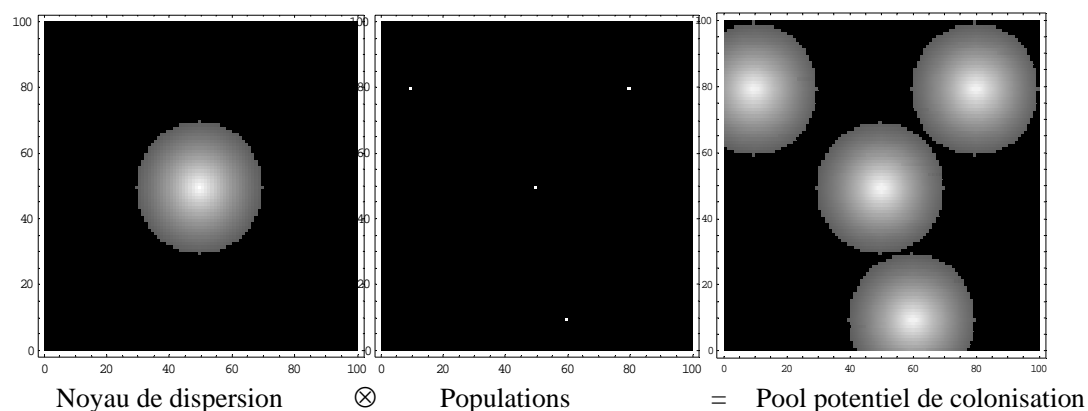
$$\text{Eq. 10} \quad p_{xy}(t+1) - p_{xy}(t) = p \otimes c'(x, y, t) \cdot [1 - p_{xy}(t)] - e \cdot p_{xy}(t)$$

C'est l'équation de base du processus stochastique de dynamique métapopulationnelle, dont les états sont les probabilités d'occupation locale à une génération donnée.

#### Encadré 11 : A propos du produit de convolution

Le produit de convolution est très utilisé en traitement du signal pour ses propriétés particulières. Le théorème fondamental est que la transformée de Fourier (ou de Laplace) d'un produit de convolution est le produit naturel des transformées de Fourier des facteurs. Ce théorème est particulièrement utile en matière de filtrage. Dans le domaine fréquentiel, on filtre un signal en supprimant une certaine gamme de fréquences : il suffit de faire le produit simple de la transformée de Fourier d'un signal par une fonction créneau représentant le filtre. Dans le domaine réel, la fonction filtrée sera le produit de convolution du signal d'origine par la transformée de Fourier inverse de la fonction de filtrage.

Dans mon travail, le produit de convolution  $p \otimes c'(x, y)$  représente l'enveloppe des colons potentiels dispersés depuis les populations occupées :



### Notations

$p_{xy}(t)$  probabilité locale d'occupation à la génération  $t$

$\bar{p}_{xy}$  espérance dans le temps de la probabilité locale quasi-stationnaire d'occupation

$\hat{p}(t)$  espérance dans l'espace de la probabilité d'occupation à la génération  $t$

$\bar{\bar{p}}$  espérance dans le temps et dans l'espace de la probabilité d'occupation

### *3.2. Stationnarité temporelle*

Le processus stochastique représenté est un processus de Markov, car les probabilités  $p_{xy}(t+1)$  d'occupation à une génération  $t+1$  ne dépend que des probabilités  $p_{xy}(t)$  à la génération précédente. Le comportement quasi-stationnaire des processus de contact, auxquels se rattache notre modèle métapopulationnel, a été étudié notamment par (Dickman and Vidigal 2002; Dickman and Martins de Oliveira 2004). La convergence des modèles de métapopulation à fonction d'incidence a été étudiée par (Ovaskainen and Hanski 2001), grâce à une représentation itérative du processus. La convergence de notre système vers une distribution quasi-stationnaire unique peut être démontrée dans ce cadre.

### Note importante

Comme l'Eq. 10 l'illustre, le processus de dynamique modélisé n'est pas linéaire. Cette non-linéarité est au cœur des propriétés singulières qui nous ont intéressés au Chapitre III, la criticalité et l'auto-organisation.

Cette non-linéarité empêche naturellement le recours à un formalisme matriciel, de type chaîne de Markov, et notamment de recourir aux théorèmes afférents permettant de caractériser leur état quasi-stationnaire. Les propriétés sur les vecteurs propres et les valeurs propres de ces chaînes seront ici sans objet. Nous rechercherons certes une possibilité de linéariser la dynamique du système (§3.4), mais cela sera assorti de conditions trop restrictives pour éclairer utilement les questions que nous posons. Nous resterons donc bien dans le cadre de l'étude d'un processus strictement non-linéaire.

Nous nous intéressons dès lors à la caractérisation de l'état quasi-stationnaire du processus. La stationnarité dans le temps sera exprimée par l'équation suivante :

$$\forall k=\{1, \dots, +\infty\}, \quad E_t(p_{xy}(t+k) - p_{xy}(t)) = 0$$

$E_t$  représente l'espérance dans le temps.

L'équation 1a donne donc à l'état stationnaire :

$$E_t(p \otimes c(x, y, t) \cdot [1 - p_{xy}(t)] - e \cdot p_{xy}(t)) = 0$$

On peut écrire : 
$$E_t(p \otimes c(x, y, t)) = \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot E_t(p_{kl}(t)) + Cov_t(c(d_{xy,kl}), p_{kl}(t))]$$

Les variables aléatoires  $c$  et  $p_{xy}$  sont indépendantes dans le temps, ce qui permet d'en déduire :

$$E_t(p \otimes c(x, y, t)) = \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot E_t(p_{kl}(t))] = E_t(p) \otimes E_t(c)(x, y) = \bar{p} \otimes \bar{c}(x, y)$$

En revanche, l'expression suivante est plus délicate à développer :

$$E_t(p \otimes c(x, y, t) \cdot p_{xy}(t)) = \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot E_t(p_{kl}(t) \cdot p_{xy}(t))] = \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot E_t(p_{kl}(t) \cdot p_{xy}(t))]$$

$$E_t(p \otimes c(x, y, t) \cdot p_{xy}(t)) = \bar{p} \otimes \bar{c}(x, y) \cdot \bar{p}_{xy}(t) + \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot Cov_t(p_{kl}(t), p_{xy}(t))]$$

La covariance temporelle de  $p_{kl}$  et  $p_{xy}$  n'est pas nécessairement nulle en raison des processus dynamiques qui les lient dans le temps.

$$Cov_t(p_{kl}(t), p_{xy}(t)) = K_t(d_{xy,kl})$$

$$\sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot Cov_t(p_{xy}(t), p_{xy}(t))] = \sum_{k=0, l=0}^{N-1} [E_t(c(d_{xy,kl})) \cdot K_t(d_{xy,kl})] = K_{cov_t}(x, y)$$

Nous discuterons plus loin l'importance possible de ce terme de l'équation stationnaire.

Eq. 11	$\bar{p} \otimes \bar{c}(x, y) \cdot [1 - \bar{p}_{xy}] - e \cdot \bar{p}_{xy} = K_{cov_t}(x, y)$
--------	---

Avec  $\bar{f}$  décrivant l'espérance dans le temps d'une variable aléatoire  $f$ .

Cette équation ne décrit pas la dynamique du système durant l'état quasi-stationnaire, mais caractérise l'espérance des probabilités locales d'occupation. Le fait de ne pas décrire réellement la dynamique temporelle interdit de décrire les effets temporaires (voir Encadré 3) induits par des perturbations contingentes : cela a été souligné au Chapitre II (§1.1).

Toutefois l'écriture complète de l'équation permet de prendre en compte, sur un temps d'estimation fini des probabilités moyennes, l'effet des corrélations spatiales et temporelles

éventuelles au travers du terme  $K_{cov_t}(x, y)$ . Ce terme de covariance temporelle ne peut être évalué directement dans la pratique, dans la mesure où l'on ne peut pas estimer les probabilités instantanées  $p_{kl}$  et  $p_{xy}$ . On peut seulement accéder aux réalisations associées, à savoir la présence-absence des populations à une date donnée. Dans tous les modèles publiés étudiant l'état quasi-stationnaire d'une métapopulation, ce terme est négligé.

#### Cas particulier : modèle champ moyen

Dans le modèle champ moyen (*mean-field*), la probabilité de colonisation n'est pas dépendante de la distance à la population source et on aurait alors :

$$K_{cov_t}(x, y) = \sum_{k=0, l=0}^{N-1} [Cov_t(p_{kl}(t), p_{xy}(t))] = (p_{xy}(t) - \bar{p}_{xy}) \cdot \sum_{k=0, l=0}^{N-1} [(p_{kl}(t) - \bar{p}_{kl})] = 0$$

$$\text{On a alors simplement: } C_t \cdot \bar{p} \cdot [1 - \bar{p}_{xy}] - e \cdot \bar{p}_{xy} = 0, \text{ et } \bar{p}_{xy} = 1 - \frac{e}{C_t} = \bar{\bar{p}}$$

Avec  $\bar{\bar{x}}$  l'espérance dans l'espace et dans le temps d'une variable aléatoire  $x$ .

### 3.3. Formulation matricielle

L'Eq. 11 peut également être écrite sous une forme matricielle. Considérons une notation des points sous la forme des nœuds d'un graphe  $X_i$  ; on remplace alors  $(x, y)$  par  $i$  et  $(k, l)$  par  $j$ . On a ainsi :

$$\left[ \sum_j c(d_{i,j}) \bar{p}_j(t) \right] \cdot [1 - \bar{p}_i] - e \cdot \bar{p}_i = K_{cov_t}(i)$$

Considérons la matrice  $C(D)$  composée des éléments  $c(d_{i,j})$ , et la transformation suivante :

$$\begin{pmatrix} \bar{p}_1 \\ \vdots \\ \bar{p}_n \end{pmatrix} \xrightarrow{P \rightarrow d(P)'} \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_n \end{pmatrix} \quad \text{On note en outre :} \quad \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\text{On a alors :} \quad P = d(P) \cdot \mathbf{1}$$

L'intérêt de cette transformation est de présenter la propriété suivante :

$$d(X) \cdot d(Y) \cdot \mathbf{1} = d(X) \cdot Y = \begin{pmatrix} X_1 * Y_1 \\ \vdots \\ X_n * Y_n \end{pmatrix}$$

Dès lors l'équation de l'état quasi-stationnaire peut s'écrire sous forme matricielle :

Eq. 12	$C(D).P-d(P).C(D).P-e.P=K_{covt}$
--------	-----------------------------------

Avec  $(C(D).P)_i = \sum_j c(d_{i,j}) \bar{p}_j(t)$  (vecteur colonne)

### 3.4. Notion de stationnarité spatiale – Linéarisation de l'équation de dynamique

L'hypothèse de stationnarité spatiale, dite hypothèse de stationnarité au second ordre, est un élément fondamental de l'analyse spatiale des variables aléatoires en écologie, car elle permet une approche analytique statistique simplifiée de la structure spatiale. Elle repose sur les conditions suivantes :

- L'espérance de  $f$  est constante dans l'espace.
- La covariance spatiale de  $f$  est indépendante de la position  $(x, y)$  :

$$Cov_d^f(x, y) = \frac{1}{n_{d_{y,kl}=d}} \sum_{d_{y,kl}=d} (f_{xy} - \bar{f}) \cdot (f_{kl} - \bar{f}) = K_d^f(x, y)$$

Nous examinons ici l'effet de cette hypothèse sur l'expression analytique de la dynamique métapopulationnelle. Sans perte de généralité, on peut écrire :

$$\bar{p}_{xy} \cdot \bar{p}_{kl} = (\bar{p}_{xy} - \bar{\bar{p}}) \cdot (\bar{p}_{kl} - \bar{\bar{p}}) - \bar{\bar{p}}^2 + \bar{\bar{p}}(\bar{p}_{xy} + \bar{p}_{kl})$$

Or 
$$Cov_d(x, y) = \frac{1}{n_{d_{y,kl}=d}} \sum_{d_{y,kl}=d} (\bar{p}_{xy} - \bar{\bar{p}}) \cdot (\bar{p}_{kl} - \bar{\bar{p}}) = K_d^{\bar{p}}(x, y)$$

Considérant la deuxième condition de l'hypothèse de stationnarité pour  $p$ , on aurait :

$$K_d^{\bar{p}}(x, y) = Cste(x, y) = K_d^{\bar{p}} \quad (\text{covariance spatiale des moyennes temporelles})$$

On considérera de même la covariance spatiale instantanée :

$$K_d^p(x, y) = \frac{1}{n_{d_{y,kl}=d}} \sum_{d_{y,kl}=d} (p_{xy} - \hat{p}) \cdot (p_{kl} - \hat{p}) = Cste(x, y) = K_d^p$$

Et on notera :  $K_{cov d}^p = \sum_d [\bar{c}(d) \cdot n_d \cdot K_d^p]$  ainsi que  $K_{cov d}^{\bar{p}} = \sum_d [\bar{c}(d) \cdot n_d \cdot K_d^{\bar{p}}]$

On a dès lors :

$$\sum_{k=0, l=0}^{N-1} [\bar{c}(d_{xy,kl}) \cdot \bar{p}'_{kl} \cdot \bar{p}_{xy}] = \sum_d [\bar{c}(d) \cdot n_d \cdot K_d^{\bar{p}}] - C_t \cdot \bar{\bar{p}}^2 + C_t \cdot \bar{\bar{p}} \cdot \bar{p}_{xy} + \bar{\bar{p}} \cdot \bar{c} \otimes \bar{p}(x, y)$$

$$\sum_{k=0, l=0}^{N-1} [\bar{c}(d_{xy,kl}) \cdot \bar{p}'_{kl} \cdot \bar{p}_{xy}] = K_{cov d}^{\bar{p}} - C_t \cdot \bar{\bar{p}}^2 + C_t \cdot \bar{\bar{p}} \cdot \bar{p}_{xy} + \bar{\bar{p}} \cdot \bar{c} \otimes \bar{p}(x, y)$$

L'Eq. 11 peut alors être réécrite de la façon suivante :

$$\bar{p} \otimes \bar{c}'(x, y) + C_t \cdot \bar{p}^2 - C_t \bar{p} \bar{p}_{xy} - \bar{p} \cdot \bar{c} \otimes \bar{p}(x, y) - e \cdot \bar{p}_{xy} = K_{cov t} + K_{cov d}^{\bar{p}}$$

Eq. 13  $(1 - \bar{p}) \cdot \bar{p} \otimes \bar{c}'(x, y) + C_t \cdot \bar{p} \cdot (\bar{p} - \bar{p}_{xy}) - e \cdot \bar{p}_{xy} = K_{cov t} + K_{cov d}^{\bar{p}}$

Supposons à présent que le taux d'extinction est constant dans l'espace :

$$\begin{aligned} \frac{\sum_{xy} (p_{xy}(t+1) - p_{xy}(t))}{n_{xy}} &= \frac{\sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl} \cdot [1 - p_{xy}(t)])}{n_{xy}} - e \cdot \hat{p}(t) \\ \hat{p}(t+1) - \hat{p}(t) &= \frac{\sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl})}{n_{xy}} - \frac{\sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl} \cdot p_{xy})}{n_{xy}} - e \cdot \hat{p}(t) \\ \hat{p}(t+1) - \hat{p}(t) &= C_t \cdot \hat{p} - \frac{\sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl} \cdot p_{xy})}{n_{xy}} - e \cdot \hat{p}(t) \end{aligned}$$

$$\text{Soit } K_{cov d}^p(t) = \sum_{kl} (c(d_{xy,kl}) \cdot (p_{kl} - \hat{p}_{kl}) \cdot (p_{xy} - \hat{p}_{xy}))$$

$$\text{et } K_{cov d}^{\bar{p}}(t) = \sum_{kl} (c(d_{xy,kl}) \cdot (\bar{p}_{kl} - \bar{p}_{kl}) \cdot (\bar{p}_{xy} - \bar{p}_{xy}))$$

$$\text{On a } \sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl} \cdot p_{xy}) = \sum_{xy} (K_{cov d}^p(t) - C_t \cdot \hat{p}^2 + C_t \cdot \hat{p} \cdot p_{xy} + \hat{p} \cdot c \otimes p(x, y))$$

$$\text{Soit } \frac{\sum_{xy} \sum_{kl} (c(d_{xy,kl}) \cdot p_{kl} \cdot p_{xy})}{n_{xy}} = K_{cov d}^p + C_t \cdot \hat{p}^2$$

$$\hat{p}(t+1) - \hat{p}(t) = C_t \cdot \hat{p} - K_{cov d}^p - C_t \cdot \hat{p}^2 - e \cdot \hat{p}(t)$$

A l'état stationnaire dans le temps, on a alors:

$$0 = C_t \cdot \bar{p} - K_{cov t} - K_{cov d}^{\bar{p}} + C_t \cdot Var_t(\hat{p}) - C_t \cdot (Var_t(\hat{p}) + \bar{p}^2) - e \cdot \bar{p}$$

$$K_{cov t} = C_t \cdot \bar{p} - K_{cov d}^{\bar{p}} - C_t \cdot \bar{p}^2 - e \cdot \bar{p}$$

Et l'Eq. 13 devient :

$$(1 - \bar{p}) \cdot \bar{p} \otimes \bar{c}'(x, y) + C_t \cdot \bar{p} \cdot (\bar{p} - \bar{p}_{xy}) - e \cdot \bar{p}_{xy} = C_t \cdot \bar{p} - C_t \cdot \bar{p}^2 - e \cdot \bar{p}$$

En se souvenant de l'expression du produit de convolution, on peut écrire également :

$$(1 - \bar{p}) \cdot \bar{p} \otimes \bar{c}'(x, y) + C_t \cdot \bar{p} \cdot (\bar{p} - \bar{p}_{xy}) - e \cdot \bar{p}_{xy} = \bar{p} \otimes \bar{c}'(x, y) \cdot (1 - \bar{p}) - e \cdot \bar{p}$$

Eq. 14	$(1 - \bar{\bar{p}}).(\bar{p}_{xy} - \bar{\bar{p}}) \otimes \bar{c}(x, y) - C_t.\bar{\bar{p}}.(\bar{p}_{xy} - \bar{\bar{p}}) - e.(\bar{p}_{xy} - \bar{\bar{p}}) = 0$ pour $e = Cste(x, y)$
--------	--

Changement de variable (centrage spatial) :  $f_{xy} = \bar{p}_{xy} - \bar{\bar{p}}$

$$(1 - \bar{\bar{p}}).f_{xy} \otimes \bar{c}(x, y) - C_t.\bar{\bar{p}}.f_{xy} - e.f_{xy} = 0$$

Le recours à l'hypothèse de stationnarité spatiale a été motivé dans mon travail analytique par le souhait d'aboutir à une version linéarisée de l'équation de dynamique métapopulationnelle. L'Eq. 14 est l'aboutissement de cette logique, la dynamique métapopulationnelle est linéaire et exprimable par une simple équation matricielle :

$$M.f_{xy} = 0 \text{ avec } M = (1 - \bar{\bar{p}}).C(D) - (C_t.\bar{\bar{p}} + e).Id \quad (\text{voir calcul de l'Eq. 12})$$

Le noyau de la matrice  $M$  (vecteurs propres dont la valeur associée est 0) détermine le(s) état(s) quasi-stationnaire(s) du système. Si le noyau est de dimension 0, alors seule la solution  $f_{xy} = 0$  existe, et la distribution des probabilités d'occupation est spatialement uniforme.

Cette approximation est donc utile et intéressante, mais son intérêt se limite à la modélisation d'un habitat non structuré dans l'espace. Si l'habitat n'est pas spatialement structuré, il y a par construction une invariance de position au sein du système : le contexte local d'habitat est le même en tout point de l'espace. L'existence d'une hétérogénéité d'habitat ne permet pas de maintenir cette hypothèse sur la probabilité *absolue* d'occupation des cellules d'habitat.

### 3.5. Prise en compte de la structure de l'habitat

Afin de prendre en compte explicitement l'existence d'une variation de la qualité de l'habitat, structurée dans l'espace, je propose d'introduire un taux de colonisation variable, dépendant de la qualité d'habitat de la cellule colonisée. Par souci de simplification, l'extinction ne dépend pas de la qualité d'habitat, et on suppose ainsi en première approximation que la colonisation dépendante de l'habitat suffit à réguler le système de manière écologiquement réaliste.



Nous reprenons donc l'Eq. 11, et écrivons :

$$\bar{p} \otimes \bar{c}(x, y) \cdot [1 - \bar{p}_{xy}] \cdot \sum_i \alpha_i \cdot h_i(x, y) - e \cdot \bar{p}_{xy} = K_{cov_t}(x, y)$$

où les  $h_i$  représentent différentes composantes de l'habitat, et  $\alpha_i$  les coefficients de qualité correspondants pour l'espèce considérée. Dans un tel modèle, les coefficients  $\alpha_i$  représentent une caractérisation en  $i$  dimensions de la niche fondamentale de l'espèce, pouvant inclure l'existence d'une variation stochastique de l'habitat.

Pour résumer la qualité globale d'habitat, on écrit :

$$h(x, y) = \sum_i \alpha_i \cdot h_i(x, y)$$

En première approximation, on peut négliger le terme de covariance temporelle, d'où :

Eq. 15 $\bar{p} \otimes \bar{c}(x, y) \cdot [1 - \bar{p}_{xy}] \cdot h(x, y) - e \cdot \bar{p}_{xy} = 0$
--

Et l'équation matricielle associée sera :

$$\text{Eq. 16} \quad d(H) \cdot (Id - d(P)) \cdot C(D) \cdot P - e \cdot P = 0$$

### 3.6. Modèle analytique spectral

Nous avons montré que l'analyse spectrale fournit une mesure spatiale pertinente pour caractériser la dynamique spatiale des métapopulations à l'état quasi-stationnaire. Suivant le principe de dualité, il existe une correspondance entre la représentation spatiale de positions  $(x, y)$ , et la représentation spectrale de fréquences  $(u, v)$ .

Cette correspondance peut être décrite analytiquement :

- Dans le contexte de l'analyse spectrale de Fourier, la transformation de Fourier est une transformation linéaire qui permet de passer de l'espace réel à l'espace spectral:

$$\begin{matrix} f(x, y) \rightarrow F(u, v) \\ \bar{p}_{xy} \rightarrow P_{uv} \end{matrix}$$

$u$  et  $v$  représentent les fréquences spatiales associées aux deux dimensions de l'espace réel ; ces dernières sont ici exprimées par les coordonnées  $x$  et  $y$ . Il est également possible de définir la transformée de Fourier dans un espace à une ou trois dimensions.

La transformation de Fourier s'écrit :

$$F(u, v) = \mathfrak{F}[f(x, y)] = \frac{1}{N-1} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(n, m) e^{i \frac{n}{N} u} e^{i \frac{m}{N} v}$$

... et la transformée inverse :

$$f(x, y) = \mathfrak{F}^{-1}[F(u, v)] = \frac{1}{N-1} \sum_{r=0}^{N-1} \sum_{s=0}^{N-1} F(r, s) e^{i \frac{r}{N} x} e^{i \frac{s}{N} y}$$

La transformée de Fourier présente entre autres propriétés notables :

$$\mathfrak{F}[f.g] = \mathfrak{F}[f] \otimes \mathfrak{F}[g]$$

$$\mathfrak{F}[f \otimes g] = \mathfrak{F}[f] \mathfrak{F}[g]$$

Dans le cas particulier d'une fonction constante, on a simplement :

$$\begin{matrix} f(x, y) \rightarrow F(u, v) \\ K \rightarrow K.\delta(u, v) \end{matrix}$$

La fonction de Dirac est nulle en tout point sauf en 0, où elle vaut 1.

Dans le domaine spectral, l'Eq. 11 devient donc :

$$\text{Eq. 17} \quad P.C - (P.C) \otimes P - e.P = K_{\text{covt}}(u, v)$$

- Nous avons vu que le principe de dualité peut également être exprimé dans le cadre de l'analyse spectrale PCNM :

Espace réel

Espace spectral

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$S_X = \begin{pmatrix} S_{X1} \\ \vdots \\ S_{Xp} \end{pmatrix}$$

Avec :

$$X_i = \sum_{j=1}^p m_{ij} . S_{X_j} = (M.S)_j$$

Rappelons maintenant la transformation spectrale PCNM :  $P = M.S_P$

L'Eq. 12 permet alors d'écrire :

$$\text{Eq. 18} \quad {}^t M . d(1 - P) . C(D) . P - e . S_P = {}^t M . K_{\text{covt}}$$

Nous pouvons donc formuler un attendu analytique sur la relation entre le spectre des probabilités d'occupation, estimé à l'état quasi-stationnaire sur une gamme de temps donnée,

et les paramètres de colonisation-extinction de la dynamique métapopulationnelle. Cet attendu peut être aussi bien formulé dans l'espace de Fourier que dans l'espace PCNM.

De la même manière, nous nous appuyons sur l'Eq. 13 pour écrire les équations duales du domaine spectral quand il existe un habitat non uniforme  $H$  :

- Espace de Fourier  $(P.C) \otimes (\mu.\delta + H) - (P.C) \otimes P \otimes (\mu.\delta + H) - \mu.P = 0$
- Espace PCNM :  $'M.d((\mu Id + H).(1 - P)).C(D).P - \mu.S_p = 0$

### 3.7. Perspectives ouvertes

Nous avons montré dans le Chapitre III l'intérêt de l'analyse spectrale comme représentation pertinente des structures spatiales émergentes. L'utilisation d'un modèle analytique dans le domaine spectral ouvre donc des perspectives intéressantes pour la compréhension de ces structures.

J'ai travaillé sur le traitement analytique de l'équation spectrale de dynamique métapopulationnelle (résultats non montrés). La complexité et la connotation très abstraite de ce traitement entravent une compréhension simple de la relation entre processus métapopulationnel et structures émergentes. Nous laissons donc ce traitement comme une perspective future, et privilégions dans la suite du travail l'étude du modèle analytique dans le domaine réel, plus approprié pour une compréhension simple des phénomènes en jeu.

## 4. Inférence de la dynamique métapopulationnelle

L'objectif fondamental de mon travail est d'accéder à la caractérisation des paramètres dynamiques de colonisation/extinction à partir de la distribution spatiale des populations et la distribution spatiale de différentes variables d'habitat. Deux cas de figure sont envisagés :

- On ne connaît que l'occupation binaire présence/absence des populations à un instant donné. C'est un cas problématique ne permettant pas de caractériser  $p$ . On peut en revanche accéder à  $\hat{p}(t)$  (espérance dans l'espace de la probabilité d'occupation à la génération  $t$ ), ou mesurer une espérance au temps  $t$  sur un ensemble de mailles agrégées localement  $\hat{p}_{loc}(t, v, w)$ , en changeant l'échelle d'échantillonnage (changement de coordonnées  $(x, y) \rightarrow (v, w)$ ).

- On connaît l'occupation locale des populations sur une séquence de générations, à un état supposé quasi-stationnaire, ce qui permet d'évaluer  $\bar{p}_{xy}$  et  $\bar{p}$ .

Plusieurs méthodes d'inférences seront discutées et testées grâce aux métapopulations simulées en habitat structuré, avec un habitat binaire (Chapitre I). L'Eq. 11 est notre relation de base (dans le domaine réel ou spectral) pour définir l'état quasi-stationnaire des métapopulations. Nous avons considéré trois approches d'inférence afin de répondre à différents objectifs :

- Méthode 1 : *Inférence directe globale sur les probabilités d'occupation*. Rechercher la meilleure adéquation des deux termes de l'équation au sens des moindres carrés.
- Méthode 2 : *Inférence directe locale sur les probabilités d'occupation*. Utiliser une estimation directe algébrique du paramètre  $r$ , dans l'esprit des équations Eq. 12 et Eq. 16.
- Méthode 3 : *Inférence indirecte sur les probabilités d'occupation*. Réaliser une régression de chaque membre de l'équation sur les variables d'habitat et déduire simultanément la caractérisation de la niche fondamentale et le paramètre de niche  $r$ .

Les deux premières méthodes impliquent de déterminer séparément l'habitat fondamental et le(s) paramètre(s) d'extinction-colonisation. La méthode 3 appréhende une estimation simultanée de ces deux composantes de la niche. Nous discutons ces trois méthodes au regard de leur qualité d'estimation, de leur robustesse, et de leur adéquation aux contraintes des données réelles.

#### 4.1. Méthode 1 : inférence globale à partir des probabilités locales moyennées dans le temps

Nous supposons dans un premier temps que la covariance temporelle des probabilités locales d'occupation est nulle dans l'espace, et que la probabilité de colonisation est dépendante de la qualité de la cellule à coloniser.

Reprenons donc l'Eq. 15 :

$$\bar{p} \otimes \bar{c}(x, y) \cdot [1 - \bar{p}_{xy}] \cdot h(x, y) - e \cdot \bar{p}_{xy} = 0$$

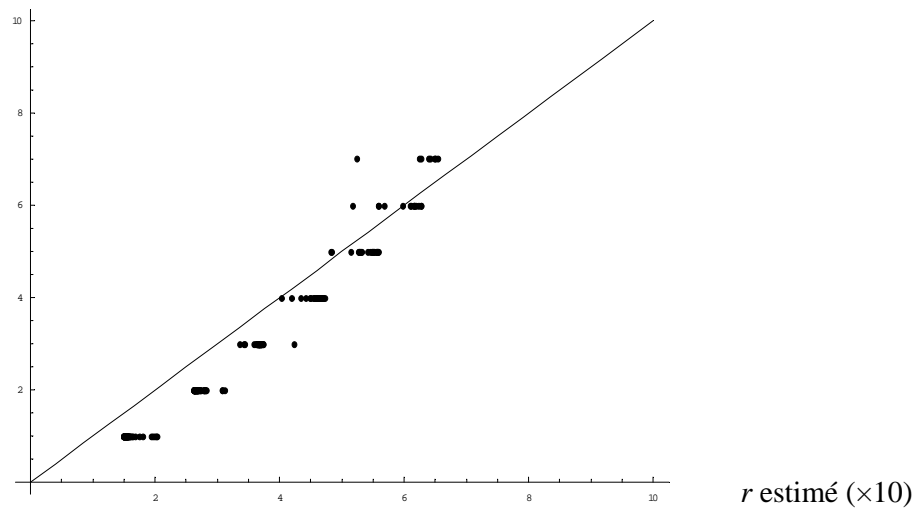
et sa contrepartie matricielle (Eq. 16) :  $d(H) \cdot (Id - d(P)) \cdot C(D) \cdot P - e \cdot P = 0$

La fonction de colonisation, sans prendre en compte le facteur dépendance d'habitat, est « normalisée » à 1, en divisant par le taux de colonisation global  $C_i$ .  $r = e / C_i$  est le paramètre

de dynamique métapopulationnelle. On ajuste donc aux moindres carrés en faisant varier le paramètre  $r$  : il s'agit de minimiser l'expression  $\sum_{x,y} \left( \bar{p} \otimes \frac{\bar{c}(x,y)}{C_t} \cdot [1 - \bar{p}_{xy}] \cdot h(x,y) - r \cdot \bar{p}_{xy} \right)^2$

Voici les résultats obtenus en appliquant cette méthode aux distributions quasi-stationnaires des métapopulations simulées en habitat binaire. La Figure 16a montre l'adéquation des valeurs du paramètre  $r$  attendues et estimées. Les valeurs attendues correspondent aux valeurs du paramètre  $r$  utilisées pour les simulations ; les valeurs estimées sont donc celles obtenues par ajustement du modèle inférentiel.

(a)  $r$  attendu ( $\times 10$ )



(b)  $r$  attendu (x 10)

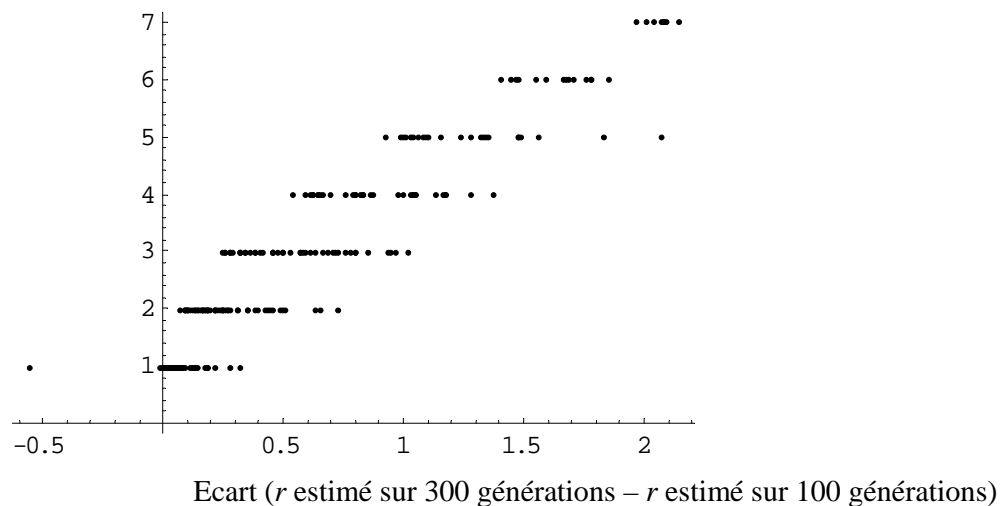


Figure 16 : Ajustement aux moindres carrés du paramètre  $r$  pour les métapopulations simulées en habitat binaire, à partir de l'équation  $\bar{p} \otimes \frac{\bar{c}(x,y)}{C_t} \cdot [1 - \bar{p}_{xy}] \cdot h(x,y) - r \cdot \bar{p}_{xy} = 0$

(a) Comparaison des valeurs attendues et estimées (sur 300 générations).

(b) Ecart d'estimation de  $r$  avec des probabilités estimées sur 100 générations et sur 300 générations (ordonnées), en fonction des valeurs attendues (abscisses).

Les valeurs estimées de  $r$  sont en général supérieures aux valeurs attendues, sauf quand on approche du seuil critique d'extinction (valeurs les plus élevées). Il ne faut pas voir en cela un effet de bord ou de structure de l'habitat, susceptible de réduire localement la capacité de colonisation effective d'une cellule donnée. En effet, cette variation spatiale de la colonisation effective est explicitement formulée par le terme de colonisation de l'Eq. 11 (produit de convolution). Nous avons toutefois vérifié que la correction de la valeur attendue de  $C_t$  par le taux de colonisation effectif réel (moyenne calculée sur l'ensemble du paysage) n'affecte pas la nature du biais de prédiction de  $r$ .

Nous observons par ailleurs une divergence de l'estimation par rapport aux valeurs attendues les plus élevées (à l'approche du seuil critique d'extinction). Cette divergence est plus rapide et plus importante quand les probabilités locales d'occupation sont estimées sur un nombre moindre de générations. La Figure 16b montre ainsi l'écart des valeurs prédites pour  $r$  avec des probabilités d'occupation estimées sur 100 et 300 générations : il existe une divergence plus rapide et plus forte de l'estimation quand la valeur attendue de  $r$  augmente, et quand le nombre de générations utilisées est plus faible.

Cette divergence est liée à l'existence du seuil critique. L'intensification de l'auto-organisation au voisinage de ce seuil induit l'existence de longueurs de corrélation spatiale et temporelle considérablement accrues. Dans ces conditions, l'estimation de  $r$  est influencée par la mauvaise estimation des probabilités locales : plus on est proche du seuil critique, plus le nombre de générations nécessaire à une estimation correcte de ces probabilités est important. En outre, le fait que l'on estime alors une valeur de  $r$  « meilleure » que la situation réelle, c'est à dire avec un taux d'extinction relatif plus faible, donne l'illusion que la dynamique de l'espèce est plus stable que la réalité : c'est le même type de processus qui est impliqué, lorsqu'on passe au-delà du seuil de survie de la métapopulation, dans la question de la *dette d'extinction*. (voir Chap. II §2)

Il est important de souligner qu'un tel biais d'estimation est une source majeure d'erreur dans le processus d'inférence, indépendamment des hypothèses liées à la modélisation proprement dite des processus. La discussion des prédictions obtenues par *Pair Approximation* pour la probabilité locale d'occupation mérite d'être envisagée dans cette perspective : contrairement à ce qui est couramment admis, l'hypothèse de l'approximation par paires pourrait être moins préjudiciable que l'effet de l'auto-organisation. Cette interprétation a déjà été évoquée par Snyder et Nisbet (2000), mais la qualité des résultats obtenus sur la prédiction des probabilités de paires montre que l'erreur d'estimation est atténuée quand on considère des grandeurs spatiales non locales.

En définitive, c'est le problème des échelles de temps et d'espace qui est posé. Que peut-on dire en effet du comportement de la métapopulation au voisinage du seuil critique, quand les distances de corrélation dans le temps et dans l'espace deviennent très grandes ? Le choix d'une amplitude spatiale et temporelle réduite empêche la possibilité d'inférer correctement les paramètres dynamiques au voisinage du seuil critique.

Nous allons voir que les deux autres méthodes d'inférence étudiées ci-après présentent une réponse différente à ce phénomène de criticalité. Le choix de la méthode d'inférence pourra alors être éclairé par la compréhension de l'importance écologique de ces différences.

#### *4.2. Méthode 2 : inférence locale à partir des probabilités moyennées dans le temps*

Nous ne modifions pas ici la nature de l'attendu analytique : covariance temporelle supposée nulle, terme de convolution calculé avec normalisation du noyau de colonisation. En revanche, la démarche d'inférence est différente : à une estimation de  $r$  unique sur l'ensemble du paysage par la méthode des moindres carrés, nous substituons ici une estimation ponctuelle de  $r$  en chaque point  $(x, y)$  de l'espace.

Cette estimation est donc basée sur la détermination locale de l'équilibre entre les événements de colonisation et d'extinction à l'état quasi-stationnaire. L'enveloppe des estimations de  $r$  sur l'ensemble de l'espace constitue un intervalle de confiance permettant de prendre en compte la sensibilité de cette estimation à la variation spatiale du contexte local d'habitat.

L'équation  $d(H).(Id - d(P)).C(D).P - e.P = 0$  permet d'expliciter simplement ce mode d'estimation. Elle devient  $d(H).(Id - d(P)).\frac{C(D)}{C_i}.P = r.P$ . Cela signifie en particulier que

$r$  est la valeur propre associée au vecteur propre  $P$  de la matrice  $d(H).(Id - d(P)).\frac{C(D)}{C_i}$ .

L'équation  $d(H).(Id - d(P)).\frac{C(D)}{C_i}.P = r.P$  donnera autant d'estimations locales de  $r$

qu'il y a de points où la probabilité d'occupation est non nulle. Il est important de souligner que le nombre de valeurs estimées, et donc la qualité de l'estimation, dépendent logiquement de l'amplitude de l'habitat *a posteriori*.

Nous testons cette méthode d'ajustement sur les distributions quasi-stationnaires des métapopulations simulées en habitat binaire (grilles 100×100). Deux échantillonnages de 2500 points sont considérés : l'un est constitué d'une grille régulière carrée, l'autre d'une distribution aléatoire uniforme dans l'espace. Les résultats obtenus dans les deux cas sont analogues. Le résultat obtenu par inférence locale apparaît différer nettement de l'estimation globale aux moindres carrés présentée au paragraphe précédent.

Nous obtenons en premier lieu une information nouvelle : celle de l'intervalle de confiance d'estimation. Il apparaît que l'amplitude de cet intervalle, en valeur absolue ou relative au centre, est plus faible quand le nombre de cellules permettant l'estimation augmente (amplitude de l'habitat *a posteriori*). L'amplitude de l'intervalle de confiance reflète essentiellement un effet d'échantillonnage. Les résultats montrés à la Figure 17 soulignent une bonne linéarité de la relation entre valeurs estimées et attendues, avec cependant un biais systématique représenté par une pente inférieure à 1.



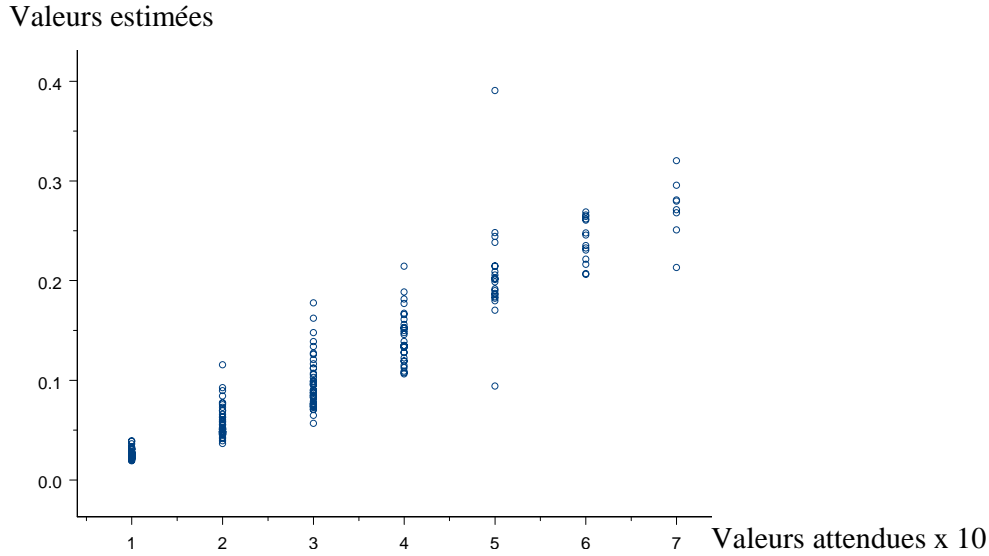


Figure 17 : Valeurs moyennes de  $r$  prédites par la méthode d'estimation locale à partir de l'estimation de  $\bar{p}_{xy}$  et l'équation  $d(H).(Id - d(P)).\frac{C(D)}{C_t}.P = r.P$ .

### Approximation locale

Etant donnée l'existence d'un processus de colonisation localisé dans l'espace et décroissant avec la distance, il existe une structure de corrélation locale entre l'état d'occupation d'une cellule et l'état de ses voisins. En première approximation, nous considérons que le voisinage d'une cellule possède une probabilité d'occupation constante, au moins à l'échelle du noyau de colonisation, en l'occurrence 5 voisinages. Dans ses conditions, nous pouvons réécrire l'équation quasi-stationnaire de la manière suivante :

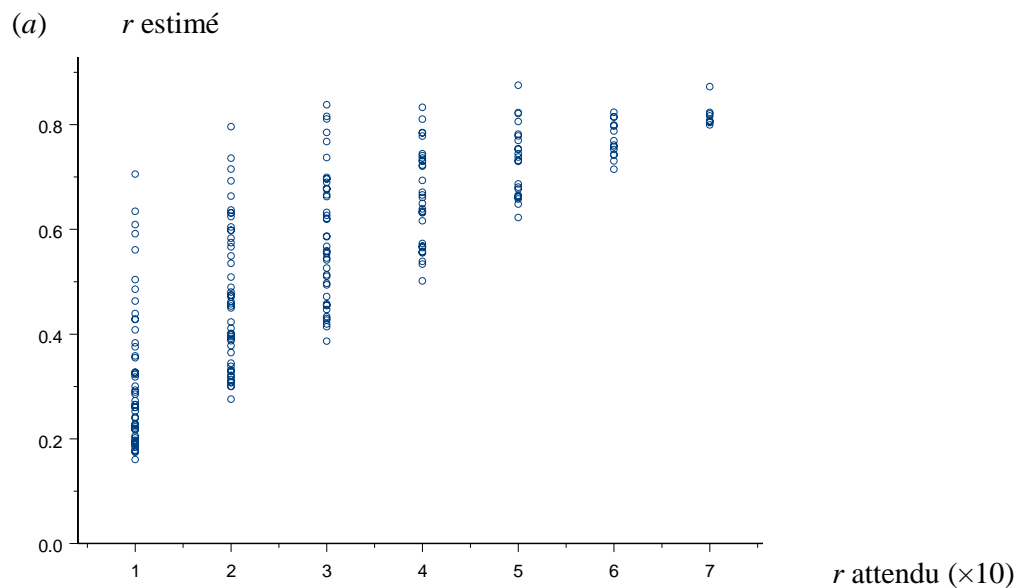
$$d(H).(Id - d(P)).\frac{C(D)}{C_t}.P = r.P \quad \text{devient} \quad d(H).(Id - d(P)).P = r.P$$

Les matrices  $d(H).(Id - d(P))$  et  $r.Id$  étant diagonales, cela implique d'identifier leurs termes diagonaux et de procéder ainsi à l'estimation locale de  $r$  simplement en posant :  $r \approx h(x, y).(1 - \bar{p}_{xy})$  pour  $\bar{p}_{xy} \neq 0$ . En habitat binaire, cela revient à traiter l'équation  $r \approx 1 - \bar{p}_{xy}$  en chaque point de l'habitat *a posteriori*.

On retrouve ainsi localement les conditions de l'équation de Levins. Cette approximation néglige de fait l'effet de la structure locale de l'habitat, et donc surestime la disponibilité des colons dans le voisinage. En conséquence, par rapport à ce qui se passe en situation réelle avec les mêmes paramètres  $e$  et  $C_t$ , la valeur théorique de  $\bar{p}_{xy}$  sera surestimée par le modèle.

Dans une logique inverse d'inférence, pour une donnée réelle  $\bar{p}_{xy}$ , le  $r$  estimé avec l'approximation locale sera donc sur-estimé. C'est en effet ce que l'on observe à la Figure 18a.

Avec une telle approximation, on constate alors une mauvaise adéquation générale entre les valeurs réelles et estimées de  $r$  (Figure 18a). Cependant, pour une carte d'habitat fixée (couple  $(p, q)$  donné), le biais de prise en compte de la structure d'habitat sera logiquement le même quand on fera varier  $r$  : pour chaque carte d'habitat, on observe ainsi une relation effectivement linéaire entre valeurs attendues et estimées de  $r$  (Figure 18b). Il se produit en fait un plafonnement des valeurs estimées quand on approche du seuil d'extinction. Cet effet de plafonnement disparaît quand la valeur estimée est corrigée par la valeur du seuil critique d'extinction (Figure 18c). Rappelons que ce seuil représente la capacité topologique d'un paysage de permettre le maintien d'une population (Ovaskainen et al. 2002). Ce que mesure  $r$  dans ce cas, c'est un taux de saturation efficace du paysage d'habitat par un organisme, indépendamment de la capacité intrinsèque d'occupation qui résulte de sa topologie.



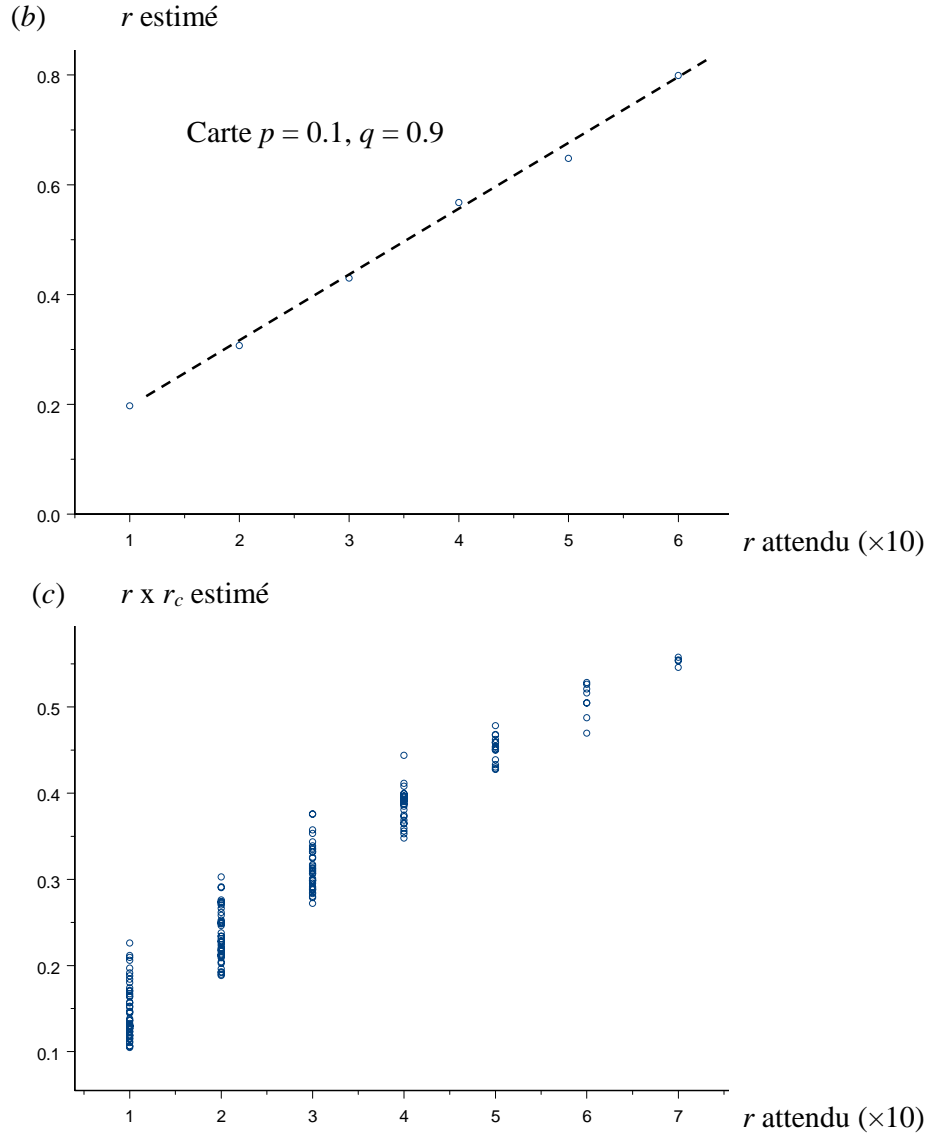


Figure 18 : Inférence locale du paramètre  $r$  sur les métapopulations simulées en habitat binaire à partir de l'estimation de  $\bar{p}_{xy}$  et l'équation  $r \approx h(x, y) \cdot (1 - \bar{p}_{xy})$ , en chaque point de l'habitat *a posteriori*. La figure *a* représente en ordonnées les valeurs de  $r$  estimées (centre des intervalles de confiance) en fonction des valeurs attendues (paramétrage des simulations). Pour une carte donnée (Fig. *b*,  $p = 0.1, q = 0.9$ ), la relation entre les valeurs estimées et attendues est linéaire. On constate cependant (Fig. *a*) que les valeurs estimées plafonnent à une valeur très inférieure aux valeurs attendues. La figure *c* montre la relation entre le produit des valeurs prédites par le seuil d'extinction en ordonnées, et les valeurs  $r$  attendues en abscisses. La linéarité entre valeurs estimées et attendues est ainsi restaurée. Il existe donc une incidence de la distance au seuil critique sur l'estimation locale de  $r$  quand on utilise l'approximation d'homogénéité locale de l'occupation.

#### 4.3. Méthode 2 : inférence locale à partir des occurrences instantanées agrégées spatialement

Dans le paragraphe précédent, nos estimations de  $r$  reposaient sur la connaissance par estimation des probabilités  $\bar{p}_{xy}$ . Cela suppose de posséder plusieurs générations de mesures de présence/absence des populations en chaque point d'échantillonnage. Cela représentera souvent une véritable gageure, car il s'agit du temps de génération des populations et non des individus.

Une solution est de procéder à l'agrégation spatiale des données d'occupation des cellules d'habitat. Dans le cas de nos simulations en habitat binaire, nous possédons au départ un maillage régulier de cellules d'habitat de dimension  $100 \times 100$ , et nous mesurons l'état d'occupation de chaque cellule. A une génération donnée de l'état quasi-stationnaire, nous proposons d'agréger ce maillage de manière à obtenir un nouveau maillage  $25 \times 25$  : chaque nouvelle maille  $(v, w)$  correspond à 16 mailles anciennes  $(x, y)$ . Nous estimons donc  $\hat{p}_{loc}(t, v, w)$  en chaque nouvelle maille  $(v, w)$  à partir de 16 données de présence-absence.

Le fait d'agréger les cellules induit une perte d'information sur la structure locale des populations, et donc sur les processus de dissémination à l'intérieur des mailles agrégées. Dans la mesure où l'on suppose que les mailles agrégées sont suffisamment petites pour présenter un niveau élevé de corrélation entre l'état des populations à l'intérieur de la maille, il est raisonnable de formuler l'hypothèse d'approximation locale que nous avons présentée dans le paragraphe précédent.

D'un autre côté, il est également possible de considérer que les mailles agrégées représentent un autre niveau d'analyse, avec des populations plus grandes, et ainsi que les probabilités agrégées dans le nouveau maillage fournissent une estimation de la probabilité d'occupation quasi-stationnaire des nouvelles mailles, c'est à dire une estimation de  $\bar{p}_{loc}(v, w)$ . Cette hypothèse nous permet de pratiquer l'estimation locale de  $r$  suivant la même méthodologie que celle présentée dans le paragraphe précédent.

Nous utilisons dès lors deux méthodes d'estimations, avec les équations

$$d(H).(Id - d(P)).\frac{C(D)}{C_t}.P = r.P \quad \text{et} \quad d(H).(Id - d(P)).P = r.P \quad (\text{approximation locale})$$

(voir Figure 19).

C'est avec l'équation  $d(H).(Id - d(P)).P = r.P$ , sans correction par le seuil critique d'extinction, que nous obtenons la meilleure estimation (Figure 19b). L'approximation locale sur l'état des populations apparaît être judicieuse. L'agrégation des données d'occupation et d'habitat pose de fait un cadre nouveau. Dans le paragraphe précédent, *l'approximation locale* était utilisée pour réaliser une *estimation locale*. Ce faisant, la méthode était handicapée par une mauvaise prise en compte de l'effet du voisinage. Nous considérons à présent une *estimation à une échelle intermédiaire*, basée sur la même *approximation locale*. **Dans ce contexte, l'agrégation des données d'occupation et des données d'habitat prend implicitement en compte l'effet local de la structure d'habitat sur l'occurrence des populations.** C'est le transfert d'échelle qui est important. Et donc il est possible d'estimer correctement le turnover d'extinction/colonisation à l'intérieur de la maille d'agrégation.

On remarque ainsi que les estimations sont bonnes, avec une variation faible, pour des valeurs de  $r$  élevées c'est à dire dans les situations où les phénomènes dynamiques sont prépondérants et l'auto-organisation forte. Or ce sont aussi les situations les plus délicates à visualiser et analyser en terme de biologie de la conservation.

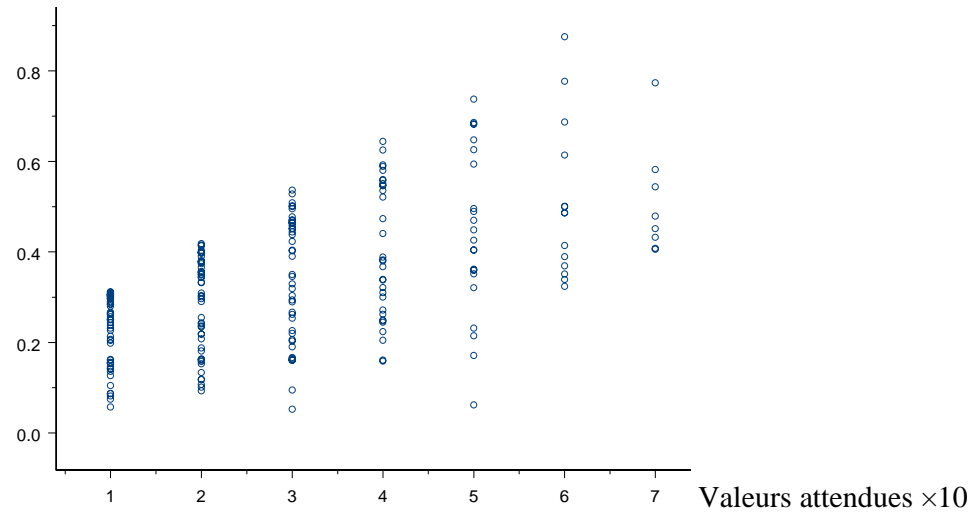
*A contrario*, l'équation  $d(H).(Id - d(P)).\frac{C(D)}{C_i}.P = r.P$  donne de mauvaises estimations

(Figure 19a). Cela peut trouver deux explications principales :

- Il n'est donc pas pertinent de considérer les probabilités  $\hat{p}_{loc}(t,v,w)$  comme de bons estimateurs de probabilités d'occupation quasi-stationnaires  $\bar{p}_{loc}(v,w)$ , pour des populations considérées à une échelle supérieure.
- La fonction de colonisation, décrivant les processus locaux, n'est pas adéquate pour représenter les flux de colonisation entre mailles agrégées. Il faudrait en ce cas la réécrire.

(a) Estimation de  $r$  à partir de l'équation  $d(H).d(1-P).\frac{C(D)}{C_t}.P = r.P$

Valeurs estimées



(b) Estimation de  $r$  à partir de l'équation  $d(H).d(1-P).P = r.P$

Valeurs estimées

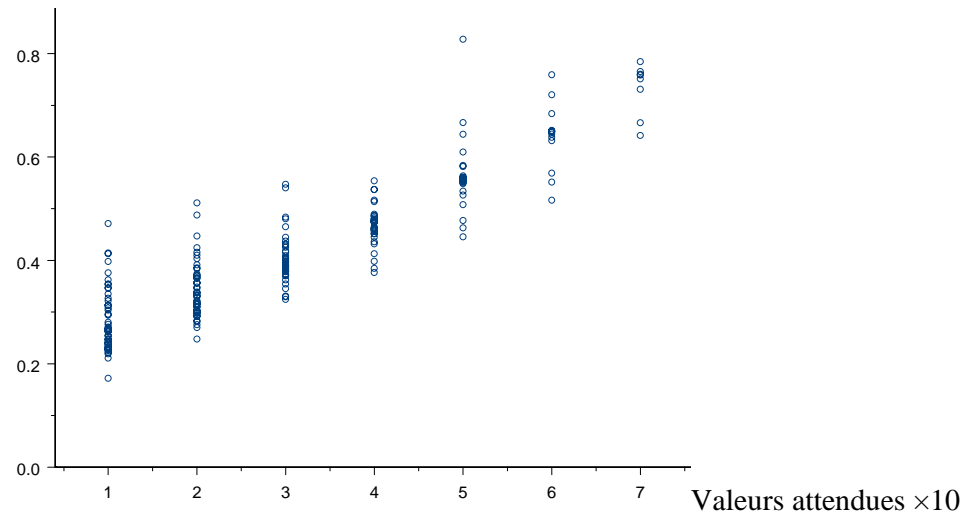


Figure 19 : Estimation du paramètre  $r$  à partir des cartes de présence absence agrégées dans l'espace, à une génération donnée de l'état quasi-stationnaire.

#### 4.4. Méthode 3 : préférences d'habitat et dynamique métapopulationnelle

Pour les méthodes 1 et 2, il est nécessaire d'avoir une connaissance préalable de l'habitat *a priori* de l'espèce. Or nous avons vu que cela représente un problème théorique et technique important. Avec notre troisième méthode d'inférence, nous partons du principe que l'habitat *a priori* est inconnu. Dans cette logique, le support analytique est l'équation locale

d'équilibre d'extinction-colonisation, avec une probabilité d'extinction dépendante de la favorabilité ou de la qualité d'habitat. La qualité de l'habitat est décrite par  $h(x,y)=\sum_i \alpha_i .h_i(x,y)$ .

Considérons un paysage binaire d'habitat. L'état quasi-stationnaire est décrit par l'équation  $\bar{p} \otimes \bar{c}(x,y).[1-\bar{p}_{xy}].h(x,y) = e.\bar{p}_{xy}$  ou encore  $\bar{p} \otimes \frac{\bar{c}}{C_t}(x,y).[1-\bar{p}_{xy}].h(x,y)=r.\bar{p}_{xy}$ . La variation de la nature de l'habitat *a priori* contraint la dynamique intrinsèque décrite par le paramètre  $r = \frac{e}{C_t}$ . Il est possible d'inférer indirectement et conjointement l'incidence de l'habitat et de la dynamique intrinsèque sur la distribution spatiale résultante, grâce à la procédure suivante :

- La régression linéaire de  $\bar{p}_{xy}$  sur les composantes d'habitat  $h_i$  donne les coefficients  $\beta_i$ , qui représentent une occupation spatiale moyenne de la  $i$ ème dimension d'habitat
- La régression linéaire de  $\bar{p} \otimes \frac{\bar{c}}{C_t}(x,y).[1-\bar{p}_{xy}].h_i(x,y)$  sur les composantes d'habitat  $h_j$  donne les coefficients  $\gamma_{ij}$ , qui représentent un taux de « tentatives de colonisation » de la  $i$ ème dimension d'habitat.
- Il existe finalement la relation suivante entre les coefficients  $\alpha_i, \beta_i$  et  $\gamma_{ij}$  :  $\sum_j \alpha_i . \gamma_{ij} = \beta_i$ .

Il est nécessaire que les composantes d'habitat  $h_i$  soient indépendantes les unes des autres, sinon la relation entre les coefficients des régressions risque d'être faussée. On considérera en général une transformation des données environnementales permettant d'obtenir une telle indépendance, par exemple une Analyse en Composantes Principales.

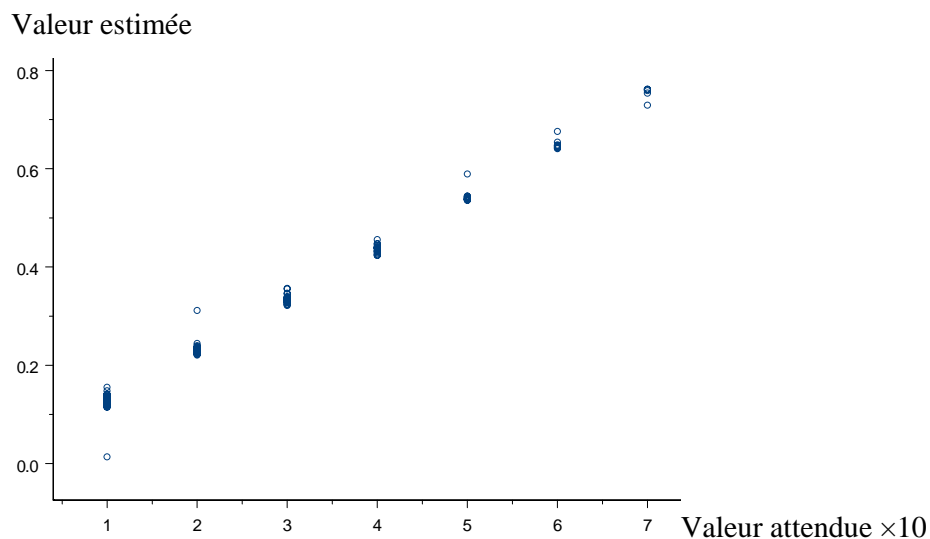
Afin d'évaluer la fiabilité de cette méthode, nous avons d'abord considéré une situation avec une variable d'habitat  $h(x,y)$  unique et binaire. L'habitat est donc connu, et il s'agit d'examiner la qualité d'estimation de  $r$ . Les données utilisées sont donc ici nos métapopulations simulées en habitat binaire, à l'état quasi-stationnaire, en échantillonnant 2500 points disposés régulièrement.

Le résultat que nous obtenons est très intéressant et révélateur (Figure 20a). Nous obtenons une excellente estimation du paramètre  $r$  quand l'habitat est faiblement dense ou au contraire très dense ( $p \leq 4$  ou ( $p = 7$  et  $q \geq 8$ ) ou  $p \geq 8$ , voir Figure 20). Il y a toute une gamme d'habitats « intermédiaires » ( $p = 5, 6$  ou ( $p = 7$  et  $q \leq 7$ )) dans leur structure ou au contraire les estimations sont extrêmement mauvaises. Il semble que cette méthode

d'estimation soit instable pour de telles configurations d'habitat, et qu'il soit alors impossible d'estimer simultanément l'habitat *a priori* et le paramètre de dynamique spatiale  $r$ . Cette procédure manque donc de robustesse.

A noter que cette méthode peut très aisément être mise en œuvre en utilisant la contrepartie spectrale de l'équation  $\bar{p} \otimes \bar{c}(x, y) \cdot [1 - \bar{p}_{xy}] \cdot h(x, y) = e \cdot \bar{p}_{xy}$ . Les régressions mettent alors en jeu les spectres et non les vecteurs bruts des probabilités d'occupation et d'habitat. Les résultats sont montrés à la Figure 20*b*. Nous constatons un comportement très analogue, avec un problème d'estimation aux valeurs intermédiaires de  $p$ , et une bonne estimation sinon. On constate cependant la présence de quelques mauvaises estimations. Il n'est pas anodin d'obtenir des qualités d'estimation globalement similaires avec des vecteurs de données réelles à 2500 composantes d'une part, et avec des vecteurs de spectres lissés à 50 composantes d'autre part : l'analyse spectrale concentre en effet l'information pertinente pour l'estimation du paramètre dynamique de l'habitat.

(a) Estimation dans le domaine réel





(b) Estimation dans le domaine spectral

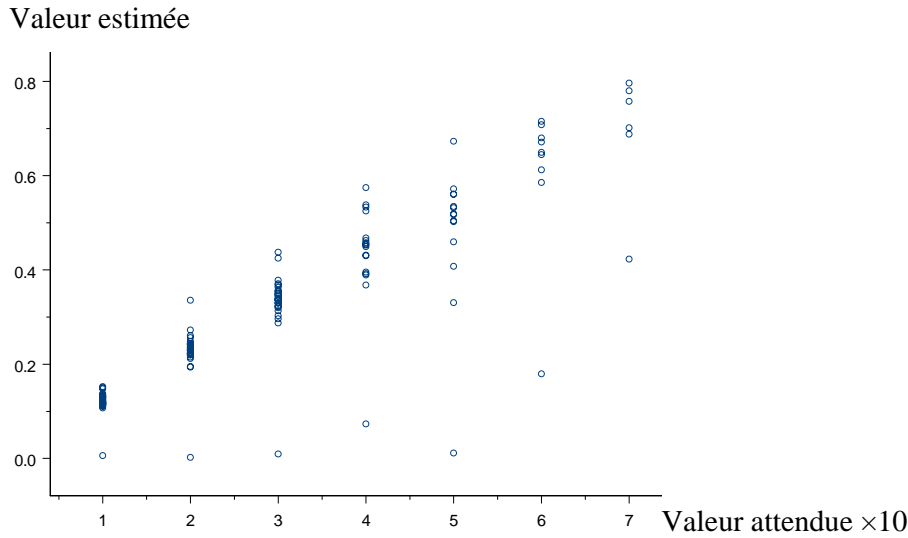


Figure 20 : Estimation de  $r$  avec la méthode d'estimation conjointe de l'habitat *a priori* et du paramètre  $r$  de dynamique spatiale. Les résultats montrés sont ceux obtenus pour une certaine gamme des paramètres d'habitat :  $p \leq 4$  ou ( $p = 7$  et  $q \geq 8$ ) ou  $p \geq 8$ . Cela représente 140 des 227 métapopulations simulées. (a) Résultat de l'estimation avec l'équation d'équilibre extinction-colonisation exprimée dans le domaine réel. (b) Résultat de l'estimation avec l'équation d'équilibre extinction-colonisation exprimée dans le domaine spectral.

4.5. Bilan : quel choix pour l'estimation à partir de données réelles ?

En définitive l'absence d'une bonne estimation avec la méthode 3 (§4.4), pour une large gamme de structures d'habitat, remet en question le principe d'estimation conjointe de la composante statique et dynamique de l'habitat. Il semble que dans de nombreuses situations, l'information contenue par la structure spatiale des métapopulations quasi-stationnaires, même moyennée dans le temps, ne permette pas de réaliser cette estimation. Un compromis possible est de mettre à profit la plus grande robustesse de la méthode 2 (§4.2 et 4.3) et donc d'estimer par cette méthode le paramètre dynamique  $r$  avec une estimation préalable imparfaite de l'habitat (par régression par exemple). C'est ce qui sera fait dans le Chapitre suivant, pour l'étude des métapopulations de plantes de la Drôme. Nous montrerons alors qu'au moins pour un habitat binaire, sans système source-puits, cette méthode fournit de bonnes estimations de  $r$ . De plus la précision de l'estimation reste particulièrement bonne pour les valeurs de  $r$  élevées, c'est à dire dans les situations où la dynamique joue le rôle le plus important et pour lesquelles les techniques classiques sont particulièrement mal adaptées.

**Ainsi le gain de robustesse d'estimation se fera au prix d'une perte de précision sur les valeurs estimées. Bien que cela puisse paraître décevant sur un plan théorique, un tel**

choix permet néanmoins d'obtenir une méthode généralisable et robuste pour estimer le paramètre  $r$  d'extinction/colonisation à partir de données d'occupation répétées dans le temps, aussi bien qu'à partir de données d'occupation à une date donnée.

## **V. Dynamique spatiale et écologie des plantes vasculaires du département de la Drôme**

Nous avons montré dans le Chapitre précédent qu'il est possible de mettre en relation les caractéristiques locales d'extinction-colonisation au sein d'une métapopulation avec les structures spatiales émergentes des populations. Il n'est pas nécessaire de rechercher une description précise et détaillées des mécanismes à l'œuvre dans cette émergence ; en revanche il est important de choisir une échelle pertinente d'estimation des paramètres dynamiques.

Nous avons en outre montré qu'un compromis était nécessaire entre l'obtention d'estimations précises des paramètres dynamiques locaux, et la robustesse de la méthode d'estimation. Nous avons ainsi pu conclure sur la proposition d'une méthode d'estimation locale du paramètre  $r$ , tolérant un niveau acceptable d'imprécision, au profit d'une plus grande robustesse.

Fort de ces conclusions, nous mettrons en œuvre cette méthode d'inférence pour estimer le paramètre dynamique  $r$  avec des distributions spatiales réelles. En l'occurrence, nous nous intéressons à la distribution des plantes vasculaires supérieures du département de la Drôme, issues du travail de Garraud (2004).

### **1. Données floristiques et environnementales du département de la Drôme**

Nous proposons dans un premier temps une présentation synthétique des données floristiques recueillies par Garraud (2004), et des données environnementales rassemblées pour l'étude des habitats potentiels. Des détails supplémentaires sur le calcul et la sélection des variables environnementales figurent en Annexe 2.

#### *1.1. Inventaire floristique*

Les données floristiques utilisées sont des relevés, c'est-à-dire des listes d'espèces, réalisés en grand partie par L. Garraud. Nous avons sélectionné les relevés présentant au moins 20 espèces, ce qui nous permet de réduire le problème des pseudo-absences. En effet, un relevé avec un faible nombre d'espèces est très probablement partiel. Dès lors il renseigne mal sur l'absence réelle des espèces non relevées. Cela motive notre choix d'un seuil de

relevé relativement élevé, à 20 espèces. Ce faisant, nous obtenons un ensemble de 4104 relevés retenus pour notre étude. La Figure 1 (p. 9) représente l'échantillonnage initial, de plus de 11000 points. La sélection que nous avons réalisée n'introduit pas d'hétérogénéité spatiale supplémentaire manifeste.

### *1.2. Données environnementales*

Nous avons regroupé des données environnementales de différentes natures. L'obtention de toutes les informations environnementales en chaque point de relevé floristique a été réalisée grâce au système d'information géographique ArcGIS. Il en résulte un tableau de 4104 lignes-relevés, qui présente en colonnes les variables environnementales et la présence-absence des 2739 taxons recensés.

- Les **données d'altitudes** sont issues du modèle numérique de terrain de l'USGS (United States Geological Survey). Ce modèle résulte du traitement de données satellitaires d'images radar. Les données sont fournies sur une image raster où les altitudes sont figurées par pixels de 30 m de côté. Le traitement de ces données avec le logiciel ArcGIS a permis d'en déduire des cartes d'expositions et de pentes, sur la base du même maillage.
- Les **données géologiques** sont issues de la numérisation des cartes géologiques du BRGM (Bureau des Recherches Géologiques et Minières), à l'échelle 1/250 000. Les données obtenues ont été traitées sous forme de polygones.
- Le **données d'occupation des sols** sont issues de la base de données *Corine Land Cover* (CLC). Le programme européen CLC a constitué un inventaire biophysique de l'occupation des terres de 27 pays européens. La surface de la plus petite unité cartographiées est de 25 hectares, mais des rubriques existent pour décrire des mosaïques plus fines. Ce traitement est basé sur l'analyse d'images satellitaires SPOT, obtenues en 1992 et 1993, et donc en très bonne adéquation avec la période d'inventaire floristique. Les données ont été fournies par l'IFEN, dans le cadre d'une convention d'usage. Ce sont des données de type polygones.
- Les **données de températures et de précipitations moyennes mensuelles**, ont été moyennées sur une période de 15 ans (janvier 1985 à décembre 1999), période choisie

pour correspondre à celle de l'inventaire floristique. Les données brutes mensuelles pour toutes les stations météorologiques de la Drôme ont été acquises via la *Climathèque* de Météo France. Elles ont été interpolées sur l'ensemble des points de relevés floristiques, en utilisant selon les cas les méthodes de krigeage et de pondération par l'inverse de la distance.

## **2. Caractères généraux de la Flore drômoise**

Dans les paragraphes 3 et 4, nous évaluerons la dynamique spatiale individuelle des espèces végétales, afin d'éclairer leur comportement écologique et biogéographique, en relation notamment avec leur biologie et leur appartenance à des groupements végétaux particuliers.

Nous proposons au préalable de caractériser la structure spatiale des communautés végétales de la Drôme. Ces caractéristiques apporteront des clefs de compréhension sur les mécanismes d'assemblage des espèces, notamment biogéographiques, et permettront donc de formuler des hypothèses pertinentes sur les comportements individuels.

Considérons la distribution des probabilités agrégées (284 mailles d'échantillonnage) pour l'ensemble des taxons recensés dans la Flore drômoise (2143 taxons retenus, les sous-espèces sont confondues). Les communautés végétales peuvent être étudiées au moyen d'une analyse multivariée réalisant des regroupements d'espèces (variables) dans l'espace des sites (objets), suivant une règle de décomposition optimale de la variance des variables. Les probabilités agrégées d'occupation sont des grandeurs quantitatives réelles, entre 0 et 1, et nous utilisons dès lors une Analyse en Composantes Principales (ACP), basée sur la matrice de variance/covariance des points d'échantillonnage.

Les résultats de cette analyse permettent d'établir une hiérarchie de composantes principales indépendantes, par ordre décroissant de variance expliquée. Ainsi les premières composantes principales reflèteront les facteurs de structuration les plus forts de la Flore drômoise. Nous examinerons ici les 5 premières composantes ; elles expriment 32.5% de la variation globale de composition des communautés. Rappelons que ces communautés sont représentées à l'échelle des mailles.

### Nota Bene

L'analyse proposée ici, volontairement simplifiée, vise à esquisser des caractéristiques générales des communautés végétales drômoises. Nous choisissons pour cela de mettre en exergue des contrastes de composition, et de les interpréter par des facteurs écologiques. Il ne s'agit pas là d'une démarche phytosociologique (voir Ewald 2003) : nous ne considérons pas la variation de composition des communautés locales, mais celle d'un cortège d'espèces défini à plus grande échelle en fonction des occurrences locales. C'est une manière de définir des *patrons de végétation* à l'échelle des mailles utilisées dans notre analyse de la Flore de la Drôme.

L'utilisation de l'Analyse en Composantes Principales présente par ailleurs des défauts techniques qui peuvent être critiqués. Nous ne les développerons pas ici, et réserverons une analyse approfondie à un traitement technique ultérieur. Il est utile cependant de mentionner que nous avons réalisé par ailleurs une Analyse Non Symétrique de Correspondances (ANSC), qui présente l'avantage de ne pas produire d'effets fer à cheval comme l'ACP, d'avoir des propriétés intéressantes de décomposition de la diversité spécifique (Pélissier et al. 2003), et de ne pas accorder trop d'importance aux espèces rares (problème posé par l'Analyse de Correspondances traditionnelle). Nous avons également réalisé une ACP et une ANSC sur variables instrumentales, contraignant l'analyse des communautés par des relations linéaires avec des variables d'habitat. Ce type d'analyse canonique nous a permis d'examiner plus précisément la relation entre les composantes principales de ces analyses et les facteurs écologiques exprimés par les variables environnementales. Cette dernière analyse est en cours et dépasse donc le cadre de la thèse.

### *2.1. De la garrigue méditerranéenne aux prairies subalpines*

La Drôme est caractérisée par un échelonnement important de la végétation depuis les formations nettement méditerranéennes dans le sud-ouest, jusqu'aux formations montagnardes et subalpines du nord-est. C'est ce contraste qui est exprimé par la première composante principale de l'analyse (voir Figure 21). Voici les noms des espèces associées aux influences les plus fortes :

- Influence montagnarde-alpine, avec domination de la hêtraie-sapinière, des pelouses et éboulis montagnards-subalpins, par ordre décroissant du poids des espèces : hêtre *Fagus sylvatica* L., sapin *Abies alba* Miller, framboisier *Rubus idaeus* L., grande gentiane *Gentiana lutea* L., fétuque vivace *Festuca laevigata* Gaudin, prénanthe pourpre

*Prenanthes purpurea* L., fraisier *Fragaria vesca* L., *Carduus medius* Gouan, myrtille *Vaccinium myrtillus* L., sycomore *Acer pseudoplatanus* L., cotonéaster du Jura *Cotoneaster jurana* Gandoger, grande berce *Heracleum sphondylium* L., raiponce en épi *Phyteuma spicatum* L., globulaire à feuilles en cœur *Globularia cordifolia* L., sorbier à grappes *Sorbus aucuparia* L., primevère coucou *Primula veris* L., sésliére bleue *Sesleria caerulea* (L.) Ard., gentiane à feuilles étroites *Gentiana angustifolia* Vill., grande ortie *Urtica dioica* L., véronique petit chêne *Veronica chamaedrys* L., laîche toujours verte *Carex sempervirens* Vill., fougère mâle *Dryopteris filix-mas* (L.) Schott, verge d'or *Solidago virgaurea* L., épicéa *Picea abies* (L.) Karsten, chèvrefeuille des montagnes *Lonicera alpigena* L.

- Influence méditerranéenne, avec domination de la garrigue, des chênaies vertes et blanches, par ordre décroissant du poids des espèces : chêne blanc *Quercus humilis* Miller, thym *Thymus vulgaris* L., garance voyageuse *Rubia peregrina* L., germandrée petit chêne *Teucrium chamaedrys* L., aphyllanthe de Montpellier *Aphyllanthes monspeliensis* L., cornouiller sanguin *Cornus sanguinea* L., chêne vert *Quercus ilex* L., chardon Rolland *Eryngium campestre* L., badasse *Dorycnium pentaphyllum* Scop., herbe au bitume *Bituminaria bituminosa* (L.) E.H. Stirton, chèvrefeuille étrusque *Lonicera etrusca* G.Santi, *Teucrium polium* L., brome érigé *Bromus erectus* Hudson, buis *Buxus sempervirens* L., amélanchier *Amelanchier ovalis* Medik., genévrier oxycèdre *Juniperus oxycedrus* L., orpin de Nice *Sedum sediforme* (Jacq.) Pau, fétuque marginée *Festuca marginata* (Hackel) K. Richter, *Prunus mahaleb* L., astragale de Montpellier *Astragalus monspessulanus* L., baguenaudier *Hippocrepis emerus* (L.) P. Lassen, *Ononis minutissima* L., *Spartium junceum* L., *Brachypodium phoenicoides* (L.) Roemer & Schultes, genêt scorpion *Genista scorpius* (L.) DC., genévrier commun *Juniperus communis* L., troène *Ligustrum vulgare* L., laîche de Haller *Carex halleriana* Asso, genêt poilu *Genista pilosa* L.

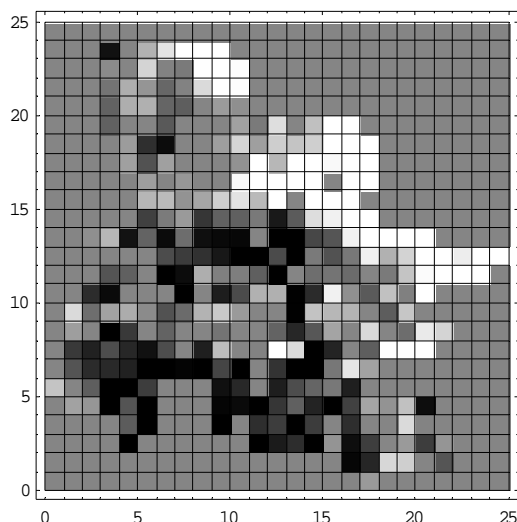


Figure 21 : 1ère composante principale de l'ACP réalisée sur l'ensemble des espèces végétales de la Drôme (variables), exprimées par leurs probabilités de présence en chaque maille de l'échantillonnage. Cette composante illustre le contraste en terme de communauté végétale entre l'influence montagnarde-alpine au nord-est et l'influence méditerranéenne au sud-ouest.

## 2.2. Composante méditerranéo-montagnarde

Un des aspects les plus remarquables de la végétation drômoise est l'existence sur une grande partie du territoire d'une végétation mêlant taxons méditerranéens et montagnards, et formant des groupements cohérents. Ce sont de surcroît des groupements calcicoles. C'est ce type de végétation qu'exprime la deuxième composante principale de l'analyse (voir Figure 22).

Parmi les espèces associées le plus fortement à cette composante, nous pouvons mentionner, par ordre décroissant des poids des espèces : buis *Buxus sempervirens* L., amélanchier *Amelanchier ovalis* Medik., lavande à feuilles étroites *Lavandula angustifolia* Miller, cytise à feuilles sessiles *Cytisophyllum sessilifolium* (L.) O.F.Lang, sorbier blanc *Sorbus aria* (L.) Crantz, genêt cendré *Genista cinerea* (Vill.) DC., pin sylvestre *Pinus sylvestris* L., anthyllide des montagnes *Anthyllis montana* L., genévrier commun *Juniperus communis* L., thym *Thymus vulgaris* L., brome érigé *Bromus erectus* Hudson, genêt poilu *Genista pilosa* L., *Achnatherum calamagrostis* (L.) P. Beauv., germandrée petit chêne *Teucrium chamaedrys* L., cupidone *Catananche caerulea* L., séslerie bleue *Sesleria caerulea* (L.) Ard., hêtre *Fagus sylvatica* L., *Avenula pratensis* (L.) Dumort., *Acer opalus* Miller, anémone hépatiche *Hepatica nobilis* Schreber, fétuque marginée *Festuca marginata* (Hackel) K. Richter, globulaire à feuilles en cœur *Globularia cordifolia* L., viorne lantane



*Viburnum lantana* L., aphyllanthe de Montpellier *Aphyllanthes monspeliensis* L., *Helianthemum oelandicum* (L.) DC., laser de France *Laserpitium gallicum* L., laîche glauque *Carex flacca* Schreber, laîche humble *Carex humilis* Leysser.

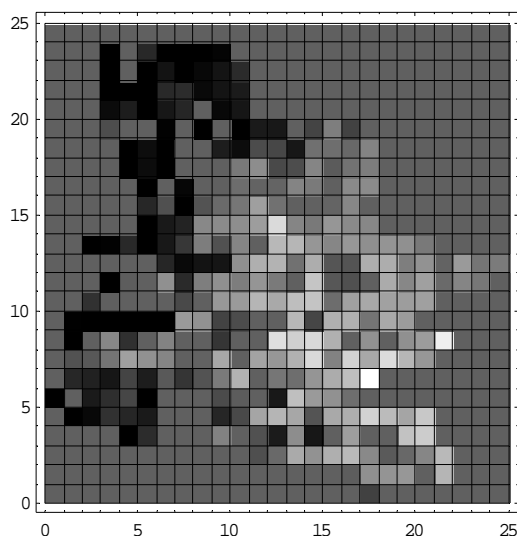


Figure 22 : 2ème composante principale de l'ACP réalisée sur l'ensemble des espèces végétales de la Drôme (variables), exprimées par leurs probabilités de présence en chaque maille de l'échantillonnage. Cette composante illustre l'influence méditerranéo-montagnarde au sein de la flore drômoise.

### 2.3. Les forêts tempérées médio-européennes

La troisième composante principale met en exergue l'existence, notamment au nord et au nord-ouest du département, de forêts à caractère tempéré, médio-européennes et de basse altitude, de type hêtraies et hêtraies-chênaies (voir Figure 23a). Nous notons en outre que cette composante est sensible aux espèces de fourrés et ourlets forestiers, ce qui traduit peut-être une importante dégradation de ces formations forestières.

Les espèces caractéristiques sont, par ordre décroissant de leurs poids : troène *Ligustrum vulgare* L., viorne lantane *Viburnum lantana* L., aubépine *Crataegus monogyna* Jacq., frêne *Fraxinus excelsior* L., noisetier *Corylus avellana* L., genévrier commun *Juniperus communis* L., pin sylvestre *Pinus sylvestris* L., lierre *Hedera helix* L., cornouiller sanguin *Cornus sanguinea* L., clématite vigne blanche *Clematis vitalba* L., *Lonicera xylosteum* L., laîche glauque *Carex flacca* Schreber, hellébore fétide *Helleborus foetidus* L., buis *Buxus sempervirens* L., érable champêtre *Acer campestre* L., chêne blanc *Quercus humilis* Miller, primevère officinale *Primula vulgaris* Hudson, brachypode des bois *Brachypodium*

*sylvaticum* (Hudson) P.Beauv., rosier des champs *Rosa arvensis* Hudson, baguenaudier *Hippocrepis emerus* (L.) P.Lassen, hêtre *Fagus sylvatica* L., sorbier blanc *Sorbus aria* (L.) Crantz, fraisier *Fragaria vesca* L., benoîte des villes *Geum urbanum* L., charme *Carpinus betulus* L., *Clinopodium vulgare* L., carotte *Daucus carota* L., genêt poilu *Genista pilosa* L., ortie *Urtica dioica* L.

Sur substrat alluvial décalcifié, ces formations forestières médio-européennes acquièrent un caractère acidiphile caractéristique (en particulier chênaie sessiliflore acidiphile), qui n'est pas exprimé par la 3<sup>ème</sup> composante principale (voir Figure 23b). En fait, l'affinité acidiphile est représenté par la 9<sup>ème</sup> composante principale, avec notamment les espèces suivantes : *Teucrium scorodonia* L., châtaignier *Castanea sativa* Miller, genêt à balais *Cytisus scoparius* (L.) Link, fougère aigle *Pteridium aquilinum* (L.) Kuhn, chêne sessile *Quercus petraea* (Mattuschka) Liebl., chèvrefeuille *Lonicera periclymenum* L., molinie *Molinia caerulea* (L.) Moench, charme *Carpinus betulus* L., houx *Ilex aquifolium* L., renoncule âcre *Ranunculus acris* L., brunelle *Prunella vulgaris* L., renoncule rampante *Ranunculus repens* L., calune *Calluna vulgaris* (L.) Hull, trèfle des prés *Trifolium pratense* L., caille-lait jaune *Galium verum* L., pulmonaire à longues feuilles *Pulmonaria longifolia* (Bast.) Boreau, millepertuis élégant *Hypericum pulchrum* L., *Holcus mollis* L., anémone sylvie *Anemone nemorosa* L.

Cela permet de souligner que les composantes principales de notre analyse expriment des influences façonnant les communautés, certes indépendantes entre elles, mais demeurant non exclusives : une communauté végétale locale peut ainsi répondre à différentes influences exprimées par plusieurs composantes principales.

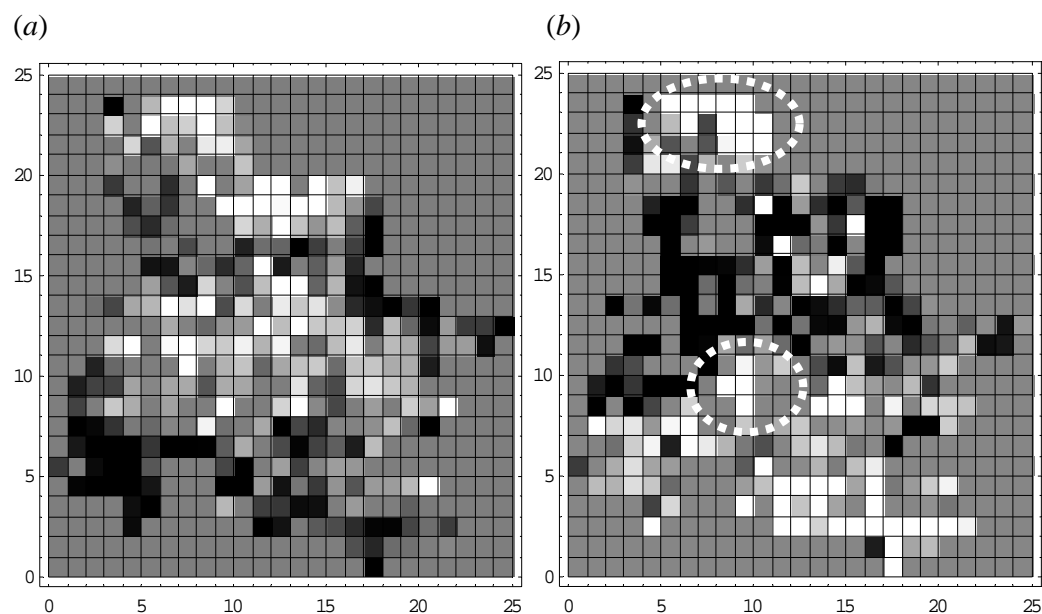


Figure 23 : (a) 3ème composante principale de l'ACP réalisée sur l'ensemble des espèces végétales de la Drôme (variables), exprimées par leurs probabilités de présence en chaque maille de l'échantillonnage. Cette composante traduit une influence forestière médio-européenne. (b) Cette influence se combine, en zones siliceuses, avec l'existence d'une flore forestière acidiphile, par exemple dans le Haut-Roubion et en Chambaran (cerclages blancs hachurés). Cette composante acidiphile est exprimée par la 9<sup>ème</sup> composante principale

#### 2.4. Vallées et zones humides

La quatrième composante principale met en exergue des taxons associés aux vallées internes du département, en particulier dans les zones humides associées, alluviales ou riveraines (Figure 24). Ces formations apparaissent notamment dans les lits du Rhône, de la Drôme, de l'Isère et du Jabron.

Les espèces caractéristiques sont, par ordre décroissant de leurs poids : carotte *Daucus carota* L., séneçon à feuilles de roquette *Senecio erucifolius* L., saule pourpre *Salix purpurea* L., peuplier noir *Populus nigra* L., verveine officinale *Verbena officinalis* L., lavande à feuilles étroites *Lavandula angustifolia* Miller, salicaire *Lythrum salicaria* L., herbe à la bûche *Cichorium intybus* L., menthe à feuilles longues *Mentha longifolia* (L.) Hudson , érigéron du Canada *Conyza canadensis* (L.) Cronq., saule blanc *Salix alba* L., *Picris hieracioides* L., jonc articulé *Juncus articulatus* L., cardère *Dipsacus fullonum* L., *Salix elaeagnos* Scop., roseau *Phragmites australis* (Cav.) Steudel, ronce bleue *Rubus caesius* L., séttaire verte *Setaria viridis* (L.) P.Beauv., cupidone *Catananche caerulea* L., lycophe d'Europe *Lycopus europaeus* L., *Achnatherum calamagrostis* (L.) P.Beauv., *Pulicaria dysenterica* (L.)

Bernh., *Bidens tripartita* L. Parmi ces espèces, notons la présence de *Conyza canadensis*, une xénophyte pour laquelle les vallées alluviales constituent des axes d'invasion : on la trouve souvent dans le lit mineur des cours d'eau.

Il n'y a pas ici d'unité écologique, étant donné qu'il y a aussi bien des espèces appréciant des contextes marneux humides que des espèces appréciant des alluvions sèches. En fait cette composante montre qu'il existe dans les vallées alluviales et les zones humides un point de rencontre entre des flores de natures différentes, et que ces communautés composites présentent dans leur ensemble une certaine constance en terme de composition.

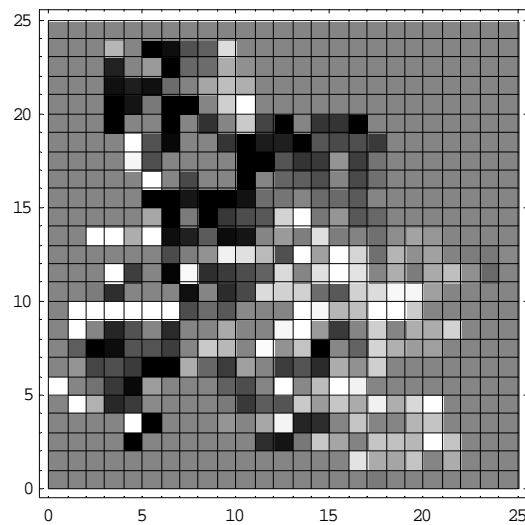


Figure 24 : 4ème composante principale de l'ACP réalisée sur l'ensemble des espèces végétales de la Drôme (variables), exprimées par leurs probabilités de présence en chaque maille de l'échantillonnage. Cette composante représente les végétations des zones humides, mais aussi des zones alluviales et des zones marneuses.

### 2.5. Le syndrome orchidées

La cinquième composante principale ne semble pas être associée à un type de végétation déterminé, mais met en exergue les espèces appartenant à une même famille botanique, celle des orchidées. Notons d'abord que l'essentiel de ces orchidées est représenté sur la quatrième composante en opposition aux espèces des vallées et zones humides : cela semble dénoter leur préférence pour les coteaux et les contextes bien drainés (en général sur calcaire). La cinquième composante exprime *a contrario* une **singularité des orchidées sans signification écologique évidente** : sont associées des espèces de pelouses sèches, de forêt méditerranéenne, de forêt tempérée ou encore de montagne.

Cette composante montre donc la singularité *biologique* des orchidées, une singularité telle que ces plantes contribuent collectivement et de manière importante à la variation de composition spécifique des communautés. Là où les réponses des autres espèces végétales correspondent à des schémas bioclimatiques et écologiques, indépendamment des parentés taxonomiques, les orchidées possèdent une réponse distincte et indépendante des autres familles botaniques.

Les espèces caractéristiques sont, par ordre décroissant de leurs poids (les noms latins des orchidées sont soulignés): listère ovale *Listera ovata* (L.) R.Br., ophrys mouche *Ophrys insectifera* L., orchis pourpre *Orchis purpurea* Hudson, orchis de Fuchs *Dactylorhiza fuchsii* (Druce) Soç, orchis brûlé *Orchis ustulata* L., orchis pyramidal *Anacamptis pyramidalis* (L.) Rich., orchis singe *Orchis simia* Lam., *Cephalanthera damasonium* (Miller) Druce, orchis moustique *Gymnadenia conopsea* (L.) R.Br., *Platanthera bifolia* (L.) L.C.M.Richard, homme pendu *Aceras anthropophorum* (L.) Aiton fil., limodore avorté *Limodorum abortivum* (L.) Swartz, ophrys bourdon *Ophrys fuciflora* (F.W. Schmidt) Moench, ophrys abeille *Ophrys apifera* Hudson, *Cephalanthera longifolia* (L.) Fritsch, ophrys araignée *Ophrys sphegodes* Miller, *Cephalanthera rubra* (L.) L.C.M.Richard, néottie nid d'oiseau *Neottia nidus-avis* (L.) L.C.M. Richard, *Ophrys pseudocolopax*, *Epipactis muelleri* Godfery, *Epipactis helleborine* (L.) Crantz, *Juniperus communis* L., *Pinus sylvestris* L., *Himantoglossum hircinum* (L.) Sprengel, *Hippocrepis comosa* L., *Viburnum lantana* L., *Genista pilosa* L., *Salvia pratensis* L.

Alors que les composantes principales précédentes recouvraient des régions bioclimatiques ou écologiques clairement déterminées et cohérentes, la distribution géographique de la composante « orchidée » est naturellement d'une autre nature (voir Figure 25). Elle met en exergue une sorte de diagonale au sein du département, qui semble donc propice à la présence des orchidées, par opposition à une frange sud-ouest et nord-est où leur présence serait déficitaire. Il est très peu probable qu'il s'agisse d'une réponse commune à un facteur du milieu physique, car nous avons souligné que ces espèces d'orchidées possèdent des écologies très différentes, et qu'aucune autre espèce ne semble répondre de cette manière.

Le facteur fédérateur serait donc une caractéristique biologique particulière et singulière des orchidées. Une hypothèse proposée est que les orchidées aient besoin pour leur symbiose racinaire de partenaires champignons possédant une distribution limitée, en particulier par des contingences biogéographiques ou des facteurs d'habitat non pris en compte et spécifiques. La présence des orchidées traduirait donc indirectement celle de ces champignons. En

supposant que la présence des autres espèces végétales de la Drôme ne soit pas liée aussi étroitement à la présence de ces champignons, cela expliquerait que l'on obtienne une composante principale associée spécifiquement aux espèces de la famille des orchidées. Toutefois cette hypothèse peut être mise en doute par le fait que d'autres espèces à symbioses racinaires très similaires (notamment Pirolacées, M-A Selloso, comm. pers.) ne ressortent pas sur cette composante.

La singularité des orchidées peut exprimer également, sans que ce soit incompatible avec l'hypothèse champignon, un comportement particulier vis à vis des activités agro-pastorales traditionnelles. En effet, si un pâturage ou une exploitation sylvicole intensive semble préjudiciable à leur maintien, on note souvent sur le terrain que les orchidées prospèrent dans des contextes très extensifs ou en déprise récente. On constate de fait sur la signature spatiale de cette composante (Figure 25) que les zones les plus marquées sont des secteurs de déprise récente, favorisée en particulier par l'influence de grands centres urbains : c'est très net pour le bassin de Valence. Les données des 4 derniers recensements agricoles ont été acquises pour le département de la Drôme, et permettront d'étudier plus précisément cette hypothèse.

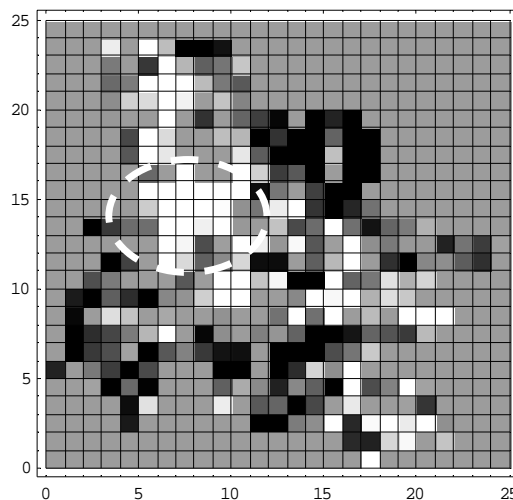


Figure 25 : 5<sup>ème</sup> composante principale de l'ACP réalisée sur l'ensemble des espèces végétales de la Drôme (variables), exprimées par leurs probabilités de présence en chaque maille de l'échantillonnage. Cette composante est positivement associée à un ensemble d'espèce appartenant à une même famille, celle des orchidées. La zone cerclée de pointillés blancs représente la plaine de Valence, qui a connu une déprise agricole récente importante.

### 3. Estimation du ratio $r$ extinction / colonisation

Les méthodes proposées au Chapitre précédent pour l'estimation du ratio extinction/colonisation  $r$  requièrent des données réelles de distribution de présence-absence qui sont agrégées soit dans le temps, soit dans l'espace. Les données que nous possédons pour le département de la Drôme ne constituent pas des répliquats temporels. Durant la période d'échantillonnage (essentiellement sur 10 ans), l'attention a été apportée à la prise en compte d'un maximum de points dans l'espace, et non au retour sur des points déjà échantillonnés.

Dès lors, c'est l'agrégation spatiale qui s'impose. Nous appliquons donc la procédure d'agrégation spatiale et d'inférence du paramètre  $r$  aux données de présence/absence des plantes vasculaires de la Drôme (voir Chapitre IV, §4.3). Les estimations obtenues serviront de support au test d'hypothèses écologiques sur ces espèces.

#### *3.1. Agrégation spatiale des données*

Considérons une grille régulière carrée comportant 25x25 mailles. Nous la superposons au graphe des relevés floristiques effectués sur la Drôme. L'amplitude verticale et horizontale de la grille est déterminée par les latitudes et longitudes extrêmes des relevés effectués dans la Drôme. Cela induit une distorsion des mailles, puisqu'une maille de la grille 25\*25 correspond en réalité à un espace géographique de 3,5 x 5,2 km environ.

Chaque maille recouvre un certain nombre de points de relevés, nombre variant d'une maille à l'autre. Nous appelons agrégation spatiale l'estimation de la probabilité d'occupation d'une maille par une espèce donnée : il s'agit de la proportion de données de présence sur l'ensemble des relevés à l'intérieur de chaque maille. A une espèce présente dans 2 relevés sur 10 est affectée une probabilité de 0.2.

Nous sélectionnons les mailles de la grille comportant au moins 4 relevés, pour limiter le biais d'estimation de la probabilité d'occupation des mailles. Le maillage résultant comporte 284 éléments (Figure 26a). Les cartes de probabilités d'occupation sont alors constituées pour chaque espèce de plante vasculaire recensée dans la Drôme (Figure 26b). Nous agrégeons également les données d'habitat. La valeur d'une variable agrégée d'habitat pour une maille donnée représente donc la valeur moyenne de cette variable sur l'ensemble des relevés à l'intérieur de la maille (Figure 26c).

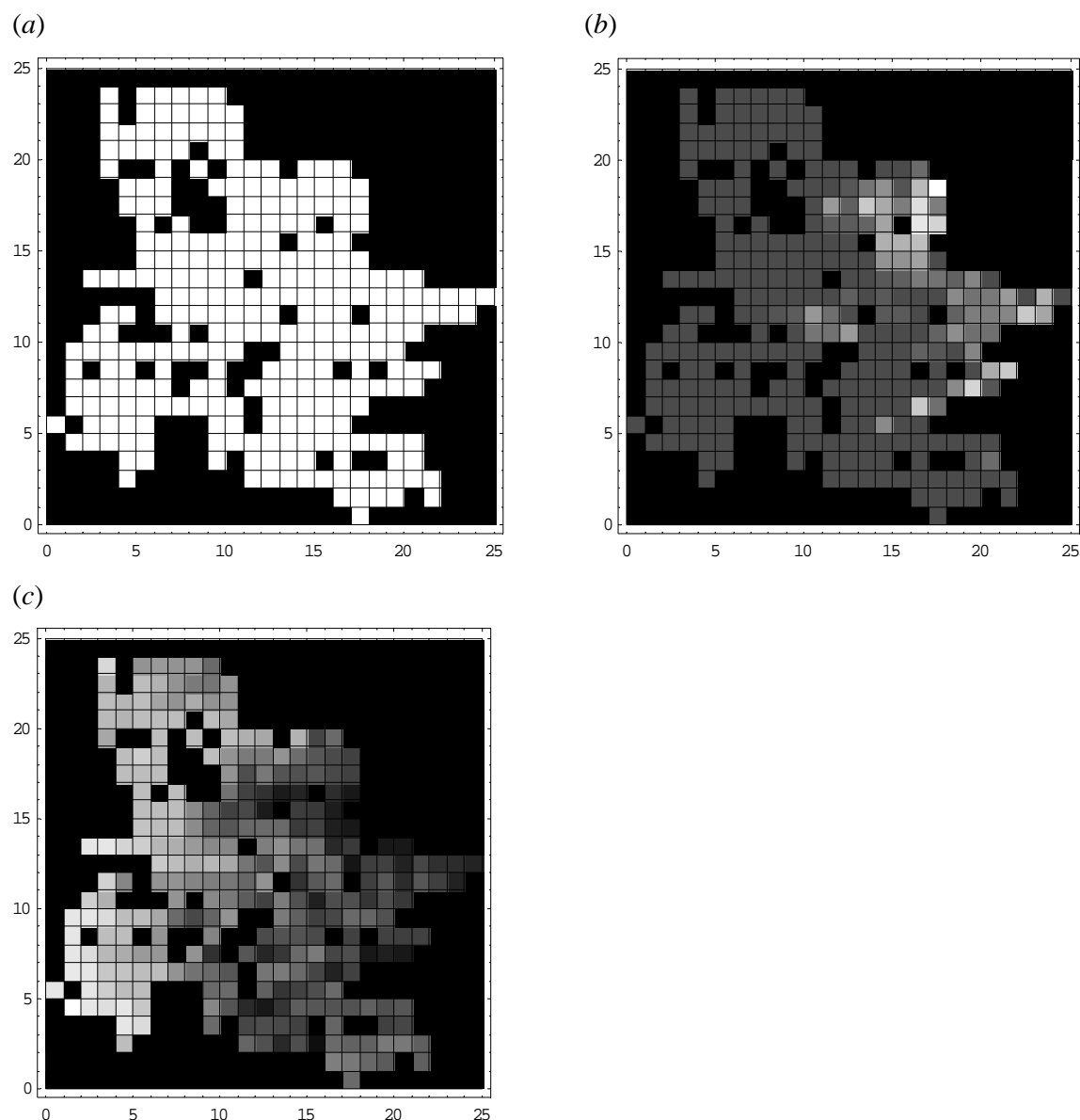


Figure 26 : Procédure d'agrégation spatiale. (a) Maillage utilisé pour agréger spatialement les données de présence-absence des plantes vasculaires du département de la Drôme. (b) Carte de probabilités agrégées d'occupation obtenue pour le sapin *Abies alba* Miller. Ce taxon montagnard est cantonné aux Préalpes, dans l'est du département. (c) Carte d'habitat agrégée, pour les données de température mensuelle moyenne en février.

### 3.2. Cartes d'habitat potentiel

Nous ne connaissons pas la niche fondamentale des espèces considérées. Nous savons en outre que la relation statistique brute entre les données de présence et les données d'habitat, par un modèle linéaire en général, est dangereusement biaisée par la non prise en compte de la dynamique spatiale des populations. Toutefois, nous avons également vu que nous n'avons



pas de solution suffisamment fiable pour entreprendre une estimation simultanée des paramètres d'habitat et du paramètre de dynamique spatiale  $r$ , car l'information donnée par la distribution spatiale des espèces peut être dans certains cas insuffisante et équivoque.

En guise de compromis, il est possible de sacrifier une partie de la précision visée par l'estimation conjointe de la niche fondamentale « réelle » et du paramètre  $r$ , en instaurant une procédure en deux temps :

- L'application d'un modèle linéaire classique permet de relier les probabilités agrégées de présence aux valeurs agrégées de variables d'habitat. On en déduit une enveloppe d'habitat potentiel, qui ne représente pas fidèlement la réalité de l'habitat fondamental mais en donne néanmoins une première estimation. Nous avons exclu l'utilisation d'une méthode autorégressive, car séparer la variation autocorrélée de la fraction non corrélée ne permet pas de séparer l'empreinte de la dynamique spatiale de l'empreinte de l'habitat : les deux parties sont étroitement liées et ne peuvent être découplées par un modèle linéaire.
- A partir de cette carte d'habitat, nous utilisons la procédure d'estimation locale du paramètre  $r$ . Cette procédure permet de définir un intervalle de confiance représentant la variation des valeurs locales estimées. Ce paramètre étant supposé être caractéristique de l'espèce et indépendant de l'habitat, il devrait être constant dans l'espace. Ainsi l'erreur d'estimation sur l'habitat *a priori* induira une variation d'estimation de  $r$ , et donc induira une perte de précision.

Nous avons donc calculé les habitats potentiels des espèces végétales du département de la Drôme à l'aide de modèles linéaires généralisés gaussiens (logiciel SPlus).

### 3.2.1. Variables d'habitat

Revenons d'abord à ce que nous appelons l'habitat d'une espèce. Nous avons introduit précédemment l'habitat comme variable de l'équation de dynamique métapopulationnelle : elle modère le taux intrinsèque d'extinction  $e$  de l'espèce. Ce dernier est lié à la capacité de persistance de l'espèce, dès lors que toutes ses exigences physiologiques sont satisfaites : il s'agit d'un taux d'extinction essentiellement démographique. Nous créons donc, par le biais de la notion d'habitat, des contraintes liées à la non-satisfaction des exigences locales de

l'espèce. Moins l'espèce rencontrera des conditions favorables à la survie individuelle, plus son taux d'extinction local sera élevé.

Cela inclut les conditions physiques du milieu, essentiellement statiques, mais on peut également l'élargir à la prise en compte des régimes de perturbation réguliers qui constitue des variables dynamiques d'habitat. C'est pourquoi nous nous intéressons également aux types d'occupations des terres par l'Homme, ce qui est un indice indirect des régimes de perturbations imposés par ses activités.

La qualité de l'estimation du paramètre-clé  $r$  dépendra de la qualité de l'estimation de l'habitat et donc de la qualité et de la pertinence des variables d'habitat. Les variables brutes que nous avons retenues sont les suivantes :

- *Variables géographiques* : latitude et longitude (en m).
- *Variables topographiques* : altitude, pente et exposition.
- *Variables climatiques* : températures mensuelles moyennes (en °C), et précipitations mensuelles moyennes (en mm). Nous avons réalisé l'interpolation des données des stations météorologiques sur l'ensemble des points de relevés.
- *Variables géologiques* : données numérisées et géoréférencées du BRGM (Bureau de recherches géologiques et minières); la nomenclature a été réduite à quatre catégories complémentaires (voir Annexe 2).
- *Variables d'occupation des terres* : ce sont des variables binaires correspondant aux rubriques de la nomenclature *Corine Land Cover* (données fournies par l'Institut Français de l'Environnement). 18 variables sont retenues (voir Annexe 2).

Le traitement de ces données, notamment la projection spatiale et le croisement géographique des informations, a été réalisé à l'aide du système d'information géographique *ArcGIS*. Pour plus de détails sur l'obtention et la sélection des variables, voir Annexe 2.

### 3.2.2. Deux représentations de l'habitat

En définitive 52 variables d'habitat sont disponibles. Elles sont plus ou moins dépendantes entre elles et de nature très variable. Pour simplifier, on a constitué deux grandes catégories :

- Les variables bio-climatiques, regroupant altitude, température, précipitation et géologie représentent les conditions du milieu physique favorables ou non à l'existence des individus.

- Les variables d'occupation des sols (*Corine Land Cover*) représentent essentiellement l'emprise de l'Homme sur la distribution des plantes, au travers du modelage profond des habitats par ses activités. C'est un indicateur des régimes de perturbation affectant la survie des populations.

Il y a une certaine dépendance entre ces groupes de variables, l'utilisation des terres par l'Homme dépendant largement des conditions bioclimatiques. On trouvera ainsi des vergers essentiellement dans les massifs méditerranéens-montagnards (Baronnies), et *a contrario* des prairies dans les secteurs montagnards-subalpins. La réponse des espèces aux variables d'habitat sera donc façonnée par l'interaction complexe de ses propres exigences bioclimatiques et des régimes de perturbation imposés par l'Homme.

Les variables étant fortement dépendantes entre elles, le calcul des habitats potentiels des espèces par régression linéaire est réalisé sur des composantes principales et non sur les variables brutes, afin que les coefficients des régressions puissent être comparés entre espèces.

Nous calculons d'une part l'habitat potentiel au sens des variables bioclimatiques, avec : (i) une composante constante, (ii) les 11 premières composantes principales des 24 variables de précipitations et de températures (incluant l'effet altitude) (iii) 3 variables de géologie, par construction indépendantes entre elles (la quatrième variable est dépendante). D'autre part nous construisons l'habitat potentiel au sens de l'occupation des sols, avec 17 variables, par construction indépendantes (la 18<sup>ème</sup> est dépendante). Enfin l'habitat potentiel global représente l'effet de l'ensemble des 52 variables initiales, et représente donc la capacité d'ajustement maximale de nos variables d'habitat. Cette dernière information donne lieu à un surparamétrage (52 variables d'habitat pour 284 points d'échantillonnage). Nous préférons donc perdre un peu de précision et gagner en pertinence en travaillant avec l'habitat potentiel bioclimatique ou l'habitat potentiel en terme de catégories d'occupation des sols, avec des degrés de liberté similaires (respectivement 15 et 18). Les deux types d'habitat potentiels sont illustrés par la Figure 27.

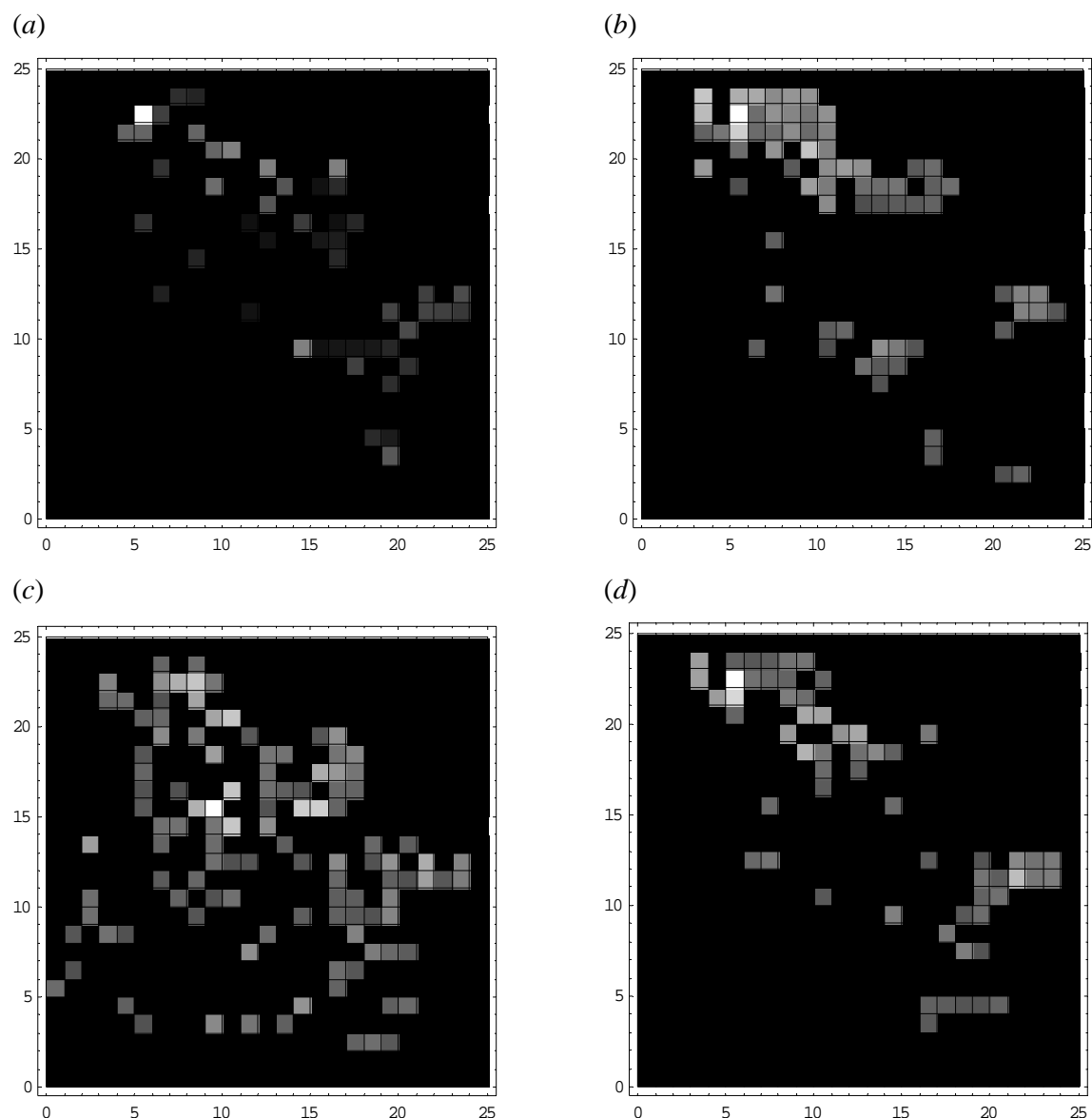


Figure 27 : Habitats potentiels calculés pour l'Herbe aux goutteux *Aegopodium podagraria*. (a) Distribution des probabilités agrégées d'occupation de l'espèce. (b) Habitat potentiel au sens des variables bioclimatiques. (c) Habitat potentiel au sens des variables d'occupation des sols. (d) Habitat potentiel avec l'ensemble des variables d'habitat. Pour alléger la représentation graphique, les valeurs des probabilités d'habitat sont tronquées à leur médiane (valeurs supérieures représentées). A noter que cette espèce est à affinité bioclimatique montagnarde (confirmé par *b*) mais est depuis longtemps cultivée par l'homme, et est donc présente en zone rudérale près des habitations (confirmé par *c*) : cette dualité est traduite par le saut qualitatif de l'ajustement mêlant les deux types de variables (*d*).

En résumé, nous allons procéder à deux séries distinctes d'inférences du paramètre  $r$  :

- Une à partir de l'habitat potentiel bioclimatique, donc négligeant la structuration des habitats par l'Homme. Dans ce cas on suppose que la structuration des habitats est directement sous dépendance bioclimatique.
- L'autre à partir de l'habitat potentiel d'occupation des sols, en considérant l'effet de l'homme sur la structure des habitats.

On obtiendra ainsi une estimation du paramètre  $r$  pour l'habitat bioclimatique ( $rBioclim$ ), une pour l'habitat CLC ( $rCLC$ ), et enfin une pour l'ensemble des 52 variables brutes (modèle surparamétré) ( $rAll$ ).

### 3.3. Validation de la procédure d'inférence

Le paramètre  $r$  est estimé localement, en chaque maille d'échantillonnage, à partir de l'équation d'incidence et de la distribution spatiale des probabilités agrégées. A titre de validation, nous utilisons le maillage agrégé des relevés floristiques de la Drôme pour estimer le paramètre  $r$  de nos métapopulations simulées en habitat binaire, à une génération quelconque de l'état quasi-stationnaire (donnée de présence-absence). Nous pouvons ainsi contrôler l'effet de la topologie de ce maillage sur la prédiction du paramètre  $r$ .

Nous présentons à la Figure 28 les résultats de l'estimation avec prise en compte de l'habitat « réel », utilisé dans la simulation (Figure 28a), et avec l'habitat potentiel estimé par régression linéaire (Figure 28b). Dans ce dernier cas, une transformation linéaire permet d'assurer une variation de l'habitat entre 0 et 1. On perd alors quelque peu de qualité d'estimation, notamment aux petites valeurs de  $r$ , mais cela demeure toutefois globalement satisfaisant.

Nous considérons dès lors que cette démarche d'estimation est fiable et applicable aux données floristiques de la Drôme. Les résultats généraux des estimations par cette méthode de  $rBioclim$ ,  $rCLC$  et  $rAll$ , pour l'ensemble des espèces présentes dans au moins 20 des 284 mailles de l'échantillonnage agrégé, sont donnés en moyennes spatiales en Annexe 3. Ces résultats seront discutés dans le paragraphe suivant.

Les différences entre ces différentes estimations, selon la définition de l'habitat, pourront être testées au moyen d'un test  $t$  sur les différences de moyennes spatiales entre estimations locales. Suivant le signe et la significativité de ces différences, nous pourrions tester notamment l'effet de l'occupation humaine sur les distributions des espèces (§4).

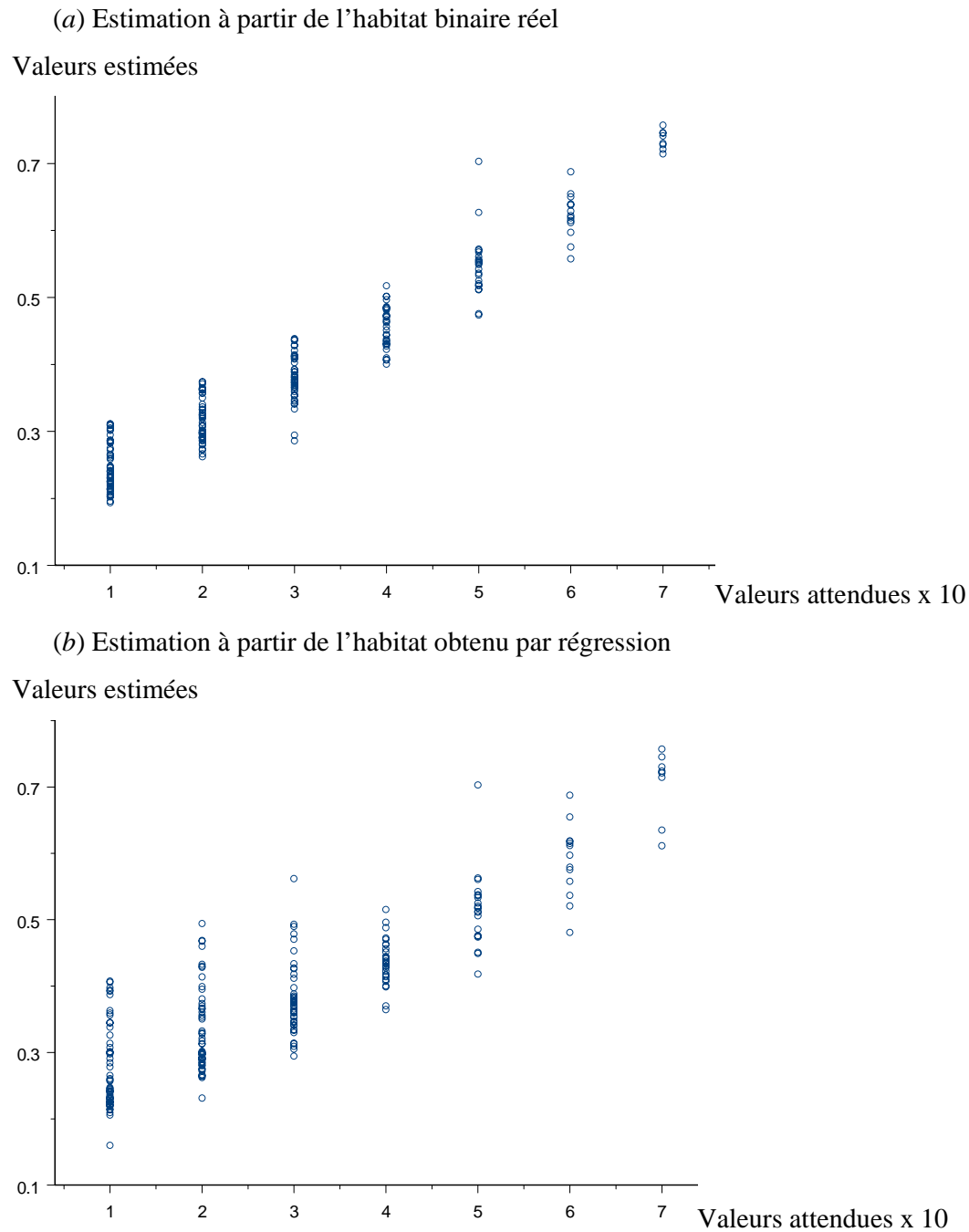


Figure 28 : Estimation du paramètre  $r$  pour les métapopulations simulées, avec une agrégation spatiale sur un graphe identique à celui utilisé pour les données floristiques de la Drôme. (a) L'habitat agrégé réel, c'est-à-dire tel qu'il est utilisé dans la simulation, est utilisé dans l'estimation. (b) L'habitat agrégé obtenu par régression sur les probabilités agrégées d'occupation est utilisé, avec une correction par la valeur maximale sur l'ensemble du graphe.

### 3.4. Normalisation de l'habitat potentiel

Nous discutons ici brièvement le choix de la normalisation de la variation de l'habitat potentiel, importante pour faire figurer correctement cette information dans le modèle d'inférence.

L'obtention de l'information d'habitat *potentiel* se fait par régression linéaire sur les probabilités d'occurrence agrégées dans l'espace. Cette formulation permet d'écrire un vecteur de « valeurs ajustées », combinaison linéaire des variables environnementales.

Dans le modèle de dynamique de métapopulation utilisé pour l'inférence (Ch. IV, §4.3), l'information habitat intervient en tant que terme correctif de la fonction de colonisation (voir Ch. IV, §3.5). Quand l'habitat vaut 1, le taux de colonisation global depuis une population vaut alors  $C_i$ . Il est donc important que la valeur maximale de l'habitat, utilisée pour l'estimation de  $r$ , vaille 1, afin que  $r = e/C_i$  représente bien pour toutes les espèces **le ratio de leur taux d'extinction local et de leur taux de colonisation maximum**.

La gamme de variation des valeurs ajustées issues de la régression linéaire de l'habitat potentiel doit ainsi être décidée en fonction des propriétés attendues du modèle de colonisation. Soit  $h_0$  le vecteur brut issu de la régression linéaire. On peut écrire :

$$h = \frac{h_0 - x}{\text{Max}(h_0) - x} \text{ avec } x \leq \text{Min}(h_0)$$

$$\text{Ainsi } \text{Max}(h) = 1 \text{ et } \text{Min}(h) = \frac{\text{Min}(h_0) - x}{\text{Max}(h_0) - x}, \text{ avec en particulier } \text{Min}(h) = 0 \text{ pour } x =$$

$\text{Min}(h_0)$ . Le paramètre  $x$  contrôle le choix de la valeur inférieure de la variation de l'habitat  $h$ .

Nous appelons *normalisation de l'habitat potentiel*, la procédure qui consiste à déterminer  $x$ . Nous avons envisagé deux options :

- Dans un premier temps, nous avons arbitrairement choisi  $x = \text{Min}(h_0)$  et donc  $\text{Min}(h) = 0$ . De cette manière, le taux de colonisation varie entre 0 et  $C_i$  sur l'ensemble du département de la Drôme. **Cela présente l'inconvénient d'imposer une gamme de variation identique pour toutes les espèces, sans intervenir toutefois sur la variation entre espèces des moyennes d'habitat.** Cette démarche induit fréquemment une certaine corrélation entre les estimations de  $r$  et de  $h$ .
- Dans un second temps, nous avons considéré que le modèle métapopulationnel avec effet d'habitat inclut dans la définition du facteur  $h$  l'effet global de l'habitat sur la dynamique de colonisation et d'extinction de l'espèce focale. Cela implique en particulier que  $r$  doit

être indépendant de  $h$ . Nous avons donc bâti une seconde méthode de renormalisation, dite *dynamique*, qui calcule pour une espèce donnée la corrélation entre  $r$  et  $h$  quand on fait varier  $x$ . On retient la valeur de  $x$  qui minimise cette corrélation, afin de satisfaire au mieux l'attendu théorique. **On supprime ainsi toute corrélation entre  $r$  et  $h$ , qui serait non signifiante et artéfactuelle au regard de notre modèle. Cela peut néanmoins occulter des phénomènes écologiquement pertinents, comme des systèmes source-puits (Dias 1996). Cette méthode modifie également de manière non triviale la variation des valeurs moyennes d'habitat entre espèces.** Cette normalisation a notamment pour effet de réduire la gamme de variation de  $r$ , à des valeurs élevées.

Un aller-retour entre le calcul de  $r$  et de  $h$ , en minimisant itérativement leur corrélation, permet de retrouver dans la deuxième méthode une logique interactive et simultanée d'estimation de l'habitat et de la dynamique des espèces. Par rapport à la première méthode, cela est loin d'être neutre d'un point de vue conceptuel.

Nous ne donnerons pas ici les résultats pour les deux méthodes. **S'agissant des observations générales issues de notre analyse de la Flore de la Drôme, nous constatons qu'elles ne sont pas qualitativement modifiées d'une méthode à l'autre.** Les résultats présentés dans la suite du travail sont ceux issus de la première méthode, avec une variation entre 0 et 1 de  $h$ .

L'existence d'une corrélation artéfactuelle fréquente entre  $r$  et  $h$  avec cette méthode affecte notamment les espèces peu fréquentes. La comparaison des résultats obtenus par les deux méthodes de normalisation nécessitera une analyse comparative plus approfondie, qui dépassera le cadre du discours de la thèse. Il en sera davantage fait état dans la rédaction d'un article méthodologique sur notre approche inférentielle.

### *3.5. Caractéristiques générales des estimations pour les végétaux de la Drôme*

Rappelons que le paramètre  $r$  représente, pour une population quelconque d'une espèce donnée, le ratio du taux local d'extinction par le taux global de colonisation des cellules voisines, par unité de temps et pour un habitat uniformément favorable. Plus  $r$  est élevé, plus l'extinction est forte relativement à la colonisation. Pour une espèce donnée, si nous considérons plusieurs estimations de  $r$  correspondant à plusieurs enveloppes possibles d'habitat potentiel, nous pouvons adopter les principes suivants :



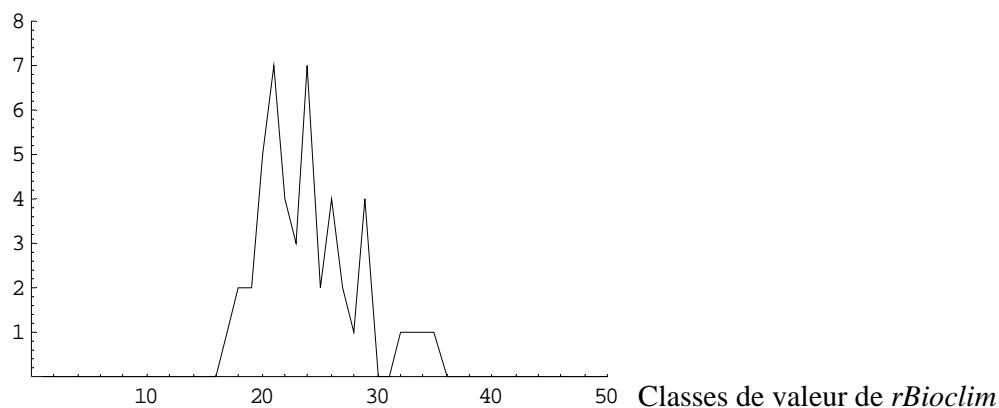
- Principe 1 :  $r$  est en théorie un paramètre intrinsèque au comportement biologique de l'espèce ; il ne varie donc pas spatialement, si ce n'est par erreur pure d'échantillonnage ou par manque de qualité d'évaluation de l'habitat réel. Dès lors, plus on est proche de l'habitat réel de l'espèce, plus on s'attend à ce que la variance d'estimation locale de  $r$  soit faible. Une première démarche est donc de considérer que l'habitat potentiel pour lequel la variance d'estimation est la plus faible est l'habitat écologiquement le plus pertinent.
- Principe 2 : Les deux types d'habitat potentiel que nous avons définis au paragraphe précédent permettent de tester la réponse écologique des espèces aux activités humaines. Pour une espèce donnée, si le paramètre  $r$  est significativement plus faible dans un habitat potentiel décrit par l'occupation anthropique des sols que dans un habitat bioclimatique, cela signifie que l'adaptation de l'espèce à son environnement anthropique prime actuellement sur ses exigences bioclimatiques. Cela implique donc que la distribution actuelle de l'espèce en question est modelée fortement par les activités humaines. Inversement, la réponse de l'espèce peut s'avérer être meilleure si l'on considère des critères d'habitat bioclimatiques, ou encore la réponse entre les deux types d'enveloppe peut ne pas être significativement différente.

Le paramètre  $r$  a été estimé pour chaque espèce sur les 284 points du maillage retenu pour l'étude des probabilités agrégées de présence dans la Drôme. Nous avons estimé le paramètre  $r$  pour l'ensemble des espèces disponibles; les sous-espèces sont regroupées, et nous en restons ici à un traitement au niveau spécifique.

Les variables d'habitat disponibles ont été définies précédemment (§3.2.1). Deux sous-ensembles sont considérés, variables bioclimatiques et variables d'utilisation des sols (Corine Land Cover, CLC). Rappelons que nous considérons tantôt les estimations de  $r$  au sens de l'habitat potentiel bioclimatique ( $rBioclim$ ), de l'habitat potentiel CLC ( $rCLC$ ), ou de l'habitat potentiel global ( $rAll$ ).

La Figure 29 illustre les distributions des estimations de  $rBioclim$  et  $rCLC$  pour l'herbe aux goutteux, *Aegopodium podagraria*, espèce que nous avons déjà considérée précédemment (voir Figure 27). Les moyennes d'estimation valent  $\bar{r}Bioclim = 0.462$  et  $\bar{r}CLC = 0.488$ .

(a) Fréquence d'estimation locale



(b) Fréquence d'estimation locale

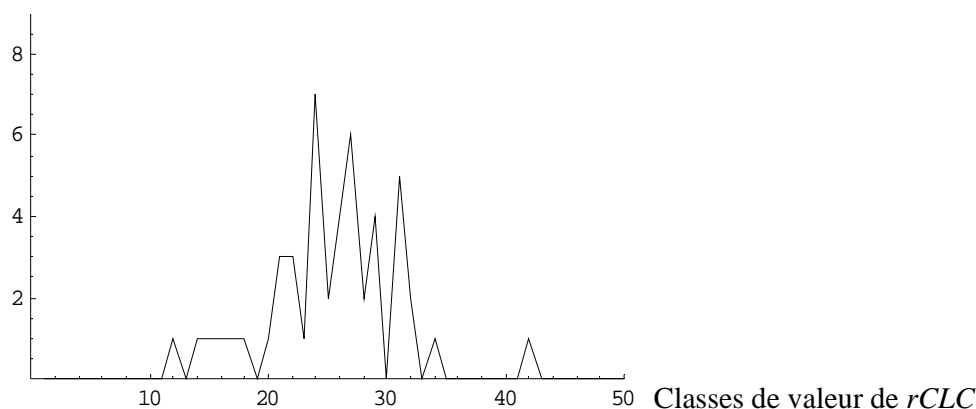


Figure 29 : Distribution globale d'estimation du paramètre  $rBioclim$  (a) et du paramètre  $rCLC$  (b), pour l'herbe aux goutteux, *Aegopodium podagraria*. Cette distribution est évaluée pour l'ensemble des estimations de  $rBioclim$  sur les mailles où la probabilité de présence de l'espèce est non nulle. Il y a 50 classes de valeurs de  $rBioclim$ , pour une variation globale de 0 à 1. L'hypothèse nulle de normalité composite des deux distributions n'est pas rejetée au sens du test de Kolmogorov-Smirnov ( $p$ -value > 0.1) (calcul réalisé avec SPlus).

Les paramètres  $rBioclim$ ,  $rCLC$  et  $rAll$  ont ainsi été évalués pour l'ensemble des espèces végétales inventoriées par Garraud (2004) dans la Drôme. Dans le souci de limiter les biais d'estimation possibles sur l'habitat potentiel et sur  $r$ , nous considérerons dans un premier temps l'ensemble des espèces présentant une probabilité agrégée d'occupation non nulle dans au moins 20 des 284 mailles considérées (voir Annexe 3).

Une remarque préliminaire importante doit être faite concernant la notion d'abondance des espèces. Le fait de travailler à une échelle spatiale intermédiaire, en « agrégeant »

l'information de l'échelon local à une échelle supérieure, implique de considérer deux aspects bien distincts :

- D'une part l'abondance des espèces au sein des mailles, en terme de densité de populations présentes. Elle est exprimée par la probabilité agrégée d'occupation calculée pour chaque espèce en chaque maille.
- D'autre part l'amplitude spatiale de l'espèce en terme de nombre de mailles occupées.

On constate à la Figure 30a qu'il existe une corrélation entre ces deux types d'abondance, mais que la faible variabilité de l'abondance moyenne intra-maille contraste avec la forte variabilité de l'abondance de mailles occupées.

Nous proposons à présent quelques commentaires généraux sur les estimations obtenues du paramètre  $r$ . La Figure 30b révèle que les estimations de  $\bar{r}CLC$  sont généralement plus faibles que  $\bar{r}Bioclim$  (position des points par rapport à la droite  $y = x$ ). Cela est confirmé par le test  $t$  très significatif entre les deux ensembles de valeurs estimées ; la valeur moyenne de  $\bar{r}CLC$  sur les espèces considérées est 0.392, celle de  $\bar{r}Bioclim$  est 0.491. Nous discuterons la signification de ce constat dans la partie suivante (§4).

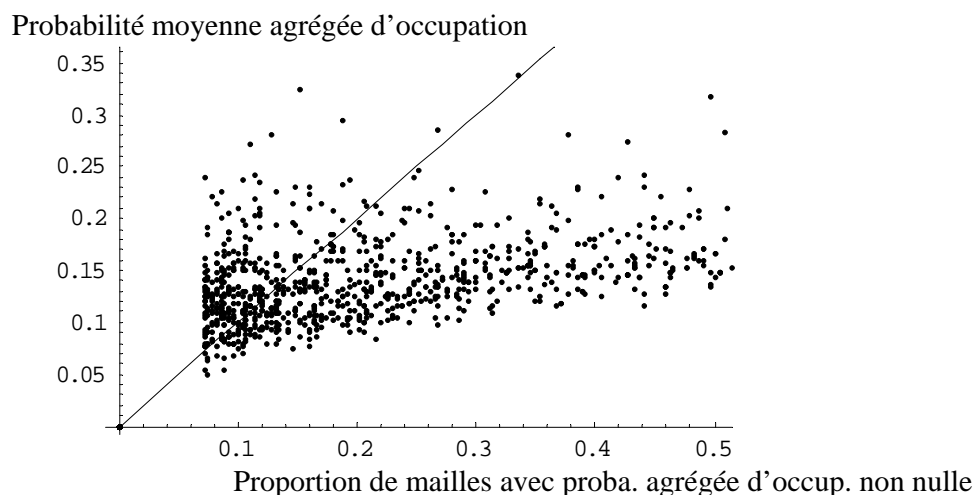
Il existe par ailleurs une corrélation positive significative entre les estimations de  $\bar{r}Bioclim$  et  $\bar{r}CLC$  (tests paramétriques et non-paramétriques effectués). Cela révèle une tendance générale entre les deux estimations. On peut également mettre en évidence un lien significatif entre  $\bar{r}Bioclim$  et  $\bar{r}CLC$  au travers de modèles régressifs généralisés (fonction *glm* de SPlus). Toutefois il y a dans de tels modèles une très forte dispersion des résidus qui suggère une large indépendance des deux estimations, une fois que la tendance générale est prise en compte. Ainsi pour la régression de  $\bar{r}Bioclim$  sur  $\bar{r}CLC$ , la déviance globale explicable est 10.435 et la déviance des résidus vaut 10.208 : la liaison linéaire entre les deux estimations n'explique que 2.17% de la déviance globale de  $\bar{r}Bioclim$ . En conclusion, les deux estimations sont en très grande partie indépendantes. Cela suggère qu'ils représentent des aspects distincts de la dynamique spatiale des espèces, et également que les habitats potentiels au sens bioclimat et usage des sols sont en grande partie indépendants. La comparaison des estimations  $\bar{r}Bioclim$  et  $\bar{r}CLC$  sera dès lors porteuse d'information écologique pertinente, dont la lecture sera l'objet de la partie 4.

Un autre élément intéressant est la corrélation non significative entre  $\bar{r}CLC$  et le nombre de mailles d'occupation non nulle (définissant l'amplitude spatiale), et *a contrario* la

corrélation négative très significative entre  $\bar{r}Bioclim$  et cette même amplitude spatiale. Cette propriété est conservée si l'on considère le nombre effectif de relevés occupés sur l'ensemble de la Drôme (échelle locale), au lieu du nombre de mailles (échelle intermédiaire).

Cela suggère que les espèces répandues, au sens de l'amplitude spatiale, saturent moins leur habitat bioclimatique que les espèces plus localisées. Cela semble traduire une opposition entre des généralistes répandus mais plus instables et des spécialistes plus localisés mais plus stables. Il est intéressant de souligner que ce dualisme n'existe pas dans l'habitat CLC : cet habitat représente des régimes de perturbations très récents, et l'émergence de compromis supposés adaptatifs entre stratégies généralistes et spécialistes serait alors encore non exprimée.

(a) Relation entre probabilité moyenne d'occupation des mailles et nombre de mailles où l'occupation est non nulle.



(b) Relation entre estimations moyennes de  $rBioclim$  et  $rCLC$ .

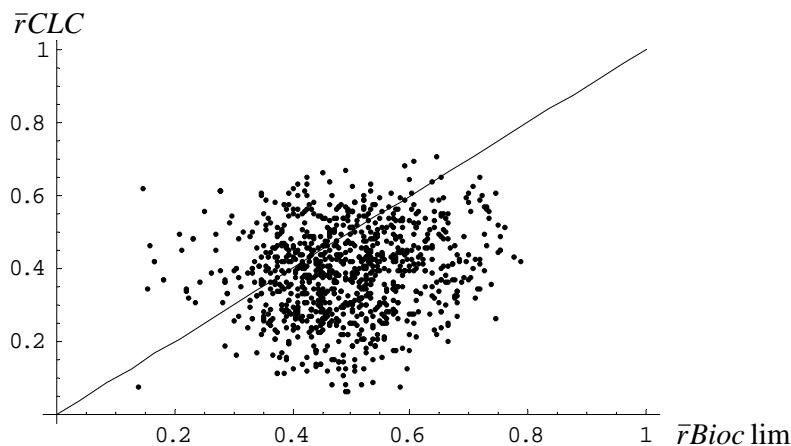


Figure 30 : Comparaison des estimations moyennes de  $rBioclim$  et  $rCLC$ ., réalisées pour le même ensemble d'espèces végétales de la Drôme (avec plus de 20 mailles agrégées avec probabilité d'occupation non nulle). Les tests de non-corrélation entre les deux estimations ( $t$ -test sur la base des coefficients de Pearson, ou  $z$ -test non paramétrique sur la base du  $\rho$  de Spearman) donnent des  $p$ -value proches de 0. Il y aurait donc corrélation entre les deux estimations. Nous notons cependant la très forte dispersion du nuage. La droite représentée est  $y = x$

#### 4. Impact des activités humaines sur la distribution spatiale des végétaux

Nous avons réalisé deux séries d'inférence du paramètre de dynamique métapopulationnelle  $r$  : l'une repose sur une vision bioclimatique de l'habitat, l'autre sur une vision de l'habitat en terme d'occupation des terres par l'homme (variables *Corine Land Cover*, CLC). Il n'y a pas de vision *a priori* meilleure qu'une autre, et il faut partir du principe que nous ne pouvons pas connaître précisément et sans erreur les exigences des espèces en terme d'habitat. En revanche, comparer les estimations obtenues pour chaque type d'habitat potentiel nous renseigne sur l'écologie des espèces, et en particulier sur leur réponse aux activités humaines actuelles.

##### 4.1. Vision relativiste de l'habitat et de la dynamique métapopulationnelle

Il est possible de comparer les estimations obtenues avec l'habitat bioclimatique et avec l'habitat CLC en construisant un intervalle de confiance à 95% des différences entre estimations locales par les deux approches ( $rBioclim - rCLC$ ). Si l'intervalle de confiance résultant contient 0, les deux estimations ne sont pas significativement différentes. Si l'intervalle est entièrement positif, alors  $rBioclim > rCLC$ . On considère alors que l'habitat CLC est un habitat plus pertinent, car permettant une meilleure dynamique de l'espèce (Principe 2). A l'inverse, si l'intervalle est entièrement négatif, alors  $rCLC > rBioclim$  et l'habitat bioclimatique est plus pertinent.

Sur les 894 espèces de plantes présentes dans au moins 20 des 284 mailles de l'échantillonnage agrégé, 565 (63.2%) correspondent à  $rBioclim > rCLC$ , et donc à une écologie spatiale fortement modelée par l'Homme, 149 (16.67%) à  $rCLC > rBioclim$ , et à une prépondérance des effets bioclimatiques, et les 180 restantes (20.13%) ne présentent pas de différence significative entre les deux modèles.

Il est possible de tester la statistique de  $\chi^2$  d'écart à la tendance générale, pour chaque famille botanique. Nous constatons que sur les 101 familles considérées, une seule diffère significativement de la tendance générale : la famille des orchidées. Le seuil de significativité

à 5% de ce test d'indépendance vaut environ 11 : les orchidées possèdent une statistique de 13.37, et la seconde famille au score le plus élevé est à seulement 8.98 (fabacées). Les orchidées ressortent donc nettement.

**La singularité de la famille des orchidées est ainsi réaffirmée de manière éclatante.** Il apparaît que l'écart ainsi testé exprime une prédominance d'espèces fortement modelées par l'Homme : 27 espèces sur 28 ! Ce fait, associé à l'existence d'une composante spécifique « orchidées » dans l'analyse des communautés (voir V.2.5), suggère que l'écologie de cette famille est très singulière, mais aussi notablement liée aux activités humaines.

Pour aller plus loin dans la confrontation des effets d'habitat potentiel, on peut estimer  $r$  en ayant recours à l'habitat « total », qui regroupe les deux groupes de variables. C'est un habitat surparamétré qui sert alors de modèle nul pour évaluer les modèles bioclimatique et CLC. On peut tester la différence entre ce modèle nul et chacun des deux autres modèles en construisant des intervalles de confiances pour  $(r_{Bioclim} - r_{All})$  et  $(r_{CLC} - r_{All})$ . Quand l'un des deux modèles n'est pas significativement différent du modèle nul, il peut être considéré comme pertinent en excluant l'autre modèle. Si en revanche les deux modèles ne sont pas significativement différents du modèle nul, cela signifie qu'ils peuvent tous deux être considérés comme pertinents. Dans ce dernier cas, le fait que la réponse de l'espèce soit confondue pour les deux modèles écologiques signifie qu'il existe une relation entre les variables bioclimatiques et les variables CLC auxquelles elle répond.

A noter que le paramètre  $r$  nul n'est pas nécessairement inférieur aux deux autres. En effet, un meilleur ajustement statistique des données de présence, avec un plus grand nombre de variables d'habitat, ne garantit pas que l'occupation du milieu par l'espèce soit plus importante au sens de la dynamique métapopulationnelle.

Il faut enfin veiller à ne pas faire une interprétation trop simpliste des différences observées entre les deux groupes de variables d'habitat. De fait, la prise en compte des activités humaines n'est pas le seul élément différenciant entre ces deux groupes :

- Les variables d'habitat bioclimatiques et CLC ne sont pas de même nature. Dans le jeu de variables bioclimatiques, les variables climatiques sont quantitatives et les variables géologiques binaires (avant agrégation spatiale). Dans le jeu de variables CLC, les variables de base sont toutes binaires. Les variables binaires auront tendance à traduire une réponse tranchée des espèces, surtout quand leur habitat est spécialisé (effet commutateur). Les variables quantitatives seront plus sensibles à l'existence de gradients

progressifs dans la réponse des espèces. Les différences de réponse entre les deux groupes de variables peuvent donc traduire la nature tranchée ou graduelle de la réponse d'une espèce à son environnement. Une réponse tranchée, plutôt qu'un effet de la structuration du milieu par l'Homme, peut représenter alors une caractéristique biologique intrinsèque de l'espèce. Cette interprétation doit être discutée au cas par cas.

- Certaines catégories d'occupation des sols (zones en eaux, sables, rochers) ne traduisent pas directement l'existence d'activités humaines, mais plutôt des contextes écologiques naturels et particuliers, non pris en compte pas les variables bioclimatiques. Certaines espèces, notamment rupicoles, répondront donc plus fortement aux variables CLC, alors que leur milieu peu accessible les préserve justement de l'emprise des activités humaines. C'est le cas par exemple de *Kernera saxatilis* ou de la silène acaule *Silene acaulis*. Toutefois, cette limite ne concernera qu'un nombre limité d'espèces spécialisées et relativement rares.

Nous analysons par la suite les estimations de  $r_{\text{Bioclim}}$  et  $r_{\text{CLC}}$  obtenues pour un ensemble d'espèces végétales de la Drôme. Nous montrerons les aspects écologiques généraux que traduisent ces estimations, en veillant à conserver une interprétation nuancée et circonstanciée des caractéristiques mesurées.

## 4.2. La fable de l'orchidée et de la fougère

### 4.2.1. La néottie nid d'oiseau *Neottia nidus-avis*

Pour illustrer ces idées, considérons l'exemple de l'orchidée *Neottia nidus-avis* (Figure 31). Ses graines légères et nombreuses facilitent sa dispersion, mais l'installation et le maintien des populations est tributaire des conditions locales, tant en terme de milieu physique que biotique, avec notamment l'existence de relations symbiotiques mycorrhiziennes (obligatoire à la germination) et peut-être une tendance saprophytique.

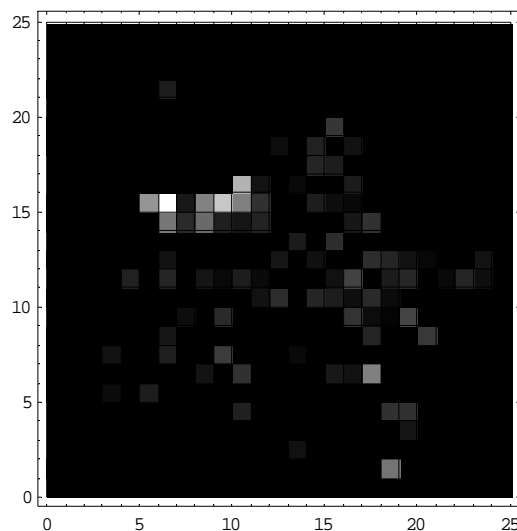
On obtient une valeur de  $r$  plus petite et une moindre variance d'estimation quand l'on considère les données CLC (occupation des sols), par rapport aux données bioclimatiques (Figure 31*b* et *c*). Le caractère montagnard de cette espèce est certes bien prédit par son habitat potentiel bioclimatique (Figure 31*d*), mais l'influence de l'homme, en terme d'exploitation forestière, est un déterminant prépondérant de sa distribution actuelle. La différence entre les prédictions de  $r$  dans les deux situations est significative au seuil de 5%.

De manière intéressante, cette prévalence des activités humaines est constatée également pour la plupart des espèces d'orchidées traitées dans notre analyse : sont concernées 27 des 28 espèces occupant au moins 20 mailles sur les 284 de l'échantillonnage agrégé. Sur l'ensemble des espèces du genre, seules deux espèces présentent une préférence bioclimatique. Elles occupent cependant moins de 10 mailles de l'échantillonnage agrégé, et ne seront dès lors pas considérées.

La préférence pour les variables CLC concerne d'autres orchidées forestières, mais aussi des orchidées de pelouses et de garrigue. La présence de ces espèces est de fait étroitement inféodée aux régimes d'exploitation pastorale. Si l'orchidée est très sensible aux régimes d'occupation par l'Homme, nous pouvons dès lors y voir deux explications non exclusives :

- L'Homme modifie la disponibilité de l'environnement biotique par rapport à une distribution sous dépendance bioclimatique : distribution artificielle des zones ouvertes et fermées, différente d'une zonation naturelle.
- La nature de l'intervention humaine (ici forestière et pastorale) modifie les interactions entre les organismes, favorisant ou non l'instauration des relations symbiotiques. L'intensité du pâturage en particulier modifiera la composition des communautés, avec des rétroactions sur les conditions micro-climatiques.

(a) Données de présence de *Neottia nidus-avis*.





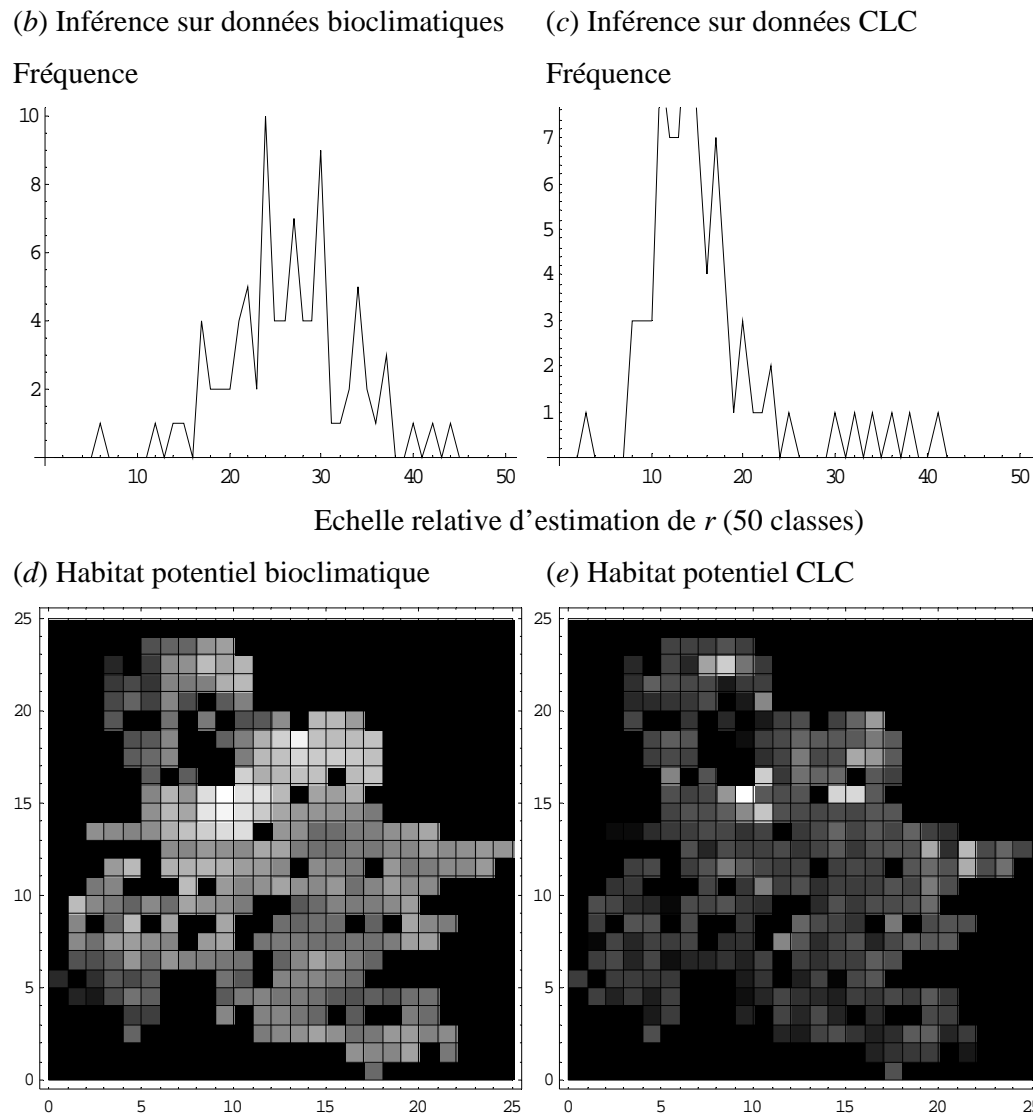


Figure 31 : Etude de la réponse de l'orchidée *Neottia nidus-avis* aux variables bioclimatiques et aux variables d'occupation des sols (CLC). La distribution des présences agrégées est donnée par (a). Les habitats potentiels bioclimatique et CLC sont donnés respectivement par (d) et (e). On obtient une meilleure estimation de  $r$  avec l'habitat potentiel CLC (c), qu'avec les variables bioclimatiques (b).

#### 4.2.2. Le capillaire noir *Asplenium adiantum-nigrum*

Le capillaire noir *Asplenium adiantum-nigrum* est une fougère forestière et saxicole médio-européenne présentant comme la néottie des propagules légères et nombreuses facilitant la dissémination. Cette espèce n'a cependant pas besoin de développer de relation symbiotique pour s'installer et persister.

A l'inverse de la néottie, il s'avère que les données bioclimatiques minimisent l'estimation du paramètre  $r$  par rapport aux données CLC (différence significative à 5%), et donc que la présence du capillaire noir dépend davantage des contraintes bioclimatiques que de la structuration de l'habitat par l'Homme (Figure 32).

De manière intéressante, d'autres espèces forestières non symbiotiques à graines très dispersives présentent également une empreinte plus forte des conditions bioclimatiques, avec par exemple : la campanule carillon *Campanula medium* (anémochorie, graines légères), la viorne lantane *Viburnum lantana* (ornithochorie, baies), l'épervière glauque *Hieracium glaucinum* (anémochorie, achènes à aigrette).

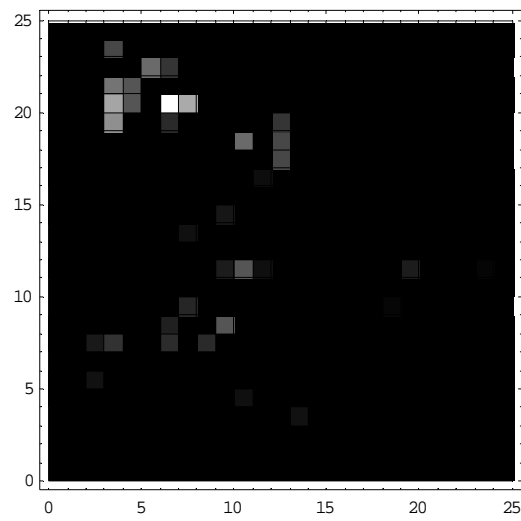
Le comportement des espèces arborées est également symptomatique de l'action de l'homme. Les espèces exploitées récemment ou actuellement pour le bois de chauffage et de construction présentent un  $r$  plus faible avec les données CLC : *Abies alba*, *Fagus sylvatica*, *Populus nigra*. En revanche les espèces à caractère pionnier sont généralement davantage déterminées par les conditions bioclimatiques : *Acer monspessulanus*, *Tilia platyphyllos*, *Fraxinus excelsior*. *Robinia pseudoacacia* fait exception, mais rappelons que cette essence exotique a été introduite récemment par l'Homme.

Les cas du chêne blanc *Quercus humilis* et du chêne vert *Quercus ilex* méritent enfin d'être développés :

- Le chêne vert répond aussi bien aux variables d'occupation des sols qu'aux variables CLC. D'une part, sa distribution a été largement façonnée par l'exploitation intensive en taillis pour le bois de chauffe. D'autre part, l'abandon de ce genre de pratiques tend à favoriser le retour d'une dépendance plus marquée aux conditions bioclimatiques. Par ailleurs, en limite climatique de son aire de distribution, l'espèce est susceptible d'être davantage sensible aux conditions bioclimatiques
- La présence du chêne blanc en revanche est nettement mieux exprimée par les conditions bioclimatiques, ce qui va dans le sens d'une distribution « plus naturelle » de cette espèce. Elle n'a de fait pas subi d'exploitation intensive par l'Homme.

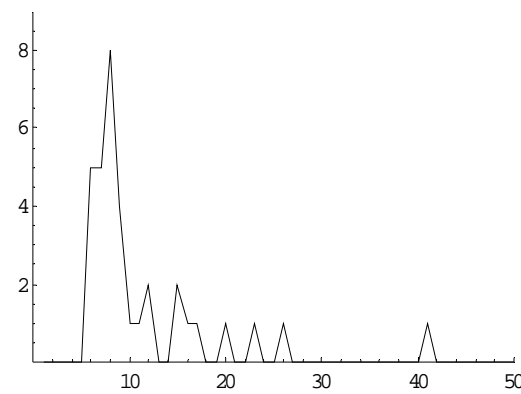
Les deux autres chênes médio-européens de la Drôme, *Quercus robur* et *Quercus petraea*, ne montrent pas de préférence significative au seuil de 5% pour les variables bioclimatiques ou CLC. L'exploitation humaine existe et a existé mais ne semblent pas avoir profondément affecté les distributions bioclimatiques supposées de ces chênes.

(a) Données de présence de *Asplenium adiantum-nigrum*.



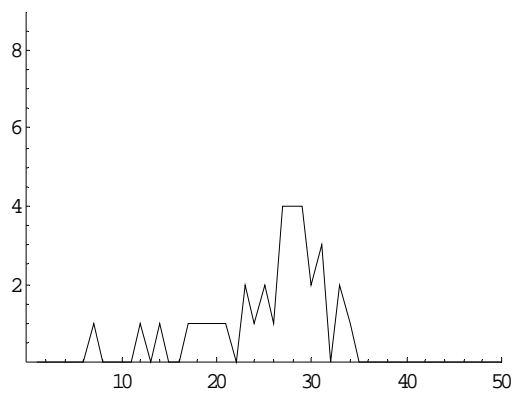
(b) Inférence sur données bioclimatiques

Fréquence



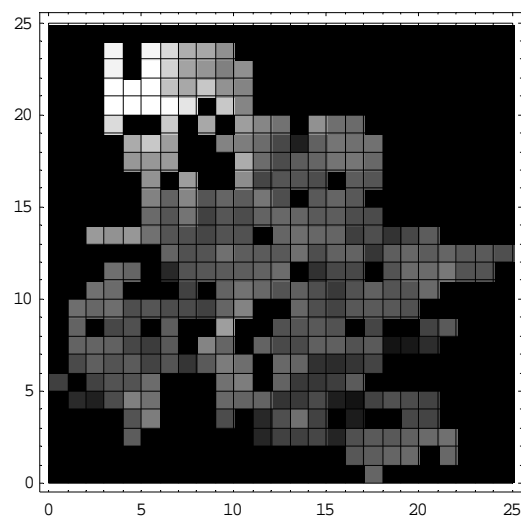
(c) Inférence sur données CLC

Fréquence



Echelle relative d'estimation de  $r$  (50 classes)

(d) Habitat potentiel bioclimatique



(e) Habitat potentiel CLC

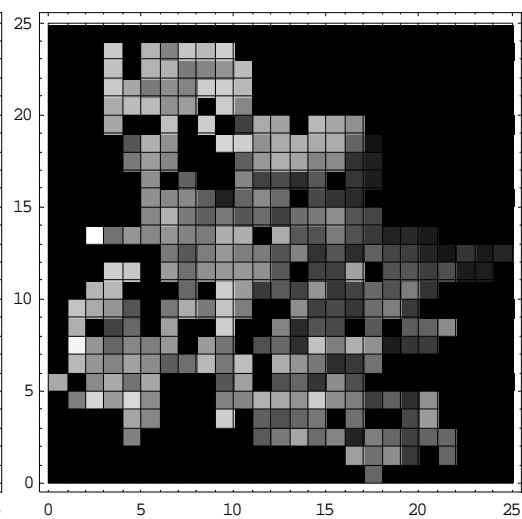


Figure 32 : Etude de la réponse de la fougère *Asplenium adiantum-nigrum* aux variables bioclimatiques et aux variables d'occupation des sols (CLC). La distribution des présences agrégées est donnée par (a). Les habitats potentiels bioclimatique et CLC sont donnés respectivement par (d) et (e). On obtient ici une meilleure estimation de  $r$  avec l'habitat bioclimatique (b), qu'avec les variables CLC (c).

#### 4.3. Syndrome de rudéralité

Les espèces rudérales sont des espèces faiblement compétitives, capable de supporter des perturbations importantes. Elles possèdent un cycle de vie court et présentent une arrivée rapide au stade reproductif.

De manière non surprenante, les espèces rudérales répondent généralement mieux aux variables d'occupation des sols qu'aux variables bioclimatiques. Leur affinité pour les milieux perturbés les rend sensibles aux régimes d'exploitation des terres par l'Homme. Citons notamment des espèces suivantes, dans l'ordre décroissant des effets les plus forts : *Sonchus arvensis*, *Anagallis arvensis*, *Setaria verticillata*, *Crepis setosa*, *Datura stramonium*, *Hypochaeris radicata*, *Fallopia convolvulus*, *Heliotropium europeum*, *Sonchus oleraceum*, *Rapistrum rugosum*, *Picris echioides*, *Kichxia spuria*. Il s'agit là de plantes annuelles, ou rarement vivaces, des premiers stades de succession secondaire. Les espèces ainsi énumérées sont associées à des perturbations directes liées à l'Homme, dans les zones urbaines, autour des aménagements routiers ou dans les cultures.

Dans certains autres cas notables cependant, des espèces présentant une affinité bien connue pour les contextes urbains et cultivés possèdent également des biotopes « primaires » dans la Drôme. Il peut s'agir notamment de pelouses écorchées et d'arènes à thérophytes. Dans ces cas, on se trouve souvent dans des configurations où les variables bioclimatiques sont de meilleurs prédicteurs de la distribution des espèces : signalons en particulier *Arenaria serpyllifolia*, *Myosotis arvensis*, *Geranium columbinum* ou *Senecio vulgaris*.

Bien sûr, cela ne signifie pas que les biotopes en question soient exempts d'action humaine : les régimes d'exploitation des prairies et forêts sont évidemment importants. Ce sont toutefois des régimes de perturbation plus anciens, moins intensifs et moins discontinus dans la durée que les pressions liées à l'agriculture et au développement urbain. Dans le contexte naturel d'une mosaïque d'habitats à des stades successionnels différents, les espèces peuvent acquérir au cours du temps une distribution tendant vers leurs exigences bioclimatiques. A titre d'hypothèse, les régimes de pâturage contrôlés par l'Homme peuvent être similaires, dans la limite de leur caractère extensif, aux régimes de perturbations naturels par des herbivores sauvages.

## 5. Epilogue

Les premiers éléments provenant de l'analyse des données de la Drôme sont fort prometteurs. De nombreuses pistes s'ouvrent à nous, ne demandant qu'à être explorées et les hypothèses écologiques à être testées. Les développements méthodologiques ont porté exclusivement sur des espèces prises individuellement. Mais les premiers résultats montrent l'existence de groupes d'espèces à comportement similaire, et les effets de la structuration des communautés pourront être prégnants.

Pourquoi pas une écologie fonctionnelle à grande échelle spatiale ? L'examen des assemblages d'espèces en fonction de leurs traits fonctionnels (Plant Functional Types, voir Lavorel et al. 2005) est une motivation fondamentale de l'écologie fonctionnelle. La recherche de règles d'assemblage transcendant les types taxonomiques, ainsi que les contingences biogéographiques et historiques, est un moyen de mesurer le poids relatif des processus écologiques dans la construction de la diversité biologique (voir par ex. Bellwood et al. 2002). Dès lors, mettre en regard un trait de dynamique spatiale déterminant la persistance des espèces à grande échelle, tel notre paramètre  $r$ , et les traits fonctionnels contrôlant leur persistance locale permettra de porter un nouveau regard sur l'écologie et l'évolution des communautés.

## Synthèse et discussion

Les questions initialement posées étaient :

- Comment mesurer de manière écologiquement pertinente la structure spatiale des populations ?
- Comment relier cette structure aux processus écologiques sous-jacents ?

Cette recherche d'une relation générale entre un processus dynamique et sa structure spatiale émergente à plus grandes échelles suscite naturellement au premier abord un certain scepticisme. On peut penser que l'information spatiale n'est pas assez riche pour rendre compte, de manière univoque, de tous les paramètres de la dynamique locale. En outre, on peut penser que la complexité du système à grande échelle entrave les possibilités de modélisation et de compréhension simple de cette relation.

Ces réserves *a priori*, je les ai fréquemment rencontrées. Toutefois elles semblent exprimer davantage l'empreinte des paradigmes dominants (voir Introduction) que de réels obstacles. Là où la culture actuelle tend à représenter précisément des mécanismes causaux au sein des systèmes écologiques, nous promouvons une utilisation judicieuse de l'incertitude (Chapitre II et Chapitre III). Là où un certain réductionnisme incite à dissocier l'action des processus à différentes échelles (Introduction et Chapitre II), nous mettons en exergue l'importance d'une prise en compte plus large et interactive des effets d'échelle (Chapitre II).

Les enseignements tirés de ces études font ainsi ressortir trois conclusions essentielles :

- *Principe d'incertitude* : il faut savoir sacrifier un peu de précision d'estimation pour gagner en robustesse et en pertinence, que ce soit dans le domaine réel ou spectral (Chapitres III et IV).
- *Importance des échelles intermédiaires* : l'estimation analytique sur des statistiques agrégées dans l'espace est optimale quand on considère une approximation locale sur les probabilités d'occupation (Chapitre IV). L'imprécision ainsi introduite réduit le biais d'estimation attendu pour un modèle analytique ne prenant pas en compte le phénomène d'auto-organisation. Le principe d'incertitude en analyse spectrale conclut de manière analogue dans le domaine spectral (Chapitre III) : on doit travailler à des échelles intermédiaires de précision sur l'estimation des fréquences, grâce au lissage.

- *Habitat potentiel* : il est nécessaire de définir une enveloppe d'habitat potentiel dans laquelle il est possible d'inférer la dynamique des espèces (Chapitres IV et V). L'estimation du paramètre dynamique est donc *relative* à une certaine conception de l'habitat de l'espèce. Dans ces conditions, différentes estimations peuvent être réalisées pour différents habitats, et leur comparaison permet le test d'hypothèses écologiques (Chapitre V).

En définitive, les éléments apportés par mon travail ne constituent pas un support théorique nouveau, mais proposent une mise en perspective originale. Cette démarche nous a permis de développer avec succès une méthode d'inférence simple, ouvrant la possibilité de tester des hypothèses inédites sur le fonctionnement des systèmes écologiques (Chapitre V). La synthèse que nous proposons ici, accompagnée de perspectives que nous espérons prometteuses, mettra en avant deux axes forts :

- D'une part, nous avons montré l'opportunité de développer une **approche diversifiée de représentation des processus dynamiques**. C'est en adoptant des angles différents d'étude du système métapopulation, auto-organisation (Chapitre II), recherche informationnelle (Chapitre III) et méthodes analytiques (Chapitre IV), que nous avons pu construire une démarche d'inférence clairement motivée.
- D'autre part, un message récurrent et crucial apparaît tout au long du travail : **l'importance d'identifier clairement les composantes dynamiques et statiques de la niche écologique**. La possibilité de définir et de quantifier deux composantes de la niche écologique, une statique et une dynamique, est au cœur de chaque chapitre, dans la réflexion sur la double criticalité (Chapitre II), à propos du découplage spectral et de la partition de variation (Chapitre III), et enfin sur la démarche d'inférence (Chapitres IV et V).

Ces axes fédèrent l'ensemble des conclusions des Chapitres, au-delà de la diversité des approches. Ils structureront donc le discours présenté en Synthèse.

## 1. Représentations de la dynamique métapopulationnelle : une question d'échelles

Nous avons mis l'accent sur la diversité des approches possibles pour l'étude de la dynamique métapopulationnelle. Malgré cette diversité, nous aboutissons à une convergence de conclusions, qui permet de formuler quelques principes généraux.

### *1.1. Etat quasi-stationnaire – auto-organisation*

La dynamique d'une métapopulation est un processus qui induit dans l'espace et dans le temps des propriétés émergentes complexes et délicates à interpréter. Dans le Chapitre II, nous avons montré qu'il est essentiel de comprendre et de prendre en compte deux aspects fondamentalement différents de cette dynamique :

- D'une part, l'existence d'un **état quasi-stationnaire déterminé** par un modèle stochastique markovien, avec une espérance d'occupation des populations déterminée sur un **temps infini**.
- D'autre part, l'existence de structures émergentes, **induites par des événements contingents**, et associées à des avalanches de colonisations et d'extinctions. Ces structures sont observables dans un **temps fini**, et possèdent des caractéristiques émergentes particulières, représentées par des lois d'échelle.

Nous avons montré que l'identification des structures émergentes induites par auto-organisation est essentielle pour la compréhension de la structure spatiale des populations (Chapitre II), et par extension pour l'inférence des paramètres de la dynamique métapopulationnelle (Chapitre IV).

### *1.2. Démarche analytique d'inférence – principe d'incertitude*

Dans le Chapitre IV (§5), nous avons considéré un modèle d'équilibre entre événements de colonisation et d'extinction de populations. A partir d'une telle représentation simple et unique du processus de dynamique métapopulationnelle, nous avons constaté que les choix de la statistique et de la méthode d'ajustement étaient critiques pour la qualité et la robustesse des estimations du paramètre  $r$ .

Voici un rappel de nos principaux résultats :

- En considérant des séries temporelles d'occupation, un **ajustement basé sur une statistique globale** (moindres carrés) (Chap. V §4.1), fournit des estimations précises de  $r$ , sauf au voisinage du seuil critique d'extinction, où le biais d'estimation augmente dramatiquement. C'est une **méthode précise, mais peu robuste**.
- En considérant toujours des séries temporelles d'occupation, l'étude de la **distribution des estimations locales de  $r$** , dont on tire la moyenne et la variance spatiales du paramètre  $r$ , fournit des estimations certes plus imprécises, mais l'imprécision n'augmente pas au voisinage du seuil d'extinction (Chap. V §4.2). Cette imprécision est homogène quelles



que soient les valeurs de  $r$ . **Cette méthode d'ajustement est donc moins précise, mais plus robuste.**

- A défaut de répliquats temporels, l'utilisation d'une **statistique agrégée dans l'espace** de présence de populations donne de bons résultats si l'on considère une approximation locale sur ces probabilités (Chap. V §4.3). Le modèle repose donc sur un transfert d'échelle du niveau local, à un niveau intermédiaire. Dans ces conditions, nous obtenons une **méthode d'estimation relativement peu précise, mais robuste.**
- Enfin l'**estimation conjointe de la qualité d'habitat et du paramètre de dynamique métapopulationnelle  $r$**  est possible à partir de séries temporelles d'occupation, mais s'avère très peu robuste (Chap. V §4.4). Dans environ la moitié des cas, une très bonne estimation des paramètres d'habitat et de dynamique est obtenue, dans les autres cas le modèle induit des estimations aberrantes. **Il n'est donc pas pertinent de rechercher une estimation simultanée de la composante statique et dynamique des espèces à partir de la distribution spatiale des populations.**

Il est évident qu'un modèle ne peut pas expliquer ou décrire l'ensemble de l'information mesurée dans un système écologique. A des fins d'inférence, la pertinence d'un modèle théorique dépendra de la capacité du modélisateur à formuler un bon compromis entre qualité d'estimation et stabilité de cette qualité d'estimation dans une gamme de situations la plus large possible. C'est ce que nous avons appelé *robustesse* du modèle.

Le principe d'incertitude doit donc guider notre démarche : il faut savoir sacrifier un peu de précision pour pouvoir gagner en stabilité et en pertinence. L'importance d'une telle attitude a été ainsi montrée clairement et doublement : d'une part dans le cadre de l'analyse spectrale (Chapitre III), d'autre part pour la modélisation en espace réel (Chapitre IV). Dans ce contexte, les tests dit de probabilité exacte sur les structures spatiales doivent être considérés avec une extrême circonspection.

### *1.3. Paradigme des échelles intermédiaires*

L'auto-organisation est un phénomène au cœur de nos résultats théoriques. Il est souvent perçu comme quelque chose d'ésotérique, difficile à comprendre et à manipuler. La recherche d'une caractérisation précise des processus sous-jacents n'était pas notre préoccupation. Simplement ce phénomène a nourri notre réflexion en orientant notre

démarche d'inférence et en façonnant notre compréhension de la dynamique métapopulationnelle (Chapitres III et IV). La leçon fondamentale est que la meilleure manière de circonvenir aux biais d'inférence induits fréquemment par de telles structures est de se placer à une bonne échelle spatiale, intermédiaire entre l'échelon local et les structures émergentes globales.

Avec un formalisme quelque peu différent, Casagrandi et Gatto (1999) ont adopté une logique similaire : une approximation locale de la structure métapopulationnelle leur a permis de placer l'échelle spatiale intermédiaire au cœur de leur modélisation, et d'obtenir ainsi des prédictions très satisfaisantes. Les approches stochastiques basées sur des statistiques agrégées globales sont *a contrario* conceptuellement et pragmatiquement inappropriées pour traiter des situations au voisinage du seuil d'extinction. Cela met en question le recours à des approches héritées de modèles markoviens de type vie-mort, où les états sont des classes globales de populations (voir Hill and Caswell 2001).

L'auto-organisation induit en particulier un problème majeur d'estimation dans le domaine temporel : l'existence de corrélations à grande échelle temporelle, notamment au voisinage du seuil critique d'extinction, engendre des biais importants dans l'estimation des probabilités moyennes temporelles. De fait, les attendus théoriques des modèles stochastiques sur les probabilités quasi-stationnaires sont asymptotiques, sur un temps infini. Or il peut exister des régimes transitoires, aussi bien que des fluctuations temporaires au sein de l'état quasi-stationnaire, sur des temps extrêmement longs (voir Encadré 3).

La notion d'échelle intermédiaire a donc également une grande importance dans le temps. A l'échelle des temps écologiques, on devra donc souvent considérer des situations transitoires et des fluctuations temporaires. Ce problème a été notamment formulé par Labra et al. (2003), dans le cadre des *meso-transients* et *super-transients*, régimes transitoires particuliers à la limite du chaos. Dans ces situations, savoir comprendre et traiter l'effet des événements contingents est important pour comprendre la dynamique évolutive des systèmes écologiques.

#### 1.4. Péroration - le démon de Maxwell

D'un point de vue thermodynamique, la formation spontanée de structures auto-organisées à des échelles intermédiaires correspond fondamentalement à un mécanisme de dissipation d'énergie, et plus fondamentalement à un mécanisme compensatoire de retour à un état

entropique proche de l'équilibre thermodynamique. L'existence de telles structures au sein des systèmes écologiques souligne l'importance de comprendre ces questions thermodynamiques pour éclairer le lien entre processus écologiques locaux et structures résultantes.

Bien connu en physique, le démon de Maxwell illustre cette idée (voir Kauffman 1993). Dans un système à deux compartiments remplis de gaz, cette créature imaginaire permet uniquement aux particules gazeuses les plus énergétiques de passer d'un compartiment à un autre via une valve. En créant un différentiel de pression entre les compartiments, le démon emmagasine de l'énergie : il peut utiliser le moment opportun le travail résultant du rétablissement de l'équilibre entre les deux compartiments. Les systèmes biologiques fonctionnent comme des démons de Maxwell en créant des gradients d'énergie importants, via notamment de nombreuses interfaces *discontinues* (membranes cellulaires, peau...).

La dissipation de l'énergie au sein de ces systèmes donne alors lieu à l'apparition de structures auto-organisées. Au sein d'un organisme, il peut s'agir d'arrangements cellulaires, de chaînes d'activation engendrant les structures particulières du pelage d'un guépard... A plus grande échelle, la dissipation d'énergie peut se manifester au travers d'événements catastrophiques spectaculaires : des épidémies, des feux de forêts... C'est un peu comme si les systèmes biologiques constituaient des barrages hydroélectriques, entravant le fonctionnement thermodynamique naturel, pour récupérer de l'énergie en relarguant de l'eau de temps en temps.

Selon Kauffman (1993), la sélection naturelle fonctionne comme un démon de Maxwell. Par le biais de l'hérédité, la sélection est un processus à mémoire qui oriente l'évolution des organismes. A une échelle « écologique » de temps et d'espace, de nombreux processus non linéaires dissipatifs sont impliqués dans le fonctionnement des êtres vivants et de la biosphère en général. Nous avons insisté sur le fait que de tels processus sont propres à engendrer de l'ordre et des structures de grandes amplitudes à partir de dynamiques stochastiques et interactives localisées. Ces structures ne possèdent pas de fonction biologique directement codée par un programme génétique. Cependant elles peuvent favoriser à grande échelle la survie d'un ensemble d'individus ou de populations. Dès lors les conditions locales et individuelles favorisant l'apparition de telles structures améliorent la survie des groupes, et indirectement la survie des individus. Les structures dissipatives auto-organisées peuvent en définitive trouver une place légitime dans l'évolution, non en tant que mécanisme mais en tant que support de la sélection, et non parce qu'elles sont sélectionnées

directement, mais parce que la sélection pourra opérer *indirectement* sur des aspects individuels propices à leur apparition. En favorisant des structures hors équilibre thermodynamique, la sélection naturelle agit alors bien comme un démon de Maxwell. Aussi structuralisme et néo-darwinisme pourraient-ils trouver, par une sage conciliation, un terrain commun à l'étude de l'évolution (Eldredge and Gould 1972). Le fait que des éléments extérieurs au génome puissent faire partie intégrante d'un système de développement évolutif focalise néanmoins l'âpre opposition entre « évo-dévo » (*evolutionary developmental biology*) et théorie des systèmes développementaux (*developmental system theory*) (Griffiths and Gray 2005). L'antagonisme stérile des dogmes demeure prévalant, engendrant des dérives idéologiques au détriment de l'esprit scientifique (voir par ex. Webster and Goodwin 1982).

Dans les métapopulations, la parabole du démon de Maxwell souligne l'importance de la nature discrète des populations. Il y a un intérieur et un extérieur dans une population, de même qu'il y a un intérieur et un extérieur dans un individu ou dans une cellule. C'est fondamentalement parce que les populations sont des entités discrètes, dans un état binaire (soi et non-soi), que notre modèle de dynamique métapopulationnelle peut présenter des phénomènes d'avalanches. Celles-ci sont alors des dissipations locales d'énergie, ou plutôt des « relargages » d'entropie, quand les configurations locales s'éloignent de l'équilibre.

Cette vision de la métapopulation peut nourrir les réflexions sur la dynamique spatiale des espèces sur des temps évolutifs. La population naît, grandit, mûrit et enfin meurt. De nouvelles populations doivent nécessairement voir le jour pour la survie à long terme de l'espèce. Le devenir d'un individu dans la population n'est alors pas le même que celui d'un individu hors de la population, ce dernier pouvant éventuellement occasionner la naissance d'une autre population. Les avalanches éventuelles peuvent faciliter les phases de propagation des organismes. Une structure discrète et close des populations occasionne des gradients d'énergie favorables à ce type d'événement.

## **2. Représentations de la niche écologique : habitats *a priori* et *a posteriori***

La clarification des concepts de niche et d'habitat, dans le cadre opérationnel de l'objectif d'inférence, est une préoccupation majeure de ma démarche (voir Introduction). Elle aboutit à une représentation duale, avec une composante statique et dynamique de la niche, qui s'avèrent dans mon travail porteuses d'informations séparables et pertinentes. La composante

dynamique exprime le processus métapopulationnel, et sa détermination fait en particulier appel aux principes énoncés précédemment en partie 1.

Les notions d'habitat *a priori* et *a posteriori* (voir Introduction) constituent la formalisation spatiale de ces idées, respectivement pour la niche statique (au sens Hutchinson 1957), et pour les phénomènes dynamiques restreignant la définition initiale de la niche (Pulliam 2000). Nous récapitulons ici les conclusions apportées sur ce thème, démontrant la valeur théorique et pratique d'une telle représentation.

### 2.1. Découplage des composantes statiques et dynamiques de la niche

La structure spatiale des métapopulations manifeste l'interaction forte entre la structure spatiale de l'habitat et la dynamique métapopulationnelle. L'idée d'un tel couplage est mise en exergue par le concept de double criticalité (Chapitre II) :

- Criticalité de type transition de phase, exprimant la nature non-linéaire de la dynamique métapopulationnelle, y compris en habitat uniforme et en modèle champ-moyen.
- Criticalité de type percolation, où la structure de l'habitat détermine les chemins possibles d'occupation dans l'espace. L'accessibilité de l'espace ne varie pas linéairement avec la densité et l'aggrégation de l'habitat, et impose un phénomène de criticalité.

Nous avons vu néanmoins qu'il était possible de dissocier les signatures spatiales de l'habitat et de la dynamique métapopulationnelle, à la faveur de l'emboîtement des processus, et grâce à l'analyse spectrale (Chapitre III). Cette analyse a permis de donner un sens tangible à la dichotomie habitat *a priori* / habitat *a posteriori*. Celle-ci se situe clairement dans une conception écologique duale des processus statiques et dynamiques, l'habitat *a priori* étant assimilé à une structure statique dans le temps et dans l'espace. Comme pour le concept de métapopulation, on dissocie le rythme de variation de l'habitat de celui des populations et métapopulations évoluant à l'intérieur.

Le phénomène de double criticalité représente ainsi quelque chose d'asymétrique, l'aspect percolation traduisant la contrainte de l'habitat *a priori* statique sur la dynamique des populations : il y a une criticalité statique et une criticalité dynamique. Dans l'analyse spectrale, c'est de fait l'habitat *a priori* qui est caractérisé par une référence d'échelle (valeur seuil de la Composante Principale associée), tandis que l'habitat *a posteriori* n'a pas d'échelle de référence propre (première Composante Principale).

L'idée de découplage n'est donc pas ici un simple artifice technique dévolu à l'inférence « propre » des paramètres statiques et dynamiques de la niche. Elle fait écho en particulier à l'idée d'auto-organisation et de structure émergente. On constate que l'on peut dissocier les signatures fréquentielles des habitats *a priori* et *a posteriori* dans une logique d'emboîtement des échelles, les populations évoluant au sein de l'habitat favorable. Cela ne signifie pas néanmoins qu'il y a une dissociation stricte d'échelle entre structures induites par l'habitat *a priori* et structures induites purement par dynamique métapopulationnelle. Ces structures possèdent des signatures spectrales à toutes les échelles.

La représentation indépendante des processus aux niveaux microscopique – mésoscopique – macroscopique que nous avons évoquée précédemment est d'une autre nature. Elle repose sur l'idée que les structures émergentes à une échelle supérieure possèdent des propriétés indépendantes des événements individuels et contingents se produisant à l'échelle inférieure. Ainsi peuvent-elles être caractérisées sans référence directe aux processus des échelles inférieures.

Dans le cadre de la représentation des habitats, ce n'est pas en dissociant les échelles d'analyse que l'on peut donner un sens aux composantes statiques et dynamiques de la niche. C'est au contraire par leurs caractéristiques « inter-échelle », avec un coefficient de loi de puissance par exemple, que ces structures peuvent être comprises en terme de composante dynamique ou statique de la niche. Cette idée reste à approfondir sur le plan théorique.

C'est finalement là que l'on peut concevoir et comprendre l'intérêt de l'analyse spectrale, propre à mettre en exergue ces structures, comme nous l'avons démontré avec l'analyse des métapopulation simulées. Elle permet en particulier de déceler les structures fractales n'ayant pas d'échelle caractéristique.

## *2.2. Notion d'habitat potentiel et modèle régressif*

Deux résultats théoriques importants expriment donc la nature de la relation entre composante dynamique et composante statique de la niche : la double criticalité (Chapitre II), et le découplage spectral (Chapitre III).

Dans le Chapitre V, on introduit l'idée d'habitat potentiel, notion en particulier consacrée par les écologues et biogéographes pour décrire statistiquement le lien entre présence des espèces et variables d'habitat (voir Guisan and Zimmermann 2000). Dès lors notre méthode d'inférence préconisée pour le traitement des données de la Drôme procède en deux étapes :

- Définition d'un habitat potentiel. On exprime une relation linéaire entre les occurrences agrégées spatialement et les variables d'habitat agrégées :  $p(x, y) = \alpha \cdot h(x, y) + \varepsilon(x, y)$ , avec  $h(x, y)$  la composante d'habitat obtenue par régression linéaire, normalisée pour une variation entre 0 et 1, et  $\varepsilon(x, y)$ , les résidus obtenus. **La régression est réalisée pour toutes les valeurs de  $p(x, y)$ , aussi bien nulles que non nulles.**

C'est une logique qui correspond à une représentation classique statique de la niche, une représentation de l'**habitat *a priori***. Toutefois il s'agit bien d'une *représentation* et non d'une *description* de l'habitat *a priori*, car elle est entachée de nombreux biais, comme nous l'avons souligné au Chapitre III. S'il s'agit là d'une différence fondamentale entre la notion d'habitat potentiel et celle d'habitat *a priori*, il demeure important de souligner que les deux notions convergent vers une prise en compte de l'espace inoccupé aussi bien qu'occupé.

- Inférence de la dynamique métapopulationnelle. La méthode d'inférence pose :

$$\forall (x, y) \text{ tel que } p(x, y) \neq 0, \quad r(x, y) = h(x, y) \cdot (1 - p(x, y)) = h(x, y) \cdot (1 - \alpha \cdot h(x, y) - \varepsilon(x, y))$$

soit encore  $\varepsilon(x, y) = 1 - h(x, y) - \frac{r(x, y)}{h(x, y)}$ , pour  $(x, y)$  tel que  $p(x, y) \neq 0$ .

En conséquence, la distribution des résidus du modèle d'habitat potentiel, **limitée spatialement aux  $(x, y)$  tels que  $p(x, y) \neq 0$** , sera de nature à exprimer au travers de la distribution de  $r(x, y)$  la composante dynamique de la niche. **Cette limitation spatiale exprime la notion d'habitat *a posteriori*.** On retrouve bien l'idée que la composante dynamique de la niche doit être étudiée dans cet espace.

Pour aller plus loin sur la formalisation des résidus  $\varepsilon(x, y)$ , on pourra examiner les propriétés de la distribution de  $r(x, y)$ . Par exemple, si l'on suppose que l'espace géographique étudié correspond à une zone d'occupation homogène, c'est-à-dire un cœur d'aire de distribution sans effet de bord ni d'invasion, alors la caractéristique  $r(x, y)$  est intrinsèque à l'espèce et indépendante de l'espace. On peut la représenter par une distribution normale d'échantillonnage, indépendante de l'espace :  $N(r_{sp}, \sigma_{sample})$ . On aboutit donc assez singulièrement à ce que la dépendance spatiale des résidus soit liée directement à celle de l'habitat potentiel :

$$\varepsilon(x, y) = 1 - h(x, y) - \frac{N(r_{sp}, \sigma_{sample})}{h(x, y)}, \text{ pour } (x, y) \text{ tel que } p(x, y) \neq 0$$

**Dans mon travail, l'habitat potentiel est une représentation préalable de l'habitat *a priori*. C'est une base nécessaire pour pouvoir inférer les caractéristiques de l'habitat *a posteriori*, exprimant la dynamique de l'organisme étudié.**

Les Chapitres II et III ont montré qu'une telle dichotomie avait un sens théorique pour la description de la structure spatiale des métapopulations, un sens écologique dont la recherche fut le premier temps de mon travail. Les Chapitres IV et V montrent que cette dichotomie a un sens pratique et opérationnel, dans la logique d'inférence qui fut le deuxième temps de mon travail.

### *2.3. Intérêt d'une définition arbitraire de l'habitat potentiel*

Pour le traitement de la dynamique des espèces végétales de la Drôme, nous avons proposé de partir d'une définition arbitraire de l'habitat potentiel. Nous avons vu que l'habitat potentiel est une représentation, et non la réalité de l'habitat *a priori*. Dans cette logique, l'habitat *a priori* est unique, mais il peut y avoir une pluralité d'habitats potentiels.

Le choix arbitraire d'une représentation est orienté par la nature des questions posées. Nous nous sommes intéressés en l'occurrence à comprendre l'incidence des activités humaines sur la dynamique des espèces végétales. Pour se faire, nous avons créé deux groupes de variables environnementales, l'une strictement bioclimatique, l'autre relative à l'utilisation des terres par l'Homme. De cette manière, il est possible de définir l'habitat potentiel des espèces au sens de chacun des groupes de variables environnementales.

Dans un second temps, nous inférons le paramètre dynamique  $r$  en confrontant les cartes d'occurrence et les cartes d'habitat potentiel. Nous avons en définitive une distribution spatiale d'estimation de  $r$  pour chacun des deux types d'habitat potentiel définis. Sur l'ensemble de l'espace, la différence moyenne d'estimation du paramètre de dynamique métapopulationnelle entre les deux types d'habitat potentiel permet de tester l'impact des activités humaines sur les dynamiques spatiales. Le critère *ad hoc* de test est qu'un habitat potentiel est plus représentatif s'il conduit *globalement* à l'inférence d'une dynamique significativement plus stable de l'espèce, c'est à dire un  $r$  plus faible.



La démarche générale, valable pour d'autres hypothèses écologiques, est en définitive la suivante :

- Nous sélectionnons des groupes de variables environnementales afin de tester une hypothèse précise, mettant en opposition deux notions d'habitat, ici sous dépendance humaine ou non.
- Nous inférons des habitats potentiels à partir de la distribution des occurrences, pour chacun de ces groupes de variables.
- Nous utilisons la distribution des occurrences dans chaque habitat potentiel pour estimer une dynamique de l'espèce relative à cet habitat. Nous avons vu que cela revenait au traitement des résidus de l'étape précédente.
- Nous testons l'hypothèse de départ en fonction des estimations de la dynamique métapopulation dans les différents habitats potentiels.

Cette approche pose donc un cadre hypothético-déductif pour le traitement d'une hypothèse écologique simple : par exemple, y-a-t-il un effet de l'Homme sur la dynamique spatiale des espèces ? Elle repose ensuite sur une démarche inférentielle, en fait une manière de mesurer la dynamique spatiale des espèces. Il en résulte une statistique de test qui permet de conclure.

La démarche pourra être aisément reproduite et généralisée dans un grand nombre de situations. Tout d'abord avec d'autres matériels biologiques, d'autre part avec d'autres questions écologiques sur l'habitat des espèces. Par exemple, il est possible de tester l'évolution du comportement dynamique des communautés d'oiseaux soumises à une évolution de la végétation, à plusieurs dates.

### **3. Perspectives**

#### *3.1. Pistes d'approfondissement méthodologique*

Le présent mémoire a été développé dans le souci de répondre à une question fondamentale, celle de la nature du lien entre structure spatiale des populations et processus écologiques sous-jacents. Le fait d'aborder plusieurs approches très différentes, de les confronter et d'arriver à des conclusions communes, a permis de montrer l'importance de quelques principes simples.

Le message a été construit de manière à guider le lecteur vers une pleine compréhension de ces principes. Dans cette logique, nous n'avons pas souhaité alourdir le contenu technique du mémoire. Un certain nombre de travaux ont été entrepris néanmoins dans le sens d'un approfondissement des méthodes proposées. Comme nous avons fait le choix de ne pas les aborder explicitement ici, nous présenterons ces axes de travail à titre de perspectives.

- Méta-populations simulées

Nous avons considéré la démarche de simulation de la dynamique méta-populationnelle (Chapitre I) comme une expérience théorique (Peck 2004) destinée à nous éclairer sur certains aspects du lien entre paramètres statique de structure d'habitat ( $p$  et  $q$ ), paramètre dynamique d'extinction/colonisation ( $r$ ), et structure spatiale résultante des populations. Dans cet esprit, nous avons conçu une plate-forme de modélisation (voir Annexe 1), destinée à permettre un test extensif des paramètres de simulation. Outre les trois paramètres  $p$ ,  $q$  et  $r$  évoqués, il s'agit de tester l'effet des conditions initiales, de la taille de la grille de paysage, d'une structure torique ou non. La plate-forme inclut la possibilité de générer de manière illimitée des cartes d'habitat aléatoires pour des paramètres  $p$  et  $q$  donnés.

Nous avons envisagé grâce à cet outil de tester extensivement ces nombreux effets possibles, mais nous nous sommes focalisés finalement dans le mémoire sur l'utilisation d'un seul jeu de cartes et un seul jeu de simulations. Une perspective importante sera donc d'évaluer les différents effets que nous venons d'évoquer.

Par ailleurs nous avons également généralisé la démarche à la prise en compte de paysages composites, dits paysages de qualité d'habitat. Il s'agit d'envisager la combinaison de l'effet de plusieurs variables binaires, et de construire ainsi davantage d'états de qualité d'habitat (par exemple 4 variables binaires = 5 états). Nous avons constaté l'existence de phénomènes particuliers de percolation qui restent à étudier et qui expriment probablement des structures auto-organisées plus complexes. Par ailleurs nous avons utilisé ces simulations pour tester les méthodes d'inférence proposées au Chapitre IV. Ces résultats non montrés par souci de simplicité pourront faire l'objet de développements futurs utiles.

- Modèle analytique spectral

Le choix d'une « bonne » mesure de la structure spatiale des populations est cruciale pour éclairer le lien entre processus dynamiques et structures spatiales résultantes. L'hypothèse

que l'analyse spectrale pouvait représenter une telle mesure pertinente, et permettre l'inférence, a été à ce titre un résultat important (voir Chapitre III).

Dès lors j'ai cherché à exprimer le découplage des signatures spectrales habitat *a priori* / *a posteriori* par un formalisme analytique. Les équations analytiques de base ont été présentées au Chapitre IV, et transposent les équations du domaine réel. Une résolution complète des équations n'a pas été possible, mais deux résolutions partielles indépendantes ont permis d'accéder à un découplage effectif mais partiel des structures d'habitat. Une perspective importante sera de poursuivre le développement de ces modèles, afin d'accéder à un attendu analytique généralisable et opérationnel pour le traitement des données réelles d'inventaires.

- Théorie des ondelettes

La dualité entre espace réel et spectral est théorique et les contraintes liées au principe d'incertitude conduisent en pratique à une rupture de ce lien (voir Chapitre III). Cela permet de mettre en exergue la véritable information spatiale écologiquement pertinente, au lieu d'exprimer de simples contingences de positions, comme nous l'avons montré dans le cadre de l'analyse PCNM (Chap. III §3.2). Toutefois la rupture de cette dualité ne permet pas de prendre en compte les effets d'hétérogénéité spatiale, et en particulier le fait que la structure spectrale puisse changer d'une partie d'une zone d'étude à une autre. C'est une limitation potentielle importante dans la perspective d'applications à des données réelles.

Conscient de ce problème, je me suis intéressé à la théorie des ondelettes, qui permet de prendre en compte une variation spatiale de la structure spectrale. La transposition des résultats semi-analytiques sur la structure spectrale des populations est apparue être peu évidente avec le traitement par ondelettes. Ce serait toutefois une piste importante pour l'avenir, en travaillant sur des estimations locales de la structure spectrale pour estimer le paramètre  $r$ . Pour l'heure, nous avons engagé une analyse descriptive de la distribution spatiale des populations d'*Adonis vernalis* sur le Causse Méjean, afin d'étudier dans quelle mesure le lien entre occurrences et variables d'habitat pouvait dépendre des échelles spatiales de la décomposition en ondelettes (Keitt and Urban 2005).

- Méthode PCNM - inférence spectrale

Nous avons montré que l'utilisation de la méthode PCNM causale, préconisée par Borcard et Legendre (2002), était de nature à exprimer des structures spatiales contingentes, non pertinentes du point de vue de la compréhension générale des processus écologiques sous-jacents (Chap. III §3). Comme l'analyse de Fourier, l'analyse PCNM spectrale permet en

revanche d'exprimer le découplage entre les signatures induites par les habitats *a priori* et *a posteriori*. Elle présente de surcroît l'avantage d'être utilisable pour l'analyse de données échantillonnées irrégulièrement dans l'espace. Nous savons qu'un échantillonnage irrégulier aléatoire n'altère pas les propriétés de découplage, mais la généralisation à des structures irrégulières plus hétérogènes sera importante. L'ACP des spectres de distributions d'oiseaux au Pic Saint-Loup, sur un échantillonnage assez homogène, nous a permis de retrouver les composantes spectrales du découplage observé sur simulations. Il semble donc qu'il y ait une réelle pertinence de ce découplage, et une réelle perspective d'inférence sur des données réelles.

- Méthode d'inférence à l'échelle intermédiaire

La méthode que nous avons mise en exergue à la fin du Chapitre IV et mise en œuvre dans le Chapitre V pose des problèmes conceptuels cruciaux que nous avons déjà évoqués dans la Synthèse (§2). Cette méthode repose sur une équation extrêmement simple qui fait un lien entre les notions d'habitat *a priori* et *a posteriori* que nous avons introduites, les modèles régressifs classiques d'estimation des habitats potentiels, et enfin les modèles classiques de dynamique métapopulationnelle.

Le point fort de cette méthode est son apparente robustesse, confirmée par la qualité des résultats obtenus dans l'étude des espèces végétales de la Drôme. Il restera à élucider et à mesurer plus en profondeur les implications théoriques et pratiques de ces résultats, qui mobilisent l'ensemble des principes mis en évidence dans mon travail. Nous espérons sur le long terme que la maturité conceptuelle qu'acquerra la méthode révélera un potentiel d'ores et déjà prometteur.

### 3.2. Inventaire floristique de la Drôme

L'inventaire floristique de Garraud (2004) représente de loin le travail de recensement naturaliste le plus important et le plus fiable en France. Nous sommes donc partis du principe que si nous ne pouvions pas accéder à partir d'une telle somme d'information à des informations intéressantes sur l'écologie des espèces, alors la démarche d'inférence était sans espoir.

Et heureusement, tout au contraire, la démarche d'inférence s'est révélée non seulement robuste aux éventuels biais d'échantillonnage, mais surtout extrêmement informative sur

l'écologie des espèces. Rappelons que nous avons dans cette optique deux gammes de résultats :

- Une analyse comparative des prédictions du paramètre  $r$  sur différents types d'habitat potentiel permet le test d'hypothèses écologiques : en l'occurrence dans notre étude nous avons testé l'effet des usages des terres par l'Homme sur la dynamique des espèces.
- Une distribution globale des estimations de  $r$  réalisées à échelle intermédiaire donne de nombreuses informations sur l'écologie des espèces. Si l'on considère comme hypothèse de départ que  $r$  est une propriété intrinsèque à l'espèce, cette distribution globale devrait être une distribution d'échantillonnage. Le test de cette hypothèse constituera un maillon intéressant et important pour la compréhension de l'écologie des espèces, en caractérisant leur dynamique au sein d'habitats potentiels bien choisis.

Les perspectives offertes par l'étude de plus de 2700 espèces végétales dans la Drôme apparaissent dès lors très prometteuses. D'ores et déjà, l'examen des résultats au cas par cas a pu révéler de prime abord de nombreuses situations intéressantes, pouvant nourrir en particulier des études solides en biologie de la conservation.

#### **4. Conclusion**

- **Métapopulations et biologie des populations**

L'idée essentielle qui a motivé et orienté mon discours a été la nécessité de ne pas travailler à une échelle « trop locale », afin d'être en mesure de prendre en compte correctement les structures émergentes induites par les processus locaux. Le jeu d'échelles qu'implique la méthode d'inférence utilisée au Chapitre V s'inscrit dans cet esprit. L'échelle spatiale intermédiaire permet d'estimer correctement l'effet des processus locaux d'extinction et de colonisation, l'échelle globale rend compte de la variation spatiale des paramètres estimés.

Dans une autre logique, l'étude des stratégies de persistance et de propagation des organismes se fait habituellement en biologie des populations à un échelon local et populationnel. On mesure et on étudie des traits individuels, et on évalue le cas échéant quels peuvent être les impacts de ces traits sur la structure spatiale des individus et/ou des populations. Mon travail se situe en somme à un échelon intermédiaire entre le champ de la biologie des populations et le champ de la biogéographie ; c'est une sorte de « biologie des métapopulations ».

Analyser et prédire le lien entre les caractéristiques individuelles, traits fonctionnels et stratégies d'histoire de vie, et la dynamique à grande échelle des espèces sera en perspective une manière passionnante d'aller au cœur de la compréhension du devenir des espèces à grande échelle. Paradoxalement peut-être, l'approche métapopulation, profondément marquée par les événements contingents, donnera la coloration écologique, tandis que l'approche locale individuelle fournira une entrée sur les processus évolutifs.

- **Métapopulations et écologie évolutive**

La question restant ouverte sera donc : dans quelle mesure les processus évolutifs régis par la sélection, la migration et la dérive au niveau individuel et populationnel peuvent-ils être influencés par les propriétés émergentes des métapopulations, qui conditionnent la survie des espèces à grande échelle (voir Gouyon 1999) ? Nous avons mis l'accent sur le problème de la détermination des caractéristiques dynamiques actuelles des espèces. Nous avons donc choisi de représenter les métapopulations dans un temps écologique. Nous avons en particulier discuté la parabole du démon de Maxwell (§1.4), et avons suggéré que les structures auto-organisées à grandes échelles de temps et d'espace pouvaient jouer un rôle dans l'évolution des espèces.

Le concept de métapopulation est fondamentalement ancré dans le champ de la génétique des populations ; nous ne développerons pas ici l'abondante bibliographie. Il a été mis en avant pour l'étude de l'évolution des stratégies de dispersion et de persistance locale (Comins et al. 1980; Olivieri 1990; Olivieri et al. 1995; Ronce et al. 2000; Ronce 2001; Kisdi 2002). Sa représentation met initialement l'accent sur les processus affectant la survie et la reproduction des individus. Le point de départ est donc une subdivision discrète des populations, jusqu'au niveau individuel, pour étudier l'opposition entre l'effet du voisinage et l'effet de la dispersion globale (Hamilton and May 1977; Comins et al. 1980). Dès lors la discussion de la prise en compte de la démographie, de la distance de dispersion, de la taille des populations, des effets successionnels et de la spatialisation de plus en plus complexifiée des populations jalonne l'évolution de la discipline.

Le changement d'échelle de définition des entités discrètes, de l'individu à la population, induit alors différents niveaux d'appréhension du problème, avec une prégnance plus ou moins forte de la démographie. Notre approche écologique du problème se situe de fait à une grande échelle d'espace et de temps, de manière à négliger résolument l'incidence de la démographie. Ce choix n'apparaît donc pas neutre dans le cadre d'une réflexion évolutive. En outre nous avons indiqué au Chapitre II (§1.2) que la prise en compte du couplage

démographique des populations était de nature à engendrer des structures dynamiques particulières, que nous n'avons pas prises en compte.

Nous avons insisté sur le fait que l'expression du lien dynamique entre populations détermine la persistance, mais aussi la structure spatiale et temporelle de la métapopulation. Il vient alors une question délicate : si l'on considère que les processus démographiques ne déterminent pas le devenir des métapopulations à grande échelle d'espace et de temps, comment peut-on envisager l'évolution des stratégies de dispersion et de persistance locale à cette échelle ? L'idée que les forces évolutives peuvent différer suivant les échelles de représentation devient cruciale (voir par ex. Olivieri 1990). Le rôle des structures auto-organisées est une singulière énigme, mais on peut difficilement l'occulter, tant elle détermine le devenir et la structure des populations.

En définitive, les caractéristiques dynamiques que nous déterminons pour chaque espèce n'auront de sens qu'à la lumière d'une compréhension des mécanismes évolutifs en jeu à cette échelle de représentation. C'est là un défi difficile mais passionnant à relever.

- **Métapopulations et écologie des communautés**

Le modèle de métapopulation sur lequel je me suis appuyé ne prend pas en compte l'effet dynamique des interactions biotiques sur le devenir individuel des espèces. De fait, toutes les variables d'habitat, y compris potentiellement la présence d'autres espèces, sont considérées comme des éléments statiques. Cependant nous avons mis en exergue dans le Chapitre V que des regroupements d'espèces en terme de dynamiques métapopulationnelles pouvaient représenter des syndromes écologiques (rudéralité) ou biologiques (effet orchidée). Des effets de structuration de communautés, dépendant des conditions écologiques ou de l'historique des activités humaines, peuvent apparaître dans notre analyse.

La lecture des effets de structuration des communautés peut ainsi se faire en considérant des syndromes de valeurs estimées pour *rCLC*, *rBioclim* et *rAll*. Nous avons évoqué dans l'Epilogue du Chapitre V (p. 180), la possibilité d'une « écologie fonctionnelle des métapopulations ». Ne serait-ce pas en effet un moyen de comprendre les déterminants de structuration des communautés à grande échelle spatiale, en considérant conjointement l'effet de l'habitat et de la dynamique des espèces ?

Ainsi pourrait-on concilier une approche purement « occurrentielle » de structuration des communautés, de type phytosociologie, et une approche purement écologique, de type analyse

fonctionnelle. Le lien entre les deux approches résiderait précisément dans la représentation de la dynamique métapopulationnelle des espèces.

- **Métapopulations et métacommunautés**

Mon travail de thèse s'arrête à l'étude de la dynamique spatiale des populations, mais une étude post-doctorale (sous la direction de Pierre Couteron) a été bâtie dans la continuité de cette approche, avec la volonté d'étudier la dynamique des communautés. Ce travail s'articule autour d'un schéma simple qui poursuit le propos de la thèse :

- Il existe actuellement un fort engouement autour des modèles neutres de métacommunautés (= communautés en réseau) (Hubbell 2001), qui expriment les notions d'équilibres mutation-dérive et migration-dérive dans le champ de l'écologie : les types d'allèles sont des espèces, le taux de mutation est un taux de spéciation. La métacommunauté est un pool d'individus fermé à grande échelle de temps et d'espace, où l'équilibre spéciation-dérive est exprimé. Les communautés au sein de la métacommunautés connaissent une dynamique interne marquée par une dérive locale et par des événements de migration depuis l'extérieur : on y exprime un équilibre migration dérive. Le modèle est neutre car toutes les espèces sont identiques, il n'y a pas de niches. Nous considérons ce modèle comme un modèle nul de départ, dont nous testons dans un premier temps les attendus. Le test sera effectué dans le contexte des forêts tropicales humides des Western Ghâts, dans le sud de l'Inde. Nous considérerons en particulier la sensibilité du modèle aux modifications des modalités de variations des paramètres, grâce à une architecture hiérarchique de modèles. Nous évaluerons la richesse de l'information disponible et la qualité d'ajustement des modèles métacommunautés neutres. On s'inscrit à ce stade dans une démarche à la fois de test d'hypothèse (falsifier le modèle nul neutre), et d'analyse informationnelle (recherche de la quantité et de la qualité de l'information décrite) (voir Introduction, §3).
- L'engouement pour cette approche s'est davantage cristallisé sur des questions techniques que sur la signification des concepts de communauté et de métacommunauté (Etienne and Olff 2004; Etienne 2005; He and Hu 2005; Etienne et al. 2006). D'une part, les communautés ne possèdent pas de délimitation spatiale précise, contrairement aux populations. Force est de constater que les applications actuelles du modèle se préoccupent fort peu de la question. D'autre part, la dérive exprimée au niveau de la



métacommunauté et des communautés individuelles porte sur les nombres d'individus (Hubbell 2001), et donc cette approche n'est pas cohérente avec la démarche métapopulationnelle, où les effets démographiques sont découplés de la dynamique des populations. Le fait de considérer des événements de dérive individuelle sur des métacommunautés avec des milliards d'individus, sur des temps très longs et avec des événements de spéciation, donne l'impression que beaucoup d'échelons intermédiaires de représentation des processus ont été oubliés. Ainsi nous confronterons l'approche métacommunauté individu-centrée à une approche population-centrée, afin de rester dans la philosophie des métapopulations et de considérer un niveau supérieur d'analyse.

- Nous envisageons donc une autre approche des métacommunautés, en tant que métapopulations couplées. La contrainte au niveau de la communauté n'est alors pas exprimée en terme de nombre d'individus mais en terme de nombre d'espèces, suivant des règles supputées d'assemblage (Diamond 1975). En outre avec notre approche (Chapitre IV & V) les taux de migration peuvent varier suivant les espèces, et les habitats potentiels peuvent être exprimés espèce par espèce. Nous resterons fidèles à une analyse spatialement réaliste des métapopulations, et mettrons donc en œuvre la méthode utilisée dans la Drôme à l'étude des espèces forestières des Western Ghâts dans le sud de l'Inde. Nous nous placerons alors dans une démarche essentiellement inférentielle. Dans un premier temps les métapopulations ne seront pas couplées, dans un second temps nous étudierons les conditions d'une contrainte locale sur le nombre d'espèces. La prise en compte de la niche se fera en modulation du terme de colonisation, comme nous l'avons montré dans le Chapitre IV. Pour le reste, la saturation et le remplacement des espèces dans les communautés seront stochastiques. Ainsi serons-nous à même d'introduire les processus neutres de migration et de dérive en même temps que la prise en compte de la notion de niche et des règles d'assemblage, le tout dans le respect du découplage des processus démographiques et métapopulationnels. Certes, nombres d'enjeux techniques restent à affronter pour cette étape d'inférence, mais les bases conceptuelles sont d'ores et déjà posées.

## Bibliographie

- Araújo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation* 96(3): 331-345.
- Arenas C, Cuadras CM (2002) Recent statistical methods based on distances. *Contributions to Science* 2(2): 183-191.
- ATILF (2004) Le Trésor de la Langue Française Informatisé (TLFi).
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101-118.
- Baguette M (2004) The classical metapopulation theory and the real, natural world: a critical appraisal. *Basic and Applied Ecology* 5: 213-224.
- Bak P (1997) *How nature works - The science of self-organized criticality*: Oxford University Press. 212 p.
- Bak P, Sneppen K (1993) Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters* 71: 4083-4086.
- Barabasi A-L, Albert R (1999) Emergence of Scaling in Random Networks. *Science* 286: 509-512.
- Barbault R (2000) *Écologie générale - Structure et fonctionnement de la biosphère* Dunod. 336 p.
- Bascompte J (2001) Aggregate Statistical Measures and Metapopulation Dynamics. *Journal of Theoretical Biology* 209: 373-379.
- Bascompte J (2003) Extinction threshold: insights from simple models. *Ann Zool Fennici* 40: 99-114.
- Bellwood DR, Wainwright PC, Fulton CJ, Hoey A (2002) Assembly rules and functional groups at global biogeographic scales. *Functional Ecology* 16: 557-562.
- Berryman AA (2002) Population : a central concept for ecology ? *Oikos* 97(3): 439-442.
- Blackburn TM, Gaston KJ (2002) Macroecology is distinct from biogeography. *Nature* 418: 723.
- Bolliger J, Sprott JC, Mladenoff DJ (2003) Self-organization and complexity in historical landscape patterns. *Oikos* 100: 514-553.

- Borcard D, Legendre P (1994) Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environmental and Ecological Statistics* 1: 37-61.
- Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153: 51-68.
- Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology* 73(3): 1045-1055.
- Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2003) Dissecting the spatial structure of ecological data at all scales. 1-35.
- Brooks CP (2003) A scalar analysis of landscape connectivity. *Oikos* 102(2): 433-439.
- Brunk GG (2000) Understanding self-organized criticality as a statistical process. *Complexity* 5(3): 26-33.
- Cain DH, Riitters K, Orvis K (1997) A multi-scale analysis of landscape statistics. *Landscape Ecology* 12: 199-212.
- Camazine S, Deneubourg J-L, Franks NR, Sneyd J, Theraulaz G et al. (2001) *Self-Organization in Biological Systems*. Princeton: Princeton University Press.
- Cantrell RS, Cosner C (2003) *Spatial Ecology via Reaction-Diffusion Equations*: Wiley. 428 p.
- Casagrandi R, Gatto M (1999) A mesoscale approach to extinction risk in fragmented habitats. *Nature* 400: 560-562.
- Chung FRK (1994) *Spectral Graph Theory*; Sciences CBotM, editor. Providence, Rhode Island: American Mathematical Society. 212 p.
- Ciliberti S, Caldarelli G, De Los Rios P, Pietronero L, Zhang Y-C (2000) Discretized Diffusion Processes. *Physical Review Letters* 85(23): 4848-4851.
- Clark JS (1998) Why Trees Migrate So Fast: Confronting Theory with Dispersal Biology and the Paleorecord. *The American Naturalist* 152(2).
- Comins HN, Hamilton WD, May RM (1980) Evolutionary stable dispersal strategies. *Journal of Theoretical Biology* 82: 205-230.
- Coveney PV (2003) Self-organization and complexity: a new age for theory, computation and experiment. *Phil Trans R Soc Lond A* 361: 1057-1079.
- Csillag F, Kabos S (2002) Wavelets, boundaries and the analysis of landscape pattern. *Ecoscience* 9(2): 177-190.

- Dale MRT, Dixon P, Fortin M-J, Legendre P, Myers DE et al. (2002) Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* 25: 558-577.
- De Candolle A (1855) *Géographie Botanique raisonnée ou exposition des faits principaux et des lois concernant la distribution géographique des plantes de l'époque actuelle*. Paris: Librairie de Victor Masson. 642 p.
- de Rosnay J (1975) *Le Macroscopie. Vers une vision globale*: Editions du Seuil.
- Desdevises Y, Legendre P, Azouzi L, Morand S (2003) Quantifying phylogenetically structured environmental variation. *Evolution* 57(11): 2647-2652.
- Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM, editors. *Ecology and evolution of communities*. Cambridge: Harvard University Press.
- Dias PC (1996) Sources and sinks in population biology. *Trends in Ecology and Evolution* 11(8): 326-330.
- Dickman R, Vidigal R (2002) Quasi-stationary distributions for stochastic processes with an absorbing state. *Journal of Physics A: Mathematical and General* 35: 1147–1166.
- Dickman R, Martins de Oliveira M (2004) Quasi-stationary simulation of the contact process. pp. 13.
- Dickman R, Munoz MA, Vespignani A, Zapperi S (2000) Paths to Self-Organized Criticality. *Brazilian Journal of Physics* 30(1): 27-41.
- Dray S, Legendre P, Peres-Neto PR (2004) Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM).
- Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S et al. (2002) A balanced view of scale in spatial statistical analysis. *Ecography* 25: 626-640.
- Ehrlén J, Eriksson O (2003) Large-scale spatial dynamics of plants: a response to Freckleton & Watkinson. *Journal of Ecology* 91(2): 316-320.
- Eldredge N, Gould SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, editor. *Models in Paleobiology*. New York: Freeman.
- Ellison A (1996) An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications* 6(4): 1036-1046.
- Ellner SP (2001) Pair Approximation for Lattice Models with Multiple Interaction Scales. *Journal of Theoretical Biology* 210: 435-447.
- Ellner SP, Sasaki A, Haraguchi Y, Matsuda H (1998) Speed of invasion in lattice population models: pair-edge approximation. *Journal of Mathematical Biology* 36: 469-484.

- Etienne RS (2000) Local Populations of Different Sizes, Mechanistic Rescue Effect and Patch Preference in the Levins Metapopulation Model. *Bulletin of Mathematical Biology* 62: 943-958.
- Etienne RS (2005) A new sampling formula for neutral biodiversity. *Ecology Letters* 8: 253-260.
- Etienne RS, Nagelkerke CJ (2002) Non-equilibria in Small Metapopulations: Comparing the Deterministic Levins Model with its Stochastic Counterpart. *Journal of Theoretical Biology* 219(4): 463-478.
- Etienne RS, Olff H (2004) A novel genealogical approach to neutral biodiversity theory. *Ecology Letters* 7: 170-175.
- Etienne RS, Latimer AM, Silander Jr. JA, Cowling RM (2006) Comment on ‘‘Neutral Ecological Theory Reveals Isolation and Rapid Speciation in a Biodiversity Hot Spot’’. *Science* 311: 610b.
- Ewald J (2003) A critique for phytosociology. *Journal of Vegetation Science* 14: 291-296.
- Felizola Diniz-Filho JA, Bini LM (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14: 177-185.
- Fisher HJ (2002) Macroecology: new, or biogeography revisited? *Nature* 417: 787.
- Fortin M-J, Boots B, Csillag F, Remmel TK (2003) On the role of spatial stochastic models in understanding landscape indices in ecology. *Oikos* 102(1): 203-212.
- Franc A (2004) Metapopulation dynamics as a contact process on a graph. *Ecological Complexity* 1: 49-63.
- Freckleton RP, Watkinson AR (2002) Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Journal of Ecology* 90: 419-434.
- Freckleton RP, Watkinson AR (2003) Are all plant populations metapopulations? *Journal of Ecology* 91(2): 321-324.
- Frigg R (2003) Self-organised criticality—what it is and what it isn’t. *Studies in History and Philosophy of Science* 34: 613-632.
- Garraud L (2004) Flore de la Drôme - Atlas écologique et floristique. Gap: Conservatoire Botanique National Alpin de Gap-Charance. 939 p.
- George EI (2000) The Variable Selection Problem. pp. 12.
- Gil L, Sornette D (1996) Landau-Ginzburg Theory of Self-Organized Criticality. *Physical Review Letters* 76(21): 3991-3994.

- Goldenfeld N, Kadanoff LP (1999) Simple Lessons from Complexity. *Science* 284: 87-89.
- Gouyon PH (1999) Sex: a pluralist approach includes species selection. (One step beyond and it's good.). *Journal of Evolutionary Biology* 12: 1029-1030.
- Griffiths PE, Gray RD (2005) Discussion: Three ways to misunderstand developmental systems theory. *Biology and Philosophy* 20: 417-425.
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135(2-3): 147-186.
- Halley JM, Hartley S, Kallimanis AS, Kunin WE, Lennon JJ et al. (2004) Uses and abuses of fractal methodology in ecology. *Ecology Letters* 7: 254-271.
- Hamilton WD, May R (1977) Dispersal in stable habitats. *Nature* 269.
- Hanski I (1998) Metapopulation dynamics. *Nature* 396: 41-49.
- Hanski I (2004) Metapopulation theory, its use and misuse. *Basic and Applied Ecology* 5: 225-229.
- Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. *Nature* 404: 755-758.
- Hanski I, Ovaskainen O (2002) Extinction Debt at Extinction Threshold. *Conservation Biology* 16(3): 666-673.
- Hanski IA, Gilpin ME, editors (1997) *Metapopulation Biology*. San Diego, California: Academic Press. 525 p.
- Harding KC, McNamara JM (2002) A Unifying Framework for Metapopulation Dynamics. *The American Naturalist* 160(2): 173-185.
- He F, Hu X-S (2005) Hubbell's fundamental biodiversity parameter and the Simpson diversity index. *Ecology Letters* 8: 386-390.
- Heikkinen RK, Luoto M, Kuussaari M, Poyry J (2005) New insights into butterfly-environment relationships using partitioning methods. *Proceedings of the Royal Society B* 272: 2203-2210.
- Hiebeler D (2000) Populations on fragmented landscapes with spatially structured heterogeneities: landscape generation and local dispersal. *Ecology* 81(6): 1629-1641.
- Hill MF, Caswell H (2001) The effects of habitat destruction in finite landscapes: a chain-binomial metapopulation model. *Oikos* 93: 321-331.
- HilleRisLambers R, Rietkerk M, Van den Bosch F, Prins HHT, de Kroon H (2001) Vegetation pattern formation in semi-arid grazing systems. *Ecology* 82(1): 50-61.
- Hinrichsen H (2000) Non-equilibrium critical phenomena and phase transitions into absorbing states. *Advances In Physics* 49(7): 815-958.

- Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton and Oxford: Princeton University Press. 375 p.
- Hui C, Li Z (2003) Dynamical complexity and metapopulation persistence. *Ecological Modelling* 164: 201-209.
- Hui C, Li Z (2004) Distribution patterns of metapopulation determined by Allee effects. *Population Ecology* 46: 55-63.
- Hutchinson GE. Concluding remarks- Cold Spring Harbor Symposia on Quantitative Biology; 1957. *Bull. of Math. Biol.* pp. 193-213.
- Ives AR, Turner MG, Pearson SM (1998) Local Explanations of Landscape Patterns: Can Analytical Approaches Approximate Simulation Models of Spatial Processes? *Ecosystems* 1: 35-51.
- Iwasa Y (2000) *Lattice Models and Pair Approximation in Ecology. The Geometry of Ecological Interaction - Simplifying Spatial Complexity*. Cambridge: Cambridge University Press. pp. 227-251.
- Johnson CR, Boerlijst C (2002) Selection at the level of the community: the importance of spatial structure. *Trends in Ecology and Evolution* 17(2): 83-90.
- Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution* 19(2): 101-108.
- Kaitala V, Ranta E (1998) Travelling wave dynamics and self-organization in a spatio-temporally structured population. *Ecology Letters* 1: 186-192.
- Kaitala V, Ranta E, Lundberg P (2001) Self-organized dynamics in spatially structured populations. *Proceedings Of The Royal Society Of London Series B-Biological Sciences* 268(1477): 1655-1660.
- Kareiva P, Wennergren U (1995) Connecting landscape patterns to ecosystem and population processes. *Nature* 373: 299-302.
- Kauffman SA (1993) *The Origins of Order - Self-Organization and Selection in Evolution*. New York - Oxford: Oxford University Press.
- Keitt TH (2000) Spectral representation of neutral landscapes. *Landscape Ecology* 15(5): 479-494.
- Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Manuscript. pp. 1-23.
- Keitt TH, Urban DL, Milne BT (1997) Detecting Critical Scales in Fragmented Landscapes. *Conservation Ecology* 1(1).
- Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography* 25: 616-625.

- Kisdi É (2002) Dispersal: Risk Spreading versus Local Adaptation. *The American Naturalist* 159(6): 579-596.
- Koenig WD (2002) Global patterns of environmental synchrony and the Moran effect. *Ecography* 25(3): 283-288.
- Labra FA, Lagos NA, Marquet PA (2003) Dispersal and transient dynamics in metapopulations. *Ecology Letters* 6(3): 197-204.
- Lande R (1987) Extinction thresholds in demographic models of territorial populations. *The American Naturalist* 130: 624-635.
- Lasserre JB, Pearce CEM (2001) On the existence of a quasistationary measure for a Markov chain. *The Annals of Probability* 29(1): 437-446.
- Lavorel S, Díaz S, Cornelissen JHC, Garnier E, Harrison SP et al. (2005) Plant functional types: are we getting any closer to the Holy Grail ? In: Canadell J, Pitelka LF, Pataki D, editors. *Terrestrial Ecosystems in a Changing World*: Springer-Verlag.
- Law R, Murrel DJ, Dieckmann U (2003) Population growth in space and time: spatial logistic equations. *Ecology* 84(1): 252-262.
- Lecoutre B (2005) Et si vous étiez un bayésien qui s'ignore? *Modulad* 32: 92-105.
- Legendre P, Legendre L (1998) *Numerical Ecology*. Amsterdam: Elsevier. 853 p.
- Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69(1): 1-24.
- Legendre P, Borcard D (2003) Quelles sont les échelles spatiales importantes dans un écosystème? In: Driesbeke J-J, Lejeune M, Saporta G, editors. *Analyse statistique de données spatiales*. TECHNIP ed. Paris.
- Legendre P, Borcard D, Peres-Neto PR (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs* Accepted.
- Leoncini I, Rivault C (2005) Could Species Segregation be a Consequence of Aggregation Processes? Example of *Periplaneta americana* (L.) and *P. fuliginosa* (Serville). *Ethology* 111: 527-540.
- Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73(6): 1943-1967.
- Levin SA (2002) Complex adaptive systems: exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society* 40(1): 3-19.
- Levin SA, Grenfell B, Hastings A, Perelson AS (1997) Mathematical and computational challenges in population biology and ecosystems science. *Science* 275(5298): 334-343.



- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15: 237-240.
- Li B-L (2000a) Fractal geometry applications in description and analysis of patch patterns and patch dynamics. *Ecological Modelling* 132: 33-50.
- Li B-L (2000b) Why is the holistic approach becoming so important in landscape ecology? *Landscape and Urban Planning* 50: 27-41.
- Li B-L (2002) Criticality, self-organized. In: El-Shaarawi AH, Piegorsch WW, editors. *Encyclopedia of Environmetrics*. Chichester: John Wiley & Sons, Ltd.
- Lichstein JW, Simons TR, Shiner SA, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72(3): 445-463.
- Liebold AM, Gurevitch J (2002) Integrating the statistical analysis of spatial data in ecology. *Ecography* 25(5): 553-557.
- Loreau M, Mouquet N, Holt RD (2003) Meta-ecosystems: a theoretical framework for a spatial ecosystem ecology. *Ecology Letters* 6(8): 673-679.
- Luisi PL (2003) Contingency and determinism. *Phil Trans R Soc Lond A* 361: 1141-1147.
- MacArthur RH (1955) Fluctuations of animal populations and a measure of community stability. *Ecology* 36: 533-536.
- Marquet PA (2002) The search for general principles in ecology. *Nature* 418: 723.
- McGill B (2003) Strong and weak tests of macroecological theory. *Oikos* 102(3): 679-685.
- Méot A, Chessel D, Sabatier R (1998) Opérateurs de voisinage et analyse des données spatio-temporelles. In: Asselain BC, editor. *Biométrie et Données spatio-temporelles*. Société Française de Biométrie ed. Rennes. pp. 38-64.
- Molofsky J, Bever JD (2004) A New Kind of Ecology? *BioScience* 54(5): 440-446.
- Murrell DJ, Purves DW, Law R (2001) Uniting pattern and process in plant ecology. *Trends in Ecology and Evolution* 16(10): 529-530.
- Nee S (2002) Thinking big in ecology. *Nature* 417: 229-230.
- O'Neill RV (2001) Is it time to bury the ecosystem concept? (with full military honors, of course!). *Ecology* 82(12): 3275-3284.
- O'Neill RV, Krummel JR, Gardner RH, Sugihara G, Jackson B et al. (1988) Indices of landscape pattern. *Landscape Ecology* 1(3): 153-162.
- Oborny B, Meszén G, Szabo G (2005) Dynamics of populations on the verge of extinction. *Oikos* 109: 291-296.

- Olivieri I (1990) The Genetics of Transient Populations: research at the metapopulation level. *Trends in Ecology and Evolution* 5: 207-210.
- Olivieri I, Michalakis Y, Gouyon PH (1995) Metapopulation genetics and the evolution of dispersal. *The American Naturalist* 146(2): 202-228.
- Ovaskainen O, Hanski I (2001) Spatially Structured Metapopulation Models: Global and Local Assessment of Metapopulation Capacity. *Theor Popul Biol* 60(4): 281-302.
- Ovaskainen O, Hanski I (2002) Transient Dynamics in Metapopulation Response to Perturbation. *Theor Popul Biol* 61(3): 285-295.
- Ovaskainen O, Sato K, Bascompte J, Hanski I (2002) Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology* 215: 95-108.
- Pascual M, Guichard F (2005) Criticality and disturbance in spatial ecological systems. *Trends In Ecology & Evolution* 20(2): 88-95.
- Pascual M, Roy M, Guichard F, Flierl G (2002) Cluster size distributions: signatures of self-organization in spatial ecologies. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences* 357(1421): 657-666.
- Peck SL (2004) Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology & Evolution* 19(10): 530-534.
- Pélissier R, Couteron P, Dray S, Sabatier D (2003) Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology* 84(1): 242-251.
- Platt T, Denman KL (1975) Spectral analysis in ecology. *Annual Review of Ecology and Systematics* 6: 189-210.
- Popper K (1959) *The Logic of Scientific Discovery* - Reprint edition (1992) Routledge publishers. 480 p.
- Prigogine I (1977) Time, structure and fluctuations. Nobel Lecture, 8 December, 1977.
- Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. *Trends in Ecology and Evolution* 20(6): 345-353.
- Pulliam HR (2000) On the relationship between niche and distribution. *Ecology Letters* 3(4): 349-361.
- Quader S, Isvaran K, Hale R, Miner B, Seavy N (2004) Nonlinear relationships and phylogenetically independent contrasts. *Journal of Evolutionary Biology* 17(3): 709-715.

- Ranta E, Lundberg P, Kaitala V, Stenseth NC (2002) On the Crest of a Population Wave. *Science* 298: 973-974.
- Reuter H, Franz H, Middelhoff U, Jopp F, Eschenbach C et al. (2005) The concepts of emergent and collective properties in individual-based models—Summary and outlook of the Bornhoved case studies. *Ecological Modelling* 186: 489-501.
- Ricotta C (2000) From theoretical ecology to statistical physics and back: self-similar landscape metrics as a synthesis of ecological diversity and geometrical complexity. *Ecological Modelling* 125(2-3): 245-253.
- Rietkerk M, Dekker SC, de Ruiter PC, van de Koppel J (2004) Self-organized patchiness and catastrophic shifts in ecosystems. *Science* 305(5692): 1926-1929.
- Ronce O (2001) Understanding plant dispersal and migration. *Trends in Ecology & Evolution* 16(12): 663-664.
- Ronce O, Perret F, Olivieri I (2000) Evolutionarily Stable Dispersal Rates Do Not Always Increase with Local Extinction Rates. *The American Naturalist* 155(4): 485-496.
- Rozdilsky ID, Chave J, Levin SA, Tilman D (2001) Towards a theoretical basis for ecosystem conservation. *Ecological Research* 16: 983-995.
- Scheffer M, Carpenter SR (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends In Ecology & Evolution* 18(12): 648-656.
- Schrodinger E (1944) What is life? The Physical Aspect of the Living Cell.
- Skar J (2003) Introduction: self-organization as an actual theme. *Phil Trans R Soc Lond A* 361: 1049-1056.
- Snyder RE, Nisbet RM (2000) Spatial structure and fluctuations in the contact process and related models. *Bulletin of Mathematical Biology* 62: 959-975.
- Solé RV, Manrubia SC, Benton M, Kauffman S, Bak P (1999) Criticality and scaling in evolutionary ecology. *Trends in Ecology and Evolution* 14(4): 156-160.
- Sornette D (1992) Critical phase transitions made self-organized : a dynamical system feedback mechanism for self-organized criticality. *J Phys I France* 2(11): 2065-2073.
- Stanley HE (1971) Introduction to phase transitions and critical phenomena: Clarendon Press.
- Stewart I (2003) Self-organization in evolution: a mathematical perspective. *Phil Trans R Soc Lond A* 361: 1101-1123.
- Svenning J-C, Skov F (2002) Mesoscale distribution of understorey plants in temperate forest (Kalø, Denmark): the importance of environment and dispersal. *Plant Ecology* 160: 169-185.

- Szabo G, Gergely H, Oborny B (2002) Generalized contact process on random environments. pp. 6.
- Taylor PJ, Haila Y (2001) Philosophy of Ecology. Encyclopedia of life sciences: MacMillan Publishers Ltd.
- Thomson NA, Ellner SP (2003) Pair-edge approximation for heterogeneous lattice population models. *Theor Popul Biol* 64(3): 271-280.
- Thuiller W (2003) BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9(10): 1353-1362.
- Thuiller W, Araújo MB, Lavorel S (2003) Generalised Models versus Classification Tree Analysis: a comparative study for predicting spatial distributions of plant species at different scales. En préparation.
- Torrence C, Compo GP (1998) A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society* 79(1): 61-78.
- Ulanowicz RE (2002) The balance between adaptability and adaptation. *BioSystems* 64(1-3): 13-22.
- Via S (2002) The Ecological Genetics of Speciation. *The American Naturalist* 159(Supplement): S1-S7.
- Vidyasagar M (2002) Nonlinear Systems Analysis. xviii + 498 p.
- Wakano JY, Maenosono S, Komoto A, Eiha N, Yamaguchi Y (2003) Self-Organized Pattern Formation of a Bacteria Colony Modeled by a Reaction Diffusion System and Nucleation Theory. *Physical Review Letters* 90(25).
- Webster G, Goodwin BC (1982) The origin of species: a structuralist approach. *Journal of Social and Biological Structures* 5(1): 15-47.
- Weitz JS (1997) Generalized Contact Processes in Ecology [Philosophy]. Princeton: A.B, Princeton University. 168 p.
- Wolfram Research I (2003) Mathematica. 5.0 ed. Champaign, Illinois: Wolfram Research, Inc.
- Wootton JT (2001) Local interactions predict large-scale pattern in empirically derived cellular automata. *Nature* 413: 841-843.
- Wu J, David JL (2002) A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling* 153(1-2): 7-26.

## Annexe 1 : Plate-forme de simulation *SimulMetapop*

La simulation de la dynamique métapopulationnelle a été réalisée de la manière la plus explicite possible et s'apparente donc à un automate cellulaire (*Cellular Automaton*). La programmation de ce processus fait donc appel à un nombre important de boucles de calculs, afin de reproduire individuellement les événements de colonisation et d'extinction dans l'espace et au cours du temps.

Les logiciels de calculs utilisant un interpréteur de commande (*Mathematica*, *Matlab*, *SPlus*) ne sont pas recommandables pour le traitement de telles boucles, car la couche logicielle d'interprétation des commandes ralentit considérablement les calculs quand un grand nombre d'instructions sont itérées.

Pour remédier à ce problème, il est possible de supprimer la couche d'interprétation de commandes en réalisant un programme compilé autonome. C'est ce que j'ai effectué, en utilisant le langage C++ et les meilleures bibliothèques de fonctions de probabilités disponibles sur Internet.

L'interface résultante est une fenêtre Windows de paramétrage et de lancement des simulations (voir

Figure 33). L'interface graphique est réalisée grâce aux bibliothèques disponibles sous Visual C++. Le gain de vitesse par rapport à *Mathematica* atteint un facteur 20. Il est possible de contourner le ralentissement induit par les processus Windows et de bénéficier de la pleine puissance de l'ordinateur grâce à la fonction « turbo ».

Une telle interface ouvre de nombreuses possibilités de tester rapidement la sensibilité des simulations à de nombreux paramètres : répliquats de cartes, variation des conditions initiales, taille de la grille, nombre de générations, bords toriques ou non... En outre, le programme est entièrement autonome, car il inclut sous sa forme compilée toutes les fonctions nécessaires aux simulations, y compris statistiques, génération de cartes aléatoires, synthèse sur les distributions quasi-stationnaires... La seule contrainte actuelle est l'environnement Windows, mais une transposition sous Unix est possible.

L'utilisation de *SimulMetapop* à des fins pédagogiques, dans le cadre par exemple d'un module d'initiation au concept de métapopulation serait donc aisée.

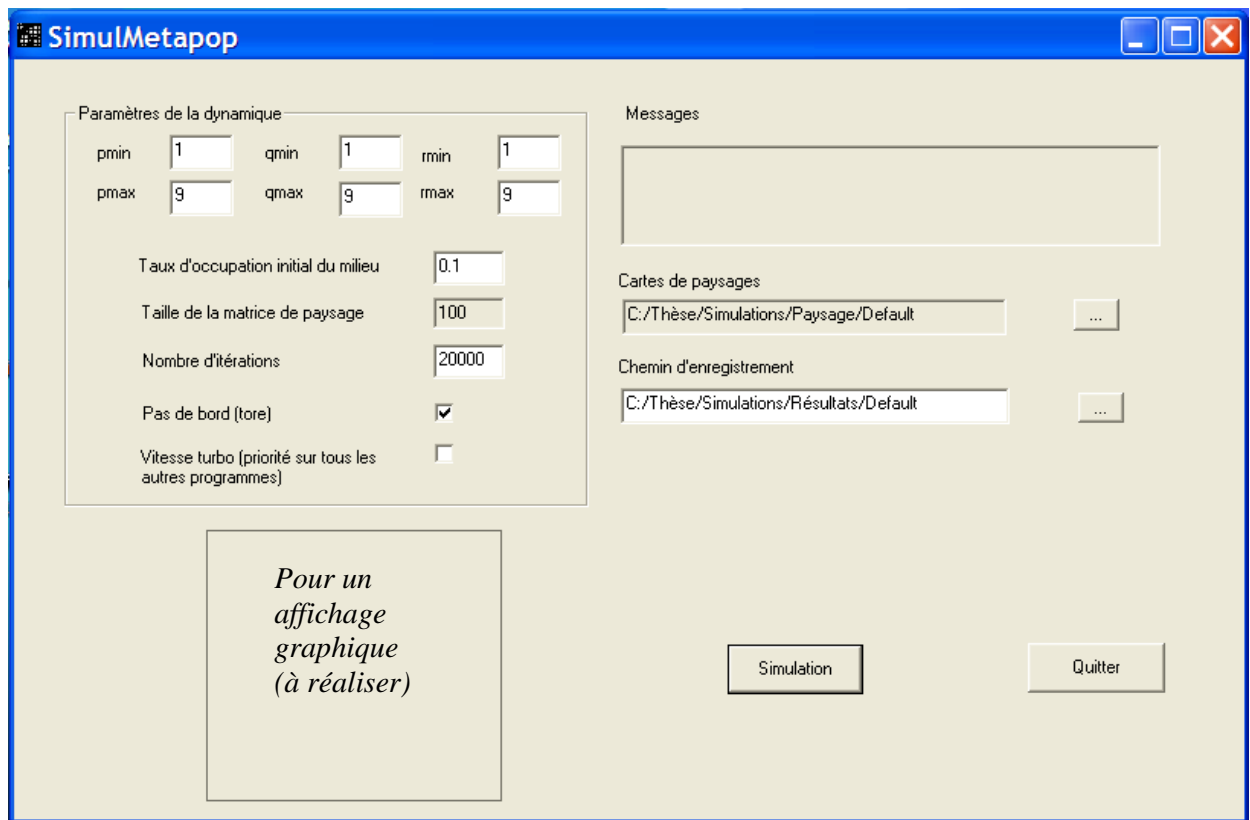


Figure 33 : Interface de simulation *SimulMetapop*. Le programme conçu sous *Visual C++* démarre avec l'ouverture d'une fenêtre Windows où le paramétrage et le lancement des simulations peuvent être réalisés simplement par l'utilisateur.

## **Annexe 2 : Sélection des variables environnementales pour le département de la Drôme**

Pour le calcul des habitats potentiels des espèces, une régression multiple est appliquée aux présence-absence agrégées des espèces, en fonction de l'ensemble des variables bioclimatiques et d'occupation des sols (voir Chapitre V). Pour éviter une inflation du nombre de variables, il convient de limiter intelligemment le nombre de catégories retenues. C'est l'objet des paragraphes suivants.

### **1. Catégories géologiques**

Les données géologiques résultent de la numérisation des cartes du BRGM, effectuée par le CBNA. Elles sont décrites par une typologie à 17 classes. On note dans la Drôme la forte dominance de roches calcaires du Secondaire, notamment datées du Crétacé. 71% des points d'échantillonnage correspondent à ce type de géologie. Les alluvions quaternaires couvrent la plaine du Rhône, mais cette zone est relativement peu échantillonnée, avec 12.5% des points. Le granite est marginal (îlot granitique de Tain l'Hermitage, 17 points sur 4349 !), et abrite une flore silicicole particulière et restreinte. Un souci est posé par l'enclave du Vaucluse, où nous n'avons pas de cartographie géologique, ce qui conduit à avoir 37 points sans information géologique. Nous négligerons l'effet de cette lacune, marginale sur l'ensemble des points (0.85% des points).

Quatre catégories écologiquement pertinentes ont été retenues : alluvions quaternaires, granite, roches sédimentaires secondaires et roches sédimentaires tertiaires. La différenciation des deux dernières classes est discutable, mais tempère la prépondérance trop forte d'une classe unique « roches sédimentaires » (86% des points). La principale contrainte sera évidemment cette très forte représentation des roches calcaires du Secondaire et Tertiaire : cela pèsera sur la délimitation des niches dans l'absolu.

Le tableau ci-après récapitule l'ensemble des classes géologiques et leur représentation dans l'ensemble des points de relevés, ainsi que dans l'échantillonnage retenu.

Libellé	Code	Nombre total	Nombre échantillon	Catégories retenues	Nouveaux effectifs
<i>Non répertorié</i>	<b>0</b>	152	37	<i>A attribuer</i>	37
Alluvions quaternaires	<b>1</b>	1199	545	Alluvions quaternaires	545
Granite	<b>2</b>	31	17	Granite	17
Pliocène marin	<b>3</b>	109	56		
Pliocène continental	<b>4</b>	105	51		
Miocène	<b>5</b>	754	300		
Miocène conglomératique	<b>6</b>	57	16	Roches sédimentaires du Tertiaire	661
Miocène Burdigalien	<b>7</b>	202	74		
Oligocène	<b>8</b>	321	117		
Eocène	<b>9</b>	139	47		
Crétacé supérieur (Turonien)	<b>10</b>	1373	496		
Crétacé inférieur (Albien)	<b>11</b>	645	218		
Crétacé inférieur (Barrémien)	<b>12</b>	1458	572		
Crétacé inférieur (Faciès urgonien)	<b>13</b>	1270	574	Roches sédimentaires du Secondaire	3089
Crétacé inférieur (BVH)	<b>14</b>	2516	961		
Jurassique supérieur	<b>15</b>	899	264		
Jurassique moyen (Dogger)	<b>16</b>	2	0		
Trias diapirique	<b>17</b>	17	4		
<b>TOTAL</b>		<b>11249</b>	<b>4349</b>		4349

## 2. Catégories d'occupation des sols

Le code *Corine Land Cover* comprend trois chiffres représentant 3 niveaux hiérarchiques : le premier niveau a cinq classes, le second 15 classes et le troisième, le plus fin, 44 classes. Le tableau ci-après résume les catégories et les effectifs associés aux points de relevés et à l'échantillonnage retenu pour la Drôme.

Suivant l'importance écologique des classes d'occupation des sols, on sélectionnera des niveaux plus ou moins fins de la nomenclature. La dernière colonne du tableau représente les 18 catégories retenues. Cela équivaut donc à 18 variables binaires complémentaires (donc 17 variables indépendantes).



		Code CLC	Ensemble relevés	Echantillon	Catégories retenues
<b>Territoires artificialisés</b>	Tissu urbain continu	111	3	0	<b>Territoires artificialisés</b>
	Tissu urbain discontinu	112	192	39	
	Zones industrielles ou commerciales	121	21	10	
	Réseaux routiers et ferroviaires	122	21	8	
	Aéroports	124	4	1	
	Extraction de matériaux	131	10	7	
	Chantiers	133	3	1	
	Equipements sportifs et de loisir	142	13	4	
	<b>SOUS-TOTAL</b>		<b>267</b>	<b>70</b>	
<b>Territoires agricoles</b>	Terres arables hors périmètre d'irrigation	211	630	207	<b>Terres arables</b>
	Périmètres irrigués en permanence	212	2	0	
	Vignobles	221	106	36	<b>Vignobles et vergers</b>
	Vergers et petits fruits	222	73	26	
	Oliveraies	223	52	20	<b>Prairies Systèmes cultureux parcellaires complexes Mosaïque agricole espaces naturels</b>
	Prairies	231	462	141	
	Systèmes cultureux et parcellaires complexes	242	972	332	
	Surfaces essentiellement agricoles, interrompues par des espaces naturels importants	243	861	335	
	<b>SOUS-TOTAL</b>		<b>3158</b>	<b>1097</b>	
<b>Forêts et milieux semi- naturels</b>	Forêts de feuillus	311	2692	1159	<b>Forêts de feuillus Forêts de conifères Forêts mélangées Pelouses et pâturages naturels Landes et broussailles Végétation sclérophylle Forêts et végétation arbustive en mutation Plages, dunes et sables Roches nues Végétation clairssemée</b>
	Forêts de conifères	312	1603	604	
	Forêts mélangées	313	1055	399	
	Pelouses et pâturages naturels	321	825	361	
	Landes et broussailles	322	626	250	
	Végétation sclérophylle	323	92	37	
	Forêts et végétation arbustive en mutation	324	422	168	
	Plages, dunes et sables	331	55	19	
	Roches nues	332	237	93	
	Végétation clairssemée	333	146	63	
	<b>SOUS-TOTAL</b>		<b>7753</b>	<b>3153</b>	
<b>Zones humides</b>	Marais intérieurs	411	4	1	<b>Zones humides</b>
<b>Surfaces en eau</b>	Cours et voies d'eau	511	44	24	<b>Surfaces en eau</b>
	Plans d'eau	512	23	4	
	<b>SOUS-TOTAL</b>		<b>67</b>	<b>28</b>	
<b>TOTAL</b>			<b>11249</b>	<b>4349</b>	

### 3. Traitement des données météorologiques

Les moyennes mensuelles de températures et de précipitations mesurées aux stations Météo-France du département de la Drôme ont été obtenues pour la période janvier 1985 – décembre 1999, via le portail Internet de la *Climathèque* (crédit spécifique CNRS Région auprès de Météo-France). Ces données doivent faire l'objet d'une interpolation spatiale, afin d'estimer des valeurs de températures et de précipitations mensuelles à chaque point d'échantillonnage.

La principale difficulté est la variation au cours du temps du nombre et de la localisation des stations ouvertes. Cela rend peu opportune une interpolation unique des données mensuelles moyennées sur la période 1985-1999, car les données par station météo peuvent être très hétérogènes.

La solution adoptée est dès lors l'interpolation des données météo mois par mois (soit 180 interpolations). Températures et précipitations ont fait l'objet d'un traitement séparé :

- Régression des données de température sur la racine carrée de l'altitude, inspirée par

l'équation hydrostatique (gaz parfaits, air immobile): 
$$k(alt - alt_0) = -\frac{R}{g}(T - T_0).(T + T_0).$$

Interpolation des résidus par krigeage, sauf pour les cas où des erreurs importantes sont constatées (en janvier 1987, décembre 1991, février 1993, août 1993, décembre 1996, janvier 1997). Pour ces derniers, une interpolation des résidus par pondération selon l'inverse de la distance est mise en œuvre.

- Régression des données de pluviométrie sur le  $\text{Log}_{10}$  de l'altitude, pour limiter les dérives de l'interpolation aux hautes altitudes. Interpolation des résidus par pondération selon l'inverse de la distance.

## **Annexe 4 : Spectral analysis of distribution maps provides insights into metapopulation dynamics**

Manuscrit soumis à *Ecological Modelling*

# **Spectral analysis of distribution maps provides insights into metapopulation dynamics**

François MUNOZ<sup>1,2</sup>, Pierre-Olivier CHEPTOU<sup>1</sup> and Finn KJELLBERG<sup>1</sup>

Affiliation: <sup>1</sup> Centre d'Ecologie Fonctionnelle et Evolutive, 1919 route de Mende, 34293 Montpellier cedex 5, France ; <sup>2</sup> French Institute of Pondicherry, 11 Saint Louis Street, Pondicherry 605 001, India.

François MUNOZ (corresponding author), [francois.m@ifpindia.org](mailto:francois.m@ifpindia.org),

Pierre-Olivier CHEPTOU, CEFE – CNRS, [pierre-olivier.cheptou@cefe.cnrs.fr](mailto:pierre-olivier.cheptou@cefe.cnrs.fr)

Finn KJELLBERG, CEFE – CNRS, [finn.kjellberg@cefe.cnrs.fr](mailto:finn.kjellberg@cefe.cnrs.fr)

Date of draft: 22/08/2006

Word count: Main text: 3286 (Global count: 4934)

Abstract: 125

## **Abstract**

Modelling the spatial dynamics of populations is a basic objective in ecology and the resulting spatial patterns are diverse and complex. In the metapopulation perspective, the spatial distribution of populations results from a colonization-extinction random process over a network of suitable habitat cells. The aim of this contribution is to show what insights Fourier spectral analysis can provide on metapopulation dynamics in a heterogeneous habitat. We simulated metapopulation dynamics in spatially structured habitat maps and investigated the steady spatial occupancy patterns using Fourier analysis. We showed that the imprint of habitat structure could be separated from the imprint of population dynamics. Hence Fourier spectral analysis provides a promising tool for inferring characteristics of metapopulation dynamics and habitat structure from field occurrence data of particular species.

## **Key words**

Metapopulation models; spatial pattern; heterogeneous habitat; spectral analysis; Fourier transform; Principal Components Analysis.

## Introduction

Much literature in ecology is devoted to the modelling of spatial population dynamics. In this respect, the metapopulation formalism has provided an important contribution. A metapopulation is defined as a set of populations perpetuating itself through a random succession of colonization and extinction events, which is assumed as widely independent from the demography inside populations (Levins, 1969) (for a debate on the assumptions, see Freckleton and Watkinson, 2002). Basically, the metapopulation model with localized colonization events constitutes a particular kind of contact process (Weitz, 1997; Franc, 2004), and therefore is expected to provide particular patterns in space and time (Snyder and Nisbet, 2000; Szabo et al., 2002). Diverse models of metapopulation have been investigated, ranging from spatially fully implicit mean field models (e.g., Levins, 1969; Lande, 1987) to fully spatially explicit models, such as cellular automata (e.g., Bascompte and Sole, 1996; Hanski and Gilpin, 1997; Keymer et al., 2000). Our purpose, here, is to account for a new and simple way to characterize the spatial structure of a metapopulation in a spatially structured environment.

Recent and popular “spatially realistic models” of metapopulations provide analytically tractable solutions to predict the future of a metapopulation in a complex, spatially structured habitat (Hanski, 1997; Moilanen and Hanski, 1998; Hanski and Ovaskainen, 2000). We investigate further the idea that some synthetic spatial statistics can provide relevant and robust insights on the spatial structure of metapopulations. To some extent, the field of landscape ecology has significantly contributed to the study of the properties of diverse spatial statistics. The purpose was to characterize the structure of

complex spatial patterns (O'Neill et al., 1988; O'Neill et al., 1989; Cain et al., 1997; Li, 2000), and to generate "neutral" models of landscape patterns as null hypotheses (Gardner et al., 1987; Gardner and O'Neill, 1991; O'Neill et al., 1992). In particular, it has been shown that spectral analysis based on Fourier transform may provide a powerful tool to analyze patterns induced by multi-scale processes (Platt and Denman, 1975; Muggleston and Renshaw, 1998; Keitt, 2000; Lundquist and Sommerfeld, 2002). Nevertheless it has never been used for analyzing the complex spatio-temporal patterns produced by metapopulation models in structured habitats.

In this paper, we shall show that spectral analysis is a valuable tool, which allows interpreting spatial patterns of metapopulations in heterogeneous habitats in terms of simple, ecologically meaningful, components. Specifically, our purpose is to address the capacity of spectral analysis to retrieve in spatial patterns the imprint of key ecological parameters of a metapopulation model. We therefore referred to a metapopulation model using three input parameters, one defining the population turnover (colonization versus extinction) and two defining the habitat structure (density and aggregation of favorable habitat patches). We then demonstrated the relevance of the Fourier spectral decomposition according to two criteria: (i) its capacity to reflect the variations of the input parameters of the metapopulation model, and (ii) its capacity to separate the effects of the different input parameters introduced into the model.

## Methods

### Simulating habitat maps

We built a 100 by 100 discrete landscape lattice, in which each cell was either suitable or unsuitable for a given species. Different kinds of "neutral", yet spatially structured landscapes, have been proposed, some of them emphasizing self-similar structures (O'Neill et al., 1992; Keitt, 2000). We considered a simple representation characterized by the density  $p$  of favourable sites along with an aggregation parameter  $q$ , which is the conditional probability for a cell to be occupied if a neighbouring cell is (see Hiebeler, 2000). Noting 0 and 1 the suitable and unsuitable cells, respectively, and  $P(00)$  the probability of an adjacent pair of cells to be 0-0,  $P(11)$  to be 1-1, and  $P(01)$  to be 0-1, maps do exist if and only if:  $P(00) = 1 - P(11) - 2.P(01) = 1 - pq - 2p(1 - q) \geq 0$

(Hiebeler, 2000). Hence a necessary and sufficient constraint on  $p$  and  $q$  is  $q \geq 2 - \frac{1}{p}$ .

Maps were generated using the C program of Hiebeler (2000). The parameters  $p$  and  $q$  were made to vary from 0.1 to 0.9, with 0.1 increments, to build the experimental set of binary habitat maps.

### Simulating metapopulation dynamics

For each habitat map, we simulated the dynamics of the given species, under the constraint that it could not persist in an unsuitable cell whereas it could be either present or absent in a suitable cell. At each time step, extinction and colonization events did occur. For a given cell at a given time, extinction events followed a binomial distribution with fixed probability  $m$ , which meant that extinction events were independent between neighbouring cells. An occupied cell provided colonizers to neighbouring cells following



an exponentially decreasing probability  $c(i) = A * \text{Exp}[-i]$ , where  $i$  was the distance of the recipient ( $i \leq 5$ ). The factor  $A$  controlled the global rain of colonizers. For simplicity sake, the first neighbourhood contained all cells in the 3x3 square around the focal cell, and hence the  $i$ th neighbourhood was a  $(2i+1) \times (2i+1)$  square ( $i = 1 \dots 5$ ), which held  $8i$  cells. Colonizers from a given cell were then expected on average to reach  $C$  cells

with  $C = \sum_{i=1, \dots, 5} 8i \times c(i)$ . The ratio  $r = \frac{m}{C}$  is the turnover parameter (see Levins, 1969), and

has been shown to be a good descriptor of metapopulation properties in spatially realistic models (Hanski and Gyllenberg, 1997). In the basic mean field formulation of metapopulation dynamics (Levins, 1969), the equilibrium density in a uniform habitat directly depends on the turnover  $m/C$  (cf. Hanski and Gilpin, 1997). This ratio also determines the stationary density in a spatially realistic formulation with dispersal limitation (Ovaskainen and Hanski, 2001), as we assumed in our simulations. Moreover a simple but important necessary condition for local population persistence in uniform habitat is  $C > m$ . We used a set of  $r$ -values ranging from 0.1 to 0.9, with 0.1 increments. The extinction rate  $m$  was set to 0.09 so that setting the  $r$ -value directly gives the

colonization function  $c(i) = \frac{m}{r} \cdot \frac{\text{Exp}(-i)}{\sum_{k=1, \dots, 5} 8.k.\text{Exp}(-k)}$ . Finally, we initiated all metapopulations

with a 10% random occupancy of the habitat.

Metapopulation dynamics was then simulated over 8000 generations, in order to reach the equilibrium density of populations (Hanski and Gilpin, 1997). We checked for temporal steadiness of density values over the last 300 generations, and referred to such equilibrium as the steady occupancy state. The model yielded a null equilibrium density of populations when the metapopulation went extinct. A set of 522 simulations was

required to cover the desired variations of parameters  $p$ ,  $q$  and  $r$ . We further made five replicate sets of simulations (522 simulations in each set), accounting for different random initial population distributions. It allowed us controlling the generality of patterns found across replicates.

### **Characterizing the steady occupancy state**

To investigate spatial properties of metapopulations at their steady occupancy state, we averaged the cell binary occupancies over  $N$  time steps. Since the estimates may depend on  $N$ , we computed such occupancy probabilities for a wide range of  $N$  values (from 1 to 300). We will call hereafter  $\bar{X} = \{\bar{X}_i\}$  this estimated spatial variable ( $i$  is the spatial index of cells  $(x_i, y_i)$  over the lattice). Each simulated map is characterized by a  $(p, q, r)$  triplet, so that we may write  $\bar{X}(N, p, q, r)$ . The variation of the three parameters  $(p, q, r)$  allows generating a wide diversity of spatial occupancy patterns (Fig. 1a). For instance, the spatial extent of the occupied area over  $N$  population generations can be very small compared to the extent of suitable habitat, even for  $N = 300$  (Fig. 1b). For all our simulations and analyses on occupancy maps, we used the *Mathematica* 5.0 software (Wolfram Research, 2003).

### **Spectral analysis**

Analysing a spatial pattern using the position information in space or time is known to imply statistical caveats, especially because of statistical dependencies according to distance (auto-correlation problems). Instead, spectral analysis allows transforming such information into statistically independent frequency classes. Intrinsically, spectral

analysis would be particularly suited to analyse the spatial aggregation of populations, under a localized extinction-colonisation process (Muggleston and Renshaw, 1996; Couteron, 2001). However the often non-intuitive nature of the results has largely hampered its use in ecology. In the following we will try to reap the advantages of spectral analysis while still producing intuitively straightforward results.

The spatially structured variable  $\bar{X}$  can be fully described either by its values  $\bar{X}_i$  at sampling points  $(x_i, y_i)$  or by a series of periodic spatial functions of frequency  $f_j$  and amplitude  $A_j$ . Hence the  $(x_i, y_i, \bar{X}_i)$  formulation and the  $(f_j, A_j)$  *spectral* formulation are equivalent (space-frequency duality). The *power spectrum* decomposition relies on this formulation and allows partitioning the variance of  $\bar{X}$  in the spectral domain. It emphasizes spatial scales at which the properties of population dynamics may express itself.

We hence conducted a 2D spatial Fourier analysis on the 100 by 100 discrete occupancy maps at the steady occupancy state (time-averaged occupancy maps  $\bar{X}(N, p, q, r)$ ), and computed the power spectra (see Appendix). The power spectrum  $PS(\bar{X})(\frac{k-1}{n}, \frac{l-1}{n})$  provided the fraction of the variance in  $\bar{X}$  that relates to the spatial frequency  $\left(f_x = \frac{k-1}{n}, f_y = \frac{l-1}{n}\right)$  (see Platt and Denman 1975).

An important and well-documented issue in spectral analysis is the necessity for spectral smoothing (e.g. Platt and Denman, 1975). Indeed the power spectrum decomposition provides as many spectral estimates as observations (here  $100 \times 100 = 10\,000$  estimates). Smoothing consists in reducing the number of spectral components by pooling them into classes, so that more than one degree of freedom is available to

estimate the smoothed spectral components. Fixing an upper limit on frequencies, according to the Nyquist criterion, allows to escape the well-documented aliasing artefact (see Platt and Denman, 1975) . See more details on these issues in the Appendix. All spectral calculations were performed using *Mathematica* 5.0 (Wolfram Research, 2003), and statistical analyses with *SPlus* 6.2 (Venables and Ripley, 2002).

### **Principal Component Analysis on spectra**

In order to describe the variation of the smoothed power spectra, we performed a Principal Component Analysis (PCA) over them, using spatial frequencies as variables and simulations as observations (see Coueron et al., 2006). We performed the analysis on all the simulations that did not go extinct. Correlations between resulting PCA components and simulation parameters  $p$ ,  $q$  and  $r$  were computed (Pearson coefficient) and compared with null expected coefficient values in the absence of correlation ( $t$ -test using the *SPlus* software). We were specifically interested in investigating the ability of spectral analysis to separate the spatial imprint of metapopulation dynamics from the spatial imprint of habitat on short timescales.

### **Results**

As expected, the model enabled to produce highly diversified spatial occupancy patterns (see some patterns in Fig. 1). We performed spectral analysis on simulated metapopulations over a wide range of  $(p, q, r)$  parameters at the steady occupancy state. For a given set of  $(p, q, r)$  values, the five replicated patterns obtained with different random initial conditions yielded very similar spectral results (not shown). Hence, we

focus the following analysis on one arbitrary reference replicate, for which 227 simulations out of 522 did not go extinct at equilibrium.

### **Population turnover vs. habitat structure**

We consider the Fourier power spectra of time-averaged metapopulation maps  $\bar{X}(N, p, q, r)$ , with  $N=300$  time steps. There are 25 frequency components, with 5 classes over the abscissa  $i$ , and 5 classes over the ordinates  $j$  (e.g. Fig. 2). The PCA was performed over all 227 simulations spectra using these 25 components as variables. The first Fourier principal component (FPC1) explained 49.7 % and the two first ones (FPC1 + FPC2) 90.8 % of total spectra variance. Hence most of the spatial variance at the steady occupancy state can be summarized by two independent gradients in the spectral domain.

The spectral signatures of principal components were characteristic (Fig. 2). FPC1 had positive, slightly decreasing loadings with respect to spatial frequencies  $I_{ij}$ . This was a scale-free component, with a power law shape. In contrast FPC2 revealed opposite contributions of high and low spatial frequencies, with negative and positive loadings, respectively. Hence FPC2 evidences long distance spatial structuring.

Importantly, FPC1 and FPC2 were correlated with the variables of interest. FPC1 was highly and significantly correlated with the metapopulation parameter  $r$  but not with the aggregation parameter  $q$ , while it was weakly though significantly linked to  $p$  (Pearson correlation coefficients) (Table 1a). Fig. 3 illustrates the relationship between  $r$  and FPC1 and FPC2. FPC2 was correlated to habitat structure through a high and

significant correlation with  $q$ , whereas it was not correlated to  $r$ . Again there was a lower correlation with  $p$ . Finally, there was a high correlation between  $q$  and  $r$ .

We conclude that FPC1 and FPC2 captured the uncorrelated part of the variation of  $q$  and  $r$  and that spectral analysis of cell occupancy at the steady occupancy state separated a first scale-free component associated with metapopulation dynamics (FPC1), and a second one associated with habitat aggregation (FPC2). Habitat density  $p$  appeared to be a confounding parameter in this sorting, as it was correlated to both principal components. This suggests that  $p$  is not a good descriptor of the habitat to investigate the imprint of the spatial habitat structure on observed population distributions.

### **Temporal sampling effects**

The results presented up to now were obtained for time-averaged metapopulation occupancy maps with  $N = 300$ . We also computed several maps with time-averaging over  $N = 1$  (binary presence-absence from a single observation), 2, 5, 10, 20, 50, 100, 150, 200 and 300 consecutive generations, in order to control for the effect of the number of sampled generations on spatial statistics and spectral patterns. The proportion of variance explained by PCA axis 1 and 2 stabilized when the number of sampled generations,  $N$ , was greater than 20 (Fig. 4), a value above which spatial features were broadly unchanged. For lower values of  $N$ , more variance was explained by FPC1, and less by FPC2, the sum of both remaining fairly constant.

The analysis of correlations between  $p$ ,  $q$ ,  $r$ , FPC1 and FPC2 showed that the decoupling property on the effects of parameters  $q$  and  $r$  was significant as soon as  $N = 2$  (Table 1b). Noticeably, the correlation of  $p$  with FPC1 increased as  $N$  decreased, so that

its confounding effect was reinforced. It is important to note that spectral decoupling no longer worked for  $N = 1$ ; and that at least two successive generations were needed to decouple features of population turnover from those of habitat spatial structure. This implied that we could not capture independent insights for both habitat structure and population dynamics with presence-absence data.

## **Discussion**

We aimed at identifying through Fourier analysis the spatial features of cell occupancy, which are driven by the structure of suitable habitat (binary representation) and those driven by metapopulation dynamics. We demonstrated that the Fourier power spectrum analysis of population occupancy patterns at equilibrium effectively allowed distinguishing between one structural feature of habitat ( $q$ ) and a parameter of population turnover ( $r$ ). Below we discuss the potential scope and limits of the approach.

## **Insights from spectral analysis**

Thanks to smoothing, the spectral analysis allowed us to remove contingent spatial structures in individual simulated habitat maps so as to only retain general features on habitat structure and population dynamics. Most of the variance among simulations spectra could be expressed on only two PCA axes, one representing populations dynamics (parameter  $r$ ), and the other representing habitat structure (parameter  $q$ ). This decoupling is a new and strong result in the field of metapopulation modelling. By pooling spectral components into classes of frequency, spectral smoothing allows

removing most of the spatial noise introduced by the stochastic nature of simulated maps, and also by the stochastic variation in metapopulation dynamics.

Interestingly, although the colonization process was very local in our modelling, namely a decreasing function over the five closest neighbors, we observed that its spatial imprint could be perceptible not only locally but also at large spatial scale. Such emergence is a signature of larger-scale dynamical features, which are known in the more general field of contact processes (Weitz, 1997; Oborny et al., 2005). For low habitat occupancy rates (large  $r$  values), large parts of the suitable habitat remained unoccupied over  $N = 300$  generations whereas the occupied part was very aggregated (see Fig. 1b left). Such self-organized large-scale spatial structures are largely independent from habitat spatial structure.

Hence we highlight that local metapopulation processes can produce patterns at large scales, so that cross-scale spatial information might be important and informative. Spectral analysis, though little used in ecology, may provide a most interesting tool to capture such ecological determinants of metapopulation spatial structure.

### **Current limits and perspectives**

We showed that sampling effects became strong for very low  $N$  values. In particular, when  $N = 1$  (presence-absence binary data), i.e. with a single temporal observation, spectral decoupling was no longer present. The Fourier decomposition technique is indeed a useful and powerful tool for analyzing continuous signals but, in a finite space, it becomes less efficient for discrete signals. Therefore the Fourier power spectrum analysis is not suited to binary presence-absence maps, and this may explain why we



failed to detect spectral decoupling in binary maps. It is an open question whether the use of more suited methods, such as the Haar wavelet decomposition (Chui, 1992), would reveal the patterns we obtained with time averaged data.

We used Fourier spectral analysis because it relies on a rather simple numerical transformation, with well-documented properties, and because very powerful algorithms are available (such as Fast Fourier Transform, FFT). Hence it is very powerful for treating large data sets on regular grids and this allowed us to draw strong conclusions. However, this method is not suited for the more frequently used irregular sampling designs, and that may be an important limitation when studying real datasets. Further studies will focus on whether other spectral techniques, such as the Principal Coordinates of Neighbor Matrices (Borcard and Legendre, 2002), may enable to detect similar patterns in situations involving irregular spatial sampling.

We noticed that a local colonization/extinction process could induce on FPC1 a characteristic scale-free pattern, probably self-organized and self-similar in space and time (see also Weitz, 1997), whereas the large-scale structure of the habitat was highlighted on FPC2. In a broader perspective, we should accordingly investigate some other kinds of neutral habitat maps, especially themselves self-similar.

Our study demonstrates that spectral analysis is a powerful and promising method for spatial ecology. Inference will be the ultimate goal. The decoupling property suggests that we can extract from species distribution spectra some unambiguous information on metapopulation dynamics. Although Fourier analysis failed with binary presence-absence data, we showed that only two colonization-extinction cycles were required to

get some characteristic spectral decoupling properties. However, because of some confounding effect of parameter  $p$ , it may be more reasonable to use at least 5 generations for capturing the spectral decoupling. With the parameters used (extinction rate  $m = 0.09$ ), this means that the decoupling is apparent as soon as a third of the available habitat experienced at least one extinction event ( $Pr(\text{at least one extinction over 5 cycles}) = 1 - (1 - 0.09)^5 = 0.38$ ). Indeed, we used a rather high extinction rate ( $m = 0.09$ ). This implies that our technique could be applied directly to the study of highly dynamic species for which data over several years are available. Practically, metapopulation data concerning butterflies (e.g., *Melitaea cinxia*, Hanski et al., 1994; *Procllossiana eunomia*, Schtickzelle et al., 2002) would meet such requirements.

In summary, our metapopulation model relied on only three ecological variables: population turnover  $r$ , habitat density  $p$  and habitat aggregation  $q$ . Although these variables were somewhat correlated because of constraints on the persistence of the metapopulations and on the spatial structure of habitat maps, we demonstrated that over 90 % of the overall variance among the Fourier spectra could be independently described by only two parameters,  $r$  and  $q$ , which are population turnover and habitat aggregation, respectively. Hence this study paves the way to get insights simultaneously on habitat spatial structure and on population dynamics from field occurrence data of real species.

## Acknowledgement

We thank P. Couteron for his useful comments on the manuscript.

## References

- Bascompte, J. and Sole, R. V., 1996. Habitat Fragmentation and Extinction Thresholds in Spatially Explicit Models. *Journal of Animal Ecology* 65, 465-473.
- Borcard, D. and Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* 153, 51-68.
- Cain, D.H., Riitters, K. and Orvis, K., 1997. A multi-scale analysis of landscape statistics. *Landscape Ecology* 12, 199-212.
- Chui, C.K., 1992. *An Introduction to Wavelets*. Academic Press Inc., San Diego.
- Couteron, P., 2001. Using spectral analysis to confront distributions of individual species with an overall periodic pattern in semi-arid vegetation. *Plant Ecology* 156, 229-243.
- Couteron, P., Barbier, N. and Gautier, D., 2006. Textural ordination based on Fourier spectral decomposition: a method to analyze and compare landscape patterns. *Landscape Ecology* 21, 555-567.
- Franc, A., 2004. Metapopulation dynamics as a contact process on a graph. *Ecological Complexity* 1, 49-63.
- Freckleton, R.P. and Watkinson, A.R., 2002. Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Journal of Ecology* 90, 419-434.
- Gardner, R.H. and O'Neill, R.V., 1991. Pattern, process, and predictability: the use of neutral models for landscape analysis. In: M.G. Turner and R.H. Gardner

- (Editors), Quantitative methods in landscape ecology. Springer-Verlag, New York, pp. 289-307.
- Gardner, R.H., Milne, B.T., Turner, M.G. and O'Neill, R.V., 1987. Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology* 1, 19-28.
- Hanski, I., 1997. Predictive and Practical Metapopulation Models: The Incidence Function Approach. In: D. Tilman and P. Kareiva (Editors), *Spatial Ecology - The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, NJ, pp. 21-45.
- Hanski, I. and Gyllenberg, M., 1997. Uniting Two General Patterns in the Distribution of Species. *Science* 275, 397-400.
- Hanski, I. and Ovaskainen, O., 2000. The metapopulation capacity of a fragmented landscape. *Nature* 404, 755-758.
- Hanski, I., Kuussaari, M. and Nieminen, M., 1994. Metapopulation structure and migration in the butterfly *Melitaea cinxia*. *Ecology* 75, 747-762.
- Hanski, I.A. and Gilpin, M.E. (Editors), 1997. *Metapopulation Biology*, 1. Academic Press, San Diego, California, 525 pp.
- Hiebeler, D., 2000. Populations on fragmented landscapes with spatially structured heterogeneities: landscape generation and local dispersal. *Ecology* 81, 1629-1641.
- Keitt, T.H., 2000. Spectral representation of neutral landscapes. *Landscape Ecology* 15, 479-494.
- Keymer, J.E., Marquet, P.A., Velasco-Hernandez, J.X. and Levin, S.A., 2000. Extinction Thresholds and Metapopulation Persistence in Dynamic Landscapes. *The American Naturalist* 156, 478-494.

- Lande, R., 1987. Extinction thresholds in demographic models of territorial populations. *The American Naturalist* 130, 624-635.
- Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15, 237-240.
- Li, B.-L., 2000. Fractal geometry applications in description and analysis of patch patterns and patch dynamics. *Ecological Modelling* 132, 33-50.
- Lundquist, J.E. and Sommerfeld, R.A., 2002. Use of fourier transforms to define landscape scales of analysis for disturbances: a case study of thinned and unthinned forest stands. *Landscape Ecology* 17, 445-454.
- Moilanen, A. and Hanski, I., 1998. Metapopulation dynamics: effects of habitat quality and landscape structure. *Ecology* 79, 2503-2515.
- Mugglestone, M.A. and Renshaw, E., 1996. A practical guide to the spectral analysis of spatial point processes. *Computational Statistics & Data Analysis* 21, 43-65.
- Mugglestone, M.A. and Renshaw, E., 1998. Detection of geological lineations on aerial photographs using two-dimensional spectral analysis. *Computers & Geosciences* 24, 771-784.
- O'Neill, R.V., Gardner, R.H. and Turner, M.G., 1992. A hierarchical neutral model for landscape analysis. *Landscape Ecology* 7, 55-61.
- O'Neill, R.V., Johnson, A.R. and King, A.W., 1989. A hierarchical framework for the analysis of scale. *Landscape Ecology* 3, 193-205.

- O'Neill, R.V., Krummel, J.R., Gardner, R.H., Sugihara, G., Jackson, B., DeAngelis, D.L., Milne, B.T., Turner, M.G., Zygmunt, B., Christensen, S.W., Dale, V.H. and Graham, R.L., 1988. Indices of landscape pattern. *Landscape Ecology* 1, 153-162.
- Oborny, B., Meszena, G. and Szabo, G., 2005. Dynamics of populations on the verge of extinction. *Oikos* 109, 291-296.
- Ovaskainen, O. and Hanski, I., 2001. Spatially Structured Metapopulation Models: Global and Local Assessment of Metapopulation Capacity. *Theoretical Population Biology* 60, 281-302.
- Platt, T. and Denman, K.L., 1975. Spectral analysis in ecology. *Annual Review of Ecology and Systematics* 6, 189-210.
- Schtickzelle, N., Le Boulenge, E. and Baguette, M., 2002. Metapopulation dynamics of the bog fritillary butterfly: demographic processes in a patchy population. *Oikos* 97, 349-360.
- Snyder, R.E. and Nisbet, R.M., 2000. Spatial structure and fluctuations in the contact process and related models. *Bulletin of Mathematical Biology*, 62, 959-975.
- Szabo, G., Gergely, H. and Oborny, B., 2002. Generalized contact process on random environments. Santa Fe Institute, Santa Fe, NM, 6 pp.
- Venables, W.N. and Ripley, B.D., 2002. *Modern Applied Statistics with S* - Fourth Edition. Springer, xi+495 pp.
- Weitz, J.S., 1997. Generalized Contact Processes in Ecology. Philosophy Thesis, A.B, Princeton University, Princeton, 168 pp.
- Wolfram Research, I., 2003. *Mathematica*. Wolfram Research, Inc., Champaign, Illinois.

## Figures captions

**Figure 1:** Examples of spatial patterns of metapopulations

a) Averaged cell occupancies  $\bar{X}(N, p, q, r)$

Three types of spatial structures are exemplified. Unoccupied habitat cells are in black. In occupied cells, time averaged occupancy probabilities (300 generations) are shown in grey levels: lighter cells have higher mean occupancy.

- Diffuse metapopulation ( $p = 0.8, q = 0.8, r = 0.7$ ): overall low local occupancy probability in a extended habitat.
- Archipelago metapopulation ( $p = 0.3, q = 0.7, r = 0.2$ ): contrast between high and low occupancy areas in a fragmented habitat.
- Uniform metapopulation ( $p = 0.9, q = 0.9, r = 0.1$ ): overall high local occupancy probability in a extended habitat.

b) Occupied habitat at the steady occupancy state

Black cells are unsuitable, grey cells represent potentially suitable sites that were never colonised during the sampling period. White cells indicate sites that were colonised at least once.

**Figure 2:** Loadings of the first two principal components FPC1 and FPC2 of the PCA on metapopulation power spectra.

Principal components loadings according to spectral frequency intervals (see smoothing procedure) are monotonous functions of frequencies. However, FPC1 is

always positively valued (white top bars), whereas FPC2 shows a threshold that contrasts high frequencies (negative loadings, grey top bars), and low frequencies (positive loadings, white top bars).

**Figure 3:** Pattern of  $r$  variation over the FCP1 and FCP2 axes. Small light dots correspond to low  $r$ -values whereas large dark dots correspond to high  $r$ -values.

**Figure 4:** Effect of temporal sampling on the analysis of the  $\bar{X}(N, p, q, r)$  maps: explained variance of PCA axis against the number  $N$  of sampled generations. Points are FPCA1 weights, triangles show FPCA2 weights. Sampling effects at low  $N$  increase the weight of FPCA1 and make the spectral decoupling less clear. Yet, as soon as  $N$  is greater than 20, weights of axis FPCA1 and FPCA2 reach an asymptote and simulation patterns over FPCA1 and FPCA2 spaces become independent of  $N$ .



## Appendix

### Fourier spectral analysis

The two-dimensional discrete Fourier transform for  $\bar{X}$  is a function depending on two dimensions of geographical frequencies, hereafter indexed  $u = \frac{k-1}{n}$  and  $v = \frac{l-1}{n}$  :

$$\mathfrak{I}(\bar{X})\left(\frac{k-1}{n}, \frac{l-1}{n}\right) = \frac{1}{n^2} \sum_{j=1}^{n^2} \bar{X}_j \cdot e^{2\pi i [(x_j-1)(k-1) + (y_j-1)(l-1)]/n}$$

It is a complex number, with  $k = \{1, \dots, n\}$  and  $l = \{1, \dots, n\}$ .

$\frac{k-1}{n}$  and  $\frac{l-1}{n}$  are spatial frequencies following respectively abscissa and ordinates.

When these frequencies are both set at 0, we have  $\mathfrak{I}(\bar{X})(0,0) = \frac{1}{n^2} \sum_{j=1}^{n^2} \bar{X}_j(r, s) = \bar{\bar{X}}$ ,

which is the general mean over space and time.

For a discrete map, the Nyquist sampling theorem states that high spatial frequency Fourier components are biased and must be set to 0. It leads to only retain the frequencies for  $k = \{1, \dots, n/2\}$  and  $l = \{1, \dots, n/2\}$  with  $n/2 = 50$  (see Platt and Denman, 1975).

The power spectrum is derived from the expression of the basic Fourier spectrum

through its modulus:  $PS(\bar{X})\left(\frac{k-1}{n}, \frac{l-1}{n}\right) = \frac{1}{2} \left| \mathfrak{I}(\bar{X})\left(\frac{k-1}{n}, \frac{l-1}{n}\right) \right|^2$ .

### Smoothing formula

Given an integer  $s$  divisor of  $n$ , we partitioned the initial spatial frequencies  $(u, v)$  into intervals  $I_{i,j}$ :

$$I_{i,j} = \left[ \frac{s(i-1)}{n}, \frac{s.i-1}{n} \right] \times \left[ \frac{s.(j-1)}{n}, \frac{s.j-1}{n} \right], \text{ for } i, j = 1, \dots, \text{Int}(n/s)$$

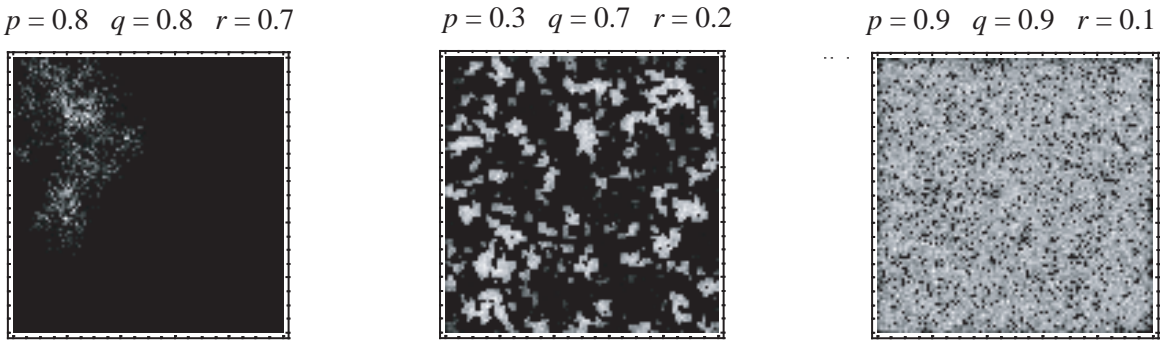
$\text{Int}$  is the integer part function (rounded to the lower integer in absolute value).

We then defined a smoothed power spectrum:  $\text{smoothPS}(\bar{X})(I_{i,j}) = \frac{\sum_{(u,v) \in I_{i,j}} \text{PS}(\bar{X})(u,v)}{k}$ .

We took  $s = 10$ .

Figure 1

(a) Local occupancy probabilities estimated over  $N = 300$  quasi-stationary generations



(b) Spatial extent of occupied habitat (white), inside the available habitat (grey), over  $N = 300$  generations

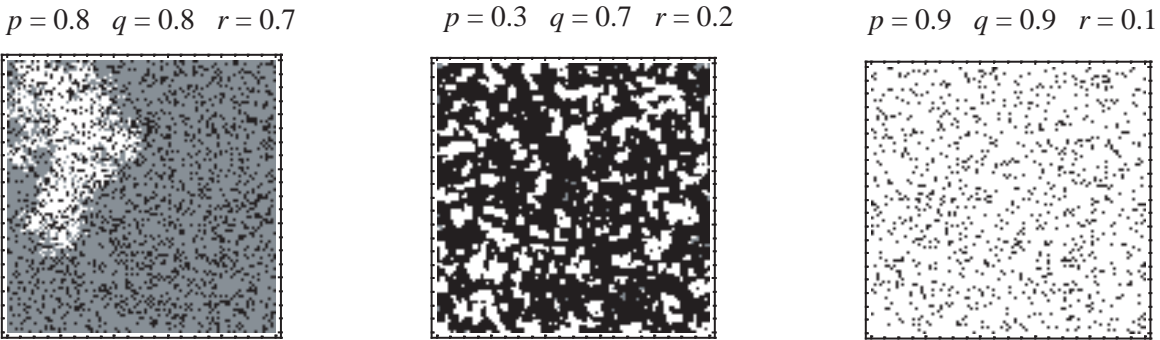


Figure 2

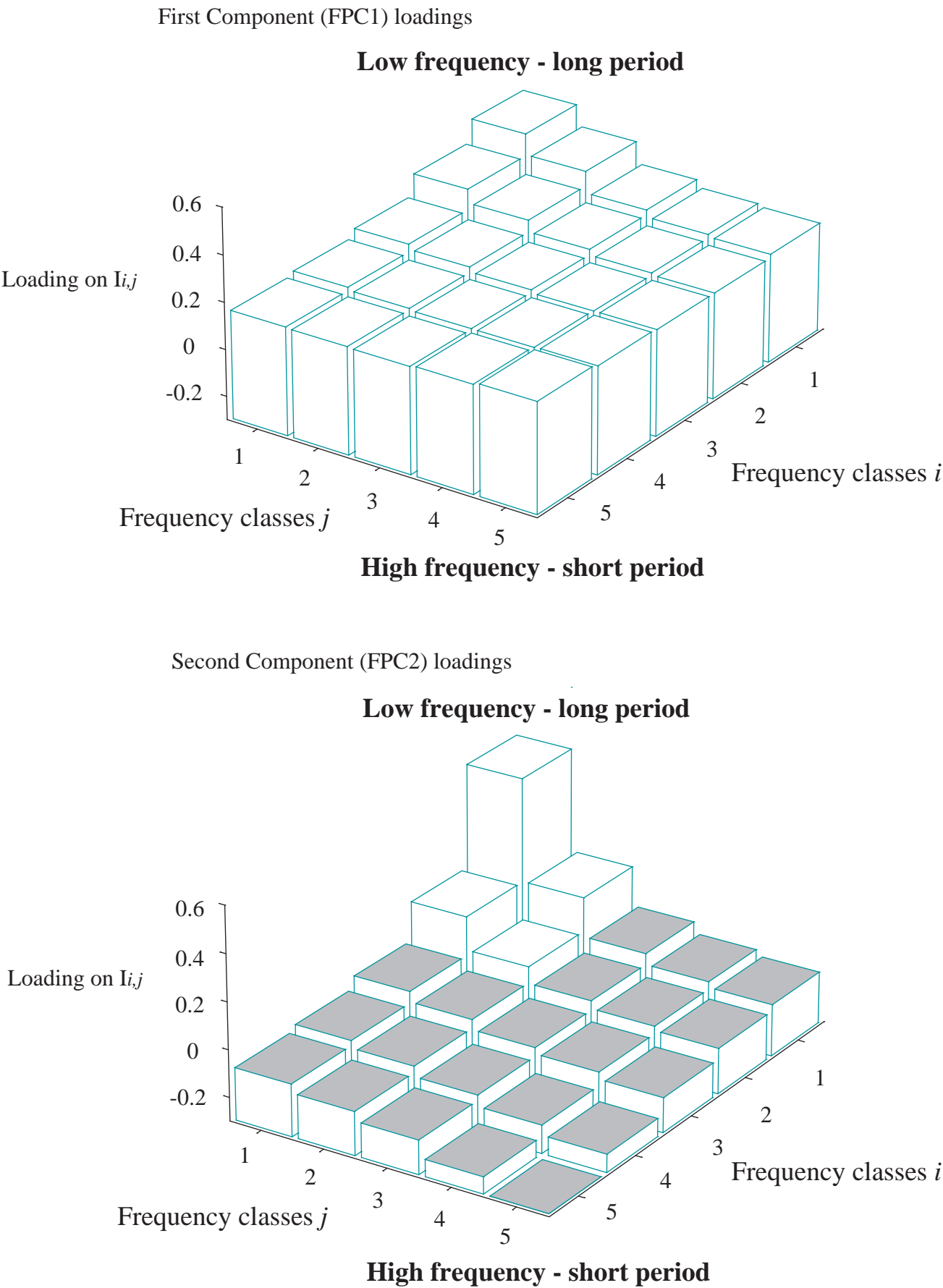


Figure 3

Variation for parameter  $r$

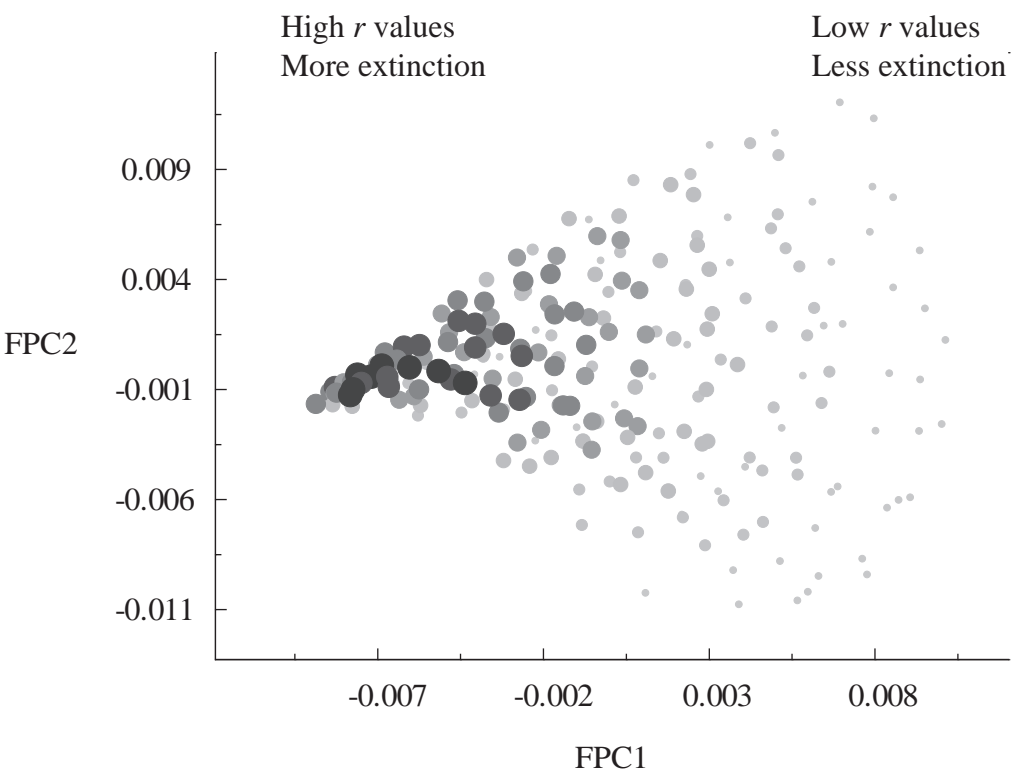
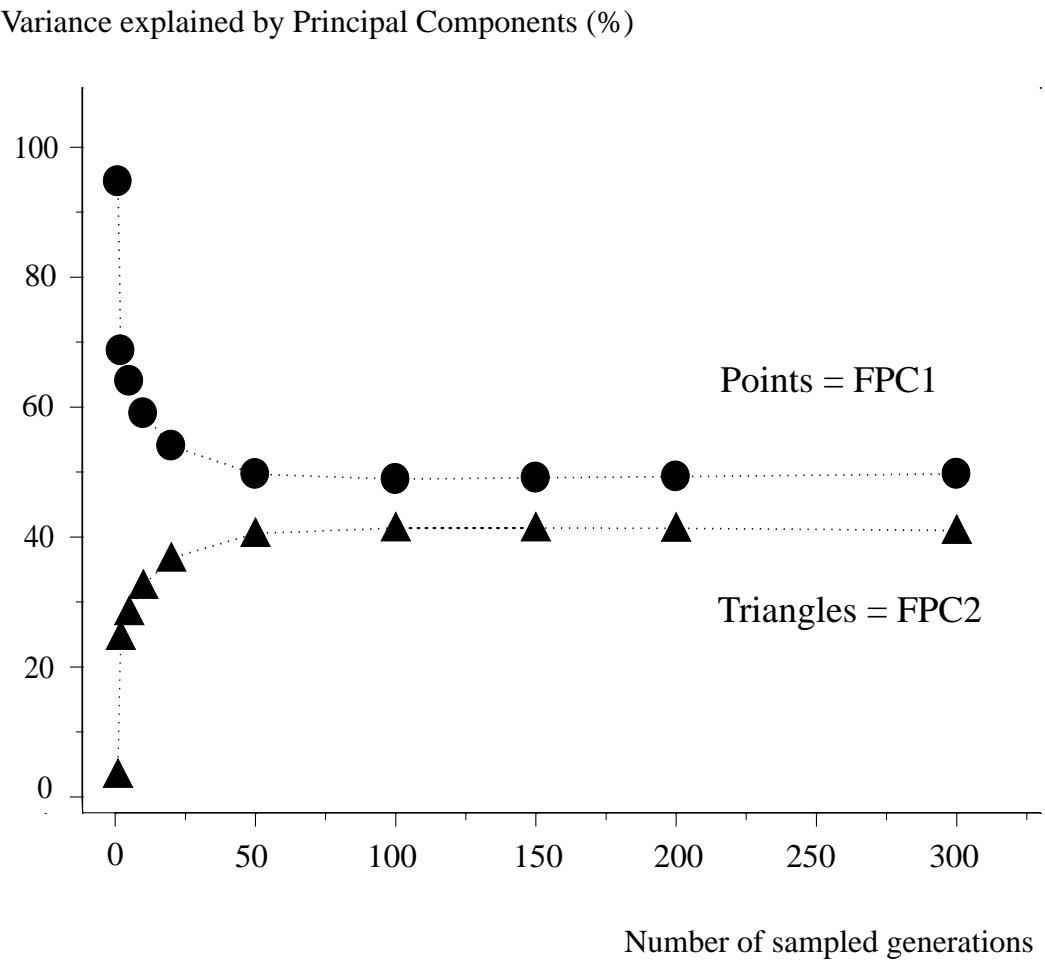


Figure 4



Tables

**Table 1:** Correlation analysis on metapopulation parameters and Principal Components for two time ranges of occupancy statistics: over 300 generations (above the diagonal) and over 2 generations (below the diagonal).

	FPC1	FPC2	<i>p</i>	<i>q</i>	<i>r</i>
FPC1	1	0	<b>0.2914***</b>	-0.0202	<b>-0.6985***</b>
FPC2	0	1	<b>0.2727***</b>	<b>-0.667***</b>	-0.069
<i>p</i>	<b>0.6220***</b>	<b>0.2513***</b>	1	<b>0.197**</b>	<b>0.2553***</b>
<i>q</i>	-0.0037	<b>-0.6189***</b>	<b>0.197**</b>	1	<b>0.4320***</b>
<i>r</i>	<b>-0.3914***</b>	-0.11	<b>0.2553***</b>	<b>0.4320***</b>	1

The occupancy probabilities are estimated over 300 generations and 2 generations. The Pearson’s correlation coefficients are tested against the null hypothesis with no correlation, using a *t*-test (computed with *SPlus*). Rejecting this hypothesis with a type I error of 5% is shown by a star \*, of 1% by two stars \*\*, and of 0.1% by three stars \*\*\*.