



UNIL | Université de Lausanne

Faculté des géosciences et de l'environnement
Institut de géographie

Etude des modèles de Whittle markoviens probabilisés

Thèse de doctorat

présentée à la Faculté des géosciences et de l'environnement
de l'Université de Lausanne par

David CARRILLO

diplômé en physique de l'Université de Genève

Jury :

M. le Prof. Klaus HOLLINGER, Université de Lausanne, Président du jury,
M. le Prof. François BAVAUD, Université de Lausanne, Directeur de thèse,
M. le Dr Sylvain SARDY, Université de Genève, Expert externe,
M. le Prof. Mikhail KANEVSKI, Université de Lausanne, Expert interne.

Thèse soutenue publiquement le 3 avril 2009.

Lausanne, 2009.



UNIL | Faculté des géosciences
et de l'environnement

Imprimatur

Vu le rapport présenté par le jury d'examen composé de

Président du colloque : M. le Professeur Klaus Holliger
Directeur de thèse : M. le Professeur François Bavaud
Expert externe : M. le Docteur Sylvain Sardy
Expert interne : M. le Professeur Mikhaïl Kanevski

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur David Carrillo

Titulaire d'un Diplôme de physicien de l'Université de Genève délivré en 1993

intitulée

Etude des modèles de Whittle markoviens probabilisés

Lausanne, le 3 avril 2009

Pour le Doyen de la Faculté des géosciences et de l'environnement



Prof. Klaus Holliger

Table des matières

1. Introduction à l'autocorrélation spatiale	7
1.1 Introduction	7
1.2 La voie des modèles spatiaux	8
1.2.1 Le modèle général	8
1.2.2 Des modèles particuliers	9
1.2.3 Le digraphe des interactions dans les modèles SAR	10
1.3 La voie des indices	12
1.3.1 Le produit croisé	12
1.3.2 Problématique : la matrice des poids spatiaux	13
1.4 Problème traité dans la thèse	14
1.4.1 Modèle de Whittle	14
1.4.2 Applications du modèle de Whittle classique	15
1.4.3 Vers plus de complexité	17
1.4.4 Le modèle de Whittle markovien probabilisé	18
1.4.5 Matrice des interactions	18
1.4.6 Perspectives pour le modèle de Whittle probabilisé	20
1.4.7 Application à un réseau percolatif	22
1.4.8 Plan de la thèse	24
2. Modèle de Whittle markovien probabilisé	27
2.1 Spécification du modèle	27
2.1.1 Hypothèses de travail	27
2.1.2 Champ spatial résultant	28
2.2 Variance du champ	29
2.2.1 Expression générale de la matrice de variance-covariance	29
2.2.2 Cas où \mathbf{M} est un paramètre	31
2.2.3 Cas où \mathbf{M} est une variable	34
2.3 Corrélation du champ	41
2.4 Effets de bord de la fenêtre d'observation-modélisation	43
2.5 Conclusion	44

3. Information de Fisher	51
3.1 Définition et rôle de l'information de Fisher	51
3.2 Information de Fisher du coefficient d'autorégression ρ	53
3.2.1 Cas général	53
3.2.2 Application à un réseau carré	54
3.2.3 Application à une triangulation de Delaunay sur un semis de points	59
3.2.4 Application sur un cas concret : les comtés d'Irlande	63
3.3 Information de Fisher du paramètre p	65
3.4 Divergence de Kullback-Leibler	66
3.5 Conclusion	67
4. Estimation ponctuelle du coefficient d'autorégression	69
4.1 Estimation de MV du coefficient d'autorégression	69
4.1.1 Fonction de vraisemblance	69
4.1.2 Estimation de MV du paramètre p	70
4.1.3 Estimation de MV du coefficient d'autorégression	70
4.1.4 Estimation par l'algorithme EM	71
4.2 Autres estimateurs du coefficient d'autorégression	74
4.2.1 Définition	74
4.2.2 Méthode par l'indice I de Moran	75
4.2.3 Méthode par l'indice I de Moran maximal	77
4.2.4 Méthode par comptage d'amas	84
4.3 Estimation bayésienne des paramètres	90
4.3.1 L'échantillonneur de Gibbs	90
4.4 Conclusion	93
5. Inférence du coefficient d'autorégression	95
5.1 Éléments de la théorie générale du test	95
5.1.1 Formulation des hypothèses du test	95
5.1.2 Règle de décision et risques	96
5.1.3 Statistiques de test	98
5.2 Test du I de Moran	99
5.2.1 Une variante du test de Cliff-Ord	99
5.2.2 Limites à la puissance du test	101
5.3 Efficacité asymptotique du test	103
5.3.1 Définition générale	103
5.3.2 Efficacité de Pitman	104
5.3.3 Application	105
5.4 Test du I de Moran	107
5.4.1 Spécification du test	107
5.4.2 Puissance pour des valeurs de ρ faibles	108
5.4.3 Puissance pour des valeurs de ρ plus élevées	109
5.5 Test du I de Moran maximal	115

5.5.1	Spécification du test	115
5.5.2	Puissance du test	115
5.6	Parallèle dans le paradigme bayésien	116
5.7	Conclusion	118
6.	Conclusion	121
A.	Notations mathématiques	125
B.	Annexes du chapitre 1	129
B.1	Entropie de Shannon	129
B.2	La régression linéaire classique	129
B.3	Théorie de la percolation	131
C.	Annexes du chapitre 2	137
C.1	Loi d'un vecteur de k variables de Bernouilli	137
D.	Annexes du chapitre 3	139
D.1	Information de Fisher de ρ au voisinage de $\rho = 0$	139
D.2	Divergence de Kullback-Leibler	140
E.	Annexes du chapitre 5	143
E.1	Moments de l'indice de Moran	143
E.2	Expression exacte de la fonction de répartition du I de Moran	149
E.3	Approximation numérique de la fonction de répartition du I de Moran	150
E.4	Approximation asymptotique de la distribution de probabilité du I de Moran	151
E.5	Autres approximations de la distribution de probabilité du I de Moran	151
E.6	Approximation de Taylor du rapport des vraisemblances	152
E.7	Base informative du test du rapport de maximum de vraisemblance	153
E.8	Efficacité de Pitman entre deux I de Moran	156
F.	Annexes du chapitre 5	161
F.1	Méthode du I de Moran maximal	161
F.2	Le diagramme de dispersion angle-Imax	164
F.2.1	Définition	164
F.2.2	Résultats	165
G.	Bibliographie	179

Introduction à l'autocorrélation spatiale

1.1 Introduction

La présente thèse s'intéresse à un aspect de l'autocorrélation spatiale à la fois sous une forme intensive et sous une forme connective. Permettons-nous tout d'abord d'introduire le concept d'autocorrélation spatiale dans la perspective du géographe en citant Vigneron (1997). *"On appelle autocorrélation spatiale le fait qu'en général deux lieux, ou deux unités spatiales quelconques, voisins ont plus tendance à se ressembler que deux lieux éloignés l'un de l'autre. Dans ce cas les individus géographiques ne peuvent être considérés comme des individus statistiques indépendants que l'on peut soumettre sans biais aux tests statistiques. Cette contrainte est souvent ignorée : sa correction impose, en effet, un renversement des façons habituelles de faire de la géographie. Mais l'autocorrélation spatiale est aussi la justification même de la géographie. Sans elle on peut même dire que la géographie n'aurait aucun intérêt scientifique. Elle signifie en effet que les valeurs ne sont pas réparties au hasard et témoigne donc de l'existence d'un certain degré de structuration de l'espace. La mesure de l'autocorrélation spatiale doit être un préalable à l'analyse statistique des données géographiques"* (Vigneron 1997).

La citation de Vigneron appelle le commentaire suivant. Lorsque l'on parle de *"lieux ou d'unités spatiales qui se ressemblent"*, c'est en référence aux attributs associées à ces lieux ou unités spatiales. A sa façon, Vigneron évoque ce que Tobler (1970) énonçait comme la première loi de la géographie : *"Everything is related to everything else, but near things are more related than distant things"*. Vigneron parle de *ressemblance*, Tobler parle de *relation*, et tous deux touchent à l'aspect observable de l'autocorrélation spatiale sensible au choix des objets spatiaux qui définissent l'espace et à leurs localisations relatives (cf. Openshaw & Taylor, 1979).

Dans l'article de réflexion (au registre de la méthodologie) de Griffith (1992) au titre évocateur de *"What is spatial autocorrelation ?"*, l'on se rend compte qu'il

n'existe en fait pas de définition claire et concise de l'autocorrélation spatiale et que le principe de Tobler ou la définition de Vigneron est une description générique de l'autocorrélation spatiale que l'on trouve fréquemment dans la littérature. L'autocorrélation spatiale peut être inhérente à la géographie, mais sa signification est contextuelle et possède plusieurs facettes. Elle peut être définie comme (Griffith, 1992) : (a) la corrélation d'une variable avec elle-même provenant de l'ordonnement géographique des données, (b) un instrument de description de la nature et de l'intensité d'une structure spatiale, (c) un indicateur de la quantité d'information latente contenue dans les données spatialisées, en particulier l'information qui s'avère toujours négligée dans les estimations statistiques classiques quand elles sont appliquées aux séries de données spatiales, (d) un outil permettant de repérer l'existence de variables significatives, mais non prises en compte dans le modèle, (e) un substitut à des données manquantes, (f) un obstacle pour l'application des méthodologies statistiques conventionnelles à des séries de données spatiales, (g) un indicateur du bien-fondé d'une partition spatiale, voire un artefact introduit par les frontières, (h) un mécanisme d'un processus spatial et (i) un effet de redistribution sur les lieux avoisinants.

1.2 La voie des modèles spatiaux

1.2.1 Le modèle général

Communément à toutes ses formes, l'autocorrélation spatiale est un concept ayant à la fois un ancrage thématique et un ancrage spatial. Elle peut être appréhendée à travers un modèle spatial dont la définition générale est un ensemble \mathcal{X} de variables aléatoires dépendantes de leur localisation spatiale (Cressie, 1993, p. 8) :

$$\mathcal{X} = \{\mathbf{X}(\mathbf{r}) | \mathbf{r} \in D\}, \quad (1.1)$$

où D est un sous-ensemble aléatoire ou constant d'un espace euclidien d -dimensionnel ($D \in \mathbb{R}^d$). Sous un aspect strictement spatial pour lequel la dimension temporelle n'existe pas, les données observées sont une réalisation unique de \mathcal{X} , amenant à considérer de facto les données comme un individu d'une *superpopulation* généralement infinie (de données). C'est en référence à cette superpopulation que l'on peut attribuer une corrélation statistique entre les éléments de D .

Nous prendrons l'habitude d'appeler \mathbf{X} un champ spatial (synonyme de modèle spatial). Dans un classement de types de données et donc de modèles, les données sont notamment *géostatistiques* si D est un sous-ensemble fixe de \mathbb{R}^d contenant un rectangle d -dimensionnel de volume non nul, et les données sont *de réseau* (régulier ou irrégulier) si D est un ensemble comptable de points de \mathbb{R}^d .

Nous entendons apporter notre attention aux données de réseau fini. Le champ spatial peut être appréhendé dans ce contexte comme un vecteur fini \mathbf{X} de dimension n :

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad X_i \in \mathcal{X} \quad \forall i, \quad (1.2)$$

dont le vecteur $\mathbf{x} = (x_1, x_2, \dots, x_n)$ dénote sa réalisation. On peut alors associer au champ spatial considéré une matrice de variance-covariance Σ de taille finie définie comme

$$\Sigma := E_{\mathbf{X}}[\mathbf{X}\mathbf{X}'], \quad (1.3)$$

où $E_{\mathbf{X}}[\cdot]$ est notre forme de noter l'espérance mathématique sur l'univers de \mathbf{X} . Elle donne une représentation fondamentale théorique de l'autocorrélation spatiale du modèle spatial. Le champ spatial qui maximalise l'entropie de Shannon (voir annexe) étant donné Σ possède une distribution gaussienne multivariée dont la densité de probabilité $f(\mathbf{x})$ est la suivante (Cover & Thomas, 1991, p. 270) :

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}\right]. \quad (1.4)$$

Ce modèle englobe un certain nombre de modèles particuliers employés dans la littérature (dont nous citerons plus loin trois familles importantes).

1.2.2 Des modèles particuliers

Pour un physicien, un modèle spatial doit avoir une composante dynamique causale. Cette composante est ce qui distingue un modèle d'un autre, même si la distribution de probabilité qui en résulte est la même. La détermination de cette composante causale est cependant moins l'affaire du statisticien pour qui finalement un modèle ne diffère d'un autre que s'il diffère en distribution de probabilité. En nous considérant a priori comme ignorant du processus physique à l'origine des données, le modèle de champ spatial est choisi comme étant soit une hypothèse de distribution de probabilité convenant le mieux aux données parmi un groupe d'autres hypothèses, soit une distribution de probabilité paramétrique ajustée aux données. Dans les deux cas, il s'agit d'une sélection de modèle.

Une méthode communément employée consiste à opposer une hypothèse contre une autre et d'en sélectionner une par le résultat d'un *test statistique*. Une autre méthode communément utilisée, lorsque les modèles sont paramétriques, consiste à sélectionner le modèle pour lequel les données sont les plus vraisemblables : le critère employé est appelé le critère du *maximum de vraisemblance*.

De nombreux modèles peuvent être associés aux mêmes données, selon l'information préalable que l'on intègre dans les hypothèses de travail. En ayant pour seule connaissance préalable la matrice de variance-covariance du champ spatial, nous entendons considérer, pour des données de réseau, des modèles de champ dont la densité de probabilité est de la forme (1.4). Dans ce registre, les modèles à relever dans le domaine de la géographie et de l'économétrie spatiale sont le modèle autorégressif simultané, en anglais *simultaneous autoregressive model* (SAR), de la forme

$$X_i = \sum_{j \neq i} g_{ij} X_j + \epsilon_i \quad (1.5)$$

et le modèle à moyenne coulissante, en anglais *moving average model* (MA), de la forme

$$X_i = \sum_{j \neq i} g_{ij} \epsilon_j + \epsilon_i, \quad (1.6)$$

pour lesquels g_{ij} est un scalaire et ϵ_i est une variable gaussienne univariée centrée de variance σ_i^2 , qui est indépendante des variables ϵ_j ($j \neq i$). Les scalaires g_{ij} sont les éléments d'une matrice \mathbf{G} donnant une mesure locale de l'interaction spatiale entre les variables X_i et X_j pour le modèle SAR ou entre les variables ϵ_i et ϵ_j pour le modèle MA. Pour ces deux modèles, la densité de probabilité du champ est (1.4), avec pour matrice de variance-covariance

$$\text{MA : } \Sigma = (\mathbf{I} + \mathbf{G})\Sigma_\epsilon(\mathbf{I} + \mathbf{G})', \quad (1.7)$$

$$\text{SAR : } \Sigma = ((\mathbf{I} - \mathbf{G})'\Sigma_\epsilon(\mathbf{I} - \mathbf{G}))^{-1}, \quad (1.8)$$

où Σ_ϵ est la matrice de variance-covariance (diagonale) du vecteur $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$:

$$\Sigma_\epsilon := E_\epsilon[\epsilon\epsilon']. \quad (1.9)$$

Un modèle SAR peut être ramené à un modèle MA, et réciproquement, sous certaines conditions algébriques (voir Whittle, 1954). Les distributions de probabilité du champ spatial issues de ces deux modèles sont équivalentes si et seulement si les matrices de variance-covariance respectives sont identiques. Pour la comparaison entre ces deux modèles, on peut se référer à Haining (1978b).

Un autre modèle mérite également d'être mentionné. Il s'agit du modèle autorégressif conditionnel, ou *conditionally autoregressive model* (CAR) en anglais, introduit par Besag (1974), qui exprime l'espérance conditionnelle d'une variable X_i du champ étant donnée la valeur des variables voisines X_j ($j \neq i$) comme la somme pondérée des valeurs des variables voisines, quelle que soit la variable X_i . Plus précisément,

$$E_{X_i}[X_i|V_i] = \sum_{j \neq i} g_{ij} X_j, \quad (1.10)$$

où V_i est l'ensemble des variables du champ sans la variable X_i et g_{ij} est un scalaire donnant une mesure de l'interaction spatiale entre la variable X_j et l'espérance conditionnelle de X_i . Pour ce modèle, la distribution de probabilité du champ est gaussienne multivariée si chaque variable du champ conditionnellement aux autres variables du champ est gaussienne. Il apparaît que n'importe quel modèle SAR gaussien peut être reformulé en un modèle CAR, mais pas nécessairement l'inverse (voir p. ex. Cressie, 1993, p. 409).

1.2.3 Le digraphe des interactions dans les modèles SAR

Les éléments non nuls de la matrice \mathbf{G} des modèles SAR peuvent être vus comme la valuation d'un digraphe si l'on prend comme convention que les éléments nuls de \mathbf{G} remplacent les valeurs indéterminées associées aux arcs manquants du digraphe. Nous appelons ce digraphe ou ce graphe un **digraphe ou graphe des**

interactions. Un choix du digraphe des interactions, dans le cas où il est possible de l'effectuer, est fondamentalement un acte de sélection de structure du modèle spatial. Pour certains cas, le choix est argumenté, mais dans d'autres cas, le choix n'est qu'une hypothèse. Le graphe des interactions, indépendamment de la valuation qu'on lui affecte, est ce qui modélise la connexité¹ du champ spatial avec la caractéristique suivante : s'il n'existe pas de chemin (au sens d'un graphe) entre le point dont est associée X_i et le point dont est associée X_j alors ces deux variables sont indépendantes l'une de l'autre conditionnellement à \mathbf{G} . Il est possible par contre que deux variables ne soient pas indépendantes l'une de l'autre conditionnellement à \mathbf{G} sans être reliée (directement) par un arc du graphe des interactions. Autrement dit, une valeur g_{ij} nulle n'est pas nécessairement équivalente à une corrélation nulle entre X_i et X_j . Ce fait distingue la nature de \mathbf{G} de celle de la matrice de variance-covariance Σ et présage des difficultés à relier directement une estimation de la covariance du champ avec une estimation des éléments de g_{ij} . Pour cette raison nous préférons voir en g_{ij} une mesure d'interaction plutôt qu'une mesure de dépendance ou d'association entre variables.

On retrouve pleinement dans \mathbf{G} l'information de connexité, sans distorsions, avec la valuation spéciale de digraphe qui consiste à donner la valeur 1 uniformément à tous ses arcs. Dans ce cas, on obtient une représentation binaire du digraphe, appelée la **matrice d'adjacence** du digraphe, dont on interprète les éléments à la fois comme quantitatifs discrets et comme catégoriels. Son rôle dans les modèles est purement structurel. On le vérifie par exemple avec la somme (1.5) qui prend la forme

$$X_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ik}X_k + \epsilon_i, \quad (1.11)$$

où a_{ij} est le terme général de la matrice d'adjacence du digraphe. L'expression (1.11) confère à la variable X_i un caractère cumulatif des valeurs des variables voisines (à un terme d'erreur près). A ce modèle est opposé fréquemment un modèle qui confère à X_i un caractère de *moyenne arithmétique* des variables voisines (à un terme d'erreur près) :

$$X_i = \frac{a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ik}X_k}{a_{i1} + a_{i2} + \dots + a_{ik}} + \epsilon_i. \quad (1.12)$$

Certains auteurs décrivent \mathbf{G} comme la forme "normalisée" en les lignes de la matrice d'adjacence du digraphe des interactions².

La normalisation par colonne de la matrice d'adjacence du digraphe est une autre façon d'entrer l'information de connexité dans les modèles précédemment

1. La connexité correspond à *l'état ou le caractère de ce qui est connexe*. Par ce terme, nous faisons référence aux interactions entre variables. Le terme *connectivité* (anglicisme) est employé dans le même sens avec un aspect quantitatif : la connectivité locale d'une variable est le nombre d'interactions provenant de celle-ci ; la connectivité globale est le rapport du nombre réel d'interactions sur le nombre maximum d'interactions possibles du champ spatial.

2. Bien que formellement applicable, le terme "normalisation" est un terme malheureux en ces circonstances car il fait perdre de vue la représentation catégorielle de la matrice d'adjacence

décrits. Pour le reste, on retient des diverses applications exposées dans la littérature que la matrice \mathbf{G} est un élément sensible (c.-à-d. soumis à critiques ou discussions) de la modélisation. Nous concevons que cela soit particulièrement le cas pour les champs spatiaux dont les variables sont associées à des unités spatiales pour lesquelles ne s'appliquent pas de manière évidente des règles de contiguïté (ou d'adjacence) spatiale et dont l'ordonnance spatiale est irrégulière. Il est en effet fort probable, en ces cas, que les écarts euclidiens entre les positions des variables participent à l'explication de l'autorégressivité spatiale (en rappelant ici la première loi de Tobler).

1.3 La voie des indices

1.3.1 Le produit croisé

Une autre manière de révéler l'ancrage à la fois thématique et spatial de l'autocorrélation spatiale, sans passer par la recherche d'un modèle spatial, peut être appliquée via des statistiques univariées idoines que l'on nomme des indices. On ne cesse d'être inventif en cette matière et il est vain d'en donner ici une liste exhaustive. Un nombre d'entre eux sont des cas particuliers de la statistique du *produit croisé* de Hubert et al. (1981) qui est de la forme

$$\Gamma_{\mathbf{w}}(\mathbf{X}) := \sum_{i=1}^n \sum_{j=1}^n w_{ij} h(X_i, X_j), \quad (1.13)$$

où w_{ij} sont des scalaires, appelés des *poids spatiaux*, traduisant le degré d'association ou de connexion spatiale entre les variables X_i et X_j et h est une fonction de X_i et X_j donnant une mesure de leur association dans une dimension différente de la dimension spatiale.

Parmi les indices issus du modèle (1.13), on relève l'indice I proposé par Moran (1950), qui, pour une réalisation $\mathbf{x} = (x_1, x_2, \dots, x_n)$ du champ spatial, est de la forme

$$I = \frac{n}{2 \sum_{ij} w_{ij}} \frac{\sum_{ij} w_{ij} z_i z_j}{\sum_i z_i^2} \quad \text{avec } z := (x_i - \bar{x}), \quad \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i \quad (1.14)$$

et l'indice c proposé par Geary (1954) qui est de la forme

$$c = \frac{(n-1) \sum_{ij} w_{ij} (z_i - z_j)^2}{4 \sum_{ij} w_{ij} \sum_i z_i^2} \quad \text{avec } z := (x_i - \bar{x}), \quad \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i, \quad (1.15)$$

où $w_{ij} = 1$ si les variables X_i et X_j sont distinctes ($i \neq j$) et si elles sont considérées comme adjacentes, $w_{ij} = 0$ sinon. L'indice I donne un rapport entre la covariance locale et la variance totale du champ spatial estimées sur l'espace, alors que l'indice c donne un rapport entre un indicateur de dispersion locale et la variance totale

des variables estimés sur l'espace. L'indice c est inspiré de la statistique d de Durbin & Watson (1950, 1951, 1971) et du ratio de von Neumann (1941) des séries temporelles.

Il existe une variété d'emplois possibles des indices. Un même indice peut être employé pour comparer plusieurs champs entre eux. Il peut être également employé pour explorer la structure d'autocorrélation en faisant varier les valeurs de w_{ij} . Notamment, en élargissant progressivement le cercle de voisinage de chaque variable, on obtient un *corrélogramme* spatial avec (1.14) ou un *variogramme* spatial avec (1.15) (en lien direct avec le corrélogramme et le variogramme que l'on définit habituellement en géostatistique). Enfin, en relation avec les modèles spatiaux, les indices peuvent servir comme statistique de test pour tester l'indépendance spatiale du champ, ou pour inférer les paramètres d'un modèle paramétrique.

1.3.2 Problématique : la matrice des poids spatiaux

Les éléments w_{ij} de (1.13) (et donc de (1.14) et (1.15)) forment les éléments d'une matrice \mathbf{W} que l'on appelle *matrice de poids spatiaux*. Elle peut être une matrice d'adjacence, représentant un graphe de voisinage, comme c'est le cas pour les définitions originelles du c de Geary ou du I de Moran, mais elle peut également posséder une plus grande variété de valeurs. La forme de \mathbf{W} dépend du cas étudié et des données qui sont disponibles. Elle est généralement spécifiée a priori comme une fonction de l'écart spatial entre les variables, mais il est généralement difficile de présenter des arguments en faveur d'une pondération particulière. Cette fonction est généralement décroissante avec la distance. Elle peut tenir compte d'autres paramètres comme par exemple la longueur de la frontière commune entre les unités spatiales lorsque les variables du champ sont associées à des unités d'une partition de l'espace. On peut appliquer une pondération spatiale plus subtile si, comme Gatrell (1979), l'on introduit le concept d'autocorrélation spatiale sur d'autres espaces que l'espace euclidien³.

Dans un contexte où l'indice est employé pour tester l'hypothèse d'indépendance spatiale, le choix arbitraire de \mathbf{W} soulève quelques problèmes. Notamment, que conclure quant à l'autocorrélation spatiale si un indice est significatif pour une matrice de poids spatiaux donnée et qu'il ne l'est pas pour une autre? Ou encore, que doit-on décider si deux matrices de poids spatiaux concurrentes aboutissent toutes les deux au rejet de l'hypothèse d'indépendance spatiale?

Obtenir la pondération a posteriori, en fonction des données, apparaît comme une solution évitant de faire ce choix. Dans ce registre nous pouvons mentionner la méthode qui consiste à choisir la pondération qui maximise une statistique de produit croisé donnée, dont spécialement le I de Moran (Kooijman, 1976, Boots & Dufournaud, 1994). Cette méthode, en revanche, pourrait ne pas être satisfai-

3. Suite à une étude de corrélogrammes sur des données transformées par technique de *multi-dimensional scaling*, l'auteur conclut à l'existence de structures autocorrélatives dans un espace dit *relatif* obtenu par transformation de la position des variables du champ spatial d'origine pour satisfaire par rapport à cet espace relatif la loi énoncée par Tobler.

sante du point de vue de l'efficacité de l'inférence statistique de modèles de champ spatiaux.

Une approche différente consiste à recourir à des effets qui se *produisent* entre unités spatiales tels que des flux de biens, de personnes, etc. entre unités spatiales (cf. par exemple Bavaud, 1998) pour calculer les poids spatiaux. En supposant détenir de telles données, la méthode oblige à concilier la forme d'interaction du champ qui est révélée par les flux avec le graphe des interactions que l'on définit en section 1.2.3.

1.4 Problème traité dans la thèse

1.4.1 Modèle de Whittle

Nous portons notre attention au modèle autorégressif simultané, et plus particulièrement au modèle de Whittle markovien *classique*. Il est de la forme autorégressive (1.5) avec

$$g_{ij} = \rho m_{ij} a_{ij}, \quad (1.16)$$

où ρ est un coefficient prenant ses valeurs dans l'intervalle ouvert $(-1, 1)$, a_{ij} est le terme général de la matrice d'adjacence \mathbf{A} du digraphe non stochastique des interactions et m_{ij} est le terme général d'une matrice \mathbf{M} satisfaisant les conditions suivantes :

$$m_{ij} \geq 0 \quad \sum_j m_{ij} \in \{0, 1\} \quad \forall i. \quad (1.17)$$

Avec les conditions (1.17), la matrice \mathbf{M} est ce que nous appelons une *matrice de Markov généralisée* dans le sens qu'elle est une matrice de Markov classique si le digraphe des interactions est fortement connexe⁴. Pour aboutir à un choix de matrice unique, nous fixons les conditions supplémentaires suivantes : (a) le terme m_{ij} est nul si et seulement si l'adjacence a_{ij} est nulle et (b) les éléments non nuls d'une même ligne de \mathbf{M} sont d'égales valeurs. La première condition a été introduite pour que le terme général m_{ij} représente une valuation non nulle des arcs du digraphe des interactions et donc que $g_{ij} = \rho m_{ij}$. La seconde condition peut découler d'un principe *ad hoc* de maximum d'entropie : si une variable est adjacente à k variables selon le graphe des interactions alors, en l'absence de toute autre information, on admet que la valuation des arcs concernés est a priori la même pour tous. En réalité, la seconde condition permet d'exprimer chaque variable X_i du champ comme la moyenne arithmétique des variables qui lui sont connectées, à une erreur ϵ_i près, amortie d'un coefficient ρ . Ceci conduit à ce que le terme m_{ij} prenne les valeurs discrètes suivantes :

$$m_{ij} = \begin{cases} \frac{1}{k} & \text{si } \sum_j a_{ij} = k > 0 \\ 0 & \text{sinon} \end{cases}. \quad (1.18)$$

4. Définition : un graphe orienté est fortement connexe si pour tout couple de sommets u, v il existe un chemin reliant u à v .

Ce choix établit un lien biunivoque entre la matrice \mathbf{M} et le graphe qui le sous-tend. En particulier, si $m_{ij} > 0$, alors on affirme qu'il y a adjacence entre la variable X_i et un ensemble de $\frac{1}{m_{ij}}$ autres variables dont la variable X_j en fait partie.

Nous désignons ρ comme étant le **coefficient d'autorégression** du modèle spatial autorégressif et nous désignons \mathbf{M} comme étant la **matrice des interactions** du modèle spatial autorégressif. Lorsque ρ tend vers 0, le modèle tend vers un modèle d'indépendance spatiale. Le modèle ainsi obtenu est appelé un modèle autorégressif d'ordre 1 ou **modèle de Whittle**, du nom de celui qui l'a proposé (Whittle, 1954). La matrice \mathbf{M} est *classiquement* un paramètre du modèle.

Pour la suite, il est supposé que les éléments ϵ_i de (1.5) sont gaussiens homos-cédastiques.

Remarque Il convient de souligner qu'en dérogeant à l'une des conditions pour le choix de la matrice des interactions, il aurait été possible de choisir un système de valuation continue du graphe des interactions tel qu'en général $m_{ij} \neq m_{ik}$. En revanche, il aurait été plus difficile d'interpréter ce système de valuation en lien avec un concept de connexité du champ. Aussi, le choix de la valuation (1.18) apparaît comme le choix le plus simple pour permettre de dégager, au travers du modèle de Whittle, un *modèle de connexité du champ*.

1.4.2 Applications du modèle de Whittle classique

Le modèle de Whittle a été initialement employé pour modéliser de manière endogène⁵ et dans une approche de grands réseaux la dépendance des variables observables telles que la production d'un champ de coton ou la production d'un champ d'orangers (Whittle, 1954). Il a été développé dans ce contexte une méthode "spectrale" d'estimation du coefficient d'autorégression. Plus tard, Ord (1975) montre que le principe du maximum de vraisemblance peut être employé pour estimer ce coefficient, plus apte pour traiter des réseaux irréguliers, à la condition de vaincre les difficultés computationnelles de la méthode. Depuis, le modèle de Whittle a été appliqué dans divers domaines de la recherche, en particulier dans le contexte très répandu de régressions linéaires. La technique de la régression linéaire se base sur l'estimation des paramètres β_i d'un modèle linéaire de la forme (voir annexe de ce chapitre) :

$$\mathbf{Y} = \sum_{i=0}^k \beta_i \mathbf{Z}_i + \boldsymbol{\eta}, \quad (1.19)$$

où \mathbf{Y} est un champ spatial que l'on veut décrire par une combinaison linéaire de variable \mathbf{Z}_i , les scalaires β_i étant les coefficients de cette combinaison et $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots)$ étant le terme d'erreur représentant la part stochastique du modèle.

5. Ce terme qualifie ce qui est *interne* au système considéré. Le modèle de Whittle est endogène parce que les variables sont décrites en relation avec elles-mêmes. Le terme opposé est "exogène". Les variables *exogènes* sont extérieures au système. La frontière entre endogène et exogène dépend de la frontière que fixe le chercheur pour déterminer son système.

Une des conditions généralement admises du modèle linéaire est que les éléments du terme d'erreur $\boldsymbol{\eta}$ sont homoscédastiques et statistiquement indépendants les uns des autres. Lorsque cette dernière condition n'est pas satisfaite, on peut alors faire intervenir le modèle de Whittle pour le terme d'erreur $\boldsymbol{\eta}$, c'est-à-dire considérer

$$\eta_i = \sum_{j \neq i} \rho m_{ij} \eta_j + \epsilon_i \quad (1.20)$$

soit pour vérifier la dépendance des résidus de la régression linéaire⁶, soit pour modéliser à l'intérieur du modèle linéaire (1.19) la corrélation des éléments du terme d'erreur $\boldsymbol{\eta}$ entre eux.

Le modèle linéaire (1.19) est souvent appelé un *modèle de perturbations spatiales* lorsque la variable $\boldsymbol{\eta}$ est autocorrélée selon un modèle de Whittle endogène. Comme montré par Ord (1975) et Doreian (1980, 1981), le principe de maximum de vraisemblance est applicable pour estimer les paramètres, mais à la condition, une fois encore, de pouvoir vaincre les difficultés computationnelles de la méthode. Celle-ci donne un estimateur des coefficients de régression qui ne coïncide généralement pas avec l'estimateur des moindres carrés classique dont on sait qu'il est non biaisé mais peut être moins efficace (voir étude de Dow et al., 1982). On trouve des exemples d'application du modèle et de la méthode aux domaines de la sociologie avec Doreian (1980, 1981), Loftin & Ward (1983), Pace & Barry (1997) et Leenders (2002), au domaine de la télédétection en hydrologie avec Eldiery & al. (2005), au domaine de l'écologie ou biogéographie avec Dark (2004), Tognelli & Kelt (2004), Bullock & Burkhart (2005), Kühn (2007) et Kissling & Carl (2008), au domaine du prix de l'immobilier ou du logement (économétrie spatiale) avec Pace (1997), Pace & Barry (1997b), Won Kim & al. (2003), Militino & al. (2004) et au domaine des flux commerciaux (économétrie spatiale) avec Porojan (2001). Les résultats tirés de ces exemples montrent l'influence parfois importante de l'autocorrélation spatiale sur l'estimation des coefficients de régression, sous condition que la matrice des interactions soit un objet donné a priori du modèle de perturbations spatiales.

Les exemples qui précèdent illustrent également qu'il est possible d'estimer le paramètre ρ par le principe de maximum de vraisemblance, plus efficacement que par le principe des moindres carrés (Ord, 1975). Dans le cas spécial où les paramètres β_i sont connus, le modèle (1.19), avec (1.20), se ramène commodément à un modèle de Whittle endogène pour la variable $\boldsymbol{\eta}$ (si elle est gaussienne en ses éléments).

Il est à relever que si on s'attend à ce qu'il existe un modèle de perturbations spatiales avec des termes de $\boldsymbol{\eta}$ spatialement indépendants pour décrire des données d'un champ, alors une estimation ou inférence d'un coefficient d'autorégression ρ non nul de (1.20) pour un choix donné de variables explicatives signifierait que le modèle n'a pas été *correctement spécifié*. Ce raisonnement amène certains auteurs

6. Malgré la condition d'indépendance des éléments du terme d'erreur $\boldsymbol{\eta}$, il est possible que les résidus soient autocorrélés. Cependant, cette corrélation disparaît asymptotiquement (voir Annexe B.2)

à considérer l'autocorrélation spatiale comme un critère de sélection de modèle. Partant, on s'évertue dans certains cas à choisir un jeu de variables explicatives minimisant ⁷ ρ .

1.4.3 Vers plus de complexité

Dans un contexte appliquant un modèle spatial autorégressif en l'absence d'une information précise sur la topologie des interactions entre variables du champ, il apparaît naturel de postuler que le terme g_{ij} du modèle autorégressif (1.5) est nul si et seulement si la position spatiale \mathbf{r}_i de la variable X_i est suffisamment proche de la position spatiale \mathbf{r}_j de la variable X_j . Mathématiquement, le postulat s'exprime ainsi :

$$g_{ij} \neq 0 \Leftrightarrow \mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) \in (0, \mathfrak{d}_l], \quad (1.21)$$

où $\mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j)$ est une *fonction de dissimilarité spatiale* (p. ex. la distance euclidienne) entre \mathbf{r}_i et \mathbf{r}_j et dont la valeur doit être entre 0 et un seuil \mathfrak{d}_l (à fixer selon le problème traité) pour que g_{ij} ne soit pas nul. Cette définition englobe également les matrices \mathbf{G} représentant un graphe de contiguïtés spatiales⁸. L'hypothèse (1.21) ne tient pas compte de la situation plus complexe, plus générale et pourtant plus réaliste où une proximité spatiale mutuelle dépassant un certain seuil est une condition nécessaire mais non suffisante pour qu'il existe une interaction.

Une manière d'appréhender le problème de la complexité de la connexité du champ consiste à probabiliser la matrice \mathbf{G} : la détermination du graphe se fait non pas en terme déterministe, mais en terme probabiliste. Ainsi, s'il n'est pas possible de vérifier (1.21), nous admettons par contre que le terme g_{ij} a une probabilité non nulle de différer de 0 si et seulement si la position spatiale \mathbf{r}_i de la variable X_i est suffisamment proche de la position spatiale \mathbf{r}_j de la variable X_j . Autrement dit, il est postulé la relation suivante :

$$\Pr[g_{ij} \neq 0] > 0 \Leftrightarrow \mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) \in (0, \mathfrak{d}_l], \quad (1.22)$$

où $\Pr[\]$ est une mesure de probabilité de l'événement indiqué dans $[\]$. La relation (1.22) permet une variante "connectiviste" de l'énoncé de Tobler : *toutes choses sont probablement reliées aux autres, mais les choses plus proches sont reliées en plus grande probabilité que les choses lointaines*⁹.

Il reste à préciser dans ce contexte si les éléments de \mathbf{G} sont vus comme des paramètres inconnus dans un univers de possibilités (chacune ayant une probabilité

7. On s'expose au risque de surparamétrisation du modèle en minimisant une autocorrélation spatiale apparente qui existe en réalité.

8. Il suffit de poser que $\mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) = 2$ si les unités spatiales étendues associées à \mathbf{r}_i et \mathbf{r}_j sont disjointes et non contiguës, $\mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) = 1$ si ces mêmes unités sont disjointes et contiguës et $\mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) = 0$ si ces mêmes unités spatiales coïncident. En posant $\mathfrak{d}_l = 1$, on obtient un graphe de contiguïté sans arêtes réflexives.

9. Il est à relever que cette formule n'affirme pas et ne contredit pas l'hypothèse que les choses plus proches sont plus fortement en interaction que les choses lointaines (comme entendu par Tobler)

de se produire) ou s'ils sont vus comme des variables du modèle spatial. La différence tient dans le fait qu'un élément considéré comme un paramètre du modèle n'introduit pas d'incertitude dans les réalisations du modèle spatial, contrairement à un élément considéré comme une variable du modèle. En cas d'incertitude sur \mathbf{G} , ces deux points de vue engendrent *deux paradigmes distincts* du traitement de \mathbf{G} . Le premier paradigme, désigné comme *paradigme bayésien*, considère \mathbf{G} comme un paramètre inconnu et confère à celui-ci une probabilité non informative sur le champ spatial avec la contrainte (1.22) traduisant l'ignorance a priori sur l'état de connexité du graphe des interactions. Le second paradigme, désigné comme *paradigme fréquentiste* (pour le distinguer du premier), associe à \mathbf{G} un digraphe des interactions qui est de nature à changer d'une réalisation à une autre du modèle, avec la contrainte (1.22). **Pour nos travaux, nous convenons du paradigme fréquentiste.** Nous effectuerons en revanche quelques parallèles dans le paradigme bayésien au long de notre étude, auquel cas nous le signalerons clairement.

1.4.4 Le modèle de Whittle markovien probabilisé

Intéressés à traiter avec la complexité des interactions entre variables, mais en étant toujours parcimonieux dans le choix du modèle, nous nous attachons au **modèle de Whittle markovien probabilisé**, de la forme

$$X_i = \sum_{j \neq i} \rho m_{ij} X_j + \epsilon_i, \quad (1.23)$$

où m_{ij} est défini par (1.17) et (1.18) avec une matrice d'adjacence du digraphe des interactions qui est *aléatoire*, de terme général a_{ij} tel que sa probabilité de valoir 1 vaut $1 - p$ si et seulement si la dissimilarité spatiale $\mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j)$ entre les positions de variables \mathbf{r}_i et \mathbf{r}_j est non nulle et inférieure à un seuil \mathfrak{d}_l :

$$\Pr[a_{ij} = 1] = 1 - p \quad \Leftrightarrow \quad \mathfrak{d}(\mathbf{r}_i, \mathbf{r}_j) \in (0, \mathfrak{d}_l], \quad (1.24)$$

où p est un paramètre dont la valeur reste à fixer. Nous convenons également que le digraphe des interactions se ramène à un graphe, c'est-à-dire :

$$a_{ij} = a_{ji}. \quad (1.25)$$

Il en résulte un champ spatial dont la distribution de probabilité dépend du graphe des interactions (induit par le choix de la fonction de dissimilarité), ainsi que de p , de ρ et de σ^2 . Le modèle de Whittle est ici employé de manière non classique étant donné qu'il possède deux sources de stochasticité : celle qui provient des variables ϵ_i (source classique) et celle qui provient des interactions spatiales.

1.4.5 Matrice des interactions

Le modèle de Whittle explique la valeur de chaque variable simultanément en fonction de la valeur des variables voisines par interaction. Dans ce cadre, le

coefficient d'autorégression ρ est ce qui module l'effet de cette interaction. Pour certaines applications, la matrice est spécifiée par avance pour refléter les propriétés relationnelles objectives entre les sites. Dans d'autres applications, la matrice des interactions est un modèle ou une hypothèse sur les interactions. Il ressort donc que le modèle de Whittle classique convient pour toutes formes de relation.

Dans beaucoup d'études, la matrice des interactions considérée est une matrice de Markov classique stationnaire dans le temps, dite *homogène*. Cela implique le point de vue suivant :

- (a) toutes les unités spatiales étudiées sont en interaction avec au moins une autre unité spatiale, quitte à ce qu'une sélection soit faite sur les unités spatiale avant de les intégrer dans le modèle ;
- (b) les unités spatiales et les interactions sont considérées comme invariantes du temps ;
- (c) chaque variable du champ est une moyenne pondérée ou arithmétique (selon les cas) d'autres variables du champ, à un terme d'erreur près.

L'homogénéité de la matrice de Markov découle de l'hypothèse que les unités spatiales et les interactions sont invariantes du temps. Cette invariance n'est pas nécessairement pertinente dans le contexte de relations autorégressives. Pour le comprendre, considérons par exemple le cas des matrices de flux migratoires de personnes entre des unités spatiales. Pour certains auteurs un tel flux est une mesure d'interactions entre unités spatiales (par ex. Bavaud, 1998), ce qui ne se vérifie qu'avec les attributs affectés des unités spatiales. En supposant connaître ces attributs, il reste à obtenir une matrice de Markov à partir de la matrice des flux. Comme simple solution, on normalise les lignes de la matrice des flux, ce qui fait que le flux d'une unité i vers une unité j se transforme en une probabilité de migration d'une unité i vers une unité j . La matrice de Markov ainsi obtenue est ensuite employée dans un modèle de Whittle markovien classique. Dans le cadre d'un processus de migration invariable dans le temps, on sait que plus le temps de mesure est long, plus précise est la mesure de cette probabilité : par la loi des grands nombres, les mesures convergent vers la matrice de Markov homogène du processus. A l'inverse, plus le temps de mesure est court, plus cette mesure de probabilité apparaît comme relativement fluctuante. Selon l'échelle de temps choisie, il apparaît alors pertinent d'expliquer la variabilité du champ spatial par une forme non homogène de la matrice markovienne des interactions, ce qu'en principe un modèle de Whittle ne permet pas de faire.

La problématique de l'homogénéité de la matrice des interactions s'applique aussi pour les matrices de Markov se basant sur la normalisation de matrices d'adjacence de digraphes. Si cette adjacence représente la contiguïté spatiale, alors la matrice des interactions dépend entièrement de la structure spatiale et de la sélection des unités spatiales pour le modèle. Or, la connexité due à la structure des unités spatiale n'est pas nécessairement la connexité qui prévaut pour les variables du champ y localisées. En écologie on est sensible à cette problématique en différenciant *landscape connectivity* et *landscape structure*, ou encore, *functional*

measure of connectivity et *structural measure of connectivity* (With & al., 1997; Goodwin, 2003). Le problème que l'on admet se produire est donc que les unités spatiales puissent entrer de manière fluctuante (dans le temps) dans le modèle autorégressif sans que la structure spatiale ne change. A l'exemple précédent des flux migratoires, si l'on mesure les flux sur des temps suffisamment courts, certains flux migratoires entre deux unités spatiales contiguës ne seront pas constatés : l'échelle de temps de mesure choisie fait que la contiguïté entre deux unités spatiales n'engendrent pas nécessairement un flux. Or, on doit admettre que la Nature n'attend pas nécessairement que deux unités spatiales entrent en interaction pour que les effets autorégressifs des autres unités se fassent sentir partout ailleurs et sur les unités spatiales en question. Une fois encore, il apparaît pertinent de considérer une matrice des interactions markovienne non homogène.

Le modèle de Whittle markovien probabilisé (1.23) est notre manière de modéliser les effets de l'inhomogénéité de la matrice de Markov du processus stochastique *dans un modèle purement spatial*, avec une vue (représentation) purement connectiviste de cette inhomogénéité (selon principe exposé en section 1.2.3).

1.4.6 Perspectives pour le modèle de Whittle probabilisé

La forme probabilisée du modèle de Whittle se contente de décrire les interactions via un paramètre p . Ceci rend le modèle approprié pour des champs spatiaux dont il est vain de connaître la configuration exacte des interactions (notamment dans le paradigme fréquentiste) et dont un concept de contiguïté entre les unités spatiales est applicable et pertinente. A notre connaissance, *le modèle de Whittle markovien probabilisé est un modèle qui n'a jamais été exploité dans l'étude de l'autocorrélation spatiale*. Formellement, le modèle de Whittle probabilisé peut se substituer au modèle classique pour les applications des auteurs mentionnés précédemment, si l'inhomogénéité de la matrice markovienne des interactions que nous décrivons y est théoriquement admissible. On aura comme différence que l'autocorrélation spatiale mesurée avec le coefficient d'autorégression ρ s'explique en regard du paramètre p des interactions et de la structure spatiale des interactions (réseau) et non plus en regard du choix d'une matrice des interactions précise. Pour dire l'essentiel, le paramètre p permet de savoir avec quel degré de fidélité la structure spatiale "reflète" le réseau d'interactions, et réciproquement. Si p ne diffère pas significativement de 0, les deux types de réseaux peuvent être considérés comme identiques et déterministes (ce qui justifierait l'emploi d'une version classique plutôt que probabilisée du modèle de Whittle). Si p diffère significativement de 0, on conclut non seulement que le champ spatial possède un réseau d'interactions qui est non stationnaire dans le temps (sa matrice de Markov est non homogène), mais aussi que **le réseau d'interactions est moins connectif et plus fragmenté que le graphe de structure spatiale des interactions**. Nous soulignons l'intérêt scientifique de ces deux derniers points.

Par la loi des grands nombres, la quantité $1 - p$ est égale à la densité des interactions dans le cas asymptotique. Cette densité est ce que nous appelons aussi

la *connectivité* du réseau. Le paramètre p permet alors de déduire en bonne approximation la connectivité du réseau d'interactions lorsque le champ est de taille suffisamment grande. Elle peut être intéressante, par exemple en économie, lorsqu'il s'agit d'expliquer l'apparition d'effets globaux à partir d'interactions individuelles (Hors, 1995)¹⁰. Il est à relever que, dans divers domaines, la connectivité est ce que l'on cherche à contrôler, comme, par exemple, dans les réseaux informatiques, les réseaux de télécommunication, les réseaux électriques, les réseaux sociaux, etc.

La *théorie de la percolation* nous montre qu'un graphe aléatoire devient statistiquement plus fragmenté lorsque p augmente. Cette théorie donne par ailleurs des mesures de cette fragmentation en lien avec le paramètre p . Dans ce contexte, en corrélation avec la connectivité du champ, le modèle de Whittle probabilisé peut trouver des applications dans les études cherchant à expliquer certains phénomènes observés en fonction du fractionnement d'un réseau. Ces études peuvent porter, par exemple, sur l'émergence d'un groupe dominant dans les réseaux sociaux (Chen et al., 2007), sur la dispersion des prix dans un réseau économique de marchés locaux (Pajot, 2003), sur la mobilité des espèces en écologie (Gilbert, 1998). Il convient de relever qu'en écologie du paysage, l'hétérogénéité spatiale fait l'objet de diverses quantifications, des indices (Li & Reynolds, 1995), et que certaines d'entre elles intègrent des mesures du fractionnement observable de la connexité du paysage, dans la nécessité d'établir des relations entre la structure spatiale et des processus écologiques. On relève en particulier la recherche d'échelle spatiale critique de la fragmentation. Cette échelle constitue la distance typique qu'une espèce parcourt avant de ressentir la fragmentation de son environnement comme un obstacle (Keitt et al., 1997). Nous voyons donc que l'apport possible du modèle de Whittle probabilisé dans ces domaines est de pouvoir confronter, avec l'observation du champ, le fractionnement du graphe de structure avec le fractionnement du graphe des interactions généralement inobservables.

Ces perspectives d'applications fondées sur l'estimation ou l'inférence du paramètre p ne doivent cependant pas faire oublier qu'elle se basent tout d'abord sur la validité d'une relation bijective, indépendante de ρ , entre la matrice des interactions et un graphe. Ce modèle d'interactions est une hypothèse de travail que l'on n'arrive pas à défendre devant certains cas. Ensuite, même si cette relation est valable, elle peut être inconsistante en regard de la valeur estimée de ρ . On peut en effet envisager qu'une estimation très précise du paramètre p soit invalidée à cause d'une valeur de ρ trop faible du point de vue scientifique¹¹ ou à cause d'une valeur

10. Hors (1995) souligne la part intensive et connective d'un réseau d'interactions (par passage d'information entre individus) en donnant l'image suivante : "*Les interactions ont un effet global dès lors que deux individus du réseau interagissent de manière significative, se transmettant effectivement de l'information, même si ils sont éloignés. On comprend dès lors que l'apparition de ces phénomènes collectifs soit conditionnelle à deux points : il faut d'une part que la perte d'information lors d'une communication ne soit pas trop importante - autrement dit que l'intensité des interactions soit suffisamment forte - et d'autre part que la perte d'information causée par les transmissions successives soit compensée par la multiplicité des chemins reliant deux individus, c'est-à-dire que le réseau soit suffisamment dense.*"

11. On parle de l'*effet taille*, un concept bien connu en Statistiques.

de ρ non significativement différente de 0 du point de vue statistique. Il devient alors tout aussi important d'étudier l'estimation de ρ conditionnellement à la valeur de p pour connaître comment évolue la consistance du graphe des interactions.

1.4.7 Application à un réseau percolatif

Le modèle de Whittle probabilisé est théoriquement applicable à des données sur un réseau irrégulier comme à des données sur un réseau régulier. Nous nous attachons cependant principalement à des données de réseau régulier, et plus précisément de réseau carré, étant donné que le graphe de la connexité est plus évident (objectif) à déterminer. Un tel choix a également des applications potentielles, étant donné que de nos jours nous disposons de moyens de mesures automatiques générant une masse importante de données géoréférencées que des systèmes d'information géographiques peuvent agréger sous forme matricielle dans l'espace. A compter également que le modèle est applicable sur les images *raster*, prises par satellite ou radar, qui ne sont finalement que des niveaux d'intensité d'une trame régulière de pixels.

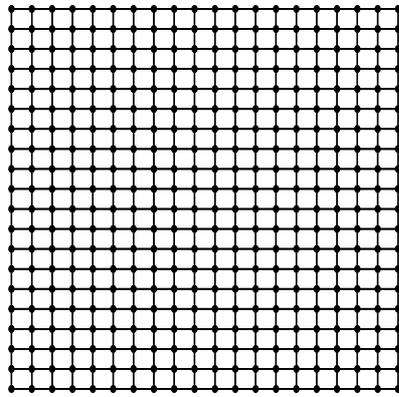
Dans le contexte d'un réseau carré, nous considérons que la relation (1.24) s'applique avec le seuil \mathfrak{d}_l égal à l'écart le plus petit entre les positions de variables distinctes du réseau, c'est-à-dire le *pas* du réseau. Il en résulte, sur un réseau carré parfait, qu'une variable interagit avec au plus 4 variables de son voisinage¹², comme illustré dans la figure 1.1 montrant des réalisations du graphe des interactions pour diverses valeurs de p .

En vertu de la relation (1.18), la complexité de la matrice des poids des interactions est entièrement induite par la complexité du graphe aléatoire des interactions. Cette complexité intervient dans la densité de probabilité du champ spatial. La figure 1.2 illustre ce fait par une réalisation du modèle de Whittle pour $p = 0.8$, $\rho = 0.7$ et $\sigma^2 = 1$ sur un réseau carré de sites. En ignorant a priori la valeur des paramètres du modèle, les deux exemples de la figure 1.2 nous amènent à nous interroger sur la manière de les déterminer, préfigurant ainsi la problématique traitée dans la présente thèse.

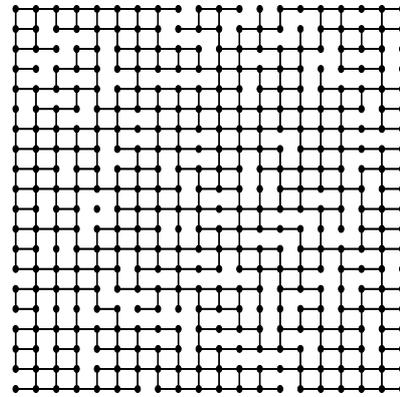
La question à laquelle on cherchera réponse est : **Quelle est la valeur de ρ (subsidiatement la valeur de p) du modèle spatial lorsqu'on observe une réalisation de \mathbf{X} mais que l'on ne parvient pas à observer conjointement la réalisation du graphe des interactions ?** Nous orientons notre recherche préférentiellement dans l'étude de l'estimation et l'inférence de ρ en fonction de p étant donné que ρ traduit traditionnellement l'autocorrélation spatiale.

Pour nous aider dans la recherche de la solution, nous partons du fait que les interactions entre les variables prennent place sur une structure de réseau d'interaction connue. Il conviendrait que ce réseau ait un fondement physique, mais il peut

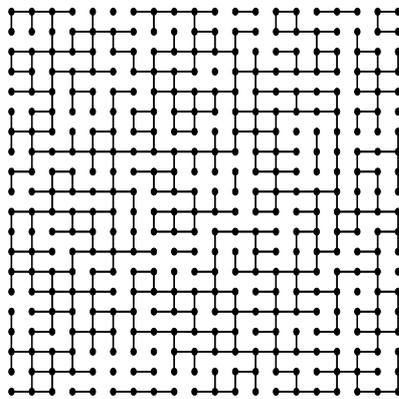
12. On rencontre souvent deux types de voisinage immédiat pour un site d'un réseau carré : le voisinage de von Neumann qui consiste en les 4 voisins les plus proches et le voisinage de Moore qui consiste en les 8 voisins les plus proches.



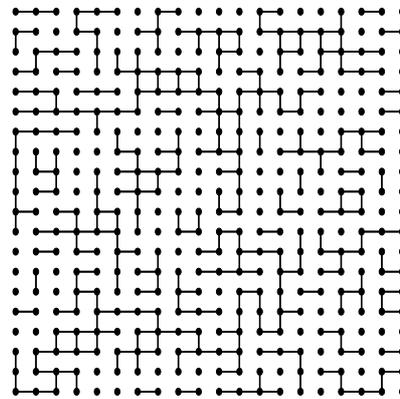
(a) $p = 0.0$



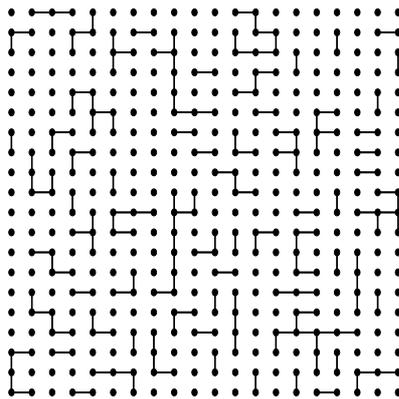
(b) $p = 0.2$



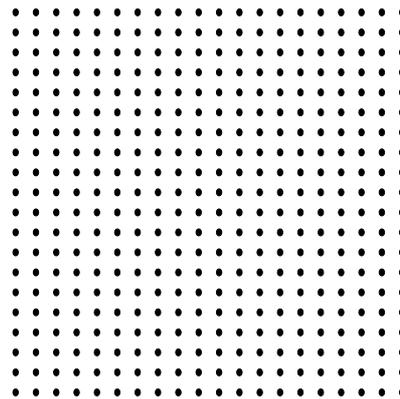
(c) $p = 0.4$



(d) $p = 0.6$



(e) $p = 0.8$



(f) $p = 1.0$

Fig. 1.1: Réalisations du graphe des interactions entre sites sur un réseau carré de 20×20 sites pour la valeur de p indiquée.

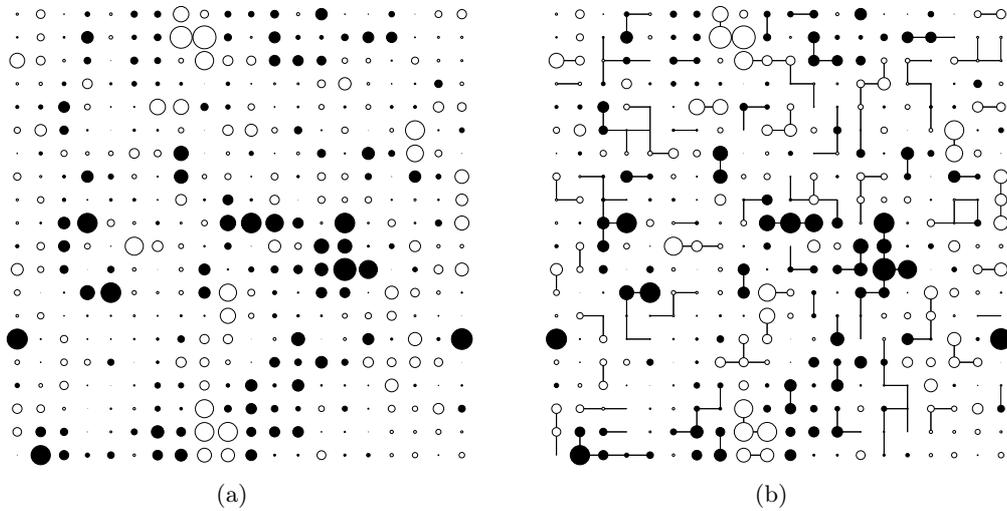


Fig. 1.2: Réalisation du couple de variables (\mathbf{X}, \mathbf{M}) avec (a) le graphe de \mathbf{M} invisible et (b) le graphe de \mathbf{M} visible. Les valeurs du champ \mathbf{X} (gaussien) sont représentés par des disques dont la couleur correspond à leur signe et le diamètre correspond proportionnellement à leur valeur absolue.

être finalement une représentation du voisinage spatial des positions des variables du champ (comme c'est déjà le cas pour des modèles de Whittle classiques).

1.4.8 Plan de la thèse

Nous organisons notre recherche de la manière suivante avec référence aux chapitres dans lesquels ses éléments seront développés.

Introduction au modèle. Nous dédions le **Chapitre 2** à la formulation détaillée des hypothèses de travail et à la connaissance approfondie du modèle de Whittle markovien probabilisé. Nous étudions sa différence par rapport à son cas particulier le modèle de Whittle classique, notamment par la variance et la corrélation du champ. Nous cherchons à être le plus illustratif possible.

L'information dans l'échantillon Nous dédions le chapitre **Chapitre 3**, à la détermination de l'information de Fisher que contient un échantillon sur ρ et la précision limite de son estimation (par un estimateur sans biais). Cette étude s'applique autant pour le modèle de Whittle probabilisé que pour le modèle de Whittle classique et permet de distinguer les différences entre ces deux modèles.

Estimation ponctuelle de ρ Nous dédions le **Chapitre 4** à des méthodes d'estimation ponctuelle du coefficient d'autorégression ρ du modèle de Whittle

markovien probabilisé. Nous étudions à cette occasion la méthode de maximum de vraisemblance contre laquelle nous opposons d'autres méthodes lorsque le graphe des interactions n'est pas observable.

Inférence des paramètres du modèle Nous dédions le **Chapitre 5** à l'inférence des paramètres p et ρ du modèle de Whittle markovien probabilisé par le test classique du I de Moran, et le test du I de Moran maximal.

Petit aide-mémoire des notations de matrices (pour la suite)

- A** représente (de manière générique) une matrice d'adjacence.
- S** est une matrice d'adjacence représentant le graphe de structure des interactions. Elle est un paramètre connu du modèle de Whittle probabilisé.
- G** représente la matrice générale des interactions. Pour la suite, nous la remplaçons par sa décomposition $\mathbf{G} = \rho\mathbf{M}$, où ρ est un scalaire.
- M** représente une matrice des interactions. Elle est une matrice de Markov généralisée. Elle est construite à partir de la matrice d'adjacence du graphe des interactions par normalisation de celle-ci. Elle est soit un paramètre du modèle de Whittle classique soit une variable du modèle de Whittle probabilisé.
- W** représente une matrice de poids spatiaux. Elle est employée dans des statistiques, par exemple un indice I de Moran.
- Σ matrice de variance-covariance du champ spatial.

Chapitre 2

Modèle de Whittle markovien probabilisé

2.1 Spécification du modèle

2.1.1 Hypothèses de travail

Le modèle de Whittle dans sa version classique ou probabilisée, pour un champ de n variables, peut être spécifié formellement sous la forme matricielle suivante

$$\mathbf{X} = \rho \mathbf{M} \mathbf{X} + \boldsymbol{\epsilon}, \quad (2.1)$$

où $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ est un vecteur aléatoire, ρ est un paramètre désigné comme le coefficient d'autorégression et \mathbf{M} est la matrice markovienne de taille $n \times n$ des poids des interactions, qui est un paramètre pour la forme classique du modèle et qui est une variable pour la forme probabilisée du modèle.

Le champ spatial \mathbf{X} qui est solution de (2.1) vaut

$$\mathbf{X} = (\mathbf{I} - \rho \mathbf{M})^{-1} \boldsymbol{\epsilon} \quad (2.2)$$

sous condition que la matrice $(\mathbf{I} - \rho \mathbf{M})$ soit inversible. Pour déterminer sa densité de probabilité, il est nécessaire de déterminer la probabilité de $\rho \mathbf{M}$. Nous formulons pour cela 4 hypothèses de travail

HT1 : Les éléments ϵ_i du vecteur aléatoire $\boldsymbol{\epsilon}$ sont des variables gaussiennes centrées, de variance σ^2 , indépendantes les uns des autres.

HT2 : Le graphe de \mathbf{M} est aléatoire et indépendant du vecteur aléatoire $\boldsymbol{\epsilon}$. Chaque réalisation de ce graphe, appelé *graphe des interactions*, résulte d'un graphe original, de matrice d'adjacence \mathbf{S} appelée *matrice de structure des interactions*, qui est détruit aléatoirement avec une probabilité p .

HT3 : La matrice aléatoire \mathbf{M} (comme paramètre ou variable) est une fonction du graphe des interactions suivant la définition (1.18) (normalisation des lignes de la matrice d’adjacence du graphe).

HT4 : Le coefficient d’autorégression ρ est un paramètre indépendant du vecteur ϵ et de la matrice \mathbf{M} (comme paramètre ou variable).

L’hypothèse HT1 est une hypothèse souvent employée dans le cadre du modèle de Whittle classique quand on ignore la distribution de l’élément stochastique ϵ . Elle ne possède pas de justification, si ce n’est que la forme gaussienne correspond à une distribution d’entropie maximale étant donnée la matrice de variance-covariance du champ.

L’hypothèse HT2 intègre l’hypothèse (1.24) (au Chapitre 1) : on considère qu’une interaction est possible entre deux variables du champ si et seulement si l’écart entre les positions spatiales de ces variables est suffisamment petit. La proximité des variables entre elles est déterminée par le graphe de structure des interactions mettant en relation les variables entre elles selon un critère de proximité spatiale. Il en découle que si s_{ij} est le terme général de la matrice de structure, alors le terme général a_{ij} de la matrice d’adjacence est une variable aléatoire de la forme

$$a_{ij} = s_{ij}b, \quad (2.3)$$

où b est une variable de Bernouilli qui vaut 0 avec une probabilité p et qui vaut 1 avec une probabilité $1-p$. Dans nos travaux, la structure des interactions principalement considérée est celle d’un *réseau carré*, mais d’autres structures d’interactions sont envisageables. La figure 1.1 (au Chapitre 1) illustre quelques réalisations du graphe des interactions sur une structure d’interactions en réseau carré.

L’hypothèse HT3 est ce qui détermine la valeur des éléments m_{ij} avec sa justification donnée au chapitre précédent.

L’hypothèse HT4 assure à ρ le rôle d’une grandeur de l’autocorrélation spatiale qui est indépendante du graphe des interactions (et donc de \mathbf{M}). Dans ce sens nous admettons que ρ est une grandeur intensive de l’autocorrélation spatiale, car elle ne dépend ni de la structure des interactions ni de l’extension spatiale du champ. Plus subtilement, l’hypothèse HT4 induit le principe que le graphe des interactions n’est pas une expression topologique de l’autocorrélation spatiale : les sites sont adjacents par hasard, sans que cela soit commandé par le même processus qui donne sa valeur à ρ , mais dès lors que deux sites sont adjacents, ceux-ci participent à l’autocorrélation spatiale.

2.1.2 Champ spatial résultant

En suivant les hypothèses HT1-HT4, les variables ϵ et \mathbf{M} sont mutuellement indépendantes. La loi de probabilité de ϵ est la loi normale de moyenne $\mathbf{0}$ et de matrice de variance-covariance diagonale $\mathbf{I}\sigma^2$, notée $\mathcal{N}_{(\mathbf{0}, \mathbf{I}\sigma^2)}$; la loi de probabilité

de \mathbf{M} , notée $\mathcal{M}_{(p,\mathbf{S})}$, est une loi discrète de paramètres \mathbf{S} et p dont \mathbf{S} est la matrice de structure des interactions et p est le taux de destruction de cette structure (suivant HT2) :

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2), \quad \mathbf{M} \sim \mathcal{M}_{(p,\mathbf{S})}. \quad (2.4)$$

La probabilité d'une réalisation \mathbf{m} de \mathbf{M} est égale à $h(\mathbf{m})$ définie par (cf. annexe de ce chapitre)

$$h(\mathbf{m}) = (1-p)^{\frac{1}{2}\tau[\mathbf{m}]} p^{\frac{1}{2}(\tau[\mathbf{S}] - \tau[\mathbf{m}])}, \quad (2.5)$$

où $\tau[\cdot]$ est un opérateur qui totalise le nombre d'éléments non nuls de la matrice donnée en argument.

Avec toutes ces conditions, on obtient un champ spatial \mathbf{X} conditionnel à \mathbf{m} qui est gaussien centré, de densité de probabilité $f(\mathbf{x}|\mathbf{m})$ de la forme

$$f(\mathbf{x}|\mathbf{m}) = \frac{1}{(\sqrt{2\pi})^n |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}\right], \quad \boldsymbol{\Sigma}^{-1} = (\mathbf{I} - \rho\mathbf{m})' \frac{1}{\sigma^2} (\mathbf{I} - \rho\mathbf{m}), \quad (2.6)$$

et un champ spatial \mathbf{X} conjoint à \mathbf{M} qui est de mesure de probabilité $g(\mathbf{x}, \mathbf{m})d\mathbf{x}$ de la forme

$$g(\mathbf{x}, \mathbf{m})d\mathbf{x} = f(\mathbf{x}|\mathbf{m})h(\mathbf{m})d\mathbf{x}. \quad (2.7)$$

Illustration La figure 2.1 illustre des réalisations indépendantes du modèle de Whittle markovien probabilisé dont la matrice de structure \mathbf{S} de taille $10^4 \times 10^4$ représente un réseau carré de 100×100 sites, et dont $\sigma^2 = 1$. Les valeurs des paramètres ρ et p sont indiquées dans la figure. Les réalisations du champ sont obtenues à partir de (2.2).

2.2 Variance du champ

2.2.1 Expression générale de la matrice de variance-covariance

Si un champ spatial \mathbf{X} suit la relation (2.1) (que la matrice des interactions \mathbf{M} soit un paramètre ou une variable), l'expression (2.2) en est la solution formelle avec pour matrice de variance-covariance $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \mathbf{E}_{\mathbf{X}}[\mathbf{X}\mathbf{X}'] = \mathbf{E}_{\mathbf{M}}\mathbf{E}_{\boldsymbol{\epsilon}}\left[\left[(\mathbf{I} - \rho\mathbf{M})^{-1}\boldsymbol{\epsilon}\right]\left[(\mathbf{I} - \rho\mathbf{M})^{-1}\boldsymbol{\epsilon}\right]'\right]. \quad (2.8)$$

L'expression n'est pas aisée à traiter mathématiquement étant données les opérations d'inversion de matrice et le calcul des espérances mathématiques. Lorsque \mathbf{M} est un paramètre au lieu d'être une variable, l'expression (2.8) se simplifie et prend la forme vu précédemment sous l'expression (1.8), plus précisément

$$\boldsymbol{\Sigma} = (\mathbf{I} - \rho\mathbf{M})^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\mathbf{I} - \rho\mathbf{M}')^{-1}, \quad \text{avec } \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathbf{I}\sigma^2, \quad (2.9)$$

conformément à la loi de $\boldsymbol{\epsilon}$ donnée dans (2.4).

La matrice de variance-covariance du champ diffère selon que la matrice \mathbf{M} est un paramètre (cas classique du modèle de Whittle) ou est une variable (modèle de Whittle probabilisé). Nous décrivons ces deux cas dans ce qui suit.

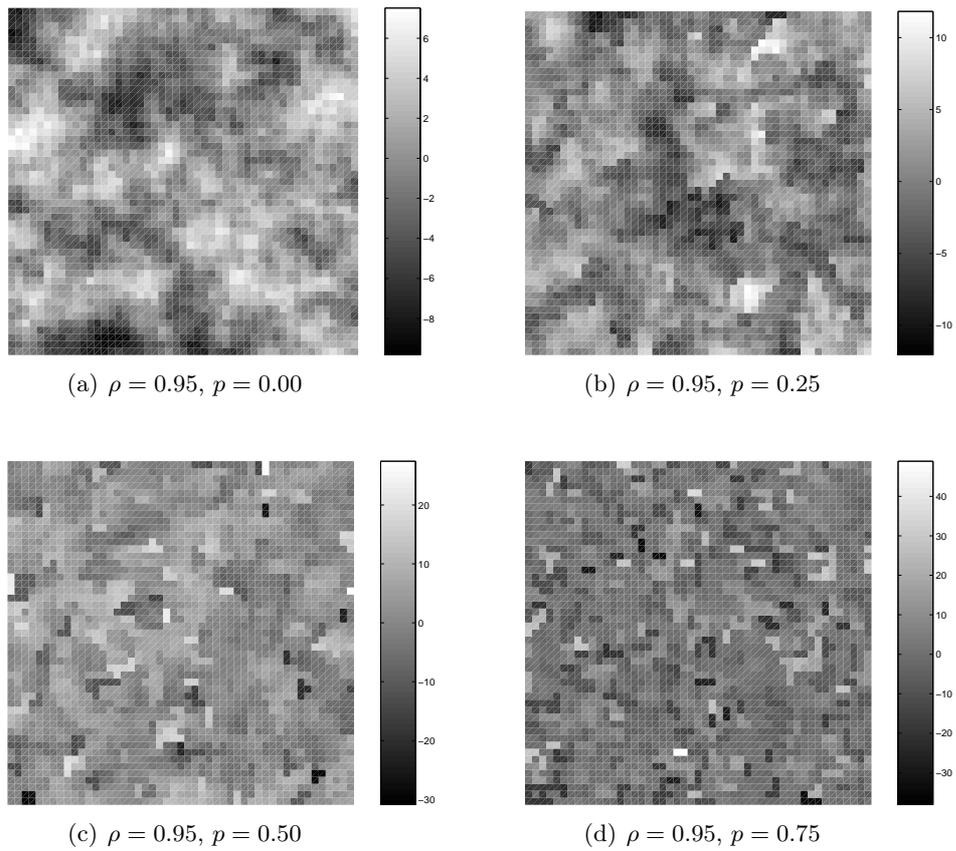


Fig. 2.1: Réalisations indépendantes d'un modèle de Whittle markovien probabilisé de 50×50 sites pour un choix de valeurs des paramètres p et ρ ($\sigma^2 = 1$).

2.2.2 Cas où \mathbf{M} est un paramètre

Espace du coefficient d'autorégression

Considérons le cas où la matrice des interactions \mathbf{M} est un paramètre du modèle de champ spatial (2.1). Dans ce cas, il s'agit d'un modèle de Whittle classique. On observe facilement que ce modèle n'est pas consistant si la matrice $(\mathbf{I} - \rho\mathbf{M})$ n'est pas inversible, ou de manière équivalente, si le déterminant de $(\mathbf{I} - \rho\mathbf{M})$ est nul. Dans ce cas, ρ^{-1} ne doit pas prendre l'une des valeurs propres de \mathbf{M} . On appelle *spectre de \mathbf{M}* l'ensemble de ces valeurs propres, noté $\text{Sp}[\mathbf{M}]$. Si \mathbf{M} est une matrice de Markov de taille $n \times n$ représentant une chaîne de Markov irréductible, alors en vertu du théorème de Perron-Frobenius (voir p. ex. MacCluer, 2000) le spectre de \mathbf{M} est composé de n valeurs complexes de module inférieur ou égal à 1 ; l'une de ces valeurs est réelle et vaut 1 avec une multiplicité algébrique égale à 1. Si \mathbf{M} est obtenue par normalisation d'une matrice d'adjacence symétrique (c'est-à-dire est de la forme (1.18)), alors toutes les valeurs propres de \mathbf{M} sont réelles (voir annexe), auquel cas le modèle de Whittle est consistant au moins pour l'intervalle ouvert $(-1, 1)$ de ρ .

Dans la littérature scientifique, on impose fréquemment au modèle le fait que ρ soit dans l'intervalle $[0,1)$ (autocorrélation positive) et/ou l'intervalle $(-1,0]$ (autocorrélation négative), quand bien même le modèle spatial puisse être consistant pour des valeurs de ρ en dehors de l'intervalle $[-1,1]$ (Kelejian & Robinson, 1995). Ce choix de l'intervalle peut être une commodité du modèle qui assure à la matrice de variance-covariance (2.9) de ne jamais être singulière. Il s'avère aussi un choix pour approcher les conditions d'un champ stationnaire au sens large.

Condition de stationnarité du champ spatial

La stationnarité au sens large d'un champ $\mathbf{X} = (X_1, X_2, \dots, X_n)$ satisfait aux deux conditions suivantes :

$$E[X_i] = \mu \quad \forall i, \tag{2.10}$$

$$\text{Cov}(X_i, X_j) = C(\mathbf{r}_i - \mathbf{r}_j), \tag{2.11}$$

où \mathbf{r}_i et \mathbf{r}_j sont les positions de X_i et X_j respectivement et $C()$ est une fonction appelée *covariogramme* qui ne dépend que de la distance relative entre les positions des variables. Il s'avère que le modèle de Whittle sur un réseau carré infini est stationnaire sous la condition $|\rho| < 1$ (Haining, 1978a, p.34). On doit s'attendre à ce que la stationnarité ne se vérifie que sur des champs idéaux dont les bords sont soit absents (champ spatial recourbé sur un tore) soit situés à l'infini (champs de taille infinie).

En réalité, la condition $|\rho| < 1$ n'assure pas la stationnarité pour n'importe quelle matrice de Markov \mathbf{M} , à commencer par le fait qu'elle est affectée par l'existence de bords physique au champ. Il est vérifié que le bord d'un champ en réseau carré parfait fini produit un dégradé de variance qui va en décroissant du bord vers

le centre du champ (Haining, 1980), comme illustré dans la figure 2.2(a). La forme du dégradé montre que pour ce type de champ, la condition $|\rho| < 1$ est à prendre comme la condition nécessaire, mais non suffisante, pour approcher des conditions de stationnarité, cette approximation s'améliorant avec la taille du réseau et l'éloignement des bords physiques.

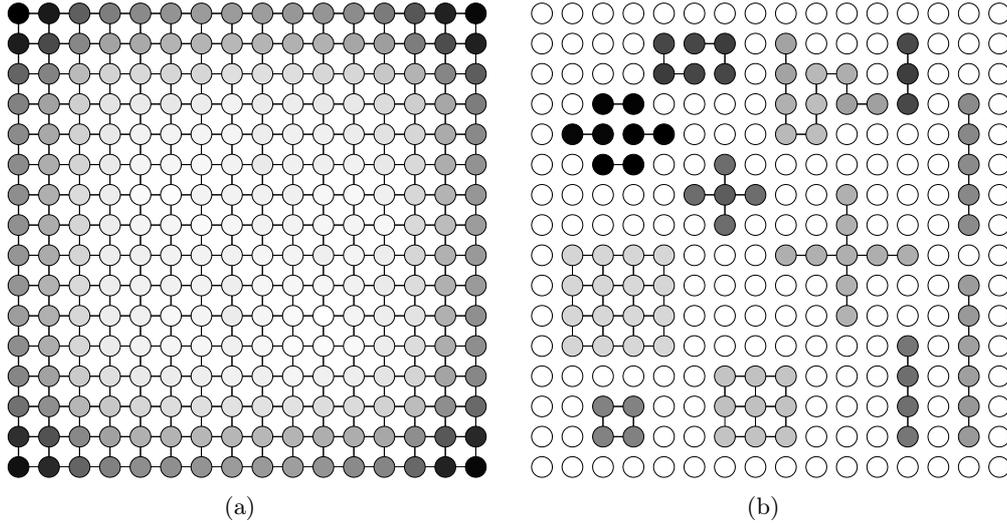


Fig. 2.2: Patterns des variance d'un modèle de Whittle markovien classique de paramètre $\rho = 0.9$, associés aux graphes indiqués, obtenus avec 10^4 réalisations du modèle spatial. Les valeurs sont traduites en niveaux de gris : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

Effet du graphe des interactions

Considérons un champ spatial dont les variables sont placées sur les noeuds d'un réseau carré. Lorsque le champ spatial est très petit, la variance dépend fortement de la géométrie du graphe des interactions du champ si ρ est suffisamment élevé, comme l'illustre la figure 2.2(b). Parmi les champs de très petites tailles induits par les sous-graphes de la figure 2.2(b), nous observons que la variance est relativement uniforme pour les variables d'un même graphe, avec une tendance à être plus élevée pour les plus petits graphes, exceptés les variables indépendantes qui, au contraire, ont les plus faibles variances. En particulier, les graphes (cliques) à 2 sites contrastent le plus en variance par rapport aux variables isolées. L'illustration laisse envisager qu'un fractionnement du graphe des interactions non seulement fait perdre de la connectivité au graphe, mais est aussi susceptible de faire augmenter la variance du champ, notamment si l'on parvient à couvrir densément l'espace de graphes d'interaction à 2 sites.

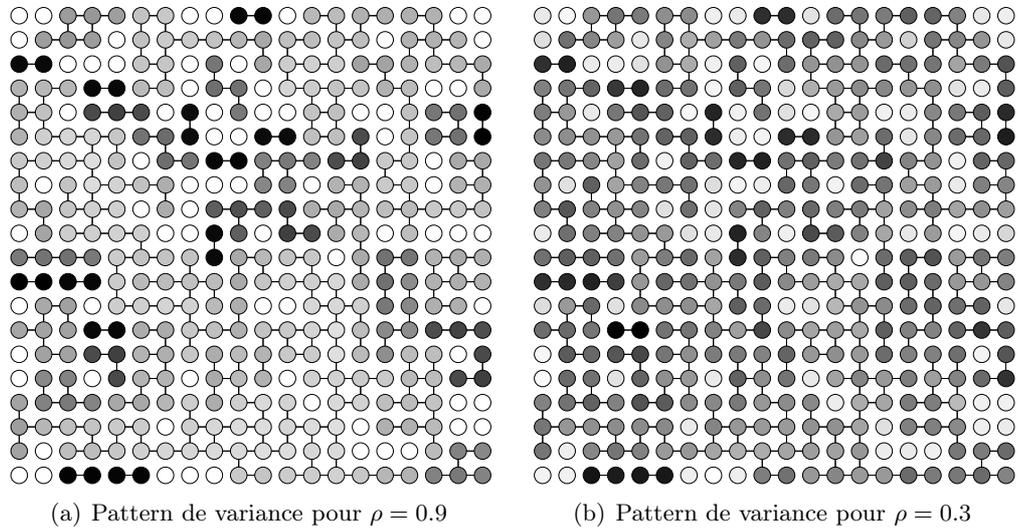


Fig. 2.3: Pattern de variance pour un champ de paramètre ρ , sur un réseau d'interactions carré détruit avec une probabilité 0.65.

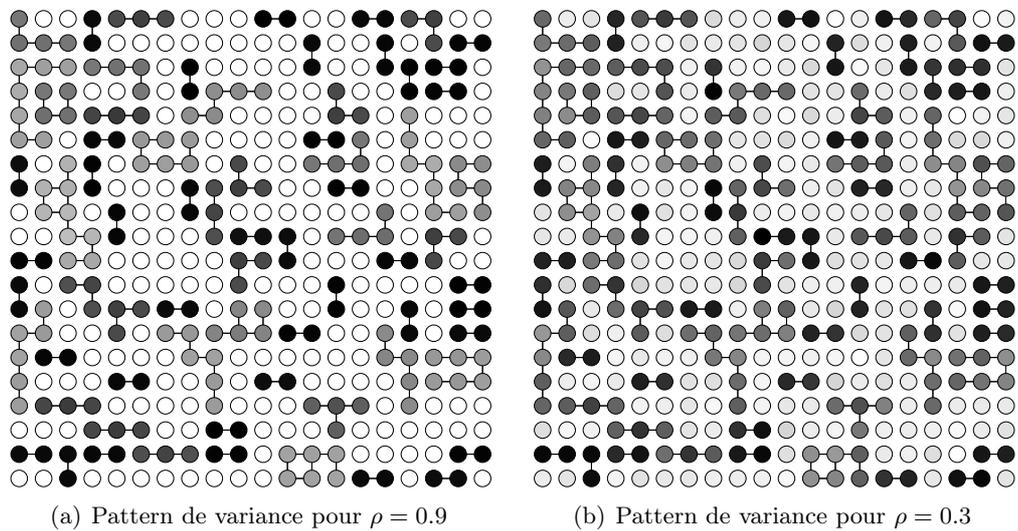


Fig. 2.4: Pattern de variance pour un champ de paramètre ρ , sur un réseau d'interactions carré détruit avec une probabilité 0.80.

En adoptant à titre d'exemple un graphe des interactions configuré de la même manière qu'un des graphes de la figure 1.1, on obtient des patterns de variances comme illustrés dans les figures 2.3 et 2.4. On peut aisément y distinguer visuellement, en certaines parties du champ, les variables qui font partie d'un même sous-graphe, la différenciation entre sous-graphes devenant plus aisée si les valeurs de p et ρ sont suffisamment élevées. Une façon d'opérer cette différenciation est d'appliquer une classification hiérarchique ascendante des données de variance. La figure 2.5 illustre le résultat d'une telle classification en se basant sur le critère de Ward¹ pour les liens entre valeurs du champ. En choisissant de ne retenir que 3 classes, on parvient à mettre en évidence en une classe les sites de connectivité locale nulle, dans une deuxième classe les cliques à 2 variables et dans la troisième classe le reste. Les deux premières classes sont particulièrement intéressants dans l'analyse de l'estimation de ρ au chapitre suivant. Retenir plus de 3 classes n'est pertinent qu'en fonction de la complexité (ou taille) des composantes connexes du graphe et de ρ .

Il reste à constater que le pattern de variance permet difficilement à lui seul de déterminer la connectivité des variables. Malgré cela, on conclut que la cartographie de la variance des variables du champ peut devenir, sous de bonnes conditions, *un appui* pour déceler directement le graphe des interactions.

Remarque Lorsqu'on possède une estimation de la matrice des variances *et des covariances* du champ, la méthode de Bhattacharjee & Jensen-Butler (2006) peut être appliquée pour déterminer la matrice des interactions et reconstituer ainsi le graphe des interactions.

2.2.3 Cas où \mathbf{M} est une variable

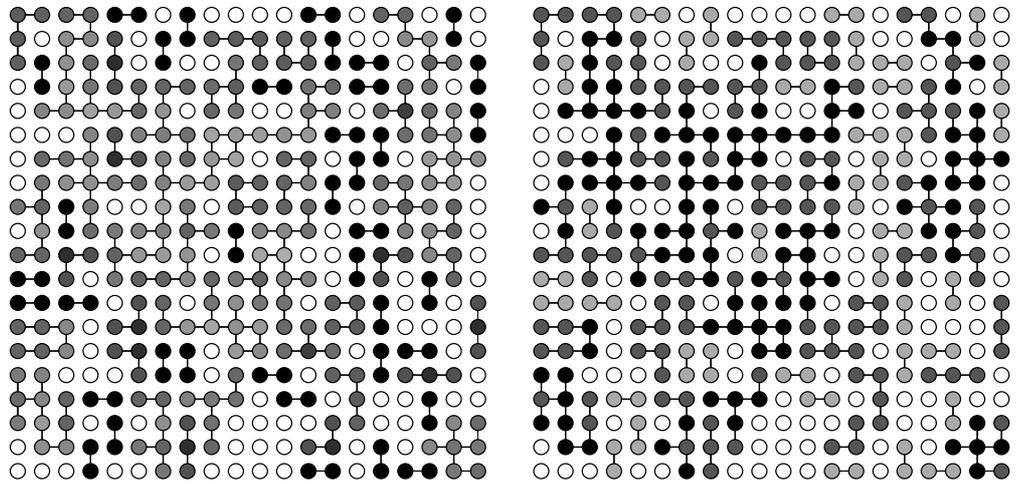
Espace du coefficient d'autorégression

Considérons maintenant le cas où la matrice des interactions de \mathbf{M} est une variable du modèle de champ spatial (2.1). Dans ce cas, il s'agit d'un modèle de Whittle markovien probabilisé. On rappelle que \mathbf{M} est une matrice de Markov aléatoire. La consistance de ce modèle est par conséquent toujours assurée pour ρ dans l'intervalle $[0, 1)$ et/ou l'intervalle $(-1, 0]$, en vertu des explications données précédemment pour le modèle de Whittle classique.

Stationnarité du champ spatial

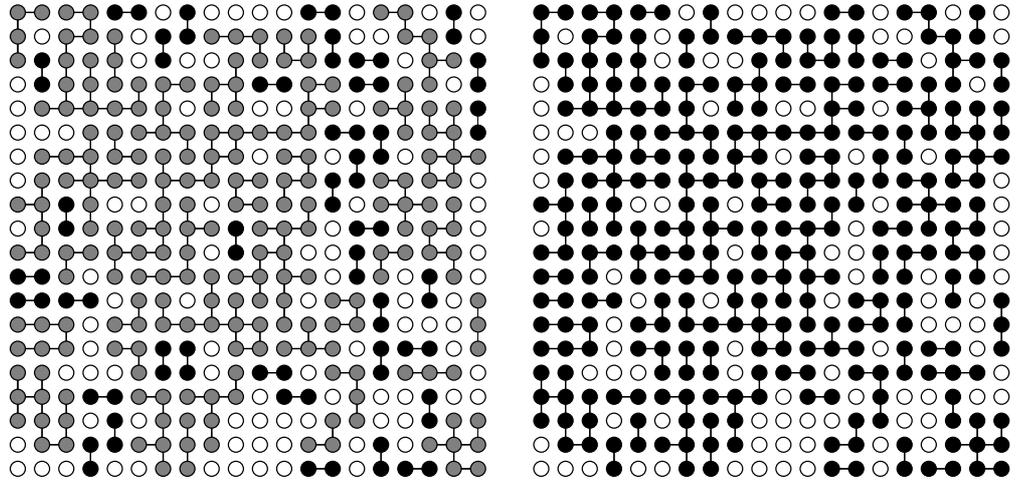
Pour ce modèle également, la condition $|\rho| < 1$ n'assure pas toujours la stationnarité du champ. Les figures 2.6(a) et 2.6(b) démontrent ce fait en reprenant le graphe des interactions de la figure 2.2(a) et 2.2(b) respectivement comme graphe de structure des interactions, et en appliquant le taux de destruction $p = 0.65$ pour

1. Le critère de Ward produit des classes relativement homogènes en variance intragroupe; voir p. ex. Lebart & al. (1995).



(a) Pattern de variances pour $\rho = 0.50$. Les valeurs sont traduites en niveaux de gris : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

(b) Classification des variances en 4 classes : le niveau de gris indique la classe. L'ordre des niveaux de gris n'a pas de signification.



(c) Classification des variances en 3 classes : le niveau de gris indique la classe. L'ordre des niveaux de gris n'a pas de signification.

(d) Classification des variances en 2 classes : le niveau de gris indique la classe. L'ordre des niveaux de gris n'a pas de signification.

Fig. 2.5: Les variances cartographiées en figure (a) sont classées en respectivement 4, 3 et 2 classes par classification hiérarchique ascendante utilisant le critère de Ward (figures (b), (c), (d)).

ces structures. On tire la même conclusion sur la pertinence de l'intervalle de ρ que dans le cas où \mathbf{M} est un paramètre.

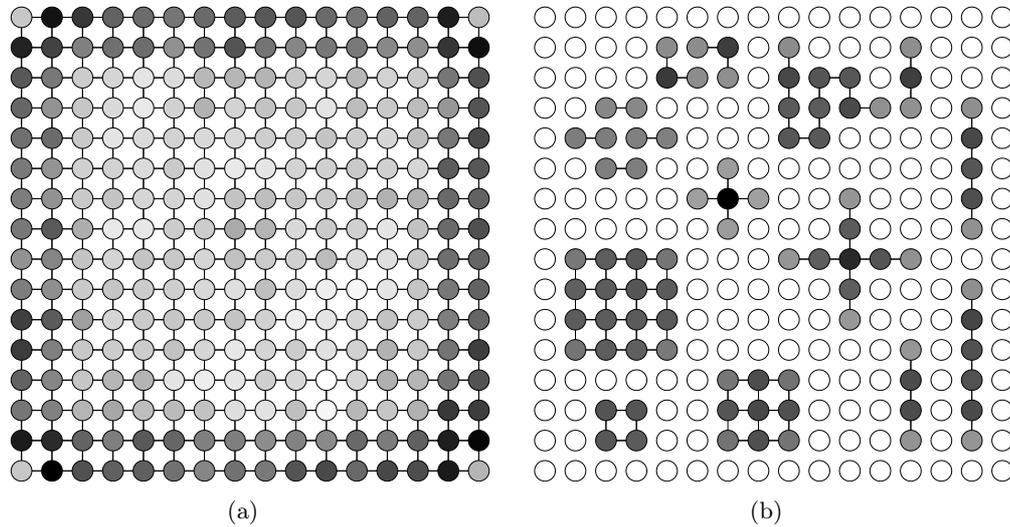
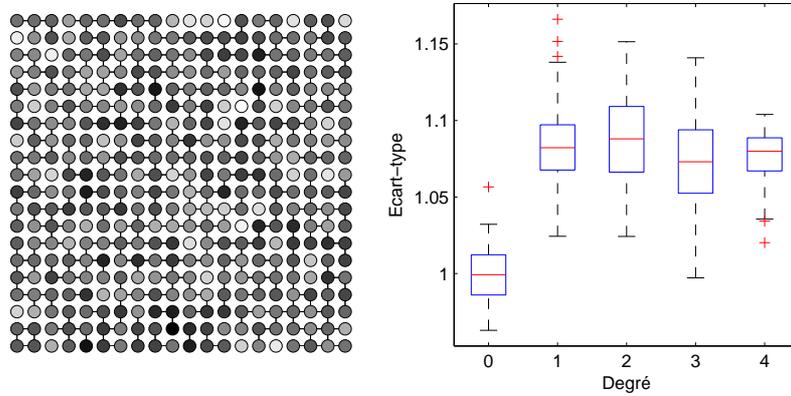


Fig. 2.6: Patterns des variance d'un modèle de Whittle markovien probabilisé de paramètres $p = 0.65$ et $\rho = 0.9$, associés aux graphes de structure indiqués, obtenus avec 10^4 réalisations du modèle spatial dans les conditions indiquées. Les valeurs sont traduites en niveaux de gris : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

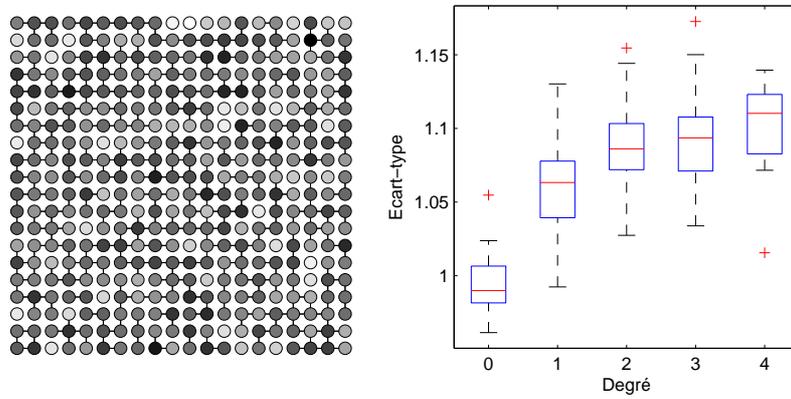
Effet des graphes de structure des interactions

En étudiant à nouveau les très petits champs, par exemple ceux associés aux graphes de structure d'interactions indiqués dans la figure 2.6(b), on observe de notables différences en comparaison avec le modèle de Whittle classique. Notamment, le pattern de variance à l'intérieur des petits champs n'est en général pas uniforme. En revanche, on relève une relative uniformité dans la variance pour les variables de connectivité locale égale à 1 (qui sont aux extrémités des champs de forme linéaire ou cruciforme). Ce fait saillant et divers autres aspects visibles du pattern de variance, sont indicateurs d'une forme de corrélation entre la variance et la connectivité locale de la variable.

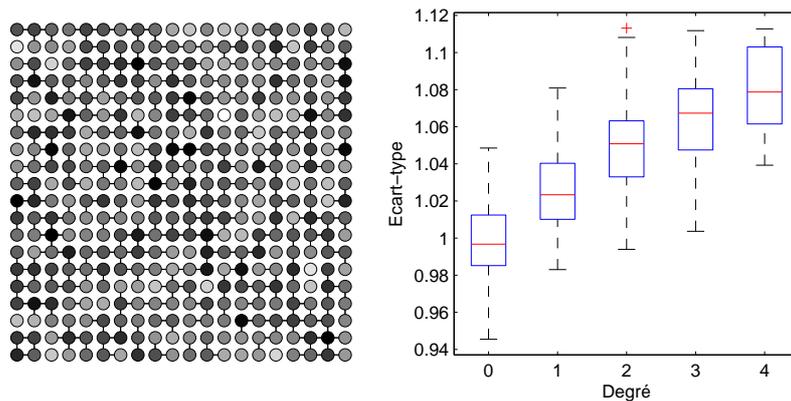
Pour avoir un aperçu de cette relation, nous considérons un modèle de Whittle probabilisé de paramètre p et dont la matrice de structure des interactions \mathbf{S} est la matrice d'adjacence d'un graphe d'un réseau carré parfait dont les arêtes ont été supprimées avec une probabilité de 0.5. Sur la base de cette structure, nous simulons une série de réalisations du champ à paramètre p constant, puis calculons la variance en chaque variable et portons les résultats dans un diagramme en boîte (boxplot). Ce type de diagramme permet de représenter la tendance centrale et la



(a) $\rho = 0.30, p = 0.20$

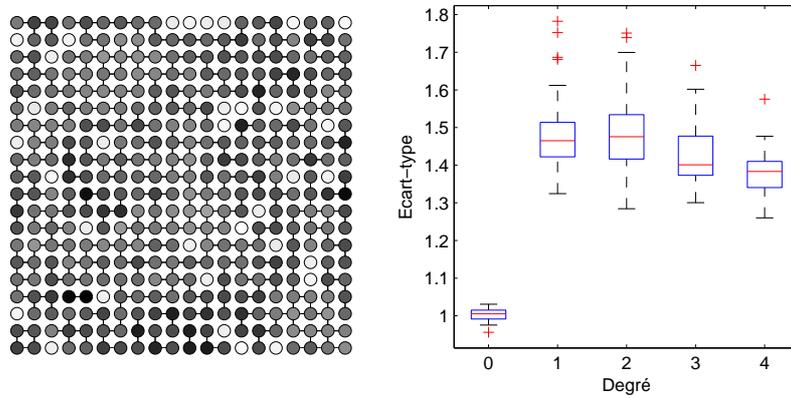


(b) $\rho = 0.30, p = 0.50$

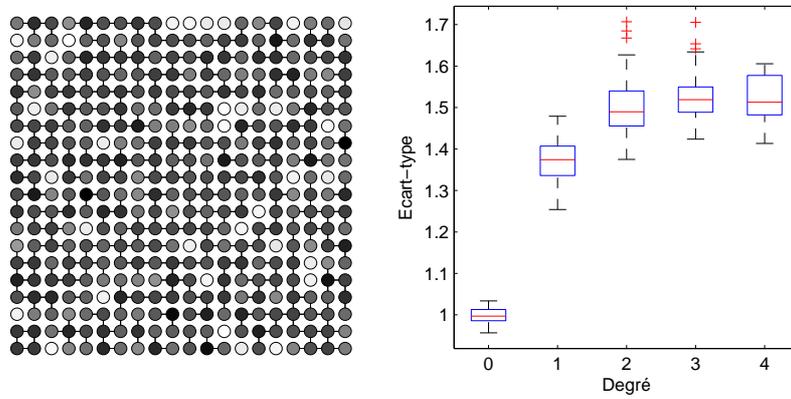


(c) $\rho = 0.30, p = 0.80$

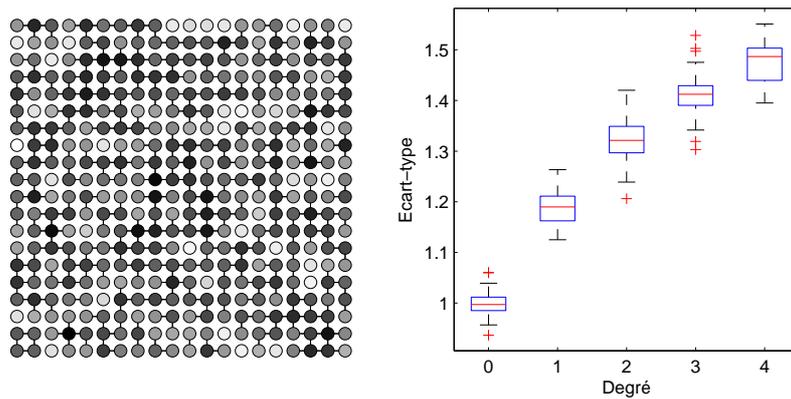
Fig. 2.7: Gauche : pattern de variances du modèle de Whittle markovien probabilisé de paramètre p et ρ et de graphe de structure y dessiné. Droite : diagrammes en boîtes caractérisant la variance des variables du champ en fonction de leur degré (connectivité locale).



(a) $\rho = 0.60, p = 0.20$

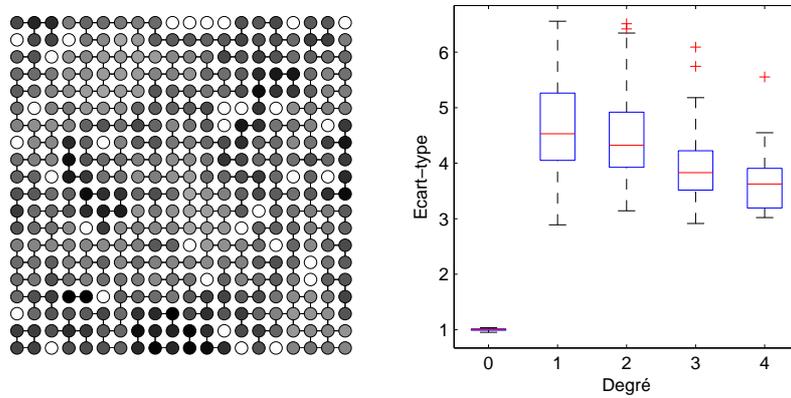


(b) $\rho = 0.60, p = 0.50$

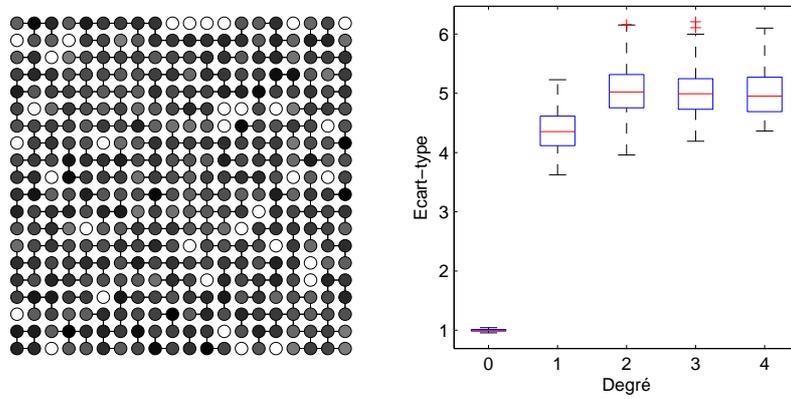


(c) $\rho = 0.60, p = 0.80$

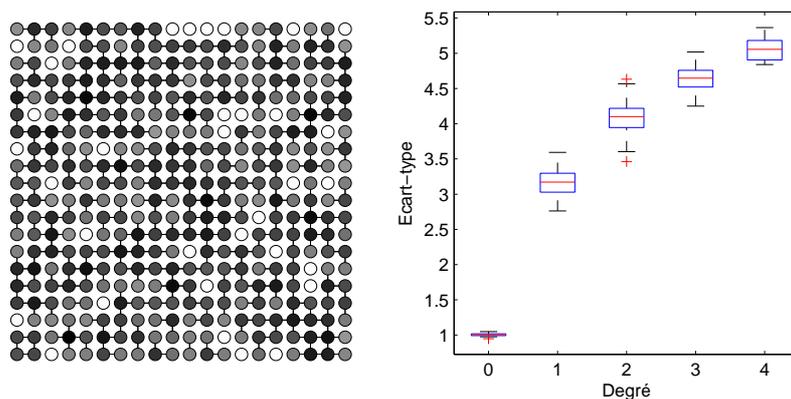
Fig. 2.8: Gauche : pattern de variances du modèle de Whittle markovien probabilisé de paramètre p et ρ et de graphe de structure y dessiné. Droite : diagrammes en boîtes caractérisant la variance des variables du champ en fonction de leur degré (connectivité locale).



(a) $\rho = 0.90, p = 0.20$



(b) $\rho = 0.90, p = 0.50$



(c) $\rho = 0.90, p = 0.80$

Fig. 2.9: Gauche : pattern de variances du modèle de Whittle markovien probabilisé de paramètre p et ρ et de graphe de structure y dessiné. Droite : diagrammes en boîtes caractérisant la variance des variables du champ en fonction de leur degré (connectivité locale).

dispersion de la variance des variables en fonction de leur connectivité locale selon le graphe de structure. Les résultats en figures 2.7 à 2.9 illustrent le pattern de variance et montrent que la variance permet de discriminer la connectivité locale du graphe des interactions, notamment si p est suffisamment élevé.

En conclusion, le pattern de variance du modèle de Whittle probabilisé, a tendance, sous de bonnes conditions, à révéler le graphe de structure des interactions par sa connectivité locale.

Variance totale du champ

La somme des variances individuelles donne la variance totale du champ. A titre illustratif, la figure 2.10 présente la variance totale en fonction de p , pour trois valeurs de ρ , d'un modèle de Whittle probabilisé de graphe de structure de la forme d'un réseau carré de 20×20 sites. La figure illustre le fait que

- (a) la variance totale augmente si ρ augmente,
- (b) la variance totale possède un extrêmuum à l'intérieur de l'intervalle (0,1) de p .

L'augmentation de variance constatée avec p croissant interpelle sur le rôle de ρ . Une remarque est émise à ce sujet en marge de la conclusion de ce chapitre.

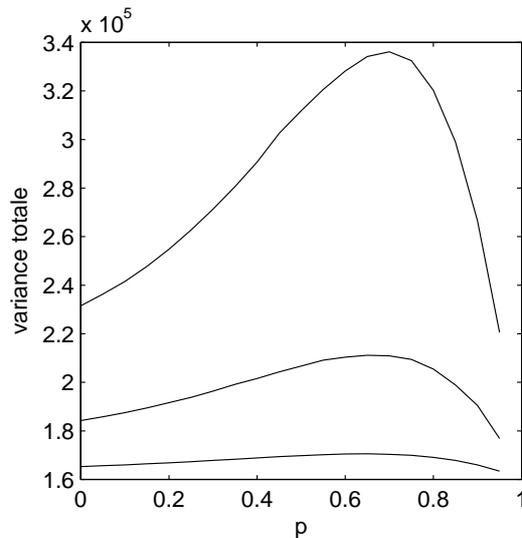


Fig. 2.10: Variance totale d'un modèle de Whittle probabilisé de graphe de structure de la forme d'un réseau carré de 20×20 sites. Les courbes sont associées, de bas vers le haut, à $\rho = 0.2, 0.4, 0.6$.

2.3 Corrélation du champ

Par définition un modèle de Whittle de paramètre $\rho = 0$ implique l'absence de corrélation entre les variables du champ. A l'inverse, lorsque ρ diffère de 0, il existe une forme de corrélation entre les variables. Les auteurs qui traitent du modèle de Whittle classique ignorent généralement la forme de relation entre le coefficient d'autorégression ρ et la corrélation entre variables du champ. Le fait qu'une valeur de ρ non nulle engendre une corrélation non nulle dans le champ et le fait que la forme du modèle de Whittle classique est proche du modèle autorégressif standard employé dans les séries temporelles ont amené certains auteurs à interpréter le paramètre ρ comme d'un coefficient d'autocorrélation, d'un paramètre d'autocorrélation ou encore d'un paramètre d'autorégression (voir Kelejian & Robinson, note 3, 1995, pour les auteurs cités en exemple). En réalité, les modèles de Whittle associées à des matrices \mathbf{M} markoviennes ont au moins deux variables du champ qui tendent vers une corrélation de 1 en valeur absolue lorsque ρ^{-1} tend vers l'une des valeurs de $\text{Sp}[\mathbf{M}]$ (Kelejian & Robinson, 1995). Pour une matrice \mathbf{M} basée sur un graphe des interactions de la forme d'un réseau carré parfait (non détruit), la totalité des couples de variables tendent vers une corrélation égale à 1 en valeur absolue lorsque ρ tend vers ± 1 . Il en est de même avec les champs de géométrie linéaire (c.-à-d. dont le graphe des interactions forme une chaîne). Ce n'est généralement pas le cas pour toutes les matrices \mathbf{M} markoviennes (voir Kelejian & Robinson, 1995, pour des exemples).

Il est intéressant de visualiser la complexité du pattern des corrélations du modèle de Whittle *classique* lorsque le graphe des interactions est aussi complexe que ceux en figure 1.1. Pour cette illustration, donnée en figure 2.11, nous nous limitons à montrer les corrélations entre variables plus proches voisines (selon la distance euclidienne) dans le contexte où le graphe des interactions est celui d'un réseau carré détruit avec une probabilité p . Les cas illustrés révèlent l'effet complexe d'augmentation et de disparition de la corrélation entre sites plus proches voisins avec le paramètre p .

Dans le cadre d'une comparaison de modèles, la figure 2.12 illustre classiquement la corrélation estimée par des simulations numériques le long d'un axe de symétrie d'un champ sur réseau carré dans le cas du modèle de Whittle classique et dans le cas du modèle de Whittle probabilisé. Elle démontre que pour un réseau carré d'interactions la corrélation décroît avec l'écart euclidien, la distance de chute de corrélation étant plus courte si ρ diminue et/ou si p augmente.

Remarque Wall (2004) constate que dans le contexte d'un modèle de Whittle classique (un modèle SAR) pour un graphe d'interactions irrégulier, le rang des corrélations entre variables du champ change avec ρ . Autrement dit, un couple de variables peut avoir une corrélation plus élevée que celle d'un autre couple de variables à une valeur donnée de ρ , mais la situation peut s'inverser à une autre valeur de ρ . Le constat est analogue avec le modèle autorégressif conditionnel (modèle CAR). Cet auteur conclut que le comportement des ces modèles est contre-

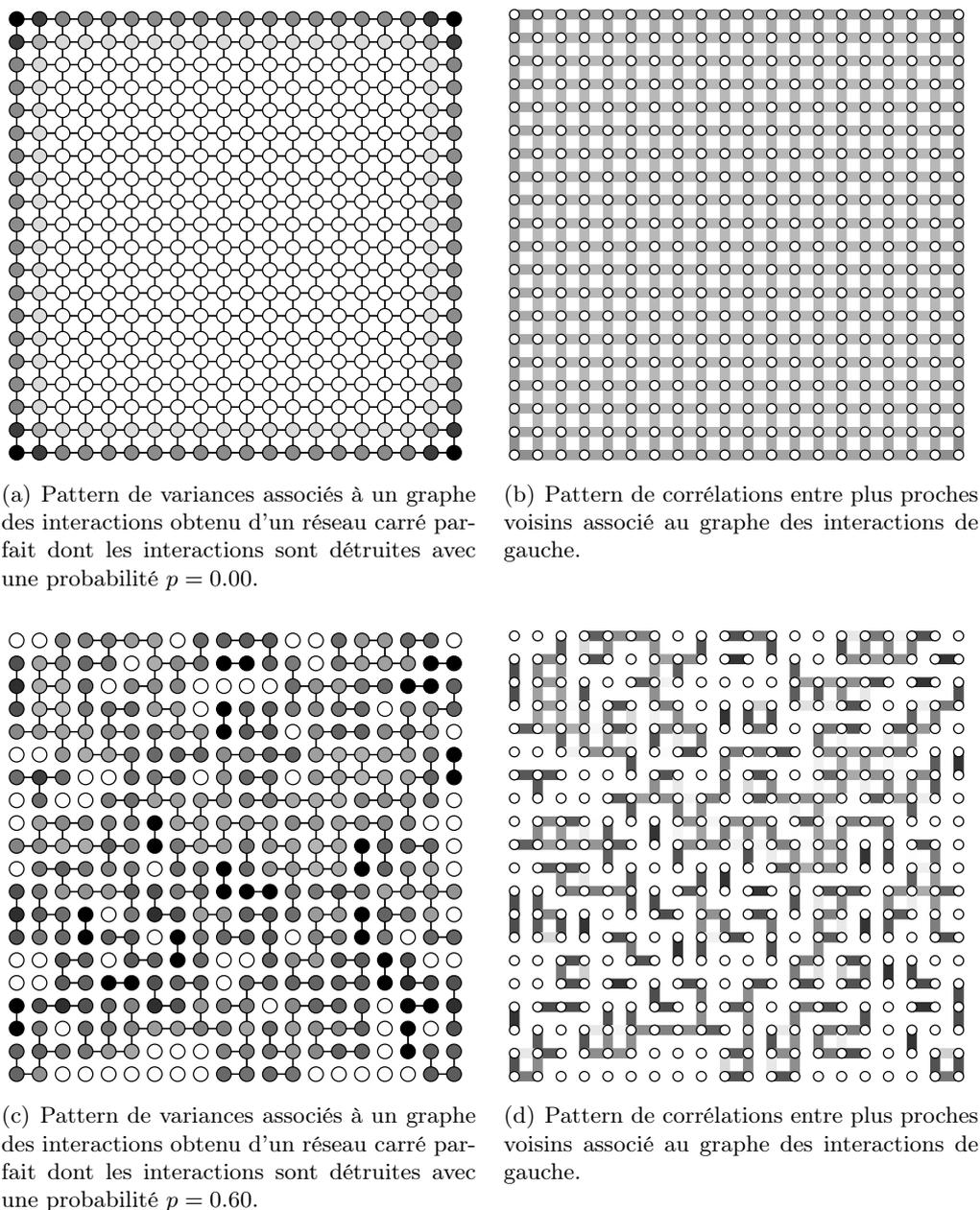


Fig. 2.11: Modèle de Whittle classique de paramètre $\rho = 0.5$. **Figures de gauche :** graphe des interactions et pattern des variances. Les valeurs des variances sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ. **Figures de droite :** pattern des corrélations entre sites plus proches voisins. Les valeurs des corrélations sont traduites en niveaux de gris des liens : blanc pour la valeur 0, noir pour la valeur 1. A noter que les corrélations pour des écarts spatiaux supérieurs ne sont pas indiqués.

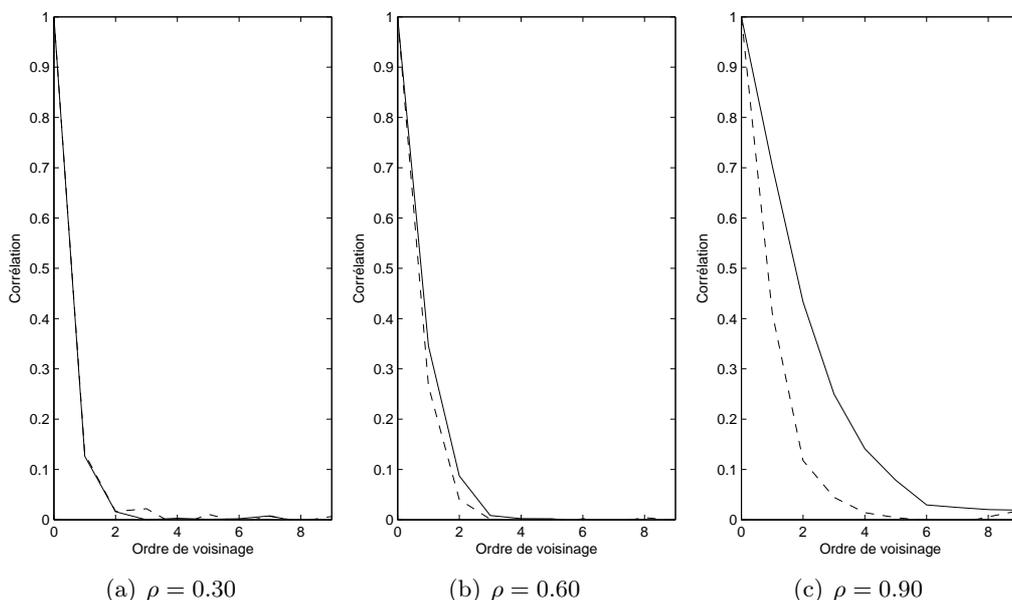


Fig. 2.12: Corrélation entre variables d'un modèle de Whittle markovien probabilisé sur réseau carré périodique de taille 100×100 sites en fonction de l'ordre de voisinage sur l'un des axes des interactions. Les résultats sont obtenus avec 10^4 réalisations du modèle aux conditions où $p = 0$ (en trait continu) ou $p = 0.65$ (en trait discontinu), $\sigma^2 = 1$ et ρ comme indiqué en bas de diagramme.

intuitif et reproche le fait que malgré la popularité des modèles SAR et CAR, ces modèles ont été ajustés aux données par divers auteurs sans comprendre leur signification.

2.4 Effets de bord de la fenêtre d'observation-modélisation

L'observation des variables s'effectue sur une partie d'un champ spatial généralement plus étendu dans l'espace, désigné ci-après comme *champ étendu*. Nous désignons le contour de cette partie comme étant la fenêtre d'observation. En définissant le contour de l'espace sur lequel s'applique le modèle de champ spatial comme étant la *fenêtre de modélisation*, il est tout à fait concevable que cette fenêtre diffère de la fenêtre d'observation, qu'elle la chevauche ou en est disjointe. Dans d'autres cas, la fenêtre d'observation et la fenêtre de modélisation coïncident. En admettant que tel est le cas pour nous, la fenêtre d'observation-modélisation détermine le périmètre en-dehors duquel les variables et le graphe des interactions du champ étendu sont ignorés. Les limites de cette fenêtre imposent des bords qui

ne sont pas les bords physiques du champ spatial étendu. De fait, elles introduisent un biais dans le traitement des quantités à estimer ou inférer. Ce problème subsiste tant que les conditions aux bords de la fenêtre d'observation-modélisation ne sont pas correctement déterminées. On compte cependant sur le fait que le biais devienne négligeable lorsque la fenêtre d'observation est de forme simple (un cercle ou un quadrat sans trous) et si le nombre de variables en bordure de fenêtre d'observation-modélisation représente une part suffisamment petite de l'ensemble des variables observées.

Dans le cas où le biais dans les estimations introduit par les limites de la fenêtre d'observation est trop grand, on peut exclure les variables trop proches du bord intérieur de la fenêtre d'observation. La partie du champ par lequel le biais est introduit peut être déterminée à l'aide de simulations numériques de la manière suivante. Si \mathbf{S} représente la matrice de structure des interactions d'un champ spatial sans bord et $\tilde{\mathbf{S}}$ sa restriction à l'intérieur de la fenêtre d'observation, alors on obtient deux champs spatiaux \mathbf{X} et $\tilde{\mathbf{X}}$ associés respectivement à \mathbf{S} et $\tilde{\mathbf{S}}$: \mathbf{X} représente le champ objectif et $\tilde{\mathbf{X}}$ représente le modèle construit dans l'ignorance des conditions aux bords externes de la fenêtre d'observation-modélisation. Partant de ces deux modèles, les variables sensiblement affectées par l'effet de bord peuvent être discriminées en chaque point à l'intérieur de la fenêtre par la différence quadratique moyenne entre la valeur qu'elles prennent si elles appartiennent au champ \mathbf{X} et la valeur qu'elles prennent si elles appartiennent au champ $\tilde{\mathbf{X}}$. Cette quantité peut être estimée, à l'aide de k réalisations du champ simulé, par l'expression

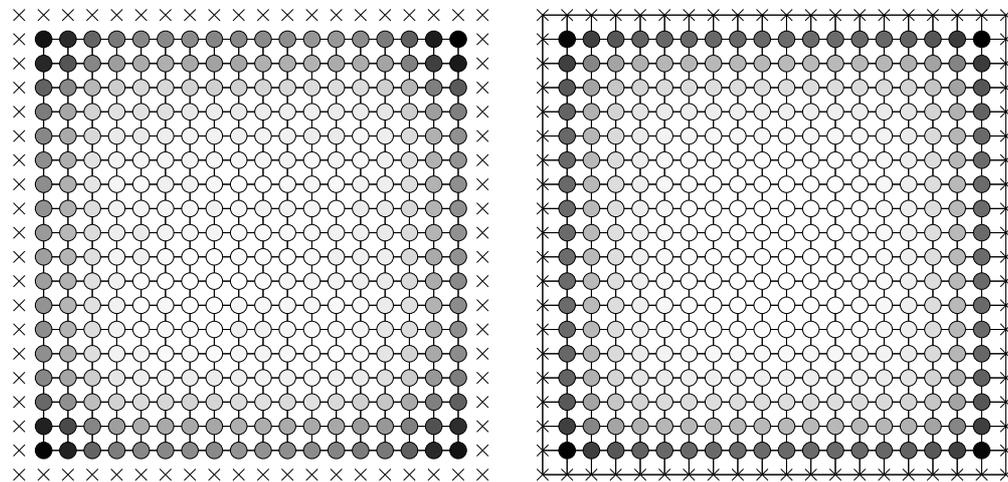
$$EQM_i = \frac{1}{k} \sum_{j=1}^k (x_{ij} - \tilde{x}_{ij})^2, \quad (2.12)$$

où x_{ij} et \tilde{x}_{ij} sont respectivement la j -ème réalisation de la variable localisée sur le i -ème site parmi les k variables localisées à l'intérieur de la fenêtre qui composent le champ \mathbf{X} , respectivement $\tilde{\mathbf{X}}$. La cartographie de la quantité EQM_i permet de délimiter la région du champ les moins affectées (respectivement les plus affectées) par les bords de la fenêtre d'observation-modélisation. Les figures 2.13 à 2.16 donnent quelques exemples de cette cartographie.

On conclut que la fragmentation du graphe de la structure des interactions réduit l'extension spatiale des sites "biaisants". Si cette extension reste toutefois critique, elle peut être exclue du modèle de Whittle employé pour modéliser le champ (en vue de l'estimation de ρ par exemple). Dans l'ignorance du graphe des interactions, la délimitation entre site "biaisant" et "non biaisant" peut évidemment être remplacée par une marge intérieure d'exclusion de sites.

2.5 Conclusion

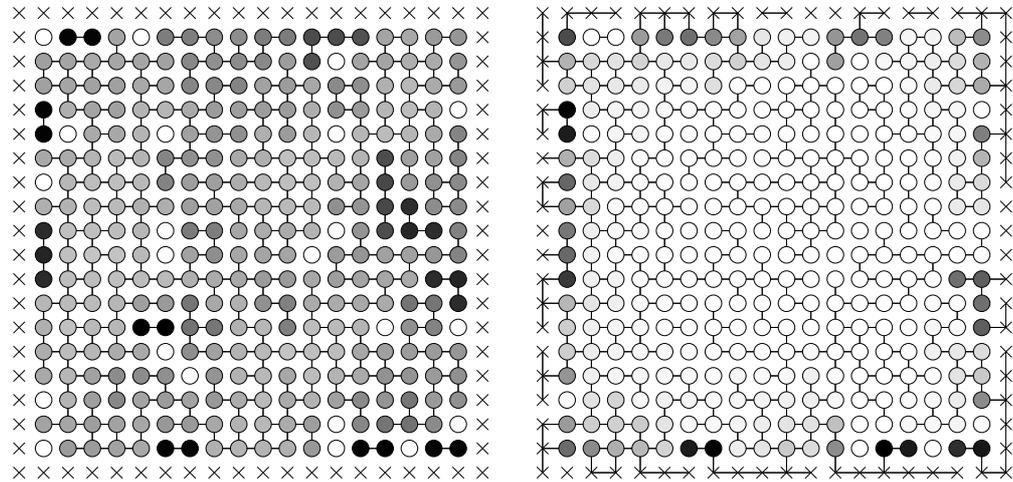
Nous rappelons que le modèle de Whittle classique est associé à un graphe des interactions déterministe alors que le modèle de Whittle probabilisé est associé à



(a) Cartographie de l'écart-type des valeurs du champ, un modèle de Whittle classique associé au graphe des interactions y indiqué. Les variables extérieures indiquées par une croix et les interactions avec celles-ci sont ignorées. Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

(b) Cartographie de $\sqrt{EQM_i}$. Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

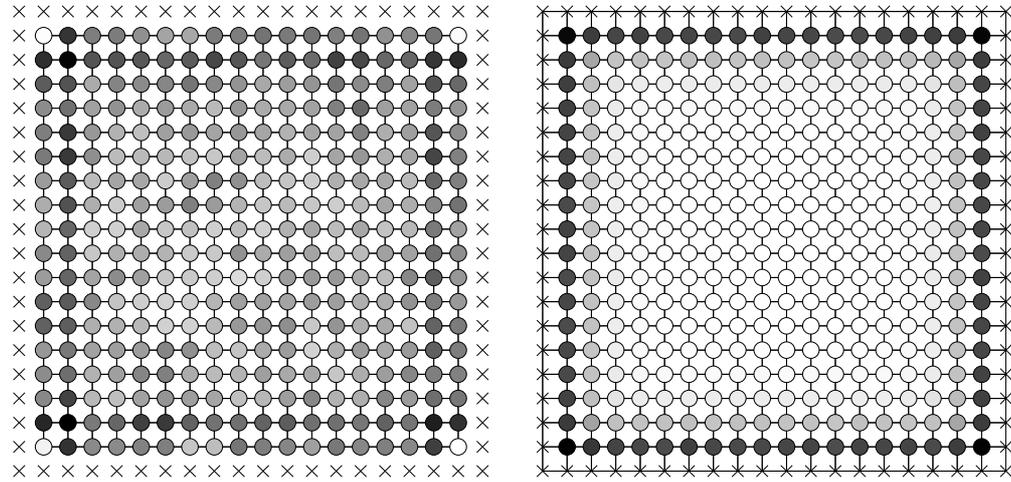
Fig. 2.13: Les ronds indiquent les variables d'un modèle de Whittle classique à l'intérieur d'une fenêtre d'observation-modélisation placée sur un champ spatial qui est un modèle de Whittle plus étendu spatialement et recourbé sur lui-même : les variables en bordure droite et gauche sont voisines, ainsi que les variables en bordure haute et basse. Les croix indiquent les variables voisines du champ étendu immédiatement en dehors de la fenêtre d'observation. Le graphe indiqué en (b) représente la structure des interactions du champ étendu (sans faire figurer les arcs joignant les bords opposés), alors que le graphe en figure (a) représente le graphe de structure que poserait l'observateur pour son modèle de champ (dans l'ignorance du reste du champ étendu). Nous illustrons ici le cas où le graphe des interactions du champ étendu est un réseau carré parfait recourbé (sur un tore) et le graphe isolé dans la fenêtre d'observation est le graphe du champ étendu tronqué par la fenêtre d'observation. Les résultats sont obtenus suite à 10^3 réalisations des modèles respectifs, avec $\rho = 0.9$.



(a) Cartographie de l'écart-type des valeurs du champ, un modèle de Whittle classique associé au graphe des interactions y indiqué. Les variables extérieures indiquées par une croix et les interactions avec celles-ci sont ignorées. Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

(b) Cartographie de la racine carrée de la quantité (2.12). Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

Fig. 2.14: Les ronds indiquent les variables d'un modèle de Whittle classique à l'intérieur d'une fenêtre d'observation-modélisation placée sur un champ spatial qui est un modèle de Whittle plus étendu spatialement et recourbé sur lui-même : les variables en bordure droite et gauche sont voisines, ainsi que les variables en bordure haute et basse. Les croix indiquent les variables voisines du champ étendu immédiatement en dehors de la fenêtre d'observation. Le graphe indiqué en (b) représente la structure des interactions du champ étendu (sans faire figurer les arcs joignant les bords opposés), alors que le graphe en figure (a) représente le graphe de structure que poserait l'observateur pour son modèle de champ (dans l'ignorance du reste du champ étendu). Nous illustrons ici le cas où le graphe des interactions du champ étendu est la réalisation d'un réseau carré parfait recourbé (sur un tore) détruit avec une probabilité de 0.5 et le graphe isolé dans la fenêtre d'observation est le graphe du champ étendu tronqué par la fenêtre d'observation. Les résultats sont obtenus suite à 10^3 réalisations des modèles respectifs, avec $\rho = 0.9$.



(a) Cartographie de l'écart-type des valeurs du champ, un modèle de Whittle probabilisé associé au graphe de structure des interactions γ indiqué. Les variables extérieures indiquées par une croix et les interactions avec celles-ci sont ignorées. Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

(b) Cartographie de la racine carrée de la quantité (2.12). Les valeurs sont traduites en niveaux de gris dans les ronds : blanc pour la valeur la plus basse, noir pour la valeur la plus élevée du champ.

Fig. 2.15: Les ronds indiquent les variables d'un modèle de Whittle probabilisé à l'intérieur d'une fenêtre d'observation-modélisation placée sur un champ spatial qui est un modèle de Whittle probabilisé plus étendu spatialement et recourbé sur lui-même : les variables en bordure droite et gauche sont voisines, ainsi que les variables en bordure haute et basse. Les croix indiquent les variables voisines du champ étendu immédiatement en dehors de la fenêtre d'observation. Le graphe indiqué en (b) représente la structure des interactions du champ étendu (sans faire figurer les arcs joignant les bords opposés), alors que le graphe en figure (a) représente le graphe de structure que poserait l'observateur pour son modèle de champ (dans l'ignorance du reste du champ étendu). Nous illustrons ici le cas où le graphe de structure des interactions du champ étendu est un réseau carré parfait recourbé (sur un tore) et le graphe isolé dans la fenêtre d'observation est le graphe de structure des interactions du champ étendu tronqué par la fenêtre d'observation. Les résultats sont obtenus suite à 10^3 réalisations des modèles respectifs, avec $\rho = 0.9$ et $p = 0.70$.

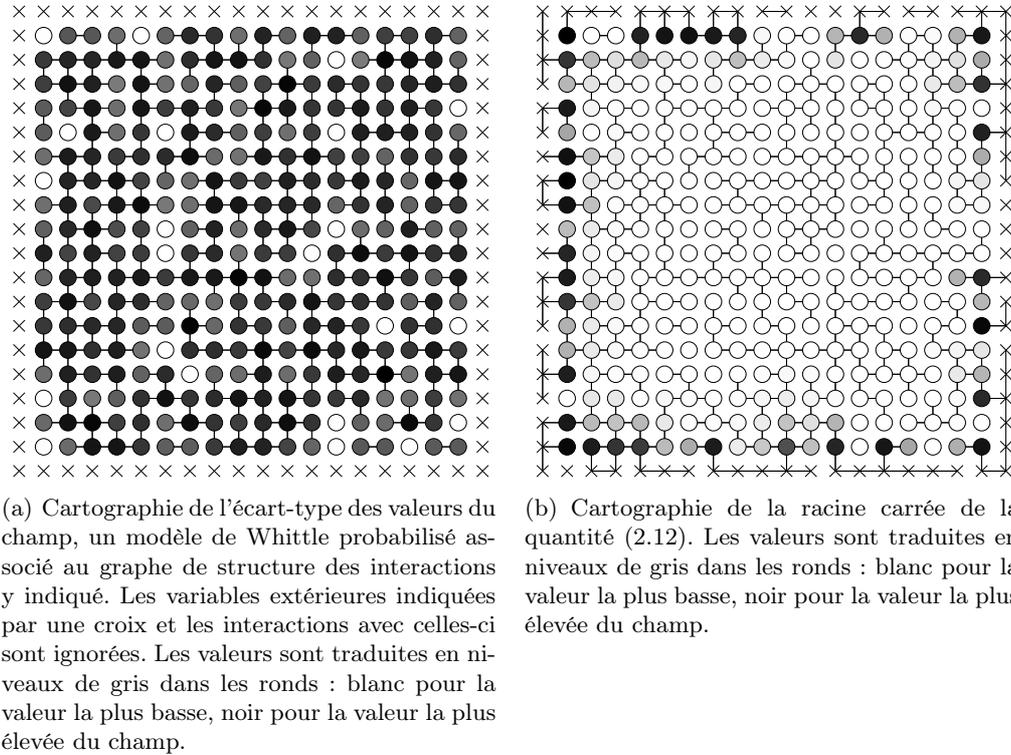


Fig. 2.16: Les ronds indiquent les variables d'un modèle de Whittle probabilisé à l'intérieur d'une fenêtre d'observation-modélisation placée sur un champ spatial qui est un modèle de Whittle probabilisé plus étendu spatialement et recourbé sur lui-même : les variables en bordure droite et gauche sont voisines, ainsi que les variables en bordure haute et basse. Les croix indiquent les variables voisines du champ étendu immédiatement en dehors de la fenêtre d'observation. Le graphe indiqué en (b) représente la structure des interactions du champ étendu (sans faire figurer les arcs joignant les bords opposés), alors que le graphe en figure (a) représente le graphe de structure que poserait l'observateur pour son modèle de champ (dans l'ignorance du reste du champ étendu). Nous illustrons ici le cas où le graphe de structure des interactions du champ étendu est la réalisation d'un réseau carré parfait recourbé (sur un tore) détruit avec une probabilité de 0.5 et le graphe isolé dans la fenêtre d'observation est le graphe de structure des interactions du champ étendu tronqué par la fenêtre d'observation. Les résultats sont obtenus suite à 10^3 réalisations des modèles respectifs, avec $\rho = 0.9$ et $p = 0.70$.

un graphe des interactions aléatoire s'inscrivant dans une structure des interactions déterministe. Les deux modèles représentent tous les deux une forme d'autorégression spatiale d'ordre 1, mais ils diffèrent quant au modèle de connexité du champ. Nous nous sommes attachés à décrire ces différences par là où on s'attend à les ressentir : dans la variance et la corrélation du champ. Nous nous sommes employés à illustrer la complexité spatiale de ces grandeurs tout en cherchant à les résumer. In fine, nous constatons les points suivants.

- Pour les deux types de modèles, on observe qu'une valeur de ρ positive traduit une augmentation de variance des variables (nulle ou strictement positive) par rapport au champ sous hypothèse d'indépendance spatiale.
- Pour les modèles de Whittle classiques, les variances individuelles des variables situées dans une même composante connexe du champ tendent à se ressembler lorsque ρ augmente. Dans ce sens, le modèle conduit à un modèle d'homoscédasticité par composantes connexes du champ. Nous avons vérifié avec de petites composantes connexes que la variance individuelle des variables en interaction dans une même composante connexe augmente lorsque la taille de cette composante connexe diminue (si l'on excepte les variables indépendantes). On relève dans ce contexte que les variables faisant partie d'une clique à deux sites ont la variance la plus élevée du champ et donc contrastent fortement avec les variables indépendantes qui ont la variance la plus basse du champ.
- Dans le cadre des modèles de Whittle probabilisés qui suivent notre modèle de connexité, les variables du champ présentent, à valeurs de ρ et de p suffisamment élevées, une corrélation positive entre variance et connectivité locale du graphe de structure.
- Pour les deux types modèles, la corrélation entre sites est décroissante de l'écart spatial des variables. A ρ donné, cette décroissance est plus rapide dans le modèle de Whittle probabilisé que dans le modèle de Whittle classique.
- La modélisation du champ par l'un des deux types de modèle à partir des observations effectuées dans une fenêtre spatiale peut introduire un biais dans l'estimation ou l'inférence de ses paramètres, à cause de la taille finie de la fenêtre d'observation. Le biais est introduit par des variables en bord intérieur de la fenêtre d'observation. Au besoin, elles peuvent être exclues du champ des estimations ou des inférences.

Remarque L'augmentation de variance du champ constatée avec la perte de connectivité du graphe est dépendante de notre modèle de connexité. Celui-ci est indépendant du coefficient d'autorégression ρ (hypothèse HT4). Cette indépendance peut être révisée dans le cas où le champ spatial étudié fait partie d'un système nécessitant une variation de ρ en même temps que la variation de la connexité du champ. Auquel cas, les courbes de variance en fonction du paramètre p peuvent avoir une toute autre allure. Un tel lien entre ρ et p traduit en fait une réduction du degré de liberté du modèle que nous considérons comme un raffinement intéressant du modèle mais hors de notre champ d'étude.

Chapitre 3

Information de Fisher

On peut se demander si nous avons bon espoir de déduire les paramètres du modèle d'autocorrélation spatiale à partir des données. En admettant que nous possédons déjà des estimateurs sans biais de ces paramètres, nous avons la possibilité de déduire théoriquement la limite à leur précision par une quantité appelée *information de Fisher*. Nous allons étudier cette quantité, principalement pour le coefficient d'autorégression ρ , et accessoirement pour le paramètre p , dans le contexte du modèle de Whittle probabilisé.

3.1 Définition et rôle de l'information de Fisher

Supposons que les données dont on dispose sont issues d'une loi \mathcal{L}_θ , dépendant d'un paramètre θ a priori inconnu. Que nous apprend en moyenne un échantillon de taille n sur ce paramètre ? Plus précisément, quelle quantité d'information l'échantillon fournit-il sur le paramètre θ ? C'est ce que la théorie de l'information au sens de Fisher permet de déterminer¹.

Pour être applicable, la théorie de l'information au sens de Fisher nécessite que la famille de lois générée par le paramètre vérifie 3 hypothèses. Désignons par Θ l'ensemble des valeurs que peut prendre θ et par Ω l'ensemble des valeurs possibles de la variable \mathbf{Z} lorsque θ décrit Θ . Désignons par $f_\theta(\mathbf{z})$ la densité de probabilité de \mathbf{Z} . Les hypothèses sont les suivantes :

1. $f_\theta(\mathbf{z}) > 0 \forall \mathbf{z} \in \Omega, \forall \theta \in \Theta$.
2. $f_\theta(\mathbf{z})$ est deux fois dérivable par rapport à θ .
3. Toute intégrale de la forme $\int f_\theta(\mathbf{z}) d\mathbf{z}$ donne le même résultat qu'elle soit dérivée deux fois par rapport à θ à l'intérieur de l'intégrale ou devant l'intégrale.

1. L'information de Fisher est généralement introduite dans tous les ouvrages de base en Statistiques qui touchent au problème de l'estimation. Pour la section 3.1, nous empruntons la présentation de Kauffmann (1994).

Ces hypothèses sont les *hypothèses de Fisher* qui nous permettent de donner la définition suivante.

Définition 1: Soit \mathbf{Z} une variable aléatoire dont la loi \mathcal{L}_θ satisfait aux hypothèses de Fisher et soit f_θ sa densité de probabilité. On appelle information au sens de Fisher fournie par la variable aléatoire \mathbf{Z} sur le paramètre θ le réel $J_{\mathbf{Z}}(\theta)$ défini comme

$$J_{\mathbf{Z}}(\theta) = \int_{\Omega} f_\theta(\mathbf{z}) \left(\frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{z}) \right)^2 d\mathbf{z}. \quad (3.1)$$

◇

La quantité (3.1) est conforme à ce que l'on est en droit d'attendre d'une mesure d'information : elle jouit de la propriété de *positivité* et de la propriété d'*additivité* :

- positivité : si la variable aléatoire \mathbf{Z} , de loi \mathcal{L}_θ , satisfait aux hypothèses de Fisher, alors $J_{\mathbf{Z}}(\theta) \geq 0$. Elle est nulle si et seulement si la loi de \mathbf{Z} ne dépend pas du paramètre θ .
- additivité : soient \mathbf{Z}_1 et \mathbf{Z}_2 deux variables aléatoires indépendantes, dont les lois respectives, \mathcal{L}_θ , \mathcal{K}_θ , dépendent d'un même paramètre θ . Si ces lois vérifient les hypothèses de Fisher, la loi du vecteur aléatoire $\vec{\mathbf{Z}} = (\mathbf{Z}_1, \mathbf{Z}_2)$ les vérifie également, et on a : $J_{\vec{\mathbf{Z}}}(\theta) = J_{\mathbf{Z}_1}(\theta) + J_{\mathbf{Z}_2}(\theta)$.

Il existe une inégalité remarquable, dite *de Frechet, Darmois, Cramer et Rao* (FDCR) relative à la variance d'estimateurs sans biais de $v(\theta)$ où v est une fonction continue de θ . Cette inégalité suppose comme condition que la loi \mathcal{L}_θ vérifie les hypothèses de Fisher et que l'information de Fisher $J_{\mathbf{Z}}(\theta)$ du paramètre θ satisfasse l'inégalité

$$J_{\mathbf{Z}}(\theta) > 0 \quad \forall \theta \in \Theta. \quad (3.2)$$

Les trois hypothèses de Fisher et la condition (3.2) comme hypothèse supplémentaire donnent les *hypothèses de Cramer et Rao* qui nous permettent de donner le théorème suivant.

Théorème 1: (Inégalité de FDCR) Soit \mathbf{Z} une variable de loi \mathcal{L}_θ satisfaisant aux hypothèses de Cramer et Rao. Soit f la densité de probabilité de \mathbf{Z} , et T un estimateur sans biais d'une fonction $v(\theta)$ de θ tel que l'intégrale $\int_{\Omega^k} T(\mathbf{z}) f_\theta(\mathbf{z}) d\mathbf{z}$ soit dérivable à l'intérieur. Alors, la fonction v est nécessairement dérivable, et l'on a :

$$\text{var}(T) \geq \frac{\left(\frac{\partial v}{\partial \theta} \right)^2(\theta)}{J_{\mathbf{Z}}(\theta)}. \quad (3.3)$$

◇

Définition 2: Si la loi \mathcal{L}_θ vérifie les hypothèses de Cramer et Rao, et si la fonction v est dérivable, on appelle borne de Frechet pour \mathcal{L}_θ et $v(\theta)$ le réel à droite de l'inégalité (3.3). ◇

Il est particulièrement intéressant de constater que cette borne dépend de l'information au sens de Fisher dont on dispose sur le paramètre θ , et qu'elle est d'autant plus petite que $J_{\mathbf{Z}}(\theta)$ est grand. Ce que nous dit l'inégalité FDCR est que l'incertitude associée à l'estimation de $\nu(\theta)$, mesurée par la variance d'un estimateur $\text{var}(T)$, peut être réduite seulement à hauteur de l'information dont on dispose. Un estimateur sans biais dont la variance atteint la borne de Fréchet pour tout $\theta \in \Theta$ est appelé un *estimateur efficace* pour l'espace des paramètres Θ . Un tel estimateur n'existe pas toujours.

Dans le cadre des modèles de Whittle probabilisés, nous adoptons $\mathbf{Z} := (\mathbf{X}, \mathbf{M})$ où \mathbf{X} est le champ spatial (2.1) et \mathbf{M} en est la matrice markovienne des interactions. Les paramètres à estimer sont ρ et p . Il nous intéresse d'analyser ces paramètres au travers de l'information au sens de Fisher et en particulier l'impact du graphe des interactions sur leur borne de Fréchet.

On relève que l'analyse qui suit part du principe que le graphe des interactions est observable, alors que nous désirons traiter principalement des situations où ce graphe n'est pas observable. Notre démarche est motivée par le désir de déterminer théoriquement l'information maximale sur ρ , en partant du principe que l'information de Fisher d'une statistique T de l'échantillon est toujours inférieure ou égale à celle apportée par l'échantillon (principe de la dégradation de l'information ; voir p. ex. Saporta, 1990). De fait

$$J_{\mathbf{X}} \leq J_{(\mathbf{X}, \mathbf{M})}. \quad (3.4)$$

L'information de Fisher des paramètres d'un modèle à partir de (\mathbf{X}, \mathbf{M}) peut servir de limite à l'information de Fisher de ces mêmes paramètres pour un estimateur sans biais ne s'appuyant que sur l'observation de \mathbf{X} (et de l'information *a priori*). A noter que cet estimateur peut ne pas exister.

3.2 Information de Fisher du coefficient d'autorégression ρ

3.2.1 Cas général

La mesure de probabilité d'une réalisation (\mathbf{x}, \mathbf{m}) du champ spatial \mathbf{X} conjointement à la matrice \mathbf{M} des interactions a été défini en section 2.1.1 et vaut $g(\mathbf{x}, \mathbf{m})d\mathbf{x} = \Pr(\mathbf{m})f(\mathbf{x}|\mathbf{m})d\mathbf{x}$, avec $f(\mathbf{x}|\mathbf{m})$ la densité de probabilité gaussienne (2.6), qui est la seule part de $g(\mathbf{x}, \mathbf{m})$ dépendante de ρ . Dans ce contexte, l'information de Fisher du coefficient d'autorégression ρ vaut :

$$J_{(\mathbf{X}, \mathbf{M})}(\rho) = E_{\mathbf{M}}[E_{\mathbf{X}}[(\frac{\partial}{\partial \rho} \ln(g(\mathbf{X}, \mathbf{M})))^2]] \quad (3.5)$$

$$= E_{\mathbf{M}}[E_{\mathbf{X}}[(\frac{\partial}{\partial \rho} \ln(f(\mathbf{X}|\mathbf{M})))^2]] \quad (3.6)$$

$$= E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(\rho)], \quad (3.7)$$

où $J_{\mathbf{X}|\mathbf{M}}(\rho)$ est l'information de Fisher de ρ contenu dans le champ \mathbf{X} conditionnellement à \mathbf{M} . En définissant la matrice \mathbf{B} comme

$$\mathbf{B} := \frac{\partial}{\partial \rho} \boldsymbol{\Sigma}^{-1} = -\frac{1}{\sigma^2}(\mathbf{m}' + \mathbf{m}) + \frac{2}{\sigma^2} \rho \mathbf{m}' \mathbf{m}, \quad (3.8)$$

où $\boldsymbol{\Sigma}$ est la matrice de variance-covariance (2.6), on obtient

$$J_{\mathbf{X}|\mathbf{m}}(\rho) = \frac{1}{2} \text{Tr}[(\mathbf{B}\boldsymbol{\Sigma})^2] + \frac{1}{4} (\text{Tr}[\mathbf{B}\boldsymbol{\Sigma}])^2. \quad (3.9)$$

Un développement de Taylor limité à l'ordre 2 de $J_{(\mathbf{X},\mathbf{M})}(\rho)$ au voisinage de $\rho = 0$ donne

$$J_{(\mathbf{X},\mathbf{M})}(\rho) = J_{(\mathbf{X},\mathbf{M})}(0) + \frac{\partial}{\partial \rho} J_{(\mathbf{X},\mathbf{M})}(0) \rho + O(\rho^2) \quad (3.10)$$

avec

$$J_{(\mathbf{X},\mathbf{M})}(0) = \sigma^4 \text{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}(\mathbf{M} + \mathbf{M}')]] \quad (3.11)$$

et

$$\frac{\partial}{\partial \rho} J_{(\mathbf{X},\mathbf{M})}(0) = 2 \text{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}^2(\mathbf{M} + \mathbf{M}')]], \quad (3.12)$$

où $\text{Tr}[\cdot]$ représente la trace de la matrice en argument. Généralement la quantité (3.12) n'est pas nulle. Elle est cependant nulle pour les réseaux qui ne permettent pas de faire de circuits de longueur de 3 arcs.

Nous illustrons le comportement de l'information de Fisher de ρ avec deux exemples de réseau : le réseau (régulier) carré et un réseau irrégulier par triangulation de Delaunay.

3.2.2 Application à un réseau carré

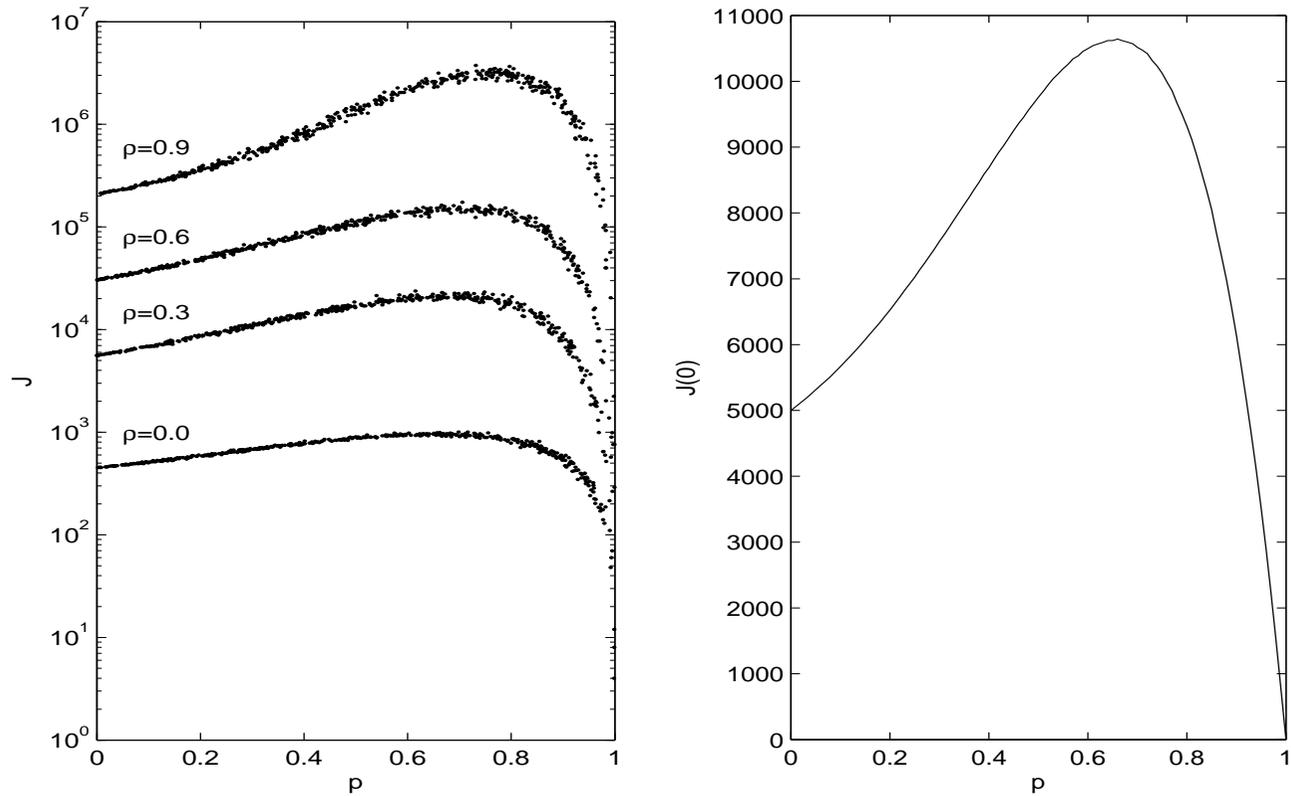
Pour un réseau carré, le terme (3.12) est nul. Par conséquent, l'information de Fisher conditionnelle à \mathbf{M} de ρ vaut

$$J_{\mathbf{X}|\mathbf{m}}(\rho) = J_{\mathbf{X}|\mathbf{m}}(0) + O(\rho^2). \quad (3.13)$$

En figure 3.1(a), nous illustrons, pour quelques valeurs de ρ , la dispersion de la variable $J_{\mathbf{X}|\mathbf{M}}(\rho)$ pour un modèle de Whittle markovien probabilisé ayant pour graphe de structure des interactions un réseau carré parfait de taille 100×100 sites (avec $\sigma^2 = 1$). On observe que la fonction $J_{(\mathbf{X},\mathbf{M})}(\rho)$ croît avec ρ (à p constant) et un fait remarquable : à ρ constant, la fonction $J_{(\mathbf{X},\mathbf{M})}(\rho)$ possède un maximum à une valeur de p largement plus grande que 0. En l'occurrence, dans notre exemple, le maximum est au voisinage de $p = 0.65$.

La dépendance non monotone avec p de $J_{(\mathbf{X},\mathbf{M})}(\rho)$ est fonction en première approximation de statistiques sur la connexité du graphe aléatoire des interactions.

Le comportement de la quantité (3.11) en fonction du paramètre p est montré dans la figure 3.1(b). En comparaison avec le diagramme de dispersion de la figure 3.1(a), on déduit que $J(0)$ affecte fortement la quantité $J_{(\mathbf{X},\mathbf{M})}(\rho)$ sur un large



(a) Valeurs de $J_{\mathbf{X}|\mathbf{m}}$ pour différentes réalisations \mathbf{m} de la matrice \mathbf{M} des interactions d'un modèle de Whittle markovien probabilisé dont le paramètre p prend des valeurs aléatoires (uniformément sur l'intervalle $(0, 1)$) et pour des valeurs de ρ indiquées sur le diagramme. Les résultats sont obtenus avec 500 réalisations du modèle spatial sur réseau carré périodique de 100×100 sites pour chaque valeur de ρ .

(b) Valeurs de $J_{\mathbf{X}|\mathbf{M}}(0)$ en fonction de p pour un modèle de Whittle markovien probabilisé. La courbe est calculée sur un échantillon de 10^4 réalisations du modèle spatial sur réseau carré périodique de 100×100 sites.

Fig. 3.1: Information de Fisher de ρ

intervalle de valeurs de ρ . Pour conclure à la dépendance en première approximation de $J_{(\mathbf{x}, \mathbf{M})}(\rho)$ (au sens de son développement de Taylor) avec des statistiques de la connexité du graphe aléatoire des interactions, il suffit d'observer que

$$J_{(\mathbf{x}, \mathbf{M})}(0) = q_1 + q_2, \quad (3.14)$$

avec

$$q_1 = \sigma^4 \mathbb{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}'\mathbf{M}]] = \sigma^4 \sum_{i=1}^4 \frac{\mathbb{E}_{\mathbf{M}}[n_i]}{i}, \quad (3.15)$$

$$q_2 = \sigma^4 \mathbb{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}^2]] = \sigma^4 \sum_{i=1}^4 \sum_{j=1}^4 \frac{\mathbb{E}_{\mathbf{M}}[n_{ij}]}{ij}, \quad (3.16)$$

où n_i est le nombre de variables du champ de connectivité locale i , c'est-à-dire interagissant avec i variables, et n_{ij} est le nombre d'interactions allant d'une variable du champ de connectivité locale i vers une variable du champ de connectivité locale j . Les expressions (3.14), avec (3.15) et (3.16), démontrent clairement que les variables de plus faible connectivité locale contribuent individuellement le plus à l'information de Fisher. De même, les paires de variables en interaction dont le produit des connectivités locales est le plus faible contribuent individuellement le plus à l'information de Fisher. Les figures 3.2 et 3.3 illustrent la courbe des éléments contributifs aux sommes dans (3.15) et (3.16) respectivement en fonction du paramètre p .

Comme on peut le voir dans la figure 3.2(a), l'effectif moyen des variables de degré nul croît avec p . Or, il convient de rappeler que les réalisations de ces variables sont des éléments de nuisance spontanés pour l'estimation de ρ , car ils ne contribuent pas, conditionnellement au graphe des interactions, à l'information de ρ . Ainsi donc, l'estimation de ρ ne peut être effectivement obtenue que d'un ensemble (variable) de variables dont le nombre moyen est une fonction monotone décroissante de p . Malgré cette tendance, l'information de Fisher moyenne croît lorsque p croît dans un voisinage large de 0. A p suffisamment grand (proche de $p = 1$), les sous-graphes (composantes) connexes de 2 sites, respectivement 1 site, qui naissent de la destruction du réseau sont les structures qui deviennent prépondérantes par leur nombre et par leur capacité, respectivement leur incapacité, précédemment décrite à informer sur ρ . Il est manifeste que l'information de Fisher de ρ est très sensible à leur nombre.

La courbe de $J_{(\mathbf{x}, \mathbf{M})}(0)$, en tant que première approximation de la courbe de $J_{(\mathbf{x}, \mathbf{M})}(\rho)$ au voisinage de $\rho = 0$, laisse très clairement comprendre que la connexité de la structure en réseau carré ne permet pas à un estimateur efficace de ρ au voisinage de $\rho = 0$ d'atteindre la meilleure efficacité absolue que permettrait en moyenne un réseau carré détruit avec une certaine probabilité à l'intérieur de l'intervalle $(0, 1)$. Si on désigne par p^* la valeur de p à l'extrémum de $J_{(\mathbf{x}, \mathbf{M})}(0)$, on peut donc à juste titre parler de graphe des interactions en moyenne trop connectif si $p < p^*$ et de trop peu connectif si $p > p^*$ à l'égard de l'efficacité d'un estimateur efficace de ρ .

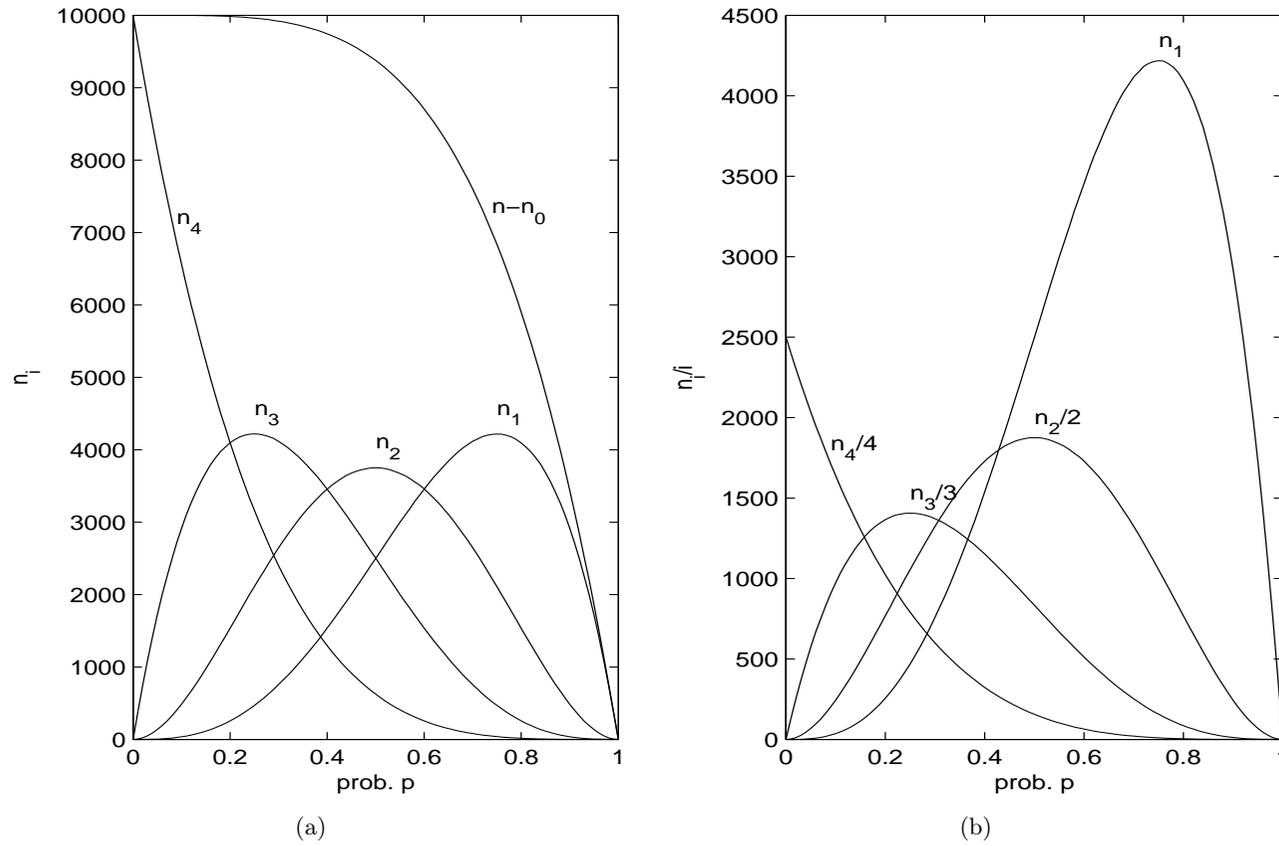


Fig. 3.2: Estimation de (en fonction de p) : (a) $E_{\mathbf{M}}[n_i]$; (b) $\frac{1}{i} E_{\mathbf{M}}[n_i]$. L'estimation de $E_{\mathbf{M}}[n_i]$ est notée ici par n_i . Les résultats sont des moyennes obtenues sur 10^3 réalisations d'un réseau carré périodique de taille 100×100 .

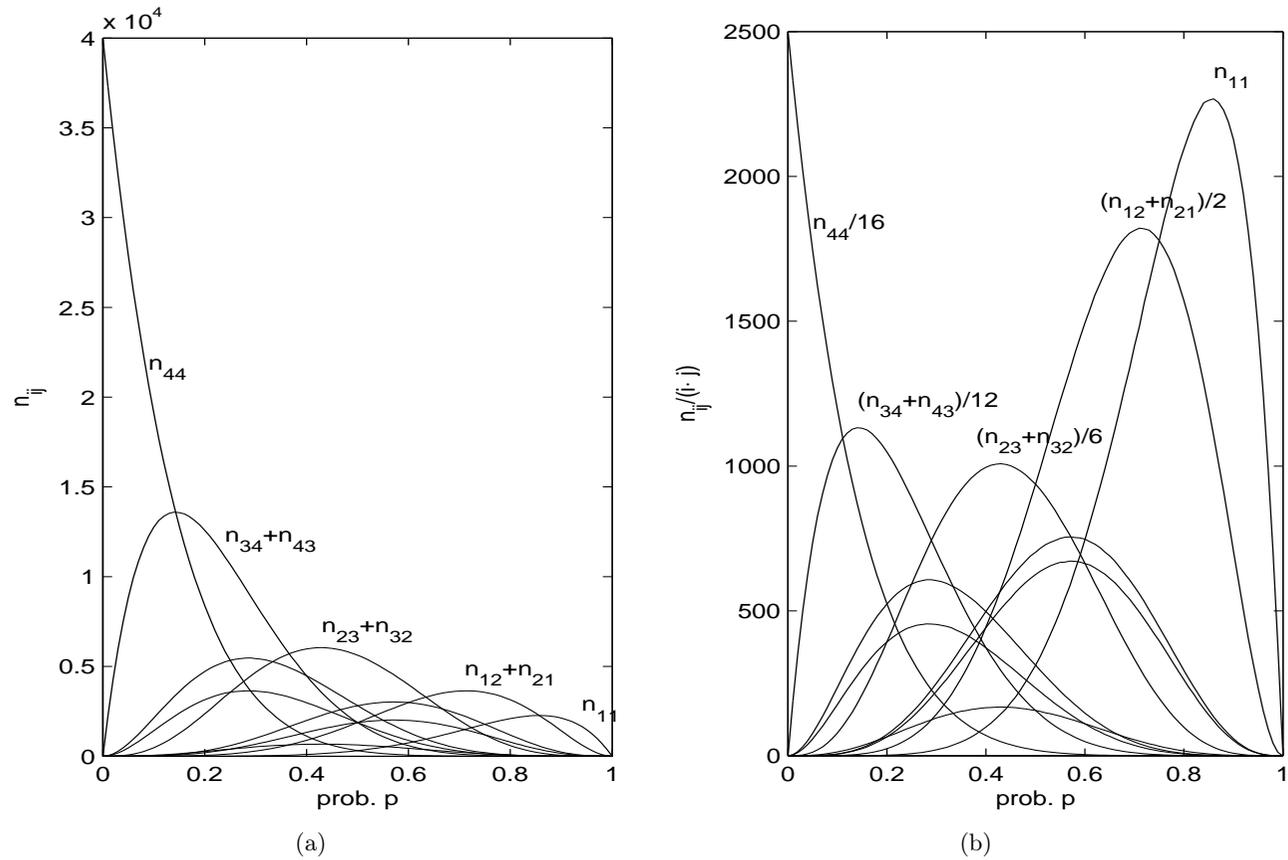


Fig. 3.3: Estimation de (en fonction de p) : (a) $E_M[n_{ii}]$ et $E_M[n_{ij} + n_{ji}]$; (b) $\frac{1}{i^2} E_M[n_{ii}]$ et $\frac{1}{ij} E_M[n_{ij} + n_{ji}]$. L'estimation de $E_M[n_{ij}]$ est notée ici par n_{ij} . Les résultats sont des moyennes obtenues sur 10^3 réalisations d'un réseau carré périodique de taille 100×100 .

Remarque Le fait que le champ ne soit pas le plus informatif sur ρ aux conditions où $p = 0$ (connectivité globale maximale) est induit par la normalisation imposée des lignes de la matrice des interactions \mathbf{M} . En effet, si dans le modèle spatial on substitue \mathbf{M} par la matrice d'adjacence du graphe des interactions, l'information de Fisher de ρ à son origine serait alors proportionnelle à $1 - p$ et donc serait la plus grande à $p = 0$.

3.2.3 Application à une triangulation de Delaunay sur un semis de points

Notre modèle de Whittle probabilisé (et son cas particulier le modèle de Whittle classique) peut également appréhender les variables qui ne sont pas régulièrement espacées dans l'espace, à supposer qu'elles interagissent entre elles selon un réseau de proche voisinage et que le champ qu'elles forment est autorégressif. Parmi les structures d'interactions possibles, nous songeons à la triangulation de Delaunay² des positions des variables du champ. Cependant, partant de l'hypothèse tout à fait admissible que certaines variables en relation géométrique par la triangulation sont en fait trop distantes l'une de l'autre pour être véritablement en interaction, nous fixons un écartement seuil des positions des variables au-delà duquel la probabilité d'une interaction est nulle. Le graphe des interactions est donc construit en deux étapes :

- (a) une triangulation complète est effectuée sur le semis de points représentant la position des variables,
- (b) les arêtes du graphe reliant deux points écartés d'une distance supérieure à l'écart seuil d_{lim} sont supprimées.

Le graphe résultant est considéré comme le graphe de structure des interactions du modèle de Whittle probabilisé. A titre de démonstration, nous générons un semis de points de manière aléatoire (la position en abscisse et la position en ordonnée étant des variables de densité de probabilité uniforme) et déterminons pour ces points le graphe de la triangulation complète. Nous notons \mathbf{S} sa matrice d'adjacence. Nous déterminons ensuite 4 écarts seuil d_{lim} dans un ordre croissant, respectivement de valeur d_1 , d_2 , d_3 et d_4 , pour construire 4 graphes de structure des interactions sur la base de \mathbf{S} avec le modus operandi qui a été énoncé. En notant les matrices d'adjacence de ces graphes \mathbf{S}_1 , \mathbf{S}_2 , \mathbf{S}_3 et \mathbf{S}_4 , les distances d_i choisies sont celles qui permettent d'obtenir le rapport q_i suivant

$$q_i := \frac{\tau[\mathbf{S}_i]}{\tau[\mathbf{S}]} = \frac{i}{4}, \quad (3.17)$$

où $\tau[\]$ est un opérateur qui compte le nombre d'éléments non nuls du vecteur ou matrice en argument. Les graphes résultants, qui s'emboîtent les uns dans les autres, sont reproduits dans la figure 3.4. Les courbes estimées de l'information

². Justifié par le fait que la triangulation de Delaunay est le graphe dual du pavage de Voronoï. Le réseau carré parfait est une triangulation de Delaunay particulière.

de Fisher (3.11) associées à ces graphes de structure sont présentées dans la figure 3.5. Les estimations ont été effectuées par simulation de Monte Carlo générant 10^3 réalisations du graphe des interactions par structure et par valeur de p allant de 0 à 1 avec un pas de 0.01.

Il apparaît que $J_{\mathbf{X},\mathbf{M}}(0)$ est sensible à la connexité du graphe de structure. Lorsque cette dernière croît en nombre d'interactions, elle affecte l'extrémum de la courbe qui se décale vers des valeurs de p plus élevées, l'amplitude de la courbe variant relativement peu. Ce décalage induit des différences dans l'efficacité de l'estimateur efficace de ρ au voisinage de $\rho = 0$. On peut notamment contempler ces différences à $p = 0$ qui correspond à des modèles de Whittle classique. Les courbes de la figure 3.5 montrent des situations où l'efficacité de l'estimateur efficace de ρ au voisinage de $\rho = 0$ est plus élevée dans l'hypothèse d'un modèle de Whittle classique que dans l'hypothèse du modèle de Whittle probabilisée, mais elles montrent aussi des situations inverses.

En notant par p_i^* la valeur du paramètre p à l'extrémum de la courbe $J_{\mathbf{X},\mathbf{M}}(0)$ associée à la matrice de structure \mathbf{S}_i , on relève qu'elle doit satisfaire la relation

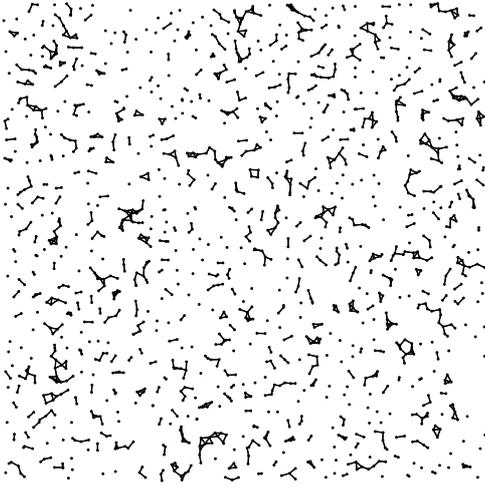
$$(1 - p_i^*)q_i = \frac{E[\tau[\mathbf{M}_i]]}{\tau[\mathbf{S}]}, \quad (3.18)$$

où \mathbf{M}_i désigne la matrice de Markov aléatoire associée à la matrice de structure des interactions \mathbf{S}_i . L'expression à droite de l'équation (3.18) porte au numérateur le nombre moyen d'interactions du champ spatial associé à la matrice de structure \mathbf{S}_i . Dans notre cas, il est observé expérimentalement que ce nombre moyen représente le 25% de la triangulation de Delaunay complète du semis de points, indépendamment de la matrice de structure \mathbf{S}_i , comme le montre le tableau 3.1 se rapportant aux courbes J_1 , J_2 , J_3 et J_4 de la figure 3.4. La quantité 3.18 n'est par contre pas invariante de la configuration du *semis de points*, comme en témoigne la courbe de comparaison associée à un réseau carré dans la figure 3.5.

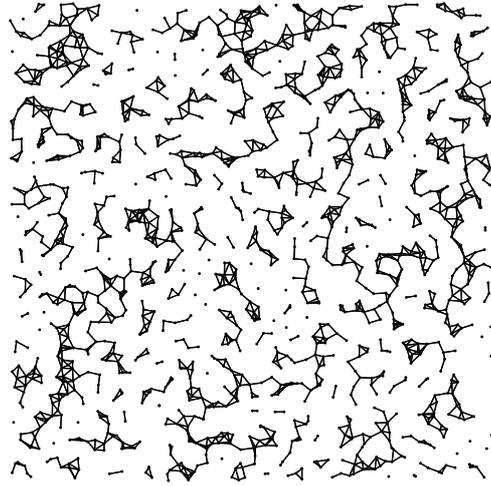
Courbe	Amplitude	q_i	p_i^*	$q_i(1 - p_i^*)$
J_1	1650	0.25	0	0.25
J_2	1589	0.50	0.53	0.24
J_3	1596	0.75	0.68	0.24
J_4	1576	1.00	0.77	0.23

Tab. 3.1: Produit $(1 - p_i^*)q_i$ par matrice de structure \mathbf{S}_i

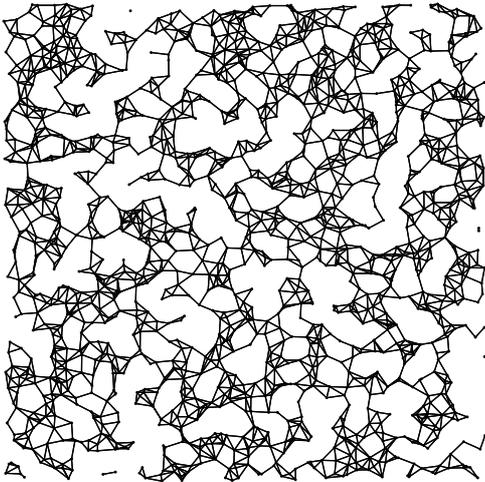
On conclut que, à semis de points donné, le champ possède une condition optimale pour l'estimation de ρ qui s'explique par le mécanisme d'augmentation-diminution de l'information décrit précédemment dans le cas du réseau carré : la connexité du champ est suffisamment faible pour bénéficier du mécanisme d'augmentation de l'information de Fisher par baisse de connectivité, sans être dominé par l'effet statistique des sites de connectivité locale nulle (non informatifs sur ρ).



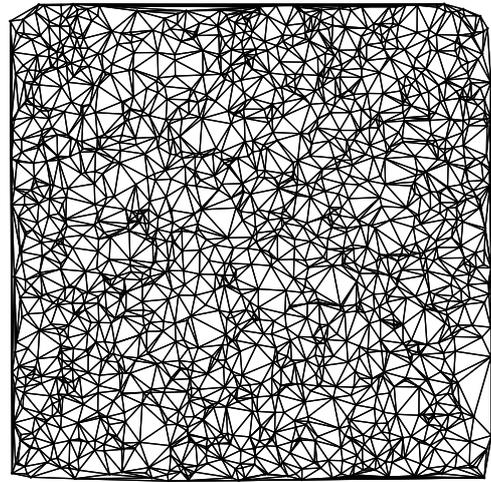
(a) Graphe de structure des interactions associé à $d_{\text{lim.}} = d_1$



(b) Graphe de structure des interactions associé à $d_{\text{lim.}} = d_2$



(c) Graphe de structure des interactions associé à $d_{\text{lim.}} = d_3$



(d) Graphe de structure des interactions associé à $d_{\text{lim.}} = d_4$

Fig. 3.4: Graphes de structure des interactions du modèle de Whittle probabilisé, sur un même semis de 2500 sites, construits comme indiqué dans le texte avec les distances seuils $d_{\text{lim.}}$ indiqués.

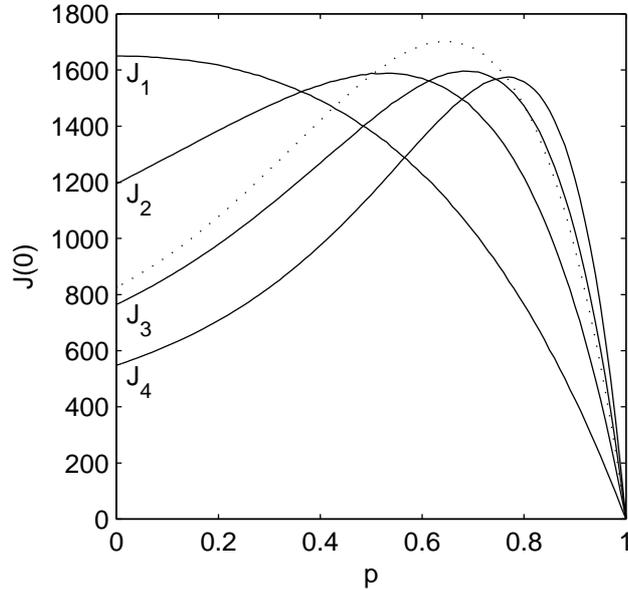


Fig. 3.5: Courbes estimées J_1 , J_2 , J_3 et J_4 de l'information de Fisher $J_{\mathbf{X},\mathbf{M}}(0)$ d'un modèle de Whittle markovien probabilisé de paramètre p et respectivement de graphe de structure d'interactions de la figure 3.4(a), 3.4(b), 3.4(c) et 3.4(d). La courbe en pointillés représente l'information de Fisher dans le cas d'un réseau carré fini avec le même nombre de sites.

Le graphe de la figure 3.4(a) donne une image de la condition optimale dont on relève que le nombre de sites de connectivité locale nulle est passablement élevé.

Parallèle dans le paradigme bayésien

En faisant l'hypothèse que le champ spatial étudié suit un modèle de Whittle classique dont le graphe des interactions est inconnu mais possède une probabilité identique à celle du graphe des interactions du modèle de Whittle probabilisé, la quantité $J_{(\mathbf{X},\mathbf{M})}(0)$ n'apparaît pas comme une quantité utile pour le modèle. Ce qui intéresse dans ce contexte est plutôt la dispersion de $J_{\mathbf{X}|\mathbf{M}}(0)$ autour de sa moyenne. Nous mesurons la dispersion de l'information de Fisher conditionnelle $J_{\mathbf{X}|\mathbf{M}}$ en calculant un intervalle inter-quantile de $J_{\mathbf{X}|\mathbf{M}}$ à des valeurs de p régulièrement espacées dans l'intervalle $[0, 1]$. Cet intervalle forme un intervalle de confiance, fixé à 95% dans notre cas. La figure 3.6 reproduit les courbes de cet indicateur de dispersion pour les 4 graphes de structure. Elles donnent les limites (inférieures et supérieures) à l'information de Fisher de ρ avec une confiance de 95% compte tenu de la probabilité a priori sur la matrice des interactions. Comme on peut le constater, ces limites suivent (à peu de différence près) la forme de croissance et décroissance de l'information de Fisher de ρ en fonction de p . Par la borne

de Fréchet, ces limites donnent aussi celles de la précision d'un estimateur sans biais efficace de ρ (s'il existe) dans l'incertitude du graphe des interactions. Cette réflexion pouvait être menée pour le cas des réseaux carrés (cf. figure 3.1(a)).

3.2.4 Application sur un cas concret : les comtés d'Irlande

Il est toujours possible d'appliquer l'analyse de l'information de Fisher pour des semis de points non uniformément répartis sur l'espace ou pour des unités spatiales irrégulièrement ordonnancées sur l'espace. Dans ce registre, on peut approcher le cas des unités administratives d'une région, d'une province, d'un pays, etc. par leur graphe de contiguïté. A titre d'exemple, on peut investiguer le graphe de contiguïté des comtés d'Irlande employé par Geary (1954), à l'origine de sa présentation de la statistique désormais connue sous le nom de c de Geary³. Le graphe, tabulé dans Cliff & Ord (1973, p. 53), est représenté en figure 3.7(a). Ce graphe exclut le comté de Dublin et met en lien les comtés de Karry et Clare malgré le bras de mer les séparant. Le graphe comporte 25 sites non isolés, autant qu'un réseau carré de taille 5×5 sites.

La figure 3.7(b) présente l'information de Fisher $J_{(\mathbf{x}, \mathbf{M})}(0)$ résultante. Elle permet de constater que la courbe d'information de Fisher basée sur la configuration des comtés de l'Irlande a un comportement avec p qui diffère de ce que l'on peut voir d'un réseau carré parfait composé des 25 unités spatiales. Les caractéristiques propres à chaque courbe sont données dans le tableau ci-dessous (dont notamment la valeur p^* du maximum de l'information). Le tableau permet de relever que le nombre moyen optimal d'interactions pour l'information de Fisher de ρ est le même pour les deux graphes de structure (cf. dernière colonne).

Interprétées à l'aide de la borne de Fréchet, ces courbes indiquent que, à coefficient d'autorégression ρ donné (suffisamment faible), dans un large intervalle de valeurs de p au voisinage de 0, un estimateur sans biais efficace sera toujours moins précis dans l'estimation de ρ avec la disposition actuelle des comtés de l'Irlande que ne permettrait un pays imaginaire dont les comtés ont pour graphe de contiguïté celui d'un réseau carré, à supposer bien évidemment que le champ étudié cadre avec un modèle de Whittle probabilisé. On envisage que cette variation de précision, d'une structure de contiguïté à une autre, peut affecter les conclusions de recherches comparatives entre régions, pays, etc. portant sur le coefficient d'autorégression ρ . De ce constat (et de ceux des applications précédentes) naît ce que nous pourrions appeler une "injustice de la structure" (des interactions).

Graphe de structure	Nbre de contiguïtés nc	p^*	$J_{(\mathbf{x}, \mathbf{M})}(0)$ à $p = p^*$	$J_{(\mathbf{x}, \mathbf{M})}(0)$ à $p = 0$	$nc(1 - p^*)$
Irlande	55	0.68	24.5	12.6	17.6
Carré 5×5	40	0.56	27.1	16.2	17.6

3. Ce graphe accompagne un jeu de données (par ex. le nombre de vaches laitières par unité de surface de pâturage et le pourcentage de citadins ou villageois par rapport à la population totale de 1951) dont le tout est devenu classique pour avoir été pris et repris par divers auteurs.

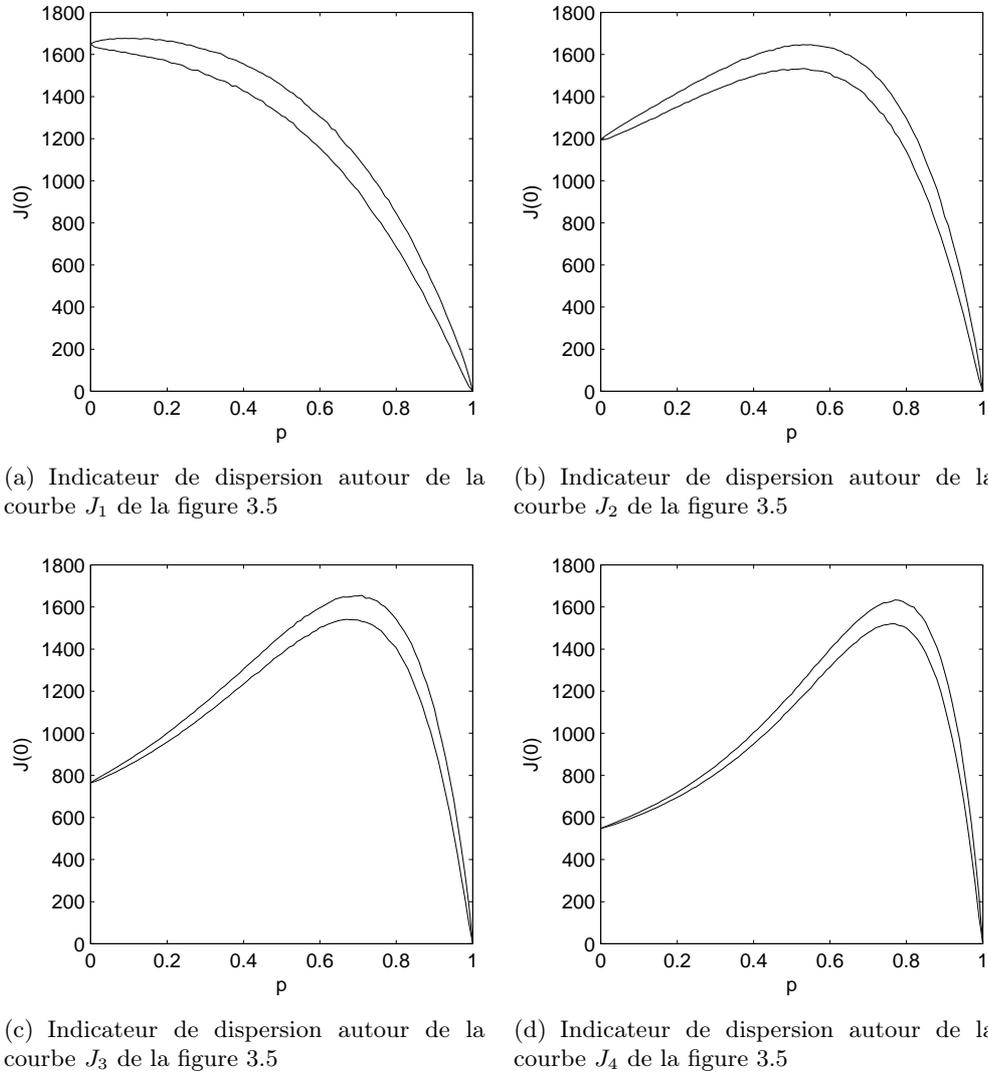
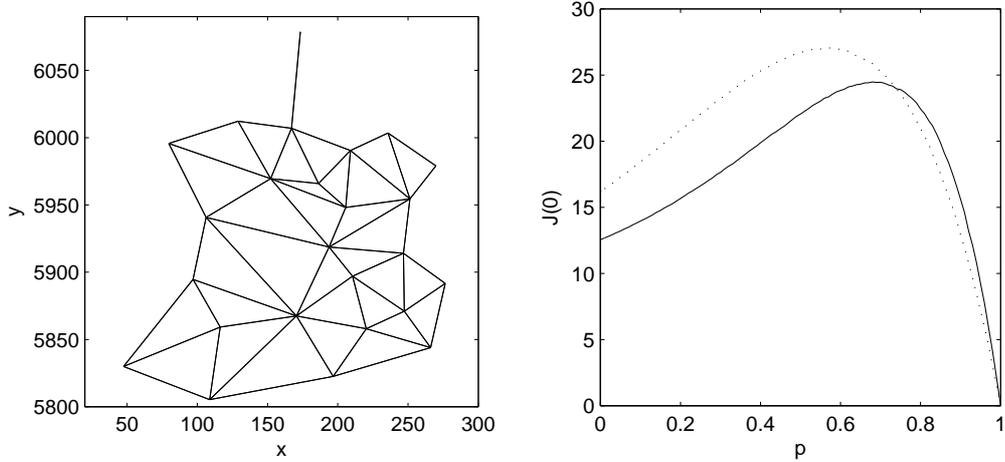


Fig. 3.6: Mesure de la dispersion de $J_{\mathbf{X}|\mathbf{M}}$ estimée de la manière suivante : pour chaque valeur de p on détermine la fonction de répartition empirique F_p des valeurs $J_{\mathbf{X}|\mathbf{M}}$ obtenus par simulation de Monte Carlo, puis on calcule les quantiles $J_{0.025}$ et $J_{0.975}$ tels que $F_p(J_{0.025}) = 0.025$ et $F_p(J_{0.975}) = 0.975$. L'écart entre ces deux quantiles contient par définition 95% des valeurs d'information de Fisher conditionnel produites par la simulation de Monte Carlo, à valeur p donnée. Chaque fonction de répartition est calculée sur 1000 individus.



(a) Graphe de contiguïté des comtés d'Irlande. (b) Information de Fisher $J_{\mathbf{X},\mathbf{M}}(0)$ associé au graphe de gauche (trait plein) et à un réseau carré de 5×5 sites (trait pointillé).

Fig. 3.7:

3.3 Information de Fisher du paramètre p

L'information de Fisher du paramètre p pour un modèle de Whittle probabilisé se calcule en observant tout d'abord que le paramètre ne figure que dans l'expression de la probabilité de la matrice \mathbf{M} . Dans ce contexte, l'information de Fisher $J_{(\mathbf{X},\mathbf{M})}(p)$ vaut

$$J_{(\mathbf{X},\mathbf{M})}(p) = \mathbf{E}_{\mathbf{M}} \left[\mathbf{E}_{\mathbf{X}} \left[\left(\frac{\partial}{\partial p} \ln(g(\mathbf{X}, \mathbf{M})) \right)^2 \right] \right] \quad (3.19)$$

$$= \mathbf{E}_{\mathbf{M}} \left[\mathbf{E}_{\mathbf{X}} \left[\left(\frac{\partial}{\partial p} \ln(\Pr(\mathbf{M})) \right)^2 \right] \right] \quad (3.20)$$

$$= \mathbf{E}_{\mathbf{M}} \left[\left(\frac{\tau[\mathbf{S}](1-p) - \tau[\mathbf{M}]}{2p(1-p)} \right)^2 \right] \quad (3.21)$$

$$= \frac{(1-p)^2(\tau[\mathbf{S}])^2 - 2\tau[\mathbf{S}]\mu_1(1-p) + \mu_2}{4p^2(1-p)^2} \quad (3.22)$$

$$= \frac{\tau[\mathbf{S}]}{4(1-p)p}, \quad (3.23)$$

où $\tau[\cdot]$ est l'opérateur précédemment introduit pour l'expression (2.5) de la probabilité de \mathbf{M} , et μ_1 et μ_2 sont respectivement le premier et deuxième moment non centré de $\tau[\mathbf{M}]$ (que l'on peut facilement calculer en faisant recours à des résultats

des distributions binomiales) :

$$\mu_1 := \mathbf{E}_{\mathbf{M}}[\tau[\mathbf{M}]] = \tau[\mathbf{S}](1 - p), \quad (3.24)$$

$$\mu_2 := \mathbf{E}_{\mathbf{M}}[(\tau[\mathbf{M}])^2] = \tau[\mathbf{S}](1 - p)p + (\tau[\mathbf{S}])^2(1 - p)^2. \quad (3.25)$$

Il en résulte que l'information de Fisher de p est indépendante de ρ et σ^2 , elle diverge aux valeurs extrêmes de p et est minimale à $p = \frac{1}{2}$ indépendamment de \mathbf{S} .

3.4 Divergence de Kullback-Leibler

La divergence de Kullback-Leibler (voir annexe du chapitre) est une fonction bien connue en théorie de l'information pour mesurer l'écart relatif entre deux fonctions de densité de probabilité. Dans le cas où ces dernières sont déterminées respectivement par la mesure de probabilité $g_0(\mathbf{x}, \mathbf{m})d\mathbf{x}$ du modèle de Whittle définie par (2.7) aux conditions où $\rho = 0$ (modèle d'indépendance spatiale) et la mesure de probabilité $g_1(\mathbf{x}, \mathbf{m})d\mathbf{x}$ de ce même modèle aux conditions où ρ prend une valeur ρ_1 non nulle, la divergence de Kullback-Leibler résultante, notée par $K(g_0, g_1)$, vaut (Kullback, 1959, chap. 9) :

$$K(g_0, g_1) = \sum_{\mathbf{m}} \left(\int_{-\infty}^{\infty} g_0(\mathbf{x}, \mathbf{m}) \ln \left(\frac{g_0(\mathbf{x}, \mathbf{m})}{g_1(\mathbf{x}, \mathbf{m})} \right) d\mathbf{x} \right) \quad (3.26)$$

$$= \mathbf{E}_{\mathbf{M}} \left[-\frac{1}{2} \ln \left(\frac{\det[\boldsymbol{\Sigma}_1^{-1}]}{\det[\boldsymbol{\Sigma}_0^{-1}]} \right) - \frac{n}{2} + \frac{1}{2} \text{Tr}[\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1}] \right], \quad (3.27)$$

où $\boldsymbol{\Sigma}_0$ et $\boldsymbol{\Sigma}_1$ sont les matrices de variance-covariance du champ \mathbf{X} sous l'hypothèse $\rho = 0$ et $\rho = \rho_1$ respectivement.

La quantité en argument de $\mathbf{E}_{\mathbf{M}}[\cdot]$ dans (3.27) peut être exprimée en une série de Taylor développée autour de $\rho = 0$ pour donner (voir annexe de ce chapitre) :

$$K(g_0, g_1) = \frac{1}{2} \mathbf{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2]] \rho^2 + \sum_{i=3}^{\infty} \frac{1}{i} \mathbf{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}^i]] \rho^i \quad (3.28)$$

qui est convergente si $|\rho| < 1$. En comparaison avec (3.11), on obtient la relation

$$K(g_0, g_1) = \frac{1}{2} J_{(\mathbf{x}, \mathbf{M})}(0) \rho^2 + O(\rho^3). \quad (3.29)$$

La relation différentielle entre la divergence de Kullback-Leibler et l'information de Fisher est en fait un résultat général⁴ (Kullback-Leibler, 1959, chap. 2).

La divergence est en première approximation (au voisinage de $\rho = 0$) en relation linéaire avec ρ^2 avec une pente proportionnelle à l'information de Fisher $J_{(\mathbf{x}, \mathbf{M})}(0)$. Ainsi donc, la divergence hérite des divers comportements possibles de $J_{(\mathbf{x}, \mathbf{M})}(0)$ observés précédemment, indépendamment de ρ . Notamment, à coefficient d'auto-régression ρ et à graphe de structure des interactions fixés, la fonction de densité de probabilité g_1 peut diverger le plus de g_0 dans des conditions où $p > 0$.

4. A remarquer que dans ce cas particulier, la trace d'une puissance impaire de \mathbf{M} est nulle, de sorte que nous obtenons la relation $K(g_0, g_1) = \frac{1}{2} J_{(\mathbf{x}, \mathbf{M})}(0) \rho^2 + O(\rho^4)$.

3.5 Conclusion

L'analyse du modèle de Whittle probabilisé par l'information de Fisher est fructueuse. Elle permet d'établir les quelques éléments ci-dessous.

- Le changement de paramètre p affecte le modèle de connexité du champ spatial qui à son tour affecte l'information de Fisher de ρ . Cette information ne possède pas nécessairement un maximum à $p = 0$ (pour un réseau carré de taille 100×100 , par exemple, le maximum se trouve autour de $p = 0.65$). Cette valeur optimale peut par contre varier avec la taille et la forme du graphe de structure. Elle ne coïncide généralement pas avec celle de l'information de Fisher du paramètre p dont elle est égale 0.5 invariablement de la structure des interactions.
- L'optimalité relevée ci-dessus est une propriété de l'échantillon et non pas de la statistique de cet échantillon. Dans ce sens, nous dénonçons une *injustice de la structure* : en premier lieu, la précision de l'estimation possède une limite théorique qui dépend intrinsèquement du réseau et non de l'Observateur. Ce dernier peut à son tour dégrader involontairement l'information de l'échantillon par le choix de sa statistique et/ou à cause de l'inobservation des interactions.
- Les mécanismes antagonistes qui contribuent à l'existence d'une valeur optimale de p pour l'information de Fisher de ρ ont été identifiés pour notre modèle de connexité du champ : il s'agit d'une part d'un effet de réduction de la connexité avec émergence de petites cliques très informatives sur ρ , dont en particulier les cliques à 2 sites qui sont les plus informatives, et d'autre part d'un effet de croissance du nombre de sites de connectivité locale nulle. Corollairement, nous avons montré l'effet produit sur l'estimation de ρ lorsque les unités spatiales de connectivité locale nulle conditionnellement au graphe des interactions sont involontairement laissées dans l'échantillon (par ignorance du graphe des interactions qui s'est réalisé).
- La divergence de Kullback-Leibler entre deux modèles de Whittle probabilisés ayant une même matrice de structure des interactions mais dont pour l'un $\rho = 0$ (modèle d'indépendance spatiale) et pour l'autre $\rho = \rho_1 > 0$ (modèle de Whittle probabilisé) indique que ceux-ci divergent proportionnellement à l'information de Fisher de ρ en $\rho = 0$ et proportionnellement à ρ_1^2 . Dans le cas où ces deux modèles sont des hypothèses sur le champ spatial étudié, on envisage que l'efficacité à sélectionner celle qui décrit le mieux les observations du champ spatial soit affectée par la valeur du paramètre p (cf. Chapitre 5).

Chapitre 4

Estimation ponctuelle du coefficient d'autorégression

4.1 Estimation de MV du coefficient d'autorégression

4.1.1 Fonction de vraisemblance

Les paramètres du modèle de Whittle markovien probabilisé peuvent être estimés par leur estimateur *de maximum de vraisemblance* (MV) : ils sont les valeurs de p , ρ et σ^2 qui rendent maximale la densité de probabilité découlant de (2.7) considérée dans ce cas comme une fonction des paramètres pour une réalisation conjointe (\mathbf{x}, \mathbf{m}) du champ \mathbf{X} et de la matrice aléatoire des interactions \mathbf{M} . Cette fonction est la *fonction de vraisemblance*, que nous notons $\mathcal{L}(\rho, \sigma^2 | \mathbf{x}, \mathbf{m})$. Son logarithme prend la forme suivante :

$$\ln \mathcal{L}(p, \rho, \sigma^2 | \mathbf{x}, \mathbf{m}) = \ln \mathcal{L}_1(\rho, \sigma^2 | \mathbf{x}, \mathbf{m}) + \ln \mathcal{L}_2(p | \mathbf{m}), \quad (4.1)$$

avec

$$\ln \mathcal{L}_1(\rho, \sigma^2 | \mathbf{x}, \mathbf{m}) = -\frac{n}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \ln(\det[\mathbf{V}^{-1}]) - \frac{1}{2\sigma^2} \mathbf{x}' \mathbf{V}^{-1} \mathbf{x}, \quad (4.2)$$

$$\ln \mathcal{L}_2(p | \mathbf{m}) = \ln(h(\mathbf{m} | p)), \quad (4.3)$$

où $h(\mathbf{m} | p)$ est la probabilité de \mathbf{m} conditionnelle à p , et \mathbf{V}^{-1} est la matrice de taille $n \times n$ définie comme

$$\mathbf{V}^{-1} := (\mathbf{I} - \rho \mathbf{m})' (\mathbf{I} - \rho \mathbf{m}). \quad (4.4)$$

La fonction (4.1) peut être réduite en substituant σ^2 par son estimateur de maximum de vraisemblance $\hat{\sigma}^2$, de la forme

$$\hat{\sigma}^2 = \frac{1}{n} \mathbf{x}' \mathbf{V}^{-1} \mathbf{x}. \quad (4.5)$$

Dans ce cas, l'estimateur de maximum de vraisemblance du coefficient d'autorégression du modèle spatial est la solution (p, ρ) qui rend maximale la fonction de vraisemblance réduite $\mathcal{L}(p, \rho, \hat{\sigma}^2 | \mathbf{x}, \mathbf{m})$.

Remarque En pratique, nous employons la décomposition LU pour calculer le déterminant de \mathbf{V}^{-1} . Il est à relever que le calcul de déterminants est un problème reconnu comme difficile dans l'analyse numérique lorsqu'il s'agit de champs de grande taille. Dans ce cas, l'estimateur de maximum de vraisemblance de ρ devrait être calculé en employant certaines techniques d'approximation comme par exemple celles de Smirnov et al. (2001), Griffith (2004) ou Celik et al. (2007).

4.1.2 Estimation de MV du paramètre p

Etant donnée la décomposition (4.1), l'estimateur de maximum de vraisemblance \hat{p} de p peut être obtenue à partir la fonction de vraisemblance $\mathcal{L}_2(p|\mathbf{x}, \mathbf{m})$. Cet estimateur prend la forme

$$\hat{p} = \frac{\tau[\mathbf{S}] - \tau[\mathbf{m}]}{\tau[\mathbf{S}]}, \quad (4.6)$$

où \mathbf{S} est la matrice de structure des interactions du champ et où $\tau[\cdot]$ est l'opérateur totalisant le nombre d'éléments non nuls de la matrice donnée en argument, comme précédemment introduit. Il en découle qu'il est difficile d'estimer p par la méthode MV exposée sans connaître la réalisation du graphe des interactions, même si p ne dépend que du nombre des interactions et non pas de la configuration exacte des interactions.

4.1.3 Estimation de MV du coefficient d'autorégression

Etant donnée la décomposition (4.1), l'estimation de maximum de vraisemblance de ρ peut être obtenue à partir la fonction de vraisemblance $\mathcal{L}_1(\rho, \sigma^2|\mathbf{x}, \mathbf{m})$ (traitée par Ord, 1975). En admettant que la matrice des interactions \mathbf{M} est de nature à ne pas pouvoir être observée, la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{m})$ peut être substituée par la fonction de vraisemblance $\mathcal{L}_1(\rho, \mathbf{m}, \hat{\sigma}^2|\mathbf{x})$ en faisant passer formellement \mathbf{m} du statut de réalisation d'une variable observable au statut de paramètre inconnu. Mais dans ce cas, pour estimer ρ , on est tenu d'estimer \mathbf{m} par son estimateur de vraisemblance, ce qui est rendu difficile par la croissance exponentielle asymptotique du nombre de degrés de liberté de \mathbf{m} . Cette difficulté pousse à estimer la matrice \mathbf{m} d'une autre manière ou bien à adopter une autre méthode d'estimation de ρ qui ne passe pas par l'estimation de \mathbf{m} .

Si nous admettons que $p = 0$, alors la fonction de vraisemblance à employer pour estimer ρ devient $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{S})$. Nous montrons qu'il en résulte une estimation biaisée de ρ lorsqu'en réalité $p \neq 0$. Ce biais est illustré dans la figure 4.1 où la matrice de structure des interactions est un réseau carré parfait de taille 30×30 sites. La figure permet de comparer l'estimation de ρ obtenue par la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{m})$ avec l'estimation de ρ obtenue par la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{S})$ et démontre que, contrairement à $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{m})$ qui permet une estimation non biaisée, la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2|\mathbf{x}, \mathbf{S})$ induit une sous-estimation de plus en plus grande du paramètre ρ lorsque p croît. De

ce point de vue, le paramètre p agit comme un *paramètre d'intensité* du coefficient d'autorégression *perçu*, justifiant le fait de considérer la connexité du champ comme un paramètre de plus de l'autocorrélation spatiale.

En observant attentivement les courbes de la figure 4.1(b), on relève que le biais introduit par l'emploi de la matrice de structure des interactions \mathbf{S} en lieu et place des matrices des interactions réalisées \mathbf{m} est faible et varie peu dans un voisinage de $p = 0$ dont le rayon dépend de ρ . Par ce constat, l'estimateur de maximum de vraisemblance de ρ employant la matrice de structure semble être tout de même un bon estimateur aux conditions que ρ et p soient suffisamment faibles. La figure 4.2 illustre la dispersion des valeurs estimées de ρ par la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{S})$ (\mathbf{S} étant la matrice d'adjacence d'un réseau carré parfait) à l'aide d'une simulation de Monte Carlo du modèle de Whittle probabilisé où ρ et p sont aléatoires, tels que

$$\rho \sim \mathcal{U}_{(0,0.90)}, \quad p \sim \mathcal{U}_{(0,0.90)}, \quad (4.7)$$

où $\mathcal{U}_{(0,0.90)}$ est la loi de distribution uniforme sur l'intervalle ouvert $(0, 0.90)$. Le diagramme de dispersion permet de confirmer que l'estimateur a un faible biais dans l'estimation du coefficient d'autorégression aux conditions que nous venons d'évoquer.

Remarque La présente étude a fait l'objet d'une publication (Carrillo, 2007). Etant donné la relation (4.6), elle présente d'une certaine manière la relation entre la densité des interactions et le biais de l'estimateur de MV de ρ . Une étude de Mizruchi & Neuman (2008), parue plus tard, indique que malgré l'emploi d'une matrice des interactions correcte, l'estimateur de MV de ρ possède un biais négatif dont la valeur absolue croît avec la valeur de ρ et la densité des interactions. Dans notre cas, ce biais, antagoniste à celui que l'on a observé en fonction de p , n'a pas été perçu significativement, ce que l'on pourrait expliquer par les faibles valeurs de densité des interactions (au sens de Mizruchi & Neuman) atteintes.

4.1.4 Estimation par l'algorithme EM

Une estimation conjointe du coefficient d'autorégression ρ et du paramètre p peut être effectuée à l'aide de l'*algorithme espérance-maximisation*, évoqué souvent sous le nom d'algorithme *EM*. Cet algorithme, présenté par Dempster et al. (1977), convient pour trouver les paramètres rendant maximale la fonction de vraisemblance d'un ensemble de données observées conjointement avec un ensemble de données manquantes (ou inobservables).

Dans le cas des modèles de Whittle markoviens probabilisés, les données observables sont en l'occurrence la réalisation \mathbf{x} d'un champ spatial \mathbf{X} , et les données inobservables sont la réalisation \mathbf{m} de la matrice des interactions \mathbf{M} du modèle spatial. Le couple $\mathbf{z} := (\mathbf{x}, \mathbf{m})$ forme les *données complètes* dont nous notons leur densité de probabilité par $\nu(\mathbf{z} | \boldsymbol{\theta})$, avec $\boldsymbol{\theta}$ un vecteur de paramètres paramétrisant la loi de probabilité de \mathbf{z} . Dans ce contexte, l'algorithme EM adapté à notre problème

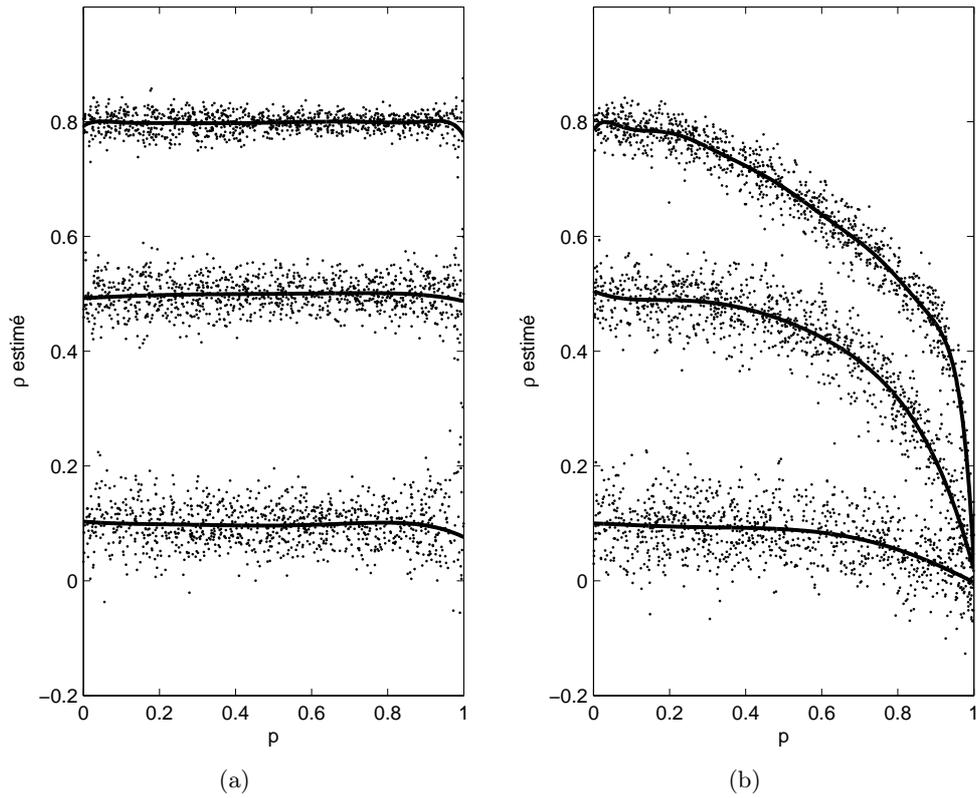


Fig. 4.1: Estimation de maximum de vraisemblance de ρ pour 1000 réalisations du couple (\mathbf{X}, \mathbf{M}) , pour un champ spatial périodique de taille 30×30 , avec la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{M})$ en figure (a) et avec la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{S})$ (\mathbf{S} étant la matrice d'adjacence d'un réseau carré parfait) en figure (b), dans des conditions où $\rho = 0.1, 0.5$ et 0.8 respectivement. Les courbes représentent, de bas vers le haut, un ajustement polynômial des points aux conditions $\rho = 0.1, 0.5$ et 0.8 respectivement.

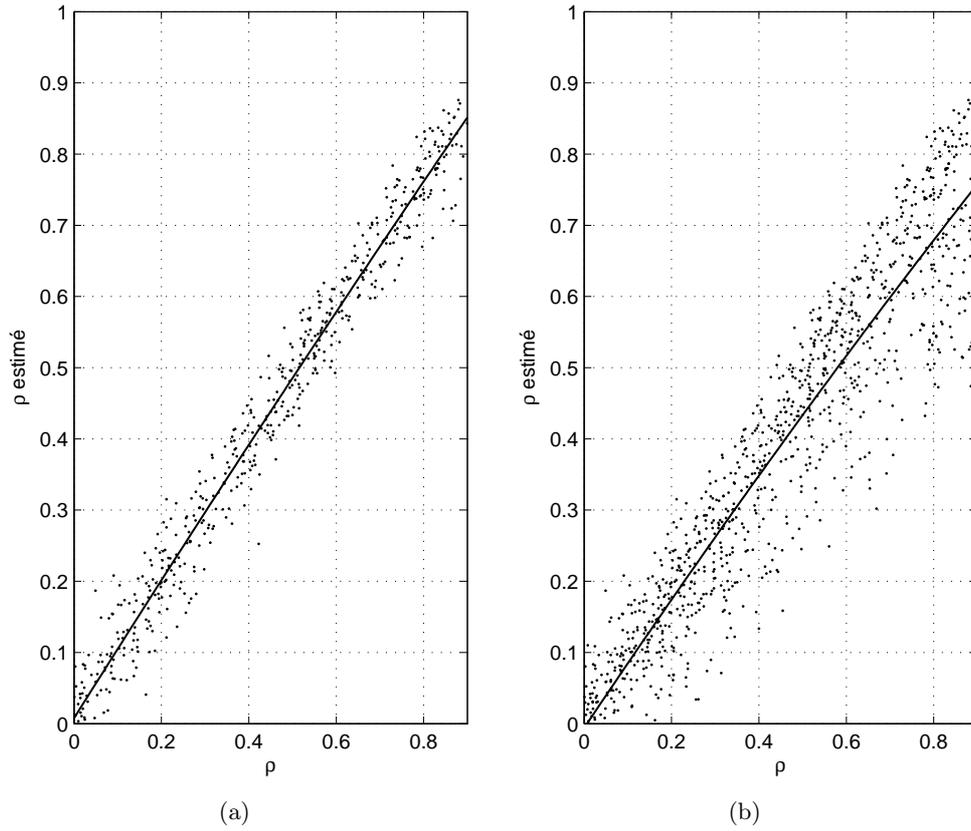


Fig. 4.2: Estimation de maximum de vraisemblance de ρ pour des réalisations du couple (\mathbf{X}, \mathbf{M}) , pour un champ spatial périodique de taille 30×30 , avec la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{S})$ (\mathbf{S} étant la matrice d'adjacence d'un réseau carré parfait). Les valeurs des paramètres p et ρ sont déterminés aléatoirement selon une distribution uniforme. Dans la figure de gauche, $\rho < 0.9$ et $p < 0.5$. Dans la figure de droite, $\rho < 0.9$ et $p < 0.8$.

se base sur la fonction

$$Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \mathbb{E} \left[\ln(\nu(\mathbf{z}|\tilde{\boldsymbol{\theta}})) | \mathbf{x}, \boldsymbol{\theta} \right] \quad (4.8)$$

(où $\mathbb{E}[\cdot]$ exprime l'espérance mathématique) et le calcul d'une suite de valeurs de paramètres $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots\}$ dont le terme $\boldsymbol{\theta}^{(j+1)}$ est induit par $\boldsymbol{\theta}^{(j)}$ de la manière suivante :

1. étape E : calculer $Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(j)})$.
 2. étape M : choisir la valeur $\boldsymbol{\theta}^{(j+1)}$ comme étant celle de $\tilde{\boldsymbol{\theta}}$ qui maximise $Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(j)})$.
- Les étapes 1 et 2 sont employées itérativement, en partant d'une valeur aléatoire $\boldsymbol{\theta}^{(0)}$.

L'utilité de cet algorithme réside dans le fait qu'il produit une suite de paramètres convergent vers la valeur du paramètre qui maximise la fonction de vraisemblance de \mathbf{x} , sans nul besoin de connaître la réalisation de la matrice des interactions \mathbf{m} . On met fin aux itérations de l'algorithme lorsqu'un critère de convergence est satisfait.

Dans le contexte du modèle de Whittle probabilisé, la quantité $Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ vaut

$$Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{M}}[\ln g(\mathbf{x}, \mathbf{m}|\tilde{\boldsymbol{\theta}}) | \mathbf{x}, \boldsymbol{\theta}] \quad (4.9)$$

$$= \sum_{\mathbf{m}} h(\mathbf{m}|\mathbf{x}, \boldsymbol{\theta}) \ln(g(\mathbf{x}, \mathbf{m}|\tilde{\boldsymbol{\theta}})), \quad (4.10)$$

où g et h sont respectivement la densité de probabilité et la probabilité que l'on retrouve dans (2.7), et

$$\boldsymbol{\theta} = (p, \rho, \sigma^2), \quad \tilde{\boldsymbol{\theta}} = (\tilde{p}, \tilde{\rho}, \tilde{\sigma}^2), \quad (4.11)$$

en admettant que la matrice de structure \mathbf{S} (supposée connue) est implicite et que $h(\mathbf{m}|\mathbf{x}, p, \rho, \sigma^2)$ se réduit à $h(\mathbf{m}|p)$.

Une difficulté de la méthode vient du calcul de (4.9) qui demande à sommer sur l'ensemble des configurations spatiales des interactions. Elle complique la détermination de $\boldsymbol{\theta}$ à l'étape M. Il est en principe suggéré de calculer (4.9) en employant un algorithme MCMC (Chib & Greenberg, 1996). Dans notre cas, cette solution comporte quelques difficultés propres aux modèles spatiaux autorégressifs (cf. méthode MCMC en section 4.3 de ce chapitre).

4.2 Autres estimateurs du coefficient d'autorégression

4.2.1 Définition

Comme nous l'avons vu, l'estimation de MV du coefficient d'autorégression est sensible à la spécification de la matrice des interactions du modèle. Si le but recherché est d'estimer le coefficient ρ sans nécessité de connaître la configuration exacte des interactions, la méthode de MV doit être abandonnée en faveur d'autres

méthodes plus robustes vis-à-vis des configurations exactes des interactions du modèle. Une voie de recherche consiste à trouver un estimateur $\hat{\rho}$, fonction du champ spatial \mathbf{X} , qui soit sensible au coefficient d'autorégression ρ , mais qui soit très peu sensible au paramètre p .

Supposons que nous traitons avec un modèle de Whittle probabilisé de la forme suivante

$$\mathbf{X} = \rho \mathbf{M} \mathbf{X} + \boldsymbol{\epsilon}, \quad \mathbf{M} \sim \mathcal{M}_{(p, \mathbf{S})}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_{(0, \mathbf{I}\sigma^2)}, \quad \rho \sim \mathcal{U}_{(0,1)}, \quad p \sim \mathcal{U}_{(0,1)}, \quad (4.12)$$

où $\mathcal{U}_{(0,1)}$ est la loi de distribution uniforme sur l'intervalle ouvert $(0, 1)$. La recherche de $\hat{\rho}$ peut être orientée vers la recherche d'une statistique $T(\mathbf{X})$ telle que sa corrélation avec ρ soit la meilleure possible, en même temps que sa corrélation avec p soit la plus faible possible. Si idéalement l'on trouve une telle statistique avec une corrélation avec ρ suffisamment proche de 1 en valeur absolue, on peut définir l'estimateur $\hat{\rho}$ comme prédicteur de $T(\mathbf{X})$, solution de

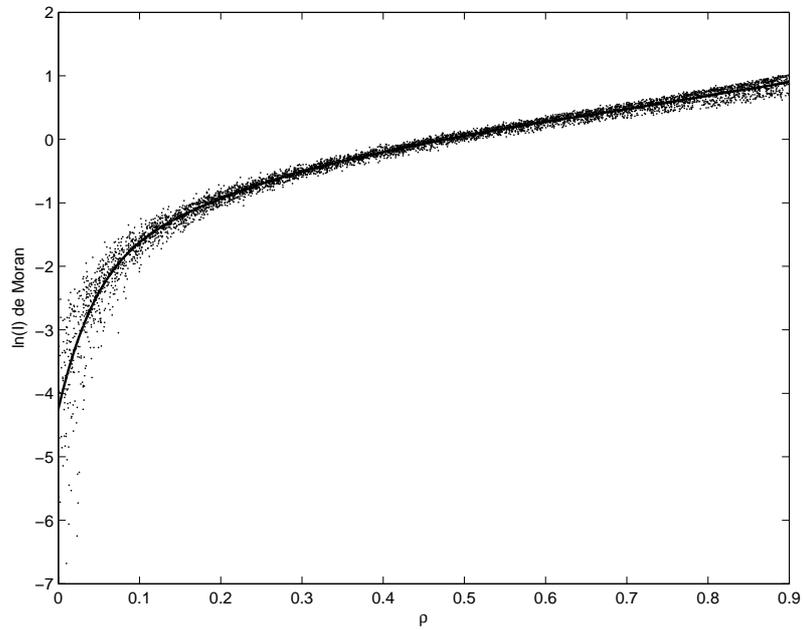
$$T(\mathbf{X}) = b_0 + b_1 \hat{\rho}, \quad (4.13)$$

où les éléments b_0 et b_1 sont des coefficients que l'on détermine à partir d'une simulation numérique effectuant une régression linéaire de ρ sur $T(\mathbf{X})$. Un tel estimateur est convergent si la corrélation converge en valeur absolue vers 1. Nous ne sommes cependant pas assurés de trouver une statistique $T(\mathbf{X})$ de corrélation satisfaisante avec ρ . Si elle existe, sa forme n'est pas évidente. Nous allons montrer ci-dessous deux pistes de recherches de $T(\mathbf{X})$ dont l'une est bien plus convaincante à exploiter que l'autre. Pour nous simplifier la tâche, nous considérons que la matrice de structure des interactions \mathbf{S} représente un réseau carré parfait de taille 100×100 sites.

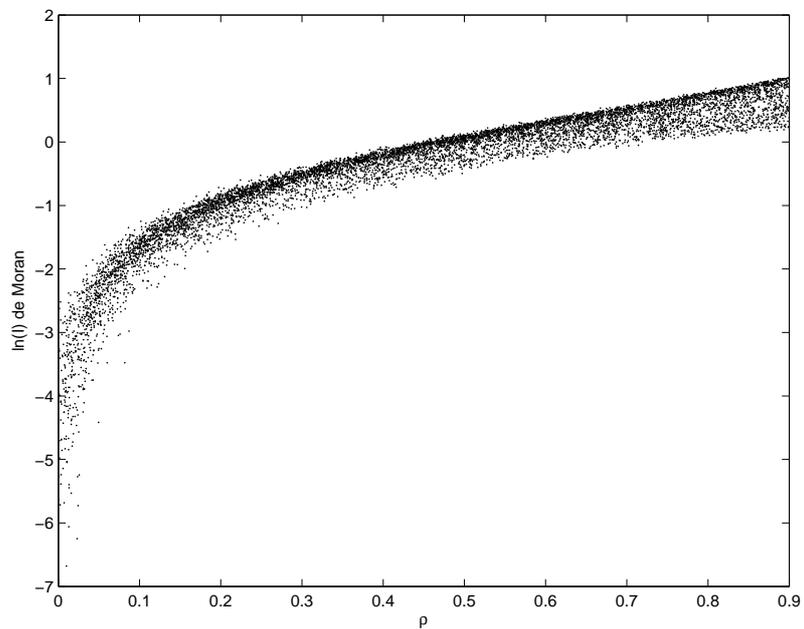
4.2.2 Méthode par l'indice I de Moran

Nous pouvons nous demander si les statistiques du produit croisé sont de bons candidats pour contribuer à l'estimation de ρ . A défaut de pouvoir étudier toutes les statistiques, nous décidons de voir si la statistique du I de Moran employant la pondération spatiale \mathbf{W}_{R1} , où \mathbf{W}_{R1} est la matrice d'adjacence d'un réseau carré parfait, est une bonne candidate. Nous réalisons pour cela un échantillon de 10^4 réalisations du modèle de Whittle probabilisé (4.12) dont il ressort que le passage du $I_{\mathbf{W}_{R1}}$ à son logarithme (en rejetant les réalisations négatives de $I_{\mathbf{W}_{R1}}$) semble améliorer la corrélation avec ρ et diminuer la corrélation avec p . Nous donnons le diagramme de dispersion de $\ln(I_{\mathbf{W}_{R1}})$ de Moran en fonction de ρ dans la figure 4.3(a) pour les individus de l'échantillon obtenus avec $p < 0.5$ et dans la figure 4.3(b) pour les individus de l'échantillon obtenus avec $p < 0.8$.

Comme on peut le constater sur les figures mentionnées ci-dessus, la relation entre $\ln(I_{\mathbf{W}_{R1}})$ et ρ n'est pas linéaire dans un intervalle de ρ au voisinage de 0. Au lieu donc de considérer la relation (4.13), nous définissons l'estimateur $\hat{\rho}$ de ρ



(a) Diagramme de dispersion de $\ln(I)$ en fonction de ρ pour 10^4 réalisations du modèle (4.12) en ayant employé cependant $p \sim \mathcal{U}_{(0,0.80)}$ et $\rho \sim \mathcal{U}_{(0,0.90)}$ et dont on a supprimé les points obtenus avec $p \geq 0.5$.



(b) Diagramme de dispersion de $\ln(I)$ en fonction de ρ pour 10^4 réalisations du modèle (4.12) en ayant employé cependant $p \sim \mathcal{U}_{(0,0.80)}$ et $\rho \sim \mathcal{U}_{(0,0.90)}$.

Fig. 4.3:

comme la solution de la relation polynomiale

$$\ln(I_{\mathbf{W}_{R1}}) = \sum_{i=0}^{\kappa} b_i \hat{\rho}^i, \quad (4.14)$$

où les coefficients b_i ($i = 1, 2, \dots, \kappa$) sont les coefficients du polynôme ajusté aux points du diagramme de dispersion de la figure 4.3(a) (qui est préféré au diagramme de la figure 4.3(b)). Un tel polynôme a été calculé et est illustré dans la figure 4.3(a). Sur la base des coefficients de ce polynôme, nous illustrons dans les figures 4.4(a) et 4.4(b) respectivement la dispersion de l'estimateur $\hat{\rho}$ à partir des valeurs $\ln(I_{\mathbf{W}_{R1}})$ du diagramme 4.3(a) et 4.3(b) respectivement, de sorte à pouvoir en apprécier la précision.

On conclut que l'estimateur de ρ par cette méthode convient mieux pour des petites valeurs de ρ et pour des valeurs de p qui ne sont pas trop élevées (disons $p < 0.5$). L'efficacité de l'estimation se dégrade lorsque ρ croît et si de surcroît les valeurs de p sont élevées.

4.2.3 Méthode par l'indice I de Moran maximal

Définition

On peut substituer au choix arbitraire de la pondération spatiale de la méthode du I de Moran une méthode de pondération spatiale ajustée aux données pour différentes échelles de distance. La méthode que nous choisissons est celle du I de Moran maximal, dont le principe consiste à déterminer des poids spatiaux w_i qui rendent maximal, pour la réalisation \mathbf{x} du champ considérée, et sous certaines contraintes, une statistique $I(\mathbf{x})$ de Moran de la forme

$$I(\mathbf{x}) = \sum_{i=1}^k w_i I_{\mathbf{A}_i}(\mathbf{x}), \quad (4.15)$$

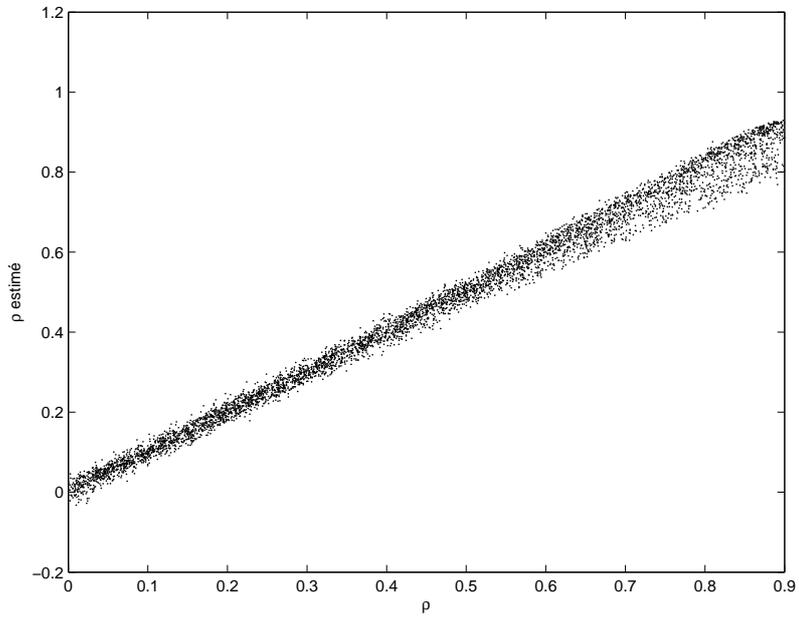
où les $I_{\mathbf{A}_i}(\mathbf{x})$ sont des indices de Moran de matrice d'adjacence constante \mathbf{A}_i . Ce principe, proposé par Kooijman (1976) pour l'étude de données écologiques, est conforme pour "révéler" une forme d'autocorrélation spatiale et donner une description de celle-ci.

Formellement, la statistique du I de Moran maximal, notée I_{\max} , est définie comme

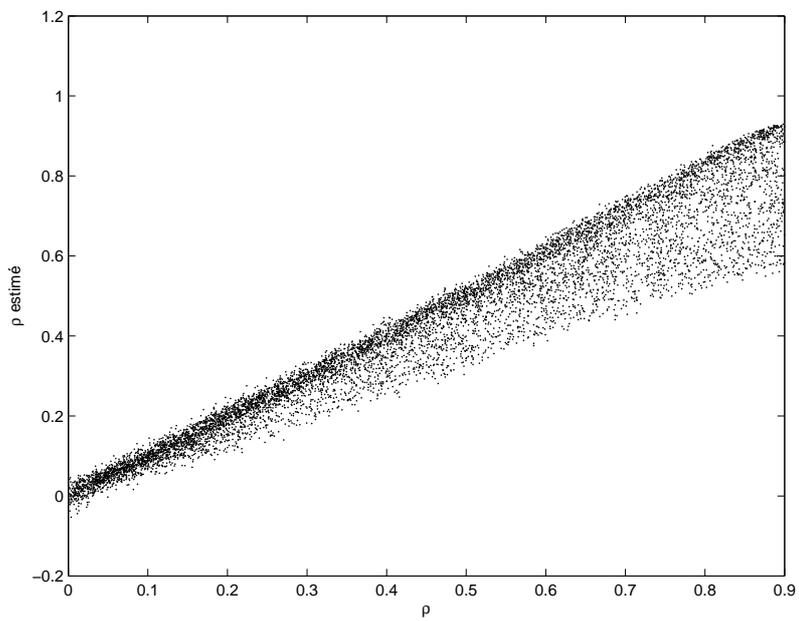
$$I_{\max}(\mathbf{X}) := \max_{\mathbf{W} \in \Omega} [I_{\mathbf{W}}(\mathbf{X})], \quad (4.16)$$

où $I_{\mathbf{W}}$ est l'indice de Moran de matrice de poids spatiaux \mathbf{W} et où Ω est l'espace de variation de \mathbf{W} . Le choix de Ω définit les contraintes imposées aux éléments de \mathbf{W} . Les contraintes choisies sont

$$[\mathbf{W} \in \Omega] \Leftrightarrow [\mathbf{W} = \sum_{i=1}^{\kappa+1} w_i \mathbf{A}_i \quad \text{et} \quad \sum_{i=1}^{\kappa+1} n_i w_i^2 = b], \quad (4.17)$$



(a) Diagramme de dispersion de $\hat{\rho}$ en fonction de ρ à partir du diagramme de dispersion de la figure 4.3(a).



(b) Diagramme de dispersion de $\hat{\rho}$ en fonction de ρ obtenu à partir du diagramme de dispersion de la figure 4.3(b).

Fig. 4.4:

où n_i est le nombre d'adjacences de \mathbf{A}_i , b est une constante arbitraire, κ est un entier fixé a priori et les \mathbf{A}_i sont des matrices d'adjacence du champ définie de la manière suivante : un site est adjacent à un autre selon \mathbf{A}_i si et seulement si les deux sites sont séparés d'une dissimilarité spatiale (par ex. une distance euclidienne) dans l'intervalle de valeurs $(\mathfrak{d}_{i-1}, \mathfrak{d}_i]$, avec la condition

$$0 = \mathfrak{d}_0 < \mathfrak{d}_1 < \mathfrak{d}_2 < \dots < \mathfrak{d}_\kappa < \mathfrak{d}_{\kappa+1}. \quad (4.18)$$

A ce titre, \mathbf{A}_1 représente le *graphe de voisinage des premiers plus proches voisins*, \mathbf{A}_2 représente le *graphe de voisinage des deuxièmes plus proches voisins*, etc. Nous appelons l'ensemble de sites qui sont adjacents selon \mathbf{A}_i un *voisinage d'ordre i* . Nous désignons la matrice \mathbf{A}_i comme la *matrice de voisinage d'ordre i* . Le choix des distances (4.18) est conçu pour partitionner l'ensemble de toutes les adjacences possibles entre sites du champ spatial, de telle sorte que l'on vérifie le résultat

$$\sum_{i=0}^{\kappa+1} \mathbf{A}_i = \mathbf{U}, \quad (4.19)$$

où $\mathbf{A}_0 = \mathbf{I}$ est la matrice de toutes les adjacences réflexives possibles et \mathbf{U} représente la matrice d'adjacence du graphe complet (incluant les arêtes réflexives). Nous formulons l'hypothèse que les variables ne sont pas corrélées avec leurs voisins au-delà d'un certain ordre de voisinage. Nous admettons dans ce contexte que \mathfrak{d}_κ représente la dissimilarité spatiale maximale pour lequel théoriquement $w_\kappa \neq 0$.

La définition (4.16) avec (4.17) constitue en fait une variante du principe de la méthode de Kooijman (1976) pour lequel son auteur imposait la condition supplémentaire $\sum_{i=1}^{\kappa} w_i = a$ (voir annexe de ce chapitre). La solution de (4.16) (calculée en annexe de ce chapitre) est

$$I_{\max}(\mathbf{x}) = \frac{\sqrt{B(\mathbf{x})}}{\mathbf{x}'\mathbf{x}}, \quad (4.20)$$

$$w_i(\mathbf{x}) = \frac{a_i(\mathbf{x})}{n_i} \frac{1}{\sqrt{B(\mathbf{x})}}, \quad 0 < i \leq \kappa, \quad (4.21)$$

où n_i est le nombre d'adjacences de \mathbf{A}_i , $a_i(\mathbf{x})$ est la forme quadratique induite par \mathbf{A}_i ,

$$a_i(\mathbf{x}) = \mathbf{x}'\mathbf{A}_i\mathbf{x}, \quad (4.22)$$

et $B(\mathbf{x})$ est la quantité définie comme

$$B(\mathbf{x}) = \sum_{i=1}^{\kappa} \frac{a_i^2(\mathbf{x})}{n_i}. \quad (4.23)$$

Pour finir de déterminer les poids spatiaux, nous devons fixer les matrices d'adjacence \mathbf{A}_i ($i = 1, 2, \dots, \kappa$). Les adjacences entre sites du champ peuvent être déterminées en fonction des distances euclidiennes qui séparent les sites. Cependant,

étant donnée la géométrie en réseau carré du champ spatial, les adjacences peuvent également être déterminées en fonction de la distance de Manhattan qui sépare les sites. En adoptant cette deuxième manière de calculer les distances, nous faisons la construction suivante : deux sites sont adjacents selon \mathbf{A}_i si et seulement si la distance de Manhattan qui les sépare est dans l'intervalle $(i - 1, i]$, $i = 1, 2, \dots, \kappa$. On obtient de cette manière une mise en classe isotrope du voisinage qui est la plus fine. Les voisinages jusqu'à l'ordre 3 sont illustrés dans la figure 4.5.

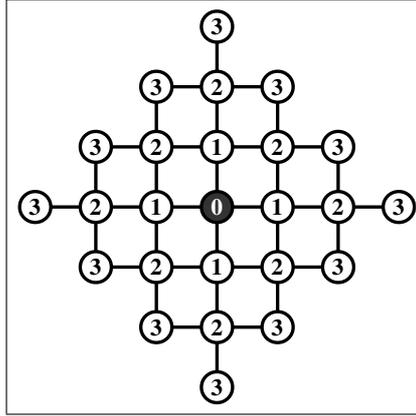


Fig. 4.5: Représentation du voisinage d'ordre 0, 1, 2 et 3 d'un site central (selon distance de Manhattan) sur un réseau carré parfait non détruit ($p = 0$) : la valeur sur chaque site indique la plus petite distance de Manhattan (écart) qui le sépare du site central (en noir).

Avec la volonté d'intégrer la structure des interactions du champ, les matrices de voisinage \mathbf{A}_i sont calculées selon l'algorithme suivant¹ :

$$\begin{aligned}
 \mathbf{A}_1 &= \mathbf{S}, \\
 \mathbf{A}_2 &= [\mathbf{A}_1^2]_g * (\mathbf{U} - \mathbf{I} - \mathbf{A}_1), \\
 \mathbf{A}_3 &= [\mathbf{A}_1^3]_g * (\mathbf{U} - \mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2), \\
 &\vdots \\
 \mathbf{A}_\kappa &= [\mathbf{A}_1^\kappa]_g * (\mathbf{U} - \mathbf{I} - \sum_{i=1}^{\kappa-1} \mathbf{A}_i),
 \end{aligned}$$

où \mathbf{S} est la matrice de structure des interactions, $[\]_g$ est un opérateur sur les matrices qui transforme chaque élément de matrice en la valeur 1 si l'élément est non nul (et laissant inchangés les éléments nuls) et la loi "*" est le produit d'Hadamard. A chaque pas de l'algorithme ci-dessus une matrice d'adjacence est

1. Pour ce calcul \mathbf{U} peut être remplacée par une matrice $\tilde{\mathbf{U}}$ moins connective, définie comme $\tilde{\mathbf{U}} = [\sum_{i=1}^{\kappa} \mathbf{S}^i]_g + \mathbf{I}$

produite établissant l'adjacence entre sites écartés d'un arc supplémentaire, sans faire de circuit (principe d'un feu qui progresse à chaque pas sans repasser par là où cela a déjà brûlé).

Estimation du coefficient d'autorégression par l'indice I_{\max} de Moran

Dans la section 4.2.2, nous avons développé une méthode simple d'estimation du coefficient ρ à partir du logarithme du I de Moran. Cette méthode peut être appliquée au logarithme du I_{\max} de Moran. En reprenant le modèle spatial (4.12), l'estimateur du coefficient ρ est dans ce cas la solution $\hat{\rho}$ telle qu'elle satisfait la relation

$$\ln(I_{\max}) = \sum_{i=0}^{\kappa} b_i \hat{\rho}^i, \quad (4.24)$$

où les coefficients b_i ($i = 1, 2, \dots, \kappa$) sont les coefficients du polynôme que l'on ajuste aux points d'un diagramme de dispersion (I_{\max}, ρ) obtenu par simulation numérique. Les figures 4.6(a) et 4.6(b) donnent un tel diagramme de dispersion (voir légende de ces figures). Un ajustement polynomial des données du diagramme de la figure 4.6(a) permet de déterminer les coefficients b_i nécessaires pour la méthode d'estimation. Sur la base des coefficients de ce polynôme, nous illustrons dans les figures 4.7(a) et 4.7(b) respectivement la dispersion de l'estimateur $\hat{\rho}$ à partir des valeurs $\ln(I_{\max})$ du diagramme 4.6(a) et 4.6(b) respectivement, de sorte à pouvoir en apprécier la précision.

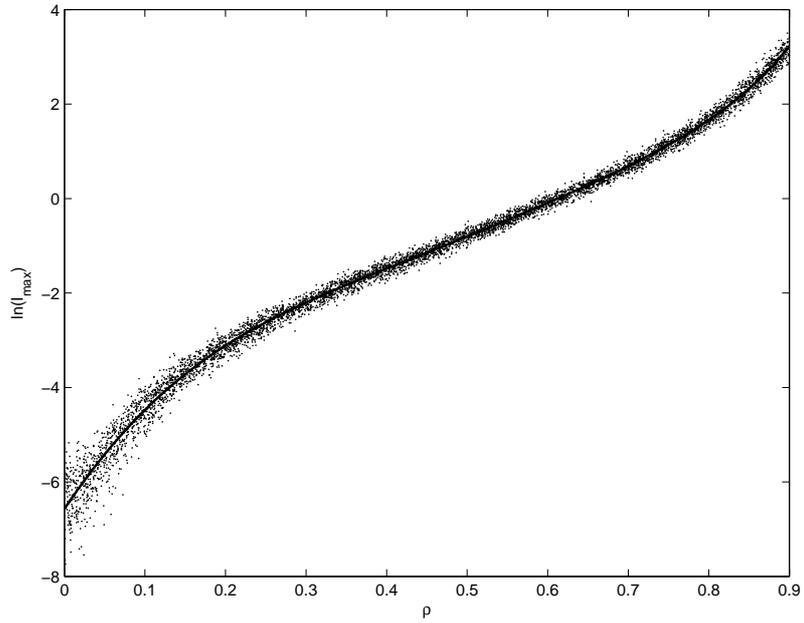
Il en ressort que la méthode d'estimation par le logarithme de I_{\max} est meilleure que la méthode d'estimation par le logarithme du I de Moran avec la pondération spatiale \mathbf{W}_{R1} (cf. section 4.2.2), notamment si $p < 0.5$. Selon les conditions sur p et sur ρ , la présente méthode peut aussi être meilleure que la méthode par comptage d'amas (cf. section 4.2.4) : c'est le cas si les valeurs de ρ et/ou de p sont suffisamment petites.

Comparaisons à $p < 0.5$

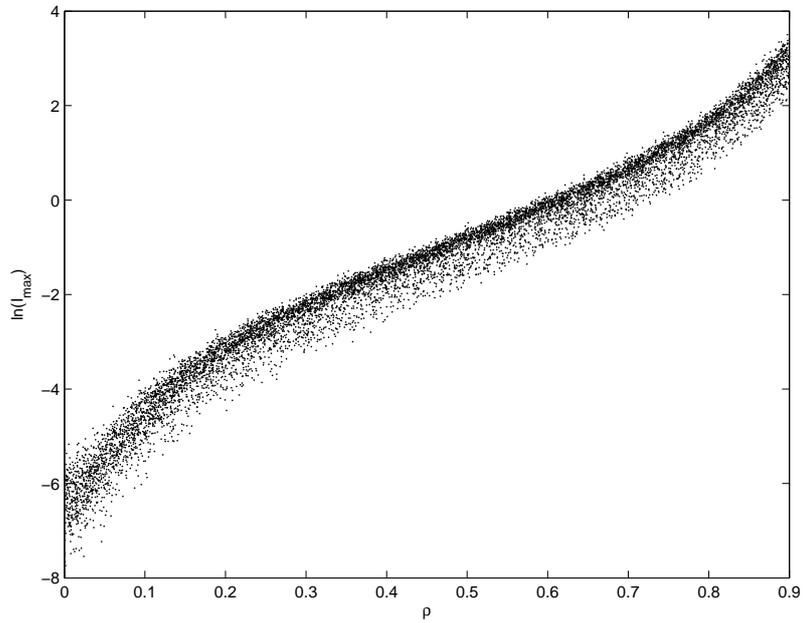
La pondération spatiale entrant dans l'indice I de Moran maximal contribue à une relation relativement robuste entre I_{\max} et ρ , si la valeur de p reste dans une marge raisonnable, disons $p < 0.5$. La méthode du I_{\max} donnant par la même occasion un corrélogramme, il conviendrait normalement d'étudier une méthode similaire d'estimation de ρ avec les poids spatiaux (4.21) entrant dans le corrélogramme. Nous ne le faisons que pour le premier poids spatial (qui est prépondérant dans le corrélogramme) sans sa normalisation et "dilaté" d'un facteur $\sqrt{n_1}$, soit

$$w_1^* = \frac{1}{\sqrt{n_1}} \mathbf{x} \mathbf{A}_1 \mathbf{x}. \quad (4.25)$$

Comme auparavant, on emploie un modèle de Whittle dont les valeurs des paramètres p et ρ sont obtenues selon une distribution uniforme (simulation de Monte Carlo). Nous limitons à $\rho < 0.8$ et $p < 0.5$. Nous voulons cette fois-ci que

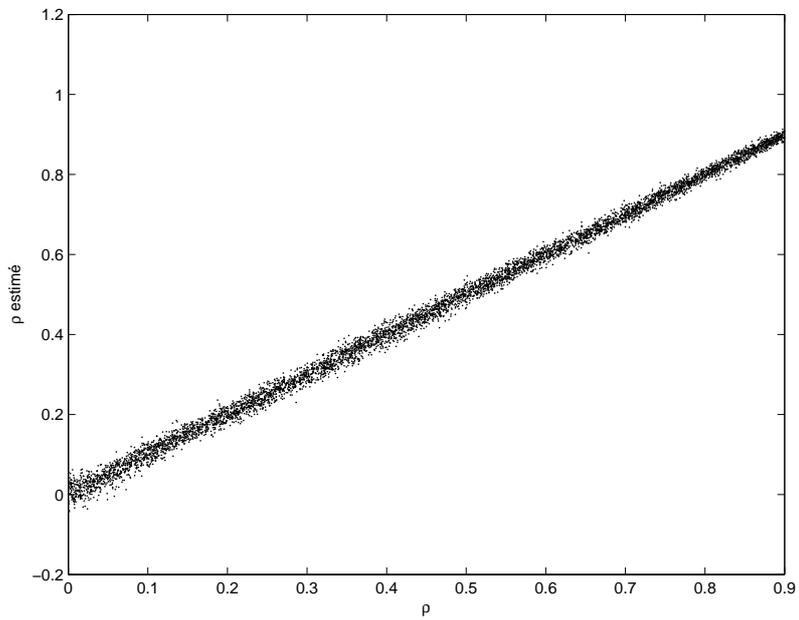


(a) Diagramme de dispersion de $\ln(I_{\max})$ en fonction de ρ pour 10^4 réalisations du modèle (4.12) en ayant employé cependant $p \sim \mathcal{U}_{(0,0.80)}$ et $\rho \sim \mathcal{U}_{(0,0.90)}$ et dont on a supprimé les points obtenus avec $p \geq 0.5$.

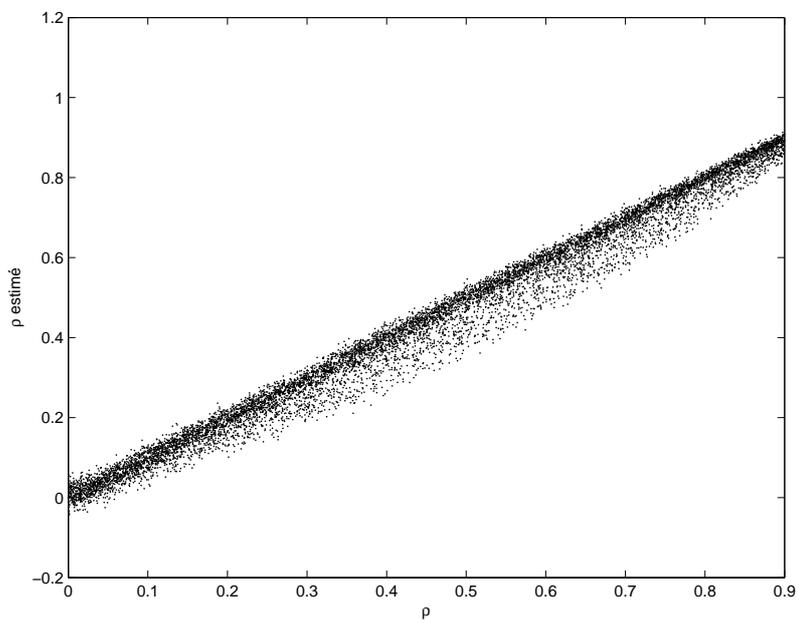


(b) Diagramme de dispersion de $\ln(I_{\max})$ en fonction de ρ pour 10^4 réalisations du modèle (4.12) en ayant employé cependant $p \sim \mathcal{U}_{(0,0.80)}$ et $\rho \sim \mathcal{U}_{(0,0.90)}$.

Fig. 4.6:



(a) Diagramme de dispersion de $\hat{\rho}$ en fonction de ρ à partir du diagramme de dispersion de la figure 4.6(a).



(b) Diagramme de dispersion de $\hat{\rho}$ en fonction de ρ à partir du diagramme de dispersion de la figure 4.6(b).

Fig. 4.7:

le graphe de structure des interactions soit produit à partir d'une triangulation de Delaunay comme illustré dans la figure 4.8(a). La méthode reste par la suite la même : on établit tout d'abord une relation entre la statistique étudiée (soit I_{\max} , respectivement w_1^*) et le coefficient ρ , puis on ajuste à cette relation un polynôme qui sert par la suite à transformer une valeur de I_{\max} , respectivement une valeur de w_1^* en une valeur estimée de ρ . Les figures 4.8(b) et 4.8(c) illustrent la relation entre les statistiques et ρ . Sur la base des coefficients du polynôme, nous illustrons dans les figures 4.9(a) et 4.9(b) respectivement la dispersion des valeurs estimées de ρ à partir de valeurs des diagrammes 4.8(b) et 4.8(c) respectivement, de sorte à pouvoir en apprécier la précision. Nous accompagnons cette comparaison avec les estimations du coefficient ρ produites par l'estimateur de maximum de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{S})$ (où \mathbf{S} est la matrice de structure des interactions) sur les mêmes simulations du champ : voir figure 4.8(c). *Il ressort de la comparaison que les estimateurs présentent des performances similaires.*

4.2.4 Méthode par comptage d'amas

Principe

Nous proposons une autre méthode pour estimer ρ . Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ le champ spatial de n variables comme défini par (4.12). Soit $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ un nouveau champ spatial, fonction de \mathbf{X} , tel que si X_i est la variable associée au site i , alors Z_i est associée à ce même site. Les variables du champ \mathbf{Z} sont à valeurs binaires déterminées à partir du champ \mathbf{X} par la définition suivante :

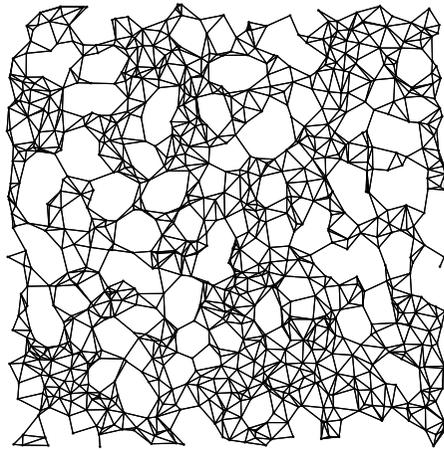
$$Z_i := \begin{cases} 1 & \text{si } \frac{X_i - \mu(\mathbf{X})}{\sqrt{\text{var}(\mathbf{X})}} > \eta, \\ 0 & \text{sinon} \end{cases}, \quad (4.26)$$

où η est un scalaire à définir, et $\mu(\mathbf{X})$ et $\text{var}(\mathbf{X})$ sont respectivement la moyenne et la variance spatiale de \mathbf{X} :

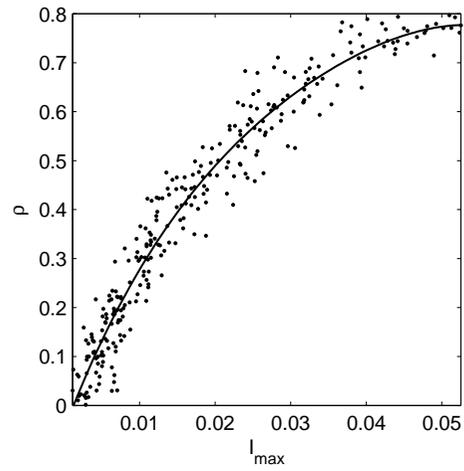
$$\mu(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{var}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu(\mathbf{X}))^2. \quad (4.27)$$

Pour chaque réalisation \mathbf{z} de \mathbf{Z} nous obtenons un champ spatial binaire avec une configuration spatiale de valeurs 0 et 1, comme illustrée par la figure 4.10. On définit dans ce contexte une fonction \mathfrak{A} telle que $\mathfrak{A}(\mathbf{z})$ donne le nombre d'amas de variables ayant réalisé la valeur 1 au sens de la théorie de la percolation de sites sur un réseau carré². Pour un bon choix de η , $\mathfrak{A}(\mathbf{z})$ peut devenir une statistique fortement corrélée avec ρ en même temps qu'elle devient faiblement corrélée avec p . Ceci est illustré dans la figure 4.11 pour un champ spatial de taille 100×100 sites. Pour cet exemple, on constate qu'il est pertinent de choisir un seuil η égal

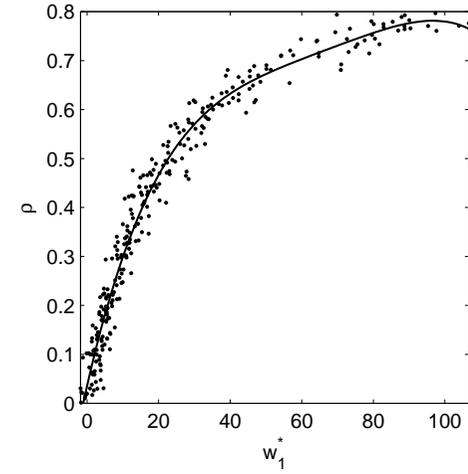
². On considère la position d'une variable comme contiguë à une autre si cette dernière est dans le voisinage de von Neumann de la première.



(a) Un semis de 900 points et sa triangulation de Delaunay dont on a supprimé 15% des arêtes par ordre décroissant de leur longueur euclidienne.

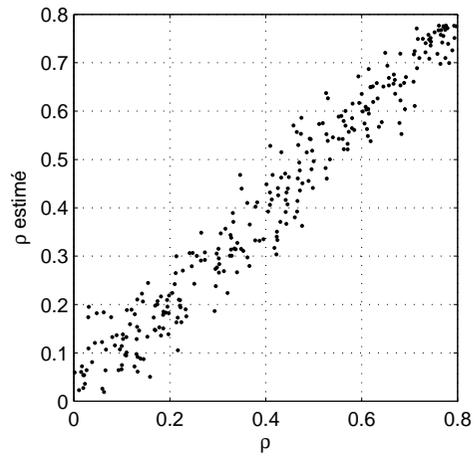


(b) Points dans le plan (I_{\max}, ρ) obtenus de réalisations du modèle de Whittle probabilisé pour des valeurs de paramètres p et ρ obtenues aléatoirement et dont le graphe de structure est en figure (a).

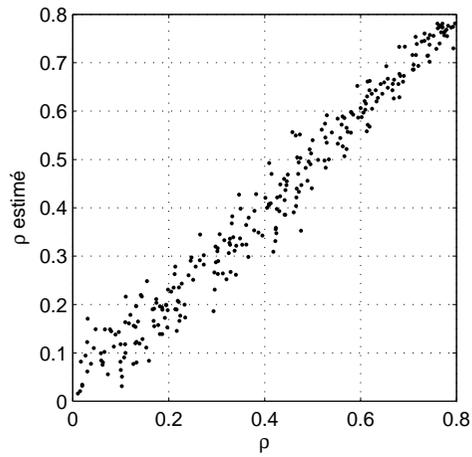


(c) Points dans le plan (w_1^*, ρ) obtenus de réalisations du modèle de Whittle probabilisé pour des valeurs de paramètres p et ρ obtenues aléatoirement et dont le graphe de structure est en figure (a).

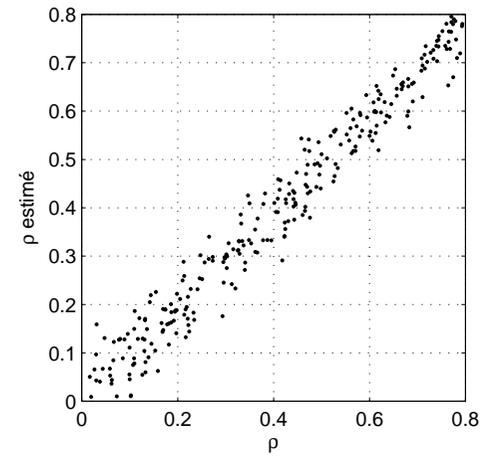
Fig. 4.8:



(a) Dispersion des estimations de ρ par la méthode du I_{\max} .

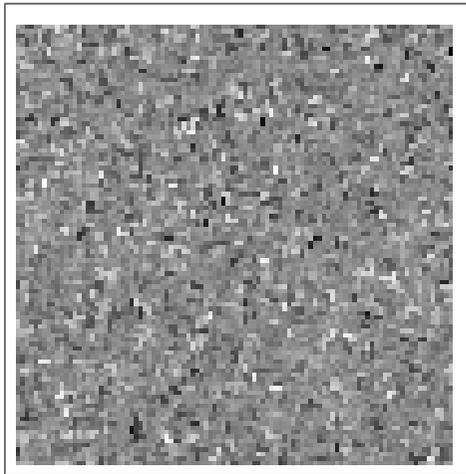


(b) Dispersion des estimations de ρ par la méthode du poids spatial w_1^*

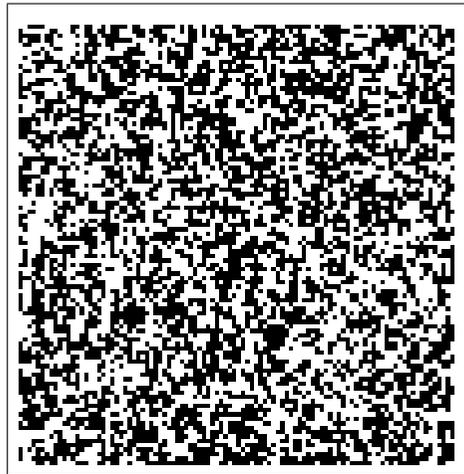


(c) Dispersion des estimations de ρ par la méthode du maximum de vraisemblance employant la fonction de vraisemblance $\mathcal{L}_1(\rho, \hat{\sigma}^2 | \mathbf{X}, \mathbf{S})$ (où \mathbf{S} est la matrice de structure des interactions)

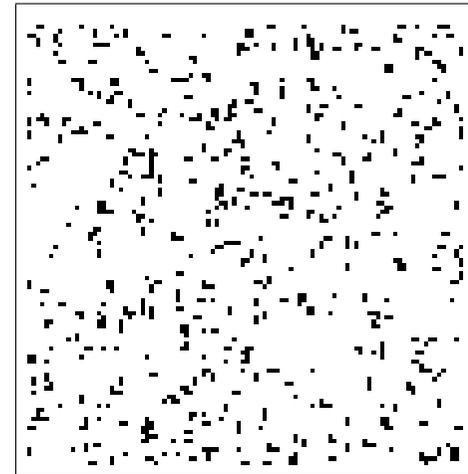
Fig. 4.9:



(a) Réalisation d'un champ spatial périodique de taille 100×100 sites conforme à un modèle de Whittle markovien probabilisé de paramètres $p = 0.80$ et $\rho = 0.80$ ($\sigma^2 = 1$).



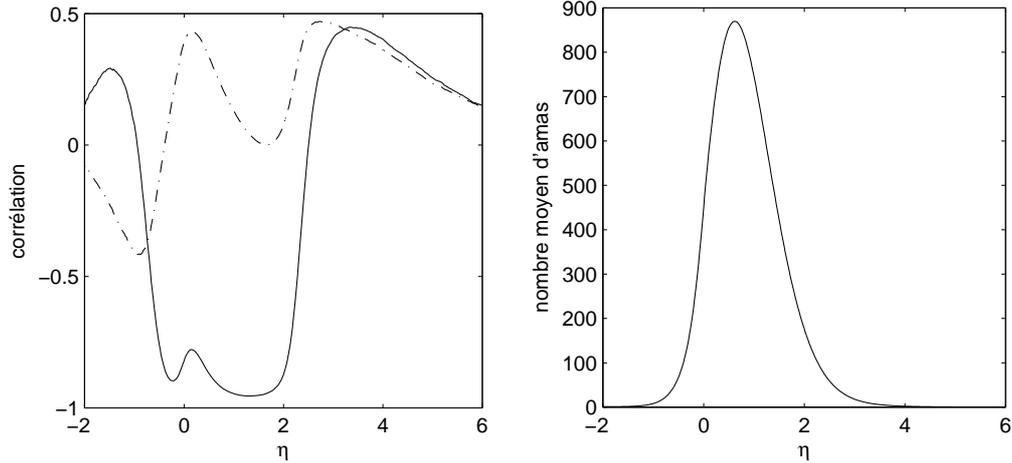
(b) Résultat de (4.26) obtenu à partir de la réalisation en figure (a), avec un seuil $\eta = 0.0$.



(c) Résultat de (4.26) obtenu à partir de la réalisation en figure (a), avec un seuil $\eta = 1.3$.

Fig. 4.10: Exemple de transformation d'un champ spatial gaussien en un champ spatial binaire suivant la définition (4.26).

à la valeur 1.3. Cependant, pour ce seuil, la relation entre $\mathfrak{A}(\mathbf{Z})$ et ρ ne semble pas tout à fait correspondre à un modèle linéaire ; une relation polynomiale d'ordre 3 comme l'illustre la figure 4.12(a) semble mieux correspondre aux données. En guise de raffinement, la définition (4.13) peut être modifiée en conséquence. Nous donnons ci-dessous un exemple d'application.



(a) Corrélation du nombre d'amas \mathfrak{A} avec ρ (en trait continu) et avec p (en trait discontinu) en fonction du seuil η dans (4.26). (b) Nombre moyen d'amas en fonction du seuil η dans (4.26).

Fig. 4.11:

Exemple

Soit le champ spatial (4.12) de 100×100 sites comme dans l'exemple précédent. Nous considérons $\eta = 1.3$ pour calculer $\mathbf{Z} = \zeta(\mathbf{X})$. Pour tenir compte de la non linéarité observée dans le diagramme de dispersion de la figure 4.12(a), nous définissons l'estimateur $\hat{\rho}$ comme un prédicteur de $\mathfrak{A}(\mathbf{Z})$, solution de la relation quadratique³

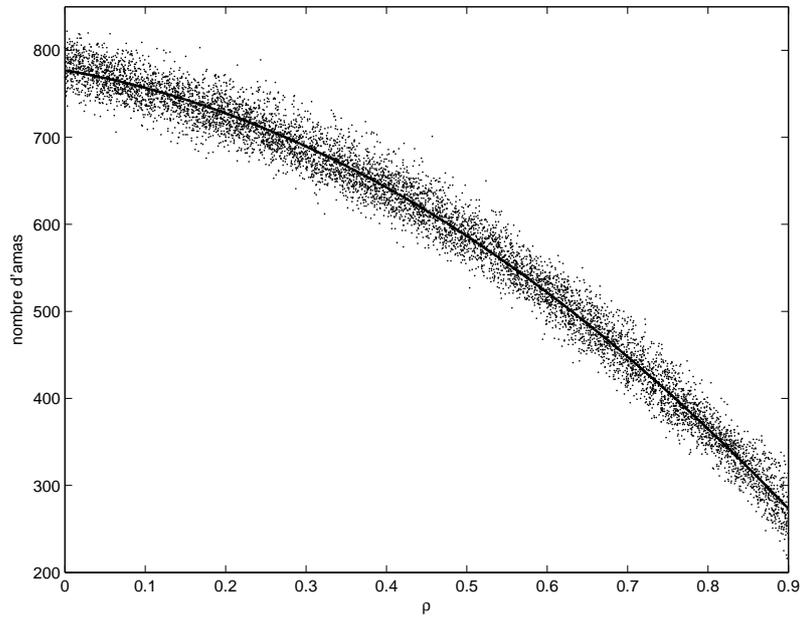
$$\mathfrak{A}(\mathbf{Z}) = b_0 + b_1 \hat{\rho} + b_2 \hat{\rho}^2, \quad (4.28)$$

où les coefficients b_0 , b_1 et b_2 sont déterminés par ajustement polynomial sur le diagramme de la figure 4.12(a). Cette solution est

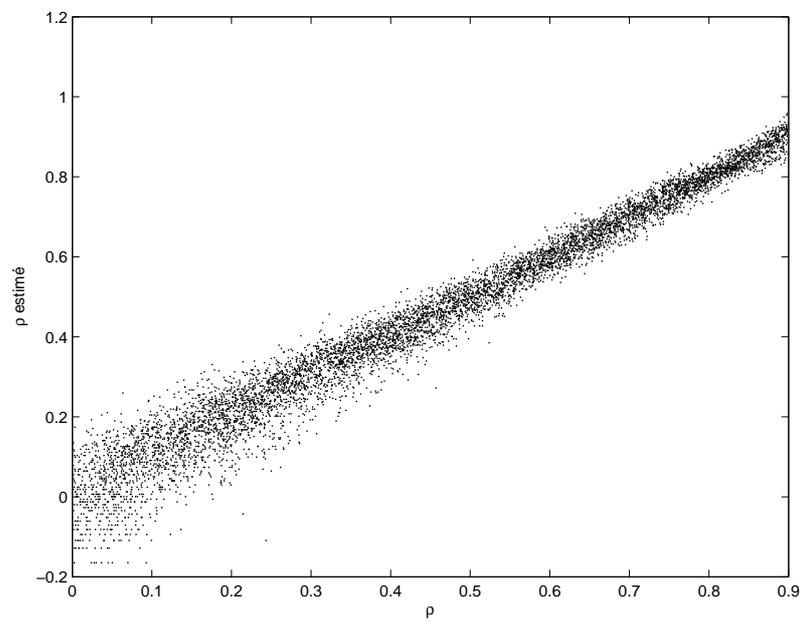
$$\hat{\rho} = -\frac{b_1 + \sqrt{b_1^2 + 4b_2(\mathfrak{A}(\zeta(\mathbf{X})) - b_0)}}{2b_2}, \quad (4.29)$$

avec $b_0 = 777.06$, $b_1 = -158.12$, $b_2 = -446.18$. Appliqué à la réalisation de champ illustrée en figure 4.10(a), on obtient $\hat{\rho} = 0.7989$, alors qu'en réalité $\rho = 0.80$ et $p = 0.80$.

3. Nous avons fait le choix de poser une relation quadratique, plutôt qu'une relation cubique (qui convient pourtant mieux), pour trouver plus aisément une solution mathématique explicite.



(a) Diagramme de dispersion du nombre d'amas \mathfrak{A} au seuil $\eta = 1.30$ en fonction de ρ pour 10^4 réalisations de (4.12) dont on a supprimé les points obtenus avec $p \geq 0.8$ ou $\rho \geq 0.9$.



(b) Diagramme de dispersion de $\hat{\rho}$ en fonction de ρ obtenu à partir du diagramme de dispersion de la figure 4.12(a).

Fig. 4.12:

Pour apprécier le degré de précision de la statistique $\hat{\rho}$, le diagramme de dispersion de la figure 4.12(a) donne les estimations de ρ via la formule (4.29) effectuées pour 10^4 réalisations du champ (4.12). On conclut que l'estimation de ρ avec cette méthode convient mieux pour des valeurs élevées de ρ . Cette méthode est globalement plus satisfaisante pour sa robustesse vis-à-vis de p que la méthode avec le I de Moran précédemment décrite.

4.3 Estimation bayésienne des paramètres

Une approche bayésienne de l'estimation des paramètres peut être employée au lieu d'une estimation ponctuelle. Dans son principe, elle consiste à associer une distribution de probabilité *a priori* à chaque paramètre inconnu et à déterminer, avec l'information que l'on dispose, leur distribution de probabilité *a posteriori*, c'est-à-dire conditionnellement aux données du champ. La méthodologie produit un résultat riche en information étant donné que d'une distribution de probabilité *a posteriori* il est toujours possible de calculer une variété d'indicateurs de tendance centrale et de dispersion.

Parmi les méthodes existantes, la simulation MCMC (Monte-Carlo Markov Chain) rencontre un grand intérêt, car elle permet à la fois de simuler un échantillon de données selon une loi paramétrique fixée à l'avance et d'estimer la valeur de ces paramètres. La méthode peut employer divers algorithmes dont les principaux sont l'algorithme de Gibbs (Geman & Geman, 1984) et l'algorithme de Metropolis-Hastings (Metropolis *et al*, 1953, Hastings, 1970). Dans ce qui suit, nous présentons sommairement l'algorithme de Gibbs, mais on peut se référer à Casella & George (1992) pour une présentation plus détaillée et illustrée de cet algorithme.

4.3.1 L'échantillonneur de Gibbs

Principe

Soient le champ spatial \mathbf{X} , la matrice des interactions \mathbf{M} , la matrice de structure des interactions \mathbf{S} et les paramètres inconnus du modèle

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (p, \rho, \sigma^2). \quad (4.30)$$

En considérant \mathbf{S} comme paramètre implicite (pour ne pas surcharger nos notations), nous notons $\nu(\mathbf{X}, \mathbf{M}, \boldsymbol{\theta})$ la densité de probabilité conjointe de \mathbf{X} , \mathbf{M} , et $\boldsymbol{\theta}$. L'expression de cette mesure est connue en se référant à (2.7). En admettant que la matrice des interactions \mathbf{M} est inobservable, elle est à estimer en même temps que les paramètres inconnus du modèle.

L'algorithme de l'échantillonneur de Gibbs consiste à associer à \mathbf{M} une suite de valeurs $\mathbf{M}^{(i)}$ et au paramètre θ_i ($i = 1, 2, 3$) une suite de valeurs $\theta_i^{(j)}$ telles que, connaissant la réalisation \mathbf{x} du champ \mathbf{X} ,

$\mathbf{M}^{(j+1)}$ est une réalisation de probabilité $\nu(\mathbf{M}|\mathbf{x}, \theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)})$,

$\theta_1^{(j+1)}$ est une réalisation de densité de probabilité $\nu(\theta_1|\mathbf{x}, \mathbf{M}^{(j)}, \theta_2^{(j)}, \theta_3^{(j)})$,
 $\theta_2^{(j+1)}$ est une réalisation de densité de probabilité $\nu(\theta_2|\mathbf{x}, \mathbf{M}^{(j)}, \theta_1^{(j)}, \theta_3^{(j)})$,
 $\theta_3^{(j+1)}$ est une réalisation de densité de probabilité $\nu(\theta_3|\mathbf{x}, \mathbf{M}^{(j)}, \theta_1^{(j)}, \theta_2^{(j)})$.

Par itération, on construit $\mathbf{M}^{(j)}$, $\theta_1^{(j)}$, $\theta_2^{(j)}$ et $\theta_3^{(j)}$ ($j > 0$), en partant de valeurs initiales $\mathbf{M}^{(0)}$, $\theta_1^{(0)}$, $\theta_2^{(0)}$ et $\theta_3^{(0)}$ aléatoires. Sous certaines conditions, l'algorithme assure que les réalisations $\mathbf{M}^{(j)}$, $\theta_1^{(j)}$, $\theta_2^{(j)}$ et $\theta_3^{(j)}$ sont issues d'une loi de probabilité dont la densité de probabilité conditionnelle à \mathbf{x} est

$$\nu(\mathbf{M}, \boldsymbol{\theta}|\mathbf{x}) \tag{4.31}$$

si j tend vers l'infini. Pour des applications numériques, on se satisfait alors de considérer j suffisamment grand.

Application à l'estimation ponctuelle des paramètres

Si l'on applique l'algorithme n fois avec les mêmes données du champ, on obtient n suites pour respectivement \mathbf{M} et θ_i ($i = 1, 2, 3$). En ne retenant que le j -ème élément de ces suites, on parvient à construire en bonne approximation un échantillon de taille n de θ_i ($i = 1, 2, 3$). On peut alors estimer l'espérance des θ_i par la moyenne empirique de leurs échantillons respectifs :

$$E[\theta_i|\mathbf{x}] \approx \frac{1}{n} \sum_{k=1}^n \theta_{ik}^{(j-1)}, \tag{4.32}$$

où $\theta_{ik}^{(j-1)}$, est le j -ème élément de la k -ème suite associée à θ_i . Il convient immédiatement de relever que l'on ne fait pas de même avec la matrice des interactions \mathbf{M} étant donné que son espérance n'est pas pertinente pour estimer la réalisation de la matrice des interactions. La question de l'estimation de \mathbf{M} doit en fait être considérée avec beaucoup de précaution. En revanche, cette estimation n'est pas nécessaire dès lors que l'attention dans les modèles de Whittle probabilisés est portée sur la détermination des paramètres.

On rappelle que dans notre cas, $\theta_1 = p$, $\theta_2 = \rho$ et $\theta_3 = \sigma^2$. Lorsque $p = 0$ avec certitude, on se place dans un contexte classique bénéficiant des exemples et des méthodes que l'on retrouve dans la littérature scientifique. LeSage (1997) a illustré l'algorithme dans le cadre d'un modèle spatial autorégressif simultané homoscédastique, ainsi que pour une variante hétéroscédastique du modèle, en supposant en revanche que la matrice des interactions est connue. L'estimation de ρ nécessite un traitement spécifique (rendant la méthode moins pratique) par le fait d'un déterminant dans le calcul de la densité de probabilité conditionnelle.

Application à l'échantillonneur de Gibbs

Dans le même ordre d'idée, la méthode EM (décrite en section 4.1.4) entre en lien avec la méthode MCMC si l'on effectue l'approximation de (4.9) par

$$Q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) \approx \frac{1}{n} \sum_{i=1}^n \nu(\mathbf{x}, \mathbf{M}^{(j-1)}|\tilde{\boldsymbol{\theta}}), \quad (4.33)$$

où $\mathbf{M}_k^{(j-1)}$, est le j -ème élément de la k -ème suite associée à \mathbf{M} .

Calcul des densités de probabilité conditionnelles

Les densités de probabilité employées par l'algorithme ne sont pas connues directement. Elles doivent être calculées à partir de ce que l'on connaît. En l'occurrence, si l'on considère que les paramètres du modèle sont indépendants les uns des autres, on peut exprimer les densités de probabilité par

$$\nu(\theta_1|\mathbf{x}, \mathbf{M}, \theta_2, \theta_3) = \frac{\nu(\mathbf{x}, \mathbf{M}|\boldsymbol{\theta})\nu(\theta_1)}{\int \nu(\mathbf{x}, \mathbf{M}|\boldsymbol{\theta})\nu(\theta_1)d\theta_1}. \quad (4.34)$$

L'équation ci-dessus est valable également après permutation circulaire des indices de $\theta_i : 1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1$. L'intégrale dans (4.34) est un terme de normalisation qui peut être difficile à calculer. Etant donné le principe employé dans la méthode MCMC, il n'est pas nécessaire de connaître le terme de normalisation, évacuant de ce fait le problème de ce calcul. Cependant, la densité de probabilité conditionnelle de $\theta_2 = \rho$ pose tout de même quelques problèmes mathématiques, à cause du déterminant dans l'expression de (2.6). On peut appliquer la solution préconisée par LeSage (1997).

En ce qui concerne \mathbf{M} , sa probabilité conditionnelle s'exprime comme

$$\nu(\mathbf{M}|\mathbf{x}, \boldsymbol{\theta}) = \frac{\nu(\mathbf{x}, \mathbf{M}|\boldsymbol{\theta})}{\sum_{\mathbf{M}} \nu(\mathbf{x}, \mathbf{m}|\boldsymbol{\theta})}, \quad (4.35)$$

où la somme porte sur les éléments \mathbf{m} de l'univers de \mathbf{M} . Dans ce cas, également, il n'est pas nécessaire de connaître le terme de normalisation pour appliquer l'algorithme. Malgré cela, l'expression au numérateur de (4.35) pose quelques problèmes mathématiques à cause du déterminant dans l'expression de (2.6) et du fait que \mathbf{M} est une variable discrète avec un nombre élevé de degrés de liberté et de contraintes (le prior de \mathbf{M} est imposé par les hypothèses de travail).

Conclusion

On conclut que la méthode est applicable aux modèles de Whittle probabilisés, mais elle possède quelques difficultés de calcul inhérentes aux modèles autorégressifs. Le calcul de (4.35) (sans nécessairement son terme de normalisation) représente le principal défi (à relever) de la méthode.

Remarque Un champ spatial peut en cacher un autre. C'est le cas lorsqu'un champ spatial observable \mathbf{Y} est fonction d'un champ spatial caché \mathbf{X} . Si ce dernier est autorégressif et possède comme mesure de probabilité (4.31), alors les paramètres du modèle peuvent être estimés sur le même principe que ce qui précède. Il faut, pour cela, déterminer la mesure de probabilité conjointe $\nu(\mathbf{X}, \mathbf{M}, p, \rho, \sigma^2 | \mathbf{Y})$ et adapter en conséquence les probabilités conditionnelles de l'algorithme de l'échantillonneur de Gibbs. On trouve par exemple dans Weir & Pettitt (1999) une illustration de la méthode pour les modèles autorégressifs conditionnels (modèles CAR) dans le cas où le champ observé présente des données binaires.

4.4 Conclusion

Plusieurs études publiées employant le modèle de Whittle classique ont utilisé l'estimateur de maximum de vraisemblance de ρ pour en déterminer la valeur. Cet estimateur a en sa faveur de bonnes qualités statistiques : il est convergent et asymptotiquement efficace. Cependant, de manière critique, l'estimateur dépend de la spécification de la matrice des interactions du modèle, ce qui rend l'usage de cet estimateur scientifiquement critiquable malgré ses qualités statistiques.

Notre étude met en lumière ce que l'on appelle l'erreur de *spécification* des poids des interactions. Cette erreur est commise lorsque l'on spécifie l'existence d'une interaction là où il n'y en a pas ou lorsque l'on ne spécifie pas l'existence d'une interaction là où elle existe en réalité. Pour notre modèle spatial, considérer la matrice de structure des interactions en lieu et place du graphe des interactions qui s'est réalisé est un cas de sur-spécification des interactions qui conduit à une sous-estimation du coefficient d'autorégression ρ . Cette sous-estimation étant croissante avec p , nous concluons que ce paramètre agit comme un paramètre d'intensité de l'autocorrélation spatiale *perçue*, malgré qu'il soit un paramètre de topologie des interactions parfaitement indépendant du coefficient d'autorégression ρ .

Ce biais dans l'estimation de maximum de vraisemblance de ρ peut être simplement corrigée, à la condition que la valeur de p dont le biais en est fonction puisse être déterminée. Or, l'estimation de maximum de vraisemblance de p passe par l'information sur le nombre des interactions, ce qui, en méconnaissance du graphe des interactions, déplace finalement le problème vers l'estimation du nombre des interactions à partir des valeurs du champ. La question de l'estimation de p simultanément à celle de ρ reste donc ouverte.

L'estimateur de maximum de vraisemblance ayant montré ses faiblesses, il était pertinent de chercher d'autres méthodes d'estimation de ρ , qui soient plus simples de calcul (plus parcimonieux aussi) et moins sensibles au graphe des interactions du champ. Nous avons développé quelques méthodes alternatives d'estimation de ρ de sorte à montrer qu'il est possible d'obtenir, dans certaines conditions, une estimation de ρ quasiment sans biais et de manière robuste vis-à-vis de p (quitte à combiner les méthodes). Les méthodes alternatives d'estimation de ρ présentées nécessitent une analyse plus approfondie, mais nous retenons tout de même que

l'estimation par la méthode du I de Moran maximal est plus attractive que celle par la méthode de I de Moran. Sous certaines conditions de p et de ρ , ses performances sont par ailleurs équivalentes à celle de l'estimateur de maximum de vraisemblance dans le cas où celui-ci sur-spécifie les interactions du champ par leur matrice de structure. Nous relevons que parmi les méthodes que nous avons développées celle par comptage d'amas est la plus robuste vis-à-vis de p et son estimation de ρ semble peu souffrir d'un biais. Pour cela, cette méthode est une bonne découverte de ce chapitre, bien que son efficacité soit moindre.

Les méthodes EM et MCMC, que nous avons sommairement introduites, permettent d'estimer les paramètres du modèle sans nécessité d'observer la matrice des interactions. Ces méthodes peuvent donc, au moyen de quelques développements mathématiques et algorithmiques supplémentaires, apporter une solution à la problématique de la spécification des poids des interactions de la méthode MV.

Chapitre 5

Inférence du coefficient d'autorégression

5.1 Éléments de la théorie générale du test

5.1.1 Formulation des hypothèses du test

Pour tester un champ spatial \mathbf{Z} , on oppose deux hypothèses sur celui-ci, nommées respectivement *hypothèse nulle* H_0 et *hypothèse alternative* H_1 , et qui sont chacune respectivement une proposition logique \mathcal{P}_0 et \mathcal{P}_1 sur la distribution de probabilité de \mathbf{Z} . Le test statistique consiste à confirmer l'une des deux hypothèses et à infirmer l'autre en fonction des données observées. Dans la théorie des tests, on distingue les *tests paramétriques* des *tests non-paramétriques* : le test est paramétrique s'il porte sur des hypothèses faites au sujet des valeurs des paramètres d'une loi de probabilité paramétrique, le test est non paramétrique sinon. Pour chacune de ces deux classes de test, il existe une sous-classe de *tests d'existence d'autocorrélation spatiale* dont le résultat nous indique seulement si le champ spatial est spatialement indépendant, avec éventuellement une information sur la loi de probabilité y associée, ou s'il ne l'est pas ; le système d'hypothèses considéré est dans ce cas

$$\begin{cases} H_0 : \mathcal{P}_0, \\ H_1 : \mathcal{P}_1 = \text{non } H_0, \end{cases} \quad (5.1)$$

où \mathcal{P}_0 affirme avec plus ou moins de détails l'indépendance spatiale.

Parmi les tests non paramétriques testant un système de la forme (5.1), les plus parcimonieux sont ceux dont on ne formule aucune hypothèse sur la loi de probabilité du champ spatial. Dans ce registre nous relevons le *test de randomization* (Sokal & Oden, 1978), pour des données quantitatives, permettant de tester (5.1) avec H_0 affirmant que "*les variables du champ spatial \mathbf{X} sont mutuellement indépendantes*". Pour des données catégorielles (chaque valeur observée appartient à une catégorie) on relève le *test du chi-deux* pour lequel on pose l'hypothèse H_0 que "*les catégo-*

ries sont mutuellement indépendantes". Les tests non paramétriques plus raffinés spécifient une loi de probabilité de référence, comme, par exemple, pour le *test de Kolmogorov-Smirnov* pour lequel H_0 affirme que "la loi de \mathbf{X} a pour fonction de répartition F ".

Les tests paramétriques d'existence d'autocorrélation spatiale ont pour hypothèse H_0 une loi de probabilité paramétrique; le système d'hypothèses considéré est alors

$$\begin{cases} H_0 : \mathbf{X} \text{ suit la loi paramétrique } \mathcal{L}_\theta \text{ de paramètre } \theta = \theta_0, \\ H_1 : \text{non } H_0, \end{cases} \quad (5.2)$$

étant entendu que les variables du champ sont indépendantes les unes des autres lorsque $\theta = \theta_0$. Cette classe de test est utile pour valider ou invalider toute prise de décision statistique se basant sur l'hypothèse d'indépendance spatiale des données. Par contre, elle ne contribue pas à préciser la forme d'autocorrélation spatiale du champ dans le cas où H_0 est rejetée. Pour pallier à cet inconvénient, l'hypothèse H_1 peut désigner un modèle d'autocorrélation spatiale et le processus de prise de décision être adapté en conséquence. Parmi les possibles systèmes opposant deux lois paramétriques, nous relevons ceux de la forme

$$\begin{cases} H_0 : \mathbf{X} \sim \mathcal{L}_\theta \quad \text{avec } \theta \in \Theta_0, \\ H_1 : \mathbf{X} \sim \mathcal{L}_\theta \quad \text{avec } \theta \in \Theta_1, \end{cases} \quad (5.3)$$

où \mathcal{L}_θ est une loi de paramètre θ dont celui-ci prend ses valeurs dans l'ensemble Θ composé de deux sous-ensembles disjoints Θ_0 et Θ_1 . L'hypothèse H_i ($i = 0, 1$) est *simple* si l'ensemble Θ_i est un singleton, sinon elle est *composite*.

5.1.2 Règle de décision et risques

Soit le système d'hypothèses paramétriques simples ou composites formelles (5.3). Les hypothèses sont *testées* sur la base d'une réalisation \mathbf{x} de \mathbf{X} à l'aide d'une *variable de décision* (ou d'une *règle de décision*) δ associée à ce test. Cette variable est une application de $\Omega_{\mathbf{X}}$, l'ensemble des valeurs que peut prendre \mathbf{X} , vers $\{0, 1\}$, définie de sorte que

$$\begin{cases} \text{si } \delta(\mathbf{x}) = 0 \text{ alors } H_0 \text{ est rejetée et } H_1 \text{ est acceptée,} \\ \text{si } \delta(\mathbf{x}) = 1 \text{ alors } H_1 \text{ est rejetée et } H_0 \text{ est acceptée.} \end{cases} \quad (5.4)$$

On distingue alors deux régions dans $\Omega_{\mathbf{X}}$:

$$\begin{cases} \mathcal{Q} = \text{ensemble des } \mathbf{x} \in \Omega_{\mathbf{X}} \text{ tels que } \delta(\mathbf{x}) = 0, \\ \bar{\mathcal{Q}} = \text{ensemble des } \mathbf{x} \in \Omega_{\mathbf{X}} \text{ tels que } \delta(\mathbf{x}) = 1, \end{cases} \quad (5.5)$$

où l'ensemble \mathcal{Q} est appelé la *région critique du test* et l'ensemble $\bar{\mathcal{Q}}$ est appelé la *région d'acceptation du test*.

Parallèlement, les hypothèses divisent l'ensemble des paramètres Θ en deux sous-ensembles mutuellement exclusifs Θ_0 et Θ_1 . Pour un élément $\theta_0 \in \Theta_0$ (c.-à-d. pour un état de la nature associé à H_0), l'erreur (ou *risque*) de première espèce $\alpha(\theta_0)$ se calcule comme

$$\alpha(\theta_0) := \Pr(\mathcal{Q}|\mathcal{L}_{\theta_0}), \quad (5.6)$$

où $\Pr(\mathcal{Q}|\mathcal{L}_{\theta})$ est une notation (abusive) pour signifier "la probabilité de l'événement \mathcal{Q} étant donnée la loi de probabilité \mathcal{L}_{θ} ".

Pour un élément $\theta_1 \in \Theta_1$, donc pour un état de la nature associé à H_1 , l'erreur (ou *risque*) de seconde espèce $\beta(\theta_1)$ se calcule comme

$$\beta(\theta_1) := \Pr(\bar{\mathcal{Q}}|\mathcal{L}_{\theta_1}). \quad (5.7)$$

La fonction de puissance d'un test est la fonction $\pi(\theta)$ telle que

$$\pi(\theta) := \Pr(\mathcal{Q}|\mathcal{L}_{\theta}), \quad \theta \in \Theta, \quad (5.8)$$

et nous appelons *puissance* toute valeur ponctuelle de la fonction de puissance.

Le niveau ou seuil α du test est défini par

$$\alpha := \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} \pi(\theta). \quad (5.9)$$

En partant de ces concepts, il suffit maintenant de développer une *méthodologie de prise de décision* en ne se référant qu'aux règles de décision ou, de manière équivalente, aux régions critiques. Pour les tests opposant des hypothèses simples, Neyman et Pearson (1933) proposent que l'on recherche la région critique qui apporte la puissance de test que nulle autre construction de région critique puisse permettre de surpasser. Cette région est qualifiée de *meilleure région critique* et un test qui lui est basé est qualifié de *test le plus puissant*. Par définition, la recherche d'une telle région critique va dans le sens d'une prise de décision optimale puisqu'elle minimise, à risque de première espèce donnée, le risque de seconde espèce. Le lemme de Neyman-Pearson (1933) donne la forme de la variable de décision optimale lorsqu'on oppose une hypothèse H_0 simple contre une hypothèse H_1 simple : elle s'exprime comme

$$\delta(\mathbf{X}) := \begin{cases} 0 & \text{si } \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})} \geq \eta_\alpha \\ 1 & \text{si } \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})} < \eta_\alpha \end{cases}, \quad (5.10)$$

où f_0 et f_1 sont les fonctions de densité de probabilité des lois de probabilité induites par les hypothèses H_0 et H_1 respectivement ; le scalaire η_α est choisi de sorte que le risque de première espèce vaille α .

Lorsque l'hypothèse H_1 est composite, le lemme de Neyman-Pearson ne s'applique pas. On s'emploie de préférence à trouver, si elle existe, la règle de décision *uniformément la plus puissante* (en abrégé, UPP)¹ qui revient à trouver la règle

1. En anglais, la règle est dite "uniformly most powerful" (en abrégé, UMP)

de décision de puissance supérieure à toute autre règle de décision pour tout état de la nature associé à l'hypothèse alternative H_1 .

Un des tests les plus fréquemment employés, lorsqu'on n'arrive pas à trouver une règle UPP, est le *test du rapport des vraisemblances maximales* (cf. par ex. Kendall & Stuart, 1973). Il consiste à rejeter l'hypothèse nulle d'un système de la forme (5.3) si

$$\frac{\sup_{\boldsymbol{\theta}_0 \in \Theta_0} [f_{\boldsymbol{\theta}_0}(\mathbf{X})]}{\sup_{\boldsymbol{\theta}_1 \in \Theta_1} [f_{\boldsymbol{\theta}_1}(\mathbf{X})]} < \eta_\alpha, \quad (5.11)$$

où $f_{\boldsymbol{\theta}}$ est la densité de probabilité de la loi $\mathcal{L}_{\boldsymbol{\theta}}$; le scalaire η_α est choisi de sorte que le risque de première espèce vaille α .

La justification des règles (5.10) et (5.11) est donnée par la théorie de l'information en annexe de ce chapitre.

5.1.3 Statistiques de test

En pratique, dans beaucoup de problèmes, plutôt que de déterminer la meilleure région critique sur le domaine $\Omega_{\mathbf{X}}$ de toutes les valeurs possibles de la variable n -dimensionnelle $\mathbf{X} \sim \mathcal{L}_{\boldsymbol{\theta}}$, on détermine la meilleure région critique sur l'ensemble des valeurs que peut prendre une statistique $T(\mathbf{X}) \sim \mathcal{G}_{\boldsymbol{\theta}}$ de dimension inférieure à n . Les hypothèses à tester sont transformées en conséquence. Pour cet usage, cette statistique est appelée une *statistique de test*. Elle est employée pour discriminer l'une des deux hypothèses du système suivant :

$$\begin{cases} H'_0 : T(\mathbf{X}) \sim \mathcal{G}_{\boldsymbol{\theta}} & \text{avec } \boldsymbol{\theta} \in \Theta_0, \\ H'_1 : T(\mathbf{X}) \sim \mathcal{G}_{\boldsymbol{\theta}} & \text{avec } \boldsymbol{\theta} \in \Theta_1, \end{cases} \quad (5.12)$$

où $\mathcal{G}_{\boldsymbol{\theta}}$ est la loi de $T(\mathbf{X})$ induite par la loi $\mathcal{L}_{\boldsymbol{\theta}}$ de \mathbf{X} .

La méthode d'inférence est analogue à ce qui précède : la variable de décision est une fonction δ définie de sorte que l'hypothèse H'_0 est rejetée en faveur de H'_1 si $\delta(T(\mathbf{X})) = 0$ et H'_1 est rejetée en faveur de H'_0 si $\delta(T(\mathbf{X})) = 1$. La région critique et la région d'acceptation du test sont définies respectivement comme

$$\begin{cases} \mathcal{Q}_T = \text{ensemble des } t \in \Omega_T \text{ tels que } \delta(t) = 0, \\ \bar{\mathcal{Q}}_T = \text{ensemble des } t \in \Omega_T \text{ tels que } \delta(t) = 1, \end{cases} \quad (5.13)$$

où l'ensemble \mathcal{Q}_T est appelé *région critique du test pour T* et l'ensemble $\bar{\mathcal{Q}}_T$ est appelé la *région d'acceptation du test pour T* . Un rejet de H'_0 équivaut à un rejet de H_0 de (5.3) si \mathcal{Q}_T , respectivement $\bar{\mathcal{Q}}_T$, est l'ensemble image de l'ensemble $\bar{\mathcal{Q}}$, respectivement \mathcal{Q} , par la fonction T associés au système (5.3). Les notions d'erreur de première et seconde espèce, ainsi que de niveau et de puissance du test précédemment définies s'appliquent à la statistique de test par analogie.

5.2 Test du I de Moran

5.2.1 Une variante du test de Cliff-Ord

Formulation des hypothèses

Conformément aux hypothèses de travail formulées au Chapitre 2, nous supposons que les données sont la réalisation d'un modèle de Whittle markovien probabilisé :

$$\mathbf{X} = \rho \mathbf{M} \mathbf{X} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_{(\mathbf{0}, \sigma^2 \mathbf{I})}, \quad \mathbf{M} \sim \mathcal{M}_{(p, \mathbf{S})}. \quad (5.14)$$

Les variables de ce modèle sont conjointement le champ spatial \mathbf{X} et la matrice des poids des interactions \mathbf{M} . La loi paramétrique $\mathcal{L}_{\boldsymbol{\theta}}$ qui découle de la donnée (5.14) est donc celle du modèle de Whittle probabilisé de paramètre $\boldsymbol{\theta} = (\mathbf{S}, p, \rho, \sigma^2)$. Nous convenons qu'une réalisation de (\mathbf{X}, \mathbf{M}) nous permet de confirmer l'une des deux hypothèses du système suivant

$$\begin{cases} H_0 : (\mathbf{X}, \mathbf{M}) \sim \mathcal{L}_{\boldsymbol{\theta}_0} & \text{avec } \boldsymbol{\theta}_0 \in \{(\mathbf{S}, p, \rho_0, \sigma_0^2) \mid \sigma_0^2 \in \mathbb{R}_+, \rho_0 = 0\}, \\ H_1 : (\mathbf{X}, \mathbf{M}) \sim \mathcal{L}_{\boldsymbol{\theta}_1} & \text{avec } \boldsymbol{\theta}_1 \in \{(\mathbf{S}, p, \rho_1, \sigma_1^2) \mid \sigma_1^2 \in \mathbb{R}_+\}, \end{cases} \quad (5.15)$$

où H_0 revient à un modèle gaussien d'indépendance spatiale.

On sait que la mesure de probabilité $g_i(\mathbf{x}, \mathbf{m}) d\mathbf{x}$ d'une réalisation (\mathbf{x}, \mathbf{m}) du couple de variables (\mathbf{X}, \mathbf{M}) s'exprime sous hypothèse H_i ($i=0,1$) comme

$$g_i(\mathbf{x}, \mathbf{m}) = f_i(\mathbf{x}|\mathbf{m}) h_i(\mathbf{m}), \quad (5.16)$$

où $f_i(\mathbf{x}|\mathbf{m})$ est la densité de probabilité gaussienne conditionnelle (2.6) sous hypothèse H_i , et $h_i(\mathbf{m}) := \Pr(\mathbf{m}|\mathcal{L}_{\boldsymbol{\theta}_i})$ est la probabilité de \mathbf{m} sous hypothèse H_i . On constate aisément que les paramètres \mathbf{S} et p sont des paramètres *muets* pour la mesure $f_0(\mathbf{x}|\mathbf{m})$ qui dans ce cas peut être notée $f_0(\mathbf{x})$. Cependant, on convient que la valeur de ces deux paramètres est pareille pour H_0 et pour H_1 . En effet, par la divergence de Kullback-Leibler entre les deux mesures de probabilité (5.16), on constate que

$$K(g_0, g_1) = \sum_{\mathbf{m}} \left(\int_{-\infty}^{\infty} g_0(\mathbf{x}, \mathbf{m}) \ln \left(\frac{g_0(\mathbf{x}, \mathbf{m})}{g_1(\mathbf{x}, \mathbf{m})} \right) d\mathbf{x} \right) \quad (5.17)$$

$$= \sum_{\mathbf{m}} \left(\int_{-\infty}^{\infty} h_0(\mathbf{m}) f_0(\mathbf{x}|\mathbf{m}) \ln \left(\frac{f_0(\mathbf{x}|\mathbf{m}) h_0(\mathbf{m})}{f_1(\mathbf{x}|\mathbf{m}) h_1(\mathbf{m})} \right) d\mathbf{x} \right) \quad (5.18)$$

$$= \sum_{\mathbf{m}} h_0(\mathbf{m}) K_{\mathbf{X}|\mathbf{m}}(f_0, f_1) + \int_{-\infty}^{\infty} f_0(\mathbf{x}) K_{\mathbf{M}}(h_0, h_1) \quad (5.19)$$

$$= E_{\mathbf{M}}[K_{\mathbf{X}|\mathbf{M}}(f_0, f_1)] + K_{\mathbf{M}}(h_0, h_1). \quad (5.20)$$

Le second terme de la dernière équation dépendant des fonctions h_i , ($i = 1, 2$). Or, si l'on admet que ce terme puisse être non nul, nous ouvrons la possibilité que l'on puisse rejeter H_0 pour ρ_1 arbitrairement proche de 0. Ce n'est pas le but recherché et nous imposons donc que $K_{\mathbf{M}}(h_0, h_1) = 0$. Ceci implique que $h_0 = h_1$.

Selon une règle de décision encore à déterminer, le rejet de H_0 permet d'inférer un champ spatial de paramètres $(\mathbf{S}, p, \rho_1, \sigma_1^2)$, avec la particularité que le paramètre σ_1^2 est a priori inconnu. En cas de non-rejet de H_0 , un champ spatial de paramètres $(\mathbf{S}, p, 0, \sigma_0^2)$ est inféré, avec la particularité que le paramètre σ_0^2 est a priori inconnu.

Variable de décision

Le système (5.15) étant composite, le lemme de Neyman-Pearson n'est pas applicable, mais nous pouvons employer le test du rapport des vraisemblances maximales. La statistique de test à considérer est de la forme (5.11) que l'on peut employer sous une forme logarithmique :

$$T(\mathbf{X}, \mathbf{M}) := \ln \left(\frac{\sup_{\sigma_1^2 \in \mathbb{R}_+} g_1(\mathbf{X}, \mathbf{M})}{\sup_{\sigma_0^2 \in \mathbb{R}_+} g_0(\mathbf{X}, \mathbf{M})} \right) \quad (5.21)$$

$$= \ln \left(\frac{\sup_{\sigma_1^2 \in \mathbb{R}_+} f_1(\mathbf{X}|\mathbf{M})}{\sup_{\sigma_0^2 \in \mathbb{R}_+} f_0(\mathbf{X})} \right). \quad (5.22)$$

Conformément à la règle de décision pour les tests de maximum de vraisemblance, on devrait rejeter H_0 si $T(\mathbf{X}, \mathbf{M}) \geq \eta$, pour un scalaire η bien choisi. Cependant, l'existence du déterminant rend la manipulation de la statistique peu pratique lorsque la taille du champ est très grande. La substitution de σ_i^2 ($i=0,1$) par son estimateur de maximum de vraisemblance, suivi d'un développement de Taylor limité autour de $\rho = 0$ de la statistique résultante permet de constater le résultat intéressant suivant (voir annexe de ce chapitre) :

$$T(\mathbf{X}, \mathbf{M}) = nI_{\mathbf{M}}(\mathbf{X})\rho_1 + O(\rho_1^2), \quad (5.23)$$

où n est la taille du champ et $I_{\mathbf{M}}$ est l'indice de Moran de matrice de poids spatiaux \mathbf{M} . Ainsi donc, la règle de décision δ s'appliquant aux hypothèses (5.15) et définie par

$$\delta(\mathbf{X}, \mathbf{M}) := \begin{cases} 0 & \text{si } I_{\mathbf{M}}(\mathbf{X}) \geq \eta \\ 1 & \text{si } I_{\mathbf{M}}(\mathbf{X}) < \eta \end{cases}, \quad (5.24)$$

approche la règle de décision du test de rapport des vraisemblances maximales lorsque la valeur de ρ_1 s'approche de 0.

Le test du I de Moran employé pour tester le système (5.15) dans le cas où $p = 0$ (et par conséquent avec \mathbf{M} considéré comme un paramètre) est parfois appelé un *test de Cliff-Ord*. Sa puissance a fait l'objet d'études, au travers de simulations numériques, de la part de Cliff & Ord (1973, 1975), Anselin & Rey (1991), Anselin & Florax (1995), Florax & Rey (1995). King (1981) démontre mathématiquement qu'il est le meilleur test invariant localement à $\rho = 0$. Tiefelsdorf (2000) le qualifie comme "*un bon test pour un usage général dans le contexte de modèles autorégressifs simultanés (SAR) et à moyenne coulissante (MA)*", tout en sachant que des meilleurs tests, plus spécifiques au problème traité, peuvent exister.

Remarque A notre connaissance, aucune étude a été faite sur la puissance du test du I de Moran dans le cadre général d'un modèle de Whittle probabilisé.

Puissance du test

Si l'état de la Nature est l'une des deux hypothèses suivantes

$$\begin{cases} H_0 : (\mathbf{X}, \mathbf{M}) \sim \mathcal{L}_{\boldsymbol{\theta}_0} & \text{avec } \boldsymbol{\theta}_0 = (\mathbf{S}, p, \rho_0, \sigma_0^2), \rho_0 = 0, \\ H_1 : (\mathbf{X}, \mathbf{M}) \sim \mathcal{L}_{\boldsymbol{\theta}_1} & \text{avec } \boldsymbol{\theta}_1 = (\mathbf{S}, p, \rho_1, \sigma_1^2), \end{cases} \quad (5.25)$$

alors la règle de décision (5.24), appliquée à (5.25), possède une fonction de puissance $\Pi(\boldsymbol{\theta})$ qui se calcule formellement par

$$\Pi(\boldsymbol{\theta}_i) := \Pr(\mathcal{Q}_{T(\mathbf{X}, \mathbf{M})} | \mathcal{L}_{\boldsymbol{\theta}_i}) \quad (i = 0, 1), \quad (5.26)$$

où $\mathcal{Q}_{T(\mathbf{X}, \mathbf{M})}$ est la région critique du test composée des réalisations potentielles (\mathbf{x}, \mathbf{m}) de (\mathbf{X}, \mathbf{M}) telles que $T(\mathbf{x}, \mathbf{m}) = I_{\mathbf{m}}(\mathbf{x}) \geq \eta$.

Il convient de relever à ce stade le rôle de la puissance d'un test. Comme formulé mathématiquement par (5.26), la puissance du test est la probabilité d'accepter l'hypothèse H_1 lorsque l'état de la Nature correspond effectivement à H_1 . Une valeur élevée de la puissance est toujours recommandée. Une valeur trop faible de puissance (au sens de celle ou celui qui applique le test) réserve aux données une erreur de seconde espèce trop élevée, dans ce sens le test a un trop faible pouvoir de discrimination entre H_0 et H_1 , et doit convaincre celle ou celui qui applique le test de repenser les divers éléments du test, notamment la taille de l'échantillon et/ou la règle de décision, dont fait partie la statistique de test.

Le principal obstacle à une étude théorique de la puissance (5.26) est la forme mathématique du I de Moran. En effet, l'indice de Moran est un rapport de deux formes quadratiques dont il est difficile de déterminer une expression analytique simple de sa distribution de probabilité (voir Tiefelsdorf (2002) pour un résumé des différentes techniques). En annexe de ce chapitre, nous contribuons modestement à donner analytiquement les deux premiers moments centrés de cette statistique et donnons quelques éléments utiles pour approcher sa distribution de probabilité.

5.2.2 Limites à la puissance du test

La puissance $\Pi(\boldsymbol{\theta})$ est difficile à calculer, mais on peut chercher à connaître une limite supérieure raisonnable qu'elle ne peut dépasser en fonction du paramètre p . En désignant par g_i la densité de probabilité du couple de variables (\mathbf{X}, \mathbf{M}) , et par v_i la densité de probabilité de la statistique $T(\mathbf{X}, \mathbf{M})$ de (\mathbf{X}, \mathbf{M}) sous les hypothèses H_i ($i=0,1$) formulée dans (5.25), on peut relever les inégalités suivantes :

$$\mathfrak{F}(\Pi(\boldsymbol{\theta}_0), \Pi(\boldsymbol{\theta}_1)) \leq K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1) \leq K_{(\mathbf{X}, \mathbf{M})}(g_0, g_1), \quad (5.27)$$

où le premier terme est défini comme

$$\mathfrak{F}(\alpha, \pi) := \alpha \ln\left(\frac{\alpha}{\pi}\right) + (1 - \alpha) \ln\left(\frac{1 - \alpha}{1 - \pi}\right) \quad (5.28)$$

et les termes $K_{(\mathbf{X}, \mathbf{M})}(g_0, g_1)$ et $K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1)$ sont les divergences de Kullback-Leibler entre les densités de probabilité g_0 et g_1 , respectivement entre les densités de probabilité v_0 et v_1 . La première inégalité à droite² de (5.27) est un résultat connu de la théorie de l'information (cf. p. ex. Kullback, 1959, théorème 4.1) qui exprime le fait qu'une fonction des données ne peut que maintenir ou dégrader l'information. La seconde inégalité³ de (5.27) est un résultat également bien connu de la théorie de l'information qui vient de la propriété de positivité et de convexité de la divergence de Kullback-Leibler (cf. p. ex. Kullback, 1959).

La fonction $\mathfrak{F}(\alpha, \pi)$ est une fonction convexe de π définie pour $\pi \in [0, 1]$ et prenant ses valeurs sur l'intervalle $[0, \infty)$. Par conséquent, borner $\mathfrak{F}(\alpha, \pi)$ à α constant revient à confiner les valeurs de π dans un intervalle continu plus restreint que l'intervalle $[0, 1]$. Le maximum de cet intervalle restreint est la plus grande valeur π^* apportant l'égalité

$$\mathfrak{F}(\alpha, \pi^*) = K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1) \quad (5.31)$$

et donne, pour $\alpha := \Pi(\boldsymbol{\theta}_0)$, une limite supérieure à la puissance du test $\Pi(\theta)$ qui est spécifique à la statistique de test employée. Selon cette dernière, la divergence $K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1)$ n'est pas aisée à calculer. Aussi, trouve-t-on une limite supérieure à la puissance du test $\Pi(\theta)$ qui est indépendante de toute statistique de test en recherchant la plus grande valeur π^* apportant l'égalité

$$\mathfrak{F}(\alpha, \pi^*) = K_{(\mathbf{X}, \mathbf{M})}(g_0, g_1). \quad (5.32)$$

Cette limite est généralement plus grande que celle définie par (5.31), mais est dans notre cas plus aisée à calculer.

Dans le cas de faibles valeurs du coefficient d'autorégression, on peut employer l'approximation à l'ordre 2 de la divergence de Kullback-Leibler dans (5.32). Dans ce cas, cela revient à rechercher la plus grande valeur π^* telle que

$$\mathfrak{F}(\alpha, \pi^*) = \frac{1}{2} J_{(\mathbf{X}, \mathbf{M})}(0) \rho^2 + O(\rho^3) \quad (5.33)$$

en négligeant le terme $O(\rho^3)$. Le comportement de cette limite supérieure de la puissance peut être ainsi aisément analysé en fonction du paramètre p , de la même manière que nous l'avons fait pour l'information de Fisher $J_{(\mathbf{X}, \mathbf{M})}(0)$ au Chapitre 2.

2. La borne peut être atteinte dans le cas suivant :

$$K_{(\mathbf{X}, \mathbf{M})}(g_0, g_1) = K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1) \Leftrightarrow \frac{g_0(\mathbf{X}, \mathbf{M})}{g_1(\mathbf{X}, \mathbf{M})} = \frac{v_0(T(\mathbf{X}, \mathbf{M}))}{v_1(T(\mathbf{X}, \mathbf{M}))} \quad (5.29)$$

3. La borne peut être atteinte dans le cas suivant :

$$K_{T(\mathbf{X}, \mathbf{M})}(v_0, v_1) = \mathfrak{F}(\Pi(\boldsymbol{\theta}_0), \Pi(\boldsymbol{\theta}_1)) \Leftrightarrow \begin{cases} \frac{v_0(T)}{v_1(T)} = \frac{\Pi(\boldsymbol{\theta}_0)}{\Pi(\boldsymbol{\theta}_1)} & \text{si } T \in \mathcal{Q}_T \\ \frac{v_0(T)}{g_1(T)} = \frac{1 - \Pi(\boldsymbol{\theta}_0)}{1 - \Pi(\boldsymbol{\theta}_1)} & \text{si } T \in \bar{\mathcal{Q}}_T \end{cases} \quad (5.30)$$

5.3 Efficacité asymptotique du test

5.3.1 Définition générale

Les critères qui permettent de qualifier un test de plus efficace qu'un autre se basent habituellement sur le risque découlant d'une règle de décision du test. En général, on procède en fixant au préalable l'erreur de première espèce et en adoptant l'erreur de deuxième espèce (ou la puissance) comme critère d'efficacité du test. Mais on peut également fixer au préalable la valeur de l'erreur de deuxième espèce (ou la puissance) et adopter l'erreur de première espèce comme critère d'efficacité du test.

Lorsqu'il s'agit de comparer l'efficacité à tester le système d'hypothèses (5.3) avec deux statistiques de test différentes sous de mêmes conditions d'observation, il existe la possibilité de fonder un critère d'*efficacité relative* en calculant le rapport des puissances induites par les deux statistiques. Une autre approche consiste à calculer le rapport des tailles minimales des échantillons respectifs qui sont nécessaires à chaque statistique de test pour induire une puissance donnée. Ce critère n'est cependant pas concis, car il dépend de l'erreur de première espèce α , de l'erreur de seconde espèce β et de la taille minimale n_1 de l'échantillon pour la statistique T_1 (ou bien de la taille minimale n_2 de l'échantillon pour la statistique T_2). L'idée est alors de considérer l'efficacité relative pour des tailles d'échantillon de plus en plus grandes et observer le rapport n_2/n_1 à la limite où les échantillons sont de taille infinie. Il y a plusieurs façons de faire croître la taille des échantillons. En admettant que la loi de probabilité du champ spatial \mathbf{Z} étudié soit une loi paramétrique \mathcal{L}_θ avec pour hypothèse H_0 " $\theta = \mathbf{0}$ " et pour hypothèse H_1 " $\theta = \theta_1 \neq \mathbf{0}$ ", nous pouvons faire croître la taille des échantillons selon l'un ou l'autre des deux principes formulés par Cochran (1952, p. 323), c'est-à-dire

- en demandant que θ_1 converge vers $\mathbf{0}$ à β et α constants tout en exigeant qu'une puissance de référence soit atteinte (dans ce cas, on nécessite effectivement de plus grands échantillons pour mieux discriminer les hypothèses), auquel cas, le rapport des tailles des échantillons n_2/n_1 converge vers une quantité appelée *efficacité de Pitman*.
- en demandant que α converge vers 0 à θ_1 et β constants tout en exigeant, là également, qu'une puissance de référence soit atteinte (dans ce cas, on nécessite de plus grands échantillons pour diminuer l'erreur de première espèce), auquel cas, le rapport des tailles des échantillons n_2/n_1 converge vers une quantité appelée *efficacité de Bahadur* (Bahadur, 1967).

Signalons que d'autres mesures d'efficacité existent comme l'efficacité de Kalenberg (1983) (qui applique les deux principes de Cochran simultanément), ou l'efficacité de Walsh (1946).

5.3.2 Efficacité de Pitman

Dans ce qui suit, nous déterminons l'efficacité de Pitman pour le cas où le paramètre θ se ramène à un scalaire θ . Les références bibliographiques des principaux résultats de cette partie sont données à la fin de cette sous-section.

Soit donc une suite de systèmes d'hypothèses simples de terme général

$$\begin{cases} H_0 : \mathbf{Z} \sim \mathcal{L}_\theta & \text{avec } \theta = 0, \\ H_1 : \mathbf{Z} \sim \mathcal{L}_\theta & \text{avec } \theta = \theta_\nu, \end{cases} \quad (5.34)$$

où \mathcal{L}_θ est une loi de probabilité de paramètre θ , θ_ν est le terme général d'une suite de paramètres, notée $\{\theta_\nu\}_{\nu=1}^\infty$, prenant ses valeurs sur un ensemble Θ continu au voisinage de $\theta = 0$ et convergeant vers $\theta = 0$. Au ν -ème élément de cette suite sont associées deux statistiques de test $T_{\nu,i}$ ($i = 1, 2$) opérant respectivement sur des échantillons spatiaux de taille $n_{\nu,i}$ (à déterminer) de sorte à avoir deux suites de tests dont on requiert qu'elles convergent chacune vers un test de niveau de signification α et de puissance $1 - \beta \in (\alpha, 1)$. Si $\{\pi_{\nu,i}(\theta_\nu)\}_{\nu=1}^\infty$ est la suite de puissances de la i -ème suite de tests y associée, alors nous définissons $n_{\nu,i}$ comme étant le nombre minimal d'observations tel que

$$\pi_{\nu,i}(0) \leq \alpha \quad \text{et} \quad \pi_{\nu,i}(\theta_\nu) \geq 1 - \beta. \quad (5.35)$$

Si elle existe, la limite du rapport

$$\mathcal{E}_P(T_1, T_2) := \lim_{\nu \rightarrow \infty} \frac{n_{\nu,2}}{n_{\nu,1}} \quad (5.36)$$

permet de comparer l'efficacité des deux suites de tests. Le rapport (5.36) est appelé *efficacité relative asymptotique* ou *efficacité de Pitman* de la première suite de tests par rapport à la seconde suite. Une efficacité relative supérieure à 1 signifie que moins d'observations sont nécessaires pour la première suite de tests qu'avec la seconde. Dans ce cas, la première suite est considérée comme la meilleure.

Sous certaines conditions, nous sommes en mesure de donner le résultat de (5.36). En gardant les mêmes notations que précédemment, nous admettons que la statistique de test $T_{\nu,i}$ ($i = 1, 2$) (opérant sur des échantillons de taille $n_{\nu,i}$) possède une distribution de probabilité qui converge en loi vers la loi normale lorsque $\nu \rightarrow \infty$, ou, corollairement, lorsque la taille $n_{\nu,i}$ de l'échantillon tend vers l'infini :

$$\lim_{\nu \rightarrow \infty} \frac{\sqrt{n_{\nu,i}}(T_{\nu,i} - \mu(T_{\nu,i}))}{\sigma(T_{\nu,i})} \sim \mathcal{N}_{(0,1)} \quad (i = 1, 2), \quad (5.37)$$

où $\mu(T_{\nu,i})$ et $\sigma(T_{\nu,i})$ sont respectivement l'espérance et l'écart-type de $T_{\nu,i}$. Dans ce cas, nous connaissons le résultat de (5.36) par le théorème suivant.

Théorème 2: *Soit une famille continue de lois de probabilité \mathcal{L}_θ paramétrée par un paramètre $\theta \geq 0$ telle que $\forall n[(\theta \rightarrow 0) \Rightarrow (\|\mathcal{L}_\theta - \mathcal{L}_0\| \rightarrow 0)]$. Soient $\{T_{n,1}\}_{n=1}^\infty$ et $\{T_{n,2}\}_{n=1}^\infty$ des suites de statistiques de test, opérant sur des échantillons de taille n ,*

qui satisfont la convergence (5.37) pour toutes suites $\{\theta_n\}_{n=1}^\infty$ qui converge vers 0 et des fonctions μ_i et σ_i telles que μ_i est dérivable en zéro et σ_i est continue en zéro, avec $\frac{d\mu_i(0)}{d\theta} > 0$ et $\sigma_i(0) > 0$. Sous ces conditions, l'efficacité relative asymptotique des tests qui rejettent l'hypothèse nulle $H_0 : \theta = 0$ pour des valeurs suffisamment grandes de $T_{n,i}$ est égale à

$$\mathcal{E}_P(T_1, T_2) = \left(\frac{\frac{d\mu_1(0)}{d\theta} \sigma_2(0)}{\sigma_1(0) \frac{d\mu_2(0)}{d\theta}} \right)^2 \quad (5.38)$$

pour toutes suites $\{\theta_n\}_{n=1}^\infty$ convergentes en $\theta = 0$, indépendantes de $\alpha > 0$ et $1 - \beta \in (\alpha, 1)$. Si les fonctions de puissance des tests basés sur les statistiques $T_{n,i}$ sont des fonctions monotones non décroissantes de $\theta \forall n$, alors la condition de normalité asymptotique des $T_{n,i}$ peut être affaiblie en la condition de normalité asymptotique pour chaque suite de terme général $\theta_n = O(\frac{1}{\sqrt{n}})$. \diamond

On trouvera la formulation et la démonstration de ce théorème dans van der Vaart (1998). A l'origine, l'approche par l'efficacité asymptotique a été introduite par Pitman (1948-1949) dans un ensemble de notes de cours (non publiées). Une preuve d'un résultat un peu plus général peut être trouvée dans Noether (1955). On trouve une discussion de l'efficacité de Pitman dans Kendall & Stuart (1973, chap. 25). L'efficacité de Pitman a été exploitée pour des modèles SAR par Cliff & Ord (1973, 1981).

5.3.3 Application

On identifie le système (5.34) au système (5.15) de sorte que \mathbf{Z} soit identifié à (\mathbf{X}, \mathbf{M}) et la loi \mathcal{L}_{θ_ν} soit identifié à celle du modèle de Whittle markovien probabilisé obtenue de la donnée (5.14) dont $\theta_\nu := \rho_\nu$ représente le terme général d'une suite de coefficients d'autorégression convergente vers 0, les autres paramètres (c'est-à-dire \mathbf{S} , p et σ^2) étant fixés et invariants sous les différentes hypothèses. On considère ensuite les deux statistiques suivantes :

$$T_1(\mathbf{X}, \mathbf{M}) := I_{\mathbf{M}}(\mathbf{X}), \quad T_2(\mathbf{X}, \mathbf{M}) := I_{\mathbf{W}}(\mathbf{X}). \quad (5.39)$$

La première statistique est l'indice de Moran qui ressort de l'approximation au premier ordre de la statistique du rapport de maximum de vraisemblance (5.23) : sa matrice de poids spatiaux coïncide avec la matrice des interactions du champ. La seconde statistique est l'indice de Moran de matrice de poids spatiaux \mathbf{W} arbitraire, cette dernière étant indépendante des variables (\mathbf{X}, \mathbf{M}) . L'efficacité de Pitman $\mathcal{E}_P(T_1, T_2)$ de l'inférence du coefficient d'autorégression ρ d'un modèle de Whittle (de paramètres \mathbf{S} , p et σ^2) vaut (voir annexe)

$$\mathcal{E}_P(T_1, T_2) = \frac{\mathbb{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2]] \text{Tr}[\mathbf{W}\mathbf{W}' + \mathbf{W}^2]}{\mathbb{E}_{\mathbf{M}}[\text{Tr}^2[\mathbf{M}'\mathbf{W} + \mathbf{M}\mathbf{W}]]}. \quad (5.40)$$

On interprète T_1 comme plus efficace que T_2 (pour le test) si $\mathcal{E}_P(T_1, T_2) > 1$; dans le cas où l'inégalité est inversée, on interprète T_1 comme moins efficace que T_2 (pour le test).

Par parcimonie, nous décidons d'identifier \mathbf{W} à une matrice d'adjacence, nous contentant ainsi de supposer un graphe (non valué) des relations d'autocorrélation spatiale entre les variables du champ. Si $\mathbf{W} := \mathbf{S}$, c'est-à-dire si \mathbf{W} est identique à la matrice de structure des interactions du modèle, on a alors

$$\mathcal{E}_P(T_1, T_2) = \frac{1}{2} J_{(\mathbf{X}, \mathbf{M})}(0) \frac{\tau[\mathbf{S}]}{(n - \mathbf{E}_{\mathbf{M}}[n_0])^2}, \quad (5.41)$$

où $J_{(\mathbf{X}, \mathbf{M})}(0)$ est l'information de Fisher (3.14) et $\tau[\cdot]$ est la fonction qui totalise le nombre d'éléments non nuls de la matrice donnée en argument.

Illustration

La figure 5.1 illustre la courbe d'efficacité de Pitman $\mathcal{E}_P(T_1, T_2)$ en fonction de p estimée par une simulation de Monte Carlo pour un modèle de Whittle dont \mathbf{S} représente un réseau carré périodique de 100×100 sites (avec $\sigma^2 = 1$). Elle montre que, asymptotiquement, la puissance avec $T_2(\mathbf{X}, \mathbf{M}) = I_{\mathbf{S}}(\mathbf{X})$ est relativement plus petite que la puissance obtenue avec la statistique $T_1(\mathbf{X}, \mathbf{M}) = I_{\mathbf{M}}(\mathbf{X})$. La divergence entre les deux puissances est une fonction monotone croissante non linéaire de p .

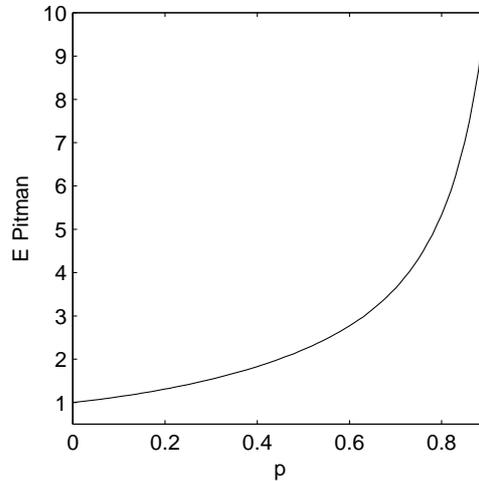


Fig. 5.1: Efficacité de Pitman (5.41) en fonction de p .

5.4 Test du I de Moran

5.4.1 Spécification du test

Dans cette partie, nous voulons calculer la puissance du test du I de Moran pour le modèle de Whittle markovien probabilisé dont sa loi \mathcal{L}_θ , rappelons-le, est paramétrique de paramètre $\theta = (\mathbf{S}, p, \rho, \sigma^2)$ conformément à la donnée (5.14). Plus précisément, nous voulons tester le système d'hypothèses simples suivant :

$$\begin{cases} H_0 : \mathcal{L}_{\theta_0} \text{ avec } \theta_0 = (\mathbf{S}, p, \rho_0, \sigma_0^2), \sigma_0^2 = 1, \rho_0 = 0, \\ H_1 : \mathcal{L}_{\theta_1} \text{ avec } \theta_1 = (\mathbf{S}, p, \rho_1, \sigma_1^2), \sigma_1^2 = 1, \end{cases} \quad (5.42)$$

où \mathbf{S} est (sous H_0 et H_1) la matrice de structure des interactions représentant un réseau carré périodique (c.-à-d. fermé sur un tore). Les paramètres testés sont p et ρ .

Nous considérons la statistique de test T_2 de (5.39), soit $I_{\mathbf{W}}(\mathbf{X})$. Ce choix permet d'analyser des cas de "mauvaise spécification" de la matrice des poids spatiaux \mathbf{W} , c'est-à-dire des cas où celle-ci ne coïncide pas avec la matrice des interactions \mathbf{M} du champ. Nous considérons les matrices de poids spatiaux suivantes (avec les notations correspondantes) :

- $\mathbf{W} = \mathbf{W}_{R1}, \mathbf{W}_{R2}, \dots, \mathbf{W}_{Rk}$, où \mathbf{W}_{Rk} est la matrice d'adjacence avec adjacence entre deux sommets si et seulement si la distance euclidienne qui les sépare est égale à k pas du réseau. Nous nous limitons aux cas $k \leq 6$. Les figures 5.2(a) et 5.3 illustrent la construction du graphe.
- $\mathbf{W} = \mathbf{W}_{B1}$, où \mathbf{W}_{B1} est la matrice d'adjacence avec adjacence entre deux sommets si et seulement si la distance euclidienne qui les sépare est égale à $\sqrt{2}$ pas du réseau. La figure 5.2(b) illustre le graphe représenté par \mathbf{W}_{B1} .
- $\mathbf{W} = \mathbf{W}_{Q1}$, où \mathbf{W}_{Q1} est la matrice d'adjacence avec adjacence entre deux sommets si et seulement si la distance euclidienne qui les sépare est supérieure à 0 et inférieure ou égale à $\sqrt{2}$ pas du réseau. La figure 5.2(c) illustre le graphe représenté par \mathbf{W}_{Q1} .

Pour la comparaison, nous considérons également la statistique T_1 de (5.39), soit $I_{\mathbf{M}}(\mathbf{X})$, qui est applicable lorsque la matrice des interactions \mathbf{M} du champ est observable.

Nous avons vu au Chapitre 2 que la corrélation entre variables d'un modèle de Whittle markovien probabilisé est fonction décroissante de la distance qui les sépare. Pour une valeur de ρ suffisamment petite, on s'attend à ce que la corrélation ne soit importante qu'à l'échelle des interactions des variables. Il en est autrement si la valeur de ρ est suffisamment grande, raison pour laquelle nous convenons de diviser l'étude de la puissance du test du I de Moran entre le cas où les valeurs de ρ sont faibles (autocorrélation spatiale faible) et le cas où les valeurs de ρ sont plus élevées.

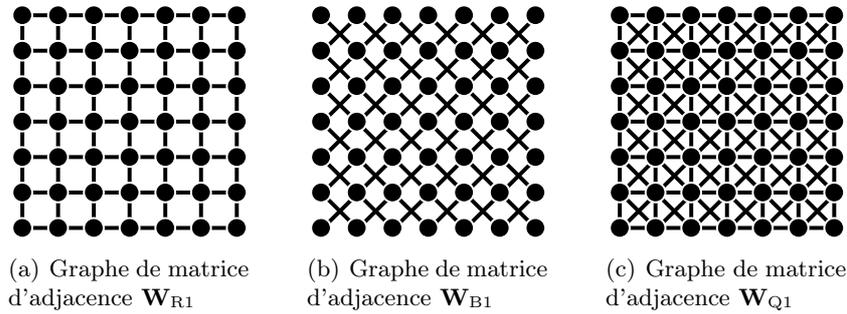


Fig. 5.2: Graphes pour la matrice des poids spatiaux.

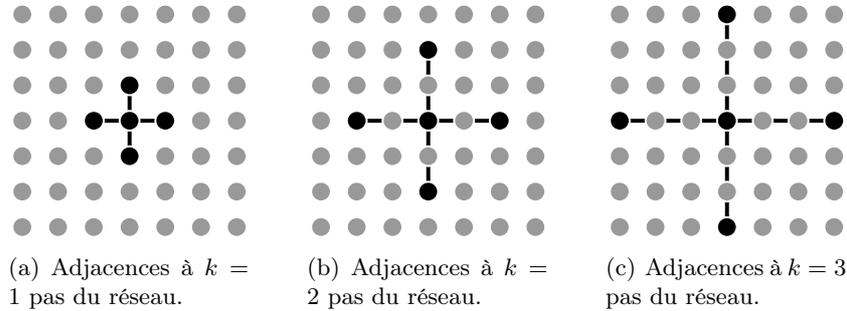


Fig. 5.3: Illustration des adjacences d'un sommet avec son voisinage selon la configuration générale induite par \mathbf{W}_{Rk} .

5.4.2 Puissance pour des valeurs de ρ faibles

Méthode

La détermination exacte de la distribution de probabilité de I étant problématique, nous déterminons la puissance du test à l'aide de simulations numériques du champ, en nous inspirant du mode opératoire de Cliff & Ord (1975). Chaque réalisation du champ est obtenue à partir d'une réalisation indépendante du vecteur des chocs $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ et en se basant sur la relation

$$\mathbf{X} = (\mathbf{I} - \rho\mathbf{M})^{-1}\boldsymbol{\epsilon}. \quad (5.43)$$

Avec $\rho_1 > 0$ dans H_1 , la puissance du test, de seuil α , est obtenue empiriquement selon les deux cas suivants :

- **Cas où la statistique de test est $T_1(\mathbf{X}, \mathbf{M}) = I_{\mathbf{M}}(\mathbf{X})$:** on calcule l'indice de Moran pour k réalisations indépendantes du couple de variables (\mathbf{X}, \mathbf{M}) sous l'hypothèse H_0 de sorte à en déterminer la fonction de répartition empirique F_{0p_1} qui dépend de p_1 . On en déduit le quantile $I_{1-\alpha}$ tel que $F_{0p_1}(I_{1-\alpha}) = 1 - \alpha$. Ensuite, on calcule un échantillon d'indices de Moran $I_{\mathbf{M}}(\mathbf{X})$ avec k' réalisations

indépendantes du couple (\mathbf{X}, \mathbf{M}) sous l'hypothèse H_1 , et on estime finalement la puissance du test en calculant la fraction des indices de Moran de cet échantillon qui sont plus grands que le quantile $I_{1-\alpha}$.

- **Cas où la statistique de test est $T_2(\mathbf{X}, \mathbf{M}) = I_{\mathbf{W}_{R1}}(\mathbf{X})$** : on effectue la même procédure que ci-dessus en employant la pondération fixe \mathbf{W}_{R1} (qui est identique à la matrice de structure \mathbf{S} des interactions) pour le calcul du I de Moran. La fonction de répartition F_{0p_1} est indépendante de p_1 dans ce cas.

Pour $\rho_1 < 0$, la procédure d'estimation de la puissance est la même, mais on adopte le fractile I_α au lieu de $I_{1-\alpha}$ et on calcule la fraction des indices de Moran de l'échantillon qui sont plus petits que le quantile I_α .

Résultats

Les courbes de puissance issues des simulations numériques pour $\rho_1 > 0$ sont illustrées dans les figures 5.4(a) et 5.4(b). Elles suffisent à décrire également la puissance pour $\rho < 0$ étant donné que la puissance apparaît dans nos simulations comme une fonction symétrique de ρ pour les deux cas de pondération spatiale considérés.

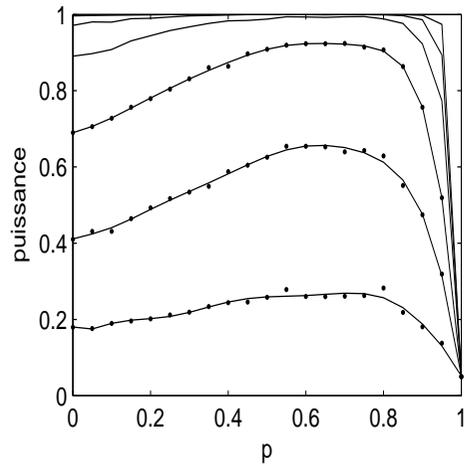
À p constant, pour la statistique de test T_1 comme pour T_2 , la puissance croît avec ρ . Par contre, à ρ constant, le comportement de la puissance pour T_1 et pour T_2 respectivement sont fondamentalement différents. Avec l'emploi de T_1 , à ρ suffisamment proche de 0, la puissance du test n'est pas une fonction monotone décroissante de p . En réalité, la puissance s'accroît sensiblement lorsque p augmente au voisinage de 0 et atteint un maximum aux conditions proches de celles qui conduisent au maximum de l'information de Fisher de ρ . Avec l'emploi de T_2 , en revanche, on observe des fonctions de puissance qui sont monotones décroissantes, montrant que la pondération spatiale employée dans la statistique de test est propre à neutraliser ce qui, par propriété de la connectivité du champ, devait apporter un gain de puissance par rapport à la configuration des interactions à $p = 0$. En l'occurrence, T_2 discrimine d'autant mieux les hypothèses du système que le champ approche de son état de connexité maximale.

La différence de résultat entre les deux statistiques est mise en évidence par la figure 5.4(c) donnant la différence (gain) de puissance découlant du passage de la statistique T_2 à la statistique T_1 . On observe que le maximum de gain de puissance suit une ligne de crête de hauteur variable sur le plan des paramètres (p, ρ) .

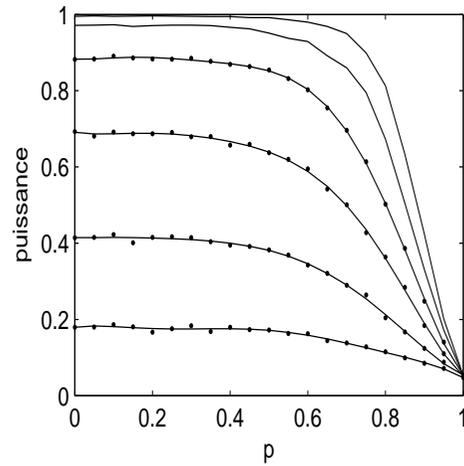
5.4.3 Puissance pour des valeurs de ρ plus élevées

Méthode et résultats

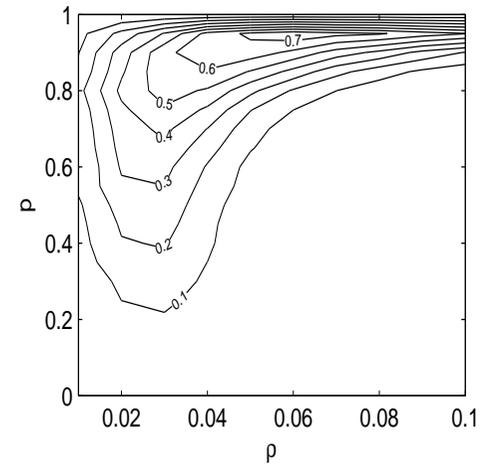
Nous calculons la puissance du test statistique employant une statistique de test $T_2(\mathbf{X}, \mathbf{M}) = I_{\mathbf{W}}(\mathbf{X})$ de la même manière que pour le cas où $\mathbf{W} = \mathbf{W}_{R1}$ précédemment décrit, avec p variant de 0 à 1 et ρ variant cette fois-ci de 0 à 0.95. Les figures 5.5 et 5.6 présentent les résultats pour \mathbf{W} identifiée respectivement à



(a) Puissance du test du I de Moran avec T_1 . Les courbes correspondent respectivement, de bas en haut, à $\rho = 0.01, 0.02, \dots, 0.06$. Les courbes passant entre les points sont des ajustements polynomiaux aux données. Chaque courbe a été obtenue avec 10^4 réalisations d'un champ spatial sur un réseau carré périodique de taille 100×100 , à $p = 0.0, 0.05, \dots, 1.0$.



(b) Puissance du test du I de Moran avec T_2 . Les courbes correspondent respectivement, de bas en haut, à $\rho = 0.01, 0.02, \dots, 0.06$. Les courbes passant entre les points sont des ajustements polynomiaux aux données. Chaque courbe a été obtenue avec 10^4 réalisations d'un champ spatial sur un réseau carré périodique de taille 100×100 , à $p = 0.0, 0.05, \dots, 1.0$.



(c) Gain de puissance du passage de T_2 à T_1 , obtenu des courbes de la figure 5.4(a) et 5.4(b), en fonction des paramètres p et ρ .

Fig. 5.4:

la matrice d'adjacence \mathbf{W}_{R1} , \mathbf{W}_{R2} , \mathbf{W}_{R3} , \mathbf{W}_{R4} , \mathbf{W}_{R5} , \mathbf{W}_{R6} , \mathbf{W}_{B1} et \mathbf{W}_{Q1} . En complément, les différences de puissance résultant du passage d'une statistique à une autre sont illustrées dans les figures 5.7 et 5.8. Ces figures appellent les commentaires qui suivent.

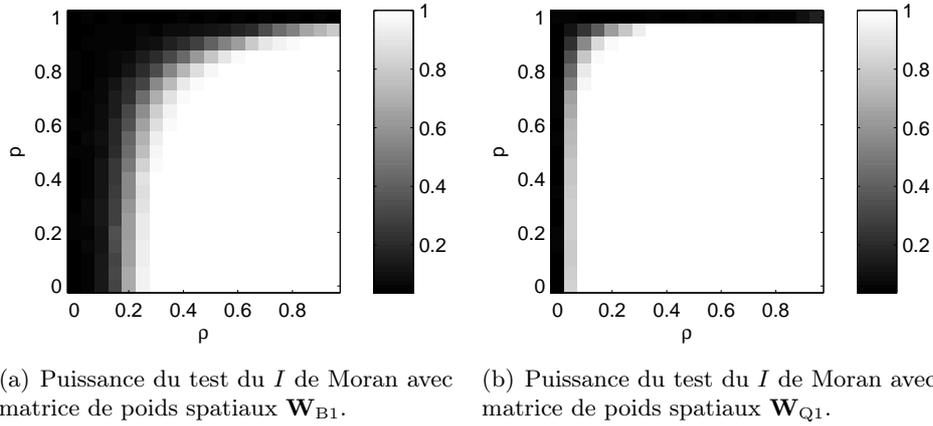


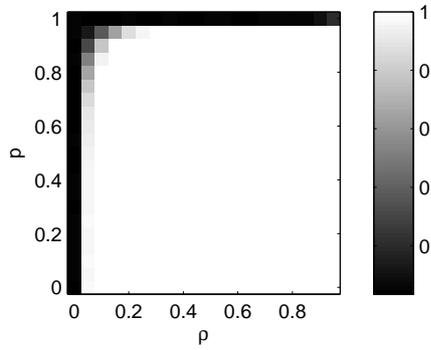
Fig. 5.5:

Commentaires

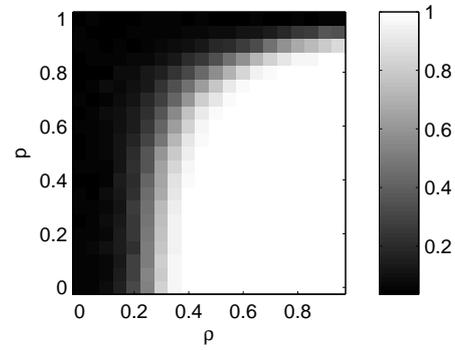
Effets conjugués de p et ρ On observe, communément à toutes les statistiques de test, que la puissance s'accroît avec ρ et décroît avec p . Nos résultats laissent apprécier comment la puissance se modifie en fonction des paramètres p et ρ , notamment comment, lorsque p décroît, il est nécessaire que le coefficient ρ augmente pour retrouver un test de même puissance (c.-à-d. pour suivre les isoplèthes de la fonction de puissance).

Effet de la distance de voisinage (corrélogramme) Pour chaque réalisation \mathbf{x} du champ \mathbf{X} , la suite d'indices de Moran $I_{\mathbf{W}_{R1}}(\mathbf{x})$ ($i = 1, 2, \dots$) constitue un corrélogramme. Nos résultats indiquent que pour toute valeur de p et de ρ la puissance du test est fonction décroissante de i , illustrant le fait que la puissance décroît avec l'écart euclidien des sites mis en adjacence au sens de la pondération spatiale. La puissance obtenue avec la pondération spatiale \mathbf{W}_{B1} corrobore ce schéma de décroissance.

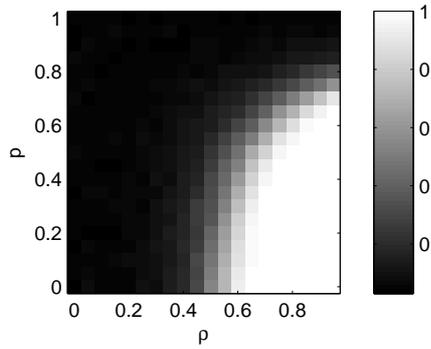
Effet de la connectivité de la pondération La pondération \mathbf{W}_{Q1} , qui est égale à la pondération cumulée $\mathbf{W}_{R1} + \mathbf{W}_{B1}$, met en relation d'adjacence les sites écartés d'une distance euclidienne de 1 ou $\sqrt{2}$ pas du réseau, et, par ce fait, la pondération \mathbf{W}_{Q1} est doublement connective. Nous observons que le passage de \mathbf{W}_{Q1} à \mathbf{W}_{R1} , respectivement de \mathbf{W}_{Q1} à \mathbf{W}_{B1} , s'accompagne d'une perte de puissance (de



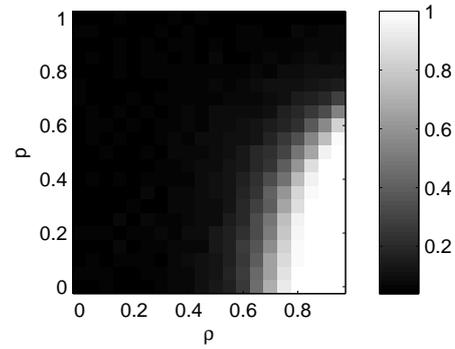
(a) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R1} .



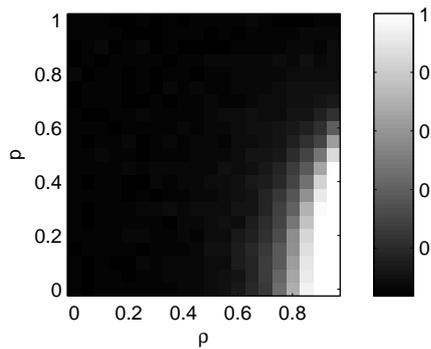
(b) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R2} .



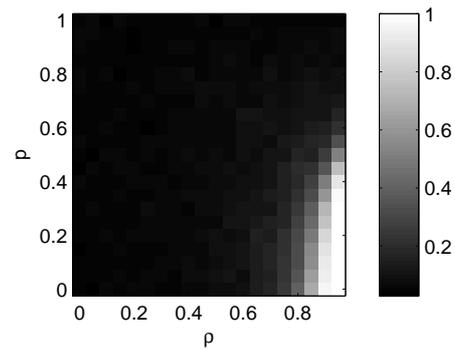
(c) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R3} .



(d) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R4} .

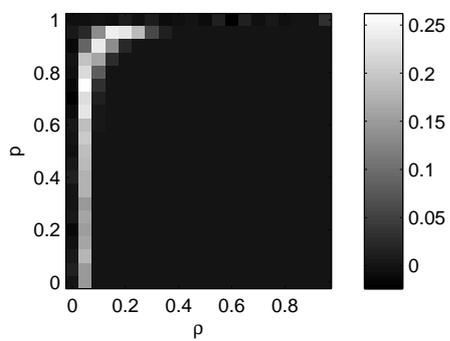


(e) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R5} .

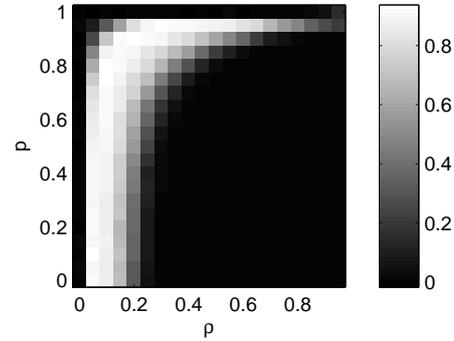


(f) Puissance du test du I de Moran avec matrice de poids spatiaux \mathbf{W}_{R6} .

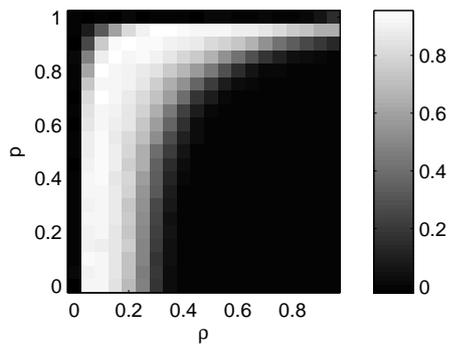
Fig. 5.6:



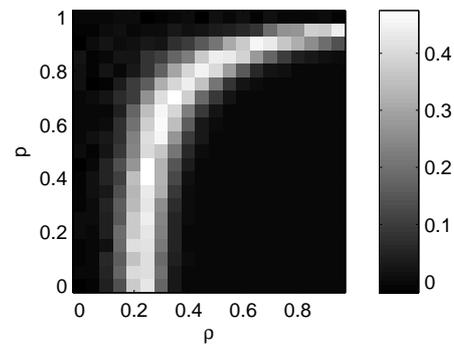
(a) Gain de puissance du test du I de Moran au passage de $I_{W_{Q1}}$ à $I_{W_{R1}}$.



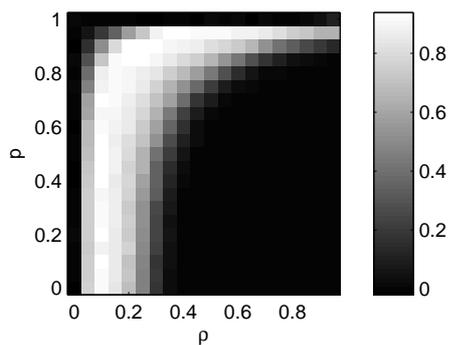
(b) Gain de puissance du test du I de Moran au passage de $I_{W_{B1}}$ à $I_{W_{R1}}$.



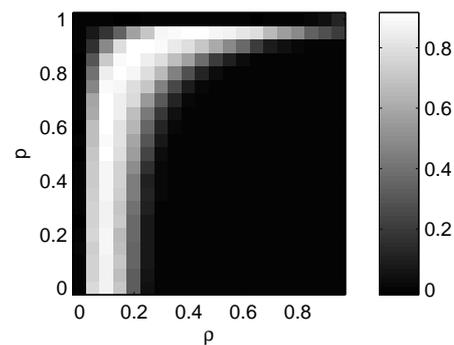
(c) Gain de puissance du test du I de Moran au passage de $I_{W_{R2}}$ à $I_{W_{R1}}$.



(d) Gain de puissance du test du I de Moran au passage de $I_{W_{R2}}$ à $I_{W_{B1}}$.



(e) Gain de puissance du test du I de Moran au passage de $I_{W_{R2}}$ à $I_{W_{Q1}}$.



(f) Gain de puissance du test du I de Moran au passage de $I_{W_{B1}}$ à $I_{W_{Q1}}$.

Fig. 5.7:

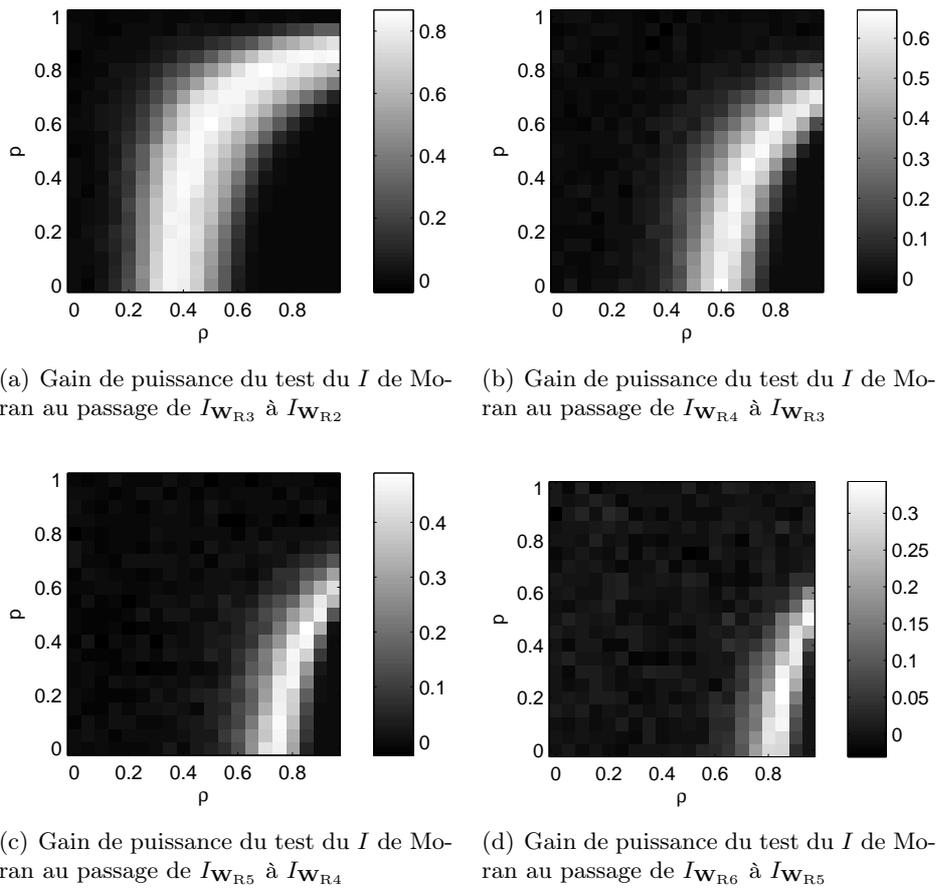


Fig. 5.8:

faible magnitude), respectivement d'un gain de puissance (très élevé pour certaines conditions de p et ρ).

5.5 Test du I de Moran maximal

5.5.1 Spécification du test

L'indice de Moran maximal, I_{\max} , a été abordé au Chapitre 4 pour estimer le coefficient d'autorégression, toutefois sans avoir abordé sa capacité de description de l'autocorrélation spatiale via ses poids spatiaux ajustés aux divers ordres de voisinage. La raison en est que l'analyse des poids spatiaux n'a de sens que si on aborde en même temps la question de sa significativité au sens de la théorie des tests statistiques (Oden, 1984). Par conséquent, nous présentons dans ce qui suit une analyse du comportement du I de Moran maximal à la fois comme statistique descriptive (dont les résultats sont en annexe de ce chapitre) et comme statistique de test. Pour finir, nous nous employons à faire, avec le I de Moran maximal, une analyse de puissance de test de manière analogue à ce qui a été fait précédemment avec le I de Moran.

Nous considérons à nouveau le système d'hypothèses (5.42) pour lequel le I de Moran maximal joue le rôle de statistique de test. Cette dernière est construite selon la méthode expliquée en section 4.2.3 avec les contraintes (4.17) et avec l'algorithme de construction des voisinages y exposé.

La méthode pour accepter ou rejeter l'hypothèse H_0 sur la base d'une observation consiste à calculer l'indice I_{\max} pour k réalisations indépendantes du couple de variables (\mathbf{X}, \mathbf{M}) sous l'hypothèse H_0 de sorte à en déterminer la fonction de répartition empirique F_0 (indépendante du paramètre p). On en déduit le quantile $I_{1-\alpha}$ tel que $F_0(I_{1-\alpha}) = 1 - \alpha$. La règle de décision du test consiste à rejeter H_0 si le I de Moran maximal observé est supérieur à $I_{1-\alpha}$.

5.5.2 Puissance du test

La fonction de puissance du test se basant sur la statistique de test du I de Moran maximal peut être estimée de manière identique à la méthode que nous avons employée pour estimer la puissance du test du I de Moran. Cette fonction est estimée sur la base du système d'hypothèses (5.42), en faisant varier les paramètres p et ρ de H_1 . La figure 5.9 montre le résultat pour un modèle dont la structure des interactions est un réseau carré de taille 100×100 sites. Pour comparaison, la différence de puissance du passage de la statistique de test $I_{\mathbf{W}_{R1}}$ à la statistique de test I_{\max} , ainsi que la différence de puissance du passage de la statistique de test $I_{\mathbf{W}_{R2}}$ à la statistique de test I_{\max} , en fonction des paramètres p et ρ , sont données dans la figure 5.10.

Il en ressort que la puissance du test du I_{\max} n'est pas meilleure que celle du test du $I_{\mathbf{W}_{R1}}$. Le passage de $I_{\mathbf{W}_{R1}}$ à I_{\max} engendre une perte de puissance au voisinage de $\rho = 0$ (selon la valeur de p) de manière analogue, mais de magnitude

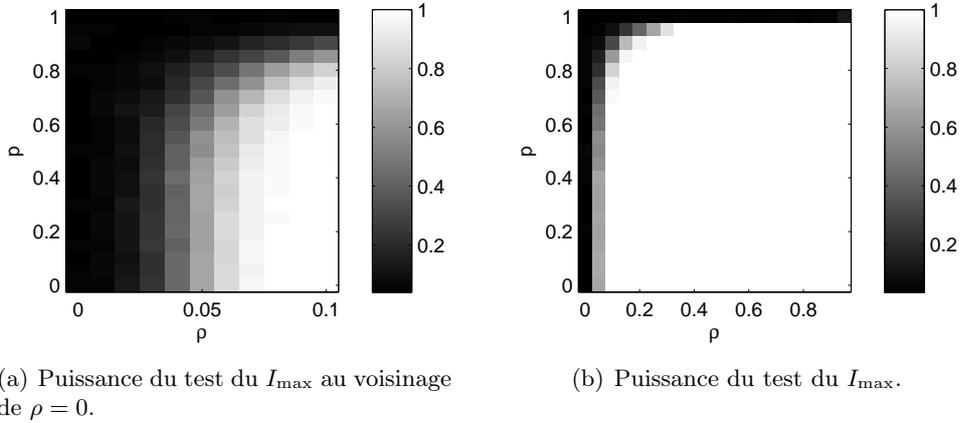


Fig. 5.9: Puissance du test pour un modèle de Whittle markovien probabilisé de 100×100 (réseau carré).

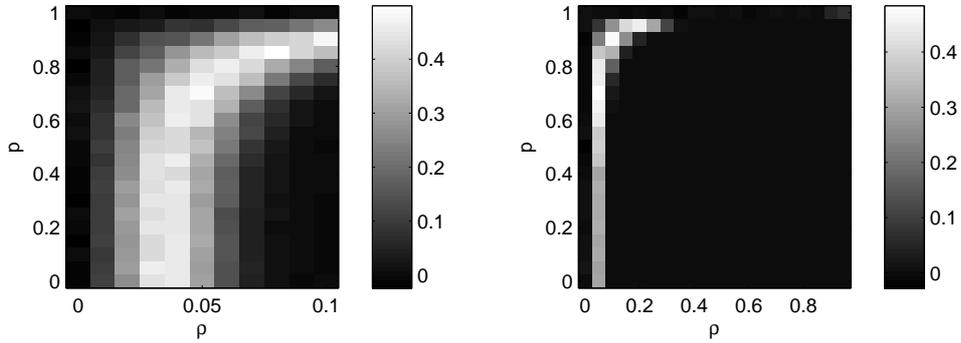
moindre, au passage de $I_{\mathbf{W}_{R1}}$ à $I_{\mathbf{W}_{Q1}}$ (cf. figure 5.7(a)). La puissance des deux tests convergent l'une vers l'autre à valeur de ρ suffisamment élevée, illustrant les conditions auxquelles une pondérations spatiale plus complexe que \mathbf{W}_{R1} permet de décrire la corrélation spatiale et de rejeter l'hypothèse d'indépendance H_0 avec tout de même autant d'efficacité qu'une pondération élémentaire comme \mathbf{W}_{R1} .

L'ajustement des poids spatiaux du I de Moran maximal semble avoir pour "coût" une perte de puissance. En contre partie il engendre une description de l'autocorrélation spatiale sous forme de corrélogramme. Le comportement des poids spatiaux sous différentes conditions de ρ et p peut être vu dans les diagrammes relégués en annexe de ce chapitre (fig. F.1 à F.13).

5.6 Parallèle dans le paradigme bayésien

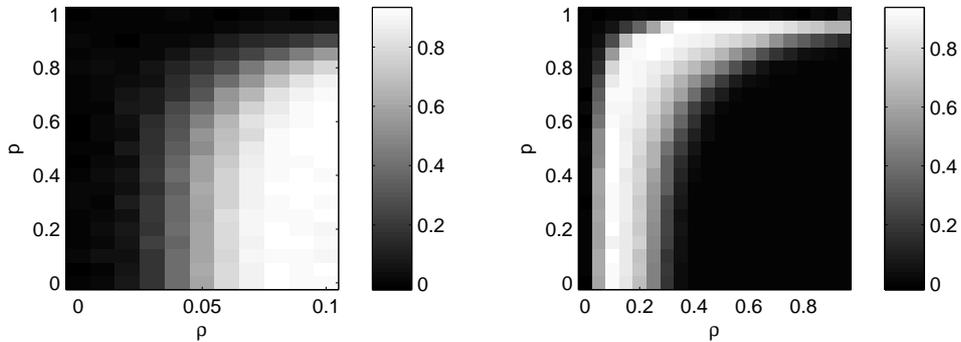
Nous relevons que notre analyse de la puissance concernant les modèles de Whittle probabilisés est utile à l'étude de l'inférence de modèles de Whittle classiques dans l'optique bayésienne où la matrice \mathbf{M} des interactions est un paramètre de valeur inconnue du modèle possédant une probabilité.

Nous considérons tout d'abord que le champ spatial suit un modèle de Whittle classique, et qu'il nous intéresse de savoir s'il est significativement autocorrélé ou non. Pour cela, on pose comme hypothèse H_0 un modèle de Whittle de paramètre $\rho = 0$, ce qui est équivalent à un modèle gaussien d'indépendance spatiale, que l'on rejette ou accepte par un test statistique. La méthodologie usuelle pour la construction du test suppose que, à seuil α du test fixé, le choix de la statistique du test soit celui qui confère au test le plus de puissance en regard de l'hypothèse alternative. En supposant que la matrice des interactions qui prévaut dans le champ est \mathbf{M} , la statistique de test devrait être choisie en fonction du système d'hypothèses



(a) Gain de puissance du test du I de Moran au passage de I_{\max} à $I_{\mathbf{W}_{R1}}$ au voisinage de $\rho = 0$.

(b) Gain de puissance du test du I de Moran au passage de I_{\max} à $I_{\mathbf{W}_{R1}}$.



(c) Gain de puissance du test du I de Moran au passage de $I_{\mathbf{W}_{R2}}$ à I_{\max} au voisinage de $\rho = 0$.

(d) Gain de puissance du test du I de Moran au passage de $I_{\mathbf{W}_{R2}}$ à I_{\max} .

Fig. 5.10: Gain de puissance du test pour un modèle de Whittle markovien probabilisé de 100×100 (réseau carré).

suivant :

$$\begin{cases} H_0 : \mathcal{L}_{\xi_0} \text{ avec } \xi_0 = (\mathbf{M}, \rho_0, \sigma_0^2), \sigma_0^2 = 1, \rho_0 = 0, \\ H_1 : \mathcal{L}_{\xi_1} \text{ avec } \xi_1 = (\mathbf{M}, \rho_1, \sigma_1^2), \sigma_1^2 = 1. \end{cases} \quad (5.44)$$

Bien que l'hypothèse H_0 soit indépendante de \mathbf{M} (à cause de la valeur nulle de ρ_0), il en est autrement avec H_1 . Dans l'optique bayésienne, ce système possède une probabilité $\Pr(\mathbf{M})$ d'être celui qu'il convient de poser pour tester la significativité de la valeur de ρ_0 . Par conséquent, après avoir choisi la règle de décision et fixé le seuil du test, l'erreur de seconde espèce associée à ce système d'hypothèse ne serait probable qu'avec une probabilité $\Pr(\mathbf{M})$. Dans ces conditions, le choix optimal de la statistique de test devrait se faire en minimisant l'erreur de seconde espèce *moyenne* calculée sur l'ensemble des systèmes d'hypothèse possibles. Aussi, la règle

de décision à choisir est celle qui minimise la quantité

$$\mathfrak{R}(\delta) := \mathbf{E}_{\mathbf{M}}[\beta(\boldsymbol{\xi}_1)], \quad (5.45)$$

où $\boldsymbol{\xi}_1 = (\mathbf{M}, \rho_1, \sigma_1^2)$ est le paramètre de l'hypothèse H_1 dont la probabilité est égale à celle de \mathbf{M} .

La quantité (5.45) est une forme dérivée de la notion de risque de Bayes de la règle de décision, qui est employée dans la théorie des jeux et est applicable aux tests statistiques (Ferguson, 1967). La quantité (5.45) permet de comparer les règles de décision entre elles, en fonction de la distribution de probabilité de \mathbf{M} a priori et d'un seuil du test fixé. Certaines règles engendrent un risque plus grand que les autres. Le principe de Bayes adapté dans notre cas consiste à choisir la règle δ qui minimise le risque $\mathfrak{R}(\delta)$. En supposant que la distribution de probabilité \mathbf{M} correspond à la distribution de probabilité objective que nous avons fixée pour la matrice des interactions de nos modèles de Whittle markoviens probabilisés, à savoir (voir section 2.1)

$$\mathbf{M} \sim \mathcal{M}_{(p, \mathbf{S})}, \quad (5.46)$$

alors les fonctions de puissance présentées précédemment pour les modèles de Whittle probabilisés sont directement applicables pour analyser le risque (5.45) dans le cas où la règle de décision est indépendante du graphe des interactions. En effet, en considérant le fait que la réalisation \mathbf{x} du champ spatial \mathbf{X} conjointement à la réalisation \mathbf{m} de la matrice des interactions d'un modèle de Whittle probabilisé ont la mesure de probabilité (2.7), à savoir $g_1(\mathbf{x}, \mathbf{m})d\mathbf{x} = f_1(\mathbf{x}|\mathbf{m})h_1(\mathbf{m})d\mathbf{x}$ sous l'hypothèse H_1 , alors l'erreur de seconde espèce de la règle de décision δ pour le système d'hypothèses (5.25) s'exprime comme

$$\beta(\boldsymbol{\theta}_1) = \sum_{\mathbf{m} \in \Omega_{\mathbf{M}}} \int_{-\infty}^{\infty} g_1(\mathbf{x}, \mathbf{m})\delta(\mathbf{x}, \mathbf{m})d\mathbf{x} \quad (5.47)$$

$$= \sum_{\mathbf{m} \in \Omega_{\mathbf{M}}} h_1(\mathbf{m}) \int_{-\infty}^{\infty} f_1(\mathbf{x}|\mathbf{m})\delta(\mathbf{x}, \mathbf{m})d\mathbf{x} \quad (5.48)$$

$$= \sum_{\mathbf{m} \in \Omega_{\mathbf{M}}} h_1(\mathbf{m})\beta(\boldsymbol{\xi}_1) \quad (5.49)$$

$$= \mathbf{E}_{\mathbf{M}}[\beta(\boldsymbol{\xi}_1)]. \quad (5.50)$$

Par conséquent, on observe que

$$\mathfrak{R}(\delta) = 1 - \pi(\boldsymbol{\theta}_1), \quad (5.51)$$

où $\pi(\boldsymbol{\theta}_1) = 1 - \beta(\boldsymbol{\theta}_1)$ est la puissance des règles de décision étudiés dans ce chapitre, à l'exception de celle qui est associée à la statistique de test $T(\mathbf{X}, \mathbf{M}) = I_{\mathbf{M}}(\mathbf{X})$.

5.7 Conclusion

Le principe méthodologique sous-tendu par nos tests statistiques est de pouvoir décider si une valeur prédéfinie, ou éventuellement un intervalle de valeurs, est

acceptable à la lumière des observations, en l'occurrence des valeurs de ρ . Les tests statistiques sont recommandés avant d'effectuer l'estimation d'une valeur de paramètre. En effet, il n'aurait pas été logique de procéder à l'estimation de ρ si l'hypothèse H_0 ne pouvait être réfutée avec les données. Cependant il faut veiller à ce que le test soit choisi de sorte à ce que la probabilité de rejeter H_0 à raison soit la plus élevée possible, ou autrement dit que sa puissance soit la plus élevée possible. L'étude que nous venons d'effectuer a présenté des résultats intéressants. Un certain nombre d'entre eux étant techniques, le lecteur est renvoyé aux figures et commentaires. Quelques éléments sont tout de même à relever.

- La différence de puissance entre deux points du plan des paramètres (ρ, p) apparaît instructif, notamment si l'un d'eux tombe sur l'axe $p = 0$. En l'occurrence, nos courbes ont permis de voir quel changement de risque se produisait en passant du cas $(\rho = \rho_1, p = p_1)$ (modèle de Whittle probabilisé) au cas $(\rho = \rho_1, p = 0)$ (modèle de Whittle classique) pour inférer ρ_1 . Selon la statistique de test choisie ne bénéficiant pas de l'observation de la connexité du champ, le changement est faible sur un large intervalle de p au voisinage de $p = 0$, ce qui autorise de choisir indifféremment le modèle de connexité.
- Les isoplèthes de puissance sur le plan des paramètres (ρ, p) sont également instructifs. Ils signifient qu'un test préparé pour accepter l'hypothèse $(\rho = \rho_1, p = 0)$ (un modèle de Whittle classique) avec une puissance désirée peut détecter avec la même puissance un effet dû au cas $(\rho = \rho'_1, p = p'_1)$ (un modèle de Whittle probabilisé) où $\rho'_1 > \rho_1$ et $p'_1 > p_1$.
- A faible régime de ρ , il existe des conditions de connexité des interactions pour lesquelles la puissance d'une règle de décision bénéficiant de l'observation de la connexité du champ est plus élevée que dans le cas où $p = 0$. Cet effet, remarquable, est produit par le mécanisme identifié au Chapitre 3 qui module l'information de Fisher de ρ . Nous observons que des statistiques de test ne bénéficiant pas de l'observation de la connexité du champ, tel que par exemple l'indice de Moran $I_{\mathbf{W}_{R1}}$, neutralise cet effet et donne une puissance (au plus équivalente) monotone décroissante de p .
- La puissance du test est sensible à la pondération spatiale choisie pour l'indice de Moran, notamment à faible régime de ρ . Le choix d'adopter, dans l'ignorance de la matrice des interactions, une pondération spatiale dont le graphe coïncide avec celui de structure des interactions s'avère le plus avantageux en terme de puissance, même si elle diffère avec la réalisations de la matrice des interactions.
- A valeur de ρ suffisamment élevée, la corrélation est apparente entre variables dont l'écart est supérieur à la distance inter-sites : il devient raisonnable de tester la corrélation spatiale à ces distances, avec cependant perte de puissance avec la distance. La plus grande distance de corrélation qu'il soit encore raisonnable de tester (selon la puissance) augmente avec ρ et diminue avec p .

Chapitre 6

Conclusion

Un modèle original

La Nature produit plusieurs formes d'autocorrélation spatiale. L'une d'elles s'observe par le fait que les mesures ou observations qui sont proches dans l'espace ont une *tendance* à être proches dans les résultats. Dans notre cas, la compréhension de ladite *tendance* passe par un concept de champ purement spatial gaussien. Nous avons convenu que ces variables font partie d'un *modèle de Whittle markovien probabilisé* - un modèle autorégressif d'ordre 1 de paramètre d'autorégression ρ et de paramètre de connexité p - qui définit chaque variable comme étant la moyenne des variables les plus proches dans l'espace, à un terme aléatoire près, suivant une connectivité locale aléatoire. On préfère ce modèle au *modèle de Whittle classique* pour deux raisons :

- la proximité spatiale entre deux variables du champ est une condition nécessaire mais non suffisante pour qu'une interaction existe entre elles.
- les interactions entre variables du champ peuvent changer dans le temps

Le modèle de Whittle probabilisé et le modèle de Whittle classique diffèrent donc par leur modèle de connexité, que l'on ressent au niveau des variances et covariances du champ, mais restent tous les deux des modèles autorégressifs d'ordre 1 avec un coefficient d'autorégression ρ .

Dans le contexte des modèles de Whittle probabilisés, nous savons à présent que la connectivité locale du champ est un facteur faisant varier l'information de Fisher de ρ dans les échantillons ; il en ressort le fait que l'estimation efficace de ρ est optimale pour des conditions sensiblement différentes de celles de la pleine connectivité. Les mécanismes élémentaires qui provoquent cette variation ont été identifiés. La contrainte apportée par le modèle de connexité qui est paramétré par une *structure des interactions* et le paramètre p , affecte l'échantillon indépendamment de la statistique employée par le Chercheur. Dans ce sens, nous relevons la potentielle *injustice de la structure* décrite au Chapitre 3 lors de comparaisons d'estimations de ρ pour différents graphes de structure d'interactions.

Une absence d'observation du graphe des interactions nous prive d'une part de l'information échantillonnale de ρ . Théoriquement possible, l'estimation de maximum de vraisemblance de la matrice des interactions demande à résoudre de sévères problèmes mathématiques lorsque le champ est de grande taille. Sans d'autres techniques, nous sommes forcés à faire des hypothèses sur la configuration des interactions. Nous avons montré que dans ce contexte l'estimateur de maximum de vraisemblance de ρ est biaisé, sous-estimant par là ce paramètre, dès le moment que l'on pose dans l'expression de la fonction de vraisemblance plus d'interactions qu'il y en a en réalité. Nous relevons à partir des résultats des estimations de ρ que le paramètre p agit comme un paramètre d'*intensité de l'autocorrélation spatiale perçue* malgré qu'il soit parfaitement indépendant de ρ . Dans ce contexte on peut employer d'autres méthodes d'estimation de ρ qui ne sont pas nécessairement moins efficaces ou plus biaisés que l'estimateur de maximum de vraisemblance "erroné". A cet effet, nous montrons deux pistes dont l'une exploite le I de Moran maximal et l'autre demande un comptage d'amas de pixels binaires d'une "image" retraitée du champ spatial. Cette deuxième piste mériterait une recherche plus approfondie.

Il convient d'attirer l'attention sur des méthodes de plus en plus populaires comme la méthode EM ou des méthodes MCMC. Elles sont déjà employées pour estimer les paramètres de modèles spatiaux autorégressifs (entre autres des modèles SAR et CAR), mais elles demandent par contre quelques affinements pour qu'elles parviennent à estimer les paramètres de notre modèle de Whittle probabilisé.

Pour les tests statistiques abordés en Chapitre 5, où l'on a spécifiquement étudié le rejet de l'hypothèse d'indépendance spatiale en faveur d'une hypothèse simple d'un modèle de Whittle probabilisé, il s'est imposé le fait que la connaissance du graphe des interactions est une condition nécessaire pour obtenir un test du I de Moran de puissance maximale (pour inférer p et ρ). Nous avons vu sous quelles conditions de ces paramètres l'on peut perdre substantiellement de cette puissance lorsque par ignorance du graphe des interactions nous faisons recours à la matrice de structure des interactions en guise de pondération spatiale fixe pour le I de Moran. Nous avons signalé cet effet comme l'effet d'une *sur-spécification* des poids spatiaux. Du point de vue de l'inférence, cette perte donne la mesure de l'importance de connaître le graphe des interactions. Avec la même méconnaissance du graphe des interactions, mais avec la même connaissance du graphe de structure des interactions, le I de Moran maximal s'avère un intéressant compromis entre statistique de test et statistique descriptive, en l'occurrence comme variante d'un covariogramme. En ce qui concerne le résultat d'autres pondérations fixes construites *a priori* pour le I de Moran, nous avons constaté le risque réel plus ou moins important de surestimer la puissance du test si p s'avère plus élevé qu'on ne le pense, *a fortiori* si l'on pense que $p = 0$.

Un nouveau type de paramètre d'autocorrélation spatiale

Dans le contexte des modèles de Whittle classiques, on reconnaît en ρ le seul paramètre qui module ou traduit (selon l'usage du modèle) l'autocorrélation spatiale,

bien qu'habituellement sa relation avec la variance et covariance du champ soit complexe et sa signification rarement examinée. En passant de la variante classique à la variante probabilisée du modèle de Whittle, le modèle de connectivité change, mais le paramètre ρ conserve un rôle de moduleur ou traducteur de l'autocorrélation spatiale, avec une relation différente mais également complexe avec le pattern de variance et covariance du champ. En revanche, nous constatons que le paramètre p , bien qu'il soit indépendant du coefficient d'autorégression joue aussi un rôle de moduleur de l'autocorrélation spatiale. Nous avons observé ses effets respectivement dans la variance totale du champ, dans la précision de l'estimation efficace de ρ , dans le biais de l'estimation de maximum de vraisemblance de ρ obtenue dans l'ignorance du graphe des interactions et dans les fonctions de puissance des tests abordés. Par conséquent, nous découvrons en le paramètre p un paramètre d'autocorrélation d'un nouveau type, modulant l'autocorrélation spatiale d'un point de vue connectiviste.

Un modèle que l'on peut affiner

Nous avons fait part de beaucoup de parcimonie dans le choix des paramètres du champ spatial. Ce choix très restrictif contraint à considérer les paramètres des sources de stochasticité du champ, ainsi que le coefficient d'autorégression, comme uniformes sur l'espace¹. La simplicité de ce modèle donne une vue synthétique et simplificatrice de l'autocorrélation spatiale qui est parfois ce que l'on attend d'un modèle. Par contre, nous convenons qu'une éventuelle inhomogénéité spatiale du champ (comme par exemple l'hétéroscédasticité des chocs, l'anisotropie des interactions, etc.) pourrait être mieux captée si l'on changeait judicieusement le paramétrage du modèle (en s'inspirant en l'occurrence des modèles SAR classiques employés dans ces conditions).

Les modifications (améliorations) que l'on pourrait apporter au modèle ne doivent cependant pas faire perdre à ce dernier la capacité de relier l'autocorrélation spatiale à des concepts de la théorie des graphes ou de la théorie de la percolation. Cette contrainte s'applique également dans le cas où l'on voudrait changer de modèle de matrice des interactions². Le modèle des interactions que nous avons traité assure de manière simple et claire ce lien, ouvrant les perspectives d'applications que nous avons ébauchées au premier chapitre. Aussi encourage-t-on de compléter l'étude de notre modèle en développant les méthodes qui parviendront à estimer le paramètre p , puis placer finalement le modèle spatial, avec les raffinements souhaités, dans le domaine de la recherche appliquée.

1. On se réfère aux paramètres σ^2 , p et ρ .

2. Changement motivé par exemple pour rendre plus pratiques les méthodes d'estimation EM ou MCMC.

Annexe **A**

Notations mathématiques

Symboles grecs minuscules

$\alpha()$	Erreur de première espèce ; p. 97
α	Niveau ou seuil du test ; p. 97
$\beta()$	Erreur de seconde espèce ; p. 97
$\delta()$	Règle ou variable de décision ; p. 96
ϵ, ϵ	Variable aléatoire gaussienne, vecteur de variables aléatoires gaussiennes
$\theta, \boldsymbol{\theta}$	Paramètre scalaire, respectivement vectoriel, d'une loi de probabilité
$\pi()$	Fonction de puissance du test ; p. 97
ρ	Coefficient d'autorégression
σ	Ecart-type
ω	Poids spatial

Symboles grecs majuscules

$\Pi()$	Fonction de puissance ; p. 101
Σ	Matrice de variance-covariance

Symboles latins

A	Matrice d'adjacence
$\text{Cov}(\cdot, \cdot)$	Covariance de (\cdot, \cdot)
\mathfrak{d}	Dissimilarité
$E[\cdot]$	Espérance mathématique de \cdot

\mathcal{E}_P	Efficacité de Pitman ; p. 104
$F()$	Fonction de répartition
$\mathfrak{F}()$	Fonction définie par (5.28) ; p. 101
\mathcal{G}	Loi de probabilité
H_0, H_1	Hypothèse nulle, hypothèse alternative d'un test
I	Matrice unité (élément neutre pour la multiplication matricielle)
I	Indice de Moran
$J()$	Information de Fisher ; p. 52
$K(., .)$	Divergence de Kullback-Leibler
\mathcal{L}	Loi de probabilité
M	Matrice des interactions, en règle générale une matrice de Markov (généralisée)
$\mathcal{M}_{(p, \mathbf{S})}$	Loi de probabilité ; cf. Définition 3 ; p. 138
$\mathcal{N}_{(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$	Loi normale de paramètre $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
p	Paramètre p de la loi $\mathcal{M}_{(p, \mathbf{S})}$
\mathcal{P}	Proposition logique
$\text{Pr}[\cdot]$	Probabilité de \cdot
$\mathcal{Q}, \mathcal{Q}_T$	Région critique du test ; p. 96
$\bar{\mathcal{Q}}, \bar{\mathcal{Q}}_T$	Région d'acceptation du test ; p. 96
\mathbb{R}, \mathbb{R}_+	Ensemble des réels, ensemble des réels positifs
$\mathfrak{R}(\delta)$	Risque de seconde espèce moyen
S	Matrice de structure des interactions (matrice d'adjacence)
$\text{Sp}[\cdot]$	Spectre de $[\cdot]$ (ensemble des valeurs propres de $[\cdot]$)
$\text{sup}[\cdot]$	Supremum de $[\cdot]$
$T()$	Statistique de test
$\text{Tr}[\cdot]$	Trace de $[\cdot]$
U	Matrice d'adjacence complète
U	Matrice orthogonale (employée dans les calculs en annexe)
$\text{var}(., .)$	Variance de $(., .)$
W	Matrice de poids spatiaux
X, x	Champ spatial, respectivement sa réalisation

Autres symboles

*	Produit d'Hadamard
$ (\cdot) $	Déterminant de (\cdot)
$\{\dots\}$	Ensemble

Formulation

$\{A B\}$	"Ensemble des éléments A tels que B "
$\forall B$	"Quel que soit B " (quantificateur universel)
$(A B)$	A sachant B

Annexe **B**

Annexes du chapitre 1

B.1 Entropie de Shannon

Entropie d'une distribution discrète

Soit une distribution de probabilité discrète $\{p_1, p_2, \dots, p_n\}$. L'entropie de Shannon S est définie comme

$$S := - \sum_{i=1}^n \ln(p_i) p_i \quad (\text{B.1})$$

avec la convention $0 \ln(0) = 0$. L'entropie est maximale si la distribution de probabilité est uniforme.

Entropie d'une distribution continue

Soit \mathbf{X} une variable aléatoire prenant ses valeurs sur un espace Ω continu et de densité de probabilité f . L'entropie de Shannon $H(\mathbf{X})$ de la variable \mathbf{X} est définie comme

$$H(\mathbf{X}) := - \int_{\Omega} f(\mathbf{x}) \ln(f(\mathbf{x})) d\mathbf{x} \quad (\text{B.2})$$

avec la convention $0 \ln(0) = 0$.

B.2 La régression linéaire classique

Soient \mathbf{y} , \mathbf{x}_1 , \mathbf{x}_2 , \dots , \mathbf{x}_p des vecteurs colonnes de scores de dimension n . Un point dans l'espace généré par ces $p + 1$ vecteurs, appelé espace des variables, représente les scores d'un individu pour chacune des variables ψ , ϕ_1 , ϕ_2 , \dots , ϕ_p dont respectivement \mathbf{y} , \mathbf{x}_1 , \mathbf{x}_2 , \dots , \mathbf{x}_p en sont une réalisation d'un échantillon de taille n .

Soit un arrangement de $\kappa + 1$ variables :

$$\psi, \phi_1, \phi_2, \dots, \phi_{\kappa} \quad (\text{B.3})$$

dont on veut exprimer la première variable par les κ suivantes. On fait l'hypothèse de régression linéaire multiple, en supposant la relation suivante :

$$\mathbb{E}[\psi|\phi_1, \phi_2, \dots, \phi_p] = \beta_0 + \sum_{i=1}^p \beta_i \phi_i. \quad (\text{B.4})$$

Cette hypothèse conduit au modèle suivant, appelé *modèle linéaire* :

$$\psi = \beta_0 + \sum_{i=1}^p \beta_i \phi_i + \epsilon, \quad (\text{B.5})$$

où ϵ est une variable aléatoire d'espérance nulle non corrélée avec les variables ϕ_i . On note σ^2 la variance de ϵ .

Soient n réalisations des variables (B.3). En notant y_i ($i = 1, 2, \dots, n$) la i -ème réalisation de la variable ψ et x_{ji} ($i = 1, 2, \dots, n$) la i -ème réalisation de la variable ψ_j ($j = 1, 2, \dots, \kappa$), on définit

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{\kappa 1} \\ 1 & x_{21} & x_{22} & \dots & x_{\kappa 2} \\ 1 & x_{31} & x_{32} & \dots & x_{\kappa 3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{\kappa n} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_\kappa \end{pmatrix}. \quad (\text{B.6})$$

Il découle du modèle linéaire que les n réalisations indépendantes de (B.3) ont la relation matricielle

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (\text{B.7})$$

où, par indépendance des réalisations,

$$\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \mathbb{1}\sigma^2. \quad (\text{B.8})$$

En règle générale, les coefficients $\beta_0, \beta_1, \dots, \beta_p$ et σ^2 sont inconnus. Dans le cas empirique, ces quantités sont estimées à partir du modèle empirique

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}, \quad (\text{B.9})$$

où $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_\kappa)'$. Le vecteur \mathbf{e} est le vecteur des résidus, formellement défini comme

$$\mathbf{e} = \mathbf{y} - \mathbf{y}^* \quad \text{où } \mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (\text{B.10})$$

La meilleure estimation de $\boldsymbol{\beta}$ est obtenue par la méthode des moindres carrés où le critère à minimiser est le carré de la norme de \mathbf{e} . En supposant que l'espace des variables est muni de la métrique \mathbf{D} , on a

$$\mathbf{y}^* = \mathbf{P}_\mathbf{X}\mathbf{y}, \quad \text{où } \mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}. \quad (\text{B.11})$$

La matrice $\mathbf{P}_{\mathbf{X}}$ est la matrice de projection \mathbf{D} -orthogonale sur le sous espace de l'espace des variables engendré par les vecteurs colonnes de \mathbf{X} . Par le fait que $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}}$, on déduit que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{y}. \quad (\text{B.12})$$

Par définition du résidu, on a

$$\mathbf{e} = \mathbf{y} - \mathbf{y}^* = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y} =: \mathbf{P}_{\mathbf{X}}^{\perp}\mathbf{y}. \quad (\text{B.13})$$

L'opérateur $\mathbf{P}_{\mathbf{X}}^{\perp}$ est la matrice de projection sur le sous-espace \mathbf{D} -orthogonal au sous espace engendré par les colonnes de \mathbf{X} . Etant donné que $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ on a donc

$$\mathbf{e} = \mathbf{P}_{\mathbf{X}}^{\perp}\boldsymbol{\epsilon}. \quad (\text{B.14})$$

Une des conséquences est que les résidus ne sont pas nécessairement indépendants, car

$$\mathbf{E}[\mathbf{e}\mathbf{e}'] = \mathbf{E}[\mathbf{P}_{\mathbf{X}}^{\perp}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{P}_{\mathbf{X}}^{\perp}] = \sigma^2\mathbf{P}_{\mathbf{X}}^{\perp}. \quad (\text{B.15})$$

En conséquence, tout test statistique d'indépendance doit tenir compte explicitement de cette dépendance structurelle entre résidus de la régression linéaire. Ceci conduit à des complications dans le calcul de la distribution de probabilité des statistiques de test. Cependant, il est à relever que l'impact de la matrice \mathbf{X} sur la matrice de projection $\mathbf{P}_{\mathbf{X}}^{\perp}$ diminue asymptotiquement, cette dernière approchant la matrice identité (Upton & Fingleton, 1985, p. 337). On retient donc que

$$\lim_{n \rightarrow \infty} \mathbf{E}[\mathbf{e}\mathbf{e}'] = \mathbf{1}\sigma^2. \quad (\text{B.16})$$

B.3 Théorie de la percolation

Introduction

Le processus de percolation a été développé en tout premier par Flory (1941) et Stockmayer (1943) pour décrire des processus de polymérisation aboutissant à la gélification d'une masse fluide. La percolation a été introduite dans la littérature mathématique par Broadbent et Hammersley (1957), par le concept de diffusion de particules fluides à travers un milieu aléatoire. En général, une telle diffusion s'entend soit comme un parcours que le fluide décide de prendre dans le milieu (processus de diffusion), soit comme un parcours dont les possibilités sont imposées par le milieu (processus percolatif). C'est ce dernier aspect qui a intéressé Broadbent et Hammersley. Ce processus est appelé *processus de percolation*, car il invoque l'idée du passage du café dans un percolateur.

La théorie classique de la percolation distingue, mais peut également mélanger, la *percolation de sites* et la *percolation de liens*. Lorsque l'on est dans le contexte de la percolation de lien, l'espace se décrit à l'aide d'un graphe dont l'ensemble des arcs est composé d'arcs dits *liens ouverts* et d'arcs dits *liens fermés*. Classiquement, l'état *ouvert* ou *fermé* est le résultat d'une épreuve indépendante de Bernouilli

Réseau	q_c sites	q_c liens
Nid d'abeille	0.6962	$1 - 2 \sin(\frac{\pi}{18}) \approx 0.65271$
Carré	0.592746	$\frac{1}{2}$
Triangulaire	$\frac{1}{2}$	$2 \sin(\frac{\pi}{18}) \approx 0.3473$

Tab. B.1: Probabilités critiques pour divers réseaux

de probabilité q , autrement dit, chaque lien est occupé avec une probabilité q , indépendamment des autres liens. Dans un contexte physique, les liens s'entendent comme toute forme de passage entre couples de points dans l'espace.

Deux sites sont connectés au sens de la théorie de la percolation s'il sont connectés au sens de la théorie des graphes. Une composante connexe du graphe constitue un *amas* ou une *île* dans la théorie de la percolation et le concept de taille s'entend au sens de la théorie des graphes. Si q est suffisamment petit, les îles sont nombreuses et leur taille est petite. A l'opposé, si q est suffisamment proche de 1, le graphe est presque complet et ne forme qu'une seule île. Il existe une probabilité critique q_c telle qu'il se produit une transition dans la structure topologique du réseau passant d'un réseau macroscopiquement déconnecté à un réseau macroscopiquement connecté, autrement dit, pour $q > q_c$, il y a au moins un chemin entre deux sommets situés chacun sur les bords opposés d'une fenêtre d'observation carrée. Il est à souligner que le concept de probabilité critique est définie dans une situation asymptotique seulement (c.-à-d. la taille du réseau tend vers l'infini).

De manière analogue à la percolation de lien, la percolation de sites se réfère à des sites *occupés* et à des sites *vacants*. On dit que deux sites plus proches voisins sont *connectés* s'ils sont tous les deux occupés. L'*amas de sites* est un amas de sites occupés ayant pour périmètre des sites vacants. Il existe une probabilité critique q_c au-dessus de laquelle il y a au moins un chemin ne passant que par des sites occupés entre deux sommets situés chacun sur les bords opposés d'une fenêtre d'observation carrée. La table B.1 donne les probabilités critiques en fonction d'un choix de structure de réseau.

Quantités importantes et exposants critiques pour des réseaux infinis

La percolation est un processus aléatoire. Par conséquent, différents réseaux percolatifs contiennent différentes îles, de taille et formes différentes. La description de la percolation passe par la mesure de propriétés moyennes. Une de ces quantités est $n_s(q)$, la proportion obtenue en divisant le nombre d'îles composées de s sites par le nombre de sites du réseau. Cette quantité n'est pas connue exactement, mais elle se comporte près de $q = q_c$ comme

$$n_s \sim s^{-\tau} \exp(-|q - q_c|^{\frac{1}{\sigma}} s) \quad (q \rightarrow q_c), \quad (\text{B.17})$$

où la valeur de τ et de σ sont données dans la table B.2.

La probabilité qu'un site fasse partie d'une île composée de s sites est $sn_s(q)$, et par conséquent, la masse moyenne d'une île est S :

$$S := \frac{\sum_s s^2 n_s(q)}{\sum_s s n_s(q)}. \quad (\text{B.18})$$

Lorsque la probabilité q est au voisinage de q_c , S diverge, au vu de (B.17), selon une loi de puissance

$$S \sim |q - q_c|^{-\gamma}, \quad (\text{B.19})$$

où la valeur de γ est donnée dans la table B.2.

Soit P_∞ la probabilité qu'un site ou un lien appartienne à l'amas infini. Pour $q < q_c$, seuls existent des amas finis et par conséquent $P_\infty = 0$. Pour $q > q_c$, P_∞ croît comme une loi de puissance :

$$P_\infty \sim |q - q_c|^\beta \quad (q > q_c), \quad (\text{B.20})$$

où la valeur de β est donnée dans la table B.2.

Le rayon de gyration est défini comme étant la moyenne quadratique des distances euclidiennes entre couples de sites d'un amas de masse s :

$$R_s^2 := \frac{1}{2} \sum_{ij} \frac{\|\mathbf{r}_i - \mathbf{r}_j\|_2^2}{s^2}. \quad (\text{B.21})$$

A partir de là, on définit la longueur de corrélation comme

$$\xi^2 := \frac{2 \sum_s R_s^2 s^2 n_s}{\sum_s s^2 n_s}. \quad (\text{B.22})$$

La taille linéaire d'un amas fini est caractérisée, en-dessous et au-dessus de q_c , par la longueur de corrélation ξ . Pour q au voisinage de q_c , on a le résultat

$$\xi \sim |q - q_c|^{-\nu}, \quad (\text{B.23})$$

où la valeur de ν est donnée dans la table B.2.

Comme on peut le constater, nombre de quantités en théorie de la percolation divergent selon une loi de puissance avec un exposant donné. Ces exposants sont appelés des *exposants critiques* et ils ont la propriété remarquable d'être universels dans le sens qu'ils ne dépendent que de la dimension d du réseau et non pas de sa structure. Un certain nombre d'entre eux est donné dans le tableau ci-dessous.

La structure des amas de la percolation peut être décrite en terme de fractale. Pour $q = q_c$, la masse $M(r)$ de sites connectés à un même amas dans un rayon r est

$$M(r) \sim \begin{cases} r^{d_f} & \text{si } r \ll \xi \\ r^d & \text{si } r \gg \xi \end{cases}, \quad (\text{B.24})$$

Exposant	$d = 2$	$d = 3$	$d = 4$
β	$\frac{5}{36}$	0.41	0.64
γ	$\frac{43}{18}$	1.80	1.44
ν	$\frac{4}{3}$	0.88	0.68
σ	$\frac{36}{91}$	0.45	0.48
τ	$\frac{187}{91}$	2.18	2.31

Tab. B.2: Exposants critiques pour diverses dimensions

où d_f est la dimension fractale de l'amas infini et d est la plus petite dimension euclidienne dans lequel réside le réseau. Les deux dimensions sont reliées par des exposants critiques :

$$d_f = d - \frac{\beta}{\nu}. \quad (\text{B.25})$$

On peut montrer que les îles finies, pour $q \leq q_c$, possèdent une dimension fractale dont la valeur est (B.25) également, tant que leur taille linéaire n'est pas plus grande que la longueur de corrélation ξ .

De même, à $q = q_c$, les chemins de longueur minimale reliant un couple de points connectés entre eux sont fractales. En effet, la longueur l de ces chemins, appelés des *géodésiques* percolatives, sont reliées à la distance euclidienne r qu'elles parcourent par la relation

$$l \sim r^{d_{\min}}, \quad (\text{B.26})$$

où d_{\min} est un exposant critique dont la valeur est indiquée dans la table B.3. En l'occurrence, on observe que $d_{\min} > 1$ pour un réseau à deux dimensions.

Dimension	$d = 2$	$d = 3$	$d \geq 6$
d_f	$\frac{91}{48}$	2.524	4
d_{\min}	1.13	1.374	2

Tab. B.3: Exposants critiques

Percolation sur des réseaux finis

Nous rappelons que les grandeurs dérivées ci-dessus sont définies sur des réseaux infinis seulement. Lorsque le réseau est fini, de surface $L \times L$, sa taille intervient dans les calculs. La théorie de la percolation distingue les situations où $L \ll \xi$ de celles où $L \gg \xi$. Lorsque la taille du réseau dépasse suffisamment la longueur de corrélation, le réseau a le comportement statistique similaire à celui d'un réseau infini. Par contre, si la longueur de corrélation est suffisamment plus grande que la taille du réseau, alors le comportement statistique du réseau est affecté par sa taille.

Si l'on considère P_∞ , on observe que

$$P_\infty \sim \begin{cases} (p - p_c)^\beta & L \gg \xi \\ L^{-\frac{\beta}{\nu}} & L \ll \xi \end{cases}. \quad (\text{B.27})$$

Un autre effet de la limitation de la taille du réseau est une translation de la probabilité critique q_c . En effet, les simulations numériques montrent, entre autres, que S et P_∞ satisfont la loi de puissance correspondante (B.18) et (B.20) dans un réseau fini si l'on translate q_c d'une certaine quantité (pour un réseau triangulaire de taille 500×500 , la probabilité critique est 0.5083 au lieu de 0.5000 ; voir fig. 14 de Stauffer et al., 1994). Ziff (1992) a étudié au moyen de simulations numériques comment la concentration critique évolue avec la taille L du réseau et donne comme résultat pour $d = 2$:

$$q_c(L) = q_c(\infty) + aL^{-1-\frac{1}{\nu}}, \quad (\text{B.28})$$

où a est une constante et $q_c(\infty)$ s'obtient par interpolation.

Annexes du chapitre 2

C.1 Loi d'un vecteur de k variables de Bernoulli

Une variable de Bernoulli est une variable B prenant la valeur 0 avec une probabilité p et la valeur 1 avec une probabilité $1 - p$. Une épreuve de Bernoulli est une réalisation de B . Si $\mathbf{B} = (B_1, B_2, \dots, B_k)$ est un vecteur de k variables de Bernoulli indépendantes, alors sa probabilité $P(\mathbf{B})$ vaut

$$P(\mathbf{B}) = (1 - p)^{\tau[\mathbf{B}]} p^{k - \tau[\mathbf{B}]}, \quad (\text{C.1})$$

où $\tau[\cdot]$ est un opérateur qui totalise le nombre d'éléments non nuls du vecteur ou de la matrice donné en argument.

Par analogie, si \mathbf{B} est une matrice carrée de k variables de Bernoulli indépendantes, sa probabilité a également la forme (C.1). Partant de là, si \mathbf{S} est une matrice d'adjacence de même taille que \mathbf{B} , alors le produit d'Hadamard¹ $\mathbf{A} = \mathbf{S} * \mathbf{B}$ est une variable aléatoire dont sa probabilité vaut

$$P(\mathbf{A}) = (1 - p)^{\tau[\mathbf{A}]} p^{(\tau[\mathbf{S}] - \tau[\mathbf{A}])}. \quad (\text{C.2})$$

En lien avec notre modèle de Whittle probabilisé, \mathbf{S} est la matrice de structure des interactions et \mathbf{A} est la matrice d'adjacence du graphe des interactions. En posant la contrainte que \mathbf{S} et \mathbf{B} sont des matrices symétriques de trace nulle, la probabilité (C.2) se transforme et devient

$$\tilde{P}(\mathbf{A}) = (1 - p)^{\frac{1}{2}\tau[\mathbf{A}]} p^{\frac{1}{2}(\tau[\mathbf{S}] - \tau[\mathbf{A}])}, \quad (\text{C.3})$$

où le terme $\frac{1}{2}$ a été ajouté pour tenir compte de la réduction de moitié du nombre de degrés de liberté des matrices \mathbf{A} et \mathbf{B} . En généralisant, toute matrice \mathbf{M} obtenue bijectivement d'une matrice \mathbf{A} satisfaisant (C.3) hérite de sa distribution de

1. Le produit d'Hadamard est défini de la manière suivante : si \mathbf{A} est une matrice de terme général a_{ij} et \mathbf{B} est une matrice de même taille de terme général b_{ij} , alors $\mathbf{A} * \mathbf{B}$ est une matrice de terme général $a_{ij}b_{ij}$.

probabilité, et donc

$$\tilde{P}(\mathbf{M}) = (1 - p)^{\frac{1}{2}\tau[\mathbf{M}]} p^{\frac{1}{2}(\tau[\mathbf{S}] - \tau[\mathbf{M}])}. \quad (\text{C.4})$$

La relation (C.4) est vérifiée pour des matrices de Markov obtenues par normalisation des lignes de \mathbf{A} . On se donne alors la définition suivante.

Définition 3: *La matrice aléatoire de Markov \mathbf{M} suit une loi de probabilité $\mathcal{M}_{(p,\mathbf{S})}$ de paramètre $p \in [0, 1]$ et de matrice d'adjacence symétrique de trace nulle \mathbf{S} si \mathbf{M} a pour probabilité (C.4).* \diamond

Compte tenu des définitions précédentes, la variable $\tau[\mathbf{M}]$ suit une loi binomiale $\mathcal{B}_{(\tau[\mathbf{S}],p)}$ de paramètre $\tau[\mathbf{S}]$ et p , autrement dit, elle a la probabilité

$$\check{P}\left(\frac{1}{2}\tau[\mathbf{M}] = k\right) = \binom{\frac{1}{2}\tau[\mathbf{S}]}{k} (1 - p)^{\frac{1}{2}\tau[\mathbf{M}]} p^{\frac{1}{2}(\tau[\mathbf{S}] - \tau[\mathbf{M}])}. \quad (\text{C.5})$$

Par conséquent, son espérance vaut

$$\mathbb{E}_{\mathbf{M}}[\tau[\mathbf{M}]] = \tau[\mathbf{S}](1 - p) \quad (\text{C.6})$$

et sa variance vaut

$$\text{var}(\tau[\mathbf{M}]) = \tau[\mathbf{S}](p)(1 - p). \quad (\text{C.7})$$

Annexes du chapitre 3

D.1 Information de Fisher de ρ au voisinage de $\rho = 0$

Développement de Taylor limité

L'information de Fisher $J_{\mathbf{X}|\mathbf{M}}(\rho)$ de ρ pour un modèle de Whittle classique s'exprime comme

$$J_{\mathbf{X}|\mathbf{M}}(\rho) = \frac{1}{2}\text{Tr}[(\mathbf{B}\boldsymbol{\Sigma})^2] + \frac{1}{4}(\text{Tr}[\mathbf{B}\boldsymbol{\Sigma}])^2, \quad (\text{D.1})$$

où

$$\mathbf{B} := \frac{\partial}{\partial \rho} \boldsymbol{\Sigma}^{-1} = -(\mathbf{M}' + \mathbf{M}) + 2\mathbf{M}'\mathbf{M}\rho \quad (\text{D.2})$$

et $\boldsymbol{\Sigma}$ est la matrice de variance-covariance du modèle.

Un développement de Taylor de $J_{\mathbf{X}|\mathbf{M}}(\rho)$ à l'ordre 1 et au voisinage de $\rho = 0$ est de la forme

$$J_{\mathbf{X}|\mathbf{M}}(\rho) = J_{\mathbf{X}|\mathbf{M}}(0) + \frac{\partial}{\partial \rho} J_{\mathbf{X}|\mathbf{M}}(0)\rho + O(\rho^2). \quad (\text{D.3})$$

En supposant que $\sigma = 1$, on obtient les quantités suivantes pour (D.3) :

$$J_{\mathbf{X}|\mathbf{M}}(0) = \text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2], \quad (\text{D.4})$$

$$\frac{\partial}{\partial \rho} J_{\mathbf{X}|\mathbf{M}}(\rho) = \text{Tr}[(\mathbf{B}\boldsymbol{\Sigma})\left(\frac{\partial}{\partial \rho} \mathbf{B}\boldsymbol{\Sigma} + \mathbf{B}\frac{\partial}{\partial \rho} \boldsymbol{\Sigma}\right)] + \frac{1}{2}\text{Tr}[\mathbf{B}\boldsymbol{\Sigma}]\text{Tr}\left[\frac{\partial}{\partial \rho} \mathbf{B}\boldsymbol{\Sigma} + \mathbf{B}\frac{\partial}{\partial \rho} \boldsymbol{\Sigma}\right] \quad (\text{D.5})$$

$$= \text{Tr}[2\mathbf{B}\boldsymbol{\Sigma}\mathbf{M}'\mathbf{M}\boldsymbol{\Sigma} - (\mathbf{B}\boldsymbol{\Sigma})^3] + \text{Tr}[\mathbf{B}\boldsymbol{\Sigma}]\text{Tr}[2\mathbf{M}'\mathbf{M}\boldsymbol{\Sigma} - (\mathbf{B}\boldsymbol{\Sigma})^2], \quad (\text{D.6})$$

$$\frac{\partial}{\partial \rho} J_{\mathbf{X}|\mathbf{M}}(0) = 2\text{Tr}[\mathbf{M}^2(\mathbf{M} + \mathbf{M}')]. \quad (\text{D.7})$$

Généralement la quantité (D.7) n'est pas nulle, comme c'est le cas pour une triangulation de Delaunay. Elle est par contre nulle pour un réseau carré.

Valeur du paramètre p au maximum de l'information de Fisher

La dérivée de $E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)]$ par rapport à p vaut

$$\frac{\partial}{\partial p} E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)] = \sum_{\mathbf{M}} J_{\mathbf{X}|\mathbf{M}}(0) \frac{\partial}{\partial p} P(\mathbf{M}) \quad (\text{D.8})$$

$$= \frac{1}{2} \sum_{\mathbf{M}} J_{\mathbf{X}|\mathbf{M}}(0) \left(\frac{\tau[\mathbf{S}] - \tau[\mathbf{M}]}{p} - \frac{\tau[\mathbf{M}]}{1-p} \right) P(\mathbf{M}). \quad (\text{D.9})$$

On obtient un extrêum de $E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)]$, si (D.9) est nul, autrement dit si

$$E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)(\tau[\mathbf{S}] - \tau[\mathbf{M}])(1-p)] = E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)\tau[\mathbf{M}]]p. \quad (\text{D.10})$$

On a pour solution p^* :

$$p^* = 1 - \frac{E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)\tau[\mathbf{M}]]}{E_{\mathbf{M}}[J_{\mathbf{X}|\mathbf{M}}(0)\tau[\mathbf{S}]]}. \quad (\text{D.11})$$

Remarquons qu'en comparant avec l'expression de l'espérance de $\tau[\mathbf{M}]$, cf. (C.6), on observe que le maximum de la courbe se produit lorsque

$$E[J_{\mathbf{X}|\mathbf{M}}]E[\tau[\mathbf{M}]] = E[J_{\mathbf{X}|\mathbf{M}}\tau[\mathbf{M}]]. \quad (\text{D.12})$$

D.2 Divergence de Kullback-Leibler

Divergence pour une distribution discrète

Soient deux distributions de probabilité discrètes $\{p_1, p_2, \dots, p_n\}$ et $\{q_1, q_2, \dots, q_n\}$ telles que $q_i = 0 \Rightarrow p_i = 0$. La *divergence de Kullback-Leibler* $K(p, q)$ entre la distribution $\{p_i\}_{i=1}^n$ et la distribution $\{q_i\}_{i=1}^n$ est définie comme

$$K(p, q) := \sum_{i=1}^n p_i \ln \left(\frac{p_i}{q_i} \right) \quad (\text{D.13})$$

avec la convention $0 \ln(0) = 0$.

Divergence pour une distribution continue

Soit \mathbf{X} une variable aléatoire prenant ses valeurs sur un espace Ω continu. Soient f et g deux densités de probabilité de \mathbf{X} avec la condition $f(\mathbf{x}) = 0 \Rightarrow g(\mathbf{x}) = 0$. La divergence ou la distance de Kullback-Leibler $K(f, g)$ entre la densité f et la densité g est définie comme

$$K_{\mathbf{X}}(f, g) := \int_{\Omega} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x}. \quad (\text{D.14})$$

Développement de Taylor de la divergence de Kullback-Leibler pour des modèles de Whittle

Soient les densités de probabilité f_i définies par

$$f_i(\mathbf{x}) := (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}' \boldsymbol{\Sigma}_i^{-1} \mathbf{x}\right) \quad (i = 0, 1), \quad (\text{D.15})$$

où $\boldsymbol{\Sigma}$ une matrice de variance-covariance de taille $n \times n$ telle que :

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\mathbf{V}_i^{-1}}{\sigma_i^2} \quad \text{avec } \mathbf{V}_i^{-1} = (\mathbf{1} - \rho_i \mathbf{M})' (\mathbf{1} - \rho_i \mathbf{M}) \quad (\text{D.16})$$

dont \mathbf{M} est une matrice de taille $n \times n$, $\rho_0 = 0$ et $\rho_1 > 0$. Nous savons que la divergence de Kullback-Leibler entre f_0 et f_1 vaut (Kullback, 1959, chap. 9)

$$K_{\mathbf{X}}(f_0, f_1) = -\frac{1}{2} \ln\left(\frac{|\boldsymbol{\Sigma}_1^{-1}|}{|\boldsymbol{\Sigma}_0^{-1}|}\right) - \frac{n}{2} + \frac{1}{2} \text{Tr}(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1}). \quad (\text{D.17})$$

Un développement de Taylor de (D.17) autour de $\rho = 0$ revient à exprimer $K_{\mathbf{X}}(f_0, f_1)$ par le développement

$$K_{\mathbf{X}}(f_0, f_1) = \sum_{i=0}^{k-1} \frac{\rho_1^i}{i!} \frac{d^i}{d\rho^i} K_{\mathbf{X}}(f_0, f_1) \Big|_{\rho=0} + O(\rho_1^k). \quad (\text{D.18})$$

Dans notre cas :

$$\boldsymbol{\Sigma}_0^{-1} := \sigma_0^{-2} \mathbf{1}, \quad (\text{D.19})$$

$$\boldsymbol{\Sigma}_1^{-1} := \sigma_1^{-2} (\mathbf{1} - \rho_1 \mathbf{M})' (\mathbf{1} - \rho_1 \mathbf{M}), \quad (\text{D.20})$$

où nous supposons que $\sigma_0^2 = \sigma_1^2$. Nous calculons facilement les quantités suivantes

$$\ln\left(\frac{|\boldsymbol{\Sigma}_1^{-1}|}{|\boldsymbol{\Sigma}_0^{-1}|}\right) = \ln |(\mathbf{1} - \rho_1 \mathbf{M})' (\mathbf{1} - \rho_1 \mathbf{M})| \quad (\text{D.21})$$

$$= \ln(|\mathbf{1} - \rho_1 \mathbf{M}|^2), \quad (\text{D.22})$$

$$\text{Tr}(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1}) = \text{Tr}[(\mathbf{1} - \rho_1 \mathbf{M})' (\mathbf{1} - \rho_1 \mathbf{M})] \quad (\text{D.23})$$

$$= \text{Tr}[\mathbf{1} - \rho_1 (\mathbf{M}' + \mathbf{M}) + \rho_1^2 (\mathbf{M}' \mathbf{M})] \quad (\text{D.24})$$

$$= n - \rho_1 \text{Tr}[\mathbf{M}' + \mathbf{M}] + \rho_1^2 \text{Tr}[\mathbf{M}' \mathbf{M}]. \quad (\text{D.25})$$

La divergence (D.17) devient

$$2K_{\mathbf{X}}(f_0, f_1) = -\ln(|\mathbf{1} - \rho_1 \mathbf{M}|^2) - \rho_1 \text{Tr}[\mathbf{M}' + \mathbf{M}] + \rho_1^2 \text{Tr}[\mathbf{M}' \mathbf{M}]. \quad (\text{D.26})$$

En passant par les valeurs propres μ_i de \mathbf{M} , on observe que

$$\ln(|\mathbf{1} - \rho_1 \mathbf{M}|^2) = 2 \ln\left(\prod_{i=1}^n (1 - \rho_1 \mu_i)\right) = 2 \sum_{i=1}^n \ln(1 - \mu_i) \quad (\text{D.27})$$

$$= 2 \sum_{i=1}^n \sum_{j=1}^{\infty} \frac{1}{j} (-\mu_i \rho_1)^j = -2 \sum_{i=j}^{\infty} \frac{1}{j} \text{Tr}[\mathbf{M}^j] \rho_1^j. \quad (\text{D.28})$$

La divergence (D.26) est donc équivalente à la série

$$2K_{\mathbf{X}}(f_0, f_1) = \text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2] \rho_1^2 + 2 \sum_{i=3}^{\infty} \frac{1}{i} \text{Tr}[\mathbf{M}^i] \rho_1^i. \quad (\text{D.29})$$

Annexe **E**

Annexes du chapitre 5

E.1 Moments de l'indice de Moran

Cas général

L'indice de Moran I pour un champ spatial \mathbf{X} est un rapport de deux formes quadratiques :

$$I = \frac{Q_1}{Q_2}, \quad Q_1 = \mathbf{X}'\mathbf{W}\mathbf{X}, \quad Q_2 = \mathbf{X}'\mathbf{X}, \quad (\text{E.1})$$

où \mathbf{W} est une matrice indépendante de \mathbf{X} . Si le k -ème moment de I existe ($k \in \mathbb{N}^*$), on a l'expression (cf. p. ex. Mathai & Provost, 1992, p.145) :

$$\mathbb{E}[I^k] = \mathbb{E}[Q_1^k Q_2^{-k}] \quad (\text{E.2})$$

$$= \mathbb{E}\left[Q_1^k \frac{1}{\Gamma(k)} \int_0^\infty s^{k-1} \exp(-Q_2 s) ds\right], \quad (\text{E.3})$$

où $\mathbb{E}[\cdot]$ exprime l'espérance mathématique et $\Gamma(k)$ est la fonction gamma de k . Sous condition de pouvoir interchanger les intégrales, on a

$$\mathbb{E}[I^k] = \frac{1}{\Gamma(k)} \int_0^\infty s^{k-1} \mathbb{E}[Q_1^k \exp(-Q_2 s)] ds. \quad (\text{E.4})$$

L'espérance est calculée en considérant un champ gaussien de densité de probabilité f

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x}\right). \quad (\text{E.5})$$

En définissant la fonction génératrice $\varphi(t_1, t_2)$ des moments mixtes par

$$\varphi(t_1, t_2) := \mathbb{E}[\exp(Q_1 t_1) \exp(Q_2 t_2)], \quad (\text{E.6})$$

l'expression (E.4) devient

$$\mathbb{E}[I^k] = \frac{1}{\Gamma(k)} \int_0^\infty t_2^{k-1} \left[\frac{\partial^k}{\partial t_1^k} \varphi(t_1, -t_2) \right] \Big|_{t_1=0} dt_2. \quad (\text{E.7})$$

Dans notre cas,

$$\varphi(t_1, -t_2) = |\boldsymbol{\Sigma}^{-\frac{1}{2}}| \int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}} \exp[\mathbf{x}'(t_1 \mathbf{W} - t_2 \mathbf{1} - \frac{1}{2} \boldsymbol{\Sigma}^{-1}) \mathbf{x}] d\mathbf{x}. \quad (\text{E.8})$$

On définit la quantité $\varphi^{(k)}(t_2)$,

$$\varphi^{(k)}(t_2) := \frac{\partial^k}{\partial t_1^k} \varphi(t_1, -t_2) \Big|_{t_1=0}, \quad (\text{E.9})$$

qui devient

$$\varphi^{(k)}(t_2) = |\boldsymbol{\Sigma}^{-\frac{1}{2}}| \int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}} (\mathbf{x}' \mathbf{W} \mathbf{x})^k \exp[-\mathbf{x}'(t_2 \mathbf{1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1}) \mathbf{x}] d\mathbf{x}. \quad (\text{E.10})$$

On opère une substitution de variable qui ne change pas la valeur de (E.10) :

$$\mathbf{x} \mapsto \mathbf{y}, \quad \mathbf{y} = \mathbf{U}' \mathbf{x}, \quad (\text{E.11})$$

où \mathbf{U} est une matrice orthogonale permettant d'obtenir la similitude $\mathbf{U} \boldsymbol{\Sigma} \mathbf{U}' = \boldsymbol{\Lambda}$, où $\boldsymbol{\Lambda}$ est une matrice diagonale comportant les valeurs propres λ_i ($i = 1, 2, \dots, n$) de $\boldsymbol{\Sigma}$ dans la diagonale. La similitude préserve le déterminant, à savoir $|\boldsymbol{\Sigma}| = |\boldsymbol{\Lambda}|$. En notant $\boldsymbol{\Theta} := \mathbf{U}' \mathbf{W} \mathbf{U}$, on obtient

$$\varphi^{(k)}(t_2) = |\boldsymbol{\Lambda}^{-\frac{1}{2}}| \int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}} (\mathbf{y}' \boldsymbol{\Theta} \mathbf{y})^k \exp[-\mathbf{y}'(t_2 \mathbf{1} + \frac{1}{2} \boldsymbol{\Lambda}^{-1}) \mathbf{y}] d\mathbf{y} \quad (\text{E.12})$$

$$= |\boldsymbol{\Lambda}^{-\frac{1}{2}}| \int_{-\infty}^{\infty} (2\pi)^{-\frac{n}{2}} \left(\sum_{ij} y_i y_j \Theta_{ij} \right)^k \exp[-\frac{1}{2} \sum_i \xi_i y_i^2] d\mathbf{y}, \quad (\text{E.13})$$

où

$$\xi_i := 2t_2 + \lambda_i^{-1}. \quad (\text{E.14})$$

Pour la suite, nous utilisons le résultat suivant :

$$h(\xi, k) := \int_{-\infty}^{\infty} y^k \exp(-\xi \frac{y^2}{2}) dy \quad (\text{E.15})$$

$$= \frac{1}{(\sqrt{\xi})^{k+1}} \int_{-\infty}^{\infty} x^k \exp(-\frac{x^2}{2}) dx \quad (\text{E.16})$$

$$= \frac{1}{2} \left(\frac{2}{\xi} \right)^{\frac{k+1}{2}} \Gamma\left(\frac{k+1}{2}\right) ((-1)^k + 1). \quad (\text{E.17})$$

Selon la parité de k , on observe que, pour $\nu \in \mathbb{N}$,

$$h(\xi, k) = \begin{cases} \left(\frac{2}{\xi}\right)^{\frac{k+1}{2}} \Gamma\left(\frac{k+1}{2}\right) & \text{si } k = 2\nu \\ 0 & \text{si } k = 2\nu + 1 \end{cases}, \quad (\text{E.18})$$

avec

$$\Gamma\left(\frac{k+1}{2}\right) = \begin{cases} \sqrt{\pi} & \text{si } k = 0 \\ \sqrt{\pi} \prod_{i=0}^{\nu-1} \left(\frac{1}{2} + i\right) & \text{si } k = 2\nu > 0 \\ \nu! & \text{si } k = 2\nu + 1 \end{cases} \quad (\text{E.19})$$

En particulier,

$$h(\xi, 0) = \sqrt{\frac{2\pi}{\xi}}. \quad (\text{E.20})$$

En définissant \mathbb{E}_k comme l'ensemble de toutes les matrices \mathbf{B} qui ont comme éléments des entiers positifs tels que $\tau[\mathbf{B}] = k$ (où $\tau[\cdot]$ est un opérateur qui totalise le nombre d'éléments non nuls de la matrice donnée en argument), puis en désignant de manière générique le terme général de chaque matrice d'adjacence \mathbf{B} par b_{ij} et, finalement, en notant par Θ_{ij} le terme général de Θ , on obtient la décomposition suivante de la forme quadratique $(\mathbf{y}'\Theta\mathbf{y})^k$:

$$(\mathbf{y}'\Theta\mathbf{y})^k = \left(\sum_{ij} y_i y_j \Theta_{ij}\right)^k \quad (\text{E.21})$$

$$= \sum_{\mathbf{B} \in \mathbb{E}_k} k! \left(\prod_{r,s} b_{rs}\right)^{-1} \prod_{i,j} y_i^{b_{ij}} y_j^{b_{ij}} \Theta_{ij}^{b_{ij}} \quad (\text{E.22})$$

$$= k! \sum_{\mathbf{B} \in \mathbb{E}_k} \left(\prod_{r,s} \frac{\Theta_{rs}^{b_{rs}}}{b_{rs}!}\right) \left(\prod_i y_i^{b_{i\bullet}}\right) \left(\prod_j y_j^{b_{\bullet j}}\right) \quad (\text{E.23})$$

$$= k! \sum_{\mathbf{B} \in \mathbb{E}_k} \left(\prod_{r,s} \frac{\Theta_{rs}^{b_{rs}}}{b_{rs}!}\right) \left(\prod_i y_i^{b_{i\bullet} + b_{\bullet i}}\right), \quad (\text{E.24})$$

où l'on a noté $b_{i\bullet} = \sum_j b_{ij}$, $b_{\bullet i} = \sum_j b_{ji}$, et où on a utilisé la généralisation de la formule du binôme de Newton, avec $\eta_{ij} := y_i y_j \Theta_{ij}$:

$$\left(\sum_{i,j} \eta_{ij}\right)^k = \sum_{\mathbf{B} \in \mathbb{E}_k} \left(\frac{k!}{\prod_{(r,s)} b_{rs}!}\right) \prod_{(i,j)} \eta_{ij}^{b_{ij}}. \quad (\text{E.25})$$

L'équation (E.13) devient donc

$$\varphi^{(k)}(t_2) = k! \frac{|\Lambda^{-\frac{1}{2}}|}{(2\pi)^{\frac{n}{2}}} \sum_{\mathbf{B} \in \mathbb{E}_k} \prod_{r,s} \frac{\Theta_{rs}^{b_{rs}}}{b_{rs}!} \prod_i h(\xi_i, b_{i\bullet} + b_{\bullet i}). \quad (\text{E.26})$$

Une fois $\varphi^{(k)}(t_2)$ déterminé, le moment d'ordre k du I de Moran s'obtient par

$$\mathbb{E}(I^k) = \frac{1}{\Gamma(k)} \int_0^\infty t_2^{k-1} \varphi^{(k)}(t_2) dt_2. \quad (\text{E.27})$$

Dans le cas particulier où \mathbf{W} et Σ commutent, ils sont simultanément diagonalisables. Dans ce cas, par le choix de l'opération de similitude appropriée, Θ est diagonale et contient les valeurs propres ω_i de \mathbf{W} . On obtient dans ce cas spécial

$$\varphi^{(k)}(t_2) = k! \frac{|\mathbf{\Lambda}^{-\frac{1}{2}}|}{(2\pi)^{\frac{n}{2}}} \sum_{\|\mathbf{b}\|_1=k} \left(\prod_j \frac{\omega_j^{b_j}}{b_j!} \right) \prod_i h(\xi_i, 2b_i), \quad (\text{E.28})$$

où l'on désigne par $\mathbf{b} = (b_1, b_2, \dots, b_n)$ un vecteur de nombres entiers.

Premier moment

On décompose la forme quadratique $\mathbf{y}'\mathbf{\Theta}\mathbf{y}$ de la manière suivante :

$$\mathbf{y}'\mathbf{\Theta}\mathbf{y} = \sum_{i,j} y_i y_j \Theta_{ij} = \sum_{i \neq j} y_i y_j \Theta_{ij} + \sum_i y_i^2 \Theta_{ii}. \quad (\text{E.29})$$

Par conséquent,

$$\begin{aligned} \varphi^{(1)}(t_2) = & \eta \sum_{i \neq j} \left(h(\xi_i, 1) h(\xi_j, 1) \Theta_{ij} \prod_{k \notin \{i,j\}} h(\xi_k, 0) \right) \\ & + \eta \sum_i \left(h(\xi_i, 2) \Theta_{ii} \prod_{j \neq i} h(\xi_j, 0) \right), \end{aligned} \quad (\text{E.30})$$

où

$$\eta = (2\pi)^{-\frac{n}{2}} |\mathbf{\Lambda}^{-\frac{1}{2}}| \quad (\text{E.31})$$

et

$$h(\xi, 0) = \sqrt{\frac{2\pi}{\xi}}, \quad h(\xi, 1) = 0, \quad h(\xi, 2) = \left(\frac{2}{\xi}\right)^{\frac{3}{2}} \frac{\sqrt{\pi}}{2}. \quad (\text{E.32})$$

Soit $\mathbf{\Xi}$ la matrice diagonale constituée des valeurs ξ_i :

$$\mathbf{\Xi} := \text{diag}(\xi_1, \xi_2, \dots, \xi_n). \quad (\text{E.33})$$

Alors

$$\varphi^{(1)}(t_2) = |\mathbf{\Lambda}^{-\frac{1}{2}}| \left(\prod_{j=1}^n \xi_j^{-\frac{1}{2}} \right) \sum_{i=1}^n \Theta_{ii} \xi_i^{-1} \quad (\text{E.34})$$

$$= (|\mathbf{\Lambda}| |\mathbf{\Xi}|)^{-\frac{1}{2}} \text{Tr}[\mathbf{\Xi}^{-1} \mathbf{\Theta}]. \quad (\text{E.35})$$

Soit la matrice $\tilde{\mathbf{\Xi}}$ obtenue par opération de similitude de $\mathbf{\Xi}$:

$$\tilde{\mathbf{\Xi}} := \mathbf{U} \mathbf{\Xi} \mathbf{U}' = 2t_2 \mathbf{1} + \mathbf{\Sigma}^{-1}. \quad (\text{E.36})$$

Le premier moment du I de Moran peut donc être obtenu avec les matrices d'origine :

$$\mathbf{E}(I) = \int_0^\infty (|\mathbf{\Lambda}| |\tilde{\mathbf{\Xi}}|)^{-\frac{1}{2}} \text{Tr}[\tilde{\mathbf{\Xi}} \mathbf{W}] dt_2 \quad (\text{E.37})$$

$$= \int_0^\infty |(2t_2 \mathbf{\Sigma} + \mathbf{1})|^{-\frac{1}{2}} \text{Tr}[(2t_2 \mathbf{1} + \mathbf{\Sigma}^{-1})^{-1} \mathbf{W}] dt_2. \quad (\text{E.38})$$

Deuxième moment

Pour calculer le deuxième moment du I de Moran, on doit calculer $\varphi^{(2)}(t_2)$, sur la base des matrices de l'ensemble \mathbb{E}_k , avec $k = 2$. En vue de simplifier le calcul, nous décomposons l'ensemble \mathbb{E}_2 en deux sous-ensembles disjoints : $\mathbb{E}_2 = \mathbb{E}_2^{\{1\}} \cup \mathbb{E}_2^{\{2\}}$ où $\mathbb{E}_2^{\{1\}}$ est l'ensemble des matrices de \mathbb{E}_2 possédant exactement un élément non nul (valant en l'occurrence 2) et $\mathbb{E}_2^{\{2\}}$ est l'ensemble des matrices de \mathbb{E}_2 possédant exactement 2 éléments non nuls (valant en l'occurrence 1). A son tour, l'ensemble $\mathbb{E}_2^{\{1\}}$ peut être décomposé en 8 sous-ensembles disjoints $\mathbb{E}_2^{\{1,m\}}$ ($m = 1, 2, \dots, 8$). En désignant respectivement par (i, j) et (k, l) la position des deux éléments non nuls dans chaque matrice de $\mathbb{E}_2^{\{1\}}$, les sous-ensembles $\mathbb{E}_2^{\{1,m\}}$ se distinguent les uns des autres par les conditions imposées à ces positions, comme indiqué dans le tableau E.1.

La quantité suivante que l'on retrouve dans (E.26),

$$h(\mathbf{B}) := 2\eta \prod_{r,s} \frac{\Theta^{b_{rs}}}{b_{rs}!} \prod_{i=1}^n h(\xi_i, b_{i\bullet} + b_{\bullet i}), \quad (\text{E.39})$$

(où η est défini en (E.31)) est non nulle uniquement pour les matrices $\mathbf{B} \in \mathbb{E}_2^{\{2\}} \cup \mathbb{E}_2^{\{1,3\}} \cup \mathbb{E}_2^{\{1,7\}}$. Par conséquent, le calcul de $\varphi^{(2)}(t_2)$ devient :

$$\varphi^{(2)}(t_2) = \sum_{\mathbf{B} \in \mathbb{E}_2^{\{2\}}} h(\mathbf{B}) + \sum_{\mathbf{B} \in \mathbb{E}_2^{\{1,3\}}} h(\mathbf{B}) + \sum_{\mathbf{B} \in \mathbb{E}_2^{\{1,7\}}} h(\mathbf{B}) \quad (\text{E.40})$$

$$= 2\eta \sum_{i=1}^n \frac{\Theta_{ii}^2}{2} h(\xi_i, 4) \prod_{j \neq i} h(\xi_j, 0) \quad (\text{E.41})$$

$$+ 2\eta \sum_{\substack{i=1 \\ j \neq i}}^n \frac{\Theta_{ij}^2}{2} h(\xi_i, 2) h(\xi_j, 2) \prod_{k \notin \{i,j\}} h(\xi_k, 0) \quad (\text{E.42})$$

$$+ 2\eta \frac{1}{2} \sum_{\substack{i=1 \\ j \neq i}}^n \Theta_{ii} \Theta_{jj} h(\xi_i, 2) h(\xi_j, 2) \prod_{k \notin \{i,j\}} h(\xi_k, 0) \quad (\text{E.43})$$

$$+ 2\eta \frac{1}{2} \sum_{\substack{i=1 \\ j \neq i}}^n \Theta_{ij} \Theta_{ji} h(\xi_i, 2) h(\xi_j, 2) \prod_{k \notin \{i,j\}} h(\xi_k, 0). \quad (\text{E.44})$$

En utilisant les résultats suivants :

$$h(\xi, 0) = \sqrt{2\pi} \xi^{-\frac{3}{2}}, \quad h(\xi, 2) = \sqrt{2\pi} \xi^{-\frac{3}{2}}, \quad h(\xi, 4) = 3\sqrt{2\pi} \xi^{-\frac{5}{2}}, \quad (\text{E.45})$$

Ensemble	Conditions sur les positions	Exemple de matrice
$\mathbb{E}_2^{\{1,1\}}$	$i = j \quad k = i$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,2\}}$	$i = j \quad l = i$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,3\}}$	$i = j \quad l = i$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,4\}}$	$i = j \quad k \neq i$ $l \neq i$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,5\}}$	$i \neq j \quad k = i$ $l \neq i$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,6\}}$	$i \neq j \quad l = i$ $k \neq j$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,7\}}$	$i \neq j \quad j = k$ $i = l$ $k \neq l$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
$\mathbb{E}_2^{\{1,8\}}$	$i \neq j \quad k \neq l$ $k \neq i$ $l \neq j$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

Tab. E.1:

on a

$$\varphi^{(2)}(t_2) = 3|\mathbf{\Lambda}^{-\frac{1}{2}}| \left(\prod_{k=1}^n \xi_k^{-\frac{1}{2}} \right) \sum_{i=1}^n \xi_i^{-2} \Theta_{ii}^2 \quad (\text{E.46})$$

$$+ |\mathbf{\Lambda}^{-\frac{1}{2}}| \left(\prod_{k=1}^n \xi_k^{-\frac{1}{2}} \right) \sum_{\substack{i=1 \\ j \neq i}}^n (\xi_i \xi_j)^{-1} (\Theta_{ij}^2 + \Theta_{ii} \Theta_{jj} + \Theta_{ij} \Theta_{ji}). \quad (\text{E.47})$$

Soit $\mathbf{\Xi}$ définie en (E.33) et sa matrice de similitude $\tilde{\mathbf{\Xi}} = 2t_2 \mathbf{1} + \mathbf{\Sigma}^{-1}$ définie par (E.36). Par conséquent, on peut exprimer le deuxième moment à partir des matrices d'origine :

$$\varphi^{(2)}(t_2) = (|\mathbf{\Lambda}| |\mathbf{\Xi}|)^{-\frac{1}{2}} (\text{Tr}[\mathbf{\Xi}^{-1} \mathbf{\Theta} \mathbf{\Xi}^{-1} \mathbf{\Theta}'] + \text{Tr}^2[\mathbf{\Xi}^{-1} \mathbf{\Theta}] + \text{Tr}[(\mathbf{\Xi}^{-1} \mathbf{\Theta})^2]) \quad (\text{E.48})$$

$$= (|\mathbf{\Sigma}| |\tilde{\mathbf{\Xi}}|)^{-\frac{1}{2}} (\text{Tr}[\tilde{\mathbf{\Xi}}^{-1} \mathbf{W} \tilde{\mathbf{\Xi}}^{-1} \mathbf{W}'] + \text{Tr}^2[\tilde{\mathbf{\Xi}}^{-1} \mathbf{W}] + \text{Tr}[(\tilde{\mathbf{\Xi}}^{-1} \mathbf{W})^2]). \quad (\text{E.49})$$

On obtient le deuxième moment du I de Moran par

$$\text{E}(I^2) = \int_0^\infty t_2 \varphi^{(2)}(t_2) dt_2. \quad (\text{E.50})$$

La variance du I de Moran s'obtient des deux premiers moments :

$$\text{var}(I) = \text{E}(I^2) - \text{E}^2(I). \quad (\text{E.51})$$

E.2 Expression exacte de la fonction de répartition du I de Moran

Soit le modèle autorégressif $\mathbf{X} = \rho \mathbf{M} \mathbf{X} + \boldsymbol{\epsilon}$ où \mathbf{X} est le champ spatial et $\mathbf{X} \sim \mathcal{N}_{(\mathbf{0}, \mathbf{\Sigma})}$, $\boldsymbol{\epsilon} \sim \mathcal{N}_{(\mathbf{0}, \mathbf{1})}$ et dont $\mathbf{\Sigma}$ est la matrice de variance-covariance de \mathbf{X} . On sait que

$$\mathbf{\Sigma} = \mathbf{L} \mathbf{L}', \quad \mathbf{X} = \mathbf{L} \boldsymbol{\epsilon}, \quad \mathbf{L} = (\mathbf{1} - \rho \mathbf{M})^{-1}. \quad (\text{E.52})$$

L'indice de Moran I , de matrice \mathbf{W} , peut être exprimé en fonction du vecteur $\boldsymbol{\epsilon}$:

$$I = \frac{\mathbf{X}' \mathbf{W} \mathbf{X}}{\mathbf{X}' \mathbf{X}} = \frac{\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{W} \mathbf{L} \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{L} \boldsymbol{\epsilon}}. \quad (\text{E.53})$$

La fonction de répartition F_I de I peut être exprimée comme

$$F_I(\eta) := \Pr(I \leq \eta) = \Pr\left(\frac{\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{W} \mathbf{L} \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{L} \boldsymbol{\epsilon}} \leq \eta\right) \quad (\text{E.54})$$

$$= \Pr(\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{W} \mathbf{L} \boldsymbol{\epsilon} \leq \eta \boldsymbol{\epsilon}' \mathbf{L}' \mathbf{L} \boldsymbol{\epsilon}) \quad (\text{E.55})$$

$$= \Pr(\boldsymbol{\epsilon}' \mathbf{L}' \mathbf{W} \mathbf{L} \boldsymbol{\epsilon} - \eta \boldsymbol{\epsilon}' \mathbf{L}' \mathbf{L} \boldsymbol{\epsilon} \leq 0) \quad (\text{E.56})$$

$$=: F_Q(0), \quad (\text{E.57})$$

où F_Q est la fonction de répartition de la forme quadratique

$$Q = \boldsymbol{\epsilon}'\mathbf{L}'\mathbf{W}\mathbf{L}\boldsymbol{\epsilon} - \eta\boldsymbol{\epsilon}'\mathbf{L}'\mathbf{L}\boldsymbol{\epsilon}. \quad (\text{E.58})$$

Il existe une matrice orthogonale \mathbf{U} permettant la décomposition spectrale suivante :

$$\mathbf{U}'(\mathbf{L}'\mathbf{W}\mathbf{L} - \eta\mathbf{L}'\mathbf{L})\mathbf{U} = \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (\text{E.59})$$

où $\boldsymbol{\Lambda}$ est la matrice diagonale contenant les $m \leq n$ valeurs propres non nulles $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ de $\mathbf{L}'\mathbf{W}\mathbf{L} - \eta\mathbf{L}'\mathbf{L}$. Si l'on définit un vecteur $\mathbf{b} := (\mathbf{b}_1, \mathbf{b}_2)'$ de même dimension que $\boldsymbol{\epsilon}$ avec \mathbf{b}_1 et \mathbf{b}_2 de dimension m et $n - m$ respectivement, tel que

$$\boldsymbol{\epsilon} = \mathbf{U}\mathbf{b}, \quad (\text{E.60})$$

alors on obtient

$$Q = \mathbf{b}'_1\boldsymbol{\Lambda}\mathbf{b}_1 = \sum_{i=1}^m \lambda_i b_i^2, \quad (\text{E.61})$$

avec des variables i.i.d. $b_i^2 \sim \chi_1^2$ (loi du chi-deux). On se retrouve dans les conditions traitées par Provost (1989a, 1989b) qui permettent d'arriver à une expression exacte de la fonction de répartition F_I . Cette expression peut être retrouvée dans Mathai & Provost (1992). Elle a malheureusement une forme compliquée qui la rend inexploitable pour nos travaux.

E.3 Approximation numérique de la fonction de répartition du I de Moran

Soient un vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n) \sim \mathcal{N}_{(\mathbf{0}, \mathbf{1})}$ et une forme quadratique associée $Q = \sum_{i=1}^n \lambda_i X_i^2$. En vertu de la distribution des éléments de \mathbf{X} , les variables X_i^2 suivent une loi du χ_1^2 (centrale). On obtient la fonction de répartition de Q en calculant d'abord sa fonction caractéristique et en appliquant une formule d'inversion (une technique bien connue en statistiques).

La fonction caractéristique de Q est $\varphi_Q(t)$:

$$\varphi_Q(t) = \prod_{j=1}^n (1 - 2it\lambda_j)^{-\frac{1}{2}}. \quad (\text{E.62})$$

La formule d'inversion possède plusieurs formes. Gil-Pelaez (1951) a dérivé la formule suivante pour la fonction de répartition de Q :

$$F_Q(\eta) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \Im[\exp(it\eta)\varphi_Q(t)] dt, \quad (\text{E.63})$$

où $\Im(z)$ signifie la partie imaginaire de z .

Imhof (1961) a montré que (E.63) peut être exprimé de la manière suivante (cf. par exemple Tiefelsdorf (2000) pour un exposé clair et détaillé) :

$$F_Q(\eta) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin(\theta u)}{u\varrho(u)} du, \quad (\text{E.64})$$

où

$$\theta(u) = \frac{1}{2} \sum_{i=1}^n \tan^{-1}[\lambda_i u] - \frac{1}{2} \eta u, \quad (\text{E.65})$$

$$\varrho(u) = \prod_{i=1}^n [1 + \lambda_i^2 u^2]^{\frac{1}{4}}. \quad (\text{E.66})$$

Il suffit ensuite d'appliquer les résultats de cette méthode à $F_Q(\eta) = \Pr(Q \leq \eta)$ avec Q défini par (E.58) pour obtenir la distribution exacte du I de Moran avec matrice de poids spatiaux constante (Tiefelsdorf & Boots, 1995).

E.4 Approximation asymptotique de la distribution de probabilité du I de Moran

Domaine de valeurs du I de Moran

L'indice de Moran, sous la forme de quotient de Rayleigh, est un rapport de formes quadratiques où la forme du dénominateur est toujours positive. Partant de là, si \mathbf{x} est une réalisation d'un champ spatial \mathbf{X} , on sait que (voir par ex. Saporta, 1990, p. 484).

$$\lambda_{\min} \leq \frac{\mathbf{x}'\mathbf{W}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \lambda_{\max}, \quad (\text{E.67})$$

où λ_{\min} et λ_{\max} sont respectivement la plus petite et la plus grande valeur propre de \mathbf{W} . Autrement dit, l'indice de Moran ne peut être ni supérieur à sa plus grande valeur propre, ni inférieur à sa plus petite valeur propre. Cette condition est toujours vraie quel que soit le champ. Il est dès lors évident qu'un indice I de Moran ne suit pas une loi normale tant que la matrice \mathbf{W} est une matrice finie. Sen (1976) démontre rigoureusement que pour certaines conditions la distribution de probabilité asymptotique du I de Moran tend vers une distribution de probabilité gaussienne. Boots & Tiefelsdorf (2000) ont étudié cette convergence pour des réseaux carrés, triangulaires et hexagonaux.

E.5 Autres approximations de la distribution de probabilité du I de Moran

En dehors de l'approximation normale asymptotique précédemment évoquée, Tiefelsdorf (2002) rapporte qu'un certain nombre d'autres méthodes d'approximation du I de Moran ont été proposées dans la littérature en géographie. Parmi

celles-ci il y a l'approche par l'*approximation de Pearson de type III*, l'approche par l'*approximation beta*, l'*approche de Durbin-Watson*, l'*approche de Henshaw*, l'approche par l'*approximation de Edgeworth* (voir Tiefelsdorf, 2002, pour les références). Dernièrement, Tiefelsdorf (2002) a traité de l'*approximation de point de selle* du I de Moran.

E.6 Approximation de Taylor du rapport des vraisemblances

Soit le rapport des vraisemblances

$$T(\mathbf{x}) = \ln \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}, \quad (\text{E.68})$$

où

$$f_i(\mathbf{x}) := (2\pi)^{-\frac{n}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}' \Sigma_i^{-1} \mathbf{x}\right) \quad (i = 0, 1) \quad (\text{E.69})$$

dont Σ une matrice de variance-covariance de taille $n \times n$ telle que

$$\Sigma_i^{-1} = \frac{\mathbf{V}_i^{-1}}{\sigma_i^2}, \quad \text{avec } \mathbf{V}_i^{-1} = (\mathbf{1} - \rho_i \mathbf{M})' (\mathbf{1} - \rho_i \mathbf{M}), \quad (\text{E.70})$$

où \mathbf{M} est une matrice de taille $n \times n$, $\rho_0 = 0$ et $\rho_1 > 0$.

On suppose que σ_0^2 et σ_1^2 sont a priori inconnus. On les estime par conséquent par le principe du maximum de vraisemblance. Un simple calcul montre que

$$\frac{\partial}{\partial \sigma^2} f_0(\mathbf{x}) = 0 \Rightarrow \sigma^2 = \frac{1}{n} \mathbf{x}' \mathbf{x}, \quad (\text{E.71})$$

$$\frac{\partial}{\partial \sigma^2} f_1(\mathbf{x}) = 0 \Rightarrow \sigma^2 = \frac{1}{n} \mathbf{x}' \mathbf{V}_1 \mathbf{x}. \quad (\text{E.72})$$

En substituant σ_0^2 et σ_1^2 par la valeur σ^2 de (E.71) et (E.72) respectivement dans l'expression de (E.69), le rapport (E.68) devient

$$T(\mathbf{x}) = \ln \left(\left(\frac{\mathbf{x}' \mathbf{V}_1^{-1} \mathbf{x}}{\mathbf{x}' \mathbf{x}} \right)^{-\frac{n}{2}} + \ln |\mathbf{V}_1|^{-\frac{1}{2}} \right). \quad (\text{E.73})$$

Le développement de Taylor au deuxième ordre de $T(\mathbf{x})$ est

$$T(\mathbf{x}) = T(\mathbf{x})|_{\rho=0} + \rho \frac{\partial}{\partial \rho} T(\mathbf{x})|_{\rho=0} + \rho \frac{\partial^2}{\partial \rho^2} T(\mathbf{x})|_{\rho=0} + O(\rho^3). \quad (\text{E.74})$$

Pour procéder aux dérivations, nous avons besoin des lemmes suivants.

Lemme 1: Soit \mathbf{B} une matrice $n \times n$ dépendant d'un paramètre ρ . Si \mathbf{B} est une fonction différentiable de ρ et est inversible étant donné ρ , alors

$$\frac{d}{d\rho} \ln(|\mathbf{B}^{-1}|) = \text{Tr}[\mathbf{B} \frac{d}{d\rho} \mathbf{B}^{-1}].$$

Lemme 2: Soit \mathbf{B} une matrice inversible d'inverse \mathbf{B}^{-1} , toutes deux étant une fonction différentiable de ρ . La dérivée de \mathbf{B}^{-1} par rapport à ρ est

$$\frac{d}{d\rho}\mathbf{B} = -\mathbf{B}\left(\frac{d\mathbf{B}^{-1}}{d\rho}\right)\mathbf{B}.$$

On obtient alors les résultats suivants :

$$\frac{\partial}{\partial\rho_1}T(\mathbf{x}) = -\frac{n}{2}\frac{\partial}{\partial\rho_1}\ln(\mathbf{x}'\mathbf{V}_1^{-1}\mathbf{x}) + \frac{1}{2}\frac{\partial}{\partial\rho_1}\ln|\mathbf{V}_1^{-1}| \quad (\text{E.75})$$

$$= -\frac{n}{2}\frac{1}{\mathbf{x}'\mathbf{V}_1^{-1}\mathbf{x}}\mathbf{x}'\frac{\partial\mathbf{V}_1^{-1}}{\partial\rho_1}\mathbf{x} + \frac{1}{2}\text{Tr}[\mathbf{V}_1\frac{\partial\mathbf{V}_1^{-1}}{\partial\rho_1}], \quad (\text{E.76})$$

$$\frac{\partial}{\partial\rho_1}T(\mathbf{x})|_{\rho_1=0} = \frac{n}{2}\frac{\mathbf{x}'(\mathbf{M}' + \mathbf{M})\mathbf{x}}{\mathbf{x}'\mathbf{x}} - \frac{1}{2}\text{Tr}[\mathbf{M}' + \mathbf{M}] \quad (\text{E.77})$$

$$= n\frac{\mathbf{x}'\mathbf{M}\mathbf{x}}{\mathbf{x}'\mathbf{x}}, \quad (\text{E.78})$$

$$\begin{aligned} \frac{\partial^2}{\partial\rho_1^2}T(\mathbf{x}) &= -\frac{n}{2}\frac{1}{(\mathbf{x}'\mathbf{V}_1^{-1}\mathbf{x})^2}[(\mathbf{x}'\frac{\partial^2\mathbf{V}_1^{-1}}{\partial\rho_1^2}\mathbf{x})(\mathbf{x}'\mathbf{V}_1^{-1}\mathbf{x}) - (\mathbf{x}'\frac{\partial\mathbf{V}_1^{-1}}{\partial\rho_1}\mathbf{x})^2] \\ &\quad + \frac{1}{2}\text{Tr}[-\mathbf{V}_1\frac{\partial\mathbf{V}_1^{-1}}{\partial\rho_1}\mathbf{V}_1\frac{\partial\mathbf{V}_1^{-1}}{\partial\rho_1} + \mathbf{V}_1\frac{\partial^2\mathbf{V}_1^{-1}}{\partial\rho_1^2}], \end{aligned} \quad (\text{E.79})$$

$$\frac{\partial^2}{\partial\rho_1^2}T(\mathbf{x})|_{\rho_1=0} = -\frac{n}{2}\left(\frac{\mathbf{x}'\mathbf{M}'\mathbf{M}\mathbf{x}}{\mathbf{x}'\mathbf{x}} + \left(\frac{\mathbf{x}'(\mathbf{M}' + \mathbf{M})\mathbf{x}}{\mathbf{x}'\mathbf{x}}\right)^2\right) + \text{Tr}[\mathbf{M}^2]. \quad (\text{E.80})$$

Le développement de Taylor de (E.68) devient

$$T(\mathbf{x}) = nI_{\mathbf{M}}(\mathbf{x})\rho_1 - (nI_{\mathbf{M}'\mathbf{M}}(\mathbf{x}) + 4nI_{\mathbf{M}}(\mathbf{x})^2 + \text{Tr}[\mathbf{M}^2])\rho_1^2 + O(\rho_1^3). \quad (\text{E.81})$$

E.7 Base informative du test du rapport de maximum de vraisemblance

Cas d'hypothèses simples

Soit un système d'hypothèses simples

$$\begin{cases} H_0 : \mathbf{X} \sim \mathcal{L}_{\boldsymbol{\theta}}, & \boldsymbol{\theta} = \boldsymbol{\theta}_0, \\ H_1 : \mathbf{X} \sim \mathcal{L}_{\boldsymbol{\theta}}, & \boldsymbol{\theta} = \boldsymbol{\theta}_1, \end{cases} \quad (\text{E.82})$$

où \mathcal{L}_θ est une loi de probabilité de densité de probabilité f_θ .

Nous voulons associer à ce système une statistique de test $T(\mathbf{X})$. Pour cela, nous voulons déterminer les densités de probabilité f_0^* et f_1^* telles que les divergences de Kullback-Leibler $K(f_0^*, f_{\theta_0})$ et $K(f_1^*, f_{\theta_1})$ soient minimales avec les contraintes respectives

$$\kappa = \int \eta(\mathbf{x}) f_i^*(\mathbf{x}) d\mathbf{x} \quad (i = 0, 1), \quad (\text{E.83})$$

où η est une fonction donnée de \mathbf{x} . Les fonctions résultantes sont (cf. Kullback, 1954; Sanov, 1957, p. 23-24)

$$f_i^*(\mathbf{x}) = \frac{\exp(\tau_i \eta(\mathbf{x})) f_{\theta_i}(\mathbf{x})}{M_{\theta_i}(\tau)}, \quad (\text{E.84})$$

où

$$M_{\theta_i}(\tau) = \int f_{\theta_i}(\mathbf{x}) \exp(\tau \eta(\mathbf{x})) d\mathbf{x} \quad (\text{E.85})$$

et où τ_i est une fonction de κ définie par la relation suivante

$$\kappa = \frac{d}{d\tau_i} \ln M_{\theta_i}(\tau_i). \quad (\text{E.86})$$

On dit que f_θ génère une *famille exponentielle de distribution de probabilité* à partir de la densité de probabilité f_θ . Les divergences minimales résultantes valent

$$K_{\mathbf{X}}(f_i^*, f_{\theta_i}) = \kappa \tau_i - \ln M_{\theta_i}(\tau_i). \quad (\text{E.87})$$

Le calcul de (E.84) dépend de la contrainte théorique (E.83). Dans l'ignorance de cette contrainte, mais en possédant, à défaut, une réalisation \mathbf{x} de la variable à observer \mathbf{X} , les divergences $K_{\mathbf{X}}(f_i^*, f_{\theta_i})$ peuvent être estimées en utilisant la valeur observée de $\eta(\mathbf{x})$ comme un estimateur de κ . L'estimateur de $\tau_i(\mathbf{x})$, que l'on note $\hat{\tau}_i(\mathbf{x})$, est obtenu via la relation

$$\eta(\mathbf{x}) = \left. \frac{d}{d\tau_i} \ln M_{\theta_i}(\tau_i) \right|_{\tau_i = \hat{\tau}_i(\mathbf{x})}. \quad (\text{E.88})$$

On obtient alors un estimateur $\hat{K}(f_i^*, f_{\theta_i} | \mathbf{x})$ de $K(f_i^*, f_{\theta_i})$ de la forme (Kullback, 1959) :

$$\hat{K}(f_i^*, f_{\theta_i} | \mathbf{x}) := \eta(\mathbf{x}) \hat{\tau}_i(\mathbf{x}) - \ln M_{\theta_i}(\hat{\tau}_i(\mathbf{x})). \quad (\text{E.89})$$

Kullback interprète la quantité (E.89) comme *une mesure de la divergence entre l'échantillon et f_{θ_i}* . Plus la valeur de $K(f_i^*, f_{\theta_i})$ est grande, moins grande est la "ressemblance" entre l'échantillon et la population de densité de probabilité f_{θ_i} .

La règle de décision qui vient naturellement est celle qui rejette H_0 si

$$T(\mathbf{x}) := \hat{K}(f_0^*, f_{\theta_0} | \mathbf{x}) - \hat{K}(f_1^*, f_{\theta_1} | \mathbf{x}) \geq \xi_\alpha, \quad (\text{E.90})$$

où ξ_α est une constante déterminée de sorte que l'erreur de première espèce vaille α .

Le choix de la statistique $\eta(\mathbf{x})$ est importante dans le sens qu'elle doit éviter de perdre l'information pertinente de l'échantillon. Nous désirons donc qu'elle soit exhaustive pour les hypothèses du test. Pour ce faire, nous invoquons le théorème suivant.

Théorème 3: *La statistique $\eta(\mathbf{x})$ est une statistique exhaustive pour la famille exponentielle (E.84) générée à partir de $f_{\boldsymbol{\theta}}$.* \diamond

Ce que ce théorème signifie est que si f_0 et f_1 sont respectivement deux densités de probabilité de \mathbf{X} de la même famille exponentielle et que g_0 et g_1 sont respectivement les densités de probabilité correspondantes de $\eta(\mathbf{X})$, alors $K(f_0, f_1) = K(g_0, g_1)$ (pas de perte d'information). Si l'on pose

$$\eta(\mathbf{x}) := \ln \frac{f_{\boldsymbol{\theta}_1}(\mathbf{x})}{f_{\boldsymbol{\theta}_0}(\mathbf{x})}, \quad (\text{E.91})$$

alors $f_{\boldsymbol{\theta}_0}$ et $f_{\boldsymbol{\theta}_1}$ sont les fonctions pour lesquels $\eta(\mathbf{x})$ est une statistique exhaustive et (cf. Kullback, 1959)

$$T(\mathbf{x}) = \ln \frac{f_{\boldsymbol{\theta}_1}(\mathbf{x})}{f_{\boldsymbol{\theta}_0}(\mathbf{x})}. \quad (\text{E.92})$$

Autrement dit, en choisissant la forme de la statistique η de sorte à satisfaire l'identité (E.91), la statistique de test T définie par (E.90) correspond au rapport des vraisemblances qui est la statistique optimale au sens du lemme de Neyman-Pearson.

Cas d'hypothèses composites

Lorsque nous avons un système d'hypothèses composites de la forme

$$\begin{cases} H_0 : \mathbf{X} \sim \mathcal{L}_{\boldsymbol{\theta}}, & \boldsymbol{\theta} \in \Theta_0, \\ H_1 : \mathbf{X} \sim \mathcal{L}_{\boldsymbol{\theta}}, & \boldsymbol{\theta} \in \Theta_1, \end{cases} \quad (\text{E.93})$$

il peut lui être associé la statistique de test T définie comme

$$T(\mathbf{x}) := \min_{\boldsymbol{\theta}_0 \in \Theta_0} \hat{K}(f_0^*, f_{\boldsymbol{\theta}_0} | \mathbf{x}) - \min_{\boldsymbol{\theta}_1 \in \Theta_1} \hat{K}(f_1^*, f_{\boldsymbol{\theta}_1} | \mathbf{x}), \quad (\text{E.94})$$

où $\hat{K}(f_1^*, f_{\boldsymbol{\theta}_1} | \mathbf{x})$ est définie par (E.89). De manière analogue au cas précédent, on pose (E.91) dans (E.94) avec les paramètres laissés libres, et on obtient

$$T(\mathbf{x}) = \min_{\boldsymbol{\theta}_0 \in \Theta_0} \max_{\boldsymbol{\theta}_1 \in \Theta_1} [\hat{K}(f^* \| f_{\boldsymbol{\theta}_0}; \mathbf{x}) - \hat{K}(f^* \| f_{\boldsymbol{\theta}_1}; \mathbf{x})] \quad (\text{E.95})$$

$$= \min_{\boldsymbol{\theta}_0 \in \Theta_0} \max_{\boldsymbol{\theta}_1 \in \Theta_1} \left[\ln \frac{f_{\boldsymbol{\theta}_1}(\mathbf{x})}{f_{\boldsymbol{\theta}_0}(\mathbf{x})} \right] \quad (\text{E.96})$$

$$= \ln \frac{\max_{\boldsymbol{\theta} \in \Theta_1} [f_{\boldsymbol{\theta}_1}(\mathbf{x})]}{\max_{\boldsymbol{\theta} \in \Theta_0} [f_{\boldsymbol{\theta}_0}(\mathbf{x})]}. \quad (\text{E.97})$$

On reconnaît le test classique du rapport de maximum de vraisemblance de Neyman-Pearson dans le cas où l'hypothèse H_0 correspond à une restriction du domaine de valeurs des paramètres de H_1 .

E.8 Efficacité de Pitman entre deux I de Moran

Soit \mathbf{M} la matrice de taille $n \times n$ des poids des interactions d'un modèle de Whittle classique. Soient $I_{\mathbf{W}_1}$ et $I_{\mathbf{W}_2}$ deux indices de Moran de matrice de poids spatiaux \mathbf{W}_1 et \mathbf{W}_2 respectivement, de même taille que \mathbf{M} , que l'on désire comparer par le critère de l'efficacité de Pitman. Cette dernière, définie en (5.38), demande seulement à calculer l'espérance de $I_{\mathbf{W}_i}$ ($i=1,2$), en fonction de ρ (en supposant qu'elle est dérivable en $\rho = 0$), et la variance de $I_{\mathbf{W}_i}$ ($i=1,2$) sous l'hypothèse H_0 .

Le calcul de $\frac{\partial \mu(\rho)}{\partial \rho}$ pour le I de Moran peut être effectué en utilisant l'expression (E.38) :

$$\mu_I(\rho) := \mathbb{E}(I_{\mathbf{W}} | H_1) = \int_0^\infty (|\Sigma \mathbf{B}|^{-1})^{\frac{1}{2}} \text{Tr}[\mathbf{B}^{-1} \mathbf{W}] dt, \quad (\text{E.98})$$

avec

$$\mathbf{A} := 2t\Sigma + \mathbf{1} \quad \mathbf{B} := 2t\mathbf{1} + \Sigma^{-1}. \quad (\text{E.99})$$

Nous appliquons la dérivée à l'intérieur de l'intégrale, c'est-à-dire sur

$$\xi(t) := (|\Sigma \mathbf{B}|^{-1})^{\frac{1}{2}} \text{Tr}[\mathbf{B}^{-1} \mathbf{W}]. \quad (\text{E.100})$$

Le développement est le suivant :

$$\frac{\partial \xi(t)}{\partial \rho} = \frac{1}{2} |\mathbf{A}|^{\frac{1}{2}} \frac{\partial |\mathbf{A}|^{-1}}{\partial \rho} \text{Tr}[\mathbf{B}^{-1} \mathbf{W}] + |\mathbf{A}|^{-\frac{1}{2}} \text{Tr}\left[\frac{\partial \mathbf{B}^{-1}}{\partial \rho} \mathbf{W}\right] \quad (\text{E.101})$$

$$= \frac{1}{2} |\mathbf{A}|^{-\frac{1}{2}} \text{Tr}\left[\mathbf{A} \frac{\partial \mathbf{A}^{-1}}{\partial \rho}\right] \text{Tr}[\mathbf{B}^{-1} \mathbf{W}] + |\mathbf{A}|^{-\frac{1}{2}} \text{Tr}\left[\frac{\partial \mathbf{B}^{-1}}{\partial \rho} \mathbf{W}\right] \quad (\text{E.102})$$

$$= - |\mathbf{A}|^{-\frac{1}{2}} \left(\frac{1}{2} \text{Tr}\left[\frac{\partial \mathbf{A}}{\partial \rho} \mathbf{A}^{-1}\right] \text{Tr}[\mathbf{B}^{-1} \mathbf{W}] + \text{Tr}\left[\left(\mathbf{B}^{-1} \frac{\partial \mathbf{B}}{\partial \rho} \mathbf{B}^{-1}\right) \mathbf{W}\right] \right) \quad (\text{E.103})$$

$$= - |2t\Sigma + \mathbf{1}|^{-\frac{1}{2}} \left(\frac{1}{2} \text{Tr}\left[2t \frac{\partial \Sigma}{\partial \rho} (2t\Sigma + \mathbf{1})^{-1}\right] \text{Tr}[(2t\mathbf{1} + \Sigma^{-1})^{-1} \mathbf{W}] \right. \\ \left. + \text{Tr}\left[\left((2t\mathbf{1} + \Sigma^{-1})^{-1} \frac{\partial \Sigma^{-1}}{\partial \rho} (2t\mathbf{1} + \Sigma^{-1})^{-1}\right) \mathbf{W}\right] \right) \quad (\text{E.104})$$

$$= - |2t\Sigma + \mathbf{1}|^{-\frac{1}{2}} \left(-\frac{1}{2} \text{Tr}\left[2t\Sigma \frac{\partial \Sigma^{-1}}{\partial \rho} (2t\mathbf{1} + \Sigma^{-1})^{-1}\right] \text{Tr}[(2t\mathbf{1} + \Sigma^{-1})^{-1} \mathbf{W}] \right. \\ \left. + \text{Tr}\left[\left((2t\mathbf{1} + \Sigma^{-1})^{-1} \frac{\partial \Sigma^{-1}}{\partial \rho} (2t\mathbf{1} + \Sigma^{-1})^{-1}\right) \mathbf{W}\right] \right) \quad (\text{E.105})$$

$$= - |\Sigma \mathbf{B}|^{-\frac{1}{2}} \left(-\frac{1}{2} \text{Tr}\left[2t\Sigma \frac{\partial \Sigma^{-1}}{\partial \rho} \mathbf{B}^{-1}\right] \text{Tr}[\mathbf{B}^{-1} \mathbf{W}] + \text{Tr}\left[\left(\mathbf{B}^{-1} \frac{\partial \Sigma^{-1}}{\partial \rho} \mathbf{B}^{-1}\right) \mathbf{W}\right] \right). \quad (\text{E.106})$$

Lorsque ρ tend vers 0, c'est-à-dire lorsque H_1 tend vers H_0 , on a les simplifications suivantes

$$\left. \frac{\partial \Sigma^{-1}}{\partial \rho} \right|_{\rho=0} = -(\mathbf{M} + \mathbf{M}'), \quad \Sigma \Big|_{\rho=0} = \mathbf{1}, \quad \mathbf{B} \Big|_{\rho=0} = (2t + 1)\mathbf{1}, \quad (\text{E.107})$$

ce qui permet d'obtenir le résultat (au point $\rho = 0$)

$$\left. \frac{\partial \xi(t)}{\partial \rho} \right|_{\rho=0} = -(2t+1)^{-\frac{n}{2}-2} (t \text{Tr}[\mathbf{M}' + \mathbf{M}] \text{Tr}[\mathbf{W}] - \text{Tr}[(\mathbf{M}' + \mathbf{M})\mathbf{W}]). \quad (\text{E.108})$$

On obtient finalement la dérivée de la moyenne au point $\rho = 0$:

$$\frac{\partial \mu(0)}{\partial \rho} = \frac{1}{n+2} \text{Tr}[(\mathbf{M}' + \mathbf{M})\mathbf{W}]. \quad (\text{E.109})$$

Pour calculer la variance du I de Moran sous l'hypothèse H_0 , on utilise (E.51) qui demande à connaître le premier et le deuxième moment de I sous l'hypothèse H_0 . Le premier moment vaut

$$\text{E}(I|H_0) = 0 \quad (\text{E.110})$$

et le deuxième moment se calcule à partir de (E.50) pour donner

$$E(I^2|H_0) = \int_0^\infty (2t+1)^{-\frac{n}{2}-2} (\text{Tr}[\mathbf{W}\mathbf{W}'] + \text{Tr}^2[\mathbf{W}] + \text{Tr}[\mathbf{W}^2]) t dt \quad (\text{E.111})$$

$$= \frac{1}{n(n+2)} (\text{Tr}[\mathbf{W}\mathbf{W}'] + \text{Tr}^2[\mathbf{W}] + \text{Tr}[\mathbf{W}^2]). \quad (\text{E.112})$$

L'efficacité de Pitman (5.38) se calcule avec ces derniers résultats pour donner

$$\mathcal{E}_P(I_{\mathbf{W}_1}, I_{\mathbf{W}_2}) = \frac{(\text{Tr}[(\mathbf{M}' + \mathbf{M})\mathbf{W}_1])^2 (\text{Tr}[\mathbf{W}_2\mathbf{W}_2'] + \text{Tr}^2[\mathbf{W}_2] + \text{Tr}[\mathbf{W}_2^2])}{(\text{Tr}[(\mathbf{M}' + \mathbf{M})\mathbf{W}_2])^2 (\text{Tr}[\mathbf{W}_1\mathbf{W}_1'] + \text{Tr}^2[\mathbf{W}_1] + \text{Tr}[\mathbf{W}_1^2])}. \quad (\text{E.113})$$

On interprète $I_{\mathbf{W}_1}$ comme plus efficace que $I_{\mathbf{W}_2}$ pour le test si $\mathcal{E}_P(I_{\mathbf{W}_1}, I_{\mathbf{W}_2}) > 1$ et $I_{\mathbf{W}_1}$ comme moins efficace que $I_{\mathbf{W}_2}$ pour le test si $\mathcal{E}_P(I_{\mathbf{W}_1}, I_{\mathbf{W}_2}) < 1$.

En prenant $\mathbf{W}_1 = \mathbf{M}$ et $\mathbf{W}_2 = \mathbf{W}$ avec $\text{Tr}[\mathbf{M}] = \text{Tr}[\mathbf{W}] = 0$, on a la simplification suivante :

$$\mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{W}}) = \frac{\text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2] \text{Tr}[\mathbf{W}\mathbf{W}' + \mathbf{W}^2]}{\text{Tr}^2[\mathbf{M}'\mathbf{W} + \mathbf{M}\mathbf{W}]} \quad (\text{E.114})$$

En considérant \mathbf{W} comme paramètre libre de cette équation, on peut déterminer la solution qui minimise (E.114). Sans perdre de généralité, on peut considérer que \mathbf{W} est symétrique¹. La matrice \mathbf{W} qui minimise (E.114) est n'importe quelle solution de la forme

$$\mathbf{W} = \eta \mathbf{M}' + (1 - \eta) \mathbf{M}, \quad \eta \in [0, 1]. \quad (\text{E.115})$$

1. En effet, si \mathbf{W}_0 est une solution (non nécessairement symétrique) qui minimise (E.114), alors $\forall \eta \in [0, 1]$

$$I_{\mathbf{W}_0} = \frac{1}{2} \frac{\mathbf{x}' \mathbf{W}_0 \mathbf{x}}{\mathbf{x}' \mathbf{x}} = \frac{\mathbf{x}' (\eta \mathbf{W}_0' + (1 - \eta) \mathbf{W}_0) \mathbf{x}}{\mathbf{x}' \mathbf{x}} = I_{\eta \mathbf{W}_0' + (1 - \eta) \mathbf{W}_0}.$$

Pour $\eta = \frac{1}{2}$, on a une solution symétrique tout aussi optimale que pour $\eta \neq \frac{1}{2}$.

auquel cas $\mathcal{E}_P = 1$. En particulier $\mathbf{W} = \mathbf{M}$ donne également $\mathcal{E}_P = 1$. Ce constat a été fait par Cliff et Ord² (1973, 1981) qui préconisent en conséquence de *toujours utiliser l'indice de Moran employant a priori une matrice des poids \mathbf{W} qui représente le pattern d'autocorrélation posé dans l'hypothèse alternative H_1* . Cette suggestion ne s'applique cependant qu'à des systèmes d'hypothèses où H_1 est une hypothèse simple.

Généralisation aux modèles de Whittle markovien probabilisés

En considérant le fait que l'expression (5.38) de l'efficacité de Pitman se base sur des espérances mathématiques sur l'univers d'un champ spatial \mathbf{X} et compte tenu du fait que le coefficient d'autorégression ρ n'est pas un paramètre de la probabilité du graphe des interactions, alors l'efficacité de Pitman du passage de la statistique de test $T_1(\mathbf{X}, \mathbf{M}) := I_{\mathbf{M}}(\mathbf{X})$ à la statistique de test $T_2(\mathbf{X}, \mathbf{M}) := I_{\mathbf{W}}(\mathbf{X})$ pour des modèles de Whittle markovien probabilisés vaut

$$\mathcal{E}_P(T_1, T_2) = \frac{\mathbb{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2]] \text{Tr}[\mathbf{W}\mathbf{W}' + \mathbf{W}^2]}{\mathbb{E}_{\mathbf{M}}[\text{Tr}^2[\mathbf{M}'\mathbf{W} + \mathbf{M}\mathbf{W}]]}. \quad (\text{E.116})$$

Application

Cas des modèles de Whittle markoviens classiques

Soit \mathbf{M} la matrice de taille $n \times n$ des poids des interactions d'un modèle de Whittle markovien classique dont le graphe des interactions est représenté par la matrice d'adjacence $\mathbf{A}_{\mathbf{M}}$. Soit \mathbf{A} une matrice d'adjacence de taille $n \times n$ dont le graphe contient le graphe de \mathbf{M} . L'efficacité de Pitman $\mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{A}})$ du passage de la pondération spatiale \mathbf{M} à la pondération spatiale \mathbf{A} pour l'indice de Moran vaut

$$1 \leq \mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{A}_{\mathbf{M}}}) \leq \mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{A}}) = \frac{1}{2} \text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2] \frac{c_{\mathbf{A}}}{(n - n_0)^2} \leq \frac{c_{\mathbf{A}}}{n - n_0}, \quad (\text{E.117})$$

où $c_{\mathbf{A}}$ est le nombre d'adjacences orientées de \mathbf{A} et n_0 est le nombre de sites sans interactions (parmi les n sites considérés). Le résultat (E.117) implique que la perte d'efficacité est bornée par la densité de connectivité $\frac{c_{\mathbf{A}}}{n - n_0}$.

En définissant $c_{\mathbf{M}}$ comme étant le nombre d'arcs du graphe de \mathbf{M} , on observe que le rapport τ des efficacités de Pitman,

$$\tau := \frac{\mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{A}_{\mathbf{M}}})}{\mathcal{E}_P(I_{\mathbf{M}}, I_{\mathbf{A}})} = \frac{c_{\mathbf{M}}}{c_{\mathbf{A}}}, \quad (\text{E.118})$$

dépend seulement du rapport de connectivité totale des matrices. On conclut que la matrice d'adjacence \mathbf{A} induit la plus grande efficacité si $\mathbf{A} = \mathbf{A}_{\mathbf{M}}$.

2. Les conditions de la preuve apportée par Cliff et Ord sont plus restrictives. Ces mêmes auteurs n'ont pas relevé dans ce contexte que la solution \mathbf{W} optimale est une famille de matrices, ce qui élargit tout de même le sens de "représentation" du pattern d'autocorrélation dans H_1 .

Cas des modèles de Whittle markoviens probabilisés

Soit \mathbf{M} la matrice de taille $n \times n$ des poids des interactions aléatoires d'un modèle de Whittle markovien classique pour un champ spatial \mathbf{X} . Soit \mathbf{A} une matrice d'adjacence constante de taille $n \times n$ dont le graphe contient chaque réalisation du graphe de \mathbf{M} . L'efficacité de Pitman $\mathcal{E}_P(T_1, T_2)$ du passage de la statistique de test $T_1(\mathbf{X}, \mathbf{M}) := I_{\mathbf{M}}(\mathbf{X})$ à la statistique de test $T_2(\mathbf{X}, \mathbf{M}) := I_{\mathbf{A}}(\mathbf{X})$ vaut

$$\mathcal{E}_P(T_1, T_2) = \frac{1}{2} \mathbb{E}_{\mathbf{M}}[\text{Tr}[\mathbf{M}'\mathbf{M} + \mathbf{M}^2]] \frac{c_{\mathbf{A}}}{\mathbb{E}_{\mathbf{M}}[(n - n_0)^2]}, \quad (\text{E.119})$$

avec $c_{\mathbf{A}}$ et n_0 définis comme précédemment.

Annexe **F**

Annexes du chapitre 5

F.1 Méthode du I de Moran maximal

Méthode à 2 contraintes (méthode de Kooijman)

On désire obtenir la valeur maximale de l'indice I de Moran de la forme

$$I = \sum_{k=0}^K \frac{a_k(\mathbf{x})w_k}{\mathbf{x}'\mathbf{x}}, \quad (\text{F.1})$$

où $a_k(\mathbf{x}) = \mathbf{x}'\mathbf{A}_k\mathbf{x}$, avec les contraintes

$$\sum_{k=1}^{\kappa} n_k w_k = a, \quad \sum_{k=1}^{\kappa} n_k w_k^2 = b. \quad (\text{F.2})$$

En utilisant la méthode des multiplicateurs de Lagrange, la quantité à maximiser est J , de la forme

$$J := \sum_{k=1}^{\kappa} \frac{a_k(\mathbf{x})w_k}{\mathbf{x}'\mathbf{x}} + \lambda \left(\sum_{k=1}^{\kappa} n_k w_k^2 - b \right) + \mu \left(\sum_{k=1}^{\kappa} n_k w_k - a \right). \quad (\text{F.3})$$

On obtient

$$\frac{\partial J}{\partial w_k} = \frac{a_k(\mathbf{x})}{\mathbf{x}'\mathbf{x}} + 2\lambda n_k w_k + \mu n_k = 0, \quad (\text{F.4})$$

$$w_k = -\frac{1}{2\lambda \mathbf{x}'\mathbf{x}} \left(\frac{a_k(\mathbf{x})}{n_k} + \mu \mathbf{x}'\mathbf{x} \right), \quad (\text{F.5})$$

$$\mu = \frac{z}{\mathbf{x}'\mathbf{x}}, \quad (\text{F.6})$$

$$\lambda = -\frac{1}{2a} \frac{A + zN}{\mathbf{x}'\mathbf{x}}, \quad (\text{F.7})$$

où

$$N(Nb - a^2)z^2 + 2A(Nb - a^2)z + A^2b - Ba^2 = 0 \quad (\text{F.8})$$

et

$$N := \sum_{k=1}^{\kappa} n_k, \quad (\text{F.9})$$

$$A := \sum_{k=1}^{\kappa} a_k(\mathbf{x}), \quad (\text{F.10})$$

$$B := \sum_{k=1}^{\kappa} \frac{a_k^2(\mathbf{x})}{n_k}. \quad (\text{F.11})$$

La solution de (F.8) est

$$z = -\frac{A}{N} \mp \frac{a}{N} \sqrt{\frac{A^2 - BN}{a^2 - bN}}. \quad (\text{F.12})$$

Par conséquent

$$\lambda = \pm \frac{1}{2\mathbf{x}'\mathbf{x}} \sqrt{\frac{A^2 - BN}{a^2 - bN}}, \quad (\text{F.13})$$

$$\mu = -\frac{1}{\mathbf{x}'\mathbf{x}} \left(\frac{A}{N} \pm \frac{a}{N} \sqrt{\frac{A^2 - BN}{a^2 - bN}} \right). \quad (\text{F.14})$$

L'indice de Moran maximal est I_{\max} , qui vaut

$$I_{\max} = -(2\lambda b + \mu a) \quad (\text{F.15})$$

$$= \frac{1}{N\mathbf{x}'\mathbf{x}} (aA + \sqrt{(A^2 - BN)(a^2 - bN)}). \quad (\text{F.16})$$

Sachant que $A^2 - BN \leq 0$ par l'inégalité de Cauchy, I_{\max} est une solution réelle si $a^2 - bN \leq 0$. En fixant a et laissant varier b , on observe que I_{\max} tend vers son minimum si $b \rightarrow \frac{a^2}{N}$.

En ce qui concerne les poids w_k , on obtient finalement,

$$w_k = -\frac{1}{2\lambda\mathbf{x}'\mathbf{x}} \left(\frac{a_k(\mathbf{x})}{n_k} + \mu\mathbf{x}'\mathbf{x} \right) \quad (\text{F.17})$$

$$= -\frac{1}{2\lambda\mathbf{x}'\mathbf{x}} \frac{a_k(\mathbf{x})}{n_k} - \frac{1}{2} \frac{\mu}{\lambda} = \frac{1}{\sqrt{C}} \left(\frac{a_k}{n_k} - \frac{A}{N} \right) + \frac{a}{N} \quad (k > 0), \quad (\text{F.18})$$

où l'on définit C par

$$C := \frac{A^2 - BN}{a^2 - bN}. \quad (\text{F.19})$$

Méthode à 1 contrainte

On désire obtenir la valeur maximale de l'indice I de Moran de la forme

$$I = \sum_{k=1}^{\kappa} \frac{a_k(\mathbf{x})w_k}{\mathbf{x}'\mathbf{x}}, \quad (\text{F.20})$$

avec la contrainte

$$\sum_{k=1}^{\kappa} n_k w_k^\gamma = b, \quad (\text{F.21})$$

où $\gamma \geq 2$.

En utilisant la méthode des multiplicateurs de Lagrange, la quantité à maximiser est J , de la forme

$$J := \sum_{k=1}^{\kappa} \frac{a_k(\mathbf{x})w_k}{\mathbf{x}'\mathbf{x}} + \lambda \left(\sum_{k=1}^{\kappa} n_k w_k^\gamma - b \right). \quad (\text{F.22})$$

On obtient ainsi une première forme de w_k ,

$$\left[\frac{\partial J}{\partial w_k} = \frac{a_k(\mathbf{x})}{\mathbf{x}'\mathbf{x}} + \lambda n_k \gamma w_k^{\gamma-1} = 0 \right] \Rightarrow \left[w_k = \left(-\frac{1}{\gamma \lambda \mathbf{x}'\mathbf{x}} \frac{a_k(\mathbf{x})}{n_k} \right)^{\frac{1}{\gamma-1}} \right], \quad (\text{F.23})$$

et substituer w_k dans la contrainte pour obtenir

$$\left(\frac{1}{\lambda} \right)^{\frac{\gamma}{\gamma-1}} = b \left(\sum_{k=0}^{\kappa} n_k \left(-\frac{b_k}{\gamma \mathbf{x}'\mathbf{x}} \right)^{\frac{\gamma}{\gamma-1}} \right)^{-1}, \quad (\text{F.24})$$

où

$$b_k := \frac{a_k(\mathbf{x})}{n_k}. \quad (\text{F.25})$$

La substitution de λ dans la première forme de w_k donne

$$w_k^\gamma = b b_k^{\frac{\gamma}{\gamma-1}} \left(\sum_{k=0}^{\kappa} n_k b_k^{\frac{\gamma}{\gamma-1}} \right)^{-1}. \quad (\text{F.26})$$

Pour $\gamma = 2$, on a

$$\lambda = \pm \frac{1}{2\mathbf{x}'\mathbf{x}} \sqrt{\frac{B}{b}}. \quad (\text{F.27})$$

Des deux possibilités ci-dessus, nous choisissons la valeur positive, étant donné que l'indice de Moran maximal est positif seulement à cette condition. Cet indice maximal, I_{\max} , vaut

$$I_{\max} = -2\lambda b = \frac{\sqrt{bB}}{\mathbf{x}'\mathbf{x}}, \quad (\text{F.28})$$

et les poids w_k valent

$$w_k = \frac{a_k(\mathbf{x})}{n_k} \sqrt{\frac{b}{B}}. \quad (\text{F.29})$$

F.2 Le diagramme de dispersion angle-Imax

F.2.1 Définition

La pondération spatiale $w_1, w_2, \dots, w_\kappa$ obtenue de l'indice de Moran maximal peut être visualisée à l'aide de diagrammes de dispersion pertinents. Si l'on effectue le rapport entre deux tels poids spatiaux, on obtient

$$\frac{w_j(\mathbf{x})}{w_i(\mathbf{x})} = \frac{C_j(\mathbf{x})}{C_i(\mathbf{x})}, \quad (\text{F.30})$$

où

$$C_i(\mathbf{x}) := \frac{a_i(\mathbf{x})}{n_i} = \frac{\mathbf{x}'\mathbf{A}_i\mathbf{x}}{n_i} \quad (\text{F.31})$$

est la covariance empirique calculée entre sites voisins à l'ordre i selon la matrice de voisinage \mathbf{A}_i . En géostatistique, on reconnaît en (F.31) les éléments d'un *covariogramme*.

Les poids spatiaux "dilatés" de la forme $\sqrt{n_i}w_i(\mathbf{x})$ obtenus à partir d'une réalisation \mathbf{x} du champ spatial peuvent être représentés par un point dans un espace euclidien de dimension κ (diagramme de dispersion). En vertu de la contrainte (4.17) sur les poids spatiaux, chaque point ainsi obtenu réside sur la surface d'une hypersphère de dimension κ . La projection de ces points dans le plan de la i -ème et j -ème dimension est un diagramme de dispersion partiel dont les points résident dans un disque. Dès lors, il est pertinent d'exprimer ces points projetés non pas en coordonnées cartésiennes mais en coordonnées polaires, chaque point représentatif d'une réalisation \mathbf{x} du champ étant alors repéré par un angle $\phi(\mathbf{x})$ et un rayon $r(\mathbf{x})$. La tangente de cet angle fournit l'information qui nous intéresse, à savoir, le rapport des poids spatiaux :

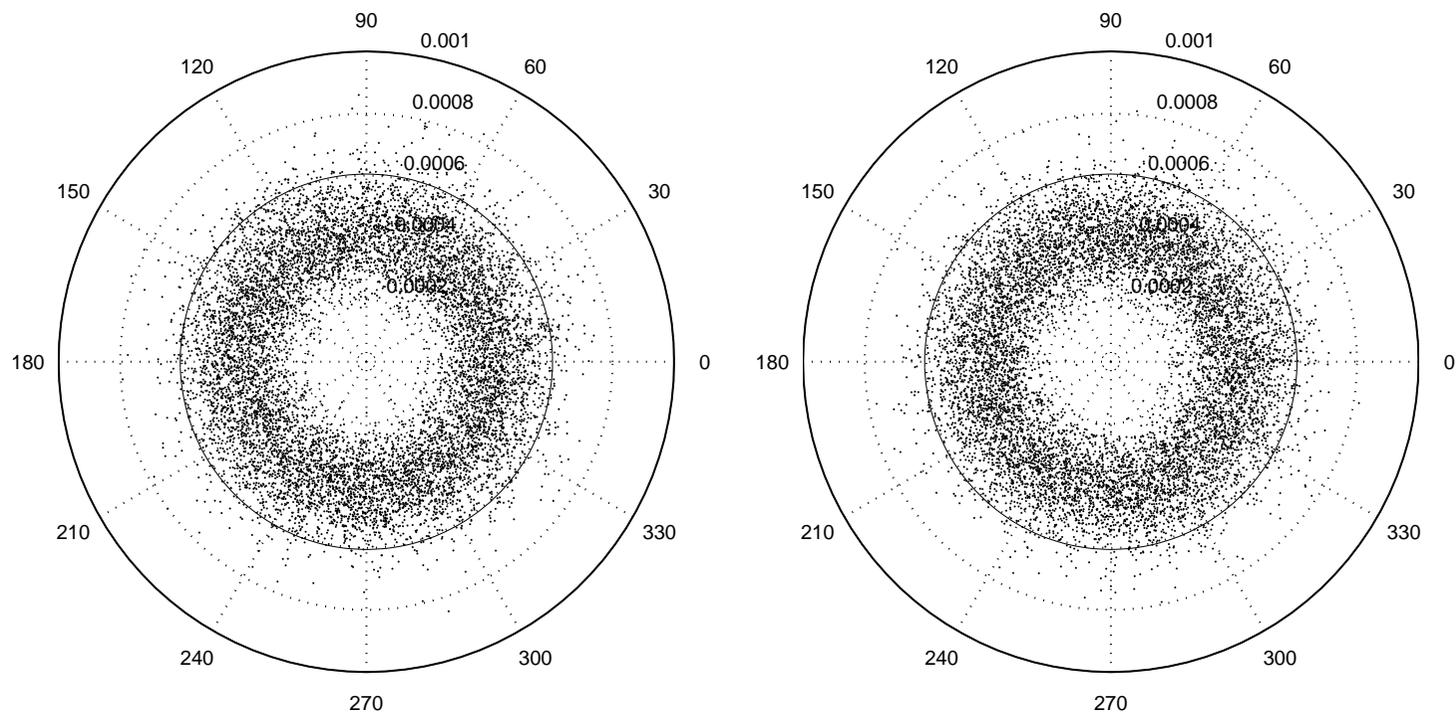
$$\tan \phi(\mathbf{x}) = \sqrt{\frac{n_j}{n_i}} \frac{w_j(\mathbf{x})}{w_i(\mathbf{x})} = \sqrt{\frac{n_j}{n_i}} \frac{C_j(\mathbf{x})}{C_i(\mathbf{x})}. \quad (\text{F.32})$$

Le rayon de l'hypersphère n'a en soi pas de signification. La variation de $r(\mathbf{x})$ dans un diagramme de dispersion partielle n'est qu'un effet de projection. Dès lors, pour toutes les projections, il peut être remplacé par n'importe quelle autre fonction positive des réalisations du champ de sorte à enrichir le diagramme de dispersion. Nous choisissons de substituer le rayon $r(\mathbf{x})$ par la valeur toujours positive de l'indice de Moran maximal $I_{\max}(\mathbf{x})$. Cette opération convertit le diagramme de dispersion partiel en un nouveau diagramme permettant de visualiser le rapport entre les poids spatiaux et de savoir en même temps si le test associé à ces poids est significatif (rejet de l'hypothèse H_0 d'indépendance spatiale) à seuil α donné. La distribution angulaire des points sont invariants des unités du champ spatial. Nous désignons ce type de diagramme un *diagramme de dispersion "angle-Imax" partiel*.

F.2.2 Résultats

Les figures F.1 à F.13 illustrent des diagrammes angle-Imax pour différentes valeurs de p et ρ d'un champ spatial de 100×100 sites (réseau carré). Nous faisons figurer sur chacun de ces diagrammes un cercle de rayon $I_{0.95}$ ($I_{0.95} = 6.048 \cdot 10^{-4}$ pour un champ de 100×100 sites) à l'intérieur duquel tout point est associé à un test non significatif (au seuil $\alpha = 0.05$) et à l'extérieur duquel, au contraire, tout point est associé à un test significatif (au seuil $\alpha = 0.05$).

Note Bien que non indiqué dans les légendes des diagrammes, les poids spatiaux w_i apparaissent dilatés d'un coefficient $\sqrt{n_i}$. Ce coefficient est fonction croissante de i .



(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.00$

(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.00$

Fig. F.1:

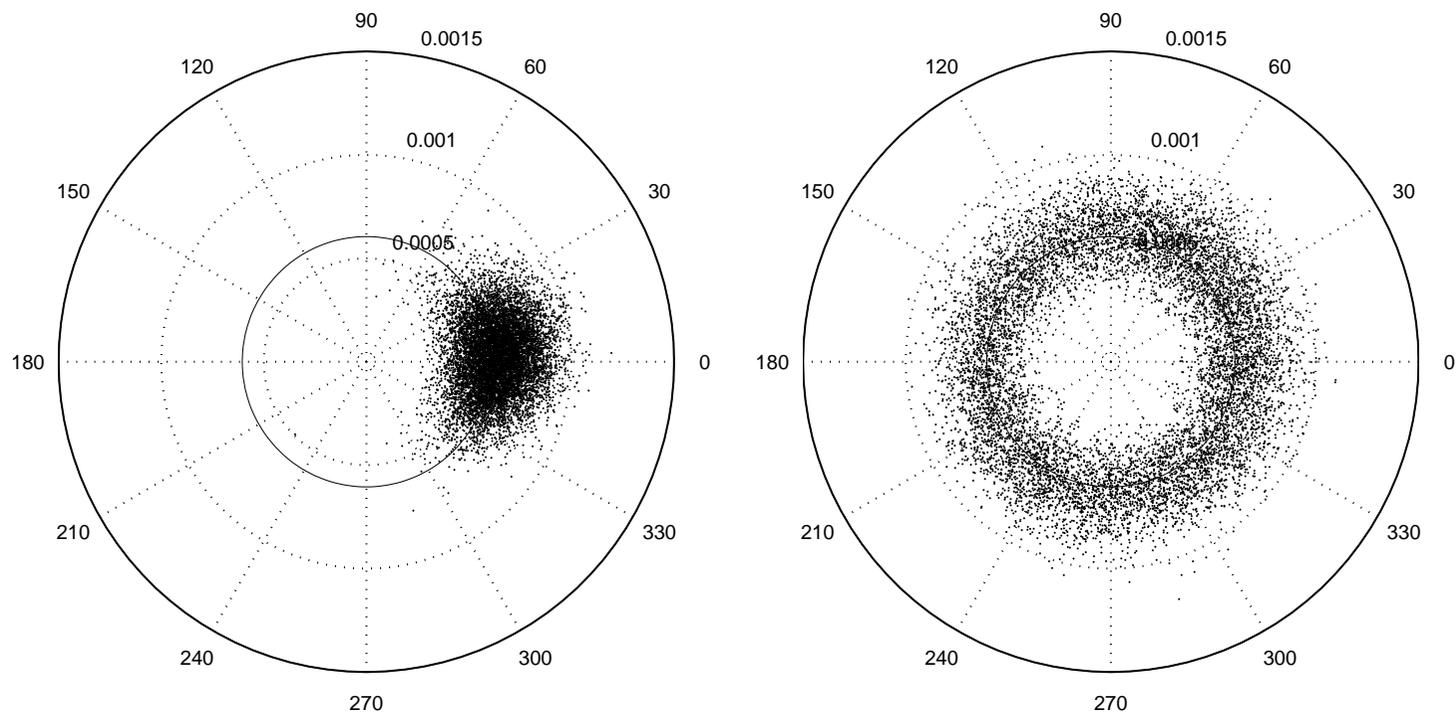
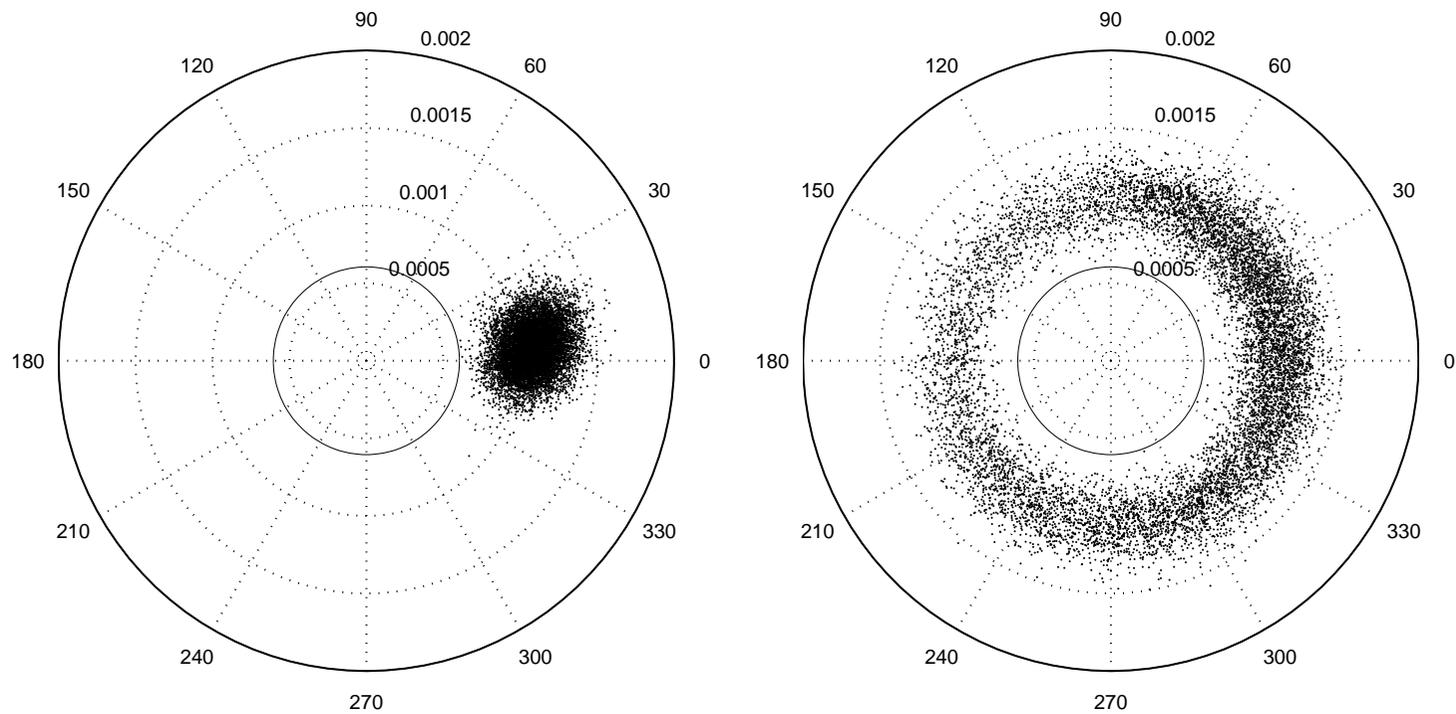


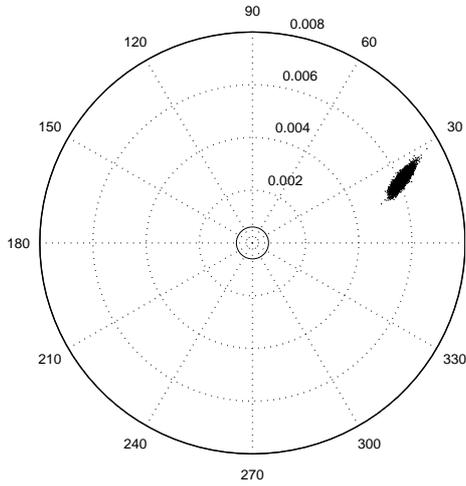
Fig. F.2:



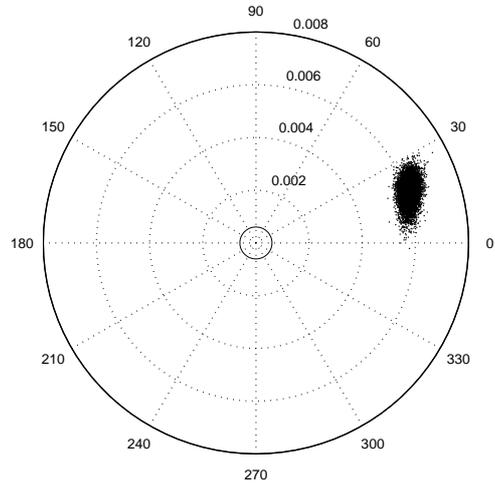
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.10$ et $p = 0.00$

(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.10$ et $p = 0.00$

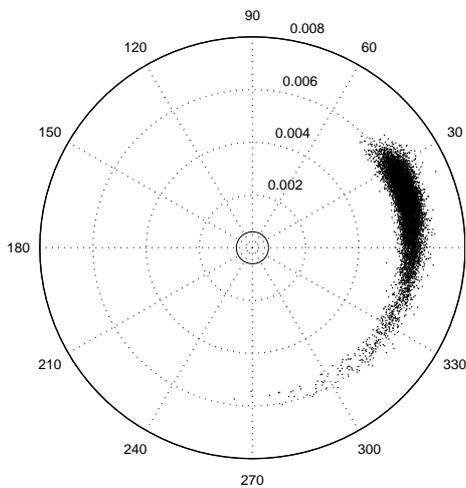
Fig. F.3:



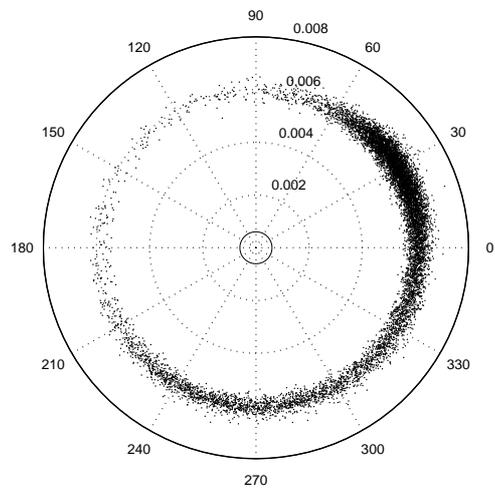
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

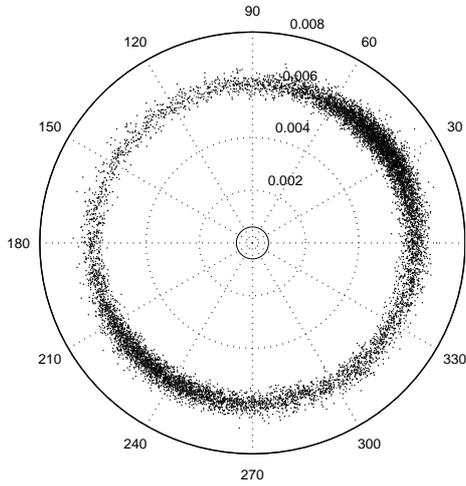


(c) Diagramme angle- I_{\max} obtenu à partir de (w_3, w_4, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

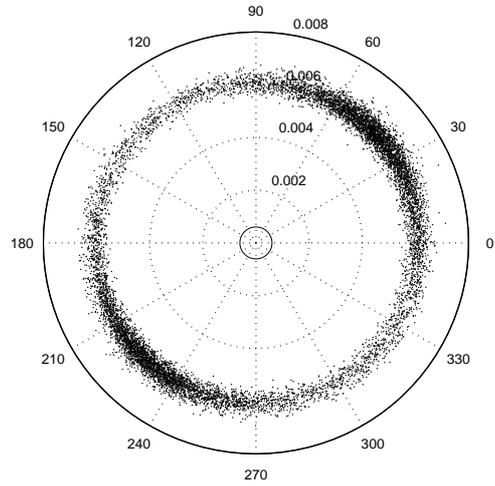


(d) Diagramme angle- I_{\max} obtenu à partir de (w_4, w_5, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

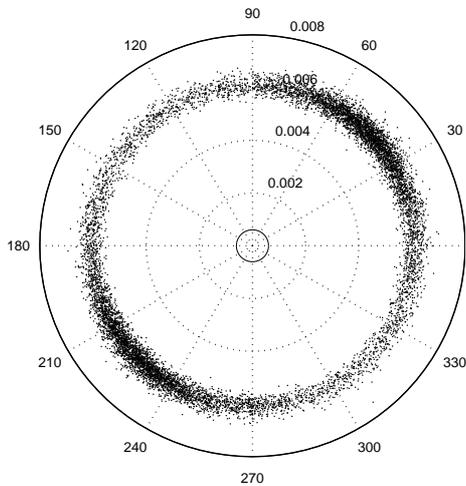
Fig. F.4:



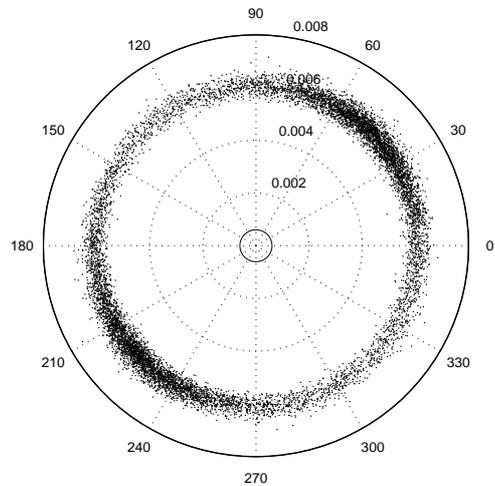
(a) Diagramme angle- I_{\max} obtenu à partir de (w_5, w_6, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_6, w_7, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

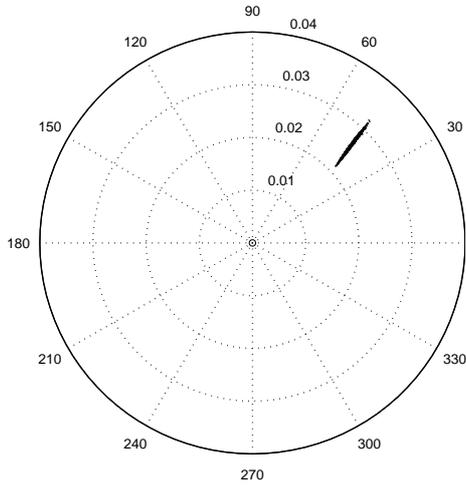


(c) Diagramme angle- I_{\max} obtenu à partir de (w_7, w_8, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

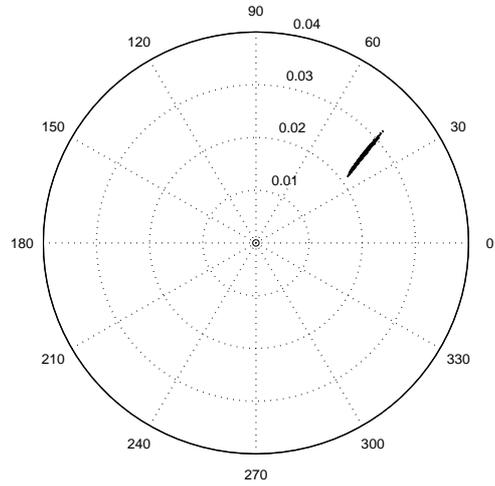


(d) Diagramme angle- I_{\max} obtenu à partir de (w_8, w_9, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.00$

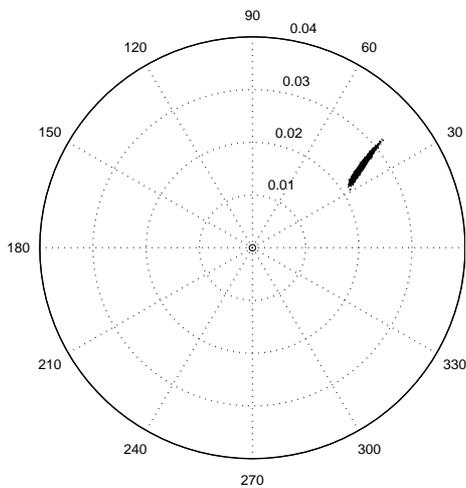
Fig. F.5:



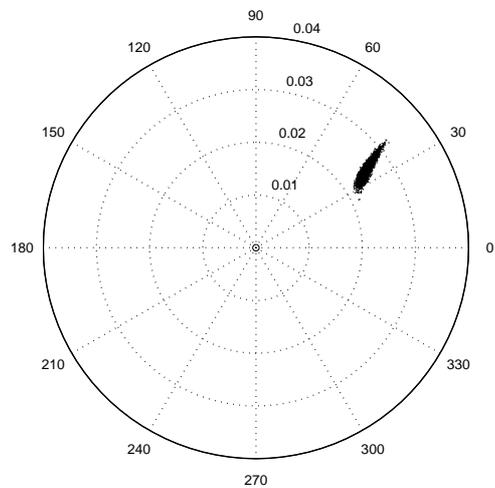
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$

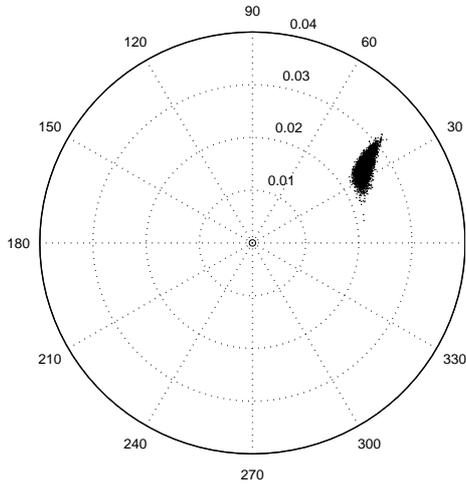


(c) Diagramme angle- I_{\max} obtenu à partir de (w_3, w_4, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$

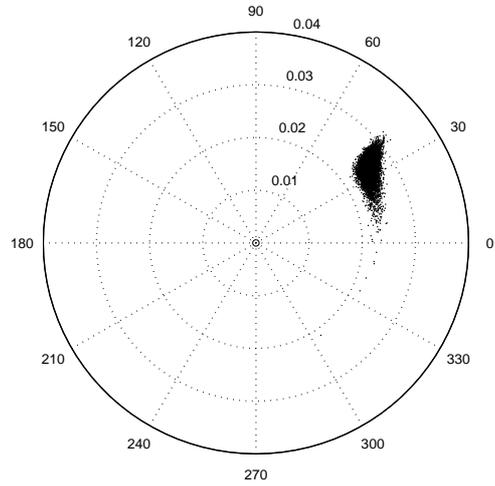


(d) Diagramme angle- I_{\max} obtenu à partir de (w_4, w_5, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$

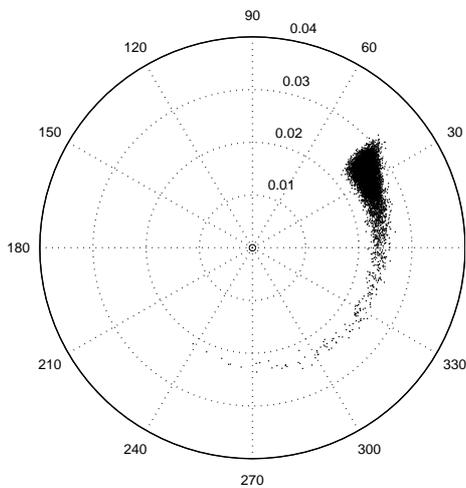
Fig. F.6:



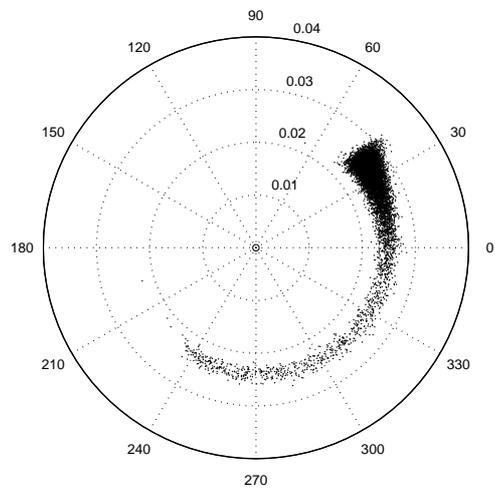
(a) Diagramme angle- I_{\max} obtenu à partir de (w_5, w_6, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_6, w_7, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$

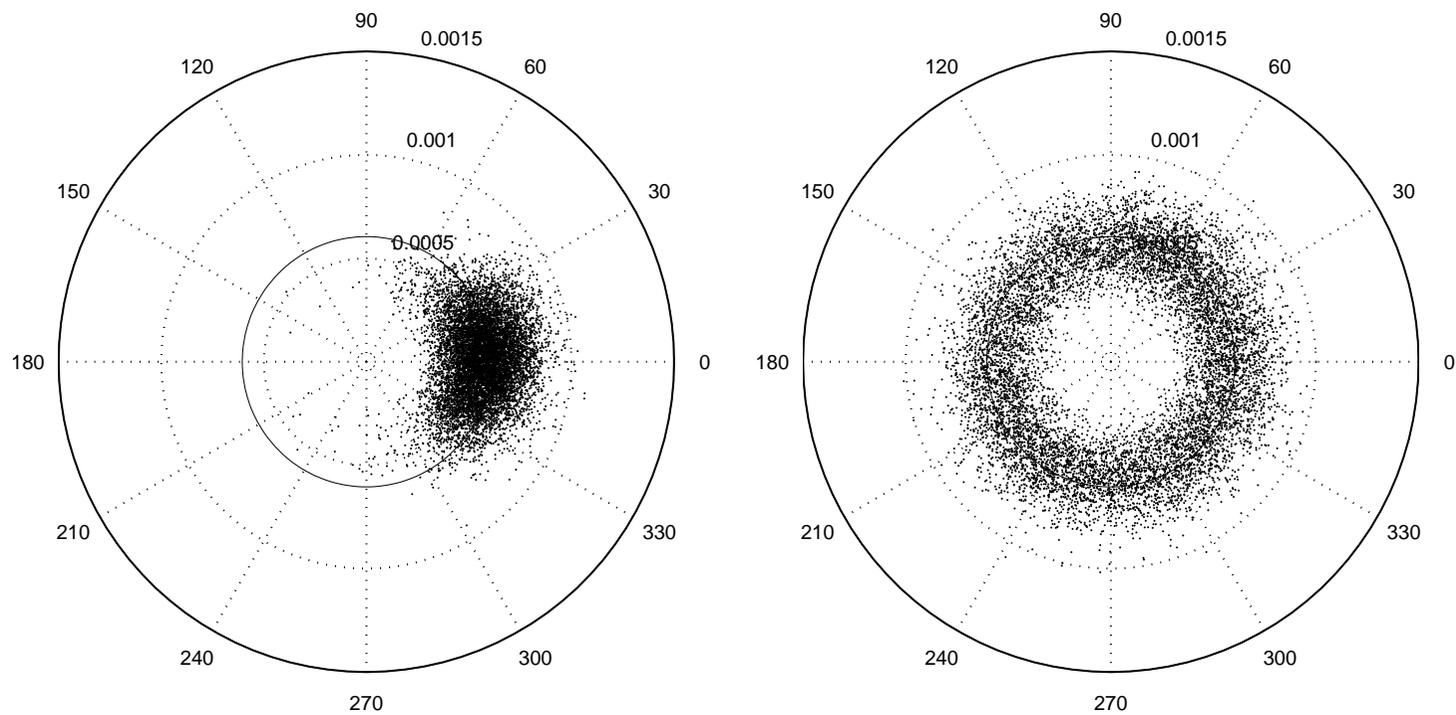


(c) Diagramme angle- I_{\max} obtenu à partir de (w_7, w_8, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$



(d) Diagramme angle- I_{\max} obtenu à partir de (w_8, w_9, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.00$

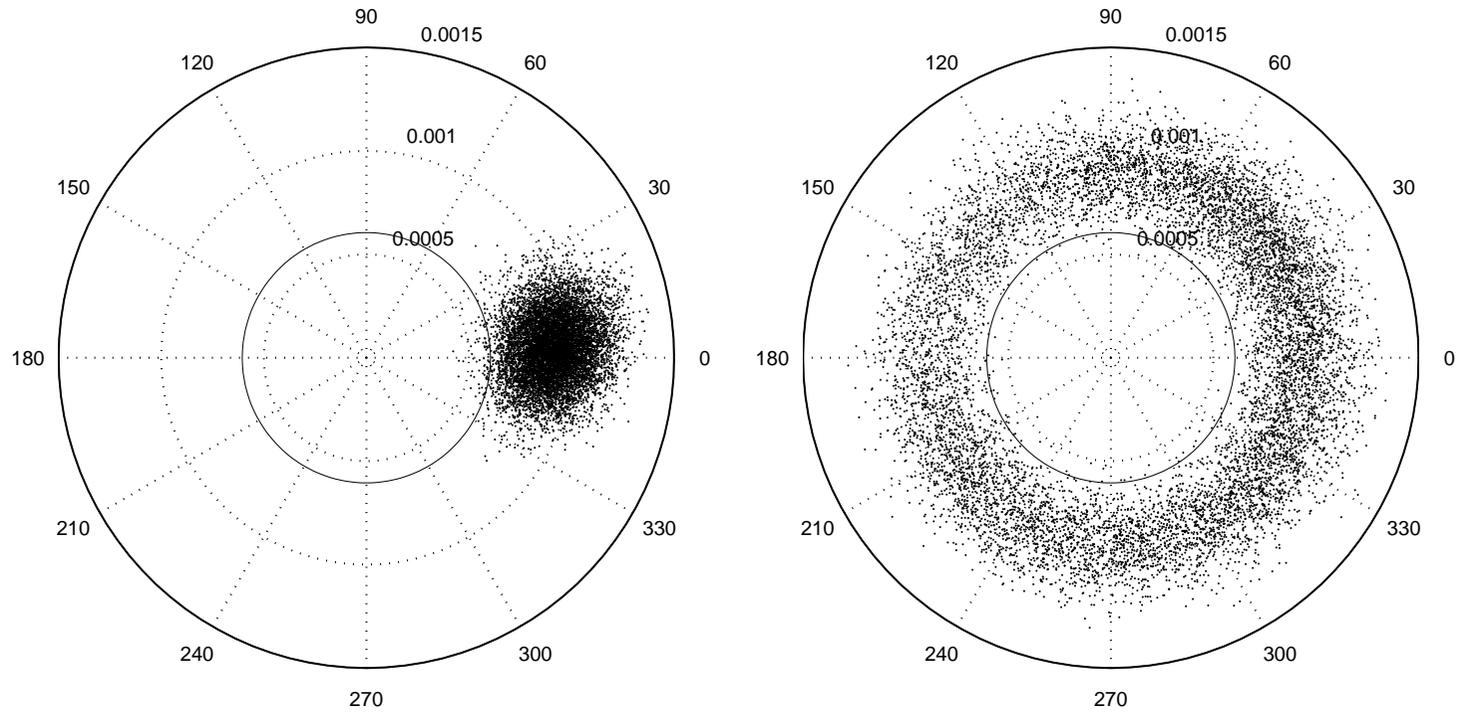
Fig. F.7:



(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.05$ et $p = 0.65$

(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.05$ et $p = 0.65$

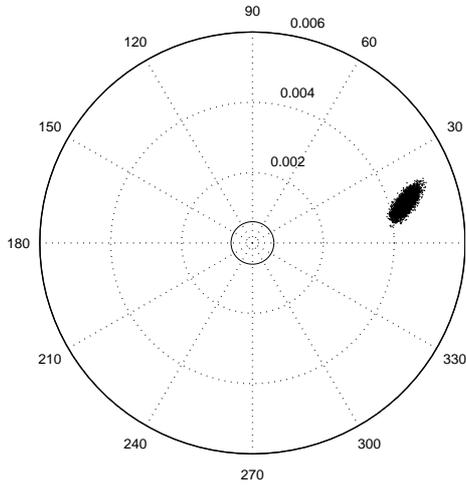
Fig. F.8:



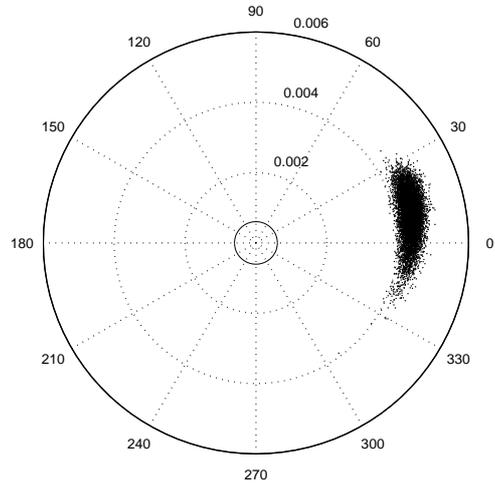
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.10$ et $p = 0.65$

(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.10$ et $p = 0.65$

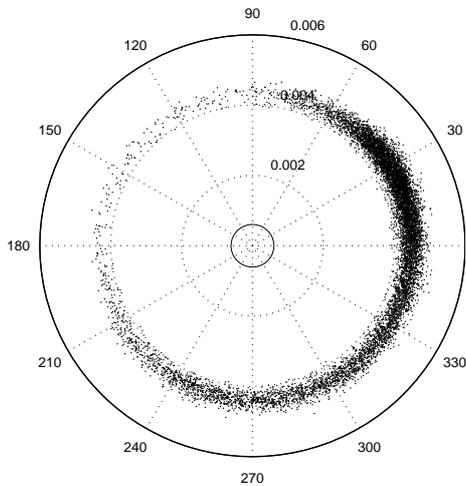
Fig. F.9:



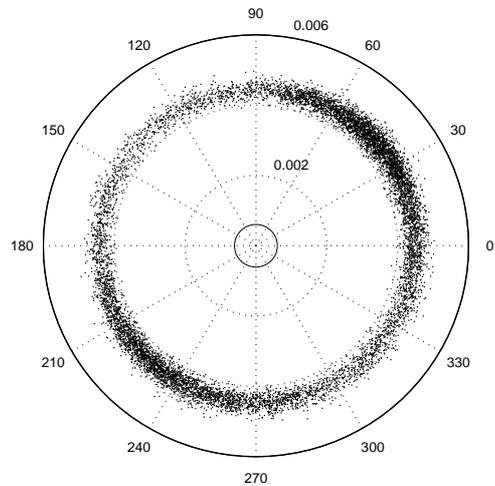
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

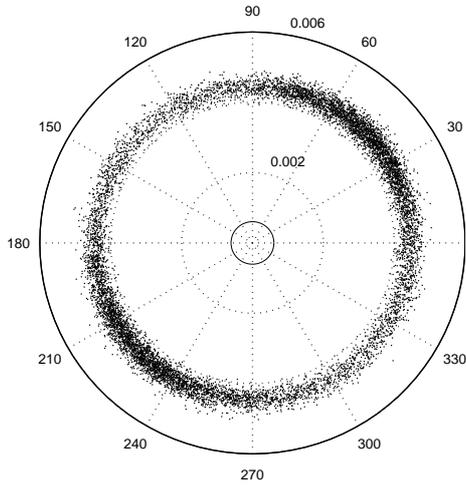


(c) Diagramme angle- I_{\max} obtenu à partir de (w_3, w_4, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

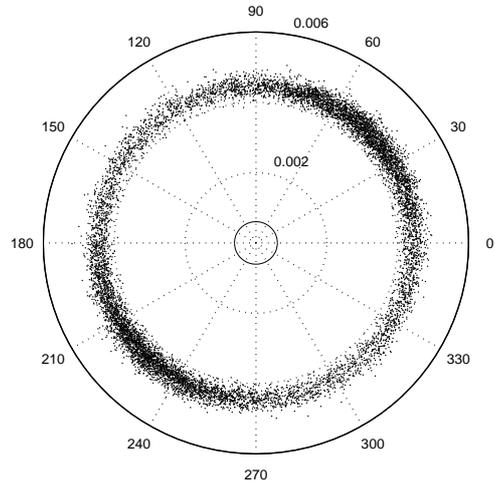


(d) Diagramme angle- I_{\max} obtenu à partir de (w_4, w_5, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

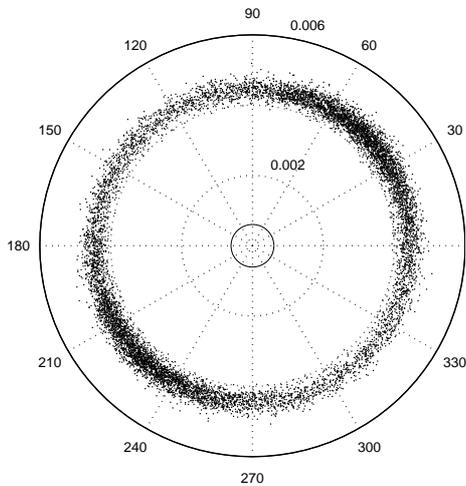
Fig. F.10:



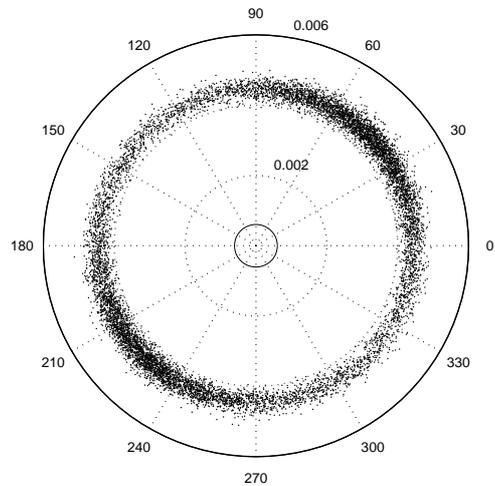
(a) Diagramme angle- I_{\max} obtenu à partir de (w_5, w_6, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_6, w_7, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

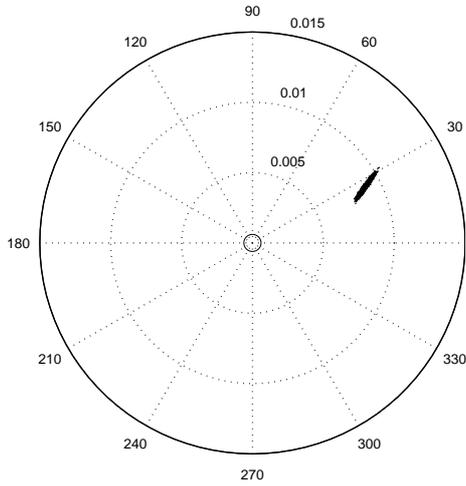


(c) Diagramme angle- I_{\max} obtenu à partir de (w_7, w_8, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

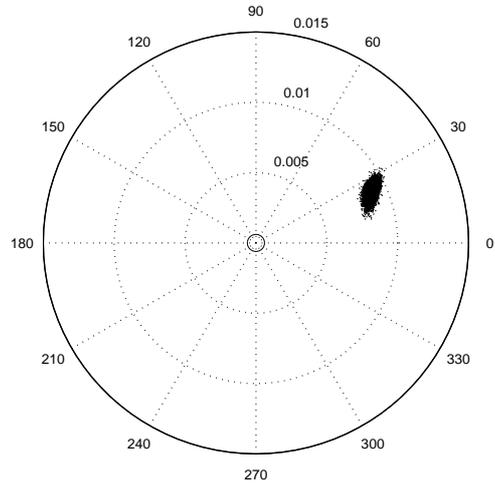


(d) Diagramme angle- I_{\max} obtenu à partir de (w_8, w_9, I_{\max}) , sous condition $\rho = 0.50$ et $p = 0.65$

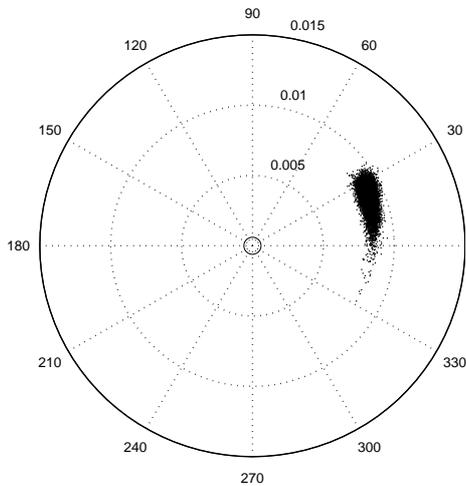
Fig. F.11:



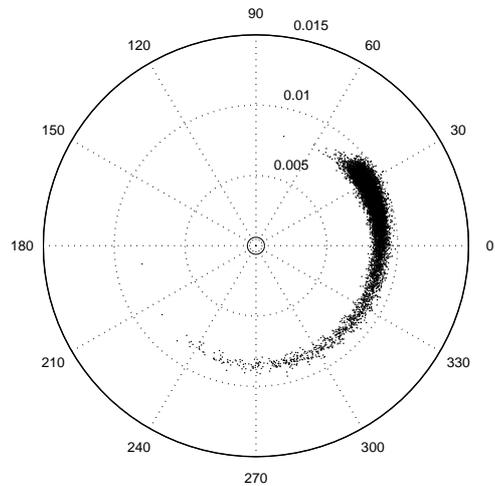
(a) Diagramme angle- I_{\max} obtenu à partir de (w_1, w_2, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_2, w_3, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$

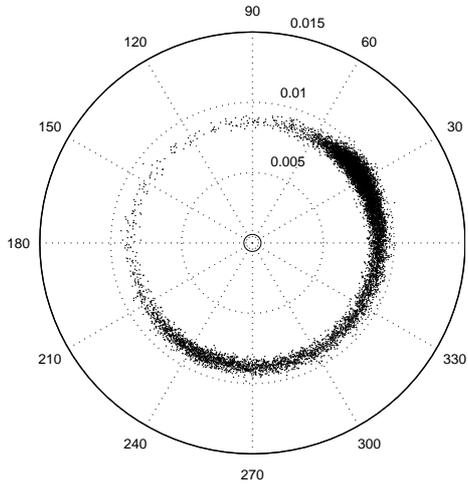


(c) Diagramme angle- I_{\max} obtenu à partir de (w_3, w_4, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$

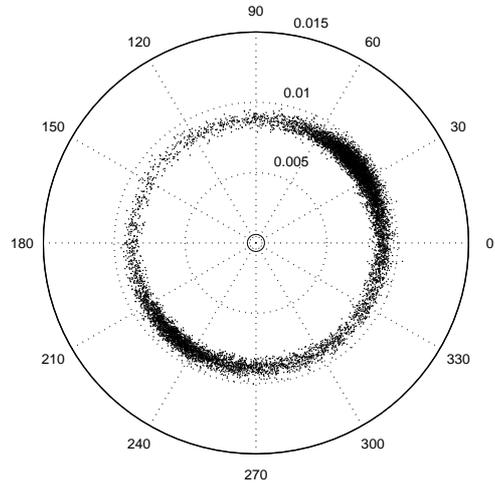


(d) Diagramme angle- I_{\max} obtenu à partir de (w_4, w_5, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$

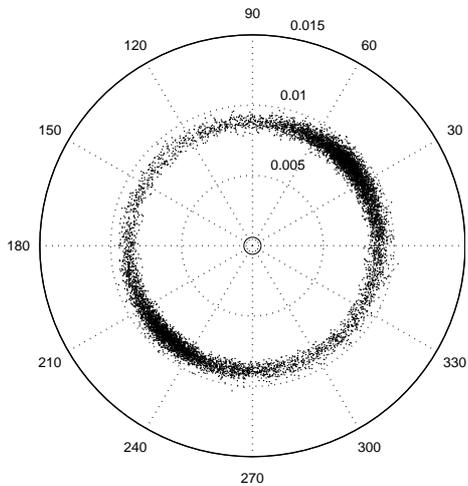
Fig. F.12:



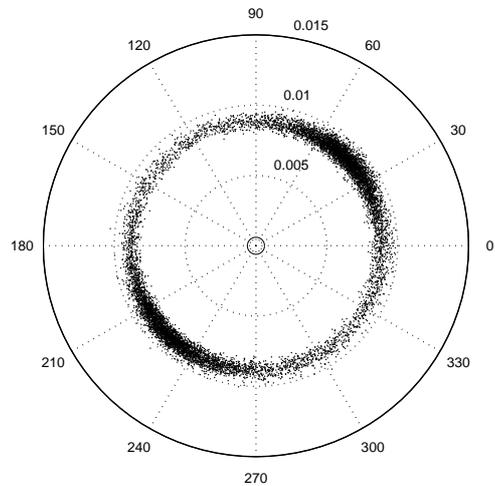
(a) Diagramme angle- I_{\max} obtenu à partir de (w_5, w_6, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$



(b) Diagramme angle- I_{\max} obtenu à partir de (w_6, w_7, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$



(c) Diagramme angle- I_{\max} obtenu à partir de (w_7, w_8, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$



(d) Diagramme angle- I_{\max} obtenu à partir de (w_8, w_9, I_{\max}) , sous condition $\rho = 0.90$ et $p = 0.65$

Fig. F.13:

Bibliographie

- Anselin, L., Rey, S. 1991, Properties of Tests for Spatial Dependence in Linear Regression Models. *Geographical Analysis*, 23, 2, 112-131.
- Anselin, L., Florax, R., 1995, Small Sample Properties of Tests for Spatial Dependence in Regression Models : Some further Results. In : *New Directions in Spatial Econometrics*, 21-74. Anselin L. & Florax R. editors. Springer Verlag, Berlin.
- Bahadur, R. R., 1967, Rates of Convergence of Estimates and Test Statistics. *The Annals of Mathematical Statistics*, 38, 2, 303-324.
- Bavaud, F., 1998, Models for Spatial Weights : A Systematic Look. *Geographical Analysis*, 30, 2, 153-171.
- Besag, J. E., 1974, Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society B*, 36, 192-225.
- Bhattacharjee, A., Jensen-Butler, C., 2006, Estimation of Spatial Weights Matrix, with an Application to Diffusion in Housing Demand. CRIEFF Discussion Paper 0519. University of St Andrew, UK.
- Boots, B., Dufournaud, C. 1994, A Programming Approach to Minimizing and Maximizing Spatial Autocorrelation Statistics. *Geographical Analysis*, 26, 1, 54-66.
- Broadbent, S. R., Hammersley, J. M., 1957, *Proc. Camb. Phil. Soc.*, 53, 629.
- Bullock, B. P., Burkhard, H. E., 2005, An Evaluation of Spatial Dependency in Juvenile Loblolly Pine Stands Using Stem Diameter. *Forest Science*, 51, 2, 102-108.
- Bunde, A., Havlin, S., 1996, *Fractals and Disordered Systems*, 2nd rev. ed., Springer, Berlin.
- Carrillo, D., 2007, Estimation et inférence du coefficient d'autorégression du modèle de Whittle sur un réseau d'interactions aléatoires faibles. *Revue Internationale de Géomatique*, 17, 3-4, 261-275.

- Casella G., George, E., 1992, Explaining the Gibbs Sampler. *The American Statistician*, 46, 3, 167-174.
- Celik, M., Kazar, B., Shekhar, S., Boley, D., Lilja, D., 2007, Technical Report n° TR 07-004. Department of Computer Science and Engineering, University of Minnesota, USA.
- Chen, Y., Paul, G., Cohen, R., Havlin, S., Borgatti, S. P., Liljeros F., Stanley, H. E., 2007, Percolation theory and fragmentation measures in social networks. *Physica A*, 378, 11-19.
- Chib, S., Greenberg, E., 1996, Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory*, 12, 3, 409-431.
- Cliff, A. D., Ord, J. K. 1973, *Spatial autocorrelation*. Pion, London.
- Cliff, A. D., Ord, J. K. 1975, The Choice of a Test for Spatial Autocorrelation, in *Display and Analysis of Spatial Data*. 54-77, editors Davis, J. C., McCullagh, M. J., John Wiley, Chichester.
- Cliff, A. D., Ord, J. K. 1981, *Spatial Processes : Models and Applications*. Pion, London.
- Cochran, W. G., 1952, The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 23, 315-345.
- Cover, T. M., Thomas, J. A., 1991, *Elements of Information Theory*. John Wiley & Sons.
- Cressie, N. A. C., 1993, *Statistics for spatial data* (rev. ed.). John Wiley & Sons, Inc., New York.
- Dark, S. J., 2004, The biogeography of invasive alien plants in California : an application of GIS and spatial regression analysis. *Diversity and Distributions*, 10, 1-9.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, B*, 39, 1, 1-38.
- Doreian, P., 1980, Linear Models with Spatially Distributed Data. *Sociological Methods & Research*, 9, 1, 29-60.
- Doreian, P., 1981, Estimating linear models with spatially distributed data. *Sociological Methodology*, 12, 359-388.
- Dow, M. M., Burton, M. L., 1982, White, D. R., Network autocorrelation : a simulation study of a foundational problem in regression and survey reserach. *Social Networks*, 4, 169-200.
- Durbin, J., Watson, G. S., 1950, Testing for Serial Correlation in Least Squares Regression - I, *Biometrika*, 37, 409-428.
- Durbin, J., Watson, G. S., 1951, Testing for Serial Correlation in Least Squares Regression - II, *Biometrika*, 38, 159-178.

- Durbin, J., Watson, G. S., 1971, Testing for Serial Correlation in Least Squares Regression - III, *Biometrika*, 58, 1-19.
- Ferguson, T. S., 1967, *Mathematical Statistics, A decision theoretic approach*. Academic Press, New York.
- Eldiery, A., Garcia, L. A., Reich, R. M., 2005, Estimating Soil Salinity from Remote Sensing Data in Corn Fields. *Proceedings of Hydrology Days 2005 (of the American Geophysical Union)*.
- Florax, R. J. G. M., Rey S., 1995, The impacts of misspecified spatial interaction in linear regression models, in *New Directions in Spatial Econometrics*, 75-95, Anselin L., Florax R. J. G. M. (Eds), Springer-Verlag, Berlin.
- Flory, P. J., 1941, *J. Am. Chem. Soc.*, 63, 3083.
- Fortin, M.-J., Boots, B., Csillag, F., Rempel, T. K., 2003, On the role of spatial stochastic models in understanding landscape indices in ecology. *Oikos*, 102, 1, 203-212.
- Gatrell, A. C., 1979, Autocorrelation in spaces. *Environment and Planning A*, 11, 507-516.
- Geary, R. C., 1954, The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5, 115-145.
- Geman, S., Geman, D., 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE transaction on pattern analysis and machine intelligence*, 6, 721-741.
- Gil-Pelaez, J., 1951, A Note on the Inversion Theorem. *Biometrika*, 38, 481-482.
- Gilbert, F., Gonzalez, A., Evans-Freke, I., 1998, Corridors maintain species richness in the fragmented landscapes of a microecosyst. *Proceedings of the Royal Society of London B*, 265, 577-582.
- Goodwin, B. J., 2003, Is landscape connectivity a dependent or independent variable? *Landscape Ecology*, 18, 687-699.
- Griffith, D. A., 1992, What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espace géographique*, 3, 265-280.
- Griffith D., 2004, Faster maximum likelihood estimation of very large spatial autoregressive models : an extension of the Smirnov-Anselin result. *Journal of Statistical Computation and Simulation*, 74, 12, 855-866.
- Haining, R. P. 1978a, Specification and estimation problems in models of spatial dependence. *Studies in geography*, Northwestern University, Evanston, IL.
- Haining, R. P., 1978b, The moving average model for spatial interaction. *Transaction and Papers, Institute of British Geographers, New Series*, 3, 202-225.
- Haining, R. P. 1990, *Spatial data analysis in the social and environmental science*. Cambridge University Press.
- Hastings, W. K., 1970, Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57,97-109.

- Hors, I., 1995, Des modèles de transition de phase en économie. *Revue économique*, 46, 3, 817-826.
- Hubert, L. J., Golledge, R. G., Costanzo, C. M., 1981, Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis*, 13, 224-233.
- Imhof, J. P. 1961, Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 3 & 4, 419-426.
- Kallenberg, W. C. M., 1983, Intermediate efficiency, theory and examples. *Ann. Statist.*, 11, 170-182.
- Kauffmann, P., 1994, *Statistique : Information, Estimation, Tests*. Dunod, Paris.
- Keitt, T., Urban, D. L., Milne, B. T., 1997, Detecting Critical Scales in Fragmented Landscapes, *Conservation Ecology*, 1, 1.
- Kelejian, H. H., Robinson, D. P., 1995, Spatial Correlation : A Suggested Alternative to the Autoregressive Model, in *New Directions in Spatial Econometrics*, 75-95, Anselin L., Florax R. J. G. M. (Eds), Springer-Verlag, Berlin.
- Kendall, M. G., Stuart, A., 1973, *The Advanced Theory of Statistics*. Vol. 2, Third Ed., Griffin, London.
- King, M. L., 1981, A small sample property of the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society B*, 43, 263-264.
- Kissling, W. D., Carl, G., 2008, Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17, 59-71.
- Kooijman, S. A. L. M., 1976, Some remarks on the statistical analysis of grids especially with respect to ecology. *Annals of Systems Research* 5, 113-132.
- Kühn, I., 2007, Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13, 66-69.
- Kullback, S., 1954, Certain inequalities in information theory and the Cramér-Rao inequality. *Ann. Math. Statist.*, Vol. 25, 745-751.
- Kullback, S., 1959, *Information theory and statistics*. John Wiley & Sons, New York. Monographie édité à nouveau en 1997 par Dover publications, Mineola.
- Lebart, L., Morineau, A., Piron, M., 1995, *Statistique exploratoire multidimensionnelle*. Dunod, Paris.
- Leenders, R. Th. A. J., 2002, Modeling social influence through network autocorrelation : constructing the weight matrix. *Social Networks*, 24, 21-47.
- LeSage, J. P., 1997, Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review*, 20, 1 & 2, 113-129.
- Li, H., Reynolds, J. F., 1995, On definition and quantification of heterogeneity. *Oikos*, 73, 2, 280-284.
- Loftin, C., Ward, S.K., 1983, A Spatial Autocorrelation Model of the Effects of Population Density on Fertility. *American Sociological Review*, 48, 1, 121-128.

- MacCluer, C. R., 2000, The Many Proofs and Applications of Perron's Theorem. *SIAM Rev.* 42 (3), pp. 487-498.
- Mathai, A. M., Provost, S. B., 1992, Quadratic forms in random variables. Marcel Dekker Inc., New York.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953, Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.
- Militino, A. F., Ugarte, M. D., Garcia-Reinaldos, L., 2004, Alternative Models for Describing Spatial Dependence among Dwelling Selling Prices. *Journal of Real Estate Finance and Economics*, 29, 2, 193-209.
- Mizruchi, M. S., Neuman, E. J., 2008, The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, 30, 190-200.
- Moran, P. A. P., 1950, Note on continuous stochastic phenomena. *Biometrika*, 37, 17-23.
- von Neumann, J., 1941, Distribution of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12, 367-395.
- Neyman, J., Pearson, E. S., 1933, On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.*, A, 231, 289.
- Noether, G. E., 1955, On a theorem of Pitman. *Annals of Mathematical Statistics*, 25, 64-68.
- Oden, N. L., 1984, Assessing the Significance of a Spatial Correlogram. *Geographical Analysis*, 16, 1, 1-16.
- Openshaw, S., Taylor, P. J., 1979, A million or so correlation coefficients : three experiments on the modifiable areal unit problem. *In* *Statistical applications in the spatial sciences*, 127-144, Wrigley, N. (ed.), Pion, London.
- Ord, K., 1975, Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70, 349, 120-126.
- Pace, R. K., 1997, Performing large-scale spatial autoregressions. *Economics Letters*, 54, 283-291.
- Pace, R. K., Barry, R., 1997, Quick computation of spatial autoregressive estimators. *Geographical Analysis*, 29, 232-247.
- Pace, R. K., Barry, R., 1997, Sparse spatial autoregression. *Statistics and Probability Letters*, 33, 291-297.
- Pajot, S., 2003, Intégration du marché global dans un système composé de marchés locaux. *Revue économique*, 54, 3, 675-686.
- Pitman, E. J. G., 1948, Lecture Notes on Non-parametric Inference, notes non publiées utilisées par l'auteur dans ses cours à l'Université de Stanford, Calif..
- Porojan, A., 2001, Trade Flows and Spatial Effects : The Gravity Model Revisited. *Open economies review*, 12, 265-280.

- Provost, S. B., 1989a, On sums of independent gamma variables. *Statistics* 20, 1-8.
- Provost, S. B., 1989b, The distribution function of some ratios of quadratic forms. In "Recent Developments in Statistical and Actuarial Science". SCITEX Publications, London, Canada, 143-154.
- Saporta, G., 1990, Probabilités, analyse des données et statistique. Editions Technip, Paris.
- Sanov, I. N., 1957, On the probability of large deviations of random variables. *Mat. Sbornik (Moscou)*, Vol. 42, No 1 (84), 11-44 (Russe). Traduction : N. C. Inst. of Statist. Mimeo Series No 192, Mar. 1958.
- Sen, A. K., 1974, Notes on spatial autocorrelation. (Non publiées).
- Sen, A. K., 1976, Large sample-size distribution of statistics used in testing for spatial correlation. *Geographical Analysis*, 8, 175-184.
- Smirnov O., Anselin L., 2001, Fast maximum likelihood estimation of very large spatial autoregressive models : a characteristic polynomial approach. *Computational Statistics & Data Analysis*, vol. 35, 301-319.
- Sokal, R. R., Oden, N. L., 1978, Spatial autocorrelation in biology, 1. Methodology. *Biological Journal of the Linnean Society*, 10, 199-228.
- Stauffer, D., Aharony, A., 1994, Introduction to Percolation Theory. Second Ed., Taylor & Francis, London.
- Stockmayer, W. H., 1943, *J. Chem. Phys.*, 11, 45.
- Tiefelsdorf, M., Boots, B., (1995), The exact distribution of Moran's I. *Environment and Planning A*, 27, 985-999.
- Tiefelsdorf, M., 2000, Modelling Spatial Processes : The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I. *Lecture Notes in Earth Sciences* 87, Springer, Berlin.
- Tiefelsdorf, M., 2002, The Saddlepoint Approximation of Moran's I's and Local Moran's I's Reference Distributions and Their Numerical Evaluation. *Geographical Analysis*, 34, 3, 187-206.
- Tobler, W. R., 1970, A computer movie simulating urban growth in the Detroit Region. *Economic Geography Supplement*, 46, 234-240.
- Tognelli, M. F., Kelt, D. A., 2004, Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, 27, 427-436.
- Upton, G., Fingleton, B. 1985, *Spatial Data Analysis by Example*, Vol 1, John Wiley & Sons, Chichester.
- van der Vaart, A. W., 1998, *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- Vigneron, E., 1997, *Géographie et statistiques, Que sais-je ?*, Presses universitaires de France, Paris.

- Wall, M. M., 2004, A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121, 311-324.
- Walsh, J. E., 1946, On the power function of the sign test for slippage of means. *Ann. Math. Statist.*, 17, 358.
- Weir, I. S., Pettitt, A. N., 1999, Spatial modelling for binary data using a hidden conditional autoregressive Gaussian process : a multivariate extension of the probit model.
- Whittle, P., 1954, On Stationary Process in the Plane. *Biometrika*, 41, 434-449.
- With, K. A., Gardner, R. H., Turner M. G., 1997, Landscape Connectivity and Population Distribution in Heterogeneous Environments. *Oikos*, 78, 151-169.
- Won Kim, C., Phipps, T. T., Anselin, L., 2003, Measuring the benefits of air quality improvement : a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45, 24-39.
- Ziff, R. M., 1992, Spanning probability in 2D percolation. *Physical Review Letter*, 69, 2670-2673.