

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de
la Recherche Scientifique

Université Mentouri de

Constantine

FACULTE DES SCIENCES EXACTES
Département de Mathématiques

N° d'ordre:

Série:

Thèse
Pour l'obtention du diplôme de
Doctorat En Sciences
En : **Mathématique**
Option : Probabilités et statistiques
Intitulée

Modélisation spatiale à changements de
régimes Markoviens

Présentée par

Diafri Houria

Devant le jury composé de:

Mme. S. Belaloui	Prof, Département de Mathématiques	U.M.C	Présidente
Mme. S. Kharfouchi	Prof, Faculté de Médecine	U.C.3	Rapporteuse
Mr. T. Hamaizia	MCA, Département de Mathématiques	U.M.C	Examineur
Mme. N. Abada	MCA, ENS Assia Djebbar	ENSC	Examinatrice

Soutenue le 05/04/2021

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de
la Recherche Scientifique

Université Mentouri de

Constantine

FACULTE DES SCIENCES EXACTES
Département de Mathématiques

N° d'ordre:

Série:

Thèse

Pour l'obtention du diplôme de

Doctorat En Sciences

En : **Mathématique**

Option : Probabilités et statistiques

Intitulée

**Modélisation spatiale à changements de
régimes Markoviens**

Présentée par

Djafri Houria

Devant le jury composé de:

Mme. S. Belaloui	Prof, Département de Mathématiques	U.M.C	Présidente
Mme. S. Kharfouchi	Prof, Faculté de Médecine	U.C.3	Rapporteuse
Mr. T. Hamaizia	MCA, Département de Mathématiques	U.M.C	Examineur
Mme. N. Abada	MCA, ENS Assia Djebbar	ENSC	Examinatrice

Soutenue le 05/04/2021

Remerciements

Je tiens tout d'abords à exprimer ma profonde gratitude à l'égard de Madame **S. Kharfouchi** professeur à l'université de Constantine 3 pour la confiance qu'elle m'a accordée en acceptant d'encadrer ma thèse et de l'avoir si bien fait. C'est, d'ailleurs, grâce à son soutien constant, à ses encouragements et à ses conseils précieux qu'elle m'a prodigué, que j'ai pu mener ce travail à bien.

Je reste très reconnaissant à Madame **S. Belaloui** professeur à l'université de Constantine 1, pour l'honneur qu'elle ma fait en acceptant de présider le jury de ce mémoire.

Je remercie aussi vivement monsieur **T. Hamaizia** professeur à l'université de Constantine 1 pour l'intérêt qu'il a bien voulu porter à ce travail et d'avoir accepté de le juger.

Je remercie également et profondément Madame **N. Abada** professeur à l'université de Constantine 3 pour l'intérêt qu'elle a bien voulu porter à ce travail et d'avoir accepté de le juger.

Table des matières

Table des figures	v
0.1 Contexte historique	2
0.2 Formulation du problème et motivation	2
0.3 L'objectif de cette thèse	3
0.4 Organisation du mémoire	4
0.5 Travaux réalisés	6
1 Généralités	7
1.1 Types de données spatiales	8
1.1.1 Données ponctuelles	8
1.1.2 Données surfaciques	9
1.1.3 Données continues	10
1.2 La dépendance spatiale	12
1.2.1 L'autocorrélation spatiale	13
1.2.2 L'hétérogénéité spatiale	15
1.3 Relations d'ordre- Ensembles indexants	17
1.3.1 Ordre total et ordre partiel	19
1.3.2 Ensembles indexants	20
1.3.3 Voisinage dans \mathbb{Z}^2	21
1.3.4 Les régions causales dans \mathbb{Z}^2	22
1.4 Notions de base	24
1.4.1 Champs aléatoires	24
1.4.2 Stationnarité	26
1.4.3 Ergodicité	29
1.5 Exemples des modèles spatiaux	30

1.5.1	Modèles autorégressifs 2D	31
1.5.2	Modèles <i>ARMA</i> spatiaux quadrantaux	35
1.5.3	Modèle <i>AR</i> spatial quadrantal	36
2	Propriétés probabilistes des processus <i>AR</i> unilatérale spatiaux	38
2.1	Introduction	38
2.2	Représentation du modèle <i>AR</i> unilatérale spatiaux	39
2.3	Stationnarité du processus <i>AR</i> spatial unilatéral	41
2.4	L'estimation dans les modèles 2D <i>AR</i> (2.1) unilatéral	44
2.4.1	Construction des matrices de poids pour le modèle autorégressif spatial du second ordre	44
2.4.2	Méthode de maximement de vraisemblance pour le modèle spatial <i>AR</i> (2.1)	47
2.4.3	Estimation spatiale Yule-Walker	50
2.4.4	Estimation spatiale conditionnelle des moindres carrés	52
2.5	L'estimation dans les modèles <i>AR</i> (p) spatial causal	55
2.5.1	L'estimateur des moindres carrés	55
2.5.2	L'estimateur de Yule-Walker	56
2.6	Consistance et normalité asymptotique des estimateurs	57
2.7	Identification de l'ordre p du modèle	59
3	Etude des champs aléatoires de Markov (<i>MRF</i>)	61
3.1	Introduction	61
3.2	Cadre théorique des champs aléatoires de Markov	63
3.2.1	Définitions	63
3.3	Champs aléatoires de Markov non causal (NMRF) :	68
3.3.1	Champs aléatoires de Gibbs	69
3.3.2	Equivalence champ aléatoire de Markov – champ aléatoire de Gibbs	70
3.4	NMRF fondamentaux	71
3.4.1	Modèle d'Ising	72
3.4.2	Modèle de Potts	73
3.5	Champ aléatoires de Markov causal (CMRF) en analyse d'image	74
3.5.1	Transformation une image 2D en une image 1D	75
3.5.2	Modèles de Markov cachés pseudo 2D	77
3.5.3	Définition CMRF	78
3.6	Modèles de Markov causaux 2D fondamentaux	80
3.6.1	Champ aléatoire de maillage de Markov	82

3.6.2	MRF unilatéral	84
3.7	Modélisation markovienne de la segmentation d'images	89
3.7.1	Segmentation	90
3.7.2	Champs de Markov cachés HMRF	90
3.7.3	Le graphe d'un champ de Markov cachée	91
3.7.4	Champs de Markov cachés à bruit indépendant	93
3.7.5	Champs de Markov cachés en segmentation d'images	94
3.7.6	Spécification de $U(x/y)$ pour le problème de segmentation d'image	95
3.8	Estimation des paramètres	97
4	Modèle 2D autorégressif à changement de régime Markovien unilatéral	
	<i>(MS - AR)</i>	100
4.1	Introduction	100
4.2	Modèles 1D MS-AR	101
4.2.1	Description du modèle	102
4.2.2	Mise en équation	102
4.3	Construction du modèle 2D MS-AR	105
4.3.1	Le champ aléatoire de Markov causal homogène	105
4.3.2	Présentation du modèle 2D MS - AR	108
4.4	Stationnarité stricte de 2D MS - AR	109
4.5	Estimation des paramètres du modèle 2D MS - AR	113
4.5.1	La fonction de vraisemblance dans les modèles 2D MS - AR	113
4.5.2	L'estimation par l'algorithme VEM	115
A	Chaînes de Markov à temps discret	119
A.1	Définitions	119
A.2	Classification des états	122
A.2.1	Communicant	122
A.2.2	Irréductible	123
A.2.3	Périodicité	124
A.2.4	Distribution initiale	124
A.2.5	Distribution stationnaire	124
A.2.6	Ergodique	125
B	Modèles de Markov cachés à temps discret	126
B.1	Présentation et caractéristiques	126
B.2	Caractéristiques des MMC	127
	Bibliographie	129

Table des figures

1.1	Détection de clusters significatifs	9
1.2	collège lui-même	11
1.3	la prédiction (à droite)	12
1.4	Les différents types d'autocorrélation	14
1.5	Taux d'hospitalisation pour la Covid-19 par département (par décile)	17
1.6	Taux de décès dus à la Covid-19 en milieu hospitalier par département (par décile)	18
1.7	Taux de surmortalité par département au 30 mars 2020 par rapport à la moyenne de 2018 et 2019 (par décile)	18
1.8	définition du passé, du présent et du futuro	22
1.9	Configurations de voisinage d'un processus AR – Les pixels considérés pour chaque type de voisinage sont les points foncés.	33
1.10	Région de AR unilatéral (4 voisin)	35
2.1	Fenêtre de prédiction pour le modèle AR spatial du premier ordre	43
3.1	Voisinage symétrique et non symétrique : (a) voisinage passé (gris clair) et futur (gris foncé). (b) voisinage symétrique de second ordre. (c) voisinage non symétrique (passé) de second ordre.	65
3.2	Cliques associées à un système de voisinage en 4-connexité et en 8-connexité	66
3.3	Orientation d'un spin à l'état d'équilibre	72
3.4	Méthodes de numérisation : (a) balayage raster, (b) balayage en zigzag, (c) balayage Peano, et (d) ordre Morton.	76
3.5	Structure de PHMM 2D.	78
3.6	Les sites ombrés sont (a) passé, (b) support causal et (c) voisins causaux associés aux modèles MMRF du troisième ordre.	79
3.7	(a) MMRF, (b) chaîne de Markov NHSP, (c) MRF unilatéral, et (d) champ aléatoire de Pickard.	81
3.8	Région rectangulaire régulier X de taille $m \times n$: $Z_{i,j}$ est la zone grise, les flèches indiquent le balayage raster	83
3.9	Configuration spatiale des pixels dans un MRF unilatéral.	84
3.10	Configuration spatiale des pixels dans un MRF unilatéral.	85

3.11	Différentes structures voisines	88
3.12	segmentation d'images.	92
3.13	La structure d'un modèle de champ de Markov caché	92
4.1	Graphe d'indépendance conditionnelle pour le modèle MS-AR d'ordre 1 . . .	103

Abréviations & Notations

AR	AutoRégressif
CM	Chaine de Markov
CMRF	Champ aléatoire de Markov causal
2D AR	AutoRégressif bidimensionnel
2D MS-AR	AutoRégressif à changement de régime Markovien bidimensionnel
EM	Expectation Maximization
GRF	Champ aléatoire de Gibbs
HMRF	Champ aléatoire de Markov caché
MRF	Champ aléatoire de Markov
MS-AR	AutoRégressif à changement de régime Markovien
NMRF	Champ aléatoire de Markov
\preceq	Ordre partiel sur \mathbb{Z}^d
\prec	Ordre lexicographique sur \mathbb{Z}^d
$S[a, b]$	Sous-ensemble de tout $x \in \mathbb{Z}^d$ tel que $a \preceq x \preceq b$ (ou $a \prec x \prec b$)
$S \langle a, b \rangle$	Le sous-ensemble $S[a, b] \setminus \{a\}$
$S[a, b \rangle$	Le sous-ensemble $S[a, b] \setminus \{b\}$

Introduction

0.1 Contexte historique

Les modèles autorégressifs $2D AR$ ont été utilisés avec succès dans un large éventail d'applications en économétrie (Holly et al (2010), Fingleton (2008)), biométrie (Shinkareva et al 2006), (Ho et al (2005)), (Macnab et Dean (2001)) et environnement (Fasso and Finazzi (2011), Fosso et al. (2007), Fuentes (2001)). Ils se sont également avérés efficaces dans le traitement du signal et de l'image. Voir (Kashyap et al (1983), Kaufman et Tekalp (1991)) et (Kokaram (2004)) pour leur utilisation dans la restauration d'image, (Mhidra et al (1993)) et (Sarkar et al (1997)) pour leur contribution à l'analyse et à la synthèse d'images texturales (Lu et Xu (1995)) et (Oe (1993)) pour leur développement et leur mise en œuvre dans la classification et la segmentation d'images,(Heitz Maitre (1987)) et (Politis (1994)) où ils se sont avérés efficaces dans la conservation de l'art et (Sharma et al (1998)) et (Hansen et Chellappa (1990)) dans lequel leur utilisation dans l'estimation du spectre est donnée.

0.2 Formulation du problème et motivation

Cependant, les modèles $2D AR$ ne permettent pas de décrire des séries spatiales avec des changements fondamentaux et des ruptures structurelles dans la dynamique des données. Un premier exemple tiré de l'économétrie spatiale est les données régionales sur les prix des logements qui partagent des points de rupture structurels les unes avec les autres, présentant une volatilité et une non-normalité accrues (e.g.,(Holly et al.(2011), Barros et al (2010), Ngene et al (2015), El-Hadidy (2019)). Un deuxième exemple tiré du traitement

d'image est lorsque les niveaux de gris des pixels d'une image richement texturée sont observés, les séries spatiales résultantes présentent une alternance de différentes dynamiques spatiales correspondant à des régimes de texture reflétant un certain changement de régime dans la dynamique régionale des données. Pour prendre en compte les ruptures structurelles survenant dans les données $2D$, nous introduisons dans cette thèse le modèle $2D$ autorégressif à changement de régime Markovien ($2D MS - AR$) qui permet la possibilité de changements soudains dans les valeurs des paramètres d'un processus $AR 2D$. Ce modèle peut être vu comme un modèle de Markov caché $2D$ (HMM) dans lequel l'observation est supposée être gérée par un processus $2D AR$ dont les paramètres dépendent d'un champ aléatoire de Markov latent. Dans la modélisation d'image markovienne, les modèles de Markov cachés $2D$ (HMM) étaient les plus utilisés. $2D HMM$ sont considérés comme une simple extension de $1D HMM$ qui fonctionne sur des données $2D$. Puisqu'il n'y a pas d'ordre naturel sur l'espace, l'extension de la série temporelle $MS - AR$ au cas $2D$ nécessite le choix d'un système spécifique de voisins un moyen simple de réaliser une telle extension consiste à ordonner artificiellement les données $2D$ d'une manière analogue à l'ordre naturel du temps pour un signal unidimensionnel, obligeant les signaux bidimensionnels à être des signaux "causaux" particuliers.

0.3 L'objectif de cette thèse

L'objectif principal de cette thèse est de construire un modèle spatial causal qui pourrait capturer diverses caractéristiques clés des données spatiales, telles qu'une croissance au niveau moyen de la volatilité régionale ou une asymétrie régionale. A cet effet, une

stratégie en deux étapes est proposée premièrement, un champ aléatoire de Markov $2D$ (MRF) est construit où la causalité imposée permet d'établir une analogie entre ce $2D$ MRF et la représentation en chaîne de Markov. Basé sur le $2D$ MRF proposé, le $2D$ $MS - AR$ est défini sous certaines hypothèses nécessaires et notations utiles. Enfin, l'estimation des paramètres du modèle est discutée ouvrant la voie à de larges perspectives d'exploitation des processus $2D$ $MS - AR$ proposés pour modéliser efficacement plusieurs phénomènes présentant une rupture structurelle dans la dépendance spatiale des données.

0.4 Organisation du mémoire

Le manuscrit de cette thèse est composé de quatre chapitres, une introduction générale et une conclusion :

Le premier chapitre est consacré aux concepts utilisés tout au long de cette thèse, ainsi aux quelques définitions et théorèmes préliminaires notamment le théorème central limite. nous considérons dans **Le deuxième chapitre** les champs AR spatiaux quadrantal étudiés par Tjostheim (1978), (1983). Il s'agit de champs $(X_t)_{t \in \mathbb{Z}^d}$ indexés par \mathbb{Z}^d et vérifiant un équation du type

$$X_t = \sum_{j \in S\langle \mathbf{0}, \mathbf{p} \rangle} \phi_j X_{t-j} + \varepsilon_t$$

où l'ensemble $S\langle \mathbf{0}, \mathbf{p} \rangle$ est défini à l'aide de l'ordre partiel usuel sur \mathbb{Z}^d et où $(\varepsilon_t)_{t \in \mathbb{Z}^d}$ est un champ de bruit blanc. Nous nous sommes particulièrement intéressés aux résultats asymptotiques des estimateurs des moindres carrés et de Yule-Walker des coefficients $\phi_p = (\phi_j)_{j \in S\langle \mathbf{0}, \mathbf{p} \rangle}$ d'un modèle $2D$ AR unilatéral obtenu par Tjostheim (1983) à l'étude probabiliste des modèles autorégressifs unilatéraux spatiaux et nous établissons les condi-

tions de stationnarité. **Le troisième chapitre** a pour but de donner une vue globale des champs aléatoires de Markov (Markov Random Fields) MRF discrets. Parce que c'est une généralisation de l'idée de chaînes de Markov unidimensionnelles ($1D MC$) au contexte 2-dimensionnel ou 3-dimensionnel. Ici, nous étudions deux modèles de MRF : Le premier modèle que nous utilisons souvent dans le contexte de phénomènes physiques, à savoir les champs de Markov non causaux ($NMRF$), ces modèles selon la théorie de Hammersley - Clifford sont équivalents aux distributions de Gibbs GRF . Les modèles MRF considérés dans la deuxième section est une extension naturelle des modèles Markov cachés (en anglais Hidden Markov Models) unidimensionnels (HMM) aux modèles MRF causaux 2-dimensionnels ($HMRF$), et le principal obstacle à cette extension est l'absence d'ordre naturel dans un réseau bidimensionnel et donc aucune idée naturelle de causalité. Dans les données spatiales générées par les images. Nous avons donc discuté de la conversion de $HMRF$ en HMM en fonction de la numérisation. La fin de ce chapitre, nous discuterons de l'estimation des paramètres du modèle $HMRF$. Dans **le quatrième chapitre**, nous présentons les modèles autorégressifs à changements de régime Markovien (Markov-Switching AutoRegressive $MS - AR$). Cette famille de modèles, initialement proposée par Hamilton (1989) pour décrire des séries chronologiques économétriques. Dans la suite de ce chapitre, on introduit une nouvelle classe de processus qui est une combinaison de modèles d'autorégressifs spatiaux $2D AR$ et de modèles causaux de champ aléatoire de Markov MRF appelé "processus $2D$ autorégressif à changement de régime Markovien unilatéral" pour les données spatiales bidimensionnelles non linéaires (Djafri et Kharfouchi (2020)). Le modèle $2D MS - AR$ est défini sous certaines hypothèses nécessaires et notations utiles. nous étudions

La stationnarité stricte et l'estimation des paramètres via l'algorithme *VEM* (Variational Expectation Maximization) est effectuée.

0.5 Travaux réalisés

Publications :

- Djafri. H et Kharfouchi. S (2021) ."On unilateral 2D Markov-Switching Autoregressive (*MS – AR*) model". Int. J. Mathematics in Operational Research. Vol. 18 n 4, No.10.1504/IJMOR.10033953

Conférences internationales :

- Djafri. H. "Modèles Autorégressifs spatiaux : Estimateurs et Applications." ICAAM, Hammamet, December 16-19,2019.

Conférences nationales :

- Djafri. H. "Aperçu sur les champs de Markov : théorie et application." CMA, Boumerdes, 2018.

Chapitre 1

Généralités

Les nombreuses années de recherche visant l'applicabilité de la théorie des processus stochastiques en une dimension à l'étude des séries temporelles ont finalement débouché sur l'implantation d'une procédure méthodologique qui on le sait, est devenue extrêmement populaire. Ce meme courant de recherche a poussé des auteurs comme Wittle(1954) à développer la théorie des processus stochastiques multidimensionnels. Lorsque limitée à 2 dimensions, cette théorie s'appelle aussi la théorie des champs aléatoires.

Le but de la statistique spatiale est d'étudier des phénomènes (températures, étude d'une population, étude des villes,...) sur un ensemble spatial $S \subset \mathbb{R}^d, d \geq 2$. Il y a une dépendance en espace de ces données. On va donc poser X , un champ aléatoire (une famille de variables aléatoires) sur S avec $X = \{X_s, s \in S\}$ composé de variables indexé par S . Un champ aléatoire est une famille de variables aléatoires sur un domaine fini $S \subset \mathbb{R}^d, d \geq 2$. On considère alors les données spatiales comme des réalisations de champs aléatoires.

Dans ce chapitre on s'intéresse à la présentation des concepts méthodologiques qui ont été développés pour permettre l'étude des champs aléatoires, ce dernier étant l'outil de probabilité de modélisation des données spatiales.

1.1 Types de données spatiales

En statistique spatiale, on associe aux observations leurs coordonnées spatiales, et ce sont celles-ci qui vont intervenir dans la modélisation statistique. Il existe trois grandes familles de données spatiales : les données ponctuelles, les données surfaciques et les données continues. Ils diffèrent d'abord par la nature de l'unité géographique attachée à chaque unité statistique, soit un lieu précis soit une surface, mais aussi par la qualité aléatoire ou non des positions spatiales.

1.1.1 Données ponctuelles

Les données spatiales ponctuelles se caractérisent par la distribution dans l'espace des observations. Le processus générateur des données génère les coordonnées géographiques associées à l'apparition d'une observation. On n'étudie pas de valeur associée à l'observation, seule compte la localisation. Il s'agit, par exemple, du lieu d'apparition d'une maladie lors d'une épidémie, ou de la répartition dans l'espace de certaines espèces d'arbres. L'analyse spatiale des données ponctuelles a pour objectif de quantifier l'écart entre la distribution spatiale des observations et une distribution complètement aléatoire dans l'espace. Si les données sont plus regroupées que si elles étaient distribuées aléatoirement sur le territoire,

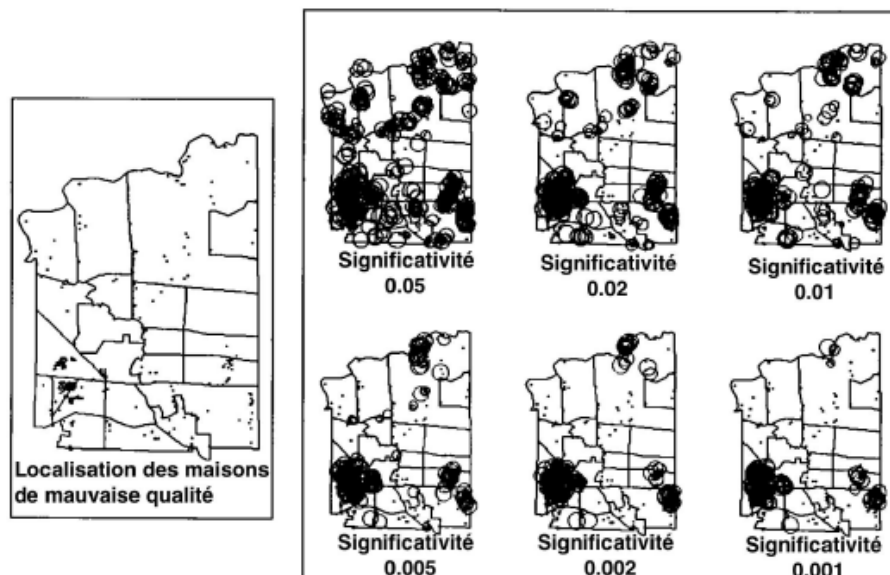


FIG. 1.1: Détection de clusters significatifs

on peut identifier des clusters et mesurer leur significativité.

Exemple : *Détection de clusters.*

Fotheringham et al. (1996) cherchent à détecter la présence de clusters significatifs de maisons inconfortables. Ils comparent la répartition spatiale de ces maisons avec la répartition qu'elles auraient si elles étaient distribuées aléatoirement parmi l'ensemble de toutes les maisons. Les hypothèses sur la distribution aléatoire dans l'espace permettent d'évaluer la significativité des regroupements de maisons

1.1.2 Données surfaciques

Pour des données surfaciques, la localisation des observations est considérée comme fixe, mais les valeurs associées sont générées suivant un processus aléatoire. Ces données

caractérisent le plus souvent une partition du territoire en zones contiguës, mais elles peuvent également être des points fixes du territoire. Il s'agit, par exemple, du PIB par région, ou du nombre de mariages par mairie. Le terme "surfacique" est donc trompeur, car ces données ne sont pas nécessairement représentées sur une surface. On s'intéresse aux relations entre les valeurs des observations voisines. L'analyse spatiale des données surfaciques commence par définir la structure de voisinage des observations puis elle quantifie l'influence qu'exercent les observations sur leurs voisines, et enfin, elle évalue la significativité de cette influence.

Exemple 1.2 : *Dépendance spatiale locale.*

GIVORD et al. (2016) cherchent à répondre à la question : "Les collègues favorisés sont-ils toujours situés dans un environnement favorisé?". Les auteurs utilisent pour cela des Indices locaux d'autocorrélation spatiale. Ces indices comparent la similarité entre le niveau social d'un collègue et celui de son environnement à la similarité qu'ils auraient si les niveaux sociaux des collègues étaient répartis aléatoirement parmi l'ensemble des collègues. Les indices locaux d'autocorrélation spatiale permettent d'identifier les collègues pour lesquels l'influence du milieu social environnant est significative. (figure 1.2).

1.1.3 Données continues

En présence de données continues, il existe une valeur pour la variable d'intérêt en tout point du territoire étudié. Les données sont générées de façon continue sur un sous ensemble de \mathbb{R}^2 . En revanche, ces données sont mesurées uniquement en un nombre discret de points. Il s'agit, par exemple, de la composition chimique du sol (utile à l'industrie minière), de la qualité de l'eau ou de l'air (pour des études sur la pollution), ou encore de

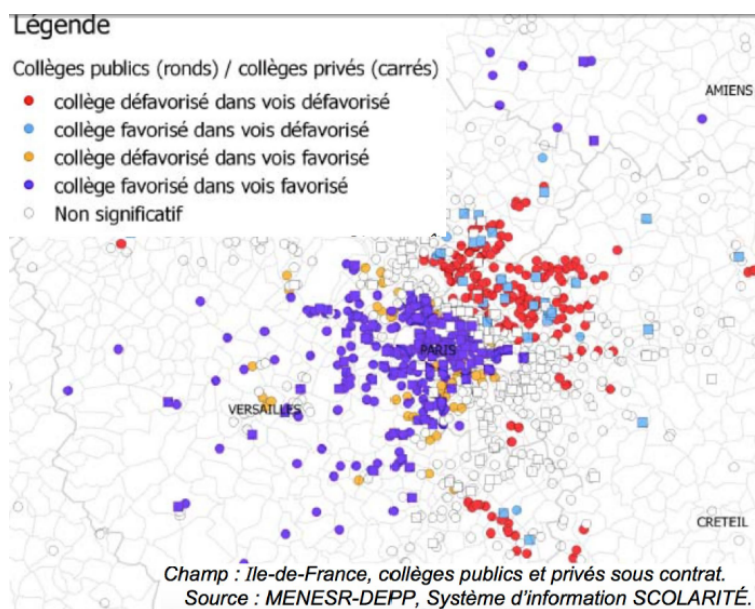


FIG. 1.2: collège lui-même

diverses variables météorologiques. L'analyse spatiale des données continues, appelée aussi géostatistique, cherche à prédire la valeur d'une variable en un point où elle n'a pas été échantillonnée, ainsi que la fiabilité de cette prédiction. La géostatistique aide également à optimiser le plan d'échantillonnage des données.

Exemple 1.2 : *Prédiction de la pollution.* Chiles et al. 2005)

Les chercheurs du groupe de travail **GeoSiPol** (Les pratiques de la géostatistique dans le domaine des sites et sols pollués) prennent en compte la structure de dépendance spatiale entre les données grâce à la technique du krigeage. Ils prédisent la quantité de polluant en des lieux où le sol n'a pas été échantillonné et quantifient l'incertitude d'estimation.

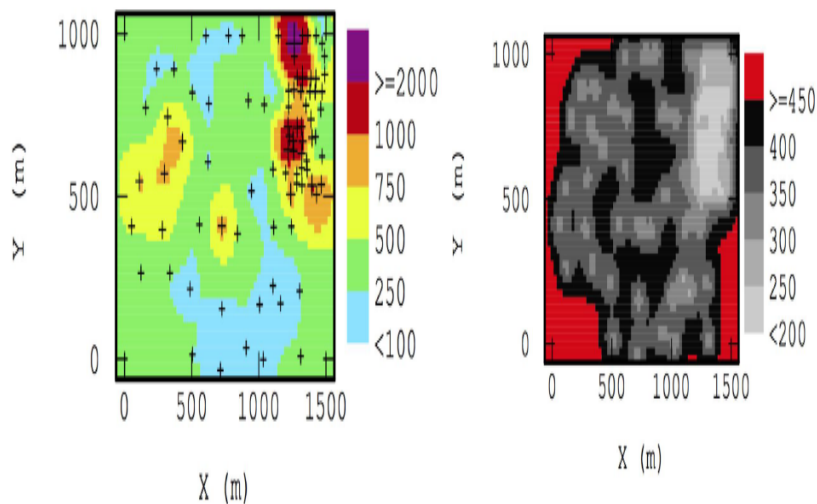


FIG. 1.3: la prédiction (à droite)

1.2 La dépendance spatiale

Généralement, la plupart des données sont des observations d'une variable mesurée en des localisations différentes réparties dans l'espace. C'est notamment le cas de la pluviométrie qui est mesurée dans chaque site d'une région ou d'une province d'un pays. Souvent, il est admis que ces données spatiales observées en coupe transversale sont indépendantes et les données sont alors traitées avec les méthodes statistiques standards. Cependant, il est important de justifier et de tester cette hypothèse d'indépendance des données spatiales puisqu'un traitement statistique de telles données qui ignore leur aspect spatial ou ne l'intègre pas de façon adéquate peut résulter en une perte d'information, des erreurs de spécifications et des estimations non convergentes et non efficaces. En effet, sur ce dernier point, il est essentiel de noter que l'analyse statistique des données spatiales n'est pas une juxtaposition de l'analyse géographique et de l'analyse statistique mais une inter-

action entre les deux. L'introduction de l'espace dans les modèles économétriques n'est ni neutre, ni immédiate, et les techniques de la statistique spatiale visent à prendre en compte la présence de deux effets spatiaux importants : l'autocorrélation spatiale qui se réfère à l'absence d'indépendance entre observations géographiques, et l'hétérogénéité spatiale qui est liée à la différenciation des variables et des comportements dans l'espace (Julie Le Gallo, 2000) .

1.2.1 L'autocorrélation spatiale

L'autocorrélation spatiale indique une relation fonctionnelle entre ce qui se passe entre une unité spatiale et ses voisines. Elle fournit une information supplémentaire par rapport aux statistiques traditionnelles, invariants par rapport à la configuration spatiale des données. L'autocorrélation spatiale est différente de l'autocorrélation temporelle dans le sens qu'elle est multidimensionnelle. Par conséquent, les techniques valables pour les séries temporelles ne sont pas directement transposables au cas spatial.

Les sources de l'autocorrélation spatiale sont multiples : interactions spatiales, contagion, externalités spatiales, effets de débordement géographique (spillover), effets d'imitation ou encore une mauvaise spécification du modèle économétrique causée par l'omission de variables elles-mêmes spatialement autocorrélées. De manière transposée, les sources d'absence d'autocorrélation sont entre autres une répartition spatiale des valeurs de la variable aléatoire, une invariance par permutation aléatoire des individus (aléas spatial), des valeurs observées en une localisation qui ne dépendent pas des valeurs observées en des localisations voisines, une égalité de degré de vraisemblance entre le schéma spatial observé et tout autre schéma spatial et une modification de la localisation des individus sans consé-

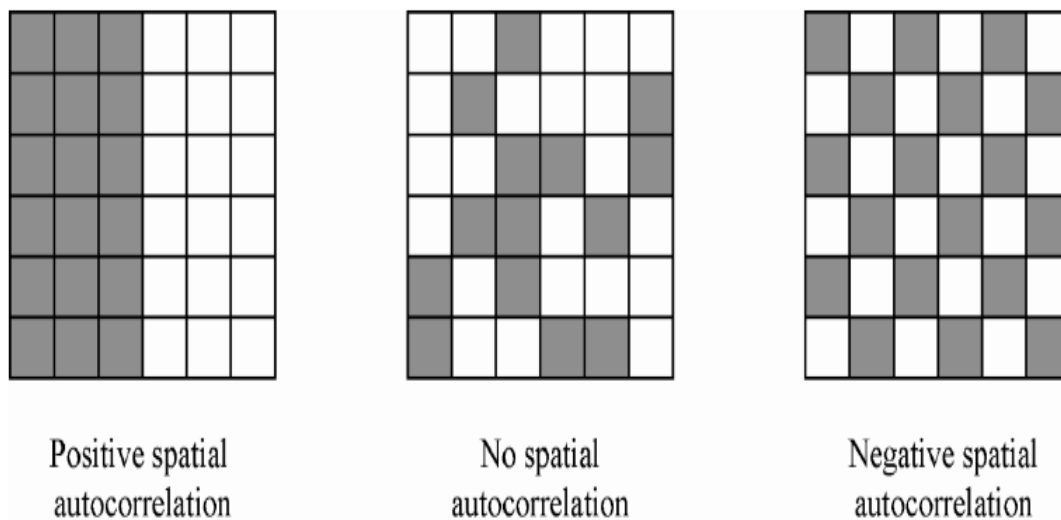


FIG. 1.4: Les différents types d'autocorrélation

quence sur le contenu informationnel des données. Il existe deux types d'autocorrélation : l'autocorrélation positive et l'autocorrélation négative. L'autocorrélation positive apparaît lorsque des valeurs similaires tendent à se concentrer dans l'espace. Elle est présente dans les processus de diffusion et contagion bien qu'il faille signaler que la diffusion tend à produire une autocorrélation spatiale positive mais que la réciproque n'est pas nécessairement vraie. Quant à l'autocorrélation négative, elle intervient lorsque les valeurs sur lesquelles porte la concentration sont dissimilaires. Enfin il y'a absence d'autocorrélation (autocorrélation spatiale nulle) lorsque deux lieux proches se ressemblent ni plus ni moins (en moyenne) que deux lieux éloignés.

1.2.2 L'hétérogénéité spatiale

L'hétérogénéité spatiale est liée à la différenciation des comportements dans l'espace. Il s'agit de l'instabilité structurelle due à une instabilité spatiale des coefficients de la régression et à la non-linéarité des formes fonctionnelles ou de l'hétéroscédasticité provenant de variables omises ou de toute autre forme de mauvaise spécification.

Ce phénomène se retrouve à plusieurs échelles : les comportements et les phénomènes économiques ne sont pas les mêmes dans le centre d'une ville et dans sa périphérie ou dans une région urbaine et dans une région rurale (Le Gallo, 2000). L'instabilité structurelle provient de l'absence de stabilité dans l'espace des comportements ou d'autres relations étudiées : les formes fonctionnelles et les paramètres varient selon leurs localisations et ne sont donc pas homogènes. Il est donc nécessaire de mobiliser des modélisations prenant en compte les caractéristiques particulières de chaque localisation de l'échantillon. S'agissant de l'hétérogénéité spatiale, dans les modèles économétriques, elle peut venir de variables manquantes ou de toute autre forme de mauvaise spécification. Par exemple, les unités spatiales elles-mêmes ne sont généralement ni de formes régulières, ni homogènes : des régions peuvent avoir des formes et des aires différentes, des niveaux de développement technologique variables, des populations plus ou moins importantes, etc.

Enfin, le traitement de l'hétérogénéité spatiale est effectué en prenant en compte l'usage de variables explicatives pour modéliser la tendance. Certaines de ces variables peuvent être spatiales de nature comme, par exemple, la distance. Cependant, il faut noter qu'il n'est pas suffisant de prendre en compte ces variables dans la moyenne pour évacuer totalement la structure spatiale du problème qui peut rester présente.

Exemple 1.3 LEVRATTO (2020)

Cet article s'intéresse à la répartition spatiale du nombre de cas de Covid-19 nécessitant une hospitalisation et du nombre de décès liés à ce virus recensés dans les départements français. Les données utilisées proviennent de l'INSEE et du ministère de la santé. Les premières analyses à ce sujet ont mis en évidence une forte hétérogénéité spatiale dans la répartition du nombre d'individus concernés entre les différents départements de France métropolitaine.

L'objectif de cette recherche est donc de contribuer à la compréhension des vecteurs de propagation de l'épidémie et d'en souligner la dimension socio-économique. Elle s'inscrit dans le champ de la géographie de la santé qui a émergé dans les années 1970 et a connu un profond renouvellement dans les années 1990 sous l'influence de travaux de recherche nombreux et variés des domaines de l'épidémiologie, des sciences de l'environnement, des sciences sociales, de la santé publique ou du management des services dans le but de rendre compte des disparités spatiales en matière d'accès aux soins, d'exposition aux maladies et de morbidité (Rosenberg, 1998 ; Curtis, 2004 ; Prior et al. 2018). Cette recherche complète aussi celles entreprises par les épidémiologistes qui pointent les facteurs individuels tels que l'âge, l'indice de masse corporelle ou l'état de santé des personnes touchées par la Covid-19 (Onder, Resa et Brusaferrò, 2020 ; Wu et McGoogan, 2020 ; Diez et Santos-Burgoa, 2020) comme le recense la base de données Worldomete

Les cartes reportées sur les figures 1.5, 1.6 et 1.7 rendent compte de l'hétérogénéité spatiale, très tôt décelée par le secteur médical, de la maladie. Les départements les plus touchés, quel que soit l'indicateur retenu, sont ceux de la région Grand Est, des Hauts-de-

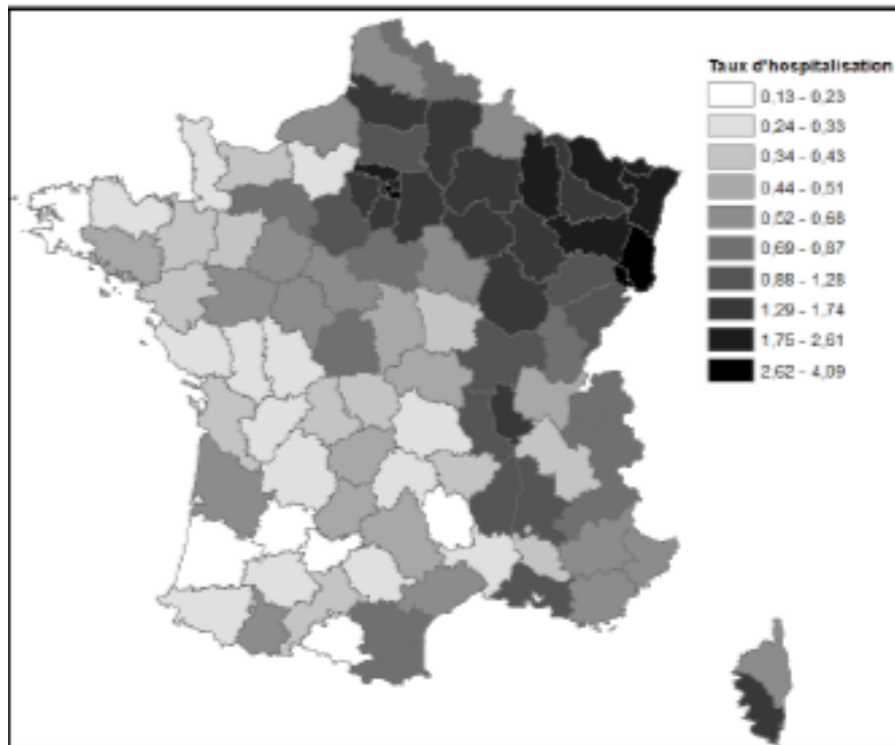


FIG. 1.5: Taux d'hospitalisation pour la Covid-19 par département (par décile)

France et de l'Ile-de-France

1.3 Relations d'ordre- Ensembles indexants

Sur \mathbb{Z}^d , nous définissons, dans le paragraphe suivant, l'ordre partiel usuel et l'ordre lexicographique. Ces notions d'ordre vont nous permettre de préciser la notion de passé et de présent considérée dans la définition de champs $2D$ $MS - AR$ unilatéral proposé dans le chapitre 04, les champs AR spatiales. Ainsi, nous pourrons définir certains ensembles utiles à la description de modèle $MS - AR$ spatiaux.

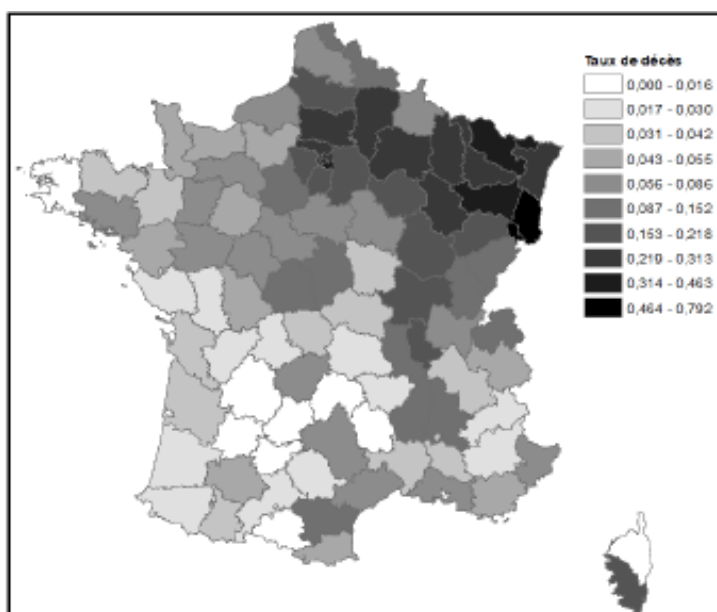


FIG. 1.6: Taux de décès dus à la Covid-19 en milieu hospitalier par département (par décile)

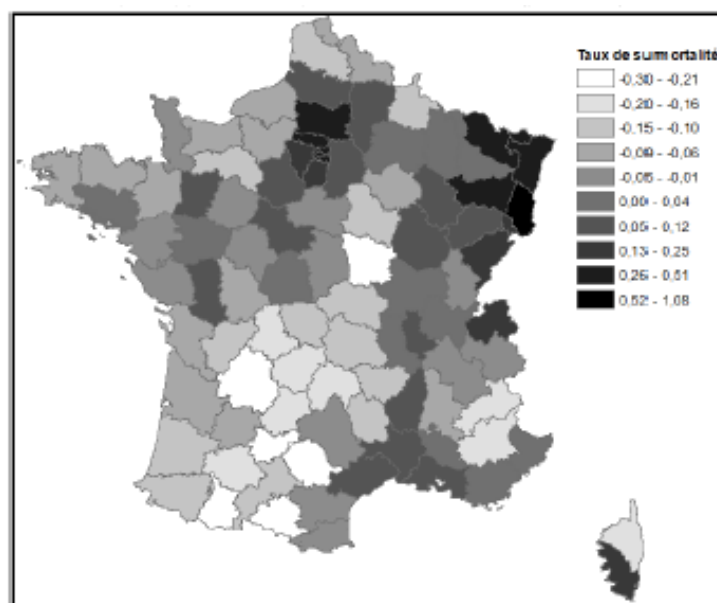


FIG. 1.7: Taux de surmortalité par département au 30 mars 2020 par rapport à la moyenne de 2018 et 2019 (par décile)

1.3.1 Ordre total et ordre partiel

Sur \mathbb{Z} , la notion d'ordre est claire. En revanche lorsque $d \geq 2$, plusieurs ordres peuvent être définis sur \mathbb{Z}^d . Nous donnons ici les deux ordres que nous utiliserons par la suite.

Définition 1 *L'ordre partiel usuel* (\preceq ou \prec) sur \mathbb{Z}^d est défini pour deux points $s = (s_1, s_2, \dots, s_d)$ et $t = (t_1, \dots, t_d)$ de \mathbb{Z}^d par

$$s \preceq t \text{ (respectively } s \prec t)$$

si pour tout $i = 1, \dots, d$

$$s_i \leq t_i \text{ (respectively } s_i < t_i)$$

Cet ordre est appelé **l'ordre des quarts de plan**.

Définition 2 *L'ordre lexicographique* (\ll) sur \mathbb{Z}^d est défini pour deux points $s = (s_1, s')$ et $t = (t_1, t')$ de \mathbb{Z}^d tels que s_1 et t_1 appartiennent à \mathbb{Z} par

$$s \ll t$$

si et seulement si $s_1 < t_1$ ou $s_1 = t_1$ et $s' \ll t'$ dans \mathbb{Z}^{d-1} .

Cet ordre est appelé l'ordre du demi-plan ou **ordre total**.

Définition 3 Soient $s = (k; l), t = (m; n)$, deux points dans \mathbb{Z}^2 . On note par $s \ll t$ l'ordre lexicographique défini par

$$s \ll t \Leftrightarrow (k < m) \text{ ou } (k = m, l \leq n) \quad (1.1)$$

Ce qui est considéré comme un cas particulier de l'ordre total.

1.3.2 Ensembles indexants

Pour deux éléments a, b de \mathbb{Z}^d , tels que $a \ll b$ et $a \neq b$ (\ll étant l'un des ordres définis dans la section précédente), on considère maintenant les sous-ensembles suivants de

$$\begin{aligned} S[a, b] &= \{x \in \mathbb{Z}^d : a \ll x \ll b\} \\ S[a, \infty] &= \{x \in \mathbb{Z}^d : a \ll x\} \\ S[\infty, a] &= \{x \in \mathbb{Z}^d : x \ll a\} \\ S\langle a, b \rangle &= S[a, b] \setminus \{a\} \\ S[a, b \rangle &= S[a, b] \setminus \{b\} \\ S\langle a, \infty \rangle &= S[a, \infty] \setminus \{a\} \end{aligned}$$

$S[a, b]$ permet de définir une notion de fenêtre spatiale qui étend la notion classique d'intervalle de temps utilisée dans la trame de temps, $S[a, \infty]$ généralise la notion de futur à un cadre spatial, enfin $S[\infty, a]$ permet de donner un équivalent à la notion de passé d'un temps d'observation lorsqu'il s'agit d'un site observé a . Enfin, pour $N \in \mathbb{N}$, nous notons \mathbf{N} l'élément de \mathbb{Z}^d dont toutes les composantes sont égales à N . Par la suite, nous choisirons

souvent $a = \mathbf{0}$ ou bien $a = \mathbf{1}$ en considérant les ensembles $S \langle \mathbf{0}, p \rangle$, $S \langle \mathbf{0}, q \rangle$ (pour définir des processus *ARMA* spatiaux) et $S[\mathbf{1}, \mathbf{N}]$ (pour N observations d'un champ spatial). Même lorsque $d > 2$ ou bien lorsque $b = \infty$ nous continuerons à appeler ces ensembles des "rectangles de \mathbb{Z}^d . "

Remarque 4 Lorsque a est fixé, on constate qu'il existe une bijection naturelle entre l'ensemble des points $b \succeq a$ de \mathbb{Z}^d et l'ensemble des rectangles $S[a, b]$ inclus dans $S[a, \infty]$.

Ceci, nous conduira alors à introduire, dans la section suivante , nous discutons du concept d'ordre et de voisinage dans \mathbb{Z}^2 .

1.3.3 Voisinage dans \mathbb{Z}^2

Dans les séries temporelles, il existe une structure de voisin naturel induite par l'ordre total existant de \mathbb{Z} (l'ensemble de toutes les valeurs passées de $t \in \mathbb{Z}$; est l'ensemble de tous les entiers inférieurs à t). Cependant, pour les points du plan, par exemple $(i, j) \in \mathbb{Z}^2$, il existe plusieurs notions de voisinage. Dans cette section, nous nous intéressons à l'identification des voisinages causales.

Notion de causalite (Whittle, 1954).

La notion de causalite est basee sur la definition du passe, du present et du futur. Pour tout point (i, j) on definit le passe relativement au point (i, j) , pour l'**ordre lexicographique** l'ensemble des points

$$\{(k, l) / k = i, l \prec j; k \prec i, -\infty \leq l \leq \infty\}$$

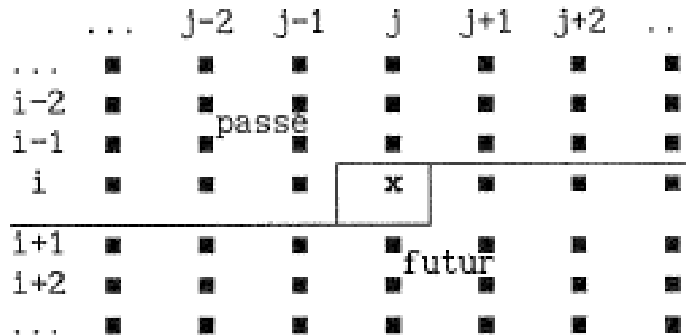


FIG. 1.8: définition du passé, du present et du futuro

et le futur l'ensemble de points

$$\{(k, l) / k = i, l \succ j; k \succ i, -\infty \leq l \leq \infty\}$$

ceci est illustré par la figure 1.8

1.3.4 Les régions causales dans \mathbb{Z}^2

Cependant, pour les points sur le plan, par exemple $(m, n) \in \mathbb{Z}^2$, il existe plusieurs notions différentes de voisinage. En général, les définitions du voisinage d'un point (m, n) sur le plan sont motivées par le système d'acquisition physique des données. C'est le cas des images qui ont été capturées par des satellites. Nous décrivons ici les définitions les plus fréquentes du quartier dans la littérature (voir Jain (1989)).

Définition 5 Pour tous $(m, n) \in \mathbb{Z}^2$:

(a) $S_1(m, n) = \{(k, l) \in \mathbb{Z}^2 : (k, l) \neq (m, n)\}$ est appelée région de prédiction non causale en (m, n) .

(b) $S_2(m, n) = \{(k, l) \in \mathbb{Z}^2 : k \prec m\} \cup \{(m, l) \in \mathbb{Z}^2 : l \neq n\}$ est appelé région de prédiction semi-causale en (m, n) .

(c) $S_3(m, n) = \{(k, l) \in \mathbb{Z}^2 : k \prec m\} \cup \{(m, l) \in \mathbb{Z}^2 : l \succ n\}$ est appelé région de prédiction causale à (m, n) (alternativement appelé demi-plan non symétrique (NSHP))

(d) $S_4(m, n) = \{(k, l) \in \mathbb{Z}^2 : k \leq m, l \leq n\} - \{(m, n)\}$ est appelée région fortement causale en (m, n) .

En général, une réalisation d'un processus stochastique défini sur le plan est observée sur l'une des régions de prédiction décrites ci-dessus. Cela motive la définition suivante.

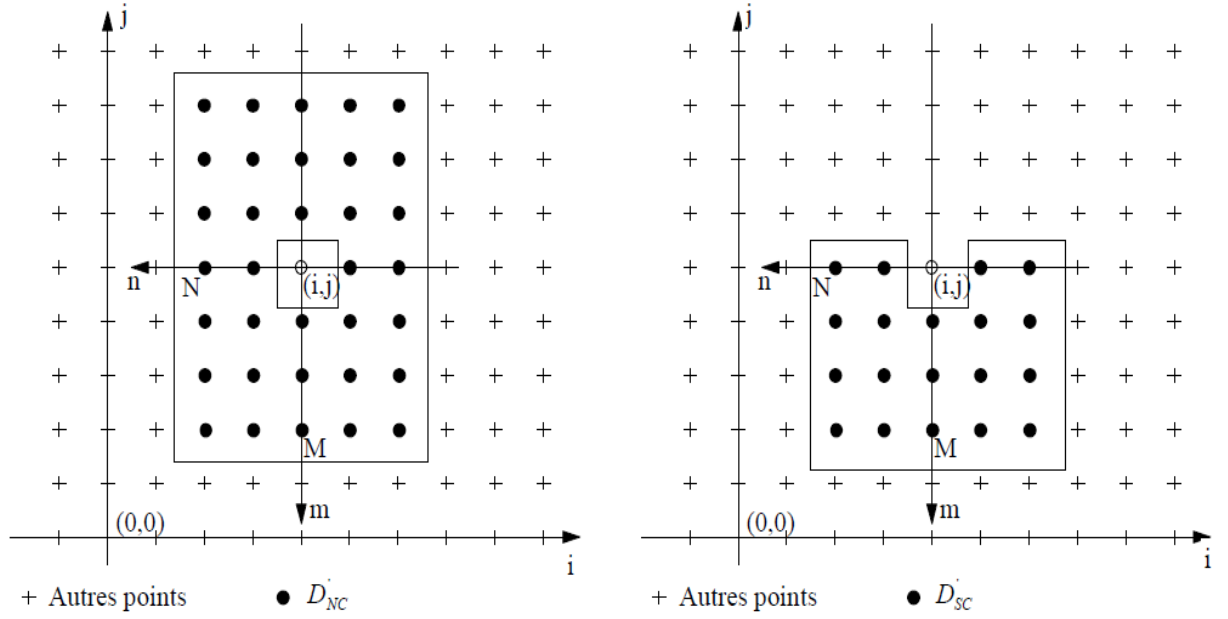
Définition 6 *Pour tous $(m, n) \in \mathbb{Z}^2$, et Pour tous $M \in \mathbb{N}$.*

(a) $W_{1,M}(m, n) = \{(k, l) \in S_1(m, n) : |k - m| \leq M, |l - n| \leq M\}$ est appelé non causal fenêtre de prédiction d'ordre M en (m, n) .

(b) $W_{2,M}(m, n) = \{(k, l) \in S_2(m, n) : m - M \leq k \prec m, |l - n| \leq M\} \cup \{(m, l) : 0 \prec |l - n| \leq M\}$ est appelée fenêtre de prédiction semi-causale d'ordre M en (m, n) .

(c) $W_{3,M}(m, n) = \{(k, l) \in S_3(m, n) : m - M \leq k \prec m, |l - n| \leq M\} \cup \{(m, l) : n \prec l \leq n + M\}$ est appelée fenêtre de prédiction causale d'ordre M en (m, n) .

(d) $W_{4,M}(m, n) = \{(k, l) \in S_4(m, n) : m - M \leq k \prec m, n - M \prec l \leq n\}$ est appelée fenêtre de prédiction fortement causale d'ordre M en (m, n) .



a. Support non-causal

b. Support semi-causal

1.4 Notions de base

Dans cette section, nous introduisons les définitions et notations que nous utiliserons tout au long de cette thèse et que nous avons principalement tirées des livres de Guyon (1995) et Geaton et Guyon (2008) et thèse doctort de Kharfouchi S (2011) .

1.4.1 Champs aléatoires

Définitions et notations

Dans la cadre de notre étude, nous noterons $S = \{s, s \in \mathbb{R}^d\}$ un sous espace euclidien de dimension fini d . Cet ensemble S , représente l'ensemble spatial des sites sur lequel est définie une famille de variables aléatoires de nombres réels $X = \{X_s, s \in S\}$. Supposons que $(\Omega, \mathcal{F}, \mathcal{P})$ un espace probabilisé et $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ un espace mesurable, avec $\mathcal{B}_{\mathbb{R}}$ la

tribu borélienne de \mathbb{R} .

Définition 7 Soit $X = \{X_s, s \in S\}$ un processus aléatoire indexé par un ensemble spatial S et à valeurs dans un espace d'états Λ .

La localisation d'un site $s \in S$ peut être soit fixée ou soit aléatoire en fonction du modèle de données. L'ensemble spatial S peut être bi-dimensionnel ($S \subseteq \mathbb{R}^2$), tri-dimensionnel ($S \subseteq \mathbb{R}^3$),

spatio-temporel (indexation d'une observation par $(s; t) \in \mathbb{R}^2 \times \mathbb{R}^+$). Par ailleurs, l'espace d'états Λ peut être de nature :

- $\Lambda \subseteq \mathbb{R}^2$ (champ gaussien)
- $\Lambda \subseteq \mathbb{R}^2$ (champ exponentiel ou Gamma)
- $\Lambda \subseteq \mathbb{R}^2$ (champ poissonien)
- $\Lambda = \{a_0, a_1, \dots, a_k\}$ (champ catégoriel)
- $\Lambda = \{0, 1\}$ (Champ binaire)

En pratique, on observe n données spatiales $\{x_{s_1}, x_{s_2}, \dots, x_{s_n}\}$, qui constituent une réalisation de X et dont les coordonnées spatiales sont $\{s_1, \dots, s_n\}$, appelées également index spatiaux. En fait, chaque site s est déterminé par ses coordonnées géographique suivant l'espace d'étude. La nature des ensembles Λ et S détermine le type de données spatiales.

Dans la suite, on considère que S est un ensemble fini de $n \geq 1$ sites. La loi de probabilité du champ aléatoire X est appelé loi spatiale. Donc la loi spatiale de X est entièrement déterminée au cas où la loi de chacune des variables $(X_{s_i})_{1 \leq i \leq n}$ est connue.

Définition 8 Un champ aléatoire est totalement caractérisé par la distribution conjointe de tout sous-ensemble fini $(X_{s_1}, X_{s_2}, \dots, X_{s_n})$ de manière cohérente, nécessitant

$$F_{s_1, \dots, s_n}(A_1 \times \dots \times A_n) = P((X_{s_1}, X_{s_2}, \dots, X_{s_n}) \in (A_1 \times \dots \times A_n))$$

où A_1, \dots, A_n sont des boréliens de \mathbb{R} une telle spécification est appelée la distribution du processus. Un champ aléatoire discret X est un processus prenant des valeurs aléatoires sur le réseau \mathbb{Z}^d .

Définition 9 Soit $(X_t)_{t \in \mathbb{R}^d}$ un champ aléatoire à valeurs réelle de carré intégrable. Sa fonction d'autocovariance γ est définie par

$$\gamma(u, v) = E(X_u - E(X_u))(X_v - E(X_v))$$

pour u, v dans \mathbb{R}^d .

1.4.2 Stationnarité

Dans les séries chronologiques, l'inférence des paramètres se fait par répétition indépendante des données. Pour les séries spatiales, en pratique, il faut inférer les propriétés de la dernière à une seule réalisation. Par exemple un épisode de pollution par l'ozone, une région agricole particulière, une épidémie végétale ... Pour parvenir à l'inférence statistique pour un événement unique, il faut jouer sur la répétition de certaines caractéristiques d'un point à un autre de l'espace, au lieu de s'appuyer sur des répétitions indépendantes d'un instant à l'autre. Ensuite, nous posons les hypothèses de stationnarité et d'ergodicité suivantes.

Stationnarité stricte

Un champ aléatoire est strictement stationnaire si toutes les lois de dimension finie sont invariantes pour toute translation c'est-à-dire $(X_{s_1}, X_{s_2}, \dots, X_{s_n})$ et $(X_{s_1+h}, X_{s_2+h}, \dots, X_{s_n+h})$ ont la même loi pour tous les n -uplets s_1, \dots, s_n et $h \in S$

Cette hypothèse difficilement vérifiable, n'est utilisée que dans certains cas rencontrés en pratique comme les champs intégrables non carrés. Il est rarement utilisé ailleurs car il a un défaut majeur : il ne dit rien sur l'existence des moments, tandis que l'espérance et la variance sont des quantités nécessaires à l'analyse statistique des données, et d'autre part il faut vérifier n supérieur au nombre des données disponibles. Nous préférons l'autre hypothèse de stationnarité du second ordre qui est plus faible.

La stationnarité de second ordre

Un champ aléatoire carré intégrable, X_s sur \mathbb{R}^d est stationnaire du second ordre si les deux premiers moments existent et sont invariants par translation :

$$E(X_s) = m \quad \forall s \in \mathbb{R}^d$$

$$Cov(X_u, X_v) = \gamma(u, v) = Cov(X_{u+\tau}, X_{v+\tau})$$

pour chaque u, v et $\tau \in \mathbb{R}^d$.

Ains $\gamma(u, v) = \gamma(u - v, 0)$ pour tout u, v dans \mathbb{R}^d pour un champ aléatoire stationnaire faible, il sera commode de redéfinir la fonction d'autocovariance en fonction d'un argument comme suit,

$$\gamma(h) = \gamma(h, 0)$$

Si X est strictement stationnaire et si $X \in L^2$, alors X est stationnaire du second ordre. L'inverse n'est pas vrai en général mais les deux concepts coïncident si X est gaussien. La fonction de covariance stationnaire $\gamma(\cdot)$ est nommée aussi **covariogramme**.

Remarque 10 *Les fonctions de covariance stationnaire vérifient les propriétés suivantes :*

1. Par la positivité nous avons $\gamma(0) \geq 0$.
2. Le covariogramme est paire : $\gamma(h) = \gamma(-h)$
3. L'inégalité de Cauchy-Schwartz implique que $|\gamma(h)| \leq \gamma(0)$

Champs aléatoires gaussiens

Définition 11 *Un processus stochastique est dit gaussien si toutes ses lois fini-dimensionnelles $\mathcal{L}(X_{s_1}, \dots, X_{s_n})$ sont gaussiennes.*

Autrement dit $(X_s)_{s \in S}$ est gaussien si et seulement si toute combinaison linéaire $a_1 X_{s_1} + \dots + a_n X_{s_n}$ suit une loi normale, pour tous $n \in \mathbb{N}^*$, $s_1, \dots, s_n \in S$ et $a_1, \dots, a_n \in \mathbb{R}$.

La loi d'un vecteur gaussien $(X_{s_1}, \dots, X_{s_n})$ est caractérisée par le vecteur des moyennes $(E[X_{s_1}], \dots, E[X_{s_n}])$ et la matrice de variance-covariance $(Cov(X_{s_i}, X_{s_j}))_{1 \leq i, j \leq n}$

Exemple 12 *Bruit Blanc Fort (BBF), BB faible (BBf), BB gaussien (BBG)*

X est un Bruit Blanc Fort (BBF) si les v.a. $\{X_s, s \in S\}$ sont centrées, indépendantes et identiquement distribuées (*i.i.d.*).

X est un Bruit Blanc faible (BBf) si les variables $\{X_s, s \in S\}$ sont centrées, $Var(X_s) = \sigma^2 < \infty$ décorrélées : $\forall u \neq v, Cov(X_u, X_v) = 0$.

X est un Bruit Blanc Gaussien (*BBG*) si les v.a $\{X_s, s \in S\}$ sont i.i.d. gaussiennes centrées.

Alors : Un *BBF* et un *BBG* sur S sont strictement stationnaire, un *BBf* sur S est stationnaire

1.4.3 Ergodicité

De plus, nous considérerons les processus stationnaires comme ergodiques, c'est-à-dire que leurs moyennes spatiales convergent vers les espérances mathématiques. On parle, alors, de fonctions aléatoires stationnaires d'ordre 2, pour lesquelles l'espérance et la covariance existent et sont stationnaires.

L'ergodicité est une propriété qui renforce la notion de stationnarité, qui surmonte la restriction d'une réalisation unique.

Définition 13 *Un champ aléatoire stationnaire X_s sur \mathbb{R}^d est dit ergodique pour la moyenne si la moyenne spatiale sur un domaine D_n (D_n est une suite croissante de convexes bornées telle que $d(D_n) \rightarrow \infty$) converge vers son espérance mathématique lorsque $n \rightarrow \infty$. C'est-à-dire, $\bar{X}_n = \frac{1}{|D_n|} \int_{D_n} X_s ds \rightarrow E(X_0) = \mu$ lorsque $n \rightarrow \infty$*

$$\bar{X}_n = \frac{1}{|D_n|} \int_{D_n} X_s ds \rightarrow E(X_0) = \mu \text{ lorsque } n \rightarrow \infty$$

où $|D_n|$ désigne le volume de D (sa mesure Lebesgue) et où la convergence est la convergence quadratique moyenne.

Exemple 14 $X = \{X_s, s \in \mathbb{Z}^d\}$, X_s i.i.d est un exemple de processus ergodique.

Après avoir introduit les notions de champs aléatoires, de stationnarité et d'ergodicité dans le cas le plus général (champs aléatoires indexés par \mathbb{R}^d), revenons au contexte de ce travail doctoral qui est le sont des champs aléatoires indexés par \mathbb{Z}^d . Sur \mathbb{Z}^d , on définit dans le paragraphe suivant, l'ordre partiel habituel et l'ordre lexicographique. Ces notions d'ordre vont nous permettre, même artificiellement, de structurer l'espace selon un certain ordre qui jouera le rôle du passé et du présent ou plus précisément de préséance et de succession des observations.

1.5 Exemples des modèles spatiaux

Les modèles spatiaux présentent des analogies avec les modèles utilisés en séries temporelles. Dans ces derniers, le temps t de $X(t)$, est une information à part entière. Il permet d'étudier la succession des observations, leur périodicité, et d'en construire des modélisations. Il en va de même pour les observations spatiales. La situation dans l'espace est une information aussi importante que la variable d'intérêt, et le modèle spatial X_s n'est pas la même chose qu'un modèle X , spécifié sans tenir compte de la position des observations. Sur le plan théorique, les modèles temporels, comme les modèles spatiaux, reposent sur des processus stochastiques. On pourra trouver des analogies entre des outils comme les variogrammes utilisés en géostatistique et les corrélogrammes. Ceci dit, cette analogie a aussi des limites. En séries temporelles, on a une direction, et des intervalles de temps permettant de faire des modélisations empiriques (moyennes mobiles) ou probabilistes (*ARMA* et dérivés). Le spatial est plus complexe puisque l'on se situe dans un plan, et en général dans des structures présentant des discontinuités.

Une Auto-régression (AR) spatiale est un modèle de données réelles défini sur un ensemble discret spatial S . L'ensemble S est régulier (i.e. $S \in \mathbb{Z}^2$) ou non, sa géométrie étant associée à un graphe orienté d'influence \mathcal{R} . La corrélation spatiale d'une AR découlera du modèle retenu. Ces modèles explicatifs sont bien adaptés à des mesures agrégées par unités spatiales : par exemple, en épidémiologie, X_s est le nombre de personnes touchées par une maladie donnée dans un canton s d'un département S ; en agronomie, X_s est le rendement d'une parcelle de culture s .

Dans la section suivante une représentation du modèle autorégressif $2D$ sur une image.

1.5.1 Modèles autorégressifs 2D

Un des aspects le plus important dans une image en niveaux de gris est la dépendance statistique des niveaux de gris des pixels. Un des modèles qui présente cette dépendance est le modèle autorégressif à deux dimensions. Ce modèle caractérise la dépendance statistique en représentant l'intensité d'un pixel comme une combinaison linéaire $2D$ de l'intensité de ses voisins et un bruit additif :

$$X_{m,n} = \sum_{k,l \in \mathcal{S}} \phi_{k,l} X_{m-k,n-l} + \varepsilon_{m,n} , \text{ avec } (k,l) \neq (0,0) \quad (1.2)$$

avec :

- k, l , l'ensemble des coefficients de prédiction ,
- $\varepsilon_{m,n}$, un bruit blanc gaussien ,
- $k, l \in \mathcal{S}$ avec \mathcal{S} , le voisinage considéré.

Il existe trois configurations de voisinage d'un processus AR [Oukil 2007] illustrées par la figure 1.9

Modèle causal

La valeur prédite ne dépend que du passé. Comme une image n'a pas de définition temporelle, mais plutôt une définition spatiale, la forme causale n'a pas de signification physique [Deguchi 86]. Son avantage provient de la simplicité d'estimation des paramètres. Il existe deux grandes formes de voisinage causal. La première est le Quarter Plane où dans l'équation 1.2, $m \in [0, M]$ et $n \in [0, N]$. La seconde se nomme le NonSymmetric Half Plane et correspond $m \in [0, M]$ et $n \in [0, N] \cup m \in [-M, 0]$ et $n \in [1, N]$

Modèle semi-causal

Le modèle est causal et non-causal à la fois. La Figure (2.1,c) illustre ce voisinage défini pour $m \in [-M, M]$ et $n \in [0, N]$.

Modèle non-causal

Le voisinage non-causal est le plus général. Il permet de n'avoir aucune contrainte sur la forme du voisinage. Cependant, l'estimation des paramètres non-causaux est plus compliquée. La figure (2.1,d) représente un voisinage non-causal défini par $m \in [-M, M]$ et $n \in [0, N]$.

Exemple 15 *Modeles autoregressifs bilateraux.*

Les modeles autoregressifs bilateraux aux quatre voisins (figure. 1.10) , Le modele s'écrit

$$X_{i,j} = \sum_{k=-m}^m \sum_{l=-n}^n d_{k,l} X_{i-k,j-l} + \epsilon_{i,j}$$

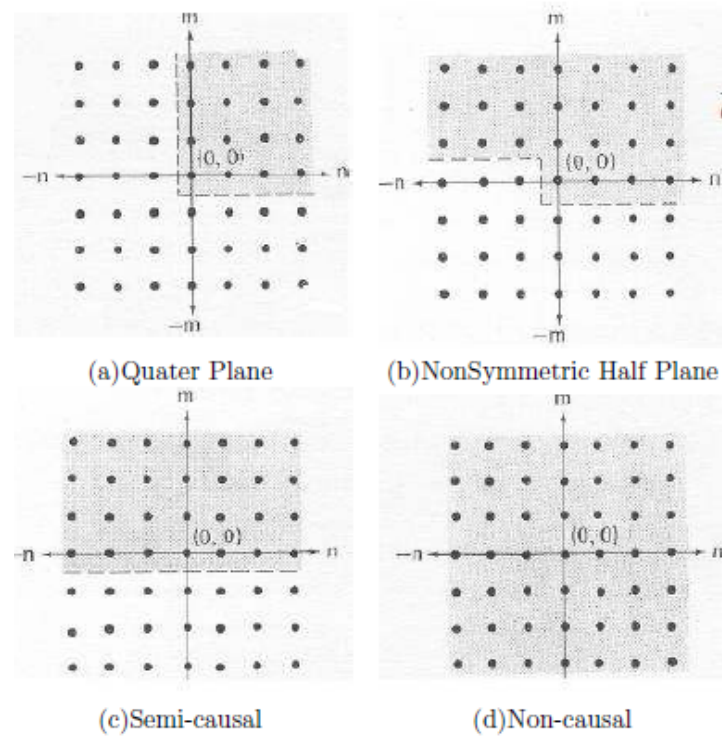


FIG. 1.9: Configurations de voisinage d'un processus AR – Les pixels considérés pour chaque type de voisinage sont les points foncés.

Avec $(k, l) \neq (0, 0)$, $m, n \in \mathbb{Z}$ et $\epsilon_{i,j}$ est une suite de variables aléatoires indépendantes et de même loi.

Exemple 16 *Modeles autoregressifs unilateraux :*

Le modele autoregressif unilaterial, au sens de Whittle, dependant des quatre voisins est represente par la figure 1.11 et par l'equation de recurrence 1.3 .

$$X_{i,j} = \alpha_1 X_{i,j-1} + \alpha_2 X_{i,j+1} + \alpha_3 X_{i-1,j} + \alpha_4 X_{i-1,j-1} + \epsilon_{i,j} \quad (1.3)$$

Avec $\alpha_1 = d_{0,1}$, $\alpha_2 = d_{1,-1}$, $\alpha_3 = d_{1,0}$, $\alpha_4 = d_{1,1}$. Où les α_i , $i = 1, 2, 3, 4$, sont les paramètres à estimer, avec les contraintes :

$$\sum |\alpha_i| < 1.$$

$$Var(X_{i,j}) = Var(\epsilon_{i,j}) = \sigma_\epsilon^2$$

On suppose que $\epsilon_{i,j}$ est non corrélé avec le passé du processus $X_{i,j}$.

Dans la section suivante, nous considérerons uniquement les modèles causaux, pour cela nous allons utiliser l'ordre lexicographique pour transformer une suite de variables à deux indices en une suite équivalente, mais à un seul indice.

Les champs que nous considérons sont des champs indexés par le réseau \mathbb{Z}^d ($d \geq 2$) muni de l'ordre partiel usuel. Plusieurs définitions de modèles *AR* spatiaux, indexés par le

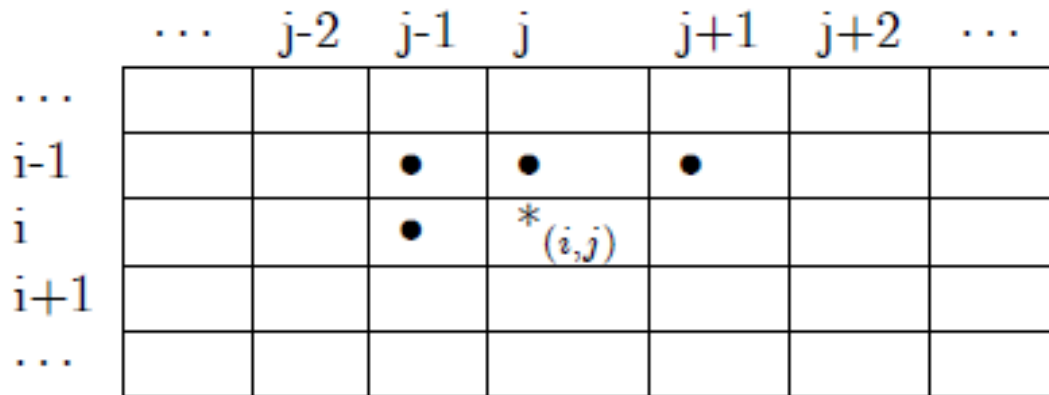


FIG. 1.10: Région de AR unilateral (4 voisin)

réseau \mathbb{Z}^d de dimension $d \geq 2$, ont été données dans la littérature. Ces définitions diffèrent principalement selon le choix des valeurs voisines dont est supposé dépendre un point du réseau. Ce choix permet de définir une structure de passé au sens large pour le modèle. Il est alors nécessaire de préciser le type de passé que nous considérons pour définir les modèles spatial *AR* ou *ARMA*.

1.5.2 Modèles *ARMA* spatiaux quadrantaux

Introduisons ici la définition de champ *ARMA* (de l'anglais Autoregressive Moving Average)

Définition 17 *Un champ de carré intégrable $(Y_t)_{t \in \mathbb{Z}^d}$ est appelé modèle *ARMA* (p, q) spatial de paramètres $p, q \in \mathbb{Z}^d$ s'il satisfait l'équation*

$$X_t = \sum_{j \in S(0,p]} \phi_j Y_{t-j} + \sum_{k \in S(0,q]} \theta_k \varepsilon_{t-j} + \varepsilon_t \quad (1.4)$$

où $(\phi_j)_{j \in S(0,p]}$ et $(\theta_k)_{k \in S(0,q]}$ désignent respectivement les paramètres de autorégressive et moyenne mobile avec $\phi_0 = \theta_0 = 1$ et $(\varepsilon_s)_{s \in \mathbb{Z}^d}$ désigne une séquence de variables aléatoires centrées indépendantes et identiques de variance σ^2 .

Notez que si $p = 0$, la somme sur $S(0,p]$ est supposé être nul et le processus est appelé champ aléatoire $MA(q)$ de moyenne mobile spatial,

$$Y_s = \sum_{k \in S(0,q]} \theta_k \varepsilon_{s-k} + \varepsilon_s \quad (1.5)$$

De même si le processus est appelé champ aléatoire $AR(p)$ autorégressif spatial, il est défini comme :

$$Y_s = \sum_{j \in S(0,p]} \phi_j Y_{s-j} + \varepsilon_s$$

Le champ aléatoire $ARMA$ est appelé causal s'il a la représentation unilatérale suivante :

$$Y_s = \sum_{j \in S[0,\infty]} \Psi_j \varepsilon_{s-j} \quad (1.6)$$

avec $\sum_{j \in S[0,\infty]} |\Psi_j| < \infty$.

1.5.3 Modèle AR spatial quadrantal

Nous rappelons la définition de modèle $AR(p)$ quadrantal de Tjøstheim (voir Tjøstheim 1978, 1983)

Définition 18 *Un champ de carré intégrable $(Y_t)_{t \in \mathbb{Z}^d}$ est appelé modèle $AR(p)$ spatial de*

paramètre p dans \mathbb{N}^d s'il satisfait l'équation

$$Y_t - \sum_{j \in S(0,p]} \phi_j Y_{t-j} = \epsilon_t \quad (1.7)$$

où $(\epsilon_t)_{t \in \mathbb{Z}^d}$ est un champ de carré intégrable stationnaire au second ordre, de variance $\sigma^2 \succ 0$ et vérifiant $E(\epsilon_t) = 0$, $E(\epsilon_t \epsilon_s) = 0$, $\forall t, s \in \mathbb{Z}^d$.

Les paramètres du modèle 1.7 sont la variance σ^2 du bruit $(\epsilon_t)_{t \in \mathbb{Z}^d}$, l'ordre p et les coefficients autorégressifs $(\phi_j)_{j \in S(0,p]}$.

Chapitre 2

Propriétés probabilistes des processus AR unilatérale spatiaux

2.1 Introduction

les champs *AR* spatiaux unilatéral a été étudié par de nombreux auteurs dont Tjøstheim ((1978),(1983), Ha et Newton (1993) et Choi (1997)). Nous nous sommes particulièrement intéressés aux résultats obtenus par Tjøstheim 1983. dans ce chapitre, on étudie le comportement asymptotique des estimateurs des moindres carrés $\hat{\alpha}_p = (\hat{\alpha}_j)_{j \in S\langle 0,p \rangle}$ et de Yule-Walker des coefficients $\alpha_p = (\alpha_j)_{j \in S\langle 0,p \rangle}$ d'un modèle *AR* quadrantal causal. Démonstre, sous certaines hypothèses, la consistance des deux estimateurs ainsi que la normalité asymptotique de l'estimateur des moindres carrés. Ce chapitre a été présenté lors Conférence internationale sur les progrès des mathématiques appliquées « ICAAM », Hammamet, December 16-19,2019, intitulée : Modèles Autorégressifs spatiaux : Estimateurs et

Applications.

2.2 Représentation du modèle AR unilatérale spatiaux

Considérons, sur un espace probabilisé (Ω, \mathcal{F}, P) ,

Définition 19 Tjostheim (1983) *Un champ AR (p) spatial causal $(Y_s)_{s \in \mathbb{Z}^d}$, avec $p \in \mathbb{Z}^d$ et $p \neq 0$, de fonction de covariance $\gamma(\cdot)$ défini par*

$$Y_t - \sum_{i \in S[\mathbf{0}, \mathbf{p}]} \alpha_i Y_{t-i} = \varepsilon_t \quad (2.1)$$

où $\varepsilon = (\varepsilon_s)_{s \in \mathbb{Z}^2}$ une famille de variables aléatoires centrées, indépendantes et identiquement distribuées et de même variance σ^2 . Les paramètres du modèle (2.1) sont la variance σ^2 du bruit $(\varepsilon_t)_{t \in \mathbb{Z}^d}$, l'ordre p et les coefficients autorégressifs $(\alpha_i)_{i \in S[\mathbf{0}, \mathbf{p}]}$.

Définition 20 *Soient $Y = (Y_t)_{t \in \mathbb{Z}^2}$ un processus spatial stationnaire au second ordre et à valeurs réelles et $\varepsilon = (\varepsilon_s)_{s \in \mathbb{Z}^2}$ une famille de variables aléatoires centrées indépendantes et identiquement distribuées et de même variance σ^2 . On dit que Y suit un modèle autorégressif spatial unilatéral AR d'ordre $\mathbf{p} = (p_1, p_2)$, si Y vérifie l'équation suivante :*

$$\sum_{i \in S[\mathbf{0}, \mathbf{p}]} \alpha_i Y_{t-i} = \varepsilon_t \quad (2.2)$$

Un cas particulier des modèles AR unilatéraux où la valeur au site (i, j) est une autorégression finie sur la région fortement causale $S_4(i, j)$ a été examinée par Tjostheim (1978). En deux dimensions, ce modèle devient

$$Y_{ij} = \sum_{k=0}^{p_1} \sum_{l=0}^{P_1} \alpha_{lk} Y_{i-k, j-l} + \varepsilon_{ij}, \quad \alpha_{0,0} = 1 \quad (2.3)$$

Ce modèle conduit à la représentation de type Wold

$$Y_{ij} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \alpha_{lk} \varepsilon_{i-k, j-l} \quad (2.4)$$

Supposons que B_1 et B_2 sont les opérateurs arrière horizontal et vertical $B_1^k Y_{i,j} = Y_{i-k, j}$ et $B_2^l Y_{i,j} = Y_{i, j-l}$, respectivement, et définissez le polynôme

$$\phi(B_1, B_2) = \sum_k \sum_l \alpha_{k,l} B_1^k B_2^l \quad \alpha_{00} = 1$$

Le modèle spatial autorégressif 2D AR est défini par

$$\phi(B_1, B_2) Y_{i,j} = \varepsilon_{i,j} \quad i, j \in \mathbb{Z}$$

Notez que l'étude de l'inversibilité de la fonction complexe $\phi(z_1, z_2)$ n'est pas un problème simple. Pour obtenir la représentation

$$Y_{ij} = \phi(B_1, B_2)^{-1} \varepsilon_{ij}$$

la fonction $\phi(z_1, z_2)^{-1}$ doit être écrite en utilisant un développement de Laurent comme

$$\phi(z_1, z_2)^{-1} = \sum \psi_{kl} z_1^k z_2^l$$

Exemple 21 *Modèle autorégressif spatial unilatéral AR($p_1, 1$)*

Un type particulier de modèles spatiaux qui a reçu beaucoup d'attention est les modèles autorégressifs unilatéraux spatiaux désignés par $AR(p_1, 1)$ et définis par

$$Y_{i,j} = \alpha_{10}Y_{i-1,j} + \alpha_{01}Y_{i,j-1} + \alpha_{11}Y_{i-1,j-1} + \dots + \alpha_{p_1 0}Y_{i-p_1,j} + \alpha_{p_1 1}Y_{i-p_1,j-1} + \varepsilon_{ij} \quad i = 1, \dots, m, j = 1, \dots, n \quad (2.5)$$

Où $\{Y_{ij}\}$ est une séquence de variables aléatoires bidimensionnelles avec une moyenne nulle et les erreurs ε_{ij} sont supposées être normalement distribuées avec une moyenne 0 et une variance commune σ^2 et α_{ij} sont des paramètres.

2.3 Stationnarité du processus AR spatial unilatéral

Dans le modèle donné par l'équation 2.3, introduit par Whittle (1954) et étudié en détail par Tjøstheim (1978, 1981, 1983), la valeur au site (i, j) est une autorégression finie sur les valeurs aux sites qui se trouvent dans le quart inférieur plan ie, Y_{ij} est une autorégression sur Y_{il} , $l \prec j$, et Y_{kl} , $k \prec i$. Tjøstheim (1978, Théorème 5.1) donne la condition de stationnarité du processus SAR unilatéral en d dimensions . Dans le plan $d = 2$. Définir le polynôme caractéristique

$$\phi(z_1, z_2) = 1 - \sum_{k=0}^{p_1} \sum_{l=0}^{p_2} \alpha_{lk} z_1^k z_2^l.$$

Le processus est stationnaire si

$$\phi(z_1, z_2) \neq 0 \quad \text{pour tous } |z_1| \leq 1, |z_2| \leq 1$$

c'est-à-dire que les racines de $\phi(z_1, z_2) = 0$ se trouvent toutes à l'extérieur du polydisque bidimensionnel. Pour le processus *SAR* unilatéral, nous avons donc une condition de stationnarité facilement vérifiée.

Définition 22 *Un champ $AR(p)$ $(Y_t)_{t \in \mathbb{Z}^d}$ est dit causal s'il admet une expression unilatérale du type*

$$Y_t = \sum_{j \in S(0, \infty]} \psi_j \varepsilon_{t-j} \quad \forall t \in \mathbb{Z} \quad (2.6)$$

avec $\sum_{j \in S(0, \infty]} |\psi_j| < \infty$.

Théorème 23 (voir *Tjøstheim 1978*) *Un champ de carré intégrable $(Y_t)_{t \in \mathbb{Z}^d}$ est appelé modèle $AR(p)$ spatial*

Une condition suffisante pour que le champ $AR(p)$ spatial soit causal est alors que le polynôme autorégressif ϕ n'admette pas de zéros dans le polydisque unité fermé $\bar{D}(0, 1)^d$ de \mathbb{C}^d .

Où $\phi(z) = 1 - \sum_{j \in S(0, p]} \alpha_j z^j \quad \forall z = (z_1, \dots, z_d) \in \mathbb{C}^d \quad \text{et} \quad \forall j = (j_1, j_2, \dots, j_d)$
 $z^j = z_1^{j_1} \dots z_d^{j_d}$.

Remarque 24 *Un champ AR causal est donc stationnaire.*

Exemple 25 *Soit le processus SAR unilatéral du premier ordre*

$$Y_{ij} = \alpha_{10} Y_{i-1, j} + \alpha_{01} Y_{i, j-1} + \alpha_{11} Y_{i-1, j-1} + \varepsilon_{i, j} \quad (2.7)$$

La fenêtre de prédiction dans ce cas est illustrée à la figure 2.1

2.4 L'estimation dans les modèles 2D $AR(2.1)$ unilatéral

2.4.1 Construction des matrices de poids pour le modèle autorégressif spatial du second ordre

Nous considérons un modèle $AR(2.1)$ spatial unilatéral non séparable, autorégressif, défini par l'équation

$$Y_{ij} = \alpha_{10}Y_{i-1,j} + \alpha_{01}Y_{i,j-1} + \alpha_{11}Y_{i-1,j-1} + \alpha_{20}Y_{i-2,j} + \alpha_{21}Y_{i-2,j-1} + \varepsilon_{ij} \quad i = 1, \dots, m, j = 1, \dots, n \quad (2.8)$$

$\{Y_{ij}\}$ est une séquence de variables aléatoires bidimensionnelles avec une moyenne nulle et les erreurs ε_{ij} sont supposées être normalement distribuées avec une moyenne 0 et une variance commune σ^2 .

En supposant que les valeurs non observées sont des zéros, et en laissant le vecteur d'observation, $\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{1n}, Y_{21}, \dots, Y_{2n}, \dots, Y_{m1}, \dots, Y_{mn})' = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m)'$ où

$$\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in})' \quad i = 1, \dots, m \quad \text{et le vecteur d'erreur } \boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{1n}, \varepsilon_{21}, \dots, \varepsilon_{2n}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{mn})'$$

$(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_m)'$ où $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})' \quad i = 1, \dots, m$

nous pouvons réécrire l'équation 2.8 sous la forme matricielle comme,

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A_2 & A_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ A_3 & A_2 & A_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_3 & A_2 & A_1 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_3 & A_2 & A_1 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \vdots \\ \mathbf{Y}_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_m \end{pmatrix} \quad (2.9)$$

où \mathbf{A}_1, A_2, A_3 sont des matrices $n \times n$ définies comme,

$$\mathbf{A}_1 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{01} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{01} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{01} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{01} & 0 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \alpha_{10} & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{11} & \alpha_{10} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{11} & \alpha_{10} & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{11} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{11} & \alpha_{10} \end{pmatrix} \quad \text{et}$$

$$\mathbf{A}_3 = \begin{pmatrix} \alpha_{20} & 0 & 0 & \cdots & 0 & 0 \\ \alpha_{21} & \alpha_{20} & 0 & \cdots & 0 & 0 \\ 0 & \alpha_{21} & \alpha_{20} & \cdots & 0 & 0 \\ 0 & 0 & \alpha_{21} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{21} & \alpha_{20} \end{pmatrix}$$

L'équation 2.9 peut être écrite de manière plus compacte comme suit :

$$\mathbf{Y} = \mathbf{A}\mathbf{Y} + \boldsymbol{\varepsilon} \quad (2.10)$$

où \mathbf{A} est la matrice $N \times N$ et $N = mn$. Il est clair que \mathbf{A} est une matrice triangulaire inférieure avec des zéros sur la diagonale principale. Alors, si on décompose \mathbf{A} en 5 matrices de telle sorte qu'il isole différents paramètres, on obtient

$$\mathbf{Y} = (\alpha_{10}\mathbf{W}_1 + \alpha_{01}\mathbf{W}_2 + \alpha_{11}\mathbf{W}_3 + \alpha_{20}\mathbf{W}_4 + \alpha_{21}\mathbf{W}_5) \mathbf{Y} + \boldsymbol{\varepsilon} \quad (2.11)$$

Avec $\mathbf{A} = \alpha_{10}\mathbf{W}_1 + \alpha_{01}\mathbf{W}_2 + \alpha_{11}\mathbf{W}_3 + \alpha_{20}\mathbf{W}_4 + \alpha_{21}\mathbf{W}_5$ et $\mathbf{W}_k, k = 1, \dots, 5$ sont les matrices $N \times N$ de poids triangulaires inférieures avec des éléments uns et zéros, données comme

$$\mathbf{W}_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{B}_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathbf{B}_1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \mathbf{B}_1 & 0 \end{pmatrix} \quad \mathbf{W}_2 = \begin{pmatrix} \mathbf{B}_2 & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{B}_2 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathbf{0} & \mathbf{B}_2 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \mathbf{0} & \mathbf{B}_2 \end{pmatrix}$$

$$\mathbf{W}_3 = \begin{pmatrix} \mathbf{0} & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{B}_2 & \mathbf{0} & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathbf{B}_2 & \mathbf{0} & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \mathbf{B}_2 & 0 \end{pmatrix} \quad \mathbf{W}_4 = \begin{pmatrix} \mathbf{0} & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & \dots & 0 & 0 & 0 \\ B_1 & \mathbf{0} & \mathbf{0} & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & B_1 & \mathbf{0} & 0 \end{pmatrix}$$

$$\text{et } \mathbf{W}_5 = \begin{pmatrix} \mathbf{0} & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & \dots & 0 & 0 & 0 \\ B_2 & \mathbf{0} & \mathbf{0} & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & B_2 & \mathbf{0} & 0 \end{pmatrix}$$

ainsi que, B_1 et B_2 sont des matrices $n \times n$ définies comme

$$B_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & & 1 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \quad \text{et } B_2 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \mathbf{1} & 0 \end{pmatrix}$$

2.4.2 Méthode de maximement de vraisemblance pour le modèle spatial

AR(2.1)

L'équation (2.11) peut alors s'écrire,

$$\mathbf{Y} = (\mathbf{I} - (\alpha_{10}\mathbf{W}_1 + \alpha_{01}\mathbf{W}_2 + \alpha_{11}\mathbf{W}_3 + \alpha_{20}\mathbf{W}_4 + \alpha_{21}\mathbf{W}_5))^{-1} \boldsymbol{\varepsilon} \quad (2.12)$$

ou

$$\mathbf{Y} = (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\varepsilon} \quad (2.13)$$

où \mathbf{I} est une matrice d'identité $N \times N$.

Par conséquent, la matrice de covariance de \mathbf{Y} , \mathbf{V} est donnée par,

$$\mathbf{V} = \sigma^2 (\mathbf{I} - \mathbf{A})^{-1} \left[(\mathbf{I} - \mathbf{A})^{-1} \right]' \quad (2.14)$$

La racine carrée du déterminant de \mathbf{V} est donnée par,

$$|\mathbf{V}|^{\frac{1}{2}} = (\sigma^2)^{\frac{N}{2}} \left| (\mathbf{I} - \mathbf{\Phi})^{-1} \right| \quad (2.15)$$

Puisque $(\mathbf{I} - \mathbf{A})$ est la matrice triangulaire inférieure avec des éléments diagonaux 1, $\left| (\mathbf{I} - \mathbf{\Phi})^{-1} \right|$. Cela mène à

$$|\mathbf{V}|^{\frac{1}{2}} = (\sigma^2)^{\frac{N}{2}} \quad (2.16)$$

Par conséquent, la fonction de vraisemblance l est donnée par,

$$\begin{aligned} l &= \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{V}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{Y}' \mathbf{V}^{-1} \mathbf{Y} \right\} \\ &= (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{Y}' \left[(\mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - \hat{\mathbf{A}})^{-1} \right]^{-1} \mathbf{Y} \right\} \\ &= (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{Y}' (\mathbf{I} - \hat{\mathbf{A}}) (\mathbf{I} - \mathbf{A}) \mathbf{Y} \right\} \end{aligned}$$

Ainsi, nous obtenons la vraisemblance logarithmique, L comme

$$L = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \mathbf{Y}' (\mathbf{I} - \hat{\mathbf{A}}) (\mathbf{I} - \mathbf{A}) \mathbf{Y} \quad (2.17)$$

Soit $\hat{\alpha} = (\alpha_{10}, \alpha_{01}, \alpha_{11}, \alpha_{20}, \alpha_{21}) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ pour chaque $i = 1, \dots, 5$

La dérivée partielle de L par rapport à α_i , $j = 1, \dots, 5$ est donné par

$$\frac{\partial L}{\partial \alpha_i} = -\frac{1}{\sigma^2} \left[-\mathbf{Y}'\mathbf{W}'_i\mathbf{Y} + \alpha_i \mathbf{Y}'\mathbf{W}'_i\mathbf{W}_i\mathbf{Y} + \sum_{\forall r \neq j} \alpha_j \mathbf{Y}'\mathbf{W}'_j\mathbf{W}_i\mathbf{Y} \right] \quad (2.18)$$

équilibrer 2.18 à zéro conduit à

$$\alpha_i \mathbf{Y}'\mathbf{W}'_i\mathbf{W}_i\mathbf{Y} + \sum_{\forall r \neq j} \alpha_j \mathbf{Y}'\mathbf{W}'_j\mathbf{W}_i\mathbf{Y} = \mathbf{Y}'\mathbf{W}'_i\mathbf{Y} \quad (2.19)$$

Par conséquent, en notant $\mathbf{Z}_i = \mathbf{W}_i\mathbf{Y}$, le maximum de vraisemblance pour $\hat{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ peut être obtenu en résolvant l'équation

$$\begin{pmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_3\mathbf{Z}_1 & \mathbf{Z}'_4\mathbf{Z}_1 & \mathbf{Z}'_5\mathbf{Z}_1 \\ \mathbf{Z}_1\mathbf{Z}_2 & \mathbf{Z}'_2\mathbf{Z}_2 & \mathbf{Z}'_3\mathbf{Z}_2 & \mathbf{Z}'_4\mathbf{Z}_2 & \mathbf{Z}'_5\mathbf{Z}_2 \\ \mathbf{Z}'_1\mathbf{Z}_3 & \mathbf{Z}'_2\mathbf{Z}_3 & \mathbf{Z}'_3\mathbf{Z}_3 & \mathbf{Z}'_4\mathbf{Z}_3 & \mathbf{Z}'_5\mathbf{Z}_3 \\ \mathbf{Z}'_1\mathbf{Z}_4 & \mathbf{Z}'_2\mathbf{Z}_4 & \mathbf{Z}'_3\mathbf{Z}_4 & \mathbf{Z}'_4\mathbf{Z}_4 & \mathbf{Z}'_5\mathbf{Z}_4 \\ \mathbf{Z}'_1\mathbf{Z}_5 & \mathbf{Z}'_2\mathbf{Z}_5 & \mathbf{Z}'_3\mathbf{Z}_5 & \mathbf{Z}'_4\mathbf{Z}_5 & \mathbf{Z}'_5\mathbf{Z}_5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} \mathbf{Y}'\mathbf{Z}_1 \\ \mathbf{Y}'\mathbf{Z}_2 \\ \mathbf{Y}'\mathbf{Z}_3 \\ \mathbf{Y}'\mathbf{Z}_4 \\ \mathbf{Y}'\mathbf{Z}_5 \end{pmatrix}$$

$$\text{Ou} \quad \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_3\mathbf{Z}_1 & \mathbf{Z}'_4\mathbf{Z}_1 & \mathbf{Z}'_5\mathbf{Z}_1 \\ \mathbf{Z}_1\mathbf{Z}_2 & \mathbf{Z}'_2\mathbf{Z}_2 & \mathbf{Z}'_3\mathbf{Z}_2 & \mathbf{Z}'_4\mathbf{Z}_2 & \mathbf{Z}'_5\mathbf{Z}_2 \\ \mathbf{Z}'_1\mathbf{Z}_3 & \mathbf{Z}'_2\mathbf{Z}_3 & \mathbf{Z}'_3\mathbf{Z}_3 & \mathbf{Z}'_4\mathbf{Z}_3 & \mathbf{Z}'_5\mathbf{Z}_3 \\ \mathbf{Z}'_1\mathbf{Z}_4 & \mathbf{Z}'_2\mathbf{Z}_4 & \mathbf{Z}'_3\mathbf{Z}_4 & \mathbf{Z}'_4\mathbf{Z}_4 & \mathbf{Z}'_5\mathbf{Z}_4 \\ \mathbf{Z}'_1\mathbf{Z}_5 & \mathbf{Z}'_2\mathbf{Z}_5 & \mathbf{Z}'_3\mathbf{Z}_5 & \mathbf{Z}'_4\mathbf{Z}_5 & \mathbf{Z}'_5\mathbf{Z}_5 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}'\mathbf{Z}_1 \\ \mathbf{Y}'\mathbf{Z}_2 \\ \mathbf{Y}'\mathbf{Z}_3 \\ \mathbf{Y}'\mathbf{Z}_4 \\ \mathbf{Y}'\mathbf{Z}_5 \end{pmatrix} \quad (2.20)$$

2.4.3 Estimation spatiale Yule-Walker

Tjøstheim (1978) a étudié la méthode Yule-Walker pour estimer les paramètres des modèles AR spatiaux. Pour les données sur une grille régulière bidimensionnelle, Y_{ij} , $i = 1, \dots, m, j = 1, \dots, n$, les autocovariances de l'échantillon au décalage (s, t) et $(s, -t)$ pour $s \geq 0$ et $t \geq 0$ sont définies respectivement comme,

$$R(s, t) = \frac{1}{mn} \sum_{i=1}^{m-s} \sum_{j=1}^{n-t} Y_{ij} Y_{i+s, j+t}$$

et

$$R(s, -t) = \frac{1}{mn} \sum_{i=1}^{m-s} \sum_{j=1}^n Y_{ij} Y_{i+s, j-t}$$

D'après ces définitions, $R(s, t) = R(-s, -t)$ et $R(-s, t) = R(s, -t)$. Pour un modèle AR unilatéral spatial (p_1, p_2) , l'analogie spatiale des équations unidimensionnelles de Yule-Walker comme dans le cas des séries chronologiques est donné comme suit :

$$R(s, t) = \sum_{k=0}^{p_1} \sum_{l=0}^{p_2} \alpha_{kl} R(s-k, t-l)$$

Ensuite, si nous définissons

$$\alpha = (\alpha_{10}, \dots, \alpha_{p_1 0}, \alpha_{01}, \dots, \alpha_{p_1 1}, \dots, \alpha_{0 p_2}, \dots, \alpha_{p_1 p_2})'$$

$$r = (R(1, 0), \dots, R(p_1, 0), R(0, 1), \dots, R(p_1, 1), \dots, R(0, p_2), \dots, R(p_1, p_2))'$$

et

$$\mathbf{R} = \begin{pmatrix} R(0,0) & R(-1,0) & \cdots & R(1-p_1,p_2) \\ R(1,0) & R(0,0) & \cdots & R(2-p_1,p_2) \\ \vdots & \vdots & \cdots & \vdots \\ R(p_1-1,p_2) & R(p_1-2,p_2) & \cdots & R(0,0) \end{pmatrix}.$$

Les estimations spatiales de Yule-Walker, $\hat{\alpha}$ peuvent être obtenues en résolvant l'équation,

$$\hat{\alpha} = \mathbf{R}^{-1}r \quad (2.21)$$

puisque notre intérêt porte sur un modèle $AR(2,1)$ spatial unilatéral tel que défini dans l'équation (2.12) ci-dessus, α , r et \mathbf{R} prendront les formes, $\alpha = (\alpha_{10}, \alpha_{20}, \alpha_{01}, \alpha_{11}, \alpha_{21})'$ et $r = (R(1,0), R(2,0), R(0,1), R(1,1), R(2,1))'$ et

$$R = \begin{pmatrix} R(0,0) & R(1,0) & R(1,-1) & R(0,1) & R(1,1) \\ R(1,0) & R(0,0) & R(2,-1) & R(1,-1) & R(0,1) \\ R(1,-1) & R(2,-1) & R(0,0) & R(1,0) & R(2,0) \\ R(0,1) & R(1,-1) & R(1,0) & R(0,0) & R(-1,0) \\ R(1,1) & R(0,1) & R(2,0) & R(1,0) & R(0,0) \end{pmatrix}$$

Ensuite, l'estimation de $\alpha = (\alpha_{10}, \alpha_{20}, \alpha_{01}, \alpha_{11}, \alpha_{21})'$ est obtenue en utilisant l'équation (2.21) ci-dessus.

2.4.4 Estimation spatiale conditionnelle des moindres carrés

Nous discutons ici de deux types de méthode d'estimation conditionnelle par les moindres carrés. Pour le type 1, comme dans la méthode du maximum de vraisemblance nous supposons que les valeurs des frontières non observées sont toutes nulles, c'est-à-dire $Y'_b = (Y_{-1,0}, \dots, Y_{-1,n}, Y_{00}, \dots, Y_{0n}, Y_{10}, \dots, Y_{m0}) = 0$. Les estimations des moindres carrés de $\alpha' = (\alpha_{10}, \alpha_{01}, \alpha_{11}, \alpha_{20}, \alpha_{21})$ sont données comme

$$\hat{\alpha}' = (X'X)^{-1} X'Y \quad (2.22)$$

Avec $\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{1n}, Y_{21}, \dots, Y_{2n}, \dots, Y_{m1}, \dots, Y_{mn})'$ et X est une matrice de dimension $(mn) \times 5$ donnée comme,

$$X = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & Y_{11} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & Y_{1,n-1} & 0 & 0 & 0 \\ Y_{11} & 0 & 0 & 0 & 0 \\ Y_{12} & Y_{21} & Y_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{1n} & Y_{2,n-1} & Y_{1,n-1} & 0 & 0 \\ Y_{21} & 0 & 0 & Y_{11} & 0 \\ Y_{22} & Y_{31} & Y_{21} & Y_{12} & Y_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{2n} & Y_{3,n-1} & Y_{2,n-1} & Y_{1n} & Y_{1,n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{m-1,1} & 0 & 0 & Y_{m-2,1} & 0 \\ Y_{m-1,2} & Y_{m,1} & Y_{m-1,1} & Y_{m-2,2} & Y_{m-2,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1} & Y_{m-2,n} & Y_{m-2,n-1} \end{pmatrix}$$

Il est évident que les estimations de l'équation 2.22 sont équivalentes aux estimations du maximum de vraisemblance données dans l'équation (14). Cela est dû à notre hypothèse précédente que

les valeurs de frontière non observées sont toutes fixées à zéro, c'est-à-dire $\mathbf{Y}'_b = (Y_{-1,0}, \dots, Y_{-1,n}, Y_{00}, \dots, Y_{0n}, Y_{10}, \dots, Y_{m0}) = 0$ ainsi que la racine carrée du déterminant

de la matrice de covariance de $\mathbf{Y}, |\mathbf{V}|^{\frac{1}{2}} (\sigma^2)^{\frac{N}{2}}$ implique aucune fonction de α' . Par conséquent, maximiser l'équation par rapport à α' est analogue à la procédure pour obtenir les estimations conditionnelles des moindres carrés comme dans l'équation 2.22.

Dans le type 2, nous obtenons les estimations conditionnelles des moindres carrés en conditionnant sur la frontière observée donnée, $\mathbf{Y}'_0 = (Y_{11}, \dots, Y_{1n}, Y_{21}, \dots, Y_{2n}, Y_{31} \dots Y_{m1})$. Ensuite, l'estimateur est donné comme suit :

$$\hat{\alpha}_0 = \left(X'_0 X_0 \right)^{-1} \mathbf{X}'_0 \mathbf{Y}_{(1)} \quad (2.23)$$

Avec

$$\mathbf{Y}'_{(1)} = (Y_{32}, \dots, Y_{3n}, Y_{42}, \dots, Y_{4n}, Y_{m2}, \dots, Y_{mn})$$

et \mathbf{X}_0 est une matrice de dimension $(m-2)(n-1) \times 5$ définie comme,

$$\mathbf{X}_0 = \begin{pmatrix} Y_{22} & Y_{31} & Y_{21} & Y_{12} & Y_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{2n} & Y_{3,n-1} & Y_{2,n-1} & Y_{1n} & Y_{1,n-1} \\ Y_{32} & Y_{41} & Y_{31} & Y_{22} & Y_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{3n} & Y_{4,n-1} & Y_{3,n-1} & Y_{2n} & Y_{2,n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{m-1,2} & Y_{m1} & Y_{m-1,1} & Y_{m-2,2} & Y_{m-2,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{m-1,n} & Y_{m,n-1} & Y_{m-1,n-1} & Y_{m-1,n} & Y_{m-1,n-1} \end{pmatrix}$$

2.5 L'estimation dans les modèles $AR(p)$ spatial causal

Étant donné un ensemble d'observations $\{Y_t, t \in S[\mathbf{1}, \mathbf{N}]\}$ du champ $(Y_t)_{t \in \mathbb{Z}^d}$, nous souhaitons estimer les coefficients $\{\alpha_j, j \in S\langle \mathbf{0}, \mathbf{p} \rangle\}$ du modèle 2.1. Pour cela, deux types d'estimateurs suivants sont considérés.

2.5.1 L'estimateur des moindres carrés

Notons $\hat{\alpha}_p = (\hat{\alpha}_{p,j})_{j \in S\langle \mathbf{0}, \mathbf{p} \rangle}$ l'estimateur des moindres carrés de $\alpha_p = (\alpha_j)_{j \in S\langle \mathbf{0}, \mathbf{p} \rangle}$ obtenu par minimisation de la quantité suivante

$$\frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} \left(Y_t - \sum_{j \in S\langle \mathbf{0}, \mathbf{p} \rangle} \alpha_j Y_{t-j} \right)^2.$$

Cette le minimisation revient à résoudre le système linéaire d'équations suivant

$$\forall j \in S\langle \mathbf{0}, \mathbf{p} \rangle, \sum_{i \in S\langle \mathbf{0}, \mathbf{p} \rangle} \hat{\alpha}_{p,i} \left(\frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} Y_{t-i} Y_{t-j} \right) = \frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} Y_t Y_{t-j} \quad (2.24)$$

En introduisant les quantités suivantes, pour tous $i, j \in S\langle \mathbf{0}, \mathbf{p} \rangle$,

$$\hat{\Gamma}_p(i, j) = \frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} Y_{t-i} Y_{t-j} \quad (2.25)$$

$$\hat{\gamma}_p(i) = \frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} Y_t Y_{t-i}. \quad (2.26)$$

le système d'équations 2.24 peut alors s'écrire sous la forme de l'équation matricielle suivante

$$\hat{\Gamma}_p \hat{\alpha}_p = \hat{\gamma}_p \quad (2.27)$$

où la matrice $\hat{\Gamma}_p$ et les vecteurs $\hat{\alpha}_p, \hat{\gamma}_p$ ont été formés en ordonnant les éléments de $S \langle 0, p \rangle$ par l'ordre lexicographique

2.5.2 L'estimateur de Yule-Walker

En multipliant successivement l'équation 2.1 par Y_{t-j} pour $j \in S \langle 0, p \rangle$ et en prenant ensuite l'espérance, l'hypothèse de causalité permet d'obtenir les équations de Yule-Walker suivantes

$$\gamma(j) = \sum_{i \in S \langle 0, p \rangle} \gamma(i-j) \alpha_i, \forall j \in S \langle 0, p \rangle \quad (2.28)$$

L'estimateur de Yule-Walker est obtenu à partir de ces équations en estimant les covariances par

$$\bar{R}(h) = \frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t+h \in S[\mathbf{1}, \mathbf{N}]}} Y_t Y_{t+h} \forall h \in S \langle 0, p \rangle \quad (2.29)$$

Plus précisément, l'estimateur de Yule-Walker $\bar{\alpha}_p = \left(\bar{\alpha}_{p,j} \right)_{j \in S \langle 0, p \rangle}$ est solution de l'équation matricielle suivante

$$\bar{\Gamma}_p \bar{\alpha}_p = \bar{\gamma}_p$$

où pour $i, j \in S \langle 0, p \rangle$,

$$\bar{\Gamma}_p(i, j) = \bar{R}(i-j)$$

$$\bar{\gamma}_p(j) = \bar{R}(j).$$

2.6 Consistance et normalité asymptotique des estimateurs

La consistance des estimateurs des moindres carrés et de Yule-Walker est obtenue par Tjostheim (1983). Ce résultat est énoncé dans le Théorème 2.4.1.

Théorème 27 (Tjostheim (1983)) *Soit $(Y_t)_{t \in \mathbb{Z}^d}$ un champ AR(p), $p \neq 0$, causal. Supposons que $(\varepsilon_t)_{t \in \mathbb{Z}^d}$ est un champ strictement stationnaire tel que $E(\varepsilon_t \varepsilon_s) = 0$ et pour tout $t \in \mathbb{Z}^d$, $E(\varepsilon_t^4) < \infty$. Si de plus pour tout $t \in \mathbb{Z}^d$,*

$$E\left(\varepsilon_t^2 \mid \bigvee_{i=1}^d \mathcal{F}^i(t_i - 1)\right) = \sigma^2 \quad p.s$$

alors les estimateurs des moindres carrés $\hat{\alpha}$ et de Yule-Walker $\bar{\alpha}$ convergent presque sûrement vers α lorsque N tend vers l'infini.

La preuve de la normalité asymptotique de l'estimateur des moindres carrés est donnée dans Tjostheim (1983). Cependant, pour la commodité du lecteur, nous donnons ci-dessous les grandes lignes de cette preuve.

Pour établir la normalité asymptotique de $\hat{\alpha}$, nous étudions la convergence en loi du vecteur $\bar{\Gamma}_p(\hat{\alpha}_p - \alpha_p)$. Les relations 2.1 et ?? permettent d'exprimer les coordonnées de ce vecteur comme suit : la i ième composante est donnée par

$$\frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} Y_{t-i} \varepsilon_t \quad (2.30)$$

Le procédé de Cramer-Wold (voir BD 91) conduit alors à étudier pour toute famille de réels $(r_u)_{u \in S(0,p)}$, la convergence en loi des statistiques

$$\frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} \sum_{u \in S(0,p)} r_u Y_{t-u} \varepsilon_t \quad (2.31)$$

En remplaçant, dans 2.31, Y_{t-u} par son expression unilatérale 2.6 induite par l'hypothèse de causalité, aux effets de causalité, aux effets de bords près, on est alors ramené pour tout $K \in \mathbb{N}^*$ grand, à l'étude de la convergence en loi des termes de la forme

$$\frac{1}{N^d} \sum_{\substack{t \in S[\mathbf{1}, \mathbf{N}] \\ t-p \in S[\mathbf{1}, \mathbf{N}]}} \sum_{u \in S\langle 0, p \rangle} r_u \sum_{v \in S[\mathbf{1}, \mathbf{K}]} \psi_v \varepsilon_{t-u-v} \varepsilon_t \quad (2.32)$$

on en déduit ainsi la normalité asymptotique de l'estimateur des moindres carrés.

Le théorème 2.2.2 énonce ce résultat de Tjostheim (1983)

Théorème 28 (Tjostheim (1983)) *sous les mêmes hypothèses que dans le théorème 2.2.1, on a*

$$N^{d/2} (\hat{\alpha}_p - \alpha_p) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(0, \sigma^2 \Gamma_p^{-1}),$$

où ${}^2\Gamma_p = (\gamma(i-j))_{i,j \in S\langle 0, p \rangle}$ et $\xrightarrow[N \rightarrow \infty]{} désigne la convergence en loi.$

Tjostheim (1983) affirmait, dans le théorème 4.2 de Tjostheim (1983), que l'estimateur de Yule-Walker était asymptotiquement normal et sans biais. Or, pour un réseau de dimension 2 et des innovations indépendantes et identiquement distribuées, Ha et Newton (1993) démontrent la normalité asymptotique de l'estimateur de Yule-Walker mais avec un biais asymptotique non nul. Dans le théorème 3.2.3 ci-dessous, nous rapploons ce résultat ainsi que l'expression du biais dans le cas $d = 2$.

Théorème 29 (Ha et Newton (1993)) *Soit $(Y_t)_{t \in \mathbb{Z}^2}$ un champ AR(p) causal avec $p \neq 0$. Supposons que $(\varepsilon_t)_{t \in \mathbb{Z}^2}$ est une famille de variables i.i.d. centrées et de variance σ^2 non nulle.*

Alors,

$$N^{d/2} \left(\bar{\alpha}_p - \alpha_p \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(\Gamma_p^{-1} \Gamma_p^* \alpha, \sigma^2 \Gamma_p^{-1})$$

où pour $(h, k) \in S\langle 0, p \rangle$,

$$\Gamma_p^*(h, k) = \begin{cases} -(k_2 + k_1) \gamma((h_1 - k_1, h_2 - k_2)) & \text{si } h_1 - k_1 \geq 0, h_2 - k_2 \geq 0 \\ -(2h_2 + k_2 + k_1) \gamma((h_1 - k_1, h_2 - k_2)) & \text{si } h_1 - k_1 \geq 0, h_2 - k_2 \leq 0 \\ -(2h_1 + k_2 - k_1) \gamma((h_1 - k_1, h_2 - k_2)) & \text{si } h_1 - k_1 \leq 0, h_2 - k_2 \geq 0 \\ -(2h_2 - k_2 + 2h_1 - k_1) \gamma((h_1 - k_1, h_2 - k_2)) & \text{si } h_1 - k_1 \leq 0, h_2 - k_2 \leq 0 \end{cases}$$

et $\Gamma_p = (\gamma(i - j))_{i, j \in S\langle 0, p \rangle}$.

2.7 Identification de l'ordre p du modèle

Considérons ici, un modèle $AR(p)$ spatial causal $(Y_t)_{t \in \mathbb{Z}^d}$ de paramètre $p \in \mathbb{Z}^d$ inconnu et d'innovations $(\varepsilon_t)_{t \in \mathbb{Z}^d}$ indépendantes et identiquement distribuées. Sur la base d'un ensemble d'observations $\{Y_t, t \in S[\mathbf{1}, \mathbf{N}]\}$, nous souhaitons déterminer l'ordre p du modèle.

Tjøstheim (1983) propose d'estimer p par \hat{p} , obtenu en minimisant la fonction critère d'information suivante

$$C(r) = \ln \hat{\sigma}^2(r) - \frac{1}{N^d} f(N^d) d(r) \quad (2.33)$$

pour des valeurs de r ordonnées selon l'ordre partiel et variant dans un rectangle $S[0, \mathbf{K}]$ suffisamment grand.

Le coefficient $d(r) = \prod_{i=1}^d (r_i + 1) - 1$ représente le nombre de paramètres auto-régressifs à estimer, le terme $\hat{\sigma}^2(r)$ est la variance résiduelle après ajustement d'un champ $AR(r)$ aux données par estimation des moindres carrés et la fonction f détermine le poids accordé au terme de pénalité de 2.33

L'utilisation de ce critère est justifié par le théorème suivant.

Théorème 30 (Tjøstheim (1983)) Soit $(Y_t)_{t \in \mathbb{Z}^d}$ un champ $AR(p)$ causal engendré par des innovations $(\varepsilon_t)_{t \in \mathbb{Z}^d}$ indépendantes et identiquement distribuées. Si $p \preceq K$ et si f est telle que $\frac{1}{x}f(x) \rightarrow \infty$ et $f(x) \rightarrow \infty$ lorsque $x \rightarrow \infty$, alors $P(\hat{p} \neq p) \rightarrow 0$ lorsque $N \rightarrow \infty$.

Chapitre 3

Etude des champs aléatoires de Markov (MRF)

Ce chapitre a été présenté lors de Congrès des Mathématiciens Algériens « CMA'2018, 12-13 mai 2018 », intitulée : « Aperçu sur les champs de Markov : théorie et application ».

3.1 Introduction

La théorie des champs aléatoires de Markov (MRF) est une branche de la théorie des probabilités pour l'analyse des dépendances spatiales ou contextuelles des phénomènes physiques. Les champs de Markov introduits par Lévy (1948), ont montré leur intérêt dans la modélisation de l'image à partir des années 70 (Besag (1974), Jai (1981)), notamment par les travaux de Geman (1984). Les modèles markoviens sont des modèles largement utilisés dans de nombreux domaines d'application, notamment en traitement d'images (restaura-

tion (Chalmond (1989), segmentation (Derin et Elliott (1987)), analyse de mouvement (Bouthemy et Franc (1993), reconstruction 3D (Hung, Cooper et Cernuschi (1991)), tomographie (Saqib, Bouman et Sauer (1998), IRM (Scherrer, Dojat, Forbes et Garbay (2007),...), analyse de textes (Metzler et Croft (2005), Bouchaffra et Meunier (1995)), reconnaissance vocale (Gravier, Sigelle et Cholle (2000)), classification de gènes (Forbes et Vignes (2006)), modélisation de musique [56]. Leur caractéristique est de fournir une modélisation explicite des dépendances entre individus via l'utilisation d'une structure de voisinage ou, de manière équivalente, d'un graphe (ou réseau) d'interactions. Le terme "individu" est ici à prendre au sens large : il s'agit de l'entité sur laquelle on fait des observations, les pixels d'une image, des gènes, des segments de texte, un signal sonore etc... Les individus étant définis sur un graphe d'interaction, par abus de langage, on les dénomme encore sites bien que leurs dépendances ne soient pas nécessairement géographiques.

Ce chapitre a pour but de donner une vue globale des champs aléatoires de Markov (Markov Random Fields) MRF discrets. Parce que c'est une généralisation de l'idée de chaînes de Markov unidimensionnelles ($1D MC$) au contexte 2-dimensionnel ou 3-dimensionnel. Ici, nous étudions deux modèles de MRF : Le premier modèle que nous utilisons souvent dans le contexte de phénomènes physique, à savoir les champs de Markov non causaux ($NMRF$), ces modèles selon la théorie de Hammersley - Clifford sont équivalents aux distributions de Gibbs GRF . Les modèle MRF considéré dans la deuxième section est une extension naturelle des modèles Markov cachés (en anglais Hidden Markov Models) unidimensionnels (HMM) aux modèles MRF causaux 2-dimensionnels ($HMRF$), et le principal obstacle à cette extension est l'absence d'ordre naturel dans un réseau bidi-

mensionnel et donc aucune idée naturelle de causalité. Dans les données spatiales générées par les images. Nous avons donc discuté de la conversion de $HMRF$ en HMM en fonction de la numérisation, la fin de ce chapitre, nous discuterons de l'estimation des paramètres du modèle $HMRF$.

Dans la première section nous commençons par rappeler quelques notions de base, tels le système de voisinage et clique. Ensuite, nous présentons la définition mathématique des champs aléatoires de Markov et distribution de Gibbs, ainsi que l'équivalence entre champs de Markov et champs de Gibbs (théorème de Hammersley – Clifford)

3.2 Cadre théorique des champs aléatoires de Markov

Dans toute la suite, on considère une grille rectangulaire S définie par : $S = \{(i, j) / 0 \leq i < m, 0 \leq j < n\}$ constituée de $n \times m$ sites désignés par leurs coordonnées (i, j) .

3.2.1 Définitions

Rappelons les définitions des champs aléatoires ainsi que des systèmes de voisinage et des cliques associées sur une grille S , Λ est l'espace d'état.

Définition 31 (Champ aléatoire) *Un champ aléatoire X défini sur S est une collection de variables aléatoires : $X = \{X_s = X_{i,j}, s = (i, j) \in S\}$ à valeurs dans Λ .*

Remarque : En analyse d'images, l'image est représentée par une grille S rectangulaire finie bidimensionnelle de taille $n \times m$, contenant un ensemble de "sites" $s = (i, j) \in S$ appelés également "pixels".

Définition 32 (Voisin). Deux sites s_i et s_j sont dits voisins s'il existe une arête $j \in V$ entre eux. L'ensemble des voisins d'un site s est noté \mathcal{V}_s .

A titre d'exemple, l'ensemble des "voisins" d'un site s peut être défini comme un ensemble de sites situé dans un rayon r : $\mathcal{V}_s = \{t \in S / d(s, t) \leq r, t \neq s\}$ où $d(s, t)$ est la distance euclidienne entre le pixel s et le pixel t et r est à valeur dans l'ensemble des entiers (Li (1995)).

Définition 33 (Système de Voisinage). $\mathcal{V} = \{\mathcal{V}_s | s \in S\}$ est un système de voisinage s'il possède les propriétés :

- $s \notin \mathcal{V}_s$ c-à-d un site n'est pas voisin à lui même,
- $\forall (s, t) \in S \quad s \in \mathcal{V}_t \Leftrightarrow t \in \mathcal{V}_s$ qui signifie que la relation de voisinage est symétrique.

(S, \mathcal{V}) est le graphe de nœuds S et d'arêtes \mathcal{V} .

Remarque : i) Un système de voisinage est généralement symétrique, le graphe de voisinage correspondant est non orienté. **ii)** Toutefois, certains modèles tels que les champs aléatoires de Markov causaux (Causal MRF) font appel à un système de voisinage non symétrique. On parlera de voisinage passé et futur (figure 3.1) :

- iii) Dans les deux cas, un site n'appartient pas à son propre voisinage : $\forall s \in S, s \notin \mathcal{V}_s$

A partir d'un système de voisinage, un système de clique est déduit.

Définition 34 Un sous-ensemble $c \subseteq S$ est un clique si chaque paire de sites dans c est voisine. Soit C l'ensemble des cliques associées à (S, \mathcal{V}) : $c \in C$ ssi

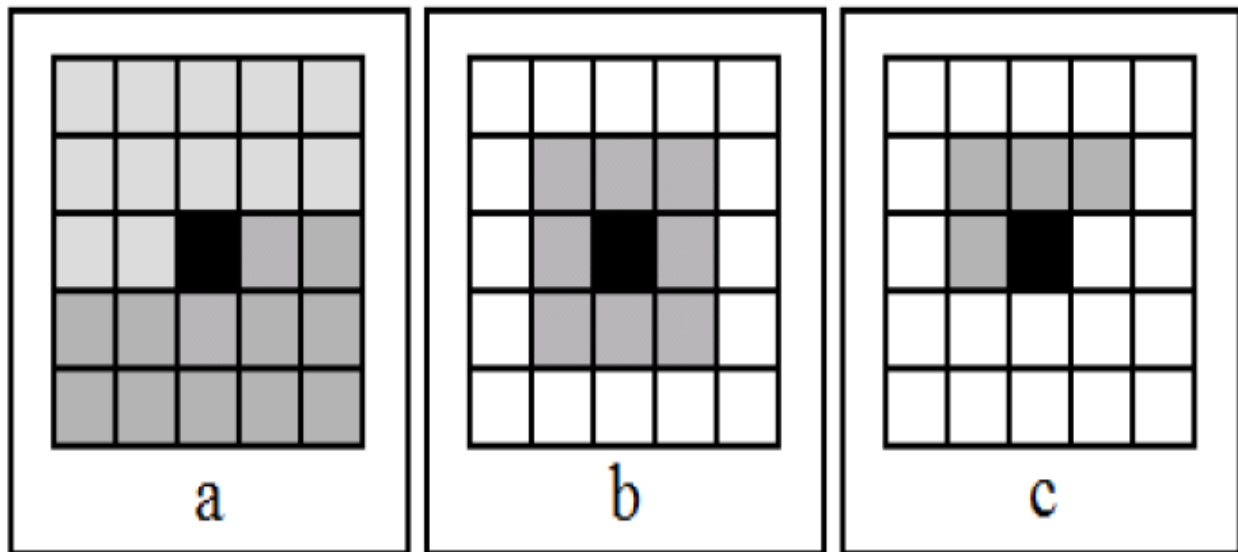


FIG. 3.1: Voisinage symétrique et non symétrique : (a) voisinage passé (gris clair) et futur (gris foncé). (b) voisinage symétrique de second ordre. (c) voisinage non symétrique (passé) de second ordre.

$$1) C = \{s\} \text{ i.e. } \text{card}\{c\} = 1,$$

$$2) \text{card}\{c\} > 1 \text{ et } (r, s) \in c \Rightarrow r \in \mathcal{V}_s \text{ et } s \in \mathcal{V}_r.$$

Exemple : les systèmes de voisinage les plus usuels, représentés en fig 3.2, sont :

- Système "4 - connexité", également appelé voisinage du premier ordre ;
- Système "8 - connexité", également appelé voisinage du second ordre.

Définition 35 (Champs aléatoires de Markov) : Un champ aléatoire $X = (X_s, s \in S)$

est dit markovien par rapport à \mathcal{V} ssi $\forall w \in \Omega, s \in S,$

$$p(X_s = w_s / X_r = w_r, r \in S - \{s\}) = p(X_s = w_s / X_r = w_r, r \in \mathcal{V}_s)$$

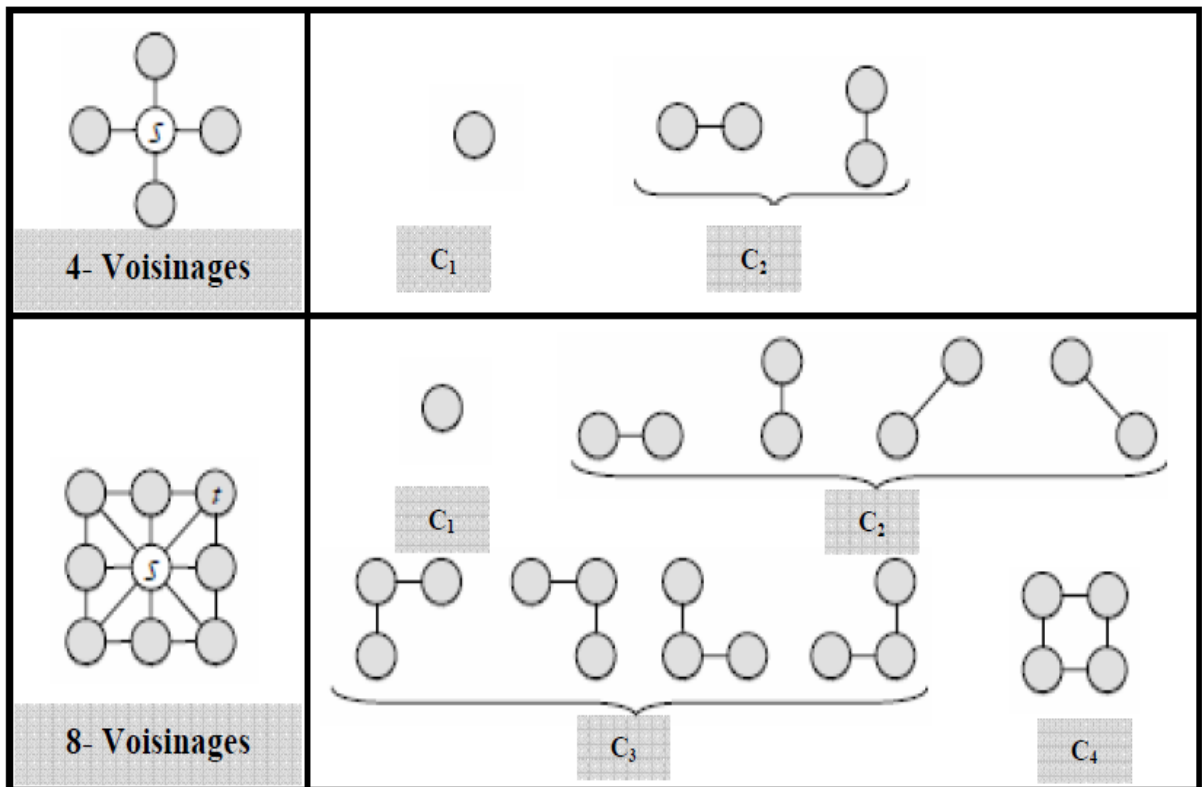
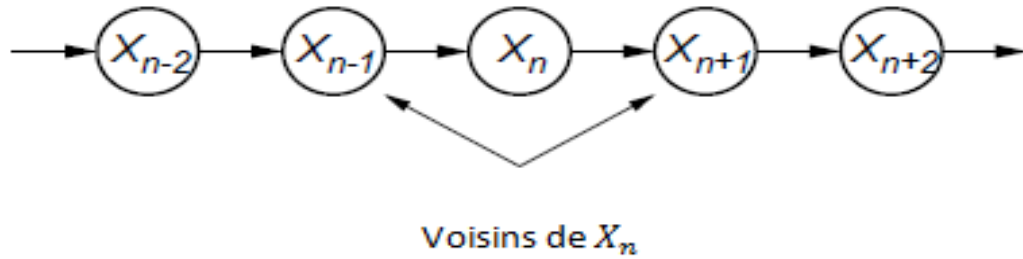


FIG. 3.2: Cliques associées à un système de voisinage en 4-connexité et en 8-connexité



Ainsi, les propriétés statistiques locales d'un champ de Markov au site s ne dépendent que de l'état d'un voisinage plus ou moins étendu défini autour de ce site et non de l'état du champ tout entier.

Dans la section suivante on étudions le cas des champs de Markov non causaux, ces modèles sont équivalents aux distributions de Gibbs GRF.

Exemple 36 *Les chaînes de Markov sont des MRF*

Les voisins de n sont $\mathcal{V}_n = \{n-1, n-2\}$

Les cliques ont la forme $c = \{n-1, n\}$

La densité a la forme

$$\begin{aligned} p(x) &= p(x_0) \prod_{n=1}^N p(x_n/x_{n-1}) \\ &= p(x_0) \exp \left\{ \sum_{n=1}^N \log p(x_n/x_{n-1}) \right\} \end{aligned}$$

Les fonctions potentielles ont la forme

$$V(x_n, x_{n-1}) = \log p(x_n/x_{n-1})$$

3.3 Champs aléatoires de Markov non causal (NMRF) :

Définition 37 *Un champ aléatoire $X = (X_s, s \in S)$ est dit markovien MRF non causal par rapport à \mathcal{V} ssi $\forall w \in \Omega, s \in S,$*

- 1) $p(X_s = w_s / X_r = w_r, r \in S - \{s\}) = p(X_s = w_s / X_r = w_r, r \in \mathcal{V}_s)$ (Markovianité),
- 2) $p(X = w) > 0$ (Contrainte de positivité)

le comportement en site s est totalement déterminé par la réalisation de ses voisins non causaux où chaque site n'est influencé que par ses sites voisins (notion d'interaction locale entre les sites). La propriété (1) signifie que la probabilité d'apparition d'une classe donnée en un site s conditionnellement au reste du champ, est égale à sa probabilité d'apparition connaissant simplement ses voisins. Utiliser la caractérisation (2) ne permet pas d'avoir une expression simple de la distribution jointe $P(X = w)$.

Notons que tous les champs positifs aléatoires peuvent être considérés comme markoviens, à condition de choisir un système de voisinage suffisamment grand [Li (1995)]. Cependant, l'intérêt de restreindre le voisinage est de permettre des temps de calculs exploitables.

Avantage :

- Dépendance locales uniquement,
- Comportement isotrope (pas de direction privilégiée).

Inconvénient :

- Probabilité $p(X)$ difficile à calculer \Rightarrow estimation des paramètres difficile.

La difficulté d'obtenir cette distribution a été facilement résolue par le théorème de Hammersley-Clifford (1971) établissant l'équivalence entre les champs de Markov et les champs de Gibbs.

3.3.1 Champs aléatoires de Gibbs

Définition 38 *Un champ aléatoire X sur S suit une distribution de Gibbs (GRF) si et seulement si il existe un système de voisinage \mathcal{V} sur S et une famille $\mathcal{V} = \{V_c, c \in C\}$ de potentiels sur les cliques du graphe (S, \mathcal{V}) tel que :*

$$P(X = x) = \frac{1}{Z} \exp(-U(x)) \quad (\text{Mesure de Gibbs})$$

avec

$$U(x) = \sum_{c \in C} V_c(x_c) \quad (\text{Fonction d'énergie})$$

et

$$Z = \sum_{X \in \Omega^{card S}} \exp(-U(x)) \quad (\text{Constante de normalisation, incalculable})$$

Une telle loi est dite une "distribution de Gibbs". Notons que V_c sont des applications de $\Omega^{card c}$ dans \mathbb{R} appelées "**fonction potentiel**" avec x_c la restriction de x à c et Z est une constante de normalisation, également appelée "**fonction de partition**" qui fait de $P(x)$ une probabilité. Etant donné que cette constante est une somme sur l'ensemble de toutes les configurations possibles du champ X , son calcul est dans la pratique généralement impossible.

Propriétés de la distribution de Gibbs :

- Un GRF est défini par une probabilité conjointe. Alors que, la définition de MRF est basée sur une probabilité conditionnelle.

- De l'analogie de la physique statistique, une température T est parfois incluse dans la distribution de Gibbs comme suit

$$P(X = x) = \frac{1}{Z} \exp\left(\frac{1}{T}U(x)\right)$$

Exemple 39 Nous allons définir l'énergie de l'image à partir des cliques, car un potentiel est associé à chaque clique dont la valeur dépend de l'état des sites qui la constituent.

L'énergie globale

U de l'image est donnée par :

$$U = \sum_{c \in C} U_c$$

et l'énergie locale en un site s dans l'image est :

$$U = \sum_{c \in C / s \in c} U_c$$

interprétée comme la somme des potentiels de toutes les cliques dont s fait parti.

3.3.2 Equivalence champ aléatoire de Markov – champ aléatoire de Gibbs

Comme nous l'avons évoqué précédemment, le champ aléatoire de Markov (*MRF*) est caractérisé par sa propriété locale (markovianité) tandis que le champ aléatoire de Gibbs (*GRF*) est caractérisé par sa propriété globale (distribution de Gibbs). Le théorème de Hammersley – Clifford (1971), dont une démonstration peut être vue dans [Bes 74], établit l'équivalence entre ces deux propriétés.

Théorème 40 (de Hammersley-Clifford) *Il y a équivalence entre les champs aléatoires de Markov et les champs aléatoires de Gibbs. X est un champ de Markov relativement à un système de voisinage \mathcal{V} ssi $P(X = x)$ est une distribution de Gibbs relativement à \mathcal{V} .*

Avantage :

- Fournit une méthode de calcul de $P(x)$.

Inconvénient :

- Les calculs de $P(x)$ sont trop importants du fait de $U(x)$ et Z , on doit donc passer par un échantillonnage.

Remarque :

- La factorisation est un lien entre la probabilité conjointe et les probabilités conditionnelles locales. Il montre que la probabilité conjointe d'un MRF peut être construite à partir de ses probabilités conditionnelles locales.
- La condition de positivité est requise pour la factorisation.

3.4 NMRF fondamentaux

Des nombreux modèles classiques sont des champs markoviens définis par leurs fonctions de potentiels et leur voisinage. Nous allons présenter deux exemples : le modèle d'Ising et le modèle de Potts, qui est la généralisation du premier pour un espace d'états plus large.

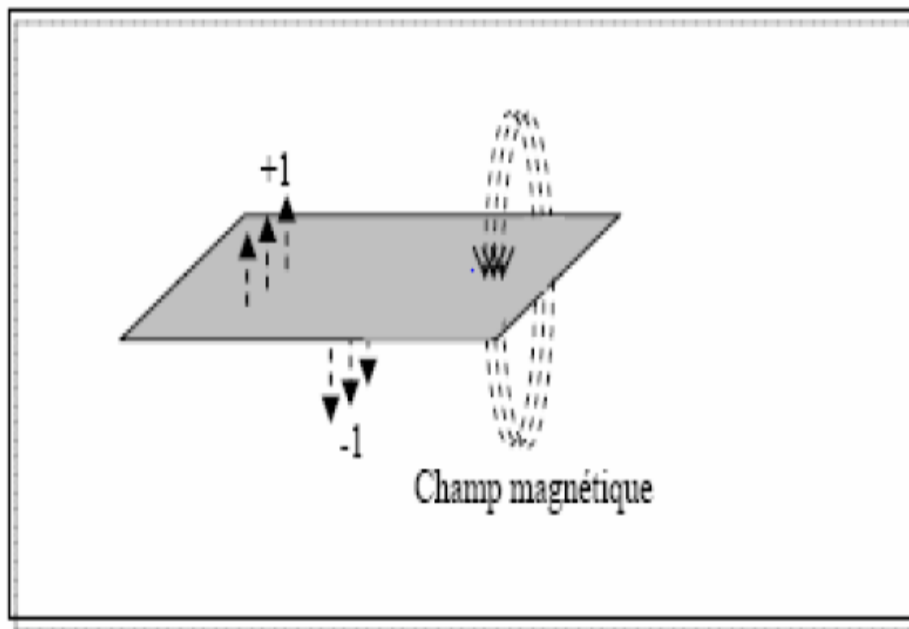


FIG. 3.3: Orientation d'un spin à l'état d'équilibre

3.4.1 Modèle d'Ising

Ce modèle, qui a été développé à partir de l'étude du ferromagnétisme en physique statistique [Isi 1925], utilise un espace binaire de descripteurs, correspondant à celui des spins $\Lambda = \{1, -1\}$ comme le montre 3.3

La fonction d'énergie pour ce modèle en tenant compte d'un voisinage 4 connexe est :

$$U_x = \sum_{c \in C_1} U_c(x_s) + \sum_{c \in C_2} U_c(x_s, x_t)$$

La clique C_1 est constituée d'un seul spin et le potentiel associé est de la forme $U_c(x_s) = -\alpha x_s$, où α est un champ magnétique externe. Le potentiel associé aux cliques

C_2 est donné par :

$$U_c(x_s, x_t) = \begin{cases} \beta & \text{si } x_t \neq x_s \\ 0 & \text{si } x_t = x_s \end{cases}$$

où β est la constante de couplage entre s et t . l'énergie totale s'écrit alors,

$$U_x = - \sum_{c \in C_1} \alpha x_s - \sum_{c \in C_2} \beta \cdot x_s, x_t$$

β régularise le modèle : lorsque $\beta > 0$, les configurations d'énergie minimale sont celles pour lesquelles les sites ont le même état, et si $\beta < 0$ les configurations privilégiées sont celles où les sites sont d'états opposés.

3.4.2 Modèle de Potts

Le modèle de Potts, [Wu (1982)], est une extension du modèle d'Ising, pour un espace de n valeurs $\{1, \dots, n\}$, et la topologie est indifféremment la 4- ou la 8-connexité dans une image

2D. Seuls les potentiels d'ordre 2 sont définis par :

$$U_c(x_s, x_t) = -\beta x_s x_t \begin{cases} \beta & \text{si } x_t \neq x_s \\ 0 & \text{si } x_t = x_s \end{cases}$$

Il est très utilisé pour la classification d'images. Dans [Sir 1993], les auteurs utilisent le modèle de Potts pour une classification en tenant compte des différentes valeurs de β . Lorsque $\beta > 0$, les configurations les plus probables correspondent à des sites voisins ayant des descripteurs égaux, ce qui dans le cas du traitement d'image donne des réalisations avec de grandes zones homogènes (au sens de la similarité entre descripteurs). La taille de ces régions est gouvernée par la valeur de la constante de couplage.

Les modèles d'écrits jusqu'à présent font partie de la famille de champ de Markov à voisinage symétrique, ou encore non causaux, puisque $j \in V_i \Leftrightarrow i \in V_j$. La difficulté des modèles non causaux réside dans la difficulté du calcul de la loi jointe $P(x)$, et ce même pour un voisinage très simple. Les champs de Markov causaux (Markov mesh models) sont construits de sorte à contourner cette difficulté. Le graphe de voisinage est supposé acyclique et orienté, la relation de voisinage n'est donc plus symétrique. La distribution du champ au pixel i ne dépend alors que d'un petit nombre des prédécesseurs de i . Il s'en suit que la loi jointe se décompose en produit de probabilités conditionnelles facilement calculables, comme dans le cas d'une chaîne de Markov .

3.5 Champ aléatoires de Markov causal (CMRF) en analyse d'image

les modèles *MRF* causaux c'est une extension aux modèles de markov cachés $1-D$ (*HMM*), dans le contexte d'un modèle de Markov caché $1-D$ (*HMM*), une observation Y_t à l'instant t est supposée être générée par un processus dont l'état X_t est caché à l'observateur. Suivant cette terminologie, un *MRF* causal est également appelé un champ aléatoire de Markov caché (*HMRF*). le principal obstacle à cette extension est le manque d'ordre naturel pour une grille bidimensionnelle et donc l'absence d'une notion naturelle de causalité dans les données d'image spatiale. En conséquence, un ordre artificiel des données d'image doit être supposé.

3.5.1 Transformation une image 2D en une image 1D

Dans le cadre des chaînes de Markov, une image $2D$ peut être facilement transformée en une image $1D$. Pour appliquer les techniques développées pour les HMM $1 - D$ aux données d'image $2D$, il est d'abord nécessaire de convertir les données d'image $2D$ en une séquence de données $1D$ ordonnée. Puisqu'il n'y a pas d'ordre causal défini naturellement pour les index des données d'image $2D$, les relations spatiales entre les pixels voisins peuvent être perdues par une telle réorganisation. Le moyen le plus simple de convertir des données d'image $2D$ en une séquence ordonnée $1D$ consiste à utiliser une méthode de balayage : commencez à partir d'un pixel de l'image et visitez l'un de ses pixels voisins les plus proches et précédemment non visités successivement jusqu'à ce que tous les pixels de n soient visités . Ensuite, les modèles de chaînes de Markov $1D$ peuvent être appliqués à la séquence ordonnée. Cependant, à la suite du balayage, la plupart des relations spatiales entre les pixels voisins, à l'exception des deux éléments consécutifs dans la séquence ordonnée, sont ignorées. Trois des nombreuses méthodes de numérisation possibles sont illustrées à la figure 3.1. Pour diverses méthodes d'analyse, les critères suivants peuvent être utilisés pour évaluer leurs performances :

- i) La séquence $1D$ numérisée doit être construite de manière à ce que les pixels spatialement proches soient aussi proches que possible,
- ii) Les directions de numérisation doivent être aussi diverses que possible.

Méthode balayage 1 le parcourt de ligne par ligne ou colonne par colonne, présentée dans la Figure 3.4

Méthode balayage 2 le parcourt de Hilbert-Peano (Pappas (1992), Peano (1990)),

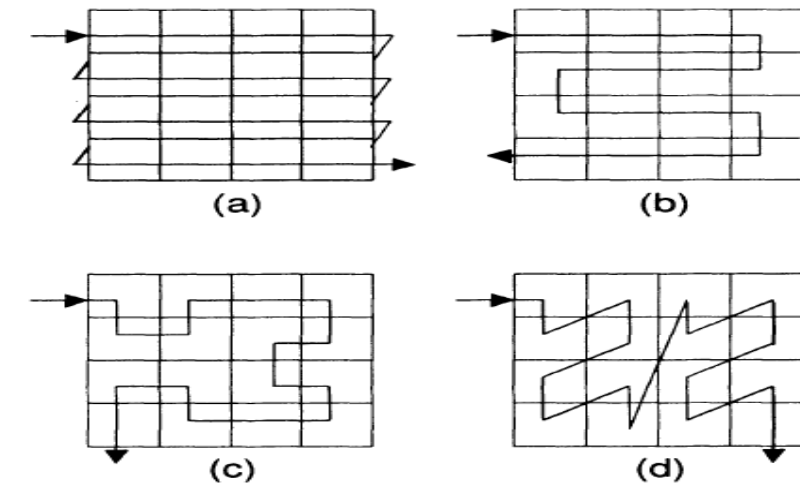


FIG. 3.4: Méthodes de numérisation : (a) balayage raster, (b) balayage en zigzag, (c) balayage Peano, et (d) ordre Morton.

présentée dans la Figure 3.4 (c).

Méthode balayage 3 l'ordre de Morton [Morton (1966)] (également appelé un Z- courbe), présentée dans la Figure 3.4 (d)

Un instant de réflexion révèle également que le scan Peano est meilleur que l'ordre Morton en ce sens qu'il n'y a pas de saut dans le scan Peano. Autrement dit, dans le balayage Peano, deux pixels adjacents sur la courbe sont également adjacents dans l'espace.

La suite du balayage, nous obtenons une séquence de variables aléatoires pour les étiquettes de classe $X_{(1 \rightarrow T)} = \{X_1, X_2, \dots, X_T\}$ et les variables aléatoires pour les données observées $Y_{(1 \rightarrow T)} = \{Y_1, Y_2, \dots, Y_T\}$ où $T = |\Omega|$. Ici, $Y_t = y_t$ peut être un niveau de gris ou un vecteur de caractéristiques pour l'étiquette de classe sous-jacente $X_t = x_t$ et l'étiquette de classe $x_t \in \{0, \dots, L_X - 1\}$ est une réalisation d'une variable aléatoire cachée X_t . Puisque ces séquences peuvent être considérées comme des séquences ordonnées dans le temps, les algorithmes pour les chaînes de Markov 1D sont directement applicables. Nous commençons

par plusieurs hypothèses :

Hypothèses :

(i) La suite $X_{(1 \rightarrow t)} = \{X_1, X_2, \dots, X_t\}$, $t \leq T$ est une chaîne de Markov homogène du premier ordre à temps discret, de sorte que

$$P(x_1, x_2, \dots, x_t) = P(X_1 = x_1) \prod_{i=2}^t P(X_i = x_i / X_{i-1} = x_{i-1})$$

(ii) Etant donné une séquence d'observation $Y_{(1 \rightarrow T)}$, on suppose qu'à chaque pas de temps $t \leq T$, $Y_t = y_t$ a été généré à partir d'une variable aléatoire cachée $X_t = x_t$ et que $\{Y_1, Y_2, \dots, Y_T\}$ sont conditionnellement indépendants des étiquettes de classe. Cela implique la factorisation suivante :

$$P(y_1, \dots, y_t / x_1, \dots, x_t) = \prod_{i=2}^t P(y_i / x_i), \quad t \leq T$$

3.5.2 Modèles de Markov cachés pseudo 2D

Une autre façon d'étendre les modèles de chaîne de Markov 1D aux données d'image 2D consiste à utiliser des pseudo modèles de Markov cachés 2D (*PHMM 2D*), qui contiennent des 1D *HMM* imbriqués [Eickeler (2002), Kuo et Agazzi (1994), Lin et Wang et Yang (1997), Nefian et Hayes (2000)]. Dans le *PHMM 2D*, on commence par un 1D *HMM* vertical de haut en bas avec des "superclasses" (ou "super-états"). Ensuite, sur chaque ligne des données d'image, un *HMM* horizontal 1D de gauche à droite est construit. Comme le montre la figure 3.5, une variable aléatoire X_i^s est adoptée pour la i ème ligne des données d'image avec $i = 0, \dots, N_1 - 1$, représentant l'une des L^s superclasses (c'est-à-dire $x_i^s \in \{0, 1, \dots, L^s\}$). Ensuite, nous avons une collection de variables aléatoires de haut en bas

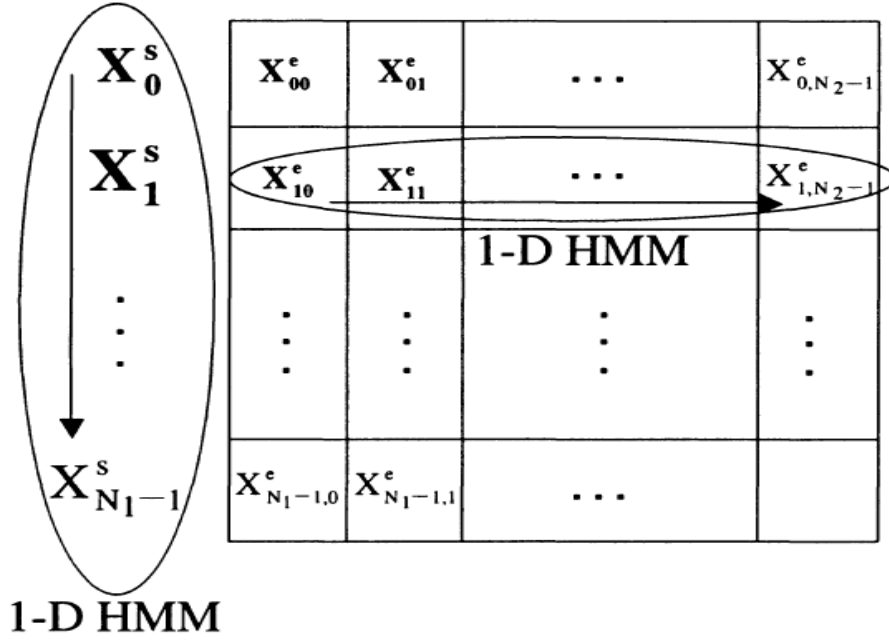


FIG. 3.5: Structure de PHMM 2D.

pour les superclasses $\mathbf{X}_{(0 \rightarrow N_1-1)}^s$ et l'observable correspondante données d'image (ou valeurs de caractéristiques) $\mathbf{Y}_{(0 \rightarrow N_1-1)}^s$:

$$\mathbf{X}_{(0 \rightarrow N_1-1)}^s = \{X_0^s, X_1^s, \dots, X_{N_1-1}^s\}$$

$$\mathbf{Y}_{(0 \rightarrow N_1-1)}^s = \{Y_0^s, Y_1^s, \dots, Y_{N_1-1}^s\}$$

3.5.3 Définition CMRF

La dépendance causale du champ aléatoire de Markov peut être définie en utilisant des ensembles de sites causaux dans une image comme suit :

$$S_{i,j}^a = \{(k,l) : k < i \text{ ou } l < j\}, S_{i,j}^b = \{(k,l) : k \leq i \text{ et } l \leq j, (k,l) \neq (i,j)\}$$

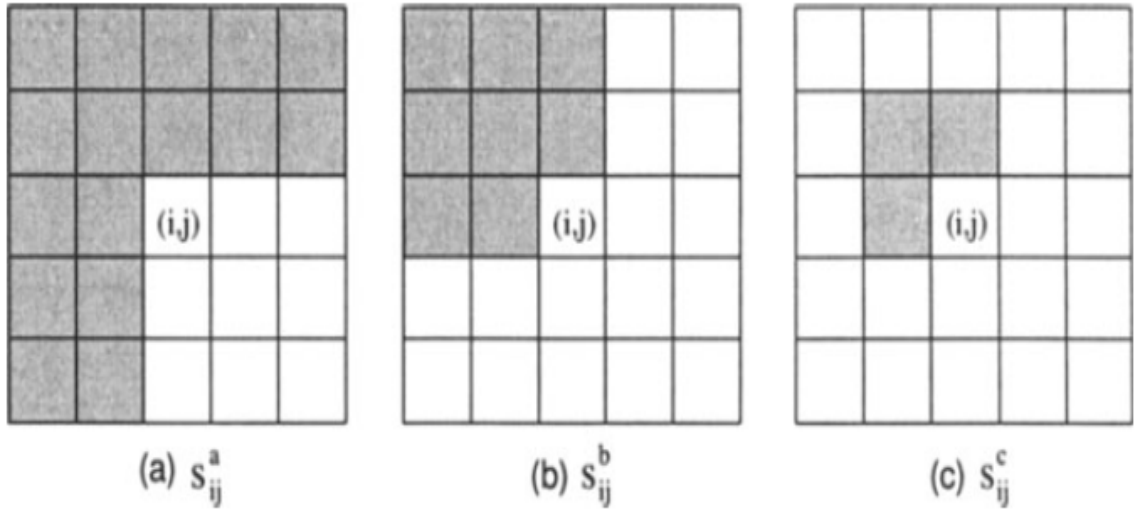


FIG. 3.6: Les sites ombrés sont (a) passé, (b) support causal et (c) voisins causaux associés aux modèles MMRF du troisième ordre.

Ici, $S_{i,j}^a$ est considéré comme le "passé" ou un ensemble de "prédécesseurs" d'une variable aléatoire $X_{i,j}$ à un pixel $(i, j) \in S$

Définition 41 *Un champ aléatoire $X = \{X_{i,j}, (i, j) \in S\}$ est dit markovien MRF causal ssi*

$$\forall x \in \Omega, P(X_{i,j} = x_{i,j} / X_{l,k} = x_{l,k}, (l, k) \in A_{i,j}) = P(X_{i,j} = x_{i,j} / X_{l,k} = x_{l,k}, (l, k) \in B_{i,j})$$

avec $B_{i,j} \subseteq A_{i,j} \subseteq \{(k, l) : k < i \text{ ou } l < j\}$. Un exemple commun est $S_{i,j}^c = B_{i,j}$.

Puisque $S_{i,j}^c \subset S_{i,j}^b \subset S_{i,j}^a$, alors

$$P(X_{i,j} = x_{i,j} / X_{l,k} = x_{l,k}, (l, k) \in S_{i,j}^a) = P(X_{i,j} = x_{i,j} / X_{l,k} = x_{l,k}, (l, k) \in S_{i,j}^c) \quad (3.1)$$

Avantage :

- Les modèles de Markov cachés $1D$ (*HMM*) sont étendus aux modèles $2D$ *MRF* Causal.

- la possibilité d'un traitement en ligne à complexité réduite pour les données d'image en $2D$.

- Disponibilité d'une variété d'outils utiles développés pour les problèmes de série Markov $1D$ tels que la reconnaissance vocale.

- Dans le cas d'une transmission d'image séquentielle, les données d'image reçues peuvent être traitées immédiatement sans attendre l'arrivée d'autres données dans l'espace d'image.

Inconvénient :

- l'absence d'un ordre naturel pour une grille bidimensionnelle et donc l'absence d'une notion naturelle de causalité dans les données d'image spatiale.

- Utiliser un arrangement artificiel plutôt qu'un ordre normal.

Remarque 42 *La distribution conjointe du MRF causal pour un support causal donné peut être calculée par une distribution conjointe avec un sous-ensemble du support causal, donnant une structure récursive.*

3.6 Modèles de Markov causaux $2D$ fondamentaux

La séquence causale obtenue par une méthode de balayage dans la section précédente exige des sacrifices importants en termes de perte d'information spatiale et structurale. Il est important de conserver autant que possible les relations spatiales entre les pixels voisins lorsqu'ils sont convertis en une séquence ordonnée $1D$. Afin d'atteindre cet

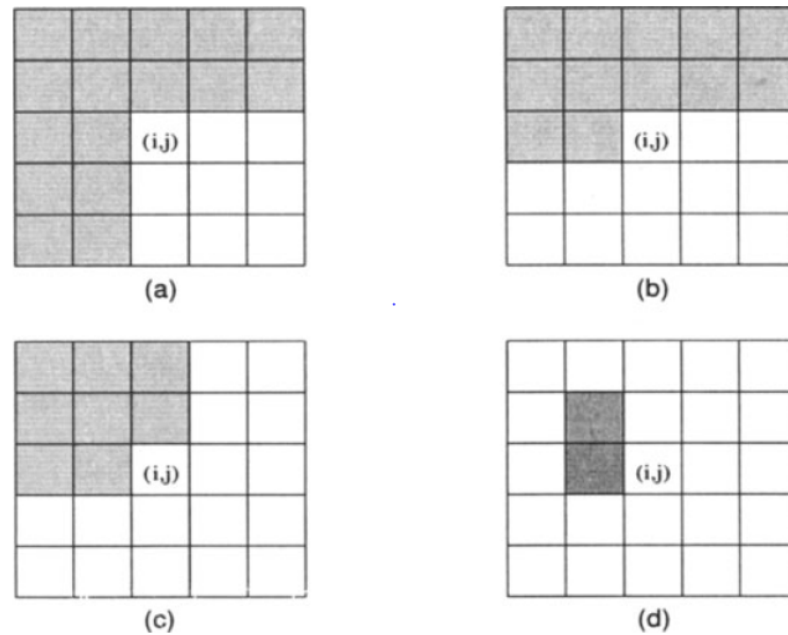


FIG. 3.7: (a) $MMRF$, (b) chaîne de Markov $NHSP$, (c) MRF unilatéral, et (d) champ aléatoire de Pickard.

objectif, des voisins causaux pour chaque pixel sont définis dans le contexte de la modélisation MRF causale. Plus précisément, la probabilité conditionnelle d'un pixel d'image étant donné le «passé» ne dépend que d'un sous-ensemble du «passé». Ici, avec une variété de dispositions pour la définition du «passé», un certain nombre de modèles causaux différents peuvent être définis [Derin et Kelly (1989)] comme un champ aléatoire de maillage de Markov (Markov Mesh Random Field) ($MMRF$) [Abend, Harley et Kanal (1965), Derin, Elliott, Cristi et Geman (1984), Kana (1980)], un demi-plan ($NSHP$) chaîne de Markov [Jeng et Woods (1987), Jeng et Woods (1991)], un MRF unilatéral (aussi appelé un MRF causal) [Chellappa (1985)]. Comme l'illustre la figure 3.6,

Seul le modèle $MMRF$ et $UMRF$ sera étudié dans cette section.

3.6.1 Champ aléatoire de maillage de Markov

On dit que le champ aléatoire $X = \{X_{i,j}, 1 \leq i \leq m \ \& \ 1 \leq j \leq n\}$ est un *MMRF* si

$$p(x_{i,j}/Z_{i,j}) = p(x_{i,j}/N_{i,j}), \quad 1 \leq i \leq m, 1 \leq j \leq n$$

où $Z_{i,j} = \{X_{a,b}, a < i \text{ ou } b < j\}$ est considéré comme les prédécesseurs de $X_{i,j}$ dans le champ aléatoire \mathbf{X} , et indiqué sur la figure 1 (a). $N_{i,j} = \{X_{a,b}, a < i \ \& \ b < j\}$ est les voisins unilatéraux de $X_{i,j}$. Abend (1965) a prouvé que $P(x)$ pour un *MMRF* de taille voisine de 2 est de la forme suivante :

$$p(x_{i,j}/Z_{i,j}) = \prod_{i=1}^m \prod_{j=1}^n p(x_{i,j}/x_{i-1,j}, x_{i,j-1}), \quad 1 \leq i \leq m, 1 \leq j \leq n$$

où l'ajustement de limite requis est réalisé en supposant que les sites à l'extérieur du réseau fini, $i = 1, j = 2$ sont égaux à zéro. Abend a prouvé que *MMRF* est également *MRF* alors que l'inverse ne tient pas nécessairement, c'est-à-dire que les *MMRF* forment une sous-classe causale de *MRF*.

D'après la définition d'un *MMRF*, la distribution conjointe de $X_{S_{i,j}^b}$ peut être exprimé comme un produit des probabilités de transition conditionnelle :

$$P(x_{S_{i,j}^b}) = \prod_{(k,l) \in S_{i,j}^b} P(x_{k,l}/x_{S_{kl}^c})$$

En exploitant cette propriété, la probabilité conjointe $P(x)$ d'un X (*MMRF*) peut être prise en compte dans un produit de probabilités de transition conditionnelle comme

$$P(x) = \prod_{(k,l) \in S_{N_1-1, N_2-1}^b} P(x_{kl}/x_{S_{kl}^c}) \quad (3.2)$$

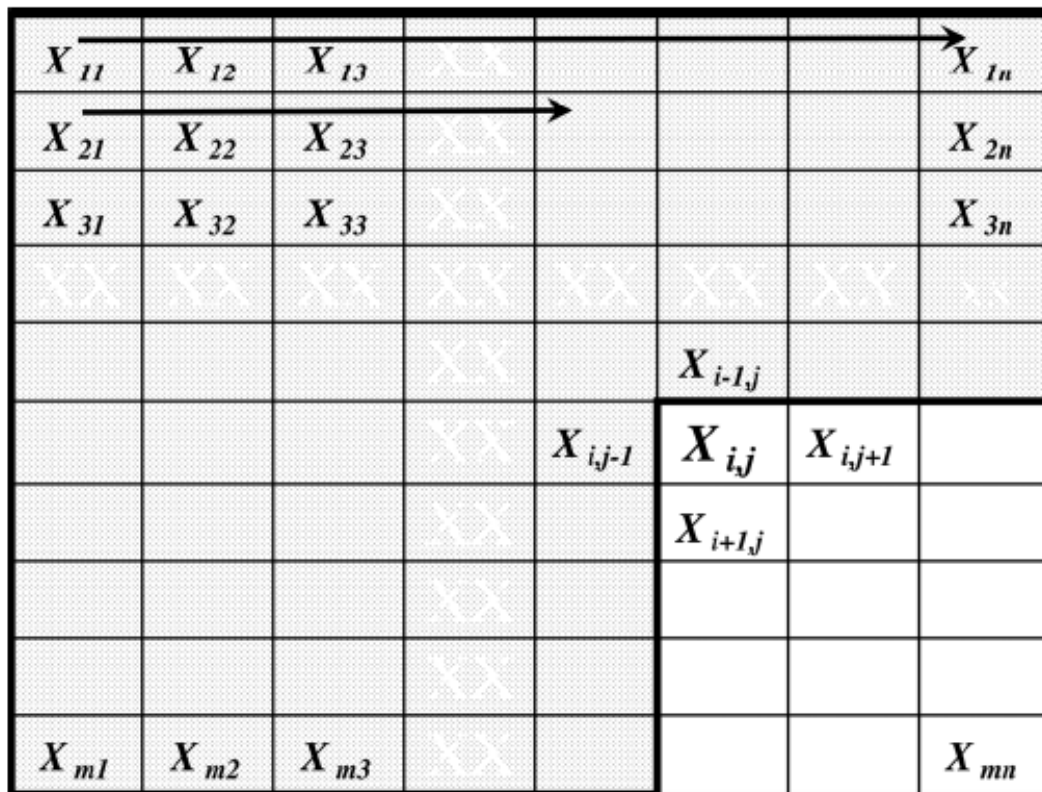


FIG. 3.8: Région rectangulaire régulier X de taille $m \times n$: $Z_{i,j}$ est la zone grise, les flèches indiquent le balayage raster

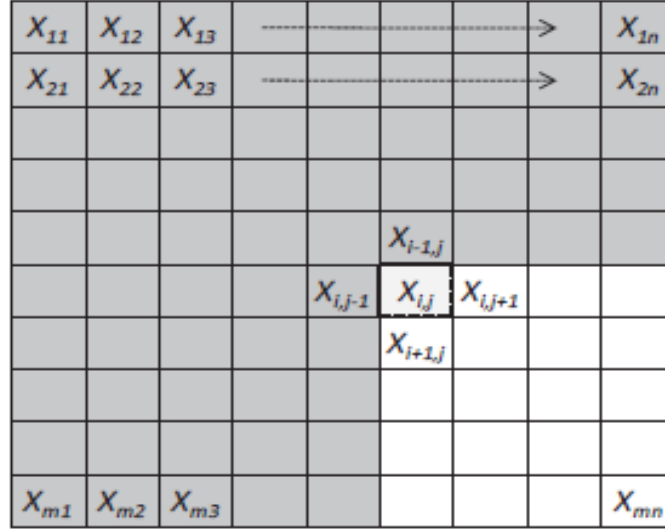


FIG. 3.9: Configuration spatiale des pixels dans un MRF unilatéral.

où $S_{N_1-1, N_2-1}^b = \Omega$. La factorisation de $P(x)$ dans 3.2 peut être vérifiée en appliquant la règle de chaîne dans le sens inverse du balayage raster et par raster et en utilisant la définition de la *MMRF* dans 3.1 :

$$\begin{aligned}
 P(x) &= P(x_{N_1-1, N_2-1}, x_{N_1-1, N_2-1-}) \\
 &= P(x_{N_1-1, N_2-1} / x_{N_1-1, N_2-1-}) P(x_{N_1-1, N_2-1-}) \\
 &= P\left(x_{N_1-1, N_2-1} / x_{S_{N_1-1, N_2-1}^c}\right) P\left(x_{N_1-1, N_2-2} / x_{S_{N_1-1, N_2-2}^c}\right) \dots P(x_{00}). \quad (3.3)
 \end{aligned}$$

où $x_{ij-} = \{x_{kl}, \forall (k, l) \in \Omega, (k, l) \neq (i, j)\}$.

3.6.2 MRF unilatéral

.Maintenant, définissons les MRF unilatéraux

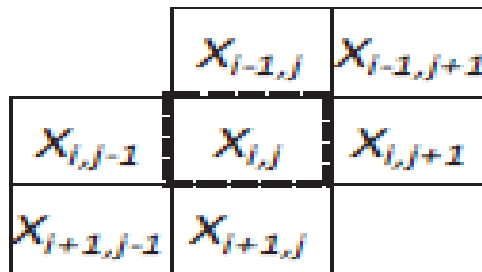


FIG. 3.10: Configuration spatiale des pixels dans un MRF unilatéral.

Le champ aléatoire X est défini comme un *MRF* unilatéral de voisinageis U_{ij} sur un réseau rectangulaire si :

$$P(x_{ij}/Z_{m,n}^{i,j}) = P(x_{ij}/x_{i-1,j}, x_{i,j-1})$$

où $Z_{m,n}^{i,j}$ est la zone grise illustrée à la figure 3. 9 et la définition est écrite en fonction de l'ordre deux. Abend (1965) a prouvé que sur la base du MRF unilatéral définition la densité de joint du réseau peut être dérivée comme suit :

$$P(X) = \prod_{i=1}^m \prod_{j=1}^n P(x_{ij}/x_{i-1,j}, x_{i,j-1})$$

Abend (1965) a également prouvé que

$$P(x_{ij}/X - \{x_{i,j}\}) = P(x_{ij}/x_{i-1,j}, x_{i-1,j+1}, x_{i,j-1}, x_{i,j+1}, x_{i+1,j-1}, x_{i+1,j}) \quad (3.4)$$

Selon l'équation (3.4), la configuration voisine résultante est illustré à la 3.10 .

Pour le moment, présentons d'autres sous-classes unilatérales de *MRF*.

A- Demi-plan non symétrique

Preuss a défini une autre sous-classe causale de MRF appelée chaîne de Markov demi-plan non symétrique (*NSHP*) [Preuss 1975]. La seule différence réside dans la définition des prédécesseurs du site $X_{i,j}$ dans lesquels

$$Z_{i,j} = \{X_{a,b/a} < i \text{ ou } (a = i \ \& \ b < j)\}$$

La relation entre *MMRF* et *NSHP* est discutée en détail dans [Jeng et Woods 1987]. Il n'y a pas de différence entre ces deux définitions, et il est facilement démontré que *MMRF* est en effet équivalent à la chaîne de Markov *NSHP*.

B- Champ aléatoire de Pickard

Une autre sous-classe unilatérale intéressante de *MRF* est le champ aléatoire de Pickard (1977) (*PRF*). Pickard a construit un treillis $n \times m$ par une cellule générique $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ayant la propriété suivante :

$$P(b/a, c) = P(b/a)$$

L'indépendance conditionnelle est également représentée par la notation abrégée ; $b \perp c/a$. Pickard (1977) montre que le champ aléatoire construit est *MMRF* et par conséquent *MRF*. Il est en outre prouvé que *PRF* est la seule classe reconnue de champ de Markov stationnaire non trivial sur des réseaux rectangulaires finis [Champagnat 1998].

C- Champ aléatoire Gibbs mutuellement compatible

Le champ aléatoire de Gibbs mutuellement compatible ($MC - GRF$) est une autre sous-classe causale de MRF [Goutsias 1989]. A (GRF) $\mathbf{X} = \{X_{i,j}, 1 \leq i \leq m, 1 \leq j \leq n\}$ est dit $MC - GRF$ si tous les sous-réseaux primaires de \mathbf{A} définis sur \mathbf{X} sont également GRF . Le sous-réseau principal \mathbf{A} est défini comme un ensemble de sites qui inclut X_{11} , et a en outre la propriété que s'il contient $X_{i,j}$, il contient également $\{X_{i-1,j}, X_{i,j-1}, X_{i-1,j-1}\}$. En termes simples, un sous-réseau principal est un ensemble de pixels qui comprend le coin supérieur gauche et contient également tous les voisins causaux immédiats de tous ses membres. Goutsias a prouvé que $MC - GRF$ et $MMRF$ sont équivalents [Goutsias 1989]. Cependant, en raison de la différence de forme algébrique des deux définitions, les contraintes sur $MC - GRF$ sont données sur la base de la fonction de potentiel global de Gibbs alors que dans $MMRF$, elle est basée sur une structure locale voisine. L'utilisation de la structure voisine locale pour analyser le comportement MRF dans $MC - GRF$ est plus faisable car il est plus facile d'appliquer ces conditions sur la fonction de Gibbs.

D- Champ aléatoire de Markov quadrilatère

Le champ aléatoire $\mathbf{X} = \{X_{i,j}, 1 \leq i \leq m, 1 \leq j \leq n\}$ est $QMRF$ avec la taille voisine de deux sur un réseau rectangulaire fini si

$$p\left(x_{i,j}/Z_{i,j}^{ul}\right) = p\left(x_{i,j}/x_{i-1,j}, x_{i,j-1}\right), 1 \leq i \leq m, 1 \leq j \leq n$$

$$p\left(x_{i,j}/Z_{i,j}^{dr}\right) = p\left(x_{i,j}/x_{i+1,j}, x_{i,j+1}\right), 1 \leq i \leq m, 1 \leq j \leq n$$

$$p\left(x_{i,j}/Z_{i,j}^{ur}\right) = p\left(x_{i,j}/x_{i-1,j}, x_{i,j+1}\right), 1 \leq i \leq m, 1 \leq j \leq n$$

$$p\left(x_{i,j}/Z_{i,j}^{dl}\right) = p\left(x_{i,j}/x_{i+1,j}, x_{i,j-1}\right), 1 \leq i \leq m, 1 \leq j \leq n$$

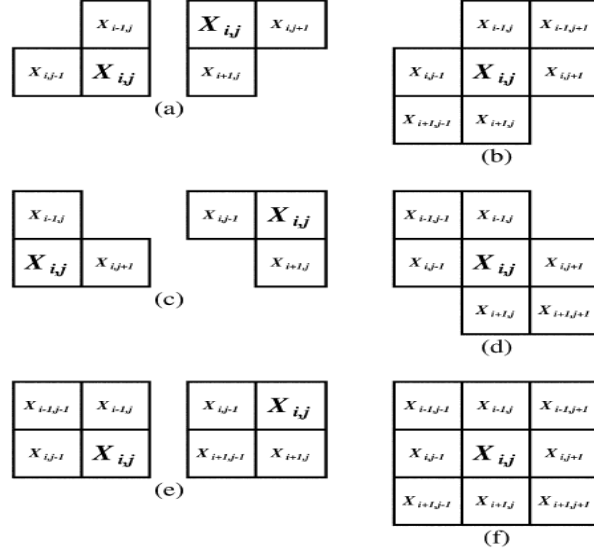


FIG. 3.11: Différentes structures voisines

où $Z_{i,j}^{ul} = \{X_{a,b}, a < i \text{ ou } b < j\}$ et $Z_{i,j}^{dr} = \{X_{a,b}, a > i \text{ ou } b > j\}$ sont représentés sur les figures 1 (a) et (b) et les deux autres prédécesseurs sont $Z_{i,j}^{ur} = \{X_{a,b}, a < i \text{ ou } b > j\}$, et

$Z_{i,j}^{dl} = \{X_{a,b}, a > i \text{ ou } b < j\}$. Pour simplifier et mettre l'accent sur l'anomalie *UMRF*, la définition *QMRF* est donnée pour un système voisin de taille deux, cependant, elle est simplement extensible à un ordre supérieur de *QMRF*.

Il convient également de souligner que les résultats de la simulation indiquent que si \mathbf{X} est *UMRF* par rapport au coin supérieur gauche du treillis, il est également *UMRF* par rapport au coin inférieur droit, et de la même manière si \mathbf{X} est *UMRF* par rapport au coin supérieur droit, c'est *UMRF* par rapport au coin inférieur gauche. Cependant, la preuve mathématique reste un problème ouvert.

- (a) *UMRF* de taille 2 par rapport au coin supérieur gauche et inférieur droit ;
 (b) *MRF* résultant des *UMRF* indiqués en (a) ; (c) *UMRF* de taille 2 par rapport au coin supérieur droit et inférieur gauche ; (d) *MRF* résultant des *UMRF* indiqués en (c) ;
 (e) *UMRF* de taille 3 par rapport au coin supérieur gauche et inférieur gauche ; (f) *MRF* résultant de l'*UMRF* de taille 3.

3.7 Modélisation markovienne de la segmentation d'images

Le contenu d'une image constitue le centre d'intérêt et la première source d'informations pour plusieurs applications en analyse d'images. Il s'agit de le décrire quantitativement au moyen de méthodes et outils appropriés. Parmi les étapes d'un processus d'analyse d'image, figure l'étape de segmentation comme étant la plus importante dans la chaîne. En effet, c'est à partir de l'image segmentée que les paramètres et les propriétés mathématiques sont mesurés. Il existe deux types de segmentation : Une segmentation supervisée où on connaît les paramètres de l'image à segmenter et une segmentation non supervisée dans le cas contraire. Ce deuxième type de segmentation est délicat, en effet, il n'existe pas de solutions complètes jusqu'à ce moment. La plupart des algorithmes de segmentation utilisés sont orientés vers l'analyse statistique de l'image observée.

Dans ce cadre, plusieurs approches et modèles sont proposés pour la résolution du problème, l'approche Bayésienne apparaît comme étant la plus intéressante. L'objet de cette section est de présenter la modélisation par les champs de Markov cachés (*HMM*), nous commençons par la définition de ce modèle, ainsi le graphe d'indépendance et champs

de Markov cachés à bruit indépendant. Nous présentons ensuite l'application de ce modèle en segmentation d'image.

3.7.1 Segmentation

La segmentation est un outil puissant qui aide à la décision. Elle permet de remonter à une réalisation de X (image cachée) à partir de l'observation des données bruitées Y . On parle dans ce contexte de champ de Markov Caché pour X ou de données incomplètes puisque Y n'est pas une réalisation de X . La segmentation d'image par le modèle de Markov Caché nécessite 3 étapes : Modélisation, Estimation, Classification.

3.7.2 Champs de Markov cachés HMRF

Soit S un ensemble de pixels de cardinal N ($N = \text{Card}(S)$), et $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$ deux champs aléatoires, l'un caché et l'autre observé. Chaque variable aléatoire X_s est à valeurs dans l'ensemble fini de classes $\Omega = \{w_1, \dots, w_k\}$ et chaque variable aléatoire Y_s est à valeurs dans \mathbb{R} . En segmentation d'images, l'objectif est de retrouver le champ inobservé X , à partir des données observées Y qui représentent la version bruitée du champ X . A cet effet, nous devons définir la loi de (X, Y) , ce qui est classiquement obtenu à partir de la loi de X et celle de Y conditionnelle à X .

Dans le cas de la modélisation par champs de Markov cachés, X est supposé markovien. Sa loi est alors donnée par :

$$P(x) = \gamma \exp(-U(x)) \tag{3.5}$$

avec γ une constante de normalisation, appelée également "fonction de partition", et U "la fonction énergie". Celle-ci se décompose en une somme de fonctions φ_c , appelées "fonctions de potentiel", associées à l'ensemble des cliques. Ainsi, $U(x)$ s'écrit :

$$U(x) = \sum_{c \in C} \varphi_c(x_c)_s \quad (3.6)$$

La fonction de partition γ est généralement impossible à calculer du fait du nombre trop important des réalisations possibles du champ X , cependant, il reste possible de simuler une réalisation de ce champ à partir de la loi conditionnelle de X_s sachant X_{V_s} .

$$P(x_s/x_{V_s}) = \frac{\exp - \sum_{c \in C} \varphi_c(x_s, x_{c-\{s\}})}{\sum_{w \in \Omega} \exp - \sum_{c \in C} \varphi_c(x_s = w, x_{c-\{s\}})}$$

Ainsi, dans le cas des champs de Markov cachés, le champ inobservé X est supposé markovien, ce qui justifie l'appellation *CMC*. La loi de X étant définie, il reste à définir la loi conditionnelle $p(y/x)$

3.7.3 Le graphe d'un champ de Markov cachée

La figure 3.4 montre le graphe d'indépendance associé aux champs de Markov cachés usuellement définis sur une grille bidimensionnelle. Les variables aléatoires X_s forment un champ de Markov caché .

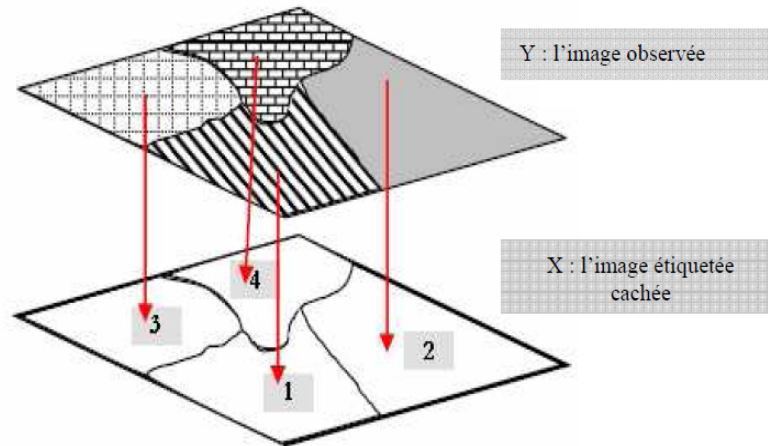


FIG. 3.12: segmentation d'images.

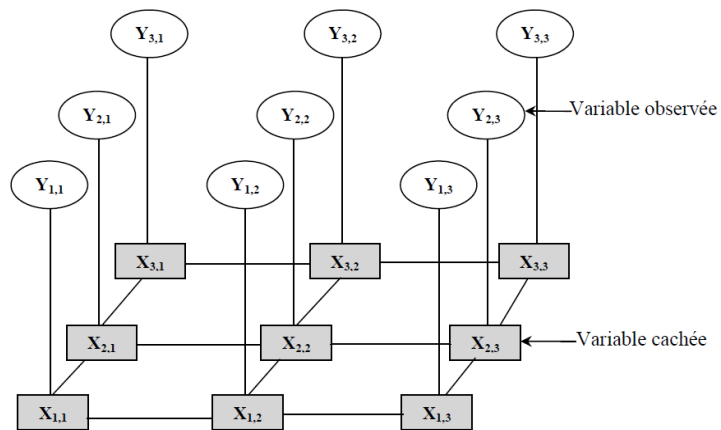


FIG. 3.13: La structure d'un modèle de champ de Markov caché

3.7.4 Champs de Markov cachés à bruit indépendant

Les champs de Markov cachés à bruit indépendant (*CMC – BI*) sont la modélisation la plus simple des *CMC*. La probabilité conditionnelle $p(y/x)$ vérifie les deux hypothèses

suivantes :

- **H1** : $P(y_s/x) = P(y_s/x_s)$ pour chaque $s \in S$
- **H2** : $P(y_s/x) = P(y_s/x_s)$

La deuxième hypothèse signifie que les Y_s sont indépendants conditionnellement à X .

Du fait de H1 et H2, la loi de $p(y/x)$ s'écrit sous la forme :

$$\begin{aligned} p(x/y) &= \prod_{s \in S} p(x_s/y_s) \\ &= \exp \left[\sum_{s \in S} \log(p(x_s/y_s)) \right] \end{aligned} \tag{3.7}$$

La loi du couple (X, Y) résulte des formules (3.5), (3.6) et (3.7), elle s'écrit comme suit :

$$\begin{aligned} p(x/y) &= p(x)p(y/x) \\ &= \gamma \exp \left[- \sum_c \varphi_{c \in C}(x_c) + \sum_{s \in S} \log(p(x_s/y_s)) \right] \end{aligned}$$

La probabilité a posteriori $p(x/y)$ du champ X est alors donnée par :

$$\begin{aligned}
p(x/y) &= \gamma \exp(-U(x, y)) \\
&= \gamma \exp \left[- \sum_c \varphi_{c \in C}(x_c) - \sum_{s \in S} \log(p(x_s/y_s)) \right]
\end{aligned}$$

avec

$$\gamma(y) = \sum_{x \in \Omega^N} \exp \left[- \sum_c \varphi_{c \in C}(x_c) + \sum_{s \in S} \log(p(x_s/y_s)) \right]$$

Cette formulation montre que la distribution a posteriori du champ X reste une distribution de Gibbs, ce qui signifie que la loi du champ caché conditionnellement à l'observation est également une loi de Markov. Cela rend possible la simulation des réalisations de X selon sa probabilité a posteriori, et permet ainsi d'utiliser différentes méthodes de segmentation bayésienne.

3.7.5 Champs de Markov cachés en segmentation d'images

Le problème posé ici est de trouver une réalisation x , modélisée par un processus markovien X , correspondant à une image d'étiquettes, chaque étiquette correspondant à un objet présent dans y : c'est le problème de segmentation d'images.

Dans un cadre bayésien, on peut rechercher la configuration \hat{x} maximisant la probabilité $P(X = x/Y = y)$, qui s'écrit :

$$P(X = x/Y = y) = \frac{P(Y = y/X = x) P(X = x)}{p(Y = y)}$$

Le premier terme du numérateur décrit la probabilité d'observation y , sachant que l'image idéal est x : il modélise donc l'acquisition de l'image sous l'hypothèse d'indépendance

des pixels,

$$P(Y = y/X = x) = \prod_{s \in S} P(Y_s = y_s/X_s = x_s)$$

Le second terme du numérateur décrit la probabilité d'existence de l'image idéale x , qui dans le cadre qui nous intéresse, répond à l'hypothèse markovienne selon un système de voisinage \mathcal{V} et un modèle ϕ dépendant de l'application (Dans le cas de segmentation).

$$P(X = x) = \frac{e^{-U(x)}}{\sum_{z \in \Omega} e^{-U(z)}}$$

Le dénominateur est constant, et en particulier indépendant de x . Ainsi

$$\begin{aligned} P(X = x/Y = y) &= KP(Y = y/X = x)P(X = x) \\ &= Ke^{\ln(P(Y=y/X=x))-U(x)} \\ &= Ke^{-U(x/y)} \end{aligned}$$

où

$$U(x/y) = - \sum P(Y_s = y_s/X_s = x_s) + \sum_{c \in C} U(x)$$

La distribution $P(X = x/Y = y)$ est donc une distribution de Gibbs, et par le théorème d'Hammersley - Clifford, le champ X conditionnellement à y est un champ de Markov.

3.7.6 Spécification de $U(x/y)$ pour le problème de segmentation d'image

Dans ce cas, X est défini sur l'espace des étiquettes, et Y sur l'espace de ses descripteurs. Le terme $P(X = x/Y = y)$ traduit donc la probabilité d'observation de l'image y connaissant l'appartenance de chaque pixel à un objet présent dans la scène. En supposant

que y_s ne dépend que de x_s , on peut écrire :

$$P(Y = y/X = x) = \prod_{s \in S} P(Y_s = y_s/X_s = x_s) = \prod_{s \in S} f_{x_s}(y_s)$$

Les valeurs de probabilité conditionnelles sont données par la fréquence d'observation des niveaux de gris pour une classe donnée, et si chaque classe k suit une distribution gaussienne de moyenne m_k et d'écart-type σ_k

$$f_{x_s}(y_s) = P(Y_s = y_s/X_s = x_s) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(y_s - m_k)^2}{2\sigma_k^2}\right]$$

Si on fait encore une fois l'hypothèse markovienne sur X et si on se restreint aux cliques d'ordre 2

$$P(X = x) = \frac{\exp\left(-\beta \sum_{s,t \in C_2} \phi(x_s, x_t)\right)}{\exp\left(-\beta \sum_{z \in \Omega, s,t \in C_2} \phi(z_s, z_t)\right)}$$

et donc :

$$U(x/y) = \sum \frac{(y_s - m_{x_s})^2}{2\sigma_{x_s}^2} + \log(2\pi) \sigma_{x_s} + \beta \sum_{s,t \in C_2} \phi(x_s, x_t)$$

où

$$\phi(x_s, x_t) = \begin{cases} 1 & \text{si } x_s \neq x_t \\ 0 & \text{si } x_s = x_t \end{cases}$$

Le terme d'ordre un exprime l'attache aux données, et le terme d'ordre deux régularise l'image. ϕ est le modèle de Potts, pour favoriser les grandes régions homogènes dans l'image.

3.8 Estimation des paramètres

L'algorithme en arrière et l'algorithme en avant peuvent également être utilisés pour estimer les paramètres associés au processus *HMM*. Généralement, dans le contexte de *1D HMM*, les «paramètres» θ se réfèrent à l'ensemble d'inconnues suivant : les probabilités d'étiquette de classe initiale, les probabilités de transition et les probabilités conditionnelles des données d'image observées avec l'étiquette de classe.

Ainsi, pour une chaîne de Markov homogène, on a un jeu de paramètres θ pour $t \in \{2, \dots, T\}$ donné par

$$\theta = \{\Pi, A, B\}$$

où Π comprend les probabilités initiales $\Pi = \{P(X_1 = m), m = 0, \dots, L_X - 1\}$, $A = \{P(X_t = l / X_{t-1} = m), m, l = 0, \dots, L_X - 1\}$, inclut les probabilités de transition, et $B = \{P(y_t / X_{t-1} = m), m = 0, \dots, L_X - 1\}$ est la fonction de vraisemblance. L'estimation du maximum de vraisemblance θ^* est alors donnée par

$$\theta^* = \arg \max_{\theta} P(y_{(1 \rightarrow T)} / \theta)$$

Puisqu'il existe également une étiquette de classe inconnue $x_{(1 \rightarrow T)}$ ainsi que θ , nous devons estimer les deux quantités à partir de $y_{(1 \rightarrow T)}$. Cependant, il est difficile de trouver des solutions optimales $x_{(1 \rightarrow T)}^*$ et θ^* qui maximisent simultanément la fonction de vraisemblance. Le problème d'estimation de paramètres avec des étiquettes cachées est bien adapté au cadre de la procédure d'espérance et de maximisation (EM) [Dempster 1977].

Considérons d'abord la probabilité de classe conjointe de $x_t = m$ et $x_{t+1} = l$ étant donné les données d'image observées $y_{(1 \rightarrow T)}$:

$$P(x_t, x_{t+1}/\mathbf{y}_{(1 \rightarrow T)}) = \frac{P(x_t, \mathbf{y}_{(1 \rightarrow t)}) P(y_{t+1}/x_{t+1}) P(\mathbf{y}_{(t+1 \rightarrow T)}/x_{t+1})}{P(\mathbf{y}_{(1 \rightarrow T)})} \quad (3.8)$$

où

$$P(\mathbf{y}_{(1 \rightarrow T)}/x_t) = P(y_t/x_t) \sum_{x_{t+1}} P(\mathbf{y}_{(t+1 \rightarrow T)}/x_{t+1}) P(x_{t+1}/x_t)$$

$$P(x_t, \mathbf{y}_{(1 \rightarrow t)}) = P(y_t/x_t) \sum_{x_{t-1}} P(x_{t-1}, \mathbf{y}_{(1 \rightarrow t-1)}) P(x_{t+1}/x_t)$$

Propriétés : L'estimateur MAP optimal x_t^* , pour l'étiquette de classe peut être obtenu comme suit

$$x_t^* = \arg \max_{x_t \in \{0, \dots, L_X - 1\}} P(x_t/\mathbf{y}_{(1 \rightarrow T)}) \quad (3.9)$$

- Le critère MAP marginal dans 3.9 peut être obtenu de manière équivalente à partir de 3.8 car

$$\sum_{x_{t-1}} P(x_t, x_{t+1}/\mathbf{y}_{(1 \rightarrow T)}) = P(x_t/\mathbf{y}_{(1 \rightarrow T)})$$

- Le nombre prévu de fois qu'une étiquette de classe particulière $x_t = m \in \{0, \dots, L_X - 1\}$ se produit est donné par

$$\sum_{t=1}^T \sum_{x_{t+1}} P(X_t = m, x_{t+1}/\mathbf{y}_{(1 \rightarrow T)})$$

- Le nombre attendu de transitions d'une classe $x_t = m \in \{0, \dots, L_X - 1\}$ à $x_{t+1} = l \in \{0, \dots, L_X - 1\}$ est

$$\sum_{t=1}^{T-1} P(X_t = m, x_{t+1} = l / \mathbf{y}_{(1 \rightarrow T)})$$

En utilisant les propriétés ci-dessus et le concept de comptage des occurrences d'événements comme probabilité, nous pouvons obtenir une estimation $\theta = \{\diamond, A, B\}$ comme suit [9, 10] :

$$\hat{P}(X_1 = m) = P(X_1 = m / \mathbf{y}_{(1 \rightarrow T)})$$

$$\hat{P}(X_{t+1} = l / X_t = m) = \frac{\sum_{t=1}^{T-1} P(X_t = m, X_{t+1} = l / \mathbf{y}_{(1 \rightarrow T)})}{\sum_{t=1}^T \sum_{x_{t+1}} P(X_t = m, x_{t+1} / \mathbf{y}_{(1 \rightarrow T)})}$$

$$\hat{P}(y_t = b / X_t = m) = \frac{\sum_{t=1}^T \sum_{y_t=b} \sum_{x_{t+1}} P(X_t = m, x_{t+1} / \mathbf{y}_{(1 \rightarrow T)})}{\sum_{t=1}^T \sum_{x_{t+1}} P(X_t = m, x_{t+1} / \mathbf{y}_{(1 \rightarrow T)})} \quad (3.10)$$

où $m, l = 0, \dots, L_X - 1, b = 0, \dots, L_Y - 1$ et la somme $\sum_{t=1}^T$ au numérateur de 3.10 ne doit être fait que pour ceux t pour lesquels $y_t = b$.

Chapitre 4

Modèle $2D$ autorégressif à changement de régime Markovien unilatéral ($MS - AR$)

Ce chapitre a fait l'objet d'un article de Djafri et Kharfouchi. S (2020) intitulé :
On unilateral $2D$ Markov-Switching Autoregressive ($MS - AR$) model.

4.1 Introduction

Les modèles autorégressifs à changement de régime Markovien $1D$ ($MS - AR$) ont été initialement introduits par Goldfeld et Quandt (1973), puis ils ont été adaptés dans le travail inspirant de Hamilton (1990) pour capturer le cycle économique de nombreuses séries chronologiques économiques qui présentent parfois des ruptures dramatiques dans leur

comportement, il est une généralisation des modèles chaîne de Markov cachée *HMM* et des modèles autorégressifs *AR*. En effet, ils combinent différents modèles autorégressifs pour décrire l'évolution du processus à différentes périodes de temps, la transition entre ces différents modèles autorégressifs étant contrôlée par une chaîne de Markov cachée comme dans *HMM*. Les modèles *MS – AR* sont présentés brièvement dans la section 4.1. Dans la suite de ce chapitre, une nouvelle classe de processus est introduite, qui est un mélange de modèles d'autorégressifs spatiaux *2D AR* et de modèles causaux de champ aléatoire de Markov *MRF* appelée "processus *2D* autorégressif à changement de régime Markovien unilatéral" , Pour les données spatiales bidimensionnelles non linéaires, est bien défini. Tout d'abord, un champ aléatoire de Markov *2D MRF* est construit pour lequel une causalité imposée permet d'établir une analogie entre ce *2D MRF* et la représentation en chaîne de Markov cachée *HMM*. Sur la base du *2D MRF* proposé, le modèle *2D MS – AR* est défini sous certaines hypothèses nécessaires et notations utiles. nous étudions la stationnarité stricte et l'estimation des paramètres via l'algorithme *VEM* (Variational Expectation Maximization) est effectuée.

4.2 Modèles *1D MS-AR*

Les modèles autorégressifs à changements de régimes ont été introduits il y a 30 ans par Hamilton (1989) en économétrie, puis ont été ensuite largement utilisés en économétrie et en traitement automatique de la parole (Krolzig (1997) et Douc et al. (2004).

4.2.1 Description du modèle

On peut définir un modèle autorégressif à changements de régime markovien ($MS - AR$) comme un processus bivarié (Z_t, Y_t) qui présente une hypothèse probabiliste dans le passage d'un régime à un autre. Le régime à la date t est déterminé par une variable inobservable Z_t modélisée par une chaîne de Markov cachée discrète du premier ordre à espace d'état fini $\mathcal{X} = \{1, \dots, M\}$, stationnaire, homogène, irréductible et ergodique indépendante de $(\varepsilon_t)_{t \in \mathbb{Z}}$ et des Y_t décalées, c'est-à-dire $(Z_t)_{t \in \mathbb{Z}}$ et $(\varepsilon_t, Y_{k+1}, k \leq t)$ sont indépendantes, tandis que l'évolution de la variable Y_t observée suit un modèle autorégressif linéaire.

Un processus $MS - AR(p)$ est alors caractérisé par les deux hypothèses d'indépendance conditionnelle ci dessous :

$$\text{i) } P(Z_t = j / Z_{t-1} = i_{t-1}, Z_{t-2} = i_{t-2}, \dots, Z_1 = i_1, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots, Y_1 = y_1) = P(Z_t = j / Z_{t-1} = i_{t-1})$$

$$\text{ii) } P(Y_t / Z_{t-1} = i_{t-1}, Z_{t-2} = i_{t-2}, \dots, Z_1 = i_1, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots, Y_1 = y_1) = P(Y_t / Z_t = s_t, Y_{t-1} = y_{t-1})$$

et $P(Y_t / S_t = s_t, Y_{t-1} = y_{t-1})$ est la probabilités d'émission .

L'indépendance conditionnelle pour le modèle $MS - AR$ d'ordre 1 (Figure 4.1)

4.2.2 Mise en équation

Un processus stochastique $(Y_t, t \in \mathbb{Z})$ à valeurs réelles, défini sur un espace de probabilité (Ω, \mathcal{A}, P) , est dit admettre une représentation autoregressive à changement de régime markovien ($MS - AR$) s'il est solution de l'équation aux différences stochastique

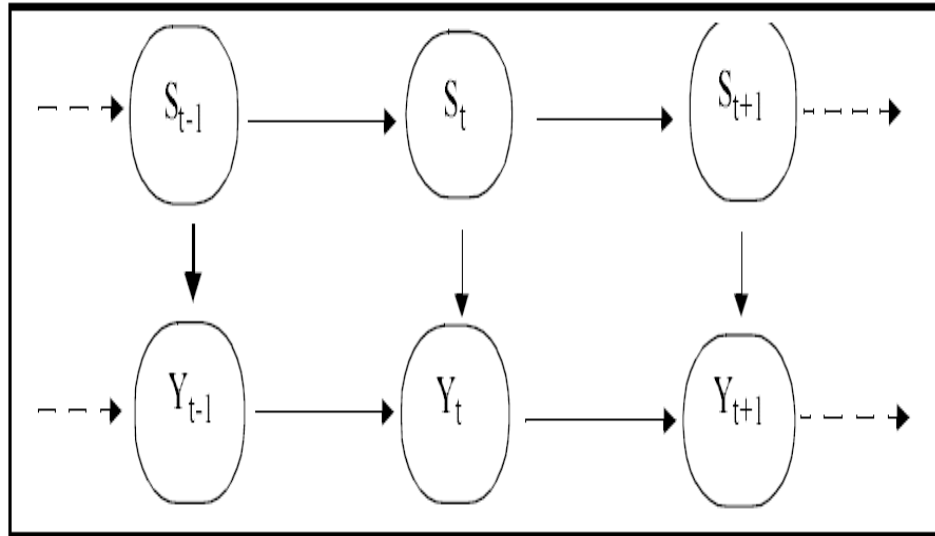


FIG. 4.1: Graphe d'indépendance conditionnelle pour le modèle MS-AR d'ordre 1

suivante (Hamilton, 1989, 1990, 1994)

$$Y_t = \sum_{i=1}^p \alpha_i(Z_t) Y_{t-i} + \varepsilon_t, \quad t \in \mathbb{Z} \quad (4.1)$$

où $(\alpha_i(Z_t))$ sont des fonctions bornées, déterministes, dépendant éventuellement d'une chaîne de Markov à espace d'état fini, et où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un processus de bruit blanc fort centré de variance finie et satisfaisant l'hypothèse suivante

$$\{\varepsilon_t \text{ et } Y_s \text{ sont indépendants pour tout } s < t\}$$

On suppose constantes les probabilités de transition de l'état i à l'état j , p_{ij} . La matrice de transition P constante associée à Z_t est alors :

$$P(Z_t = j / Z_{t-1} = i_{t-1}, Z_{t-2} = i_{t-2}, \dots, Z_1 = i_1) = P(Z_t = j / Z_{t-1} = i_{t-1}) = p_{ij}$$

et

$$P = \begin{pmatrix} P(Z_t = 1/Z_{t-1} = 1) & \cdots & P(Z_t = M/Z_{t-1} = 1) \\ \vdots & \ddots & \vdots \\ P(Z_t = 1/Z_{t-1} = M) & \cdots & P(Z_t = M/Z_{t-1} = M) \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix}$$

La matrice de transition P est stochastique car elle vérifie les propriétés suivantes :

$$\begin{cases} 1) & 0 \leq p_{ij} \leq 1, \forall i, j \in \{1, \dots, M\} \\ 2) & \sum_{j=1}^M p_{ij} = 1, \forall i \in \{1, \dots, M\} \end{cases}$$

La propriété (2) découle du fait que les évènements $P(Z_t = 1/Z_{t-1} = i), \dots, P(Z_t = M/Z_{t-1} = i)$

forment une partition de l'univers Ω .

MS(M) – AR(p) sous la forme matricielle

On peut réécrire l'équation 4.1 sous la forme

$$\bar{\mathbf{Y}}_t = A^{(Z_t)} \bar{\mathbf{Y}}_{t-1} + E_t \quad (4.2)$$

avec

$$E_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad A^{(z)} = \begin{pmatrix} \alpha_1^{(z)} & \alpha_2^{(z)} & \cdots & \alpha_{p-1}^{(z)} & \alpha_p^{(z)} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \text{où } z \in \Lambda, \quad \bar{\mathbf{Y}}_t = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}$$

L'application du modèle 1D $MS - AR$ sur des données réelles permet d'obtenir

les estimateurs des probabilités associées aux régimes inobservés conditionnellement à l'ensemble d'information disponible et donc de fournir une indication sur l'historique de ces régimes. On distingue alors trois types de probabilité : la probabilité associée au filtrage (Filtered probabilities), $P(Z_t/Y_t)$, la probabilité lissée (Smoothed probabilities), $P(Z_t/Y_T)$,

et les probabilités prévues (predicted probabilities), $P(Z_t/Y_{t-1})$, où Y_t désigne l'historique des réalisations de la variable Y jusqu'à la date t pour t variant de 0 à T .

L'estimation du modèle se fait en utilisant l'algorithme *EM*. Dans l'étape « Expectation », une inférence est constituée sur la chaîne de Markov cachée pour un ensemble de paramètres à estimer fixé. Dans l'étape de « Maximisation », les paramètres de la chaîne de Markov cachée inférée sont réestimés. Ces deux étapes sont répétées jusqu'à la convergence de l'algorithme. L'utilisation de l'algorithme EM permet ainsi d'estimer les paramètres propres à chaque régime, la matrice de transition et les probabilités lissés (inférence optimal de la chaîne de Markov).

4.3 Construction du modèle $2D$ *MS-AR*

Pour incorporer les informations spatiales dans les modèles *MS-AR*, on propose dans cette section un processus $2D$ *MS-AR*. Ce modèle est une extension naturelle du modèle *MS-AR* original de Hamilton au cas $2D$. C'est aussi, simplement, un modèle $2D$ *AR* unilatéral (définir en chapitre 2) dont les paramètres sont gérés par des états latents. Ces états sont eux-mêmes aléatoires et sont supposés suivre un champ aléatoire de Markov causal, homogène et discret proposé défini dans la sous-section suivante.

4.3.1 Le champ aléatoire de Markov causal homogène

Les champs aléatoires de Markov ont été incorporés dans le traitement d'image en tant qu'instrument fondamental de sa boîte à outils, permettant l'émergence d'une nouvelle classe de modèles qui occupent une position centrale dans l'analyse d'images. Dans ce cas,

selon le théorème de Hammersley-Clifford, le MRF sous-jacent doit suivre une distribution de Gibbs avec une fonction d'énergie appropriée. Cependant, la non-causalité de ces modèles conduit généralement à des algorithmes d'inférence itérative coûteux. Pour contourner cette difficulté, nous considérons dans cet article une alternative intéressante à ces spécifications MRF non causales en ordonnant artificiellement les données $2D$ d'une manière analogue à l'ordre naturel du temps pour un signal unidimensionnel, obligeant les signaux bidimensionnels à être "causaux" particuliers signaux. La causalité permet d'établir une analogie entre les $2D MRF$ et la représentation en chaîne de Markov.

Cette analogie permet d'utiliser des outils utiles développés pour les problèmes de chaîne de Markov tels que l'algorithme EM utilisé pour l'estimation des paramètres du modèle. La manière la plus simple de construire une telle notion de causalité dans le cas $2D$ est d'utiliser des méthodes de balayage qui convertissent les données de signal $2D$ en une séquence de données $1D$ ordonnée. Nous sélectionnons comme méthode de balayage le balayage raster correspondant à l'ordre lexicographique sur \mathbb{Z}^2 . Par exemple, dans le traitement d'image, le balayage de trame garantit que la séquence $1D$ obtenue sera agencée de sorte que les pixels spatialement proches soient aussi proches que possible.

Considérons ensuite un ensemble fini $S \subset \mathbb{Z}^2$ représentant les emplacements (ou sites) des données. L'utilisation de l'ordre **lexicographique** (\ll) sur \mathbb{Z}^2 fournit une notion de passé et de présent considéré sur S lors de l'introduction de la définition du champ aléatoire de Markov proposé. L'ordre lexicographique (\ll) sur \mathbb{Z}^2 est défini pour deux sites (i, j) et (i', j') de \mathbb{Z}^2 par l'équation 1.1, auquel cas on dit que le site (i, j) est avant (i', j') . A partir de cet ordre, nous pouvons définir quelques ensembles utiles pour la description du

modèle $2D$ $MS - AR$ proposé. Pour deux éléments a, b de \mathbb{Z}^2 tels que $a \ll b$, considérons les sous-ensembles suivants de \mathbb{Z}^2 :

$$\begin{aligned} S[\mathbf{a}, \mathbf{b}] &= \{x \in \mathbb{Z}^2 / \mathbf{a} \ll x \ll \mathbf{b}\}; \quad S|\mathbf{a}, \mathbf{b}] = S[\mathbf{a}, \mathbf{b}] \setminus \{\mathbf{a}\} \\ S[\mathbf{a}, +\infty[&= \{x \in \mathbb{Z}^2 / \mathbf{a} \ll x\}; \quad S|\mathbf{a}, +\infty[= S[\mathbf{a}, +\infty[\setminus \{\mathbf{a}\}. \end{aligned}$$

Or, pour chaque site $\mathbf{s} \in S$ ($S \subset \mathbb{Z}^2$), soit $Z_{\mathbf{s}}$ désigne la variable aléatoire associée et $z_{\mathbf{s}}$ la réalisation de $Z_{\mathbf{s}}$. $z_{\mathbf{s}}$ prend ses valeurs dans un espace d'états Λ (dans le cas d'une image, Λ est en général l'ensemble des niveaux de gris de l'image). Tout au long de ce chapitre, nous noterons $P(z_{\mathbf{s}})$ la distribution $P(Z_{\mathbf{s}} = z_{\mathbf{s}})$ ainsi que toute distribution conditionnelle ou conjointe en découlant : par exemple $P(Z_{\mathbf{s}} = z_{\mathbf{s}} \setminus Z_{\mathbf{s}'} = z_{\mathbf{s}'} : \mathbf{s}' \in S[\mathbf{a}, \mathbf{b}]) = P(z_{\mathbf{s}} \setminus z_{\mathbf{s}'} : \mathbf{s}' \in S[\mathbf{a}, \mathbf{b}])$. Soit la suite $\alpha_{(\mathbf{a} \rightarrow \mathbf{b})} = \{\alpha_{\mathbf{r}_1}, \dots, \alpha_{\mathbf{r}_{\overleftarrow{\mathbf{ab}}}}\}$ l'ensemble $\{\alpha_{\mathbf{s}} : \mathbf{s} \in S[\mathbf{a}, \mathbf{b}]\}$ arrangé dans l'ordre lexicographique, où $\overleftarrow{\mathbf{ab}} = \text{Card}(S[\mathbf{a}, \mathbf{b}])$.

Définition 43 $(Z_{\mathbf{s}})_{\mathbf{s} \in S[\mathbf{1}, \mathbf{n}]}$ est dit être un champ aléatoire markovien causal homogène du premier ordre si pour tout $\mathbf{s} \in S[\mathbf{1}, \mathbf{n}]$ la suite ordonnée $\mathbf{Z}_{(1 \rightarrow \mathbf{s})}$ est une chaîne de Markov homogène du premier ordre à temps discret, de sorte que :

$$P(z_{\mathbf{r}_1}, \dots, z_{\mathbf{r}_{\overleftarrow{\mathbf{1s}}}}) = P(z_{\mathbf{r}_1}) \prod_{k=2}^{\overleftarrow{\mathbf{1s}}} P(z_{\mathbf{r}_k} / z_{\mathbf{r}_{k-1}}). \quad (4.3)$$

où $\overleftarrow{\mathbf{1s}} = \text{Card}(S[\mathbf{1}, \mathbf{s}])$

Remarque 44 Comme Hamilton (1990), lorsque nous considérons la dynamique markovienne de premier ordre à valeurs scalaires, nous pouvons facilement utiliser les résultats

obtenus pour un système à valeurs vectorielles ou d'ordre supérieur qui peut toujours être réécrit comme un processus de premier ordre à valeurs scalaires .

4.3.2 Présentation du modèle 2D MS – AR

Soit S tout sous-ensemble fini non vide de \mathbb{Z}^2 , $(Y(\mathbf{t}), \mathbf{t} \in S)$ désignent le champ aléatoire associé.

Définition 45 On dit que $Y(\mathbf{t})_{\mathbf{t} \in S}$ est un modèle 2D autorégressif à changement de régime Markovien causal d'ordre $\mathbf{p} = (p_1, p_2)$, sur un espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$, (2D MS-AR(\mathbf{p})) s'il a la représentation d'équation de différence suivante :

$$Y(\mathbf{t}) = \sum_{\mathbf{k} \in S]0, \mathbf{p}] } a_{\mathbf{k}}(Z_{\mathbf{t}}) Y(\mathbf{t} - \mathbf{k}) + \varepsilon(\mathbf{t}), \quad (4.4)$$

$\varepsilon(\mathbf{t})$ est une suite de variables aléatoires i.i.d de moyennes nulles et de variance σ^2 . Les fonctions $a_{\mathbf{k}}(\cdot)$ Dépendent du champ aléatoire causal de Markov latent du premier ordre $(Z_{\mathbf{t}})_{\mathbf{t} \in S}$ défini sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$.

De plus, $(Z_{\mathbf{t}})_{\mathbf{t} \in S}$ est supposé être homogène, stationnaire, irréductible et indépendant de $\varepsilon(\mathbf{t})_{\mathbf{t} \in S}$ et retardé $Y(\mathbf{t})$ i.e., $Z_{\mathbf{t}}$ et $\{\varepsilon(\mathbf{t}), Y(\mathbf{k}), \mathbf{k} \prec \mathbf{t}\}$ sont indépendants.

Matrice de transition

Soit $(Z_{\mathbf{t}})_{\mathbf{t} \in S}$ un champ aléatoire causal de Markov homogène du premier ordre, à l'espace d'états $\Lambda = \{1, \dots, L\}$ fini. La matrice de probabilité de transition $\mathbf{P} = (p_{ij})_{1 \leq i, j \leq L}$ donnée par

$$p_{ij} = P(Z_{\mathbf{r}_k} = j / Z_{\mathbf{r}_{k-1}} = i), k = 1, \dots, \overleftarrow{\mathbf{1}, \mathbf{s}}, \forall s \in S$$

où $\sum_{i=1}^L p_{ij} = 1$ pour tout $\mathbf{s} \in S$.

D'après Hamilton (1990), la loi de probabilité régissant les états initiaux non observés $Z_{(\mathbf{1} \rightarrow \mathbf{p})}$, $\mathbf{k} \in S] \mathbf{0}, \mathbf{p}]$ est tirée d'une distribution de probabilité distincte dont le paramètre ρ est un vecteur $(L^K \times 1)$ sans rapport avec \mathbf{P} et $a_{\mathbf{k}}(\cdot)$, où $K = \overleftrightarrow{\mathbf{0}, \mathbf{p}} - 1$ est la cardinalité de l'ensemble $S] \mathbf{0}, \mathbf{p}]$. Les éléments de ρ totalisent l'unité et ils sont donnés par :

$$P(\mathbf{z}_{(\mathbf{1} \rightarrow \mathbf{p})} / y_{(\mathbf{1} \rightarrow \mathbf{p})}) = \rho_{\mathbf{z}_{(\mathbf{1} \rightarrow \mathbf{p})}}, \mathbf{z}_{(\mathbf{1} \rightarrow \mathbf{p})} \in \Lambda^K. \quad (4.5)$$

Dans la partie d'estimation, l'hypothèse que les observations sont des solutions non anticipatives strictement stationnaires de 4.4 est nécessaire. Par conséquent, on étudie la stationnarité que le modèle $2D MS - AR(p)$ en section suivante

4.4 Stationnarité stricte de $2D MS - AR$

Notre attention se concentre maintenant sur les conditions pour que le modèle $2D MS - AR(p)$ soit strictement stationnaire. Les matrices de transition $T^{n,m}(i,j)$ définies ci-dessous et les produits de la théorie des matrices aléatoires seront utilisés pour montrer que la condition nécessaire pour une stationnarité stricte est **la négativité d'un coefficient associé** à une séquence appropriée des matrices stationnaires et ergodiques.

Rappelons tout d'abord que pour chaque matrice $m \times m$, $A = (a_{ij})$, la norme $\|A\|$ de A est définie par $\|A\| = \max_i \sum_j |a_{ij}|$ et pour tout $x > 0$, $\log^+ x = \max\{\log x, 0\}$.

Représentation espace-état du modèle $2D MS - AR$

Nous proposons maintenant de représenter le modèle $2D MS - AR(p)$ défini par l'équation 4.4 par une version d'espace d'états. Cela peut être fait, pour $t = (i, j)$, en

définissant les vecteurs $(p_1 + 1)(p_2 + 1) \times 1$.

$$\underline{Y}(i, j) = (y_0(i, j), \dots, y_{p_1}(i, j))' \text{ et } \underline{\varepsilon}(i, j) = (\varepsilon(i, j), 0, \dots, 0),$$

Avec $y_k(i, j) = (Y(i - k, j), Y(i - k, j - 1), \dots, Y(i - k, j - p_2))$, $k = 0, \dots, p_1$ et en considérant $\mathcal{A}_1(i, j)$ les matrices $(p_1 + 1)(p_2 + 1) \times (p_1 + 1)(p_2 + 1)$ et $\mathcal{A}_2(i, j)$ définies ci-dessous. Avec ces notations, l'équation 4.4 peut être réécrite sous la forme :

$$\underline{Y}(i, j) = \mathcal{A}_1(i, j)\underline{Y}(i, j - 1) + \mathcal{A}_2(i, j)\underline{Y}(i - 1, j) + \underline{\varepsilon}(i, j), \quad (4.6)$$

où

$$\mathcal{A}_1(i, j) = \begin{pmatrix} \underline{a}_0(i, j) & \underline{a}_1(i, j) & & & \underline{a}_{p_1}(i, j) \\ \mathcal{J} & O_{p_2 \times (p_2 + 1)} & \cdots & \cdots & O_{p_2 \times (p_2 + 1)} \\ O_{1 \times (p_2 + 1)} & \cdots & \cdots & & O_{1 \times (p_2 + 1)} \\ O_{p_2 \times (p_2 + 1)} & \mathcal{J} & O_{p_2 \times (p_2 + 1)} & \cdots & O_{p_2 \times (p_2 + 1)} \\ O_{1 \times (p_2 + 1)} & \cdots & \ddots & & O_{1 \times (p_2 + 1)} \\ \vdots & & & \ddots & \vdots \\ O_{p_2 \times (p_2 + 1)} & \cdots & & O_{p_2 \times (p_2 + 1)} & \mathcal{J} \end{pmatrix},$$

$$\mathcal{A}_2(i, j) = \begin{pmatrix} \bar{a}_1(i, j) & \cdots & \bar{a}_{p_1}(i, j) & \bar{0} \\ O_{p_2 \times (p_2 + 1)} & & & \vdots \\ \mathcal{T}_{(p_2 + 1) \times (p_2 + 1)} & \cdots & O_{(p_2 + 1) \times (p_2 + 1)} & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ O_{(p_2 + 1) \times (p_2 + 1)} & \cdots & \mathcal{T}_{(p_2 + 1) \times (p_2 + 1)} & O_{(p_2 + 1) \times (p_2 + 1)} \end{pmatrix}$$

Où $\underline{a}_k(i, j)$, $\bar{a}_s(i, j)$ et $\bar{0}$ sont $(p_2 + 1)$ vecteurs tels que,

$$\underline{a}_k(i, j) = (a_{k1}(i, j), \dots, a_{kp_2}(i, j), 0), \text{ et } \bar{a}_s(i, j) = (a_{s0}(i, j), 0, \dots, 0), 0 \leq k \leq p_1, 1 \leq s \leq p_1 \text{ et } \bar{0} = (0, \dots)$$

$O_{p \times q}$ désigne la matrice nulle d'ordre $p \times q$. \mathcal{J} et $\mathcal{T}_{r \times s}$ sont définis par

$$\mathcal{J} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}_{p_2 \times (p_2+1)}, \text{ et } \mathcal{T}_{r \times s} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}_{r \times s}$$

Dans ce cas,

Théorème 46 *Considérons le modèle 4.4 avec représentation en espace d'états 4.6 et supposons que $E \log^+ \{\|\mathcal{A}_1(i, j)\| + \|\mathcal{A}_2(i, j)\|\} < \infty$. Si presque sûrement (p.s.)*

$$\rho < 0, \tag{4.7}$$

où

$$\rho = \limsup_k \frac{1}{k} \log \prod_{l=0}^{k-1} \sum_{s=0}^l (\|\mathcal{A}_1(i-s, j-k+l)\| + \|\mathcal{A}_2(i-s, j-k+s)\|),$$

alors pour tout $(i, j) \in \mathbb{Z}^2$,

$$\underline{Y} = \sum_{k \geq 0} \sum_{l=0}^k T^{l, k-l}(i, j) \underline{\varepsilon}(i-l, j-k+l),$$

converge absolument p.s. et le processus $\{\underline{H}'\underline{Y}(i, j), (i, j) \in \mathbb{Z}^2\}$ est l'unique solution strictement stationnaire de 4.4. Les matrices de transition $T^{n, m}(i, j)$ sont définies comme suit :

i) $T^{0,1}(i, j) = \mathcal{A}_1(i, j)$, $T^{1,0}(i, j) = \mathcal{A}_2(i, j)$ et $T^{n,m}(i, j) = \mathcal{A}_1(i, j) T^{n,m-1}(i, j-1) + \mathcal{A}_2(i, j) T^{n-1,m}(i-1, j)$

ii) $T^{0,0}(i, j) = I_{(P+Q)}$ et $T^{-n,m}(i, j) = T^{n,-m}(i, j) = 0_{(P+Q)}$

et le vecteur $\underline{H} = \left(1, \underline{Q}'_{(P-1)}\right)'$

Proof. En itérant l'équation 4.6, il est facile de voir que $\sum_{k \geq 0} \sum_{l=0}^k W_k(i, j)$, où

$$W_k(i, j) = T^{l, k-l}(i, j) \underline{\varepsilon}(i-l, j-k+l)$$

est une solution de 4.6 à condition que cette série de vecteurs aléatoires soit absolument convergente. Pour établir cette convergence, nous utilisons le critère de Cauchy appliqué à la série $\sum_{k=0}^{\infty} \|W_k(i, j)\|$. D'après lemme de Borel-Cantelli, on a

$$P\left(\limsup |\varepsilon(i-l, j-k+l)|^{\frac{1}{k}} > x\right) = 0$$

pour tout $x > 1$ et tout $l \in \overline{0.k}$. Par contre, en utilisant les récursions i) -ii), on obtient presque sûrement pour tout k

$$\|W_k(i, j)\| \leq \prod_{l=0}^{k-1} \sum_{s=0}^l (\|\mathcal{A}_1(i-s, j-k+l)\| + \|\mathcal{A}_2(i-s, j-k+s)\|),$$

alors

$$\limsup \frac{1}{k} \log \|W_k(i, j)\| \leq \rho.$$

Par le théorème ergodique sous-additif (voir par exemple Cohen [23]) ρ existe avec probabilité un sous la condition $E \log^+ \{\|\mathcal{A}_1(i, j)\| + \|\mathcal{A}_2(i, j)\|\} < \infty$, est non aléatoire, finie et a une espérance qui satisfait $|\rho| = E |\rho|$. Donc quand $\rho < 0$, $\|W_k(i, j)\|^{\frac{1}{k}} \rightarrow 0$ a.s et le résultat découle du test de racine de Cauchy.

Prouvons maintenant l'unicité, supposons qu'il y ait deux solutions stationnaires de probabilité bornées à 4.6 $\underline{X}(i, j)$ et $\underline{Y}(i, j)$ et poser $\underline{U}(i, j) = \underline{X}(i, j) - \underline{Y}(i, j)$. Alors $\underline{U}(i, j)$ satisfait $\underline{U}(i, j) = \mathcal{A}_1(i, j)\underline{U}(i, j-1) + \mathcal{A}_2(i, j)\underline{U}(i-1, j)$. Donc, pour tout $k \geq 1$,

$$\underline{U}(i, j) = \sum_{l=0}^k T^{l, k-l}(i, j) \underline{U}(i-l, j-k+l).$$

Par conséquent

$$\|\underline{U}(i, j)\| = \sum_{l=0}^k \left\| T^{l, k-l}(i, j) \right\| \|\underline{U}(i-l, j-k+l)\|,$$

et puisque $(\underline{U}(i, j), (i, j) \in \mathbb{Z}^2)$ est borné en probabilité, alors le membre droit de l'inégalité ci-dessus converge vers 0 comme $k \rightarrow \infty$ et donc $\underline{X}(i, j) = \underline{Y}(i, j)$, p.s. ■

Remarque 47 *La quantité peut être considérée comme la contrepartie 2D de l'exposant de Lyapunov associé à la série temporelle autorégressive généralisée proposée dans (Brandt [16])*

4.5 Estimation des paramètres du modèle 2D MS – AR

4.5.1 La fonction de vraisemblance dans les modèles 2D MS – AR

Supposons que l'on observe $n_1 n_2$ variables aléatoires $\mathbf{Y} = (Y_{(1 \rightarrow \mathbf{n})})$, où $\mathbf{n} = (n_1, n_2)$, générées à partir du 2D MS – AR (p) pour lequel il existe un total de L régimes possibles à partir desquels une observation particulière peut être tirée $Y(i, j)$. Il est clair que nos observations sont incomplètes puisque nous ne savons pas de quelle composante vient l'observation $Y(i, j)$, $(i, j) \in S[\mathbf{1}, \mathbf{n}]$, Soit $\mathbf{Z} = (Z_{(1 \rightarrow \mathbf{n})})'$ et $\mathbf{z} = (z_{(1 \rightarrow \mathbf{n})})'$ l'état non observé

$(n_1 n_2 \times 1)$ pour l'échantillon entier et sa réalisation respectivement. Le paramètre inconnu est rassemblé dans

$$\underline{\theta} = \left(\rho, \underline{\mathbf{a}}'_i, \underline{p}_i, 1 \leq i \leq L \right)$$

où $\underline{\mathbf{a}}_i = \left(a_1(i), \dots, a_{\overleftarrow{\mathbf{1}, \mathbf{p}}}(i) \right)'$ et $\underline{p}_i = (p_{ji}, 1 \leq j \leq L \text{ et } i \neq j)$ (du fait des contraintes $\sum_{j=1}^L p_{ij} = 1$ pour tout i). La vraie valeur du paramètre est notée $\underline{\theta}_0$. Enfin, on note la fonction de densité des innovations $\varepsilon(i, j)$ par $f(\cdot)$.

Ensuite, la distribution conjointe conditionnelle de \mathbf{Y} et \mathbf{Z} si \mathbf{Z} était observable étant donné $Y_{\mathbf{1} \rightarrow \mathbf{p}}$ ressemblerait à :

$$P_{\mathbf{n}}(\mathbf{Y}, \mathbf{Z}, \underline{\theta}) \equiv P\left(Y_{((p_1, p_2+1) \rightarrow \mathbf{n})}, Z_{((p_1, p_2) \rightarrow \mathbf{n})} \setminus Y_{(1 \rightarrow \mathbf{p})}, \underline{\theta}\right) = \prod_{t=\overleftarrow{\mathbf{1}, \mathbf{p}+1}}^{\overleftarrow{\mathbf{1}, \mathbf{n}}} p_{z_{r_t} z_{r_{t-1}}} f(\varepsilon_{\underline{\theta}}(r_t)) \times \rho_{\mathbf{z}_{(1 \rightarrow \mathbf{p})}},$$

où $(\varepsilon_{\underline{\theta}}(\mathbf{t}), \mathbf{t} \in S[(\mathbf{1}, \mathbf{n})])$ est le processus déterminé récursivement par, $\varepsilon_{\underline{\theta}}(\mathbf{t}) = Y(\mathbf{t}) - \sum_{\mathbf{k} \in S[\mathbf{0}, \mathbf{p}]} a_{\mathbf{k}}(Z_{\mathbf{t}}) Y(\mathbf{t} - \mathbf{k})$, à partir de là, la fonction de vraisemblance marginale, $L_{\mathbf{n}}(\underline{\theta})$, serait simplement la somme de la vraisemblance conditionnelle conjointe $P_{\mathbf{n}}(\mathbf{X}, \mathbf{Z}, \underline{\theta})$ sur toutes les configurations possibles de $(z_{r_1}, \dots, z_{r_{\overleftarrow{\mathbf{1}, \mathbf{n}}}})$:

$$L_{\mathbf{n}}(\underline{\theta}) = \sum_{(z_{r_1}, \dots, z_{r_{\overleftarrow{\mathbf{1}, \mathbf{n}}}}) \in \Lambda^{\overleftarrow{\mathbf{1}, \mathbf{n}}}} \left\{ \prod_{t=\overleftarrow{\mathbf{1}, \mathbf{p}+1}}^{\overleftarrow{\mathbf{1}, \mathbf{n}}} p_{z_{r_t} z_{r_{t-1}}} f(\varepsilon_{\underline{\theta}}(r_t)) \times \rho_{\mathbf{z}_{(1 \rightarrow \mathbf{p})}} \right\}.$$

En raison de cette sommation, $L_{\mathbf{n}}(\underline{\theta})$ devient intraitable car beaucoup de voxels sont impliqués. Par conséquent, l'estimation des paramètres du modèle $2D \text{ MS} - AR$ ne peut pas être réalisée par MLE et serait effectuée via l'algorithme VEM d'espérance variationnelle-maximisation, également connu sous le nom de Bayes variationnelle.

4.5.2 L'estimation par l'algorithme *VEM*

L'algorithme *VEM* est une généralisation de l'algorithme *EM* discuté et utilisé dans de récents manuels de statistiques informatiques tels que [13]. Dans cet algorithme, de nouvelles estimations de log-vraisemblance maximale du vecteur de paramètres sont définies en minimisant l'expression $Q(\underline{\theta}^{j+1}, \underline{\theta}^j, \mathbf{Y})$ pour désigner la log-vraisemblance attendue et donnée par 4.8. Cela se fait alternativement en donnant, après avoir assigné le choix d'une estimation initiale au vecteur de paramètres $\underline{\theta}^0$, à mettre à jour les règles appelées étape *VE* et étape *VM*,

$$Q(\underline{\theta}^{j+1}, \underline{\theta}^j, \mathbf{Y}) = \sum_{z_{(1 \rightarrow \mathbf{n})} \in \Lambda^{\overleftarrow{1 \rightarrow \mathbf{n}}}} P_{\mathbf{n}}(\mathbf{Y}, z_{(1 \rightarrow \mathbf{n})}, \underline{\theta}^j) \log P_{\mathbf{n}}(\mathbf{Y}, z_{(1 \rightarrow \mathbf{n})}, \underline{\theta}^{j+1}). \quad (4.8)$$

Dans notre étape *VE*, des probabilités d'état filtrées et lissées sont produites progressivement. L'étape *VM* est formellement équivalente à l'étape *M* et évalue de nouvelles estimations log-vraisemblables maximales du vecteur de paramètres sur la base des étapes précédentes en utilisant 4.8 donnant une solution unique.

Dans notre cas, l'algorithme *EM* peut se résumer comme suit :

Etape 1 : (étape d'initialisation)

- Choisissez une estimation initiale $\underline{\theta}^0$

Etape 2 : (*E - step*)

- Soit $\hat{\underline{\theta}}^j$ le paramètre estimé à la j -ième itération. Si on note $P(z_{\mathbf{r}_t} = k \mid \mathbf{Y}, \hat{\underline{\theta}}^j)$

par $\tau_t(k)$ i.e.

$$\tau_t(k) = P\left(z_{\mathbf{r}_t} = k \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right),$$

$\tau_t(k)$ peut être dérivé en évaluant :

• **Les probabilités filtrées**

Basé sur la règle de Bayes pour $t = 1, \dots, n_1 n_2$, itérer sur des équations

$$P\left(z_{\mathbf{r}_t} = k \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_t}, \widehat{\underline{\theta}}^j\right) = \frac{P\left(z_{\mathbf{r}_t} = k \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_{t-1}}\right) f\left(\varepsilon_{\underline{\theta}^j}(\mathbf{r}_t)\right)}{\sum_{i=1}^L P\left(z_{\mathbf{r}_t} = i \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_{t-1}}\right) f\left(f\left(\varepsilon_{\underline{\theta}^j}(\mathbf{r}_t)\right)\right)},$$

et

$$P\left(z_{\mathbf{r}_{t+1}} = k \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_t}, \widehat{\underline{\theta}}^j\right) = \sum_{i=1}^L p_{ik} P\left(z_{\mathbf{r}_t} = i \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_t}, \widehat{\underline{\theta}}^j\right),$$

jusqu'à ce que $P\left(z_{\overleftarrow{\mathbf{r}_{1:n}}} = k \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right)$ soit calculé.

pour $t = 2, \dots, \overleftarrow{\mathbf{r}_{1:n}} - 1$ et $\tau = t + 1, \dots, \overleftarrow{\mathbf{r}_{1:n}}$

$$P\left(z_{\mathbf{r}_\tau} = k, z_{\mathbf{r}_t} = i \mid Y_{\mathbf{r}_1}, Y_{\mathbf{r}_2}, \dots, Y_{\mathbf{r}_\tau}, \widehat{\underline{\theta}}^j\right) = \frac{\sum_{l=1}^L P\left(z_{\mathbf{r}_{\tau-1}} = l, z_{\mathbf{r}_t} = i \mid Y_{\mathbf{r}_1}, \dots, Y_{\mathbf{r}_{\tau-1}}\right) f\left(\varepsilon_{\underline{\theta}^j}(\mathbf{r}_t)\right) p_{lk}}{P\left(Y_{\mathbf{r}_\tau} \mid Y_{\mathbf{r}_1}, \dots, Y_{\mathbf{r}_{\tau-1}}, \widehat{\underline{\theta}}^j\right)},$$

jusqu'à ce que $P\left(z_{\overleftarrow{\mathbf{r}_{1:n}}} = k, z_{\mathbf{r}_t} = i \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right)$ soit atteint.

• **Les probabilités lissées**

Pour $t = 1, \dots, \overleftarrow{\mathbf{r}_{1:n}} - 1$, itérer sur les équations

$$P\left(z_{\mathbf{r}_t} = i \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right) = \sum_{k=1}^L P\left(z_{\overleftarrow{\mathbf{r}_{1:n}}} = k, z_{\mathbf{r}_t} = i \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right).$$

Etape 3 : ($M - step$)

Dans cette étape, le paramètre de mise à jour estime $\widehat{\underline{\theta}}^{j+1}$ à la $j+1$ -ème itération est obtenu en maximisant la fonction $Q(\underline{\theta}, \underline{\theta}^j, \mathbf{Y})$ par rapport à $\underline{\theta}$. Donc 4.8 est maximisé en prenant $\widehat{\underline{\theta}}^{j+1}$ pour satisfaire

$$\bullet \hat{\rho}_{\mathbf{z}_{(1 \rightarrow \mathbf{p})}}^{j+1} = P\left(\mathbf{z}_{\mathbf{r}_1} = i_1, \dots, \mathbf{z}_{\mathbf{r}_{\overleftarrow{\mathbf{1}, \mathbf{p}}}} = i_{\overleftarrow{\mathbf{1}, \mathbf{p}}} \mid \mathbf{Y}, \widehat{\underline{\theta}}^j\right), i_1, \dots, i_{\overleftarrow{\mathbf{1}, \mathbf{p}}} = 1, \dots, L, \quad (4.9)$$

$$\bullet \hat{p}_{lk}^{j+1} = \frac{\sum_{t=\overleftarrow{\mathbf{1}, \mathbf{p}+1}}^{\overleftarrow{\mathbf{1}, \mathbf{n}}} \tau_t(k) \frac{p_{lk}^j \times P(z_{\mathbf{r}_{t-1}}=l \mid Y_{\mathbf{r}_1}, \dots, Y_{\mathbf{r}_{t-1}}, \widehat{\underline{\theta}}^j)}{P(z_{\mathbf{r}_t}=k \mid Y_{\mathbf{r}_1}, \dots, Y_{\mathbf{r}_{t-1}}, \widehat{\underline{\theta}}^j)}}{\sum_{t=\overleftarrow{\mathbf{1}, \mathbf{p}+1}}^{\overleftarrow{\mathbf{1}, \mathbf{n}}} \tau_{t-1}(l)}, l, k = 1, \dots, L, \quad (4.10)$$

$$\bullet \sum_{t=\overleftarrow{\mathbf{1}, \mathbf{p}+1}}^{\overleftarrow{\mathbf{1}, \mathbf{n}}} \tau_t(k) \left. \frac{\partial \log f(\varepsilon_{\underline{\theta}^j}(\mathbf{r}_t))}{\partial \mathbf{a}_k} \right|_{\mathbf{a}_k = \mathbf{a}_i^{j+1}} = 0, k = 1, \dots, L, \quad (4.11)$$

Etape 4 : Itérer vers la convergence.

à partir de $\widehat{\underline{\theta}}^0$, l'étape 2 produit les premières probabilités d'état lissées sur $\widehat{\underline{\theta}}^0$, tandis que la partie de l'étape 3 produit une estimation de paramètre de mise à jour $\widehat{\underline{\theta}}^1$ conditionnelle aux probabilités d'état lissé obtenues à l'étape 2. Les étapes 2 et 3 sont alternées de manière répétée jusqu'à la convergence (étape 4).

Remarque 48 : (convergence de l'algorithme VEM)

Par construction, le VEM dans notre cas s'assure qu'il ne fait que passer, de la valeur de $\widehat{\underline{\theta}}^{j+1}$ à la valeur de $\underline{\theta}^j$, alors la fonction de vraisemblance donnée par 4.8 est une fonction croissante (c'est-à-dire $Q(\widehat{\underline{\theta}}^{j+1}, \underline{\theta}^j, \mathbf{Y}) \geq Q(\underline{\theta}^j, \underline{\theta}^j, \mathbf{Y})$ avec égalité seulement si $\widehat{\underline{\theta}}^{j+1} = \underline{\theta}^j$). Par conséquent, la convergence de notre séquence d'estimation de paramètre

VEM peut être énoncée dans une manière assez similaire à celle du modèle $1D MS - AR$ donné dans Hamilton (1990).

Annexe A

Chaînes de Markov à temps discret

Cette annexe a pour objectifs de rappeler les définitions de base concernant les chaînes de Markov à temps discret. Les éléments exposés dans ce chapitre sont, de façon générale, tirés des articles et des ouvrages suivants : Rabiner (1989), Rabiner et Juang (1993), Weigend et Shi (1997), Mamon et Elliott (2007) et Bhar et Hamori (2004) et Howard (1971).

A.1 Définitions

Une chaîne de Markov est un réseau spécifique de variables aléatoires décrit par un système dynamique d'états de transitions.

Définition 49 : Soit $(S_t)_{t \in \mathbb{Z}}$ une suite de v.a $(\Omega, F, P) \rightarrow (\mathbb{S}, P(\mathbb{S}))$ est dite chaîne de Markov si $\forall t \in \mathbb{Z}, \forall s_0, s_1, \dots, s_t \in \mathbb{S}$ on a

$$P(S_t = s_t / S_{t-1} = s_{t-1}, \dots, S_0 = s_0) = P(S_t = s_t / S_{t-1} = s_{t-1})$$

On peut dire d'un processus S qu'il est markovien si son état actuel fournit toute l'information nécessaire pour connaître son évolution future. Sa distribution dans le futur étant donné le présent et le passé ne dépend que du présent. On parle alors d'absence de mémoire. L'ensemble \mathbb{S} est appelé l'espace des états.

La probabilité de transition

La probabilité de transition de l'état j à l'état i est notée p_{ij} tq : $p_{ij} = P(S_t = j / S_{t-1} = i)$
 $\forall i, j \in \mathbb{S}, t \in \mathbb{Z}$ et $p_{ij}^{(n)} = P(S_t = j / S_{t-n} = i)$ est appelée la probabilité de transition de l'état j à l'état i

à n^{ieme} étape.

Définition 50 Une chaîne de Markov est dite **homogène** lorsque la probabilité de transition ne dépend pas de t , c'est à dire : $p_{ij} = P(S_t = j / S_{t-1} = i) = P(S_1 = j / S_0 = i)$,
 $\forall t \in \mathbb{Z}$

Remarque 51 On dira de plus que la chaîne de Markov est homogène si $p_{ij}^{(n)} = p_{ij}$,
 $\forall n \geq 1$.

La définition permet alors de caractériser une chaîne de Markov homogène à l'aide d'une matrice de transition et d'un vecteur d'états initiaux.

Matrice de transition

On appelle une matrice de transition toute matrice \mathbb{P} tel que $\mathbb{P} = (p_{ij})_{(i,j \in \mathbb{S})}$, telle que $0 \leq p_{ij} \leq 1$ et $\sum_{j \in \mathbb{S}} p_{ij} = 1$. Elle est dite stochastique.

Exemple 52 Supposons l'espace d'états $\mathbb{S} = \{1, 2, 3\}$ avec des probabilités de transitions $p_{11} = p_{12} = p_{13} = \frac{1}{3}$, $p_{21} = 0$; $p_{22} = p_{23} = \frac{1}{2}$, $p_{31} = p_{32} = 0$ et $p_{33} = 1$. La matrice de

transition est :

$$\mathbb{P} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

Les états sont représentés par des cercles numérotés et les probabilités de transitions positives par des flèches.

Exemple 53 *L'étude statistique du phénomène de la pluie sur une ville donnée a été, convenablement, caractérisée ou décrit par les chaînes de Markov à 2 états :*

L'état 1 représente les jours secs.

L'état 2 simule les jours pluvieux.

Sous les 2 hypothèses suivantes :

Si le jour actuel est pluvieux alors la probabilité que le lendemain sera pluvieux est de $p_{22} = 0.8$.

Si le jour actuel est sec, alors la probabilité que le lendemain sera pluvieux est de $p_{12} = 0.4$.

De sorte que l'on peut écrire la matrice de transition comme suit :

$$\mathbb{P} = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

L'analyse de cette matrice de transition donne ce qui suit :

1) $P_{11} = 0.6$ caractérise l'état que si aujourd'hui est sec, alors le lendemain restera sec avec la probabilité de 0.6

2) $p_{21} = 0.2$ représente l'état que si aujourd'hui est pluvieux, alors le lendemain sera sec avec la probabilité de 0.2.

Théorème 54 $\forall n \geq 0$ $\mathbb{P}^{(n)} = \mathbb{P}^n$ la matrice $\mathbb{P}^{(n)}$ est appelée la matrice de transition en n étapes $\mathbb{P}^{(n)} = \left(p_{ij}^{(n)} \right)_{(i,j \in \mathbb{S})}$.

Remarque 55 On a évidemment $p_{ij}^{(1)} = p_{ij}$ et $\mathbb{P}^{(1)} = \mathbb{P}$. Convention $\mathbb{P}^{(1)} = I$ c-à-d

$$p_{ij}^{(1)} = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases} .$$

Théorème 56 Equation de Chapman –Kolmogorov

Pour tout état $i, j \in \mathbb{S}$, pour tout $n, m \in \mathbb{N}$ et $k \in [0, n]$, on a l'égalité $p_{ij}^{(n+m)} = p_{ik}^{(m)} p_{kj}^{(n)}$.

En notation matricielle on a $\mathbb{P}^{(m+n)} = \mathbb{P}^{(m)}\mathbb{P}^{(n)}$.

A.2 Classification des états

Il existe différents types d'états. Leur classification permet de mieux étudier les propriétés asymptotiques des chaînes de Markov. Maintenant, classifions les états possibles selon diverses caractéristiques.

A.2.1 Communiquent

Définition 57 Un état i est dit **récurrent** si, $P(\exists n \geq 1, S_n = i / S_0 = i) = 1$. C'est à dire que la probabilité d'un éventuel retour à l'état i vaut 1, sachant que la chaîne a commencé à l'état i . Sinon, on dit que l'état est **transitoire** ; .

Un état i est **récurrent nul** si, $\rho_i = E[R_i / S_0 = i] = \infty$ où $R_i = \min \{r : S_r = i\}$. Sinon, l'état i est **récurrent positif**.

Définition 58 Soient $i, j \in \mathbb{S}$ et $n \geq 1$ un entier. La probabilité du premier temps de passage à l'état j , au n -ième pas, sachant que le processus démarre à l'état i est définie par :

$$f_{ij}^{(n)} = P(S_n = j, S_k \neq j, k = 1, \dots, n-1 / S_0 = i) = P(R_i = n / S_0 = i) \quad \forall n \geq 1.$$

Par convention $f_{ij}^{(0)} = 0$.

Définition 59 Soient i et j deux états de \mathbb{S} . L'état j est **accessible** depuis l'état i , noté $i \rightarrow j$, si $\exists n \in \mathbb{N}, p_{ij}^{(n)} = P(S_n = j / S_0 = i) > 0$

On dit que les états i et j **communiquent** si ils sont tous deux accessibles l'un de l'autre. On note $i \longleftrightarrow j$

Proposition 60 Les états i et j communiquent ssi il existe $n > 0$ et $m > 0$, tel que :
 $p_{ij}^{(n)} > 0$ et $p_{ji}^{(m)} > 0$

Proposition 61 La relation \longleftrightarrow est une relation d'équivalence sur \mathbb{S} . Elle est

- réflexive : i communique avec i ;
- symétrique : $\forall i, j \in \mathbb{S}, i \longleftrightarrow j$ si $j \longleftrightarrow i$;
- transitive : $\forall i, j, k \in \mathbb{S}$, si $i \longleftrightarrow j$ et $j \longleftrightarrow k$ alors $i \longleftrightarrow k$.

Définition 62 Les états i est absorbant ssi $p_{ii} = 1$ (Ou a alors $p_{ij} = 0, \forall i \neq j$)

A.2.2 Irréductible

Définition : Une chaîne de Markov $(\mathbb{S}_t)_{t \in \mathbb{Z}}$ à espace d'état \mathbb{S} et la matrice de transition \mathbb{P} est dite irréductible si pour tous $i, j \in \mathbb{S}$ nous avons que $i \longleftrightarrow j$. Sinon la chaîne est dite réductible.

Proposition : Une chaîne de Markov est irréductible ssi toutes ses paires d'états communiquent.

A.2.3 Périodicité

Définition 63 La période d'un état $i \in \mathbb{S}$ notée $d(i)$ est l'entier définie par :

$$d(i) = \text{PGCD} \left\{ n \geq 1 : p_{ii}^{(n)} > 0 \right\}$$

On dit que i est périodique si $d(i) > 1$, sinon il est apériodique.

A.2.4 Distribution initiale

La distribution des états d'une chaîne de Markov après n transitions est notée $\pi(n)$ tq : $\pi(n) = P(S_n = i)$ et la distribution initiale $\pi(0)$ est $\pi(0) = P(S_0 = i) \quad \forall i \in \mathbb{S}$

Théorème 64 Soit \mathbb{P} la matrice de transition d'une chaîne de Markov et $\pi(0)$ la distribution initiale,

pour tout $n \geq 1$ on a :

$$\begin{cases} \pi(n) = \pi(0)\mathbb{P}^{(n)} \\ \pi(n) = \pi(n-1)\mathbb{P} \end{cases}$$

A.2.5 Distribution stationnaire

Définition 65 Un vecteur $\pi = \{\pi_i\}_{i \in \mathbb{S}}$ sur \mathbb{S} est stationnaire pour la chaîne de Markov S si

$$(i) \pi_i \geq 0, \forall i \in \mathbb{S} \text{ et } \sum_{i \in \mathbb{S}} \pi_i = 1$$

$$(ii) \pi_j = \sum_{i \in \mathbb{S}} \pi_i p_{ij} \quad \forall j \in \mathbb{S} \text{ ou en notation matricielle } \pi = \pi \mathbb{P}.$$

Remarque 66 *La stationnarité de cette distribution se voit en itérant l'égalité de l'assertion (ii), soit $\pi \mathbb{P}^2 = (\pi \mathbb{P}) \mathbb{P} = \pi \mathbb{P} = \pi \mathbb{P}$. De la même manière, on a $\pi \mathbb{P}^n = \pi$, $\forall n \in \mathbb{N}$.*

A.2.6 Ergodique

Une chaîne de Markov irréductible et apériodique est dite ergodique lorsque tous ses états sont récurrents positifs et apériodique, et non-ergodique lorsque tous ses états sont transitoires ou récurrents nuls.

Théorème 67 (ergodique) *Soit une chaîne de Markov irréductible et apériodique $S = \{S_n\}$ $n \geq 1$. Alors*

$$1. \lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{f_{ij}}{\rho_i}$$

2. Sous les mêmes conditions,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum p_{ij}^{(n)} = \frac{f_{ij}}{\rho_i}, \text{ pour tous les entiers } i, j \text{ dans } \mathbb{S}.$$

Annexe B

Modèles de Markov cachés à temps discret

B.1 Présentation et caractéristiques

Certains phénomènes peuvent se décrire adéquatement par des chaînes de Markov. Cela suppose alors que les états et les probabilités de transition sont bien connus. Toutefois il est fréquent d'observer une fonction de ces états, et plus généralement une variable aléatoire associée aux états. On parle alors de Modèles de Markov Cachés (MMC). On peut les décrire comme des fonctions probabilistes d'une chaîne de Markov.

Brièvement, un *MMC* à temps discret peut se définir comme une modélisation doublement stochastique : un processus dit «caché» parfaitement modélisé par une chaîne de Markov discrète et un processus observable dont la distribution dépend des états du processus caché. on s'intéresse particulièrement aux MMC de premier ordre.

B.2 Caractéristiques des MMC

Nous introduisons les caractéristiques des MMC. Pour faire cela, considérons les éléments notés N , M , A , B , μ et λ définie comme suit :

1. λ est un ensemble paramétrique,
2. N est le nombre d'états cachés dans le modèle. On note l'ensemble d'états cachés par $S = \{S_1, \dots, S_N\}$ et l'état au temps t par y_t ,
3. M est le nombre de symboles distincts observables par états. On note ces symboles par o_k où $k = 1, 2, \dots, M$, et l'observation au temps t par O_t ;
4. $A = \{a_{ij}\}$ est la matrice de transition des états cachés où

$$a_{ij} = P\left(y_{t+1} = S_j \mid y_t = S_i, \tilde{\theta}\right) \text{ avec } i = 1, 2, \dots, N \text{ et } j = 1, 2, \dots, M$$

5. La matrice B contient les probabilités d'observer au temps t le symbole k sachant qu'au même instant le modèle est dans l'état caché S_i .

6. $\mu = \{\mu_i\}$ la distribution de l'état initial du modèle où

$$\mu_i = P\left(y_1 = S_i \mid \tilde{\theta}\right), \quad i = 1, \dots, N$$

Ce vecteur contient la probabilité qu'au moment initial ($t = 1$), le modèle se trouve dans l'état caché S_i .

Un MMC est un quintuplet $\tilde{\theta}$ qui se définit par

$$\tilde{\theta} = (N, M, \mu, A, B)$$

Soit T le nombre d'observations. On définit respectivement par $O(T) = O_1O_2\dots O_T$ et $y(T) = y_1y_2\dots y_T$ des séquences d'observations ainsi que d'états cachés pouvant être obtenues à partir du modèle $\tilde{\theta}$

Remarque 68 *La notation des MMC est très souvent réduite au triplet $\tilde{\theta} = (\mu, A, B)$ car A est une matrice $N \times N$, et B une matrice $N \times M$.*

Bibliographie

- [1] Abend, K. Harley, T.J and Kanal, L.N. (1965). "Classification of binary random patterns," IEEE Trans. Information Theory. IT-11, pp.538-544.
- [2] Abdullah, S. and Shitan, M. (2009). Some explicit conditions for a stationary representation of unilateral second-order spatial ARMA model, Pertanika J. Sci. Technol. **17** , no. 1, 163–171.
- [3] Amdaoud. M , Arcuri. G , Levratto. N « Covid-19 : analyse spatiale de l'influence des facteurs socio-économiques sur la prévalence et les conséquences de l'épidémie dans les départements français ». Codes JEL : C21, I14, R23 EconomiX, université Paris Nanterre.
- [4] Bartlett, M. S. (1971). Physical nearest-neighbour models and non-linear time-series, J. Appl. Probab. **8**, no. 2, 222–232.
- [5] Basli, **S. and Reinsel, G. C. (1994)** , Regression models with spatially correlated errors, Journal of the American Statistical Association : Theory and Method, **89**(425), pp. 88-99.
- [6] Basu, S. and Reinsel, G. C. (1992). **A note on** properties of spatial Yule-Walker estimators, Journal of Statistical Computing and Simulation, **41**, pp. 243-255.

- [7] Basu, S. and Reinsel, G. C. (1993). Properties of the spatial unilateral first order ARMA model, *Advances in Applied Probability*, **25**, pp. 631-648.
- [8] Basu, S. and Reinsel, G. C. (1994). Regression models with spatially correlated errors, *J. Amer. Statist. Assoc.* **89** , no. 425, 88–99.
- [9] Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society B*, **36**, pp. 192-236.
- [10] Besag, J. E. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments, *J. Roy. Statist. Soc. Ser. B* **61** , no. 4, 691–746
- [11] Bhar et . Hamori, S. (2004) " Hidden Markov Models : Applications to Financial Economies". Springe
- [12] Bhattacharyya, B. B. Khalil T. M. and Richardson, G. D. (1996) Gauss–Newton estimation of parameters for a spatial autoregression model, *Statist. Probab. Lett.* **28** , no. 2, 173–179.
- [13] Bishop, C.M. (2006) ' Pattern recognition and machine learning', Springer.
- [14] Barros, C. P. Chen, Z. and Gil-Alana, L. A. (2013). Long memory in the housing price indices in China. *Asian Journal of Empirical Research*, **3** (7), 785–807.
- [15] Bouchaffra, D. and Meunier, Markovia, J. (1995). An Random Field approach to information retrieval, in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 997– 1002.
- [16] Chalmond, B. (1989). An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recognition*, **22**, pp. 747–762.

- [17] Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability*, **18**, 211-220.
- [18] Bustos, O. H. Fraiman, R. and Yohai, V. J. (1984) Asymptotic behavior of the estimates based on residual autocovariances for ARMA models, in : *Robust and Nonlinear Time Series Analysis*, Springer, New York , 26–49.
- [19] Bustos, O. Ojeda, S and Vallejos, R. (2009). "Spatial ARMA models and its applications to image filtering" *Brazilian Journal of Probability and Statistics*. **23**, No. 2, 141–165
- [20] Champagnat, F. Idier, J. and Goussard, Y. (1998). "Stationary Markov random fields on a finite rectangular lattice," *IEEE Trans. Inf. Theory*. **44**, no. 11, pp. 2901–2916, Nov.
- [21] Chalmond, B. (1989). An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recognition*, **22** , pp. 747–762.
- [22] Chellappa, R. , (1985) "Two-dimensional discrete Gaussian Markov random field models for image processing," *Progress in Pattern Recognition 2*, L.N.Kanal and A. Rosenfeld, Eds. Elsevier Science Publishers B.V.
- [23] Chellappa, R. and Jain, A. (1993). *Markov Random Fields : theory and applications*, Academic Press, inc.
- [24] Chen, J. Chen, Z. (2019) A new optimal multi-product (Q, R, SS) policy with multivariate Markov stochastic demand forecasting model. *International Journal of Mathematics in Operational Research (IJMOR)*, **14** No.1 DOI : 10.1504/IJMOR.2019.096980

- [25] Cohen, Joel, E. (1979). Ergodic theorems in demography. Bull. Amer. Math. Soc. (N.S.) **1** (2), 275–295.
- [26] Cressie, N. A. C. (1993), Statistics for Spatial Data, Revised Edition, Wiley, New York.
- [27] Cullis, B. R and A. C. Gleeson, (1991). Spatial analysis of field experiments-an extension to two dimensions, Biometrics **47**, 1449–1460
- [28] Demster, A.P., Lair, N.M., Rubin, D.B. (1977) 'Maximum likelihood from incomplete data via the E.M. Algorithm. J. R', Stat. Soc. Ser. B Stat, **39**, pp.1–38.
- [29] Derin, H., H. Elliott, R. Cristi, and D. Geman, (1984). "Bayes smoothing algorithms for segmentation of binary images modelled by Markov random fields," IEEE Tran. Pattern Analysis and Machine Intelligence, **6**. pp.707-720, Nov.
- [30] Derin H. and Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, **9** , pp. 39–55.
- [31] Devijver, P.A. and M.M. Dekesel, (1987). "Learning the parameters of a hidden Markov random field image model : a simple example," Pattern Recognition Theory and Applications, Edited by P.A. Devijver and J. Kittler, NATO ASI Series, **vol.F30**, Springer-Verlag.
- [32] Djafri. H et Kharfouchi. S (2021) ."On unilateral 2D Markov-Switching Autoregressive 132 (MS-AR) model". Int. J. Mathematics in Operational Research. Vol. 18 n 4, No. 10.1504/IJMOR.10033953

- [33] Drapatz, M.(2016) Strictly stationary solutions of spatial ARMA equations, *Ann. Inst. Statist. Math.* **68** , no. 2, 385–412.
- [34] Eickeler, S.(2002) "Face database retrieval using pseudo 2D hidden Markov models," *Proc. of the fifth IEEE Int. Con.,{ on Auto. Face and Gesture Recog.,*
- [35] El-Hadidy, M., A., A. (2019). Generalised linear search plan for a D-dimensional random walk target. *International Journal of Mathematics in Operational Research (IJMOR)*, **15**, No. 2, DOI : 10.1504/IJMOR.2019.101622
- [36] Etchison T., Pantula S. G., and Brownie C. (1994). Partial autocorr elation function for spatial processes. *Statist. Probab. Lett.*, **21**(1) : 9-19.
- [37] Fass' o, A. & Finazzi, F. (2011). Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics*, **22** (6), 735–748.
- [38] Fass' o, A., Cameletti, M. & Nicolis, O. (2007). Air quality monitoring using heterogeneous networks. *Environmetrics*, **18** (3), 245–264.
- [39] Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis*, **3** (1), 27–44.
- [40] Forbes, F. and Vignes, M. (2006). Integrated Markov models for clustering genes combining individual features and pairwise relationships, in 4th workshop on Statistical methods for post-genomic data, .
- [41] Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, **12** (5), 469–483.
- [42] Gaetan,C.andGuyon,X.(2008) *Mod elisationetstatistiquespatiale*. Springer.

- [43] : Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, **6**, 721–741.
- [44] Genton, M. G. and Koul, H. L. (2008). Minimum distance inference in unilateral autoregressive lattice processes, *Statist. Sinica* **18** , 617–631
- [45] Goldfeld, S.M. and Quandt, R.E. (1973). A Markov Model for Switching Regressions. *Journal of Econometrics*, **1**, 3-16.
- [46] Goryainov, V. B. (2011). Least-modules estimates for spatial autoregression coefficients, *J. Comput. Syst. Sci. Int.* **50** , no. 4, 565–572.
- [47] Goutsias, J. (1989). “Mutually compatible Gibbs random fields,” *IEEE Trans. Inf. Theory*, **35**, pp. 1233–1249.
- [48] Gravier, G. Sigelle, M. and Cholle, G. (2000). A Markov Random Field model for speech recognition, in *International Conference on Pattern Recognition*.
- [49] Guyon, X. (1995). *Random... eldsonanetwork. Probabilityanditsapplications*. Springer, NewYork.
- [50] Hammersley, J. M., & Clifford, P. (1971), Markov fields on finite graphs and lattices.
- [51] Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *J. Econometrics*, **45**, 39–70.
- [52] Hansen, R. R. and Chellappa, R. (1990). Noncausal 2-D spectrum estimation for direction finding. *IEEE Trans. Inf. Theory*, IT-**36** (1), 108–125.
- [53] Heitz, F., and Maitre, H. (1987). Application of autoregressive models to fine arts painting analysis. *Signal pprocessing*. **13**, 1-14.

- [54] Hilbert, D. 1991 "Über die stetige abbildung einer linie aufFlächenstück," *Math., Ann.*, **38**, pp.459-460.
- [55] Holly, S., Pesaran, M. H. & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics* **158**, 160–173.
- [56] Holly, S., Pesaran, H. M., & Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, **60**(1), 2–23.
- [57] Ho, M.-H. R., Ombao, H. & Shumway, R. (2005). A state-space approach to modelling brain dynamics. *Statistica Sinica*, 407–425.
- [58] Hung, Y. Cooper, D. and Cernuschi-Frias, B. (1991). Asymptotic Bayesian surface estimation using an image sequence, *International Journal of Computer Vision*, **6** , pp. 105–132.
- [59] Ising. E. (1925) Beitrag zur theorie des ferromagnetims. *Zeitschrift fur Physik*, 31 :253–258.
- [60] Jeng, F.-C. and Woods, J.W. (1987). "On the relationship of the Markov mesh to the NSHP Markov chain," *Pattern Recognition Letters*. **5**, pp.273-279,
- [61] Jeng, F.C. and J.W. Woods, "Compound Gauss-Markov random fields for image estimation," *IEEE Tran. Signal Processing*. **39**, no.3, pp.683-{}97, March 1991
- [62] Julie Le Gallo. 2000. Econométrie spatiale 1 -autocorrélation spatiale. LATEC - Document de travail - Economie (1991-2003) 2000-05, LATEC, Laboratoire d'Analyse et des Techniques Economiques, CNRS UMR 5118, Université de Bourgogne. **2**, **3**, 20, 22.

- [63] Kana!, L.N.(1980). "Markov mesh models," in Image Modeling, New York : Academic, pp.239-243.
- [64] Kashyap, R. L.and Chellappa, R. (1983)."Estimation and choice of neighbors in spatial-interaction models of images," IEEE Trans. Inf. Theory, IT-**29** (1),60–72.
- [65] Kaufman, H. and Tekalp,A. M. (Jan. 1991). "Survey of estimation techniques in image restoration," IEEE Control Syst. Mag. **11**, no. **1**, pp. 16–24, .
- [66] Kharfouchi. S. (**2008**). " Statistical Inferencein Spatial Bilinear Processes"
- [67] Kokaram, A. (2004). A statistical framework for picture reconstruction using 2D AR models. Image Vision Computing. **22**, 631-648.
- [68] Kuo, S.-S. and O. Agazzi, (August 1994) "Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models," IEEE Tran. Pattern Analysis and Machine Intelligence, .**16**, no. 8, pp.842-848.
- [69] Li, S. Z. (1995)."Markov Random Field Modelling in Computer Vision", Springer.
- [70] Lin, H.-C., Wang, L.-L and Yang S.-N. ,(1997) "Color image retrieval based on hidden Markov models," IEEE Trans. Image Processing, **6**, no.2, pp.332-339.
- [71] Lu, S. W., and Xu, H. (1995).Textured image segmentation using autoregressive model and artificial neural network," Pattern Recognition, **28** (12), 1807–1817.
- [72] MacNab, Y. C., & Dean, C. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. Biometrics, **57** (3), 949–956.
- [73] Mamon, R.S et Elliott, R. J (2007) " Hidden Markov Models in Finance. Springer Science+Business Media, LLC.

- [74] McLachlan G. J. and Krishnan, T. (1997). EM Algorithm and Extensions, Wiley.
- [75] Metzler, D. and Croft, W. B. (2005). A Markov Random Field model for term dependencies, in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 472–479.
- [76] Mhidra, H., Brochard, J., and Leard, M. (1993). AR models and bidimensional discrete moments applied to texture modeling and recognition. *Pattern Recognition*, **26** (5), 721–726.
- [77] Morton, G.M. (1966). "A computer oriented geodetic data base and a new technique in file sequencing," Tech. Rep. IBM, Ottawa, Ont., Canada.
- [78] Nefian, A.V. and M.H. Hayes III, (2000) "Maximum likelihood training of the embedded HMM for face detection and recognition," Proc. of IEEE Int. Conf on Image Processing. **1**, pp.33-36.
- [79] Ngene, G. M., and Lambert, C. A., and Darrat, A. F. (2015). Testing Long Memory in the Presence of Structural Breaks : An Application to Regional and National Housing Markets. *J Real Estate Finan Econ*, **50**, 465–483.
- [80] Ni, Y., Wang, K., and Zhao, L. (2017) A Markov decision process model of allocating emergency medical resource among multi-priority injuries. *International Journal of Mathematics in Operational Research (IJMOR)*. **10**, No. 1 DOI : 10.1504/IJMOR.2017.080738.
- [81] Oe, S. (1993). Texture segmentation method using two-dimensional AR model and Kullback information. *Pattern Recognition*, **26** (12), 237–244.
- [82] Oukil. A (2007). Analyse variographique, modélisation et synthèse de textures appli-

quées aux image numériques. Thèse de doctorat, Université des science et de la technologie 'Houari Boumediene', Algérie,

- [83] Peano, G. (1890). "Sur une courbe qui remplit toute une aire plane," *Math., Ann.* **36**, pp.157-160.
- [84] Pickard, D. K. (1977). "A curious binary lattice process," *J. Appl. Probab.* **14**, pp. 717–731.
- [85] Pickens, J. and Iliopoulos, C. (2005). Markov Random Fields and Maximum Entropy modeling for music information retrieval, in *Proceedings of the 6th International Conference on Music Information Retrieval*, pp. 207–214.
- [86] Politis, D. N. (1994). A simple information theoretic proof of the maximum entropy property of some Gaussian random fields," *IEEE Trans. Image Process.*, **3** (6), 865–868.
- [87] Preuss, D.(1975). "Two-dimensional facsimile source coding based on a Morkov model," *NTZ* **28**, pp. 358–368.
- [88] RABINER , L.R (1989) " A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **14** p. 257-286.
- [89] Rabiner, L. R.et Juang, B.H (1986) " An introduction to hidden markov models". *IEEE ASSP Magazine*, **16** (3). 4-16.
- [90] Saquib,S. Bouman, C. and Sauer, K. (1998). ML parameter estimation for Markov Random Fields, with applications to Bayesian tomography, *IEEE Transactions on Image Processing*, **7**, pp. 1029–1044.
- [91] Sarkar, A., Sharma, K. M. S., and Sonak, R. V. (1997). A new approach for subset

- 2-D AR model identification for describing textures. *IEEE Trans. Image Process.*, **6** (3), 407–413.
- [92] Scherrer, B. Dojat, M. Forbes, F. and Garbay, Locs, C. (2007). Local Cooperative Unified Segmentation of MRI brain scans, in MICCAI, Brisbane, Australia.
- [93] Sharma, G., and Chellappa, R. (1986). Two-dimensional spectrum estimation using noncausal autoregressive models. *IEEE Trans. Inf. Theory*, IT-**32** (2), 268–275.
- [94] Shinkareva, S. V., Ombao, H. C., Sutton, B. P., Mohanty, A. & Miller, G. A. (2006). Classification of functional brain images with a spatio-temporal dissimilarity map. *NeuroImage*, **33** (1), 63–71.
- [95] Tjostheim, D. (1978). Statistical spatial séries modelling. *Advances in Appl. Probability*, **10**(1) :130-154.
- [96] Tjostheim, D. (1983). Statistical spatial séries modelling. *IL Some further results on unilatéral lattice processes. Adv. in Appl. Probab.*, **15**(3) :562-584.
- [97] Weigend, A. S et S. Shi ,S. (1997) "Taking time seriously : Hidden markov experts applied to financial engineering". *Proceedings of the IEEE/IAFE*,**18** p. 244-252, .
- [98] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. New York, 7.
- [99] Whittle. P.(1954). On stationary processes in the plane. *Biometrika*, **41**, 434–449, . 35, 63, 90.
- [100] Xavier Guyon and Carlo Gaetan. (2008)*Modélisation et statistique spatiales*. Springer . ISBN : 978-3-540-36710-1.

Résumé

Cette thèse aborde le problème de la perte d'informations et de caractéristiques importantes dans la modélisation des données spatiales en construisant un modèle spatial causal qui peut capturer les principales caractéristiques de ces données. Après avoir fait le tour de la question, une stratégie en deux étapes a été proposée : premièrement, un champ aléatoire de Markov 2D (MRF) est généré où la causalité imposée permet d'établir une analogie entre ce champ MRF bidimensionnel et une représentation en chaîne de Markov. Ensuite, sur la base du champ MRF 2D proposé, le modèle 2D MS-AR est défini selon certaines hypothèses essentielles et symboles utiles.

Enfin, l'estimation des paramètres du modèle est discutée, ce qui ouvre la voie à de larges perspectives pour exploiter les processus 2D MS-AR proposés pour modéliser efficacement plusieurs phénomènes qui présentent une discontinuité structurelle dans la dépendance spatiale des données.

Mots clés : 2D-AR, Traitement d'image, Modèles spatiaux, Processus autorégressifs unilatéraux spatiaux, MRF causal, 2D MS-AR.

Abstract

This thesis deals with the problem of loss of important information and features in spatial data modeling by building a causal spatial model that can capture the various main characteristics of these data. After a through discussion, a two-step strategy has been proposed : first, a 2D Markov Random Field (MRF) is generated where imposed causation allows to establish an analogy between this 2D MRF and a Markov chain representation ; secondly, based on the proposed 2D MRF, 2D MS-AR is defined according to some essential assumptions and useful symbols.

Finally, estimation of model parameters is discussed, which opens the way for broad perspectives to exploit the proposed 2D MS-AR processes to efficiently model several phenomena that present a structural discontinuity in the spatial dependence of the data.

Keywords : AR-2D Models, Image Processing , Spacial Models, Spatial unilateral autoregressive processes, MRF causal, 2D MS-AR.

ملخص

- تتناول هذه الرسالة مشكلة فقدان المعلومات الهامة في نمذجة البيانات المكانية. من خلال بناء نموذج مكاني سببي يمكنه التقاط الخصائص الرئيسية المختلفة لهذه البيانات بعد مناقشة، تم اقتراح استراتيجية من خطوتين: أولاً، يتم إنشاء حقل ماركوف العشوائي ثنائي الأبعاد (MRF) حيث تسمح السببية المفروضة بإقامة تشابه بين هذا ثنائي الأبعاد MRF وتمثيل سلسلة ماركوف ثانياً، بناء على 2D MFR المقترح، يتم تعريف 2D MS-AR وفقاً لبعض الافتراضات الأساسية والرموز المفيدة.
- أخيراً، تمت مناقشة تقدير معلمات النموذج، مما يفتح الطريق لوجهات نظر واسعة لاستغلال عمليات MS-AR المقترحة ثنائية الأبعاد لنمذجة العديد من الظواهر بكفاءة والتي تمثل انقطاعاً هيكلياً في الاعتماد المكاني للبيانات.
- **كلمات المفتاحية:** نموذج الانحدار الذاتي- عمليات الانحدار الذاتي المكاني من جانب واحد- النماذج المكانية - الانحدار التلقائي المكاني- حقل ماركوف العشوائي السببي- معالجة الصور