Outils et modèles pour l'étude de quelques risques spatiaux et en réseaux : application aux extrêmes climatiques et à la contagion en finance

Erwan Koch

ISFA UNIVERSITÉ LYON 1
ÉCOLE DOCTORALE SCIENCES ÉCONOMIQUES ET DE GESTION

OUTILS ET MODÈLES POUR L'ÉTUDE DE QUELQUES RISQUES SPATIAUX ET EN RÉSEAUX : APPLICATION AUX EXTRÊMES CLIMATIQUES ET A LA CONTAGION EN FINANCE

TOOLS AND MODELS FOR THE STUDY OF SOME SPATIAL AND NETWORK RISKS : APPLICATION TO CLIMATE EXTREMES AND CONTAGION IN FINANCE

THÈSE

pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITÉ CLAUDE BERNARD LYON I
Spécialité : MATHÉMATIQUES APPLIQUÉES

présentée et soutenue publiquement par

Erwan KOCH

2 Juillet 2014

JURY

Directeurs de thèse	**Christian Robert** Professeur, ISFA, Université Lyon 1 **Pierre Ribereau** Maître de Conférences, ISFA, Université Lyon 1
Rapporteurs	**Hansjörg Albrecher** Professeur, HEC Lausanne **Jean-Noël Bacro** Professeur, Université Montpellier 2
Examinateurs	**Anne-Laure Fougères** Professeur, Université Lyon 1 **Christian Gouriéroux** Professeur, Université de Toronto et Université Paris-Dauphine **Ragnar Norberg** Chaire CARDIF Management de la modélisation

L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs.

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. François-Noël GILLY**
Vice-président du Conseil d'Administration	M. le Professeur Hamda BEN HADID
Vice-président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur Philippe LALLE
Vice-président du Conseil Scientifique	M. le Professeur Germain GILLET
Directeur Général des Services	M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. MATILLON
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme. la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. FLEURY
Département Chimie Biochimie	Directeur : Mme Caroline FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. Georges TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. Jean-Claude PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. P. FOURNIER
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Remerciements

Mes remerciements vont en premier lieu à mes directeurs de thèse, Christian et Pierre, tant pour leur disponibilité que pour la liberté qu'ils m'ont donnée dans le choix de mes sujets de recherche. Leurs encouragements ainsi que leurs remarques et critiques toujours très constructives m'ont été précieux. Un grand merci également à Paul Embrechts pour le temps qu'il m'a consacré ainsi que ses nombreux commentaires extrêmement pertinents. Un chaleureux merci à Philippe Naveau, co-auteur de l'un des travaux, qui a su me transmettre son optimisme, sa passion pour les statistiques et son goût pour les îles paradisiaques. Merci également à Sylvie Joussaume de m'avoir aiguillé vers ce projet de thèse. Naturellement, je remercie le Ministère de l'Ecologie, qui, via le projet MIRACCLE (Mesures et Indicateurs de Risques Adaptés au Changement Climatique), a financé cette thèse et m'a permis de participer à de nombreuses et passionnantes conférences.

Ma profonde reconnaissance va aussi au Laboratoire de Finance-Assurance du CREST qui m'a hébergé durant ces années de thèse, m'offrant des conditions de travail royales. En particulier, j'adresse tous mes remerciements à Christian Gouriéroux et Jean-Michel Zakoian ainsi qu'à Luciano Campi, Jean-David Fermanian, Christian Francq, Gaëlle Le Fol, Bertrand Villeneuve et Olivier Wintenberger pour leurs formidables qualités humaines, nos enrichissantes discussions scientifiques et l'intérêt manifesté à l'égard de mon travail. Ils sont pour moi de véritables exemples. Un grand merci également au Laboratoire SAF pour son accueil lors de mes visites.

Naturellement, je souhaite remercier chaleureusement Hansjörg Albrecher et Jean-Noël Bacro pour avoir accepté de rapporter ma thèse. Leur constante bonne humeur et leurs qualités scientifiques forcent mon admiration. Toutes mes amitiés respectueuses vont à Anne-Laure Fougères, Christian Gouriéroux et Ragnar Norberg qui ont bien voulu siéger dans mon jury.

Un grand merci également à mes amis doctorants du CREST : Céline, pour son enthousiasme et ses précieux conseils au début de ma thèse ; Giuseppe, les deux Guillaume, Gulten, Isabelle, Jean-Cyprien, Jérémy, Mathilde, Sophie et Yang pour des moments de détente fort sympathiques. Je remercie en particulier Jean-Cyprien pour une première collaboration très enrichissante. Je remercie également sincèrement l'ensemble des doctorants du Laboratoire SAF et plus particulièrement leur représentant, Julien, pour son dévouement.

Evidemment, un grand merci à toute ma famille : mes parents, ma grand-mère, mon frère et sa famille, ... qui ont toujours su me soutenir, m'encourager et m'entourer d'affection. Naturellement, un grand merci à tous les amis fidèles qui ont su m'accompagner pendant cette période. Parmi mes amis à deux pattes, je pense notamment à Randy, Henry, Trudy et toute la bande. Je n'oublie pas mes amis à quatre pattes, Ambre, Noisette, Stellou et Paddy. Enfin, un doux merci à Hélène, qui a partagé avec moi la dernière partie de cette thèse, qui m'a toujours empli de son affection, attention et délices culinaires et qui m'a transmis son infaillible motivation.

Résumé

Cette thèse s'attache à développer des outils et modèles adaptés à l'étude de certains risques spatiaux et en réseaux. Elle est divisée en cinq chapitres. Le premier consiste en une introduction générale, contenant l'état de l'art au sein duquel s'inscrivent les différents travaux, ainsi que les principaux résultats obtenus.

Le Chapitre 2 propose un nouveau générateur de précipitations multi-site. Il est important de disposer de modèles capables de produire des séries de précipitations statistiquement réalistes. Alors que les modèles précédemment introduits dans la littérature concernent essentiellement les précipitations journalières, nous développons un modèle horaire. Il n'implique qu'une seule équation et introduit ainsi une dépendance entre occurrence et intensité, processus souvent considérés comme indépendants dans la littérature. Il comporte un facteur commun prenant en compte les conditions atmosphériques grande échelle et un terme de contagion auto-régressif multivarié, représentant la propagation locale des pluies. Malgré sa relative simplicité, ce modèle reproduit très bien les intensités, les durées de sécheresse ainsi que la dépendance spatiale dans le cas de la Bretagne Nord.

Dans le Chapitre 3, nous proposons une méthode d'estimation des processus max-stables, basée sur des techniques de vraisemblance simulée. Les processus max-stables sont très adaptés à la modélisation statistique des extrêmes spatiaux mais leur estimation s'avère délicate. En effet, la densité multivariée n'a pas de forme explicite et les méthodes d'estimation standards liées à la vraisemblance ne peuvent donc pas être appliquées. Sous des hypothèses adéquates, notre estimateur est efficace quand le nombre d'observations temporelles et le nombre de simulations tendent vers l'infini. Cette approche par simulation peut être utilisée pour de nombreuses classes de processus max-stables et peut fournir de meilleurs résultats que les méthodes actuelles utilisant la vraisemblance composite, notamment dans le cas où seules quelques observations temporelles sont disponibles et où la dépendance spatiale est importante.

Le Chapitre 4 s'intéresse aux mesures de risque dans un cadre spatial, dans le but de prendre en compte les caractéristiques spatiales des processus environnementaux. L'objectif principal est la quantification de la diversification spatiale ainsi que de la sensibilité du risque à certaines caractéristiques spatiales, question cruciale pour les autorités ainsi que les compagnies d'assurance. Pour ce faire, le concept de mesure de risque spatiale est introduit et une réflexion sur une axiomatique adaptée est proposée. Nous construisons et étudions deux exemples de telle mesure, fondés sur la perte agrégée (spatialement) due à des événements extrêmes. La variable environnementale est modélisée par différents types de processus max-stables.

Le Chapitre 5 propose un modèle expliquant les interconnexions entre institutions financières. De tels modèles d'interconnexions endogènes sont très utiles pour évaluer l'impact systémique de chocs économiques ou encore les conséquences potentielles d'un changement de régulation. Notre modèle est fondé sur l'idée que les interconnexions entre institutions financières proviennent d'un choix de diversification. Le bilan des différentes institutions (hétérogènes) est totalement endogène et découle de la maximisation de leur utilité espérée. Nous comparons le réseau obtenu avec un réseau stylisé et montrons que la diversification apparaît effectivement comme un motif plausible. Enfin, à l'aide de ce modèle, nous analysons l'impact de changements réglementaires.

Mots-clés : Contagion ; Diversification ; Extrêmes spatiaux ; Facteur commun ; Générateur de précipitations ; Maximum de vraisemblance simulée non paramétrique ; Mesures de risque ; Modèle spatio-temporel ; Processus max-stables ; Réseaux financiers ; Risque systémique.

Abstract

This thesis aims at developing tools and models that are relevant for the study of some spatial risks and risks in networks. The thesis is divided into five chapters. The first one is a general introduction containing the state of the art related to each study as well as the main results.

Chapter 2 develops a new multi-site precipitation generator. It is crucial to dispose of models able to produce statistically realistic precipitation series. Whereas previously introduced models in the literature deal with daily precipitation, we develop a hourly model. The latter involves only one equation and thus introduces dependence between occurrence and intensity; the aforementioned literature assumes that these processes are independent. Our model contains a common factor taking large scale atmospheric conditions into account and a multivariate autoregressive contagion term accounting for local propagation of rainfall. Despite its relative simplicity, this model shows an impressive ability to reproduce real intensities, lengths of dry periods as well as the spatial dependence structure.

In Chapter 3, we propose an estimation method for max-stable processes, based on simulated likelihood techniques. Max-stable processes are ideally suited for the statistical modeling of spatial extremes but their inference is difficult. Indeed the multivariate density function is not available and thus standard likelihood-based estimation methods cannot be applied. Under appropriate assumptions, our estimator is efficient as both the temporal dimension and the number of simulation draws tend towards infinity. This approach by simulation can be used for many classes of max-stable processes and can provide better results than composite-based methods, especially in the case where only a few temporal observations are available and the spatial dependence is high.

Chapter 4 considers risk measures in a spatial framework in order to account for the spatial features of environmental processes. The main aim is to study the spatial diversification and more generally the sensitivity of risk with respect to the spatial region features, which are crucial issues for authorities and insurance companies. To this purpose, the notion of "spatial risk measure" is introduced and an axiomatic approach adapted to the spatial context is proposed. We build and study two examples of such risk measures, based on the spatially aggregated loss due to extreme events. The environmental variable is modeled using different types of max-stable processes.

Chapter 5 proposes a model explaining the interconnections between financial institutions. Such models dealing with endogenous interconnections are very useful to assess the systemic impact of economic crises or the consequences of regulatory changes. Our model is based on the assumption that interconnections across financial institutions come from a diversification motive. The balance sheet of the different heterogenous institutions is totally endogenous and stems from the maximization of their expected utility. We compare the obtained network with a stylized one and show that the diversification appears as a plausible motive. Finally, using our model, we study the impact of regulatory changes.

Key words: Contagion; Common factor; Diversification; Financial networks; Max-stable processes; Non parametric maximum simulated likelihood inference; Precipitation generator; Spatial extremes; Spatial-temporal model; Risk measures; Systemic risk.

Contents

Chapitre 1

Introduction

Cette thèse porte principalement sur l'étude de quelques outils pour l'évaluation de risques spatiaux et en réseaux. Une part importante des résultats présentés peut être appliquée à la gestion des risques environnementaux, notamment catastrophiques. Ainsi, dans la première partie de cette introduction générale, nous proposons un bref état de l'art sur le changement climatique (et notamment l'évolution des extrêmes climatiques) ainsi que quelques considérations sur les impacts potentiels des catastrophes naturelles ainsi que du changement climatique. Nous y décrivons ensuite brièvement quelques méthodes de couverture des risques catastrophiques. Nous y montrons enfin dans quelles mesures les outils développés dans cette thèse peuvent s'avérer utiles pour la modélisation de certains de ces impacts. Les parties suivantes présentent l'état de l'art relatif à chaque chapitre ainsi que les principaux résultats qui y sont obtenus.

1.1 Changement climatique, événements extrêmes et impacts

1.1.1 Changement climatique et catastrophes naturelles

L'année 2013 a été marquée par le typhon Haiyan qui a fait au moins 6000 victimes aux Philippines. Depuis le début des mesures météorologiques systématiques, il s'agit du cyclone le plus puissant ayant touché terre et probablement également du plus puissant jamais enregistré. Juste avant son entrée sur les Philippines, des vents moyens (sur dix minutes) de 275 $km.h^{-1}$ et des rafales de l'ordre de 315 $km.h^{-1}$ ont été mesurés.

Dans un contexte de changement climatique, certains événements extrêmes tendent à être de plus en plus fréquents (GIEC 2013). La Figure 1.1 montre une claire augmentation du nombre de catastrophes météorologiques[1] depuis les années 1970[2].

[1]Dans le rapport *sigma*, un événement est considéré comme une catastrophe si les dommages assurés, les pertes économiques totales ou le nombre de victimes dépassent un certain seuil. Par exemple, le seuil pour les dommages assurés dus à une catastrophe maritime est de 19.3 millions de USD.

[2]Notons toutefois que cette augmentation est peut-être due en partie à un recensement plus complet et systématique de ce type d'événements.

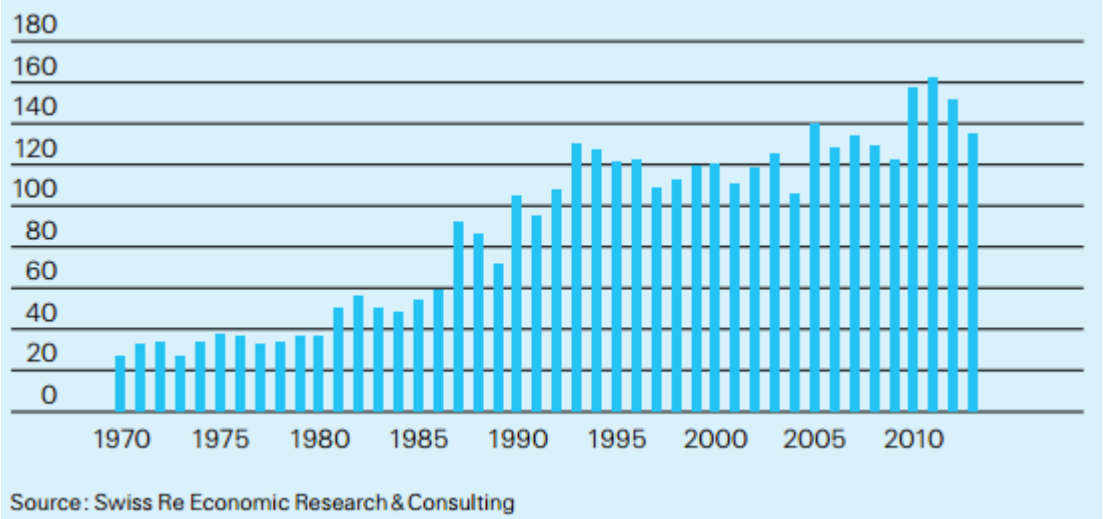


FIGURE 1.1 – Nombre de catastrophes météorologiques, 1970-2013, selon le rapport *sigma* 2013.

Par ailleurs, la population ainsi que la richesse par habitant sont en constante augmentation. L'urbanisation s'accélère et les villes en expansion sont souvent situées dans des régions côtières. Ainsi, les nouveaux quartiers des pays en développement se trouvent le plus souvent dans des zones exposées aux inondations, alors que la construction des infrastructures préventives ne parvient pas à suivre le rythme effréné de l'urbanisation. De surcroît, du fait de la complexité accrue des processus de production industrielle, les pertes catastrophiques dans ce secteur sont plus élevées que par le passé. Enfin, la dégradation des sols, la déforestation ou encore les modifications d'affectation des sols constituent autant de facteurs aggravant l'impact des phénomènes météorologiques extrêmes. En conséquence, on observe une nette tendance à la hausse des dommages économiques correspondants, comme le montre la Figure 1.2.

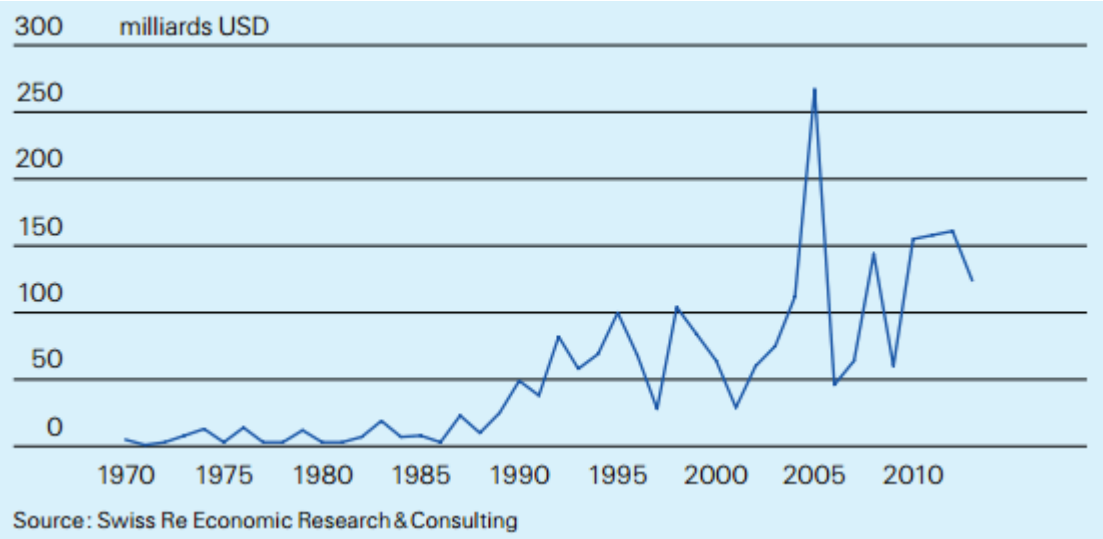


FIGURE 1.2 – Dommages économiques dus à des événements météorologiques extrêmes, 1970-2013, selon le rapport *sigma* 2013.

Si l'on s'intéresse aux dommages assurés, la tendance est également à la hausse, y compris lorsque l'on inclut les séismes/tsunamis et catastrophes techniques. L'année 2011

détient le record en ce qui concerne à la fois les pertes économiques et les pertes assurées, avec respectivement environ 380 et 135 milliards de dollars. La Figure 1.3 représente les pertes assurées imputables aux catastrophes entre 1970 et 2013 ainsi que la proportion attribuable à chaque type d'événement.

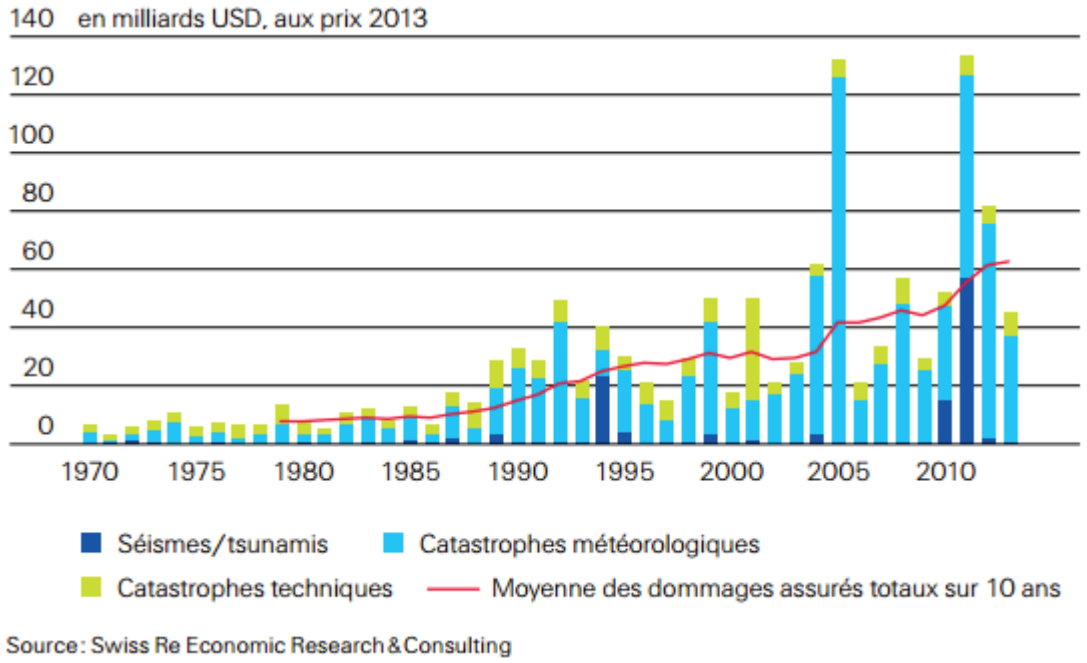


FIGURE 1.3 – Dommages assurés dus à des catastrophes, 1970-2013, selon le rapport *sigma* 2013.

Notons cependant que le taux de croissance des dommages assurés est inférieur à celui des dommages économiques totaux.

Le changement climatique peut être interprété comme les conséquences d'un choc exogène (comme les émissions de CO_2 par exemple). Ainsi, il peut s'apparenter à la réponse endogène (via différents mécanismes d'amplification et de contagion) à un choc exogène. On peut également directement voir le changement climatique comme un choc exogène au système économique, qui engendre diverses conséquences par des mécanismes de contagion. Le thème du choc exogène initial et de ses conséquences endogènes, notamment via des phénomènes de contagion, sera abordé à plusieurs reprises dans cette thèse, notamment dans les Chapitres 2 et 5.

Le rapport spécial du GIEC sur les phénomènes extrêmes (2012) décrit de manière détaillée les différentes projections concernant les changements relatifs aux extrêmes climatiques. Nous pouvons en tirer les trois points principaux suivants. Il est très probable (probabilité entre 90 et 100%) que la durée ainsi que la fréquence et/ou l'intensité des épisodes de canicules augmente dans la plupart des régions. Par ailleurs, il est probable (probabilité entre 66 et 100%) que la fréquence des épisodes de fortes précipitations (ou au moins la part de celles-ci dans la pluviosité totale) soit en augmentation dans de nombreuses régions. En effet, la hausse des températures a tendance à accroître le taux d'humidité atmosphérique et donc à dynamiser le cycle de l'eau. Enfin, il est probable que la vitesse maximale du vent dans les cyclones tropicaux augmente. En revanche, la fréquence de ces derniers pourrait rester constante voire diminuer. Nous notons que les projections au sujet des extrêmes de précipitations et de vent sont plus incertaines que

celles concernant les températures.

En résumé, il est probable que la fréquence, l'intensité ainsi que l'étendue spatiale et temporelle de la plupart des phénomènes météorologiques extrêmes soient en augmentation. Les dommages économiques et assurés associés devraient donc également augmenter. En conséquence, leurs prévisions constituent un challenge majeur de ce siècle.

1.1.2 Impact des catastrophes naturelles

Comme nous venons de le voir, les catastrophes naturelles engendrent des dommages humains et économiques de plus en plus importants. Une part importante de ces risques est transférée aux compagnies d'assurance. En France, ce transfert s'effectue via des contrats classiques de type Multi-Risques Habitations, Automobile et Multi-Risques Entreprises qui comportent une garantie Cat Nat. Les catastrophes naturelles ont donc potentiellement de très forts impacts sur les compagnies d'assurance, et même plus généralement sur le système "banques-compagnies d'assurance".

Avant de préciser ces conséquences, il convient de définir les principaux types de risques encourus par les institutions financières :

- Le risque de marché, lié à la modification de la valeur d'une position financière due aux changements de prix de ses composantes (actions, obligations, taux de change, matières premières, ...) ;
- Le risque de crédit, risque de ne pas recevoir les paiements prévus du fait du défaut de l'emprunteur ;
- Le risque opérationnel, risque de pertes liées à des défaillances de processus internes (personnes ou systèmes) ou externes. Celles-ci peuvent être causées par une erreur humaine, une fraude, un incendie, ...
- Le risque de liquidité, lié au fait qu'un actif ne puisse pas être vendu à son juste prix du fait de la difficulté de trouver un acheteur. On parle également de risque de liquidité pour désigner un risque de trésorerie non lié à un défaut de solvabilité.
- Le risque de modèle, lié à l'utilisation d'un modèle inexact ou sous des hypothèses erronées. Un exemple typique est l'utilisation du modèle de Black-Scholes dans le cas de rendements non Gaussiens.
- Le risque de souscription, pris par un assureur via les contrats d'assurance souscrits.

Dans le cas des banques, les événements extrêmes engendrent un risque de marché (actifs immobiliers détenus frappés par une catastrophe), un risque opérationnel (risque de détérioration des installations), un risque de crédit (si un emprunteur possède des actifs, typiquement immobiliers, touchés par une catastrophe) et un risque de liquidité (si un pourvoyeur de liquidité tel un assureur est durement touché par une catastrophe).

Dans le cas des compagnies d'assurance, de tels événements génèrent également un risque de marché, opérationnel, de crédit et de liquidité mais surtout un risque de souscription (du fait des risques intrinsèques à chaque contrat) et de modèle (si le risque de catastrophes est mal évalué). La réforme de la réglementation prudentielle européenne dans le domaine de l'assurance, Solvabilité 2, impose une meilleure évaluation des différents risques, et notamment ceux liés aux événements climatiques. D'où la nécessité d'élaborer des mesures de risque adaptées à ce type d'événements.

Outre l'impact à l'échelle d'une compagnie, il convient de ne pas omettre les problématiques de risque systémique, c'est-à-dire à l'échelle du système financier dans son ensemble.

1.1.3 Impacts à long terme du changement climatique

Contrairement aux risques météorologiques liés à des événements particuliers se produisant sur des périodes courtes, les risques liés au changement climatique font référence à des tendances sur le long terme. L'augmentation de la fréquence des catastrophes naturelles en fait partie. Cependant, outre cet aspect, le changement climatique pourrait s'accompagner d'effets indirects. Les secteurs les plus concernés sont ceux de l'énergie, des transports, ou encore de l'agriculture. En effet, le coût des énergies fossiles risque d'augmenter, notamment du fait de leur taxation afin de limiter les émissions. L'exploitation agricole pourrait progressivement devenir plus délicate dans certains pays (du fait d'une sécheresse accrue), d'où une augmentation des coûts de la nourriture. Par ailleurs, la pénurie en eau pourrait encore s'accentuer dans les pays en développement mais aussi dans certains pays développés (du bassin Méditerranéen par exemple), ce qui pourrait engendrer d'importants flux migratoires.

Concernant les conséquences sur la santé, les températures plus élevées pourraient favoriser la prolifération de certaines maladies infectieuses. Actuellement, les moustiques porteurs de maladies ne peuvent se reproduire aux hautes latitudes car il y fait trop froid. Néanmoins, ceci pourrait évoluer dans le cas d'une augmentation significative de la température. Des précipitations en hausse risquent également d'augmenter le nombre de parcelles d'eau stagnante, nécessaires à leur reproduction. De surcroît, de fréquentes inondations peuvent fragiliser la qualité de l'eau. De manière générale, l'Organisation Mondiale de la Santé a observé l'apparition d'une trentaine de nouvelles maladies depuis 1976 ainsi que la réapparition de certaines maladies. La mondialisation est naturellement un facteur de prolifération important. Par ailleurs, le printemps a tendance à survenir de plus en plus tôt et la production de pollen est favorisée par l'augmentation de la concentration en CO_2. Cela serait à l'origine d'allergies de plus en plus nombreuses.

Enfin, le changement climatique pourrait s'accompagner d'une importante réduction de la biodiversité. En effet, les conditions de vie risquent de ne plus être supportables et certaines espèces ne seront pas capables d'atteindre les régions adéquates. Le Living Planet Index, indicateur de l'état de la biodiversité, a chuté de 28% entre 1970 et 2007.

Même si certains outils de cette thèse pourraient être utilisés pour la quantification de certains impacts à long terme (notamment le modèle de contagion du Chapitre 2 dans le cas de la prolifération de maladies infectieuses), ils sont plus naturellement applicables à l'étude des impacts directs des catastrophes naturelles sur les populations et le système financier. Par ailleurs, les impacts du changement climatique mentionnés précédemment sont hypothétiques et très difficilement quantifiables avec précision, du fait en premier lieu de l'incertitude sur les projections climatiques.

1.1.4 Couverture du risque catastrophique

Afin de couvrir les risques extrêmes, les assureurs ont le choix entre deux types de transfert de risque :

- La réassurance traditionnelle : proportionnelle ou en excès de seuil ;

- Les méthodes alternatives : CAT bonds, CAT options, ILW's, Risks swaps, ...

En effet, certaines catastrophes peuvent engendrer des coûts supérieurs à la capacité du marché de la réassurance. L'ouragan Andrew en 1992 et le tremblement de terre à Northridge (près de Los Angeles) en 1994 ont réduit le capital de beaucoup de compagnies de réassurance et diminué la capacité de souscription du marché de la réassurance. Ainsi, l'augmentation de la demande de réassurance CAT conjuguée à une baisse de l'offre a engendré une hausse des primes de réassurance. Les assureurs se sont alors tournés vers les marchés financiers, via de nouveaux titres dont le pay-off est indexé sur la réalisation d'événements particuliers, les "event-linked securities". Outre les CAT bonds, on trouve notamment les CAT options, ILW's ou encore les risks swaps. Cummins (2008) propose une description complète de ces différents produits ainsi que de l'évolution de leurs marchés respectifs. L'obligation catastrophe (ou CAT bond) constitue l'actif le plus populaire. Il s'agit d'une obligation à taux flottant (émise par la compagnie souhaitant se couvrir), dont les coupons et/ou le principal ne sont pas remboursés en cas de catastrophe naturelle spécifiée dans le contrat. Le seuil de déclenchement peut être de type indemnitaire (basé sur la perte de la compagnie émettrice), indiciel (basé sur la perte agrégée du secteur, une perte modélisée ou un indice paramétrique) ou encore hybride. Ce type de produit est totalement collatéralisé et donc le risque de crédit s'avère beaucoup plus faible que dans le cas de la réassurance traditionnelle. Par ailleurs, les CAT bonds offrent une protection à plus long terme que la réassurance (généralement jusqu'à cinq ans). Enfin, ces produits sont généralement attrayants pour les investisseurs car ils fournissent un rendement élevé (Libor auquel s'ajoute un spread généralement de l'odre de 3 à 20%) et sont décorrélés du marché actions. Pour une discussion sur les avantages des "event-linked securities" du point de vue de l'investisseur, voir Litzenberger *et al.* (1996).

Malgré leur intérêt, les CAT bonds n'ont pas connu le développement flamboyant escompté. Ceci provient en partie d'un manque de confiance des investisseurs. Les méthodes de "pricing" de ce type de produit doivent encore être très largement améliorées. Finn et Lane (1997) proposent une riche discussion sur les liens entre le "pricing" de l'assurance catastrophe et celui des options traditionnelles. Par analogie avec la notion de volatilité implicite, ils définissent la notion de distribution de perte implicite comme étant la distribution permettant d'expliquer au mieux les prix de réassurance observés. Cox et Pedersen (2000) développent un modèle de prix des CAT bonds qui tient compte des problèmes de marché incomplet en présence de catastrophes. Lee et Yu (2002) s'intéressent à un autre aspect, celui de la prise en compte du caractère stochatique des taux d'intérêt dans le prix des CAT bonds. Albrecher *et al.* (2003, 2004) proposent des méthodes d'approximation des intégrales apparaissant dans les prix de CAT bonds, fondées sur des techniques de Quasi-Monte Carlo. Un autre courant de littérature s'attache à comprendre précisément pourquoi le marché des CAT bonds n'explose pas plus rapidement (voir par exemple Barrieu et Loubergé (2009)).

Certains outils développés ou utilisés dans cette thèse (dans les Chapitres 2 à 4) peuvent s'avérer utiles en vue d'améliorer le pricing de ce type de produits, notamment via une meilleure modélisation des pertes. Un travail en cours est brièvement décrit en conclusion.

Avant de proposer un résumé des différents chapitres, nous expliquons dans quelle mesure les outils développés dans cette thèse peuvent permettre la modélisation de certains impacts mentionnés précédemment.

1.1.5 Modélisation des impacts dans cette thèse

Afin de quantifier l'impact des catastrophes naturelles, la première étape consiste en la modélisation du comportement statistique des événements environnementaux extrêmes. Par nature, les phénomènes météorologiques et plus généralement environnementaux sont spatiaux et il est crucial de prendre cette caractéristique en compte. Par exemple, de fortes pluies auront des conséquences bien plus importantes si elles sont spatialement étendues que si elles n'affectent qu'une seule localité. L'aspect spatial s'avère omniprésent dans cette thèse.

Les statistiques spatiales se sont principalement développées à partir des années 1960. On les classe en général en trois grandes catégories. La géostatistique concerne l'étude des données spatiales modélisées par un champs aléatoire (ou processus stochastique) continu. Lorsque les données sont intrinsèquement liées à un réseau, il est certes possible d'utiliser le formalisme de la géostatistique, mais celui des champs Markoviens est souvent plus adapté. Enfin, lorsque ce sont les coordonnées de certains événements qui portent l'information principale (points d'impacts de foudre, épicentres de séismes) et non une variable mesurée spatialement, la théorie des processus de Poisson est particulièrement adaptée.

La géostatistique s'est principalement développée afin de proposer des méthodes d'estimation des réserves en gisements miniers. L'ouvrage par Matheron (1965) en pose les fondements. La théorie des valeurs extrêmes, dont le premier résultat date de 1928, s'est parallèlement principalement développée à partir des années 1970, à la suite de la thèse de doctorat de de Haan et en partie du fait de l'émergence de nombreuses applications potentielles. La théorie des extrêmes spatiaux (et notamment celle des processus max-stables) s'est développée à partir des années 80/90 et se trouve aux confins de la théorie des valeurs extrêmes et de la géostatistique.

Les outils développés dans les Chapitres 2 et 3 peuvent s'avérer utiles pour la modélisation des variables environnementales. Le Chapitre 2 s'intéresse à la modélisation spatio-temporelle des précipitations et utilise plutôt le formalisme des champs Markoviens. Le modèle proposé s'inspire de techniques utilisées dans l'étude de la contagion dans les réseaux bancaires. Ce modèle s'intéresse à la distribution des précipitations dans son ensemble. A l'inverse, le Chapitre 3 ne s'intéresse qu'au maximum et propose un nouvel outil d'estimation des processus max-stables. Notons cependant que les applications potentielles du Chapitre 3 ne se limitent pas aux phénomènes environnementaux.

La seconde étape implique de convertir les variables environnementales en dommages économiques. Cela est fait en leur appliquant tout d'abord une fonction de dommage (ou de vulnérabilité) qui donne la propension d'un point donné à subir des dommages. Il convient ensuite de pondérer ce résultat par l'exposition, qui indique la sensibilité de ce point aux catastrophes naturelles. L'exposition tient compte de facteurs démographiques (densité d'habitants), économiques (richesse par habitants, structure des bâtiments, ...) et orographiques. Cette démarche de quantification des coûts économiques est adoptée dans le Chapitre 4, qui s'intéresse aux mesures de risque dans un cadre spatial, c'est-à-dire prenant en compte la nature spatiale des phénomènes environnementaux. La notion de mesure de risque fournit un résumé de la distribution des dommages potentiels.

Le Chapitre 5 est un peu isolé dans le sens où il traite des liens entre institutions financières. Il possède néanmoins quelques similarités avec le Chapitre 2. Tout d'abord, l'approche spatiale adoptée est de type "réseau" et non "géostatistique". Un autre point commun réside dans le fait qu'un choc exogène initial s'accompagne de conséquences

endogènes via des phénomènes de contagion. Dans le cas du générateur de précipitations développé dans le Chapitre 2, une perturbation exogène grande échelle (baisse de la pression, augmentation du taux d'humidité) engendre des phénomènes à l'échelle locale, typiquement des précipitations qui se propagent dans l'espace et le temps par contagion. Dans le cas du modèle de réseau bancaire développé dans le Chapitre 5, un choc exogène global (par exemple une crise économique) peut entraîner des conséquences locales (typiquement la faillite d'une institution) qui se propagent à d'autres institutions également par contagion.

Par ailleurs, le Chapitre 4 propose des outils adaptés à l'étude du risque spécifique à une institution. Cependant, ce chapitre ne traite pas du lien entre compagnies d'assurance, et les outils correspondants ne permettent pas de quantifier la contagion de manière naturelle. Ainsi, le risque systémique (i.e. à l'échelle d'un réseau d'institutions financières pris dans son ensemble) ne peut être abordé que de manière partielle. Le Chapitre 5, en revanche, fournit des outils d'étude de la contagion bancaire et s'avère donc plus adapté au risque systémique. Notons qu'un épisode systémique pour le réseau bancaire est en général engendré par une crise économique (et non une catastrophe naturelle). Néanmoins, l'étude des conséquences systémiques potentielles d'une catastrophe naturelle constitue un travail en cours et sera évoquée en conclusion générale de cette thèse.

1.2 Résultats du Chapitre 2

Dans le Chapitre 2, issu d'un article[3] co-écrit avec Philippe Naveau (LSCE,CNRS), nous introduisons un nouveau type de générateur de précipitations.

Dans une première partie de cette section, nous définissons les générateurs de précipitation et décrivons l'état de l'art correspondant. Nous introduisons ensuite notre modèle ainsi que les principaux résultats. Enfin, nous menons une large discussion sur d'autres modèles que nous avons développés et testés. Nous expliquons notamment les raisons qui nous ont amenés à choisir le modèle décrit dans ce chapitre.

1.2.1 Etat de l'art sur les générateurs de précipitations

Un générateur de temps est un modèle statistique permettant de simuler de manière "réaliste" les principales grandeurs climatiques : précipitations, températures et vitesses de vent par exemple. Le terme "réaliste" doit être compris au sens statistique. Ainsi, les valeurs générées ne se doivent pas de correspondre aux observations date par date mais sont censées exhiber un comportement statistique proche de celui observé. De ce fait, les générateurs de temps ne peuvent généralement pas être utilisés comme outils de prévision mais revêtent un grand intérêt pour plusieurs raisons. Ils sont apparus dans les années 60 en hydrologie, avec le papier de Gabriel et Neumann (1962). Dans la suite, nous ne nous intéressons qu'aux générateurs de précipitations. Ceux-ci permettent de générer des précipitations en des lieux et à des dates pour lesquels aucune donnée n'est disponible. Ils offrent notamment la possibilité de simuler des précipitations réalistes à l'échelle locale connaissant les informations à grande échelle (downscaling). Mais ils sont utiles également dans le cas où des données historiques sont disponibles. En effet, simuler des précipitations dans le futur en tirant aléatoirement dans l'historique ne permet pas de prendre en compte les futures évolutions climatiques. A l'inverse, un modèle tenant

[3] A paraître dans la revue "Advances in Water Resources".

compte des informations climatiques à grande échelle peut générer des précipitations futures cohérentes avec les conditions climatiques.

Ainsi, les générateurs apparaissent très utiles pour quantifier l'impact économique des phénomènes hydrologiques. Typiquement, les simulations de précipitations sont très utilisées comme entrées des modèles d'inondation, de ruissellement, d'érosion, de production agricole ou encore d'approvisionnement en eau des écosystèmes.

Enfin, outre son intérêt en "Risk-management", la modélisation statistique des précipitations s'avère extrêmement intéressante du point de vue mathématique. Entre autres, les précipitations présentent un caractère spatio-temporel naturel et mêlent deux types d'événements (sécheresse et humidité), offrant ainsi un grand nombre de challenges tant en termes de modélisation que d'inférence.

La littérature sur les générateurs de précipitation s'est tout d'abord focalisée sur la modélisation en un seul site. La plupart des modèles correspondants font intervenir séparément deux composantes : le processus d'occurrence (présence de pluie ou non) et le processus d'intensité. L'occurrence est généralement modélisée à l'aide d'une chaîne de Markov (Katz, 1977). Le plus souvent, une chaîne de Markov du premier ordre et à deux états est utilisée (voir par exemple Richardson (1981)). La modélisation de l'intensité repose en général sur des distributions exponentielles (Richardson, 1981), des distributions gamma (Stern et Coe, 1984) ou des mélanges d'exponentielles (Woolhiser et Pegram, 1979). Ces modèles sont donc paramétriques et assez facilement estimables. Néanmoins, comme on va le voir en détails, ils présentent certains défauts.

Les méthodes non paramétriques proposent une alternative intéressante. Elles n'imposent aucune densité de probabilité et, au contraire, ne sont guidées que par les données (Wilks et Wilby, 1999). Ces méthodes sont principalement de quatre types. L'utilisation des distributions empiriques est la plus simple (voir par exemple Semenov et Porter (1995)). Le deuxième type de méthode s'appuie sur les réseaux de neurones. Trigo et Palutikof (1999) l'appliquent à la génération de séries de températures. Le troisième, développé par Rajagopalan *et al.* (1997), utilise des estimateurs à noyaux multivariés. Enfin, le ré-échantillonnage (ou bootstrap) fournit une alternative simple et efficace aux méthodes à noyaux (voir par exemple Young (1994)). Cette approche a ensuite été améliorée (k plus proches voisins) par Lall et Sharma (1996) et appliquée par Rajagopalan et Lall (1999) au cas de la simulation multivariée (plusieurs variables météorologiques).

Des épisodes de pluies extrêmes très localisés ou au contraire très étendus n'ont absolument pas les mêmes conséquences économiques. Ainsi, dans le cas des modèles d'inondation par exemple, il est nécessaire de prendre en compte la nature spatiale des précipitations. Ceci est impossible avec les modèles uni-sites décrits précédemment. Ainsi, depuis la fin des années 1990, la littérature s'intéresse principalement aux modèles spatio-temporels, dans le but de tenir compte des dépendances à la fois spatiales et temporelles. Plusieurs approches ont été proposées et, comme dans le cas uni-site, elles se divisent entre les modèles paramétriques et les modèles non paramétriques.

Les modèles non paramétriques sont essentiellement de deux types : purement statistiques ou à "état météorologique". Les méthodes purement statistiques incluent des processus à chaîne dépendante généralisés (Zheng et Katz, 2008; Zheng *et al.*, 2010), la transformation de la distribution de l'intensité en loi normale (via une fonction puissance) (Sansó et Guenni, 2000; Yang *et al.*, 2005) et les approches fondées sur des copules (Bárdossy et Pegram, 2009). Le papier fondateur de Wilks (1998) est à l'origine de nombreuses approches actuelles pour la modélisation multi-site. Cette méthode présente l'avantage de décrire la dépendance spatiale observée tout en permettant de la flexibilité dans la modé-

lisation de la dépendance temporelle. De plus, les paramètres du modèle étant considérés comme des champs aléatoires, cette approche permet de simuler en des lieux où aucune observation n'est disponible ainsi que de quantifier l'incertitude sur les paramètres. Cette approche a été étendue dans les travaux de Wilks (2009) et Kleiber *et al.* (2012). Kleiber *et al.* (2012) utilisent des processus Gaussiens à la fois pour l'intensité et les paramètres. Cette approche se différencie de celle adoptée par Wilks (2009) en incluant les modèles d'occurrence et d'intensité dans un cadre de type GLM.

Un défaut majeur de ces modèles purement statistiques réside dans le fait qu'ils ne peuvent généralement produire des simulations réalistes que sous des conditions climatiques semblables à celles observées sur la période utilisée pour l'inférence du modèle. Les modèles à "état météorologique" ont été introduits pour pallier cette difficulté. Ils utilisent les données atmosphériques disponibles et tentent de représenter explicitement certains processus physiques à l'origine des précipitations. Parmi ces modèles, on trouve notamment ceux de Hay *et al.* (1991), Bardossy et Plate (1992) et Kidson (1994). Des cas particuliers sont les modèles à chaînes de Markov cachées pour l'occurrence (Hughes et Guttorp, 1994; Hughes *et al.*, 1999) et l'intensité (Ailliot *et al.*, 2009; Charles *et al.*, 1996). Dans le dernier cas, les précipitations dépendent d'un processus latent suivant une chaîne de Markov et dont les probabilités de transition varient en fonction des conditions atmosphériques. Néanmoins, les modèles à "état météorologique" font souvent intervenir davantage de paramètres que les modèles purement statistiques et sont donc plus difficiles à estimer.

Dans le contexte spatial, la plupart des méthodes non paramétriques utilisent du ré-échantillonnage basé sur les plus proches voisins. Buishand et Brandsma (2001) ainsi que Yates *et al.* (2003) étendent la technique du bootstrap à k plus proches voisins à la génération multi-site. Les réseaux de neurones artificiels sont également parfois utilisés. (Cannon, 2008). Enfin, Apipattanavis *et al.* (2007) développent un modèle impliquant une première étape paramétrique fondée sur une chaîne de Markov du premier ordre à trois états et une seconde non paramétrique utilisant du ré-échantillonnage à k-plus-proches voisins.

Dans ce chapitre, nous ne considérons que des modèles paramétriques. La plupart des modèles introduits dans la littérature ont pour but de générer des précipitations journalières et non horaires. Il est en général plus délicat d'obtenir une structure réaliste pour les précipitations horaires du fait de la persistance temporelle ainsi que des phénomènes complexes se produisant à l'échelle locale.

Dans les approches traditionnelles (voir par exemple Kleiber *et al.* (2012)), on simule d'abord le processus d'occurrence. Le processus d'intensité est ensuite simulé aux lieux et dates où il pleut. Ainsi, en général, les processus d'occurrence et d'intensité sont indépendants. Les défenseurs de ce type d'hypothèse prétendent que les mécanismes physiques à l'origine de ces deux processus sont très différents. En réalité, cette indépendance s'avère très éloignée de la réalité, notamment à l'échelle horaire. La probabilité d'occurrence à la date t est clairement fonction croissante de l'intensité à la date $t-1$.

Une méthode permettant d'obtenir de la dépendance entre ces deux processus est de s'appuyer sur un champ aléatoire caché, par exemple Gaussien, noté Z_t dans la suite. Il y a occurrence quand Z_t est supérieur à un certain seuil ; l'intensité est ensuite obtenue par une transformation de Z_t. Cette approche est notamment adoptée par Allard et Bourotte (2013). On comprend que si Z_t possède de la persistance temporelle, alors plus Z_{t-1} est élevé, plus Z_t l'est (avec une grande probabilité). Ainsi, plus l'intensité en $t-1$ (fonction croissante de Z_{t-1}) est élevée, plus la probabilité d'occurrence en t est élevée. La relation

croissante entre intensité et probabilité d'occurrence s'obtient donc si le processus caché sous-jacent possède de la persistance temporelle et si l'intensité est fonction croissante de ce processus. Le modèle développé par Kleiber *et al.* (2012) pourrait s'adapter à ce cas mais l'inférence s'avère relativement compliquée dans un tel cadre. Par ailleurs, dans toutes ces approches, le processus Z_t est difficilement interprétable, ce qui n'est pas très satisfaisant.

Afin de s'affranchir de ces difficultés, nous proposons un modèle en une seule équation, ce qui crée automatiquement de la dépendance entre les processus d'occurrence et d'intensité. Par ailleurs, nous utilisons un facteur commun prenant en compte les informations atmosphériques à grande échelle. Le modèle proposé s'interprète donc plus facilement physiquement que ceux utilisant des processus cachés. Enfin, il évite le défaut de certains modèles purement statistiques mentionnés précédemment, qui ne tiennent pas compte des informations à grande échelle.

Dans la littérature décrite précédemment, peu de modèles relient petite et grande échelle, que l'échelle considérée soit spatiale ou temporelle. Le Chapitre 5 de cette thèse s'intéresse au risque systémique et à la contagion en finance et assurance. Le générateur de précipitations proposé ici s'inspire de ce courant de littérature. L'analyse du risque systémique dans le système financier implique de modéliser la dépendance entre institutions. Un premier type de dépendance provient de la présence de facteurs communs observables ou inobservables. Ceux-ci sont nommés facteurs systématiques (ou "frailties"). Typiquement, le risque de contrats d'assurance vie dépend généralement d'une augmentation incertaine de la durée de vie, le risque de longévité. Ce facteur est inobservable et commun à l'ensemble des compagnies d'assurance. Comme la régulation le souligne depuis 2008, une seconde source de dépendance provient de phénomènes de contagion entre institutions, liés aux interconnexions. Ces aspects seront étudiés en détails dans le Chapitre 5. Typiquement, la faillite d'une compagnie peut provoquer une perte pour les institutions créditrices et actionnaires et même éventuellement leur faillite. Des faillites peuvent alors subvenir en cascade.

Les premiers modèles de contagion en finance proviennent des modèles introduits en épidémiologie par Bailey (1953, 1957) et Kendall (1956) et réintroduits dans le modèle d'infection statique de Davis et Lo (2001). Les modèles intégrant à la fois les effets de facteur commun et de contagion sont introduits entre autres par Frey et Backhaus (2003), Giesecke et Weber (2004, 2006), Azizpour *et al.* (2008) et Gagliardini et Gouriéroux (2013). Il est très important de réussir à découpler ces deux sources de dépendance. Néanmoins, cette question s'avère très délicate. Manski (1993) décrit le "problème de réflexion". La magnitude relative des chocs exogènes au système et de la contagion (qui est endogène au système) doit être mesurée afin de clarifier leurs rôles distincts. Ceci est fait dans Gagliardini et Gouriéroux (2013).

Le générateur de pluie développé dans ce chapitre contient à la fois un facteur commun et un terme de contagion. L'analogie avec le système financier est la suivante. La région au sein de laquelle on souhaite générer des précipitations constitue notre système, équivalent du système financier. Chaque station d'observation correspond à une institution. Le facteur commun tient compte des conditions atmosphériques grande échelle, qui peuvent être considérées en première approximation comme exogènes au système. Elles sont l'équivalent des conditions exogènes au système financier, typiquement les conditions macro-économiques. Une fois les déterminants grande échelle fixés, des processus physiques locaux sont impliqués, et ce à plus haute fréquence. La propagation des précipitations d'un site à l'autre peut être interprétée comme de la contagion entre deux sites.

Un tel phénomène dépend des caractéristiques orographiques locales. Par exemple, la présence de montagnes ou rivières peut engendrer des directions de propagation privilégiées pour les averses ou orages. Ces caractéristiques locales sont l'équivalent des participations croisées en dette et en action dans le système financier. Notons que la notion de grande échelle ou échelle locale est liée à des phénomènes basse/haute fréquence. Notre modèle établit un lien entre les deux échelles et peut s'avérer utile pour le downscaling. Contrairement aux modèles financiers décrits précédemment, le facteur commun est observable. De plus, contrairement à la plupart des modèles à facteur commun et contagion, le facteur commun n'est pas additif mais apparaît dans la volatilité du modèle. Il s'agit donc d'un modèle hétéroscédastique. Les modèles hétéroscédastiques sont très répandus en finance. Voir par exemple la très abondante littérature sur les processus ARCH/GARCH définis par Engle (1982) et généralisés par Bollerslev (1986).

1.2.2 Les résultats présentés dans ce chapitre

Nous présentons dans cette partie le modèle développé dans ce chapitre ainsi que les principaux résultats obtenus.

Le modèle

Soit M le nombre de stations météorologiques et $P_{m,t}$ la quantité de précipitation enregistrée à la m-ième station météorologique pendant la t-ième heure. Notre générateur multi-site est défini par

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} \geq u, \\ 0 & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} < u, \end{cases} \tag{1.1}$$

où $u > 0$ et le produit scalaire

$$\mathbf{B}_m^{'}\mathbf{P}_{t-1} = \beta_{m,1}P_{1,t-1} + \cdots + \beta_{M,1}P_{M,t-1}$$

entre le vecteur $\mathbf{P}_{t-1} = (P_{1,t-1}, \ldots, P_{M,t-1})^{'}$ et $\mathbf{B}_m = (\beta_{m,1}, \ldots, \beta_{m,M})^{'}$ représente le vecteur auto-régressif multivarié d'ordre 1 qui capture l'effet local dynamique. Il décrit comment le comportement du voisinage impacte la station m une heure plus tard. Les $M \times M$ coefficients autorégressifs inconnus $\beta_{i,j}$ peuvent être concaténés dans une matrice, $\mathbf{B}$, ayant pour lignes les $\mathbf{B}_m^{'}$.

La variable aléatoire $\epsilon_{m,t}$ caractérise la variabilité locale. Nous la relions au facteur commun grande échelle $\mathbf{F}_t$, de la manière suivante

$$\epsilon_{m,t} \sim \mathcal{N}\left(0, \exp(\theta' \mathbf{F}_t)\right) \text{ (conditionnellement à } \mathbf{F}_t), \tag{1.2}$$

où $\mathcal{N}(\mu, \sigma)$ désigne la distribution normale de moyenne μ et de déviation standard σ, $\theta \in \mathbb{R}^d$ pour $d \in \mathbb{N}^*$, $(\mathbf{F}_t)_{t=1,\ldots,T} \in \mathbb{R}^d$. Pour tout $t \in 1, \ldots, T$, Pour $t = 1, \ldots, T$ et $m = 1, \ldots, M$, les $\{\epsilon_{m,t}\}$ sont indépendants conditionnellement à $\mathbf{F}_t$ $(t = 1, \ldots, T)$. Le vecteur $\mathbf{F}_t$ est observable et composé de d variables atmosphériques explicatives.

Le seuil u permet de générer des périodes sèches. La longueur des périodes sèches et humides est déterminée par la volatilité du bruit, c'est-à-dire par le terme $\exp(\theta' \mathbf{F}_t)$. Imaginons qu'il ne pleuvait nul part à la date $t-1$. Le terme de contagion est alors égal à 0. Si les conditions grande échelle ne sont pas favorables à un incrément de précipitation important, alors la volatilité est faible et il est peu probable que le bruit atteigne le seuil u. Ainsi, la probabilité que les conditions soient sèches dans toutes les stations en t est très

élevée. Le caractère hétéroscédastique du bruit permet donc de reproduire la persistance temporelle (due à une longue mémoire) des périodes sèches. Dans le cas de conditions grande-échelle favorables à un incrément élevé, la volatilité est élevée et il est donc très probable que le bruit atteigne le seuil. Dans ce cas, des précipitations sont générées et la période sèche se termine. Notons que la persistance temporelle n'est pas toujours bien représentée dans la plupart des modèles introduits dans la littérature, ou en tout cas peu évaluée.

Un avantage du modèle (1.1) est de propager de la même façon les précipitations nulles et non nulles (à la fois dans l'espace et le temps), ce qui crée une dépendance naturelle entre occurrence et intensité. S'il ne pleuvait pas (ou peu) dans un voisinage de la station m à la date $t-1$ et que la volatilité est faible, il est très probable que le temps soit sec en m et à la date t. Dans le cas contraire, si il pleuvait beaucoup en $t-1$, il est très probable que ce soit toujours le cas en t du fait de la contagion. Ce modèle introduit donc une relation clairement croissante entre l'intensité en $t-1$ et la probabilité d'occurrence en t. Cela est tout à fait conforme à la réalité, notamment dans le cas des précipitations horaires.

Remarquons qu'il pourrait être pertinent d'introduire une dépendance par rapport à davantage de lags : typiquement $P_{m,t-2}$, $P_{m,t-3}$ (pour rendre compte des longs épisodes de précipitation) ou même $P_{m,t-24}$ (en raison des cycles diurnes). Néanmoins, nous supposons que ces ordres sont inclus dans le $\mathbf{F}_t$ qui contient l'information grande échelle et basse fréquence.

Une approche classique dans les modèles à facteur commun et contagion en finance serait de prendre $\mathbf{F}_t$ comme étant inobservable. Ce dernier est alors statique (les $(\mathbf{F}_t)_{t=1,\dots,T}$ sont i.i.d.) ou dynamique. Cependant, dans le contexte météorologique/climatologique, un facteur commun inobservable ne serait pas nécessairement pertinent. En effet, nous avons à notre disposition un grand nombre d'observations pour différentes variables représentatives des conditions grande échelle, par exemple la pression ou encore le taux d'humidité. Les processus physiques sont de mieux en mieux connus et il est donc possible de proposer des variables pertinentes pour expliquer les précipitations. Une telle connaissance précise des mécanismes n'est pas toujours accessible en finance. De plus, comme précisé en introduction, la prise en compte des conditions atmosphériques à grande échelle s'avère indispensable pour simuler des précipitations réalistes.

Comme décrit ci-après, en se fondant sur la littérature existante, nous aurions pu choisir comme facteur commun un indice tel que NAO (North Atlantic Oscillation) ou ENSO (El Nino Southern Oscillation). Néanmoins, s'ils représentent de très bons indicateurs à l'échelle mondiale, ils ne contiennent qu'un résumé assez pauvre des informations grande échelle dans le cas de notre région d'intérêt et ne permettent pas d'exploiter la masse d'observations disponibles (pour différents types de variables) à l'échelle de la région étudiée. Une autre possibilité aurait été de relier $\mathbf{F}_t$ à un régime de temps défini par un indice ou des covariables.

Afin de limiter le sur-apprentissage (et donc d'améliorer la capacité de généralisation ou la robustesse de notre modèle), un nombre limité de covariables doit être utilisé pour construire le $\mathbf{F}_t$. De surcroît, étant donné que le modèle (1.1) doit être un générateur de temps, les futures valeurs des covariables doivent pouvoir être facilement simulées, par exemple en prenant les sorties de GCM (Global Climate Model). Les covariables finalement retenues sont les moyennes spatiales de la température, de la pression au niveau de la mer et du taux d'humidité, notées respectivement $\bar{T}_t$, $\bar{P}_t$ et $\bar{H}_t$.

Estimation

L'estimation du modèle se fait par maximum de vraisemblance. La fonction de vraisemblance conditionnelle au facteur commun, notée $L_u(\mathbf{B}, \theta_0, \ldots, \theta_d)$ pour un u donné, est donnée par la proposition suivante.

Proposition 1.1. *Nous avons, conditionnellement au facteur commun* $(\mathbf{F}_t)_{t=1,\ldots,T}$

$$L_u(\mathbf{B}, \theta_0, \ldots, \theta_d) = \sum_{t=1}^{T}\sum_{m=1}^{M}\left\{\mathbf{I}_{\{P_{m,t}\geq u\}}\log\left[\frac{1}{\exp(\theta'\ \mathbf{F}_t)}\phi\left(\frac{P_{m,t}-\mathbf{B}'_m\mathbf{P}_{t-1}}{\exp(\theta'\ \mathbf{F}_t)}\right)\right]+\mathbf{I}_{\{P_{m,t}=0\}}\log\left[\Phi\left(\frac{u-\mathbf{B}'_m\mathbf{P}_{t-1}}{\exp(\theta'\ \mathbf{F}_t)}\right)\right]\right\}, \tag{1.3}$$

où $\Phi(.)$ *et* $\phi(.)$ *représentent respectivement la fonction de répartition et la densité de probabilité de la loi normale standard.*

Nous proposons une procédure d'estimation automatique pour le seuil u qui consiste, pour un jeu d'estimations de $\mathbf{B}$, $\theta_0, \ldots, \theta_d$, à minimiser l'écart entre la longueur moyenne des périodes de sécheresse simulées et celle des plages réelles. Ainsi, on peut itérativement estimer u et $\mathbf{B}$, $\theta_0, \ldots, \theta_d$ jusqu'à stabilisation des valeurs. Les intervalles de confiance sont obtenus par la méthode de la "profile" log vraisemblance. La méthode d'inférence proposée fournit des résultats satisfaisants, notamment à partir d'une taille d'échantillon de l'ordre de $T = 1000$.

Dans le cas d'un grand nombre de sites, le nombre de coefficients serait extrêmement important et donc l'estimation par maximum de vraisemblance très longue. Afin de pallier cette difficulté, une possibilité est d'imposer une structure paramétrique à la matrice $\mathbf{B}$, avec des coefficients β_{mn} fonctions de la distance d_{mn}, par exemple $\beta_{mn} = \exp(\beta_0 - \beta_1\ d_{nm})$. Cependant, la distance euclidienne n'est en général pas adaptée en raison d'effets orographiques importants, de la présence de rivières ou de la proximité de la mer. Elaborer la bonne distance constitue un problème à part entière.

Application à la Bretagne Nord

Comme détaillé dans le Chapitre 2, le modèle (1.1) fournit des résultats très satisfaisants. En particulier, la dynamique des précipitations simulées est conforme à la réalité. Ici, on entend par dynamique la persistance temporelle, à l'origine de clusters de précipitations et d'heures sèches. Si les covariables sont connues, notre modèle peut d'ailleurs être utilisé pour effectuer de la prévision. Notons que la dynamique des séries générées n'est pas toujours évaluée dans les travaux de la littérature. Les QQ plots d'intensité présentés dans les papiers peuvent être satisfaisants sans que la dynamique ne soit bien représentée.

Notre modèle étant capable de reproduire la dynamique, il reproduit également de manière très satisfaisante la distribution des intensités de précipitations ainsi que celle des durées de sécheresse. Notons également que les principales caractéristiques de la structure de corrélation spatiale sont relativement bien représentées, notamment son caractère asymétrique. Ce dernier vient de la prédominance des régimes perturbés ouest-est. Néanmoins, la corrélation spatiale semble être assez largement améliorée par l'ajout d'un ou plusieurs prédicteurs (la vitesse du vent par exemple). Notons également que notre modèle reproduit plutÃť bien les différentes probabilités de transition et que son pouvoir prédictif est assez bon.

Il convient de souligner que la performance de notre modèle est extrêmement satisfaisante non seulement sur la base d'apprentissage (qui a servi à estimer le modèle) mais aussi sur une base de validation totalement indépendante.

1.2.3 Développement de modèles apparentés

Dans le modèle (1.1), le seuil u doit être suffisamment élevé afin que la distribution des durées de plages sèches soit en accord avec la réalité. Ceci provient du fait que la volatilité σ_t n'est pas suffisamment faible, ce qui engendre parfois de petites précipitations parasites. Dans le cas de la Bretagne Nord, $u = 0.7$, ce qui signifie que notre modèle ne reproduit pas les pluies d'intensité inférieure à 0.7 mm. Afin de pallier cette difficulté, nous avons tenté différentes possibilités.

La variabilité du modèle (1.1) s'écrit : $\sigma_t = \exp(\theta' \mathbf{F}_t) = \exp(\theta_1 + \theta_2 \bar{T}_t + \theta_3 \bar{Pr}_t + \theta_4 \bar{H}_t)$. Afin de bénéficier de plages où la volatilité est suffisamment faible, nous pouvons proposer la forme suivante :

$$\sigma_t = \begin{cases} \exp(\theta_1 + \theta_2 \bar{T}_t + \theta_3 \bar{Pr}_t + \theta_4 \bar{H}_t) & \text{si } \bar{T}_t > -5°C, \bar{P}_t < 1025\ hPa \text{ et } \bar{H}_t > 70\% \\ 0.01 & \text{sinon} \end{cases}.$$

L'idée est d'imposer une volatilité arbitrairement faible dès lors que les covariables prennent des valeurs peu propices à des précipitations. Ce modèle fournit des résultats plutôt satisfaisants tant en termes d'intensité, de longueur des périodes sèches et de dépendance spatiale. Néanmoins, le choix des conditions sur les covariables revêt un caractère arbitraire et subjectif qui limite la capacité de généralisation du modèle. Pourquoi imposer $\bar{P}_t < 1025\ hPa$ et non pas $\bar{P}_t < 1030\ hPa$? Pourquoi $\bar{H}_t > 70\%$ et non $\bar{H}_t > 80\%$? De tels choix requièrent des connaissances a priori et s'avèrent très dépendants de la région considérée. Pour ces raisons, nous n'avons pas retenu ce modèle.

Une deuxième possibilité est de considérer le modèle (1.1) avec un seuil u_t fonction du temps. Nous nous sommes intéressés au cas où le seuil est une fonction du facteur commun, $u(\mathbf{F}_t)$. Cela permet un seuil fonction décroissante du facteur commun, c'est-à-dire d'autant plus élevé que les covariables grande échelle ne sont pas favorables à des précipitations. Néanmoins, dans le cas de la Bretagne Nord, le seuil est en général élevé, conduisant à des intensités trop faibles et à des plages sèches trop longues.

Dans le même registre, nous avons testé le modèle suivant, composé de deux sous-modèles :
Si $\sigma_{t-1} \leq \sigma^*$, nous définissons

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} \geq u_1, \\ 0 & \text{si } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} < u_1, \end{cases}$$

Si $\sigma_{t-1} > \sigma^*$,

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} \geq u_2, \\ 0 & \text{si } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} < u_2, \end{cases}$$

où σ^*, u_1 et u_2 sont des seuils à définir et vérifiant $u_1 > u_2$. Comme précédemment, l'idée est de considérer un seuil plus élevé lorsque la volatilité est faible, i.e. lorsque les conditions grande échelle ne sont pas favorables à des précipitations. Cela permet de limiter les précipitations parasites déjà mentionnées et donc d'augmenter la durée des périodes de sécheresse. L'application de ce modèle à la Bretagne Nord avec $\sigma^* = 0.5$,

$u_1 = 0.7$ et $u_2 = 0.2$ conduit à des résultats satisfaisants, notamment en ce qui concerne les intensités et les périodes sèches.

Cependant, comparé au modèle (1.1), ce modèle impose le choix de deux paramètres supplémentaires, σ^* et u_2, difficiles à estimer de manière automatique. Leur choix peut donc apparaître artificiel. De surcroît, l'interprétation du fait que l'on considère deux sous-modèles associés à deux seuils différents n'est pas très claire, même si cela fait penser à un changement de régime.

Dans le même ordre d'idée, nous avons développé le modèle suivant.
Si $P_{m,t-1} = 0$,

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} \geq u_1, \\ 0 & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} < u_1. \end{cases}$$

Si $P_{m,t-1} \neq 0$,

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} \geq u_2, \\ 0 & \text{si } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} < u_2, \end{cases}$$

où, comme précédemment, u_1 et u_2 sont des seuils vérifiant $u_1 > u_2$. Un seuil plus élevé lorsque les conditions étaient sèches à la date $t-1$ permet d'introduire de la persistance dans les périodes sèches. Ce modèle fournit d'excellents résultats pour $u_1 = 0.6$ et $u_2 = 0.2$. Contrairement au modèle (1.1), il peut générer des précipitations à partir de $0.2\ mm$. Il a cependant légèrement tendance à sous-estimer les précipitations extrêmes. De plus, il impose le choix d'un seuil supplémentaire, u_2, difficilement estimable et interprétable.

Finalement, pour les raisons décrites, nous avons choisi de présenter dans ce chapitre le modèle (1.1). Il permet d'exposer de manière assez simple l'idée principale suivante : les précipitations sont reliées aux précipitations passées (contagion) et la variabilité de leurs incréments est fonction de covariables grande échelle. Comme expliqué ci-après, nous pensons de surcroît que la question du seuil pourrait être en partie réglée en remplaçant la relation (1.2) par une relation non linéaire.

1.2.4 Démarche ayant conduit à cette classe de modèles

L'élaboration de la classe de modèles présentés dans le paragraphe précédent ainsi que du modèle final (1.1) s'est dessinée de manière très progressive. Avant de décrire notre démarche de modélisation, il nous faut introduire un processus de régime de temps $(R_t)_{t=1,\ldots,T}$ qui suit une chaîne de Markov du premier ordre à quelques états. Le nombre d'états ainsi que leurs caractéristiques sont naturellement liés à la localisation de la région ainsi qu'à ses dimensions. Le régime peut dépendre d'un indice sous-jacent comme l'indice NAO ou ENSO ou alors de covariables.

Les régimes de temps principaux observés en France sont les suivants :

- Régime d'ouest classique, accompagné de vents d'ouest et de précipitations se produisant de manière assez étendue du fait de perturbations organisées ;
- Régime de sud, accompagné d'un vent de sud et de précipitations peu organisées ;
- Régime d'est, accompagné de vents d'est continentaux et d'un temps généralement sec.

On comprend donc que dans le cas français, le régime pourrait par exemple dépendre de la direction des vents dominants. A l'échelle européenne, les deux régimes principaux correspondent aux deux état quasi-stationnaires du jet-stream et sont décrits ci-après :

- Régime zonal, responsable d'un grande nombre de dépressions et en particulier des tempêtes hivernales ;
- Régime de blocage, responsable des vagues de froid intenses déferlant de la Scandinavie vers l'Europe de l'ouest.

Le premier modèle testé découplait l'occurrence (notée O) et l'intensité (notée I) comme suit

$$P_{m,t} = I_{m,t}\ O_{m,t}, \tag{1.4}$$

où la variable $O_{m,t} = 1$ s'il pleut au site m pendant la t-ième heure et 0 sinon. L'idée du modèle à facteur commun-contagion se retrouvait dans le processus $I_{m,t}$ d'équation

$$I_{m,t} = \exp\left(\mu + \alpha\ F_t + \mathbf{B}_m^{'} \log \mathbf{I}_{t-1} + \epsilon_{m,t}\right), \tag{1.5}$$

où $\mu \in \mathbb{R}$, $F_t \in \mathbb{R}$, $\alpha \in \mathbb{R}^+$, et les $(\epsilon_{m,t})_{m=1,\dots M\ ;\ t=1,\dots T}$ sont i.i.d. (temporellement et spatialement) de moyenne nulle et d'écart-type constant. μ est un paramètre de position, F_t est un facteur commun à spécifier, α la sensibilité par rapport à ce facteur et $\mathbf{B}$ la matrice contenant les paramètres auto-régressifs, comme dans (1.1). L'exponentielle permet d'assurer la positivité stricte des intensités. Contrairement au modèle (1.1), le facteur commun n'apparaît pas dans la volatilité mais est additif. De ce fait, si l'on prend le logarithme de l'Equation (1.5), on remarque que le logarithme de l'intensité suit quasiment exactement le modèle à facteur commun et contagion considéré par Manski (1993).

Nous avons testé différentes spécifications pour F_t :

- Facteur commun inobservable i.i.d. de loi normale ou de Student ;
- Facteur commun observable, constitué de variables explicatives précédemment mentionnées ;
- Facteur commun dépendant du régime de temps R_t, par exemple de la forme :

$$F_t = FR(R_t) + V_t, \tag{1.6}$$

  où FR est une fonction déterministe qui associe une valeur de précipitation à chaque état. Les valeurs correspondantes sont des paramètres qui doivent être estimés. Les $(V_t)_{t=1,\dots,T}$ représentent un bruit et sont supposés i.i.d. Gaussiens standards spatialement et temporellement.

Notons que nous n'avons pas testé de facteur commun inobservable possédant une dynamique. En effet, l'inférence des modèles à facteur commun est numériquement délicate dans ce cas. Il est nécessaire d'intégrer par rapport à la trajectoire du facteur commun et donc d'évaluer une intégrale multiple dont la dimension est égale au nombre de dates. Par ailleurs, comme précisé précédemment, un facteur commun observable apparaît plus pertinent dans le contexte météorologique.

Concernant le processus d'occurrence, une première possibilité consiste à considérer une chaîne de Markov et à introduire de la dépendance à l'aide d'une matrice de transition locale et dynamique

$$T_{m,t} = \begin{pmatrix} p_{00}(R_t, O_{n,t-1}, n \in \{1,\dots,M\}) & p_{01} \\ p_{10}(R_t, O_{n,t-1}, n \in \{1,\dots,M\}) & p_{11} \end{pmatrix}, \tag{1.7}$$

où p_{xy} est la probabilité de transition de l'état x à l'état y, 0 étant l'état sec et 1 l'état humide. Les probabilités de transition dépendent du régime de temps R_t ainsi que de l'occurrence à tous les sites à la date précédente.

Une autre idée consiste à obtenir $O_{m,t}$ en appliquant une fonction déterministe aux occurrences de la date précédente. Ainsi

$$O_{m,t} = k_{R(t)}(O_{n,t-1},\ \mathbf{d}_{nm},\ n\ \in \{1,\ldots,M\}), \tag{1.8}$$

où $\mathbf{d}_{nm}$ représente le vecteur séparant les sites n and m. La contagion du site n vers le site m dépend en effet à la fois de la distance entre les deux sites mais également de la direction matérialisée par leurs positions respectives. Notons que la fonction k est indexée par $R(t)$. La propagation spatiale des occurrences est en effet dépendante de processus conditionnés par le régime de temps (par exemple le vent dominant). Cette approche apparaît riche et pertinente mais s'avère relativement délicate à mettre en pratique.

La dernière possibilité testée consiste en l'utilisation de modèles "agent-based". Notons $\mathbf{MO}_t$ la matrice d'occurrence en tous les sites à la date t, représentés conformément à leur position géographique. La matrice $\mathbf{MO}_t$ est réinitialisée à chaque changement de régime de temps, l'état initial étant dépendant du régime de temps en question. Afin de modéliser l'éventuelle propagation des précipitations, nous utilisons les modèles de type "agent-based", issus de la littérature concernant le recensement et la migration des populations. Ce courant étudie comment les populations des différentes espèces évoluent dans le temps. Une règle de propagation est associée à chaque régime de temps. Dans le cas d'un régime d'ouest classique, des précipitations sont observées sous la perturbation, qui se propage d'ouest en est. Ainsi, l'occurrence peut être initialisée à 1 pour le site le plus à l'ouest et à 0 pour les autres. Considérons que le régime d'ouest apparaît à la date t_0 et qu'il prévaut pendant 2 périodes. Ainsi,

$$\mathbf{MO}_{t_0} = \begin{pmatrix}1 & 0 & 0\end{pmatrix}\ ,\ \mathbf{MO}_{t_0+1} = \begin{pmatrix}0 & 1 & 0\end{pmatrix}\ ,\ \mathbf{MO}_{t_0+2} = \begin{pmatrix}0 & 0 & 1\end{pmatrix}.$$

Dans le cas d'un régime continental est-ouest classique, le temps est généralement sec et donc

$$\mathbf{MO}_{t_0} = \begin{pmatrix}0 & 0 & 0\end{pmatrix}\ ,\ \mathbf{MO}_{t_0+1} = \begin{pmatrix}0 & 0 & 0\end{pmatrix}\ ,\ \ldots$$

Dans le cas d'un régime de sud, les précipitations ne sont pas aussi bien organisées que dans le cas des deux régimes précédents. Ainsi $\mathbf{MO}_t$ n'est pas uniforme par partie. Des règles probabilistes doivent être ajoutées afin de compléter la description.

Remarquons que l'intérêt d'un F_t de la forme (1.6) réside dans le fait qu'à la fois le facteur commun et le processus d'occurrence dépendent du régime de temps R_t, ce qui introduit une dépendance entre les processus d'intensité et d'occurrence. De manière générale, les idées introduites dans le modèle (1.4) sont intéressantes et riches mais elles se sont avérées assez difficiles à mettre en pratique. Notamment, une difficulté intrinsèque au modèle (1.4) réside dans le fait que le processus I n'est observable que lorsque la variable d'occurrence vaut 1, ce qui engendre des difficultés d'inférence. Cela pose également des difficultés pour simuler le modèle. Quelle valeur choisir pour réinitialiser la récurrence (1.5) après une période sèche ? Par ailleurs, ce modèle implique un grand nombre de composantes.

Ainsi, nous avons cherché un modèle plus parcimonieux, mais restant réaliste. L'idée la plus simple permettant de s'affranchir de la complexité du couple occurrence/intensité

était alors de ne considérer qu'une seule équation et d'y appliquer un seuil. Nous avons conservé une relation du type contagion-frailty, aboutissant au modèle :

$$P_{m,t} = \begin{cases} \alpha F_t + \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} & \text{si } \alpha F_t + \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} \geq u, \\ 0 & \text{si } \alpha F_t + \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} < u, \end{cases} \tag{1.9}$$

où les ϵ_{mt} suivent une loi à paramètres indépendants du temps. Nous avons notamment testé la loi de Pareto. Celle-ci peut présenter un intérêt dans le cas d'un paramètre d'échelle assez faible comparé au seuil (afin que celui-ci ne soit pas souvent dépassé) et d'un paramètre de forme modéré (afin de générer suffisamment d'extrêmes).

Nous souhaitions construire un F_t observable, fonction des covariables disponibles. Les scatter plots des données de précipitation versus les différentes variables explicatives potentielles (température, taux d'humidité, vitesse de vent, ...) font apparaître des relations bilatérales non linéaires. En effet, on n'observe pas de véritable nuage de points, notamment en raison de nets effets de seuil. Les précipitations ne sont par exemple enregistrées que dans une certaine plage de température et d'humidité (d'où l'idée précédemment évoquée d'introduire des effets de seuil dans le modèle (1.1)). Une relation linéaire aurait été surprenante compte-tenu de la grande complexité des phénomènes physiques à l'origine des précipitations. De surcroît, les précipitations ainsi que les variables explicatives ne sont pas distribuées de manière Gaussienne. Pour ces raisons, une régression linéaire n'est pas adaptée et il apparaît nécessaire d'exprimer le facteur commun comme fonction non linéaire des covariables. Une possibilité consiste à utiliser le modèle GAM (Generalized Additive Model). Pour une description détaillée, voir par exemple l'ouvrage de Hastie et Tibshirani (1990). Dans ce cadre, l'équation de récurrence du modèle (1.9) s'écrit

$$P_{m,t} = f_1(\bar{T}_t) + f_2(\bar{P}_t) + f_3(\bar{H}_t) + \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t}. \tag{1.10}$$

Comparé aux modèles GLM (Generalized Linear Model), le modèle GAM présente l'avantage d'attribuer une fonction de réponse spécifique à chaque covariable. De plus ces fonctions peuvent être non paramétriques, ce qui laisse "parler" les données. Nous avons implémenté et estimé ce modèle avec des splines (voir par exemple Smith (1979)), à l'aide du package GAM de R. Néanmoins, la procédure d'estimation GAM n'est pas adaptée aux modèles à seuil. Ainsi, l'estimation du modèle (1.9) nécessite plusieurs étapes. Les fonctions f_i ainsi que la matrice $\mathbf{B}$ sont en effet estimées une première fois par la procédure GAM appliquée à (1.10). Il convient ensuite de réestimer la matrice $\mathbf{B}$ en utilisant la vraisemblance exacte du modèle seuillé, en considérant les f_i donnés par la première étape. S'ensuit alors une nouvelle estimation des fonctions f_i en prenant comme variable de réponse la quantité $P_{m,t} - \mathbf{B}_m^{'}\mathbf{P}_{t-1}$ (et non plus $P_{m,t}$), et ainsi de suite ...

Le F_t estimé était en général très faible, conduisant à des intensités générées par notre modèle trop faibles. On alors eu l'idée, dans (1.9), d'introduire le facteur commun F_t non plus comme terme additif mais dans l'écart-type (les $\epsilon_{m,t}$ étant désormais Gaussiens). Un tel modèle (sans le seuil) peut être estimé grâce au package VGAM de R. Néanmoins, sous contrainte d'un terme additif nul, VGAM n'est pas en mesure de traiter le cas d'une volatilité fonction non linéaire des covariables. Dans le cas linéaire, la vraisemblance du modèle seuillé est facilement connue et notre modèle s'estime facilement par maximum de vraisemblance.

Cette longue démarche a finalement conduit au modèle (1.1), que l'on a fini par choisir après avoir développé les modèles apparentés présentés en Section 1.2.3. Comme décrit précédemment, le modèle (1.1) s'avère très performant dans le cas de la Bretagne Nord.

Néanmoins, nous pensons que l'utilisation d'une relation non linéaire pour la volatilité permettrait d'abaisser le seuil et donc de reproduire les précipitations de très faible intensité. L'implémentation du GAM dans le cas du modèle seuillé est en cours.

1.3 Résultats du Chapitre 3

Ce chapitre présente une nouvelle méthode d'estimation des processus max-stables, basée sur des techniques économétriques. Les processus max-stables apparaissent comme une extension de la théorie des valeurs extrêmes multivariée à la dimension infinie. Avant de décrire les principaux résultats de ce chapitre et de s'intéresser spécifiquement aux processus max-stables, il est naturel d'introduire quelques résultats importants de la théorie des valeurs extrêmes univariée et multivariée.

1.3.1 La théorie des extrêmes univariée

Les premiers résultats concernant la distribution asymptotique du maximum d'un échantillon de variables aléatoires i.i.d. sont dus à Fisher et Tippett (1928). Ils ont ensuite été formalisés par Gnedenko (1943) de la manière suivante :

Théorème 1.1. *Soient X_1, X_2, ...X_T des variables aléatoires i.i.d. de distribution F. Soit $M_T = \max_{t\in 1,...,T} X_t$. S'il existe deux suites de réels $\{a_t > 0\}$ et $\{b_t\}$ telles que*

$$\mathbb{P}\left(\frac{M_T - b_t}{a_t} < z\right) \to G(z), \quad \textit{pour } T \to \infty,$$

où G est non dégénérée, alors G appartient à l'une des 3 familles suivantes :

$$I.\ G(z) = \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\} - \infty < z < +\infty \tag{1.11}$$

$$II.\ G(z) = \begin{cases} 0 & z \le \mu, \\ \exp\left[-\left(\frac{z-\mu}{\sigma}\right)^{-\alpha}\right] & z > \mu, \end{cases} \tag{1.12}$$

$$III.\ G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-\mu}{\sigma}\right)^{\alpha}\right]\right\} & z < \mu, \\ 1 & z \ge \mu, \end{cases} \tag{1.13}$$

où μ, $\sigma > 0$ et $\alpha > 0$ sont respectivement des paramètres de position, d'échelle et de forme.

Les familles I, II et III sont respectivement connues sous le nom de distributions de Gumbel, Fréchet et Weibull. Ces distributions sont respectivement à queue légère, lourde et bornée supérieurement.

La formulation du Théorème 1.1 n'est pas très pratique pour l'inférence étant donné qu'elle impose le choix d'une famille au préalable. Heureusement, Von Mises (1954) et Jenkinson (1955) ont séparément introduit la Generalized Extreme Value Distribution (GEV), définie par

$$G(z) = \exp\left(-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right) \text{ pour } z \text{ vérifiant } 1 + \xi\left(\frac{z-\mu}{\sigma}\right) > 0. \tag{1.14}$$

ξ est le paramètre de forme[4] et caractérise le type de la loi du maximum :

- $\xi < 0$ correspond à la distribution the Weibull (III) ;
- $\xi = 0$ correspond à la distribution de Gumbel (II) ;

[4]Notons que ξ est égal à l'inverse du coefficient α introduit dans le Théorème 1.1.

- $\xi > 0$ correspond à la distribution de Fréchet (I).

Suivant la loi du maximum, on dit que la distribution F appartient au domaine d'attraction de la loi de Weibull, de Gumble ou de Fréchet.

Brièvement, le Théorème 1.1 provient du fait que si la loi asymptotique du maximum existe, alors elle est nécessairement max-stable, ce qui s'écrit :

Définition 1.1. *Une distribution G est dite max-stable si elle vérifie, pour deux suites de réels $\{c_t > 0\}$ et $\{d_t\}$,*

$$G^t(c_t\ z + d_t) = G(z),\ \ z \in \mathbb{R}, t \in \mathbb{N}. \tag{1.15}$$

Cela signifie que la distribution est stable par l'opérateur "maximum". Il apparaît ensuite qu'une distribution max-stable univariée est nécessairement de l'un des trois types donnés par le Théorème 1.1.

Notons que les hypothèses du Théorème 1.1 peuvent être relâchées au cas des séries stationnaires (donc avec dépendance). Voir par exemple Leadbetter *et al.* (1989) pour une description détaillée.

Théorème 1.2. *Soient $X_1, \ldots, X_T$ un processus stationnaire et $X_1^*, \ldots, X_T^*$ une suite de variables i.i.d. de même distribution marginale. Notons $M_T = \max(X_1, \ldots, X_T)$ et $M_T^* = \max(X_1^*, \ldots, X_T^*)$. Sous certaines conditions de régularité,*

$$\mathbb{P}\left(\frac{M_T^* - b_t}{a_t} < z\right) \to G_1(z) \text{ quand } T \to \infty, \text{ pour deux suites normalisantes } \{a_t > 0\} \text{ et } \{b_t\},$$

où G_1 est non dégénérée, si et seulement si

$$\mathbb{P}\left(\frac{M_T - b_t}{a_t} < z\right) \to G_2(z) \text{ quand } T \to \infty,$$

où

$$G_2(z) = G_1(z)^\theta,$$

pour une constante θ vérifiant $0 < \theta \leq 1$.

La constante θ est appelée indice extrêmal. Dans le cas de variables dépendantes, les extrêmes forment généralement des clusters dont la taille moyenne est égale à θ^{-1}. Le résultat du Théorème 1.2 est particulièrement important car il élargit la portée pratique de la GEV, l'hypothèse de variables i.i.d. du Théorème 1.1 étant rarement vérifiée en pratique.

La GEV possède une forme totalement paramétrique, ce qui présente un avantage pour l'inférence. Une large littérature concerne l'estimation des paramètres de la GEV. La méthode du maximum de vraisemblance a par exemple été étudiée par Prescott et Walden (1980) et Hosking (1985). Elle offre de la flexibilité mais l'estimateur correspondant ne possède pas toujours les propriétés asymptotiques classiques car le domaine de définition est fonction des paramètres. Smith (1985) étudie cet aspect en détails. Parmi les approches concurrentes, on peut notamment citer celles fondées sur les moments pondérés (Hosking *et al.*, 1985) ou encore les statistiques d'ordres (de Haan, 1990). Il convient tout de même de préciser que quelle que soit la méthode, l'incertitude sur les paramètres est importante si le nombre de données est insuffisant.

Si cette théorie univariée est très utile dans le cas d'un seul site, elle ne permet pas de prendre en compte la dépendance entre différents sites. Il s'avère donc nécessaire de recourir à d'autres outils en vue de quantifier les risques environnementaux, spatiaux par nature.

1.3.2 La théorie des extrêmes multivariée

La théorie des extrêmes multivariée se concentre sur les maxima composante par composante d'un vecteur. Si la théorie probabiliste est bien établie, les aspects statistiques sont encore en profond développement.

Soient $(Z_{1,t}, \ldots, Z_{M,t})'$, $t = 1, \ldots, T$, une suite de vecteurs aléatoires M-dimensionnels i.i.d. et $M_T = (M_{1,T}, \ldots, M_{M,T})'$ le vecteur des maxima de chaque composante, i.e. $M_{m,T} = \max(Z_{m,1}, \ldots, Z_{m,T})$. Si l'on suppose qu'il existe des suites normalisantes $\{a_{m,T} > 0\}$, $\{b_{m,T}\}$ telles que la distribution du maximum de chaque composante soit non dégénérée, alors la distribution limite du vecteur renormalisé

$$\left(\frac{M_{1,T} - b_{1,T}}{a_{1,T}}, \ldots, \frac{M_{M,T} - b_{M,T}}{a_{M,T}} \right) \tag{1.16}$$

peut être caractérisée (voir Théorèmes 1.3 et 1.4).

Notons tout d'abord que chaque composante possède une distribution limite de type GEV. Sans perte de généralité, nous supposons en général que les marges suivent une loi de Fréchet standard, obtenue par transformation des marges véritables. Si une variable aléatoire Y suit la $GEV(\mu, \sigma, \xi)$, alors $Z = \left[1 + \xi \left(\frac{Y - \mu}{\sigma} \right) \right]^{\frac{1}{\xi}}$ a pour distribution une Fréchet standard, de fonction de répartition $F(z) = \exp(-1/z)$. Resnick (1987) montre que le domaine d'attraction est préservé par toute transformation monotone croissante des marges. Cette standardisation est similaire à la théorie des copules. Néanmoins, ici la référence n'est pas la loi uniforme mais une loi suggérée par la théorie des valeurs extrêmes.

D'un point de vue probabiliste, les distributions multivariées sont bien caractérisées ; voir par exemple Resnick (1987). On compte deux représentations fondamentales pour la distribution des extrêmes multivariés (Resnick, 1987), appelée "Multivariate Generalized Extreme Values Distribution" (MGEV).

Théorème 1.3. *Une distribution multivariée G est une distribution limite du vecteur aléatoire* (1.16) *si et seulement si*

$$G(z_1, \ldots, z_M) = \exp[-V(z_1, \ldots, z_M)] \quad \textit{pour} \quad (z_1, \ldots, z_M)' \in \mathbb{R}^M, \tag{1.17}$$

où la fonction V est appelée la mesure exponentielle (Resnick, 1987, page 268) et vérifie

$$V(\infty, \ldots, \infty, z_m, \infty, \ldots, \infty) = \frac{1}{z_m} \quad \forall m = 1, \ldots, M, \tag{1.18}$$

et

$$V(h\ z_1, \ldots, h\ z_M) = h^{-1} V(z_1, \ldots, z_M). \tag{1.19}$$

La première propriété assure des marges de Fréchet standards tandis que la deuxième impose que la fonction V soit homogène d'ordre -1.

Remarque 1.1. *Deux cas particuliers intéressants sont*

$$V(z_1, \ldots, z_M) = \frac{1}{z_1} + \cdots + \frac{1}{z_M} \quad \textit{et} \quad V(z_1, \ldots, z_M) = \frac{1}{\min(z_1, \ldots, z_M)},$$

correspondant respectivement à la dépendance parfaite et à l'indépendance parfaite.

Notons que toute distribution MGEV est max-stable (voir Resnick (1987)). Du fait de l'homogénéité de V, les MGEV possèdent différentes représentations spectrales. La plus connue est due à Pickands (1981) et s'écrit

$$V(z_1,\dots,z_M) = \int_{S_M} \max_{m=1,\dots,M} \left(\frac{w_m}{z_m}\right) dH(w_1,\dots,w_M), \tag{1.20}$$

où H est une mesure positive définie sur le simplexe de dimension $M-1$, S_M, défini par $S_M = \{\mathbf{w} \in {\mathbb{R}_+}^M \mid w_1 + \cdots + w_M = 1\}$. H désigne la mesure spectrale (encore appelée angulaire) et détermine la structure de dépendance du vecteur aléatoire (1.16). Elle vérifie la condition de moments

$$\int_{S_M} w_m \, dH(w_1,\dots,w_M) = 1, \quad m = 1,\dots,M, \tag{1.21}$$

imposée par (1.18). A la différence du cas univarié, la distribution limite G ne possède pas de forme paramétrique simple car V est une fonction quelconque vérifiant (1.18) et (1.19) (ou de manière équivalente H est une fonction quelconque vérifiant (1.21)). Une solution est de considérer des familles paramétriques denses dans l'espace général des fonctions H.

Dans le cas bivarié, la famille Gumbel logistique est souvent utilisée. Elle est définie par

$$G(z_1,z_2) = \exp\left[-(z_1^{-\frac{1}{\alpha}} + z_2^{-\frac{1}{\alpha}})\right]^{\alpha}, \tag{1.22}$$

où $z_1 > 0$, $z_2 > 0$ et $0 < \alpha < 1$. Cette famille est particulièrement appréciée car elle peut modéliser à la fois la dépendance et l'indépendance totales à l'aide de l'unique paramètre α (respectivement $\alpha \to 0$ et $\alpha = 1$). Elle manque cependant de flexibilité.

Une autre approche consiste à utiliser des méthodes non paramétriques mais il s'avère délicat de contraindre un estimateur à vérifier une condition du type (1.21) (Coles *et al.*, 2001). De plus, du fait du fléau de la dimension, ces méthodes se sont essentiellement cantonnées au cas bivarié (Fougères, 2004; Boldi et Davison, 2007; Einmahl et Segers, 2009).

Le deuxième type de représentation des distributions des extrêmes multivariés G ayant des marges de Fréchet standard est donné par le théorème suivant.

Théorème 1.4. *Une distribution multivariée G est une distribution limite du vecteur aléatoire* (1.16) *si et seulement si*

$$G(z_1,\dots,z_M) = \exp\left(-A(w_1,\dots,w_M)\sum_{m=1}^{M} x_m^{-1}\right),$$

où $w_m = \dfrac{z_m^{-1}}{\sum_{k=1}^{M} z_k^{-1}}, m = 1,\dots,M$, *et $A(.)$ est une fonction convexe sur S_M satisfaisant* $\forall (w_1,\dots,w_M) \in S_M, \max(w_1,\dots,w_M) \le A(w_1,\dots,w_M) \le 1$.

A est appelée fonction de dépendance (Pickands, 1981).

Comme dans le cas univarié, l'hypothèse i.i.d. peut être relâchée. Sous certaines conditions de mélange (Leadbetter *et al.*, 1989), les distributions des extrêmes multivariés peuvent encore être limites dans le cas des séries stationnaires, ce qui élargit la portée pratique des Théorèmes 1.3 et 1.4.

Même lorsque l'on s'est ramené à une forme paramétrique, l'inférence s'avère délicate. Considérons par exemple la représentation du Théorème 1.3. L'estimation par maximum de vraisemblance implique alors la différentiation de l'expression $\exp[-V(z_1,\ldots,z_M)]$ par rapport à $z_1,\ldots,z_M$. Le nombre de termes de la dérivée est égal au nombre de partitions de M et il y a donc explosion combinatoire. A titre indicatif, pour $M = 10$, le nombre de termes est proche de 10^5. Ainsi, cette méthode n'est pas applicable en général, à l'exception de certains cas permettant des simplifications.

Si la théorie multivariée permet de prendre en compte la dépendance entre différents sites, elle ne permet pas d'effectuer d'"interpolation" ou d'agrégation spatiale, d'où l'intérêt de l'extension à la dimension infinie, présentée ci-après.

1.3.3 Les processus max-stables

Définition et représentation spectrale

Il est crucial pour les applications d'être en mesure de caractériser le comportement du maximum d'un processus stochastique, par exemple une variable météorologique. Les processus max-stables constituent un outil très adéquat car ils correspondent à la théorie des extrêmes multivariés quand le nombre de sites tend vers l'infini. Ainsi, ils se situent aux confins de la théorie des valeurs extrêmes et de la géostatistique.

Considérons un processus $V(\mathbf{x})$, $\mathbf{x} \in E \subset \mathbb{R}^d$ (représentant par exemple la vitesse du vent), à trajectoire continue. Nous souhaitons caractériser le processus résultant lorsque l'on considère le maximum temporel en tous points de l'espace, c'est-à-dire

$$\left\{ \frac{\max_{n=1}^{N} V_n(\mathbf{x}) - b_N(\mathbf{x})}{a_N(\mathbf{x})} \right\}_{\mathbf{x}\in E}, \tag{1.23}$$

où les $V_n(\mathbf{x})$ sont des réplications i.i.d. du processus $V(\mathbf{x})$ et où $\{a_N(\mathbf{x}) > 0\}$ et $\{b_N(\mathbf{x})\}$ sont des suites de fonctions. Notons que l'on est souvent intéressé par les maxima annuels. Dans ce cas, n est le n-ième jour et $N = 365$. de Haan (1984) a montré que la limite quand $N \to \infty$ d'un tel processus est nécessairement max-stable (si elle est non dégénérée).

Définition 1.2. *Un processus stochastique $\{S(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ à trajectoires continues est dit max-stable s'il existe $\{c_T(\mathbf{x}) > 0\}$ et $\{b_T(\mathbf{x})\}$ telles que si $S_1(\mathbf{x}),\ldots,S_T(\mathbf{x})$ sont des réplications i.i.d. de $S(\mathbf{x})$, alors*

$$\left\{ \frac{\max_{t=1}^{T} S_t(.) - b_T(.)}{a_T(.)} \right\} \stackrel{d}{=} \{S(.)\}. \tag{1.24}$$

Par rapport à la théorie multivariée, les processus max-stables permettent d'obtenir la loi conditionnelle à des observations dans tout l'espace. Ils permettent donc par exemple de déterminer la probabilité que le maximum dépasse un certain niveau en un lieu donné, connaissant les observations en des sites voisins. Ils offrent également la possibilité d'effectuer de l'agrégation spatiale, ce qui est fait dans le Chapitre 4. Les processus max-stables contiennent la totalité de l'information sur la dépendance spatiale. Ainsi, comme nous le verrons dans le Chapitre 4, une mesure de risque basée sur de tels processus permet de tenir compte de la dépendance entre les différents contrats souscrits par une compagnie d'assurance.

Notons que dans le Chapitre 3, nous nous intéressons directement au processus max-stable des maxima. Ainsi, les observations haute-fréquence de la variable ne jouent plus

aucun rôle. Il convient de noter que l'échelle temporelle est très différente entre les Equations (1.23) et (1.24). Si $S(\mathbf{x})$ correspond à des maxima annuels, t désigne la t-ième année d'observation et T est le nombre d'années d'observations. Typiquement, T est de l'ordre de 30 ans. Supposer que le processus des maxima suit un processus max-stable suppose que la limite $N \to \infty$ (dans (1.23)) a été atteinte, c'est à dire, dans le cas de maxima annuels, que le nombre de jours dans une année est suffisant. Cela suppose également que les maxima d'un jour sur l'autre soient considérés comme indépendants.

Une conséquence directe de la Définition 1.2 est que les distributions marginales unidimensionnelles et multi-dimensionnelles sont max-stables et appartiennent respectivement aux classes GEV et MGEV. On a, $\forall \mathbf{x} \in \mathbb{R}^d, S(\mathbf{x}) \sim GEV\left[\mu(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x})\right]$ où $\mu(\mathbf{x})$, $\sigma(\mathbf{x})$ et $\xi(\mathbf{x})$ sont respectivement les paramètres de position, d'échelle et de forme au lieu $\mathbf{x}$. Si la région d'étude est étendue, la distribution marginale varie au sein de la région. La variation des paramètres marginaux est due à des effets spatiaux régionaux. Elle peut être en grande partie caractérisée par des covariables comme la latitude ou l'altitude, ce qui génère une première source de dépendance. Une fois que l'effet des marges a été pris en compte, subsiste un effet spatial local lié à l'étendue des différents événements extrêmes. Cela crée alors une dépendance "résiduelle" qui provient du fait que différents sites sont affectés par le même phénomène.

Cette distinction entre effets régionaux et locaux s'apparente à la décomposition des séries temporelles en une moyenne (tendance et effets saisonniers) et un bruit stationnaire décrit par une structure de covariance (Brockwell et Davis, 2009, Par. 1.3.3). L'effet régional peut être assimilé à une caractéristique climatique (dont l'échelle de temps est de l'ordre de quelques décennies) alors que l'effet local peut être considéré comme météorologique. Les paramètres marginaux reflètent un comportement moyen de long-terme, notamment dû à l'orographie, et varient sur une échelle spatiale assez importante. En revanche, les phénomènes météorologiques spatiaux sont, particulièrement dans le cas des extrêmes, souvent plus localisés (typiquement les épisodes de vent violent).

Dans ce chapitre, nous nous affranchissons de l'effet des marges, ce qui revient en quelque sorte à standardiser le climat à travers la région. Nous ne considérons que la dépendance "résiduelle" de type météorologique. Les marges choisies sont de type Fréchet standard. En effet, comme dans le cas multivarié, on peut s'y ramener sans perte de généralité via la transformation

$$Z(\mathbf{x}) = \left[1 + \xi\left(\frac{G(\mathbf{x}) - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]^{\frac{1}{\xi(\mathbf{x})}}.$$

De telles marges imposent $a_T(\mathbf{x}) = T$ et $b_T(\mathbf{x}) = 0$ dans (1.24), qui devient $\{T^{-1} \max_{t=1,\ldots,T} Z_t(.)\} \stackrel{d}{=} \{Z\}$. Notons qu'une telle transformation implique une première étape lors de laquelle les paramètres de la distribution marginale doivent être estimés en chaque site[5].

La définition de la max-stabilité donnée par (1.24) définit parfaitement une certaine classe de processus mais ne nous dit pas comment les construire et les simuler. Ceci est permis grâce aux représentations spectrales de de Haan (1984) et Schlather (2002). On note $C^+(\mathbb{R}^d)$ l'ensemble des processus stochastiques positifs et à trajectoires continues sur $\mathbb{R}^d$.

Première représentation (de Haan, 1984) :
Soient $\{(\xi_i, \mathbf{c_i})\}_{i\geq 1}$ les points d'un processus ponctuel de Poisson, Π, sur $(0, +\infty) \times \mathbb{R}^d$

[5]Celle-ci est laissée de côté dans le Chapitre 3.

ayant pour mesure d'intensité $d\Lambda(\xi, \mathbf{c}) = \xi^{-2}\ d\xi\ \nu(d\mathbf{c})$, où ν est une σ-finie sur $\mathbb{R}^d$. Si $Z(\mathbf{x})$ est un processus max-stable dans $C^+(\mathbb{R}^d)$ et ayant des marges de Fréchet standards, alors il existe une famille de fonctions continues non-négatives $\{f_{\mathbf{x}}(\mathbf{c}) : \mathbf{c} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d\}$ vérifiant

- pour tout $\mathbf{x} \in \mathbb{R}^d$, $\displaystyle\int_{\mathbb{R}^d} f_{\mathbf{x}}(\mathbf{c})\ d\mathbf{c} = 1$,
- pour tout compact $K \subset \mathbb{R}$, $\displaystyle\int_{\mathbb{R}^d} \sup_{\mathbf{x}\in K} f_{\mathbf{x}}(\mathbf{c})\ \nu(d\mathbf{c}) < +\infty$,

telle que

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \overset{d}{=} \{\max_{i\geq 1} \xi_i f_{\mathbf{x}}(\mathbf{c_i})\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{1.25}$$

Inversement, tout processus défini par (1.25) est un processus max-stable ayant des marges de Fréchet standards.

La preuve du sens direct se trouve dans de Haan (1984) dans le cas $\mathbf{x} \in \mathbb{R}_+$ et $\mathbf{c} \in [0, 1]$ ainsi que dans de Haan et Ferreira (2007). Dans le dernier cas, la preuve est donnée pour $\mathbf{x} \in \mathbb{R}$, $\mathbf{c} \in\ [0, 1]$ et ν étant la mesure de Lebesgue. L'implication contraire est facile à prouver.

Deuxième représentation (Penrose, 1992; Schlather, 2002; de Haan et Ferreira, 2007) :
Soient $\{\xi_i\}_{i\geq 1}$ les points d'un processus ponctuel de Poisson sur $(0, +\infty)$, d'intensité $d\Lambda(\xi) = \xi^{-2}d\xi$ et $Y_1, \ldots, Y_n$ des réplications i.i.d. d'un processus stochastique tel que $\mathbb{E}(\max\{0, Y(\mathbf{x})\}) = 1, \forall \mathbf{x} \in \mathbb{R}^d$ et $\mathbb{E}(\sup_{\mathbf{x}\in K} \max[0, Y(\mathbf{x})]) < +\infty$ pour tout compact $K \subset \mathbb{R}^d$. Alors tout processus max-stable dans $C^+(\mathbb{R}^d)$ et ayant des marges de Fréchet standards peut s'écrire

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \overset{d}{=} \left\{\max_i \{\xi_i \max(0, Y_i(\mathbf{x}))\}\right\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{1.26}$$

Inversement, tout processus défini par (1.26) est un processus max-stable ayant des marges de Fréchet standards.

Remarque 1.2. *Si Y est un processus stationnaire, alors le processus défini par* (1.26) *est stationnaire. La sationnarité est ici entendue au sens stricte (stationnarité du premier ordre) : il y a invariance par translation de toutes les distributions jointes finidimensionnelles.*

Dans le papier de Schlather (2002), on trouve l'implication contraire. Pour une preuve du fait que tout processus max-stable dans $C^+(\mathbb{R}^d)$ ayant des marges de Fréchet standards peut s'écrire comme dans (1.26), on réfère à de Haan et Ferreira (2007). Notons que de Haan et Ferreira (2007) considèrent $[0, 1]$ au lieu de $\mathbb{R}^d$, par souci de simplicité.

Différents modèles de processus max-stables

Le modèle de Smith :
Dans un manuscrit non publié, Smith (1990) utilise la représentation spectrale (1.25) afin de proposer un modèle paramétrique pour les extrêmes spatiaux. Il considère le cas

particulier où ν est la mesure de Lebesgue sur $\mathbb{R}^d$ et $f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma_0}}(\mathbf{x}-\mathbf{c})$ où $f_{\mathbf{\Sigma_0}}$ est la densité d'une loi normale d-dimensionnelle de moyenne $\mathbf{0}$ et de matrice de covariance $\mathbf{\Sigma_0}$:

$$f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma_0}}(\mathbf{x}-\mathbf{c}) = (2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma_0}|^{-\frac{d}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{c})'\mathbf{\Sigma_0}^{-1}(\mathbf{c}-\mathbf{x})\right]. \tag{1.27}$$

Ainsi, ce modèle est totalement paramétrique, le paramètre étant la matrice de covariance $\mathbf{\Sigma_0}$. Cette dernière contient la totalité de l'information relative à la structure de dépendance spatiale. Le principal avantage de ce modèle réside dans son interprétation en termes de processus de tempêtes (Smith, 1990), la forme de ces tempêtes étant régie par la matrice de covariance. Le deuxième avantage est que la densité trivariée est connue explicitement (Genton *et al.*, 2011), contrairement au modèle de Schlather, décrit ci-après. La Figure 1.4 montre une réalisation du processus de Smith (vue du dessus) dans un cas isotrope, ce qui se matérialise par des cercles.

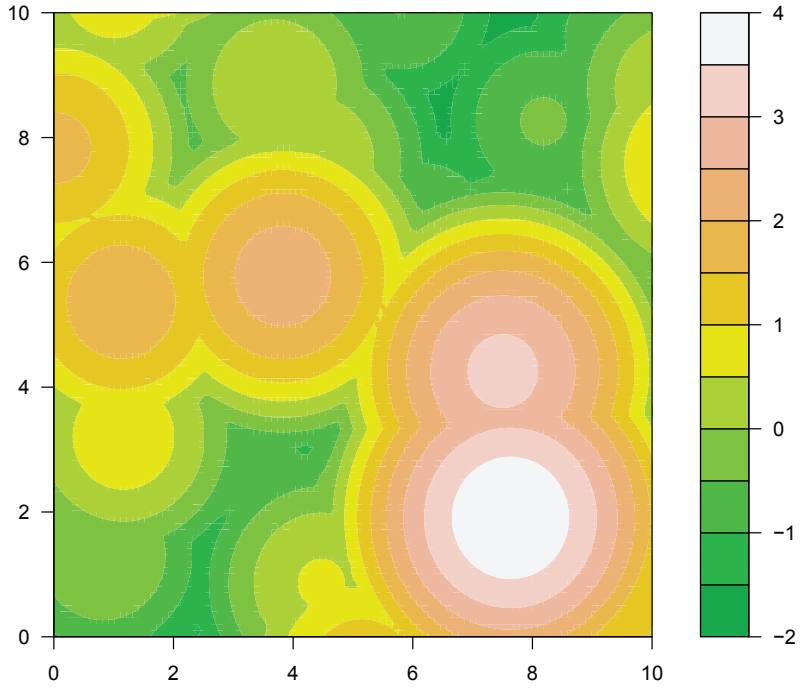


FIGURE 1.4 – Exemple de réalisation du processus de Smith (échelle logarithmique).

Le modèle de Schlather :
Schlather (2002) propose de poser $Y(\mathbf{x}) = \sqrt{2\pi}\ \epsilon(\mathbf{x})$ dans (1.26), où $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ est un processus Gaussien standard ayant pour fonction de corrélation $\rho(h)$. L'ensemble des fonctions de corrélation provenant de la géostatistique peuvent être utilisées, ce qui permet de modéliser une grande variété de comportements. La Figure 1.5 propose une réalisation du processus de Schlather.

Le processus Gaussien géométrique :
Le processus de Schlather ne permet pas de rendre compte de l'indépendance totale. De

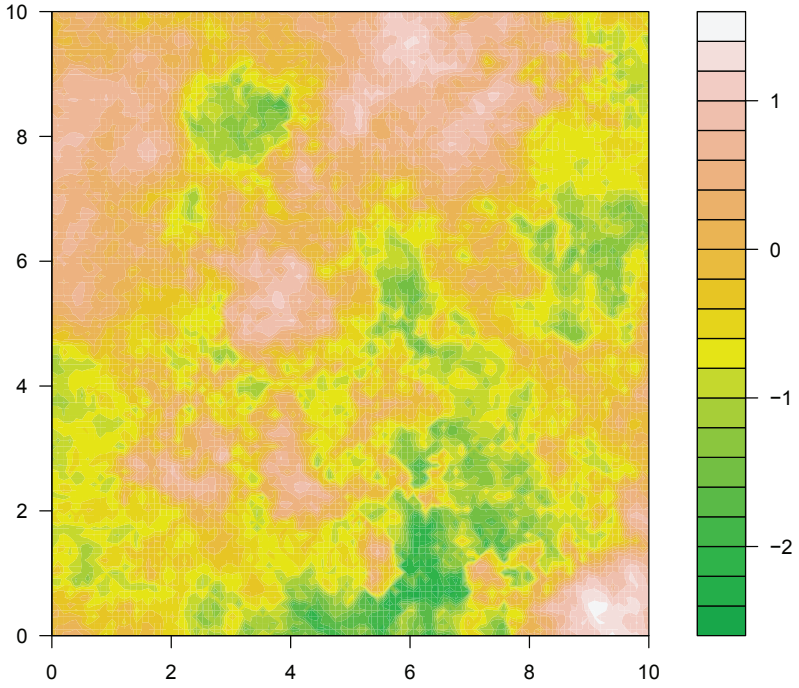


FIGURE 1.5 – Exemple de réalisation du processus de Schlather (échelle logarithmique).

ce fait, Davison (2003) a introduit le modèle Gaussien géométrique. Dans (1.26), il prend pour $Y(\mathbf{x})$ un processus log normal et non Gaussien :

$$Y(\mathbf{x}) = \exp\left(\sigma\epsilon(\mathbf{x}) - \frac{\sigma^2}{2}\right),$$

où $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ est un processus Gaussien standard de variance σ^2 et de fonction de corrélation $\rho(.)$.

Le processus de Brown-Resnick :

Le processus Gaussien géométrique est un cas particulier d'un modèle introduit par Brown et Resnick (1977). Kabluchko *et al.* (2009) introduisent une généralisation de ce dernier en considérant dans (1.26) :

$$Y(\mathbf{x}) = \exp\left(W(\mathbf{x}) - \frac{\sigma^2(\mathbf{x})}{2}\right),$$

où $\{W(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ est un processus Gaussien de moyenne nulle à incréments stationnaires et
$\sigma^2(\mathbf{x}) = \mathrm{Var}[W(\mathbf{x})], \forall \mathbf{x} \in \mathbb{R}^d$. Le processus W et donc le processus de Brown-Resnick correspondant sont totalement caractérisés par la variance $\sigma^2(\mathbf{x})$ et le semi-variogramme, défini par

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{Var}[W(\mathbf{x}+\mathbf{h}) - W(\mathbf{x})], \mathbf{h} \in \mathbb{R}^d. \tag{1.28}$$

Notons que le processus de Smith et le processus Gaussien géométrique sont des cas particuliers du processus de Brown-Resnick, respectivement de variogrammes $\gamma(\mathbf{h}) = \frac{\|\mathbf{h}\|^2}{2}$ et $\gamma(\mathbf{h}) = \sigma^2[1 - \rho(\|\mathbf{h}\|)]$.

Une mesure de dépendance spatiale

Une connaissance complète de la structure de dépendance entre les composantes d'un vecteur max-stable nécessite de connaître la distribution complète via une spécification de la mesure exponentielle $V(.)$ (voir Théorème 1.3) ou de la mesure angulaire $H(.)$ (voir (1.20)). Néanmoins, cette spécification est lourde, notamment dans le cas d'une dimension élevée. Ainsi, différentes mesures résumant la dépendance extrêmale ont été introduites. L'une d'entre elles, le coefficient extrêmal Θ (Schlather et Tawn, 2003) est défini par

$$\mathbb{P}(Z_1 \leq z, \ldots, Z_M \leq z) = \exp\left(-\frac{V(1,\ldots,1)}{z}\right) = \exp\left(-\frac{\Theta}{z}\right). \tag{1.29}$$

La première égalité provient de la propriété d'homogénéité de la mesure exponentielle V (voir (1.19)). Seule la probabilité que toutes les variables $Z_1, \ldots, Z_M$ soient inférieures à z est prise en compte et le coefficient extrêmal n'est donc pas aussi riche que la distribution jointe complète. Cependant, il s'agit d'une mesure intéressante du degré de dépendance et son interprétation est aisée. Le coefficient extrêmal mesure en effet le nombre de sites indépendants : $\Theta = 1$ correspond à la dépendance totale alors que $\Theta = M$ correspond à l'indépendance totale.

Dans le cas bivarié, on a la propriété suivante, $\lim_{z\to\infty} P(Z_2 > z | Z_1 > z) = 2 - \Theta$, qui fournit la probabilité que l'une des variables prenne une valeur élevée sachant que c'est le cas pour l'autre variable.

Les propriétés d'ergodicité et de mélange des processus max-stables sont étroitement liées au coefficient extrêmal bivarié. Kabluchko et Schlather (2010) s'intéressent aux propriétés de mélange des processus max-infiniment divisibles, dont les processus max-stables représentent une sous-classe. Dans le cas max-stable, ils introduisent le coefficient $\{r(h)\}_{h\in\mathbb{R}}$ défini par $r(h) = 2 - \Theta(h)$, où $\Theta(h)$ est le coefficient extrêmal et h correspond à la distance entre les deux sites considérés. Leur Théorème 3.1 stipule que pour un processus max-stable stationnaire et mesurable $\{Z(x)\}_{x\in\mathbb{R}}$ ayant des marges de Fréchet standards, Z est mélangeant (voir Chapitre 4 pour la définition) si et seulement si $\lim_{h\to\infty} \Theta(h) = 2$. Dans le Théorème 3.2, ils montrent que Z est ergodique si et seulement si $\lim_{l\to\infty} \frac{1}{l}\int_{h=1}^{l} r(h) = 0$. On peut en déduire par extension à $\mathbb{R}^d$ que le processus de Smith est mélangeant sur $\mathbb{R}^d$ alors que le processus de Schlather n'est ni ergodique, ni mélangeant sur $\mathbb{R}^d$. Le processus de Brown-Resnick est mélangeant si et seulement si son variogramme vérifie $\lim_{\|\mathbf{h}\|\to\infty} \gamma(\mathbf{h}) = +\infty$.

Les méthodes d'inférence de type vraisemblance composite

Comme nous l'avons vu, un certain nombre de résultats probabilistes de portée importante ont été établis dans le cas des processus max-stables. Cependant, la théorie des processus max-stables est encore en profond développement, notamment les aspects statistiques. Les processus max-stables proposent en effet encore énormément de défis. Par exemple, leur inférence s'avère particulièrement délicate, ce qui n'est pas surprenant quand on sait

que c'était déjà le cas avec la théorie multivariée. En effet, en général, la densité multidimensionnelle (dans l'espace) ne peut être calculée analytiquement en grande dimension[6]. En revanche, dans le cas des modèles précédemment mentionnés (Smith, Schlather, Brown-Resnick), elle peut être calculée en dimension 2 ou 3. Cela a conduit à l'utilisation de la vraisemblance composite. L'approche par vraisemblance composite dépasse largement le cadre des processus max-stables. Soit un modèle statistique paramétrique de fonction de densité $\{l(\mathbf{z};\mathbf{\Sigma}) : \mathbf{z} \in \mathcal{Z}, \mathbf{\Sigma} \in \mathcal{U}$, où $\mathcal{Z} \subseteq \mathbb{R}^M, \mathcal{U} \subseteq \mathbb{R}^q, M \geq 1$ et $q \geq 1$. Soit un ensemble d'événements $\{A_i \in \mathcal{F}, i \in I \subseteq \mathbb{N}\}$, où $\mathcal{F}$ est une σ algèbre sur $\mathcal{Z}$. La log vraisemblance composite pondérée (WCL) (Lindsay, 1988), est définie par

$$L_C(\mathbf{\Sigma}) = \sum_{i \in I} w_i \log l(\mathbf{z} \in A_i \,; \mathbf{\Sigma}),$$

où $l(\mathbf{z} \in A_i; \Sigma) = l(\{z_m \in \mathcal{Z} : z_m \in A_i\} \,; \mathbf{\Sigma})$, $\mathbf{z} = (z_1, \ldots, z_M)'$ et $\{w_i, i \in I\}$ est un ensemble de poids. Il est naturel de considérer l'estimateur WCL qui maximise L_C de manière globale.

Padoan *et al.* (2010) définissent la log vraisemblance composite par paires

$$L_{CP}(\mathbf{\Sigma}) = \sum_{t=1}^{T} \sum_{m_1 < m_2} w_{m_1,m_2} \log \left[l(\mathbf{z}_t(\mathbf{x}_{m_1}), \mathbf{z}_t(\mathbf{x}_{m_2}); \mathbf{\Sigma}) \right],$$

où les $\mathbf{z}_t, t = 1, \ldots, T$, sont les observations du processus. Genton *et al.* (2011) et Huser et Davison (2013) utilisent une vraisemblance basée sur des triplets, respectivement dans les cas du processus de Smith et de Brown-Resnick.

Si les variables aléatoires $\mathbf{Z}_t, t = 1, \ldots, T$, sont i.i.d., alors sous les conditions de régularité standards, l'estimateur correspondant est fortement convergent et asymptotiquement normal quand $T \to +\infty$ et M est fixé (Lindsay, 1988; Varin et Vidoni, 2005) :

$$\sqrt{T}(\hat{\mathbf{\Sigma}}_\mathbf{C} - \mathbf{\Sigma_0}) \xrightarrow[T \to \infty]{d} N(\mathbf{0}, \tilde{\mathbf{\Omega}}^{-1}),$$

où $\tilde{\mathbf{\Omega}} = \mathbf{\Omega_d}\ \mathbf{\Omega}^{-1}\ \mathbf{\Omega_d}$, avec

$$\mathbf{\Omega} = V_{\mathbf{\Sigma_0}} \left[\frac{\partial \log L_C(\mathbf{\Sigma_0})}{\partial \mathbf{\Sigma}} \right] \text{ et } \mathbf{\Omega_d} = E_{\mathbf{\Sigma_0}} \left[-\frac{\partial^2 \log L_C(\mathbf{\Sigma_0})}{\partial \mathbf{\Sigma} \partial \mathbf{\Sigma}'} \right] .$$

Des poids égaux sont le plus souvent utilisés mais des gains d'efficacité non négligeables peuvent être obtenus en utilisant des poids inégaux appropriés. Sang et Genton (2013) proposent la "tapered" vraisemblance composite. Les poids sont égaux à 1 pour des paires ou triplets de sites dont la distance est inférieure au "taper range" γ_s, et égaux à 0 sinon. Le "taper range" optimal est déterminé en minimisant le déterminant ou la trace de la matrice $\tilde{\mathbf{\Omega}}$.

Les principaux avantages de ces méthodes basées sur la vraisemblance composite résident dans leur flexibilité et leur temps de calcul modéré. Cependant, les paires ou triplets sont considérés comme indépendants, ce qui conduit à une détérioration spatiale. Le modèle est alors mal-spécifié, nous avons $\mathbf{\Omega_d} \neq \mathbf{\Omega}$ et la borne d'efficacité de Fréchet-Darmois-Cramer-Rao n'est donc pas atteinte[7]. Ainsi, en pratique, les méthodes de type composite nécessitent un grand nombre d'observations temporelles, notamment quand la corrélation spatiale est élevée.

[6]Notons toutefois que Bienvenüe et Robert (2014) parviennent à la caractériser dans certains cas.

[7]à part dans dans le cas exotique d'un processus spatial complètement indépendant.

La méthode développée dans ce chapitre

Nous proposons une méthode fournissant des estimateurs efficaces. Soient $\mathbf{z}_1, \ldots, \mathbf{z}_T$ des observations temporelles d'un processus max-stable paramétrique (par exemple le modèle de Smith, Schlather ou Brown-Resnick) en M sites. Les paramètres à estimer sont contenus dans la matrice $\mathbf{\Sigma_0}$ dont les dimensions dépendent du processus max-stable considéré. Ainsi, $\forall t$, $\mathbf{z}_t$ est un vecteur de dimension M : $(z_t^1, \ldots, z_t^M)$. La log-vraisemblance associée s'écrit

$$L_T(\mathbf{\Sigma}) = \frac{1}{T} \sum_{t=1}^{T} \log l(\mathbf{z}_t, \mathbf{\Sigma}),$$

où $l(\mathbf{z}_t, \mathbf{\Sigma})$ est la densité de $\mathbf{z}_t$.

Etant donné que la densité multivariée ne possède généralement pas de forme fermée, l'estimateur du maximum de vraisemblance ne peut pas être déterminé. Néanmoins, la plupart des processus max-stables peuvent être simulés et donc notre idée est d'approcher la densité multivariée inconnue à l'aide d'un estimateur à noyau multivarié (Silverman, 1986; Scott, 1992). Soient S simulations i.i.d. du processus considéré, $\mathbf{z}_s^S(\mathbf{\Sigma}), s = 1, \ldots, S$, la densité simulée $l^S(.,.)$ s'écrit

$$l^S(\mathbf{z}_t, \mathbf{\Sigma}) = \frac{1}{S} \sum_{s=1}^{S} K_{\mathbf{H}}(\mathbf{z}_t - \mathbf{z}_s^S(\mathbf{\Sigma})),$$

où $K_{\mathbf{H}}$ est un noyau associé à la matrice des fenêtres $\mathbf{H}$ de dimension $M \times M$. Afin de simplifier la présentation, nous considérons dans ce chapitre que $\mathbf{H} = h\ \mathbf{I}$ où h est une fenêtre telle que $h \to 0$ quand $S \to \infty$ et $\mathbf{I}$ est la matrice identité.

La log-vraisemblance simulée est

$$\tilde{L}_T^S(\mathbf{\Sigma}) = \frac{1}{T} \sum_{t=1}^{T} \log l^S(\mathbf{z}_t, \mathbf{\Sigma}).$$

Cependant, le logarithme ayant une dérivée infinie en 0, des erreurs d'estimation sur de faibles valeurs de la densité sont fortement amplifiées par le passage au logarithme. Ainsi, les plus petites valeurs de $l^S(\mathbf{Z}_t, \mathbf{\Sigma})$ doivent être éliminées. La log-vraisemblance simulée non paramétrique est donc définie, comme dans Fermanian et Salanié (2004), par

$$L_T^S(\mathbf{\Sigma}) = \sum_{t=1}^{T} \tau_S[l^S(\mathbf{z}_t, \mathbf{\Sigma})] \log l^S(\mathbf{z}_t, \mathbf{\Sigma}), \tag{1.30}$$

où la fonction $\tau_S(.)$ vérifie pour $\delta \geq 0$

$$\tau_S(u) = \begin{cases} 0 & \text{si } |u| < h^\delta; \\ \text{suffisamment régulière} & \text{si } |u| \in [h^\delta, 2h^\delta]; \\ 1 & \text{si } |u| > 2h^\delta. \end{cases}$$

L'estimateur du maximum de vraisemblance non paramétrique est alors défini par

$$\hat{\mathbf{\Sigma}}_{\mathbf{T}}^{\mathbf{S}} = \arg\max_{\mathbf{\Sigma}} L_T^S(\mathbf{\Sigma}). \tag{1.31}$$

Cette méthode est assez générale puisqu'elle peut être appliquée à n'importe quel processus max-stable dès lors que l'on sait le simuler. Néanmoins, dans ce chapitre, nous nous concentrons sur le processus de Smith, défini par (1.25) et (1.27).

Dans ce cadre, les deux principaux résultats de ce papier sont donnés dans les théorèmes suivants. Pour des raisons de parcimonie, nous ne détaillons pas ici les différentes hypothèses techniques ainsi que les idées de la démonstration. Nous renvoyons le lecteur au coeur du Chapitre 3.

Théorème 1.5. *Sous certaines hypothèses de régularité,* $\hat{\mathbf{\Sigma}}_{\mathbf{T}}^{\mathbf{S}}$ *est fortement convergent. Presque sûrement,*

$$\lim_{S,T\to\infty} \hat{\mathbf{\Sigma}}_{\mathbf{T}}^{\mathbf{S}} = \mathbf{\Sigma_0}.$$

Théorème 1.6. *Sous certaines hypothèses de régularité,* $\hat{\mathbf{\Sigma}}_{\mathbf{T}}^{\mathbf{S}}$ *est asymptotiquement normal et asymptotiquement efficace :*

$$\sqrt{T}(\hat{\mathbf{\Sigma}}_{\mathbf{T}}^{\mathbf{S}} - \mathbf{\Sigma_0}) \xrightarrow[S,T\to\infty]{d} N(0, \mathbf{\Omega}),$$

où $\mathbf{\Omega}$ *est la matrice de covariance asymptotique de l'estimateur du maximum de vraisemblance exact.*

Comme détaillé dans le chapitre, le choix de la matrice de fenêtre $\mathbf{H}$ a constitué une difficulté pratique très importante dans cette étude. Après avoir testé de nombreuses méthodes, nous préconisons l'utilisation d'une fenêtre variable obtenue par la méthode des k plus proches voisins (Loftsgaarden et Quesenberry, 1965).

Nous évaluons notre méthode sur une étude par simulations, en comparant notre estimateur avec l'estimateur composite par paires. Nous montrons un gain relativement significatif, notamment lorsqu'un faible nombre d'observations temporelles est disponible. En particulier, notre estimateur tire mieux parti d'un grand nombre de sites d'observation, étant donné qu'il s'affranchit de l'hypothèse d'indépendance spatiale partielle intrinsèque aux méthodes fondées sur la vraisemblance composite.

1.4 Résultats du Chapitre 4

Ce chapitre propose d'étudier les mesures de risque dans un cadre spatial. Avant de décrire notre approche ainsi que les résultats principaux, nous proposons une brève revue de la littérature générale sur les mesures de risque.

1.4.1 Mesures de risque classiques et premières extensions

Soit $(\Omega, \mathcal{F}, P_r)$ un espace de probabilité. Nous considérons des variables aléatoires réelles L définies sur l'espace mesurable $(\Omega, \mathcal{F})$.

Définition 1.3. *Une mesure de risque est une fonction* Π *qui associe à toute variable aléatoire (risque)* L *un nombre positif,* $\Pi(L)$*, représentant une somme d'argent pour que le risque* L *soit acceptable.*

$\Pi(L)$ constitue un résumé de la distribution de L et son but est de quantifier la dangerosité du risque L. Les mesures de risque sont très utiles :

- pour la tarification des contrats d'assurance. La valeur $\Pi(L)$ fournit une réponse à la question : quel montant l'assureur doit-il demander pour accepter de prendre le risque L ? Ainsi, $\Pi(L)$ peut par exemple correspondre à la prime d'assurance.
- pour le calcul du montant de fonds propres à détenir. $\Pi(L)$ peut être le capital à détenir pour que le risque associé au portefeuille global de l'assureur, L, soit acceptable pour un contrôleur interne ou externe.

Une large littérature se consacre aux mesures de risque. Le plus souvent, on exige un certain nombre de propriétés de ces mesures de risque. Les axiomes les plus courants sont les suivants : pas de chargement inutile, chargement positif, pas de chargement injustifié, objectivité, translation, sous-additivité, additivité pour les risques comonotones, homogénéité positive, monotonie, invariance de type 1, invariance de type 2, convexité, convergence en loi (voir Denuit *et al.* (2006)).

Dans ce chapitre, nous ne considérons que quelques-uns de ces axiomes. Depuis l'article fondateur de Artzner *et al.* (1999), la notion de mesure de risque cohérente est couramment utilisée. Celle-ci est définie de la manière suivante.

Définition 1.4. *Une mesure de risque est dite cohérente si elle satisfait les axiomes :*

1. *Translation :* $\forall c \in \mathbb{R}, \Pi(L + c) = \Pi(L) + c$ *;*
2. *Sous-additivité : Pour tous risques* L_1 *et* L_2*, alors* $\Pi(L_1 + L_2) \leq \Pi(L_1) + \Pi(L_2)$ *;*
3. *Homogénéité positive :* $\forall \lambda > 0, \Pi(\lambda L) = \lambda \Pi(L)$ *;*
4. *Monotonie : Pour tous risques* L_1 *et* L_2*,* $P(L_1 \leq L_2) = 1 \Longrightarrow \Pi(L_1) \leq \Pi(L_2)$.

Il apparaît assez naturel de demander ce type de propriétés à des mesures de risque. Typiquement, l'axiome de translation signifie que les frais fixes associés à la souscription du risque L doivent être ajoutés à la valeur de la prime ou encore que l'augmentation du capital à détenir provenant de l'ajout d'un risque déterministe est égale à la valeur de ce risque. La sous-additivité signifie que la prime nécessaire pour assurer la somme de deux risques est inférieure à la somme des primes individuelles, ce qui est assez naturel

si l'on suppose qu'il y a diversification. L'homogénéité positive assure une cohérence par changement de devise. Enfin, la monotonie stipule que si la perte potentielle L_1 est inférieure à la perte L_2 dans tous les états du monde, alors la prime correspondant à L_1 doit être inférieure à celle correspondant à L_2.

La notion de mesure de risque cohérente a ensuite été développée par Delbaen (2000) et étendue au cas des mesures de risque convexes par Frittelli et Rosazza Gianin (2002) et Föllmer *et al.* (2004). L'axiome de convexité signifie que pour tous risques L_1 et L_2 et toute constante $\alpha \in [0;1]$,

$$\Pi(\alpha L_1 + (1-\alpha)L_2) \leq \alpha\Pi(L_1) + (1-\alpha)\Pi(L_2). \tag{1.32}$$

Les mesures de risque convexes sont définies ci-après.

Définition 1.5. *Une mesure de risque est dite convexe si elle satisfait les axiomes : 1. Translation ; 2. Convexité ; 3. Monotonie.*

La classe des mesures de risque convexes englobe celle des mesures cohérentes, étant donné que l'homogénéité positive et la sous-additivité impliquent la convexité.

Tous les papiers précédemment cités concernent des mesures de risque statiques : le but est de quantifier, à un moment donné, le risque associé à une position future incertaine. Une extension naturelle s'intéresse au cadre conditionnel, introduit par Bion-Nadal (2004) et Detlefsen et Scandolo (2005). L'idée est de prendre en compte l'information disponible au moment de l'évaluation du risque. Dans un cadre dynamique, la mesure du risque peut ainsi être actualisée en utilisant la nouvelle information disponible, ce qui conduit à la notion de mesure de risque dynamique. Afin de formaliser ce concept, considérons un cadre discret $t = 1, \ldots, T$ et un espace de probabilité filtré $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Pour tout t, notons $L_t^\infty = L^\infty(\Omega, \mathcal{F}_t, P)$ l'espace des variables aléatoires $\mathcal{F}_t$-mesurables essentiellement bornées et $L^\infty = L^\infty(\Omega, \mathcal{F}_T, P)$.

Définition 1.6. *Une mesure de risque conditionnelle Π_t attribue à tout risque L une variable aléatoire $\mathcal{F}_t$-mesurable qui quantifie le risque de la position L conditionnellement à l'information $\mathcal{F}_t$.*

Une définition rigoureuse des mesures de risque conditionnelles convexes a été donnée par Detlefsen et Scandolo (2005) :

Définition 1.7. *Une fonction $\Pi_t : L^\infty \mapsto L_t^\infty$ est une mesure de risque conditionnelle convexe si elle satisfait, pour tous L_1, $L_2 \in L_t^\infty$:*

1. *Invariance par translation conditionnelle : $\forall c_t \in L_t^\infty$, $\Pi_t(L + c_t) = \Pi_t(L) + c_t$;*
2. *Monotonie : $L_1 \leq L_2 \Rightarrow \Pi_t(L_1) \leq \Pi_t(L_2)$;*
3. *Convexité conditionnelle :*
   $\forall \lambda \in L_t^\infty, 0 \leq \lambda \leq 1, \Pi_t(\lambda L_1 + (1-\lambda)L_2) \leq \lambda\Pi_t(L_1) + (1-\lambda)\Pi_t(L_2)$;
4. *Normalisation : $\Pi_t(0) = 0$.*

Elle est dite conditionnelle cohérente si elle satisfait en plus la propriété d'homogénéité positive conditionnelle : $\forall \lambda \in L_t^\infty, \lambda \geq 0, \Pi_t(\lambda L) = \lambda\Pi_t(L)$.

Nous pouvons désormais formellement introduire la notion de mesure de risque dynamique.

Définition 1.8. *Une suite* $(\Pi_t)_{t=1...T}$ *est une mesure de risque dynamique convexe (respectivement cohérente) si* Π_t *est une mesure de risque conditionnelle convexe (respectivement cohérente) pour tout* $t = 1, \ldots, T$.

La littérature concernant les mesures de risque dynamiques cohérentes est assez étendue : voir par exemple Riedel (2004), Delbaen (2006) et Artzner *et al.* (2007). Les mesures de risque dynamiques convexes sont également considérées dans de nombreux papiers, par exemple Frittelli et Rosazza Gianin (2004), Cheridito *et al.* (2006), Bion-Nadal (2006), Föllmer et Penner (2006) et Klöppel et Schweizer (2007). Le cas particulier des g-espérances ou des équations différentielles rétrogrades stochastiques a notamment été étudié par Peng (2004), Rosazza Gianin (2006) et Barrieu et Karoui (2007).

Une question cruciale dans le cadre dynamique concerne la relation entre les évaluations du risque à différents instants. Notamment, les jugements fondés sur la mesure de risque ne doivent pas se contredire dans le temps. Ainsi, différentes notions de consistance temporelle ont été introduites. Nous remarquons que cette notion de consistance s'avère très délicate à adapter dans le contexte spatial.

Au lieu de faire varier l'instant d'évaluation du risque comme dans le cadre dynamique, une autre approche liée à la dimension temporelle consiste à faire varier l'horizon : il s'agit de l'étude de la structure par termes des mesures de risque. Soit $L_{t,t+h}$ la perte d'une institution financière dans l'intervalle $[t, t+h]$. La recherche d'une propriété d'homogénéité par rapport à l'horizon temporel h revient à comparer $\Pi(L_{t,t+\lambda h})$ et $\Pi(L_{t,t+h})$, pour tout $\lambda > 0$. La littérature fournit quelques résultats, en particulier dans le cas de la Value at Risk (VaR) et de l'Expected Shortfall (ES). Par exemple, si les variations de prix sont i.i.d. et Gaussiennes, nous avons $\mathrm{VaR}_{t,t+\lambda}(\alpha) = \sqrt{\lambda}\ \mathrm{VaR}_{t,t+1}(\alpha)$, où $\mathrm{VaR}_{t,t+h}(\alpha)$ est la VaR de $L_{t,t+h}$, au niveau $\alpha > 0$. A part dans des cas simples, aucune formule analytique ne relie en général $\Pi(L_{t,t+\lambda h})$ et $\Pi(L_{t,t+h})$. En recourant à des méthodes numériques, Guidolin et Timmermann (2006) effectuent une comparaison de la structure par termes de la VaR et de l'ES sous différents modèles économétriques.

L'approche développée dans ce chapitre s'apparente à une espèce d'équivalent spatial de la structure par termes des mesures de risque. En notre connaissance, seuls Föllmer (2014) et Föllmer et Klüppelberg (2014) invoquent le terme de mesure de risque spatial. Néanmoins, leur approche est de type "réseaux" et non géostatistique. Elle peut être adaptée aux interactions entre institutions financières et est donc en lien avec les outils que l'on développe dans le Chapitre 5.

1.4.2 L'approche spatiale développée dans ce chapitre

Dans le cas d'une compagnie d'assurance, la perte totale liée au portefeuille (notée L) associé à un aléa[8] donné (par exemple le vent) dépend évidemment de la localisation géographique des contrats ainsi que de l'aléa en question. Dans ce chapitre, nous nous proposons de clairement faire apparaître ces deux composantes dans la définition de la mesure de risque. Le but est ensuite d'étudier la sensibilité de la mesure par rapport à la composante spatiale.

Considérons l'ensemble $\mathcal{A} \subset \mathbb{R}^2$ de tous les ensembles mesurables de $\mathbb{R}^2$ dont la mesure de Lebesgue est strictement positive : $\mathcal{A} = \{A : A \subset \mathbb{R}^2 \text{ and } |A| > 0\}$, où $|.|$ désigne la mesure de Lebesgue dans $\mathbb{R}^2$. Soit $\mathcal{P}$ une famille de champs aléatoires sur $\mathcal{A}$ représentant le coût économique dû à un phénomène environnemental. Comme nous allons le voir, les

[8]On trouve parfois le terme de péril, totalement équivalent.

membres de cette famille résultent de la composition d'une fonction de dommage avec un processus stochastique représentant le phénomène environnemental. Soit Π une mesure de risque classique au sens de la Définition 1.3.

Le moyen le plus simple et intuitif de construire une telle mesure serait d'intégrer une mesure de risque usuelle (par exemple la VaR), définie sur une région $A \in \mathbb{R}^2$. Notons $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ le processus du coût économique dû à un aléa environnemental particulier (par exemple le vent). Si le processus $C_P(\mathbf{x})$ (de loi notée P) est strictement stationnaire, sa distribution est indépendante de $\mathbf{x}$ et donc pour toute mesure de risque $\Pi(.)$ définie dans la Définition 1.3,

$$R_\Pi(A,P) = \frac{1}{|A|}\int_A \Pi[C_P(\mathbf{x})]\ d\mathbf{x} = \Pi[C(\mathbf{0})],$$

ce qui signifie que la mesure de risque spatiale $R_\Pi(A,P)$ se réduit à la mesure de risque classique associée à un seul site. Cette démarche ne permet donc pas de prendre en compte la dépendance spatiale et s'avère donc plutôt inintéressante du point de vue pratique.

Nous choisissons donc une alternative consistant à étudier les mesures de risque classiques appliquées à la perte spatialement agrégée normalisée sur la région A.

Définition 1.9. *Pour $A \in \mathcal{A}$ et $P \in \mathcal{P}$, la perte spatialement agrégée normalisée est définie par*

$$L_N(A,P) = \frac{\int_A C_P(\mathbf{x})\ d\mathbf{x}}{|A|}, \tag{1.33}$$

où le processus stochastique $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ a la distribution P.

Le risque $L_N(A,P)$ défini par (4.2) prend totalement en compte la structure de dépendance du processus $C(\mathbf{x})$. La mesure de risque spatiale introduite dans ce chapitre est définie de la manière suivante :

Définition 1.10. *Une mesure de risque spatiale est une fonction R_Π qui associe à tout ensemble $A \in \mathcal{A}$ et tout type de champs aléatoire $P \subset \mathcal{P}$ un nombre réel :*

$$\begin{array}{rccc} R_\Pi & : & \mathcal{A}\times\mathcal{P} & \to & \mathbb{R} \\ & & (A,P) & \mapsto & R_\Pi(A,P) = \Pi[L_N(A,P)], \end{array}$$

où $L_N(A,P)$ a est définie par (4.2).

Si la distribution P est fixée, alors R_Π résume la dangerosité associée à une certaine région. Nous comprenons que les propriétés spatiales (i.e. par rapport à l'espace) de $R_\Pi(A,P)$ dépendent à la fois de la mesure de risque classique choisie, Π, et de l'aléa décrit par P.

Nous proposons l'approche axiomatique suivante dans le cas de la perte normalisée.

Définition 1.11. *Dans les propositions suivantes, P est quelconque et appartient à $\mathcal{P}$. Π est une mesure de risque classique.*

1. ***Invariance par translation spatiale :*** *Pour un vecteur $\mathbf{c} \in \mathbb{R}^2$, si l'on note $A+\mathbf{c}$ la région translatée de $\mathbf{c}$, $R_\Pi(A+\mathbf{c},P) = R_\Pi(A,P)$*
2. ***Sous-additivité spatiale :*** $\forall\ A_1, A_2 \in \mathcal{A},\ R_\Pi(A_1 \cup A_2, P) \leq \min[R_\Pi(A_1,P), R_\Pi(A_2,P)]$

3. ***Homogénéité spatiale asymptotique d'ordre*** $-\boldsymbol{\alpha}, \boldsymbol{\alpha} \in \mathbb{R}$ ***:*** $\forall A \in \mathcal{A}$,

$$R_\Pi(\lambda A, P) \underset{\lambda \to \infty}{\sim} \left(K_1 + \frac{K_2}{\lambda^\alpha}\right) R_\Pi(A, P), \tag{1.34}$$

où λA représente la région obtenue en appliquant une homothétie de rapport λ à A par rapport à son centre, $K_1 \in \mathbb{R}$ et $K_2 > 0$.

4. ***Anti-monotonicité spatiale*** *:*
$\forall A_1, A_2 \in \mathcal{A},\ A_1 \subset A_2 \Rightarrow R_\Pi(A_2, P) \leq R_\Pi(A_1, P)$.

Nous montrons que les axiomes 2 et 4 sont parfaitement équivalents.

Le point crucial à comprendre est que ces propriétés caractérisent la mesure de risque non pas par rapport à l'aléa sous-jacent (de loi P) mais par rapport à l'espace. Par exemple, dans le cas de la sous-additivité spatiale, notre approche compare la mesure du risque associé à l'union ensembliste de deux régions spatiales avec la mesure du risque de chaque région prise séparément. Ainsi, alors que l'approche classique comparerait $\Pi\left[L_N(A, P_1) + L_N(A, P_2)\right]$ avec $\Pi\left[L_N(A, P_1)\right] + \Pi\left[L_N(A, P_2)\right]$, notre approche compare $\Pi\left[L_N(A_1 \cup A_2, P)\right]$ avec $\Pi\left[L_N(A_1, P)\right]$ et $\Pi\left[L_N(A_2, P)\right]$. Dans le premier cas, la région est fixe mais deux sources d'aléa sont considérées, via P_1 et P_2. Dans le second, deux régions sont considérées mais seule une source d'aléa est présente.

L'axiome d'invariance par translation apparaît naturel dans le cas d'un processus de dommage économique stationnaire. Les axiomes de sous-additivité, homogénéité spatiale et anti-monotonicité spatiale traduisent le fait qu'il y a diversification spatiale : une compagnie d'assurance a intérêt à souscrire des contrats sur une région plus étendue ou dans différentes régions.

Dans la suite du chapitre, nous construisons et étudions deux exemples de mesures de risque spatiales. Afin d'élaborer une telle mesure, il convient de se donner un modèle pour le coût économique $C_P(\mathbf{x})$, sachant que nous sommes intéressés par les risques extrêmes. Afin de quantifier C_P, la première étape consiste donc à modéliser le comportement des maxima des variables environnementales d'intérêt. Compte-tenu de leur étendue spatiale naturelle, les processus max-stables (voir Section 1.3.3) s'avèrent très adaptés. Comme mentionné précédemment, ils permettent une agrégation spatiale et prennent totalement en compte la dépendance spatiale. L'agrégation apparaît évidemment nécessaire en vue de définir $L_N(A, P)$ selon (4.2).

La seconde étape concerne le modèle convertissant la variable environnementale en un coût de dommage économique pour les autorités ou les assureurs. Ce modèle fait intervenir une fonction de dommage ainsi que l'exposition spécifique à chaque localité, cette dernière dépendant des conditions démographiques (densité d'habitants), économiques (structure des bâtiments, coût de construction, richesse par habitant) et orographiques (une localité risque d'être davantage affectée si elle est située au sommet d'une colline). Afin de simplifier les calculs, nous considérons une exposition uniformément égale à 1. Ainsi, le modèle de coût économique se réduit à

$$C_P(\mathbf{x}) = D\left[Z(\mathbf{x})\right], \tag{1.35}$$

où Z représente l'aléa environnemental et $D(.)$ est la fonction de dommage. Du fait de la complexité des processus max-stables, il est en général délicat d'effectuer des calculs analytiques. Ainsi, dans les exemples développés, nous nous concentrons sur des mesures de

risque classiques assez simples, telles l'espérance, la variance et la VaR. L'espérance n'est pas très intéressante car elle ne fournit aucune information sur la dépendance spatiale.

Dans la suite, par souci de simplicité, nous omettons la dépendance en P dans les notations : $R_\Pi(A, P)$ et $L_N(A, P)$ seront donc respectivement notés $R_\Pi(A)$ et $L_N(A)$. La distribution P sera toujours clairement explicité. Par ailleurs, compte-tenu de la complexité des expressions obtenues, nous nous focalisons principalement, dans cette introduction, sur des descriptions qualitatives et des interprétations des résultats. Nous renvoyons le lecteur au coeur du Chapitre 4 pour les détails techniques.

1.4.3 Un premier exemple fondé sur les dépassements de seuil

Le premier exemple traité se fonde sur la fonction de dommage

$$D\left[Z(\mathbf{x})\right] = \mathbf{I}_{Z(\mathbf{x})>u}, \tag{1.36}$$

où $u > 0$. La perte totale est donc définie par $L_N(A) = \dfrac{1}{|A|}\displaystyle\int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\, d\mathbf{x}$.

Cette fonction de dommage apparaît adaptée aux conséquences humaines des vagues de chaleur. u peut en effet être défini comme un seuil de dangerosité au delà duquel les populations sont en péril.

Nous considérons tout d'abord $R_1(A) = \mathbb{E}\left[L_N(A)\right]$, qui n'est autre que la prime actuarielle, et montrons que celle-ci ne dépend pas de la région considérée. Ce résultat provient de la stationnarité du processus et s'apparente au cas d'un portefeuille d'assurance constitué de risques homogènes.

La variance

Nous étudions ensuite la variance de la perte agrégée normalisée, $R_2(A) = \mathrm{Var}\left[L_N(A)\right]$. Contrairement à l'espérance, cette mesure de risque permet la prise en compte de la dépendance spatiale dans l'évaluation du risque. Elle apparaît très utile, que ce soit pour la gestion du risque spatial ou une meilleure compréhension des propriétés des processus max-stables. Nous explicitons notamment la fonction $\lambda \mapsto R_2(\lambda A)$, et nous nous intéressons plus particulièrement au cas des processus max-stables isotropes pour des régions A étant des disques ou des carrés. $R_2(\lambda A)$ est fonction décroissante du coefficient extrêmal $\Theta(\lambda h)$. Pour tout h positif, $\Theta(\lambda h)$ est fonction croissante de λ et donc $R_2(\lambda A)$ est fonction décroissante de λ, ce qui signifie qu'il y a diversification spatiale. L'indépendance asymptotique, correspondant à $\lim_{h\to\infty}\Theta(h) = 2$, s'avère particulièrement intéressante. Dans ce cas, $\lim_{\lambda\to\infty} R_2(\lambda A) = 0$ et il y a donc diversification spatiale totale. Ce résultat est étroitement lié aux propriétés de mélange et d'ergodicité des processus max-stables, décrites dans la Section 1.3.3.

Néanmoins, les expressions que nous obtenons pour $R_2(\lambda A)$ font intervenir une intégrale ne possédant pas de forme fermée et nous devons recourir à une analyse numérique. Nous étudions l'évolution de $R_2(\lambda A)$ en fonction de λ pour les processus de Smith, Schlather et le processus Gaussien géométrique, introduits dans la Section 1.3.3. Dans les deux derniers cas, nous comparons différents types de fonctions de corrélation $\rho(.)$ (voir Section 1.3.3). En fonction du modèle considéré et de la fonction de corrélation, la diversification spatiale est plus ou moins rapide (plus rapide dans le cas du processus de Smith) et peut être totale (Smith) ou partielle (Schlather et processus Gaussien géométrique). Il est intéressant de noter que le processus de Brown-Resnick (dont les processus de Smith

et Gaussiens géométriques sont des cas particuliers) permet de modéliser différents types de diversification spatiale. Nous introduisons également un nouveau modèle de processus max-stable, le processus à "tube", et étudions sa densité bivariée ainsi que son coefficient extrêmal. Ce processus permet une diversification totale et plus rapide que les processus précédemment décrits. Notamment, lorsque le rayon des tubes tend vers 0, il tend vers un processus totalement indépendant.

Il convient de noter que la "vitesse" de diversification (i.e. la taille de région minimale permettant d'atteindre un certain niveau de variance) constitue une information importante pour une compagnie d'assurance ou de réassurance.

En utilisant un résultat sur les "excursion sets" (Spodarev, 2014, Theorem 4), nous montrons un théorème central limite pour la perte normalisée.

Théorème 1.7. *Dans le cas du processus de Smith, du processus de Brown-Resnick dont le variogramme vérifie* $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ *avec* $a > 0$ *et du processus "tube", nous avons*

$$\lambda\left(L_N(\lambda A) - \left[1 - \exp\left(-\frac{1}{u}\right)\right]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \text{ pour } \lambda \to \infty,$$

où

$$\sigma^2 = \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}.$$

Nous montrons que la mesure de risque $R_2(A)$ vérifie l'axiome de translation spatiale, de sous-additivité spatiale et d'anti-monotonicité spatiale, pour tout processus max-stable ayant des marges de Fréchet unitaires. Par ailleurs, elle vérifie la propriété d'homogénéité spatiale d'ordre -2 dans le cas du processus de Smith, du processus "tube" et d'un processus de Brown-Resnick dont le variogramme vérifie $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ avec $a > 0$.

La Value at Risk

Il apparaît impossible d'obtenir une expression analytique pour la VaR de $L_N(A)$, et ce même à une intégrale près. Ainsi, notre étude utilise des techniques numériques. Une réalisation de $L_N(A)$ peut être approchée par des techniques de discrétisation de Riemann. Pour chaque type de schéma numérique, l'erreur peut être quantifiée. En répétant cette opération un grand nombre de fois, une approximation de la distribution de $L_N(A)$ est obtenue et la VaR approchée peut être calculée en prenant le quantile empirique. L'incertitude peut alors être estimée par des techniques de ré-échantillonnage.

Dans le cas du processus de Smith, du processus de Brown-Resnick dont le variogramme vérifie $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ avec $a > 0$ et du processus "tube", le Théorème 1.7 permet de déduire que la quantité $R_{3,1-\alpha}(A) = VaR_{1-\alpha}(L_N(A))$ vérifie l'axiome d'homogénéité spatiale asymptotique d'ordre -1.

1.4.4 Un deuxième exemple fondé sur la fonction puissance

Dans cette partie, nous considérons la fonction de dommage $D[Z(\mathbf{x})] = Z(\mathbf{x})^\beta$ et donc $L_N(A) = \dfrac{1}{|A|}\displaystyle\int_A Z(\mathbf{x})^\beta d\mathbf{x}$. Une telle fonction est adaptée aux dommages causés par le vent (Klawa *et al.*, 2003; Pinto *et al.*, 2007; Donat *et al.*, 2011). Sauf mention contraire,

nous nous focalisons principalement dans cette partie sur le processus de Smith. Néanmoins, l'extension des résultats proposés à un processus de Brown-Resnick plus général est possible.

Nous introduisons tout d'abord une mesure de dépendance adaptée aux dommages causés par le vent, la corrélation Corr $\left(Z(\mathbf{x}_1)^{\beta_1}, Z(\mathbf{x}_2)^{\beta_2}\right)$, où $\mathbf{x}_1$ et $\mathbf{x}_2$ sont deux sites. Nous étudions cette mesure et obtenons une formule à une intégrale près. Nous analysons numériquement l'évolution de la quantité Corr $\left(Z(\mathbf{x}_1)^{\beta}, Z(\mathbf{x}_2)^{\beta}\right)$ en fonction de β, et montrons qu'à part le cas limite $\beta \to 0$, la corrélation des dommages dépend relativement peu du facteur β. En revanche, elle décroît rapidement quand la distance a entre les sites $\mathbf{x}_1$ et $\mathbf{x}_2$ augmente.

Nous nous intéressons ensuite à quelques mesures de risque spatiales fondées sur la perte normalisée $L_N(A)$ et principalement à la variance.

En utilisant l'expression de $\mathbb{E}\left[Z(\mathbf{x}_1)^{\beta}, Z(\mathbf{x}_2)^{\beta}\right]$ obtenue lors de l'étude de la corrélation entre les dommages, nous étudions la mesure de risque $R_2(A) = \mathrm{Var}[L_N(A)]$. Nous obtenons une expression sous forme intégrale de $R_2(\lambda A)$ en fonction de λ. Comme dans le cas de la fonction de dommage à dépassement de seuil, la diversification est totale et très rapide.

En utilisant les propriétés de mélange des processus max-stables (Kabluchko et Schlather, 2010) ainsi que le Théorème 1 de Gorodetskii (1987), nous pouvons montrer le résultat suivant[9].

Théorème 1.8. *Dans le cas du processus de Smith, nous avons*

$$\lambda\left[L_N(\lambda A) - \Gamma(1-\beta)\right] \xrightarrow{d} \mathcal{N}(0, \sigma^2), \text{ pour } \lambda \to \infty,$$

où

$$\sigma^2 = \int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^{\beta} Z(\mathbf{x})^{\beta}] - [\Gamma(1-\beta)]^2\right) d\mathbf{x}.$$

Enfin, dans le cas du processus de Smith, $R_2(A)$ vérifie les axiomes de translation spatiale, de sous-additivité spatiale, d'anti-monotonicité spatiale et d'homogénéité spatiale asymptotique d'ordre -2.

En ce qui concerne la VaR, comme pour la fonction de dommage à seuil, il est très délicat d'obtenir une expression analytique, et ce même à une intégrale près. Ainsi, la même procédure que dans la Section 1.4.3 doit être adoptée. Comme dans la partie précédente, le Théorème 1.8 permet de déduire que la quantité $R_{3,1-\alpha}(A) = VaR_{1-\alpha}(L_N(A))$ vérifie l'axiome d'homogénéité spatiale asymptotique d'ordre -1.

[9] à condition d'admettre l'extension des propriétés de mélange de $\mathbb{R}$ à $\mathbb{R}^d$; voir la discussion de la Section 1.3.3.

1.5 Résultats du Chapitre 5

Le Chapitre 5 est issu d'un article co-écrit avec Jean-Cyprien Héam (ACPR et CREST).

Dans le chapitre précédent, nous avons présenté des outils pouvant être utilisés afin d'analyser le risque à l'échelle d'une institution prise indépendamment des autres (risque à l'échelle micro). Les mesures de risque sont notamment très utiles pour calculer le capital réglementaire à détenir en vue de faire face à un choc exogène de grande ampleur, par exemple une crise économique (engendrant une chute de la croissance) dans le cas d'une banque ou une catastrophe naturelle dans le cas d'une compagnie d'assurance.

Néanmoins, un tel choc peut se propager à l'ensemble du système par des mécanismes de contagion, dont les canaux sont les interconnexions entre institutions, notamment sous forme de participations croisées en dette et en actions. Ainsi, la faillite d'une institution générée par le choc initial peut fragiliser ou mettre en péril d'autres institutions, éventuellement non affectées par ce même choc initial. Un choc exogène isolé peut donc avoir un impact systémique. Si les régulateurs ou praticiens se sont pendant longtemps limités à la quantification du risque spécifique à chaque institution via des mesures de risque comme la VaR ou l'ES (voir Chapitre 4), le risque systémique fait l'objet d'une attention croissante depuis environ deux décennies (Kaufman, 1994).

En vue de quantifier cet impact, il est crucial de savoir précisément comment se mettent en place les mécanismes de contagion. La littérature s'est principalement concentrée sur l'impact de chocs économiques (crises) mais dans le cas des compagnies d'assurance, il est très intéressant d'analyser comment l'impact d'une catastrophe naturelle peut se propager par des mécanismes de contagion. Ceci constitue un travail en cours et sera brièvement présenté en conclusion de cette thèse. Dans ce chapitre, nous nous intéressons principalement au cas des banques. La littérature sur la contagion bancaire est très étendue. Kaufman (1994) décrit ses caractéristiques propres et montre comment des chocs peuvent rapidement se propager à l'ensemble du secteur bancaire et au-delà, engendrant potentiellement des conséquences sur l'économie réelle. Un certain nombre d'études empiriques, répertoriées par Upper (2007), ont simulé la propagation d'un choc dans un réseau interbancaire. Le courant de littérature incarné par Eisenberg et Noe (2001) ainsi que Demange (2012) et Gouriéroux *et al.* (2012) est plutôt méthodologique et fournit les outils quantitatifs nécessaires pour comprendre rigoureusement comment un choc exogène se propage dans un réseau. Ces travaux cherchent à calculer les bilans d'institutions reliées entre elles par des participations croisées en actions et en dette données de manière exogène. A l'équilibre, il existe un unique vecteur de paiements entre institutions respectant l'égalité actif-passif de chaque institution. Suite à un choc exogène, un nouvel équilibre peut être également calculé, ce qui permet de quantifier la contagion.

Les conditions propices à des conséquences systémiques ainsi qu'à la contagion ont été assez largement étudiées. Pour Diamond et Dybvig (1983), les épisodes systémiques peuvent provenir de crises de liquidité. Si les banques ne possèdent pas suffisamment de liquidité, elles se retrouvent, dans le cas de retraits massifs (bank runs), dans l'obligation de liquider des actifs de long terme (donc illiquides) à perte. Cela peut entraîner des faillites. Allen et Gale (2000) proposent une extension de ce modèle au cas d'une économie constituée de banques situées dans des régions hétérogènes. Par ailleurs, les déposants peuvent exiger leurs dépôts plus ou moins rapidement, ce qui peut générer un choc de liquidité pour les banques. L'existence du marché des prêts interbancaires a pour but de créer une réserve (pool) de liquidité permettant de mutualiser ce risque. A chaque période, la liquidité transite des zones où elle est en excès vers celles où elle est

insuffisante, ce qui évite aux banques d'avoir à liquider des actifs illiquides à des prix dépréciés. Néanmoins, un défaut de liquidité dans une certaine région peut s'étendre à d'autres régions par l'intermédiaire de ce marché. Les auteurs montrent que l'ampleur de la contagion dépend de la structure du réseau. Elle est relativement faible dans le cas d'un réseau complet (où toutes les banques sont reliées entre elles) et forte dans le cas d'un réseau incomplet. En effet, un nombre élevé de banques s'avère nécessaire pour absorber le choc de liquidité initial. Elliott *et al.* (2012) ainsi que Acemoglu *et al.* (2013) étendent le modèle de Allen et Gale (2000) et aboutissent à des conclusions similaires sur les structures de réseau favorisant la contagion. Ce courant de littérature fournit des informations qualitatives sur l'ampleur de la contagion en fonction des caractéristiques du réseau. Néanmoins, contrairement au courant incarné par Eisenberg et Noe (2001) et ses successeurs précédemment cités, ils ne modélisent pas les bilans de manière détaillée et s'appuient plutôt sur un équilibre "qualitatif" entre institutions.

Dans la grande majorité des papiers précédemment cités, le réseau est cristallisé. Ainsi, les stress-tests (analyse des conséquences de la détérioration brutale de certains indicateurs économiques) sont le plus souvent réalisés à réseaux fixés. Il en va de même pour l'étude des répercussions de la faillite de l'une des banques du réseau (voir par exemple Gouriéroux *et al.* (2012)). Or, en réalité, sous des conditions économiques radicalement différentes ou en cas de faillite d'une institution, il est probable que le réseau soit modifié. Néanmoins, l'hypothèse de cristallisation est en général justifiée par le fait que les effets des chocs sont analysés à court terme et donc que le réseau n'a pas le temps de s'adapter. Notons que l'hypothèse de cristallisation est également le plus souvent faite dans les études de sensibilité à des évolutions réglementaires. Les régulateurs étudient les variations du "bien-être" économique (en un sens défini ci-après) en fonction des principaux paramètres réglementaires, en faisant comme si le réseau ne s'adaptait pas à des modifications de ceux-ci. Une telle hypothèse est plus difficilement acceptable dans ce type d'étude étant donné que c'est justement l'impact sur le long terme que l'on cherche à analyser.

On comprend donc qu'il est nécessaire, au moins dans certains cas, de relâcher cette hypothèse de cristallisation du réseau. Ainsi, dans ce chapitre, nous proposons un modèle dans lequel tous les éléments du bilan des institutions, et donc notamment les interconnexions, sont endogènes. Les paramètres exogènes représentent les conditions économiques ainsi que les grandeurs réglementaires. Elaborer un modèle endogène nécessite d'expliquer pourquoi les institutions financières ont tendance à s'interconnecter. Ce phénomène d'interconnexions semble de plus en plus notable, que ce soit entre banques, compagnies d'assurance et également entre banques et compagnies d'assurance[10]. Sachant que les institutions financières sont en concurrence (notamment les banques en vue d'attirer les dépôts des clients), ceci peut paraître paradoxal.

L'une des explications réside dans le fait que les interconnexions permettent de partager une réserve de cash servant à mutualiser le risque de liquidité (Holmstrom et Tirole, 1996; Tirole, 2010; Rochet, 2004). Comme précédemment décrit, pour Allen et Gale (2000), le marché des prêts interbancaires de court terme permet d'absorber certains chocs de liquidité. Le risque de liquidité est particulièrement important dans le secteur bancaire car les banques sont exposées au phénomène de runs (Diamond et Dybvig, 1983) et manipulent des montants très importants (Rochet, 2004). Néanmoins, le problème de liquidité n'est pas spécifique au secteur bancaire, étant donné que toute entreprise (indus-

[10]Avec l'émergence des bancassureurs, la distinction entre banques et compagnies d'assurance est d'ailleurs de plus en plus floue.

trielle ou autre) fait face à des flux de cash entrants et sortants asynchrones. Par ailleurs, de nombreuses études empiriques montrent que les banques ne sont pas seulement interconnectées sur le court terme mais également sur le long terme. Par exemple, Upper et Worms (2004) constatent que la moitié des prêts interbancaires allemands possède une maturité supérieure à quatre ans. La liquidité est un phénomène de court terme et ne peut donc pas expliquer ces interconnexions de long terme. Une deuxième raison possible est l'intégration horizontale : plusieurs institutions proposent un produit joint et se partagent donc un ensemble de clients. Afin de rassurer leurs partenaires respectifs sur leur bonne foi, ils créent des liens sous forme de participations croisées. Il arrive également que deux groupes importants créent une filiale commune sur des sujets spécialisés, afin à terme de se séparer de certains services de manière jointe. D'autres motivations possibles d'interconnexion sont l'intégration verticale (par exemple le transfert de risque d'un assureur à un réassureur) ou l'égo de managers souhaitant maximiser leur contrôle. Enfin, la diversification apparaît comme un motif plausible qui mérite d'être testé. Dans la réalité, les différentes causes possibles jouent probablement chacune un rôle mais nous nous proposons ici de ne considérer que la diversification. Cette dernière comprend deux aspects. Banques et compagnies d'assurance peuvent appartenir à différents types de "business models", issus de processus de spécialisation. Ainsi cette diversité d'institutions entraîne une grande variété d'actions et obligations disponibles sur le marché. S'interconnecter permet donc indirectement de profiter des "business models" et des clients spécifiques des autres institutions, et donc éventuellement de profiter indirectement de leurs investissements à haut rendement. Le deuxième aspect concerne la réduction de la variance et donc du risque, conformément à la théorie du portefeuille classique (Markowitz, 1952).

Si les réseaux endogènes ont intensément été étudiés en sociologie (pour une revue de littérature, voir Goyal (2012) ou Jackson (2010)), la finance représente un nouveau champ d'application. Comme nous l'avons déjà mentionné, généralement, les interconnexions entre institutions sont considérées comme données de manière exogène. Ceci est notamment le cas dans les papiers appliqués comme Cifuentes *et al.* (2005), Arinaminpathy *et al.* (2012) ou Anand *et al.* (2012). Les réseaux endogènes viennent du papier fondateur d'Allen et Gale (2000). Des modèles plus récents ont été proposés par Elliott *et al.* (2012) ou encore Acemoglu *et al.* (2013). Ils s'inspirent de la microéconomie ainsi que de la théorie des jeux. Néanmoins, les institutions financières décrites par ce courant de littérature calculent uniquement les interconnexions : les autres éléments de leur bilan sont totalement exogènes. Cette hypothèse est acceptable dans le cas d'interconnexions de court terme mais pas de long terme. Ainsi, nous nous démarquons de cette tendance par le fait que nous considérons un bilan totalement endogène. A notre connaissance, seuls Bluhm *et al.* (2013) proposent une optimisation du bilan complet (à part la dette). Enfin, il convient de souligner le fait que la plupart des articles mentionnés ne considèrent que des interconnexions sous forme de dette. Inspirés par Gouriéroux *et al.* (2012), nous introduisons également des participations en actions.

1.5.1 Principaux éléments du bilan et benchmark

Nous considérons un ensemble $\{i = 1, \ldots, n\}$ d'institutions financières. L'institution i a accès à un actif illiquide externe (extérieur au réseau) spécifique, noté Ax_i et à un actif liquide très peu risqué, $A\ell_i$, assimilable à du cash. Ax_i s'apparente à un prêt illiquide et ne peut être échangé sur le marché. Le rendement net de Ax_i et sa réalisation sont respectivement notés R_i et r_i. Celui de $A\ell_i$, r_{rf}. La fonction de répartition des $R_i, i =$

$1, \ldots, n$, et leur densité sont respectivement notées F_R et f_R. L'institution i détient par ailleurs des actions et de la dette de l'institution j $(j = 1, \ldots, n)$ respectivement en proportions π_{ij} et γ_{ij}.

Le passif est constitué des capitaux propres (apportés par les investisseurs) ainsi que de la dette, dont les valeurs comptables sont respectivement notées K_i et L_i. La dette nominale est notée L_i^*. Les actions et obligations étant échangées sur le marché secondaire, nous introduisons les valeurs de marché des capitaux propres et de la dette, $\mathcal{K}_i$ et $\mathcal{L}_i$. Toutes les institutions émettent de la dette selon la même courbe de taux notée $r_D(.)$. En revanche, chaque institution possède un paramètre de transformation de maturité qui lui est propre, noté $\omega_i \in [0, 1]$. Le cas $\omega_i = 0$ correspond à une absence de transformation de maturité (la maturité des moyens de financement au passif est égale à celle des prêts à l'actif) alors que $\omega_i = 1$ correspond à une transformation totale (cas par exemple quand la dette est uniquement constituée de dépôts). Le coût de financement est fonction décroissante de ω_i (financement à plus court terme) alors qu'à l'inverse le risque liquidité en est fonction croissante.

Notons que le jeu de paramètres du modèle peut être interprété à la fois pour des banques ou des compagnies d'assurance. Ainsi, en théorie, notre modèle se prête au cas d'un réseau constitué de banques et de compagnies d'assurance. Au lieu de se spécialiser dans une classe d'actif Ax_i, les compagnies d'assurance se spécialisent dans une classe de risque. Dans ce contexte, ω_i ne doit plus être interprété comme une transformation de maturité mais comme une sévérité moyenne de la classe de risques assurée par la compagnie i. Dans la suite, nous ne considérons que le cas des banques.

Nous proposons un modèle à une période. A la date $t = 0$, chaque institution optimise l'espérance d'utilité de ses capitaux propres à l'instant $t = 1$. Les dates sont notées en exposant et entre parenthèses. Les bilans de la banque i en $t = 0$ et $t = 1$ sont respectivement présentés dans les tableaux 1.1 and 1.2.

		Actif	Passif		
détentions croisées interbancaires en actions	$\leftrightarrow$	$\pi_{i,1}\mathcal{K}_1^{(0)}$ $\vdots$ $\pi_{i,n}\mathcal{K}_n^{(0)}$	L_i^*	$\leftrightarrow$	dette
prêts interbancaires	$\leftrightarrow$	$\gamma_{i,1}\mathcal{L}_1^{(0)}$ $\vdots$ $\gamma_{i,n}\mathcal{L}_n^{(0)}$	$K_i^{(0)}$	$\leftrightarrow$	capitaux propres
actifs illiquides externes	$\leftrightarrow$	$Ax_i^{(0)}$			
cash	$\leftrightarrow$	$A\ell_i^{(0)}$			

TABLE 1.1 – Bilan de l'institution i en $t = 0$.

		Actif	Passif		
détentions croisées interbancaires en actions	$\leftrightarrow$	$\pi_{i,1}K_1^{(1)}$ $\vdots$ $\pi_{i,n}K_n^{(1)}$	$L_i^{(1)}$	$\leftrightarrow$	dette
prêts interbancaires	$\leftrightarrow$	$\gamma_{i,1}L_1^{(1)}$ $\vdots$ $\gamma_{i,n}L_n^{(1)}$	$K_i^{(1)}$	$\leftrightarrow$	capitaux propres
actifs illiquides externes	$\leftrightarrow$	$Ax_i^{(1)}$			
cash	$\leftrightarrow$	$A\ell_i^{(1)}$			

TABLE 1.2 – Bilan de l'institution i en $t = 1$.

Il est important de noter que les capitaux propres et la dette des autres institutions (à l'actif) sont "pricés" à la valeur de marché en $t = 0$, alors qu'en $t = 1$, la valeur comptable réalisée est utilisée. Ceci garantit l'absence d'opportunités d'arbitrage.

Selon le modèle Value-of-the-Firm (Merton, 1974), la valeur de la dette L_i et celle des capitaux propres K_i Ãă toutes dates doivent vérifier les conditions d'équilibre

$$K_i = \max\Big(\sum_{j=1}^{n}\pi_{i,j}K_j + \sum_{j=1}^{n}\gamma_{i,j}L_j + A\ell_i + Ax_i - L_i^*, 0\Big),$$

$$L_i = \min\Big(\sum_{j=1}^{n}\pi_{i,j}K_j + \sum_{j=1}^{n}\gamma_{i,j}L_j + A\ell_i + Ax_i, L_i^*\Big).$$

La Proposition 2 dans Gouriéroux *et al.* (2012) stipule que quelles que soient les valeurs de Ax_i, $A\ell_i$, π_{ij}, γ_{ij} et L_i^* (vérifiant certaines hypothèses), le réseau existe théoriquement (pour un jeu unique de valeurs K_i et L_i, $i = 1, \ldots, n$). Cela implique donc que l'on a le droit d'optimiser les différentes grandeurs du bilan.

Cette optimisation est effectuée sous contraintes réglementaires. La contrainte de solvabilité est

$$K_i^{(0)} \geq k_i^A Ax_i^{(0)} + k^\pi \sum_{j=1}^{n}\pi_{i,j}\mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^{n}\gamma_{i,j}\mathcal{L}_j^{(0)},$$

où k_i^A, k^π et k^γ sont des paramètres réglementaires (poids du risque) associés respectivement aux actifs externes et détentions croisées en actions et en dette, vérifiant $0 < k_i, k^\pi, k^\gamma < 1$. Cette contrainte signifie que les capitaux propres doivent être suffisants. La contrainte de liquidité s'écrit

$$A\ell_i^{(0)} \geq k^L\ l(\omega_i,\ L_i^*),$$

où l est une fonction croissante (par rapport aux deux variables) et k_L vérifie $0 < k^L < 1$. Elle assure un matelas de cash pour faire face à un éventuel choc de liquidité.

Benchmark :

Le réseau obtenu grâce à notre modèle sera confronté avec un réseau type. Pour des raisons de confidentialité, nous ne pouvons nous appuyer sur les données détaillées d'un réseau réel. Ainsi, nous présentons des faits stylisés fondés sur des données agrégées.

Fait stylisé 1 : Les actifs externes Ax_i représentent généralement 90 à 95% des actifs totaux. Les capitaux propres K_i en représentent environ 5%.

Fait stylisé 2 : Un réseau composé de banques hétérogènes en taille présente généralement une structure de type "coeur-périphérie". Les matrices contenant les π_{ij} et les γ_{ij}, Π et Γ, présentent alors une structure par blocs avec beaucoup de zéros.

Fait stylisé 3 : Un réseau composé de grandes banques homogènes est en général complet. Ainsi Π et Γ contiennent très peu de coefficients nuls.

Fait stylisé 4 : Dans le cas des grandes banques, les participations croisées en dette représentent la majeure partie des interconnexions (entre 80 et 90%).

1.5.2 Le programme d'optimisation individuelle

Nous décrivons la formation du réseau en deux étapes. Dans la première, nous expliquons comment une institution optimise son propre bilan (et notamment ses interconnexions), connaissant le bilan des autres institutions du réseau. Dans la seconde, nous présentons une démarche itérative permettant de former le réseau en se fondant sur les comportements individuels décrits à la première étape. Dans cette partie, nous nous concentrons sur la première étape.

Une difficulté importante provient du fait que les choix des différentes institutions sont inter-dépendants. A moins de s'affranchir des problèmes d'information (information incomplète, anticipations,...), il apparaît extrêmement délicat de modéliser la totalité du bilan via un équilibre de Nash. Babus (2007) et Acemoglu *et al.* (2013) utilisent un équilibre de Nash, mais les éléments du bilan autres que les interconnexions sont exogènes. Une telle démarche pose problème dans notre cas étant donné que nous considérons des interconnexions de long terme.

Nous adoptons donc la démarche simplifiée suivante. Chaque institution suppose que l'actif des autres banques est seulement composé d'actif externe (c'est-à-dire d'aucune interconnexion). Ainsi, la banque i suppose que la banque j n'est pas interconnectée avec elle. Cela implique que la banque i peut supposer que ses choix n'ont pas d'impact sur la banque j et qu'elle n'a donc pas à anticiper les réactions futures de la banque j à ses propres choix. En résumé, chaque banque optimise son bilan de manière juste, sans prendre en compte les futures réactions des autres. Cette hypothèse apparaît fondée, et ce pour trois raisons. Tout d'abord, les données d'interconnexions ne sont pas publiques. Deuxièmement, les actions et obligations étant en grande partie échangées sur le marché secondaire, chaque institution ne connaît pas les détenteurs de sa dette et de ses actions et ne peut donc reconstituer les données d'interconnexion. Enfin, les interconnexions représentent à peine 5 à 10% du total des actifs, ce qui signifie que chaque interconnexion ne doit pas en représenter beaucoup plus que 0.5 à 1%. La réaction des partenaires ne porte donc que sur une proportion faible du total des actifs et peut donc être négligée[11].

Du fait de cette hypothèse d'absence d'interconnexions, il convient de corriger le total actif des partenaires afin de rétablir l'équilibre de leur bilan. Ceci est fait via les facteurs

[11]Notons toutefois que l'on aurait pu considérer un vrai équilibre de Nash, à condition de s'affranchir des problèmes d'information. Ce que l'on fait en est d'ailleurs très proche étant donné que nous considérons la fonction de meilleure réponse aux actions des autres.

$\kappa_j = \dfrac{L_j^{(0)} + K_j^{(0)}}{Ax_j^{(0)} + A\ell_j^{(0)}}$. Ainsi, les capitaux propres de l'institution i à la date $t = 1$ s'écrivent

$$K_i^{(1)} = \max\Big[Ax_i^{(1)} + A\ell_i^{(1)} + \sum_{j=1}^{n} \pi_{i,j} \max\Big(\kappa_j(Ax_j^{(1)} + A\ell_j^{(1)}) - L_j^{*(1)}, 0\Big) + \sum_{j=1}^{n} \gamma_{i,j} \min\Big(\kappa_j(Ax_j^{(1)} + A\ell_j^{(1)}), L_j^{*(1)}\Big) - (1 + r_D(\omega_i))L_i^{(0)}, 0\Big].$$

Chaque institution est gérée dans l'intérêt de ses investisseurs. Ceux-ci sont risque adverses et possèdent une fonction d'utilité notée u_i. Ils sont dotés d'un capital initial $K_i^{(0)}$. Nous notons respectivement $1-c_i^{\pi}$ et $1-c_i^{\gamma}$ le flottant (la part échangeable sur le marché) des actions et de la dette de l'institution j.
Le programme d'optimisation $\mathcal{P}_i$ de l'institution i s'écrit alors

$$\mathcal{P}_i := \begin{cases} \max\limits_{\substack{Ax_i^{(0)}, A\ell_i^{(0)} \\ L_i^{(0)}, \omega_i \\ \pi_{i,1}, \dots, \pi_{i,n} \\ \gamma_{i,1}, \dots, \gamma_{i,n}}} & \mathbb{E}_0[u_i(K_i^{(1)})] \\ \text{tel que} & Ax_i^{(0)} + A\ell_i^{(0)} + \sum\limits_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + \sum\limits_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} = K_i^{(0)} + L_i^{(0)} \quad (BC) \\ & K_i^{(0)} \geq k_i^A \; Ax_i^{(0)} + k^{\pi} \sum\limits_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + k^{\gamma} \sum\limits_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} \quad (SC) \\ & A\ell_i^{(0)} \geq k^L \; l(\omega_i, \; L_i^{(0)}) \quad (LC) \\ & Ax_i^{(0)} \geq 0, A\ell_i^{(0)} \geq 0, L_i^{(0)} \geq 0 \\ & \omega_i \in [0,1] \\ & \forall j \in \{1, \dots, n\}, 0 \leq \pi_{i,j} \leq 1 - c_j^{\pi}, 0 \leq \gamma_{i,j} \leq 1 - c_j^{\gamma}. \end{cases}$$

1.5.3 Existence, unicité et caractérisation de la solution

Sous certaines hypothèses mentionnées dans la proposition suivante, $\mathcal{P}_i$ admet une solution.

Théorème 1.9 (Existence de la solution de $\mathcal{P}_i$)**.** *Si*

- ($A1$) *Les investisseurs négligent les interconnexions de leurs contreparties,*
- ($A2$) *La fonction d'utilité u_i est continue et croissante,*
- ($A3$) *La fonction F_R est continue. De plus, f_R est strictement positive sur $[a, +\infty)^n$, pour une constante $a \in \mathbb{R}$.*
- ($A4$) *La courbe de taux r_D est continue et strictement supérieure au taux sans risque r_{rf},*

alors $\mathcal{P}_i$ admet une solution.

La question de l'unicité de la solution de ce programme pose de nombreuses difficultés. Une stratégie classique pour montrer l'unicité d'un problème de maximisation consiste à prouver que la fonction objective est strictement concave sur un ensemble fermé convexe. Dans notre cas, la fonction objective s'écrit $\mathbb{E}_0[u_i(K_i^{(1)})]$. Si l'on introduit la position de la banque i, $P_i^{(1)}$, comme l'écart entre son actif et son passif, on a alors, par définition des capitaux propres, $K_i^{(1)} = \max(P_i^{(1)}, 0)$. Ainsi, du fait de la responsabilité limitée des actionnaires, la fonction $u_i \; o \; K_i^{(1)}$ est non différentiable et non concave pour les fonctions d'utilité u_i standards. Du fait de l'opération d'intégration, il est possible que la fonction $\mathbb{E}_0[u_i(K_i^{(1)})]$ soit strictement concave même si $u_i \; o \; K_i^{(1)}$ n'est pas concave. Néanmoins, compte tenu de la dimension du problème $\mathcal{P}_i$, étudier la stricte concavité de $\mathbb{E}_0[u_i(K_i^{(1)})]$ impose des calculs d'intégrale infaisables à part dans des cas où F_R est très simple. Par ailleurs, sans hypothèse particulière sur u_i, la stricte concavité de $\mathbb{E}_0[u_i(K_i^{(1)})]$ n'est probablement vraie que pour des fonctions F_R particulières. Pour toutes ces raisons, en ce qui concerne la stricte concavité, nous ne travaillons pas sur la fonction $\mathbb{E}_0[u_i(K_i^{(1)})]$ mais directement sur la fonction $u_i \; o \; K_i^{(1)}$. Cette dernière n'étant pas strictement concave, nous proposons une approximation de $K_i^{(1)}$ qui la rende concave. $K_i^{(1)}$ est remplacé par $v(P_i^{(1)})$ où v est une fonction régulière à définir. Nous définissons ainsi le programme d'optimisation approché $\mathcal{P}_i^{'}$ où la fonction objectif $\mathbb{E}_0[u_i(K_i^{(1)})]$ est remplacée par $\mathbb{E}_0\{u_i[v(P_i^{(1)})]\}$. Les conditions d'unicité du programme $\mathcal{P}_i^{'}$ sont données dans la proposition suivante.

Théorème 1.10 (Existence et unicité de $\mathcal{P}_i'$)**.** *Sous les hypothèses précédentes* $(A1)$*,* $(A2)$*,* $(A3)$*,* $(A4)$ *et les hypothèses :*

- $(A5)$ *la composée de la fonction de transformation* v *et de la fonction d'utilité* u_i *est strictement concave :* $\forall P \in \mathbb{R}, (u_i \; o \; v)''(P) < 0$ *;*
- $(A6)$ *la courbe de taux d'intérêt est strictement concave :* $\forall \omega_i \in [0,1], r_D''(\omega_i) < 0$ *;*
- $(A7)$ *l'intérêt sur la dette vérifie* $\forall \omega_i \in [0,1], r_D'(\omega_i) \neq 0$ *;*
- $(A8)$ *la fonction* l *dans* (LC) *vérifie :*

$$\frac{\partial^2 l}{\partial \omega_i^2} \geq 0 \; et \; \frac{\partial^2 l}{\partial \omega_i^2} \frac{\partial^2 l}{\partial {L_i^{(0)}}^2} \geq \left(\frac{\partial^2 l}{\partial \omega_i \partial L_i^{(0)}} \right)^2 ;$$

alors $\mathcal{P}_i^{'}$ *admet une unique solution au sens suivant. Si toutes les variables de contrôle apparaissant à l'actif de la banque* i *sont fixées à l'exception d'une seule, notée* $Ac_i^{(0)}$*, alors il y a unicité du triplet optimal* $(Ac_i^{(0)}, \; L_i^{(0)}, \; \omega_i)$*.*

Ce résultat signifie également que les grands éléments du bilan sont uniques, à savoir le triplet $(A_i^{(0)}, L_i^{(0)}, \omega_i)$, où $A_i^{(0)}$ désigne le total des actifs de l'institution i. Du fait de l'opération d'intégration, il est possible que toutes les variables de contrôle soient uniques. Mais compte-tenu des difficultés techniques déjà évoquées, ceci ne peut être vérifié que numériquement.

La proposition suivante fournit les spécifications que nous préconisons, en accord avec les hypothèses de la Proposition 1.10.

Corollaire 1.1 (Existence et unicité du programme d'optimisation spécifique)**.** *Outres les hypothèses* $(A1)$ *à* $(A4)$*, considérons :*

- *une fonction d'utilité logarithmique*

$$u_i(x) = \log(x);$$

- *la fonction d'approximation de la responsabilité limitée des actionnaires suivante :*

$$v(P) = \log\left(\exp(P) + 1\right);$$

- *une contrainte de liquidité exponentielle :*

$$l(\omega, L) = \exp(\omega)\exp(L);$$

- *la courbe de taux d'intérêt suivante :*

$$r_D(\omega) = \alpha - \beta\exp(\omega) \qquad \textit{pour } \omega \in [0,1].$$

Dans ce cas, le programme d'optimisation associé $\mathcal{P}_i'$ possède une unique solution (au sens précédent).

Les Propositions 1.9, 1.10 et 1.1 fournissent des résultats théoriques forts et utiles. Néanmoins, elles ne donnent aucune indication sur la forme de la solution et notamment les interconnexions.

Nous montrons sur un cas simplifié que sous certaines conditions, les interconnexions peuvent être non nulles. Ce résultat est établi en utilisant les conditions de Karush, Kuhn et Tucker (KKT). Néanmoins, compte-tenu du grand nombre de variables de contrôle et de contraintes, il est impossible de déterminer la solution de manière explicite. Ainsi, nous décomposons le problème : nous considérons d'abord un agent neutre au risque sans responsabilité limitée, puis un agent averse au risque et enfin l'impact de la responsabilité limitée est pris en compte. Dans cette analyse, compte-tenu de la difficulté du problème, nous considérons que ω_i est fixé de manière exogène.

Le cas d'un agent neutre au risque est régi par la proposition suivante.

Théorème 1.11. *Soit le programme d'optimisation suivant :*

$$\mathcal{P}_{\mathcal{RNG}} := \left\{ \begin{array}{rl} \max & \left(Ax_i\mathbb{E}(R_{g,i}) + \sum_{j=1}^n \pi_{ij}\mathbb{E}(K_j^{(1)}) + \sum_{j=1}^n \gamma_{ij}\mathbb{E}(L_j^{(1)})\right) \\ Ax_i, \pi_{ij}, \gamma_{ij} & \\ \textit{tel que} & k^A Ax_i + k^\pi \sum_{j=1}^n \pi_{ij}\mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^n \gamma_{ij}\mathcal{L}_j^{(0)} \leq 1 \\ & Ax \geq 0 \\ & 0 \leq \pi_{ij} \leq c^\pi \\ & 0 \leq \gamma_{ij} \leq c^\gamma \end{array} \right. .$$

Afin de déterminer la solution de ce problème, il faut trier dans l'ordre décroissant les rendements suivants (rapportés à leur poids réglementaire) : $\dfrac{\mathbb{E}(R_{g,i})}{k^A}$, $\dfrac{\mathbb{E}(K_j^{(1)})}{k^\pi\mathcal{K}_j^{(0)}}$ $(j = 1, \ldots, n)$, $\dfrac{\mathbb{E}(L_j^{(1)})}{k^\gamma\mathcal{L}_j^{(0)}}$ $(j = 1, \ldots, n)$. *La solution optimale consiste à investir le plus possible dans l'actif ayant le rendement (rapporté au poids réglementaire) le plus élevé. Lorsque cet actif n'est plus disponible, il est alors optimal d'investir dans le second le plus rentable, et ainsi de suite. Ceci doit être répété jusqu'à saturation de la contrainte de solvabilité.*

Notons que la contrainte d'équilibre du bilan n'apparaît pas dans le programme $\mathcal{P}_{\mathcal{RNG}}$. En effet, cette contrainte d'égalité peut être supprimée en écrivant par exemple la dette $L_i^{(0)}$ en fonction des autres variables de contrôle (voir le coeur du chapitre pour davantage de détails).

Cette proposition conduit à un réseau très structuré et directionnel. Dans le cas d'un agent adverse au risque, du fait de son désir de réduire le risque et donc la variance, l'investissement est plus diversifié (y compris dans des actifs moins rentables que d'autres). On s'attend donc à un réseau assez complet, ce qui sera vérifié numériquement. Cela est vrai principalement dans le cas où les actifs spécifiques Ax_i sont suffisamment risqués. Des corrélations négatives entre les actifs spécifiques sont également favorables à un réseau complet.

Enfin, la prise en compte de la responsabilité limitée permet de voir l'impact de la nature différente des obligations et des actions. Considérons un cas simplifié à deux banques notées 1 et 2. Si la corrélation ρ entre Ax_1 et Ax_2 est positive, alors chaque banque préfère investir dans les actions de l'autre banque plutôt que dans ses obligations. Dans le cas d'une corrélation négative, l'inverse se produit. La responsabilité limitée a donc un impact sur le choix entre actions et obligations.

1.5.4 Formation du réseau, calibration et étude numérique

Nous venons d'analyser le programme d'optimisation d'une institution connaissant l'état des autres institutions. La seconde étape consiste à créer le réseau total en utilisant le programme d'optimisation individuel. Pour ce faire, nous utilisons un algorithme itératif. A chaque étape, une institution unique optimise son bilan connaissant l'état du réseau obtenu à l'étape précédente. Ce procédé est répété jusqu'à l'équilibre. Numériquement, on constate que cet algorithme converge rapidement vers un équilibre. Notons que s'il y a équilibre, il est nécessairement unique du fait de la Proposition 1.10.

Pour l'étude numérique, nous choisissons la calibration suivante. Nous prenons $n = 2$, $K_1^{(0)} = K_2^{(0)} = 1$ et $u_i(x) = \log(x)$. Compte-tenu de la complexité du modèle et par conséquent des difficultés de calibration, nous excluons ω_i des variables de contrôle. A la place de $r_D(\omega_1)$ et $r_D(\omega_2)$, nous prenons donc $r_{D,1} = r_{D,2} \in \{0, 1\%\}$. Pour les mêmes raisons, nous imposons $A\ell_1 = A\ell_2 = 0$ et excluons la contrainte de liquidité (LC). En ce qui concerne les poids du risque, nous comparons 8 jeux réglementaires possibles. Conformément au modèle de Merton (1974), nous choisissons une dynamique log normale pour le rendement brut des actifs spécifiques sous-jacents. Afin de calibrer les paramètres de la loi normale sous-jacente, à savoir μ_1, μ_2, σ_1 et σ_2 et ρ, nous prenons des valeurs de rendements nets de l'ordre de 1% en accord avec le rapport BHC (Bank Holding Company), et fixons les probabilités de défaut en autarcie à $PD_1 = PD_2 = 0.1\%$. Le coefficient ρ impacte fortement le réseau obtenu et nous le faisons donc varier entre -1 et 1.

Les valeurs de marché des capitaux propres et de la dette sont calculées en prenant l'espérance actualisée de leur valeur en $t = 1$, sous la probabilité physique. En effet, l'actif sous-jacent Ax est illiquide et l'existence d'une unique probabilité risque neutre n'est donc pas assurée. Cependant, dans ce contexte, les actions et obligations ont une rentabilité moyenne égale au taux sans risque sous la probabilité physique, ce qui ne les rend pas particulièrement attractifs. Il y a donc principalement diversification pour des raisons de réduction de la variance.

Pour des valeurs suffisamment négatives de ρ, notre modèle fournit des résultats

proches de ceux observés dans la réalité en termes de proportion d'actifs inter-bancaires dans le bilan total. Néanmoins, les interconnexions via les obligations apparaissent faibles et, pour $\rho > 0$, notre modèle a tendance à sous-estimer l'ensemble des interconnexions. Ceci provient en partie du fait que le prix des actions et obligations ait été établi sous la probabilité physique, ce qui rend ces instruments relativement peu attractifs. On peut s'attendre à des résultats plus réalistes dans le cas d'un "pricing" sous la probabilité risque-neutre.

1.5.5 Application au calcul du bien-être économique

En guise d'application, nous étudions l'évolution du bien-être économique (welfare) en fonction des paramètres réglementaires. L'un des avantages d'un modèle de réseau endogène est qu'il permet de savoir comment le réseau réagirait sous des conditions réglementaires différentes des conditions actuelles. Nous adaptons l'analyse de welfare menée par Repullo et Suarez (2013). Le welfare W est défini comme la somme des contributions de toutes les banques. La contribution est positive si la banque est en vie en $t = 1$ (car elle peut alors octroyer des prêts Ax qui financent l'économie réelle) et négative si elle est en défaut (car l'Etat doit rembourser les dépôts garantis). Nous calculons W pour chaque jeu de paramètres réglementaires et montrons que le bien-être est supérieur dans le cas de réglementations favorisant les interconnexions.

Bibliographie

Acemoglu, D., Ozdaglar, A. et Tahbaz-Salehi, A. (2013). Systemic risk and stability in financial networks. Rapport technique, National Bureau of Economic Research.

Ailliot, P., Thompson, C. et Thomson, P. (2009). Space–time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 58(3):405–426.

Albrecher, H., Hartinger, J. et Tichy, R. F. (2003). Multivariate approximation methods for the pricing of catastrophe-linked bonds. *International Series of Numercial Mathematics*, pages 21–40.

Albrecher, H., Hartinger, J. et Tichy, R. F. (2004). QMC techniques for CAT bond pricing. *Monte Carlo Meth. and Appl.*, 10(3):197–211.

Allard, D. et Bourotte, M. (2013). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *Stochastic Environmental Research and Risk Assessment*, Submitted.

Allen, F. et Gale, D. (2000). Financial contagion. *Journal of political economy*, 108(1): 1–33.

Anand, K., Gai, P., Kapadia, S., Brennan, S. et Willison, M. (2012). A network model of financial system resilience.

Apipattanavis, S., Podestá, G., Rajagopalan, B. et Katz, R. W. (2007). A semiparametric multivariate and multisite weather generator. *Water Resources Research*, 43(11).

Arinaminpathy, N., Kapadia, S. et May, R. M. (2012). Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences*, 109(45):18338–18343.

Artzner, P., Delbaen, F., Eber, J.-M. et Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. et Ku, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, 152(1):5–22.

Azizpour, S., Giesecke, K. *et al.* (2008). Self-exciting corporate defaults : contagion vs. frailty. *Manuscript, Stanford University.*

Babus, A. (2007). The formation of financial networks.

Bailey, N. T. (1953). The total size of a general stochastic epidemic. *Biometrika*, pages 177–185.

Bailey, N. T. (1957). *The mathematical theory of epidemics*. Griffin London.

Bárdossy, A. et Pegram, G. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, 13(12):2299–2314.

Bardossy, A. et Plate, E. J. (1992). Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28(5):1247–1259.

Barrieu, P. et Karoui, N. E. (2007). Pricing, hedging and optimally designing derivatives via minimization of risk measures. *arXiv preprint arXiv :0708.0948*.

Barrieu, P. et Loubergé, H. (2009). Hybrid cat bonds. *Journal of Risk and Insurance*, 76(3):547–578.

Bienvenüe, A. et Robert, C. Y. (2014). Likelihood based inference for high-dimensional extreme value distributions. *arXiv preprint arXiv :1403.0065*.

Bion-Nadal, J. (2004). Conditional risk measure and robust representation of convex conditional risk measures. *CMAP Preprint*, 557.

Bion-Nadal, J. (2006). Dynamic risk measuring : discrete time in a context of uncertainty, and continuous time on a probability space. *CMAP Preprint*, 596.

Bluhm, M., Faia, E. et Krahnen, J. P. (2013). Endogenous banks' networks, cascades and systemic risk. Rapport technique, Mimeo, Goethe University.

Boldi, M.-O. et Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69(2):217–229.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Brockwell, P. J. et Davis, R. A. (2009). *Time series : theory and methods*. Springer.

Brown, B. M. et Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Buishand, T. A. et Brandsma, T. (2001). Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest-neighbor resampling. *Water Resources Research*, 37(11):2761–2776.

Cannon, A. J. (2008). Probabilistic multisite precipitation downscaling by an expanded Bernoulli-gamma density network. *Journal of Hydrometeorology*, 9(6):1284–1300.

Charles, S. P., Hughes, J., Bates, B. et Lyons, T. (1996). Assessing downscaling models for atmospheric circulation| local precipitation linkage. *In Proceedings of the International Conference on Water Resources and Environmental Research : Towards the 21st Century, Tokyo, Japan*.

Cheridito, P., Delbaen, F. et Kupper, M. (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11(3):57–106.

Cifuentes, R., Ferrucci, G. et Shin, H. S. (2005). Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2-3):556–566.

Coles, S., Bawa, J., Trenner, L. et Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Cox, S. H. et Pedersen, H. W. (2000). Catastrophe risk bonds. *North American Actuarial Journal*, 4(4):56–82.

Cummins, J. D. (2008). Cat bonds and other risk-linked securities : State of the market and recent developments. *Risk Management and Insurance Review*, 11(1):23–47.

Davis, M. et Lo, V. (2001). Modelling default correlation in bond portfolios. *Mastering risk*, 2:141–151.

Davison, A. C. (2003). *Statistical models*. Cambridge University Press.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. (1990). Fighting the arch–enemy with mathematics. *Statistica neerlandica*, 44(2):45–68.

de Haan, L. et Ferreira, A. (2007). *Extreme value theory : an introduction*. Springer.

Delbaen, F. (2000). Coherent risk measures. *Blätter der DGVFM*, 24(4):733–739.

Delbaen, F. (2006). The structure of m–stable sets and in particular of the set of risk neutral measures. *In In Memoriam Paul-André Meyer*, pages 215–258. Springer.

Demange, G. (2012). Contagion in financial networks : a threat index.

Denuit, M., Dhaene, J., Goovaerts, M. et Kaas, R. (2006). *Actuarial theory for dependent risks : measures, orders and models*. John Wiley & Sons.

Detlefsen, K. et Scandolo, G. (2005). Conditional and dynamic convex risk measures. *Finance and Stochastics*, 9(4):539–561.

Diamond, D. W. et Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *The journal of political economy*, pages 401–419.

Donat, M., Pardowitz, T., Leckebusch, G., Ulbrich, U. et Burghoff, O. (2011). High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Natural Hazards and Earth System Science*, 11(10):2821–2833.

Einmahl, J. H. et Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, pages 2953–2989.

Eisenberg, L. et Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.

Elliott, M., Golub, B. et Jackson, M. (2012). Financial networks and contagion. *Available at SSRN 2175056*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica : Journal of the Econometric Society*, pages 987–1007.

Fermanian, J.-D. et Salanié, B. (2004). A nonparametric simulated maximum likelihood estimation method. *Econometric Theory*, 20(04):701–734.

Finn, J. et Lane, M. (1997). The perfume of the premium... or pricing insurance derivatives. *In Proceedings of the 1995 Bowles Symposium on Securitization of Risk*, pages 27–35.

Fisher, R. A. et Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *In Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.

Föllmer, H. (2014). Spatial risk measures and their local specification : The locally law-invariant case.

Föllmer, H. et Klüppelberg, C. (2014). Spatial risk measures : Local specification and boundary risk.

Föllmer, H. et Penner, I. (2006). Convex risk measures and the dynamics of their penalty functions. *Statistics & Decisions*, 24(1/2006):61–96.

Föllmer, H., Schied, A. et Lyons, T. J. (2004). Stochastic finance. An introduction in discrete time. *The Mathematical Intelligencer*, 26(4):67–68.

Fougères, A.-L. (2004). Multivariate extremes. *Monographs on Statistics and Applied Probability*, 99:373–388.

Frey, R. et Backhaus, J. (2003). Interacting defaults and counterparty risk : a Markovian approach. *University of Leipzig, Leipzig, Germany*.

Frittelli, M. et Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486.

Frittelli, M. et Rosazza Gianin, E. (2004). Dynamic convex risk measures. *Risk measures for the 21st century*, pages 227–248.

Gabriel, K. et Neumann, J. (1962). A Markov chain model for rainfall occurrence at tel aviv. *Quart. J. R. met. Soc.*, 88:90–95.

Gagliardini, P. et Gouriéroux, C. (2013). Correlated risks vs contagion in stochastic transition models. *Journal of Economic Dynamics and Control*.

Genton, M. G., Ma, Y. et Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

Giesecke, K. et Weber, S. (2004). Cyclical correlations, credit contagion, and portfolio losses. *Journal of Banking & Finance*, 28(12):3009–3036.

Giesecke, K. et Weber, S. (2006). Credit contagion and aggregate losses. *Journal of Economic Dynamics and Control*, 30(5):741–767.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *The Annals of Mathematics*, 44(3):423–453.

Gorodetskii, V. (1987). Moment inequalities and the central limit theorem for integrals of random fields with mixing. *Journal of Soviet Mathematics*, 36(4):461–467.

Gouriéroux, C., Héam, J.-C. et Monfort, A. (2012). Bilateral exposures and systemic solvency risk. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(4): 1273–1309.

Goyal, S. (2012). *Connections : an introduction to the economics of networks*. Princeton University Press.

Guidolin, M. et Timmermann, A. (2006). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, 131(1):285–308.

Hastie, T. J. et Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC Press.

Hay, L. E., McCabe, G. J., Wolock, D. M. et Ayers, M. A. (1991). Simulation of precipitation by weather type analysis. *Water Resources Research*, 27(4):493–501.

Holmstrom, B. et Tirole, J. (1996). Private and public supply of liquidity. Rapport technique, National Bureau of Economic Research.

Hosking, J. (1985). Algorithm as 215 : Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics*, pages 301–310.

Hosking, J., Wallis, J. R. et Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Hughes, J. P. et Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, 30(5):1535–1546.

Hughes, J. P., Guttorp, P. et Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 48(1):15–30.

Huser, R. et Davison, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.

Kabluchko, Z. et Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Processes and their Applications*, 120(3):281–295.

Kabluchko, Z., Schlather, M. et de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, pages 2042–2065.

Katz, R. W. (1977). Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, 16(7):671–676.

Kaufman, G. G. (1994). Bank contagion : A review of the theory and evidence. *Journal of Financial Services Research*, 8(2):123–150.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. *In Proc. 3rd Berkeley Symp. Math. Statist. Prob*, volume 4, pages 149–165.

Kidson, J. W. (1994). Relationship of New Zealand daily and monthly weather patterns to synoptic weather types. *International journal of climatology*, 14(7):723–737.

Klawa, M., Ulbrich, U. *et al.* (2003). A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Science*, 3(6):725–732.

Kleiber, W., Katz, R. W. et Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Resources Research*, 48(1).

Klöppel, S. et Schweizer, M. (2007). Dynamic indifference valuation via convex risk measures. *Mathematical Finance*, 17(4):599–627.

Lall, U. et Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3):679–693.

Leadbetter, M. R., Lindgren, G. et Rootzén, H. (1989). Extremes and related properties of random sequences and processes.

Lee, J.-P. et Yu, M.-T. (2002). Pricing default-risky cat bonds with moral hazard and basis risk. *Journal of Risk and Insurance*, 69(1):25–44.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39.

Litzenberger, R. H., Beaglehole, D. R. et Reynolds, C. E. (1996). Assessing catastrophe reinsurance-linked securities as a new asset class. *Journal of Portfolio Management*, 23:76–86.

Loftsgaarden, D. O. et Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, pages 1049–1051.

Manski, C. F. (1993). Identification of endogenous social effects : The reflection problem. *The review of economic studies*, 60(3):531–542.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

Matheron, G. (1965). *Les variables régionalisées et leur estimation*. Thèse de doctorat.

Merton, R. C. (1974). On the pricing of corporate debt : The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Padoan, S. A., Ribatet, M. et Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Peng, S. (2004). Nonlinear expectations, nonlinear evaluations and risk measures. *In Stochastic methods in finance*, pages 165–253. Springer.

Penrose, M. D. (1992). Semi-min-stable processes. *The Annals of Probability*, 20(3): 1450–1463.

Pickands, J. (1981). Multivariate extreme value distributions. *In Proceedings 43rd Session International Statistical Institute*, volume 2, pages 859–878.

Pinto, J., Fröhlich, E., Leckebusch, G. et Ulbrich, U. (2007). Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM. *Natural Hazards and Earth System Science*, 7(1):165–175.

Prescott, P. et Walden, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Rajagopalan, B. et Lall, U. (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, 35(10):3089–3101.

Rajagopalan, B., Lall, U., Tarboton, D. et Bowles, D. (1997). Multivariate nonparametric resampling scheme for generation of daily weather variables. *Stochastic Hydrology and Hydraulics*, 11(1):65–93.

Repullo, R. et Suarez, J. (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies*, 26(2):452–490.

Resnick, S. (1987). Extreme values, regular variation, and point processes.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic processes and their applications*, 112(2):185–200.

Rochet, J.-C. (2004). Macroeconomic shocks and banking supervision. *Journal of Financial Stability*, 1(1):93–110.

Rosazza Gianin, E. (2006). Risk measures via g-expectations. *Insurance : Mathematics and Economics*, 39(1):19–34.

Sang, H. et Genton, M. G. (2013). Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*.

Sansó, B. et Guenni, L. (2000). A nonstationary multisite model for rainfall. *Journal of the American Statistical Association*, 95(452):1089–1100.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. et Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values : Properties and inference. *Biometrika*, 90(1):139–156.

Scott, D. W. (1992). Multivariate density estimation, 317 pp.

Semenov, M. A. et Porter, J. (1995). Climatic variability and the modelling of crop yields. *Agricultural and forest meteorology*, 73(3):265–283.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician*, 33(2):57–62.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript, University of North Carolina*.

Spodarev, E. (2014). Limit theorems for excursion sets of stationary random fields. *In Modern Stochastics and Applications*, pages 221–241. Springer.

Stern, R. et Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)*, pages 1–34.

Tirole, J. (2010). *The theory of corporate finance*. Princeton University Press.

Trigo, R. M. et Palutikof, J. P. (1999). Simulation of daily temperatures for climate change scenarios over Portugal : a neural network model approach. *Climate Research*, 13(1):45–59.

Upper, C. (2007). Using counterfactual simulations to assess the danger of contagion in interbank markets.

Upper, C. et Worms, A. (2004). Estimating bilateral exposures in the german interbank market : Is there a danger of contagion ? *European Economic Review*, 48(4):827–849.

Varin, C. et Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Von Mises, R. (1954). La distribution de la plus grande de n valeurs, in selected papers, volume ii, vol. 44 of american mathematical society. *Providence, RI*.

Wilks, D. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191.

Wilks, D. S. (2009). A gridded multisite weather generator and synchronization to observed weather data. *Water Resources Research*, 45(10):W10419.

Wilks, D. S. et Wilby, R. L. (1999). The weather generation game : a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357.

Woolhiser, D. A. et Pegram, G. (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *Journal of Applied Meteorology*, 18:34–42.

Yang, C., Chandler, R., Isham, V. et Wheater, H. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, 41(11).

Yates, D., Gangopadhyay, S., Rajagopalan, B. et Strzepek, K. (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research*, 39(7).

Young, K. C. (1994). A multivariate chain model for simulating climatic parameters from daily data. *Journal of Applied Meteorology*, 33:661–671.

Zheng, X. et Katz, R. W. (2008). Simulation of spatial dependence in daily rainfall using multisite generators. *Water Resources Research*, 44(9):W09403.

Zheng, X., Renwick, J. et Clark, A. (2010). Simulation of multisite precipitation using an extended chain-dependent process. *Water Resources Research*, 46(1).

Chapter 2

A multi-site precipitation generator based on a frailty-contagion approach

Accurate stochastic simulations of hourly precipitation are needed for impact studies at local spatial scales. Statistically, hourly precipitation data represent a difficult challenge. They are non-negative, skewed, heavy tailed, contain a lot of zeros (dry hours) and they have complex temporal structures (e.g., long persistence of dry episodes). Inspired by frailty-contagion approaches used in finance and insurance, we propose a multi-site precipitation simulator that, given appropriate regional atmospheric variables, can simultaneously handle dry events and heavy rainfall periods. One advantage of our model is its conceptual simplicity in its dynamical structure. In particular, the temporal variability is represented by a common factor based on a few classical atmospheric covariates like temperatures, pressures and others. Our inference approach is tested on simulated data and applied on measurements made in the northern part of French Brittany.

Key words: Common factor; Contagion; Precipitation simulators; Spatial-temporal dependence; Weather generators.

2.1 Introduction

Stochastic weather generators (WG) are statistical models that aim at simulating quickly and realistically random sequences of atmospheric variables like temperatures, precipitation and wind speeds. Historically, weather generators started in hydrological sciences in the sixties and seventies. In 1962, Gabriel and Neumann proposed a Markov model for daily precipitation occurrences. Since then, a strong research effort in modeling precipitation distributions has sustained in the hydrological and statistical communities. This attention towards precipitation WGs can be explained by, at least, two different reasons. These stochastic simulations can be used in assessment studies, especially these linked to water resources managements. As one of the initial drivers in impact studies, simulated times series of precipitation can play a fundamental role in exploring some part of the sensitivity of floods, erosion and crops models. The second reason, very attractive to applied statisticians, is that modeling accurately precipitation distributions in space and time has always been an intriguing mathematical challenge. From a probabilistic point of view, a sequence of precipitation mixes two types of events (dry or wet) and the rainfall intensity represents a strong departure from the Gaussian territory (positive, skewed and sometime heavy tailed). Spatially and temporally, events can be strongly correlated

within dry or wet episodes. Still today, the spatio-temporal dynamics of precipitation is difficult to model statistically. Numerous elaborate parametric approaches have been proposed in recent years (see, e.g. Furrer and Katz, 2008; Lennartsson et al., 2008; Vrac et al., 2007; Allard, 2012) (non-parametric approaches also exist but we will not discuss them here).

From a probabilistic point of view, most of daily precipitation generators decouple the occurrence and intensity processes. For example, Kleiber et al. (2012) first generated spatially and temporally correlated rainfall occurrences and then, at locations with positive precipitation, simulated their intensities. This strategy is classical in the WG literature (see, e.g. Katz, 1977; Richardson, 1981). It is mathematically convenient, in terms of inference, to frame the estimation scheme into a clear two step algorithm. But the hypothesis that the occurrence process can be simulated before the intensity process, i.e. independently of past, present or future rainfall amounts, could be challenged, especially at the hourly scale. If very large (small) rainfall amounts have been observed at time t, it seems more likely that a wet (dry) hour will follow at time $t+1$. To quantify this naive reasoning, we need to introduce the example that will lead all our discussions in this article. We will work with winter (DJF) hourly precipitation recorded in the northern part of Brittany in France at three stations from 1995 to 2011, see Figure 2.1. In a nutshell,

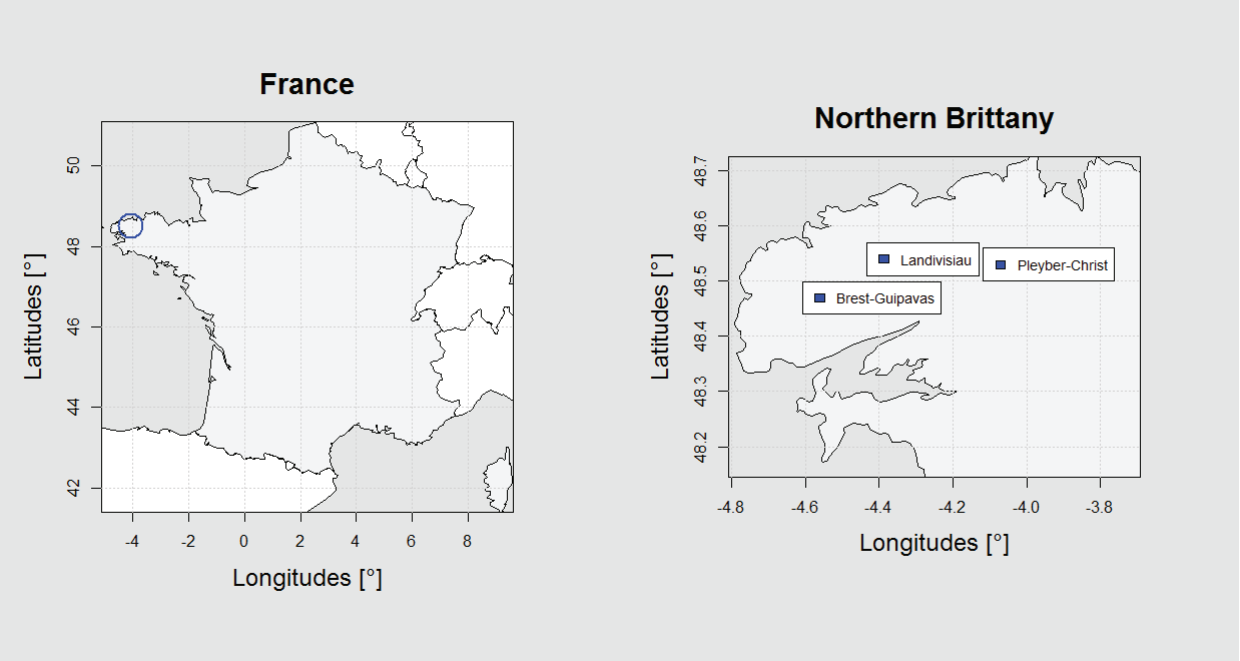


Figure 2.1: Three weather stations in northern Brittany (France) with hourly winter precipitation records from 1995 to 2011.

winter precipitation in this region are generally due to large scale perturbations coming from the Atlantic ocean.

Coming back to the question of simulating the occurrence process independently of the intensity, Table 2.1 shows that the probability of rainfall occurrence at time $t+1$ increases in function of the rainfall intensity at date t for each station. To couple the occurrence and intensity processes, a possible approach, see e.g. Ailliot et al. (2009) and Kleiber et al. (2012) who discussed this issue on p.4, is to first generate a Gaussian random field, say Z_t. Secondly, a censoring mechanism applied to this field produces rainfall

	Weather stations in Fig. 2.1		
Intensity range (mm)	Brest-Guipavas	Landivisiau	Pleyber-Christ
0.2-0.8	0.63	0.65	0.68
0.8-1.4	0.74	0.75	0.79
1.4-2	0.81	0.82	0.82
2-4	0.90	0.91	0.91
> 4	0.94	0.94	0.94
Chi-Square test p-value	8.40×10^{-64}	4.57×10^{-59}	4.39×10^{-43}

Table 2.1: Rain occurrence probability at time $t+1$ given the rainfall intensity at t. Each row represents a specific range of hourly rainfall intensity. The Chi-Square p-values test the independence hypothesis.

occurrences, i.e. wet events occur whenever Z_t is above a given threshold. Then, rainfall intensities are obtained by transforming the Gaussian values above the threshold into positive quantities (e.g., see Allard and Bourotte, 2013). The resulting rainfall intensity distribution depends on the choice of the transform function that may have different flavors (see, e.g. Wilks, 1998; Furrer and Katz, 2008; Lennartsson et al., 2008; Vrac et al., 2007; Allard, 2012). Concerning the correlation structure, the hidden process Z_t, by construction, drives both the intensity and occurrence processes that have become dependent. Another strategy (Serinaldi and Kilsby, 2014) is to directly fit a discrete-continuous at-site mixture distribution with two components (zero or rainfall amount). The parameters of these two components are then modeled by spatio-temporal processes (General Additive Model (GAM) or General Linear Model (GLM)).

This idea of working with an unobserved “seed" process is mathematically appealing but it adds some difficulties. In terms of statistical inference, it makes the estimation difficult. Concerning the interpretation, the latent process and its associated parameters are not necessary simple to explain. Although it has been a classical strategy to introduce hidden quantities in weather generators, for example see the abundant WG literature concerning hidden Markov structures (see reviews like Wilks and Wilby, 1999; Srikanthan et al., 2001; Ailliot et al., 2014), it would be a clear improvement, in terms of interpretation, if multi-site precipitation generators could bypass the hypothesis of hidden processes and directly model raw observations. This implies that additional information characterizing the dynamical structure of the weather system has to be integrated in a different way.

A well-maintained weather station recording hourly precipitation is rarely climatologically isolated in the sense that atmospheric variables like temperatures, pressures or others should be available in the same region. Most likely, the weather station itself may monitor this type of data. In this context, our plan is to develop the following two simple ideas in order to propose a new multi-site precipitation generator:

(A) the main driver of current precipitation at a single location are precipitation recorded one hour early at all sites;

(B) the second source of information comes from atmospheric variables like temperatures and others that jointly impact all sites by dynamically driving the variability of the residuals obtained from step (A).

Although resembling to existing concepts, this two step roadmap differs from the aforementioned studies by a few but important points. We will not represent the precipitation

range by a finite mixture of distributions based on numerous weather types, each one representing different atmospheric conditions. Neither an hidden Gaussian process nor a function to transform a Gaussian variable into a Gamma type rainfall intensity are needed.

Our point (A) is not novel and it corresponds to the well-known idea of an auto-regressive process (see, for example the Markov chain in Katz, 1977). Here, we will use a basic multivariate auto-regressive representation (e.g., see Davis et al., 2014; Grimaldi et al., 2005). Point (B), modeling the residuals as a dynamical function of atmospheric variables, could be considered as more innovative in the WG literature. It is closely linked to the research developed in the statistical downscaling literature (e.g., see Maraun et al., 2010; Wilks, 2010, 2012). Downscaling approaches have mainly focused redon how to make the connection between circulation patterns and local atmospheric variables at the daily scale (e.g., see Vrac and Naveau, 2007; Flecher et al., 2010). Besides a different temporal scale (hourly in our case), our "spatial" component differs from downscaling. Our "common" signal, ordinary atmospheric variables averaged over our three sites in northern Brittany, corresponds to an information shared by all the sites, but it does not represent a regional feature over hundred of kilometers. The common point with downscaling is rather the idea that precipitation simulators performances can be improved if some appropriate explanatory variables can be injected in the statistical model.

The most interesting point of our article resides in combining ideas (A) and (B). From this coupling, a very simple model can handle precipitation in northern Brittany at the hourly scale, not only in terms of moderate rainfalls but also in terms of intense dry periods and even heavy precipitation. Before describing in detail our model, see Section 2.2, we need to say that our modeling strategy has been inspired by tools coming from two research fields not directly linked to geosciences: systemic risk in insurance and epidemic models.

Analyzing systemic risk in the financial system involves modeling the dependence between institutions. A first kind of dependence comes from common risk factors. These are called systematic factors, or frailties. For instance, the risk of life insurance contracts depends on the general uncertain increase of human lifetime, called longevity risk. As highlighted by the regulation since 2008, the second main source of dependence stems from contagion phenomenon between institutions. The contagion effect is due to the interconnections between banks and insurance companies by means of their debt and shares participations. Typically, the failure of a company will imply losses for its lenders, and maybe the failure of some of them. Then failure of some lenders of the lenders can occur, and so on. Originally, contagion models in finance come from the epidemic model introduced by Bailey (1953, 1957), Kendall (1956) and reintroduced in the so-called infectious model used in a static framework by Davis and Lo (2001). Specifications including both frailty and contagion effects are introduced among others in Frey and Backhaus (2003), Giesecke and Weber (2004, 2006), Azizpour et al. (2008), Gagliardini and Gouriéroux (2013).

Coming back to hourly precipitation data, our weather system in Brittany also contains both a common factor (frailty term) and a contagion term. The analogy with the financial system is the following. Each weather station corresponds to an institution. Our common factor represents large scale conditions that are, in some sense, exogenous to our system. For example, a storm front coming from the Altantic ocean is not directly produced by the weather system observed in Brittany. It is similar to an exogenous shock affecting the whole financial system. Then once large scale processes are set then local

physical processes are involved. These processes depend on the local characteristics. For instance, presence of mountains or flows can originate a privilegiate direction for thunderstorms propagation. It is some kind of contagion from one site to another. These local interactions between sites are the equivalent of cross-holdings at the origin of contagion in the financial system. Contrary to the usual assumptions used in finance, the common factor in climatology can be observable, covariates as temperatures, winds, pressures, and so on are also recorded at the weather stations and they represent valuable information that we have to take into account. Another link with financial statistics resides in the assumption that rainfall variability will be modeled by an heteroscedastic variance, a classical feature in econometrics. See e.g. the ARCH/GARCH model literature.

The remaining of the paper is organized as follows. Our heteroscedastic multi-site rainfall generator is detailed in Section 2.2. Then Section 2.3 presents our inference method based on maximum likelihood. Section 2.4 assesses our model's performance on simulations and summarizes our case study results dealing with the northern part of Brittany. Finally a discussion is given in Section 2.5.

2.2 A heteroscedastic multi-site rainfall generator

Denote M the number of weather stations and $P_{m,t}$ the precipitation amount at station m recorded during the t^{th} hour. Our multi-site model based on (A) and (B) is defined as follows

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} & \text{if } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} \geq u, \\ 0 & \text{if } \mathbf{B}_m^{'}\mathbf{P}_{t-1} + \epsilon_{m,t} < u, \end{cases} \tag{2.1}$$

where the scalar product

$$\mathbf{B}_m^{'}\mathbf{P}_{t-1} = \beta_{m,1}P_{1,t-1} + \cdots + \beta_{m,M}P_{M,t-1}$$

between the vector $\mathbf{P}_{t-1} = (P_{1,t-1}, \ldots, P_{M,t-1})^{'}$ and $\mathbf{B}_m = (\beta_{m,1}, \ldots, \beta_{m,M})^{'}$ represents the multivariate auto-regressive vector of order one that captures the dynamical local-scale effect, i.e. how the neighborhood behavior affects station m one hour later. The $M \times M$ unknown auto-regressive coefficients $\beta_{i,j}$ can be concatenated into a matrix, say $\mathbf{B}$, composed of the rows $\mathbf{B}_m^{'}$. For our Brittany example, large scale perturbations coming from the Atlantic ocean should make the matrix $\mathbf{B}$ asymmetric. The Brest weather station should influence the two stations eastward, while the converse should not be true.

The random variable $\epsilon_{m,t}$ in (2.1) corresponds to the local variability driven by a few atmospheric variables and $\epsilon_{m,t}$ is simply modeled by a sequence of zero-mean Gaussian independent random variables. The non-stationarity comes from the standard deviation of $\epsilon_{m,t}$, $\sigma_t > 0$, that varies with time t in a log linear fashion

$$\log \sigma_t = \theta^{'}\mathbf{F}_t, \tag{2.2}$$

where $\mathbf{F}_t$ represents a vector of d atmospheric explanatory variables at time t. The vector of scalars θ corresponds to the unknown regression terms of the log-linear model in Equation (2.2). Concerning the spatio-temporal features of (2.2), neither θ and $\mathbf{F}_t$ change from site to site, they only contain a spatially pooled information shared by all sites, and only $\mathbf{F}_t$ varies with time. With a financial vocabulary, an economist to describe (2.1) and (2.2) will speak of a heteroscedastic model with a volatility driven by the "frailty term" (exogenous common factor) $\mathbf{F}_t$.

To ensure non-negative precipitation and produce dry events, (2.1) contains the condition that rainfall only occur if the quantity $\mathbf{B}_m^{'}\mathbf{P}_{t-1}+\epsilon_{m,t}$ is greater than some positive fixed threshold u. This could be loosely interpreted as saying that the rainfall generated at time t, $\mathbf{B}_m^{'}\mathbf{P}_{t-1}+\epsilon_{m,t}$, has to be large enough to be recorded by the m^{th} weather station. This view emphasizes the fundamental role of having a temporally varying σ_t.

Imagine that all stations are dry at time $t-1$. All elements of the vector $\mathbf{P}_{t-1}$ would then be equal to zero. In this case, the only way to produce large (small) rainfall at time t is to randomly generate a large (small) $\epsilon_{m,t}$. This can be done by choosing a large (small) volatility, σ_t, or equivalently a large (small) frailty term $\mathbf{F}_t$. In other words, large scale conditions are captured by $\mathbf{F}_t$ that drives the volatility of the system, and ultimately drives precipitation occurrences and strengths. This implies that heavy rainfall behavior, as well as dry period persistence, can be reproduced only if an adequate vector $\mathbf{F}_t$ is chosen.

Compared to Kleiber et al. (2012), the dependence between occurrence and intensity is directly built in (2.1). If it had rained a lot at time $t-1$, then $\mathbf{P}_{t-1}$ is large, and it is very likely to have a wet hour at time t because of (2.1). In the opposite case, if $\mathbf{P}_{t-1}$ is equal to zero, it is rather unlikely to have a wet hour at time t, unless the volatility is large. So, current rainfall occurrences are correlated with the intensity value observed one hour early.

Concerning the rain intensity distribution itself, our hypothesis that all $\epsilon_{m,t}$ are normally distributed does not imply that our simulated rainfall will follow a truncated Gaussian density, and consequently this does not mean that heavy tailed behavior cannot be simulated. The non-stationarity of σ_t explains this phenomenon. At any time, σ_t can take a large value and therefore simulated precipitation resembles to a complex infinite mixture of truncated Gaussian random variables with a wide range of standard deviation. A consequence of this is that we do not need, at least in our example, to apply a power transform to go back and forth between the Gaussian world and the rainfall values like in Allard and Bourotte (2013) and Ailliot et al. (2009). Another reason is that $\epsilon_{m,t}$ does not represent raw precipitation but a type of increment between two consecutive hours, see (2.1).

As emphasized in the previous paragraphs, the choice of $\mathbf{F}_t$ is paramount in the overall capacity of our model to accurately simulate hourly precipitation. In order to limit overfitting, the dimension of d should not be too high and, to make our model useful, the type of covariates within $\mathbf{F}_t$ should be easy to obtain in most rainfall applications. For example, the components of $\mathbf{F}_t$ in our northern Brittany case are hourly temperatures, pressures at the sea level, and humidity, respectively. For each atmospheric variable, we spatially average hourly values recorded over our three sites to get $\mathbf{F}_t$. Before closing this section, we would like to emphasize that the choice of the threshold u plays an important role in accurately modeling the length of dry periods. The coming two sections will illustrate this point.

2.3 Inference

Given the vector $\mathbf{F}_t$ and the threshold u, our inference scheme is based on maximizing the likelihood (ML) of model (2.1) with respect to the matrix of auto-regressive parameters $\mathbf{B}$ and the regression coefficients $(\theta_0,\ldots,\theta_d)$ where θ_0 represents the intercept. The log-likelihood function denoted $L_u(\mathbf{B},\theta_0,\ldots,\theta_d)$ for a given u can be written as (see the proof

in the Appendix)

$$L_u(\mathbf{B}, \theta_0, \ldots, \theta_d)$$
$$= \sum_{t=2}^{T} \sum_{m=1}^{M} \left\{ \mathbf{I}_{\{P_{m,t} \geq u\}} \log \left[\frac{1}{\sigma_t} \phi \left(\frac{P_{m,t} - \mathbf{B}_m' \mathbf{P}_{t-1}}{\sigma_t} \right) \right] + \mathbf{I}_{\{P_{m,t}=0\}} \log \left[\Phi \left(\frac{u - \mathbf{B}_m' \mathbf{P}_{t-1}}{\sigma_t} \right) \right] \right\}, \tag{2.3}$$

where $\Phi(.)$ and $\phi(.)$ represent the cumulative distribution function and the probability distribution function of the standardized normal distribution, respectively. The vector $(\theta_0, \ldots, \theta_d)'$ appears throughout σ_t, see (2.2). The indicator function $\mathbf{I}_{\{P_{m,t} \geq u\}}$, equal to one if $P_{m,t} \geq u$ and zero otherwise, corresponds to our condition in (2.1) that generates either a dry or wet hour at station m. These indicator functions and the absence of a closed form for $\Phi(.)$ make the derivation of explicit ML estimates impossible, these values can only be obtained numerically. Our confidence intervals are derived by using a profile log-likelihood approach (Pawitan, 2001).

The above ML approach assumed that the threshold u has been correctly chosen. As our weather station precision is 0.2 mm, the scalar u is iteratively set to small values close to this instrument precision limit, say $0.2, 0.3, 0.4, 0.5, 0.6$ and 0.7. To find the optimal u from this set, we arbitrarily chose a first guest, say 0.5 mm, and implement our ML inference that provides estimates for $(\mathbf{B}, \theta_0, \ldots, \theta_d)$. For this set of ML estimates, we can simulate hourly precipitation for different values of u and then, choose the value of u minimizing the difference between the observed mean length of dry periods and the mean length of dry periods of the simulated precipitation. This process can be repeated with the new value u and we stop when the estimated threshold remains unchanged. To conclude on the choice of u, we note that, by construction, model (2.1) will never simulate values above zero but smaller than u. Hence, if our fixed u is little bit greater than the instrument precision of 0.2 mm, then very small but positive precipitation in Brittany are not considered in our ML optimization scheme, see (2.3). This technical detail does not play an important role in our overall modeling strategy but allowing u to be greater than 0.2 mm improves substantially our capacity to reproduce dry period lengths.

2.4 Application

For our simulations and Brittany example, a learning set of $10,000$ hours and a validation set of $26,816$ hours are used to assess the performance of our estimation scheme. Figure 2.2 summarizes our inference and validation scheme.

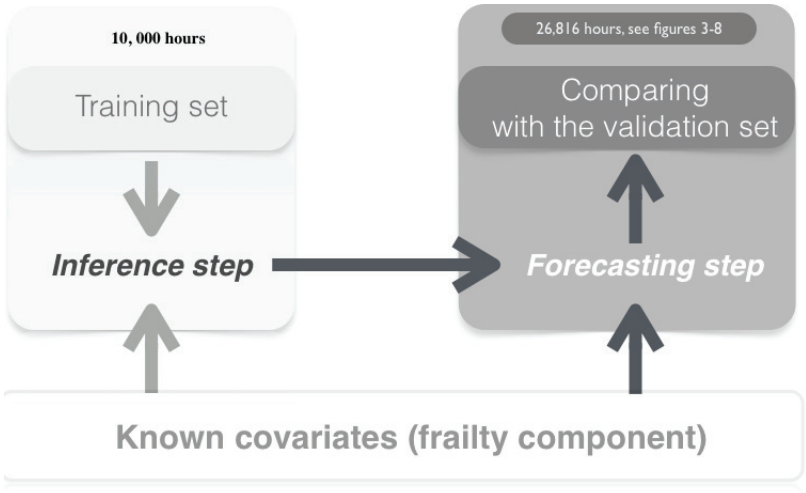


Figure 2.2: Inference and validation scheme: a learning set of $10,000$ hours is used to infer the parameters of our model. Then, assuming that the covariates are known in the future, we predict during the next $26,816$ hours and compare forecasted rainfall (so-called out-of-sample predictions) with precipitation recorded in Brittany. A few visual summaries of this forecast are shown in figures 3-8.

2.4.1 Simulations

To test our inference scheme, we simulate 100 independent replicas for our model (2.1) with parameters that are chosen to be similar to our Brittany example, see the first column of Table 2.2. For two sample sizes (100 and 1000), the parameter estimates are derived by a classical ML approach based on (2.3) and indicates that inferring 13 parameters with a small sample may be done with caution. The empirical mean, standard deviation and relative error derived from our 100 replicas are shown in Table 2.2. As expected, the sample length plays an important role in the inference. For a size of 100, the estimation can be difficult. The corresponding results can seem strange. They are due to the fact that for some simulations, the estimate falls very far away from the true value, deteriorating the global statistics. In contrast, working with a sample size of 1000 provides accurate estimates. In practice, the sample size for hourly data is often long, i.e. more than 1000 hours. For our Brittany example, we have inferred our parameters with $10,000$ observations. In our simulations, the inference accuracy with a sample size of $10,000$ points is slightly better to the one with $1,000$ observations, except for $\theta_0 = 30.63$ which is much improved (the bias and stdev become -0.005 and 0.59, respectively).

2.4.2 Hourly precipitation in northern Brittany

The estimated auto-regressive coefficients β_{ij} of the matrix $\mathbf{B}$ with the 95% confidence intervals and coverage probability are displayed in Table 2.3. The first column makes it clear that the weather station of Brest has a strong influence in the two eastward stations. In particular, the station of Landivisiau has a larger coefficient with Brest (0.47) than with itself (0.25). In contrast, the two stations of Landivisiau and Pleyber-Christ, which are close to each other, do not have an impact on Brest. This pattern corresponds to the expected behavior for this region where rainfall comes from the Atlantic side, see Figure

Sample size	100			1000		
Parameter values	Bias	Stdev	Relative error	Bias	Stdev	Relative error
$\beta_{11} = 0.65$	-2.08	17.71	-3.19	0.00	0.05	0.00
$\beta_{12} = -0.08$	-8.74	22.97	114.37	-0.01	0.06	0.07
$\beta_{13} = 0.11$	-13.30	30.62	-122.71	0.00	0.06	0.03
$\beta_{21} = 0.47$	-3.44	17.94	-7.37	0.01	0.05	0.02
$\beta_{22} = 0.25$	-5.57	13.83	-22.36	0.00	0.04	-0.01
$\beta_{23} = 0.02$	-7.76	20.43	-435.58	0.00	0.06	-0.09
$\beta_{31} = 0.22$	-8.64	27.02	-39.95	0.00	0.05	-0.01
$\beta_{32} = 0.10$	-7.17	21.43	-75.40	0.00	0.05	0.02
$\beta_{33} = 0.36$	-8.18	24.75	-22.65	-0.01	0.05	-0.03
$\theta_0 = 30.63$	-12.36	146.13	-0.40	0.19	1.54	0.01
$\theta_1 = 0.07$	0.07	0.36	0.96	0.00	0.01	-0.01
$\theta_2 = 0.03$	-0.06	0.14	-0.27	-0.07	0.00	0.01
$\theta_3 = 0.03$	0.03	0.08	0.92	0.00	0.00	0.02

Table 2.2: Inference assessment by simulations for two sample sizes of 100 and 1000. The empirical bias, standard deviation and relative error are derived from 100 independent replicas simulated from (2.1) with parameters given in the first column.

2.1. This implies that Brest would be the first one to be impacted by a westerly front, and then the two eastward stations will be hit by the storm later on.

Matrix B	Brest-Guipavas	Landivisiau	Pleyber-Christ
Brest-Guipavas	0.65 [0.59 ; 0.74] (87)	-0.08 [-0.15 ; 0.01] (86)	0.11 [0.02 ; 0.19] (89)
Landivisiau	0.47 [0.41 ; 0.53] (81)	0.25 [0.17 ; 0.32] (92)	0.02 [-0.06 ; 0.09] (91)
Pleyber-Christ	0.22 [0.16 ; 0.27] (77)	0.10 [0.02 ; 0.17] (82)	0.36 [0.30 ; 0.43] (80)

Table 2.3: Estimated auto-regressive coefficients β_{ij} in (2.1) for the three weather stations plotted in Figure 2.1. The intervals represent the 95% confidence intervals and the number between brackets corresponds to the 95% coverage probability.

Concerning the three atmospheric variables in $\mathbf{F}_t$ (temperatures, pressures at the sea level, and humidity) that drive our variability σ_t, Table 2.4 provides the respective estimated regression coefficients from (2.2) with their associated 95% confidence intervals and coverage probability. The values of $\hat{\theta}_j$ for $j = 1, 2, 3$ are small but their 95% confidence intervals do not contain zero. So, we keep these three explanatory variables in our rainfall generator.

To visualize the predictive capacity of our model that has been fitted on the training set, we can generate synthetic hourly precipitation trajectories and compare them to the one kept in our validation period. For each weather station shown in Figure 2.1, Figure 2.3 compares observed rainfall (dark gray) with the simulated one (light gray) during 500 hours of the validation period, see Figure 2.2. Note that we have chosen a relative short period for the sake of visibility. However, the conclusions that can be drawn here are true whatever the period considered in the validation set (figures available upon request).

Atmospheric variables	Estimates	Confidence interval	Coverage probability
Regression intercept	$\hat{\theta}_0 = 30.626$	[28.02 ; 32.32]	80
Temperature	$\hat{\theta}_1 = 0.070$	[0.064; 0.076]	38
Seal level pressure	$\hat{\theta}_2 = -0.034$	[-0.037 ; -0.031]	94
Humidity	$\hat{\theta}_3 = 0.028$	[0.022; 0.036]	94

Table 2.4: Estimated regression coefficients in (2.2) and corresponding 95% confidence intervals in function of our three explanatory atmospheric variables.

In this exercise, we assume that the three atmospheric variables in $\mathbf{F}_t$, temperature, sea level pressure and humidity, are known at the regional scale.

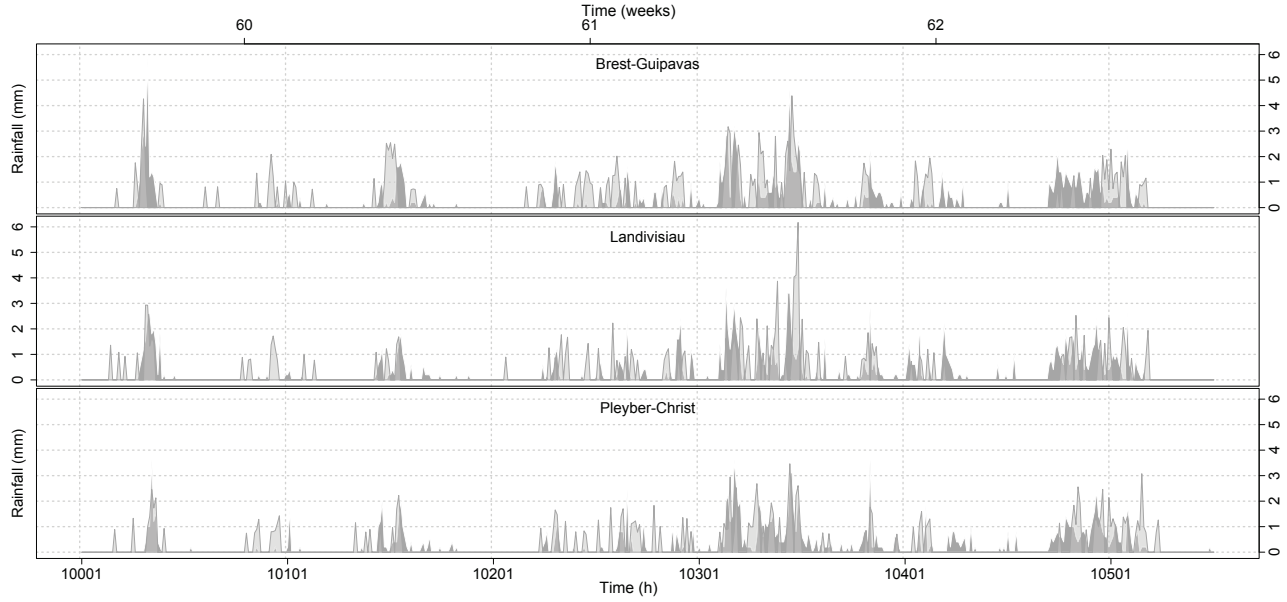


Figure 2.3: Each panel corresponds to one of our three weather stations shown in Figure 2.1. The dark and light gray color represent recorded precipitation and forecasted rainfall, respectively. Our predicted values are obtained on the validation period according to the scheme shown in Figure 2.2. The regional factor σ_t in (2.2) is the only quantity supposed to be known for the forecast. In particular, we do not use the precipitation recorded at time t to forecast the following hour or the following week.

The simulated rainfall appear to reproduce well the dynamical structure of real precipitation in the sense that clusters of simulated rainfall seem to be temporally and spatially synchronized with the observations. This indicates that combining the spatial factor $\mathbf{F}_t$ with a multivariate auto-regressive structure drives accurately the dynamics of the system. Consequently, our model, conditionally on the common factor, cannot only be used as a rainfall generator but it can also carry out predictions. One can also notice in Figure 2.3 that real rainfall (left panel) have a jigsaw pattern. This is due to the recording device that is discrete in nature. Recorded precipitation amounts are added by increment of 0.2 mm (the precision of the instrument). This mechanical feature is not built in our statistical model defined by (2.1) because we view this as an undesirable characteristic of measured precipitation. This discrepancy explains the difference of "granularity" between the two panels. This phenomenon is particularly important for small and medium precip-

itation amounts. To illustrate this, we show the quantile-quantile plot between recorded and simulated rainfall intensities in Figure 2.4. The median and the 98% confidence interval in gray, obtained by parametric bootstrap (over 1000 simulated out-sample series), have been added. Rainfall amounts above 3 mm are well captured by our statistical

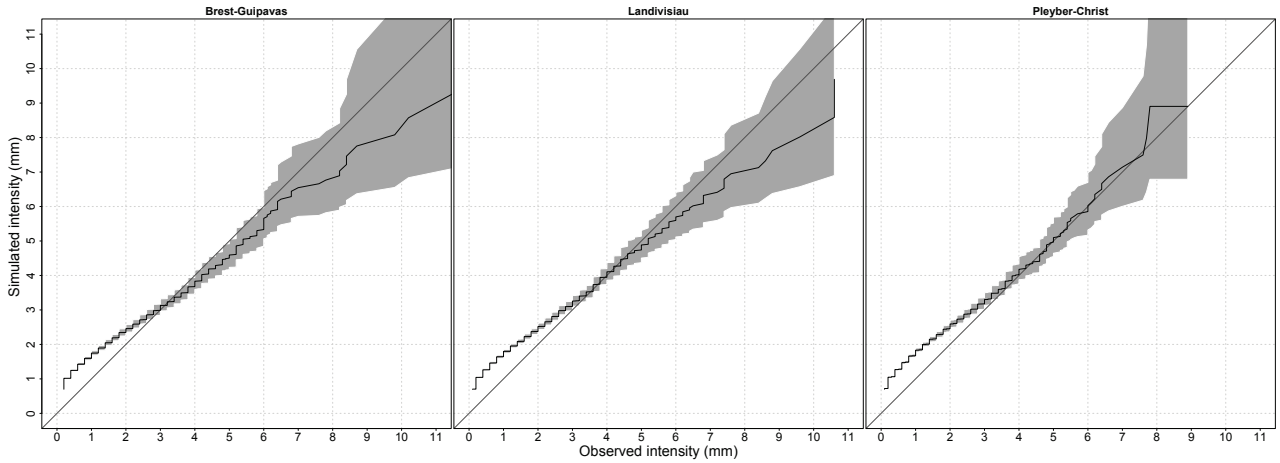


Figure 2.4: Out-sample quantile-quantile plot between observed rainfall amount (x-axis) and simulated one (y-axis) from model (2.1). Each panel corresponds to one of our three weather stations shown in Figure 2.1. The gray color corresponds to the 98% confidence interval and the solid line to the median.

model, this is particularly true for extreme intensities. Precipitation under 3 mm appear to be more difficult to reproduce and their intensities are slightly overestimated. This may be due to the nature of the recording process, the jigsaw pattern due to the instrument precision, and also to be the choice of our threshold u in model (2.1). The latter was optimized ($u = 0.7$ mm) to capture accurately dry episodes. Figure 2.5 shows that this objective has been basically reached. Both long and short dry episodes appear to be

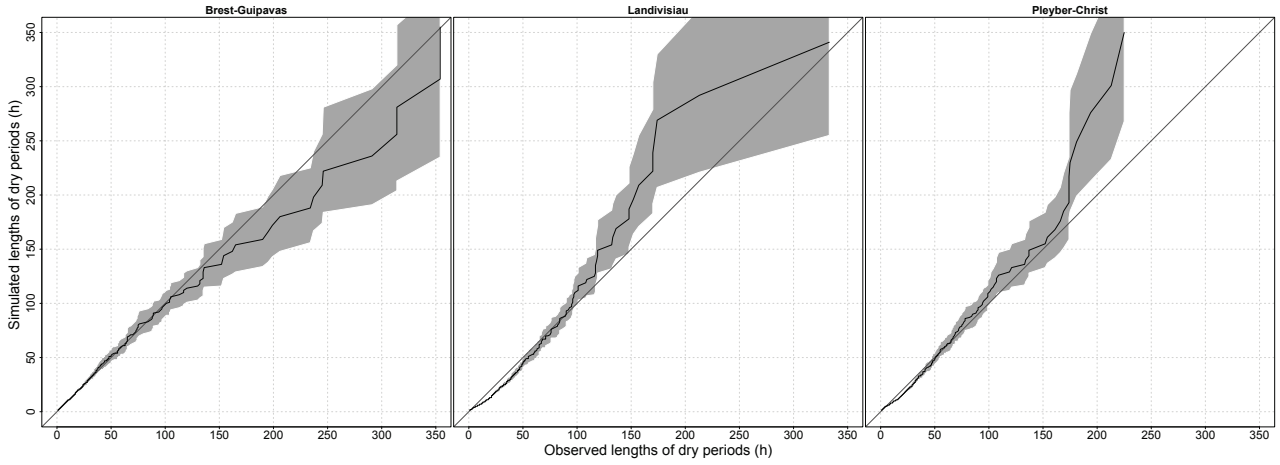


Figure 2.5: Out-sample quantile-quantile plot between observed dry periods length (x-axis) and simulated one (y-axis) from model (2.1). The gray color corresponds to the 98% confidence interval and the solid line to the median.

simulated accurately by our statistical model, especially at the Brest-Guipavas station.

To go one step further in analyzing our predictive out-sample, the probability of having a wet hour given that the preceding hour was also wet is displayed on the left panels of Figure 2.6, as well as the probability of moving from a dry hour to a wet one (right panels). Overall, these transition probabilities appear to give reasonable values, the black point corresponding to the observed estimate and the density to the distribution from simulated out-samples. This is particularly true for the dry to wet transition. Let us note that the other transition probabilities stem directly from these due to the relations $\mathbb{P}(\text{Wet}|\text{Wet}) + \mathbb{P}(\text{Dry}|\text{Wet}) = 1$ and $\mathbb{P}(\text{Wet}|\text{Dry}) + \mathbb{P}(\text{Dry}|\text{Dry}) = 1$.

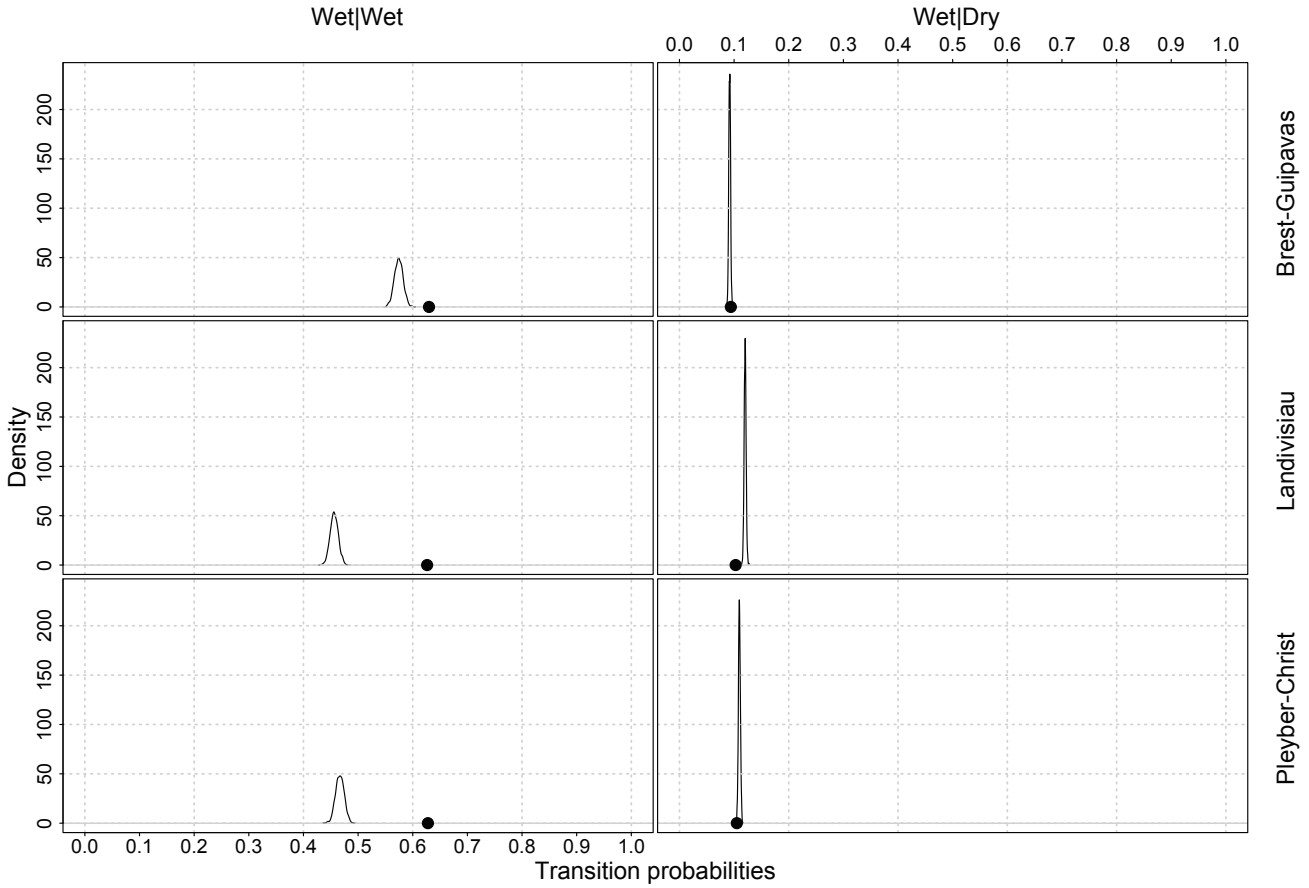


Figure 2.6: Each row represents a weather station. The left and right panels display the out-sample probability of having a wet hour given that the preceding hour was wet (left) and dry (right). The black points represent the observed estimates whereas the curves correspond to the density of estimates over 1000 out-sample simulated series.

Concerning the temporal memory, the top panel of Figure 2.7 shows the out-sample auto-correlations at different time lags of one hour for the Brest-Guipavas station. This graph indicates that this local short term persistence (from one to five hours) is very well reproduced. The lower two panels focus on the out-sample cross-correlations between two pairs of stations, Landivisiau-Brest and Pleyber-Brest, respectively. These correlations are more difficult to reproduce: the observed ones (dotted lines) are above the ones computed on our out-sample. However, we capture the temporal asymmetry. The underestimation of correlations at lag -1, 0 and 1 stems partly from the fact that the $\{\epsilon_{m,t}\}_{m=1,\dots,M}$ are i.i.d. This point will be further discussed in conclusion.

In order to analyze the prediction ability of our model, we have computed the following odd-ratios: the out-sample probability of predicting a dry hour while this hour is indeed

dry and the out-sample probability of predicting a wet hour while this hour is indeed wet. In order to carry out a comparison, we have considered as a reference the simple model that gives as a prediction for a future horizon the value observed now. In Figure 2.8, we observe that from a certain value of the horizon, our model outperforms the naive prediction scheme. In some sense, this test allows to see approximately at which moment the frailty factor takes over the contagion term.

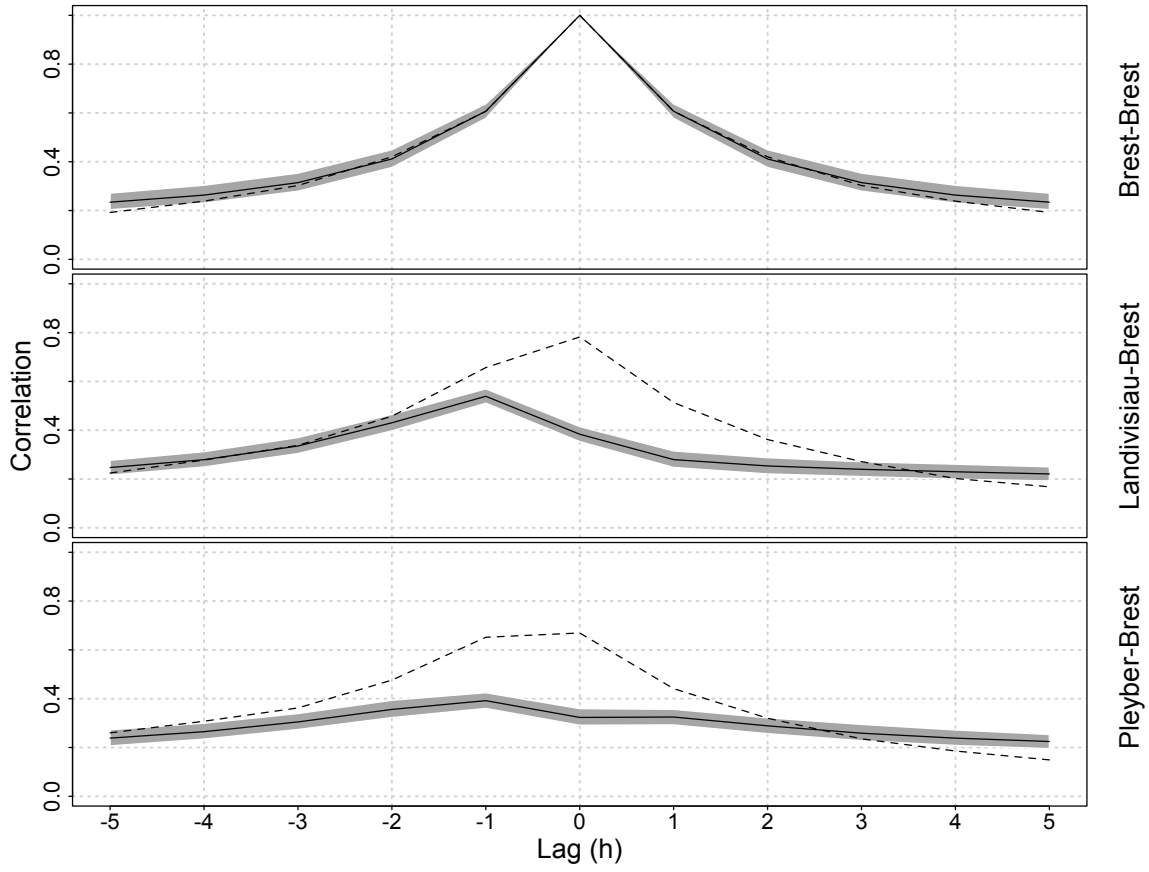


Figure 2.7: These panels correspond to out-sample correlations at different lags. On top, the auto-correlation at Brest-Guipavas. On the middle, the cross-correlation between Landivisiau and Brest-Guipavas and on bottom the cross-correlation between Pleyber-Christ and Brest-Guipavas. The gray color corresponds to the 98% confidence interval, the solid line to the median and the dashed one to the observed correlations.

In our example, the frailty factor plays an important role in reproducing accurately temporal dynamics, rainfall intensity and dry period persistence. To test this (figures available upon request), our model was fitted without the frailty component and the aforementioned features were not adequately reproduced in such an instance. We have also fitted our model without the contagion term (figures available upon request). The performance is quite good, although the persistence of wet periods seems to be underestimated.

We would like to emphasize that figures 2.3-2.8 have been obtained on a validation set, totally different from the training one used to fit the model. This is reassuring and, albeit a few adjustments or extensions, we can expect that the basic ideas developed here could be relevant for other regions. These aspects are discussed in our conclusion.

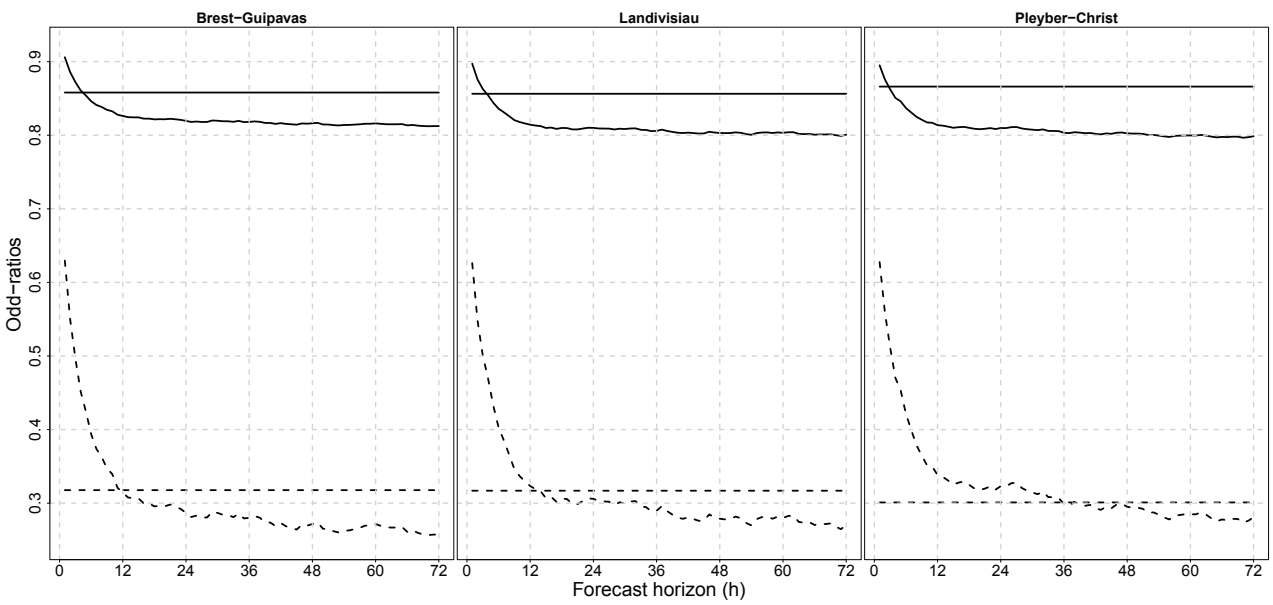


Figure 2.8: Each panel corresponds to one of our three weather stations shown in Figure 2.1. In solid are displayed the estimated probabilities to predict a dry hour while this hour is effectively dry. The constant line corresponds to the estimate obtained with our model. The curve corresponds to the model predicting for a given horizon the same value as now. In dashed, the same concerns the probabilities to predict a wet hour while this hour is effectively wet.

Finally, we have also applied our model on daily precipitations measured at the same stations (see Figure 2.1). Our model's performance is at least as good as in the hourly case. Corresponding figures are available upon request. To give perspective, we compared our model with the standard model by Wilks (1998). Our model gives similar results concerning the general statistical properties (distributions of intensity and dry periods lengths, transition probabilities, cross-correlations, ...). Due to the absence of covariates, Wilks' model does not capture the dynamic structure, and especially the synchronicity between observed and simulated rainfalls. As already mentioned, these features are well reproduced by our model.

2.5 Concluding remarks

Our main objective was to show that a "simple" multivariate auto-regressive model with a heteroscedastic variance driven by a few well-chosen atmospheric covariates can provide an interesting blueprint to generate dry episodes, medium and heavy rainfall. From there, it is easy to extend this work by adding or modifying a few elements of our model. For example, we could imagine to replace the spatially independent random Gaussian noise $\epsilon_{m,t}$ by a multivariate Gaussian vector with a covariance matrix that could represent some spatial dependence among the different weather stations. One can even think about a multivariate student or elliptical distribution in order to provide heavier tails. This may be needed in regions with very heavy rainfall. Our northern Brittany example is known for frequent rainfall episodes but with rather low intensities, say compared to the South of France.

Another possible road for improvements would be to replace the linearity assumption in our regression model (2.2) by a non-parametric relationship (like a GAM, (e.g., Serinaldi and Kilsby, 2014)) that will allow to capture some non-linear behavior. Although it was not really needed for our Brittany example, it is likely that the link between rainfall variability and temperatures could be more complex than our linear model for other regions.

It is also true that our multi-site statistical model is not a pure precipitation weather generator in the sense that we need more than precipitation data to fit and run our model. A few atmospheric covariates are necessary to drive our rainfall variability (for simulation of temperature time series, see e.g. Huong Hoang et al. (2009) and Hoang et al. (2011)). This limitation could be viewed as an advantage in the context of climate change studies. By letting our spatially averaged atmospheric covariates being driven by a numerical model, we could explore how precipitation change under different forcings. This is obviously related to downscaling themes and it would be interesting to pursue this path in future work. In this context, selecting appropriate covariates could be even more relevant, especially for large regions where scale indices like the North Atlantic Oscillation (NAO) Index or the El Nino Southern Oscillation (ENSO) Index could be useful.

One possible drawback of our approach may reside in our small number of sites. Modeling tens or even hundreds of stations, instead of three, will lead to computationally difficulties because of the $M \times M$ size of the matrix $\mathbf{B}$. This could be solved by imposing a parametric structure on the β_{ij}, for example they could decrease with the distance with respect to other stations or even set to zero for far apart stations. Finding a suitable distance is not trivial and should depend on orographic and other physical features.

To conclude, there are many different ways to fine tune and extend our approach. Basically, this will strongly depend on the application at hand and more research is needed to clearly see if this framework can be generalized.

2.6 Appendix: Likelihood computation

Proof. We consider that the common factor's path $\mathbf{F}_t$ $(t = 1, \dots, T)$ is given. Therefore we omit $\mathbf{F}_t$ in the following. Let us denote by I_{t-1} the information available at time $t-1$, $I_{t-1} = (P_{1,t-1}, \dots, P_{M,t-1}, P_{1,t-2}, \dots, P_{M,t-2}, \dots)$ and by h the density function. We can write by using a sequential argument that

$$h(P_{1,t}, \dots, P_{M,t}\ ; t \in 1, \dots, T) = \prod_{t=2}^{T} h(P_{1,t}, \dots, P_{M,t} | P_{1,t-1}, \dots, P_{M,t-1}). \tag{2.4}$$

Since the $\epsilon_{m,t}, m = 1, \dots, M$ are i.i.d. (conditionally on $\mathbf{F}_t$), the variables

$$\mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t}, m = 1, \dots, M$$

are independent. Thus it is the same for the variables

$$\mathbf{I}_{\mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} > u}, m = 1, \dots, M$$

and therefore for the product

$$P_{m,t} = \left(\mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} \right) \mathbf{I}_{\mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} > u}, m = 1, \dots, M.$$

Therefore the measurement equation given by model (2.1) gives that conditionally on (I_{t-1}), the variables $P_{1,t}, \dots, P_{M,t}$ are independent, yielding

$$h(P_{1,t}, \dots, P_{M,t} | P_{1,t-1}, \dots, P_{M,t-1}) = \prod_{m=1}^{M} h(P_{m,t} | P_{1,t-1}, \dots, P_{M,t-1}). \tag{2.5}$$

Combining (2.4) and (2.5), we obtain

$$h(P_{1,t}, \dots, P_{M,t}; t \in 1, \dots, T) = \prod_{t=2}^{T} \prod_{m=1}^{M} h(P_{m,t} | P_{1,t-1}, \dots, P_{M,t-1}). \tag{2.6}$$

Let us recall that

$$P_{m,t} = \begin{cases} \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} & \text{if } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} \geq u, \\ 0 & \text{if } \mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} < u. \end{cases}$$

Due to this threshold mechanism, $P_{m,t}$ is a mixture of a discrete random variable and a continuous one. Let us denote by h_d the density of the discrete part and by h_c the density of the continuous one. We have

$$hc(P_{m,t} | P_{1,t-1}, \dots, P_{M,t-1}) = \frac{1}{\sigma_t} \phi \left(\frac{P_{m,t} - \mathbf{B}_m^{'} \mathbf{P}_{t-1}}{\sigma_t} \right)$$

and

$$\begin{aligned} hd(P_{m,t} | P_{1,t-1}, \dots, P_{M,t-1}) &= \mathbb{P} \left[P_{m,t} = 0 \ | P_{1,t-1}, \dots, P_{M,t-1} \right] \\ &= \mathbb{P} \left[\mathbf{B}_m^{'} \mathbf{P}_{t-1} + \epsilon_{m,t} < u \right] \\ &= \mathbb{P} \left[\epsilon_{m,t} < u - \mathbf{B}_m^{'} \mathbf{P}_{t-1} \right] \\ &= \Phi \left(\frac{u - \mathbf{B}_m^{'} \mathbf{P}_{t-1}}{\sigma_t} \right). \end{aligned}$$

Using (2.6), the log-density is written

$$\begin{aligned}
&\log\left[h(P_{1,t},\ldots,P_{M,t},t=1,\ldots,T)\right]\\
&=\sum_{t=2}^{T}\sum_{m=1}^{M}\log\left[hc(P_{m,t}|P_{1,t-1},\ldots,P_{M,t-1})\ \mathbf{I}_{\{P_{m,t}\geq u\}}+hd(P_{m,t}|P_{1,t-1},\ldots,P_{M,t-1})\ \mathbf{I}_{\{P_{m,t}=0\}}\right]\\
&=\sum_{t=2}^{T}\sum_{m=1}^{M}\mathbf{I}_{\{P_{m,t}\geq u\}}\log\left[hc(P_{m,t}|P_{1,t-1},\ldots,P_{M,t-1})\right]+\mathbf{I}_{\{P_{m,t}=0\}}\ \log\left[hd(P_{m,t}|P_{1,t-1},\ldots,P_{M,t-1})\right].
\end{aligned}$$

That finally yields the log-likelihood function

$$\begin{aligned}
&L_u(\mathbf{B},\theta_0,\ldots,\theta_d)\\
&=\sum_{t=2}^{T}\sum_{m=1}^{M}\left\{\mathbf{I}_{\{P_{m,t}\geq u\}}\log\left[\frac{1}{\sigma_t}\phi\left(\frac{P_{m,t}-\mathbf{B}_m'\mathbf{P}_{t-1}}{\sigma_t}\right)\right]+\mathbf{I}_{\{P_{m,t}=0\}}\log\left[\Phi\left(\frac{u-\mathbf{B}_m'\mathbf{P}_{t-1}}{\sigma_t}\right)\right]\right\},
\end{aligned}$$

completing the proof. □

Bibliography

Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2014). Stochastic weather generators: an overview of weather type models. *Submitted*.

Ailliot, P., Thompson, C., and Thomson, P. (2009). Space time modeling of precipitation using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 58(3):405–426.

Allard, D. (2012). Modeling spatial and spatio-temporal non Gaussian processes. In Porcu, E., Montero, J.-M., and Schlather, M., editors, *Advances and Challenges in Space-time Modelling of Natural Events*, Lecture Notes in Statistics, pages 141–164. Springer Berlin Heidelberg.

Allard, D. and Bourotte, M. (2013). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *Stochastic Environmental Research and Risk Assessment*, Submitted.

Azizpour, S., Giesecke, K., et al. (2008). Self-exciting corporate defaults: contagion vs. frailty. *Manuscript, Stanford University*.

Bailey, N. T. (1953). The total size of a general stochastic epidemic. *Biometrika*, 40(2):177–185.

Bailey, N. T. (1957). *The mathematical theory of epidemics*. Griffin London.

Davis, M. and Lo, V. (2001). Modelling default correlation in bond portfolios. *Mastering risk*, 2:141–151.

Davis, R. A., Zang, P., and Zheng, T. (2014). Sparse vector autoregressive modeling. *Submitted*.

Flecher, C., Naveau, P., Allard, D., and Brisson, N. (2010). A stochastic daily weather generator for skewed data. *Water Resources Research*, 46(7).

Frey, R. and Backhaus, J. (2003). Interacting defaults and counterparty risk: a Markovian approach. *University of Leipzig, Leipzig, Germany*.

Furrer, E. and Katz, R. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, 44(12).

Gabriel, K. and Neumann, J. (1962). A Markov chain model for rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88:90–95.

Gagliardini, P. and Gouriéroux, C. (2013). Correlated risks vs contagion in stochastic transition models. *Journal of Economic Dynamics and Control*, 37(11):2241–2269.

Giesecke, K. and Weber, S. (2004). Cyclical correlations, credit contagion, and portfolio losses. *Journal of Banking & Finance*, 28(12):3009–3036.

Giesecke, K. and Weber, S. (2006). Credit contagion and aggregate losses. *Journal of Economic Dynamics and Control*, 30(5):741–767.

Grimaldi, S., Serinaldi, F., and Tallerini, C. (2005). Multivariate linear parametric models applied to daily rainfall time series. *Advances in Geosciences*, 2(2):87–92.

Hoang, T., Dacunha-Castelle, D., and Benmenzer, G. (2011). Estimation of a diffusion model with trends taking into account the extremes. Application to temperature in France. *Environmetrics*, 22(3):464–479.

Huong Hoang, T., Parey, S., and Dacunha-Castelle, D. (2009). Multidimensional trends: The example of temperature. *The European Physical Journal-Special Topics*, 174(1):113–124.

Katz, R. (1977). Precipitation as a chain-dependant process. *Journal of Applied Meteorology*, 16:671–676.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 149–165.

Kleiber, W., Katz, R., and Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Ressources Research*, 48(1).

Lennartsson, J., Baxevani, A., and Chen, D. (2008). Modelling precipitation in Sweden using multiple step Markov chains and a composite model. *Journal of Hydrology*, 363(1):42–59.

Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3).

Pawitan, Y. (2001). *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press.

Richardson, C. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Serinaldi, F. and Kilsby, C. G. (2014). Simulating daily rainfall fields over large areas for collective risk estimation. *Journal of Hydrology*, 512:285–302.

Srikanthan, R., McMahon, T., et al. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences Discussions*, 5(4):653–670.

Vrac, M. and Naveau, P. (2007). Stochastic downscaling of precipitation: from dry events to heavy rainfalls. *Water Resources Research*, 43(7).

Vrac, M., Stein, M., and Hayhoe, K. (2007). Statistical downscaling of precipitation through non homogeneous stochastic weather typing. *Climate Research*, 34:169–184.

Wilks, D. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191.

Wilks, D. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 1(6):898–907.

Wilks, D. (2012). Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3):267–278.

Wilks, D. and Wilby, R. (1999). The weather generation game : a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357.

Chapter 3

Estimation of max-stable processes by simulated maximum likelihood

Max-stable processes are very appropriate for the statistical modeling of spatial extremes. Nevertheless their inference is difficult. Indeed, the multivariate density function is not available and thus standard likelihood-based estimation methods cannot be applied. The commonly used method - based on composite likelihood - is flexible and requires a relatively low computational cost. However, it leads to non efficient estimators. In this study an approach based on nonparametric simulated maximum likelihood is developed. We take advantage of the possibility of simulating many max-stable models and propose to approximate the multivariate density (in space) using kernel methods. Our estimator is efficient when both the temporal dimension and the number of simulations tend towards infinity. This approach can be used for many subclasses of max-stable processes and provides better results than composite-based methods, especially in the case where only a few temporal observations of the process are available and the spatial dependence is high. However, due to the curse of dimensionality, the observation sites have to be separated into subgroups when they are too numerous. Finally, the methodology is examined on simulated data.

Key words: Extreme value theory; Max-stable processes; Nonparametric maximum simulated likelihood estimator; Spatial extremes; Spatial dependence.

3.1 Introduction

In the context of climate change, extreme events tend to be more frequent. Moreover, a constant growth of population and wealth is observed and the penetration rate of insurance is increasing. As a consequence, 2011 holds the record in what concerns both economic losses and insured losses. Both the insurance and reinsurance industry are very sensitive to natural disasters (see e.g. Embrechts et al. (1997)). In order to quantify their impacts, a first step is to characterize the behavior of the maxima of the relevant environmental variables at each point of the region under study. Due to the natural spatial extent of environmental variables, max-stable processes (de Haan, 1984; de Haan and Pickands, 1986; Resnick, 1987) are ideally suited for modeling purposes. Indeed the distribution of the spatial process resulting from taking the temporal maxima (renormalized) at each point of the space is necessarily max-stable, when the maximum is taken over an infinite number of i.i.d. observations (de Haan, 1984). Max-stable processes are

a natural extension of multivariate extreme value distribution to the level of stochastic processes. Their interest, compared with componentwise maxima, lies especially in the possibility of spatial aggregation. Max-stable processes completely account for the spatial dependence structure and provide a methodology to discern spatial patterns.

The seminal paper by de Haan (1984) provides spectral representations of max-stable processes. These probabilistic constructions allow for the building of some relevant subclasses of max-stable processes: the Smith model (Smith, 1990), the Schlather model (Schlather, 2002), the Brown-Resnick model (Kabluchko et al., 2009). These different models offer a large variety of spatial dependence structures and have become very useful for applications.

However, the inference of max-stable processes is very difficult. Indeed, since the multivariate density function is not available in dimension higher than 2 or 3, standard likelihood-based estimation methods cannot be applied. Mainly two methodologies of inference can be found in the literature. In both cases, asymptotic results involve a number of temporal observations tending towards infinity, whereas the number of sites is fixed. The first approach (Smith, 1990) is based on least squares and aims at minimizing the error between the estimated extremal coefficient and the analytical one. However, this approach suffers from two drawbacks. As typically observations are far from being Gaussian, the least squares estimator is not efficient in the sense given by the Cramér-Rao lower-bound criterion. Moreover, one needs a preliminary transformation to standard Fréchet marginals. Thus the practical use of this methodology involves two separate steps.

A more flexible and more accurate approach is based on the maximization of the composite likelihood (Lindsay, 1988; Varin and Vidoni, 2005). The density function of the Schlather model is only known in the bivariate case whereas the density functions of the Smith and Brown-Resnick models are known in the trivariate case. Padoan et al. (2010) and Genton et al. (2011) respectively implement pairwise and triplewise likelihood estimation in the context of the Smith process. Huser and Davison (2013) carry out a triplewise estimation for the Brown-Resnick process. The composite likelihood estimator is consistent and asymptotically normal but not efficient (in the sense of the Cramér-Rao bound) as the temporal dimension tends towards infinity. Indeed spatial dependence is in part ignored, implying that the method cannot completely take advantage of a high number of observation sites. Therefore an accurate estimation requires a high number of temporal observations of the maxima. Such a long historical sample is often unavailable in applications - especially in climatology- whereas observation stations can be numerous.

In this paper, we propose to compute the log-likelihood function by taking advantage of the possibility of simulating max-stable random fields. Schlather (2002) and Oesting et al. (2012) propose accurate algorithms, respectively in the cases of the Smith and the Schlather models, and the Brown-Resnick model. Inspired by Fermanian and Salanié (2004), we approximate the unknown multivariate density using kernel methods. Silverman (1986) and Scott (1992) show that kernel density estimation can be a useful tool in the multivariate case. However, in high dimension, some practical challenges must be overcome and will be discussed in details in the following. Our estimator is derived by maximizing the non parametric simulated maximum likelihood function. This approach is general since it can be used as soon as the process can be simulated.

The idea of using simulated likelihood methodologies in econometrics goes back to Lerman and Manski (1981), when the technique was applied to estimate choice probabilities. Simulation-based inferences techniques have then been widely studied in econometrics. For a general overview, see e.g. Gouriéroux and Monfort (1997). Such estimation

methods can be split in three general classes. The first one gathers methods that are general-purpose but not asymptotically efficient, even as the number of simulation draws tends towards infinity fast enough. The methods of simulated moments (McFadden, 1989; Pakes and Pollard, 1989), simulated pseudo-maximum likelihood (Laroque and Salanié, 1994) and indirect inference (Gouriéroux et al., 1993; Smith, 1993) all belong to this category. Methods of the second type rely on simulating the likelihood (see e.g. Lee (1995)) or the score function (see e.g. Hajivassiliou and McFadden (1998)) using Monte-Carlo techniques. However, these methods can only be used in particular cases since the likelihood usually cannot be written as a function of mathematical expectations. Especially such methodologies cannot be applied easily in the case of max-stable processes. The third category consists of asymptotically efficient and general-purpose methods. The method of efficient moments (Gallant and Tauchen, 1996) and the non parametric approach by Fermanian and Salanié (2004) are among these. The first one involves the issue of the choice of moments. In this paper we adapt the approach by Fermanian and Salanié (2004) to the case of max-stable processes. To the best of our knowledge, it is the first time a simulation-based methodology is used in the context of extremes.

The remaining of the article is organized as follows. Section 2 is a brief reminder of key results concerning max-stable processes. In Section 3, the estimation method is described and the main theoretical results are stated for the Smith process. We show especially that our estimator is consistent, asymptotically normal and efficient as both the temporal dimension and the number of simulations tend towards infinity. Section 4 deals in details with practical implementation challenges stemming from estimation in high dimension. In Section 5, we use simulated data to compare our method's performance with that of the pairwise methodology. Our estimator performs better, in particular when only a few temporal observations of the process are available and when the spatial dependence is high. Section 6 concludes. All proofs are gathered in Appendix.

3.2 Max-stable processes, inference and simulation

3.2.1 Definition and representation

A stochastic process $G(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$ with continuous sample paths is said max-stable if there exist $\{a_T(\mathbf{x})\}$ and $\{b_T(\mathbf{x}) > 0\}$ such that if $G_1(\mathbf{x}), \ldots, G_T(\mathbf{x})$ are i.i.d. replications of $G(\mathbf{x})$:

$$\frac{\max_{t=1}^{T} G_t(.) - b_T(.)}{a_T(.)} \stackrel{d}{=} G(.). \tag{3.1}$$

As a direct consequence the one-dimensional marginal distributions are max-stable and then belong to the class of generalized extreme value distributions: $\forall \mathbf{x} \in \mathbb{R}^d \; G(\mathbf{x}) \sim GEV\left(\mu(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x})\right)$ where $\mu(\mathbf{x})$, $\sigma(\mathbf{x})$ and $\xi(\mathbf{x})$ respectively denote the location, scale and shape parameters at location $\mathbf{x}$. In the following, we consider standard Fréchet max-stable processes (denoted by Z), i.e. satisfying $\mathbb{P}(Z(\mathbf{x}) \leq z) = \exp\left(-\frac{1}{z}\right), \forall \mathbf{x} \in \mathbb{R}^d$ and $\forall z > 0$, implying that $a_T(\mathbf{x}) = T$ and $b_T(\mathbf{x}) = 0$. Hence (3.1) becomes $T^{-1} \max_{t=1,\ldots,T} Z_t(.) \stackrel{d}{=} Z$. A process with standard Fréchet margins will be referred to as a simple max-stable process in the following.

The definition of max-stability provides a well-defined class of processes but it does not suggest how to construct and simulate them. A probabilistic construction of max-stable

processes is given by de Haan (1984) via the spectral representation. A second representation is due to Schlather (2002).We denote by $C^+(\mathbb{R}^d)$ the set of positive processes having continuous sample paths on $\mathbb{R}^d$.

First representation (de Haan, 1984):
Let $\{(\xi_i, \mathbf{c_i})\}_{i\geq 1}$ be the points of a Poisson point process on $(0, +\infty) \times \mathbb{R}^d$ with intensity measure $d\Lambda(\xi, \mathbf{c}) = \xi^{-2} d\xi \nu(d\mathbf{c})$, where ν is a σ-finite measure on $\mathbb{R}^d$. If $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a simple max-stable process in $C^+(\mathbb{R}^d)$, then there exists a family of continuous non negative functions $\{f_{\mathbf{x}}(\mathbf{c}) : \mathbf{c} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d\}$ satisfying

- for each $\mathbf{x} \in \mathbb{R}^d$, $\displaystyle\int_{\mathbb{R}^d} f_{\mathbf{x}}(\mathbf{c})\ \nu(d\mathbf{c}) = 1$,
- for each compact $K \subset \mathbb{R}$, $\displaystyle\int_{\mathbb{R}^d} \sup_{\mathbf{x}\in K} f_{\mathbf{x}}(\mathbf{c})\ \nu(d\mathbf{c}) < +\infty$,

such that

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \stackrel{d}{=} \{\max_{i\geq 1} \xi_i f_{\mathbf{x}}(\mathbf{c_i})\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{3.2}$$

Conversely, each process defined by the right-hand side of (3.2) is a simple max-stable process.

The proof of the direct statement can be found in de Haan (1984) for $\mathbf{x} \in \mathbb{R}_+$ and $\mathbf{c} \in [0, 1]$ as well as in de Haan and Ferreira (2007). In the latter case, the proof is given for $\mathbf{x} \in \mathbb{R}$, $\mathbf{c} \in\ [0, 1]$ and ν being the Lebesgue measure. The converse statement is not difficult to prove.

Second representation (Penrose, 1992; Schlather, 2002; de Haan and Ferreira, 2007):
Let $\{\xi_i\}_{i\geq 1}$ be the points of a Poisson point process on $(0, +\infty)$, with intensity $d\Lambda(\xi) = \xi^{-2} d\xi$ and $Y_1(\mathbf{x}), Y_2(\mathbf{x}), \ldots$ i.i.d. replications of a stochastic process $\{Y(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ such that $\mathbb{E}(\max[0, Y(\mathbf{x})]) = 1$ for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbb{E}(\sup_{\mathbf{x}\in K} \max[0, Y(\mathbf{x})]) < +\infty$ for any compact $K \subset \mathbb{R}^d$. Then any simple max-stable process in $C^+(\mathbb{R}^d)$ can be written as

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \stackrel{d}{=} \{\max_{i\geq 1} \{\xi_i \max[0, Y_i(\mathbf{x})]\}\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{3.3}$$

Conversely, each process defined by the right-hand side of (3.3) is a simple max-stable process.

Remark 3.1. *If Y is a strictly stationary process, then* (3.3) *defines a strictly stationary process.*

In the paper by Schlather (2002), we find the converse statement. For a proof of the fact that any simple max-stable process in $C^+(\mathbb{R}^d)$ can be written as in (3.3), we refer to de Haan and Ferreira (2007). Note that de Haan and Ferreira (2007) consider $[0, 1]$ instead of $\mathbb{R}^d$ for convenience.

3.2.2 Different classes of max-stable processes

The Smith model:
In an unpublished manuscript, Smith (1990) uses the spectral representation (3.2) to provide a parametric model for spatial extremes. He considers a particular setting where

ν is the Lebesgue measure on $\mathbb{R}^d$ and $f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma_0}}(\mathbf{x}-\mathbf{c})$ with $f_{\mathbf{\Sigma_0}}$ the density of a d-variate normal law with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Sigma_0}$:

$$f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma_0}}(\mathbf{x}-\mathbf{c}) = (2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma_0}|^{-\frac{d}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{c})'\mathbf{\Sigma_0}^{-1}(\mathbf{c}-\mathbf{x})\right]. \tag{3.4}$$

The parameter is the covariance matrix $\mathbf{\Sigma}$, which contains all the information about the spatial dependence structure. A nice feature of this model lies in its nice interpretation in terms of rainfall-storm process (Smith, 1990), the shape of these storms being driven by the covariance matrix. Moreover, in the case $d = 2$, the trivariate density (the density of an observation at 3 sites) can be explicitly written (see e.g. Genton et al. (2011)) contrary to the Schlather model, presented immediately hereafter.

The Schlather model:
Schlather (2002) proposes to set $Y(\mathbf{x}) = \sqrt{2\pi}\ \epsilon(\mathbf{x})$ in (3.3), where $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a stationary standard Gaussian process with any correlation function $\rho(.)$. All correlation functions stemming from the geostatistical literature can be used, allowing for a rich diversity of behaviors. In that paper, the following correlation families will be considered and compared:

Whittle-Matern:	$\rho(h) = \frac{2^{1-c_2}}{\Gamma(c_2)}\left(\frac{h}{c_1}\right)^{c_2} K_{c_2}\left(\frac{h}{c_1}\right),\ c_1 > 0,\ c_2 > 0,$
Exponential:	$\rho(h) = \exp\left[-\frac{h}{c_1}\right],\ c_1 > 0,$
Cauchy:	$\rho(h) = \left[1+\left(\frac{h}{c_1}\right)^2\right]^{-c_2},\ c_1 > 0,\ c_2 > 0,$
Powered ewponential:	$\rho(h) = \exp\left[-\left(\frac{h}{c_1}\right)^{c_2}\right],\ c_1 > 0,\ 0 < c_2 < 2,$

where c_1 and c_2 are the range and the smooth parameters, Γ is the Gamma function and K_{c_2} is the modified Bessel function of the third kind with order c_2.

The geometric Gaussian model:
Independence is unreachable in the case of the Schlather model. To deal with this issue, Davison (2003) introduces the geometric Gaussian model. He takes in (3.3) $\{Y(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ a log normal process and not a Gaussian process:

$$Y(\mathbf{x}) = \exp\left(\sigma\epsilon(\mathbf{x}) - \frac{\sigma^2}{2}\right),$$

where $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a standard Gaussian process with variance σ^2 and correlation function $\rho(.)$.

The Brown-Resnick model:
The geometric Gaussian process is a particular case of a model introduced by Brown and Resnick (1977). Kabluchko et al. (2009) introduce a generalization of the latter model, which is referred to as the Brown-Resnick model, by taking in (3.3):

$$Y(\mathbf{x}) = \exp\left(W(\mathbf{x}) - \frac{\sigma^2(\mathbf{x})}{2}\right),$$

where $\{W(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a zero-mean Gaussian process with stationary increments and $\sigma^2(\mathbf{x}) = \mathrm{Var}[W(\mathbf{x})],\ \forall\mathbf{x}\in\mathbb{R}^d$. The process W and therefore the resulting Brown-Resnick

process are completely characterized by the variance $\sigma(\mathbf{x})$ and the semi-variogram, defined by

$$\gamma(\mathbf{h}) = \frac{1}{2}\text{Var}[W(\mathbf{x}+\mathbf{h}) - W(\mathbf{x})], \forall \mathbf{h} \in \mathbb{R}^d, \tag{3.5}$$

where Var stands for the variance.

3.2.3 Composite likelihood-based inference

Consider a parametric statistical model with density function $\{l(\mathbf{z}; \boldsymbol{\Sigma}) : \mathbf{z} \in \mathcal{Z}, \boldsymbol{\Sigma} \in \Theta\}$, where $\mathcal{Z} \subseteq \mathbb{R}^M, \Theta \subseteq \mathbb{R}^q, M \geq 1$ and $q \geq 1$. Consider a set of events $\{A_i \in \mathcal{F}, i \in I \subseteq \mathbb{N}\}$ where $\mathcal{F}$ is a σ algebra on $\mathcal{Z}$. The weighted composite log-likelihood (WCL) (Lindsay, 1988) is defined by

$$L_C(\boldsymbol{\Sigma}) = \sum_{i \in I} w_i \log l(\mathbf{z} \in A_i \, ; \boldsymbol{\Sigma}),$$

where $l(\mathbf{z} \in A_i; \boldsymbol{\Sigma}) = l(\{z_m \in \mathcal{Z} : z_m \in A_i\} \, ; \boldsymbol{\Sigma})$, $\mathbf{z} = (z_1, \ldots, z_M)'$ and $\{w_i, i \in I\}$ is a set of weights. Then it is natural to consider the WCL estimator as the global maximizer of L_C.

Padoan et al. (2010) consider pairs of sites and define the composite pairwise log-likelihood

$$L_{CP}(\boldsymbol{\Sigma}) = \sum_{t=1}^{T} \sum_{m_1 < m_2} w_{m_1,m_2} \log \left[l(\mathbf{z}_t(\mathbf{x}_{m_1}), \mathbf{z}_t(\mathbf{x}_{m_2}); \boldsymbol{\Sigma}) \right],$$

where the $\mathbf{z}_t, t = 1, \ldots, T$, are observations of the process. Genton et al. (2011) and Huser and Davison (2013) consider a triplewise likelihood respectively in the cases of the Smith process and the Brown-Resnick process.

If the $\mathbf{Z}_t, t = 1, \ldots, T$, are i.i.d replications, then under the usual regularity conditions, the corresponding estimator is strongly consistent and asymptotically normal as $T \to +\infty$ and M is fixed (Lindsay, 1988; Varin and Vidoni, 2005):

$$\sqrt{T}(\hat{\boldsymbol{\Sigma}}_{\mathbf{C}} - \boldsymbol{\Sigma}_{\mathbf{0}}) \xrightarrow[T \to \infty]{d} N(\mathbf{0}, \tilde{\boldsymbol{\Omega}}^{-1}),$$

where the information matrix is $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_{\mathbf{d}} \, \boldsymbol{\Omega}^{-1} \, \boldsymbol{\Omega}_{\mathbf{d}}$, with

$$\boldsymbol{\Omega} = V_{\boldsymbol{\Sigma}_{\mathbf{0}}} \left[\frac{\partial \log L_C(\boldsymbol{\Sigma}_{\mathbf{0}})}{\partial \boldsymbol{\Sigma}} \right] \text{ and } \boldsymbol{\Omega}_{\mathbf{d}} = \mathbb{E}_{\boldsymbol{\Sigma}_{\mathbf{0}}} \left[-\frac{\partial^2 \log L_C(\boldsymbol{\Sigma}_{\mathbf{0}})}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'} \right].$$

If equal weights are by far the most widely used, efficiency gains can be obtained by using appropriate unequal weights. Sang and Genton (2013) propose the tapered composite likelihood. Weights are equal to 1 for pairs (or triplets) of sites whose distance is lower than the so-called taper range γ_s, and equal to 0 otherwise. The optimal taper range is determined by minimizing either the determinant or the trace of $\tilde{\boldsymbol{\Omega}}$.

The main advantages of these composite-based methods lie in their flexibility and relatively low computation time. However, the pairs or triplets are considered as independent, leading to a spatial deterioration. There is mispecification, $\boldsymbol{\Omega}_{\mathbf{d}} \neq \boldsymbol{\Omega}$ and the Cramér-Rao efficiency bound is not reached.[1] In practice, composite-based approaches require a high number of temporal observations to lead to accurate results, especially when the process under study has a high spatial correlation. The method proposed in that paper can provide more accurate estimates but at the price of a higher computing time.

[1] apart from the exotic case of a completely independent spatial process.

3.2.4 Simulation

Adequate simulation algorithms have been proposed for max-stable processes. See Schlather (2002) for the Smith and the Schlather process and Oesting et al. (2012) for the Brown-Resnick process. de Haan's representation involves the maximum over an infinite number of copies of a random function but for simulation purpose the number of simulations is necessarily finite. However, Schlather (2002) shows that it is possible under certain conditions to get exact simulations with a finite sampling size. If these conditions are not met, the approximation is still good.

3.3 Estimation by simulated maximum likelihood

3.3.1 The non parametric maximum likelihood estimator

Assume that we have T temporal observations $\mathbf{z}_1, \ldots, \mathbf{z}_T$ of the logarithm of a parametric max-stable process (e.g. the Smith, Schlather or Brown-Resnick model) at M sites. The corresponding parameters to estimate are gathered in the matrix $\mathbf{\Sigma_0}$ whose dimensions depend on the max-stable model considered. So $\forall t$, $\mathbf{z}_t$ is a vector of dimension M: $(z_t^1, \ldots, z_t^M)$. The associated log-likelihood is

$$L_T(\mathbf{\Sigma}) = \frac{1}{T} \sum_{t=1}^{T} \log l(\mathbf{z}_t, \mathbf{\Sigma}),$$

denoting $l(\mathbf{z}_t, \mathbf{\Sigma})$ the density of $\mathbf{z}_t$.

As we already mentioned, the multivariate density function is generally not available, making impossible to compute the maximum likelihood estimator $\tilde{\mathbf{\Sigma}}_\mathbf{T}$. Since many max-stable processes can be simulated, the idea is to approximate the unknown density by a multivariate kernel estimator based on S i.i.d. simulations $\mathbf{z}_s^S(\mathbf{\Sigma}), s = 1, \ldots, S$. The simulated density function l^S is then written

$$l^S(\mathbf{z}_t, \mathbf{\Sigma}) = \frac{1}{S} \sum_{s=1}^{S} K_\mathbf{H}(\mathbf{z}_t - \mathbf{z}_s^S(\mathbf{\Sigma})),$$

where $K_\mathbf{H}$ is a kernel function associated with the bandwidth matrix $\mathbf{H}$ of dimension $M \times M$. To simplify the presentation we consider that $\mathbf{H} = h\ \mathbf{I}$, h being a bandwidth such that $h \to 0$ as $S \to \infty$ and $\mathbf{I}$ the identity matrix.

The simulated log-likelihood is written

$$\tilde{L}_T^S(\mathbf{\Sigma}) = \frac{1}{T} \sum_{t=1}^{T} \log l^S(\mathbf{z}_t, \mathbf{\Sigma}).$$

However, since the logarithm has an infinite derivative in 0, estimation errors on small values of the density are strongly amplified by the logarithmic transformation. Therefore the smallest values of $l^S(\mathbf{Z}_t, \mathbf{\Sigma})$ have to be trimmed. Finally the nonparametric simulated log-likelihood is defined as (Fermanian and Salanié, 2004)

$$L_T^S(\mathbf{\Sigma}) = \sum_{t=1}^{T} \tau_S[l^S(\mathbf{z}_t, \mathbf{\Sigma})] \log l^S(\mathbf{z}_t, \mathbf{\Sigma}), \tag{3.6}$$

where function $\tau_S(.)$ verifies for $\delta \geq 0$

$$\tau_S(u) = \begin{cases} 0 & \text{if } |u| < h^\delta, \\ \text{sufficiently regular} & \text{if } |u| \in [h^\delta, 2h^\delta], \\ 1 & \text{if } |u| > 2h^\delta. \end{cases}$$

The size of the "penalty frame" $[-h^\delta, h^\delta]$ is decreasing for h decreasing, i.e. for S increasing. This is logical since S increasing corresponds to a better precision.
Finally, the Non Parametric Simulated Maximum Likelihood (NPSML) estimator is defined as

$$\hat{\mathbf{\Sigma}}_T^S = \arg\max_{\mathbf{\Sigma}} L_T^S(\mathbf{\Sigma}). \tag{3.7}$$

3.3.2 Consistency and asymptotic normality

In the following we focus on the Smith process, denoted by $Z(\mathbf{x})$. The true covariance matrix is denoted $\mathbf{\Sigma_0}$. Assume that $\mathbf{z}_1, \ldots, \mathbf{z}_T$ are observations of the logarithm of the Smith model. According to (3.2) and (3.4), the process is written

$$Z(\mathbf{x}) = \log \left[\max_i \xi_i \ f_{\mathbf{\Sigma_0}}(\mathbf{x} - \mathbf{c}_i) \right] = \log \xi_{n_\mathbf{x}} \ + \log f_{\mathbf{\Sigma_0}}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}),$$

where $n_\mathbf{x}$ is the almost everywhere (a.e.) unique random index realizing the maximum at point $\mathbf{x}$. Note that $\mathbf{\Sigma_0}$ is symmetric and therefore contains only 3 distinct parameters in the case $d = 2$. We consider that $\mathbf{\Sigma_0}$ belongs to a compact set $\mathcal{K} \subset Sdp_{2\times 2}$ where $Sdp_{2\times 2}$ designs the set of symmetric positive definite matrices of dimension $(2, 2)$.

Let us just summarize the main notations: T is the number of temporal observations, t the t-th observation year, M the number of observation sites and S the number of simulation draws.

We study the properties of our estimator where T tends towards infinity whereas the number M of observation sites is fixed. S tends naturally towards infinity. Recall that $\mathbf{H} = h\ \mathbf{I}$, h being a bandwidth such that $h \to 0$ as $S \to \infty$. In the following, h must be understood as a function of S. Before providing the main results, we state some assumptions and lemmas. Fermanian and Salanié (2004) consider a general econometric model whose reduced form is written $\mathbf{y} = g(\mathbf{v}, \theta, \epsilon)$, where g is a twice differentiable function with respect to $\mathbf{v}$ and θ, $\mathbf{y}$ the endogenous variable to be explained, $\mathbf{v}$ the exogenous variable and ϵ the model noise. They provide asymptotic results about the NPSML estimator, under a high number of assumptions both on the kernel and the bandwidth and on model g. The lemmas we propose further check the assumptions relative to the model, in the particular case of the Smith process. The M sites $\mathbf{x}_1, \ldots, \mathbf{x_M}$ being fixed, there is no exogenous variable in our case. The Smith model can be written

$$\begin{pmatrix} Z(\mathbf{x}_1) \\ \vdots \\ Z(\mathbf{x_M}) \end{pmatrix} = \begin{pmatrix} \log \xi_{n_{x_1}} \ + \log f_{\mathbf{\Sigma_0}}(\mathbf{x}_1 - \mathbf{c}_{\mathbf{n_{x_1}}}) \\ \vdots \\ \log \xi_{n_{x_M}} \ + \log f_{\mathbf{\Sigma_0}}(\mathbf{x_M} - \mathbf{c}_{\mathbf{n_{x_M}}}) \end{pmatrix} = g(\mathbf{\Sigma_0}, \epsilon), \tag{3.8}$$

where $\epsilon = (\xi_{n_{\mathbf{x}_m}}, \mathbf{c}_{n_{\mathbf{x}_m}}, m = 1, \ldots, M)$.

In order to use the results by Fermanian and Salanié (2004), it is necessary to check that function g is twice differentiable with respect to $\mathbf{\Sigma}$. This result is precisely stated in Proposition 3.1.

Proposition 3.1. *For a fixed realization ω (fixed $(\xi_i, \mathbf{c_i}, i \geq 1)$), there exists a negligible set $\mathcal{N} \subset \mathbb{R}^2$ such that for all $(\mathbf{x_1}, \ldots, \mathbf{x_M}) \in \{\mathbb{R}^2/\mathcal{N}\}^M$, there exists a neighborhood $V_0(\mathbf{x_1}, \ldots, \mathbf{x_M}; \omega)$ of $\mathbf{\Sigma_0}$ in which the function g is twice differentiable with respect to $\mathbf{\Sigma}$.*

Proposition 3.1 states that the probability that the observation locations $\mathbf{x}_1, \ldots, \mathbf{x_M}$ belong to the set of points at which g is not twice differentiable is equal to zero.

The three next assumptions concern the choice of the kernel and the bandwidth.

Assumption 3.1. *The kernel K is twice continuously differentiable. Denote by ρ the kernel order.*

Assumption 3.1 mainly allows differentiating the simulated log-likelihood. The proof of asymptotic normality is based on a Taylor development of the simulated log-likelihood at the second order and then requires the existence of its second derivative.

Assumption 3.2. *Let us recall that h must be understood as a function of S tending to 0 as S tends to infinity. Then we assume that*

$$\delta < \rho \quad \text{and} \quad \lim_{S\to\infty} \frac{Sh^{M+2\delta}}{\log S} = +\infty. \tag{3.9}$$

The condition $\delta < \rho$ is natural. Accuracy is an increasing function of the order ρ. Thus, if a higher order is considered, smaller density values can be accepted. Then the trimming frame has to be smaller, corresponding to a higher δ. Trimming aside, Assumption 3.2 is the equivalent of the usual assumption ensuring convergence of the multivariate kernel density estimator, $\lim_{S\to\infty} Sh^M = +\infty$.

Assumption 3.3. *We assume that*

$$\lim_{S,T\to\infty} T^{\frac{1}{2}} h^{\rho-\delta} \log h = 0 \tag{3.10}$$

and

$$\lim_{S,T\to\infty} Th^{-2(M+\delta+1)} \log^2 h \frac{\log S}{S} = 0. \tag{3.11}$$

Assumptions 3.2 and 3.3 impose conditions on the rate of convergence of the bandwidth h to zero as S and T tend towards infinity.

The three next assumptions are related to the real underlying densities and log-likelihoods.

Assumption 3.4. *Denote by $Vech$ the half-vectorization operator converting a matrix into a vector containing its unique elements. The maximum likelihood estimator $\tilde{\mathbf{\Sigma}}_\mathbf{T}$ is consistent, asymptotically normal and asymptotically efficient. The true parameter $\mathbf{\Sigma_0}$ belongs to the interior of $\mathcal{K}$. More precisely, we assume that for some positive definite matrix $\mathbf{\Omega}$,*

$$-\frac{\partial^2 L_T}{\partial Vech(\mathbf{\Sigma})\partial Vech(\mathbf{\Sigma})'}(\mathbf{\Sigma}^*) \xrightarrow[T\to\infty]{P} \mathbf{\Omega}$$

uniformly with respect to $\mathbf{\Sigma}^*$ in a neighborhood of $\mathbf{\Sigma_0}$, and that

$$T^{\frac{1}{2}} \frac{\partial L_T}{\partial Vech(\mathbf{\Sigma})}(\mathbf{\Sigma_0}) \xrightarrow[T\to\infty]{P} \mathcal{N}(0, \mathbf{\Omega}).$$

Assumption 3.4 represents classical conditions ensuring consistency, asymptotic normality and efficiency of the true maximum likelihood estimator.

Assumption 3.5. *The functions* $l(\mathbf{z}, \mathbf{\Sigma})$ *and* $\dfrac{\partial^\rho l(\mathbf{z}, \mathbf{\Sigma})}{\partial \mathbf{z}^\rho}$ *are bounded above on* $\mathbb{R}^M \times \mathcal{K}$. *Moreover* $\exists \beta > 1$ *such that almost surely (a.s.),*

$$\frac{1}{T}\sum_{t=1}^{T} |\log l(\mathbf{Z}_t, \mathbf{\Sigma})|^\beta \quad \textit{is convergent uniformly with respect to } \mathbf{\Sigma} \in \mathcal{K}.$$

$$\textit{Moreover, } \mathbb{E}\left[\sup_{\mathbf{\Sigma}\in\mathcal{K}} \left\| \frac{\partial l(\mathbf{Z}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}} \right\|\right] < \infty. \tag{3.12}$$

Assumption 3.6. *There exist* $\gamma > 1$ *and* $\gamma' > 1$ *such that*

$$\mathbb{E}\left[\sup_{\mathbf{\Sigma}\in\mathcal{K}} \left\| \frac{\partial \log l(\mathbf{Z}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}} \right\|^\gamma\right] < \infty \textit{ and } \mathbb{E}\left[\sup_{\mathbf{\Sigma}\in\mathcal{K}} \left\| \frac{\partial l(\mathbf{Z}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}} \right\|^{\gamma'}\right] < \infty.$$

Moreover, the function

$$\frac{\partial^{\rho+1} l(\mathbf{z}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma} \partial \mathbf{z}^\rho} \quad \textit{is bounded over in } R^M \times \mathcal{K}.$$

Assumptions 3.5 and 3.6 say that the true multivariate density is regular, i.e. has no exotic form.

Assumption 3.7. *Let us assume that*

$$\lim_{S,T\to\infty} \left(T^{\frac{1}{2}} |\log h|\right)^{\frac{\gamma}{\gamma-1}} \mathbb{P}\left(\inf_{\mathbf{\Sigma}\in V_0} l(\mathbf{Z}, \mathbf{\Sigma}) \leq 2h^\delta\right) = 0.$$

Assumption 3.7 aims at controlling the frequency of trimming. Clearly it holds when h tends towards zero fast enough.

As in Fermanian and Salanié (2004), we assume that S is a power of T and that h is a power of S, i.e. $S = [C_1 T^a]$ and $h = [C_2 S^{-b}]$ with $a, b, C_1, C_2 > 0$, where $[\,]$ designs the floor function. We have the following result.

Proposition 3.2. *Assumptions 3.2, 3.3 and 3.7 respectively impose*

$$b < \frac{1}{M + 2\delta},$$

$$ab > \frac{1}{2(\rho - \delta)} \quad \textit{and} \quad 1 + 2ab(M + \delta + 1) < a,$$

$$ab > \frac{\gamma}{\delta(\gamma - 1)}.$$

Note that the asymptotically optimal bandwidth for kernel estimation corresponds to $b = \frac{1}{M+2\rho}$. By choosing, $\rho > \delta$, this bandwidth satisfies the condition of Assumption 3.2: $b < \frac{1}{M+2\delta}$.

Moreover, note that conditions of Proposition 3.2 are compatible. Choose the kernel order $\rho > \delta$ and some $C > \max\left(\dfrac{1}{2(\rho-\delta)}; \dfrac{\gamma}{\delta(\gamma-1)}\right)$. Choosing a and b on the hyperbola

$ab = C$ in the zone of large a and small b allows satisfying the 3 implications. Thus our conditions define a non-degenerate region in the plan (a, b). Moreover this region intersects the line $b = \frac{1}{M+2\rho}$ - corresponding to the optimal bandwidth - if $\rho - \delta$ is large enough. This can imply using higher-order kernels ($\rho > 2$).

We now state the aforementioned lemmas.

Lemma 3.1. *We have*

$$\mathbb{E}\left[\sup_{\mathbf{\Sigma}\in V_0}\left\|\frac{\partial g}{\partial \mathbf{\Sigma}}(\mathbf{\Sigma},\epsilon)\right\|\right] < +\infty,$$

where the neighborhood V_0 has been defined in Proposition 3.1.

Lemma 3.2. *i) $\exists p_0 \in \mathbb{R} > 4$ such that*

$$\mathbb{E}\left[\left(\sup_{\mathbf{\Sigma}\in V_0}\left\|\frac{\partial g}{\partial \mathbf{\Sigma}}(\mathbf{\Sigma},\epsilon)\right\|\right)^{p_0}\right] < +\infty.$$

ii) Moreover,

$$\mathbb{E}\left[\sup_{\mathbf{\Sigma}\in V_0}\left\|\frac{\partial^2 g}{\partial^2 \mathbf{\Sigma}}(\mathbf{\Sigma},\epsilon)\right\|\right] < +\infty.$$

iii) Finally, for some $\zeta > 1$,

$$\mathbb{E}\left[\sup_{\mathbf{\Sigma}\in V_0}\left\|\frac{\partial g}{\partial \mathbf{\Sigma}}\right\|^{\zeta}\right] < +\infty.$$

Lemma 3.3. *There exists a positive $\nu > 0$ such that*

$$\lim_{S\to+\infty} \mathbb{P}(\|\mathbf{Z}\| > S^{\nu}) \log h = 0.$$

Lemma 3.4. *For some $\nu > 0$,*

$$\left[T^{\frac{\gamma}{2\gamma-1}} + \left(\frac{T^{\frac{1}{2}}}{h^{\delta}}\right)^{\frac{\gamma'}{\gamma'-1}} + \left(\frac{T^{\frac{1}{2}}|\log h|}{h^{M+1+\delta}}\right)^{\frac{\zeta}{\zeta-1}}\right] \mathbb{P}(\|\mathbf{Z}\| > S^{\nu})$$

tends to zero as S and T tend to infinity, where γ and γ' (resp. ζ) are introduced in Assumption 3.6 and Lemma 3.2.

We can now state the two main results of this paper, giving strong properties of the non parametric maximum likelihood estimator in the case of the Smith process.

Theorem 3.1. *Under Assumptions 3.1, 3.2, 3.4 and 3.5, $\hat{\mathbf{\Sigma}}_T^S$ is strongly consistent. Almost surely,*

$$\lim_{S,T\to\infty} \hat{\mathbf{\Sigma}}_T^S = \mathbf{\Sigma_0}.$$

Theorem 3.2. *Under Assumptions 3.1- 3.7, $\hat{\mathbf{\Sigma}}_T^S$ is asymptotically normal and asymptotically efficient:*

$$\sqrt{T}(\hat{\mathbf{\Sigma}}_T^S - \mathbf{\Sigma_0}) \xrightarrow[S,T\to\infty]{d} N(0, \mathbf{\Omega}),$$

where $\mathbf{\Omega}$ is the asymptotic covariance matrix of the exact maximum likelihood estimator.

Remark 3.2. *Note that the convergence rate of S to infinity is irrelevant for consistency.*

3.4 Practical implementation

3.4.1 Kernel and bandwidth selection

The performance is not very sensitive to the kernel's choice (Wand and Jones, 1993; Duong and Hazelton, 2005). The latter must only be sufficiently regular. In practical applications, a Gaussian kernel (with covariance matrix being the identity) is used:

$$K(\mathbf{z}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{z}'\ \mathbf{z}\right),\ \forall \mathbf{z} \in \mathbb{R}^M \tag{3.13}$$

A kernel estimator of the multivariate density is used. Rapidly computed and high-quality density estimates are available in the univariate case (Silverman, 1986; Hall and Marron, 1988; Wand et al., 1991). But considering the curse of dimensionality, if the kernel choice has no real importance, a good choice of the bandwidth matrix is crucial (Wand and Jones, 1993; Duong and Hazelton, 2005). This bandwidth choice is indeed the main issue of implementation of our method. In this paper, the high dimension considered (number of sites M) as well as the challenging multivariate density shape of max-stable processes lead us to use a variable bandwidth, more precisely the k-nearest neighbor estimator (Loftsgaarden and Quesenberry, 1965). Adaptive kernel estimators (with variable bandwidths) indeed alleviate the lack of local adaptivity of kernels as the Gaussian one (Botev et al., 2010). Moreover, in high dimension ($M \geq 4$), variable bandwidths perform better than fixed ones (Terrell and Scott, 1992). A detailed discussion about different bandwidth choices is provided in Appendix 3.8.

3.4.2 A method to manage the curse of dimensionality

Due to the curse of dimensionality, density estimation in dimension higher than 6 or 7 requires in practice too large a number of simulation draws S (e.g. $S >> 10^6$ or 10^7). In order to circumvent this issue, our idea is to consider densities of lower dimension by preliminary using a clustering algorithm. We use the Partitioning Around Medoids (PAM) algorithm by Kaufman and Rousseeuw (1987). Like the K-means, it is a partitional algorithm breaking the dataset into groups by minimizing a specific error related to some distance. However, in the case of K-means, each cluster is represented by an artificially created entity (e.g. the barycenter) - the centroid- , while in the case of PAM, the representative is a dataset's member - the medoid. The algorithm aims at minimizing the overall dissimilarity between the representatives of each cluster and the cluster's members.

We then add the density contributions of all clusters (composite approach), assuming independence between clusters. Thus, in order to minimize the mispecification, it is natural to consider clusters that are as independent as possible. The distance used for the clustering should therefore be representative of the spatial dependence in extremes. An adequate choice is the F-madogram (Cooley et al., 2006; Bernard et al., 2013), defined as

$$d(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2}\mathbb{E}\left[|F(Z(\mathbf{x}_1)) - F(Z(\mathbf{x}_2))|\right], \tag{3.14}$$

where $Z(.)$ is a stationary max-stable random field with unit Fréchet margins and $F(z) = \exp\left(-\frac{1}{z}\right)$. Indeed, it is an extension of the variogram (Cressie and Cassie, 1993)

that suits particularly the extremes. The relation between this distance and the extremal coefficient is given by

$$d(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2}\left(\frac{\theta(\mathbf{x}_1 - \mathbf{x}_2) - 1}{\theta(\mathbf{x}_1 - \mathbf{x}_2) + 1}\right), \tag{3.15}$$

where the extremal coefficient (Schlather and Tawn, 2003) is defined by

$$\mathbb{P}[Z(\mathbf{x}_1) \leq z, Z(\mathbf{x}_2) \leq z] = \exp\left(\frac{-\theta(\mathbf{x}_1 - \mathbf{x}_2)}{z}\right)$$

and represents the number of independent sites. In the case of the Smith model, $\theta(\mathbf{x}_1 - \mathbf{x}_2) = 2\Phi\left(\frac{a(\mathbf{x}_1, \mathbf{x}_2)}{2}\right)$, where $a(\mathbf{x}_1, \mathbf{x}_2)$ is the euclidean distance for the scalar product associated to $\mathbf{\Sigma_0}$, defined by $a(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)' {\mathbf{\Sigma_0}}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}$. Using (3.15), we obtain

$$d(\mathbf{x}_1, \mathbf{x}_2) = \frac{2\Phi\left(\frac{a(\mathbf{x}_1, \mathbf{x}_2)}{2}\right) - 1}{4\Phi\left(\frac{a(\mathbf{x}_1, \mathbf{x}_2)}{2}\right) + 2}, \tag{3.16}$$

showing that the F-madogram distance is strictly increasing with respect to the euclidean distance and thus that the F-madogram distance is in one-to-one relationship with the euclidean distance. Thus in theory it is equivalent to apply the PAM using the F-madogram distance and using the modified euclidean distance. However, the euclidean distance a is unknown since $\mathbf{\Sigma_0}$ is unknown. A solution is to estimate the F-madogram. Equation (3.14) directly suggests the following estimator:

$$\hat{d}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2T}\sum_{t=1}^{T} |F(z_t(\mathbf{x}_1)) - F(z_t(\mathbf{x}_2))|, \tag{3.17}$$

where $z_t(\mathbf{x}_1)$ and $z_t(\mathbf{x}_2)$ are the t-th observations of the random field at locations $\mathbf{x}_1$ and $\mathbf{x}_2$.[2] Note that a sufficient number of temporal observations T is required to get good results i.e. to obtain the same clustering using the analytic F-madogram and its estimate.

In applications, clusters containing up to around 7 locations can be built. Note that the clusters' size can be controlled by imposing the number of clusters. The simulated composite likelihood is defined as

$$LC_T^S(\mathbf{\Sigma}) = \sum_{t=1}^{T}\sum_{c=1}^{C} w_c \; \tau_S[l^S(\mathbf{z}_{t,c}, \mathbf{\Sigma})] \;\; \log[l^S(\mathbf{z}_{t,c}, \mathbf{\Sigma})], \tag{3.18}$$

where C is the number of clusters, $\mathbf{z}_{t,c}$ the vector containing the values of the process at locations in the c-th cluster and w_c is the weight attributed to the c-th cluster. Then the Maximum Simulated Composite Likelihood (MSCL) estimator is computed.

Compared to the pairwise and triplewise log-likelihood-based approaches, this method has the advantage to partially relax the assumption of spatial independence. Indeed, in our case, the independence between clusters is based on a real dependence measure and the cluster size is allowed to be much higher than 2 or 3.

[2] If isotropy is assumed, then it might be better to use a "binned" version of this estimator (Cooley et al., 2006).

Remark 3.3. *Note that the computational cost is considerably reduced in comparison with a situation where all groups of a given size would be considered (e.g. all groups containing 6 locations).*

Finally, the clustering can be based on the euclidean distance a, using the analytical formula of the F-madogram given by (3.16). Then the F-madogram does not need to be estimated. Nevertheless, since this distance is unknown, an iterative procedure must be carried out. At first, a can be computed assuming $\mathbf{\Sigma_0} = \mathbf{I}$. After clustering, a first estimate of $\mathbf{\Sigma_0}$ is obtained using the MSCL approach. Thus a new value of a is obtained and a more accurate clustering can be performed, yielding a second estimate of $\mathbf{\Sigma_0}$. This can be repeated until stabilization of $\mathbf{\Sigma_0}$. Note that this approach is highly time-consuming.

3.5 Results on simulated data

In this section, we assess the performance of our methodology on simulated data. Our estimator is compared to the pairwise likelihood-based one (Padoan et al., 2010). The M site locations are uniformly generated over a 40×40 region. We simulate a Smith max-stable random field[3] whose covariance matrix is

$$\mathbf{\Sigma_0} = \begin{pmatrix} 200 & 150 \\ 150 & 300 \end{pmatrix},$$

and take the logarithm. We consider the same covariance matrix as in Padoan et al. (2010). The estimator performance is assessed for different values of T and M. Our routines are written in R and C++ using the interface Rcpp. The use of C++ speeds up the algorithm by a factor 60.

In all following experiments, the number of simulation draws S is equal to 10^6 and the number of neighbors used to compute the bandwidth is $k = 5$. The statistics (average bias, average relative error and standard deviation) presented in the following tables are computed over 20 replications. The Nelder-Mead algorithm has been used for the optimization.

True	**Simulated likelihood**			**Pairwise likelihood**		
	Bias	Relative error	Stdev	Bias	Relative error	Stdev
Cov11=200	44.0	0.22	95.8	3480579.2	17402.9	15305579.7
Cov22=300	13.8	0.05	51.6	197631.8	658.8	807980.5
Cov12=150	-38.4	-0.26	79.4	815456.1	5436.4	3515785.2

Table 3.1: Comparison of the performance of the simulated likelihood/pairwise likelihood in the case $T = 5$ and $M = 5$.

As shown in Table 3.1, the pairwise likelihood estimator performs very badly when the number of temporal observations T and the number of sites M are low. Indeed, the estimator falls very often very far from the true value, leading to the explosion of the bias

[3]using the SpatialExtremes R package by Ribatet and Padoan (2008).

and the standard deviation. The simulated likelihood-based estimator performs much better. As expected, when T increases, the performance of both methods is improved, both in terms of bias and standard deviation. As shown in Table 3.2 for $T = 30$ and $M = 5$, the simulated likelihood still seems to perform better but composite-based estimates are much more reliable than in the previous case.

True	**Simulated likelihood**			**Pairwise likelihood**		
	Bias	Relative error	Stdev	Bias	Relative error	Stdev
Cov11=200	11.7	0.06	22.5	11.8	0.06	61.8
Cov22=300	13.7	0.05	20.0	44.3	0.15	182.9
Cov12=150	-2.8	-0.02	12.7	13.1	0.09	109.3

Table 3.2: Comparison of the performance of the simulated likelihood/pairwise likelihood in the case $T = 30$ and $M = 5$.

Table 3.3 has been obtained using the PAM algorithm with a number of clusters $C = 6$. The simulated likelihood performs much better than the pairwise one, both in terms of bias and standard deviation. Our estimator takes better advantage of a high number of observation sites, since the assumption of spatial independence is relaxed compared to the pairwise-based one.

True	**Simulated likelihood**			**Pairwise likelihood**		
	Bias	Relative error	Stdev	Bias	Relative error	Stdev
Cov11=200	27.4	0.14	20.8	531.4	2.66	667.5
Cov22=300	17.4	0.06	29.3	1001.2	3.34	931.8
Cov12=150	-20.2	-0.13	25.7	418.1	2.79	698.9

Table 3.3: Comparison of the performance of the simulated likelihood/pairwise likelihood in the case $T = 1$ and $M = 30$.

Remark 3.4. *The simulated log-likelihood is sometimes not very smooth. A solution can be to use the Accelerated Random Search (ARS)(Appel et al., 2004) to find a good initial point and then to carry out the optimization using the Nelder-Mead algorithm.*

3.6 Conclusion

This paper proposes a new estimator for max-stable processes, that can be applied to many classes of max-stable processes. The precision level can be controlled via the number of simulation draws S. Theoretically, in the case of the Smith process, the corresponding estimator is consistent and asymptotically normal and efficient as both the temporal dimension and the number of simulation draws tend towards infinity. However, this relies on some assumptions on the real likelihood (especially its existence). An extension of this result to the Schlather and the Brown-Resnick processes would be of interest.

First simulation results are promising. Especially, our estimator seems to perform much better than the pairwise-based one when only a few temporal observations are available. This case is not anecdotal since it happens relatively often in many environmental studies. We expect the gain of our estimator to be an increasing function of the spatial dependence. However, these results need to be confirmed by further investigations. The comparison with the tripletwise estimator constitutes ongoing work.

Finally, we assumed the margins to be standard Fréchet. This implies a first step at which the GEV parameters have to be estimated at each location. The incorporation of the uncertainty due to this step would be interesting.

3.7 Appendix: Proofs

For Proposition 3.1

Proof. To simplify the presentation, we prove the result for one component but it can be extended. We first show that the set of points at which the maximum is realized by at least 2 distinct storms has an empty interior. It is sufficient to consider the case of 2 distinct storms. The sets respectively corresponding to 3, 4, ... different storms are indeed included in the one for 2 storms. Let us thus consider the set:

$$\mathcal{I} = \left\{ \mathbf{x} \in \mathbb{R}^2, \exists (m_{\mathbf{x}}, n_{\mathbf{x}}) \in \mathbb{N}^2, m_{\mathbf{x}} \neq n_{\mathbf{x}}, \max_i \xi_i \ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_i) = \xi_{n_{\mathbf{x}}} \ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) = \xi_{m_{\mathbf{x}}} \ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}) \right\}$$

A point $\mathbf{x} \in \mathcal{I}$ is characterized as follows

$$\begin{aligned}
\mathbf{x} \in \mathcal{I} &\iff \xi_{n_{\mathbf{x}}} \ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) = \xi_{m_{\mathbf{x}}} \ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}) \\
&\iff (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) - 2 \log \xi_{n_{\mathbf{x}}} = (\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}) - 2 \log \xi_{m_{\mathbf{x}}} \\
&\iff \|\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}\|^2_{\mathbf{\Sigma^{-1}}} - 2 \log \xi_{n_{\mathbf{x}}} = \|\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}\|^2_{\mathbf{\Sigma^{-1}}} - 2 \log \xi_{m_{\mathbf{x}}} \\
&\iff (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}).(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) - 2 \log \xi_{n_{\mathbf{x}}} = (\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}).(\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}) - 2 \log \xi_{m_{\mathbf{x}}}. \qquad (3.19)
\end{aligned}$$

where the dot denotes the scalar product associated to $\mathbf{\Sigma^{-1}}$ and $\|.\|_{\mathbf{\Sigma^{-1}}}$ is the induced norm.
Let us consider a vector $\mathbf{h} \in \mathbb{R}^2$ such that $\mathbf{x} + \mathbf{h}$ is in the neighborhood of $\mathbf{x}$, in the sense that $m_{\mathbf{x}} = m_{\mathbf{x}+\mathbf{h}}$ and $n_{\mathbf{x}} = n_{\mathbf{x}+\mathbf{h}}$. Using (3.19), we obtain

$$\begin{aligned}
\mathbf{x} + \mathbf{h} \in \mathcal{I} &\iff \|\mathbf{x} + \mathbf{h} - \mathbf{c}_{n_{\mathbf{x}}}\|^2_{\mathbf{\Sigma^{-1}}} - 2 \log \xi_{n_{\mathbf{x}}} = \|\mathbf{x} + \mathbf{h} - \mathbf{c}_{m_{\mathbf{x}}}\|^2_{\mathbf{\Sigma^{-1}}} - 2 \log \xi_{m_{\mathbf{x}}} \\
&\iff (\mathbf{x} + \mathbf{h} - \mathbf{c}_{n_{\mathbf{x}}}).(\mathbf{x} + \mathbf{h} - \mathbf{c}_{n_{\mathbf{x}}}) - 2 \log \xi_{n_{\mathbf{x}}} = (\mathbf{x} + \mathbf{h} - \mathbf{c}_{m_{\mathbf{x}}}).(\mathbf{x} + \mathbf{h} - \mathbf{c}_{m_{\mathbf{x}}}) - 2 \log \xi_{m_{\mathbf{x}}} \\
&\iff (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}).(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) + 2(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}).\mathbf{h} + \mathbf{h}.\mathbf{h} - 2 \log \xi_{n_{\mathbf{x}}} = (\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}).(\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}) \\
&\qquad + 2(\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}).\mathbf{h} + \mathbf{h}.\mathbf{h} - 2 \log \xi_{m_{\mathbf{x}}} \\
&\iff (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}).\mathbf{h} = (\mathbf{x} - \mathbf{c}_{m_{\mathbf{x}}}).\mathbf{h} \\
&\iff (\mathbf{c}_{m_{\mathbf{x}}} - \mathbf{c}_{n_{\mathbf{x}}}).\mathbf{h} = 0.
\end{aligned}$$

Therefore, $\mathbf{x} + \mathbf{h} \in \mathcal{I}$ if and only if $\mathbf{h}$ is orthogonal to the vector $(\mathbf{c}_{m_{\mathbf{x}}} - \mathbf{c}_{n_{\mathbf{x}}})$. Thus only one direction is suitable, showing that there is not any ball around $\mathbf{x}$ belonging to $\mathcal{I}$. That proves that the interior of $\mathcal{I}$ is empty and thus that the measure of $\mathcal{I}$ is equal to zero.

Let us now consider the set $\mathcal{D} = \mathbb{R}^2 / \mathcal{I}$. We show that for all $\mathbf{x} \in \mathcal{D}$, there exists a neighborhood V_0 of $\mathbf{\Sigma_0}$ in which each component of function g is differentiable with respect to $\mathbf{\Sigma}$. V_0 depends on $\mathbf{x}$. Remark that the points $\log [\xi_i f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_i)]$ are the points of a Poisson point process on $[-\infty; +\infty]$. We study its intensity measure. For a fixed $\mathbf{x} \in \mathbb{R}^2$, consider the set:

$$\mathcal{P}_\alpha = \left\{ (\xi, \mathbf{c}) \in \mathbb{R}_+ \times \mathbb{R}^2; \log[\xi f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}) > \alpha] \right\} \text{ for } \alpha \in \mathbb{R}.$$

The study of this set is equivalent $\forall \mathbf{x} \in \mathbb{R}^2$. Therefore, for the sake of simplicity, we consider $\mathbf{x} = \mathbf{0}$. We have

$$\begin{aligned}
\log \xi + \log f_{\mathbf{\Sigma}}(\mathbf{c}) > \alpha &\iff \log \xi - \log(2\pi) - \log \left(\sqrt{\det(\mathbf{\Sigma})} \right) - \frac{1}{2} \|\mathbf{c}\|^2_{\mathbf{\Sigma^{-1}}} > \alpha \\
&\iff \|\mathbf{c}\|^2_{\mathbf{\Sigma^{-1}}} < 2(\log \xi - K),
\end{aligned}$$

by introducing the constant $K = \log(2\pi) + \log\left(\sqrt{\det(\mathbf{\Sigma})}\right) + \alpha$. We know that the $(\xi_i, \mathbf{c}_i)$ are the points of a Poisson process (denoted in the following by N) on $\mathbb{R}_+ \times \mathbb{R}^2$ with intensity $\xi^{-2} d\xi d\nu$. The intensity measure of $\mathcal{P}_\alpha$ is then

$$\mu\{\mathcal{P}_\alpha\} = \int_{\xi > \exp(K)} \left(\int_{\|\mathbf{c}\|^2_{\mathbf{\Sigma^{-1}}} < 2(\log\xi - K)} d\mathbf{c} \right) \xi^{-2} d\xi = \int_{\xi > \exp(K)} \pi\sqrt{v_1 v_2}\; 2(\log\xi - K)\; \xi^{-2} d\xi$$

where v_1, v_2 are the two eigenvalues of $\mathbf{\Sigma}$. Indeed we used the fact that

$$\int_{\|\mathbf{c}\|^2_{\mathbf{\Sigma^{-1}}} < 2(\log\xi - K)} d\mathbf{c} \quad \text{is the area of the ellipse whose equation is } \|\mathbf{c}\|^2_{\mathbf{\Sigma^{-1}}} = 2(\log\xi - K).$$

By an appropriate variable change, we observe that this ellipse has main axes equal to $\sqrt{v_1 2(\log\xi - K)}$ and $\sqrt{v_2 2(\log\xi - K)}$ and thus an area equal to $\pi\sqrt{v_1 v_2}\; 2(\log\xi - K)$.

Therefore:

$$\begin{aligned}
\mu\{\mathcal{P}_\alpha\} &= 2\pi\sqrt{v_1 v_2} \int_{\xi > \exp(K)} (\log\xi - K)\; \xi^{-2} d\xi \\
&= 2\pi\sqrt{v_1 v_2} \left(\left[-\frac{1}{\xi}(\log\xi - K) \right]_{\exp(K)}^{+\infty} + \int_{\exp(K)}^{+\infty} \xi^{-2}\; d\xi \right) \text{(after integrating by parts)} \\
&= 2\pi\sqrt{v_1 v_2} \left[-\frac{1}{\xi} \right]_{\exp(K)}^{+\infty} = 2\pi\sqrt{v_1 v_2}\exp(-K) = \exp(-\alpha), \qquad (3.20)
\end{aligned}$$

by using the fact that $\det(\mathbf{\Sigma}) = v_1 v_2$. Therefore the intensity measure of $\mathcal{P}$ is finite, meaning that there is no accumulation of points above the level α. Especially there is no accumulation around the maximum.

Finally, using especially the fact that the set of points $\{i, \xi_i > 1\}$ is almost surely finite, it can be shown that the order of the points is not modified in a neighborhood of the true parameter (storms centered very far away cannot contribute to the maximum if the variation of $\mathbf{\Sigma}$ is sufficiently low).

□

For Lemma 3.1

Proof. We first show that each component of function g has its derivative bounded by a variable with finite expectation. Recall that (see (3.8)) the $m-th$ component of function g is written

$$g_m(\mathbf{\Sigma}, \epsilon) = \log\; \left[\max_i \xi_i\; f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_i) \right] = \log\xi_{n_{\mathbf{x}_m}} + \log f_{\mathbf{\Sigma}}(\mathbf{x}_m - \mathbf{c}_{n_{\mathbf{x}_m}}),$$

where $n_{\mathbf{x}_m}$ is the index (random) realizing the maximum at point $\mathbf{x}_m$. To simplify the notation in the following, we omit the index m relative to $\mathbf{x}_m$.

Function $f_{\mathbf{\Sigma}}$ being the multivariate normal density, we have

$$f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) = \frac{1}{(2\pi)^{\frac{d}{2}}\det(\mathbf{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) \right), \text{ giving}$$

$$\log f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log[\det(\mathbf{\Sigma})] - \frac{1}{2}(\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{n_{\mathbf{x}}}).$$

Therefore the derivative of g_m with respect to the matrix $\mathbf{\Sigma}$ is:

$$\frac{\partial g_m(\mathbf{\Sigma}, \epsilon)}{\partial \mathbf{\Sigma}} = -\frac{1}{2}\left(\frac{\partial \log[\det(\mathbf{\Sigma})]}{\partial \mathbf{\Sigma}} + \frac{\partial (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})}{\partial \mathbf{\Sigma}}\right)$$

Formula 11.7 in Dwyer (1967) gives, for any symmetric matrix $\mathbf{\Sigma}$,

$$\frac{\partial \log[\det(\mathbf{\Sigma})]}{\partial \mathbf{\Sigma}} = \mathbf{\Sigma}^{-1}.$$

Moreover Equation (11.8) in Dwyer (1967) provides, for any symmetric matrix $\mathbf{\Sigma}$,

$$\frac{\partial (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma^{-1}} (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})}{\partial \mathbf{\Sigma}} = -\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}.$$

Combining the two previous expressions we finally obtain

$$\frac{\partial g_m(\mathbf{\Sigma}, \epsilon)}{\partial \mathbf{\Sigma}} = -\frac{1}{2}\left(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}\right).$$

Therefore, for any norm,

$$\left\| \frac{\partial g_m(\mathbf{\Sigma}, \epsilon)}{\partial \mathbf{\Sigma}} \right\| \leq \frac{1}{2}\left(\|\mathbf{\Sigma}^{-1}\| + \|\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}\|\right). \tag{3.21}$$

All norms are equivalent in finite dimension. For the sake of simplicity, we choose the spectral norm, denoted by $\|.\|_2$. Denoting $A = \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})$, we have $\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1} = AA'$. By definition of the spectral norm,

$$\|AA'\|_2 = \sqrt{\lambda_{max}[(AA')'AA']} = \sqrt{\lambda_{max}[(AA')^2]}.$$

where $\lambda_{max}(.)$ denotes the highest eigenvalue. AA' is symmetric and therefore diagonalizable. Moreover AA' is of rank 1 and thus has a null eigenvalue. The other eigenvalue is $A'A$, since $(AA')A = A(A'A) = (A'A)A$. That yields $\lambda_{max}[AA'] = A'A$. Since AA' is diagonalizable, $\lambda_{max}[(AA')^2] = \lambda_{max}[AA']^2$. Finally, $\|AA'\|_2 = A'A$ i.e.

$$\|\mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}\|_2 = (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-2}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}). \tag{3.22}$$

$\mathbf{\Sigma}^{-1}$ and $\mathbf{\Sigma}^{-2}$ being symmetric definite positive,

$$(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \geq \lambda_{min}(\mathbf{\Sigma}^{-1})\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^2 \text{ and}$$

$$(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-2}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \leq \lambda_{max}(\mathbf{\Sigma}^{-2})\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^2, \text{ yielding}$$

$$(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-2}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \leq \frac{\lambda_{max}(\mathbf{\Sigma}^{-2})}{\lambda_{min}(\mathbf{\Sigma}^{-1})}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}). \tag{3.23}$$

It is therefore sufficient to control the term: $(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})$.
As the center of the storm realizing the max at point $\mathbf{x}$, $\mathbf{c}_{n_\mathbf{x}}$ is characterized by

$$\begin{aligned}
&\xi_{n_\mathbf{x}}\ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \geq \xi_i\ f_{\mathbf{\Sigma}}(\mathbf{x} - \mathbf{c}_i)\ \ \forall i \\
&\qquad \iff \log \xi_{n_\mathbf{x}} - \frac{1}{2}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma^{-1}}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \geq \log \xi_i - \frac{1}{2}(\mathbf{x} - \mathbf{c}_i)' \mathbf{\Sigma^{-1}}(\mathbf{x} - \mathbf{c}_i)\ \forall i \\
&\qquad \iff (\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})' \mathbf{\Sigma^{-1}}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}) \leq 2 \log \frac{\xi_{n_\mathbf{x}}}{\xi_i} + (\mathbf{x} - \mathbf{c}_i)' \mathbf{\Sigma^{-1}}(\mathbf{x} - \mathbf{c}_i)\ \forall i \\
&\qquad \iff \|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^2_{\mathbf{\Sigma^{-1}}} \leq 2 \log \frac{\xi_{n_\mathbf{x}}}{\xi_i} + \|\mathbf{x} - \mathbf{c}_i\|^2_{\mathbf{\Sigma^{-1}}}\ \ \forall i.
\end{aligned}$$

Note that for 2 independent real-valued random variables U and V, $U+V \geq \lambda \Longrightarrow U \geq \frac{\lambda}{2}$ or $V \geq \frac{\lambda}{2}$, giving $\mathbb{P}(U+V \geq \lambda) \leq \mathbb{P}\left(U \geq \frac{\lambda}{2}\right) + \mathbb{P}\left(V \geq \frac{\lambda}{2}\right)$. Thus

$$\begin{aligned}\mathbb{P}(\|\mathbf{x}-\mathbf{c}_{n_{\mathbf{x}}}\|^2_{\mathbf{\Sigma^{-1}}} \geq \lambda) &\leq \mathbb{P}(\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}^{-1}} - 2\log\xi_i + 2\log\xi_{n_{\mathbf{x}}} \geq \lambda,\ \forall i) \\ &\leq \mathbb{P}\left(\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}^{-1}} - 2\log\xi_i \geq \frac{\lambda}{2},\ \forall i\right) + \mathbb{P}\left(2\log\xi_{n_{\mathbf{x}}} \geq \frac{\lambda}{2}\right).\end{aligned} \tag{3.24}$$

Let us first deal with the first term: Since the $(\xi_i, \mathbf{c}_i)$ are the points of a Poisson process on $\mathbb{R}_+ \times \mathbb{R}^2$ with intensity $\xi^{-2}\, d\xi\, d\nu$, we have

$$\mathbb{P}\left(\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \geq \frac{\lambda}{2},\ \forall i\right) = \exp\left(-\mu\left\{\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \leq \frac{\lambda}{2},\ \forall i\right\}\right), \text{ where}$$

$$\begin{aligned}\mu\left\{\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \leq \frac{\lambda}{2},\ \forall i\right\} &= \int_{e^{-\frac{\lambda}{4}}}^{+\infty}\left(\int_{\|\mathbf{x}-\mathbf{c}\|^2_{\mathbf{\Sigma}} \leq \frac{\lambda}{2}+2\log\xi} d\mathbf{c}\right)\xi^{-2}d\xi \\ &= \sqrt{v_1 v_2}\pi\int_{e^{-\frac{\lambda}{4}}}^{+\infty}\left(\frac{\lambda}{2}+2\log\xi\right)\,\xi^{-2}d\xi\end{aligned}$$

Making the variable change $u = \frac{\lambda}{2} + 2\log\xi$, yielding $\xi = \exp\left(\frac{u}{2}-\frac{\lambda}{4}\right)$ and $d\xi = \exp\left(\frac{u}{2}-\frac{\lambda}{4}\right)\frac{du}{2}$, we obtain

$$\begin{aligned}\mu\left\{\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \leq \frac{\lambda}{2},\ \forall i\right\} &= \frac{\sqrt{v_1 v_2}\pi}{2}\int_{e^{-\frac{\lambda}{4}}}^{+\infty} u\exp\left(-\frac{u}{2}+\frac{\lambda}{4}\right)du \\ &= \frac{\sqrt{v_1 v_2}\pi}{2}\exp\left(\frac{\lambda}{4}\right)\int_{e^{-\frac{\lambda}{4}}}^{+\infty} u\exp\left(-\frac{u}{2}\right)du.\end{aligned}$$

Integrating by parts, we obtain

$$\begin{aligned}\int_{e^{-\frac{\lambda}{4}}}^{+\infty} u\exp\left(-\frac{u}{2}\right)du &= \left[-2\exp\left(-\frac{u}{2}\right)u\right]_{\exp\left(-\frac{\lambda}{4}\right)}^{+\infty} + 2\int_{\exp\left(-\frac{\lambda}{4}\right)}^{+\infty}\exp\left(-\frac{u}{2}\right)du \\ &= 2\exp\left[-\frac{\exp\left(-\frac{\lambda}{4}\right)}{2}\right]\exp\left(-\frac{\lambda}{4}\right) + 4\exp\left[-\frac{\exp\left(-\frac{\lambda}{4}\right)}{2}\right] \\ &= 2\exp\left[-\frac{\exp\left(-\frac{\lambda}{4}\right)}{2}\right]\left(\exp\left(-\frac{\lambda}{4}\right)+2\right), \text{ giving}\end{aligned}$$

$$\begin{aligned}\mu\left\{\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \leq \frac{\lambda}{2},\ \forall i\right\} &= \sqrt{v_1 v_2}\pi\exp\left[-\frac{\exp\left(-\frac{\lambda}{4}\right)}{2}\right]\left[1+2\exp\left(\frac{\lambda}{4}\right)\right] \\ &\underset{\lambda\to\infty}{\sim} \sqrt{v_1 v_2}\pi\left(1-\frac{1}{2}\exp\left(-\frac{\lambda}{4}\right)\right)\left[1+2\exp\left(\frac{\lambda}{4}\right)\right] \\ &\underset{\lambda\to\infty}{\sim} \sqrt{v_1 v_2}\pi\left[2\exp\left(\frac{\lambda}{4}\right)-\frac{1}{2}\exp\left(-\frac{\lambda}{4}\right)\right], \text{ yielding}\end{aligned}$$

$$\mathbb{P}\left(\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}} - 2\log\xi_i \geq \frac{\lambda}{2},\ \forall i\right) \underset{\lambda\to\infty}{\sim} \exp\left\{-\sqrt{v_1v_2}\pi\left[2\exp\left(\frac{\lambda}{4}\right) - \frac{1}{2}\exp\left(-\frac{\lambda}{4}\right)\right]\right\}$$
$$\underset{\lambda\to\infty}{\sim} \exp\left[-2\sqrt{v_1v_2}\pi\exp\left(\frac{\lambda}{4}\right)\right]\left[\sqrt{v_1v_2}\pi + \frac{\sqrt{v_1v_2}\pi}{2}\exp\left(-\frac{\lambda}{4}\right)\right], \tag{3.25}$$

which is integrable.

Let us now deal with the second term, i.e.

$$\mathbb{P}\left[2\log\xi_{n_\mathbf{x}} \geq \frac{\lambda}{2}\right] = \mathbb{P}\left[\frac{1}{\xi_{n_\mathbf{x}}} \leq \exp\left(-\frac{\lambda}{4}\right)\right].$$

A Poisson process on $\mathbb{R}_+{}^*$ can be represented as the sum of i.i.d. standard exponential random variables. Moreover the mapping $\xi \to \xi^{-1}$ to the points of a Poisson process yields a new Poisson process of intensity $\xi^{-2}d\xi$. We thus can write

$$\min_i \frac{1}{\xi_i} \overset{d}{=} Exp(1),$$

where $Exp(1)$ denotes the standard exponential distribution. Moreover, $\dfrac{1}{\xi_{n_\mathbf{x}}} \geq \min_i \dfrac{1}{\xi_i}$, implying

$$\mathbb{P}\left[\frac{1}{\xi_{n_\mathbf{x}}} \leq \exp\left(-\frac{\lambda}{4}\right)\right] \leq \mathbb{P}\left[\min_i \frac{1}{\xi_i} \leq \exp\left(-\frac{\lambda}{4}\right)\right] = 1-\exp\left[\exp\left(-\frac{\lambda}{4}\right)\right] \underset{\lambda\to\infty}{\sim} \exp\left(-\frac{\lambda}{4}\right),$$

$$\text{giving} \quad \mathbb{P}\left(2\log\xi_{n_\mathbf{x}} \geq \frac{\lambda}{2}\right) \underset{\lambda\to\infty}{\sim} \exp\left(-\frac{\lambda}{4}\right) \text{ which is integrable.} \tag{3.26}$$

Finally, writing the expectation as the integral of the survival function and using (3.24), we have

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}-\mathbf{c}_{n_\mathbf{x}}\|^2_{\mathbf{\Sigma}^{-1}}] &= \int_0^{+\infty} \mathbb{P}(\|\mathbf{x}-\mathbf{c}_{n_\mathbf{x}}\|^2_{\mathbf{\Sigma}^{-1}} \geq \lambda)\, d\lambda \\ &\leq \int_0^{+\infty} \mathbb{P}\left(\|\mathbf{x}-\mathbf{c}_i\|^2_{\mathbf{\Sigma}^{-1}} - 2\log\xi_i \geq \frac{\lambda}{2},\ \forall i\right) d\lambda + \int_0^{+\infty} \mathbb{P}\left(2\log\xi_{n_\mathbf{x}} \geq \frac{\lambda}{2}\right) d\lambda.\end{aligned}$$

Using (3.25) and (3.26), $\mathbb{E}[\|\mathbf{x}-\mathbf{c}_{n_\mathbf{x}}\|^2_{\mathbf{\Sigma}}]$ is finite. Therefore, using (3.22) and (3.23), $\mathbb{E}[\|\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{c}_{n_\mathbf{x}})(\mathbf{x}-\mathbf{c}_{n_\mathbf{x}})'\mathbf{\Sigma}^{-1}\|]$ also and finally, using (3.21), $\mathbb{E}\left[\left\|\frac{\partial g_m(\mathbf{\Sigma},\epsilon)}{\partial \mathbf{\Sigma}}\right\|\right]$ also. This is true for all m and thus it is also true for $\mathbb{E}\left[\left\|\frac{\partial g(\mathbf{\Sigma},\epsilon)}{\partial \mathbf{\Sigma}}\right\|\right]$.

Carrying out exactly the same reasoning on the sup for $\mathbf{\Sigma} \in V_0$, we can deduce that $\mathbb{E}\left[\sup_{\mathbf{\Sigma}\in V_0}\left\|\frac{\partial g(\mathbf{\Sigma},\epsilon)}{\partial \mathbf{\Sigma}}\right\|\right] < +\infty$.

Note that here, the derivatives are written with respect to the matrix $\mathbf{\Sigma}$ whereas in Fermanian and Salanié (2004), they are written with respect to a vector. But the result is not modified. Indeed by choosing the norm being equal to the sum of absolute values of all elements of the matrix (respectively the vector), it is clear that if each element is bounded, then the norm of the matrix (respectively the vector) is also bounded. □

For Lemma 3.2

Proof. i) Similarly to Lemma 3.1, it is sufficient to show the result for each component. Let us consider a whole number $p_0 \geq 5$. Using (3.21), we have

$$\left\| \frac{\partial g_m(\boldsymbol{\Sigma}, \epsilon)}{\partial \boldsymbol{\Sigma}} \right\|^{p_0} \leq \frac{1}{2^{p_0}} \left(\|\boldsymbol{\Sigma}^{-1}\| + \|\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})'\boldsymbol{\Sigma}^{-1}\| \right)^{p_0}.$$

The binomial theorem then tells us that it is sufficient to control the expectation of each term
$\|\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})(\mathbf{x} - \mathbf{c}_{n_\mathbf{x}})'\boldsymbol{\Sigma}^{-1}\|^p$, for $p = 1, \ldots, p_0$. To this purpose, using (3.22) and (3.23), we know that it is sufficient to control the expectation of $\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^{2p}_{\boldsymbol{\Sigma}^{-1}}$. Note that

$$\mathbb{P}(\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^{2p}_{\boldsymbol{\Sigma}^{-1}} \geq \lambda) = \mathbb{P}\left(\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^{2}_{\boldsymbol{\Sigma}^{-1}} \geq \lambda^{\frac{1}{p}} \right).$$

Exactly the same computations than in Lemma 3.2 with $\lambda^{\frac{1}{p}}$ instead of λ lead to the integrability of $\mathbb{P}\left(\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^{2}_{\boldsymbol{\Sigma}^{-1}} \geq \lambda^{\frac{1}{p}} \right)$ and therefore of $\mathbb{P}(\|\mathbf{x} - \mathbf{c}_{n_\mathbf{x}}\|^{2p}_{\boldsymbol{\Sigma}^{-1}} \geq \lambda)$, showing the result.

ii) The proof is very similar to the one of Lemma 3.1. Taking the derivative with respect to $\boldsymbol{\Sigma}$ of $\frac{\partial g_m(\boldsymbol{\Sigma}, \epsilon)}{\partial \boldsymbol{\Sigma}}$ lets appear Kronecker products whose norms can be bounded in the same way than in Lemma 3.1.

iii) It is clear that

$$\sup_{\boldsymbol{\Sigma}} \left(\left\| \frac{\partial g}{\partial \boldsymbol{\Sigma}} \right\|^{\zeta} \right) = \left(\sup_{\boldsymbol{\Sigma}} \left\| \frac{\partial g}{\partial \boldsymbol{\Sigma}} \right\| \right)^{\zeta}$$

and therefore this points stems directly from i), where the result has been shown for $\zeta > 4$.

□

For Lemma 3.3

Proof. By considering the infinity norm, we have

$$\mathbb{P}(\|\mathbf{Z}\| > S^\nu) = 1 - \mathbb{P}(\|\mathbf{Z}\| \leq S^\nu) = 1 - \mathbb{P}\left[\max(Z_1, \ldots, Z_M) \leq S^\nu\right].$$

Moreover, using the fact that $\mathbf{Z}$ has standard Gumbel margins, we know that

$$\begin{aligned} \mathbb{P}\left[\max(Z_1, \ldots, Z_M) \leq S^\nu\right] = \mathbb{P}\left(Z_1 \leq S^\nu\right)^{\theta(Z_1, \ldots, Z_M)} &= \exp[-\exp(-S^\nu)]^{\theta(Z_1, \ldots, Z_M)} \\ &= \exp\left[-\theta(Z_1, \ldots, Z_M)\exp(-S^\nu)\right], \end{aligned} \tag{3.27}$$

where $\theta(Z_1, \ldots, Z_M)$ is the extremal coefficient. Then, since $h = C_2 S^{-b}$, we obtain

$$\begin{aligned} \mathbb{P}(\|\mathbf{Z}\| > S^\nu)\log h &= \left(1 - \exp\left[-\theta(Z_1, \ldots, Z_M)\exp(-S^\nu)\right]\right)\log h \\ &\sim -b\ \theta(Z_1, \ldots, Z_M)\exp(-S^\nu)\log S \end{aligned}$$

Thus, by taking $\nu > 0$,

$$\lim_{S \to +\infty} \mathbb{P}(\|\mathbf{Z}\| > S^\nu)\log h = 0,$$

completing the proof. □

For Lemma 3.4

Proof. Since $S = C_1 T^a$, we have $T^{\frac{\gamma}{2\gamma-1}} = d_1 S^{D_1}$, where $d_1 = C_1^{-\frac{\gamma}{a(2\gamma-1)}}$ and $D_1 = \frac{\gamma}{a(2\gamma-1)}$.

In a similar way, $h = C_2 S^{-b}$ gives

$$\left(\frac{T^{\frac{1}{2}}}{h^{\delta}}\right)^{\frac{\gamma'}{\gamma'-1}} = d_2 S^{D_2} \text{ for constants } d_2 \text{ and } D_2.$$

Idem, concerning the third term, we have

$$\left(\frac{T^{\frac{1}{2}}|\log h|}{h^{M+1+\delta}}\right)^{\frac{\zeta}{\zeta-1}} = d_3 S^{D_3} |\log(C_2) - b\log S|^{\frac{\zeta}{\zeta-1}} \text{ for constants } d_3 \text{ and } D_3.$$

Using (3.27), we obtain

$$\begin{aligned}&\left[T^{\frac{\gamma}{2\gamma-1}} + \left(\frac{T^{\frac{1}{2}}}{h^{\delta}}\right)^{\frac{\gamma'}{\gamma'-1}} + \left(\frac{T^{\frac{1}{2}}|\log h|}{h^{M+1+\delta}}\right)^{\frac{\zeta}{\zeta-1}}\right] \mathbb{P}(\|\mathbf{Z}\| > S^{\nu}))\\ &\sim \left(d_1 S^{D_1} + d_2 S^{D_2} + d_3 S^{D_3} |\log(C_2) - b\log S|^{\frac{\zeta}{\zeta-1}}\right) \theta(Z_1, \ldots, Z_M) \exp(-S^{\nu})\end{aligned}$$

For all $\nu > 0$, the last expression tends towards 0 as S tends towards infinity, showing the result. □

For Theorem 3.1

Proof. Note that our Lemma 3.3 shows that Assumption $T1$ in Fermanian and Salanié (2004) is verified in the particular case of the max-stable Smith process. Similarly, our Lemma 3.1 precisely shows that the Assumption $M1$ in Fermanian and Salanié (2004) is verified in our particular case. Indeed, in our case, the s_0 in Fermanian and Salanié (2004) can be set to 0 since there is no exogenous variable. Finally their condition reduces to the one we stated. Moreover, Assumptions $K1$, $L1$, $L2$ and $R1$ in Fermanian and Salanié (2004) are respectively our Assumptions 3.1, 3.4, 3.5 and 3.2.

We can then directly apply the Theorem 2.1 in Fermanian and Salanié (2004), stating that under Assumptions $K1$, $L1$, $L2$, $T1$, $M1$ and $R1$, $\hat{\mathbf{\Sigma}}_T^S$ is strongly consistent. Almost surely,

$$\lim_{S,T\to\infty} \hat{\mathbf{\Sigma}}_T^S = \mathbf{\Sigma_0}.$$

□

For Theorem 3.2

Proof. As previously mentioned, our Lemma 3.1 shows that the Assumption $M1$ in Fermanian and Salanié (2004) is verified in our particular case. Similarly Lemma 3.2 shows that Assumption $M2$ is verified. Indeed, as in the case of Assumption $M1$, the r_0 and s_1 in conditions i) and ii) of Fermanian and Salanié (2004) can bet set to zeros in our case. Then their conditions reduce to the ones we stated.

Moreover, Assumptions $K1$, $L1-L3$, $R1-R3$ and $T1-T2$ are respectively our Assumptions 3.1, 3.4, 3.5, 3.6, 3.2, 3.3 and 3.7. Thus we can directly apply thorem 2.2 in Fermanian and Salanié (2004), stating that under Assumptions $K1$, $M1$ and $M2$, $L1-L3$, $R1-R3$ and $T1$ and $T2$, $\hat{\mathbf{\Sigma}}_T^S$ is asymptotically normal and asymptotically efficient:

$$\sqrt{T}(\hat{\mathbf{\Sigma}}_T^S - \mathbf{\Sigma_0}) \underset{S,T\to\infty}{\xrightarrow{d}} N(\mathbf{0}, \mathbf{\Omega}),$$

where $\mathbf{\Omega}$ is the asymptotic covariance matrix of the exact maximum likelihood estimator. □

For Proposition 3.2

Proof. Since $h = C_2 S^{-b}$, we have

$$\frac{Sh^{M+2\delta}}{\log S} = \frac{S(C_2^{-b})^{M+2\delta}}{\log S} = \frac{SC_2^{M+2\delta}S^{-b(M+2\delta)}}{\log S} = \frac{C_2^{M+2\delta}\ S^{1-b(M+2\delta)}}{\log S}.$$

Therefore (3.9) of Assumption 3.2 implies $1 - b(M+2\delta) > 0$ i.e. $b < \dfrac{1}{M+2\delta}$.

Using $C_1 T^a$, we have $h = C_2(C_1T^a)^{-b} = C_2C_1^{-b}T^{-ab} = C\ T^{-ab}$ and $S = \left(\dfrac{h}{C_2}\right)^{-\frac{1}{b}}$, where $C = C_2C_1^{-b}$. Then

$$\begin{aligned} T^{\frac{1}{2}}h^{\rho-\delta}\log h = T^{\frac{1}{2}}(C\ T^{-ab})^{\rho-\delta}\log(C\ T^{-ab}) &= T^{\frac{1}{2}}C^{\rho-\delta}(T^{-ab})^{\rho-\delta}(\log C\ - ab\log T) \\ &= C^{\rho-\delta}\log C\ T^{\frac{1}{2}-ab(\rho-\delta)} - C^{\rho-\delta}\ ab\ T^{\frac{1}{2}-ab(\rho-\delta)}\log T. \end{aligned}$$

Thus (3.10) of Assumption 3.3 implies: $\dfrac{1}{2} - ab(\rho-\delta) < 0$, i.e. $ab > \dfrac{1}{2(\rho-\delta)}$.

We have moreover

$$T = \left(\frac{h}{C}\right)^{-\frac{1}{ab}}, \tag{3.28}$$

giving

$$\begin{aligned} Th^{-2(M+\delta+1)}\log^2 h\frac{\log S}{S} &= \left(\frac{h}{C}\right)^{-\frac{1}{ab}} h^{-2(M+\delta+1)}\log^2 h \times \left(-\frac{1}{b}\right)\frac{\log\left(\frac{h}{C_2}\right)}{\left(\frac{h}{C_2}\right)^{\frac{-1}{b}}} \\ &= Kh^{-\left(\frac{1}{ab}+2(M+\delta+1)-\frac{1}{b}\right)}\log^2 h \times\ (\log h - \log C_2). \end{aligned}$$

Therefore (3.10) of Assumption 3.3 imposes moreover

$$\frac{1}{ab} + 2(M+\delta+1) - \frac{1}{b} < 0 \quad \text{i.e.} \quad 1 + 2ab(M+\delta+1) < a.$$

Moreover, the density of a standard Gumbel is

$$l(z) = \exp[-z - \exp(-z)].$$

Thus

$$\begin{aligned}\mathbb{P}(l(Z) \leq 2h^\delta) &= \mathbb{P}(\exp[-(Z+\exp(-Z)]) \leq 2h^\delta) = \mathbb{P}[Z+\exp(-Z) \geq -\log(2h^\delta)] \\ &= \mathbb{P}[Z+\exp(-Z) \geq -\log(2h^\delta)] \\ &\leq \mathbb{P}\left[Z \geq -\frac{\log(2h^\delta)}{2}\right] + \mathbb{P}\left[\exp(-Z) \geq -\frac{\log(2h^\delta)}{2}\right].\end{aligned}$$

Let us deal with the first term:

$$\begin{aligned}\mathbb{P}\left[Z \geq -\frac{\log(2h^\delta)}{2}\right] = 1 - \mathbb{P}\left[Z \leq -\frac{\log(2h^\delta)}{2}\right] &= 1 - \exp\left[-\exp\left(\frac{\log(2h^\delta)}{2}\right)\right] \\ &\underset{h\to 0}{\sim} \exp\left(\frac{\log(2h^\delta)}{2}\right) = \sqrt{2}\; h^{\frac{\delta}{2}}.\end{aligned}$$

Concerning the second term,

$$\begin{aligned}&\mathbb{P}[\exp(-Z) \geq -\log(2h^\delta)] = \mathbb{P}\left[\frac{1}{\exp(Z)} \geq -\frac{\log(2h^\delta)}{2}\right] = \mathbb{P}\left[\exp(Z) \leq -\frac{2}{\log(2h^\delta)}\right] \\ &(\text{ since for } h \to 0, \frac{-2}{\log(2h^\delta)} > 0) \\ &= \exp\left[\frac{\log(2h^\delta)}{2}\right] \text{ since } \exp(Z) \text{ is a unit Fréchet(1) variable} \\ &= \sqrt{2}\; h^{\frac{\delta}{2}}.\end{aligned}$$

Using (3.28), we obtain

$$(T^{\frac{1}{2}}|\log h|)^{\frac{\gamma}{\gamma-1}} = C^{\frac{1}{ab}}|\log h|^{\frac{\gamma}{\gamma-1}} h^{-\frac{\gamma}{2ab(\gamma-1)}}.$$

Thus, the expression

$$\left(T^{\frac{1}{2}}|\log h|\right)^{\frac{\gamma}{\gamma-1}} \mathbb{P}\left(l(\mathbf{z}) \leq 2h^\delta\right) \text{ is dominated by a term equivalent to}$$

$$2\sqrt{2}\; C^{\frac{1}{ab}} h^{\frac{\delta}{2}} |\log h|^{\frac{\gamma}{\gamma-1}} h^{-\frac{\gamma}{2ab(\gamma-1)}}.$$

Moreover,

$$2\sqrt{2}\; C^{\frac{1}{ab}} h^{\frac{\delta}{2}} |\log h|^{\frac{\gamma}{\gamma-1}} h^{-\frac{\gamma}{2ab(\gamma-1)}} \underset{h\to 0}{\to} 0 \Longleftrightarrow \delta > \frac{\gamma}{ab(\gamma-1)},$$

giving that

$$ab > \frac{\gamma}{\delta(\gamma-1)}$$

is a sufficient condition for Assumption 3.7 to be verify in the univariate case. □

3.8 Appendix: Bandwidth selection

In the univariate case, the Asymptotic Means Integrated Squared Error (AMISE) optimal bandwidth is written (see e.g. Wand and Jones (1994))

$$h_{AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(l'') S}\right]^{\frac{1}{5}}, \tag{3.29}$$

where K is a general kernel function, l the density to estimate, $R(g) = \int g(z)^2 dz \ \forall g \in L_2(\mathbb{R})$, and $\mu_2(K) = \int z^2 K(z) dz$.

A normal scale (or normal reference rule) bandwidth selector involves using a bandwidth that is optimal for the normal density having the same scale as the one estimated on the underlying density. If l is the density of the standard normal distribution, $R(l'')$ can easily be computed and the corresponding optimal bandwidth is

$$h_{AMISE,N} = \left[\frac{8\pi^{\frac{1}{2}} R(K)}{3\mu_2(K)^2 S} \right]^{\frac{1}{5}} \sigma. \tag{3.30}$$

Finally, a normal scale bandwidth is obtained by replacing σ by its estimate $\hat{\sigma}$. In the case of a Gaussian kernel, $R(K)$ is easily computed, yielding the following approximation (Scott rule): $h_{AMISE,N} = S^{-\frac{1}{5}} \hat{\sigma}$. The extension to the multivariate case is direct:

$$\mathbf{H} = S^{-\frac{1}{M+4}} \hat{\mathbf{\Sigma}}_{\mathbf{l}}^{\frac{1}{2}},$$

where $\mathbf{\Sigma_l}$ is the covariance matrix of the multivariate normal distribution.

However, the normal reference rule gives very poor results and we have tested the Gumbel distribution as reference (which is exactly the marginal distribution of the process under study). In this case, $l(z) = \exp\{-[z + \exp(-z)]\}$, yielding easily

$$l''(z) = \exp\{-[z + \exp(-z)]\}[(\exp(-z) - 1)^2 - \exp(-z)].$$

A numerical integration gives $R(l'') \sim 0.25$ and $h_{AMISE,Gumbel}$ can be computed using (3.29). However, the results are very poor in the multivariate case, even when considering the exponent $\frac{1}{M+4}$ to take the dimension M into account.

Thus the idea was to use the exact asymptotic optimal bandwidth in higher dimension than 1. In dimension 2, the diagonal AMISE optimal bandwidth has the expression (Wand and Jones, 1993)

$$\mathbf{H_{diag,AMISE}} = 4\pi \ S \ A(l)^{-\frac{1}{6}} \ \mathrm{diag}(R, R^{-1}), \tag{3.31}$$

where

$$A(l) = \int \frac{\partial^4 l}{\partial z_1^2 \partial z_2^2} l(z_1, z_2) \, dz_1 \, dz_2 + \left[\left(\int \frac{\partial^4 l}{\partial z_1^4} l(z_1, z_2) \ dz_1 \ dz_2 \right) \left(\int \frac{\partial^4 l}{\partial z_2^4} l(z_1, z_2) \ dz_1 \ dz_2 \right) \right]^{\frac{1}{2}} \text{ and}$$

$$R = \left[\left(\int \frac{\partial^4 l}{\partial z_1^4} l(z_1, z_2) \ dz_1 \ dz_2 \right) \left(\int \frac{\partial^4 l}{\partial z_2^4} l(z_1, z_2) \ dz_1 \ dz_2 \right) \right]^{\frac{1}{8}}.$$

Using the expression given in Padoan et al. (2010), the bivariate density of the logarithm of the Smith process is easily obtained:

$$\begin{aligned} l(z_1, z_2) = \exp\left[-\frac{\Phi(w)}{\exp(z_1)} - \frac{\Phi(v)}{\exp(z_2)} \right] \Bigg[& \left(\frac{\phi(w)}{a \exp(z_1)} + \frac{\Phi(w)}{\exp(z_1)} - \frac{\phi(v)}{a \exp(z_2)} \right) \\ & \times \left(\frac{\phi(v)}{a \exp(z_2)} + \frac{\Phi(v)}{\exp(z_2)} - \frac{\phi(w)}{a \exp(z_1)} \right) \\ & + \left(\frac{w\phi(v)}{a^2 \exp(z_2)} + \frac{v\phi(w)}{a^2 \exp(z_1)} \right) \Bigg], \end{aligned}$$

where Φ and ϕ are respectively the cumulative distribution function and the density of a standard normal,

$$a = \sqrt{(\mathbf{x}_2 - \mathbf{x}_1)' \ \mathbf{\Sigma_0}^{-1}(\mathbf{x}_2 - \mathbf{x}_1)}, \quad w = \frac{a}{2} + \frac{1}{a}(z_2 - z_1) \text{ and } \ v = a - w.$$

Computations involved by (3.31) are not tractable due to the complexity of l, meaning that it is hopeless to obtain the optimal diagonal bandwidth matrix for $M = 2$ and a fortiori the full optimal matrix as well as optimal matrices in higher dimensions.

Finally, for all these reasons and since variable bandwidths are known to perform better than fixed ones in high dimension (Terrell and Scott, 1992), we made the choice and we advise to use such variable bandwiths, especially the nearest neighbor estimator by Loftsgaarden and Quesenberry (1965). The considered distance in their approach is the standard euclidean distance of $\mathbb{R}^M$.

Remark 3.5. *Note that in order to gain flexibility, we tested the nearest neighbor approach site by site. However, the distance used in this case is the euclidean distance in dimension 1. Therefore, the corresponding estimator does not perform well in dimension higher than 2.*

Bibliography

Appel, M., Labarre, R., and Radulovic, D. (2004). On accelerated random search. *SIAM Journal on Optimization*, 14(3):708–731.

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20).

Botev, Z., Grotowski, J., Kroese, D., et al. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in probability and statistics*, pages 373–390. Springer.

Cressie, N. A. and Cassie, N. A. (1993). *Statistics for spatial data*, volume 900. Wiley New York.

Davison, A. C. (2003). *Statistical models*. Cambridge University Press.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer.

de Haan, L. and Pickands, J. (1986). Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 72(4):477–492.

Duong, T. and Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32(3):485–506.

Dwyer, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*, volume 33. Springer.

Fermanian, J.-D. and Salanié, B. (2004). A nonparametric simulated maximum likelihood estimation method. *Econometric Theory*, 20(04):701–734.

Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12(04):657–681.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

Gouriéroux, C. and Monfort, A. (1997). *Simulation-based econometric methods*. Oxford University Press.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118.

Hajivassiliou, V. A. and McFadden, D. L. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, pages 863–896.

Hall, P. and Marron, J. (1988). Variable window width kernel estimates of probability densities. *Probability Theory and Related Fields*, 80(1):37–49.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, pages 2042–2065.

Kaufman, L. and Rousseeuw, P. (1987). Clustering by means of medoids.

Laroque, G. and Salanié, B. (1994). Estimating the canonical disequilibrium model: Asymptotic theory and finite sample properties. *Journal of Econometrics*, 62(2):165–210.

Lee, L.-F. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 11:437–437.

Lerman, S. and Manski, C. (1981). On the use of simulated frequencies to approximate choice probabilities. *Structural analysis of discrete data with econometric applications*, 10:305–319.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, pages 1049–1051.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026.

Oesting, M., Kabluchko, Z., and Schlather, M. (2012). Simulation of Brown-Resnick processes. *Extremes*, 15(1):89–107.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pages 1027–1057.

Penrose, M. D. (1992). Semi-min-stable processes. *The Annals of Probability*, 20(3):1450–1463.

Resnick, S. (1987). Extreme values, regular variation, and point processes.

Ribatet, M. and Padoan, S. (2008). Spatialextremes: An R package for modelling spatial extremes.

Sang, H. and Genton, M. G. (2013). Tapered composite likelihood for spatial max-stable models. *Spatial Statistics*.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Scott, D. W. (1992). Multivariate density estimation, 317 pp.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(S1):S63–S84.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript, University of North Carolina*.

Terrell, G. R. and Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wand, M. and Jones, M. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, 88(422):520–528.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*, volume 60. Crc Press.

Wand, M. P., Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86(414):343–353.

Chapter 4

Spatial risk measures and applications to max-stable processes

The risk of extreme environmental events is of great importance for both the authorities and the insurance industry. This paper concerns risk measures in a spatial setting, in order to introduce the spatial features of damages stemming from environmental events in the measure of the risk. We develop a new concept of spatial risk measure, based on the spatially aggregated loss over the region of interest, and propose an adapted set of axioms for these spatial risk measures. These axioms quantify the sensitivity of the risk measure with respect to the space and are especially linked to spatial diversification. In order to model the loss underlying our definition of spatial risk measure, we apply a damage function to the environmental variable considered. The latter is assumed to follow a max-stable process, very well suited for the modeling of extreme spatial events. Two damage functions are considered, respectively adapted to temperatures and wind speeds. The theoretical properties of the resulting examples of spatial risk measures are studied and some interpretations in terms of insurance are provided.

Key words: Extreme value theory; Diversification; Max-stable processes; Mixing properties; Risk measures; Spatial dependence.

4.1 Introduction

It is of prime importance for authorities as well as for (re)insurance companies to take the spatial features of environmental risks into account. For authorities, it is crucial to be able to detect the risky zones: is it safer to build houses in this area or somewhere else? In the same way, an insurance/reinsurance company has to choose its portfolio size as well as its geographical zone of activity. The (re)insurance company wants to know where it is safer to underwrite contracts, if it should underwrite contracts on a small or a vast region, if it has an interest in being active in different regions. The last two issues are obviously related to spatial diversification. Thus, tools (especially risk measures) able to quantitatively deal with spatial diversification are needed.

The notion of risk measure has been widely studied in the literature. A risk measure Π is a mapping from a cone of random variables to the real numbers, that satisfies some axioms. The seminal paper by Artzner et al. (1999) introduced the concept of coherent risk measure, which was then generalized to the convex case by Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002). This static framework for risk measures was

then extended to the conditional and the dynamic setting. For a detailed overview about conditional and dynamic risk measures, we refer to Acciaio and Penner (2011). The most often used risk measure in the regulatory context is the Value at Risk (VaR).

The aforementioned risk measures are univariate. In $\mathbb{R}$, the natural order allows to easily define the notion of quantile and therefore the VaR. However, in dimension higher than 2, the lack of such a natural order thwarts a straightforward generalization of the notion of quantile. This is why many different definitions of multivariate quantiles have emerged in the literature. One can especially distinguish multivariate quantile functions based on depth functions (see e.g. Zuo and Serfling (2000)), multivariate quantiles based on norm minimization, also called geometric quantiles (see e.g. Chaudhuri (1996)) and multivariate quantiles as inversions of mappings (see e.g. Koltchinskii (1997)). For a very detailed review on multivariate quantiles, we refer to Serfling (2002). Regarding the extension of the the VaR to the multivariate setting, see e.g. Embrechts and Puccetti (2006) as well as Cousin and Di Bernardino (2013).

To the best of our knowledge, only Föllmer (2014) and Föllmer and Klüppelberg (2014) use the expression *spatial risk measure*. At each node of a network of financial institutions, they carry out a local conditional risk assessment in the sense that the risk measure applied takes into account the situation at the other nodes. The main issue they raise is whether the local risk assessments can be aggregated in a consistent way in order to provide a global risk measure.

Since risk measures initially appeared to deal with financial risks, they do not make explicit, in an insurance context, the influence of the region where the contracts were underwritten. However, in an insurance or a reinsurance portfolio, this particular region has a huge impact on the risk undertaken by the company. In this paper, we introduce a new notion of spatial risk measure by explicitly disentangling, in the measure of the risk, the spatial region and the hazard generating losses over this region. Then, we study how the measure of the risk is expected to evolve with respect to some of the features of the spatial region, such as its location and its size. This leads to a set of axioms adapted to the spatial context. Contrary to the axioms proposed by Artzner et al. (1999), we study the sensitivity of the measure of the risk with respect to space. For instance, our spatial homogeneity axiom relates the measure of the risk associated to a given region with that of this region to which an homothety was applied.

The analogy in a times series context is the sensitivity of the measure of risk with respect to the time horizon and is referred to as the term structure of risk measures. It is linked to temporal diversification and is of course of interest for banks as well (re)insurance companies. Let $L_{t,t+h}$ be the loss of a financial institution within the period $[t, t+h]$. For example, the homogeneity property with respect to the time horizon involves comparing $\Pi(L_{t,t+\lambda h})$ and $\Pi(L_{t,t+h})$, for all $\lambda > 0$. The literature gives some results about this term structure in the particular cases of the VaR and the Expected Shortfall (ES). If the prices variations are independent, identically distributed and Gaussian, we have $\mathrm{VaR}_{t,t+\lambda}(\alpha) = \sqrt{\lambda}\ \mathrm{VaR}_{t,t+1}(\alpha)$, where $\mathrm{VaR}_{t,t+h}(\alpha)$ is the VaR relative to the loss $L_{t,t+h}$, at the level $\alpha > 0$. If the prices variations follow an autoregressive process of order 1, an analytic expression is also available. Apart from these two cases, only a few offer a closed formula. Guidolin and Timmermann (2006) carry out a comparison of the term structure of risk measures such as the VaR and the ES under different econometric models. Except for quite simple models, closed formulas are not available and thus they use simulation methods.

Let us denote by $A \subset \mathbb{R}^2$ the region under consideration and by $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ the

process of the economic (or insured) cost due to a particular environmental hazard (e.g. wind). The easiest approach to build spatial risk measures is to integrate the classical existing risk measures (e.g. the variance or the VaR) over A, i.e. to consider $\frac{1}{|A|}\int_A \Pi[C_P(\mathbf{x})]\, d\mathbf{x}$. However, if the process C_P is strictly stationary, then the distribution of $C_P(\mathbf{x})$ is independent of $\mathbf{x}$ and thus for any risk measure Π, the previous quantity is equal to $\Pi[C(\mathbf{0})]$. The corresponding spatial risk measure reduces to the classical risk measure associated to a single site, meaning that this approach does not account for the spatial dependence structure of the process of cost. In order to overcome this defect, we define our spatial risk measure by applying a univariate static risk measure to the normalized aggregated loss over A.

After having defined our notion of spatial risk measure and the corresponding set of axioms, we introduce a model for the process of the cost. This model involves a mapping of the environmental variable under consideration to an economic (or insured) loss via a damage function. In a context of climate change some extreme events tend to be more and more frequent; see e.g. SwissRe (2014). It is of prime importance for authorities as well as for the insurance industry to assess the risk of natural disasters. In the second case, a precise assessment of the risk of extreme events is crucial in order to satisfy capital requirements relative to Solvency II. Therefore, due to the spatial feature of the environmental events, we model the process of the environmental variable using max-stable processes, which constitute an extension of the extreme value theory to the level of stochastic processes (de Haan, 1984; de Haan and Pickands, 1986; Resnick, 1987).

The remaining of the paper is organized as follows. Our concept of spatial risk measure and its corresponding set of axioms are introduced in Section 4.2. Then Section 4.3 describes our model for the economic (or insured) loss. Section 4.4 studies a concrete example of spatial risk measure adapted to heatwaves and Section 4.5 does the same in the case of windstorms. Section 4.6 concludes. All proofs are gathered in the Appendix.

4.2 Spatial risk measures

Let us denote by $\mathcal{A}$ the set of all measurable subsets of $\mathbb{R}^2$ whose Lebesgue measure is positive: $\mathcal{A} = \{A \text{ measurable} : A \subset \mathbb{R}^2 \text{ and } |A| > 0\}$, where $|.|$ denotes the Lebesgue measure in $\mathbb{R}^2$. Denote by $\mathcal{P}$ a family of distributions of stochastic processes on $\mathbb{R}^2$ having continuous sample paths. Each process represents the economic or insured loss caused by the events belonging to specified classes and occurring during a given time period, say $[0, T_L]$. In the following, T_L is considered as fixed and does not appear anymore for the sake of parsimony in the notations. The events considered here have a spatial extent and thus it is natural to consider a loss process on $\mathbb{R}^2$. Each class of events (e.g. heat wave, hurricane, earthquake, hail storm) will be referred to as a hazard in the following. Let $\mathcal{L}$ be the set of all positive-valued and bounded random variables and $\mathcal{L}_\Pi$ the set of all real-valued and bounded random variables, both defined on a measurable space $(\Omega, \mathcal{F})$. A risk measure typically will be some function $\Pi : \mathcal{L}_\Pi \mapsto \mathbb{R}$.

4.2.1 Definitions

We first give the definition of the spatially aggregated loss, which allows to disentangle the contribution of the space and the contribution of the hazards. Indeed, in the case of

an insurance company, the total loss in a portfolio of risks depends both on the region where the policies have been underwritten and on the hazards covered in these policies.

Definition 4.1 (Spatially aggregated loss)**.** *For $A \in \mathcal{A}$ such that $|A| < +\infty$ and $P \in \mathcal{P}$, the spatially aggregated loss over region A associated with the hazards whose financial costs are characterized by distribution P is defined as follows:*

$$L(A, P) = \int_A C_P(\mathbf{x})\ d\mathbf{x}, \tag{4.1}$$

where the stochastic process $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ has distribution P.

The integral (4.1) exists since $|A| < +\infty$ and the process C_P has continuous sample paths. Moreover, $L(A, P) \in \mathcal{L}$ due to the stochastic and positive nature of the process C_P. The random variable $L(A, P)$ corresponds to the total economic (or insured) loss over region A due to specified hazards and is therefore of interest for spatial risk management. As explained below, it seems more relevant, for both theoretical study as well as practical interpretation, to consider the normalized version of the spatially aggregated loss.

Definition 4.2 (Normalized spatially aggregated loss)**.** *For $A \in \mathcal{A}$ and $P \in \mathcal{P}$, the normalized spatially aggregated loss is defined by*

$$L_N(A, P) = \frac{\displaystyle\int_A C_P(\mathbf{x})\ d\mathbf{x}}{|A|}, \tag{4.2}$$

where the stochastic process $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ has distribution P.

The normalized spatially aggregated loss is a loss per surface unit and can be interpreted in a discrete setting as the loss per insurance policy.

Using the concept introduced in Definition 4.2 , we now define our notion of spatial risk measure, which makes explicit the contribution of the space in the risk measurement.

Definition 4.3 (Spatial risk measure)**.** *A spatial risk measure is a function R_Π that associates to any region $A \in \mathcal{A}$ and to any distribution $P \in \mathcal{P}$ a real number:*

$$\begin{array}{rccc} R_\Pi & : & \mathcal{A}\times\mathcal{P} & \to & \mathbb{R} \\ & & (A, P) & \mapsto & R_\Pi(A, P) = \Pi[L_N(A, P)], \end{array}$$

where $L_N(A, P)$ is defined in (4.2)*.*

If the distribution P of the economic (or insured) loss process is given, then the function $R_\Pi(., P)$ summarizes, for any region belonging to $\mathcal{A}$, the risk caused by the hazards characterized by P. In the following, $R_\Pi(., P)$ will be referred to as the spatial risk measure induced by P. The above definition takes the spatial dependence structure of the process C_P into account, except in the trivial case of the expectation.

It appears now natural to analyze how $R_\Pi(A, P)$ evolves with respect to A for a given P. Some desirable properties of $R_\Pi(., P)$ are described in the set of axioms presented below. The spatial properties of $R_\Pi(., P)$ depend both on the risk measure Π (variance, Value at Risk (VaR), Expected Shortfall (ES), ...) and the probabilistic properties of the economic loss process characterized by P.

4.2.2 A set of axioms for spatial risk measures

This section provides a set of axioms in the context of spatial risk measures as introduced above. These axioms concern the spatial risk measures properties with respect to the space and not to the economic loss distribution, the latter being considered as given by the problem at hand.

Definition 4.4 (Set of axioms for spatial risk measures)**.** *For a fixed $P \in \mathcal{P}$, we define the following axioms for the spatial risk measure induced by P:*

1. ***Spatial invariance under translation:*** $\forall \mathbf{v} \in \mathbb{R}^2$ *and* $\forall A \in \mathcal{A},\ R_\Pi(A + \mathbf{v}, P) = R_\Pi(A, P)$*, where $A + \mathbf{v}$ denotes the region A translated by the vector $\mathbf{v}$;*
2. ***Spatial sub-additivity:*** $\forall A_1, A_2 \in \mathcal{A},\ R_\Pi(A_1 \cup A_2, P) \leq \min[R_\Pi(A_1, P), R_\Pi(A_2, P)]$*;*
3. ***Asymptotic spatial homogeneity of order $-\alpha, \alpha \in \mathbb{R}$:*** $\forall A \in \mathcal{A}$,
$$R_\Pi(\lambda A, P) \underset{\lambda \to \infty}{\sim} \left(K_1 + \frac{K_2}{\lambda^\alpha}\right) R_\Pi(A, P), \tag{4.3}$$
where λA is the area obtained by applying an homothety of rate λ to A with respect to its center, $K_1 \in \mathbb{R}$ and $K_2 > 0$;
4. ***Spatial anti-monotonicity:*** $\forall A_1, A_2 \in \mathcal{A},\ A_1 \subset A_2 \Rightarrow R_\Pi(A_2, P) \leq R_\Pi(A_1, P)$.

It is easy to derive the two following statements.

Proposition 4.1. *The properties of spatial sub-additivity and spatial anti-monotonicity are equivalent.*

Proposition 4.2. *In the case of a strictly stationary process C_P, there is spatial invariance under translation.*

As stated in Proposition 4.2, the axiom of spatial invariance under translation is natural when the the process $\{C_P(\mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^2}$ is strictly stationary. Our spatial sub-additivity axiom means that the risk associated to the normalized spatially aggregated loss is lower when considering the union of two regions instead of only one of these. It indicates that there is spatial diversification, which appears as a natural property. If this axiom is satisfied, an insurance company has interest to underwrite policies on both regions A_1 and A_2 since it decreases its risk per policy. Obviously, the spatial anti-monotonicity axiom is also linked to the concept of spatial diversification. As we will see in the following, the axiom of asymptotic spatial homogeneity of order $-\alpha$ can be satisfied especially for Π being the variance or the VaR. When region A is multiplied by the factor λ, its area is multiplied by λ^2. The analogy in a discrete setting would be to multiply the number of insurance contracts by n^2. In the case of a portfolio composed of independent and identically distributed (i.i.d.) risks, this procedure decreases the variance by a factor n^2. By analogy, we can expect $K_1 = 0$ and $\alpha = 2$ in (4.3). This axiom constitutes a suggestion of spatial diversification behavior but other types of homogeneity properties could be introduced.

Though there are some links between our notion of spatial risk measures and financial risk measures as for instance summarized in Föllmer and Schied (2004), the inclusion of the space and the process C_p in Definition 4.3 sets our approach rather aside.

In order to build concrete examples of spatial risk measures, we need a model for the economic loss process C_P. Such a model is developed in the following section.

4.3 A model for the economic (or insured) loss process

4.3.1 The model

Our model for the economic loss process $\{C_P(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ requires two components. The first one concerns the loss generating hazards. We assume that the economic loss is only due to a unique class of events, i.e. to a unique hazard. In the following, we consider a natural hazard (e.g. a heat wave, a windstorm, an earthquake ...) described by the stochastic process of an environmental variable (respectively the temperature, the windspeed, the magnitude, ...), denoted by $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$.

The second component involves a model mapping the natural hazard into a damage and thus an economic cost. That model requires both the destruction percentage and the exposure at each location. The destruction percentage is obtained by applying a damage function (also referred to as vulnerability curve in the literature), denoted by $D_{\mathbf{x}}(.)$, to the natural hazard. The damage function depends on $\mathbf{x}$ since it is specific to the type of building at location $\mathbf{x}$ (see e.g. Khanduri and Morrow (2003) in the case of wind). For instance, a wooden structure will be much more easily damaged than a structure in reinforced concrete. Furthermore, note that the damage function is specific to the type of hazard considered. Thus, in order to be sufficiently realistic, the damage function should be indexed by the type of hazard considered. This is not done for the sake of notational simplicity.

The exposure process, denoted by $\{E(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$, can be considered as deterministic and involves especially the demographic conditions (density of inhabitants), the economic conditions (building structure, building cost, wealth per inhabitant) and the topographic conditions. If insured losses are of interest then the penetration rate of insurance should also be taken into account. However, this latter aspect is neglected here.

Then, the destruction percentage must be multiplied by the exposure, yielding the following model for the economic loss at location $\mathbf{x}$:

$$C_P(\mathbf{x}) = E(\mathbf{x})\ D_{\mathbf{x}}\left[Z(\mathbf{x})\right]. \tag{4.4}$$

Remark 4.1. *The presence of P in the right-hand term of* (4.4) *is implicit : the distribution P of the process $C_P(\mathbf{x})$ indeed depends on the three components of the right-hand term.*

Remark 4.2. *The temporal dimension does not appear explicitly in Model* (4.4)*. This could be addressed as follows*

$$C_P(\mathbf{x}) = \int_0^{T_L} E(\mathbf{x},t)\ D_{\mathbf{x},t}\left[Z(\mathbf{x},t)\right] dt,$$

where the environmental process, the damage function and the exposure depend on time. The damage function and the exposure should depend on time since the destruction percentage and the exposure at time t depend on what has been destroyed during the interval $[0,t)$. However, this latter aspect is very difficult to take into account and is often not dealt with in the literature about vulnerability models; see e.g. Khanduri and Morrow (2003) in the case of wind. Therefore, it is neglected in the following. Moreover, due to the difficulty of obtaining analytical results in the remaining of the paper, we ignore the temporal aspect of the environmental process. However, as explained further, the duration of the environmental events can be taken into account in our approach. The damage

function, even if independent of time, should also account for this duration. For instance, repeatedly applied wind gusts on a structure can lead to its fatigue and eventually to its destruction. For the sake of simplicity, this aspect will be ignored, as it is generally done in the literature about vulnerability curves related to windstorms or earthquakes; see e.g. Khanduri and Morrow (2003) in the case of wind.

Remark 4.3. *Model* (4.4) *involves only one environmental variable. However, in the case of floods, the damages on buildings depend both on the water level and the water velocity. Thus, in that case, two stochastic processes* $\{Z_1(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ *and* $\{Z_2(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ *should be considered.*

Due to the complexity of computations in the following, we consider that the damage function is common to all locations (thus denoted $D[.]$) and that the exposure is uniformly equal to unity. Finally, the model for the economic loss reduces to

$$C_P(\mathbf{x}) = D\left[Z(\mathbf{x})\right]. \tag{4.5}$$

Two types of damage function will successively be considered in the following:

$$D[Z(\mathbf{x})] = \mathbf{I}_{\{Z(\mathbf{x})>u\}}, \text{ where } u > 0, \tag{4.6}$$

and

$$D[Z(\mathbf{x})] = Z(\mathbf{x})^{\beta}, \text{ where } \beta > 0, \tag{4.7}$$

respectively in Sections 4.4 and 4.5. As detailed further, these two damage functions are adapted to heat waves and windstorms, respectively.

An important focus of our paper is the modelling of economic losses stemming from extreme events; this is particularly relevant for authorities as well as insurance and reinsurance companies. Therefore, at each point of the space, we consider $Z(\mathbf{x})$ being the temporal (e.g. yearly) maximum of the considered environmental variable at location $\mathbf{x}$. As explained in the next section, in that case, max-stable processes (de Haan, 1984; de Haan and Pickands, 1986; Resnick, 1987) are ideally suited for modeling purposes. For the remaining of the paper, we will assume the process $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ to be max-stable.

Definition 4.5. *(Max-stable process). For* $d \in \mathbb{N}^*$*, a stochastic process* $\{G(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ *with continuous sample paths is said to be max-stable if there exist sequences of continuous functions* $\{a_T(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} > 0$ *and* $\{b_T(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ *such that if* $G_t(\mathbf{x}), t = 1, \ldots, T$ *are i.i.d. replications of* $G(\mathbf{x})$*,*

$$\left\{\frac{\max_{t=1}^{T} G_t(\mathbf{x}) - b_T(\mathbf{x})}{a_T(\mathbf{x})}\right\}_{\mathbf{x}\in\mathbb{R}^d} \overset{d}{=} \{G(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d},$$

where $\overset{d}{=}$ *stands for equality in distribution.*

The notion of max-stability indicates that the distribution is invariant when applying the max operator with a suitable normalization. Max-stable processes arise as a natural extension of multivariate extreme value distributions to the level of stochastic processes. They are at the crossroads of geostatistics and extreme value theory. The next section provides a description of their main features.

4.3.2 A short introduction to max-stable processes

Motivation

Let us consider i.i.d. replications $T_i(\mathbf{x}), i = 1, \ldots, n$ of a stochastic process $\{T(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ having continuous sample paths. In the case where $d \in \{1, 2, 3\}$, $T(\mathbf{x})$ can for instance be some environmental variable. Let $\{c_n(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} > 0$ and $\{d_n(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ be sequences of continuous functions. Then de Haan (1984) shows that if there exists a non degenerate process $\{G(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ such that

$$\left\{\frac{\max_{i=1}^n T_i(\mathbf{x}) - d_n(\mathbf{x})}{c_n(\mathbf{x})}\right\}_{\mathbf{x}\in\mathbb{R}^d} \xrightarrow{d} \{G(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}, \text{ for } n \to \infty, \tag{4.8}$$

then $G(\mathbf{x})$ is necessarily max-stable. Therefore, max-stable processes are very well suited to characterize the joint behavior of the temporal maxima at all points of the space. Choosing our environmental process $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ to be max-stable implies that the loss given by (4.1) corresponds to the spatial aggregation of the losses caused by the worst events happening at each location during the time period considered.

By considering the maximum, one could think that one does not take into account the duration during which this maximum has occurred. This issue is solved by choosing data at the appropriate time scale. For instance, in the case of wind, instead of using the 3 seconds average for the gusts, we can use the 10 seconds average or even an average over a few hours. Thus the duration of events could be implicitly incorporated in (4.5) if the damage function was able to account for the event duration.

For practical purposes, the number n of observations over which the maxima are taken depends on the length T_L of the time period considered. We assume that the limit in (4.8) has been reached. This assumption of course needs some statistical justification in the examples analyzed. Classically, $T_L = 1$ year and $n = 365$. However, an insurer can be more interested in the loss due to some particular events than in the one corresponding to the worst event in the year. In that case, we can for instance take $T_L = 1$ week. It is still relevant to use max-stable processes in this context, since a particular event (e.g. a storm) realizes the maximum of a specific variable over a week.

Comments about the definition

As a direct consequence of Definition 4.5, the one-dimensional marginal distributions are max-stable and hence belong to the class of the generalized extreme value distributions (GEV):

$$\forall \mathbf{x} \in \mathbb{R}^d,\ G(\mathbf{x}) \sim GEV\left[\mu(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x})\right],$$

where $\{\mu(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$, $\{\sigma(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ and $\{\xi(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ are respectively the deterministic processes of location, scale and shape parameters. For a detailed overview about extreme value theory, we refer to Embrechts et al. (1997), Coles (2001), Beirlant et al. (2006) and de Haan and Ferreira (2007). Most regions under study are quite large and thus the marginal distribution generally varies over the region. The variation of the univariate distribution parameters can be viewed as regional spatial effects. In order to simplify the computations, we only consider in this paper the dependence remaining after accounting for the marginal effects. This residual dependence is related to the spatial extent of individual extreme events and comes from the fact that multiple sites are affected by the same event. Max-stable models introduced further aim at modeling this residual dependence. The margins are assumed to be standard Fréchet, i.e.

$\forall \mathbf{x} \in \mathbb{R}^d,\ \forall z > 0, \mathbb{P}(Z(\mathbf{x}) \leq z) = \exp\left(-\frac{1}{z}\right)$; see e.g. Smith (1990). A max-stable process having standard Fréchet margins will be referred to as a simple max-stable process in the following.

Spectral representations and max-stable models

A probabilistic construction of max-stable processes is given by de Haan (1984) via the spectral representation. A second representation is due to Penrose (1992) and Schlather (2002). We denote by $C^+(\mathbb{R}^d)$ the set of positive processes having continuous sample paths on $\mathbb{R}^d$.

First representation (de Haan, 1984):
Let $\{(\xi_i, \mathbf{c_i})\}_{i\geq 1}$ be the points of a Poisson point process on $(0, +\infty) \times \mathbb{R}^d$ with intensity measure $d\Lambda(\xi, \mathbf{c}) = \xi^{-2} d\xi \nu(d\mathbf{c})$, where ν is a σ-finite measure on $\mathbb{R}^d$. If $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a simple max-stable process in $C^+(\mathbb{R}^d)$, then there exists a family of continuous non negative functions $\{f_{\mathbf{x}}(\mathbf{c}) : \mathbf{c} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d\}$ satisfying

- for each $\mathbf{x} \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} f_{\mathbf{x}}(\mathbf{c})\ \nu(d\mathbf{c}) = 1$,
- for each compact $K \subset \mathbb{R}$, $\int_{\mathbb{R}^d} \sup_{\mathbf{x}\in K} f_{\mathbf{x}}(\mathbf{c})\ \nu(d\mathbf{c}) < +\infty$,

such that

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \stackrel{d}{=} \{\max_{i\geq 1} \xi_i f_{\mathbf{x}}(\mathbf{c_i})\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{4.9}$$

Conversely, each process defined by the right-hand side of(4.9) is a simple max-stable process.

The proof of the direct statement can be found in de Haan (1984) for $\mathbf{x} \in \mathbb{R}_+$ and $\mathbf{c} \in [0, 1]$ as well as in de Haan and Ferreira (2007). In the latter case, the proof is given for $\mathbf{x} \in \mathbb{R}$, $\mathbf{c} \in\ [0, 1]$ and ν being the Lebesgue measure. The converse statement is not difficult to prove.

Second representation (Penrose, 1992; Schlather, 2002; de Haan and Ferreira, 2007):
Let $\{\xi_i\}_{i\geq 1}$ be the points of a Poisson point process on $(0, +\infty)$, with intensity $d\Lambda(\xi) = \xi^{-2} d\xi$ and $Y_1(\mathbf{x}), Y_2(\mathbf{x}), \ldots$ i.i.d. replications of a stochastic process $\{Y(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ such that $\mathbb{E}(\max[0, Y(\mathbf{x})]) = 1$ for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbb{E}(\sup_{\mathbf{x}\in K} \max[0, Y(\mathbf{x})]) < +\infty$ for any compact $K \subset \mathbb{R}^d$. Then any simple max-stable process in $C^+(\mathbb{R}^d)$ can be written as

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} \stackrel{d}{=} \{\max_{i\geq 1}\{\xi_i \max[0, Y_i(\mathbf{x})]\}\}_{\mathbf{x}\in\mathbb{R}^d}. \tag{4.10}$$

Conversely, each process defined by the right-hand side of (4.10) is a simple max-stable process.

Remark 4.4. *If Y is a strictly stationary process, then* (4.10) *defines a strictly stationary process.*

In the paper by Schlather (2002), we find the converse statement. For a proof of the fact that any simple max-stable process in $C^+(\mathbb{R}^d)$ can be written as in (4.10), we refer

to de Haan and Ferreira (2007). Note that de Haan and Ferreira (2007) consider $[0,1]$ instead of $\mathbb{R}^d$ for convenience.

These two representations have led to different models for max-stable processes, presented in the following.

The Smith model:
In an unpublished manuscript, Smith (1990) uses the spectral representation (4.9) to provide a parametric model for max-stable processes. He considers a particular setting where ν is the Lebesgue measure on $\mathbb{R}^d$ and $f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma}}(\mathbf{x}-\mathbf{c})$, where $f_{\mathbf{\Sigma}}$ is the density of a d-variate normal law with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$:

$$f_{\mathbf{x}}(\mathbf{c}) = f_{\mathbf{\Sigma}}(\mathbf{x}-\mathbf{c}) = (2\pi)^{-\frac{d}{2}}|\mathbf{\Sigma}|^{-\frac{d}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{c})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{c})\right]. \tag{4.11}$$

The parameter is the covariance matrix $\mathbf{\Sigma}$, which contains all the information about the spatial dependence structure. A nice feature of this model lies in its nice interpretation in terms of rainfall-storm processes (Smith, 1990), the shape of these storms being driven by the covariance matrix. Moreover, in the case $d = 2$, the trivariate density (the density of an observation at 3 sites) can be explicitly written (see e.g. Genton et al. (2011)) contrary to the Schlather model below.

The Schlather model:
Schlather (2002) proposes to set $Y(\mathbf{x}) = \sqrt{2\pi}\epsilon(\mathbf{x})$ in (4.10), where $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a stationary standard Gaussian process with any correlation function $\rho(.)$. All correlation functions stemming from the geostatistical literature can be used, allowing for a rich diversity of behaviors. We will contrast and compare the following correlation families:

Whittle-Matern:	$\rho(h) = \frac{2^{1-c_2}}{\Gamma(c_2)}\left(\frac{h}{c_1}\right)^{c_2} K_{c_2}\left(\frac{h}{c_1}\right),\ c_1>0,\ c_2>0,$
Exponential:	$\rho(h) = \exp\left[-\frac{h}{c_1}\right],\ c_1>0,$
Cauchy:	$\rho(h) = \left[1+\left(\frac{h}{c_1}\right)^2\right]^{-c_2},\ c_1>0,\ c_2>0,$
Powered ewponential:	$\rho(h) = \exp\left[-\left(\frac{h}{c_1}\right)^{c_2}\right],\ c_1>0,\ 0<c_2<2,$

where c_1 and c_2 are the range and the smoothing parameters, Γ is the Gamma function and K_{c_2} is the modified Bessel function of the third kind of order c_2.

The geometric Gaussian model:
Independence is unreachable in the case of the Schlather model (see Section 4.4). To deal with this issue, Davison (2003) introduces the geometric Gaussian model. He takes in (4.10) $\{Y(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ a log normal process and not a Gaussian process:

$$Y(\mathbf{x}) = \exp\left(\sigma\epsilon(\mathbf{x}) - \frac{\sigma^2}{2}\right),$$

where $\{\epsilon(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a standard Gaussian process with variance σ^2 and correlation function $\rho(.)$.

The Brown-Resnick model:
The geometric Gaussian process is a particular case of a model introduced by Brown and Resnick (1977). Kabluchko et al. (2009) introduce a generalization of the latter model, which they referred to as the Brown-Resnick model, by taking in (4.10):

$$Y(\mathbf{x}) = \exp\left(W(\mathbf{x}) - \frac{\sigma^2(\mathbf{x})}{2}\right),$$

where $\{W(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is a zero-mean Gaussian process with stationary increments and $\sigma^2(\mathbf{x}) = \text{var}[W(\mathbf{x})],\ \forall \mathbf{x} \in \mathbb{R}^d$. The process W and therefore the resulting Brown-Resnick process are completely characterized by the variance $\sigma(\mathbf{x})$ and the semi-variogram, defined by

$$\gamma(\mathbf{h}) = \frac{1}{2}\text{var}[W(\mathbf{x}+\mathbf{h}) - W(\mathbf{x})], \forall \mathbf{h} \in \mathbb{R}^d, \tag{4.12}$$

where var stands for the variance. Note that both the Smith process and the geometric Gaussian process are particular cases of the Brown-Resnick process. This is clear in the case of the geometric Gaussian process since the standard Gaussian process has stationary increments. In the case of the Smith process, see e.g. Yuen and Stoev (2013), p.6.

Extremal coefficient

The extremal coefficient (Schlather and Tawn, 2003) is a measure of spatial dependence for max-stable processes and will play an important role in the study of concrete examples of spatial risk measures carried out in Sections 4.4 and 4.5 . In the case of M locations $(\mathbf{x}_1, \ldots, \mathbf{x}_M)$, it is denoted by $\Theta(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ and is defined by

$$\mathbb{P}\left(Z(\mathbf{x}_1) \leq u, \ldots, Z(\mathbf{x}_M) \leq u\right) = \exp\left(-\frac{\Theta(\mathbf{x}_1, \ldots, \mathbf{x}_M)}{u}\right). \tag{4.13}$$

In the case $M = 2$, if Z is strictly stationary, then $\Theta(\mathbf{x}_1, \mathbf{x}_2)$ only depends on the vector $\mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2$ and is denoted by $\Theta(\mathbf{h})$. Moreover, if Z is isotropic, $\Theta(\mathbf{x}_1, \mathbf{x}_2)$ only depends on $h = \|\mathbf{h}\|$, the Euclidean distance between sites $\mathbf{x}_1$ and $\mathbf{x}_2$, and is denoted by $\Theta(h)$.

Ergodicity, mixing properties and extremal coefficient

Results about spatial diversification we present in Section 4.4 are expressed in terms of the extremal coefficient. Moreover, spatial diversification is linked with notions of ergodicity and mixing. Thus, in order to link our results with the existing literature about mixing properties of max-stable processes, let us briefly discuss the established links between the extremal coefficient behavior and the properties of ergodicity and mixing.

We first recall basic concepts of ergodicity and mixing. Let us consider a stationary stochastic process $\{R(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ such that $\mathbb{E}[R(\mathbf{x})] < +\infty$ and denote $\mathcal{F}_A$ the σ-field generated by the random variables $\{R(\mathbf{x}) : \mathbf{x} \in A\}$, for some region $A \subset \mathbb{R}^d$.

Definition 4.6. *(Mean-ergodicity). The stochastic process $\{R(\mathbf{x}, \omega)\}_{\mathbf{x}\in\mathbb{R}^d}$ is said mean-ergodic if, for $A \subset \mathbb{R}^d$,*

$$\lim_{|A|\to\infty} \mathbb{E}\left[\left(\frac{1}{|A|}\int_A R(\mathbf{x}, \omega)\ d\mathbf{x} - \mu\right)^2\right] = 0,$$

where $\mu = \mathbb{E}(R)$.

In other words, the spatial mean (for one fixed realization ω) converges to the expectation in quadratic mean. This property means that for a region with infinite size, only one realization of the process is sufficient to compute its expectation.

Definition 4.7 (Strong mixing)**.** *The α-mixing coefficient (or strong mixing coefficient by Rosenblatt (1956)) between the σ-fields $\mathcal{F}_{A_1}$ and $\mathcal{F}_{A_2}$ is defined by*

$$\alpha(A_1, A_2) = \sup\{|\mathbb{P}(S_1 \cap S_2) - \mathbb{P}(S_1)\mathbb{P}(S_2)| : S_1 \in \mathcal{F}_{A_1}, S_2 \in \mathcal{F}_{A_2}\}.$$

The process R is said to be strongly mixing if $\alpha(A_1, A_2)$ tends to 0 as $d(A_1, A_2)$ tends to $+\infty$, where $d(A_1, A_2)$ is the distance between A_1 and A_2.

Other equivalent definitions can be found, e.g. in Kabluchko and Schlather (2010) (Definition 2.1, Bullet 3). Their definition in the case of $\mathbb{R}$ can be easily extended to $\mathbb{R}^d$.

First results about ergodicity of max-stable processes are due to Weintraub (1991). Then Stoev (2010) uses the extremal integral representation to derive necessary and sufficient conditions for mixing of max-stable processes. Kabluchko and Schlather (2010) extend these results to the class of max-infinitely divisible processes. Dombry and Eyi-Minko (2012) find a simple upper bound for the β-mixing coefficient (or absolute regularity coefficient by Volkonskii and Rozanov (1959)), $\beta(A_1, A_2)$, between two subsets A_1 and A_2. They derive a central limit theorem and show the asymptotic normality of three estimators of the extremal coefficient.

Kabluchko and Schlather (2010) study mixing properties of the class of the max-infinitely divisible processes that encompasses max-stable processes. In the max-stable case, they introduce the dependence coefficient $\{r(h)\}_{h\in\mathbb{R}}$ defined by $r(h) = 2 - \Theta(h)$, where $\Theta(h)$ is the extremal coefficient. Their Theorem 3.1 states that for a stationary measurable simple max-stable process $\{Z(x)\}_{x\in\mathbb{R}}$, Z is strongly mixing if and only if $\lim_{h\to\infty} r(h) = 0$. In their Theorem 3.2, they show that Z is ergodic if and only if $\lim_{l\to\infty} \frac{1}{l}\int_{h=1}^{l} r(h)\, dh = 0$. These results could be extended to $\mathbb{R}^d$.

To close this section, let us consider the process $\{H(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d} = \{D[Z(\mathbf{x})]\}_{\mathbf{x}\in\mathbb{R}^d}$, where D is any function.

Lemma 4.1. *If the process Z is strongly mixing (respectively ergodic), then the process H is strongly mixing (respectively ergodic).*

That means that the mixing (respectively ergodic) properties of the environmental process are also valid for the economic loss process.

4.4 Example based on the threshold damage function

In this section, we consider the normalized spatially aggregated loss obtained by combining (4.2), (4.5) and (4.6):

$$L_N(A, P) = \frac{1}{|A|}\int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\, d\mathbf{x}, \tag{4.14}$$

where $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ is a simple max-stable process and u is a positive threshold. This quantity is particularly interesting when analyzing for instance the impact of high temperatures on the distribution network of electricity as well as on populations, typically as

in the case of the European heatwave in 2003. If the threshold is well chosen, $L_N(A,P)$ represents indeed the proportion of the surface at which the temperature exceeds a dangerous threshold for the electric cables or for populations.

Note that the spatially aggregated loss $\int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\, d\mathbf{x}$ corresponds to the area, or the so-called intrinsic volume, of the excursion set $E_u(Z,A) = \{\mathbf{x} \in A : Z(\mathbf{x}) \geq u\}$. Excursion sets of stochastic processes have been widely studied in the literature. Especially Lévy processes, diffusions, stable and Gaussian processes have been investigated; see e.g. Berman (1992), references in Ivanov et al. (2013) and Spodarev (2014) for an overview.

The dependence of $L_N(.,.)$ with respect to P lies in the distribution of process Z. In the following, the different max-stable models introduced in Section 4.3.2 will be considered. In each case, the model used will be explicitly indicated and therefore, we make the dependence in P implicit in $L_N(A,P)$: from now on, $L_N(A,P)$ is denoted $L_N(A)$. In the same way, the risk measure $R_\Pi(A,P)$ is denoted $R_\Pi(A)$.

The case of the expectation $R_1(A) = \mathbb{E}\left[L_N(A)\right]$ is trivial. Here only high losses are considered, so $R_1(.)$ can for instance be the actuarial premium of reinsurance contracts. Using the linearity property of the expectation, we immediately show that $\forall A \in \mathcal{A},\ R_1(A) = 1 - \exp\left(-\frac{1}{u}\right)$, meaning that this premium does not depend on the region considered (and therefore on its size). It stems directly from the fact that the process has standardized margins. This is similar to the case of an insurance portfolio composed of homogeneous risks. The expectation is not a very useful risk measure since it does not involve any information relative to the variability. Moreover, due to linearity, it does not account for the spatial dependence of the loss process.

In the following, we study the case of the variance in details, before providing some insights about the VaR.

4.4.1 The variance

We consider in this section the quantity $R_2(A) = \mathrm{var}\left[L_N(A)\right]$. The variance allows taking into account part of the spatial dependence in the risk assessment. Therefore, its study is interesting for both the risk management of extreme spatial events and the understanding of some properties of max-stable processes. Moreover, variance is of prime interest for (re)insurance companies since it controls the variability of the normalized portfolio's loss around the expected one. As we will see, $R_2(.)$ is linked with the notions of spatial diversification, ergodicity and mixing.

In the following, our aim is to study whether $R_2(.)$ satisfies the axioms presented in Definition 4.4. Before, we study in details the function $\lambda \mapsto R_2(\lambda A)$. It is of course related to the property of asymptotic spatial homogeneity but we are also interested in the behaviour of $R_2(\lambda A)$ for finite values of λ. This study can be of practical relevance for the (re)insurance industry and moreover the results obtained will be used to prove the axioms of spatial anti-monotonicity and spatial sub-additivity.

General results about spatial homogeneity

The expression of $R_2(\lambda A)$ in the general case is given in the next theorem.

Theorem 4.1. *In the case of a simple max-stable process, $\forall A \in \mathcal{A}$, we have*

$$R_2(\lambda A) = \frac{1}{\lambda^4 |A|^2} \int_{\lambda A} \int_{\lambda A} \left[\exp\left(-\frac{\Theta(\mathbf{x}, \mathbf{y})}{u} \right) - \exp\left(-\frac{2}{u} \right) \right] \, d\mathbf{x} \, d\mathbf{y}. \tag{4.15}$$

From Theorem 4.1, we can derive the behavior of $R_2(\lambda A)$ in the case of isotropic max-stable processes, when region A is either a disk or a square. The result is given in the next corollary.

Corollary 4.1. *Consider an isotropic simple max-stable process having $\{\Theta(h)\}_{h \in \mathbb{R}_+}$ as extremal coefficient function, and $A \in \mathcal{A}$. Then:*

1. ***If* A *is a disk*** *with radius R, then*

$$R_2(\lambda A) = -\exp\left(-\frac{2}{u} \right) + \int_0^{2R} f_d(h, R) \exp\left(-\frac{\Theta(\lambda h)}{u} \right) \, dh, \tag{4.16}$$

where f_d is the density of the distance between two points uniformly distributed on A, given by

$$f_d(h, R) = \frac{2h}{R^2} \left(\frac{2}{\pi} \arccos\left(\frac{h}{2R} \right) - \frac{h}{\pi R} \sqrt{1 - \frac{h^2}{4R^2}} \right).$$

2. ***If* A *is a square*** *with side R, then*

$$R_2(\lambda A) = -\exp\left(-\frac{2}{u} \right) + \int_0^{\sqrt{2}R} f_s(h, R) \exp\left(-\frac{\Theta(\lambda h)}{u} \right) \, dh, \tag{4.17}$$

where f_s is the density of the distance between two points uniformly distributed on A, given by:
For $h \in [0, R]$,

$$f_s(h, R) = \frac{2\pi h}{R^2} - \frac{8h^2}{R^3} + \frac{2h^3}{R^4}.$$

For $h \in [R, R\sqrt{2}]$,

$$f_s(h, R) = \left(-2 - b + 3\sqrt{b-1} + \frac{b+1}{\sqrt{b-1}} + 2 \arcsin\left(\frac{2-b}{b} \right) - \frac{4}{b\sqrt{1 - \frac{(2-b)^2}{b^2}}} \right) \frac{2h}{R^2},$$

where $b = \dfrac{h^2}{R^2}$.

3. ***In both cases****, $R_2(\lambda A)$ converges as $\lambda \to \infty$ to the limiting risk measure given by*

$$-\exp\left(-\frac{2}{u} \right) + \lim_{\lambda \to \infty} \exp\left(-\frac{\Theta(\lambda h)}{u} \right). \tag{4.18}$$

Remark 4.5. *Expressions* (4.16) *and* (4.17) *could have the same structure for other types of region A (see the proof of Corollary 4.1). However, the densities f_d and f_s should be replaced by the adequate one. The appropriate density can be computed using the approach described in Moltchanov (2012) but it may not be so obvious in some cases.*

The following corollary directly follows from Corollary 4.1.

Corollary 4.2. *In the case of perfect dependence, i.e.* $\forall h, \Theta(h) = 1$*, we have*

$$R_2(A) = \exp\left(-\frac{1}{u}\right) - \exp\left(-\frac{2}{u}\right).$$

Apart from the case of perfect dependence, the function $\Theta(\lambda h)$ is strictly increasing with respect to λ, giving that $R_2(\lambda A)$ is strictly decreasing with respect to λ. Consequently, there is spatial diversification. Corollary 4.1 offers an interesting tool for the insurance industry since it allows determining the dimension of the geographical area required to reach a low variance level. As shown in the following corollary, in some cases, the diversification can be total.

Corollary 4.3. *In the case of asymptotic independence, i.e.* $\lim_{h\to\infty} \Theta(h) = 2$*, we have*

$$\lim_{\lambda\to\infty} R_2(\lambda A) = 0.$$

This result is not surprising. Indeed, as mentioned in Section 4.3.2, $\lim_{h\to\infty} \Theta(h) = 2$ implies that the process $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ is mixing, if we accept the extension from $\mathbb{R}$ to $\mathbb{R}^d$. Then every transformation of Z is also mixing (see Lemma 4.1), giving that $\mathbf{I}_{\{Z(\mathbf{x})>u\}}$ is mixing and therefore ergodic and mean-ergodic, which is equivalent to Corollary 4.3. In terms of insurance, Corollary 4.3 states that the spatial diversification can be total. If there is asymptotic spatial independence and if the insurance company can underwrite policies on a sufficiently large region, then the corresponding portfolio is "equivalent" to a portfolio containing i.i.d. risks.

Corollary 4.1 shows that the decrease of $R_2(\lambda A)$ as λ is increasing is mainly driven by the extremal coefficient $\Theta(.)$; the latter of course depends on the max-stable model under consideration. Our aim in the following is to study the influence of the factor λ for different max-stable models. However, the integrals appearing in Corollary 4.1 have no closed form and so we use a Riemann approximation.

Homogeneity for max-stable models already introduced in the literature

Using the results of Corollary 4.1, we study the behavior of the function $\lambda \mapsto R_2(\lambda A)$ in the case of the parametric models of max-stable processes introduced in Section 4.3.2. Without loss of generality, we set $R = 1$. Another issue concerns the choice of the threshold u. Clearly, $R_2(A)$ decreases as u increases. Nevertheless, u has no influence on the shape of the function $\lambda \mapsto R_2(\lambda A)$ and we choose $u = 1$.

The Smith model

For two locations $\mathbf{x}_1$ and $\mathbf{x}_2$, the extremal coefficient is given by

$$\Theta(\mathbf{x}_1 - \mathbf{x}_2) = 2\Phi\left(\frac{\sqrt{(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_1 - \mathbf{x}_2)}}{2}\right),$$

where Φ denotes the distribution function of the standard Gaussian random variable. The Smith process is isotropic only if $\boldsymbol{\Sigma}$ is proportional to the identity matrix. Without loss of generality, let $\boldsymbol{\Sigma}$ be the identity matrix. In this case, we have

$$\Theta(h) = 2\Phi\left(\frac{h}{2}\right).$$

Therefore, $\lim_{h\to\infty}\Theta(h) = 2$ and Corollary 4.3 gives that $\lim_{\lambda\to\infty} R_2(\lambda A) = 0$, meaning that the spatial diversification is total. For A being a disk, we observe in Figure 4.1 that $R_2(\lambda A)$ rapidly decreases to the limiting risk measure when λ increases.

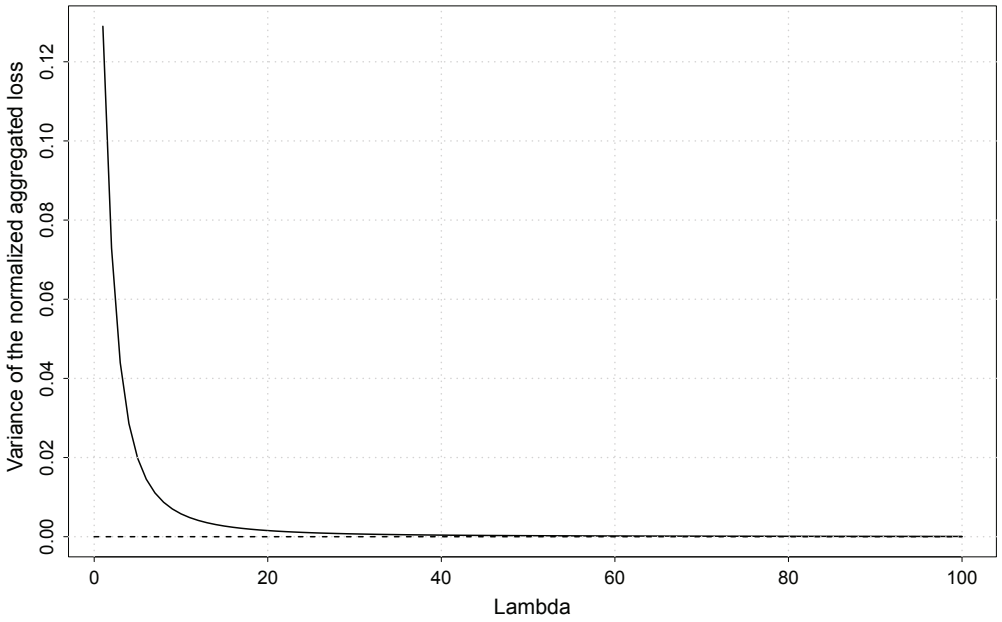


Figure 4.1: The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ in the case of the Smith model, with A being a disk. The dashed line represents the limiting risk measure.

The Schlather process

The extremal coefficient is given by $\Theta(\mathbf{x}_1 - \mathbf{x}_2) = 1 + \sqrt{\dfrac{1-\rho(\mathbf{x}_1 - \mathbf{x}_2)}{2}}$. There is isotropy if and only if the correlation function ρ is isotropic. In that case,

$$\Theta(h) = 1 + \sqrt{\frac{1-\rho(h)}{2}}.$$

Therefore, if $\lim_{h\to\infty}\rho(h) = 0$, we have $\lim_{\lambda\to\infty}\Theta(\lambda h) = 1 + \sqrt{\dfrac{1}{2}}$ and (4.18) gives that

$$\lim_{\lambda\to\infty} R_2(\lambda A) = -\exp\left(-\frac{2}{u}\right) + \exp\left(-\frac{1+\sqrt{\frac{1}{2}}}{u}\right),$$

which is different from zero.

The limiting risk measure is positive, showing that the process $\mathbf{I}_{\{Z(\mathbf{x})>u\}}$ is not mean-ergodic. This is consistent with the fact that the Schlather process is neither mixing nor ergodic (see Section 4.3.2). In terms of insurance, this result means that the spatial diversification is never total: there is always some kind of residual common risk factor.

We set the smoothing parameter $c_2 = 0.5$ and we compare two values of the range parameter: $c_1 = 1$ and $c_1 = 10$.

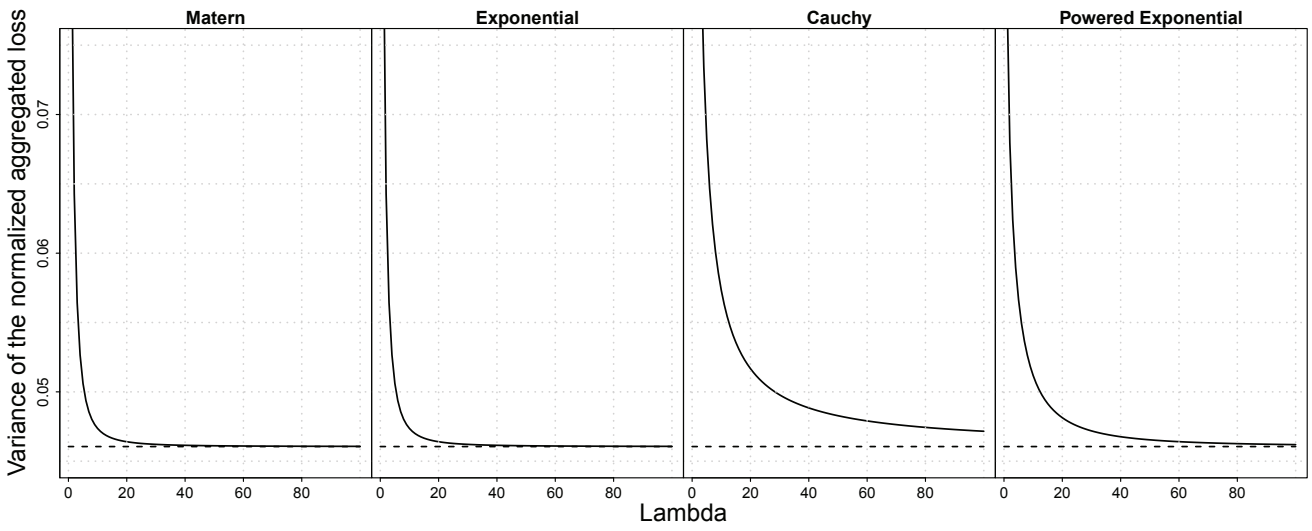


Figure 4.2: Each panel correponds to a different correlation function having a range parameter $c_1 = 1$. The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ in the case of the Schlather model, with A being a disk. The dashed line represents the limiting risk measure.

Figures 4.2 and 4.3 show that the speed of decrease of $R_2(\lambda A)$ to the limiting risk measure depends on the type of correlation function. This decrease is much slower in the cases of the Cauchy and the powered exponential functions. Therefore, the Schlather process allows for a large variety of spatial diversification behaviors.

Globally, we observe that the decrease to the limiting risk measure is slower than in the case of the Smith model. This comparison must be done with equivalent characteristic distances of the spatial correlation, meaning that the eigenvalues of $\mathbf{\Sigma}$ must be equal to c_1, which is the case for $c_1 = 1$. As expected, we see in Figure 4.3 that the decrease is slower when increasing the range parameter c_1.

The geometric Gaussian process

The extremal coefficient is given by $\Theta(\mathbf{x}_1 - \mathbf{x}_2) = 2\Phi\left(\sqrt{\frac{\sigma^2(1-\rho(\mathbf{x}_1-\mathbf{x}_2))}{2}}\right)$. Consequently, if the function ρ is isotropic, the extremal coefficient becomes

$$\Theta(h) = 2\Phi\left(\sqrt{\frac{\sigma^2(1-\rho(h))}{2}}\right).$$

Therefore, $\lim_{\lambda\to+\infty}\Theta(\lambda h) = 2\Phi\left(\sqrt{\frac{\sigma^2}{2}}\right)$, giving

$$\lim_{\lambda\to\infty} R_2(\lambda A) = -\exp\left(-\frac{2}{u}\right) + \exp\left(-\frac{2\Phi\left(\sqrt{\frac{\sigma^2}{2}}\right)}{u}\right),$$

which is different from zero. From Figure 4.4, we draw very similar conclusions as those related to the Schlather model. The only difference consists in the value of the limiting risk measure.

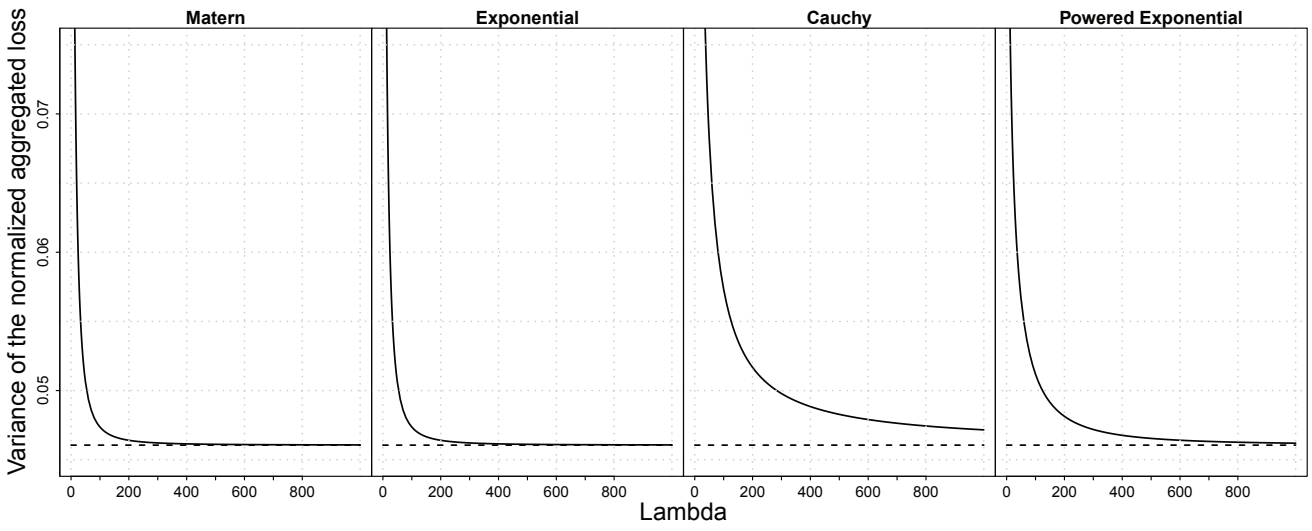


Figure 4.3: Each panel correponds to a different correlation function having a range parameter $c_1 = 10$. The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ in the case of the Schlather model, with A being a disk. The dashed line represents the limiting risk measure.

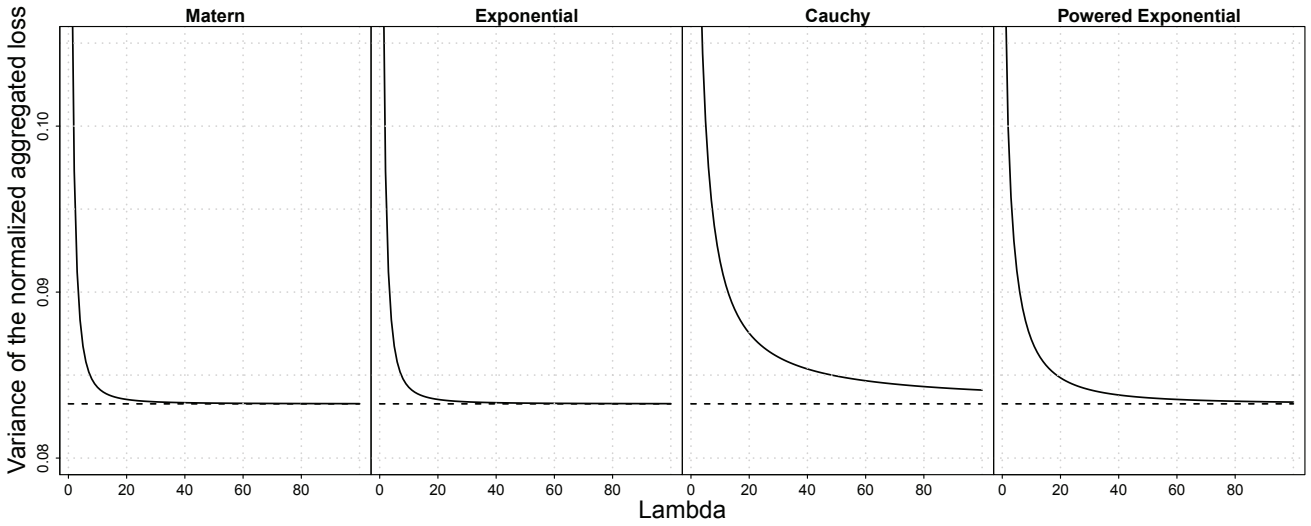


Figure 4.4: Each panel correponds to a different correlation function having a range parameter $c_1 = 1$. The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ in the case of the geometric Gaussian model, with A being a disk. The dashed line represents the limiting risk measure.

Homogeneity for a new max-stable model: the tube model

In order to allow for a faster spatial diversification, we introduce a new max-stable model, the tube model, defined hereafter.

Definition 4.8. *(The tube model). The tube model is defined using the spectral representation* (4.9)*, with ν being the Lebesgue measure:*

$$\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2} = \{\max_{i\geq 1} \xi_i f_0(\mathbf{c_i}, \mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2},$$

where $f_0(\mathbf{c}, \mathbf{x}) = f_0(\mathbf{c} - \mathbf{x}) = h_b \, \mathbf{I}_{\{\|\mathbf{c}-\mathbf{x}\|<R_b\}}$, *with* $R_b > 0$ *and* $h_b = \dfrac{1}{\pi R_b^2}$.

The last condition stems from the fact that f_0 must be a density, imposing $\pi R_b^2 h_b = 1$. The density f_0 has the shape of a tube of height h_b centered at point $\mathbf{c}$ and with radius R_b. The extremal coefficient is given in the next proposition and is depicted in Figure 4.5.

Proposition 4.3. *The extremal coefficient of the tube model is given by*

$$\Theta(h) = \begin{cases} 2\left[1 - h_b\left(R_b^2 \arcsin\left(\dfrac{\sqrt{4R_b^2 - h^2}}{\sqrt{2}R_b}\right) - \dfrac{h}{2\sqrt{2}}\sqrt{4R_b^2 - h^2}\right)\right] & \text{if } h \leq 2R_b, \\ 2 & \text{if } h > 2R_b. \end{cases}$$

An interesting property stems from the fact that the extremal coefficient $\Theta(h)$ reaches 2 whenever $h \geq 2R_b$, meaning that there is spatial independence at a finite distance. That explains why the spatial diversification is faster than in the case of the previously introduced models, as can be observed in Figure 4.6. Obviously, $\lim_{h\to\infty} \Theta(h) = 2$, and

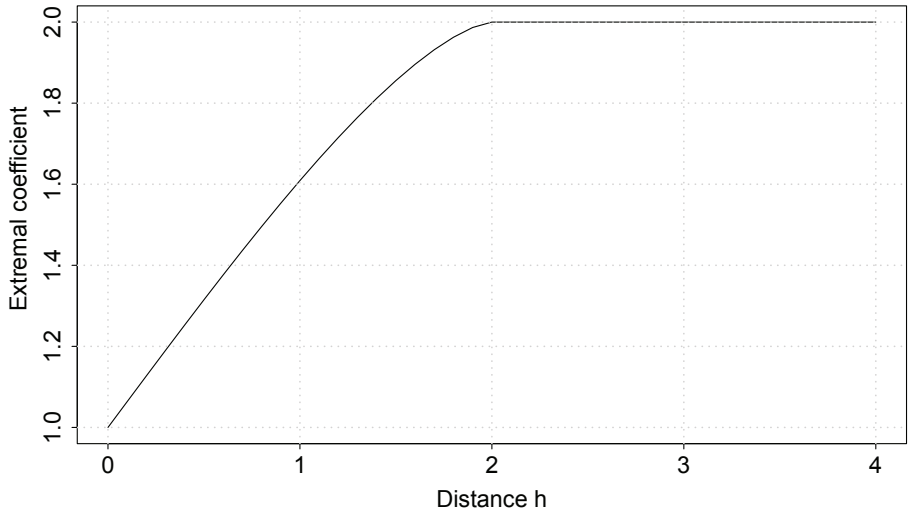


Figure 4.5: Extremal coefficient function of the tube model.

Corollary 4.3 yields $\lim_{\lambda\to\infty} R_2(\lambda A) = 0$.

Moreover, due to the spatial independence at finite distance, we have the following proposition.

Proposition 4.4. *In the case of the tube model, we have*
$\forall \lambda > 0, \lim_{R_b\to 0} R_2(\lambda A) = R_2(A) = 0$.

The limit process arising as R_b tends to 0 corresponds to the case of perfect independence.

To summarize, a comparison of the 4 models considered (Smith, Schlather, geometric Gaussian and tube) is provided in Figure 4.7. In the cases of the Schlather and the geometric Gaussian models, a Cauchy correlation function has been used. These different processes show a large variety of behaviors, both in terms of speed and "completeness"

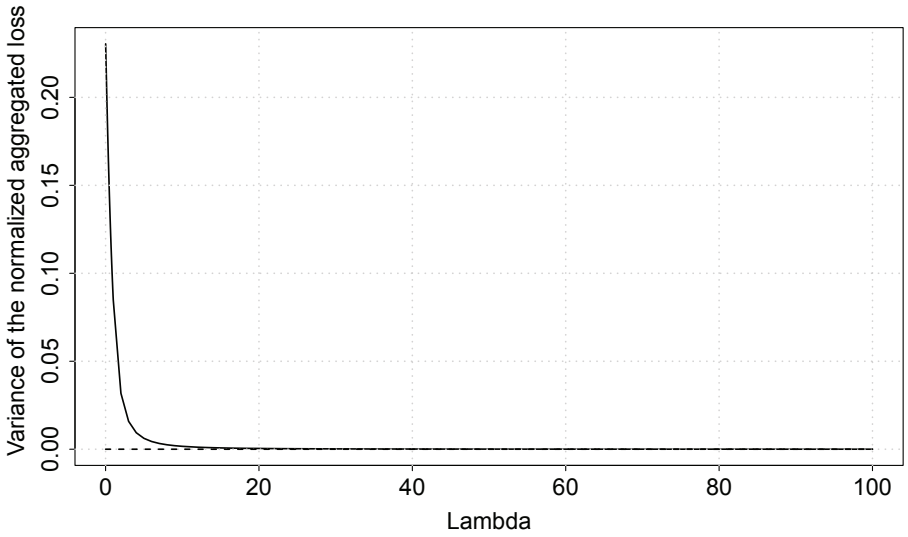


Figure 4.6: The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ in the case of the tube model, with $R_b = 1$ and A being a disk. The dashed line represents the limiting risk measure.

of spatial diversification. The Brown-Resnick process itself includes various types of behaviors. The two particular cases considered here (the Smith model and the geometric Gaussian model) are very different. In terms of spatial diversification, the optimal strategy for an insurance company highly depends on the type of max-stable process driving extreme events.

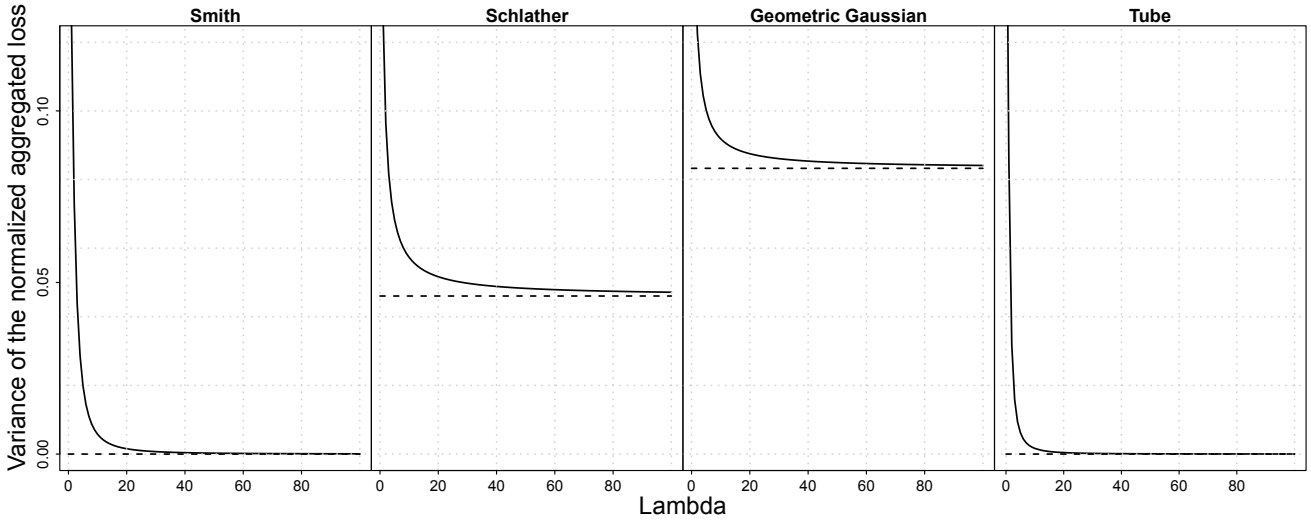


Figure 4.7: Each panel correponds to a different max-stable model. The solid line corresponds to the evolution of $R_2(\lambda A)$ with respect to λ, where A is a disk. The dashed line represents the limiting risk measure.

Remark 4.6. *It is important to note that our analysis has been carried out with a standardized dependence structure: the eigenvalues of* $\mathbf{\Sigma}$ *as well as the range parameter* c_1 *are equal to 1. However, in a real case study, the characteristic dimension of the area*

required to reach a given level of variance depends on the real values of $\mathbf{\Sigma}$ *and* c_1 *on the region of interest.*

Remark 4.7. *We have performed similar calculations for A being a square; the conclusions are very similar. The only difference consists in the fact that the spatial diversification is slightly slower.*

Central limit theorem and axioms

Using Theorem 4.1, we can show that

$$\lim_{\lambda\to\infty} \lambda^2 R_2(\lambda A) = \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}.$$

This result suggests the existence of a central limit theorem under some additional assumptions.

Since mixing conditions are generally rather difficult to check, Spodarev (2014) proposes a central limit theorem for excursion sets based on the concept of association. Using his result, we can derive the following theorem.

Theorem 4.2. *In the cases of the Smith model, the Brown-Resnick model whose variogram satisfies* $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ *with* $a > 0$*, and the tube model, we have, for all* $A \in \mathcal{A}$*,*

$$\lambda\left(L_N(\lambda A) - \left[1 - \exp\left(-\frac{1}{u}\right)\right]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \text{ for } \lambda \to \infty,$$

where

$$\sigma^2 = \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}.$$

Note that this result can also be obtained using mixing conditions obtained by Dombry and Eyi-Minko (2012) as well as general results about central limit theorems for strongly mixing random fields obtained for example by Gorodetskii (1987).

The following theorem shows that the different axioms introduced in Section 4.2.2 are satisfied in the case of some max-stable processes.

Theorem 4.3. *1. For any strictly stationary max-stable process with standard Fréchet margins, the spatial risk measure* R_2 *satisfies the following axioms:*

- *(a) Spatial invariance under translation;*
- *(b) Spatial sub-additivity when the two regions are both a disk or a square;*
- *(c) Spatial anti-monotonicity when the two regions are both a disk or a square.*

2. In the cases of the Smith model, the Brown-Resnick model whose variogram satisfies $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ *with* $a > 0$*, and the tube model, we have asymptotic spatial homogeneity of order* -2*, with*

$$K_1 = 0 \qquad \text{and} \qquad K_2 = \frac{\displaystyle\int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}}{R_2(A)}.$$

4.4.2 The Value at Risk

We focus here on $R_{3,1-\alpha}(A) = VaR_{1-\alpha}[L_N(A)]$, where $VaR_{1-\alpha}$ is the VaR at the level $1-\alpha$, for α small. It seems impossible to derive formulas (even up to an integral) for the VaR of $L_N(A)$. Therefore, we evaluate it by using Monte-Carlo techniques. The process $Z(\mathbf{x})$ is simulated on a grid containing different locations $\mathbf{x}_m \in A \in \mathcal{A}, m = 1, \ldots, M$. Then one realization of the normalized loss can be approximated by a Riemann sum. Different approximation methods are available, the most efficient being probably the trapeze method. This Riemann-based approach has the advantage of providing the convergence rate of the approximated realization to the real one, using classical results on the discretization error.

Then, by generating a number S of independent approximated replications of the random variable $L_N(A)$, an approximation of its distribution is obtained. Finally an approximation of the VaR can be obtained by taking the empirical quantile of this distribution. The uncertainty on the VaR stemming from this second step can be quantified by classical bootstrap methods. The VaR can be computed on many sub-samples of the sample $s = 1, \ldots, S$, yielding the empirical variance of the corresponding estimates. Under classical regularity conditions, confidence intervals on the VaR can be obtained.

To illustrate the procedure, the evolution of $R_{3,0.9}(\lambda A)$ with respect to λ in the case of the Smith model is displayed in Figure 4.8, for different values of M and S, where M is the number of sites in region A. Note that region λA contains $\lambda^2 M$ sites. The same

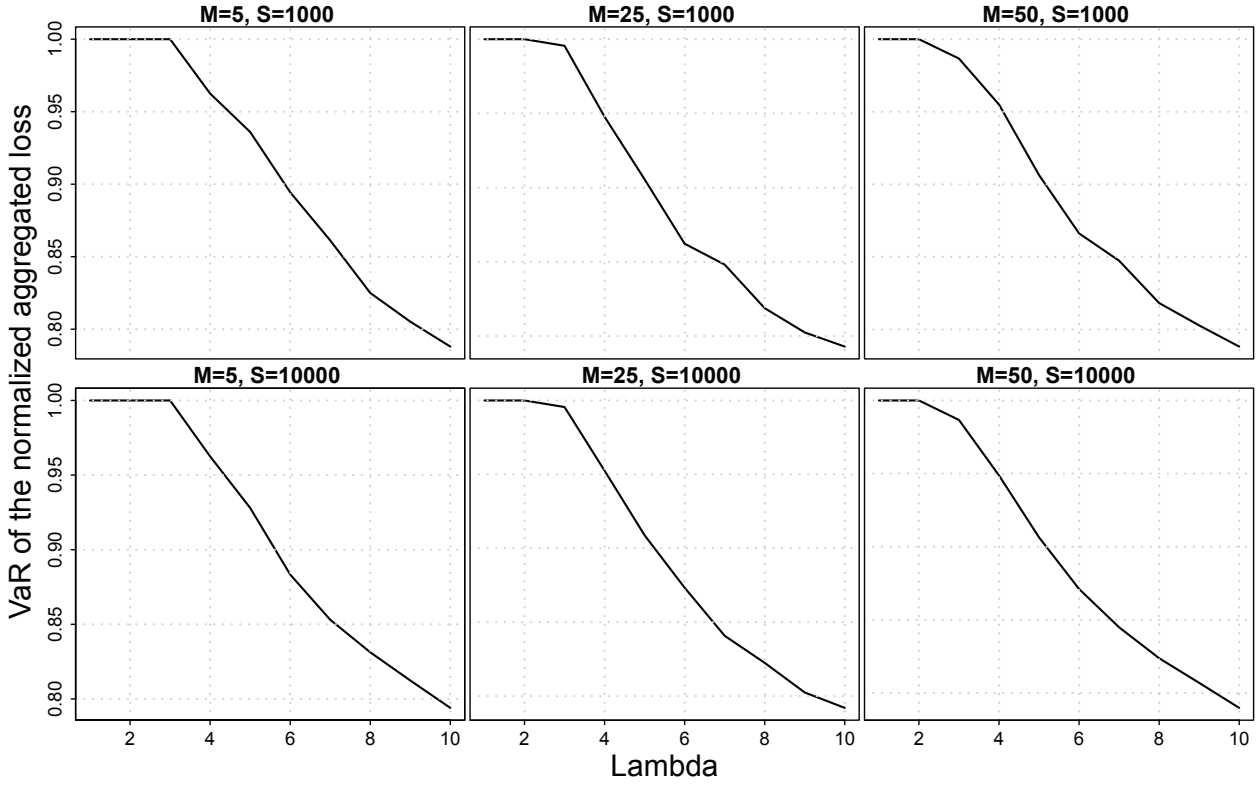


Figure 4.8: $R_{3,0.9}(\lambda A)$ with respect to λ in the case of the Smith process, for different values of M and S, where A is a square of side 1.

kind of spatial diversification is obtained in the cases of the Schlather and the geometric

Gaussian processes.

Though we do not have any explicit formula for $R_{3,1-\alpha}(\lambda A)$ for finite values of λ, by taking advantage of Theorem 4.2, we know the asymptotic behavior (when $\lambda \to \infty$) of $R_{3,1-\alpha}(\lambda A)$ for some max-stable models. This is presented in the following theorem.

Theorem 4.4. *In the cases of the Smith model, the Brown-Resnick model whose variogram satisfies* $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ *with* $a > 0$*, and the tube model,* $R_{3,1-\alpha}$ *satisfies the axiom of asymptotic spatial homogeneity of order -1, with*

$$K_1 = \frac{1 - \exp\left(-\frac{1}{u}\right)}{R_{3,1-\alpha}(A)} \quad \text{and} \quad K_2 = \frac{q_{1-\alpha}\sqrt{\int_{\mathbb{R}^2}\left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}}}{R_{3,1-\alpha}(A)},$$

where $q_{1-\alpha}$ *is the quantile at the level* $1-\alpha$ *of the standard Gaussian distribution.*

Regarding the other axioms, due to Proposition 4.2, we can state that for any strictly stationary max-stable process, $R_{3,1-\alpha}$ is invariant under translation. But since we don't have any formula for the VaR for non-asymptotic values of λ, we cannot say anything about spatial sub-additivity and spatial anti-monotonicity. In the case of the variance, Corollary 4.1 is used to prove the spatial sub-additivity and the spatial anti-monotonicity.

4.5 Example based on the power damage function

We consider in this section the normalized spatially aggregated loss obtained by combining (4.2), (4.5) and (4.7):

$$L_N(A) = \frac{1}{|A|}\int_A Z(\mathbf{x})^\beta \, d\mathbf{x}. \tag{4.19}$$

As in Section 4.4, the process $\{Z(\mathbf{x}\}_{\mathbf{x}\in\mathbb{R}^2}$ is simple max-stable. Here we mainly focus on the Smith process.

The damage function $Z(\mathbf{x})^\beta$ is particularly adapted in the case of wind hazard. The literature about damage functions related to wind is abundant. Such damages are generally taken as a power of the wind speed. For instance, in Klawa and Ulbrich (2003), Pinto et al. (2007) or Donat et al. (2011), the loss damage is proportional to the third power of the wind speed. Their motivation lies in the fact that the energy is related to the cube of the wind speed. Nevertheless, the relationship between damage and energy is not necessarily linear (see e.g. Prettenthaler et al. (2012)). Another approach followed by Dorland et al. (1999) and Prettenthaler et al. (2012) is to model damage as the exponential of the wind speed.

Before studying the variance and the VaR, we introduce a dependence measure for the damages related to wind. This tool will be useful in order to study R_2; it is also interesting in itself.

4.5.1 A new spatial dependence measure for damages

In the literature, there exist many dependence measures for max-stable processes, such as the extremal coefficient, the madogram (Matheron, 1987), the F-madogram (Cooley et al., 2006) and the λ-madogram (Naveau et al., 2009). Here we propose a spatial dependence measure for damages due to windstorms and not for the environmental variable itself.

We consider the quantity Corr $\left[Z_1^{\beta_1}, Z_2^{\beta_2}\right]$, where $Z_1 = Z(\mathbf{x}_1)$ and $Z_2 = Z(\mathbf{x}_2)$ for two sites $\mathbf{x}_1$ and $\mathbf{x}_2$, β_1 and β_2 are two positive coefficients and Corr is the correlation.

We denote by $\mathcal{F}(\nu, \sigma_f, m_f)$ the Fréchet distribution with decay, scale and location parameters ν, $\sigma_f > 0$ and m_f, respectively. By definition, a random variable Z follows $\mathcal{F}(\nu, s_f, m_f)$ if

$$\mathbb{P}(Z \leq z) = \begin{cases} 0 & \text{if } z \leq m_f, \\ \exp\left[-\left(\dfrac{\sigma_f}{z - m_f}\right)^{\nu}\right] & \text{if } z > m_f. \end{cases}$$

The following lemma gives the condition on β such that Z^β has a second order moment.

Lemma 4.2. *Let Z be a random variable following $\mathcal{F}(1, 1, 0)$; then Z^β has a second order moment if and only if $\beta < \dfrac{1}{2}$. Moreover, $\mathbb{E}(Z^\beta) = \Gamma(1 - \beta)$.*

In the case of the Smith process, we have the following result.

Theorem 4.5. *Let $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ be a Smith process with covariance matrix $\mathbf{\Sigma}$ and consider two sites $\mathbf{x}_1$ and $\mathbf{x}_2$. We denote $\mathbf{h} = (\mathbf{x}_2 - \mathbf{x}_1)'$ and $h = \sqrt{\mathbf{h}'\mathbf{\Sigma}^{-1}\mathbf{h}}$.*
For β_1 and $\beta_2 < \frac{1}{2}$, we have

$$\begin{aligned} &\mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] \\ &= \int_0^{+\infty} \theta^{\beta_2} \left[C_2(\theta)\ C_1(\theta)^{\beta_1+\beta_2-2}\ \Gamma(2 - \beta_1 - \beta_2) + C_3(\theta)\ C_1(\theta)^{\beta_1+\beta_2-1}\ \Gamma(1 - \beta_1 - \beta_2)\right]\ d\theta, \end{aligned} \tag{4.20}$$

where

$$C_1(\theta) = \left(\Phi\left(\frac{h}{2} + \frac{\log(\theta)}{h}\right) + \frac{\Phi\left(\frac{h}{2} - \frac{\log(\theta)}{h}\right)}{\theta}\right),$$

$$\begin{aligned} C_2(\theta) = &\left(\Phi\left(\frac{h}{2} + \frac{\log(\theta)}{h}\right) + \frac{\phi\left(\frac{h}{2} + \frac{\log(\theta)}{h}\right)}{h} - \frac{\phi\left(\frac{h}{2} - \frac{\log(\theta)}{h}\right)}{h\theta}\right) \\ &\times \left(\frac{\Phi\left(\frac{h}{2} - \frac{\log(\theta)}{h}\right)}{\theta^2} + \frac{\phi\left(\frac{h}{2} - \frac{\log(\theta)}{h}\right)}{h\theta^2} - \frac{\phi\left(\frac{h}{2} + \frac{\log(\theta)}{h}\right)}{h\theta}\right), \end{aligned}$$

$$C_3(\theta) = \left(\frac{(\frac{h}{2} - \frac{\log(\theta)}{h})\ \phi\left(\frac{h}{2} + \frac{\log(\theta)}{h}\right)}{h^2\theta} + \frac{(\frac{h}{2} + \log(\theta))\ \phi\left(\frac{h}{2} - \frac{\log(\theta)}{h}\right)}{h^2\theta^2}\right),$$

where Φ and ϕ denote the distribution function and the density of the standard Gaussian random variable, respectively.

Remark 4.8. *Due to the similarity of the bivariate density, the same kind of expression can be obtained in the cases of the geometric Gaussian process and some other specific Brown-Resnick processes. It is more difficult in the case of the Schlather process, due to its more complex bivariate density.*

Using Theorem 4.5, we obtain the result of interest regarding the correlation between damages.

Corollary 4.4. *Under the same conditions as in Theorem 4.5, we have*

$$
\begin{aligned}
&Corr\left(Z_1^{\beta_1}, Z_2^{\beta_2}\right) \\
&= \frac{\displaystyle\int_0^{+\infty} \theta^{\beta_2} \left[C_2(\theta)\ C_1(\theta)^{\beta_1+\beta_2-2}\ \Gamma(2-\beta_1-\beta_2) + C_3(\theta)\ C_1(\theta)^{\beta_1+\beta_2-1}\ \Gamma(1-\beta_1-\beta_2)\right]\ d\theta}{\sqrt{\Gamma\left(1-2\beta_1\right) - \left(\Gamma\left(1-\beta_1\right)\right)^2}\sqrt{\Gamma\left(1-2\beta_2\right) - \left(\Gamma\left(1-\beta_2\right)\right)^2}} \\
&\quad - \frac{\Gamma[1-\beta_1]\Gamma[1-\beta_1]}{\sqrt{\Gamma\left(1-2\beta_1\right) - \left(\Gamma\left(1-\beta_1\right)\right)^2}\sqrt{\Gamma\left(1-2\beta_2\right) - \left(\Gamma\left(1-\beta_2\right)\right)^2}}.
\end{aligned}
$$

By setting $\beta_1 = \beta_2$ the following corollary follows.

Corollary 4.5. *Under the same conditions as in Theorem 4.5, we have*

$$
\begin{aligned}
&Corr\left(Z_1^{\beta}, Z_2^{\beta}\right) \\
&= \frac{\int_0^{+\infty} \theta^{\beta} \left[C_2(\theta)\ C_1(\theta)^{2\beta-2}\ \Gamma(2-2\beta) + C_3(\theta)\ C_1(\theta)^{2\beta-1}\ \Gamma(1-2\beta)\right]\ d\theta - \left(\Gamma[1-\beta]\right)^2}{\Gamma\left(1-2\beta\right) - \left(\Gamma\left(1-\beta\right)\right)^2}.
\end{aligned}
$$

The integral appearing in Corollary 4.5 has no closed form and therefore a numerical approximation is required. We have tested both Riemann and Monte-Carlo methods and obtained similar results.

The evolution of $\text{Corr}\left(Z_1^{\beta}, Z_2^{\beta}\right)$ with respect to the distance h and the coefficient β is depicted in Figure 4.9. For β tending towards 0, we observe an abrupt increase of the correlation to 1. However, for very small values of β, there is some numerical instability in the computation of the integral, explaining the values slightly larger than 1. For β varying in $[0.01, 0.5)$, we observe that the correlation between damages is not very sensitive to the damage coefficient β. This may be relevant for practical situations. Secondly, we observe that the correlation rapidly decreases when the distance h between the two sites increases.

4.5.2 The risk measures

Using Lemma 4.2 as well as the linearity of the expectation, we immediately obtain that $\forall A \in \mathcal{A}, R_1(A) = \mathbb{E}\left[L_N(A)\right] = \Gamma(1-\beta)$. The conclusions are the same than in the case of the threshold damage function. We study the variance $R_2(A)$ and the VaR $R_{3,1-\alpha}(A)$ in further details.

The variance

We focus in this section on the quantity $R_2(A) = \text{var}\left[L_N(A)\right]$. As in Section 4.4, we first study in detail the function $\lambda \mapsto R_2(\lambda A)$ and then look at the different axioms introduced in Definition 4.4.

Some results about spatial homogeneity

The expression of $R_2(\lambda A)$ in the general case is given in the next theorem.

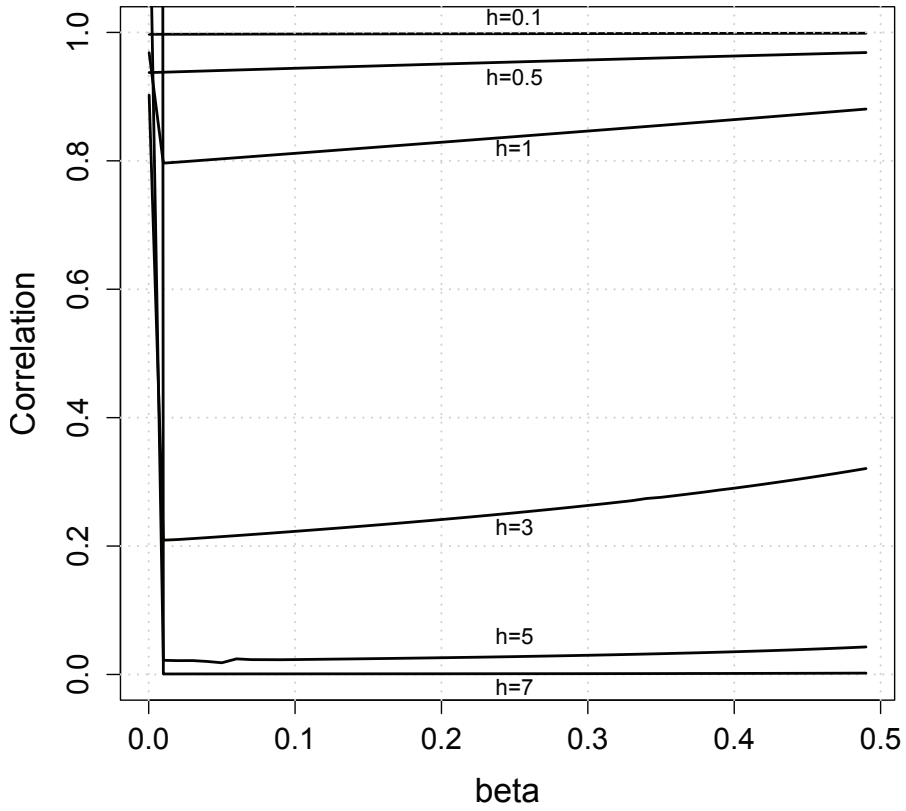


Figure 4.9: Correlation between the damages due to wind at two sites, in the case of the Smith model, with respect to β and for various values of the distance h between the two sites.

Theorem 4.6. *In the case of a simple max-stable process, $\forall A \in \mathcal{A}$, we have that*

$$R_2(\lambda A) = \frac{1}{\lambda^4 |A|^2} \int_{\lambda A} \int_{\lambda A} \left[\mathbb{E}\left[Z(\mathbf{x})^\beta \; Z(\mathbf{y})^\beta \right] - \left[\Gamma(1-\beta) \right]^2 \right] d\mathbf{x} \; d\mathbf{y}. \tag{4.21}$$

Using Theorem 4.6 and taking advantage of the expression of $\mathbb{E}\left[Z(\mathbf{x})^\beta Z(\mathbf{y})^\beta \right]$ obtained in Theorem 4.5, we can derive the behavior of $R_2(\lambda A)$ in the case of the isotropic Smith process, when the region A is either a disk or a square. The result is given in the next corollary.

Corollary 4.6. *Assume that Z is an isotropic Smith process with standard Fréchet margins and that $A \in \mathcal{A}$. Then:*

1. ***If* A *is a disk*** *with radius R, we have*

$$R_2(\lambda A) = -\left[\Gamma(1-\beta) \right]^2 + \int_{h=0}^{2R} f_d(h, R) \; g(\lambda h) \; dh,$$

where f_d has been defined in Corollary 4.1,

2. ***If* $\mathbf{A}$ *is a square*** *with side R, we have*

$$R_2(\lambda A) = -\left[\Gamma(1-\beta)\right]^2 + \int_{h=0}^{R\sqrt{2}} f_s(h,R)\ g(\lambda h)\ dh,$$

where f_s has been defined in Corollary 4.1, where

$$\begin{aligned} g(h) &= \mathbb{E}\left[Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta\right] \\ &= \int_0^{+\infty} \theta^\beta \left[C_2(\theta)\ C_1(\theta)^{2\beta-2}\ \Gamma(2-2\beta) + C_3(\theta)\ C_1(\theta)^{2\beta-1}\ \Gamma(1-2\beta)\right]\ d\theta \end{aligned}$$

(by setting $\beta_1 = \beta_2$ in Theorem 4.5) and $h = \sqrt{\mathbf{h}'\boldsymbol{\Sigma}^{-1}\mathbf{h}}$, with $\mathbf{h} = \mathbf{x} - \mathbf{y}$.
The dependence in h of the function g appears via the dependence in h of the functions $C_1(.)$, $C_2(.)$ and $C_3(.)$ defined in Theorem 4.5.

3. ***In both cases****, $R_2(\lambda A)$ converges as $\lambda \to \infty$ to the limiting risk measure given by*

$$-\left[\Gamma(1-\beta)\right]^2 + \lim_{\lambda\to\infty} g(\lambda h) = 0.$$

Thus, the spatial diversification is total.

We can show that the function $h \mapsto g(\lambda h)$ is decreasing with respect to λ, meaning that there is spatial diversification. The different integrals appearing in Corollary 4.6 have no closed form. Therefore, we use a Riemann approximation. The coefficient β is set randomly to 0.49 but its value has no strong influence on the shape of the curves. The identity matrix is taken as covariance matrix of the Smith process. Figure 4.10 displays the same type of rapid decrease of $R_2(\lambda A)$ with respect to λ as in the case of the threshold damage function. The decrease is a little slower in the case of the square.

Central limit theorem and axioms

Using mixing conditions obtained by Dombry and Eyi-Minko (2012) and general results about central limit theorems for strongly mixing random fields (see e.g. Gorodetskii (1987)), we can prove the following theorem.

Theorem 4.7. *In the case of the Smith model, we have*

$$\lambda\left[L_N(\lambda A) - \Gamma(1-\beta)\right] \xrightarrow{d} \mathcal{N}(0,\sigma^2),\ \textit{for}\ \lambda\to\infty,$$

where

$$\sigma^2 = \int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta] - \left[\Gamma(1-\beta)\right]^2\right) d\mathbf{x},$$

where $\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta]$ is given in Theorem 4.5.

Finally, since the function $h \mapsto g(\lambda h)$ (appearing in Corollary 4.6) is decreasing with respect to λ, we obtain the following proposition, in exactly the same way as in the proof of Theorem 4.3.

Theorem 4.8. *In the case of the Smith process, the spatial risk measure R_2 satisfies the following axioms:*

1. *Spatial invariance under translation;*

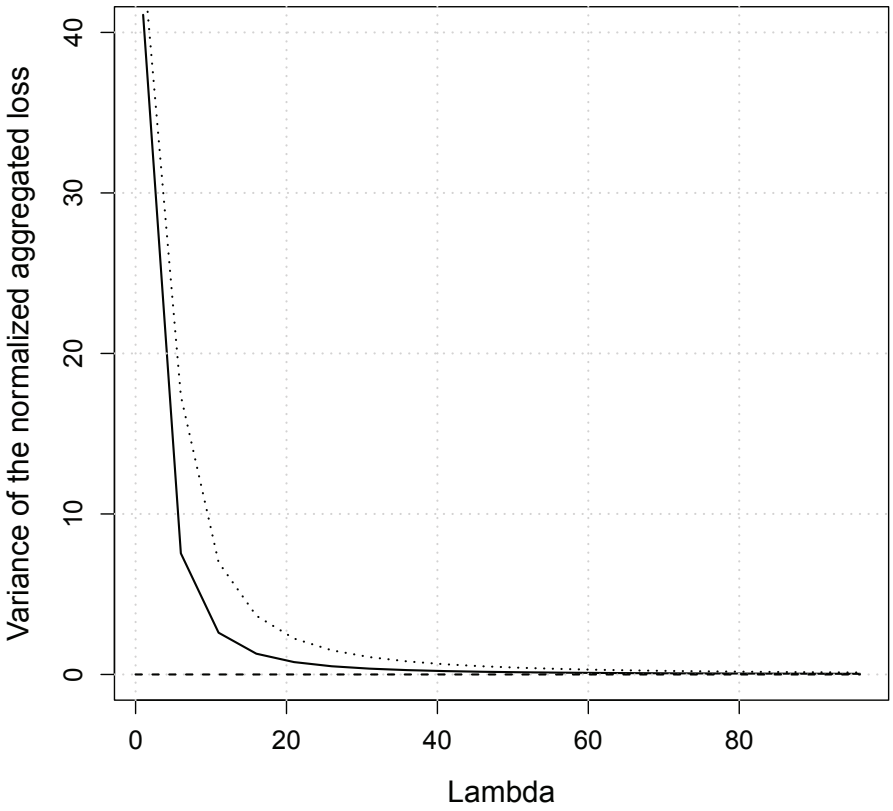


Figure 4.10: $R_2(\lambda A)$ with respect to λ in the case of the Smith process, for $\beta = 0.49$. The solid line corresponds to a disk with radius $R = 1$ whereas the dotted line corresponds to a square with side $R = 1$.

2. *Spatial sub-additivity when the two regions are both a disk or a square;*

3. *Asymptotic spatial homogeneity of order* -2*, with*

$$K_1 = 0 \qquad \textit{and} \qquad K_2 = \frac{\displaystyle\int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta] - [\Gamma(1-\beta)]^2\right) d\mathbf{x}}{R_2(A)},$$

   where $\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta]$ *is given in Theorem 4.5;*

4. *Spatial anti-monotonicity when the two regions are both a disk or a square.*

The Value at Risk

As in the case of the threshold damage function, it seems very difficult to find a closed formula for the VaR $R_{3,1-\alpha}(\lambda A)$, even up to an integral. The same type of approximation as in Section 4.4.2 can be used, leading to the same kind of graphs. This approach is numerically rather time consuming.

However, using Theorem 4.7, we can show the following result, exactly in the same way as in the proof of Theorem 4.4.

Theorem 4.9. *In the case of the Smith model, $R_{3,1-\alpha}$ satisfies the axiom of asymptotic spatial homogeneity of order -1 with*

$$K_1 = \frac{\Gamma(1-\beta)}{R_{3,1-\alpha}(A)} \qquad \text{and} \qquad K_2 = \frac{q_{1-\alpha}\sqrt{\displaystyle\int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta] - [\Gamma(1-\beta)]^2\right) d\mathbf{x}}}{R_{3,1-\alpha}(A)},$$

where $\mathbb{E}[Z(\mathbf{0})^\beta Z(\mathbf{x})^\beta]$ is given in Theorem 4.5.

Regarding the other axioms, we can make the same comments as in the case of Theorem 4.4.

4.6 Conclusion

This paper introduces a new notion of spatial risk measure, based on the normalized spatially aggregated loss, and proposes a set of axioms adapted to the spatial context. Contrary to the classical literature, our axiomatic approach aims at quantifying the sensitivity of the risk measurement with respect to space. The idea is to propose a new framework for risk measures as well as relevant tools for public authorities and the (re)insurance industry. The concept of spatial risk measure can be an interesting tool for spatial risk assessment and management. Characterizing all risk measures satisfying the proposed axioms or even proposing other adapted axioms could be relevant and useful.

In order to develop concrete examples of spatial risk measures, we propose a model that maps the process of the environmental variable generating loss into an economic damage, via a damage function. In this paper, we are mainly interested in risks related to extreme environmental events. Therefore, we model the environmental process using max-stability. Two damage functions, respectively adapted to heatwaves and windstorms, are considered. Theoretical properties of our spatial risk measures are derived for classical max-stable models as well as for a simple max-stable model introduced in this article, the tube model. We show that in the case of the variance, these risk measures satisfy the axioms proposed, for some underlying models. Furthermore, an interpretation in terms of insurance is provided.

On a practical point of view, the construction of the spatial risk measures introduced here involves the following steps:

1. Fit several max-stable models (Smith, Schlather, Brown-Resnick, tube) on historical data. Fitting methods can be based for instance on the non parametric simulated maximum likelihood estimator or on a composite likelihood (see Padoan et al. (2010));

2. Choose the best max-stable model via a model selection based e.g. on the Akaike Information Criterion (Akaike, 1974) or the likelihood ratio statistic (Davison, 2003);

3. Choose the model converting the environmental hazard into economic losses;

4. Compute the risk measure as explained.

Note that we have considered processes with standard Fréchet margins, which constitutes a classical assumption in the literature. However, considering more realistic margins would be an improvement. If Z is standard Fréchet, then $Y = \mu + \sigma \left(\frac{Z^\xi - 1}{\xi} \right)$ follows a $GEV(\mu, \sigma, \xi)$. For the sake of simplicity, we can consider that $\mu = 0$ and $\sigma = 1$. Ongoing work consists in the study of $\frac{1}{|A|} \int_A \left(\frac{Z(\mathbf{x})^\xi - 1}{\xi} \right)^\beta d\mathbf{x} = \frac{1}{|A|\ \xi^\beta} \int_A (Z(\mathbf{x})^\xi - 1)^\beta \ d\mathbf{x}$, where the process $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^2}$ has standard Fréchet margins. In the case of wind hazard, practical studies generally show that the marginal shape parameter ξ is slightly negative or null (see e.g. Perrin et al. (2006)), although the results depend on the type of measure (average speed, gusts, ...). A possibility consists in considering the limiting case $\xi \to 0$, $\left(\frac{Z^\xi - 1}{\xi} \right) \underset{\xi \to 0}{\sim} \log Z$. An extension to non stationary max-stable processes (involving varying margin parameters) could also be of interest.

Examples of spatial risk measures developed in our paper include only one environmental hazard (e.g. a heatwave or a windstorm). However, it would be possible to extend this approach to more sources of hazard. Let us denote by $Z_1, \ldots, Z_k$ the spatial processes of the maxima of k environmental variables. In this case, the loss on region A would be given by

$$L(A, P) = \int_A C_P(\mathbf{x})\ d\mathbf{x} = \int_A D[Z_1(\mathbf{x}), \ldots, Z_k(\mathbf{x})]\ d\mathbf{x},$$

where the damage function D is k-variate.

4.7 Appendix: Proofs

For Proposition 4.1

Proof. We first show that spatial sub-additivity implies spatial anti-monotonicity. Let $A_1, A_2 \in \mathcal{A}$, with $A_1 \subset A_2$. We have

$$R_\Pi(A_2, P) = R_\Pi(A_1 \cup A_2, P) \leq \min[R_\Pi(A_1, P), R_\Pi(A_2, P)] \leq R_\Pi(A_1, P).$$

We now prove that spatial anti-monotonicity implies spatial sub-additivity. Let $A_1, A_2 \subset \mathcal{A}$. We have $A_1 \subset A_1 \cup A_2$ and $A_2 \subset A_1 \cup A_2$, giving that

$$R_\Pi(A_1 \cup A_2, P) \leq R_\Pi(A_1, P) \text{ and } R_\Pi(A_1 \cup A_2, P) \leq R_\Pi(A_2, P).$$

Therefore, $R_\Pi(A_1 \cup A_2, P) \leq \min[R_\Pi(A_1, P), R_\Pi(A_2, P)]$, showing the result. □

For Proposition 4.2

Proof. Using the fact that $|A + \mathbf{v}| = |A|$ and the change of variable $\mathbf{y} = \mathbf{x} - \mathbf{v}$, we have

$$R_\Pi(A + \mathbf{v}, P) = \Pi\left[\frac{1}{|A + \mathbf{v}|}\int_{A+\mathbf{v}} C_P(\mathbf{x})\, d\mathbf{x}\right] = \Pi\left[\frac{1}{|A|}\int_A C_P(\mathbf{y} + \mathbf{v})\, d\mathbf{y}\right]. \quad (4.22)$$

Due to the strict stationarity of process C_P, we have

$$\forall \mathbf{x}, \mathbf{y} \text{ and } \mathbf{v} \in \mathbb{R}^2, C_P(\mathbf{x}) \stackrel{d}{=} C_P(\mathbf{y} + \mathbf{v}), \text{ yielding}$$

$$\Pi\left[\frac{1}{|A|}\int_A C_P(\mathbf{y} + \mathbf{v})\, d\mathbf{y}\right] = \Pi\left[\frac{1}{|A|}\int_A C_P(\mathbf{x})\, d\mathbf{x}\right] = R_\Pi(A, P). \quad (4.23)$$

The combination of (4.22) and (4.23) provides the result. □

For Lemma 4.1

Proof. Mixing and ergodicity are properties of the σ-algebra generated by the underlying process. Since the σ-algebra associated to a function of the process Z is smaller than that associated to Z, one can say that process H is mixing (respectively ergodic) whatever the function D. □

For Theorem 4.1

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[(L(A))^2\right] = \mathbb{E}\left[\left(\int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\, d\mathbf{x}\right)^2\right] &= \mathbb{E}\left[\int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\, d\mathbf{x} \int_A \mathbf{I}_{\{Z(\mathbf{y})>u\}}\, d\mathbf{y}\right] \\
&= \mathbb{E}\left[\int_A \int_A \mathbf{I}_{\{Z(\mathbf{x})>u\}}\mathbf{I}_{\{Z(\mathbf{y})>u\}}\, d\mathbf{x}\, d\mathbf{y}\right] \\
&= \mathbb{E}\left[\int_A \int_A \mathbf{I}_{\{Z(\mathbf{x})>u, Z(\mathbf{y})>u\}}\, d\mathbf{x}\, d\mathbf{y}\right] \\
&= \int_A \int_A \mathbb{P}(Z(\mathbf{x}) > u, Z(\mathbf{y}) > u)\, d\mathbf{x}\, d\mathbf{y}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbb{P}(Z(\mathbf{x}) > u, Z(\mathbf{y}) > u) \\
&= 1 - \mathbb{P}(Z(\mathbf{x}) \leq u, Z(\mathbf{y}) \leq u) - [\mathbb{P}(Z(\mathbf{x}) \leq u) - \mathbb{P}(Z(\mathbf{x}) \leq u, Z(\mathbf{y}) \leq u)] - [\mathbb{P}(Z(\mathbf{y}) \leq u) \\
&- \mathbb{P}(Z(\mathbf{x}) \leq u, Z(\mathbf{y}) \leq u)] \\
&= 1 + \mathbb{P}(Z(\mathbf{x}) \leq u, Z(\mathbf{x}) \leq u) - \mathbb{P}(Z(\mathbf{x}) \leq u) - \mathbb{P}(Z(\mathbf{y}) \leq u) \\
&= 1 + \mathbb{P}(Z(\mathbf{x}) \leq u, Z(\mathbf{y}) \leq u) - 2\exp\left(-\frac{1}{u}\right),
\end{aligned}$$

yielding

$$\mathbb{E}\left[(L(A))^2\right] = |A|^2\left(1 - 2\exp\left(-\frac{1}{u}\right)\right) + \int_A\int_A \exp\left(-\frac{\Theta(\mathbf{x},\mathbf{y})}{u}\right)\, d\mathbf{x}\, d\mathbf{y}.$$

Therefore

$$\begin{aligned}
&R_2(A) \\
&= \text{var}\left[\frac{L(A)}{|A|}\right] \\
&= \frac{1}{|A|^2}\ \left(\mathbb{E}\left[(L(A))^2\right] - [\mathbb{E}(L(A))]^2\right) \\
&= \frac{1}{|A|^2}\left[|A|^2\left(1 - 2\exp\left(-\frac{1}{u}\right)\right) + \int_A\int_A \exp\left(-\frac{\Theta(\mathbf{x},\mathbf{y})}{u}\right)\, d\mathbf{x}\, d\mathbf{y} - |A|^2\left(1 - \exp\left(-\frac{1}{u}\right)\right)\right] \\
&= -\exp\left(-\frac{2}{u}\right) + \frac{1}{|A|^2}\int_A\int_A \exp\left(-\frac{\Theta(\mathbf{x},\mathbf{y})}{u}\right)\, d\mathbf{x}\, d\mathbf{y} \\
&= \frac{1}{|A|^2}\int_A\int_A \left[\exp\left(-\frac{\Theta(\mathbf{x},\mathbf{y})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right]\, d\mathbf{x}\, d\mathbf{y}.
\end{aligned}$$

The result is obtained by replacing A by λA. □

For Corollary 4.1

Proof. Let us show the result in the case of A being a disk with radius R. The density between 2 points uniformly distributed on A is denoted $f_d(h, R)$. Let us denote by N_p the number of pairs of infinitesimal squares $(d\mathbf{x}, d\mathbf{y})$, such that $\|\mathbf{x} - \mathbf{y}\| \in [h, h + dh]$ and by N_t the total number of pairs. The definition directly yields

$$f_d(h, R)\ dh = \mathbb{P}\left(\|\mathbf{x} - \mathbf{y}\| \in [h, h + dh]\right) = \frac{N_p}{N_t}.$$

Moreover, we have $N_t = \dfrac{|A|^2}{|d\mathbf{x}||d\mathbf{y}|}$, where $|d\mathbf{x}|$ and $|d\mathbf{y}|$ are the areas of infinitesimal squares $d\mathbf{x}$ and $d\mathbf{y}$. That gives $N_p = f_d(h, R)\ dh\ \dfrac{|A|^2}{|d\mathbf{x}||d\mathbf{y}|}$. We denote by $Area(h)$ the total area corresponding to the infinitesimal squares $(d\mathbf{x}, d\mathbf{y})$, such that $\|\mathbf{x} - \mathbf{y}\| \in [h, h + dh]$. We have

$$\text{Area}(h) = N_p\ |d\mathbf{x}||d\mathbf{y}| = |A|^2 f_d(h, R)dh.$$

For any function g depending only on the Euclidean distance $h = \|\mathbf{x} - \mathbf{y}\|$, we have

$$\int_A \int_A g(\mathbf{x}, \mathbf{y}) \; d\mathbf{x} \; d\mathbf{y} = \int_0^{2R} Area(h) \; g(h).$$

Therefore

$$\int_A \int_A g(\mathbf{x}, \mathbf{y}) \; d\mathbf{x} \; d\mathbf{y} = \int_0^{2R} |A^2| \; f_d(h, R) \; g(h) \; dh. \tag{4.24}$$

Applying this result to $g(h) = \exp\left(-\dfrac{\Theta(h)}{u}\right)$ and using (4.15), we obtain

$$R_2(A) = -\exp\left(-\frac{2}{u}\right) + \int_0^{2R} f_d(h, R) \; \exp\left(-\frac{\Theta(h)}{u}\right) \; dh.$$

Moreover, Moltchanov (2012) shows that

$$f_d(h, R) = \frac{2h}{R^2}\left(\frac{2}{\pi}\arccos\left(\frac{h}{2R}\right) - \frac{h}{\pi R}\sqrt{1 - \frac{h^2}{4R^2}}\right), \text{ for } 0 \leq h \leq 2R, \tag{4.25}$$

yielding

$$f_d(\lambda h, \lambda R) = \frac{2h}{R^2} \; \frac{1}{\lambda} \; \left(\frac{2}{\pi}\arccos\left(\frac{h}{2R}\right) - \frac{h}{\pi R}\sqrt{1 - \frac{h^2}{4R^2}}\right) = \frac{1}{\lambda} \; f_d(h, R).$$

Therefore, by the change of variable $h_d = \dfrac{h}{\lambda}$, we obtain

$$\begin{aligned}
&\int_0^{2R} f_d(h, R) \; \exp\left(-\frac{\Theta(h)}{u}\right) \; dh = \int_0^{2\lambda R} f(h, \lambda R) \; g(h) \; dh \\
&= \int_0^{2R} f(\lambda h_d, \lambda R) \; g(h_d) \; \lambda \; dh_d = \int_0^{2R} f(h_d, R) \; g(\lambda h_d) \; dh_d.
\end{aligned}$$

Finally

$$R_2(\lambda A) = -\exp\left(-\frac{2}{u}\right) + \int_0^{2R} f_d(h, R) \; \exp\left(-\frac{\Theta(\lambda h)}{u}\right) \; dh.$$

and the result is established.

The same reasoning yields the result in the case of a square. Let us now compute the density $f_s(h, R)$ in the case of a square with side R. Moltchanov (2012) shows that the distribution function of the distance between 2 points uniformly drawn on a square with side R is written

$$F_s(h, R) = \begin{cases} \dfrac{\pi h^2}{R^2} - \dfrac{8h^3}{3R^3} + \dfrac{h^4}{2B^4} & \text{if } h \in [0, R], \\ \dfrac{1}{3} - 2b - \dfrac{b^2}{2} + \dfrac{2}{3}\sqrt{(b-1)^3} + 2\sqrt{b-1} + 2b\sqrt{b-1} + 2b\arcsin\left(\dfrac{2-b}{b}\right) & \text{if } h \in [R, R\sqrt{2}], \end{cases}$$

where $b = \dfrac{h^2}{R^2}$.

Therefore, for $h \in [0, R]$, the density is directly

$$f_s(h, R) = \frac{2\pi h}{R^2} - \frac{8h^2}{R^3} + \frac{2h^3}{R^4}.$$

For $h \in [R, R\sqrt{2}]$, we obtain

$$F'(b) = -2 - b + \sqrt{b-1} + \frac{1}{\sqrt{b-1}} + 2\sqrt{b-1} + \frac{b}{\sqrt{b-1}} + 2\arcsin\left(\frac{2-b}{b}\right) - \frac{4}{b\sqrt{1-\frac{(2-b)^2}{b^2}}}$$

$$= -2 - b + 3\sqrt{b-1} + \frac{b+1}{\sqrt{b-1}} + 2\arcsin\left(\frac{2-b}{b}\right) - \frac{4}{b\sqrt{1-\frac{(2-b)^2}{b^2}}}, \text{ giving}$$

$$= \left(-2 - b + 3\sqrt{b-1} + \frac{b+1}{\sqrt{b-1}} + 2\arcsin\left(\frac{2-b}{b}\right) - \frac{4}{b\sqrt{1-\frac{(2-b)^2}{b^2}}}\right)\frac{2h}{R^2}.$$

□

For Proposition 4.3

Proof. Since the $\{(\xi_i, \mathbf{c_i})\}_{i\geq 1}$ are the points of Poisson point process on $(0, +\infty) \times \mathbb{R}^2$ with intensity measure $d\Lambda(\xi, \mathbf{c}) = \xi^{-2} d\xi d\mathbf{c}$, we have

$$\begin{aligned}
&-\log[\mathbb{P}(Z(\mathbf{x}_1) \leq z_1, Z(\mathbf{x}_2) \leq z_2)] \\
&= \int_{\mathbb{R}^2} \int_{\min\left(\frac{z_1}{f_0(\mathbf{x}-\mathbf{x}_1)}, \frac{z_2}{f_0(\mathbf{x}-\mathbf{x}_2)}\right)}^{+\infty} \xi^{-2} d\xi \; d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \left[-\frac{1}{\xi}\right]_{\min\left(\frac{z_1}{f_0(\mathbf{x}-\mathbf{x}_1)}, \frac{z_2}{f_0(\mathbf{x}-\mathbf{x}_2)}\right)}^{+\infty} d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \max\left(\frac{f_0(\mathbf{x}-\mathbf{x}_1)}{z_1}, \frac{f_0(\mathbf{x}-\mathbf{x}_2)}{z_2}\right) d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \frac{f_0(\mathbf{x}-\mathbf{x}_1)}{z_1} \mathbf{I}_{\left\{\frac{f_0(\mathbf{x}-\mathbf{x}_1)}{z_1} > \frac{f_0(\mathbf{x}-\mathbf{x}_2)}{z_2}\right\}} d\mathbf{x} \\
&\quad + \int_{\mathbb{R}^2} \frac{f_0(\mathbf{x}-\mathbf{x}_2)}{z_2} \mathbf{I}_{\left\{\frac{f_0(\mathbf{x}-\mathbf{x}_2)}{z_2} \geq \frac{f_0(\mathbf{x}-\mathbf{x}_1)}{z_1}\right\}} d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \frac{f_0(\mathbf{x})}{z_1} \mathbf{I}_{\left\{\frac{f_0(\mathbf{x})}{z_1} > \frac{f_0(\mathbf{x}+\mathbf{x}_1-\mathbf{x}_2)}{z_2}\right\}} d\mathbf{x} + \int_{\mathbb{R}^2} \frac{f_0(\mathbf{x})}{z_2} \mathbf{I}_{\left\{\frac{f_0(\mathbf{x})}{z_2} \geq \frac{f_0(\mathbf{x}+\mathbf{x}_2-\mathbf{x}_1)}{z_1}\right\}} d\mathbf{x} \\
&= \frac{1}{z_1}\mathbb{P}\left(\frac{f_0(\mathbf{X})}{z_1} > \frac{f_0(\mathbf{X}+\mathbf{x}_1-\mathbf{x}_2)}{z_2}\right) + \frac{1}{z_2}\mathbb{P}\left(\frac{f_0(\mathbf{X})}{z_2} \geq \frac{f_0(\mathbf{X}+\mathbf{x}_2-\mathbf{x}_1)}{z_1}\right), \qquad (4.26)
\end{aligned}$$

where $\mathbf{X}$ is a random vector having density f_0.

Let us compute the first probability and denote by E_1 the event $\left\{\frac{f_0(\mathbf{X})}{z_1} > \frac{f_0(\mathbf{X}+\mathbf{x}_1-\mathbf{x}_2)}{z_2}\right\}$. We have

$$\begin{aligned}
E_1 &= \left\{z_2 \, \mathbf{I}_{\{\|\mathbf{X}\| \leq R_b\}} > z_1 \, \mathbf{I}_{\{\|\mathbf{X}-\mathbf{x}_2+\mathbf{x}_1\| \leq R_b\}}\right\} \\
&= \left\{\|\mathbf{X}\| \leq R_b \text{ and } z_2 > z_1 \text{ if } \|\mathbf{X}-\mathbf{x}_2+\mathbf{x}_1\| \leq R_b\right\}.
\end{aligned}$$

Thus, if $z_2 > z_1$, $E_1 = \{\|\mathbf{X}\| \leq R_b\}$, giving $\mathbb{P}(E_1) = \mathbb{P}(\|\mathbf{X}\| \leq R_b) = 1$. Indeed $\mathbf{X}$ has density f_0 and then $\|\mathbf{X}\| \leq R_b$ almost everywhere.

If $z_1 \geq z_2$,

$$\begin{aligned} E_1 &= \{\|\mathbf{X}\| \leq R_b \text{ and } \|\mathbf{X} - \mathbf{x}_2 + \mathbf{x}_1\| > R_b\} \\ &= \{\|\mathbf{X} - \mathbf{x}_2 + \mathbf{x}_1\| > R_b\}, \end{aligned}$$

since $\|\mathbf{X}\| \leq R_b$ is necessarily satisfied a.e.. Therefore,

$$\begin{aligned} \mathbb{P}(E_1) = \int_{\mathbb{R}^2} \mathbf{I}_{\{\|\mathbf{x}-\mathbf{x}_2+\mathbf{x}_1\|>R_b\}}\, \mathbf{I}_{\{|\mathbf{x}\|\leq R_b\}}\, d\mathbf{x} &= h_b \int_{\mathbb{R}^2} \mathbf{I}_{\{\|\mathbf{x}\|\leq R_b \,\cap\, \|\mathbf{x}-(\mathbf{x}_2-\mathbf{x}_1)\|>R_b\}}\, d\mathbf{x} \\ &= h_b\, (\pi R_b^2 - A_{int}(h)), \end{aligned}$$

where $A_{int}(h)$ is the area of the intersection between the base of the tube of center $\mathbf{0}$ and the one of the tube of center $(\mathbf{x}_2 - \mathbf{x}_1)$ and $h = \|\mathbf{x}_2 - \mathbf{x}_1\|$. Note that the area of the intersection between the base of the tube of center $\mathbf{0}$ and the one of the tube of center $(\mathbf{x}_1 - \mathbf{x}_2)$ is also equal to $A_{int}(h)$.

Let us compute the second probability and denote by E_2 the event $\left\{\dfrac{f_0(\mathbf{X})}{z_2} \geq \dfrac{f_0(\mathbf{X}+\mathbf{x}_2-\mathbf{x}_1)}{z_1}\right\}$. We have that

$$\begin{aligned} E_2 &= \left\{z_1\, \mathbf{I}_{\{\|\mathbf{X}\|\leq R_b\}} \geq z_2\, \mathbf{I}_{\{\|\mathbf{X}-\mathbf{x}_1+\mathbf{x}_2\|\leq R_b\}}\right\} \\ &= \{\|\mathbf{X}\| \leq R_b \text{ and } z_1 \geq z_2 \text{ if } \|\mathbf{X} - \mathbf{x}_1 + \mathbf{x}_2\| \leq R_b\} \end{aligned}$$

Thus, if $z_2 > z_1$, $E_2 = \{\|\mathbf{X} - \mathbf{x}_1 + \mathbf{x}_2\| > R_b\}$, giving $\mathbb{P}(E_2) = \mathbb{P}\left(\|\mathbf{X} - \mathbf{x}_1 + \mathbf{x}_2\| > R_b\right) = h_b(\pi R_b^2 - A_{int}(h))$.
If $z_1 \geq z_2$, $E_2 = \{\|\mathbf{X}\| \leq R_b\}$ if $\|\mathbf{X} - \mathbf{x}_1 + \mathbf{x}_2\| \leq R_b$ and is always satisfied otherwise, yielding $\mathbb{P}(E_2) = 1$ since $\{\|\mathbf{X}\| \leq R_b\}$ is necessarily satisfied.

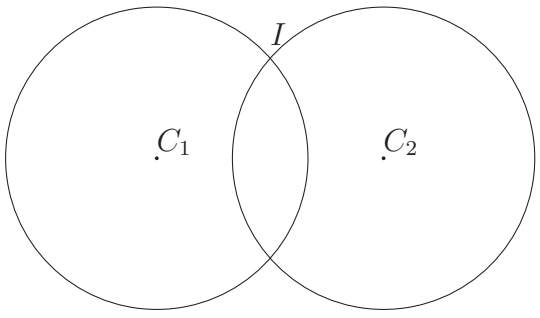
By definition of the extremal coefficient,

$$\Theta(\mathbf{x}_1 - \mathbf{x}_2) = -\log[\mathbb{P}(Z(\mathbf{x}_1) \leq u, Z(\mathbf{x}_2) \leq u]\; u.$$

Therefore, using (4.26), we obtain

$$\Theta(\mathbf{x}_1 - \mathbf{x}_2) = u\left(\frac{h_b}{u}\left(\pi R_b^2 - A_{int}(h)\right) + \frac{1}{u}\right) = h_b(\pi R_b^2 - A_{int}(h)) + 1 = 2 - h_b\; A_{int}(h). \tag{4.27}$$

Since we consider the L_2 norm, the bases of the tubes are circular. Let us then compute the intersection between two discs, respectively with radius R_b and centers C_1 and C_2 at a distance h. This intersection is not empty if and only if $h \leq 2R_b$. Let us consider this case, represented in the following picture:



By using Héron's formula, the area of the triangle IC_1C_2, denoted by A_T is given by

$$A_T = \sqrt{p(p-R_b)(p-R_b)(p-h)} \quad \text{where } p = \frac{1}{2}(2R_b + h). \tag{4.28}$$

Furthermore, by denoting by H the height of the triangle IC_1C_2, we have $A_T = \dfrac{hH}{2}$, giving

$$H = \frac{2A_T}{h}. \tag{4.29}$$

Denote α and β the angles $I\hat{C_1}C_2$ and $I\hat{C_2}C_1$, respectively. We have $\sin\alpha = \sin\beta = \dfrac{H}{R_b}$, yielding, using (4.29),

$$\alpha = \beta = \arcsin\left(\frac{H}{R_b}\right) = \arcsin\left(\frac{2A_T}{hR_b}\right). \tag{4.30}$$

Denote by S the area of angular sectors delimited respectively by the angles α and β. We have

$$S = \frac{\alpha R_b^2}{2} \tag{4.31}$$

Combining (4.28), (4.30) and (4.31), we obtain

$$A_{int}(h) = 2(2S - A_T) = 2\left[R_b^2 \arcsin\left(\frac{2\sqrt{p(p-R_b)^2(p-h)}}{hR_b}\right) - \sqrt{p(p-R_b)^2(p-h)}\right].$$

Since

$$\sqrt{p(p-R_b)^2(p-h)} = \frac{1}{4}\sqrt{(2R_b+h)h^2(2R_b-h)} = \frac{h}{4}\sqrt{4R_b^2-h^2},$$

we finally obtain

$$A_{int}(h) = \begin{cases} 2\left(R_b^2 \arcsin\left(\frac{\sqrt{4R_b^2-h^2}}{2R_b}\right) - \frac{h}{2}\sqrt{4R_b^2-h^2}\right) & \text{if } h \le 2R_b, \\ 0 & \text{if } h > 2R_b. \end{cases} \tag{4.32}$$

The combination of (4.27) and (4.32) yields the result. □

For Proposition 4.4

Proof. We consider the case of A being a disk but the proof is exactly the same in the case of a square. We have

$$\begin{aligned} &R_2(\lambda A) \\ &= -\exp\left(-\frac{2}{u}\right) + \int_0^{2R} f_d(h,R)\exp\left(-\frac{\Theta(\lambda h)}{u}\right)\, dh \\ &= -\exp\left(-\frac{2}{u}\right) + \int_0^{2R_b} f_d(h,R)\exp\left(-\frac{\Theta(\lambda h)}{u}\right)\, dh + \int_{2R_b}^{2R} f_d(h,R)\exp\left(-\frac{\Theta(\lambda h)}{u}\right)\, dh. \end{aligned}$$

Moreover, for $\lambda > 1$ and $h > 2R_b$, we have $\Theta(\lambda h) = \Theta(h) = 2$. Therefore,

$$\begin{aligned}
&R_2(\lambda A)\\
&= \int_0^{2R_b} f_d(h,R) \exp\left(-\frac{\Theta(\lambda h)}{u}\right)\, dh - \exp\left(-\frac{2}{u}\right) + \int_{h=0}^{2R} f_d(h,R) \exp\left(-\frac{\Theta(h)}{u}\right)\, dh\\
&\quad - \int_0^{2R_b} f_d(h,R) \exp\left(-\frac{\Theta(h)}{u}\right)\, dh\\
&= R_2(A) + \frac{1}{|A|^2} \int_0^{R_b} f_d(h,R) \left[\exp\left(-\frac{\Theta(\lambda h)}{u}\right) - \exp\left(-\frac{\Theta(h)}{u}\right)\right] dh,
\end{aligned}$$

When R_b tends to 0, the second term vanishes and we obtain $R_2(\lambda A) = R_2(A)$. Moreover, if $R_b = 0$, $\forall h, \forall \lambda > 1, \Theta(\lambda h) = 2$. Thus, (4.18) gives that $\forall \lambda > 0, R_2(\lambda A) = 0$. □

For Theorem 4.2

Proof. We denote by Cov the covariance. A random field $\{X(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$ is called associated if

$$\mathrm{Cov}(f(X_I), g(X_I)) \geq 0$$

for any discrete finite subset $I \subset \mathbb{R}^d$ and for any bounded coordinatewise non-decreasing functions $f : \mathbb{R}^{\mathrm{card}(I)} \mapsto \mathbb{R}$, $g : \mathbb{R}^{\mathrm{card}(I)} \mapsto \mathbb{R}$ (card stands for cardinality), where $X_I = \{X(\mathbf{x}) : \mathbf{x} \in I\}$. Max-stable processes are associated and therefore positively associated (see Spodarev (2014)). Moreover, we have

$$\begin{aligned}
\sigma^2 &= \int_{\mathbb{R}^2} \mathrm{Cov}\left(\mathbf{I}_{\{Z(\mathbf{0})>u\}}, \mathbf{I}_{\{Z(\mathbf{x})>u\}}\right) d\mathbf{x}\\
&= \int_{\mathbb{R}^2} \mathbb{E}\left(\mathbf{I}_{\{Z(\mathbf{0})>u\}}\mathbf{I}_{\{Z(\mathbf{x})>u\}}\right) - \mathbb{E}(\mathbf{I}_{\{Z(\mathbf{0})>u\}})^2\, d\mathbf{x}\\
&= \int_{\mathbb{R}^2} \mathbb{P}(Z(\mathbf{0}) > u, Z(\mathbf{x}) > u) - \left[1 - \exp\left(-\frac{1}{u}\right)\right]^2 d\mathbf{x}\\
&= \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}.
\end{aligned}$$

Firstly, note that in the case of the max-stable models considered, $\Theta(\mathbf{x}) < 2$, i.e. $\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) > \exp\left(-\frac{2}{u}\right)$, on a set whose Lebesgue measure is positive. Therefore,

$$\int_{\mathbb{R}^2} \mathrm{Cov}\left(\mathbf{I}_{\{Z(\mathbf{0})>u\}}, \mathbf{I}_{\{Z(\mathbf{x})>u\}}\right) d\mathbf{x} > 0.$$

We now show that this integral converges.
In the case of the Smith model, recall that

$$\Theta(\mathbf{x}) = 2\Phi\left(\frac{\|\mathbf{x}\|}{2}\right). \tag{4.33}$$

Using (4.33), we have

$$\begin{aligned}
\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right) &= \exp\left(-\frac{2}{u}\right)\exp\left(-\frac{\Theta(\mathbf{x})}{u}+\frac{2}{u}\right) - \exp\left(-\frac{2}{u}\right) \\
&= \exp\left(-\frac{2}{u}\right)\left(\exp\left(\frac{2-\Theta(\mathbf{x})}{u}\right)-1\right) \\
&\underset{\|\mathbf{x}\|\to\infty}{\sim} \exp\left(-\frac{2}{u}\right)\left(\frac{2-\Theta(\mathbf{x})}{u}\right) \\
&= \frac{2\exp\left(-\frac{2}{u}\right)}{u}\left[1-\Phi\left(\frac{\|\mathbf{x}\|}{2}\right)\right] \\
&\underset{\|\mathbf{x}\|\to\infty}{\sim} \frac{2\exp\left(-\frac{2}{u}\right)}{u}\frac{\exp\left(-\frac{\|\mathbf{x}\|^2}{8}\right)}{\frac{\|\mathbf{x}\|}{2}} \\
&= \frac{4\exp\left(-\frac{2}{u}\right)}{u}\frac{\exp\left(-\frac{\|\mathbf{x}\|^2}{8}\right)}{\|\mathbf{x}\|},
\end{aligned}$$

which is clearly convergent.

In the case of the Brown-Resnick model whose variogram $\gamma(\mathbf{h})$ satisfies $\gamma(\mathbf{h}) \underset{\|\mathbf{h}\|\to\infty}{\sim} \|\mathbf{h}\|^a$ with $a > 0$, the convergence is obtained in exactly the same way. Indeed, we have $\Theta(\mathbf{x}) = 2\Phi\left(\sqrt{\frac{\gamma(\mathbf{x})}{2}}\right)$.

In the case of the tube model, we have $\Theta(\mathbf{x}) = 2$ for $\|\mathbf{x}\| \geq 2R_b$. Therefore, the term $\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)$ has a compact support and is integrable.

Finally, in all cases mentioned, we have $\sigma^2 < +\infty$. By applying Theorem 7 in Spodarev (2014), we obtain the result. □

For Theorem 4.3

Proof.

1. (a) Spatial invariance under translation:
   Since the max-stable process $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ is assumed to be strictly stationary, the same is true for the process $C_P(\mathbf{x}) = \mathbf{I}_{\{Z(\mathbf{x})>u\}}$. Therefore the invariance under translation directly follows from Proposition 4.2.

   (b) Spatial anti-monotonicity when the two regions are both a disk or a square:
   Let us consider two regions A_1 and A_2 being both a disk or a square and such that $A_1 \subset A_2$. Due to the spatial invariance under translation, the region A_2 can be translated to region A_2', where A_2' corresponds to the region obtained by an homothety of A_1, whose center is the center of A_1 and factor is denoted $\lambda > 1$. Thus $R_2(A_2) = R_2(A_2') = R_2(\lambda A_1)$. By Corollary 4.1, we know that $R_2(\lambda A)$ is a decreasing function of λ, giving that $R_2(\lambda A_1) \leq R_2(A_1)$. Therefore $R_2(A_2) \leq R_2(A_1)$ and the spatial anti-monotonicity is shown.

   (c) Spatial sub-additivity when the two regions are both a disk or a square:
   Due to Proposition 4.1, the spatial sub-additivity directly follows from the spatial anti-monotonicity.

2. Theorem 4.2 shows that $R_2(\lambda A) \underset{\lambda\to\infty}{\sim} \dfrac{\sigma^2}{\lambda^2}$. Therefore, by setting $K_1 = 0$ and $K_2 = \dfrac{\sigma^2}{R_2(A)}$, the axiom of spatial asymptotic homogeneity of order -2 is satisfied (see Definition 4.4).

□

For Theorem 4.4

Proof. Theorem 4.2 gives that

$$\lambda\left(L_N(\lambda A) - m\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \text{ for } \lambda \to \infty,$$

where

$$m = \left[1 - \exp\left(-\frac{1}{u}\right)\right] \text{ and } \sigma^2 = \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\Theta(\mathbf{x})}{u}\right) - \exp\left(-\frac{2}{u}\right)\right] d\mathbf{x}.$$

Therefore,

$$L_N(\lambda A) \xrightarrow{d} \mathcal{N}\left(m, \frac{\sigma^2}{\lambda^2}\right), \text{ for } \lambda \to \infty.$$

Thus, by definition of the convergence in distribution and using the classical formula of the VaR of a Gaussian random variable, we have

$$\begin{aligned} R_{3,1-\alpha}(\lambda A) = VaR_{1-\alpha}[L_N(\lambda A)] = m + \frac{\sigma}{\lambda} q_{1-\alpha} &= R_{3,1-\alpha}(A)\left(\frac{m}{R_{3,1-\alpha}(A)} + \frac{\sigma q_{1-\alpha}}{R_{3,1-\alpha}(A)}\frac{1}{\lambda}\right) \\ &= \left(K_1 + \frac{K_2}{\lambda}\right) R_{3,1-\alpha}(A), \end{aligned}$$

where $K_1 = \dfrac{m}{R_{3,1-\alpha}(A)}$ and $K_2 = \dfrac{\sigma q_{1-\alpha}}{R_{3,1-\alpha}(A)}$. □

For Lemma 4.2

Proof. In the case of a Fréchet distribution $\mathcal{F}(\nu, 1, 0)$, it is well known that the moment of order k, denoted by μ_k, exists if and only if $k < \nu$ and that $\mu_k = \Gamma\left(1 - \dfrac{k}{\nu}\right)$. In the case of a random variable $Z \sim \mathcal{F}(1, 1, 0)$, we have that for $\beta > 0$,

$$\mathbb{P}(Z \le z) = \exp\left(-\frac{1}{z}\right), \text{ giving } \mathbb{P}(Z^\beta \le z^\beta) = \exp\left(-\frac{1}{z}\right).$$

Then, by setting $U = Z^\beta$ and $u = z^\beta$,

$$\mathbb{P}(U \le u) = \exp\left(-\frac{1}{u^{\frac{1}{\beta}}}\right),$$

giving $Z^\beta \sim \mathcal{F}\left(\dfrac{1}{\beta}, 1, 0\right)$. Since $\mathcal{F}(\nu, 1, 0)$ has a second order moment if and only if $\nu > 2$, Z^β has a second order moment if and only if $\dfrac{1}{\beta} > 2$ i.e. $\beta < \dfrac{1}{2}$. Moreover, $\mathbb{E}(Z^\beta) = \Gamma(1 - \beta)$. □

For Theorem 4.5

Proof. We have

$$\mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] = \int_0^{+\infty} \int_0^{+\infty} z_1^{\beta_1} z_2^{\beta_2} h(z_1, z_2)\ dz_1\ dz_2.$$

In order to take advantage of the decomposition intensity/radius of the multivariate extreme value distributions, it is relevant to make the following change of variable:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} u \\ \theta\ u \end{pmatrix} = \begin{pmatrix} \Phi_1[u,\theta] \\ \Phi_2[u,\theta] \end{pmatrix}.$$

The corresponding Jacobian matrix is written

$$J_\Phi(u,\theta) = \begin{pmatrix} \dfrac{\partial \Phi_1}{\partial u} & \dfrac{\partial \Phi_1}{\partial \theta} \\ \dfrac{\partial \Phi_2}{\partial u} & \dfrac{\partial \Phi_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \theta & u \end{pmatrix},$$

yielding $|J_\Phi(u,\theta)| = u$.
Therefore, denoting by $a(z_1, z_2) = z_1^{\beta_1} z_2^{\beta_2} h(z_1, z_2)$, we have

$$\begin{aligned}
\mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] &= \int_0^{+\infty} \int_0^{+\infty} z_1^{\beta_1} z_2^{\beta_2} h(z_1, z_2)\ dz_1\ dz_2 \\
&= \int_0^{+\infty} \int_0^{+\infty} a(z_1, z_2)\ dz_1\ dz_2 \\
&= \int \int_{\Phi^{-1}([0,+\infty]^2)} a[\Phi(u,\theta)]\ |J_\Phi(u,\theta)|\ du\ d\theta \\
&= \int_0^{+\infty} \int_0^{+\infty} u^{\beta_1} \theta^{\beta_2} u^{\beta_2} h(u, \theta u)\ u\ du\ d\theta \\
&= \int_0^{+\infty} \int_0^{+\infty} u^{\beta_1+\beta_2+1} \theta^{\beta_2} h(u, \theta u)\ du\ d\theta. \qquad (4.34)
\end{aligned}$$

In the case of the Smith process, Padoan et al. (2010) provide

$$h(z_1, z_2) = \exp\left[-\frac{\Phi(w)}{z_1} - \frac{\Phi(v)}{z_2}\right] \times \left[\left(\frac{\Phi(w)}{z_1^2} + \frac{\phi(w)}{h z_1^2} - \frac{\phi(v)}{h z_1 z_2}\right) \times \left(\frac{\Phi(v)}{z_2^2} + \frac{\phi(v)}{h z_2^2} - \frac{\phi(w)}{h z_1 z_2}\right) + \left(\frac{v\phi(w)}{h^2 z_1^2 z_2} + \frac{w\phi(v)}{h^2 z_1 z_2^2}\right)\right], \qquad (4.35)$$

where

$$w = w(z_2, z_1) = \frac{h}{2} + \frac{\log\left(\dfrac{z_2}{z_1}\right)}{h} \text{ and } v = v(z_1, z_2) = \frac{h}{2} - \frac{\log\left(\dfrac{z_2}{z_1}\right)}{h}.$$

We have

$$w(z_1, \theta z_2) = \frac{h}{2} + \frac{\log(\theta)}{h} \text{ and } v(z_1, \theta z_1) = \frac{h}{2} - \frac{\log(\theta)}{h}.$$

From (4.35), it follows that

$$
\begin{aligned}
&h(z_1, \theta z_1) \\
&= \exp\left[-\frac{\Phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{z_1} - \frac{\Phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{\theta z_1}\right] \times \left[\left(\frac{\Phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{z_1^2} + \frac{\phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h z_1^2} - \frac{\phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h\theta z_1^2}\right)\right. \\
&\quad \times \left(\frac{\Phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{\theta^2 z_1^2} + \frac{\phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h\theta^2 z_1^2} - \frac{\phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h\theta z_1^2}\right) \\
&\quad \left. + \left(\frac{(\frac{h}{2}-\frac{\log(\theta)}{h})\ \phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h^2\theta z_1^3} + \frac{(\frac{h}{2}+\frac{\log(\theta)}{h})\ \phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h^2\theta^2 z_1^3}\right)\right] \\
&= \exp\left[-\frac{1}{z_1}\left(\Phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right) + \frac{\Phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{\theta}\right)\right] \times \left[\frac{1}{z_1^4}\left(\Phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right) + \frac{\phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h}\right.\right. \\
&\quad \left. - \frac{\phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h\theta}\right) \times \left(\frac{\Phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{\theta^2} + \frac{\phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h\theta^2} - \frac{\phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h\theta}\right) \\
&\quad \left. + \frac{1}{z_1^3}\left(\frac{(\frac{h}{2}-\frac{\log(\theta)}{h})\ \phi\left(\frac{h}{2}+\frac{\log(\theta)}{h}\right)}{h^2\theta} + \frac{(\frac{h}{2}+\frac{\log(\theta)}{h})\ \phi\left(\frac{h}{2}-\frac{\log(\theta)}{h}\right)}{h^2\theta^2}\right)\right] \\
&= \exp\left[-\frac{C_1(\theta)}{z_1}\right] \times \left[\frac{C_2(\theta)}{z_1^4} + \frac{C_3(\theta)}{z_1^3}\right].
\end{aligned}
$$

Thus, using (4.34), we obtain

$$
\begin{aligned}
\mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] &= \int_0^{+\infty}\int_0^{+\infty} u^{\beta_1+\beta_2+1}\theta^{\beta_2} h(u, \theta u)\ du\ d\theta \\
&= \int_0^{+\infty} \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2+1} \exp\left[-\frac{C_1(\theta)}{u}\right] \times \left[\frac{C_2(\theta)}{u^4} + \frac{C_3(\theta)}{u^3}\right]\ du\right) d\theta \\
&= \int_0^{+\infty} \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2+1} \exp\left[-\frac{C_1(\theta)}{u}\right] \times \left[\frac{C_2(\theta)}{u^4} + \frac{C_3(\theta)}{u^3}\right]\ du\right) d\theta \\
&= \int_0^{+\infty} C_2(\theta)\ \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2-3} \exp\left[-\frac{C_1(\theta)}{u}\right]\ du\right) d\theta \\
&\quad + \int_0^{+\infty} C_3(\theta)\ \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2-2} \exp\left[-\frac{C_1(\theta)}{u}\right]\ du\right) d\theta \\
&= \int_0^{+\infty} C_2(\theta)\ \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2-1}\ \frac{1}{u^2}\ \exp\left[-\frac{C_1(\theta)}{u}\right]\ du\right) d\theta \\
&\quad + \int_0^{+\infty} C_3(\theta)\ \theta^{\beta_2}\left(\int_0^{+\infty} u^{\beta_1+\beta_2}\ \frac{1}{u^2} \exp\left[-\frac{C_1(\theta)}{u}\right]\ du\right) d\theta \\
&= \int_0^{+\infty} \frac{C_2(\theta)}{C_1(\theta)}\ \theta^{\beta_2}\ \mu_{\beta_1+\beta_2-1}\left(\mathcal{F}(1, C_1(\theta))\right)\ d\theta \\
&\quad + \int_0^{+\infty} \frac{C_3(\theta)}{C_1(\theta)}\ \theta^{\beta_2}\ \mu_{\beta_1+\beta_2}\left(\mathcal{F}(1, C_1(\theta))\right)\ d\theta.
\end{aligned}
$$

For a random variable X following $\mathcal{F}(1, \sigma_f, 0)$ where σ_f is the scale parameter, the density is written

$$f(x) = \frac{\sigma_f}{x^2} \exp\left(-\frac{\sigma_f}{x}\right), \quad \text{yielding the following moments of order } k$$

$$\mu_k = \int_0^{+\infty} x^k \frac{\sigma_f}{x^2} \exp\left(-\frac{\sigma_f}{x}\right) dx = \int_{+\infty}^{0} \frac{\sigma_f^k}{t^k} \frac{\sigma_f t^2}{\sigma_f^2} \exp(-t) \frac{-\sigma_f}{t^2}\, dt = \sigma_f^k \int_0^{+\infty} t^{-k} \exp(-t)\, dt$$
$$= \sigma_f^k \Gamma(1-k).$$

Therefore

$$\mu_{\beta_1+\beta_2-1}(\mathcal{F}(1, C_1(\theta)) = C_1(\theta)^{\beta_1+\beta_2-1}\Gamma(2-\beta_1-\beta_2) \text{ and}$$
$$\mu_{\beta_1+\beta_2}(\mathcal{F}(1, C_1(\theta)) = C_1(\theta)^{\beta_1+\beta_2}\Gamma(1-\beta_1-\beta_2),$$

completing the proof. □

For Corollary 4.4

Proof. We have

$$\text{Cov}\left(Z_1^{\beta_1}, Z_2^{\beta_2}\right) = \mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] - \mathbb{E}\left[Z_1^{\beta_1}\right] \mathbb{E}\left[Z_2^{\beta_2}\right] = \mathbb{E}\left[Z_1^{\beta} Z_2^{\beta}\right] - \Gamma[1-\beta_1]\ \Gamma[1-\beta_2],$$

yielding

$$\text{Corr}\left(Z_1^{\beta_1}, Z_2^{\beta_2}\right) = \frac{\mathbb{E}\left[Z_1^{\beta_1} Z_2^{\beta_2}\right] - \Gamma[1-\beta_1]\ \Gamma[1-\beta_2]}{\sigma(Z^{\beta_1})\sigma(Z^{\beta_2})}.$$

We know that if $Z \rightsquigarrow \mathcal{F}(1,1,0)$, we have $Z^\beta \rightsquigarrow \mathcal{F}\left(\frac{1}{\beta}, 1, 0\right)$ and therefore

$$\mu_2(Z^\beta) = \Gamma\left(1-2\beta\right),$$

giving

$$\text{var}(Z^\beta) = \Gamma\left(1-2\beta\right) - \left[\Gamma\left(1-\beta\right)\right]^2 \tag{4.36}$$

The result is obtained combining (4.20) and (4.36). □

For Theorem 4.6

Proof. We have

$$\mathbb{E}[L(A)^2] = \mathbb{E}\left[\int_A Z(\mathbf{x})^\beta d\mathbf{x} \int_A Z(\mathbf{y})^\beta d\mathbf{y}\right] = \mathbb{E}\left[\int_A \int_A Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta d\mathbf{x}\ d\mathbf{y}\right]$$
$$= \int_A \int_A \mathbb{E}\left[Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta\right] d\mathbf{x}\ d\mathbf{y}.$$

Therefore,

$$R_2(A) = \text{var}\left[\frac{L(A)}{|A|}\right] = \frac{1}{|A|^2}\ \left(\mathbb{E}\left[(L(A))^2\right] - [\mathbb{E}(L(A))]^2\right)$$
$$= \frac{1}{|A|^2} \int_A \int_A \mathbb{E}\left[Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta\right] d\mathbf{x}\ d\mathbf{y} - \Gamma(1-\beta)^2.$$
$$= \frac{1}{|A|^2} \int_A \int_A \left[\mathbb{E}\left[Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta\right] - \Gamma(1-\beta)^2\right] d\mathbf{x}\ d\mathbf{y}.$$

The result is obtained upon replacing A by λA. □

For Corollary 4.6

Proof. Using (4.21), applying the same arguments than in the proof of Corollary 4.1 (for example (4.24)) and taking advantage of the expression of $\mathbb{E}\left[Z(\mathbf{x})^\beta\ Z(\mathbf{y})^\beta\right]$ given by (4.20), we obtain the first two items.

Regarding the last one, we must show that $\lim_{h\to\infty} g(h) = [\Gamma(1-\beta)]^2$. We first study the behavior of functions C_1, C_2 and C_3 defined in Theorem 4.5. It is easy to show that $\lim_{h\to\infty} C_1(\theta) = 1+\frac{1}{\theta}$, $\lim_{h\to\infty} C_2(\theta) = \frac{1}{\theta^2}$ and $\lim_{h\to\infty} C_3(\theta) = 0$. Moreover, for h higher than a given threshold, we can upper-bound the integrand of the integral in (4.20) by a function depending only on θ. Thus, we are in the conditions allowing to permute the limit and the integral. Hence,

$$\begin{aligned}\lim_{h\to\infty} g(h) &= \int_0^{+\infty} \theta^\beta \left[\frac{1}{\theta^2}\left(1+\frac{1}{\theta}\right)^{2(\beta-1)} \Gamma(2-2\beta)\right] d\theta \\ &= \Gamma(2-2\beta)\int_0^{+\infty} \theta^\beta \left[\frac{1}{\theta^2}\left(\frac{\theta+1}{\theta}\right)^{2(\beta-1)}\right] d\theta \\ &= \Gamma(2-2\beta)\int_0^{+\infty} \theta^{-\beta}(\theta+1)^{2(\beta-1)}\ d\theta. \end{aligned} \tag{4.37}$$

Moreover, it is well known that

$$B(1-\beta,1-\beta) = \frac{\Gamma(1-\beta)\Gamma(1-\beta)}{\Gamma[2(1-\beta)]} = \frac{[\Gamma(1-\beta)]^2}{\Gamma(2-2\beta)}, \tag{4.38}$$

where B stands for the beta distribution. By definition of the beta distribution, and then using the change of variable $\theta = \frac{u}{1-u} \Leftrightarrow u = \theta(1-u) \Leftrightarrow u = \frac{\theta}{1+\theta}$ (giving $du = \frac{1+\theta-\theta}{(1+\theta)^2}\ d\theta = \frac{d\theta}{(1+\theta)^2}$), we also have

$$\begin{aligned} B(1-\beta,1-\beta) = \int_0^1 u^{-\beta}(1-u)^{-\beta} &= \int_0^{+\infty} \left(\frac{\theta}{1+\theta}\right)^{-\beta}\left(\frac{1}{1+\theta}\right)^{-\beta}\frac{d\theta}{(1+\theta)^2} \\ &= \int_0^{+\infty} \theta^{-\beta}(\theta+1)^{2(\beta-1)}\ d\theta. \end{aligned} \tag{4.39}$$

The combination of (4.38) and (4.39) yields

$$\Gamma(2-2\beta)\int_0^{+\infty} \theta^{-\beta}(\theta+1)^{2(\beta-1)}\ d\theta = [\Gamma(1-\beta)]^2,$$

and (4.37) gives the result.

□

For Theorem 4.7

Proof. Let $\{Z(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^2}$ be the Smith process. Thus, Z is strongly mixing and applying Lemma 4.1, Z^β is strongly mixing too. We now show that the three conditions of Theorem 1 in Gorodetskii (1987) are satisfied.

Firstly, we have

$$\sup_{\mathbf{x}} \mathbb{E}\left[|Z(\mathbf{x})^{\beta}|^{s}\right] = \sup_{\mathbf{x}} \mathbb{E}\left[Z(\mathbf{x})^{\beta s}\right],$$

which is finite if $\beta s < 1$ i.e. $s < \frac{1}{\beta}$. Recall that $\beta < \frac{1}{2}$ i.e. $\frac{1}{\beta} > 2$. Thus, it is easy to find some $s > 2$ such that

$$\sup_{\mathbf{x}} \mathbb{E}\left[|Z(\mathbf{x})^{\beta}|^{s}\right] < +\infty.$$

Secondly, the majoration of the β-mixing coefficient given by Dombry and Eyi-Minko (2012) is the following:

$$\beta(\mathbf{x}_1, \mathbf{x}_2) \leq 4[2 - \Theta(h)], \tag{4.40}$$

where $h = \|\mathbf{x}_2 - \mathbf{x}_1\|$. Moreover, it is well known that

$$\alpha(\mathbf{x}_1, \mathbf{x}_2) \leq \frac{1}{2}\beta(\mathbf{x}_1, \mathbf{x}_2). \tag{4.41}$$

Let us denote $\alpha(\mathbf{x}_1, \mathbf{x}_2)$ and $\beta(\mathbf{x}_1, \mathbf{x}_2)$ by $\alpha(h)$ and $\beta(h)$, respectively. Combining (4.40) and (4.41), we have

$$\begin{aligned}\int_0^{+\infty} h[\alpha(h)]^{\frac{s-2}{s}} dh &\leq \int_0^{+\infty} h \frac{1}{2^{\frac{s-2}{2}}} [\beta(h)]^{\frac{s-2}{s}} dh \\ &\leq \int_0^{+\infty} h 2^{\frac{s-2}{2}} [2 - \Theta(h)]^{\frac{s-2}{s}} dh.\end{aligned}$$

Now, using (4.33), we obtain

$$h2^{\frac{s-2}{2}}[2 - \Theta(h)]^{\frac{s-2}{s}} = h2^{\frac{s-2}{2}} \left(2\left[1 - \Phi\left(\frac{h}{2}\right)\right]\right)^{\frac{s-2}{s}} \underset{h \to +\infty}{\sim} h2^{\frac{s-2}{2}} \left[\frac{4\exp\left(-\frac{h^2}{8}\right)}{h}\right]^{\frac{s-2}{s}},$$

which is integrable. Thus, $\int_0^{+\infty} h[\alpha(h)]^{\frac{s-2}{s}} dh < +\infty$.

Finally, we have

$$\begin{aligned}\sigma^2 &= \int_{\mathbb{R}^2} \mathrm{Cov}\left(Z(\mathbf{0})^{\beta}, Z(\mathbf{x})^{\beta}\right) d\mathbf{x} \\ &= \int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^{\beta} Z(\mathbf{x})^{\beta}] - \mathbb{E}[Z(\mathbf{0})^{\beta}]^2\right) d\mathbf{x} \\ &= \int_{\mathbb{R}^2} \left(\mathbb{E}[Z(\mathbf{0})^{\beta} Z(\mathbf{x})^{\beta}] - [\Gamma(1-\beta)]^2\right) d\mathbf{x}.\end{aligned}$$

Firstly, we can show that $\mathrm{Cov}\left(Z(\mathbf{0})^{\beta}, Z(\mathbf{x})^{\beta}\right) > 0$. Moreover, we have that $\|Z(\mathbf{0})^{\beta}\|_q = \mathbb{E}[Z(\mathbf{0})^{\beta q}]^{\frac{1}{q}} < +\infty$ if and only if $\beta q < 1$ i.e. $q < \frac{1}{\beta}$. For the same reason, $\|Z(\mathbf{0})^{\beta}\|_r$ is defined if and only if $r < \frac{1}{\beta}$. Davydov's inequality yields that for $q > 1, r > 1$ such that $\frac{1}{q} + \frac{1}{r} = 1 - \frac{1}{p}$, we have

$$|\mathrm{Cov}\left(Z(\mathbf{0})^{\beta}, Z(\mathbf{x})^{\beta}\right)| \leq 2p[2\alpha(\|\mathbf{x}\|)]^{\frac{1}{p}} \|Z(\mathbf{0})^{\beta}\|_q \|Z(\mathbf{x})^{\beta}\|_r. \tag{4.42}$$

Recall that $\frac{1}{\beta} > 2$. Thus the conditions $q, r > 1$ and $q, r < \frac{1}{\beta}$ are compatible.

Combining (4.33), (4.40), (4.41) and (4.42), we obtain

$$\begin{aligned}
|\mathrm{Cov}\left(Z(\mathbf{0})^\beta, Z(\mathbf{x})^\beta\right)| &\leq 2p\ 8^{\frac{1}{p}}(2-\Theta(\|\mathbf{x}\|))^{\frac{1}{p}}\|Z(\mathbf{0})^\beta\|_q\|Z(\mathbf{x})^\beta\|_r \\
&= 2p\ 16^{\frac{1}{p}}\left(1-\Phi\left(\frac{\|\mathbf{x}\|}{2}\right)\right)\|Z(\mathbf{0})^\beta\|_q\|Z(\mathbf{x})^\beta\|_r \\
&\underset{\|\mathbf{x}\|\to\infty}{\sim} 4p\ 16^{\frac{1}{p}}\exp\left(-\frac{\|\mathbf{x}\|^2}{8}\right)\|Z(\mathbf{0})^\beta\|_q\|Z(\mathbf{x})^\beta\|_r.
\end{aligned}$$

Therefore, $\sigma^2 < +\infty$, proving the result. □

Bibliography

Acciaio, B. and Penner, I. (2011). Dynamic risk measures. In *Advanced Mathematical Methods for Finance*, pages 1–34. Springer.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

Berman, S. (1992). *Sojourns and Extremes of Stochastic Processes*. CRC Press.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. *Dependence in Probability and Statistics, Lecture Notes in Statistics*, 187:373–390.

Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of Value-at-Risk. *Journal of Multivariate Analysis*, 119:32–46.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

de Haan, L. (1984). A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer.

de Haan, L. and Pickands, J. (1986). Stationary min-stable stochastic processes. *Probability Theory and Related Fields*, 72(4):477–492.

Dombry, C. and Eyi-Minko, F. (2012). Strong mixing properties of max-infinitely divisible random fields. *Stochastic Processes and their Applications*, 122(11):3790–3811.

Donat, M., Pardowitz, T., Leckebusch, G., Ulbrich, U., and Burghoff, O. (2011). High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Natural Hazards and Earth System Science*, 11(10):2821–2833.

Dorland, C., Tol, R. S., and Palutikof, J. P. (1999). Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change. *Climatic Change*, 43(3):513–535.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, volume 33. Springer.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526–547.

Föllmer, H. (2014). Spatial risk measures and their local specification: the locally law-invariant case. *Statistics & Risk Modeling*, 31(1):79–101.

Föllmer, H. and Klüppelberg, C. (2014). Spatial risk measures: Local specification and boundary risk.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447.

Föllmer, H. and Schied, A. (2004). *Stochastic Finance: An Introduction in Discrete Time*, volume 27. de Gruyter Studies in Mathematics.

Frittelli, M. and Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

Gorodetskii, V. (1987). Moment inequalities and the central limit theorem for integrals of random fields with mixing. *Journal of Soviet Mathematics*, 36(4):461–467.

Guidolin, M. and Timmermann, A. (2006). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, 131(1):285–308.

Ivanov, A. V., Leonenko, N., Ruiz-Medina, M. D., Savich, I. N., et al. (2013). Limit theorems for weighted nonlinear transformations of Gaussian stationary processes with singular spectra. *The Annals of Probability*, 41(2):1088–1114.

Kabluchko, Z. and Schlather, M. (2010). Ergodic properties of max-infinitely divisible processes. *Stochastic Processes and their Applications*, 120(3):281–295.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065.

Khanduri, A. and Morrow, G. (2003). Vulnerability of buildings to windstorms and insurance loss estimation. *Journal of Wind Engineering and Industrial Aerodynamics*, 91(4):455–467.

Klawa, M. and Ulbrich, U. (2003). A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Science*, 3(6):725–732.

Koltchinskii, V. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25(2):435–477.

Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif ? *Résumé*, 5:2.

Moltchanov, D. (2012). Distance distributions in random networks. *Ad Hoc Networks*, 10(6):1146–1166.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Penrose, M. D. (1992). Semi-min-stable processes. *The Annals of Probability*, 20(3):1450–1463.

Perrin, O., Rootzén, H., and Taesler, R. (2006). A discussion of statistical methods used to estimate extreme wind speeds. *Theoretical and Applied Climatology*, 85(3):203–215.

Pinto, J., Fröhlich, E., Leckebusch, G., and Ulbrich, U. (2007). Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM. *Natural Hazards and Earth System Science*, 7(1):165–175.

Prettenthaler, F., Albrecher, H., Köberl, J., and Kortschak, D. (2012). Risk and insurability of storm damages to residential buildings in Austria. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 37(2):340–364.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished manuscript, University of North Carolina*.

Spodarev, E. (2014). Limit theorems for excursion sets of stationary random fields. In *Modern Stochastics and Applications*, volume 90, pages 221–241. Springer.

Stoev, S. A. (2010). Max-stable processes: Representations, ergodic properties and statistical applications. *Dependence in Probability and Statistics, Lecture Notes in Statistics*, 200:21–42.

SwissRe (2014). Natural catastrophes and man-made disasters in 2013: large losses from floods and hail; Haiyan hits the Philippines. *Swiss Re Sigma Report*.

Volkonskii, V. and Rozanov, Y. A. (1959). Some limit theorems for random functions. I. *Theory of Probability and Its Applications*, 4(2):178–197.

Weintraub, K. S. (1991). Sample and ergodic properties of some min-stable processes. *The Annals of Probability*, 19(2):706–723.

Yuen, R. and Stoev, S. (2013). Crps m-estimation for max-stable models. *Extremes*, 17(3):387–410.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.

Chapter 5

Diversification and endogenous financial networks

We test the hypothesis that interconnections across financial institutions can be explained by a diversification motive. This idea stems from the empirical evidence of the existence of long-term exposures that cannot be explained by a liquidity motive (maturity or currency mismatch). We model endogenous interconnections of heterogenous financial institutions facing regulatory constraints using a maximization of their expected utility. Both theoretical and simulation-based results are compared to a stylized genuine financial network. The diversification motive appears to plausibly explain interconnections among key players. Using our model, the impact of regulation on interconnections between banks -currently discussed at the Basel Committee on Banking Supervision- is analyzed.

Key words: Diversification; Financial networks; Regulation; Solvency; Systemic risk.

5.1 Introduction

The behavior of financial institutions, namely banks and insurance companies, constitutes a paradox. On the one hand, they oppose one another in a competition to collect deposits as one may expect for firms in a common sector. In this perspective, the distress of one institution seems good news for the others since there is room for increasing market shares. However, on the other hand, financial institutions need to cooperate. For instance, the withdrawal of a bank from the short term interbank market means that a source of liquidity vanishes, triggering setbacks for other banks. In this case, one financial institution's distress is definitely bad news for the other ones. Thus, even if they are in competition, banks cooperate, insurance companies cooperate and last but not least, banks cooperate with insurance companies. The last point has been ever more significant during the recent years. A support of this cooperation is the interconnections they develop between each other.

In a short-term view, interconnections mirror the resolution of the liquidity needs. As any other firms, banks and insurance companies face asynchronous in-flows and out-flows of cash. One solution is that every institution has its own cash buffer. Alternatively, institutions can create a liquidity pool by sharing their cash to mutualize the liquidity risk (Holmstrom and Tirole, 1996; Tirole, 2010; Rochet, 2004). Allen and Gale (2000) explicitly link the interconnectedness of banks to liquidity shocks. Besides the asynchro-

Table 1:
Interbank assets and liabilities by maturity and broad bank categories
(end December 1998)

(1): excl. interbank assets and liabilities vis-à-vis foreign banks, building societies and Bundesbank
(2): all interbank assets and liabilities

			maturity				
			daily	> 1 day & < 3 mths	≥ 3 mths & ≤ 1 yr	> 1 yr & < 4 yrs	≥ 4 yrs
all banks	**liabilities**	**(1)**	**11.7**	**13.2**	**14.5**	**5.6**	**55.0**
		(2)	*12.6*	*24.3*	*17.4*	*4.5*	*41.1*
	assets	**(1)**	**10.6**	**14.1**	**17.2**	**5.6**	**52.5**
		(2)	*11.2*	*18.8*	*18.0*	*6.0*	*46.0*

Figure 5.1: Extract of Table 1 in Upper and Worms (2004).

nism of in-flows and out-flows, the liquidity risk is particularly salient since banks are exposed to runs (Diamond and Dybvig, 1983) and operate large gross transactions in payment systems (Rochet and Tirole, 1996). Indeed flows between institutions are not netted.

However, one may argue that this analysis is not specific to banks and insurance companies since every firm actually faces asynchronous in-flows and out-flows. Liquidity concerns are not the only cause of interconnections between financial institutions. Moreover, there is evidence in the literature that banks are interconnected not only in the short term but also in the long run. For instance, according to Upper and Worms (2004), half of German interbank lending is composed of loans whose maturity is over 4 years (see Figure 5.1). According to Table 1 in Alves et al. (2013), interbank assets with residual maturity larger than one year account for about 50% of total interbank assets at the European level.[1] These long term exposures cannot be explained by a liquidity motive since liquidity is a short term phenomenon. Other possible reasons are horizontal integration (share of a pool of customers via joint products), vertical integration (e.g. risk transfer between insurance and reinsurance companies), ego of top managers aiming at increasing their control of the market and last but not least diversification. Of course, in practice, the network formation stems from a combination of all these motives. However, for reasons explained further, diversification appears as a very important motive. Therefore, in this paper, we consider that these long-term exposures are accounted for by a diversification principle, in a sense that will be defined in the following.

The diversification principle is supported by the existence of various business models for banks and insurance companies. The diversity of institutions leads to a diversity of debts and shares available for the other institutions as assets. In the case of insurance companies, there is a clear-cut distinction between mutual funds and profit-oriented insurance companies. The banking sector regroups heterogenous institutions from mutual saving banks to commercial banks. Moreover, the bankassurance business model blurs the separation between banking and insurance activities. This variety can be explained by the different preferences of stakeholders or by historical patterns. Investors who have the same risk aversion gather and form an institution. This heterogeneity can also be linked to a specialization process. For instance, a mutual savings bank funded by farmers is very efficient in granting loans to farmers who in turn favor this bank since it provides

[1]The existence of long-term interconnections, through loans or shares, is also reported for other countries such as Canada (see Table 3 in Gauthier et al. (2012)) or France (Fourel et al., 2013).

the fairest interest rate. This auto-selection mechanism leads to a situation close to a local monopoly. We then understand that for a specific bank, getting interconnected to other institutions is a way to get access to their specific markets. Considering specific markets implicitly assumes that these are not perfectly correlated: for example retail differs from trading. Similarly, insurance companies also specialize in specific risk classes.[2] Thus being interconnected to other insurance companies allows diversifying one's risk portfolio. All this supports the fact that the diversification principle may explain long-term interconnections among banks and insurance companies.

In order to properly model banking and insurance activities, one has to keep in mind that the banking and insurance sectors are characterized by a very specific production process as well as a heavy regulation. The core activity of a bank consists of the selection of profitable loans and in the management of the resulting maturity transformation. Banks screen potential entrepreneurs for reliable projects and fairly price the resulting interest rate charged. At the same time, they manage the maturity gap: loans to entrepreneurs are long-term assets whereas deposits and issued bonds constitute short-term debt on the liability side. Information is also key to the core activity of insurance (e.g. damage insurance): the insurer has to efficiently assess the riskiness of the potential policyholder and to deduce the corresponding premium. Strictly speaking the insurance company does not provide maturity transformation. However, its production cycle is reversed: it first collects premia and cushions losses when claims occur. The regulation of the banking and insurance sectors is crucial to maintain people's confidence in the system. In order to avoid bank runs, it is necessary that depositors consider their deposit as safe. Likewise, if policy-holders are not confident in the capacity of the insurer to honor its commitments, they will make other insurance choices. A solvency ratio is imposed to banks and insurance companies: in the case of banks, the ratio compares the riskiness of granted loans with own funds, while for insurance companies, the ratio balances the riskiness of insured risks and the collected premia.
Our paper has two main objectives. The first objective is to test whether a diversification motive is a plausible cause for interconnectedness across financial institutions. To do so, we build a model where interconnections are endogenous choices of financial institutions resulting from the maximization of their expected utility. After deriving some theoretical and simulation-based features of the resulting network, we compare these features to those of a stylized financial network (benchmark) based on empirical evidence. The second objective is to fairly assess the impact of regulation on interconnections based on our model.

The cornerstone of this paper is the modeling of the endogenous balance sheet of a financial institution, especially interconnections. Endogenous networks have been intensively analyzed in sociology or socio-economics (for a survey, see Goyal (2012) or Jackson and Zenou (2013)). However, finance yields a new field of application. Usually, interconnections among financial institutions are considered as given, especially in applied papers (see among others Cifuentes et al. (2005), Arinaminpathy et al. (2012), or Anand et al. (2013)). Endogenous financial networks stem from the seminal paper by Allen and Gale (2000). For instance, Babus (2007) models interconnections across banks as the result of an insurance motive: interconnections represent a means of protection against contagion. More recently, Acemoglu et al. (2013) focus on the short-term interbank market

[2]For instance, in the US health insurance sector, specialized institutions exist. The Federal Employees Health Benefits (FEHB) Program is dedicated to federal employees.

and model the network formation in the case of risk-neutral banks being able to renegotiate their claims in a case of distress. Elliott et al. (2014) make a case of showing the incentives that may drive financial network formations. Important insights are brought by this strand of literature inspired by microeconomics and game theory analysis.[3] Nevertheless, in this field, financial institutions only compute their interconnections: the remaining elements of their balance sheet are completely exogenous. This assumption seems suitable in a short-term perspective but not anymore when considering long-term interconnections. Therefore, by including more endogenous balance sheet items than the sole interconnections, we distance ourselves from this strand of research. To the best of our knowledge, the unique paper that considers a complete balance sheet optimization (apart from the debt) is Bluhm et al. (2013). They propose a dynamic network formation with risk-neutral banks. Using a specific "trial and error" process, the authors first compute the volume of interbank assets (that corresponds to the network's importance) and second its allocation (that corresponds to the network's shape). The allocation is carried out using a matching algorithm based on the strict indifference of banks. In contrast, our paper considers heterogeneously risk-averse banks which explicitly get interconnected to specific counterparts. Last but not least, almost all papers only consider lending (or debt securities) whereas, based on Gouriéroux et al. (2012), our paper also considers shares. This feature cannot be neglected in a long run perspective since financial institutions can take cross-shareholdings.

The paper is organized as follows. Section 2 falls into two parts. First, the production process of banks and insurance companies and the regulatory constraints are described. Secondly, we introduce the financial network benchmark. Section 3 presents the theoretical results. After describing the optimization program of financial institutions, we show the existence of an equilibrium and discuss the conditions for its uniqueness. We show that interconnections are usually optimal for financial institutions. These theoretical properties allow to characterize the shape of the network stemming from a diversification motive. Therefore, we compare the shape of a genuine interbank network to a diversification-based one. In Section 4, we first present the computational methodology and the calibration choices. Then we show some simulation results which lead us to assess the proximity of the obtained network to the benchmark network both in terms of balance sheet volume and support of interconnections (debt securities or cross-shareholdings). Section 5 provides an analysis of financial interconnections with respect to financial regulation. Elaborating on Repullo and Suarez (2013), we first show how to fairly analyze interconnectedness and then compare different regulatory frameworks. Section 6 concludes. All proofs are gathered in the Appendix.

5.2 Balance sheet structure and network benchmark

In this section, we first describe the economic setup which corresponds to the technology of financial institutions. We introduce the different elements of their balance sheet as well as the regulatory constraints. We then present the stylized network to be later used as a benchmark.

[3]See among others Cohen-Cole et al. (2011), Gofman (2012), Farboodi (2014) or Georg (2014).

5.2.1 Bank and insurance business

Each bank has access to a specific class of external illiquid assets and each insurance company specializes in one specific class of risk. These classes can be interpreted as main banking (respectively insurance) activities such as, for instance, trading, commercial loans, mortgage loans, sovereign loans (respectively e.g. property insurance, liability insurance, life insurance).
The tight relationship between a specific class of assets (respectively risks) and a specific institution has to be interpreted as a consequence of costly portfolio management by investors followed by a specialization process. By portfolio management, we mean the screening process. For banks, that means selecting promising entrepreneurs to finance and offering a fair interest rate. In the case of insurance companies, it means organizing the mutualization of risks, i.e. finding the adequate premium with respect to the policyholder's risk profile. The specialization process strengthens the efficiency of managing a specific portfolio. Due to auto-selection of customers, specialization triggers further specialization.

Asset side

Bank i's specific asset book value is labeled Ax_i, for $i = 1, \dots, n$ (we consider n financial institutions). This asset is some illiquid loan and therefore cannot be exchanged on a market. Thus, no market value can be defined and only its book value is considered in the following. We denote R_i and r_i the net return of Ax_i and its realization, respectively. The distribution function of the returns $R_1, \dots, R_n$ is denoted F_R: $F_R(r_1, \dots, r_n) = \mathbb{P}(R_1 \leq r_1, \dots, R_n \leq r_n)$. The corresponding density is denoted by f_R.
Banks have access to another external asset, denoted by $A\ell_i$. Its return, deterministic and assumed to be common to all institutions, is denoted r_{rf}. Here, $A\ell_i$ is a very liquid and low-risk asset (for instance AAA bonds or S&P 500 shares), the management of which does not require high technical skills. In the following, $A\ell_i$ will be assimilated to cash, which does not require any screening. We assume that insurance companies' external assets are only composed of $A\ell_i$. Insurance companies are indeed assumed not to have the same capacity of selecting promising innovators as banks, and therefore do not own any specific asset.
Besides, Institution i can buy shares or debt securities issued by Institution j in proportions $\pi_{i,j}$ and $\gamma_{i,j}$, respectively.

Liability side

The liability side is composed of equity (that is brought by investors) and nominal debt, whose book values are respectively denoted by K_i and L_i^* for Institution i. Since equity and debt securities will be traded on the the secondary market, it is necessary to introduce their market values, respectively denoted by $\mathcal{K}_i$ and $\mathcal{L}_i$.

In the case of banks, L_i^* includes different types of debts (deposits and bonds of various maturities) considered as homogeneous in terms of seniority.[4] Banks issue debt along a common yield curve. In other words, bank debt securities are considered risky (the interest rate curve is above the risk free yield curve) but have a common degree of risk (the same rating, say). Despite this common feature, Bank i chooses its own degree of maturity transformation $\omega_i \in [0, 1]$. Let us denote by T_{Li} the average of maturities of all

[4]For various seniority levels, see Gouriéroux et al. (2013).

types of debts and by T_{Ai} the maturity of the assets. Then, ω_i is defined as $\omega_i = 1 - \dfrac{T_{Li}}{T_{Ai}}$. For instance deposits can be seen as every day re-funded overnight loans by households to banks and therefore their maturity is equal to 0, yielding $\omega_i = 1$. On the opposite, a debt whose maturity equals the asset maturity corresponds to $\omega_i = 0$. Banks usually assume that their short-term debt will be rolled over. However, it is not always the case, especially during crises. If a bank is only funded by deposits ($\omega_i = 1$), it may happen that all depositors suddenly quit, causing a funding liquidity shock. The same can happen in the case of debt issued with bonds if investors decide not to roll over. In the extreme opposite case ($\omega_i = 0$), there is no possible liquidity shock (but there is no maturity transformation). Banking activity is precisely profitable due to maturity transformation since the interest rate corresponding to long term lending (asset side) is larger than the one corresponding to short term borrowing. In our model, the interest rate charged on the debt of Bank i is deterministic, depends on ω_i and is denoted by $r_D(\omega_i)$.

In the case of insurance companies, the nominal debt L_i^* mostly corresponds to technical provisions relative to the underwritten risks. Therefore, ω_i can no longer be interpreted as a degree of maturity transformation but as the mean severity of claims. Thus, we do not have necessarily $\omega_i \in [0, 1]$ anymore. Contrary to banks, the liability side of an insurer is stochastic. For instance, in line with standard ruin models (see e.g. Asmussen and Albrecher, 2010), ω_i could be the parameter of the Pareto distribution in a claims model. Of course, the collected premia directly reflect the risk profile of the insurance contracts.

The balance sheet of Bank i is represented at the initial date and the end date in Tables 5.1 and 5.2, respectively. The dates are represented by an upper-scripted index in parenthesis.

		Asset	Liability		
interbank cross-shareholdings	$\leftrightarrow$	$\pi_{i,1}\mathcal{K}_1^{(0)}$ $\vdots$ $\pi_{i,n}\mathcal{K}_n^{(0)}$	L_i^*	$\leftrightarrow$	debt
interbank lending	$\leftrightarrow$	$\gamma_{i,1}\mathcal{L}_1^{(0)}$ $\vdots$ $\gamma_{i,n}\mathcal{L}_n^{(0)}$	$K_i^{(0)}$	$\leftrightarrow$	value of the firm
external assets	$\leftrightarrow$	$Ax_i^{(0)}$			
cash	$\leftrightarrow$	$A\ell_i^{(0)}$			

Table 5.1: Balance sheet of Bank i at the initial date $t = 0$.

		Asset	Liability		
interbank cross-shareholdings	$\leftrightarrow$	$\pi_{i,1}K_1^{(1)}$ $\vdots$ $\pi_{i,n}K_n^{(1)}$	$L_i^{(1)}$	$\leftrightarrow$	debt
interbank lending	$\leftrightarrow$	$\gamma_{i,1}L_1^{(1)}$ $\vdots$ $\gamma_{i,n}L_n^{(1)}$	$K_i^{(1)}$	$\leftrightarrow$	value of the firm
external assets	$\leftrightarrow$	$Ax_i^{(1)}$			
cash	$\leftrightarrow$	$A\ell_i^{(1)}$			

Table 5.2: Balance sheet of Bank i at the end date $t = 1$.

It is important to note that the equity and the debt of the other institutions (on the asset side) must be priced at the market value at $t = 0$. At time $t = 1$, the book value can be considered.

In line with the Value-of-the-Firm model (Merton, 1974), the value of debt L_i and equity K_i at any date are linked through the following equilibrium equations

$$K_i = \max\Big(\sum_{j=1}^{n} \pi_{i,j}K_j + \sum_{j=1}^{n} \gamma_{i,j}L_j + A\ell_i + Ax_i - L_i^*, 0\Big), \text{ for } i = 1, \ldots, n, \text{ and} \quad (5.1)$$

$$L_i = \min\Big(\sum_{j=1}^{n} \pi_{i,j}K_j + \sum_{j=1}^{n} \gamma_{i,j}L_j + A\ell_i + Ax_i, L_i^*\Big) \text{ for } i = 1, \ldots, n, \quad (5.2)$$

These $2n$ equations define a liquidation equilibrium. Equation (5.1) corresponds to the simple accounting definition of equity as the net value of assets over debts. Equation (5.2) is very similar to (5.1) and directly follows from Merton's model: the debt value is the minimum between the asset value and the nominal debt.

Proposition 2 in Gouriéroux et al. (2012) states that these equations define a suitable liquidation equilibrium (see Proposition 5.5 in Appendix 5.8.1). The cornerstone of our approach will consist in optimizing the balance sheet items of the financial institutions (apart from the equity which is exogenous). Proposition 5.5 states that whatever the balance sheet composition of each institution (whatever the values of Ax_i, $A\ell_i$, π_{ij}, γ_{ij} and L_i^* satisfying Assumptions $(A1')$, $(A2')$ and $(A3')$ in Proposition 5.5), the network obtained can theoretically exist (under suitable unique values for K_i and L_i, $i = 1, \ldots, n$). In particular, our optimized network exists and thus the approach we develop in this paper can be carried out.

Note that although Gouriéroux et al. (2012) do not consider any maturity, Proposition 5.5 still holds true in the presence of ω_i. It is sufficient to replace L_i^* by $L_i^*(1 + r_D(\omega_i))$ in the proof.

5.2.2 Regulatory constraints

In line with the usual Basel regulation (see e.g. BCBS, 2011, Section I)[5], the solvency constraint for Institution i is written

$$K_i^{(0)} \geq k_i^A A x_i^{(0)} + k^\pi \sum_{j=1}^n \pi_{i,j} \mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^n \gamma_{i,j} \mathcal{L}_j^{(0)}, \tag{5.3}$$

where k_i^A and k^π are regulatory parameters (risk weights) for external assets and interfinancial shareholdings and debtholdings, satisfying $0 < k_i^A, k^\pi, k^\gamma < 1$. The parameter relative to the external assets is specific to each institution whereas those relative to interfinancial assets are common within a specific sector (banking or insurance business). This constraint means that the equity must be higher than the risk-weighted assets and aims at ensuring the existence of a sufficient capital buffer to avoid losses for creditors in most cases.

Note that in the case of insurance companies, (5.3) corresponds to the Solvency I regulatory framework (see CEC, 1979)[6], apart from the term corresponding to interconnections. Since Solvency II is not implemented so far, we choose not to consider it in our modeling. Moreover, let us emphasize that the weights of banks differ from those of insurance companies. In the case of an insurer, the constraints on k^π and k^γ can be relaxed to $0 \leq k_i^\pi, k_i^\gamma < 1$.

Even if we do not focus on liquidity shocks, we introduce a liquidity constraint:

$$A\ell_i^{(0)} \geq k^L \; l(\omega_i, \; L_i^*), \tag{5.4}$$

with l being some increasing function with respect to both variables which will be characterized further and k_L satisfying $0 < k^L < 1$. This constraint aims at ensuring a sufficient liquid assets buffer to face exposure to liquidity risk (maturity transformation in the case of banks and claims in the case of insurance companies) stylized by ω_i and L_i. Note that this constraint is similar to the Basel III Liquidity Coverage Ratio (see BCBS, 2013).

5.2.3 Summary of the optimization framework

In short, both banks and insurance companies select their balance sheet items under restrictions (class of assets for banks and class of risks for insurance companies) and regulatory constraints. Their business model is reflected through a size variable and an intensity variable: the size is the total credit granted for a bank and the total of individual risks covered for an insurance company, while the intensity is the degree of maturity transformation for a bank and the claims' severity for an insurance company.

We emphasize that our modeling allows to take the specificities of banks and insurance companies into account in a unified way. The same parameters allow interpretation in terms of banks as well as of insurance companies. However, as we mentioned, the nature of the debt L_i^* and that of the maturity ω_i are different when considering a bank or an insurance company. In the following, we will mainly focus on banks.

[5]BCBS means Basel Committee on Banking Supervision.

[6]CEC means the Council of the European Communities.

5.2.4 Network Benchmark

Our testing principle is to compare the network obtained through our modeling and a stylized network, so-called benchmark. In this part, we describe this stylized network along three dimensions. First, we provide the main aggregate items of a bank's balance sheet. Thus, we will be able to check if, apart from interbank assets, the obtained balance sheet composition is close to a real one. Second, we focus on the network shape. This level provides a qualitative assessment of interconnections. Last, the size of interconnections along instruments in a typical banking network is described. This last level provides a quantitative assessment of interconnections. We restrict the analysis to interbank networks in industrial countries, typically the United-States, Canada or Europe. We identify four stylized facts that characterize an interbank network.

Main aggregate items of a bank's balance sheet

We consider the Bank Holding Company Performance Report Peer Group Data, published by the Federal Financial Institutions Examination Council, that provides the structure of asset and liability sides for banks above $10 billion (from 69 banks in 12/2008 to 90 in 12/2012). Figure 5.2 provides the composition of the asset side and the leverage for these banks. Corresponding informations are summarized in the following stylized fact:

Stylized fact 1: For a typical bank, the external assets (Ax_i) represent about 95% of its total assets while its equity (K_i) represents about 5% of its total assets.

BHCPR PEER GROUP DATA — PERCENT COMPOSITION OF ASSETS AND LOAN MIX — PEER GROUP 01

PERCENT OF TOTAL ASSETS	12/31/2012	12/31/2011	12/31/2010	12/31/2009	12/31/2008
Real Estate Loans	32.33	32.56	34.18	37.26	40.98
Commercial and Industrial Loans	11.71	11.70	11.77	12.18	13.26
Loans to Individuals	4.94	6.00	5.68	5.34	4.89
Loans to Depository Institutions	0.10	0.14	0.07	0.05	0.08
Agricultural Loans	0.12	0.15	0.17	0.17	0.19
Other Loans and Leases	4.07	3.49	3.40	3.20	3.01
Net Loans and Leases	57.96	57.64	57.51	59.71	63.39
Debt Securities Over 1 Year	15.92	16.67	16.23	15.06	13.34
Mutual Funds and Equity Securities	0.12	0.12	0.18	0.19	0.17
Subtotal	74.97	75.86	75.08	75.78	77.57
Interest-Bearing Bank Balances	4.20	3.96	4.00	4.51	2.92
Federal Funds Sold & Reverse Repos	1.13	1.30	1.27	0.78	0.80
Debt Securities 1 Year or Less	3.03	3.07	3.03	3.02	3.10
Trading Assets	0.98	1.38	1.43	0.95	1.33
Total Earning Assets	87.84	88.19	87.62	87.54	87.74
Non-Int Cash and Due From Dep Inst	1.52	1.52	1.33	1.72	1.92
Other Real Estate Owned	0.22	0.32	0.37	0.32	0.22
All Other Assets	10.46	10.18	10.96	10.69	10.18
Total Liabilities/Equity	30.15	32.75	36.76	37.77	31.65

Comment: interbank assets are mostly concentrated in highlighted lines.

Figure 5.2: Excerpt of the Bank Holding Company Performance Report Peer Group Data between 12/2008 and 12/2012. Source: www.ffiec.gov.

Network shape

National interbank networks[7] are usually characterized by a core-periphery structure (Craig and Von Peter, 2014). The core is composed of large banks highly interconnected.

[7]See Furfine (2003) for USA, Wells (2002) for UK, Upper and Worms (2004) for Germany, Lublóy (2005) for Hungary, van Lelyveld and Liedorp (2006) for the Netherlands, Degryse and Nguyen (2007) for Belgium, Toivanen (2009) for Finland, Gauthier et al. (2012) for Canada, Mistrulli (2011) for Italy and Fourel et al. (2013) for France.

The periphery is composed of smaller banks which are connected to core banks only. Figure 5.3 represents a typical national interbank network. Note that at the international level, the core-periphery structure is much less clear among major banks (Alves et al., 2013). A complete structure seems more representative of the reality. These observations are summarized in the following two stylized facts:

Stylized fact 2: For a network composed of banks heterogeneous in size, a core-periphery structure is ideally expected. In other words, matrices $\mathbf{\Pi} = (\pi_{ij})_{i,j=1,\dots,n}$ and $\mathbf{\Gamma} = (\gamma_{ij})_{i,j=1,\dots,n}$ present a block structure with a majority of zeros.

Stylized fact 3: For a network composed of large banks homogeneous in size, a complete structure is ideally expected. In other words, $\mathbf{\Pi}$ and $\mathbf{\Gamma}$ have few zero coefficients.

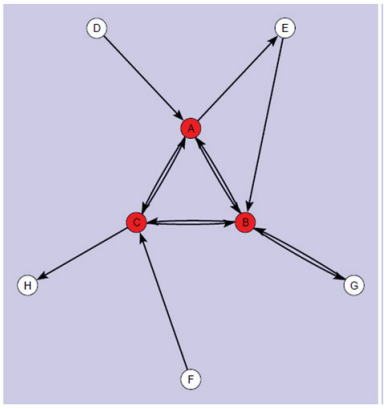


Figure 5.3: Core-Periphery structure. The core is composed of banks A to C while the periphery is composed of banks D to H. Source: Figure 1 in Craig and Von Peter (2014).

Interconnections size and support

As mentioned above, total interbank assets account for about 5% of total assets. However, data concerning the relative importance of the different instruments are scarce. At the European level (at the end of 2011), according to Table 1 in Alves et al. (2013), credit claims (direct credit from one bank to another) and debt securities represent 90% of exposures. The remainder is composed of "other assets". For the 6 largest Canadian banks (as at May 2008), there is a factor 4 between exposure through traditional lending and exposure through cross-shareholdings, as reported in Table 3 in Gauthier et al. (2012).

Stylized fact 4: In the case of large banks, lending exposures represent a major part of exposures (between 80% and 90%). In other words, $\mathbf{\Gamma L}^* \approx \alpha \times (\mathbf{\Pi K} + \mathbf{\Gamma L}^*)$, where $\mathbf{K} = (K_i)_{i=1,\dots,n}$, $\mathbf{L}^* = (L_i^*)_{i=1,\dots,n}$ and $\alpha \in [80\%, 90\%]$. However, cross-shareholdings can not be neglected.[8]

[8]It is paramount to take the relative weight of share securities into account since they are more risky than debt/lending: shareholders lose as soon as the financial institution has losses while a debt holder is only affected if the losses of the financial institution are above its equity. For contagion analysis, cross-shareholdings cannot be neglected.

5.3 Model, theoretical properties and network shape

We model the network formation in two steps. The first one -dealt with in this section- concerns the modeling of the behavior of one institution, the state of the others being given. The aim is to determine how a financial institution defines its balance sheet and especially the interconnections knowing the main balance sheet elements of the other ones. For instance, how does a new bank get interconnected to previously existing ones? Or how does a bank adapt its balance sheet to modifications of the structure of others? The second step concerns the whole network formation using the modeling of individual behaviors and will be considered in Section 5.4.

Based on the framework introduced in the previous section, a one-period model is built. Banks are risk-averse agents optimizing their balance sheet structure for the shareholder's interest at the initial date $t = 0$. The horizon is the final date $t = 1$.

The assumption that interconnections represent a long-term choice is a cornerstone of our analysis. Interconnections are not motivated by any liquidity features: they correspond to optimal choices in the long-run. Including liquidity-motivated interconnections that stem from daily work of Asset Liability managers, as well as the interactions between short-term and long-term interconnections, constitutes an ongoing work of ours.

A very important concern is the problem of reflexivity: how to technically manage the fact that the choices of financial institutions are interdependent? The main issue is that a complete Nash equilibrium modeling of the whole balance sheet structure -interconnections, external assets and debt- is clearly wishful thinking. It triggers difficulties, especially with respect to privately available information and anticipation formation. Note that in models with Nash equilibrium such as in Babus (2007) or Acemoglu et al. (2013), choices are only taken at the level of interconnections: all the other components of the balance sheet are exogenous. This scope is arguably adapted in a short-term framework but is clearly unsuitable from a long-term perspective. In order to circumvent a complete game theoretic model, we adopt some simplifying assumptions backed up by practical considerations.

5.3.1 Modeling strategy

We choose an efficient, albeit simple strategy: each financial institution is assuming that the asset side of the other financial institutions is only composed of their external assets. This implies that the institution optimizing its balance sheet is not taking into account the future reactions of the other financial institutions. In this perspective, the optimization program is not strategic. Apart from simplifying the resulting optimization program, this strategy corresponds to sound assumptions for each financial institution and this for several reasons.

Firstly, the information set used in the optimization program is very close to the genuinely available one. Actually, bilateral exposures are private information. Publicly available information for any major financial institution are the detailed income statement and the balance sheet. For instance, return-on-asset, return-on-equity, cash, total interbank assets, loans on the asset side, debt and equity on the liability side are easily extracted from the public financial communication of firms or published reports (see Appendix 5.7 for an excerpt of the Consolidated Financial Statements for Bank Holding Companies (BHCs) of Bank of America published by the Federal Financial Institutions

Examination Council[9]). Secondly, note that a large part of debt securities and shares are traded on the secondary market. Therefore, Bank i cannot know exactly who its creditors and shareholders are: Bank i knows its asset side but not the repartition of its liability side. The part of tradable shares is called the floating equity. By analogy, we call the floating debt the part of the debt traded on the secondary market.

Lastly, the absence of anticipation of reaction constitutes an approximation. As previously mentioned, there is no information on bilateral exposures. However, total interbank assets represent about 5% or 10% of total assets.[10] Each bilateral exposure should be much smaller: 0.5% of total assets seems a reasonable upper bound. Therefore, when a new bank gets interconnected, the new interconnections do not significantly modify its balance sheet. It may trigger a reaction from its own counterparts but the effects can be neglected by comparison to the risk borne in the external assets for instance. As we will see in the simulation results, the reaction of counterparts only has a light influence on each institution, leading to a rapid stabilization of the network. This provides an indication that this assumption of absence of anticipation can be accepted as a first step towards building more realistic models.

Then this assumption allows us to derive in the next subsection some strong and tractable theoretical results.

5.3.2 Optimization program

Bank i is managed for the benefits of its investors (i.e. shareholders) who are risk-averse and endowed with an initial capital $K_i^{(0)}$. The risk-aversion of the investors of Bank i is represented by a utility function u_i. We denote $1 - c_j^\pi$ (respectively $1 - c_j^\gamma$) the floating equity (respectively debt) of Bank j, for $j = 1, \dots, n$.

In line with our modeling strategy, we scale the total assets of Bank j by $\kappa_j = \dfrac{L_j^{(0)} + K_j^{(0)}}{Ax_j^{(0)} + A\ell_j^{(0)}}$. These scaling factors compensate for the fact that we consider that the counterparts are not interconnected. Thus, we get the following approximation for the equity of Bank i at time $t = 1$:

$$K_i^{(1)} = \max\Big[Ax_i^{(1)} + A\ell_i^{(1)} + \sum_{j=1}^{n} \pi_{i,j} \max\Big(\kappa_j(Ax_j^{(1)} + A\ell_j^{(1)}) - L_j^{*(1)}, 0\Big) + \sum_{j=1}^{n} \gamma_{i,j} \min\Big(\kappa_j(Ax_j^{(1)} + A\ell_j^{(1)}), L_j^{*(1)}\Big) - [1 + r_D(\omega_i)]L_i^{(0)}, 0\Big]. \tag{5.5}$$

[9] http://www.ffiec.gov/nicpubweb/content/help/HelpFinancialReport.htm

[10] For instance, on June 30, 2013, the proportion of interbank assets in the total assets is 3.4% for Bank of America, 13% for JPM, 8.40% for Citigroup 8.3% for Wells Fargo, according to the Consolidated Financial Statements for BHCs.

If we denote by $\mathbb{E}_0$ the expectation computed at time $t = 0$, the optimization program $\mathcal{P}_i$ of Bank i is

$$\mathcal{P}_i := \left\{ \begin{array}{rll} \max & \mathbb{E}_0\left[u_i\left(K_i^{(1)}\right)\right] & \\ Ax_i^{(0)}, A\ell_i^{(0)} & & \\ L_i^{(0)}, \omega_i & & \\ \pi_{i,1}, \ldots, \pi_{i,n} & & \\ \gamma_{i,1}, \ldots, \gamma_{i,n} & & \\ \text{such that (s.t.)} & Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} = K_i^{(0)} + L_i^{(0)} & (BC) \\ & K_i^{(0)} \geq k_i^A \; Ax_i^{(0)} + k^\pi \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} & (SC) \\ & A\ell_i^{(0)} \geq k^L \; l(\omega_i, \; L_i^{(0)}) & (LC) \\ & Ax_i^{(0)} \geq 0, A\ell_i^{(0)} \geq 0, L_i^{(0)} \geq 0 & \\ & \omega_i \in [0,1] & \\ & \forall j \in \{1, \ldots, n\}, 0 \leq \pi_{i,j} \leq 1 - c_j^\pi, 0 \leq \gamma_{i,j} \leq 1 - c_j^\gamma. & \end{array} \right.$$

The constraint (BC) ensures the balance sheet equilibrium at the initial date. Note that this constraint allows the network resulting from our formation process (see Section 5.4) to satisfy (5.1) for each institution. The inequalities (SC) and (LC) are respectively the regulatory solvency and liquidity constraints presented in Section 5.2.2. BC, SC and LC stand for Balance sheet Constraint, Solvency Constraint and Liquidity Constraint, respectively.

5.3.3 Solution analysis

We define the position P_i of Bank i as the difference between its total assets (denoted by A_i) and its nominal debt. Therefore, at time $t = 1$, $P_i^{(1)} = A_i^{(1)} - L_i^{*(1)}$. If this difference is positive, the position is simply the equity; if the difference is negative, the position is the loss for creditors (while the equity is equal to zero in this situation). P can be interpreted as the profit-and-loss.

The uniqueness of the solution usually requires the strict concavity of the objective function. The concavity of $u_i \circ K_i^{(1)}$ (where $\circ$ denotes the composition operator) is not a necessary condition since we could expect that the integration operation makes the expectation strictly concave even if $u_i \circ K_i^{(1)}$ is not strictly concave everywhere (see the Appendix for more details). Moreover, it would impose conditions on F_R. Thus, we look for conditions on $u_i \circ K_i^{(1)}$. Due to their limited liability, shareholders aim at maximizing the expected utility of the equity. The latter is defined as $K_i^{(1)} = \max(P_i^{(1)}, 0)$, making $u_i \circ K_i^{(1)}$ non-differentiable and introducing a level shape. An unfortunate consequence is that for standard utility functions u_i, $u_i \circ K_i^{(1)}$ is not strictly concave and not even concave. Then our strategy is to approximate the real equity by a function $v(P_i^{(1)})$ to obtain the concavity. From an economic perspective, it is satisfactory to consider a transformation of the equity, as we will see in the following. Therefore, we decompose the analysis of $\mathcal{P}_i$ into two steps. Firstly, we show that under mild assumptions there exists a solution (Theorem 5.1). Secondly, we transform the optimization program $\mathcal{P}_i$ into a close one ($\mathcal{P}_i'$) for which existence and uniqueness are ensured (Theorem 5.2).

Analysis of the exact optimization program

Contrary to usual optimization programs where the total wealth is exogenous, increasing wealth by issuing debt is allowed in $\mathcal{P}_i$. Therefore, intuitively, the main difficulty in showing the existence of a solution is to show that Bank i has no gain in issuing an infinite amount of debt. The argument is as follows. The equity is exogenously fixed. Therefore, (SC) implies that the total value of risky assets is bounded. Thus, starting from a specific amount of debt, the funding obtained by issuing more debt is necessarily invested in the risk free liquid asset. But since the interest rate charged on the debt is higher than the risk free rate, it is not profitable to issue debt to invest in liquid assets. In other words, banks are expected to invest in risky assets: granting credit is the core activity of banks.
All this goes to state the following proposition:

Theorem 5.1 (Existence of a solution to $\mathcal{P}_i$)**.** *If*

- $(A1)$ *the investors neglect interconnections among their counterparts;*
- $(A2)$ *the utility function u_i is continuous and strictly increasing;*
- $(A3)$ *the distribution function F_R is continuous. Moreover, the density f_R is strictly positive on $[a, +\infty)^n$, for some $a \in \mathbb{R}$;*
- $(A4)$ *the yield curve, r_D, is continuous and strictly higher than the risk free rate;*

then there exists a solution to $\mathcal{P}_i$.

Assumption $(A1)$ is both a technical assumption and a way to reflect the restricted information available for each agent. Assumptions $(A2)$, $(A3)$ and $(A4)$ are very common in the literature and not restrictive.

Analysis of the approximated optimization program

As stressed before, it appears impossible to establish the uniqueness for $\mathcal{P}_i$ except in particular cases of simple models for F_R. We therefore consider an optimization problem $\mathcal{P}'_i$ where the sole difference with $\mathcal{P}_i$ is that the objective function is the expected utility of a strictly increasing transformation (denoted by v) of the position of Bank i, $P_i^{(1)}$. Considering the position directly makes things easier. However, it means not taking into account the limited liability which has some important implications. Indeed, it plays the role of a protection against extreme events for the managers: they are impacted by regular shocks but not by extreme ones. Some phenomena cannot be explained by

macro-economic models ignoring limited liability. The optimization program $\mathcal{P}_i'$ is

$$\mathcal{P}_i' := \left\{ \begin{array}{rll} \max & \mathbb{E}_0\left\{u_i\left[v(P_i^{(1)})\right]\right\} & \\ Ax_i^{(0)}, A\ell_i^{(0)} & & \\ L_i^{(0)}, \omega_i & & \\ \pi_{i,1}, \ldots, \pi_{i,n} & & \\ \gamma_{i,1}, \ldots, \gamma_{i,n} & & \\ \text{s.t.} & Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} = K_i^{(0)} + L_i^{(0)} & (BC) \\ & K_i^{(0)} \geq k_i^A\ Ax_i^{(0)} + k^\pi \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} & (SC) \\ & A\ell_i^{(0)} \geq k^L\ l(\omega_i,\ L_i^{(0)}) & (LC) \\ & Ax_i^{(0)} \geq 0, A\ell_i^{(0)} \geq 0, L_i^{(0)} \geq 0 & \\ & \omega_i \in [0,1] & \\ & \forall j \in \{1, \ldots, n\}, 0 \leq \pi_{i,j} \leq 1 - c_j^\pi, 0 \leq \gamma_{i,j} \leq 1 - c_j^\gamma. & \end{array} \right.$$

With this specification, the level aspect of the limited liability is removed and the transformation v ensures flexibility. For instance, with $v = Id$, one considers the usual maximization of the expected utility of profits. Alternatively, v can be chosen to closely fit the design of the limited liability of shareholders while relaxing their complete indifference for loss magnitude. In the latter case, $\mathcal{P}_i'$ is very close to $\mathcal{P}_i$.

In short, the argument for the existence of a solution of $\mathcal{P}_i'$ is similar to the argument for the existence of a solution of $\mathcal{P}_i$. The uniqueness mainly stems from the strict concavity of the objective function we obtain by adjusting v. However, the strict convexity of the constraints is necessary, imposing restrictions on the functional form of (LC) (see the proof for details). The following theorem provides the result regarding uniqueness:

Theorem 5.2 (Existence and uniqueness of a solution to $\mathcal{P}_i'$)**.**
Under $(A1)$, $(A2)$, $(A3)$, $(A4)$ *and the extra assumptions:*

- $(A5)$ *the composition of the transformation function* v *and the utility function* u_i *is strictly concave:* $\forall P \in \mathbb{R}, (u_i\ o\ v)''(P) < 0$*;*
- $(A6)$ *the interest rate on debt is strictly concave:* $\forall \omega_i \in [0,1], r_D''(\omega_i) < 0$*;*
- $(A7)$ *the interest rate on debt satisfies* $\forall \omega_i \in [0,1], r_D'(\omega_i) \neq 0$*;*
- $(A8)$ *the function* l *in* (LC) *satisfies*

$$\frac{\partial^2 l}{\partial \omega_i^2} \geq 0 \quad \textit{and} \quad \frac{\partial^2 l}{\partial \omega_i^2}\frac{\partial^2 l}{\partial {L_i^{(0)}}^2} \geq \left(\frac{\partial^2 l}{\partial \omega_i \partial L_i^{(0)}}\right)^2;$$

there exists a unique solution to $\mathcal{P}_i'$ *in the following sense. If all control variables appearing on the asset side of Bank* i *are fixed apart from one variable, denoted by* $Ac_i^{(0)}$*, then there is uniqueness of the triplet* $(Ac_i^{(0)},\ L_i^{(0)},\ \omega_i)$*.*

Note that the result of Theorem 5.2 is equivalent to saying that the main balance sheet items are unique. Indeed, the value of total assets $A_i^{(0)}$, the degree of maturity

transformation ω_i and the debt $L_i^{(0)}$ are unique. Due to the high number of control variables on the asset side and the complexity of the problem, it seems impossible to prove the uniqueness of all control variables (see the Appendix for more details). The uniqueness for all control variables will be verified on simulations.

Approximation properties

As mentioned before, the transformation function v gives room for flexibility. Lemma 5.1 provides two specifications satisfying $(A5)$, corresponding respectively to the position and a very good approximation of the equity.

Lemma 5.1 (Some specifications of v and u_i)**.**

- ***i)*** *If* $\forall P \in \mathbb{R}, v(P) = P$*, then* $(A5)$ *reduces to* $u_i'' < 0$*.*
- ***ii)*** *If* $\forall P \in \mathbb{R}, v(P) = \log(\exp(P) + 1)$*, then* $(A5)$ *is satisfied for the utility function* $u_i = \log$*.*

The approximation corresponding to $v(P) = \log(\exp(P) + 1)$ is shown in Figure 5.4. As we can see, the approximation error is very low. In the perspective of maximizing the utility, this function is probably even more satisfactory than the real equity. Indeed the utility of the equity is equal to zero whatever the position if the position is negative. In reality, one may think that the bank's managers prefer a light insolvency situation to a large one, for example for the sake of reputation. It is be difficult to find funding to build a new project after letting an institution in a state of large insolvency. Our approximation function is strictly increasing and therefore takes this aspect into account. This is especially true for position values not too far away from the insolvency point.

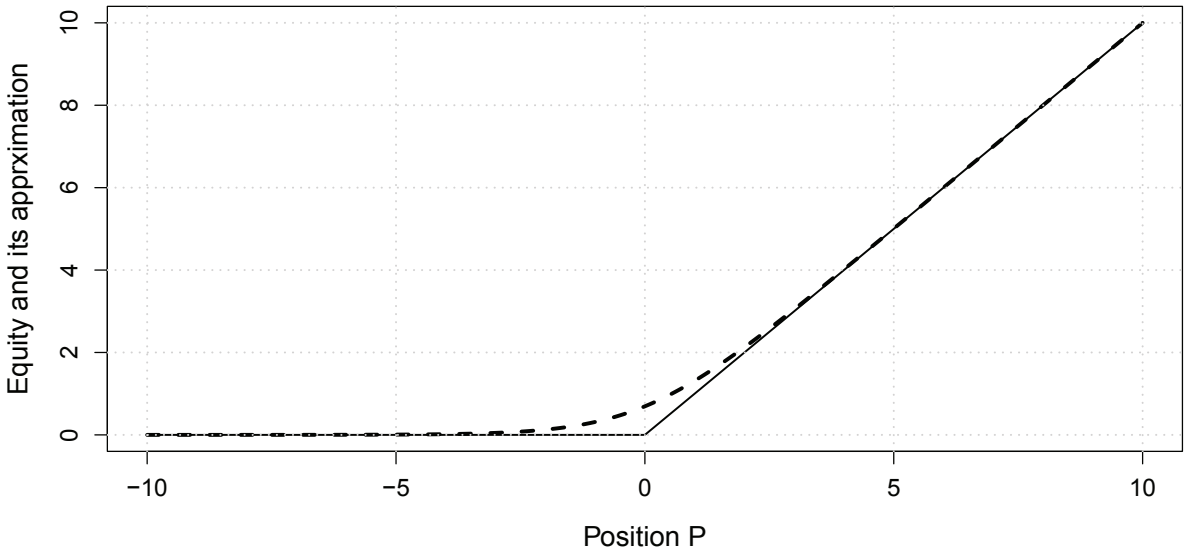


Figure 5.4: The solid line represents the equity and the dashed line displays the approximated equity using funtion $v(P) = \log(\exp(P) + 1)$.

Lemmas 5.2 and 5.3 provide a specification for the interest rate curve r_D and the function l appearing in (RC), respectively satisfying $(A6)$ and $(A8)$.

Lemma 5.2 (Specification of function r_D)**.**
An interest rate curve of the form

$$r_D(\omega) = \alpha - \beta \exp(\omega), \qquad \textit{for } \omega \in [0,1], \tag{5.6}$$

satisfies $(A6)$.

Lemma 5.3 (Specification of function l)**.**
The function defined by

$$l(\omega, L) = \exp(\omega)\exp(L), \qquad \textit{for } \omega \in [0,1] \textit{ and } L \in \mathbb{R}_+,$$

satisfies $(A8)$.

Choice

Previous theoretical results provide different suitable specifications (especially of the function v) leading to a unique solution of the optimization program. In order to clarify the presentation, let us make a clear recommendation of choice. The following result is directly derived from Theorem 5.2 and Lemmas 5.1, 5.3 and 5.2.

Corollary 5.1 (Existence and uniqueness to a solution of a specific optimization program)**.**
Additionally to $(A1) - (A4)$*, let us consider:*

- *a logarithmic utility function*

$$u_i(x) = \log(x), \qquad \textit{for } x \in \mathbb{R};$$

- *the following approximation of the limited liability of shareholders:*

$$v(P) = \log\left(\exp(P) + 1\right), \qquad \textit{for } P \in \mathbb{R};$$

- *the following liquidity constraint:*

$$l(\omega, L) = \exp(\omega)\exp(L), \qquad \textit{for } \omega \in [0,1] \textit{ and } L \in \mathbb{R}_+;$$

- *the following interest rate curve:*

$$r_D(\omega) = \alpha - \beta \exp(\omega), \qquad \textit{for } \omega \in [0,1].$$

Then, the associated optimization program $\mathcal{P}_i'$ has a unique solution.

To conclude this section, let us emphasize that all parameters and variables required to perform the optimization can be obtained via publicly available data.

5.3.4 Optimal interconnections

Previous theoretical results ensure that the bank's maximization program has a (unique) solution. However, we did not characterize this solution, in particular the interconnections. In this part, we show that under some conditions, it is optimal for a bank to get interconnected. In this section, in order to simplify the presentation and to explain the main features, we do not take into account the control variables $A\ell_i$ and ω_i, as well as the liquidity constraint (LC).

In order to start, let us consider a simplified case of a portfolio composed of a quantity Ax and a quantity π of assets having respectively random variables R_g and R_g^π as gross returns, under a solvency constraint.[11] The penalization weights are respectively k^A and k^π. The corresponding optimization program is

$$\mathcal{P}_{\mathcal{RA}} := \left\{ \begin{array}{rl} \max\limits_{Ax,\,\pi} & \mathbb{E}\left[u(AxR_g + \pi R_g^\pi)\right] \\ \text{s.t.} & k^A Ax + k^\pi \pi \leq 1 \\ & Ax \geq 0 \\ & 0 \leq \pi \leq 1 \end{array} \right. .$$

The Karush, Kuhn and Tucker (KKT) Theorem (Karush, 1939; Kuhn and Tucker, 1951) allows to derive the following proposition.

Proposition 5.1. *For the sake of simplicity, we denote* $f = \mathbb{E}\left[u(AxR_g + \pi R_g^\pi)\right]$. *Under the condition*

$$\forall Ax \in \mathbb{R}_+ \text{ and } \forall \pi \in [0,1], \frac{\dfrac{\partial f}{\partial Ax}(Ax,\pi)}{k^A} < \frac{\dfrac{\partial f}{\partial \pi}(Ax,\pi)}{k^\pi},$$

the optimal π^* *is different from 0.*

This shows that under the condition that the derivative of the expected utility with respect to π (relative to its corresponding weight) is higher than the one with respect to Ax, the optimal π^* is strictly positive. Proposition 5.1 does not provide the solution but gives an indication that interconnections can be strictly positive under some conditions. This result can be generalized to a higher number of assets. Note that this illustrative program does not contain any equality constraint. However, such a constraint can be trimmed by replacing one control variable in function of the others. That reduces the problem's dimension. This point will be further detailed in the following.

Due to the high complexity of our optimization problem (high dimension and high number of constraints), the KKT conditions are very numerous and therefore it seems impossible to derive the solution in a closed form. We decompose the analysis in different steps. We first consider a risk-neutral agent maximizing the value of its portfolio without limited liability. Secondly, we consider the case of a risk-averse agent and finally the limited liability is taken into account.

Risk-neutral agent without limited liability:
In the risk-neutral case, the utility function is the identity function. Therefore, we can

[11] For the sake of simplicity, the product $\pi\mathcal{K}$ of the complete program has been simplified into π.

consider the following optimization program:

$$\mathcal{P}_{\mathcal{RN}} := \left\{ \begin{array}{rl} \underset{Ax,\,\pi}{\max} & \left(Ax\mathbb{E}(R_g) + \pi\mathbb{E}(K^{(1)})\right) \\ \text{s.t.} & k^A Ax + k^\pi \pi\ \mathcal{K}^{(0)} \leq 1 \\ & Ax \geq 0 \\ & \pi \geq 0 \end{array} \right. ,$$

where $K^{(1)}$ is the equity value (book value) of another institution at time $t = 1$ and $\mathcal{K}^{(0)}$ is the equity value (market value) of this institution at time $t = 0$.

By using the same type of argument as in Proposition 5.1, it is easy to show that if $\mathbb{E}(R_g) > 0$ or $\dfrac{\mathbb{E}(K^{(1)})}{\mathcal{K}^{(0)}} > 0$, then

- if $\dfrac{\mathbb{E}(R_g)}{k^A} > \dfrac{\mathbb{E}(K^{(1)})}{k^\pi \mathcal{K}^{(0)}}$, the unique solution is $\left(Ax^* = \dfrac{1}{k^A}, \pi^* = 0\right)$;
- if $\dfrac{\mathbb{E}(R_g)}{k^A} < \dfrac{\mathbb{E}(K^{(1)})}{k^\pi \mathcal{K}^{(0)}}$, the unique solution is $\left(Ax^* = 0, \pi^* = \dfrac{1}{k^\pi \mathcal{K}^{(0)}}\right)$;
- if $\dfrac{\mathbb{E}(R_g)}{k^A} = \dfrac{\mathbb{E}(K^{(1)})}{k^\pi \mathcal{K}^{(0)}}$, the solution is not unique.

Therefore, due to the solvency constraint, a risk-neutral agent only invests in the asset having the highest return with respect to its specific regulatory weight in the solvency constraint.

Let us now consider the case where a limit to the availability is introduced: the constraint $\pi \geq 0$ is replaced by $0 \leq \pi \leq c$. In this case, if $\dfrac{\mathbb{E}(R_g)}{k^A} < \dfrac{\mathbb{E}(K^{(1)})}{k^\pi \mathcal{K}^{(0)}}$, $\pi^* = \min\left(c, \dfrac{1}{k^\pi \mathcal{K}^{(0)}}\right)$. Therefore, if $c < \dfrac{1}{k^\pi \mathcal{K}^{(0)}}$, investing all in $\mathcal{K}^{(0)}$ does not bind the solvency constraint. In this case (and if $\mathbb{E}(R_g) > 0$), an investment in Ax completes the portfolio. This result can be easily generalized to the case of n institutions and where it is possible to invest in the debt $L_j, j = 1, \ldots, n$, of the other institutions. This is done in the following theorem.

Theorem 5.3. *Let us consider the following optimization program:*

$$\mathcal{P}_{\mathcal{RNG}} = \left\{ \begin{array}{rl} \underset{Ax_i,\,\pi_{ij},\,\gamma_{ij}}{\max} & \left(Ax_i\mathbb{E}(R_{g,i}) + \sum_{j=1}^n \pi_{ij}\mathbb{E}(K_j^{(1)}) + \sum_{j=1}^n \gamma_{ij}\mathbb{E}(L_j^{(1)})\right) \\ s.t. & k^A Ax_i + k^\pi \sum_{j=1}^n \pi_{ij}\mathcal{K}_j^{(0)} + k^\gamma \sum_{j=1}^n \gamma_{ij}\mathcal{L}_j^{(0)} \leq 1 \\ & Ax \geq 0 \\ & 0 \leq \pi_{ij} \leq c^\pi \\ & 0 \leq \gamma_{ij} \leq c^\gamma \end{array} \right. .$$

To find this problem's solution, let us sort in decreasing order the following returns (relative to their penalty weight): $\dfrac{\mathbb{E}(R_{g,i})}{k^A}$, $\dfrac{\mathbb{E}(K_j^{(1)})}{k^\pi \mathcal{K}_j^{(0)}}$ $(j = 1, \ldots, n)$, $\dfrac{\mathbb{E}(L_j^{(1)})}{k^\gamma \mathcal{L}_j^{(0)}}$ $(j = 1, \ldots, n)$. *The optimal solution consists in investing as much as possible in the asset having the highest return with respect to its regulatory weight. When this asset is not available anymore, it is better to invest as much as possible in the second one, and so on. This is repeated until the solvency constrained is binding.*

Risk-averse agent without limited liability
A risk-averse agent aims at decreasing the variance of its portfolio. To this purpose, it is necessary to diversify. Therefore, in this case, we can expect an investment in many assets, contrary to the "binary" investment described previously. This is confirmed by numerical experiments.

Agent with limited liability
In the previous considerations, we did not take into account the limited liability as well as the fact that equity and debt have very different features. Therefore, we could not see the implications of the fact that the π_{ij} and the γ_{ij} are related to very different instruments. To pinpoint these implications, let us consider a stylized set-up with two banks. One can identify four situations in which Bank 1 (or 2) is either solvent or in default. Table 5.3 reports these 4 states. Let us focus on the impact of limited liability for Bank 1. We assume that Bank 1 builds interconnections with Bank 2 anyway (for example in order to reduce its variance) and we discuss the distribution among shares and debt securities.

	Bank 2 in default	Bank 2 solvent
Bank 1 in default	e_{11}	e_{12}
Bank 1 solvent	e_{21}	e_{22}

Table 5.3: Banks' states.

The expected utility of Bank 1 is written as follows

$$\mathbb{E}(U_1) = \mathbb{P}(e_{11})\ PO(e_{11}) + \mathbb{P}(e_{12})\ PO(e_{12}) + \mathbb{P}(e_{21})\ PO(e_{21}) + \mathbb{P}(e_{22})\ PO(e_{22}),$$

where $\mathbb{P}(e)$ is the probability of being in state e and $PO(e)$ the associated payoff for Bank 1. Due to limited liability, $PO(e_{11}) = PO(e_{12}) = 0$. Thus

$$\mathbb{E}(U_1) = \mathbb{P}(e_{21})\ PO(e_{21}) + \mathbb{P}(e_{22})\ PO(e_{22}).$$

In the state e_{21}, Bank 2 defaults, meaning that its equity is equal to zero. It is therefore more interesting to invest in its debt. In the state e_{22}, Bank 2 is solvent. Thus, if the equity of Bank 2 has a higher return than its debt with respect to their regulatory weights, Bank 1 prefers investing in the share securities of Bank 2, thus increasing the π_{12}. If the correlation ρ between the external assets of both banks is highly positive, both banks are likely to be solvent and to default simultaneously. That means that $P(e_{21})$ is very low, giving: $\mathbb{E}(U_1) \approx \mathbb{P}(e_{22})\ PO(e_{22})$. In this situation, Bank 1 prefers investing in share securities. On the contrary, if the correlation ρ between the external assets of both Banks is highly negative, Bank 2 is likely to default when Bank 1 is solvent. In this case $\mathbb{E}(U_1) \approx \mathbb{P}(e_{21})\ PO(e_{21})$ and Bank 1 prefers investing in debt securities.

It is important to understand that the asymmetry between the cases $\rho > 0$ and $\rho < 0$ is due to the limited liability feature. Indeed, let us assume that Bank 1 has no limited liability and thus is not indifferent to losses. If ρ is highly positive, $\mathbb{E}(U_1) \approx \mathbb{P}(e_{11})\ PO(e_{11}) + \mathbb{P}(e_{22})\ PO(e_{22})$. In state e_{11}, Bank 2 defaults and it is better to invest in its debt whereas in state e_{12}, it is better to invest in its shares. Therefore, it can be appropriate to invest in both instruments and thus the asymmetry disappears. The same happens for a highly negative ρ.

However, keep in mind that this set-up is too minimal to show all the implications of the limited liability.

5.3.5 Cost of funding

In the considerations of Section 5.3.4, we assumed that the agent owns a sufficient amount of wealth to invest until the solvency constraint is binding. However, the capital $K_i^{(0)}$ is very low compared to the total assets to invest (due to the regulatory weight values). Thus, once the total capital has been used, the institution must raise debt in order to continue to invest. Returns of shares and debt securities must be netted by the cost of funding. To make the investment attractive (in terms of net returns), the cost of raising debt should be lower than the returns of shares and debt securities.

Let us now state some results about the returns of investments in shares and debt securities issued by other institutions, compared to their funding cost. For the sake of simplicity of the interpretation, before stating the result for general functions u_i and v, we propose a result in the case where u_i and v are the identity functions. It corresponds to the case of a risk-neutral institution maximizing its position $P_i^{(1)}$.

Proposition 5.2 (Returns against opportunity cost, in the case of a risk-neutral institution maximizing the expectation of its position)**.**

- *The expected return of a share issued by Bank j is larger than the cost of funding of Bank i if and only if*

$$\int_{\frac{-b_j}{a_j}}^{+\infty} (a_j + b_j r_j) f_{R,j}(r_j)\ dr_j > [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)}, \tag{5.7}$$

  where $a_j = \kappa_j A x_j^{(0)}$, $b_j = \kappa_j \left(A x_j^{(0)} + A\ell_j^{(0)}(1 + r_f) \right) - L_j^*[1 + r_D(\omega_j)]$ and $f_{R,j}$ is the marginal density of the net return of the external asset of Bank j.

- *The expected return of the debt issued by Bank j is higher than the cost of funding of Bank i if and only if*

$$\int_{-\infty}^{\frac{L_j^*[1+r_D(\omega_j)]-b_j}{a_j}} (a_j r_j + b_j) f_{R,j}(r_j)\ dr_j + L_j^* c_j [1 + r_D(\omega_j)] > [1 + r_D(\omega_i)] \mathcal{L}_j^{(0)},$$

  where $c_j = \mathbb{P}\left(r_j > \frac{L_j^{*(1)} - b_j}{a_j} \right)$.

Proposition 5.3 (Returns against opportunity cost, in the general case of an institution maximizing the expectation of the utility of its equity)**.**

- *The expected return of a share issued by Bank j is larger than the cost of funding of Bank i if and only if*

$$\int_{-\frac{b_j}{a_j}}^{+\infty} (a_j r_j + b_j)\ w(r_j)\ dr_j > [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\ dr_j,$$

  where

$$w(r_j) = \int_{r_1=-\infty}^{+\infty} \dots \int_{r_{j-1}} \int_{r_{j+1}} \dots \int_{r_n} h_{i1}(r_1, \dots, r_j, \dots, r_n)\ f_R(r_1, \dots, r_n)\ dr_n \dots dr_{j+1}\ dr_{j-1} \dots dr_1,$$

where

$$h_{i1}(r_1, \ldots, r_j, \ldots, r_n) = \frac{\partial(u_i \circ v)}{\partial P_i^{(1)}}.$$

- *The expected return of the debt issued by Bank j is higher than the cost of funding of Bank i if and only if*

$$\int_{-\infty}^{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}} (a_j r_j + d_j) w(r_j)\ dr_j + L_j^*[1 + r_D(\omega_j)]) \int_{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}}^{+\infty} w(r_j)\ dr_j$$

$$> [1 + r_D(\omega_i)]\mathcal{L}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\ dr_j,$$

where $d_j = \kappa_j \left(A x_j^{(0)} + A \ell_j^{(0)} (1 + r_f) \right)$ *and* $w(r_j)$ *has been defined above.*

Equation (5.7) corresponds to the fact that $\mathbb{E}\left[K_j^{(1)}\right] - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)} > 0$. Note that in this formula, the return of only Bank j matters. It can be beneficial for Bank i to increase its participation in Bank j if the return on equity of Bank j is higher than the interest rate that Bank i must pay for its debt.

In the general case, the same type of inequality as (5.7) is obtained. However, it takes the marginal utility (up to function v) into account via $w(r_j)$. For interpretation purpose, let us assume that $v = Id$. For a given value of r_j, the algebraic gain of increasing the participation π_{ij} must be weighted by the marginal utility, which depends on the returns of all institutions. Integrating this marginal utility with respect to all returns $r_1, \ldots, r_n$ apart from r_j yields the term $w(r_j)$. The risk aversion of Bank i is embedded in the term $w(r_j)$.

The same type of argument applies in the case of the debt.

5.3.6 Testing of the diversification motive: the network shape

Let us now compare the consequences of Theorem 5.3 and Stylized Facts 2 and 3 on the network shape, and discuss the impact of risk-aversion and limited liability.

A risk-neutral bank with unlimited liability gets interconnected to others by strict mechanical behaviors: it seeks sequentially for the highest returns until binding the solvency constraint. Consequently, the network shape is very structured and directive since everyone gets interconnected in the same direction. Thus, in such a case, there is no general shape.[12] In other words, with risk-neutral banks and unlimited liability, the diversification motive cannot provide interesting results.

In the case of risk-averse banks, the interconnections tend to shape a complete network. Institutions carry out a diversification to decrease the variance, in addition to their aim of obtaining higher returns. Note that a diversified portfolio has a lower variance than a concentrated one.[13] Therefore, even if all institutions have similar returns, it can be beneficial to get interconnected. To significantly benefit from the diversification, the variance reduction must be high enough: situations where the specific assets are not

[12] Nevertheless, with a particular set of returns, a star network can occur.

[13] If X and Y are two random variables with mean μ, variance σ^2 and correlation $\rho < 1$, then $\mathbb{E}(X + Y) = 2\mu = \mathbb{E}(2X)$ whereas $\mathrm{Var}(X + Y) = 2(1 + \rho)\sigma^2 < 4\sigma^2 = \mathrm{Var}(2X)$.

almost non-risky and/or where the correlation is negative are prone to yield a complete network structure. These findings will be confirmed numerically in the next section. The limited liability feature can modify the balance between shares and debt securities.

When considering risk-averse banks, the diversification motive generates complete financial networks, such as those usually observed among major institutions. Therefore, we cannot rule out diversification as explaining interconnections between key financial players.[14]

5.4 Network formation and simulation results

In this section, we derive simulation results in order to assess the relevance of the diversification motive for the financial network formation. First, we present the specification that we use and our calibration strategy. Second, we develop a network formation process taking advantage of the strong and tractable theoretical results obtained in the previous section. Then, optimal choices for one financial institution and regarding the whole network are analyzed.

5.4.1 Specifications

For the sake of simplicity, two banks are considered, i.e. $n = 2$. Each institution is endowed with a capital amount of 1, i.e. $K_i^{(0)} = 1, i = 1, 2$. Both institutions have $x \mapsto \log(x)$ as utility function. An initial capital of 1 implies that the equity value $K_i^{(1)}, i = 1, 2$, at the optimization horizon is about 1. Therefore, the objective function is close to be linear over the most likely area, meaning that the banks are only slightly risk-averse.

In order to properly understand the main features of our model, we exclude $A\ell$ and ω from the control variables. The interest rates paid by the two financial institutions, denoted by $r_{D,1}$ and $r_{D,2}$ are therefore fixed. Moreover, the risk-free interest rate is set to zero: $r_{rf} = 0$.

Finally, note that the expectations are computed using Monte-Carlo techniques; 100 000 simulations ensure a good precision.

5.4.2 Calibration strategy

The gross returns on external assets follow a bivariate log-normal distribution:

$$\begin{pmatrix} \log\left(\dfrac{Ax_1^{(1)}}{Ax_1^{(0)}}\right) \\ \log\left(\dfrac{Ax_2^{(1)}}{Ax_2^{(0)}}\right) \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]. \tag{5.8}$$

In order to calibrate the mean parameter, we consider the income statement in the Consolidated Financial Statements for BHCs (reporting form FR Y-9C) for banks over \$10 billion. Between 12/31/2010 and 12/31/2012, the (annual) net income varies from 0.51% to 0.71% of the total assets. We round this value, considering that on average the net

[14] Note that our approach has no clue on the relevance of the other motives mentioned in Introduction. We simply show that diversification provides consistent results with empirical observations.

income of our banks is equal to 1%. Over the same period, the interest expenses represent between 0.74% and 1.07% of the total assets.[15] We basically consider that the cost of debt ($r_{D,1}$ and $r_{D,2}$) varies between 0% and 1%. Finally, the expected return of the external assets for Bank i is equal to $1\% + r_{D,i}\ell_i$, where ℓ_i is the Bank i's ratio of debt over total assets. For the variance parameter, a probability of default of 0.1% is in line with the current rating of major banks. We combine the informations relative to the net income and the probability of default to compute the parameters μ_i and $\sigma_i, i = 1, 2$ (see Appendix 5.11 for details). The parameter ρ lies between -0.9 to 0.9. A negative ρ can be interpreted as a sign of competition between the two banks or as the fact that banks operate in different markets (or geographical areas). Meanwhile, a positive ρ could be interpreted as an underlying common factor affecting both banks.

We consider the Basel 2 regulation. This regulation does not provide a unique set of values for the risk weights k_i^A, k^π and k^γ. If the external assets correspond to a retail activity (i.e. loans to households), loans to unrated firms (i.e. small firms) or quoted shares, the required capital is equal to 6%, 8% or 23.2% of the total exposure, respectively. For debt securities issued by banks, the required capital is equal to 1.6% (when AAA or AA rated) or 4% (when A rated). Lastly, as discussed in Repullo and Suarez (2013), there is a factor between the regulatory capital and the (accounting) equity, that varies from 1 to 2. For the sake of simplicity, we consider that the regulatory capital is either equal to the equity or to a half of the equity. Bottom line, we have 8 possible sets of risk weights.

5.4.3 Discussion about the pricing of shares and debt securities

Recall that the position of Bank 1 at time $t = 1$ is as follows if Bank 2 is solvent:

$$\begin{aligned}
P_1^{(1)} &= Ax_1^{(0)}(1+r_1) + \pi_{12}[\kappa_2 Ax_2^{(0)}(1+r_2) - L_2^*(1+r_{D,2})] + \gamma_{12}L_2^*(1+r_{D,2}) \\
&\quad - \left(Ax_1^{(0)} + \pi_{12}\mathcal{K}_2^{(0)} + \gamma_{12}\mathcal{L}_2^{(0)} - K_2^{(0)}\right)(1+r_{D,1}) \\
&= Ax_1^{(0)}(r_1 - r_{D,1}) + \pi_{12}\left[\kappa_2 Ax_2^{(0)}(1+r_2) - L_2^*(1+r_{D,2}) - \mathcal{K}_2^{(0)}(1+r_{D,1})\right] \\
&\quad + \gamma_{12}\left[L_2^*(1+r_{D,2}) - \mathcal{L}_2^{(0)}(1+r_{D,1})\right] + K_2^{(0)}(1+r_{D,1}). \qquad (5.9)
\end{aligned}$$

The terms $\mathcal{K}_2^{(0)}$ and $\mathcal{L}_2^{(0)}$ are respectively the market values of the share securities and debt securities issued by Bank 2 at time $t = 0$. In a complete market and with the usual assumptions, the price of an asset would be the discounted expected payoff under the risk-neutral probability:

$$\mathcal{K}_2^{(0)} = \frac{\mathbb{E}_{RN}[K_2^{(1)}|\mathcal{F}_0]}{1+r_{rf}},$$

where $\mathcal{F}_0$ denotes the available information at time $t = 0$. Since $\forall t \in [0,1], K_2^{(t)} = \max\left[\kappa_2 Ax_2^{(t)} - L_2^*(1+r_{D,2}), 0\right]$, K_2 appears as a call option whose underlying is Ax_2 and whose strike is $L_2^*(1+r_{D,2})$. However, since Ax is the price of an illiquid asset, it is difficult to argue that there exists a unique probability (the risk-neutral probability) that makes Ax a martingale. Therefore, we choose to consider that the price is the discounted

[15] In this paper, we consider that the total assets are equal to the earning assets and to the average assets.

expected payoff under the physical probability. The corresponding prices $\mathcal{K}_2^{(0)}$ and $\mathcal{L}_2^{(0)}$ are given in the following proposition.

Proposition 5.4. *If we assume that* $\log\left(Ax_i^{(1)}/Ax_i^{(0)}\right) \sim \mathcal{N}(\mu_i, \sigma_i^2)$*, then the expected equity and debt values of Bank* i *are*

$$\begin{aligned}
\mathbb{E}_0\left(K_i^{(1)}\right) &= \kappa_i A x_i^{(0)} e^{\mu_i + \frac{1}{2}\sigma_i^2}\left[1 - \Phi(\tilde{u} - \sigma_i)\right] - L_i^{*(1)}\left[1 - \Phi(\tilde{u})\right], \\
\mathbb{E}_0\left(L_i^{(1)}\right) &= \kappa_i A x_i^{(0)} e^{\mu_i + \frac{1}{2}\sigma_i^2}\Phi\left(\tilde{u} - \sigma_i\right) + \left[1 - \Phi(\tilde{u})\right],
\end{aligned}$$

where $\tilde{u} = \dfrac{1}{\sigma_i}\left(\log\left(\dfrac{L_i^{*(1)}}{\kappa_i A x_i^{(0)}}\right) - \mu_i\right)$, $L_i^{*(1)} = L_i^{*(0)}(1 + r_{D,i})$ *and* Φ *is the distribution function of the standard Gaussian variable.*

In order to understand some implications of our pricing choice, consider a situation where all returns are deterministic and $r_2 > r_{D,2}$. In such a framework, we have

$$\mathcal{K}_2^{(0)} = \frac{\kappa_2 A x_2(1 + r_2) - L_2^*(1 + r_{D,2})}{1 + r_{rf}} \quad \text{and} \quad \mathcal{L}_2^{(0)} = \frac{L_2^*(1 + r_{D,2})}{1 + r_{rf}}.$$

Therefore, injecting these prices in (5.9), we obtain

$$\begin{aligned}
P_1^{(1)} = {} & Ax_1(r_1 - r_{D,1}) + \pi_{12}\left[\left(\kappa_2 A x_2(1 + r_2) - L_2^*(1 + r_{D,2})\right)\left(1 - \frac{1 + r_{D,1}}{1 + r_{rf}}\right)\right] \\
& + \gamma_{12}\left[L_2^*(1 + r_{D,2})\left(1 - \frac{1 + r_{D,1}}{1 + r_{rf}}\right)\right] + K_2^{(0)}(1 + r_{D,1}). \qquad (5.10)
\end{aligned}$$

Generally, we have $r_{D,1} > r_{rf}$, meaning that the factors of π_{12} and γ_{12} are negative and thus that the net yields on shares and debt securities are negative. Therefore, for a risk-neutral agent (i.e. not interested in variance reduction), it would not be optimal to invest in shares and debt securities. That stems partly from the fact that we have priced these instruments using the physical probability. Under the latter probability, the shares and debt securities yield in average the risk-free rate. This feature could of course be challenged. Note that we should pay attention to the interpretations based on (5.10) since (5.10) only gives the expression of the position in a very simplified case. Equation (5.10) must only be considered as an indication.

Contrary to the share and debt security prices, the initial value of Ax_1 does not take the future returns into account. As we already mentioned, Ax_1 is an illiquid asset that cannot be exchanged on the market. Therefore, the assumption of absence of arbitrage is not necessarily satisfied and we price Ax_1 using its book value. Since generally $r_1 > r_{D,1}$, the specific asset Ax_1 provides a positive return. This is logical since getting positive returns via maturity transformation constitutes the core business of banks. However, in the pricing of $\mathcal{K}_2^{(0)}$, we consider the future returns of Ax_2. This asymmetry can be discussed but it is difficult to find an ideal solution given the close link between a market asset ($\mathcal{K}_2^{(0)}$) and an illiquid asset (Ax_2) in our model.

5.4.4 Methodology for the network formation

The optimization programs $\mathcal{P}_i$ and $\mathcal{P}_i'$ presented in Section 5.3 allow computing the balance sheet of an institution, knowing the state of the others. Here the aim is to build a complete network using this individual optimization program. To this purpose, we operate in a sequential way until an equilibrium in the network is reached.

We propose to use an iterative game. At each step, one institution optimizes its balance sheet taking into account the state of the network obtained at the previous step. Thanks to Corollary 5.1, there exists only one network at each step. The procedure is as follows[16]:

1. Bank 1 optimizes its balance sheet on Ax_1 and L_1^*. Quantities $\pi_{1,2}$ and $\gamma_{1,2}$ are forced to be equal to zero since at the initialization step, Bank 2's balance sheet is totally unknown;

2. Bank 2 optimizes its balance sheet on Ax_2, L_2^*, $\pi_{2,1}$ and $\gamma_{2,1}$ given Bank 1's balance sheet from step 1;

3. Bank 1 optimizes its balance sheet on Ax_1 ,L_1^*, $\pi_{1,2}$ and $\gamma_{1,2}$ given Bank 2's balance sheet from step 2. $\pi_{1,2}$ and $\gamma_{1,2}$ are optimized for the first time;

4. Bank 2 optimizes its balance sheet on Ax_2, L_2^*, $\pi_{2,1}$ and $\gamma_{2,1}$ given Bank 1's balance sheet from step 3;

5. Bank 1 optimizes its balance sheet on Ax_1 ,L_1^*, $\pi_{1,2}$ and $\gamma_{1,2}$ given Bank 2's balance sheet from the previous step;

6. and so on.

For further details, see Appendix 5.10.

Theoretically, this procedure may be endless. However, in less than 10 steps, the variations of the control variables from one step to the next are lower than 1% and we consider that the final situation constitutes an equilibrium. Moreover, if we accept the numerical argument for the existence of the limit-network, we can affirm its uniqueness. Indeed, if at each step the network is unique, then its final state is necessarily unique. It is interesting to note that this method is inspired by the classical methodology used to determine a Nash equilibrium (in the sense that no institution has any interest in deviating from its current state). However, further investigations would be required to know if the network obtained by our method effectively corresponds to a Nash equilibrium.

Last but not least, it is important to check that the obtained network is consistent in the sense that it satisfies (5.1) and (5.2). Firstly, at time $t = 0$, all banks considered in the network are solvent; otherwise they would disappear from the network. That means that the initial debt equals the contractual one: $L_i^{(0)} = L_i^*, i = 1, \ldots, n$. Therefore, (5.2) is automatically satisfied for each institution. Moreover, at each step, being a constraint of the optimization program, (5.1) is satisfied for the bank optimizing its balance sheet. If preliminary, this step has impacts on the other banks' balance sheets and (5.1) is not exactly satisfied anymore for them. Nevertheless, after some iterations, the network does not evolve from one step to the next (due to the convergence), implying that (5.1) is

[16]Note that this formation process can be applied in the general framework of Section 5.3 but is here presented using the previously mentioned specification.

satisfied for all institutions. These two points show that the obtained network is actually consistent.

This sequential algorithm could appear a little artificial but it is actually close to what happens in reality. An example of a real formation process of a network is as follows:

1. Consider an initial situation where there is no bank;
2. A first bank, denoted by B_1, is created during year $t = 0$. Since there are no other banks, there are no possible interconnections. Thus, B_1 optimizes Ax_1 and L_1. On January 1st of year $t = 1$, B_1 publishes its balance sheet;
3. Imagine that on January 3rd, a second bank B_2 is created. B_2 knows Ax_1 and L_1 and then can solve the optimization program to determine Ax_2, L_2, $\pi_{2,1}$ and γ_{21}. Once proportions $\pi_{2,1}$ and $\gamma_{2,1}$ have been determined, B_2 can buy on the secondary market shares and bonds issued by B_1 in these proportions;
4. On June 1st, B_1 and B_2 publish their balance sheets (apart from interconnections). Since the balance sheet of B_1 did not evolve since January 1st, B_2 has no new optimization to carry out. On the other hand, B_1 discovers for the first time informations relative to B_2: Ax_2 and L_2. Then B_1 optimizes its balance sheet and thus obtains Ax_1, L_1, π_{12} and γ_{12}. B_1 can buy on the secondary market shares and bonds issued by B_2;
5. On January 1st of year $t = 2$, balance sheets of B_1 and B_2 are published. The balance sheet of B_2 did not change and thus B_1 has no optimization to do. On the other hand, B_2 must adapt to the new balance sheet of B_1;
6. and so on.

After such iterations, one may think that there is convergence to an equilibrium in the network. Balance sheets of B_1 and B_2 do not evolve a lot from one step to the next.

5.4.5 Simulation results about the optimal choice for one institution

Let us here focus on the second step of the iterative game where Bank 2 optimizes its whole balance sheet (knowing the choice of Bank 1 at step 1). We assume that Bank 1's external assets are equal to 10. We present the sensitivity of the optimal choices of external assets Ax_2, nominal debt L_2^* and interconnections $\pi_{2,1}$ and $\gamma_{2,1}$, with respect to the regulatory parameters and correlation ρ. Our computations were carried out under various debt-issuing conditions (not costly with $r_{D,1} = r_{D,2} = r_{rf} = 0$, both costly with $r_{D,1} = r_{D,2} = 1\% > r_{rf} = 0$ and only one costly with $r_{D,1} = 1\% > r_{D,2} = r_{rf} = 0$) and we observe that the results are independent of these conditions. In each set-up, we consider the 8 sets of risk-weights and we let the correlation parameter vary between -0.9 and $+0.9$.

The corresponding results are summarized in Table 5.4. First, we observe that interconnections based on debt securities are never used. A direct consequence is that the risk weight on debt, k_γ, has no impact on the balance sheet and thus does not appear in Table 5.4. Second, interconnections based on share securities are used only when the correlation is lower than -0.3 (independently of the interest rates) and when the associated

risk weight is equal to 23.2%. They linearly decrease from about 45% to 0% between $\rho = -0.9$ and $\rho = -0.3$. Third, the solvency constraint is binding. The optimal external assets represent about $1/k^A$. The last row-block displays the ratio of interbank assets over the total assets: when interconnections are present, their proportion in the total assets is in line with the stylized facts.

These results could be interpreted as follows. First, the bank plays its core business: it invests as much as it can in its external assets. Then, if the regulation is not too strict and if the competitor's results are sufficiently anti-correlated, the bank opts for diversification: it slightly lowers its external assets to buy share securities issued by the competitor. Debt securities are not used since their net returns are negative (as a consequence of the pricing specification described in Section 5.4.3) and "nearly" deterministic (due to the low probability of default).

	k^π	k^A	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$
Ax	23.2%	6%	14	15	16
	23.2%	8%	11	12	11
	46.4%	12%	8	8	8
	46.4%	16%	6	6	6
π	23.2%	6%	45	25	0
(%)	23.2%	8%	45	25	0
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0
γ	23.2%	6%	0	0	0
(%)	23.2%	8%	0	0	0
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0
IBA/TA	23.2%	6%	3.1	1.6	0
(%)	23.2%	8%	3.9	2.0	0
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0

Table 5.4: Stylized results for the optimal choice of one institution, when $r_{D,1} = r_{D,2} = 0$.

5.4.6 Iterative game results

The iterative game reaches an equilibrium in less than 5 steps. The features pictured in the analysis of the behavior of one institution are still present. Especially, results are robust to the debt-issuing conditions.

Both institutions have the same balance sheet, whose composition is given in Table 5.5. Results are very similar to those for one institution only (Table 5.4). In particular, the proportion of interbank assets in the total assets is in agreement with the sylized facts. Note that for $\rho = -0.9$ and $\rho = -0.6$, the values of γ_{12} and γ_{21} are close to 10^{-4}. However, we have reported 0 since such low values do not have any economic meaning.

Let us state that these results have been obtained using $\kappa_i = 1, i = 1, 2$, in order to avoid numerical instability. Indeed, if the values of κ_i become too large, it makes no sense anymore to assume that the asset side of the other banks is only composed of their external assets.

	k^π	k^A	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$
Ax	23.2%	6%	15	15	16
	23.2%	8%	11	12	12
	46.4%	12%	8	8	8
	46.4%	16%	6	6	6
π	23.2%	6%	70	45	16
(%)	23.2%	8%	60	35	6
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0
γ	23.2%	6%	0	0	0
(%)	23.2%	8%	0	0	0
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0
IBA/TA	23.2%	6%	3.2	2.3	1
(%)	23.2%	8%	3.6	2.4	0.5
	46.4%	12%	0	0	0
	46.4%	16%	0	0	0

Table 5.5: Stylized results for the iterative game, when $r_{D,1} = r_{D,2} = 0$.

5.4.7 Testing the diversification motive

Regarding the capacity of the diversification motive to account for interconnections, the previous results provide a quantitative assessment completing the qualitative arguments developed in Section 3. The key result is that when returns on specific assets are anti-correlated, diversification leads to interconnections with reasonable size in terms of proportion of the total assets. However, debt securities are never used, meaning that interconnections are only supported by share securities. This portfolio composition contrasts with empirical findings.

Nevertheless, it is important to emphasize that in our simulation study, the choice of pricing shares and debt securities under the physical probability has large impacts. As explained in Section 5.4.3, it implies that the net yields of shares and bonds are negative. Therefore, in this framework, interconnections only allow for variance reduction but not for gain opportunity. We can expect this feature to be modified if the pricing is done under the risk-neutral probability. Interconnections in both shares and debt securities could then be observed, even for values of ρ larger than -0.3. The study of the risk-neutral specification constitutes an ongoing work. In some sense, these two types of specification for the pricing allow disentangling the two aims of the diversification: variance reduction and opportunity.

The latter discussion shows that our model seems promising but that results are very sensitive to the different possible specifications. Moreover, one feature that is not included in our model for the sake of simplicity may partly explain the discrepancy regarding debt securities. In reality, there are additional constraints -apart from the required capital- imposed to large shareholders, such as mandatory public communication. These constraints could discourage banks to invest in shares and could instead lead to higher investments in debt securities.

5.5 Application: impact of interconnectedness regulation

The diversification motive has proven an interesting explanation of the bank size (Stylized Fact 1), the network shape (Stylized Facts 2 and 3) and the composition of interconnections (Stylized Fact 4). Previous results concern the initial network resulting from banks' choices based on their expectations. Due to the endogenous feature of interconnections, we can build some plausible counterfactual scenarios, allowing to analyze the impact of regulation on the welfare at time $t = 1$.

5.5.1 Assessing interconnections

The interconnectedness across financial institutions has become a key concern of supervisors and regulatory authorities. Currently, long-term interbank exposures are covered by two main requirements. The first one concerns the solvency required capital for the interconnections, as for any other assets. It imposes a constraint on the total interbank exposure. The second one concerns "large" single exposures and imposes the risk-weighted exposure to be lower than a fraction of the equity.[17] Currently, the Basel Committee considers that an exposure is large if above 5% (instead of 10%) of the equity and to impose that the risk-weighted exposure ($k^{\pi}\pi_{ij}K_j + k^{\gamma}\gamma_{ij}L_j$ for the exposition to Bank j) has to be lower than 25% of the equity (see BCBS, 2014, Section II and Section IV.B). These requirements are valid for any type of exposure (e.g. corporate or sovereign) but the weights can vary with respect to the type. Moreover, the Basel Committee proposes to introduce tighter rules about interbank exposures for the G-SIBs (Global Systematically Important Banks). An upper bound between 10% and 15% instead of 25 % is in discussion (see BCBS, 2014, Section V). These tighter rules about interbank exposures aim at reducing the risk of contagion.

These different aspects show that interconnectedness is generally assessed in a negative way. Actually, supervisors are primarily concerned with excessive risks and therefore either analyze the effects of interconnections under depressed scenarios (stress-test approach) or build indicators in order to monitor the current fragility of the financial sectors. In both approaches, interconnectedness usually means contagion only. For instance, the seminal papers about network stress-tests -such as Furfine (2003) on US data or Upper and Worms (2004) on German data- sequentially consider the effects in their national banking sectors of the default of each bank. From their point of view, interconnected banks are likely to trigger defaults or to go bankrupt due to contagion.

Nevertheless, these analyses are not built on counterfactuals. They certainly give informative insights about what could happen within the current network in the case of defaults of some institutions or difficult macroeconomic conditions. However, since the network reaction is not taken into account, such studies do not really provide any clue on the way to obtain a more resilient network structure. Moreover, note that the question of regulation impact has hardly been addressed quantitatively, even in the case of a crystallized network.

The endogenous nature of interconnections in our model precisely allows us to study how the network reacts to tough macroeconomic conditions or to assess the impact of regulation on interbank exposures, for instance of the regulation in discussion at the

[17] We do not distinguish equity, own funds and regulatory capital.

Basel Committee. In the following, we focus on the impact of regulatory changes. To do so, we consider our 8 sets of regulatory weights associated to interbank exposure (k^A, k^π and k^γ).[18] For each specific set, the initial network is derived using our formation process. This step accounts for the diversification motive. Then we simulate returns of the external assets and examine the network at time $t = 1$. Let us emphasize that the shocks are properly propagated through the real interconnections.[19] The unique set of values K_i and L_i (see Proposition 5.5) is determined using the algorithm described in Appendix 5.12. This allows us to carry out a fair assessment of contagion. To do so, we build a welfare indicator including an explicit concern for the real economy and examine its sensitivity to the regulatory set of weights.

5.5.2 Welfare analysis

We adapt the welfare analysis by Repullo and Suarez (2013) to assess the impact of the regulatory parameters on the real economy.

The contribution of one bank is either negative or positive. When a bank defaults, its contribution is negative and proportional to the loss on its debt. This feature encompasses the cost of deposit insurance. When a bank is solvent, its contribution is the volume of external assets, i.e. the lendings provided to the real economy. This component captures the capacity to finance the real economy. The contribution of Bank i is written

$$w_i = -c \left(L_i^{*(1)} - L_i^{(1)} \right) + Ax_i^{(1)},$$

where c is the social cost for deposit insurance (in Repullo and Suarez (2013), c varies in $[0, 60\%]$).

Our welfare indicator W is the ratio of the contribution of all banks over the initial lending to the real economy:

$$W = \frac{w_1 + w_2}{Ax_1^{(0)} + Ax_2^{(0)}} = \frac{Ax_1^{(1)} + Ax_2^{(1)} - c\left(L_1^{*(1)} - L_1^{(1)} + L_2^{*(1)} - L_2^{(1)} \right)}{Ax_1^{(0)} + Ax_2^{(0)}}.$$

For $c = 0$, the welfare is given in Table 5.6. When there are interconnections, the welfare is higher than 1, indicating an increase of the banking capacity to lend to the real economy. In contrast, when there is no interconnection, the value of the external assets decreases. A complete analysis of the impact of interconnections would require further studies. However, these results suggest that the interconnections stemming from diversification are beneficial for the real economy.

[18] In reality only 4 since with the specification chosen, k^γ has no impact.

[19] Contrary to the assumption -used in the individual optimization program- that banks do not consider interconnections of their counterparts.

	k^π	k^A	$\rho = -0.9$	$\rho = -0.6$	$\rho = -0.3$
Sum of contributions	23.2%	6%	29.9	30.9	32.4
	23.2%	8%	22.8	23.6	24.8
	46.4%	12%	15.6	15.6	15.6
	46.4%	16%	11.9	11.9	11.9
Welfare (%)	23.2%	6%	101.0	101.0	101.0
	23.2%	8%	101.0	101.0	100.8
	46.4%	12%	93.4	93.4	93.4
	46.4%	16%	95.6	95.6	95.5

Table 5.6: Welfare.

5.6 Concluding remarks

A diversification motive appears as a sound candidate to account for long-term exposures across financial institutions. The first aim of this paper is to test this assumption.

To this purpose, we build a model of financial network in which the balance sheets of all institutions (including interconnections) are totally endogenous apart from the equity. The network formation process involves two components. The first one explains how a bank optimizes its balance sheet knowing the state of the other banks in the network. We prove the existence and partial uniqueness of the solution of this optimization. The second part shows how to form the network using the individual optimization program. The existence and unicity of this network are shown by numerical arguments. An important feature of our model is its ability to account for the main features of the banking and the insurance business with the same set of parameters. Nevertheless, we focus in this paper on the banking business.

Secondly, the characteristics of the resulting network are compared to features usually observed. As to the shape of the network, we theoretically find that the diversification motive leads to a network close to those observed across big banks. Regarding the size and support of the interconnections, we show that a correct magnitude is reached under standard calibration. Moreover, the results are sensitive to some specifications, for example the pricing method of shares and debt securities.

The fact that our network is totally endogenous allows studying how it adapts to regulatory changes. Thus, the second aim is to apply our model to fairly assess the impact of regulation on interbank exposures. To this purpose we study the evolution of the welfare with respect to the regulatory weights k^A and k^π. We observe that the welfare is higher under regulations favoring interconnections.

Ongoing work includes the complete study in the case of insurance companies and the extension to short-term interconnections. An exhaustive sensitivity analysis of the obtained network with respect to macroeconomic parameters like the returns's means as well as other specifications -e.g. concerning the pricing of shares and debt securities- are also under study. Finally, a simulation exercise in the case of 3 or 4 banks would also be of great interest.

5.7 Appendix: Example of public information on banks' balance sheets

Dollar Amounts in Thousands			BHCK	
Assets				
1. Cash and balances due from depository institutions:				
a. Noninterest-bearing balances and currency and coin[1]			0081	37192000
b. Interest-bearing balances:[2]				
(1) In U.S. offices			0395	54802000
(2) In foreign offices, Edge and Agreement subsidiaries, and IBFs			0397	16469000
2. Securities:				
a. Held-to-maturity securities (from Schedule HC-B, column A)			1754	54922000
b. Available-for-sale securities (from Schedule HC-B, column D)			1773	243227000
3. Federal funds sold and securities purchased under agreements to resell:				
a. Federal funds sold in domestic offices		BHDM	B987	0
b. Securities purchased under agreements to resell[3]		BHCK	B989	236749000
4. Loans and lease financing receivables:				
a. Loans and leases held for sale			5369	14549000
b. Loans and leases, net of unearned income	B528	946724000		
c. LESS: Allowance for loan and lease losses	3123	21235000		
d. Loans and leases, net of unearned income and allowance for loan and lease losses (item 4.b minus 4.c)			B529	925489000
5. Trading assets (from Schedule HC-D)			3545	283242000
6. Premises and fixed assets (including capitalized leases)			2145	10836000
7. Other real estate owned (from Schedule HC-M)			2150	2221000
8. Investments in unconsolidated subsidiaries and associated companies			2130	5306000
9. Direct and indirect investments in real estate ventures			3656	6348000
10. Intangible assets:				
a. Goodwill			3163	69930000
b. Other intangible assets (from Schedule HC-M)			0426	11943000
11. Other assets (from Schedule HC-F)			2160	152461000
12. Total assets (sum of items 1 through 11)			2170	2125686000

Dollar Amounts in Thousands	BHDM	
Liabilities		
13. Deposits:		
a. In domestic offices (from Schedule HC-E):		
(1) Noninterest-bearing[1]	6631	355864000
(2) Interest-bearing	6636	653607000
b. In foreign offices, Edge and Agreement subsidiaries, and IBFs:	BHFN	
(1) Noninterest-bearing	6631	5816000
(2) Interest-bearing	6636	67785000
14. Federal funds purchased and securities sold under agreements to repurchase:	BHDM	
a. Federal funds purchased in domestic offices[2]	B993	166000
	BHCK	
b. Securities sold under agreements to repurchase[3]	B995	232442000
15. Trading liabilities (from Schedule HC-D)	3548	126986000
16. Other borrowed money (includes mortgage indebtedness and obligations under capitalized leases) (from Schedule HC-M)	3190	270160000
17. Not applicable		
18. Not applicable		
19. a. Subordinated notes and debentures[4]	4062	31100000
b. Subordinated notes payable to unconsolidated trusts issuing trust preferred securities, and trust preferred securities issued by consolidated special purpose entities	C699	7695000
20. Other liabilities (from Schedule HC-G)	2750	142964000
21. Total liabilities (sum of items 13 through 20)	2948	1894585000
22. Not applicable		
Equity Capital		
Holding Company Equity Capital		
23. Perpetual preferred stock and related surplus	3283	14241000
24. Common stock (par value)	3230	107000
25. Surplus (exclude all surplus related to preferred stock)	3240	157085000
26. a. Retained earnings	3247	67308000
b. Accumulated other comprehensive income[5]	B530	-7709000
c. Other equity capital components[6]	A130	0
27. a. Total holding company equity capital (sum of items 23 through 26.c)	3210	231032000
b. Noncontrolling (minority) interests in consolidated subsidiaries	3000	69000
28. Total equity capital (sum of items 27.a and 27.b)	G105	231101000
29. Total liabilities and equity capital (sum of items 21 and 28)	3300	2125686000

Figure 5.5: Excerpt of the Consolidated Financial Statements for BHCs of Bank of America at 06/30/2013. Source: www.ffiec.gov.

5.8 Appendix: The model of Gouriéroux et al. (2012)

In this part, we expose the model of Gouriéroux et al. (2012), that provides the conditions defining an equilibrium between n financial institutions intertwined through shares and

debt securities.

5.8.1 Existence and uniqueness of the equilibrium

Proposition 5.5. *Let us denote by* $\mathbf{K} = (K_i)_{i=1,\dots,n}$, $\mathbf{L} = (L_i)_{i=1,\dots,n}$, $\mathbf{L}^* = (L_i^*)_{i=1,\dots,n}$, $\mathbf{Ax} = (Ax_i)_{i=1,\dots,n}$ *and* $\boldsymbol{A\ell} = (A\ell_i)_{i=1,\dots,n}$. *There exists a unique liquidation equilibrium, that is a unique set of values for* $\mathbf{K}$ *and* $\mathbf{L}$ *for any given values of* $\mathbf{L}^*$, $\mathbf{Ax}$, $\boldsymbol{A\ell}$ *if for all* $i, j = 1, \dots, n$*:*

- $(A1')$ *we have* $\pi_{i,j} \geq 0$, $\gamma_{i,j} \geq 0$*;*
- $(A2')$ *we have* $Ax_i \geq 0$, $A\ell_i \geq 0$, $L_i^* \geq 0$*;*
- $(A3')$ *we have* $\sum_{i=1}^{n} \pi_{i,j} < 1$, $\sum_{i=1}^{n} \gamma_{i,j} < 1$.

Proof. See Gouriéroux et al. (2012). □

Assumptions $(A1')$ and $(A2')$ define a proper space for the parameters: all elements composing the balance sheet must obviously be non-negative.

Assumption $(A3')$ means that some shareholders and creditors do not belong to the perimeter of the selected financial institutions. In practice, the first part of $(A3')$ is generally satisfied providing that we consider consolidated groups. Indeed, the empirical evidence in the studies by Gauthier et al. (2012) and Alves et al. (2013) clearly shows that $\sum_{i=1}^{n} \pi_{i,j} < 1$. The constraint on the γ_{ij} is largely satisfied since core deposits (deposits from external agents) represent approximately 55% of a bank's debt.

5.8.2 Case of two financial institutions

For illustrative purposes, let us consider a network of two institutions whose balance sheets are shown in Table 5.7. In such a case the equilibrium equations (5.1)-(5.2) are

$$\left\{\begin{array}{rl} K_1 = & \max\Big(\pi_{1,1}K_1 + \pi_{1,2}K_2 + \gamma_{1,2}L_2 + A\ell_1 + Ax_1 - L_1^*, 0\Big), \\ L_1 = & \min\Big(\pi_{1,1}K_1 + \pi_{1,2}K_2 + \gamma_{1,2}L_2 + A\ell_1 + Ax_1, L_1^*\Big), \\ K_2 = & \max\Big(\pi_{2,1}K_1 + \pi_{2,2}K_2 + \gamma_{2,1}L_1 + A\ell_2 + Ax_2 - L_2^*, 0\Big), \\ L_2 = & \min\Big(\pi_{2,1}K_1 + \pi_{2,2}K_2 + \gamma_{2,1}L_1 + A\ell_2 + Ax_2, L_2^*\Big). \end{array}\right. \tag{5.11}$$

One can identify 4 regimes depending on the situations of Institutions 1 and 2, respectively. These regimes, represented in Figure 5.6, are:

- Regime 1: both Institutions 1 and 2 are solvent;
- Regime 2: both Institutions 1 and 2 default;
- Regime 3: Institution 1 defaults while Institution 2 is solvent;
- Regime 4: Institution 1 is solvent while Institution 2 defaults.

Institution 1	
Asset	Liability
$\pi_{1,1}K_1$	L_1
$\pi_{1,2}K_2$	K_1
$\gamma_{1,2}L_2$	
$A\ell_1$	
Ax_1	

Institution 2	
Asset	Liability
$\pi_{2,1}K_1$	L_2
$\pi_{2,2}K_2$	K_2
$\gamma_{2,1}L_1$	
$A\ell_2$	
Ax_2	

Table 5.7: Balance sheets of Institutions 1 and 2.

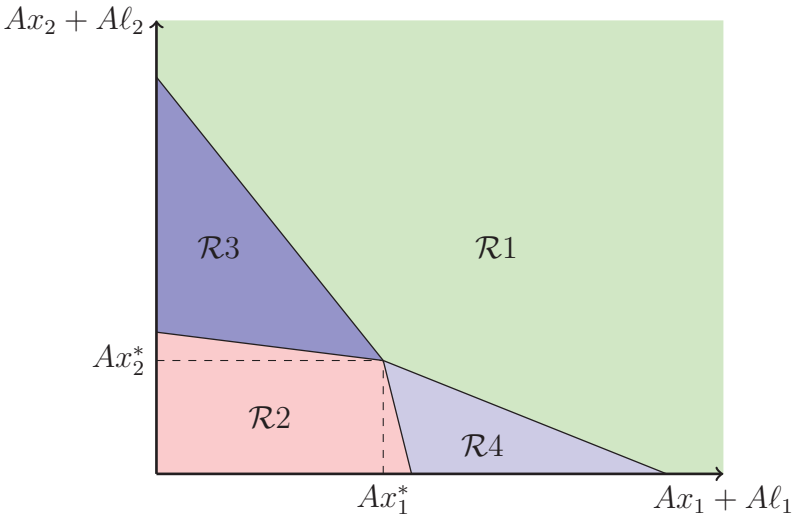


Figure 5.6: Regimes.

Figure 5.6 motivates the existence of interconnections between institutions. In a situation without interconnections, the 4 regimes would be defined by rectangles. Here the bounds deviate due to the presence of interconnections. In the case where the external assets of Institution 2, $Ax_2+A\ell_2$, are just above the limit value Ax_2^*, if it is interconnected and if $Ax_1 + A\ell_1$ is low, then Institution 2 can default ($\mathcal{R}_2$ is larger in the presence of interconnections). In this case, interconnections have a negative effect since the predicament of Institution 1 negatively impacts Institution 2 by contagion. When $Ax_2 + A\ell_2$ is very low, Institution 2 necessarily defaults if not linked to Institution 1. However, if Institution 2 owns shares of Institution 1, Institution 2 can survive if the external assets' value of Institution 1 is sufficient ($\mathcal{R}_1$ is larger in the presence of interconnections). In such a case, Institution 2 takes advantage of the high yield investments of Institution 1. Thus, we understand that the impacts of interconnections are not necessarily negative and must be fairly assessed.

5.9 Appendix: Proofs

For Theorem 5.1

Proof. For $i = 1, \dots, n$, let us denote the vectors of all control variables by $\mathbf{X}$. We have

$$\mathbf{X} = \left(Ax_i^{(0)}, A\ell_i^{(0)}, L_i^{(0)}, \omega_i, \pi_{i,1}, \dots, \pi_{i,n}, \gamma_{i,1}, \dots, \gamma_{i,n} \right)' \in \mathcal{X}_{ad},$$

where $\mathcal{X}_{ad}$ is the admissible space satisfying all constraints of $\mathcal{P}_i$. From now on, for the sake of notational simplicity, we omit the dependence in i of the vectors containing the control variables.

The proof relies on the Weierstrass Theorem: a continuous function on a compact set reaches its bounds. Therefore, we first show the continuity of the objective function and then the compactness of the admissible set $\mathcal{X}_{ad}$.

Continuity of the objective function

Under $(A2)$ and $(A3)$, both u_i and F_R are continuous. Therefore, the expectation is also continuous and the objective function is continuous.

Compactness of the admissible set $\mathcal{X}_{ad}$

To prove the compactness of $\mathcal{X}_{ad}$, we show that it is a closed and a bounded set. Before, we prove that $\mathcal{X}_{ad}$ is not empty.

$\mathcal{X}_{ad}$ is non-empty:

Let us consider the vector of parameters $\mathbf{X}_0$ defined as

$$\mathbf{X}_0 = \left(K_i^{(0)} - k^L l(0,0), k^L l(0,0), 0, \dots, 0 \right)'.$$

All constraints apart from $Ax_i \geq 0$, (BC), (SC) and (LC) are obviously satisfied. The inequality $Ax_i \geq 0$ imposes that $K_i^0 \geq k^L l(0,0)$ which is not restrictive due to the low value of k^L and the fact that $l(0,0)$ can be taken equal to one. The constraint (BC) reduces to $K_i^{(0)} - k^L l(0,0) + k^L l(0,0) = K_i^{(0)}$ and is thus satisfied. The constraint (SC) is written

$$\begin{aligned} K_i^{(0)} \geq k_i^A A x_i^{(0)} &\iff K_i^{(0)} \geq k_i^A [K_i^{(0)} - k^L l(0,0)] \\ &\iff K_i^{(0)}(1 - k_i^A) \geq -k_i^A \; k^L l(0,0). \end{aligned}$$

Due to the inequality $k_i^A < 1$ and the positivity of k_i^A, k^L and function l, the left hand term is positive whereas the right one is negative, giving that (SC) is satisfied.
Thus $\mathbf{X}_0$ belongs to the admissible set $\mathcal{X}_{ad}$, which is therefore not empty.

$\mathcal{X}_{ad}$ is a closed set:

In order to show that the admissible set $\mathcal{X}_{ad}$ is a closed set, we show that it is the intersection of closed sets.

i) The constraint (BC) can be written

$$Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} - K_i^{(0)} - L_i^{(0)} = 0.$$

The corresponding admissible space is the reciprocal image of the singleton $\{0\}$, which is a closed set of $\mathbb{R}$, by a continuous function. Therefore, (BC) defines a closed set.

ii) The constraint (SC) is derived in

$$k_i^A \; Ax_i^{(0)} + k^{\pi} \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + k^{\gamma} \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} - K_i^{(0)} \leq 0.$$

The corresponding admissible space is the reciprocal image of $[-\infty,\ 0]$, which is a closed set of $\mathbb{R}$, by a continuous function. Therefore, (SC) defines a closed set.

iv) The constraint (LC) is derived in

$$k^L \; l(\omega_i,\ L_i^{(0)}) - A\ell_i^{(0)} \leq 0.$$

The corresponding admissible space is the reciprocal image of $[-\infty,\ 0]$, which is a closed set of $\mathbb{R}$, by a continuous function. Therefore, (LC) defines a closed set.

v) The positivity constraints ($Ax_i^{(0)} \geq 0$, $A\ell_i^{(0)} \geq 0$ and $L_i^{(0)} \geq 0$) also define closed sets, as the reciprocal images of $[0, +\infty]$, which is a closed set of $\mathbb{R}$, by a continuous function.

vi) The constraints $\omega_i \in [0,1]$, $0 \leq \pi_{i,j} \leq 1 - c_j^{\pi}$, $0 \leq \gamma_{i,j} \leq 1 - c_j^{\gamma}$ $(\forall j \in \{1, \ldots, n\})$ define a closed admissible set as the reciprocal images of $[0,1], [0, 1 - c_j^{\pi}]$ and $[0, 1 - c_j^{\gamma}]$, which are closed sets of $\mathbb{R}$, by a continuous function.

The admissible set $\mathcal{X}_{ad}$ is the intersection of the admissible sets defined by each constraint. Moreover, an intersection of closed sets is a closed set. Thus, combining points i) to vi), we obtain that $\mathcal{X}_{ad}$ is a closed set.

$\mathcal{X}_{ad}$ is a bounded set:

Let us show that the admissible set is bounded.
Conditions $0 \leq \pi_{i,j} \leq 1 - c_j^{\pi}$ and $0 \leq \gamma_{i,j} \leq 1 - c_j^{\gamma}$ $(\forall j \in \{1, \ldots, n\})$ show that all the $\pi_{i,j}$ and $\gamma_{i,j}$ are bounded. The same is true for $\omega_i \in [0,1]$. Let us now prove that $A\ell_i$, Ax_i and L_i are bounded.

i) Bound for $A\ell_i$

The combination of the constraints $L_i^{(0)} \geq 0$ and (BC) implies that the institution invests at least all its own capital.

If $L_i^{(0)} = 0$, $K_i^{(0)}$ is an upper-bound for $A\ell_i^{(0)}$.

Let us now consider the case $L_i^{(0)} > 0$. The constraint (BC) can be used to express the debt as a function of other control variables,

$$L_i^{(0)} = Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j}\mathcal{K}_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j}\mathcal{L}_j^{(0)} - K_i^{(0)}.$$

Using this last equation, one can express $P_i^{(1)}$ as a function of other control variables:

$$\begin{aligned}
P_i^{(1)} &= Ax_i^{(0)}(1+r_i) + A\ell_i^{(0)}(1+r_{rf}) \\
&+ \sum_{j=1}^{n} \pi_{i,j} \max\left(\kappa_j[Ax_j^{(0)}(1+r_j) + A\ell_j^{(0)}(1+r_{rf})] - L_j^{*(1)}, 0\right) \\
&+ \sum_{j=1}^{n} \gamma_{i,j} \min\left(\kappa_j[Ax_j^{(0)}(1+r_j) + A\ell_j^{(0)}(1+r_{rf})], L_j^{*(1)}\right) \\
&- [1+r_D(\omega_i)]\Big(Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j}K_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j}L_j^{(0)} - K_i^{(0)}\Big) \\
&= Al_i^{(0)}[r_{rf} - r_D(w_i)] + d(\mathbf{X}_{-A\ell}, \mathbf{r}), \qquad (5.12)
\end{aligned}$$

where $\mathbf{X}_{-A\ell} = \left(Ax_i^{(0)}, L_i^{(0)}, \omega_i, \pi_{i,1}, \ldots, \pi_{i,n}, \gamma_{i,1}, \ldots, \gamma_{i,n}\right)'$ is the vector of all the control variables apart from $A\ell_i^{(0)}$, $\mathbf{r} = (r_1, \ldots, r_n)'$ is the vector of the realized net returns of the external assets and d is some function. The position $P_i^{(1)}$ is a function of $A\ell_i^{(0)}$, $\mathbf{X}_{-A\ell}$ and $\mathbf{r}$, from now on denoted by $P_i^{(1)}(A\ell_i^{(0)}, \mathbf{X}_{-A\ell}, \mathbf{r})$. Assumption $(A4)$ states that $r_D(\omega_i) > r_{rf}$, giving that $P_i^{(1)}(.,.,.)$ is strictly decreasing with respect to $A\ell_i^{(0)}$.

Let us consider a value $V_1 > K_i^{(0)}$ for $A\ell_i^{(0)}$. From (5.12), we see that, for all admissible $\mathbf{X}_{-A\ell}$, there exists a set $\varepsilon_1, \ldots, \varepsilon_n$ of values such that, if $r_k \geq \varepsilon_k, k = 1, \ldots, n$, then $P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r}) > 0$. For a second value V_2 such that $K_i^{(0)} \leq V_2 < V_1$, we have for all admissible $\mathbf{X}_{-A\ell}$

$$P_i^{(1)}(V_2, \mathbf{X}_{-A\ell}, \mathbf{r}) > P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r}). \qquad (5.13)$$

Therefore, if $r_k \geq \varepsilon_k, k = 1, \ldots, n$, we have

$$P_i^{(1)}(V_2, \mathbf{X}_{-A\ell}, \mathbf{r}) > 0. \qquad (5.14)$$

Now, let us compare the expected utility at $A\ell_i^{(0)} = V_1$ and $A\ell_i^{(0)} = V_2$. We have

$$\begin{aligned}
\mathbb{E}\left[u_i\left(K_i^{(1)}\right)\right](V_1, \mathbf{X}_{-A\ell}) &= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} u_i\left(\max\left[P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r}), 0\right]\right) f_R(\mathbf{r})\, d\mathbf{r} \\
&= \int_{-\infty}^{\varepsilon_1} \ldots \int_{-\infty}^{\varepsilon_n} u_i\left(\max\left[P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r}), 0\right]\right) f_R(\mathbf{r})\, d\mathbf{r} \\
&\quad + \int_{\varepsilon_1}^{+\infty} \ldots \int_{\varepsilon_n}^{+\infty} u_i\left[P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r})\right] f_R(\mathbf{r})\, d\mathbf{r}. \qquad (5.15)
\end{aligned}$$

By the same decomposition and using (5.14), we obtain, for $A\ell_i^{(0)} = V_2 > V_1$,

$$\begin{aligned}
\mathbb{E}\left[u_i\left(K_i^{(1)}\right)\right](V_2, \mathbf{X}_{-A\ell}) &= \int_{-\infty}^{\varepsilon_1} \ldots \int_{-\infty}^{\varepsilon_n} u_i\left(\max\left[P_i^{(1)}(V_2, \mathbf{X}_{-A\ell}, \mathbf{r}), 0\right]\right) f_R(\mathbf{r})\, d\mathbf{r} \\
&\quad + \int_{\varepsilon_1}^{+\infty} \ldots \int_{\varepsilon_n}^{+\infty} u_i\left[P_i^{(1)}(V_2, \mathbf{X}_{-A\ell}, \mathbf{r})\right] f_R(\mathbf{r})\, d\mathbf{r}. \qquad (5.16)
\end{aligned}$$

Using (5.13), we have, for $\mathbf{r} \in (-\infty, \varepsilon_1] \times \cdots \times (-\infty, \varepsilon_n]$,

$$\max\left[P_i^{(1)}(V_2, \mathbf{X}_{-A\ell}, \mathbf{r}), 0\right] \geq \max\left[P_i^{(1)}(V_1, \mathbf{X}_{-A\ell}, \mathbf{r}), 0\right],$$

and, since u_i is strictly increasing ($(A2)$),

$$u_i\left(\max\left[P_i^{(1)}(V_2,\mathbf{X}_{-A\ell},\mathbf{r}),0\right]\right)\geq u_i\left(\max\left[P_i^{(1)}(V_1,\mathbf{X}_{-A\ell},\mathbf{r}),0\right]\right).$$

Using (5.13), we have, for all $\mathbf{r}\in[\varepsilon_1,+\infty)\times\cdots\times[\varepsilon_n,+\infty)$,

$$P_i^{(1)}(V_2,\mathbf{X}_{-A\ell},\mathbf{r})>P_i^{(1)}(V_1,\mathbf{X}_{-A\ell},\mathbf{r}),$$

and, since u_i is strictly increasing,

$$u_i\left[P_i^{(1)}(V_2,\mathbf{X}_{-A\ell},\mathbf{r})\right]>u_i\left[P_i^{(1)}(V_1,\mathbf{X}_{-A\ell},\mathbf{r})\right].$$

Moreover, there exists $a\in\mathbb{R}$, such that, for all $\mathbf{r}\in[a,+\infty)^n$, $f_R(\mathbf{r})>0$ ($(A3)$).

Therefore, combining (5.15) and (5.16) yields

$$\forall \text{ admissible } \mathbf{X}_{-A\ell},\quad \mathbb{E}\left[u_i\left(K_i^{(1)}\right)\right](V_1,\mathbf{X}_{-A\ell})<\mathbb{E}\left[u_i\left(K_i^{(1)}\right)\right](V_2,\mathbf{X}_{-A\ell}),$$

meaning that, for $A\ell_i^{(0)}\geq K_i^{(0)}$, the objective function is strictly decreasing with respect to $A\ell_i^{(0)}$. Consequently, $\mathcal{P}_i$ is equivalent if we upper-bound the space of $A\ell_i$. Moreover, since $A\ell_i$ is lower-bounded by 0, $A\ell_i$ is bounded.

ii) Bounds for Ax_i and L_i

Let us recall that (SC) is written

$$K_i^{(0)}\geq k_i^A Ax_i^{(0)}+k^\pi\sum_{j=1}^n\pi_{i,j}\mathcal{K}_j^{(0)}+k^\gamma\sum_{j=1}^n\gamma_{i,j}\mathcal{L}_j^{(0)}.$$

The equity $K_i^{(0)}$ is fixed as an endowment. Moreover, in the right hand term of (SC), all components are positive. Thus, it imposes that each term is bounded. Therefore, $k_i^A Ax_i$ is upper-bounded and, since $k_i^A>0$ by assumption, Ax_i is upper-bounded. Moreover, $Ax_i\geq 0$ and thus Ax_i is bounded.

Using the fact that $k^\pi>0$ and $k^\gamma>0$, we also obtain that both $\sum_{j=1}^n\pi_{i,j}K_j^{(0)}$ and $\sum_{j=1}^n\gamma_{i,j}L_j^{(0)}$ are upper-bounded. Let us recall that (BC) gives

$$L_i^{(0)}=Ax_i^{(0)}+A\ell_i^{(0)}+\sum_{j=1}^n\pi_{i,j}\mathcal{K}_j^{(0)}+\sum_{j=1}^n\gamma_{i,j}\mathcal{L}_j^{(0)}-K_i^{(0)},$$

implying that $L_i^{(0)}$ is upper-bounded since all terms in the right part of the equation are upper-bounded. Moreover, since $L_i^{(0)}\geq 0$ by assumption, $L_i^{(0)}$ is bounded.

Existence

To summarize, the admissible set is not empty. It is also closed and bounded, and therefore compact. The objective function is continuous and the Weierstrass Theorem ensures the existence of a solution. □

For Theorem 5.2

Proof.

Existence

The existence can be shown exactly in the same way as for Theorem 5.1.

Uniqueness

The uniqueness is based on a fundamental theorem of optimization: a strictly concave function on a closed convex set admits a unique maximum. We first show that the admissible set is convex and then that the objective function is strictly concave.

Convexity of the admissible set

As before, we denote

$$\mathbf{X} = \left(Ax_i^{(0)}, A\ell_i^{(0)}, L_i^{(0)}, \omega_i, \pi_{i,1}, \ldots, \pi_{i,n}, \gamma_{i,1}, \ldots, \gamma_{i,n} \right)' \in \mathcal{X}_{ad},$$

where $\mathcal{X}_{ad}$ is the admissible space of $\mathcal{P}_i'$.

Let us show that each constraint of $\mathcal{P}_i'$ defines a convex set. All constraints excluding (LC) involve linear functions of the control variables and thus each of these constraints obviously defines a convex set.

The constraint (LC) requires more attention. For the sake of notational simplicity, let us denote by $x = \omega_i$, $y = L_i^{(0)}$ and $z = A\ell_i^{(0)}$. The constraint (LC) can therefore be re-written $z > l(x, y)$. The corresponding set is the epigraph of the function l. The epigraph is convex if and only if l is convex, i.e. if and only if the Hessian of l, $\mathbf{H_l}$, is semi definite positive. By definition, we have

$$\mathbf{H_l} = \begin{pmatrix} \dfrac{\partial^2 l}{\partial x^2} & \dfrac{\partial^2 l}{\partial x \partial y} \\ \dfrac{\partial^2 l}{\partial x \partial y} & \dfrac{\partial^2 l}{\partial y^2} \end{pmatrix}.$$

The Sylvester's criterion states that a matrix is semi definite positive if and only if all its leading principal minors are positive, i.e.

$$\frac{\partial^2 l}{\partial x^2} \geq 0 \quad \text{and} \quad \frac{\partial^2 l}{\partial x^2}\frac{\partial^2 l}{\partial y^2} \geq \left(\frac{\partial^2 l}{\partial x \partial y} \right)^2.$$

Thus, under $(A8)$, (LC) defines a convex set and finally all constraints define a convex set. Since the intersection of convex sets is a convex set, $\mathcal{X}_{ad}$ is a convex set.

We want to show that there is uniqueness of the solution of the optimization of the triple $(Ac_i^{(0)}, L_i^{(0)}, \omega_i)$, where $Ac_i^{(0)}$ is one of the variables appearing on the asset side, i.e. among $Ax_i^{(0)}$, $A\ell_i^{(0)}$, $\pi_{i,1}, \ldots, \pi_{i,n}, \gamma_{i,1}, \ldots, \gamma_{i,n}$. Let us denote

$$\mathbf{X3} = \left(Ac_i^{(0)}, L_i^{(0)}, \omega_i \right)' \in \mathcal{X}3_{ad},$$

where $\mathcal{X}3_{ad}$ is the admissible set of the three-dimensional optimization program. By using the same arguments as for $\mathcal{X}_{ad}$, $\mathcal{X}3_{ad}$ defines a convex set, whatever the control variable $Ac_i^{(0)}$ that is chosen. Moreover, note that one can show that $\mathcal{X}3_{ad}$ is a closed set, as for Theorem 5.1.

Expectation and underlying objective function

In the following, we generally denote the position by $P_i^{(1)}(\mathbf{X3}, \mathbf{r})$ but sometimes we omit the arguments $\mathbf{X3}$ and $\mathbf{r}$ for simplicity. The strict concavity of $u_i\left[v\left(P_i^{(1)}\right)\right]$ is a sufficient condition to obtain the strict concavity of $\mathbb{E}\left\{u_i\left[v\left(P_i^{(1)}\right)\right]\right\}$ with respect to $\mathbf{X3}$. Indeed, let us assume that $u_i\left[v\left(P_i^{(1)}\right)\right]$ is strictly concave. Combining the latter assumption with the fact that f_R is strictly positive on $[a, +\infty)^n$ $((A3))$, we get, for all $(\mathbf{X3}_1, \mathbf{X3}_2) \in \mathcal{X}3_{ad}^2$ and for all $\lambda \in [0, 1]$,

$$
\begin{aligned}
&\mathbb{E}\left\{u_i\left[v\left(P_i^{(1)}\right)\right]\right\}(\lambda\mathbf{X3}_1 + (1-\lambda)\mathbf{X3}_2) \\
&= \int_{\mathbb{R}^n} u_i\left(v\left[P_i^{(1)}\left(\lambda\mathbf{X3}_1 + (1-\lambda)\mathbf{X3}_2, \mathbf{r}\right)\right]\right) f_R(\mathbf{r})\, d\mathbf{r} \\
&> \int_{\mathbb{R}^n} \left[\lambda\, u_i\left(v\left[P_i^{(1)}(\mathbf{X3}_1, \mathbf{r})\right]\right) + (1-\lambda)\, u_i\left(v\left[P_i^{(1)}(\mathbf{X3}_2, \mathbf{r})\right]\right)\right] f_R(\mathbf{r})\, d\mathbf{r} \\
&= \lambda\, \mathbb{E}\left\{u_i\left[v\left(P_i^{(1)}\right)\right]\right\}(\mathbf{X3}_1) + (1-\lambda)\mathbb{E}\left\{u_i\left[v\left(P_i^{(1)}\right)\right]\right\}(\mathbf{X3}_2),
\end{aligned}
$$

showing the strict concavity of the expected utility.

Strict concavity of the underlying objective function

We now focus on $u_i\left[v\left(P_i^{(1)}\right)\right]$. We consider that only one control variable is free on the asset side. For the sake of notational simplicity, we denote by $x_1 = Ac_i^{(0)}$, $x_2 = \omega_i$ and $x_3 = L_i^{(0)}$. Here we interpret $P_i^{(1)}$ as the function defined by

$$
\begin{array}{rccc}
P_i^{(1)} & : & \mathbb{R}^+ \times [0,1] \times \mathbb{R}^+ & \to & \mathbb{R} \\
 & & \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & \mapsto & t(x_1) - [1 + r_D(x_2)]x_3,
\end{array}
$$

where $t(.)$ is a linear transformation mapping the control variable chosen into the value of the total assets $Ax_i^{(0)}(1+r_i) + Al_i^{(0)}(1+r_{rf}) + \sum_{j=1}^n \pi_{i,j}\mathcal{K}_j^{(0)} + \sum_{j=1}^n \gamma_{i,j}\mathcal{L}_j^{(0)}$. Let us denote by g the function $u_i \circ v \circ P_i^{(1)}$. We now study the strict concavity of g. We denote by $m = u_i \circ v$, yielding $g = m \circ P_i^{(1)}$. The function g is strictly concave if and only if its Hessian matrix $\mathbf{H_g}$ is definite negative. We have

$$
\mathbf{H_g} = m'' \begin{pmatrix}
1 & -x_3\, r_D'(x_2) & -[1 + r_D(x_2)] \\
-x_3\, r_D'(x_2) & -x_3\left[\frac{m'}{m''} r_D''(x_2) + r_D'^2(x_2)x_3\right] & r_D'(x_2)\left[-\frac{m'}{m''} + x_3[1 + r_D(x_2)]\right] \\
-[1 + r_D(x_2)] & r_D'(x_2)\left[-\frac{m'}{m''} + x_3[1 + r_D(x_2)]\right] & [1 + r_D(x_2)]^2
\end{pmatrix}.
$$

The Sylvester's criterion states that $\mathbf{H_g}$ is definite negative if and only if all its leading principal minors are strictly negative. Let us now study the three corresponding minors.

i) First minor

The first minor is

$$
\mathrm{Det}_1 = |m''|.
$$

According to the Sylvester's criterion, $m'' < 0$ is imposed.

ii) Second minor

The second minor is

$$\mathrm{Det}_2 = m''^2 \times \left[-x_3 \left[\frac{m'}{m''} r''_D(x_2) + r'^2_D(x_2) x_3 \right] + x_3^2 \; r'^2_D(x_2) \right].$$

Thus, the Sylvester's condition imposes, $\forall x_2 \in [0,1]$ and $x_3 \in \mathbb{R}^+$,

$$\begin{aligned} & x_3 \left[m' m'' r''_D(x_2) + m''^2 r'^2_D(x_2) x_3 \right] > x_3^2 r'^2_D(x_2) m''^2 \\ \iff \quad & m' m'' r''_D(x_2) + m''^2 r'^2_D(x_2) x_3 > x_3 r'^2_D(x_2) m''^2 \\ \iff \quad & m' m'' r''_D(x_2) > 0 \\ \iff \quad & r''_D(x_2) < 0, \end{aligned}$$

since $m' > 0$ by assumption (u and v are strictly increasing so $m = u \circ v$ is strictly increasing as well) and the previous condition (see i) imposes $m'' < 0$.

iii) Third minor

We compute the third minor using Sarrus' rule. We obtain

$$\begin{aligned} \mathrm{Det}_3 = m''^3 \Bigg\{ & - x_3 \left[\frac{m'}{m''} r''_D(x_2) + r'^2_D(x_2) x_3 \right] [1 + r_D(x_2)]^2 \\ & + 2(-x_3 \; r'_D(x_2)) \; r'_D(x_2) \left[-\frac{m'}{m''} + x_3 [1 + r_D(x_2)] \right] \times (-[1 + r_D(x_2)]) \\ & - \Bigg[[1 + r_D(x_2)]^2 \left(- x_3 \left[\frac{m'}{m''} r''_D(x_2) + r'^2_D(x_2) x_3 \right] \right) + r'_D(x_2) \left[-\frac{m'}{m''} + x_3 [1 + r_D(x_2)] \right]^2 \\ & + x_3^2 \; r'^2_D(x_2) \; [1 + r_D(x_2)]^2 \Bigg] \Bigg\} \\ = m''^3 \Bigg[& \left\{ 2 x_3 \; r'^2_D(x_2) [1 + r_D(x_2)] + r'^2_D(x_2) \left(-\frac{m'}{m''} + x_3 \, [1 + r_D(x_2)] \right) \right\} \; \left(- \frac{m'}{m''} + x_3 [1 + r_D(x_2)] \right) \\ & + x_3^2 \; r'^2_D(x_2)) \; [1 + r_D(x_2)]^2 \Bigg]. \end{aligned}$$

Considering $m'' < 0$ (see ii) and $m' > 0$ (by assumption), we have $\dfrac{m'}{m''} < 0$. Thus, assuming $\forall x_2 \in [0,1], r'_D(x_2) \neq 0$, all terms in the brackets are strictly positive. Moreover $m''^3 < 0$ and thus the condition $\mathrm{Det}_3 < 0$ is satisfied.

Summary

The following assumptions

- $m''(x) < 0$ ($(A5)$);
- $r''_D < 0$ ($(A6)$);
- $r'_D \neq 0$ ($(A7)$);

are sufficient to ensure that the Hessian matrix of g is definite negative and therefore that g is strictly concave with respect to the control variable $Ac_i^{(0)}$, the debt $L_i^{(0)}$ and the maturity transformation ω_i.

Finally, under $(A5)$, $(A6)$, $(A7)$ and $(A8)$, the objective function $\mathbb{E}\left\{u_i\left[v\left(P_i^{(1)}\right)\right]\right\}$ is strictly concave on a closed convex set, showing the uniqueness. □

Remark 5.1. *Let us now come back to the choice of working directly on the integrand. Even if $u_i \circ v$ is not strictly concave everywhere, one may certainly expect the strict concavity to come from the integration with respect to the realized returns $\mathbf{r}$ (for some appropriate densities f_R). However, as we have shown, studying the concavity of a multivariate function involves studying its Hessian and this is already quite complicated in the case of the integrand. The Hessian matrix of the expected utility implies much more complicated expressions, especially products of integral, apart from the first leading minor. The condition on this first leading minor is written as follows:*

$$\int_{\mathbb{R}^n} (u_i \circ v \circ P_i^{(1)})''(\mathbf{X}, \mathbf{r}) f_R(\mathbf{r})\ d\mathbf{r} > 0.$$

Thus, even in the case of the first leading minor, it seems difficult to obtain results except in particular cases of very simple density functions f_R. Moreover, the study of the uniqueness of all control variables (and thus the study of the strict concavity with respect to all control variables) would require the study of a high dimensional Hessian, which is very difficult.

For Lemma 5.1

Proof. **i)** We consider the function defined by $\forall P \in \mathbb{R}, v(P) = P$.
We have $v'(P) = 1$ and $v''(P) = 0$. Thus, $(u_i \circ v)'(P) = u_i'[v(P)]v'(P) = u_i'(P)$, giving $(u_i \circ v)''(P) = u_i''(P)$. Therefore, $(A5)$ imposes $\forall P \in \mathbb{R}, u_i''(P) < 0$.

ii) Here, we consider the function defined by $\forall P \in \mathbb{R}, v(P) = \log(\exp(P) + 1)$.
We have, for all $P \in \mathbb{R}$,

$$v'(P) = \frac{e^P}{e^P + 1} \text{ and } v''(P) = \frac{e^P(e^P + 1) - e^P\ e^P}{(e^P + 1)^2} = \frac{e^P}{(e^P + 1)^2}.$$

Let us study the function $h = u_i \circ v = \log \circ v$. We have

$$h'(P) = \frac{v'(P)}{v(P)},$$

and thus

$$\begin{aligned} h''(P) &= \frac{v''(P)\ v(P) - v'^2(P)}{v^2(P)} \\ &= \frac{e^P}{(e^P + 1)^2} \frac{1}{\log(e^P + 1)} - \frac{e^P e^P}{(e^P + 1)^2} \frac{1}{[\log(e^P + 1)]^2} \\ &= \frac{e^P}{(e^P + 1)^2} \frac{1}{\log(e^P + 1)} \left(1 - \frac{e^P}{\log(e^P + 1)}\right). \end{aligned}$$

The first two factors are positive whereas the third one is negative (since $\forall x \in \mathbb{R}_+^*, \log(1 + x) < x$). Consequently, $\forall P \in \mathbb{R},\ h''(P) < 0$. Hence the result. □

For Lemma 5.3

Proof. We consider the function defined by $\forall \omega \in [0,1]$ and $\forall L \in \mathbb{R}^+, l(\omega, L) = \exp(\omega)\exp(L)$. We have

$$\forall \omega \in [0,1] \text{ and } \forall L \in \mathbb{R}^+, \frac{\partial^2 l}{\partial \omega^2} = \exp(\omega)\exp(L) > 0 \quad \text{and}$$

$$\frac{\partial^2 l}{\partial \omega^2}\frac{\partial^2 l}{\partial L^2} = [\exp(\omega)\exp(L)]^2 = \left(\frac{\partial^2 l}{\partial \omega \partial L}\right)^2.$$

That shows that $(A8)$ is satisfied. □

For Proposition 5.1

Proof. The proof is based on the Karuch, Kuhn, Tucker (KKT) Theorem, which provides necessary conditions on a local optimum of an optimization problem under equality and inequality constraints. We show that assuming $\pi^* = 0$ leads to a contradiction.

The KKT Theorem states that there exist coefficients $\mu_i \geq 0$ such that a local maximum (Ax^*, π^*) is a local maximum of the objective function $\mathcal{L}^a$, defined as

$$\mathcal{L}^a = f - \mu_1(k^A Ax + k^\pi \pi - 1) + \mu_2 Ax + \mu_3 \pi - \mu_4(\pi - 1),$$

where f is the initial objective function, i.e. $\mathbb{E}\left[u(AxR_g + \pi R_g^\pi)\right]$. Moreover, the μ_i coefficients satisfy

$$\forall i, \mu_i C_i = 0,$$

where C_i is the i-th constraint.

At a local optimum, the KKT conditions are

$$\begin{cases} \dfrac{\partial f}{\partial Ax} - \mu_1 k^A + \mu_2 = 0 \\ \dfrac{\partial f}{\partial \pi} + \mu_3 - \mu_1 k^\pi - \mu_4 = 0 \\ \mu_1(k^A Ax^* + k^\pi \pi^* - 1) = 0 \\ \mu_2 Ax^* = 0 \\ \mu_3 \pi^* = 0 \\ \mu_4(\pi^* - 1) = 0 \end{cases}.$$

We now assume that $\pi^* = 0$. The last equation directly provides $\mu_4 = 0$. Since f is strictly increasing, Ax^* is necessarily strictly positive (such Ax^* is compatible with the constraints). Therefore, we have $\mu_2 = 0$. Thus, the first equation provides

$$\mu_1 = \frac{\partial f}{\partial Ax}\frac{1}{k^A}.$$

Injecting this result into the second equation gives

$$\mu_3 = \frac{\partial f}{\partial Ax}\frac{k^\pi}{k^A} - \frac{\partial f}{\partial \pi} < 0 \text{ (by assumption).} \tag{5.17}$$

Equation (5.17) is in contradiction with the KKT theorem, stating that $\forall i, \mu_i \geq 0$. Therefore, $\pi^* \neq 0$. □

For Proposition 5.2

Proof. First, let us recall that

$$
\begin{aligned}
P_i^{(1)} &= Ax_i^{(1)} + A\ell_i^{(1)} + \sum_{j=1}^{n} \pi_{i,j} K_j^{(1)} + \sum_{j=1}^{n} \gamma_{i,j} L_j^{(1)} \\
&\quad - [1 + r_D(\omega_i)]\Big(Ax_i^{(0)} + A\ell_i^{(0)} + \sum_{j=1}^{n} \pi_{i,j} \mathcal{K}_j^{(0)} + \sum_{j=1}^{n} \gamma_{i,j} \mathcal{L}_j^{(0)} - K_i^{(0)} \Big) \\
&= Ax_i^{(1)} + A\ell_i^{(1)} - [1 + r_D(\omega_i)](Ax_i^{(0)} + A\ell_i^{(0)} - K_i^{(0)}) + \sum_{j=1}^{n} \pi_{i,j} \Big(K_j^{(1)} - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)} \Big) \\
&\quad + \sum_{j=1}^{n} \gamma_{i,j} \Big(L_j^{(1)} - [1 + r_D(\omega_i)] \mathcal{L}_j^{(0)} \Big).
\end{aligned}
$$

We have $u_i \circ v = Id$. Then the derivative of the objective function with respect to $\pi_{i,j}$ is written

$$
\frac{\partial \mathbb{E}[u_i \circ v(P_i^{(1)})]}{\partial \pi_{ij}} = \frac{\partial \mathbb{E}(P_i^{(1)})}{\partial \pi_{ij}} = \mathbb{E}\left[K_j^{(1)} - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)} \right] = \mathbb{E}\left[K_j^{(1)} \right] - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)}, \tag{5.18}
$$

where

$$
K_j^{(1)} = \max\left(\kappa_j \left(Ax_j^{(1)} + A\ell_j^{(1)} \right) - L_j^{*(1)}, 0 \right).
$$

Let us now explicit the latter expression:

$$
\begin{aligned}
\kappa_j \left(Ax_j^{(1)} + A\ell_j^{(1)} \right) - L_j^{*(1)} &= \kappa_j \left(Ax_j^{(0)}(1 + r_j) + A\ell_j^{(0)}(1 + r_{rf}) \right) - L_j^{*}[1 + r_D(\omega_j)] \\
&= a_j r_j + b_j,
\end{aligned}
$$

by denoting $a_j = \kappa_j A x_j^{(0)}$ and $b_j = \kappa_j \left(Ax_j^{(0)} + A\ell_j^{(0)}(1 + r_{rf}) \right) - L_j^{*}[1 + r_D(\omega_j)]$
Then,

$$
\mathbb{E}\left[K_j^{(1)} \right] = \mathbb{E}\left[\max\left(a_j r_j + b_j, 0 \right) \right] = \int_{\frac{-b_j}{a_j}}^{+\infty} (a_j r_j + b_j)\ f_{R,j}(r_j)\ dr_j. \tag{5.19}
$$

Combining (5.18) and (5.19), we obtain

$$
\frac{\partial \mathbb{E}[u_i \circ v(P_i^{(1)})]}{\partial \pi_{ij}} > 0 \Longleftrightarrow \int_{\frac{-b_j}{a_j}}^{+\infty} (a_j r_j + b_j)\ f_{R,j}(r_j) > [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)}.
$$

The derivative with respect to γ_{ij} is written

$$
\begin{aligned}
\frac{\partial \mathbb{E}(P_i^{(1)})}{\partial \gamma_{ij}} &= \mathbb{E}[L_j^{(1)}] - [1 + r_D(\omega_i)]\mathcal{L}_j^{(0)} \\
&= \mathbb{E}\left[\min\left(a_j r_j + b_j, L_j^{*(1)} \right) \right] - [1 + r_D(\omega_i)]\mathcal{L}_j^{(0)} \\
&= \int_{-\infty}^{\frac{L_j^{*(1)} - b_j}{a_j}} (a_j r_j + b_j) f_{R,j}(r_j)\ dr_j + L_j^{*(1)} \mathbb{P}\left(r_j > \frac{L_j^{*(1)} - b_j}{a_j} \right) - [1 + r_D(\omega_i)]\mathcal{L}_j^{(0)} \\
&= \int_{-\infty}^{\frac{L_j^{*(1)} - b_j}{a_j}} (a_j r_j + b_j) f_{R,j}(r_j)\ dr_j + L_j^{*} c_j\ [1 + r_D(\omega_j)] - [1 + r_D(\omega_i)]\mathcal{L}_j^{(0)},
\end{aligned}
$$

by denoting $c_j = \mathbb{P}\left(r_j > \frac{L_j^{*(1)} - b_j}{a_j}\right)$. Finally,

$$\frac{\partial \mathbb{E}[u_i \circ v(P_i^{(1)})]}{\partial \pi_{ij}} > 0 \Longleftrightarrow \int_{-\infty}^{\frac{L_j^*[1+r_D(\omega_j)]-b_j}{a_j}} (a_j r_j + b_j) f_{R,j}(r_j)\, dr_j + L_j^* c_j\, [1+r_D(\omega_j)] > [1+r_D(\omega_i)]\mathcal{L}_j^{(0)}.$$

□

For Proposition 5.3

Proof. Let us at first consider the derivative with respect to π_{ij}. We have

$$\frac{\partial \left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \pi_{ij}} = \frac{\partial (u_i \circ v)}{\partial P_i^{(1)}} \frac{\partial P_i^{(1)}}{\partial \pi_{ij}}.$$

The first term $\dfrac{\partial (u_i \circ v)}{\partial P_i^{(1)}}$ can be interpreted as some kind of marginal utility (with the utility being composed with function v). It depends on the returns of all banks connected to Bank i and not only on the return of Bank j. Let us denote

$$h_{i1}(r_1, \ldots, r_j, \ldots, r_n) = \frac{\partial (u_i \circ v)}{\partial P_i^{(1)}}.$$

Moreover, we have

$$\frac{\partial P_i^{(1)}}{\partial \pi_{ij}} = K_j^{(1)} - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)} = \max\left(a_j r_j + b_j, 0\right) - [1 + r_D(\omega_i)]\ \mathcal{K}_j^{(0)}.$$

Let us introduce

$$h_{i2}(r_j) = \frac{\partial P_i^{(1)}}{\partial \pi_{ij}}.$$

Thus,

$$\frac{\partial \mathbb{E}\left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \pi_{ij}}$$
$$= \mathbb{E}\left[\frac{\partial u_i \circ v\left(P_i^{(1)}\right)}{\partial \pi_{ij}}\right]$$
$$= \int_{r_1=-\infty}^{+\infty} \dots \int_{r_j} \dots \int_{r_n} h_{i1}(r_1, \dots, r_j, \dots, r_n)\ h_{i2}(r_j) f_R(r_1, \dots, r_n)\ dr_n \dots dr_j \dots dr_1$$
$$= \int_{r_j=-\infty}^{+\infty} \left[\int_{r_1} \dots \int_{r_{j-1}} \int_{r_{j+1}} \dots \int_{r_n} h_{i1}(r_1, \dots, r_j, \dots, r_n) h_{i2}(r_j) f_R(r_1, \dots, r_n)\ dr_n \dots dr_{j+1}\ dr_{j-1} \dots dr_1\right]$$
$$= \int_{r_j=-\infty}^{+\infty} h_{i2}(r_j) \left[\int_{r_1} \dots \int_{r_{j-1}} \int_{r_{j+1}} \dots \int_{r_n} h_{i1}(r_1, \dots, r_j, \dots, r_n) f_R(r_1, \dots, r_n)\ dr_n \dots dr_{j+1}\ dr_{j-1} \dots dr_1\right]$$
$$= \int_{-\infty}^{+\infty} h_{i2}(r_j) w(r_j)\ dr_j$$
$$= \int_{-\frac{b_j}{a_j}}^{+\infty} (a_j r_j + b_j) w(r_j)\ dr_j - \int_{-\infty}^{+\infty} [1 + r_D(\omega_i)] \mathcal{K}_j^{(0)} w(r_j)\ dr_j$$
$$= \int_{-\frac{b_j}{a_j}}^{+\infty} (a_j r_j + b_j) w(r_j)\ dr_j - [1 + r_D(\omega_i)] \mathcal{K}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\ dr_j,$$

where

$$w(r_j)$$
$$= \int_{r_1=-\infty}^{+\infty} \dots \int_{r_{j-1}} \int_{r_{j+1}} \dots \int_{r_n} h_{i1}(r_1, \dots, r_j, \dots, r_n)\ f_R(r_1, \dots, r_n)\ dr_n \dots dr_{j+1}\ dr_{j-1} \dots dr_1.$$

Therefore,

$$\frac{\partial \mathbb{E}\left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \pi_{ij}} > 0 \Longleftrightarrow \int_{-\frac{b_j}{a_j}}^{+\infty} (a_j r_j + b_j) w(r_j)\ dr_j > [1 + r_D(\omega_i)] \mathcal{K}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\ dr_j.$$

Let us now consider the case of γ_{ij}. As in the previous case, the corresponding derivative is written

$$\frac{\partial \left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \gamma_{ij}} = \frac{\partial (u\ o\ v)}{\partial P_i^{(1)}} \frac{\partial P_i^{(1)}}{\partial \gamma_{ij}}.$$

The first term is equal to $h_{i1}(r_j)$ and the second is denoted $h_{i3}(r_j)$. The same computation as in the case of π_{ij} yields

$$\frac{\partial \mathbb{E}\left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \gamma_{ij}} = \int_{-\infty}^{+\infty} h_{i3}(r_j) w(r_j)\ dr_j.$$

We have

$$h_{i3}(r_j) = L_j^{(1)} - [1 + r_D(\omega_i)] \mathcal{L}_j^{(0)} = \min(a_j r_j + d_j, L_j^*[1 + r_D(\omega_j)]) - [1 + r_D(\omega_i)] \mathcal{L}_j^{(0)},$$

where $d_j = \kappa_j \left(Ax_j^{(0)} + A\ell_j^{(0)}(1+r_{rf}) \right)$. Finally

$$\frac{\partial \mathbb{E}\left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \gamma_{ij}} = \int_{-\infty}^{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}} (a_j r_j + d_j) w(r_j)\, dr_j + L_j^*[1+r_D(\omega_j)]) \int_{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}}^{+\infty} w(r_j)\, dr_j$$
$$- [1+r_D(\omega_i)]\mathcal{L}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\, dr_j,$$

giving that

$$\frac{\partial \mathbb{E}\left[u_i \circ v\left(P_i^{(1)}\right)\right]}{\partial \gamma_{ij}} > 0$$
$$\iff \int_{-\infty}^{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}} (a_j r_j + d_j) w(r_j)\, dr_j + L_j^*[1+r_D(\omega_j)]) \int_{\frac{L_j^*[1+r_D(\omega_j)]-d_j}{a_j}}^{+\infty} w(r_j)\, dr_j$$
$$> [1+r_D(\omega_i)]\mathcal{L}_j^{(0)} \int_{-\infty}^{+\infty} w(r_j)\, dr_j.$$

□

For Proposition 5.4

Proof. Recall that we consider the following dynamics for $Ax_i, i = 1, 2$:

$$\log\left(\frac{Ax_i^{(1)}}{Ax_i^{(0)}}\right) \sim \mathcal{N}(\mu_i, \sigma_i) \quad \text{i.e. } Ax_i^{(1)} = Ax_i^{(0)} e^{\mu_i + \sigma_i U}, \text{ where } U \sim \mathcal{N}(0,1).$$

We have $K_i^{(1)} = \max(\kappa_i Ax_i^{(1)} - L_i^{*(1)}, 0)$ and $L_i^{(1)} = \min(\kappa_i Ax_i^{(1)}, L_i^{*(1)})$.
We define $\tilde{u}$ such that $\kappa_i Ax_i^{(0)} e^{\mu_i + \sigma_i \tilde{u}} = L_i^{*(1)}$, i.e. $\tilde{u} = \frac{1}{\sigma_i}\left(\log\left(\frac{L_i^{*(1)}}{\kappa_i Ax_i^{(0)}}\right) - \mu_i\right)$. Let us denote by ϕ the density of the standard Gaussian variable. We have

$$\begin{aligned}
\mathbb{E}_0\left(K_i^{(1)}\right) &= \mathbb{E}_0\left[\max(\kappa_i Ax_i^{(1)} - L_i^{*(1)}, 0)\right] \\
&= \int_{\tilde{u}}^{+\infty} \left(\kappa_i Ax_i^{(0)} e^{\mu_i + \sigma_i u} - L_i^{*(1)}\right) \phi(u)\, du \\
&= \kappa_i Ax_i^{(0)} e^{\mu_i} \int_{\tilde{u}}^{+\infty} e^{\sigma_i u} \phi(u)\, du - L_i^{*(1)} \int_{\tilde{u}}^{+\infty} \phi(u)\, du \\
&= \kappa_i Ax_i^{(0)} e^{\mu_i} \int_{\tilde{u}}^{+\infty} e^{\sigma_i u} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du - L_i^{*(1)} \left[1 - \Phi(\tilde{u})\right] \\
&= \kappa_i Ax_i^{(0)} e^{\mu_i} \int_{\tilde{u}}^{+\infty} e^{\frac{1}{2}\sigma_i^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-\sigma_i)^2} du - L_i^{*(1)} \left[1 - \Phi(\tilde{u})\right] \\
&= \kappa_i Ax_i^{(0)} e^{\mu_i + \frac{1}{2}\sigma_i^2} \int_{\tilde{u}-\sigma_i}^{+\infty} \phi(v)\, dv - L_i^{*(1)} \left[1 - \Phi(\tilde{u})\right] \text{ (by the change of variable } v = u - \sigma_i) \\
&= \kappa_i Ax_i^{(0)} e^{\mu_i + \frac{1}{2}\sigma_i^2} \left[1 - \Phi(\tilde{u} - \sigma_i)\right] - L_i^{*(1)} \left[1 - \Phi(\tilde{u})\right].
\end{aligned}$$

In the same way,

$$\begin{aligned}
\mathbb{E}_0\left(L_i^{(1)}\right) &= \mathbb{E}_0\left[\min(\kappa_i A x_i^{(1)}, L_i^{*(1)})\right] \\
&= \int_{-\infty}^{\tilde{u}} \kappa_i A x_i^{(0)} e^{\mu_i+\sigma_i u}\phi(u)\ du + \int_{\tilde{u}}^{+\infty} L_i^{*(1)}\phi(u)\ du \\
&= \kappa_i A x_i^{(0)} e^{\mu_i+\frac{1}{2}\sigma_i^2}\int_{-\infty}^{\tilde{u}-\sigma_i}\phi(v)\ dv + L_i^{*(1)}\left[1-\Phi(\tilde{u})\right] \\
&\quad \text{(using the same trick than in the case of } K_i^{(1)}\text{)} \\
&= \kappa_i A x_i^{(0)} e^{\mu_i+\frac{1}{2}\sigma_i^2}\Phi\left(\tilde{u}-\sigma_i\right) + L_i^{*(1)}\left[1-\Phi(\tilde{u})\right].
\end{aligned}$$

□

5.10 Appendix: Algorithm of network formation

In the case of 2 institutions ($n = 2$), the algorithm of network formation is the following:

1. **Optimization for Bank 1 without interconnections**. Indeed in this first step, $K_2^{(0)}$ and $L_2^{(0)}$ are not known.
   We then have to optimize $E\left\{u_1\left[v\left(P_1^{(1)}(Ax_1^{(0)}, L_1^{(0)})\right)\right]\right\}$, where

   $$P_1^{(1)} = Ax_1^{(0)}(1+r_1) - [1+r_{D,1}]L_1^{(0)}.$$

   This step provides $Ax_1^{(0)}$ and $L_1^{(0)}$.

2. **Optimization for Bank 2 with interconnections**. We have

   $$\begin{aligned}
   P_2^{(1)} &= Ax_2^{(0)}(1+r_2) + \pi_{2,1}\max\left(\kappa_1 Ax_1^{(0)}(1+r_1) - L_1^{(0)}\ [1+r_{D,1}], 0\right) \\
   &+ \gamma_{2,1}\min\left(\kappa_1 Ax_1^{(0)}(1+r_1), L_1^{(0)}[1+r_{D,1}]\right) - [1+r_{D,2}]L_2^{(0)},
   \end{aligned}$$

   where $\kappa_1 = \dfrac{L_1^{(0)}+K_1^{(0)}}{Ax_1^{(0)}}$ is the scaling factor compensating the absence of interconnections (it keeps the balance sheet of Bank 1 balanced). Since $K_1^{(0)}$ has been obtained at step 1 under the assumption that Bank 1 is not interconnected, here $\kappa_1 = 1$. But this will be corrected in further iterations.
   This step gives $Ax_2^{(0)}, L_2^{(0)}, \pi_{2,1}$ and $\gamma_{2,1}$.

3. **Optimization for Bank 1 with interconnections**. We have

   $$\begin{aligned}
   P_1^{(1)} &= Ax_1^{(0)}(1+r_1) + \pi_{1,2}\max\left(\kappa_2 Ax_2^{(0)}(1+r_2) - L_2^{(0)}[1+r_{D,2}], 0\right) \\
   &+ \gamma_{1,2}\ \min\left(\kappa_2 Ax_2^{(0)}(1+r_2), L_2^{(0)}[1+r_{D,2})]\right) - [1+r_{D,1}]L_1^{(0)}
   \end{aligned}$$

   where $\kappa_2 = \dfrac{L_2^{(0)}+K_2^{(0)}}{Ax_2^{(0)}}$. This step provides $Ax_1^{(0)}, L_1^{(0)}, \pi_{1,2}$ and $\gamma_{1,2}$.

4. **New optimization for Bank 2 with interconnections**.
   Note that at this step, $\kappa_1 = \dfrac{L_1^{(0)} + K_1^{(0)}}{Ax_1^{(0)}} > 1$, since at the previous step, the optimization has be done for Bank 1 being interconnected.

5. **New optimization for Bank 1 with interconnections**.

Further iterations can be carried out if the variation in the estimates from one step to the next is higher than a predefined threshold.

5.11 Appendix: Calibration of external assets returns

Given the values of the mean net returns and the probability of default, let us derive the corresponding values of μ_i and σ_i, for $i = 1, 2$. We denote by GR_i and NR_i the gross and the net return of Bank i, respectively. They satisfy the relationship $NR_i = GR_i - 1$. Thus, since the gross returns are log-normal,

$$\mathbb{E}(NR_i) = \mathbb{E}(GR_i) - 1 = \exp\left(\mu_i + \frac{\sigma_i^2}{2}\right) - 1.$$

If we denote by m_i the empirical mean of the net return, we then have

$$m_i = \exp\left(\mu_i + \frac{\sigma_i^2}{2}\right) - 1,$$

that gives

$$\mu_i = \log(1 + m_i) - \frac{\sigma_i^2}{2}. \tag{5.20}$$

We need a second equation to find μ_i and σ_i. We could use the expression

$$\mathrm{Var}(NR_i) = \mathrm{Var}(GR_i) = (\exp(\sigma_i^2) - 1)\ \exp(2\mu_i + \sigma_i^2)$$

i.e., by denoting v_i the empirical variance of RN_i,

$$v_i = (\exp(\sigma_i^2) - 1)\ \exp(2\mu_i + \sigma_i^2).$$

However, it is difficult to find reliable values for v_i. If we consider banks' data, only one return is available per year and thus the estimation of the variance is inaccurate. Another possibility is to compute the variance of the net returns of an index like the CAC 40. However, such an index is not representative of the external assets of a financial institution since it only contains shares. Moreover, it does not take the hedging strategy of the institution into account.

Therefore, we choose to derive the needed equation from the probability of default. This quantity is indeed easier to obtain. Actually, the usual rating for large banks corresponds to a probability of default of about 0.1%. Considering an autarkic stylized bank with debt L_i and a total asset A_i, whose gross returns are log-normal of parameter (μ_i, σ_i), the probability of default is

$$PD = \Phi\left(\frac{\log\left(\dfrac{L_i}{A_i}\right) - \mu_i}{\sigma_i}\right). \tag{5.21}$$

Using in (5.21) the expression of μ_i in (5.20), we obtain

$$PD = \Phi\left(\frac{\log\left(\frac{L_i}{A_i}\right) - \log(1+m_i) + \frac{\sigma_i^2}{2}}{\sigma_i}\right) = \Phi\left(\frac{\log\left(\frac{L_i}{A_i(1+m_i)}\right) + \frac{\sigma_i^2}{2}}{\sigma_i}\right).$$

If we denote by p the empirical probability of default, the equation to solve is

$$p = \Phi\left(\frac{\log\left(\frac{L_i}{A_i(1+m_i)}\right) + \frac{\sigma_i^2}{2}}{\sigma_i}\right) \iff \frac{\sigma_i^2}{2} - \sigma_i\, \Phi^{-1}(p) + \log\left(\frac{L_i}{A_i(1+m_i)}\right) = 0.$$

This is a quadratic equation with discriminant $\Delta = [\Phi^{-1}(p)]^2 - 2\log\left(\frac{L_i}{A_i(1+m_i)}\right)$. With chosen values of A_i, L_i and m_i, we know that $\Delta > 0$ and thus $\sigma_i = \Phi^{-1}(p) + \sqrt{\Delta}$, since the other solution is strictly negative and thus unsuitable for a volatility. Finally the implied volatility is written

$$\sigma_i = \Phi^{-1}(p) + \sqrt{[\Phi^{-1}(p)]^2 - 2\log\left(\frac{L_i}{A_i(1+m_i)}\right)}. \tag{5.22}$$

We then obtain μ_i using (5.20).

5.12 Appendix: Algorithm of equilibrium computation

The computation of the equilibrium involving n financial institutions requires to solve up to 2^n linear systems with a brutal force approach (see Gouriéroux et al. (2012) for details), implying a total complexity in $O(n^3 \times 2^n)$. The cubic term stems from the resolution of a linear system that requires to invert a $n \times n$ matrix. Only a little gain can be obtained on this term. The exponential term comes from testing each possible situation: each institution is either solvent or in default.

Instead, in order to deal with the exponential term, we adopt an heuristic algorithm. The key idea is to test the 2^n potential regimes in a "proper" order and to use the existence and uniqueness property to stop the algorithm as soon as one feasible solution is computed. Since interconnections are small, the way of sorting the regimes relies on the situation without interconnections.

To do so, let us define Regime r by $\mathbf{d}^r = (d_1^r, \ldots, d_n^r)'$, where $d_i^r = -1$ if Institution i is in default and 1 otherwise (for $i = 1, \ldots, n$). We define a weight vector $\mathbf{w} = (w_1, \ldots, w_n)$, where $w_i = (Ax_i + A\ell_i - L_i^*)/L_i^*$ (for $i = 1, \ldots, n$). Note that $\mathbf{w}$ depends on some known inputs only and can therefore be easily computed. When w_i is positive, the external assets of financial institution i are higher than its nominal debt. Therefore, whatever the situations of other financial institutions, Institution i is always solvent at the equilibrium. On the contrary, when w_i is negative, the financial institution needs a sufficient amount of inter-financial assets to be solvent. In that case, since interconnections are assumed to be small, the (absolute) value of w_i indicates the likelihood (in a *non-statistical* sense)

of default of Institution i. One can associate to Regime r a score, given by $\mathbf{w}.\mathbf{d}^r$, which measures the likelihood of Regime r. For instance, if $\mathbf{w}$ contains many negative values, we might think that the equilibrium lies in a regime with a lot of institutions in default. Thus, a regime with many values of d_i^r equal to -1 will be likely and will be associated to a high score.

Actually, the regime with the highest score can easily be derived from $\mathbf{w}$. This regime, labeled $\underline{r}$, is defined by $d_i^{\underline{r}} = \mathbf{I}_{\{w_i>0\}} - \mathbf{I}_{\{w_i\leq 0\}}$, for $i = 1, \ldots, n$. If $w_i \leq 0$, it is likely that Institution i is in default and thus we set $d_i^{\underline{r}} = -1$. The contrary is true when $w_i > 0$. We test this most likely regime. If it corresponds to the solution, we have finished. If not, one can switch the components of $\mathbf{d}^{\underline{r}}$ one by one to get new regimes with high scores. It is important to keep in mind that assuming the default of an institution with positive weight is dead-end. While no solution was found, this mechanism of building new regimes can be carried on until having sorted all the potential regimes apart from the ones for which there exists i such that $w_i > 0$ and $d_i = -1$.

The complexity (in the worst case) of this algorithm is in $O(n^3 \times 2^{n-p})$, where $p = \#\{i : w_i > 0\}$, with $\#$ standing for the cardinal. Thus, we still have an exponential term. However, the expectation of the number of regimes to be tested before finding the solution is much lower than in the case of the brutal force approach. The algorithm performs well in practice. For example, with 10 financial institutions having log-normal returns and random interconnections, the equilibrium lies in the 10 first tested regimes in almost all cases.

NB: If one remains concerned by exploring all the regimes (implying keeping the exponential term in the complexity), one solution is to stop the search after an arbitrary number of regimes (for instance n). When the exploration approach is stopped, a pure numerical approach can be carried out, in order to solve the system $(\mathbf{K}, \mathbf{L})' = q[(\mathbf{K}, \mathbf{L})']$, where the function q is defined using Equations 5.1 and 5.2. For instance, routines to minimize or to find the zeros of $k[(\mathbf{K}, \mathbf{L})'] = q[(\mathbf{K}, \mathbf{L})'] - (\mathbf{K}, \mathbf{L})'$ can be used.

Bibliography

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2013). Systemic risk and stability in financial networks. *National Bureau of Economic Research Working Paper Series*, 18727.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Alves, I., Ferrari, S., Franchini, P., Héam, J.-C., Jurca, P., Langfield, S., Laviola, S., Liedorp, F., Sanchez, A., Tavolaro, S., and Vuillemy, G. (2013). The structure and resilience of the European interbank market. *European Systemic Risk Board (ESRB) Occasional Paper Series*, 3.

Anand, K., Gai, P., Kapadia, S., Brennan, S., and Willison, M. (2013). A network model of financial system resilience. *Journal of Economic Behavior and Organization*, 85:219–235.

Arinaminpathy, N., Kapadia, S., and May, R. M. (2012). Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45):18338–18343.

Asmussen, S. and Albrecher, H. (2010). *Ruin probabilities*, volume 14. World Scientific.

Babus, A. (2007). The formation of financial networks. *Fondazione Eni Enrico Mattei Working Paper Series*, 129.

BCBS (2011). Basel III: A global regulatory framework for more resilient banks and banking systems.

BCBS (2013). Basel III: The liquidity coverage ratio and liquidity risk monitoring tools.

BCBS (2014). Supervisory framework for measuring and controlling large exposures.

Bluhm, M., Faia, E., and Krahnen, J. P. (2013). Endogenous banks' networks, cascades and systemic risk. *SAFE Working Paper Series*, 12.

CEC (1979). First council directive of 5 march 1979 on the coordination of laws, regulations and administrative provisions relating to the taking up and pursuit of the business of direct life assurance. *Official Journal of the European Communities*, 79/267/EEC.

Cifuentes, R., Ferrucci, G., and Shin, H. S. (2005). Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2):556–566.

Cohen-Cole, E., Patacchini, E., and Zenou, Y. (2011). Systemic risk and network formation in the interbank market. *CEPR Discussion Paper Series*, 8332.

Craig, B. R. and Von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347.

Degryse, H. and Nguyen, G. (2007). Interbank exposures: An empirical examination of contagion risk in the Belgian banking system. *International Journal of Central Banking*, 3(2):123–171.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *The journal of political economy*, 91(3):401–419.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–3153.

Farboodi, M. (2014). Intermediation and voluntary exposure to counterparty risk. *Chicago Booth School of Business, Mimeo*.

Fourel, V., Héam, J.-C., Salakhova, D., and Tavolaro (2013). Domino effects when banks hoard liquidity: the French network. *Banque de France Working Paper Series*, 432.

Furfine, C. H. (2003). Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking*, 35(1):111–128.

Gauthier, C., Lehar, A., and Souissi, M. (2012). Macroprudential capital requirements and systemic risk. *Journal of Financial Intermediation*, 21(4):594–618.

Georg, C.-P. (2014). Contagious herding and endogenous network formation in financial networks. *ECB Working Paper Series*, 1700.

Gofman, M. (2012). A network-based analysis of over-the-counter markets. *AFA 2012 Chicago Meetings Paper*.

Gouriéroux, C., Héam, J.-C., and Monfort, A. (2012). Bilateral exposures and systemic solvency risk. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(4):1273–1309.

Gouriéroux, C., Héam, J.-C., and Monfort, A. (2013). Liquidation equilibrium with seniority and hidden CDO. *Journal of Banking and Finance*, 37(12):5261–5274.

Goyal, S. (2012). *Connections: an introduction to the economics of networks*. Princeton University Press.

Holmstrom, B. and Tirole, J. (1996). Private and public supply of liquidity. *National Bureau of Economic Research Working Paper Series*, 5817.

Jackson, M. and Zenou, Y. (2013). *Economic Analyses of Social Networks*. The International Library of Critical Writings in Economics, London: Edward Elgar Publishing.

Karush, W. (1939). *Minima of functions of several variables with inequalities as side constraints*. PhD thesis, Master's thesis, Dept. of Mathematics, Univ. of Chicago.

Kuhn, H. and Tucker, A. (1951). Proceedings of 2nd berkeley symposium.

Lublóy, Á. (2005). Domino effect in the Hungarian interbank market. *Hungarian Economic Review*, 52(4):377–401.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates*. *The Journal of Finance*, 29(2):449–470.

Mistrulli, P. E. (2011). Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. *Journal of Banking and Finance*, 35(5):1114–1127.

Repullo, R. and Suarez, J. (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies*, 26(2):452–490.

Rochet, J.-C. (2004). Macroeconomic shocks and banking supervision. *Journal of Financial Stability*, 1(1):93–110.

Rochet, J.-C. and Tirole, J. (1996). Controlling risk in payment systems. *Journal of Money, Credit and Banking*, 28(4):832–862.

Tirole, J. (2010). *The theory of corporate finance*. Princeton University Press.

Toivanen, M. (2009). Financial interlinkages and risk of contagion in the Finnish interbank market. *Bank of Finland Research Discussion Paper Series*, 6.

Upper, C. and Worms, A. (2004). Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review*, 48(4):827–849.

van Lelyveld, I. and Liedorp, F. (2006). Interbank contagion in the Dutch banking sector: A sensitivity analysis. *International Journal of Central Banking*, 2(2):99–133.

Wells, S. (2002). UK interbank exposures: systemic risk implications. *Financial Stability Review*, 13(12):175–182.

Chapter 6

Conclusion

6.1 Proposal introduction

As a conclusion, I provide a brief research proposal for the future, entitled: how to manage human and economic impacts of natural disasters? Typhoon Haiyan which recently hit the Philippines unfortunately reminds everyone of the potentially devastating effects of such extreme events on both populations and economy. Moreover, in a climate change context, some extremes tend to be more frequent, according to the IPCC. The aim of my project is first to model natural disasters, second their human and economic impacts and third to find the best combination of solutions to respond to them. The points mentioned here concern extensions and improvements of what has been described in this thesis.

As to modeling natural disasters I propose a method that allows accounting for spatial extent, nonstationarity and temporal dependence of extremes of environmental variables. An approach to model precipitation and earthquakes is provided. These issues are dealt with in Section 6.2.

Concerning the modeling of their impacts, some aspects are given in Section 6.3.

As to the response, insurance constitutes a first type of solution. Insurance companies transfer a part of their underwritten risk to reinsurers.

However, the insurance/reinsurance system capacity can be exceeded in case of major events (Litzenberger et al., 1996). Some financial securities have therefore been designed to transfer part of the extreme risk to the financial markets. I propose to specify the combination of risk transfer strategies that best protects policyholders and more generally the whole economy against natural disasters. A fair pricing of previously introduced securities is proposed, which is an important issue to increase their development. Insurance solutions for developing countries (e.g. micro-insurance) will also be considered. Koch (2010) proposes an insurance for Moroccan farmers based on climatic indices. A second type of solution is risk reduction, involving adaptation and prevention. The best compromise between insurance and risk reduction has to be defined. Related issues are detailed in Section 6.4.

6.2 Modeling of the environmental variables

6.2.1 Spatial extent, nonstationnarity and dependence in extremes

In order to model the statistical behavior of natural disasters, it is of prime interest to model the maxima of environmental variables. As explained before, due to the spatial

extent of these, max-stable processes are ideally suited to attack this problem. Max-stable processes completely account for the spatial dependence structure and also have the potential to model dependence across time.

However, only a few papers relax the assumptions of temporal independence and stationarity. In a time series context, Davis and Resnick (1989) consider max-autoregressive moving averages (MARMA) processes which are examples of general max-stable processes. Davis et al. (1993), Zhang and Smith (2010) and Naveau et al. (2011) have a similar approach. Davis et al. (2013) propose an inference method for max-stable processes in space and time. Inspired by this strand of literature, our aim in this part of the project is to find new max-stable models accounting for other types of dependence. For example, it could be interesting to study the behavior of the process: $Z_t(\mathbf{x}) = \max[Z_{t-1}(\mathbf{x}), W(\mathbf{x})]$ where W is a known process. For instance W can be a Smith process or a Schlather process. The idea is to find the stationary distribution of Z with respect to the law of W. Here the max must be considered as an operator to be built: it is not necessarily the maximum of the two processes at the same point $\mathbf{x}$. Previously mentioned models and methods will be applied to environmental data, leading to a better understanding of their temporal structure.

6.2.2 Spatio-temporal modeling of precipitation

Whereas previously introduced models in the literature deal with daily precipitation, Chapter 2 proposes a multi-site hourly precipitation generator. This model involves a common factor and a contagion term. The common factor appears in the noise volatility and involves linear combinations of covariates like humidity or pressure.

As a natural extension, we propose to consider nonparametric functions of the covariates, thus accounting for nonlinear features of the relationship between covariates and precipitation. The issue of interpolation at locations at which observations are unavailable will also be addressed. Last but not least, by plugging the outputs of Global Climate Models into the model, the evolution of precipitation in a climate change context will be studied. Variable selection and analysis of the probabilistic properties of the model are also of interest.

6.2.3 Modeling earthquakes using MINAR processes

Boudreault and Charpentier (2011) use a bivariate Integer-Valued Autoregressive process of order 1, INAR(1), to model earthquakes counts. They consider a full matrix in the thinning operator (Steutel and Van Harn, 1979) which allows modeling contagion between pairs of tectonic plates. However, they only take into account the contagion from one plate and do not model the magnitude. We here propose to consider a Multivariate INAR (MINAR)(1) of any dimension and to jointly model counts and magnitudes, obtaining potentially prolific results for applications in risk management. It would also be an advance from a theoretical point of view since MINAR processes have never been considered in dimension greater than 2.

6.3 Spatial risk measures and impact modeling

Chapter 4 introduces the concept of spatial risk measure and proposes an axiomatic approach adapted to the spatial context. However, the examples of such risk measures that

are developed are based on spatially stationary max-stable processes and the exposure is uniformly equal to unity. In this project, more realistic margins will be taken into account, the stationarity assumption will be relaxed, other processes will be considered and a more realistic exposure will be introduced.

Considering several risks simultaneously (e.g. due to wind and precipitation) would also be of major interest. Note that precipitation and earthquakes could be modeled according to Sections 6.2.2 and 6.2.3.

6.4 Optimal risk transfer

6.4.1 Systemic risk in insurance and reinsurance

Hurricane Andrew which hit Florida in 1992 caused the bankruptcy of 11 insurance companies and left 930 thousand policyholders without any cover options. We propose in this part to analyse how a natural disaster can affect the whole economy by contagion. This problem of systemic risk in insurance/reinsurance has been hardly addressed in the literature so far.

We will first focus on the insurance/reinsurance perimeter. In the case of a low reinsurance rate, the climatic shock (systematic risk) would first weaken each insurance company and even cause the bankruptcy of some of them. Secondly, due to the links between insurance companies (in terms of cross debt and shareholdings) there would be contagion effects: the previous bankruptcies could trigger a vicious circle of new bankruptcies. Thus reinsurance companies would be systemically risky due to a lack of coverage. In the case of high reinsurance rates, the capacity of reinsurers could be exceeded, generating defaults. Insurance companies could also go bankrupt through contagion. These two cases suggest the probable existence of an optimal cover rate (for the entire system). Bernard and Tian (2009) show that insurers have no incentives to protect themselves against extreme losses under regulatory requirements based on tail risk measures. Using this strand of literature as well as network theory (see the approach by Blanchet and Shi (2012)), we propose to specify a regulation minimizing systemic risk. The model developed in Chapter 5 will be extended to the case of insurance companies, allowing to analyse how the long term interconnections are modified by a natural disaster, and therefore to quantify the long term systemic consequences of a natural disaster.

However, systemic risk originating from the insurance/reinsurance sector can extend to a wider perimeter due to high interconnections between sectors (Cummins and Weiss, 2010; Acharya, 2009; Billio et al., 2010). Billio et al. (2010) conclude that a liquidity shock to one sector propagates to other sectors eventually triggering defaults and systemic events. Inspired by Gouriéroux et al. (2012), we propose in Chapter 5 a model of endogenous financial network formation, which we will use to analyse the systemic features of the insurance/reinsurance sector including all types of financial institutions. In order to investigate the potential spread of a lack of liquidity from the insurance/reinsurance sector to the banking sector and the global economy we will extend the study developed in Chapter 5 to the case of short term interconnections (liquidity exchanges).

6.4.2 CAT bonds pricing

As explained in Chapter 1, CAT bonds are the most famous example of event-linked securities. Surprisingly, such products did not meet investors' expectations due to a lack

of confidence. A crucial issue to alleviate this problem is to find their fair price, requiring a suitable modeling of the loss. As explained by Finn and Lane (1997), losses are generally assumed to follow a particular distribution which is fitted on available data. Nevertheless, this strategy does not account for changes in factors driving the losses amount. Here we propose to model physical hazards and demographic and economic factors separately as in Section 6.3. This allows including future trends in each factor and not arbitrarily in the loss distribution parameters. The issue of creating more adapted securities will also be addressed and finally the analysis described in Section 6.4.1 will be carried out including such products. This allows determining the best combination of transfer solutions. The latter problem cannot be studied independently of risk reduction.

Bibliography

Acharya, V. V. (2009). *On the financial regulation of insurance companies*. PhD thesis, London Business School.

Bernard, C. and Tian, W. (2009). Optimal reinsurance arrangements under tail risk measures. *Journal of risk and insurance*, 76(3):709–725.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2010). Econometric measures of systemic risk in the finance and insurance sectors. Technical report, National Bureau of Economic Research.

Blanchet, J. and Shi, Y. (2012). Stochastic risk networks: Modeling, analysis and efficient Monte Carlo. *Analysis and Efficient Monte Carlo (February 29, 2012)*.

Boudreault, M. and Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv preprint arXiv:1112.0929*.

Cummins, J. D. and Weiss, M. A. (2010). Systemic risk and the US insurance sector. *Temple University, September*, 14.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5):791–819.

Davis, R. A. and Resnick, S. I. (1989). Basic properties and prediction of max-ARMA processes. *Advances in applied probability*, 21(4):781–803.

Davis, R. A., Resnick, S. I., et al. (1993). Prediction of stationary max-stable processes. *The Annals of Applied Probability*, 3(2):497–525.

Finn, J. and Lane, M. (1997). The perfume of the premium... or pricing insurance derivatives. In *Proceedings of the 1995 Bowles Symposium on Securitization of Risk*, pages 27–35.

Gouriéroux, C., Héam, J.-C., and Monfort, A. (2012). Bilateral exposures and systemic solvency risk. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(4):1273–1309.

Koch, E. (2010). Etude de faisabilité d'une assurance rendement basée sur indice climatique.

Litzenberger, R. H., Beaglehole, D. R., and Reynolds, C. E. (1996). Assessing catastrophe reinsurance-linked securities as a new asset class. *Journal of Portfolio Management*, 23:76–86.

Naveau, P., Zhang, Z., and Zhu, B. (2011). An extension of max autoregressive models. *Statistics and Its Interface*, 4:253–266.

Steutel, F. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, pages 893–899.

Zhang, Z. and Smith, R. L. (2010). On the estimation and application of max-stable processes. *Journal of Statistical Planning and Inference*, 140(5):1135–1153.