



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC

*_*_*_*_*

HAUT COMMISSARIAT AU PLAN

*_*_*_*_*_*_*_*

INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

INSEA



Projet de Fin d'Etudes

**Etude d'un produit d'assurance
paramétrique contre le risque
d'inondation au Maroc**

Préparé par : **M. Saïd IDIR**

Sous la direction de : **M. Khalil SAID (INSEA)**
M. Abderrahim OULIDI (FSEC)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Filière : Actuariat-Finance

Devant le jury composé de :

- **M. Khalil SAID (INSEA)**
- **M. Driss EFFINA (INSEA)**
- **M. Abderrahim OULIDI (FSEC)**

Juin 2021 / PFE N° 19

Dédicace

“

*À ma mère et mon père qui m'ont accompagné et soutenu
tout au long de mon parcours, je leur dédie ce modeste
travail pour leurs prières tout au long de ma construction
et pour leurs conseils qui continuent et continueront à me
guider,*

*À mes chers frères et soeurs qui m'ont toujours soutenu et
encouragé, Je vous souhaite le succès et le bonheur dans
votre vie,*

À tous les membres de ma grande famille,

À tous mes chers amis de l'INSEA et d'ailleurs,

Merci.

”

- Said

Remerciements

En préambule à ce mémoire, je souhaiterais adresser mes remerciements les plus sincères à toutes les personnes qui ont pris part à ce projet, à celles qui m'ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire.

Je tiens à remercier tout particulièrement mon encadrant externe **M. OULIDI Abderrahim** à qui je dédie mes plus profonds respect et estime pour l'aide qu'il m'a apporté, pour son encadrement, sa disponibilité, ses précieux conseils et son encouragement tout au long de mon stage.

Je tiens à remercier également mon professeur et encadrant interne **M. SAID Khalil** pour sa disponibilité, la qualité de son suivi ainsi que pour tous ses conseils.

Je tiens aussi à adresser mes plus sincères remerciements à **M. Chaffai Abderrahim**, directeur de FSEC pour m'avoir accueilli dans son honorable organisme et pour son encouragement.

Je tiens également à remercier **M. Effina Driss** pour avoir accepté d'examiner travail.

Mes remerciements s'adressent également au corps professoral et au corps administratif de l'INSEA pour leurs grands efforts et pour avoir contribué à notre formation.

Merci à tous et à toutes.

Résumé

Suite à des évènements catastrophiques qu'a connu Maroc lors de la dernière décennie, le Fonds de Solidarité contre les Evènements Catastrophiques (FSEC) a vu le jour. La mission principale du FSEC consiste à intervenir pour indemniser toute personne touchée par une catastrophe et ne disposant pas d'une assurance privée.

Les ressources de FSEC proviennent principalement d'une taxe parafiscale instituée par voie réglementaire. Ces ressources peuvent rapidement s'avérer très insuffisantes si un séisme de l'ampleur de celui qui a touché par exemple en 2004 la région d'AL Hoceima ou une inondation comme celle qui a touché la ville de Mohammedia en novembre 2002. L'amélioration de la connaissance et de la modélisation des risques naturels, l'élaboration de stratégies optimales et efficaces de transferts de ces risques, demeurent une mesure incontournable et une action prioritaire.

Pour remplir sa mission conformément aux dispositions de la loi et assurer sa résilience aux risques, Le FSEC a ainsi fait recours au transfert de risque via une solution paramétrique qui couvre les pertes causées par les inondations. Ce produit d'assurance paramétrique de couverture du risque de pluie extrême se base sur la détection des pluies extrêmes en fonction d'un indice agrégeant les précipitations sur l'ensemble des provinces du royaume. Ce produit fera l'objet d'étude de ce mémoire.

A l'inverse d'une assurance classique, dans le contexte d'une assurance paramétrique, les indemnités ne sont pas reversées en fonction des pertes réelles mais selon la mesure d'un indice paramétrique défini dans le contrat. L'objectif est de définir l'indice et caractériser la distribution de probabilité de cet indice paramétrique afin de pouvoir tarifier le produit.

Mots clés : Assurance Paramétrique, Extreme Value Theory, Generalized Pareto Distribution, Inondations, Loi 110-14.

Note de synthèse

Le Maroc est fortement exposé aux catastrophes naturelles comme les inondations. Ces Évènements peuvent menacer l'économie et la population d'une nation entière. En effet selon le Global Climate Change Risk Index 2016 la perte annuelle moyenne estimée causée par les catastrophes naturelles est d'au moins 7,8 milliards MAD, dont environ 60% des pertes sont causées par les inondations). Afin de se prémunir contre ce risque, le Maroc a abordé les risques des catastrophes naturelles dans le cadre d'un **projet** coordonné, inclusif, participatif impliquant toutes les parties prenantes concernées. Ce projet a permis à la mise en place de deux régimes de couverture des risques catastrophique : EVCAT une couverture assurantielle et le FSEC pour couvrir les victimes de catastrophes non assurées.

Les ressources de FSEC proviennent principalement d'une taxe parafiscale instituée par voie réglementaire (Loi 110-14). Ces ressources peuvent rapidement s'avérer très insuffisantes pour faire face à une inondation comme celle qui a touché la ville de Mohammadia en novembre 2002. Le FSEC a ainsi fait recours au transfert de risque via une solution paramétrique qui couvre les pertes causées par les inondations. L'assurance paramétrique présente plusieurs avantages qui la rend un instrument clé de transfert de risque catastrophique, en effet elle permet de supprimer l'alea morale et le problème d'antisélection, et elle assure la rapidité des paiements. Notre étude se limitera à l'analyse de risque dans les provinces les plus exposées, à savoir Tanger, Tétouan, Kénitra, Salé, Rabat, Mohammadia, Casablanca.

1. Base de données

Le but principal de ce projet est de tarifier la police d'assurance paramétrique contre le risque d'inondation au Maroc. Pour y arriver il faut passer par plusieurs étapes :

- Construire un indice paramétrique pour la police d'assurance paramétrique,
- Définir une fonction de paiements
- Modéliser la dynamique de l'indice
- Tarifier le produit par simulation

1.1 Données des précipitations

La construction de notre indice se basera sur les précipitations, Un historique de précipitations de 20 ans (de 2000 jusqu'à 2020) issue de de produit GPM (GPM - Global Precipitation Measurement) de la NASA est utilisé dans la modélisation de la dynamique de l'indice dans chacune des provinces.

1.2 Exposition

L'impact des inondations ne dépend pas seulement des caractéristiques de l'inondations mais aussi de la valeur immobilière exposée à ce risque. La richesse exposée au risque dans les provinces sera aussi utilisée dans la construction de l'indice. Le FSEC dispose d'une base de données qui comporte l'exposition au risque dans chaque province.

2. Construction et modélisation de l'indice paramétrique

2.1 Construction de l'indice paramétrique

L'étape fondamentale dans la mise en place d'une assurance paramétrique est la construction de l'indice paramétrique. La construction de l'indice passe par 4 étapes :

1. Sur chacune des Provinces du Maroc, le cumul des 3 derniers jours de pluie est calculé jour après jour (cumul glissant). L'évènement se déclenche lorsque ces précipitations cumulées sur 3 jour dépassent 200 mm et se termine lorsqu'il redescend en-dessous de ce seuil.

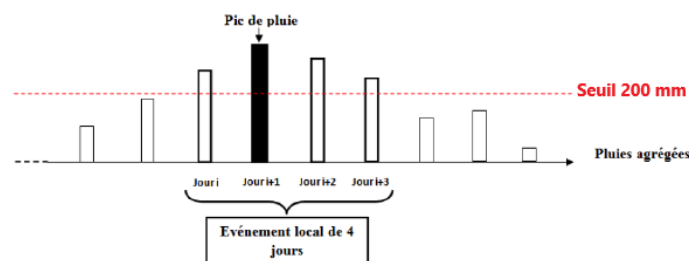


FIG. 1 : Détection des événements extrêmes

2. Sur chacun des événements détectés, seul le pic des pluies cumulées sur 3 jours est utilisé pour calculer un indice de perte via une courbe de vulnérabilité qui relie la quantité de pluie à un taux de dommage
3. Pour prendre en considération les différences en exposition entre les provinces il faut appliquer un taux d'exposition au taux de dommage.

4. L'indice nationale est la somme des indices provinciaux.

$$indice\ nationale = \sum_{i=1}^7 I_i$$

2.2 Fonction de paiement

La fonction de paiement permet de fixer les indemnités à partir de la valeur de l'indice paramétrique mesurée après la survenance de l'inondation. Notre fonction de paiement est définie comme suit :

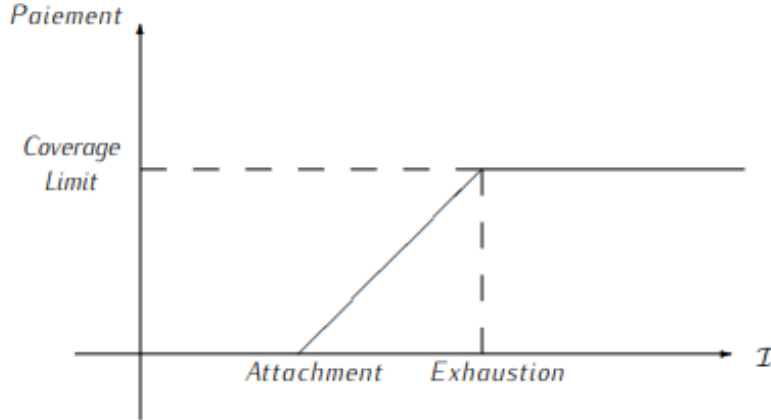


FIG. 2 : Fonction de Paiement

La borne inférieure, communément appelée Attachement est fixée de façon que tous les sinistres ayant un indice inférieur à cette borne ne soient pas pris en charge. La frontière supérieure est usuellement nommée Exhaustion. L'indemnisation est bornée supérieurement par le Coverage Limit (capacité). Cette limite est atteinte lorsque l'indice vaut Exhaustion.

$$Paiement = \begin{cases} 0 & \text{if } \mathcal{I} < Attachment \\ \frac{\mathcal{I} - Attachment}{Exhaustion - Attachment} \times Coverage\ Limit & \text{si } Attachment \leq \mathcal{I} \leq Exhaustion \\ Coverage\ Limit & \text{si } \mathcal{I} > Exhaustion \end{cases}$$

2.3 Modélisation de la dynamique de l'indice paramétrique

L'objectif est de déterminer une prime pure nationale, ceci passe par la modélisation de la dynamique des précipitations. Le choix des distributions pour ajuster la loi des précipitations est guidé par la considération que les inondations n'arrivent que si les précipitations sont extrêmes. Ainsi notre travail est axé sur l'étude des précipitations extrêmes, particulièrement le cumul des précipitations sur 3 jours. Nous faisons recours à La Théorie des Valeurs Extrêmes. C'est une branche des statistiques traitant le comportement de la distribution des maxima d'une suite de variables aléatoires, il s'agit d'un outil pratique pour évaluer les risques d'évènements rares tels que les inondations.

Cette discipline se base sur la détermination d'un certain seuil (u) à partir duquel les précipitations sont considérées extrêmes puis modéliser les dépassements de ce seuil par la Distribution de Pareto Généralisée (GPD). Cette approche est appliquée aux cumuls des précipitations dans les sept provinces.

La loi de Pareto Généralisée $GPD(\xi, \tilde{\sigma})$ est définie par la fonction de répartition :

pour $\xi \neq 0$:

$$G_{\xi, \tilde{\sigma}}(x) = (1 - (1 + \xi \frac{x}{\tilde{\sigma}})^{-1/\xi})$$

pour $\xi = 0$:

$$G_{\xi, \tilde{\sigma}}(x) = (1 - \exp(-\frac{x}{\tilde{\sigma}}))$$

Ci-dessous le QQ-plot dans le cas de Rabat

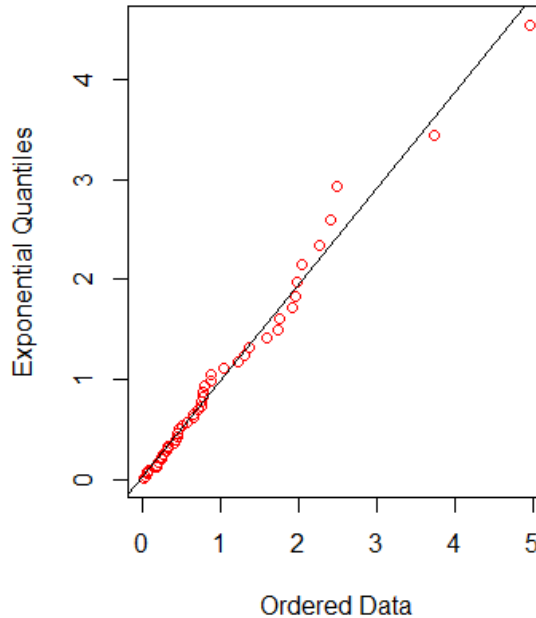


FIG. 3 : QQ-plot rabat

Cette modélisation par une loi GPD est un succès. Cela nous permet de générer des événements et de construire une base de données d'indices simulées pour procéder à la tarification.

2.4 Modélisation de la dépendance

La tarification du produit se fera par deux approches, dans un premier temps nous considérons que les indices des provinces sont indépendants (c'est une hypothèse forte vu que les provinces sont géographiquement proches), dans ce cas la prime nationale n'est que la somme des primes locales. Dans un deuxième temps nous tarifions le produit en

tenant compte de la dépendance, L'indice nationale est défini comme la somme des indices nationaux (somme de variables aléatoires). Pour déterminer la dynamique de l'indice nationale il faut passer par la modélisation de la dépendance par les copules. Après des diagnostics graphiques et tests statistiques la copule de Student et la copule Gaussienne sont retenues, pourtant ces copules présentent des limites car ils ne modélisent pas la dépendance extrême.

Le nombre de province (sept provinces) nous empêche d'utiliser les copules extrêmes (copule de Joe, copule de Gumbel) vu que ces copules sont paramétrisées par un seul paramètre, or il est imprudent de modéliser la dépendance entre sept variables par un seul paramètre. Pour y remédier nous avons divisé notre portefeuille de provinces en trois classes selon le coefficient de corrélation linéaire (voir l'annexe VI) et la proximité géographique, ainsi nous construisons les classes suivantes :

- A= {Rabat, Salé, Kenitra}
- B= {Casablanca, Mohammadia}
- C= {Tanger, Tétouan}.

Nous supposons que les trois classes sont indépendantes, la dépendance au sein de chaque classe est modélisée par une copule extrême.

3. Tarification

La prime pure est définie, mathématiquement, comme l'espérance des indemnités. Pour tarifier le produit, nous allons utiliser les modélisations des indices de pluie provinciaux et nationaux établies.

$$prime\ pure = E[N] \times E[paiements]$$

3.1 Première approche, hypothèse d'indépendance

La prime locale dépend essentiellement de la variable aléatoire ($X_j/X > 200$) où X sont les précipitations agrégées sur la province considérée. C'est pour cette raison que nous allons utiliser la modélisation des dépassements de seuils mise en place. Nous simulons les dépassements de seuil pour en déduire le montant de prime exigé pour chaque province. Dans ce cas la prime nationale s'élève à 188.23 Millions MAD.

3.2 Deuxième approche, hypothèse de dépendance

Pour déterminer la prime pure au niveau national, il faut connaître la distribution de l'indice national. Pour cela, nous avons simulé la distribution multivariée des indices locaux (I_1, \dots, I_7) en passant par les copules et le théorème de Sklar pour en déduire l'indice nationale $I_{national} = \sum_{i=1}^7 I_i$. Le tableau ci-dessous résume les résultats selon la copule utilisée.

Copule	prime pure nationale en Millions MAD
Copule gaussienne	276.38
Copule de Student	277.17
Copule extrême	290,84

TAB. 1 : Prime pure nationale selon le type de copule

Table des matières

Dédicace	I
Remerciements	II
Résumé	III
Note de synthèse	IV
Introduction générale	1
1 Présentation générale des concepts	3
I Introduction	4
II Fonds de solidarité contre les événements catastrophiques - FSEC	5
II.1 Contexte	5
III L'assurance paramétrique	9
III.1 Généralités	9
III.2 Risque de base	10
IV Présensation des données	11
2 Modélisation de la sinistralité catastrophique	14
I Introduction	15
II Deux approches de la modélisation des phénomènes naturels	16
II.1 Modélisation déterministe	16
II.2 Modélisation probabiliste	16
III Modélisation stochastique de la sinistralité	17
III.1 L'aléa	18
III.2 La vulnérabilité	18
III.3 Les dommages	19
IV Produit MnhPRA	19
IV.1 Modélisation des taux de dommages	21
IV.2 Analyse des pertes	23
IV.3 Modélisation des pertes annuelles	26
IV.4 Sensibilité de la perte nette au taux d'inhabitabilité	28
V Conclusion	30
3 Construction de l'indice	31
I l'indice paramétrique	32

II	Fonction de vulnérabilité	34
III	Exposition	35
IV	Fonction de paiement	36
4	Modélisation des précipitations extrêmes	38
I	Introduction	39
II	Théorie des valeurs extrêmes	39
II.1	Estimation des paramètres et calibration	42
II.2	Analyse par les dépassements d'un seuil élevé	44
II.3	Détermination de seuil de la loi de Pareto Généralisée	45
II.4	Estimation des paramètres de la loi de Pareto Généralisée	46
II.5	Return level estimation	48
II.6	Problème de dépendance	49
III	Application aux indices provinciaux	51
IV	Théorie des copules	53
IV.1	les Familles de copules :	54
IV.2	Tau de Kendall et Rho de Spearman	57
V	La modélisation des dépendances avec les copules	58
V.1	L'estimation et la calibration à partir des données	58
V.2	Méthode du maximum de vraisemblance exacte	58
V.3	Méthode des fonctions d'inférence des marginales (IFM)	59
V.4	Méthode de pseudo-maximum de vraisemblance (maximum de vraisemblance canonique CML)	59
VI	Application aux données	60
VI.1	Tests d'adéquation	63
5	Tarification	65
I	Principe de calcul de la prime pure	66
II	Application aux provinces	68
III	Prime nationale	71
IV	Comparaison des primes pures locales et de la prime nationale	72
V	Sensibilité au paramètre la fonction de vulnérabilité	73
VI	Sensibilité au choix de la copule	74
	Conclusion et perspectives	77
	Annexes	80
I	Les principaux aspects de la loi 110-14	81
II	Schéma d'indemnisation	83
III	Test de Kolmogorov-Smirnov	85
IV	Choix des seuils de modélisation	86
V	Paramètres GPD Estimés & Expositions	90

VI Matrice de corrélation estimée 91

Table des figures

1	Détection des événements extrêmes	V
2	Fonction de Paiement	VI
3	QQ-plot rabat	VII
1.1	Occurrence des désastres naturels par type de désastre.	4
1.2	Carte historique de la localisation des inondations au nord du Maroc, Source : données du ministère de l'environnement	5
1.3	Les polices concernées par l'obligation d'extension aux événements catas- trophiques	7
1.4	Processus de déroulement des différentes étapes depuis l'avènement de la catastrophe jusqu'à l'indemnisation éventuelle des sinistrés.	8
1.5	Schéma classique d'assurance paramétrique.	10
1.6	Histogramme des précipitations journalières(Donnés de la province Rabat).	12
1.7	Moyenne des précipitations mensuelles (en mm).	12
2.1	Modèle probabiliste.	18
2.2	Courbes de dommages développées par MnhPRA.	20
2.3	Données simulées par le produit MnhPRA.	21
2.4	Histogramme des taux de dommages.	22
2.5	Boxplot des taux de dommages.	22
2.6	QQplot taux de dommages – loi log normale	23
2.7	Pertes nettes.	26
2.8	QQ-plot des pertes brutes, seuil 10%.	27
2.9	QQ-plot des pertes nettes, seuil 10%.	27
2.10	Sensibilité des pertes nettes aux seuils d'inhabitabilité.	28
2.11	Calcul des expositions	29
2.12	Sensibilité des expositions aux seuils d'inhabitabilité.	30
3.1	Etude des cumuls des précipitations	32
3.2	Agrégation des pluies journalières sur 3 jours	33
3.3	Détection des événements extrêmes	33
3.4	Fonction de vulnérabilité.	34
3.5	Carte des valeurs exposées par province	35
3.6	Fonction de Paiement	36
4.1	Densités de la loi des valeurs extrêmes pour différentes valeurs de ξ	42
4.2	Illustration de la dépendance des extrêmes (données : précipitations Rabat)	50
4.3	Espérance résiduelles des dépassements de seuils des pluies agrégées sur 3 jours (province Rabat)	51

4.4	Hill plot (donnés Rabat)	52
4.5	Graphe Quantile-Quantile '(Rabat)	53
4.6	Nuage de points de cumules des précipitations	60
4.7	Nuage de points des données transformées	61
4.8	Copules elliptiques	62
4.9	Copules extrêmes	62
4.10	Nuage de points des données simulées	63
4.11	Goodness of fit test	64
5.1	Convergence d'après la loi forte des grands nombres.	67
5.2	Histogramme des indices paramétriques de Rabat.	69
5.3	Histogramme de l'indice paramétrique national	71
5.4	Sensibilité de la prime pure nationale au choix de l'attachement	72
5.5	Goodness of fit , copule t de Student	75
5.6	Contour plots de Gumbel copula et Joe copula (données Casablanca- Mo- hammadia)	75
5.7	Goodness of fit , copule de Joe	76
5.8	Goodness of fit , copule de Gumbel	76
II.1	Schéma d'indemnisation	84
IV.1	Modélisation- province Salé	87
IV.2	Modélisation- province Kenitra	87
IV.3	Modélisation- province Casablanca	88
IV.4	Modélisation- province Mohammedia	88
IV.5	Modélisation province Tanger Assilah	88
IV.6	Modélisation province Tétouane	89

Liste des tableaux

1	Prime pure nationale selon le type de copule	IX
2.1	Avantages et inconvénients des deux approches de la modélisation des phénomènes naturels.	17
2.2	Statistiques descriptives des pertes en millions de DH	26
2.3	Pertes estimées en millions de DH pour un seuil de 10%	28
4.1	Estimations des paramètres	52
5.1	Les primes avec TD =5%	70
5.2	Statistiques descriptives du cumul des précipitations sur trois jour (en mm)	70
5.3	Les primes avec TD= 10%	70
5.4	Sensibilité des primes au paramètre de la fonction de vulnérabilité	74
5.5	Primes pures, modélisation des dépendances au sien des régions	76
V.1	Estimation des paramètres GPD	90
V.2	Expositions de quelques provinces	90
VI.1	Matrice de corrélation estimée	91

Liste des sigles et acronymes

GEV	<i>Generalized Extreme Value.</i>
GPD	<i>Generalized Pareto Distribution.</i>
FSEC	<i>Fonds de Solidarité contre les Evenments Catastrophiques</i>
MHPV	<i>Ministère de l'Habitat et de la Politique de la Ville</i>
MnhPRA	<i>Morocco natural hazards Probabilistic Risk Analysis</i>

Introduction générale

Contexte

Aujourd'hui, les effets du changement climatique sont une réalité. Des catastrophes naturelles plus fortes et plus fréquentes, détruisent des maisons et les infrastructures à un rythme record et mettent en danger des systèmes alimentaires entiers. Les enjeux du changement climatique tels que la chaleur extrême, les inondations et autres catastrophes naturelles dominent les rapports publiés par des organisations internationales telles que le forum économique mondial. En effet les chercheurs, décideurs et gouvernements du monde entier cherchent à comprendre la nature de ces changements et les défis auxquels ils seront confrontés. En raison de sa position géographique et de ses conditions climatiques, le Maroc reste potentiellement confronté à une série varié de risques de catastrophes naturelles, de plus en plus récurrents et qui sont exacerbés par ce changement climatique. Ces risques engendrent généralement des dégâts économiques et sociaux lourds affectant souvent des populations et des territoires vulnérables. En témoignent les inondations de novembre 2002 à Mohammedia ; le séisme de février 2004 à Al Hoceima ; les inondations de novembre 2014 à Guelmim. Dans ce contexte, un fonds dénommé Fonds de Solidarité contre les Évènements Catastrophiques (FSEC) a été créé en vertu de la loi n° 110-14. Ce fonds indemniserà les victimes non assurées, notamment les ménages vulnérables qui ne peuvent pas s'offrir une assurance privé (EV CAT).

Selon une analyse actuarielle préliminaire menée en collaboration avec la Banque mondiale, les ménages assurés pourraient recevoir des prestations d'assurance de plusieurs millions de dirhams chaque année en cas de catastrophes naturelles telles que d'inondations et les tremblements de terre, les ménages touchés non assurés pourraient recevoir ces prestations en cas de catastrophe par le biais du FSEC. En outre, un tel programme devrait assurer des paiements plus rapidement, puisque la structure de paiement est déjà prédéfinie. Certaines analyses économiques montrent que le paiement immédiat d'espèces à la suite d'une catastrophe a des avantages sociaux importants, en particulier pour les ménages pauvres ; 1 dollar EU versé à la suite d'une catastrophe vaut au moins 2,5 dollars EU payés plusieurs semaines plus tard (DANIEL J. CLARKE 2013) d'où la nécessité des polices d'assurances paramétriques qui présente l'avantage de la rapidité des paiements.

Problématique

Les principales recettes du FSEC parviennent de produits des taxes parafiscales instituées à son profit par voie réglementaire, ces recettes sont limitées, le FSEC a ainsi fait recours au transfert du risque pour assurer sa résilience.

Comment transférer le risque d'inondations au Maroc via un produit d'assurance paramétrique? Comment modéliser la dynamique de l'indice paramétrique sous-jacent dans le but d'établir un profil de risque du royaume pour pouvoir tarifier le produit?

Organisation du mémoire

Ce mémoire est organisé en cinq chapitres :

Le premier chapitre du rapport définit le contexte de la gestion des risques au Maroc et souligne l'importance de passer d'actions isolées de gestion des risques à des efforts délibérés pour coordonner les actions au niveau national et local à travers différents types de risques. Il couvre également l'assurance paramétrique en tant qu'outil efficace de gestion des risques.

Le deuxième chapitre décrit en détail la méthodologie de la modélisation de la sinistralité catastrophique, la présentation du produit MnHPRA et conclut sur une analyse approfondie des pertes auxquelles le FSEC fera face.

Le troisième chapitre sera consacré à la méthodologie de construction de l'indice paramétrique utilisé dans la détection et la tarification du risque d'inondations au Maroc.

Le quatrième chapitre exposera les outils mathématiques nécessaires pour la résolution du problème posé, à savoir principalement la théorie des valeurs extrêmes et la théorie des copules qui seront ensuite implémentés informatiquement et illustrés par une application sur nos données.

Le dernier chapitre analysera de manière prospective la police d'assurance paramétrique en calculant sa prime. Cette prime sera confrontée à des sensibilités afin de rendre compte des limites des hypothèses.

Chapitre 1

Présentation générale des concepts

I Introduction

Les inondations sont les catastrophes naturelles les plus fréquentes dans le monde. Selon le Centre de Recherche sur l'Epidémiologie des Désastres (CRED), les inondations sont les catastrophes naturelles qui touchent le plus grand nombre d'individus, elles représentent près de 43% de toutes les catastrophes naturelles répertoriées de 1994 à 2019 (Voir figure 1.1). Durant ces deux dernières décennies, les inondations ont causé la mort de près de 158 000 personnes à travers le monde et ont touché à différents niveaux plus de 2,3 milliards. Bien que le nombre de décès dus aux inondations ait considérablement diminué depuis le début des années 1980, les pertes économiques associées à ces catastrophes sont de plus en plus élevées. Elles se chiffrent en 23 milliards de dollars (US) par année. A titre de comparaison, cela correspond à plus de 140% du PIB annuel des États-Unis d'Amérique (Jha et al., 2012). Les inondations sont ainsi considérées comme une des menaces les plus importantes à la sécurité des populations et au développement économique des pays qui y sont exposés, en effet de tels événements peuvent saper le progrès réalisé en matière de la réduction de la pauvreté et de développement.

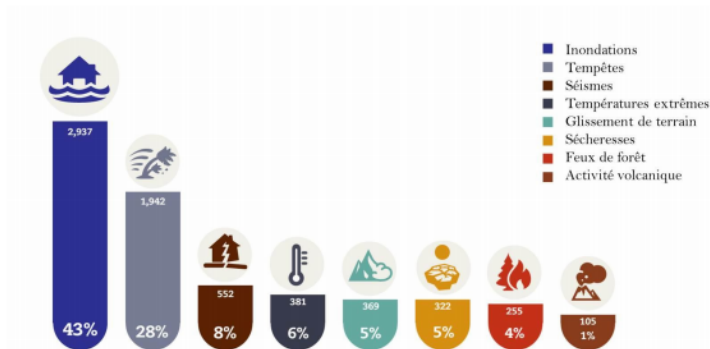


FIG. 1.1 : Occurrence des désastres naturels par type de désastre.

Les inondations causent souvent des dommages non atténués, en particulier dans les pays à faible revenu où la protection contre les inondations est encore moins développée. En effet près de 95% des décès et près de 60% des pertes économiques occasionnées par les inondations sont enregistrés dans des pays en développement, Ce lourd bilan est en grande partie dû à une densité de population de plus en plus importante à l'intérieur même ou à proximité directe des plaines inondables, combinée à une forte vulnérabilité des populations exposées à cet aléa (habitat non adapté, pauvreté de la population, manque de politiques et de mesures de préparation et d'adaptation..).

Le Maroc se trouve lui aussi régulièrement en proie à ce genre de phénomène. De 1950 à 2020, le pays a été touché par des dizaines inondations majeures, tuant des centaines de personnes, affectant des milliers d'autres et occasionnant des milliards de dirhams de dégâts.

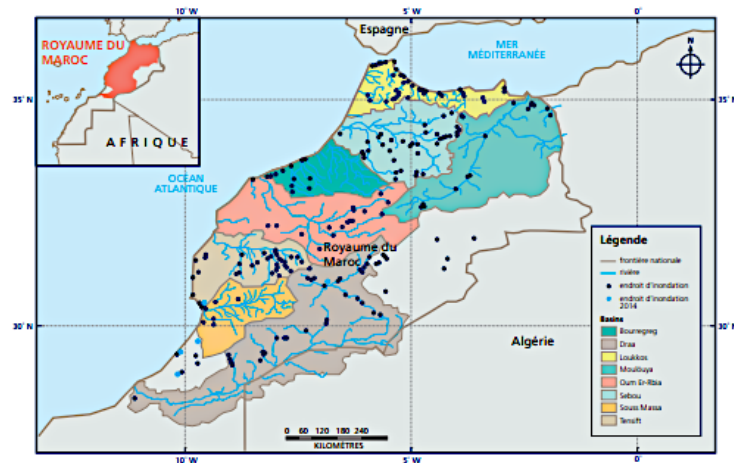


FIG. 1.2 : Carte historique de la localisation des inondations au nord du Maroc, Source : données du ministère de l'environnement

De nombreuses études ont déjà montré que l'augmentation des événements météorologiques extrêmes liés au réchauffement climatique entraînera une augmentation significative des inondations dans les prochaines années, et cela avec un fort niveau de confiance. Les risques liés aux inondations, qui constituent déjà une réalité bien présente pour beaucoup de pays, devraient donc se trouver renforcés dans les années à venir. Depuis une vingtaine d'années déjà, le nombre d'inondations a considérablement augmenté, affectant ainsi un nombre croissant d'individus et occasionnant une hausse importante des dégâts économiques et structurels. Par conséquent, la gestion des risques d'inondations devient de plus en plus importante pour les gouvernements, qui doivent formuler des stratégies et des construire des outils pour prévoir, gérer et s'adapter au mieux à ces risques.

II Fonds de solidarité contre les événements catastrophiques - FSEC

II.1 Contexte

Comme les autres pays de la région, le Maroc est vulnérable à un nombre varié de risques importants et peut être victime de catastrophes naturelles majeures. Les catastrophes naturelles de grande ampleur peuvent avoir un impact considérable sur le budget et l'économie du Maroc, En effet selon le Global Climate Change Risk Index 2016 la perte annuelle moyenne estimée causée par les catastrophes naturelles est d'au moins 7,8 milliards de MAD, dont environ 60% des pertes sont causées par les inondations).

Conscient que les catastrophes naturelles peuvent avoir des conséquences graves sur l'homme et l'économie, le Maroc a beaucoup investi ces dernières années dans l'analyse de sa vulnérabilité aux risques naturels majeurs afin de mieux évaluer et gérer ces risques. Cependant jusqu'à récemment, la gestion des risques au Maroc a été dominée par une approche fragmentée et principalement passive donnant la priorité à la réponse aux crises et reste très sectorielle. Après chaque catastrophe, un nouveau plan d'urgence est mis en

place dans la zone touchée, entraînant ainsi un déséquilibre dans la gestion du problème au niveau national. Cette approche de gestion des risques par plusieurs agences sans coordination empêche le Maroc de mettre en place une approche globale, plus efficace, qui intégrerait l'ensemble des éléments du problème. Pour remédier à ces problèmes, le gouvernement a donc formulé un plan global de réforme de la gestion des risques de catastrophes naturelles finançant ainsi efficacement les risques résiduels qui ne peuvent être atténués à moindre coût. Le programme du gouvernement comprend la consolidation et le développement des activités de gestion des risques du Maroc, visant notamment à améliorer le cadre institutionnel et la capacité du Maroc en matière de gestion des risques de catastrophes naturelles, à accroître les investissements dans la réduction des risques de catastrophes naturelles et à protéger les ménages vulnérables en fournissant une assurance contre les risques de catastrophe. Le plan d'action se concentrera sur les communautés pauvres qui sont souvent les moins préparées à faire face aux risques. En mettant en place un régime global d'assurance contre les risques de catastrophe, le programme contribuera à mobiliser le secteur privé. Plus précisément, le plan introduit un partenariat public-privé pour le financement des catastrophes en vertu duquel le secteur de l'assurance privée fournira une assurance (obligatoire : EVCAT) contre les risques de catastrophe naturelle pour les ménages assurés, élargissant ainsi le marché national de l'assurance et créant de nouvelles opportunités commerciales. Par ailleurs, la nouvelle orientation de la gestion des risques marocaine remplace la réponse a posteriori par une culture de prévention a priori.

Le gouvernement a ainsi adopté La loi actuel (loi n° 110-14) qui régit la gestion des risques catastrophiques au Maroc, la loi vise à couvrir les ménages pauvres contre les catastrophes naturelles via la mise en place de deux régimes de couverture des risques catastrophique : une couverture assurantielle obligatoire (EVCAT) ; et un système allocataire au profit des personnes physiques ne disposant pas de couverture par la mise en place d'un Fonds de solidarité dédié en cas de catastrophes (le FSEC, Fonds de solidarité contre les événements catastrophiques). L'article 17 de ladite loi stipule qu'un des objectifs assignés au FSEC est l'indemnisation des victimes d'événements catastrophiques. Au sens de cette loi, est considéré comme événement catastrophique tout fait générateur de dommages directs survenus au Maroc, ayant pour origine déterminante l'action d'intensité anormale d'un agent naturel ou l'action violente de l'homme (Article 3) (Les principaux aspects de la loi 110-14 en Annexe VI). En vertu de cette loi, les ménages pauvres, qui sont principalement non assurés sont en droit de recevoir une compensation financière auprès du fonds de solidarité (FSEC) pour les dommages corporels (y compris l'incapacité permanente et le décès) et la perte de la résidence principale et de son utilisation causée par des catastrophes.

Le projet repose sur la capacité financière du secteur national de l'assurance soutenue par le marché international de la réassurance, pour financer les pertes dues aux catastrophes par le biais de primes d'assurance supplémentaires (obligatoires). Afin de protéger le secteur de l'assurance contre les grands sinistres catastrophiques et que ces primes puissent être supportées par les assurés, la législation fixe un plafond global d'indemnisation et que l'indemnisation sera proportionnellement réduite en cas de dépassement de ce plafond. Cette approche propose de combiner une assurance privée contre les risques catastrophiques à travers la couverture obligatoire au sein des contrats d'assurance dommages, automobile et responsabilité civile...(la figure ci-dessous résume les catégories

Chapitre 1. Présentation générale des concepts

concernées par l'obligation d'extension aux événements catastrophiques) et un système de couverture de base qui s'appuie sur le principe de solidarité destiné aux ménages vulnérables qui ne peuvent pas s'offrir une assurance privée (FSEC).

Catégories d'assurance	Taxe parafiscale (Oui/Non)	Prime EVCAT (Oui/Non)	Taux de prime EVCAT Prévu sur l'arrêt (Oui/Non)
Responsabilité civile générale			
R.C résultant de l'emploi de véhicules fluviaux et maritimes	Oui	Non	NC
R.C résultant de l'emploi de véhicules aériens	OUI	Non	NC
Autres responsabilités civiles (1)	Oui	Oui	2,00%
Incendie et éléments naturels			
Incendie	Oui	Oui	8,00%
Éléments naturels	Oui	Oui	8,00%
Pertes pécuniaires	Oui	Non	NC
Assurances des risques techniques			
Tous risques chantiers	Oui	Oui	8,00%
Tous risques montage	Oui	Oui	8,00%
Bris de machines	Oui	Oui	8,00%
Tous risques informatiques	Oui	Oui	8,00%
Pertes pécuniaires	Oui	Non	NC
Responsabilité civile décennale	Oui	Non	NC
Transport			
Maritime Corps	Oui	Non	NC
Maritime facultés	Oui	Non	NC
Marchandises transportées par voie terrestre	Oui	Oui	8,00%
Aviation corps	Oui	Non	NC
Autres opérations			
Vol	Oui	Oui	8,00%
Grêle ou gelée (2)	Oui	Oui	8,00%
Mortalité du bétail	Oui	Oui	8,00%
Assurance récolte (2)	Oui	Oui	8,00%
Protection juridique	Oui	Non	NC
Assurances pour pertes pécuniaires	Oui	Non	NC
Autres (3)	Oui	Oui	8,00%
Assistance - crédit - caution			
Assistance	Oui	Non	NC
Assurance crédit	Oui	Non	NC
Assurance caution	Oui	Non	NC

FIG. 1.3 : Les polices concernées par l'obligation d'extension aux événements catastrophiques

Le Maroc s'est alors engagé dans une dynamique de renforcement de ses politiques publiques de gestion intégrée des risques (GIR) de catastrophes naturelles, qui fait de l'amélioration de la connaissance et de la **modélisation** des risques naturels sa priorité. La modélisation des risques naturels nécessite de définir et de mettre en oeuvre un ensemble multidisciplinaire de connaissances et de mesures qualitatives et quantitatives afin d'agir sur **l'aléa** naturel, sur **l'exposition** et **la vulnérabilité** à leurs effets et l'estimation des **dommages** potentiels. La modélisation des risques naturels, par des modèles déterministes et probabilistes, reste une mesure incontournable afin de bien comprendre l'aléa naturel, estimer les dommages potentiels et établir in fine une résilience financière, à travers l'élaboration de stratégies efficaces optimales de **transfert de risque**.

La figure ci-dessous illustre le processus de déroulement prévu des différentes étapes (article 6 de la loi) depuis l'avènement de la catastrophe jusqu'à l'indemnisation éventuelle des sinistrés. Le schéma détaillé d'indemnisation en annexe

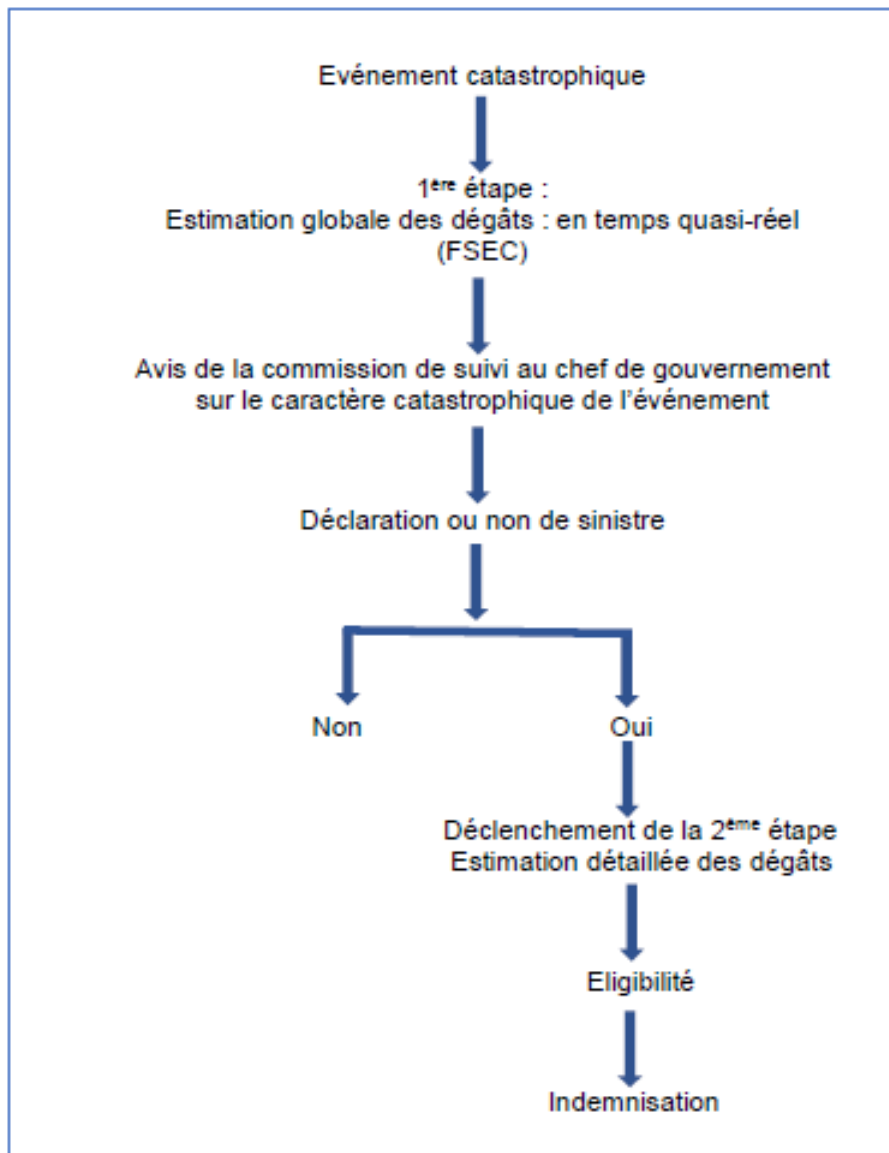


FIG. 1.4 : Processus de déroulement des différentes étapes depuis l'avènement de la catastrophe jusqu'à l'indemnisation éventuelle des sinistrés.

Dans le cadre du projet du FSEC, l'élément déclencheur de la catastrophe naturelle et les éléments pouvant influencer son ampleur, bien qu'ils soient importants pour la compréhension de la genèse de l'événement et de ses conséquences, ne seront pas pris en compte dans l'évaluation des dégâts et de leurs indemnisations éventuelles. Il s'agit d'un programme de solidarité destiné à venir en aide à tout citoyen exposé à un risque dont il n'a aucune couverture assurantielle qui peut prendre en charge une partie ou la totalité les dégâts qui peuvent résulter de ce risque.

III L'assurance paramétrique

III.1 Généralités

Le transfert des risques est un instrument clé pour gérer les risques de dommages climatiques. Le transfert de risque implique qu'un agent transmet un risque à un autre en payant une prime contre d'éventuelles pertes. La principale forme de transfert de risque est l'assurance, dans laquelle l'agent paie une prime en échange de garantie d'indemnisation contre des éventualités précises. Traditionnellement, l'assurance a été basée sur les pertes, c'est-à-dire que les assurés sont indemnisés à la suite des résultats d'un processus de règlement des sinistres dans lequel les pertes réelles sont identifiées.

La structure de l'assurance traditionnelle, la rend particulièrement vulnérable à trois problèmes. Premièrement, l'aléa moral ; se produit lorsque la couverture d'assurance encourage l'assuré à prendre plus de risques. Par exemple, un conducteur qui a souscrit un contrat d'assurance automobile pourrait être incité à conduire plus imprudemment, se sentant en sécurité sachant que tout accident serait couvert par son assureur. Deuxièmement, l'antisélection ; se produit lorsque les mauvais risques (les plus exposés au risque) se couvrent à un taux plus élevé que les bons risques qui sont moins exposés au risque. Cela est dû à l'asymétrie de l'information entre l'assureur et l'assuré. L'antisélection peut menacer l'intégrité financière du système d'assurance. Le troisième inconvénient est le temps d'évaluation des dommages, en effet le temps d'évaluation des sinistres peut durer plusieurs mois pour le risque de catastrophe naturelle lorsque des territoires entiers sont dévastés.

Récemment, en réponse aux coûts et aux difficultés pratiques inhérentes à la détermination des pertes dans le cas des événements catastrophiques, le secteur de l'assurance a développé une nouvelle forme d'assurance basée sur un indice plutôt qu'une perte, l'assurance paramétrique. Dans le cadre de l'assurance paramétrique les assurés ne sont pas indemnisés sur la base des pertes, mais plutôt sur la base de la valeur d'un indice prédéterminé. Les précipitations, par exemple, sont couramment utilisées comme indice qui quantifie des dommages engendrés par les inondations, si l'indice dépasse un seuil défini à l'avance, les indemnités sont déclenchées. Cet indice est fonction de facteurs physiques qui ne sont pas objets à des manipulations humaines.

La plupart des régimes d'assurance paramétrique sont principalement destinés à protéger les agriculteurs et les populations vulnérables des pays en développement contre les pertes dues aux catastrophes naturelles, généralement conçus comme des mécanismes de réponse aux chocs météorologiques. Ce type de régime présente trois avantages fondamentaux :

- Par rapport à l'assurance traditionnelle, l'assurance paramétrique réduit le problème de l'aléa moral, car les conditions sont fixées à l'avance et connues de tous dans le contrat. Le versement des indemnités est basé sur un indice lié à des variables exogènes aux deux agents donc les assurés sont moins incités à prendre des risques supplémentaires.
- En faisant passer la base des paiements de pertes individuelles sur lesquelles les assureurs ont relativement peu d'informations, à des valeurs d'indice objectives et

accessibles au public liées à des décaissements prédéfinis, le problème de sélection adverse est réduit. Les assurés peuvent modéliser le risque avec plus de précision et évaluer les produits de manière plus appropriée pour contribuer à garantir la solvabilité financière. L'assurance paramétrique implique une plus grande symétrie d'informations et donc un portefeuille de risques plus diversifié.

- L'assurance paramétrique implique des coûts de transaction plus faibles et assure la rapidité des paiements. Traditionnellement, l'évaluation des sinistres est souvent un processus long et coûteux. L'assurance paramétrique, en revanche, n'implique pas des enquêtes longues sur les pertes, car les indemnités se calculent à base de l'indice. En effet, le temps moyen d'indemnisation pour une assurance paramétrique s'élève à deux semaines, alors qu'il nécessite plusieurs mois (voir années dans certaines cas) pour évaluer les dommages dans le cas d'une assurance classique, ce point est crucial, concrètement les pays pauvres subissant des catastrophes naturelles se retrouvent désemparés et font face à de sérieux problèmes de liquidité. L'assurance paramétrique permet un déclenchement des indemnités immédiat en appliquant la règle préétablie. Cela permet au pays de disposer rapidement des liquidités nécessaires pour subvenir aux besoins les plus urgents. Ci-dessous le schéma classique d'assurance paramétrique.

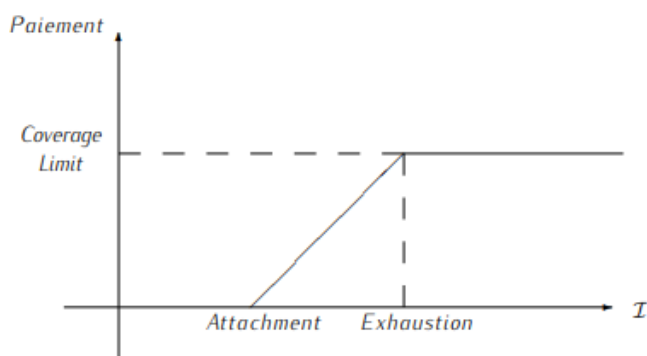


FIG. 1.5 : Schéma classique d'assurance paramétrique.

Pour se couvrir contre un aléa climatique, l'assuré paie une prime et attend une indemnisation qui n'est pas fonction des pertes réalisées mais fonction de l'évolution de l'indice.

III.2 Risque de base

L'assurance paramétrique introduit un nouveau problème appelé risque de base. Le risque de base est défini comme l'écart potentiel entre les indemnités versées (sur la base des valeurs d'indice) et le coût économique des dégâts réels au sol. Il est clair qu'il n'existe aucun risque de base pour les contrats d'assurance classique, car l'indemnisation n'est rien d'autre que le coût réel du sinistre. L'introduction de l'assurance paramétrique réduit le temps d'évaluation des dommages, mais augmente le risque de base par essence. Le risque de base concerne principalement les assurés qui subissent les dommages mais ne reçoivent aucune compensation car la police d'assurance n'a pas été déclenchée ou

bien l'indemnisation ne représente qu'une portion minimale des dégâts réels subis. d'où la nécessité de choisir un indice qui est fortement corrélé aux pertes réelles.

IV Présensation des données

Les données de précipitations GPM (Global Precipitation Measurement) ont été utilisées comme données principales . Le satellite GPM a été lancé en coopération entre les États-Unis et le Japon NASA et JAXA (Japanese Aerospace Exploration Agency) pour fournir des données pluviométriques spatio-temporelles détaillées sur ensemble des continents. La mission GPM contribue à faire progresser notre compréhension des cycles de l'eau et de l'énergie de la Terre, améliore la prévision des événements extrêmes qui provoquent des catastrophes naturelles et étend les capacités actuelles d'utilisation des informations sur les précipitations satellitaires au profit direct de la société. En effet les données GPM ont été largement utilisées pour diverses applications dans le monde telles que météorologiques, hydrologiques, agricoles et la gestion catastrophes. Le produit IMERG, (Integrated Multi-satellite Retrievals for GPM), est un produit de précipitation par satellite unifié par la NASA pour estimer les précipitations de surface sur la majeure partie du globe. Ce produit fournit la meilleure estimation des précipitations à partir de plusieurs capteurs à bord de plusieurs satellites sur une grille de 25 km x 25 km quotidiennement de données pluviométriques. Ces données pluviométriques sont disponibles en ligne dans plusieurs services de données en ligne géré par la NASA. Sur la base de la disponibilité des données, nous avons utilisé des données de précipitations quotidiennes de 2000 jusqu'à 2020.

Cette étude ne concerne que l'axe Casablanca-Tanger, notamment les provinces de Casablanca, Mohammadia, Rabat, Sale, Kenitra, Tétouan et Tanger. ce choix se justifie par la difficulté de traitement des données de toutes les provinces du Maroc et par la concentration de la population ainsi que l'activité économique dans cet axe, selon les données de l'HCP, ces trois régions ont crée 58% de la richesse nationale en 2020.

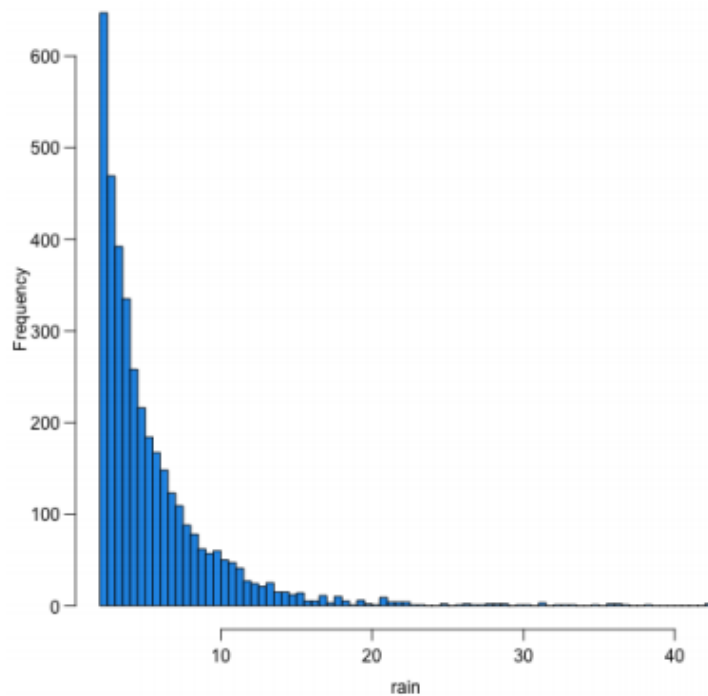


FIG. 1.6 : Histogramme des précipitations journalières (Données de la province Rabat).

Le Maroc est un pays de la région subtropicale du nord-ouest de l'Afrique. Il se caractérise par un climat qui varie d'une région à l'autre, en effet, les zones littorales bénéficient d'un climat tempéré, tandis que les régions du sud et de l'est du pays ont un climat désertique. Le climat du royaume a de nombreuses nuances : méditerranéen au Nord, océanique à l'Ouest, continental à l'intérieur des terres et saharien au Sud. Le climat varie aussi en fonction des saisons.

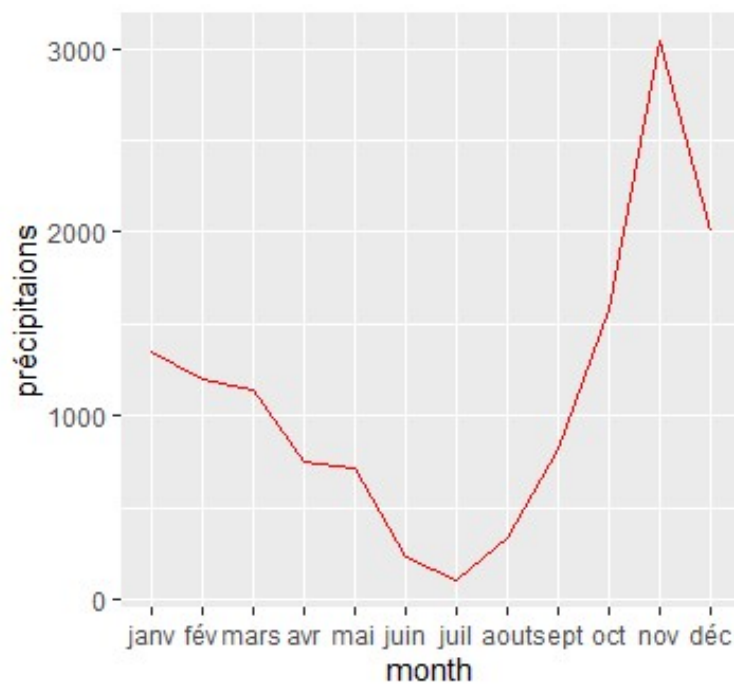


FIG. 1.7 : Moyenne des précipitations mensuelles (en mm).

D'après le graphique ci-dessus, il est clair que le climat marocain est marqué par une bi-saisonnalité, une saison humide (pluvieuse) durant laquelle les précipitations sont beaucoup plus importantes. Au Maroc cette saison s'étale de septembre jusqu'à Mars et une saison sèche ,plus calme sèche de Mars à septembre.

Chapitre 2

Modélisation de la sinistralité catastrophique

I Introduction

Les catastrophes naturelles telles que les tremblements de terre, les inondations, les tsunamis ou les sécheresses constituent une source potentielle de grandes pertes pour le Maroc. Même les économies les plus avancées technologiquement ont du mal à faire face à ces événements : le tremblement de terre et le tsunami au Japon (20.000 morts) 2011 ; le séisme de 2008 en Chine (80.000 morts) et le passage de l'ouragan Katrina 2005 aux Etats-Unis (a coûté 150 milliards de dollars US) sont des exemples parlants à cet égard. Chacun de ces événements s'est produit soudainement et a été initialement désigné "à faible probabilité" et "sans précédent" par la suite ce genre de phénomène naturel est très difficile à prévoir. En effet, l'alerte pour un événement climatique est donnée au mieux quelques jours avant la survenance de ce dernier. Pour d'autres périls comme le séisme ou le tsunami, il est impossible d'anticiper à l'avance l'événement.

La leçon à tirer est simple mais vitale ; les catastrophes naturelles se produisent, et une histoire récente de l'absence de telles catastrophes ne constitue pas une indication de diminution du risque. En fait, la quiétude récente est plus généralement un prélude à un risque imminent. L'historique des aléas naturels au Maroc est relativement court et pas toujours bien documenté. Ci-dessus un résumé d'historique des catastrophes naturelles, qui indique que les risques hydro-météorologiques, notamment les inondations ont touché plus de personnes que tout autre phénomène, et ont causé le plus de pertes économiques, tandis que les tremblements de terre ont causé le plus de pertes en vies humaines, et constituent également une source importante de pertes économiques. Les événements spécifiques retenus se déclinent comme suit :

Inondations :

- 2002 : Mohammedia et Berrechid : 63 morts, des maisons effondrées, des centaines de maisons inondées, des centaines d'hectares de terres agricoles touchées.
- 2008 : Tanger, 30 morts, des pertes économiques (zone industrielle).
- 2009 : Al Gharb, 400 maisons détruites, 100.000 hectares de terres détruites avec le coût des dégâts s'élevant à plus de 1 milliard DH.
- 2014 :Guelmim, 47 morts, les dommages ont été évalués à plus de 6 milliards de dirhams.

Tremblements de terre :

- 1960 : Agadir, magnitude 5.7, 12.000 morts.
- 2004 : Al-Hoceima : 628 morts, 926 blessés, 12.367 maisons effondrées.

Ces événements n'étant pas prévisibles, les assureurs et réassureurs peuvent néanmoins évaluer en moyenne la fréquence et la sévérité des sinistres catastrophes futurs sur une longue période. Ainsi, dans ce chapitre nous allons présenter les deux modèles

permettant l'évaluation de la sinistralité des phénomènes naturels, présenter le produit MnhPRA et analyser en profondeur les résultats de modèle, notamment les pertes pour estimer les fonds à allouer pour pouvoir indemniser les victimes des inondations.

II Deux approches de la modélisation des phénomènes naturels

II.1 Modélisation déterministe

Le modèle déterministe est la solution la plus simple pour simuler les sinistres provoqués par des catastrophes naturelles. Cette méthode repose sur l'analyse de l'historique des phénomènes naturels afin d'évaluer la fréquence et la sévérité de chaque péril. Des lois de distributions spécifiques sont appliquées, comme une loi de comptage de type Poisson ou Binomiale négative pour la fréquence et une loi de type Log-Normal, Weibull ou Pareto pour la sévérité. L'analyse de l'historique permet d'estimer les paramètres des lois..

Ce type de modélisation recourt à des événements majeurs présents dans l'historique que l'on rapporte aux valeurs assurées d'aujourd'hui par une mise en "As-If". L'une des principales difficultés de l'analyse en As-If des périls naturels est le suivi de l'exposition des biens assurés. En effet, l'exposition géographique des biens assurés est un facteur clé dans la modélisation de la sinistralité. Par exemple, un bien situé à proximité d'un 'Oued' sera plus sensible aux inondations qu'un bien plus en retrait. Le portefeuille des assureurs évolue au cours du temps. Ainsi, une inondation présente dans l'historique peut provoquer des dégâts plus importants si elle se reproduisait aujourd'hui. Le manque de données historiques est un autre inconvénient majeur de cette méthode. En effet, les assureurs et réassureurs se sont constitués un historique d'une trentaine voire d'une quarantaine d'années. Or, cette période ne semble pas assez longue pour avoir un historique suffisamment exhaustif. Il se peut par exemple qu'aucun événement extrême ne se soit produit dans le passé.

II.2 Modélisation probabiliste

Alors que le modèle déterministe repose sur un historique de moins d'un demi-siècle, le modèle probabiliste consiste à simuler l'ensemble des événements possibles sur une période de plusieurs milliers d'années. La modélisation permet de définir une liste de sinistres catastrophes avec la probabilité de survenance et la charge moyenne de chacun de ces événements. Les modèles probabilistes ont été développés par des sociétés spécialistes du climat dont les plus connues sur le marché de l'assurance sont : Applied Insurance Research (AIR), Risk Management Solutions (RMS) et EQECAT. Chacun de ces modèles s'appuie sur trois modules que nous présenterons

- L'aléa
- La vulnérabilité.

Modèle	Avantages	Inconvénients
Modèle déterministe	<ul style="list-style-type: none"> • Délai réduit d'obtention du modèle • Sinistralité historique 	<ul style="list-style-type: none"> • Manque de données historiques • Le passé n'est pas forcément représentatif du futur
Modèle probabiliste	<ul style="list-style-type: none"> • Evènements absents de l'historique • Prise en compte de la vulnérabilité • Prise en compte de l'évolution de l'exposition 	<ul style="list-style-type: none"> • Complexité du modèle • Nécessite beaucoup de données précises

TAB. 2.1 : Avantages et inconvénients des deux approches de la modélisation des phénomènes naturels.

- Les dommages.

Les deux modélisation présentent des avantages et des inconvénients que nous allons exposer.

III Modélisation stochastique de la sinistralité

L'objectif de la modélisation probabiliste est de constituer un jeu d'événements ou les périls qui ne sont jamais survenus, mais la probabilité de survenance n'est pour autant pas nulle. Ce modèle repose sur trois modules, qui sont la pierre angulaire de la modélisation des sinistres. Le schéma ci-dessous illustre l'articulation des trois modules.

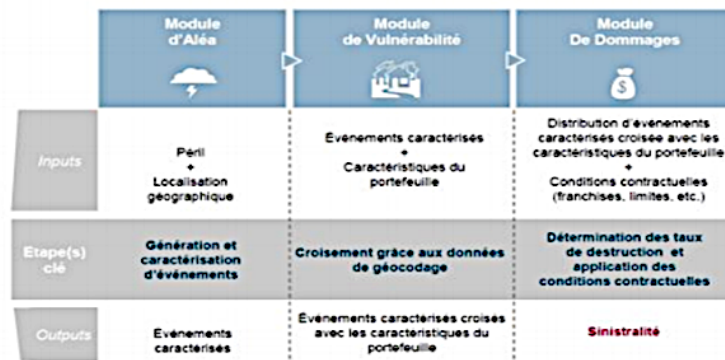


FIG. 2.1 : Modèle probabiliste.

III.1 L'aléa

Le module aléa a pour objectif de simuler des événements fictifs, mais réalistes, afin de mesurer quel pourrait être l'impact d'événements plus extrêmes que ceux disponibles à l'historique de l'ensemble des périls pouvant se réaliser. Ainsi, le module se base dans un premier temps sur l'historique des sinistres antérieurs avant de faire varier certains paramètres comme la localisation géographique et l'intensité. ces paramètres sont variés selon les études de spécialistes du climat. Pour les inondations par exemple, l'intensité est mesurée par la hauteur d'eau. En règle générale, l'intensité diminue à mesure qu'on s'éloigne du centre de l'événement. L'étude de l'intensité suivant la zone géographique est importante pour les assureurs car elle est étroitement liée aux dommages constatés. Le module aléa doit être alimenté par des données sur les périls antérieurs ainsi que sur la localisation géographique. Après traitement du modèle, le module aléa constitue un jeu d'événements par périls étudiés. L'ensemble de ces événements sont ainsi susceptibles de se réaliser sur une période de plusieurs milliers d'années. Intervient ensuite le module vulnérabilité, consistant à déterminer le coût.

III.2 La vulnérabilité

L'intensité et la localisation géographique d'un péril naturel ne sont pas les seuls paramètres permettant de mesurer l'ampleur des dégâts. En effet, suivant le type de construction des bâtiments, leur ancienneté, leur hauteur ainsi que leur contenu, la charge totale peut varier entre deux événements de même intensité. La vulnérabilité s'appuie sur l'estimation des valeurs assurées déterminées par des données clients et des sources nationales. (Pour le Maroc, ces sources sont la Direction Générale des Impôts, le Haut-commissariat au Plan, le ministère de l'habitat et de la politique de la ville (MHPV)).

- Les valeurs liées au bâtiment. La méthode utilisée pour l'estimation consiste à multiplier un prix de reconstruction, avec la surface réelle ou estimée m^2 .
- La notion de contenu. Cette notion est très importante dans l'estimation des dommages, en effet le FSEC ne dédommage que les pertes en contenant, Cette notion est déterminée soit par une estimation des biens, soit par une valeur forfaitaire définie à l'échelle nationale.

Dans ce cadre, le module vulnérabilité a pour objectif de déterminer le taux moyen de sinistralité en fonction de l'intensité du péril. Le taux moyen de sinistralité est défini comme le montant total du sinistre par rapport à la valeur totale de tous les objets d'assurance (donc y compris ceux qui n'ont subi aucun dommage) dans la zone considérée.

Ce module permet de tracer une courbe de vulnérabilité pour chaque catégorie de risque (commerces, maisons individuelles, industries, etc.). Ces courbes expriment le rapport entre l'intensité et le taux moyen de sinistralité. Suivant les différentes catégories de risque étudiés, le modèle prend également en compte les fluctuations de la vulnérabilité liées par exemple à la qualité des matériaux de construction. Ainsi, la charge totale d'un événement sera une variable aléatoire. Le module vulnérabilité doit être alimenté par des informations détaillées sur les biens du portefeuille assuré. L'articulation entre les modules aléa et vulnérabilité permet de déterminer le montant brut moyen pour chaque événement.

III.3 Les dommages

Le modèle de dommages a pour objectif de calculer le montant des pertes à partir de l'aléa et de la vulnérabilité, en effet la modélisation des périls naturels doit permettre aux assureurs d'évaluer le montant des pertes futures. Ainsi, les conditions d'assurances comme les garanties prises en charge et le niveau de franchise sont des éléments à prendre en compte. A cette fin, l'objectif de ce troisième module est de calculer, à partir du montant brut d'un sinistre, la charge nette pour l'assureur. Après traitement des trois modules sur le portefeuille d'une société, le modèle fournit une liste d'événements pouvant se réaliser.

IV Produit MnhPRA

Dans le cadre du partenariat Banque mondiale-Maroc, une évaluation de risques à été réalisée pour les catastrophes naturelles par la création de MnhPRA ("Morocco natural hazards Probabilistic Risk Analysis" ou "l'Analyse Probabiliste des Risques de Catastrophes Naturelles au Maroc"). Il s'agit d'un modèle catastrophe (Cat) probabiliste et multi-périls (séismes, inondations, tsunamis, sécheresses, glissement de terrain) permet d'analyser l'exposition à différents aléas, à l'échelle régionale, provinciale et communale en utilisant une approche probabiliste avancée de l'aléa (ou modèles de catastrophes).

Ce produit est basé sur une approche probabiliste. Une approche probabiliste de modélisation des catastrophes offre plus de valeur qu'une approche déterministe, puisque elle inclut tous les événements qui peuvent causer des dommages, et elle génère une analyse détaillée des périodes de retour sur la base de modèles avancés de risques. Les trois modules de base pour un modèle de catastrophes sont : l'aléa, la vulnérabilité et la perte, tel qu'illustré ci-après :

Le premier module comprend deux parties principales. La première partie aborde l'apparition et la fréquence des événements en élaborant d'abord une série d'événements stochastiques simulés caractérisant les événements observés ou scientifiquement modélisés

et leurs probabilités d'occurrence. La seconde partie du module de risque calcule la gravité des événements à chaque site de la région d'étude ; il s'agit dans ce cas de l'ensemble des communes de royaume.

Le deuxième module quantifie la vulnérabilité, ou la manière dont les propriétés ou les infrastructures à risque vont réagir physiquement sous les événements générés dans le module de risques. Les fonctions de la vulnérabilité représentent les relations entre l'intensité des risques (par exemple, la hauteur d'eau pour les inondations) sur le site et le niveau des dommages. (voir figure 2.2).

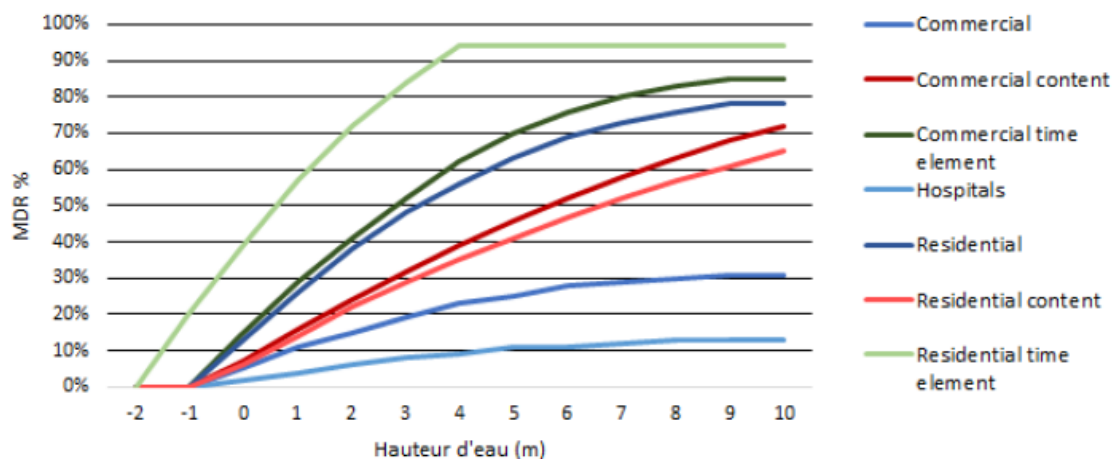


FIG. 2.2 : Courbes de dommages développées par MnhPRA.

Le dernier module couvre les pertes : Cette partie décrit le modèle de dommage mis au point pour le FSEC. Dans le cadre de la mise en place du régime d'allocations aux victimes des inondations, On suppose pour l'instant que la seule incertitude est la survenance d'un événement. Il s'agit, par exemple, de la survenance d'inondations à Casablanca qui suit un certain processus mathématique qui a été spécifié dans le module de risques. Compte tenu d'un événement qui a eu lieu, la deuxième partie du module de risques calcule la gravité de l'événement sur l'ensemble de la ville. Le module de vulnérabilité calcule ensuite le ratio moyen de dégâts (taux de dommage) à chaque site étant donné les caractéristiques spécifiées de site. Enfin, le module de perte examine les dégâts selon les valeurs d'exposition et calcule les pertes totales dans ce site.

On dispose d'une base de données contenant des données simulées, générés à partir du produit MnhPRA. Pour chaque commune nous avons 10 000 simulations d'inondations, nous cherchons à déterminer le profil des pertes annuelles qui seraient subis par le FSEC, nous cherchons à mesurer le risque subit par le FSEC et calculer les pertes associées à des événements catastrophiques de différentes périodes de retour pour permettre au FSEC d'allouer les ressources nécessaires pour honorer ces engagements vis-vis les citoyens qui seront touchés par ces inondations. Ci-dessous un extrait de la base de données

Commune	Class	Event.Year	Damage	Province	Type	EXP_ADJ	EXP_AVG	CNT_ADJ	DR
1	RES	1	0.01106753	1	Urban	820.1243	217376.4	3772.829	1.792081e-05
1	RES	2	0.00000000	1	Urban	820.1243	217376.4	3772.829	0.000000e+00
1	RES	3	0.00000000	1	Urban	820.1243	217376.4	3772.829	0.000000e+00
1	RES	4	39.17881020	1	Urban	820.1243	217376.4	3772.829	6.343925e-02
1	RES	5	33.68699860	1	Urban	820.1243	217376.4	3772.829	5.454678e-02
1	RES	6	36.63730660	1	Urban	820.1243	217376.4	3772.829	5.932398e-02
1	RES	7	30.29502210	1	Urban	820.1243	217376.4	3772.829	4.905441e-02
1	RES	8	29.81528040	1	Urban	820.1243	217376.4	3772.829	4.827760e-02
1	RES	9	0.00000000	1	Urban	820.1243	217376.4	3772.829	0.000000e+00
1	RES	10	0.00000000	1	Urban	820.1243	217376.4	3772.829	0.000000e+00
1	RES	11	34.92116770	1	Urban	820.1243	217376.4	3772.829	5.654517e-02
1	RES	12	31.89756660	1	Urban	820.1243	217376.4	3772.829	5.164929e-02
1	RES	13	0.00000000	1	Urban	820.1243	217376.4	3772.829	0.000000e+00
1	RES	14	31.13259830	1	Urban	820.1243	217376.4	3772.829	5.041063e-02

FIG. 2.3 : Données simulées par le produit MnhPRA.

- La variable Commune spécifie le code de la commune ;
- La variable Class représente la classe de risque (bâtiments résidentiels, installations essentielles, mosquées, infrastructures commerciales, réseaux, agriculture), notre base de données ne concerne que les bâtiments résidentiels ;
- La variable Event.Year représente l'année de simulation ;
- La variable Damage : la perte due aux inondations dans la commune (en millions de DH) ;
- La variable EXP_ADJ est l'exposition (en millions de DH) de la commune exposée au risque ;
- La variable EXP_AVR est l'exposition moyenne ;
- La variable CNT_ADJ représente le nombre de bâtiment exposés au risque dans la commune ;
- La variable DR (Damage ratio) le taux de dommages de la commune généré par le module de vulnérabilité.

IV.1 Modélisation des taux de dommages

dans cette partie nous analyserons en détail la variable DR (Damage ratio) afin de déterminer le profil de risque du Maroc, rappelons que notre étude ne concerne que les bâtiments résidentiels donc les pertes qui seront calculée ultérieurement sont les pertes des résidences qui sont touchées par les inondations.

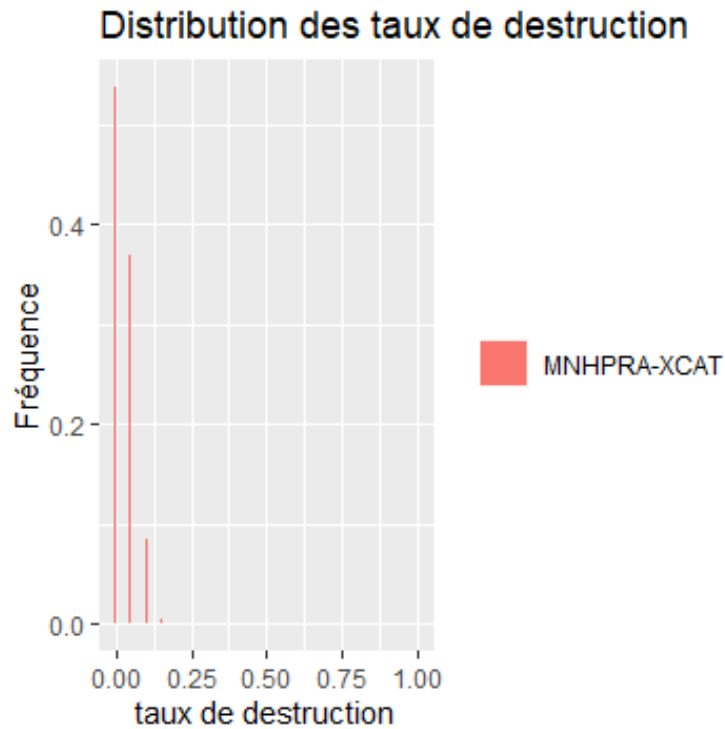


FIG. 2.4 : Histogramme des taux de dommages.

L'histogramme exhibe une tendance clé des taux de dommages, on effet on remarque que les taux de dommages pour le modèle MnhPRA sont concentrés entre 0 et 15%, ce constat est confirmé par le boxplot ci-dessous.

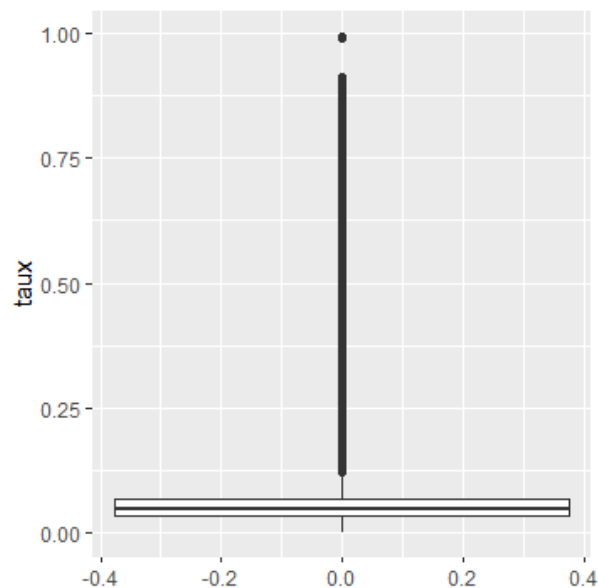


FIG. 2.5 : Boxplot des taux de dommages.

L'histogramme des taux des de dommages est symétrique vers la droite, donc une distribution normale ne fournirait pas un bon ajustement. Nous allons faire recours à d'autres distributions tel que la distribution log normale, gamma et Gumbel.

Définition 2.1. Une variable aléatoire X à valeurs dans $]0, +\infty]$ suit la loi log-normale de paramètres (μ, σ) si $Y = \log(X)$ suit la loi normale de paramètres (μ, σ) .

Après transformation de données, la moyenne et l'écart-type de \log (DR) sont calculés et on en déduit la moyenne et la variance des taux de dommages. On détermine les paramètres de la distribution log normale à partir de la moyenne et de la variance (méthode des moments).

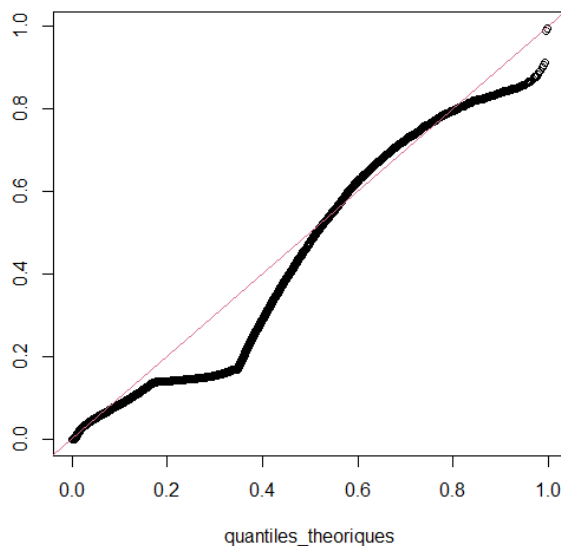


FIG. 2.6 : QQplot taux de dommages – loi log normale .

D'après le graphe QQ-plot il est clair que la distribution log normale ne permet pas l'ajustement des taux de dommages, pour confirmer ce constat nous faisons recours au test de Kolmogorov-Smirnov, le test retourne une P-value de 0.0029 très faible (inférieur à 5%) donc la loi log-normale ne fournit pas un bon ajustement aux taux de dommages.

En se basant sur le test de Kolmogorov-Smirnov, nous avons conclu que la loi log-normale ne permet l'ajustement des taux de dommages, en effet nos échantillons sont de larges tailles (3 millions observations pour MnhPRA), la P-value mesure la distance entre les données et l'hypothèse nulle à l'aide de l'estimation du paramètre d'intérêt. La distance est généralement mesurée en unités de l'écart-type par rapport à cette estimation. Les estimateurs cohérents ont des erreurs standard qui diminuent à mesure que la taille de l'échantillon augmente. Avec un très grand échantillon, l'erreur standard devient extrêmement faible, de sorte que même des distances minuscules entre l'estimation et l'hypothèse nulle deviennent statistiquement significatives. Les lois gamma et Gumbel sont aussi rejetées par le test de Kolmogorov-Smirnov.

IV.2 Analyse des pertes

La modélisation des catastrophes naturelles a pour objectif l'estimation et la prédiction des conséquences financières des catastrophes, comme décrit dans la première

partie le modèle d'inondation repose sur trois modules distincts. Le dernier module est le module de dommages qui permet d'estimer les montants des pertes à partir du croisement entre l'aléa et la vulnérabilité. Ce modèle est souvent calibré à partir des données de sinistralité historiques le résultat du modèle de dommages est une estimation des pertes à l'échelle du bien, d'une entité administrative (commune, province, etc.) ou encore du pays. Dans le cadre du produit MnhPRA l'étude a été faite sur les communes (en se basant sur l'hypothèse que les bâtiments ont les mêmes caractéristiques au niveau de la même commune)

IV.2.1 Méthode de calcul des pertes subies par le FSEC

Les pertes qui figurent dans la base de données sont des pertes brutes c'est à dire les pertes observées réellement sur le sol, or le FSEC n'indemnise pas toutes les pertes. Dans le cadre de la mise en place du régime d'allocations aux victimes des inondations, le schéma d'indemnisation suivant a été mis en place :

- Lorsqu'un propriétaire voit sa maison touchée, le montant de l'indemnité est calculé comme étant la somme de la perte de résidence. La perte de résidence correspond au minimum entre la valeur des dommages D , 70% coût de reconstruction à neuf et une limite de 250000 MAD. notons que le FSEC ne dédommage que les pertes en contenant.(Article 40 loi 110-14)
- Il faut noter que pour que le bien soit indemnisé, il faut qu'il soit considéré comme inhabitable, le taux de dommage de sa résidence devra dépasser un seuil bien précis nommé le seuil d'inhabitabilité fixé par l'ACAPS. Si le taux de dommage est en dessous de ce seuil , les pertes sont à des victimes! Le taux de dommages minimal (c'est-à-dire le seuil d'inhabitabilité) pour les dommages matériels RES est fixé (par l'ACAPS) à 10% donc, on s'intéresse uniquement aux dommages dont les taux de destruction est supérieur ou 'égale à 10% ;

Ainsi l'indemnité la réhabilitation de la résidence principale est donnée par

$$Indemnité = \min(D , 70\% \times A , B) \quad (2.1)$$

avec

- D : Valeur des dommages occasionnés à la résidence principale ;
- A : Coût de reconstruction à neuf d'une partie ou de la totalité de la résidence principale rendue inhabitable ;
- B : Montant fixé par l'administration après avis de l'ACAPS. Ce montant ne peut être inférieur à 250 000 dirhams.

On disposant des données sur des communes, les dommages dans une commune sont calculés par la formule suivante :

$$Dommage = \text{taux de destruction} \times \text{Exposition commune} \quad (2.2)$$

Les pertes couvertes par la couverture CAT sont réparties entre la population assurée et celle non-assurée selon les proportions d'assurés déterminées par l'ACAPS. Le FSEC ne dédommage que les pertes subies par les non assurés. Donc les pertes nettes assurées par FSEC dans une commune i et une année de simulation j en supposant l'homogénéité entre les résidences de la même commune (Il s'agit ici d'une hypothèse très forte, mais en l'absence de granularité plus fine du modèle, il n'est pas possible a priori d'avoir une connaissance sur la vulnérabilité des résidences), sont données par la formule suivante :

$$(2.3) \quad \text{Loss}(i, j) = \min(n_i \times 250000, 0.7 \times \text{exposition}_i \times C, DR_{i,j} \times \text{exposition}_i \times C) \times PNA$$

Avec :

- n_i est le nombre de résidences dans la commune i
- exposition_i est l'exposition de la commune i
- C pourcentage de la valeur du contenant, $C = 0.62$
- $DR_{i,j}$ taux de dommage de l'évènement qui a touché la commune i à l'année j .
- PNA est le pourcentage des non-assurés, $PNA = 0.975$.

Pour déterminer une perte annuelle de tout le territoire dans une année j , il suffit d'agréger les pertes des communes

$$\text{Pertes sur tout le territoire national} = \sum_{i=1}^n \text{pertes commune } i = \sum_{i=1}^n \text{Loss}_{i,j} \quad (2.4)$$

Nous données sont classées en année de simulation, le produit MnhPRA nous fournit des données sur 10000 simulations, on calcule une perte annuelle pour chaque année de simulation en agrégeant les pertes de toutes les communes, finalement on aura un échantillon de 10000 pertes annuelles nette (après application des franchises).

Année 1	Année 2	Année i	Année 10000
Perte commune 1	Perte commune 1		Perte commune 1		Perte commune 1
Perte commune 2	Perte commune 2		Perte commune 2		Perte commune 2
.....
Perte commune i	Perte commune i		Perte commune i		Perte commune i
.....
Perte commune x	Perte commune x		Perte commune x		Perte commune x
Perte annuelle 1	Perte annuelle 2		Perte annuelle i		Perte annuelle 10000

FIG. 2.7 : Pertes nettes.

Avec la perte la perte annuelle i est donnée par la formule 2.4 appliquée aux pertes de l'année i . Les pertes brutes sont calculées de la même manière mais sans application de franchises.

$$Loss_{i,j} = dommage_{i,j} \tag{2.5}$$

Ci-dessous quelques statistiques descriptives de ces pertes

statistique	Pertes brutes (MMAD)	Pertes nettes (MMAD)	nette / brute
Min	23.03	20.53	89.14%
1 er quartile	1476.15	42.68	2.85%
Mediane	2029.18	62.26	3.06%
Moyenne	2047.19	88	4.3%
3-ème quartile	2499.91	93.05	3.72%
Max	17480.65	5080	29.06%

TAB. 2.2 : Statistiques descriptives des pertes en millions de DH

La perte moyenne annuelle assurée par le FSEC est de 88 Millions de DH, et la perte maximale s'élève à 5080 millions de DH. Cependant la perte brute moyenne annuelle est de 2047.19 millions de DH donc le FSEC n'assure au moyen que $88/2047.19 = 4.3\%$ des pertes réellement engendrées par les inondations.

IV.3 Modélisation des pertes annuelles

Afin de déterminer les fonds à allouer pour faire face à des évènements catastrophiques (inondations) de différentes périodes de retour, on modélise la loi des pertes brutes et des pertes nettes par la loi log-normale, puis on calcule les quantiles des probabilités qui correspondent à chaque période de retour. On modélise la loi des pertes par une loi log-normale, les paramètres de la distribution sont estimés par la méthode des moments.

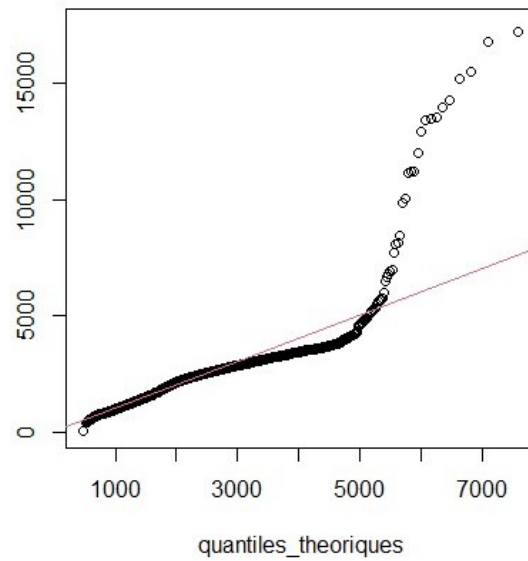


FIG. 2.8 : QQ-plot des pertes brutes, seuil 10%.

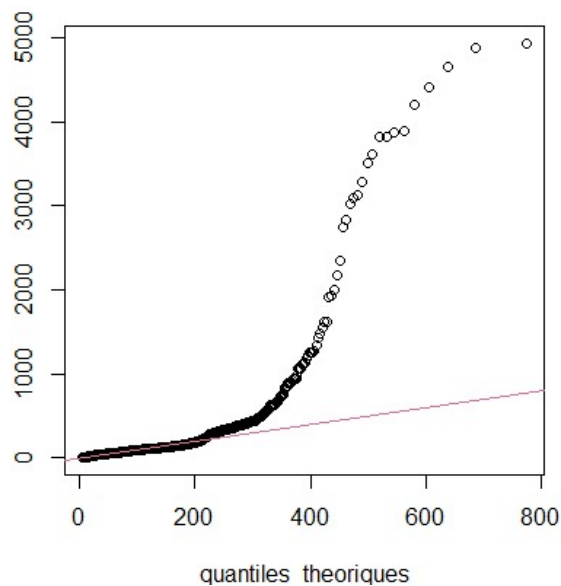


FIG. 2.9 : QQ-plot des pertes nettes, seuil 10%.

Selon les graphes QQ-plot la loi log normale sous-estime les pertes pour les périodes de retour élevées, pourtant pour les périodes de retour moyennes, inférieur à 100 ans, (ce sont ces périodes de retour qui intéressent le FSEC) la loi log normale fournit un bon ajustement. Ci- dessous les pertes correspondantes aux évènements catastrophiques de différentes périodes de retour.

Période retour (ans)	Pertes brutes	Pertes nettes	nettes /brutes
10	3074.789	151.138	4.915%
15	3335.393	175.120	5.250%
20	3517.912	192.851	5.482%
25	3658.622	207.041	5.859%
50	4093.477	253.718	6.158%
100	4528.604	304.630	6.726%

TAB. 2.3 : Pertes estimées en millions de DH pour un seuil de 10%

Pour un évènement de période de retour de 10 ans, on estime que la perte brute serait de l'ordre de 3074.789 millions de DH, pour faire face à cet évènement le FSEC doit allouer un montant de 151.138 millions de DH, soit 4.9% de la perte brute.

IV.4 Sensibilité de la perte nette au taux d'inhabitabilité

L'indemnisation n'intervient que pour les habitations inhabitables, un seuil d'inhabitabilité S est utilisé. Ce seuil correspond à un taux de destruction en dessous duquel le montant des dommages est nul :

$$Domage(DR < S) = 0$$

Dans le calcul des pertes nettes nous avons considéré que ce taux de destruction S est de 10%. Ce seuil de 10% n'est pas fixée par la loi donc l'ACAPS a le droit de choisir un autre seuil en fonction des circonstances. Dans cette partie nous allons étudier la sensibilité de la perte moyenne annuelle aux seuils d'inhabitabilité. On calcule les pertes nettes pour différents seuils d'inhabitabilité, ci-dessous la courbe des pertes nettes en fonction des seuils d'inhabitabilité.

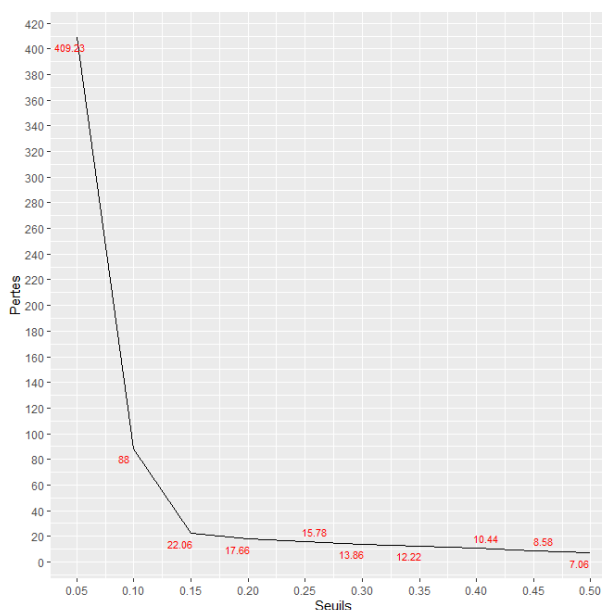


FIG. 2.10 : Sensibilité des pertes nettes aux seuils d'inhabitabilité.

Les pertes nettes sont très sensibles aux seuils d'inhabitabilité, la perte nette pour un seuil de 5% est de 409 Millions DH, en augmentant le seuil de 10% la perte chute de manière drastique à 88 millions de DH, en effet avec le seuil d'inhabitabilité à 10%, un grand nombre de taux de destruction sont éliminés, ce qui explique que les montants simulés soient plus faibles. En augmentant le seuil la perte nette continue à diminuer, pourtant à partir de seuil de 15% la diminution de la perte est faible et commence à se stabiliser. Ces résultats sont extrêmement importants en ce qui concerne la réassurance, on cherche à garder les 'bons risques' c'est-à-dire la partie dans laquelle les pertes sont moins volatiles et de céder aux réassureurs la partie dans laquelle les pertes sont plus volatiles.

La perte ne permet pas une analyse complète du risque, en effet la comparaison entre les pertes de deux communes dont les expositions sont différentes est erronée car les valeurs des biens exposés aux risques ne sont pas les mêmes. Donc il faut tenir compte de l'exposition.

il faut tenir compte de taux de dommage lors du calcul des expositions annuelles, une transformation de données est effectuée. Pour une commune i et une année de simulation j , l'exposition est donnée par :

$$\widehat{exposition}_{(i,j)} = \begin{cases} exposition_{(i,j)} & \text{si } DR \geq S \\ 0 & \text{sinon} \end{cases} \quad (2.6)$$

On calcule l'exposition annuelle de tout le territoire, pour une année j , l'exposition de tout le Maroc est donnée par :

$$\widehat{exposition}_{(.,j)} = \sum_{i=1}^n \widehat{exposition}_{(i,j)} \quad (2.7)$$

Année 1	Année 2	Année j	Année 10000
$\widehat{exposition}_{1,1}$	$\widehat{exposition}_{1,2}$		$\widehat{exposition}_{1,j}$		$\widehat{exposition}_{1,10000}$
$\widehat{exposition}_{2,1}$	$\widehat{exposition}_{2,2}$		$\widehat{exposition}_{2,j}$		$\widehat{exposition}_{2,10000}$
.....
$\widehat{exposition}_{i,1}$	$\widehat{exposition}_{i,2}$		$\widehat{exposition}_{i,j}$		$\widehat{exposition}_{i,10000}$
.....
$\widehat{exposition}_{x,1}$	$\widehat{exposition}_{x,2}$		$\widehat{exposition}_{x,j}$		$\widehat{exposition}_{x,10000}$
$\widehat{exposition}_{.,1}$	$\widehat{exposition}_{.,2}$		$\widehat{exposition}_{.,j}$		$\widehat{exposition}_{.,10000}$

FIG. 2.11 : Calcul des expositions .

Finalement on calcule l'exposition annuelle moyenne :

$$\widehat{exposition} = \sum_{j=1}^{10000} \frac{\widehat{exposition}_{.,j}}{10000} \quad (2.8)$$

On pose E_0 l'exposition brute qui correspond à un seuil S nulle des dommages, on cherche à étudier la variation des expositions par rapport à l'exposition brute en fonction des

seuils, on trace la courbe $\frac{\text{exposition}}{E_0}$ pour les différents seuils d'inhabitabilité.

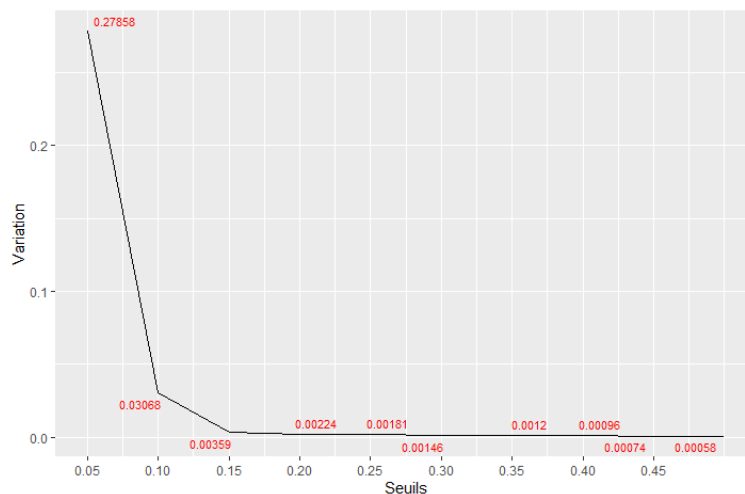


FIG. 2.12 : Sensibilité des expositions aux seuils d'inhabitabilité.

En passant d'un seuil d'inhabitabilité de 0% à un seuil de 5%, on perd déjà 72% de l'exposition brute, au seuil de 10% on ne garde que 3% de l'exposition totale. Ce qui explique le fait que le FSEC n'assure en moyen que $88/2047.19 = 4.3\%$ des pertes réellement engendrées par les inondations. On remarque que l'exposition réagit au changement du seuil de la même manière que les pertes, une chute de l'exposition en passant d'un seuil de 5% à un seuil de 15% puis une diminution moins forte en passant par les autres seuils.

V Conclusion

Dans cette partie nous avons exposé les modèles utilisés pour la modélisation des aléas catastrophiques notamment les inondations, nous avons également présenté l'outil MnhPRA qui a permis la modélisation des catastrophes au Maroc. Cet outil nous permet d'estimer les pertes brutes causées par les inondations ainsi que les pertes du FSEC. Compte tenu des données disponibles, un certain nombre de simplifications et d'hypothèses ont été posées (homogénéité de risque au sein des communes...). Ces limites constituent des pistes d'amélioration et d'ajustement qui pourront être réalisées au cours de la vie du modèle. Nous allons désormais présenter comment on peut construire un indice paramétrique qui nous permettra de céder le risque via une police d'assurance paramétrique.

Chapitre 3

Construction de l'indice

I l'indice paramétrique

L'étape fondamentale dans la mise en place d'une assurance paramétrique est la construction de l'indice paramétrique. Puisqu'il n'existe pas d'expertise avant indemnisation, l'indice a pour vocation d'être fortement corrélé aux pertes réellement réalisées, donc d'être clairement identifié, externe, objectif, facilement mesurable et relié à un long historique de données. Parmi les indices utilisés pour la mise en place d'une assurance paramétrique, on retient le cumul de pluie (cumul sur une période à déterminer). On peut construire un indice de cumul de pluie en ne prenant en compte que les pluies en deçà d'un seuil donné. (cette méthodologie a été adoptée dans la conception d'un produit d'assurance paramétrique contre le risque de pluie torrentielle au profit des pays de Caraïbes voir mémoire (FLORENT RITLENG 2014)).

Nous disposons d'une base de données des précipitations sur 30 min des sept provinces (Tanger, Tétouan, Kénitra, Salé, Rabat, Casablanca, Mohammadia) de 2000 jusqu'à 2020. Pour minimiser le risque il faut choisir un indice corrélé aux pertes réelles. Pour les inondations l'indice communément utilisé est lié au cumul des précipitations sur plusieurs jours. Après l'étude de plusieurs cumuls (cumul sur une journée jusqu'au cumul sur 5 jours) nous avons conclu que le cumul sur trois jours permet de mieux exhiber les inondations. En effet lors d'analyse de données des événements historiques nous avons remarqué qu'après la troisième journée l'intensité des précipitations devient très faible, ce qui a été confirmé dans les inondations de Mohammadia en 2002 (voir le graphique ci-dessous)

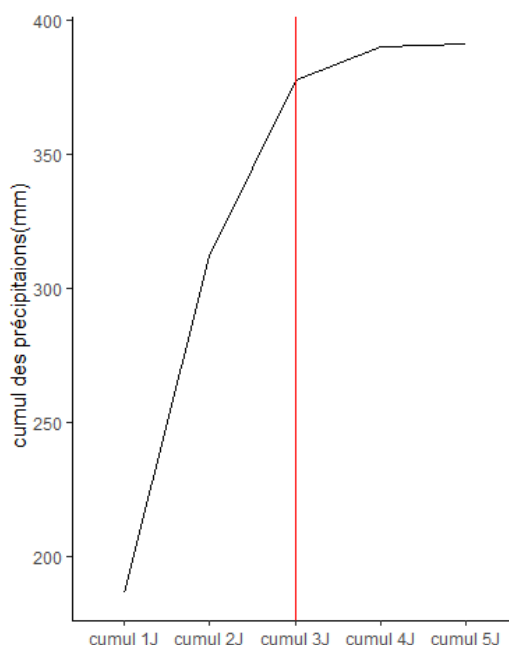


FIG. 3.1 : Etude des cumuls des précipitations

La convergence du cumul précipitations après trois jours s'explique par le fait que le nombre de jours consécutifs présentant de fortes précipitations ne dépasse pas 3 jours, donc si le cumul converge après 3 jours pour tous les événements de fortes précipitations on choisira le cumul sur jours comme base de notre indice.

Un évènement d'inondations est défini comme la période continue dont le cumul des précipitations sur trois jours dépasse un certain seuil (à déterminer). Le Maroc ne connaît pas des inondations chaque année, les deux dernières décennies on a vécu les inondations Mohammadia 2002, celles de Gharb en 2009, les inondations de Guillemim en 2014, donc une période de retour de 5 ans (voir 6 ans) or on estime que l'occurrence de ces évènements augmentera dans le futur à cause des changements climatiques, alors nous avons choisi un seuil qui représente les précipitations de période de retour de 2ans. ce seuil correspond à un cumul des précipitations sur trois jours de 200 mm (ce seuil a été déterminé après la calibration de la loi du cumul des précipitations sur 3 jours par la théorie des valeurs extrêmes, voir partie Return level estimation).

La construction de l'indice passe par 4 étapes :

1. Sur chacune des Provinces du Maroc, le cumul des 3 derniers jours de pluie est calculé jour après jour,

$$\text{Pluie agrégée sur 3 jours} = \sum_{j=0}^2 \text{pluie journalière}_j$$

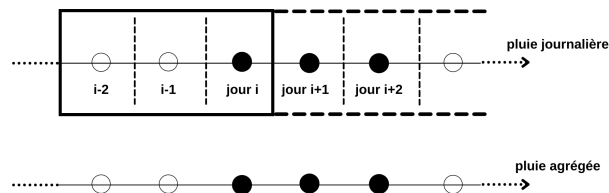


FIG. 3.2 : Agrégation des pluies journalières sur 3 jours

L'évènement se déclenche lorsque ces précipitations cumulées sur 3 jour dépassent 200 mm et se termine lorsqu'il redescend en-dessous de ce seuil. Ce seuil de 200 mm correspond à des évènements dont la période de retour est de 2 ans (voir graphique ci-dessous) ;

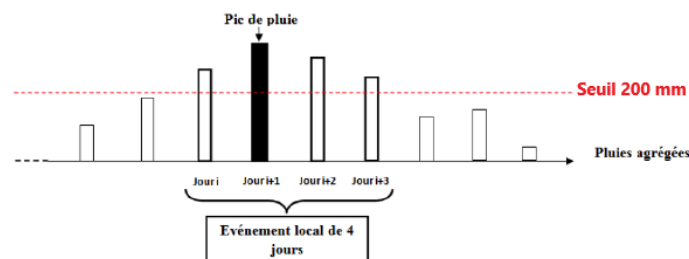


FIG. 3.3 : Détection des évènements extrêmes

2. Sur chacun des évènements détectés, seul le pic des pluies cumulées sur 3 jours est utilisé pour calculer un indice de perte via une courbe de vulnérabilité qui relie la quantité de pluie à un taux de dommage ;

3. Pour tenir compte des différences en exposition entre les provinces, il faut appliquer un taux d'exposition au taux de dommage. En effet par un taux de dommages de 10% à Casablanca n'est pas équivalent à un taux de 10% à Mohammadia, car les deux provinces n'ont pas la même exposition au risque, les pertes à Casablanca serait supérieur à celles de Mohammadia. La part d'exposition de la province géographique par rapport à l'exposition globale est donc appliquée au taux de vulnérabilité de chaque événement local afin de calculer l'indice de perte de la province sur l'événement local.
4. Pour calculer l'indice national, les indices provinciaux sont additionnés. Au niveau national, un événement est défini comme une période continue où l'indice local est positif dans au moins une des provinces.

$$indice\ national = \sum_{i=1}^7 I_i$$

II Fonction de vulnérabilité

Une fois que les événements provinciaux ont été détectés, nous calculons les indices de pertes associés. Le calcul de l'indice de perte se base seulement sur le pic du cumul pluie de l'événement considéré qui est appliqué à une fonction de vulnérabilité pour déterminer un taux de vulnérabilité (taux de destruction). La seule information que nous connaissons est qu'elle est croissante (lorsque l'intensité des précipitations augmente, le taux de destruction augmente), nous avons alors choisie une fonction linéaire simple pour calculer les indices de perte, la figure 3.4.

Par manque de données sur les pertes réelles, cette fonction a volontairement été choisie positive à partir de 200 mm : à tous les événements détectés sera reversée une indemnité. Le maximum du cumul des précipitations sur 3 jours dans toutes les provinces est de 498.14 mm (survenu à Tanger), nous avons choisi le niveau de 1000 mm de pluie pour atteindre un seuil de destruction de 100 %, Nous reviendrons sur ce point pour tester la sensibilité de la prime à ce paramètre.

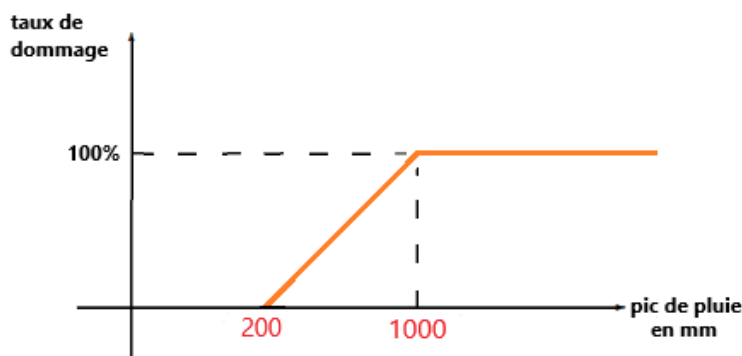


FIG. 3.4 : Fonction de vulnérabilité.

III Exposition

La vulnérabilité seule ne permet pas d'expliquer les pertes, un taux de dommage de 10% à Casablanca n'engendre pas forcément les mêmes pertes qu'un un taux de dommage de 10% à Mohammadia par exemple, car la valeur des biens exposés au risque n'est pas la même. On rappelle qu'après le calcul de taux de dommage à partir de pic des précipitations via la fonction de vulnérabilité, afin de relativiser il faut le multiplier par un taux d'exposition pour trouver l'indice paramétrique.

Le FESC possède une base de données d'exposition, estimées à partir d'outil Mnh-PRA, cet outil permet d'estimer les expositions aux pertes sur chacune des provinces. Les expositions se basent sur des statistiques économiques et démographiques. (voir tableau V.2 en Annexe)

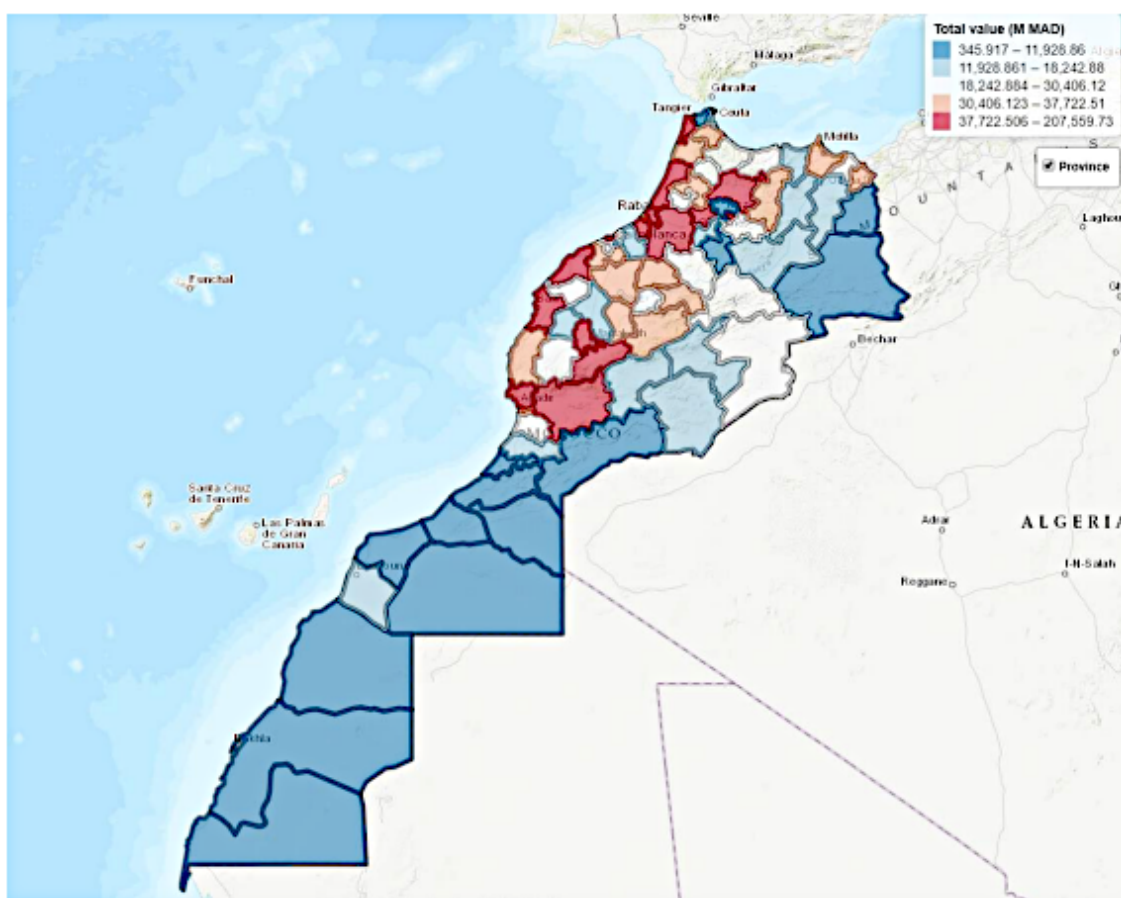


FIG. 3.5 : Carte des valeurs exposées par province

La figure ci-dessus, représentant la carte des valeurs exposées par province, montre que la répartition des valeurs exposées sur le territoire Marocain est inégale. Ce sont essentiellement les grandes métropoles qui concentrent les valeurs les plus élevées, notamment les provinces que nous avons choisie dans le cadre de ce mémoire.

Nous calculerons donc un taux d'exposition (TE) pour chaque province :

$$TE_i = \frac{Exp_i}{\sum_{i=1}^n Exp_i} \quad (3.1)$$

avec Exp_i est l'exposition de la province i et n le nombre totale des provinces.

IV Fonction de paiement

la fonction de paiement permet de fixer les indemnités payées en fonction de l'intensité de la catastrophe. Par intensité, on désigne des facteurs physiques comme les précipitations, c'est justement l'indice paramétrique qui va mesurer l'intensité de la catastrophe. Afin minimiser le risque de base, cette fonction devrait être calibrée à partir de la fonction de dommage mais par absence de données sur les dommages causés par les évènements historiques, on choisit une fonction de paiement linéaire. Et pour tenir compte du profil de risque des provinces, chaque province aura sa propre fonction de paiement.

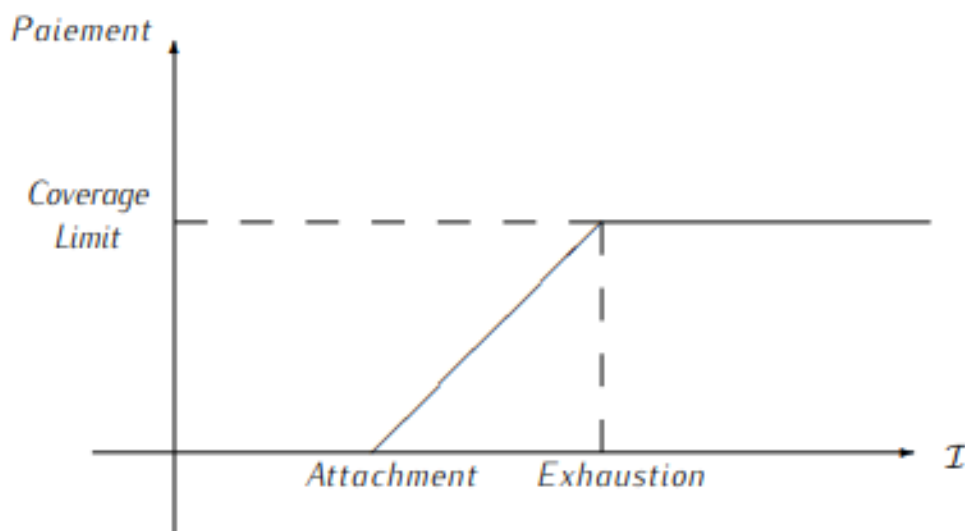


FIG. 3.6 : Fonction de Paiement

La borne inférieure, communément appelée Attachement est fixée de façon que tous les sinistres ayant un indice inférieur à cette borne ne soient pas pris en charge. Pour le FSEC on fixe une borne supérieure afin de limiter le montant de la prime. La frontière supérieure est usuellement nommée Exhaustion. Ces limites sont définies comme des quantiles bien précis de la distribution de probabilité de l'indice paramétrique.

$$\text{Paiement} = \begin{cases} 0 & \text{si } \text{Indice} \leq \text{Attachment} \\ \frac{\text{Indice} - \text{Attachment}}{\text{Exhaustion} - \text{Attachment}} \times \text{Coverage limit} & \text{si } \text{Attachment} < \text{Indice} < \text{Exhaustion} \\ \text{Coverage limit} & \text{si } \text{Indice} \geq \text{Exhaustion} \end{cases} \quad (3.2)$$

Attachment correspond à un seuil dès que l'indice le franchit, la police d'assurance se déclenche. L'ACAPS a fixé un taux de dommages minimal de 10% à partir duquel le propriétaire de la résidence a droit à l'indemnisation, pourtant ce seuil n'est pas fixé dans la loi et l'ACAPS peut toujours choisir un seuil plus faible, une raison de ne pas choisir ce seuil est le fait que le taux de 10% est fixé pour les résidences, l'hypothèse de

l'homogénéité des risques au sien de la commune peut être acceptable mais au sien de la même province est loin d'être réaliste. Donc on choisit 5% comme seuil.

Attachement de la province $j = 5\% \times \text{taux exposition}_j$

L'indemnisation est bornée supérieurement par le Coverage Limit. Cette limite est atteinte lorsque l'indice vaut Exhaustion, autrement dit le coverage limite est la perte à laquelle le FSEC devra faire face lorsque l'indice vaut Exhaustion. La borne Exhaustion correspondra à un quantile précis des indices.

Chapitre 4

Modélisation des précipitations extrêmes

I Introduction

On s'intéresse dans l'étude des valeurs extrêmes aux grandes valeurs d'un échantillon de variables aléatoires afin de caractériser sa loi. Cette approche vient pour compléter la statistique classique qui se concentre généralement sur l'étude du comportement des variables aléatoires autour de leurs valeurs moyennes. Une théorie dédiée à l'étude de ces valeurs particulières a été développée depuis les années 70 et suscite l'intérêt de nombreux statisticiens, ingénieurs et scientifiques tant le champs d'applications qu'elle touche est vaste. Il s'agit dans cette théorie de caractériser le comportement des queues de distribution à l'aide de modèles spécifiques permettant un bon ajustement des extrêmes de la distribution. On s'intéresse alors aux événements rares, qui se produisent avec des faibles probabilité. La théorie des extrêmes offre une classe d'outils permettant une extrapolation dans les queues de distribution, à partir des valeurs maximales observées, afin de prédire l'apparition de valeurs non observées. Les champs d'applications de cette classe d'outils est de plus en plus vaste. On retrouve la théorie des extrêmes dans les domaines suivants :

- Finance : gestion de portefeuille, évaluation du risque sur les marchés, prédiction de crises ;
- Environnement : évolution du climat, concentration de la pollution, vitesse du vent, surveillance des crues de cours d'eau.

II Théorie des valeurs extrêmes

Il s'agit dans l'étude des valeurs extrêmes d'analyser l'épaisseur des queues de distributions, ou encore d'étudier les plus grandes observations d'un échantillon pour caractériser sa loi initiale. Ainsi, la théorie des extrêmes vient en complément de la théorie statistique classique où il est plus commun d'étudier le comportement d'une distribution autour de sa moyenne plutôt que dans le domaine des observations extrêmes souvent appelées événements rares. Nous allons voir que toute la théorie des extrêmes est fondée sur un équivalent au théorème central limite mais pour les queues de distribution.

L'étude des extrêmes d'un processus se base sur l'analyse du maximum d'un échantillon de taille n donnée $M_n = \max(X_1, X_2, \dots, X_n)$ où X_1, \dots, X_n sont des variables aléatoires indépendantes identiquement distribuées ayant pour fonction de répartition F . En théorie, la distribution de M_n peut être obtenue en fonction de F : soit $M_n = \max((X_i)_{i \geq 1})$.

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x)$$

$$P(M_n \leq x) = P(X_1 \leq x) \times P(X_2 \leq x) \times \dots \times P(X_n \leq x)$$

$$P(M_n \leq x) = P(X_1 \leq x)^n = (F(x))^n$$

En pratique, F est inconnue, et la relation précédente n'est donc pas utilisable directement, pourtant on peut toujours calibrer la loi des X_i et en déduire celle de M_n . Or les lois utilisées pour la modélisation des X_i sont des lois classiques : la loi exponentielle, la loi Gamma, la loi de Weibull, la loi Log-normale ... Le problème c'est que ces lois classiques négligent trop souvent et fortement le poids des queues de distribution, celles-ci n'en

représentant qu'une portion minime. Les calibrations usuelles et centrales permettent d'être proche de la réalité "en moyenne". Nous cherchons quant à nous à être proche de la réalité en "quantile élevé". Les maximas exigent une calibration propre à eux. Une autre alternative présuppose que l'on conserve F comme inconnue, et que l'on approche cette distribution par des distributions limites estimées dans le domaine des observations extrêmes.

Notons x^F le point extrémal de $F, x^F = \sup(x|F(x) \leq 1)$. Le support de F peut être borné ($x^F \leq \infty$:loi uniforme, loi beta ...) ou infini ($x^F = \infty$:loi de pareto), autrement dit, la distribution asymptotique de M_n est dégénérée ; il est donc nécessaire de normaliser la suite.

Le premier théorème de la théorie des valeurs extrêmes précise les lois asymptotiques que peut suivre le maximum normalisé d'une suite de variables aléatoires IID.

Théorème(Fisher - Tippett) S'il existe des suites de réels a_n et b_n telles que, quand $n \rightarrow \infty$, $P((M_n - a_n)/b_n \leq x) \rightarrow H_\xi(x)$ pour une distribution non-dégénérée $H_\xi(x)$, alors $H_\xi(x)$ est une distribution de Fréchet($\xi > 0$), de Gumbel($\xi = 0$) ou Weibull ($\xi < 0$). avec :

Fréchet($\xi > 0$) :lorsque $\xi > 0$, de telles distributions possèdent des queues lourdes et la convergence de F^n vers la loi limite se fait très lentement.

$$\Phi_\xi(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\xi}) & \text{sinon} \end{cases} \quad (4.1)$$

Weibull ($\xi < 0$) : lorsque $\xi < 0$, dans ce cas la queue de la distribution sera très mince.

$$\Psi_\xi(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-(-x^{-\xi})) & \text{sinon} \end{cases} \quad (4.2)$$

Gumbel($\xi = 0$) : lorsque $\xi = 0$, la distribution H_0 est appelée distribution de Gumbel. Le support de cette loi est \mathbb{R} et dans ce cas les queues de distribution sont légères et décroissent de manière exponentielle. Son expression est donnée par :

$$\Lambda_\xi(x) = \exp(-\exp(-x))$$

On dit que F appartient au domaine d'attraction de Fréchet (respectivement de Gumbel ou de Weibull) lorsque la loi limite des maximas suit une loi de Fréchet (respectivement de Gumbel ou de Weibull).

Ces trois lois caractérisent chacune un comportement spécifique des valeurs extrêmes. On les distingue à travers deux caractéristiques :

- La vitesse de décroissance de la probabilité des grandes valeurs. Si la décroissance est rapide, de type exponentielle, alors la loi des extrêmes est la loi de Gumbel. Si au contraire, elle est plus lente, de type puissance, on se réfère au second critère :
- L'étendue de variation des données. Si celle-ci est bornée, alors la loi des extrêmes est la loi de Weibull, si ce n'est pas le cas, c'est la loi de Fréchet.

Ce théorème donne un résultat particulièrement intéressant puisqu'il permet, quelle que soit la loi de la variable parente, de déterminer la distribution du maximum sous certaines hypothèses généralement admises dans les modélisations actuarielles. Le comportement des queues de distribution de la plupart des lois usuelles peut donc s'expliquer par une seule loi définie ci-dessous dont un unique paramètre ξ en contrôle la lourdeur.

Pour résoudre des problèmes statistiques, (VON MISES 1954) et (JENKINSON 1955) ont proposé la loi d'extrémum généralisée (notée **GEV** pour *Generalized Extreme Value*) qui unifie les distributions de Fréchet, Weibull et Gumbel

Définition 2(*loi d'extrémum généralisée*) : la loi d'extrémum généralisé $G(\mu, \sigma, \xi)$ est définie par la fonction de répartition :

pour $\xi \neq 0$:

$$G(\mu, \sigma, \xi) = \exp\left(-\left(1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

pour $\xi = 0$:

$$G(\mu, \sigma, \xi) = \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right)$$

Cette distribution possède trois paramètres :

- μ le paramètre de localisation, qui est un indicateur de la taille moyenne des extrêmes dans la distribution parente ;
- σ paramètre de dispersion, il indique l'étalement des extrêmes ;
- ξ l'indice de queue qui donne une indication de la forme ou le poids des queues de la distribution limite. Plus ce paramètre de queue diminue plus la distribution considérée est à queue épaisse.

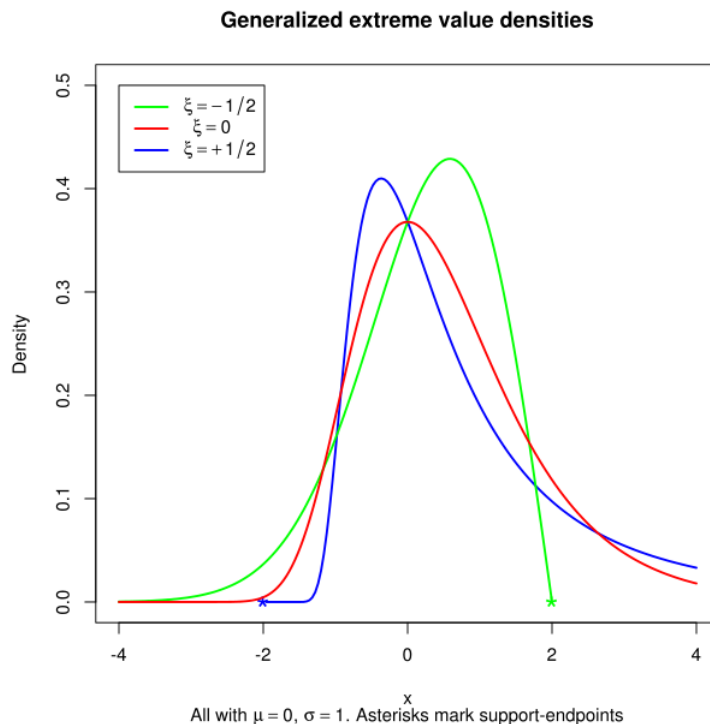


FIG. 4.1 : Densités de la loi des valeurs extrêmes pour différentes valeurs de ξ

Définition 1 (*Domaine d'attraction*) : F appartient au domaine d'attraction de G ($F \in D(G)$) s'il existe deux suites (a_n) et (b_n) telles que la convergence précédente (Fisher-Tippett)) ait lieu.

En pratique un grand nombre de distributions vérifient le théorème. Les lois usuelles peuvent être classées selon leur domaine d'attraction :

- **Weibull** ($\xi < 0$) : Uniforme, Bêta
- **Fréchet** ($\xi > 0$) : Cauchy, Student, Pareto
- **Gumbel** ($\xi = 0$) : Normale, Gamma, Exponentielle

II.1 Estimation des paramètres et calibration

On donne la forme généralisée de la loi des valeurs extrêmes GEV :

pour $\xi \neq 0$:

$$G(\mu, \sigma, \xi) = \exp\left(-\left(1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

pour $\xi = 0$:

$$G(\mu, \sigma, \xi) = \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right)$$

On cherche à estimer le paramètre $\theta = (\mu, \sigma, \xi)$. La première difficulté technique vient du fait qu'on ne peut pas affirmer que la distribution de X suit exactement la

loi GEV, seul son maximum suit cette hypothèse. Or le maximum ne représente qu'une seule observation il devient donc difficile de calibrer la loi et de déterminer le paramètre. Pour pallier ce problème, Gumbel a introduit en 1958 (GUMBEL 1958) une méthode "block maxima", l'idée est de découper les observations en m échantillons de taille K . Sur chacun de ces blocs, on prend en compte le maximum et on obtient donc m valeurs, ces valeurs sont donc traitées comme un échantillon de la variable M_n , puis de calibrer la loi sur cet échantillon. nous pouvons désormais utiliser nos outils classiques pour estimer les paramètres. On utilise donc la méthode de maximum de vraisemblance pour estimer $\theta = (\mu, \sigma, \xi)$.

Hypothèses

Nous considérons une suite de variables indépendantes et identiquement distribuées (IID) X_1, X_2, \dots, X_N de loi F inconnue. Ce sont nos sinistres. Nous les rangeons dans l'ordre chronologique. On divise cet échantillon de taille N en k blocs de taille n .

Si $k \times n > N$ alors le dernier bloc contiendra un nombre de valeurs inférieur à n . Pour chaque bloc, nous retenons le coût de sinistre maximum observé. Ainsi, nous obtenons un vecteur de k maximums Y_1, Y_2, \dots, Y_k .

notons $\theta = (\mu, \sigma, \xi)$ et m_1, \dots, m_k les maxima issus des sous-échantillons. La log-vraisemblance prend une expression différente selon la valeur du paramètre de queue :

pour $\xi \neq 0$:

$$L(\theta, m_1, \dots, m_k) = -K \ln(\sigma) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^k \ln\left(1 + \xi \left(\frac{m_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^k \left(1 + \xi \left(\frac{m_i - \mu}{\sigma}\right)\right)^{-1/\xi}$$

pour $\xi = 0$:

$$L(\theta, m_1, \dots, m_k) = -K \ln(\sigma) - \sum_{i=1}^k \left(\frac{m_i - \mu}{\sigma}\right) - \sum_{i=1}^k \exp\left(-\frac{m_i - \mu}{\sigma}\right)$$

Aucune solution analytique aux équations annulant les dérivées n'existe, il faut donc avoir recours à une résolution numérique. Le calcul du maximum de vraisemblance ne pose pas de difficultés (EMBRECHTS P. 2001), et l'estimateur conserve les propriétés classiques de l'Estimateur du Maximum de Vraisemblance (EMV) tant que $\xi > -1/2$ (SMITH 1985)

La GEV étant une loi limite pour le maximum, il est important que les blocs utilisés pour modéliser la loi soient de taille suffisamment grande. La largeur des blocs réduit cependant l'information disponible. Considérons par exemple les maximas annuels d'une série de pluies journalières, les blocs sont les années. Du fait de l'échantillonnage, il peut arriver qu'une année contienne plusieurs valeurs extrêmes, et qu'une autre année n'en contienne aucune. Dans le premier cas de figure, une seule réalisation de valeur extrême est conservée, on perd des maximas potentiels. Ce phénomène de « superposition de maxima » s'accroît avec la taille des blocs. Dans le deuxième cas de figure, le maximum n'est pas une réalisation de valeur extrême. Un autre point faible de la méthode block-maxima réside dans la perte d'information induite par le choix de modéliser le maximum uniquement. En

effet on souhaite avoir une distribution des valeurs extrêmes sans s'arrêter au maximum. Pour pallier ces défauts, il conviendrait de pouvoir modéliser les valeurs les plus élevées sans tenir compte de leur position relative au sein de l'échantillon en considérant toutes les valeurs au-dessus d'un certain seuil par exemple.

II.2 Analyse par les dépassements d'un seuil élevé

La deuxième approche de la théorie des valeurs extrêmes est l'étude du comportement des valeurs observées au-delà d'un seuil.

Considérons toujours un échantillon IID, de fonction de répartition F . Il s'agit d'étudier l'échantillon X_i pour tout i tel que $X_i > u$.

Définition 3 (*Loi des excès*) : Pour un seuil u fixé, nous notons F_u la fonction de répartition de la loi des excès, définie par

$$F_u(x) = P(X - u \leq x | X > u)$$

Le second théorème fondamental de la théorie des valeurs extrêmes indique que la convergence du maximum de variables aléatoires IID correctement normalisées est équivalente à la convergence de la distribution des excès vers une loi de Pareto généralisée (notée GPD pour Generalized Pareto Distribution)

Définition 4 (*Loi de Pareto généralisée*) : La loi de Pareto généralisée $GPD(\xi, \tilde{\sigma})$ est définie par la fonction de répartition :

pour $\xi \neq 0$:

$$G_{\xi, \tilde{\sigma}}(x) = (1 - (1 + \xi \frac{x}{\tilde{\sigma}})^{-1/\xi})$$

pour $\xi = 0$:

$$G_{\xi, \tilde{\sigma}}(x) = (1 - \exp(-\frac{x}{\tilde{\sigma}}))$$

Cette définition permet de faire la correspondance entre les lois GEV et GPD. En effet, le paramètre de forme ξ reste toujours le même. Les lois GPD se réduisent à deux paramètres $\tilde{\sigma}$ et ξ , il n'y a plus le terme de localisation μ comme dans les loi GEV. Toutefois, il existe une relation entre les μ , σ d'une GEV et $\tilde{\sigma}$ de la GPD correspondante.

$$\tilde{\sigma} = \sigma + \xi(u - \mu)$$

Ainsi, si une loi GPD est adaptée au problème de dépassement de seuil u_0 , alors les excès au delà du $u \geq u_0$ devraient suivre aussi une loi GPD. Le paramètre de forme ξ est identique pour les deux distributions. Cependant, le paramètre d'échelle σ pour les excès au-delà de $u \geq u_0$ est donné par :

$$\sigma(u) = \sigma(u_0) + \xi(u - u_0)$$

σ varie donc linéairement en fonction de u , cette relation n'est valable que si $\xi \neq 0$.

Illustration :

soient X_1, X_2, \dots, X_n n réalisations IID d'une distribution F . Soit u un seuil assez grand on note $N_u = \{X_i | X_i > u\}$ et les Y_i les excès correspondants. Au-delà d'un seuil u , le théorème de Balkema-de Haan-Pickands montre que la fonction de distribution des excès F_u peut être approchée par $G_{\xi, \tilde{\sigma}}$

La principale problématique consiste désormais à déterminer le seuil u , il existe différentes méthodes pour choisir le seuil u au-dessus duquel les valeurs sont modélisées par une loi GPD. La sélection de seuil peut s'avérer délicate car il faut faire un compromis entre le biais de modèle et la variance du modèle. En effet, une trop faible valeur de seuil u peut introduire un biais important dans le modèle tandis qu'un seuil trop élevé générera trop peu d'observations sur lesquelles le modèle peut être estimé, donc une variance trop forte.

II.3 Détermination de seuil de la loi de Pareto Généralisée

Le choix du seuil u est très important car il induit une grande variabilité dans l'estimation des paramètres de la loi des dépassements. Il existe plusieurs approches pour le choix du seuil u . En effet, le seuil doit être suffisamment élevé pour remplir le caractère asymptotique du modèle, mais pas trop grand, afin de maintenir un nombre suffisant d'observations pour estimer correctement les paramètres du modèle.

II.3.1 La méthode des excès moyens

Lorsqu'une variable aléatoire suit une loi de Pareto généralisée, sa fonction moyenne des excès (notée MEF pour Mean Excess Function) présente une propriété que nous allons utiliser pour déterminer le seuil des valeurs extrêmes afin de sélectionner le seuil adéquat, que nous présentons ici.

Définition (la fonction des excès moyens) : Soit X une variable aléatoire, on définit la fonction des excès moyens, notée $e(u)$, par :

$$e(u) = E(X - u | X > u)$$

Propriété : Soit X une variable aléatoire qui une suit $G(\xi, \sigma), E(X) < \infty$. si $\xi < 1$ alors :

$$e(u) = \frac{\xi}{1 - \xi} u + \frac{\sigma}{1 - \xi}$$

En pratique, si N_u est le nombre d'observations au-dessus du seuil u , on a l'estimateur empirique suivant de $e(u)$:

$$\hat{e}(u) = \frac{1}{N_u} \sum_{i=1}^n (X_i - u) 1_{\{X_i > u\}}$$

$e(u)$ est donc linéaire en u , cette propriété nous permet une détermination graphique du seuil u à partir du graphe des estimateurs $\hat{e}(u)$ calculés pour tous les seuils,

nous cherchons la valeur u à partir de laquelle la courbe a une forme linéaire. En effet, si le Mean Excess plot semble avoir un comportement linéaire au dessus d'une certaine valeur positive de u , cela signifie que les excès au-delà de ce seuil suivent une distribution de pareto généralisée.

II.3.2 Hill Plot

Pour cette méthode il faut calculer un estimateur de ξ (valable pour $\xi > 0$). Lorsque nous utilisons tout l'échantillon pour cela, sa valeur est biaisée car seuls les dépassements au-dessus d'un seuil élevé suivent une loi GPD. L'idée est de calculer cet estimateur avec un grand échantillon et d'enlever au fur et à mesure les observations les plus petites, jusqu'à ce que l'estimateur se stabilise. Le point où il se stabilise est notre seuil. l'estimateur de Hill que nous allons utiliser est construit à partir des $k+1$ plus grandes valeurs de l'échantillon, autrement dit celles qui dépassent la statistique d'ordre $X_{n-k,n}$ on dit que celle-ci joue le rôle de seuil.

l'estimateur de Hill est donné par la formule :

$$\widehat{H}_{k,n} = \frac{1}{k} \sum_{i=1}^k (\log(X_{n-i+1,n}) - \log(X_{n-k,n}))$$

L'estimateur de Hill est consistant et asymptotiquement convergent, mais ces propriétés pour des petits échantillons ne sont pas bien connues. Cet estimateur est par ailleurs très sensible au choix de u . Cependant, malgré ces réserves, l'estimateur de Hill est considéré comme un bon estimateur de l'indice de queue.

Le graphique de Hill ou *Hill plot* se construit à la fois sur le nombre d'observations et l'estimation de l'indice de queue communément appelé l'estimateur de Hill. Il permet de déterminer le nombre d'observations à partir duquel se forment les valeurs extrêmes. La valeur de l'observation correspondante à ce nombre constitue le seuil.

Il faut noter que l'estimateur de Hill ne s'applique qu'aux distributions à queues épaisses c'est à dire les distributions ayant un indice de queue positif. Cela constitue une limite d'application du graphique de Hill si nous avons à faire à distributions normales ou des distributions à queues fines ayant un indice queue inférieur à zéro. En utilisant cette méthodologie, le choix du seuil se fait en observant la stabilité de la valeur de l'indice de queue. Par conséquent, la stabilité de l'estimateur de Hill est d'une grande importance pour déterminer un seuil optimal. En effet, le nombre d'excès correspondant au seuil est déterminé par l'observation sur le graphique de la zone de stabilité (zone où la valeur de l'indice de queue ne varie pas assez) de la valeur de l'indice de queue. Le nombre d'observations à retenir correspond au début de la zone de stabilité.

II.4 Estimation des paramètres de la loi de Pareto Généralisée

Une fois que le seuil u est sélectionné, les paramètres d'une loi $GPD(\sigma, \xi)$ peuvent être estimés par maximum de vraisemblance ou par la méthode des moments.

II.4.1 Méthode de maximum de vraisemblance

: Les paramètres ξ et σ sont généralement estimés par les méthodes de maximum de vraisemblance

La fonction de log-vraisemblance pour un échantillon de n excès Y, \dots, Y_n IID suivant une loi de Pareto Généralisée est donnée par (SMITH 1985) :

pour $\xi \neq 0$:

$$L(Y, \xi, \sigma) = -n \ln(\sigma) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^n \ln\left(1 + \xi \frac{y_i}{\sigma}\right)$$

pour $\xi = 0$:

$$L(Y, 0, \sigma) = -n \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^n y_i$$

Le problème de maximisation de la log-vraisemblance nécessite d'avoir recours à des méthodes numériques, type algorithme de Newton-Raphson.

II.4.2 Méthode des moments

La méthode des moments est la technique la plus simple pour ajuster une loi paramétrique. En général, les estimateurs calculés par cette méthode ne sont pas optimaux, mais ont l'avantage d'être rapides à implémenter et ne demandent pas beaucoup de temps de calcul.

La méthode des moments consiste à déterminer les estimateurs des paramètres de la loi en faisant coïncider les moments m_j de la loi avec leurs estimateurs empiriques. Pour estimer ξ et σ dans le cas de la loi de Pareto généralisée, on utilise les deux premiers moments de la loi pour disposer de deux équations. Ces deux moments s'expriment ainsi en fonction des paramètres de la loi :

$$E(Y) = \frac{\sigma}{1-\xi}$$

$$V(Y) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}$$

Ces deux quantités existent uniquement si $\xi < 1/2$. on en déduit les estimateurs de la méthode des moments, sous l'hypothèse $\xi < 1/2$:

$$\hat{\xi} = \frac{\bar{Y}^2 - s^2}{2s^2}$$

$$\hat{\sigma} = \frac{\bar{Y}(\bar{Y}^2/s^2 + 1)}{2}$$

avec :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

II.5 Return level estimation

L'intérêt ne réside généralement pas dans les estimations des paramètres GPD eux-mêmes, mais dans la façon dont nous pouvons utiliser le modèle ajusté pour estimer d'autres quantités, tel que les précipitations ayant des périodes de retour 10 ans, pour déterminer le dimensionnement d'un réseau d'assainissement qui permettra la protection contre ce genre d'évènement. Ces quantités, dans une terminologie de valeurs extrêmes, sont généralement appelées "return levels".

Période de retour Soit X , une variable aléatoire, F sa fonction de répartition avec : $F(x) = P(X \leq x)$, on pose $S(x) = 1 - F(x) = P(X > x)$ la fonction de survie.

Par définition la période de retour d'un évènement est définie par :

$$PDR(x) = \frac{1}{S(x)}$$

$PDR(x)$ représente la période au bout de laquelle la valeur x peut être dépassé.

$$P(X < x) = 1 - \frac{1}{PDR(x)}$$

Donc le niveau de retour x peut être calculé comme suit

$$Return\ Level = x = quantile\left(1 - \frac{1}{PDR(x)}\right)$$

nous pouvons estimer r -ans "return level" z_r , dans le cas des dépassements de seuils, pour n'importe quelle période en posant :

$$Pr(X \geq u + y | X \geq u) = \frac{1}{r}$$

Avec

$$Pr(X \geq u + y | X \geq u) = \left[1 + \frac{\hat{\xi} \times y}{\hat{\sigma}}\right]^{-\frac{1}{\xi}}$$

or pour $\xi \neq 0$

$$Pr(X \geq u + y | X \geq u) = \frac{Pr(X \geq u + y)}{Pr(X \geq u)}$$

$$Pr(X \geq u + y) = Pr(X \geq u + y | X \geq u) \times Pr(X \geq u)$$

on remplace $Pr(X \geq u + y | X \geq u)$ par son expression , on aura,

$$Pr(X \geq u + y) = \hat{\lambda} \times \left[1 + \frac{\hat{\xi} \times y}{\hat{\sigma}}\right]^{-\frac{1}{\xi}}$$

avec $\hat{\lambda} = Pr(X \geq u)$ estimé comme le taux empirique de dépassement des seuils. Ainsi, une estimation du niveau z_t qui est dépassé en moyenne une fois chaque t observations est obtenue comme solution de

$$\hat{\lambda} \times \left[1 + \frac{\hat{\xi} \times (z_t - u)}{\hat{\sigma}} \right]^{\frac{-1}{\hat{\xi}}} = \frac{1}{t}$$

donc pour $\xi \neq 0$

$$\hat{z}_t = u + \frac{\hat{\sigma}}{\hat{\xi}} [(t\hat{\lambda})^{\hat{\xi}} - 1]$$

et pour $\xi = 0$

$$\hat{z}_t = u + \hat{\sigma} \log(t\hat{\lambda})$$

Par construction, z_t est le niveau de retour t -observation; toutefois, il est souvent plus commode de donner des "return levels" sur une échelle annuelle de sorte que le r -year "return level" est le niveau espéré à être dépassé une fois tous les r ans. S'il y a n observations par an, cela correspond au "return level" de t -observation avec $t = r \times n$.

II.6 Problème de dépendance

Il a été montré que les paramètres de la distribution limite sont affectés par la dépendance dans la série. Nous pouvons l'examiner en comparant les distributions des maximas d'une série stationnaire et d'une série de variables indépendantes ayant la même distribution marginale. Il est habituel de supposer une condition qui limite l'étendue des dépendances à des niveaux extrêmes, de sorte que les événements $X_i \geq u$ et $X_j \geq u$ (avec $j > i$) sont approximativement indépendants, à condition que u soit suffisamment élevé et que les points temporels i et j soient suffisamment éloignés. En d'autres termes, les événements extrêmes sont considérés indépendants si la durée entre leur survenance est suffisamment élevée. Dans l'ajustement des précipitations journalières on peut supposer l'indépendance des extrêmes, cette hypothèse d'indépendance entre les précipitations journalières semble acceptable, cependant l'indépendance des précipitations agrégées ne paraît pas être recevable. Rappelons que pour construire notre indice paramétrique, il est nécessaire de calculer l'agrégation des pluies 3 jours. Si on prend par exemple le cas des précipitations cumulées sur 3 jours, si le cumul sur $[i, i+3]$ dépasse un seuil S , il est fort probable que le cumul sur $[i+1, i+4]$ dépassera aussi ce seuil S , de même pour le cumul sur $[i+2, i+5]$. (voir figure 4.2)

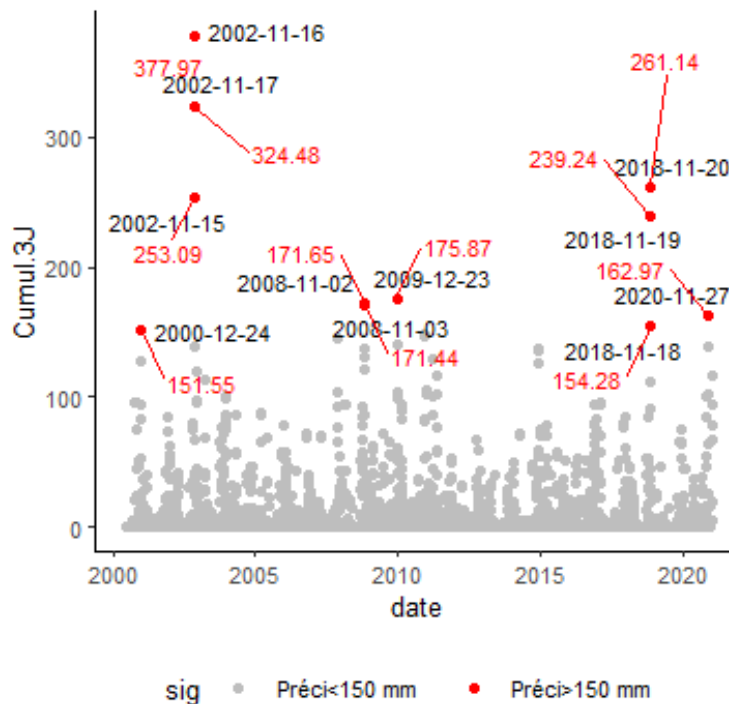


FIG. 4.2 : Illustration de la dépendance des extrêmes (données : précipitations Rabat) .

Les points en rouge représentent les précipitations cumulées sur 3 jours dont le cumul dépasse le seuil de 150 mm (choisi arbitrairement), on peut bien remarquer que les extrêmes ont tendance à se regrouper, ce qui implique qu'une journée dont le cumul est extrêmement élevé est susceptible d'être suivie d'une autre, donc l'hypothèse d'indépendance n'est plus valable, la dépendance rend la log-vraisemblance invalide. De plus, il n'y a pas de théorie générale pour fournir une vraisemblance comparative qui incorpore la dépendance dans les observations. Donc un changement en pratique est nécessaire.

Diverses suggestions, avec différents degrés de sophistication, ont été adoptées pour traiter le problème des dépassements dépendants. La méthode qui est largement adoptée est le 'declustering' qui correspond à un filtrage des observations dépendantes pour obtenir un ensemble de dépassements de seuil approximativement indépendants. La démarche est la suivante :

- utiliser une règle empirique pour définir des groupes de dépassements. Un cluster peut être défini comme une période continue de dépassement d'un seuil (cette définition colle à la définition d'un événement d'inondations). Il est possible d'introduire une contrainte sur l'espacement temporel minimal des clusters afin de pouvoir considérer les événements indépendants ;
- identifier l'excédent maximum au sein de chaque groupe ;
- supposer que les maxima des groupes sont indépendants ;
- ajustement de la distribution de Pareto Généralisée aux maxima du groupe.

La méthode est simple, mais elle a ses limites. En particulier, les résultats peuvent être sensible aux choix arbitraires faits dans la détermination des clusters et il y-a sans doute une perte d'informations en ne gardant que le maximum de cluster.

III Application aux indices provinciaux

Dans un premier temps, nous allons devoir trouver un seuil pour modéliser notre loi GPD. Ce seuil doit être assez grand pour que l'approximation par la loi GPD soit fondée, mais pas trop grand pour qu'il reste suffisamment d'observations au-dessus du niveau u pour estimer les deux paramètres de la loi GPD. Pour déterminer ce seuil, nous allons nous aider d'un graphique, il s'agit de "Mean Residual Plot". Pour tracer ces graphiques et pour ajuster la loi GPD, nous avons utilisé le package (extension) `ismev` du logiciel R. Ci-après, le graphique du "Mean Residual Plot".

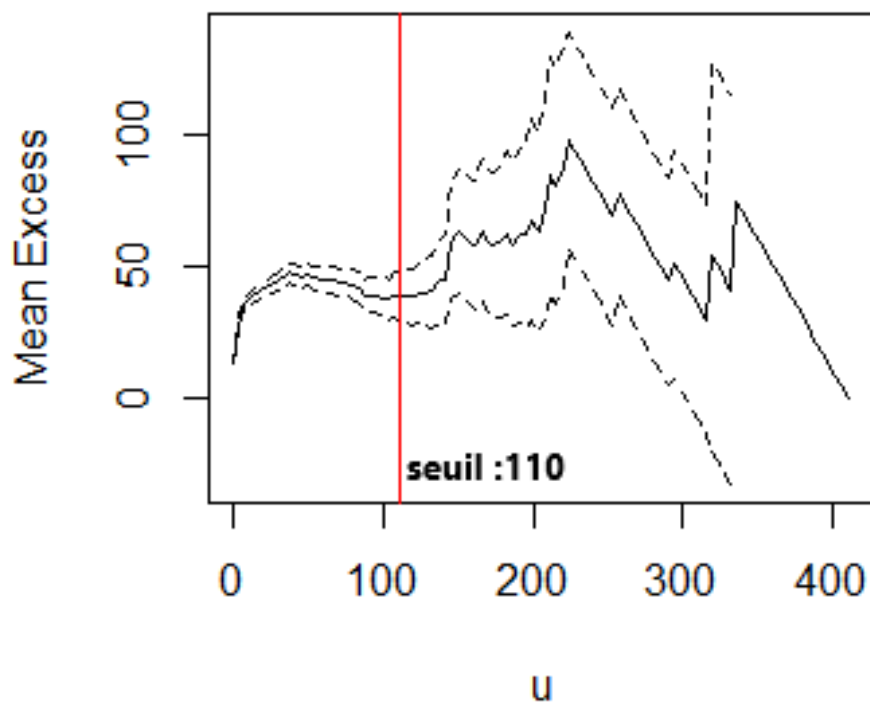


FIG. 4.3 : Espérance résiduelles des dépassements de seuils des pluies agrégées sur 3 jours (province Rabat)

A partir de ce graphique, nous devons trouver le seuil à partir duquel le "Mean Residual Plot" est approximativement linéaire. Il est difficile de choisir un unique seuil de modélisation. Pour le choix du seuil, nous avons décidé de choisir le seuil $u = 110$ mm pour notre modélisation. Pour vérifier que ce seuil n'est pas aberrant, nous affichons les

paramètres d'échelle en fonction du seuil (Hill plot 4.4). En utilisant cette méthodologie, le choix du seuil se fait en observant la stabilité de la valeur de l'indice de queue. En effet, le nombre d'excès correspondant au seuil est déterminé par l'observation sur le graphique de la zone de stabilité (zone où la valeur de l'indice de queue ne varie pas assez) de la valeur de l'indice de queue. Le nombre d'observations (seuil) à retenir correspond au début de la zone de stabilité.

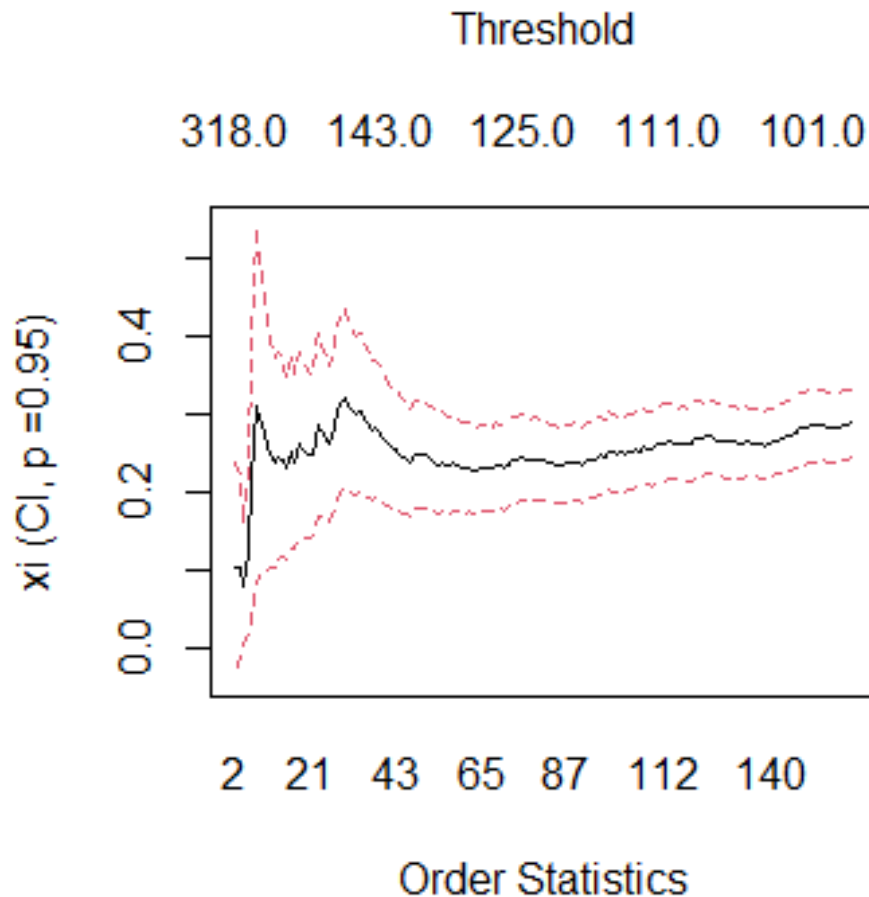


FIG. 4.4 : Hill plot (donnés Rabat)

Le résultat de l'estimateur de Hill est en cohérence avec celui de la mean excess function : nous obtenons une estimation du seuil autour de 110 mm. Nous conservons donc le résultat de l'estimation : nous définirons les précipitations extrêmes comme le cumul des précipitations sur 3 jours supérieur à 110 mm. Pour l'ajustement de la GPD, nous avons utilisé la méthode du maximum de vraisemblance. Ci-dessous, le tableau récapitulatif des paramètres estimés.

Seuil de modélisation en mm	Paramètre de forme ξ	aramètre d'échelle σ
110	0.265	30.108

TAB. 4.1 : Estimations des paramètres

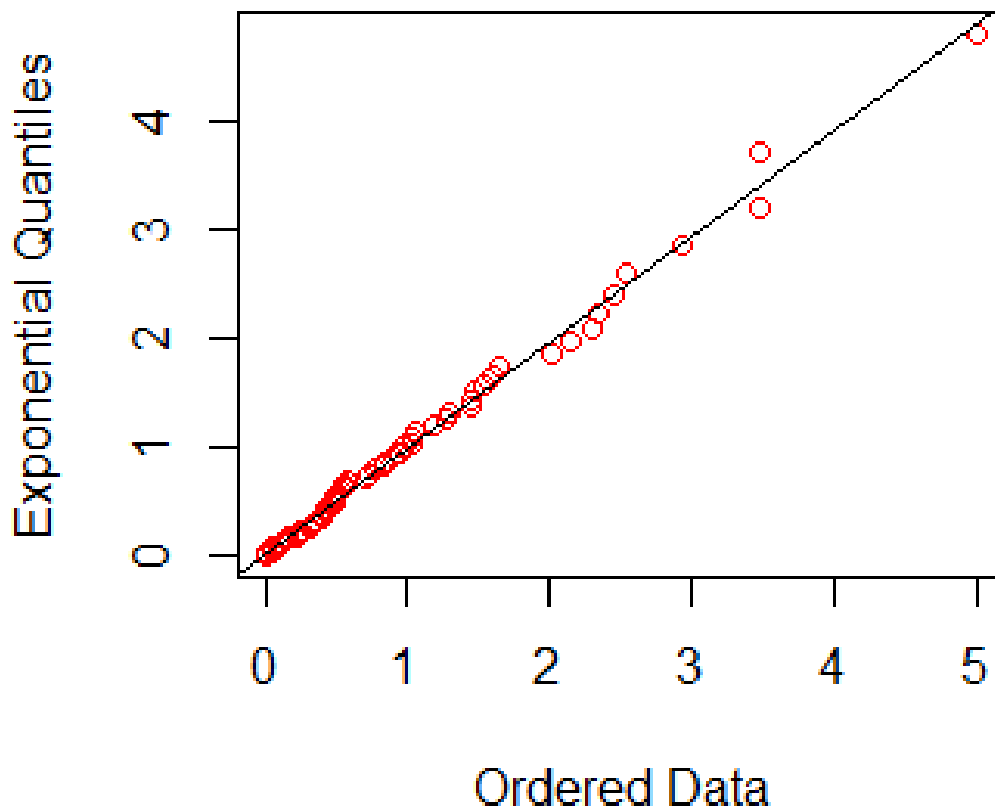


FIG. 4.5 : Graphe Quantile-Quantile '(Rabat)

Le graphe QQ-plot est fortement linéaire ce qui indique un bon ajustement du modèle GPD à nos données, maintenant que nous avons ajusté le modèle GPD nous pouvons l'utiliser pour satisfaire nos objectifs.

pour les autres provinces, le tableau V.1 en Annexe résume les seuils choisis ainsi que l'estimation des paramètres. L'annexe VI regroupe les graphiques (mean excess plots et les QQ-plots) des autres provinces.

IV Théorie des copules

Définition 4.1. Une copule multivariée $C : [0,1]^d \rightarrow [0,1]$ est une fonction de répartition dont les lois marginales sont uniforme sur $[0,1]$

Théorème 4.1. Une copule multivariée $C : [0,1]^d \rightarrow [0,1]$ est une fonction qui vérifie :

- Pour tout $u_1, \dots, u_d \in [0,1]^d : C(u_1, \dots, 0, \dots, u_d) = 0$

- Pour tout $u_1, \dots, u_d \in [0, 1]^d : C(1, \dots, u_i, \dots, 1) = u_i$

Le théorème est une traduction directe de la définition de la copule

$$C(u_1, u_2, \dots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d) \quad (4.3)$$

avec U_1, \dots, U_d sont des variables aléatoires uniformes.

Définition 4.2 (Densité d'une copule). Pour tout $u_1, \dots, u_d \in [0, 1]^d$, on définit la densité associée à la copule multivariée par :

$$C(u_1, \dots, u_d) = \frac{d^d C(u_1, \dots, u_d)}{du_1 \dots du_d}$$

Pour obtenir la distribution d'un vecteur aléatoire de dimension d on peut utiliser la relation suivante :

$$f(u_1, \dots, u_d) = C(u_1, \dots, u_d) \prod_{i=1}^d f_i(u_i)$$

où $f_i(u_i)$ sont les distributions marginales du vecteur.

Théorème 4.2 (Théorème de Sklar (1959)). Le théorème de Sklar est le théorème fondamental de la théorie des copules car il fournit un moyen d'analyser la structure de dépendance des distributions multivariées sans étudier les distributions marginales. Le théorème de Sklar pour les copules multivariées est défini comme ci-dessous :

Si C est une copule, et F_1, \dots, F_d d fonctions de répartition univariées, alors pour tout $(x_1, \dots, x_d) \in \mathbb{R}^d$, on a :

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (4.4)$$

est une fonction de répartition dans $F(F_1, \dots, F_d)$ (classe de Fréchet)

Réciproquement, pour tout $F \in F(F_1, \dots, F_d)$, il existe une copule C telle que :

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (4.5)$$

Si en plus les F_i sont continues, C est unique.

IV.1 les Familles de copules :

Il existe un grand nombre de familles de copules, nous allons introduire deux types de familles de copules, les copules elliptiques et les copules archimédiennes. Les copules elliptiques sont la famille la plus utilisée car elles sont simples à calibrer.

IV.1.1 Les copules elliptiques :

Les copules gaussiennes et les copules de Student proviennent toutes deux de la famille des copules elliptiques. Les copules elliptiques sont simplement les copules des

distributions elliptiques. L'avantage des copules elliptiques se résume au fait que leurs paramètres tels que la corrélation peuvent être facilement ajustés à partir des données.

La copule gaussienne multivariée a été décrite dans (CHERUBINI 2004). Il est défini comme suit.

Définition 4.3. *soit ϕ_{Σ} la distribution normale multivariée normalisée avec une matrice de corrélation R tel que R est une matrice n -dimensionnelle, symétrique, définie positive avec $\text{diag}(\Sigma) = (1, 1, \dots, 1)^T$, la copule gaussienne multivariée est définie comme suit :*

$$C_{\Sigma}(u_1, \dots, u_n) = \phi_{\Sigma}(\phi^{-1}(u_1), \dots, \phi^{-1}(u_n))$$

où ϕ est la fonction de répartition de la loi normale centrée réduite.

À partir du théorème de Sklar, nous pouvons noter que la copule gaussienne génère la fonction de distribution normale conjointe centrée réduite si les marginaux sont normaux centrés réduits. Les copules gaussiennes sont généralement utilisées pour modéliser les dépendances de corrélation linéaire.

La copule de Student est définie comme suit ;

Soit $t_{\Sigma, v}$ la distribution multivariée standardisée de Student avec Σ est la matrice de corrélation et v est le degré de liberté. i.e

$$t_{\Sigma, v}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \frac{\Gamma(\frac{v+n}{2})}{\Gamma(\frac{v}{2})(v\pi)^{\frac{n}{2}}} (I + \frac{1}{v} X^T \Sigma^{-1} X)^{-\frac{v+1}{2}} dx_1 \dots dx_n$$

Ensuite, la copule t de Student multivariée est définie comme suit :

$$T_{\Sigma, v}(u_1, \dots, u_n) = t_{\Sigma, v}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_n)) = \int_{-\infty}^{t_v^{-1}(u_1)} \dots \int_{-\infty}^{t_v^{-1}(u_n)} \frac{\Gamma(\frac{v+n}{2})}{\Gamma(\frac{v}{2})(v\pi)^{\frac{n}{2}}} (I + \frac{1}{v} X^T \Sigma^{-1} X)^{-\frac{v+1}{2}} dx_1 \dots dx_n$$

où t_v^{-1} est l'inverse de la fonction de distribution univariée de la distribution de Student avec v degrés de liberté. Lorsque le degré de liberté de la copule de Student tend à l'infini, la copule de Student converge vers la copule gaussienne. On peut dire que pour une grande de degré de liberté la copule de Student se rapproche de la copule gaussienne.

IV.1.2 Les copules archimédiennes :

La différence entre les copules elliptiques et archimédiennes réside dans le fait que les copules archimédiennes ne dérivent pas des fonctions de distribution multivariées en utilisant le théorème de Sklar. D'un point de vue pratique, les copules archimédiennes sont utiles car il est possible de générer un certain nombre de copules en interpolant certaines copules. Il existe une grande variété de copules archimédiennes comme la copule de Clayton, la copule de Gumbel et la copule de Frank.

Définition 4.4 (le générateur). *soit $\Phi : I = [0, 1] \rightarrow R^+$ une fonction continue, décroissante et convexe telle que $\Phi(1) = 0$, la fonction Φ est appelée un générateur.*

si $\Phi(0) = +\infty$, Φ est dite un générateur strict. on définit pour tout u dans $[0, \infty]$, le pseudo-inverse Φ^{-1} de Φ par

$$\Phi^{-1}(u) = \begin{cases} \Phi^{-1}(u) & \text{si } 0 \leq u \leq \Phi(0) \\ 0 & \text{sinon.} \end{cases}$$

Théorème 4.3 (Théorème de Kimberling). *soit Φ un générateur, et C la fonction définie de $[0, 1]^2 \rightarrow [0, 1]$ par*

$$C(u, v) = \Phi^{-1}(\Phi(u) + \Phi(v))$$

C est une copule si et seulement si le générateur Φ est convexe.

Les propriétés suivantes nous permettront de définir une copule archimédienne multivariée.

Propriété 4.1 (Propriété : symétrie et association). *Soit C une copule archimédienne de générateur Φ , donc pour tout u, v dans $[0, 1]$:*

- *C est symétrique i.e $C(u, v) = C(v, u)$*
- *C est associative i.e $C(C(u, v), w) = C(u, C(v, w))$*

Les copules archimédiennes ont un inconvénient par rapport aux copules gaussiennes. En réalité elles ont une structure de dépendance très limitée car tous les marginaux sont identiques du point de vue du Théorème de Kimberling. Les marginaux sont des variables aléatoires $U(0,1)$.

IV.1.3 Copule de Gumbel

Le générateur de la copule de Gumbel est donné par :

$$\Phi_{Gumbel}(u) = (-\ln(u))^\alpha$$

avec $\alpha > 1$. On a aussi

$$\Phi_{Gumbel}(u)^{-1} = \exp(-u^{\frac{1}{\alpha}})$$

Définition 4.5 (copule de Gumbel). *Soit $u = (u_1, \dots, u_n)$ un vecteur dans $[0, 1]^n$, la copule multivariée de Gumbel est donnée par : $C_\alpha^{Gumbel}(u) = \exp((-\sum_{i=1}^n (-\ln(u_i))^\alpha)^{\frac{1}{\alpha}})$*

Le générateur de la copule de Frank est donné par :

$$\Phi_{Frank}(u) = \ln\left(\frac{\exp(-\alpha u) - 1}{\exp(-\alpha) - 1}\right)$$

Définition 4.6 (copule de de Frank). Soit $u = (u_1, \dots, u_n)$ un vecteur dans $[0, 1]^n$, la copule multivariée de de Frank est donnée par :

$$C_\alpha^{Frank}(u) = \frac{-1}{\alpha} \ln\left(1 + \frac{\prod_{i=1}^n (\exp(-\alpha u_i) - 1)}{\exp(-\alpha) - 1}\right)$$

IV.2 Tau de Kendall et Rho de Spearman

Le Tau de Kendall et Rho de Spearman sont deux paramètres qui mesurent la dépendance. Ils peuvent être considéré comme des alternatives au coefficient de corrélation linéaire. Avant de traiter le Tau de Kendall et le Rho de Spearman, nous aimerions rappeler la définition de la concordance et de la corrélation linéaire.

La concordance se produit lorsque la probabilité d'avoir des « grandes » valeurs ou « petites » valeurs de deux variables aléatoires X et Y sont élevés, tandis que la probabilité d'avoir « grandes » (ou « petites ») valeurs de X avec des « petites » (ou « grandes ») valeurs de Y est faible. Pour deux variables aléatoires X et Y , la corrélation linéaire ρ_{XY} est définie comme suit :

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Définition 4.7 (Tau de Kendall). Soit X et Y deux variables aléatoires de copule C , et soit $F_X(u)^{-1}, F_Y(v)^{-1}$ les fonctions quantiles, avec u et v les quantiles définies sur $[0, 1]=I$. le tau de Kendall est défini comme suit :

$$\tau(X, Y) = 4 \int_I \int_I C(u, v) dC(u, v)$$

On peut montrer que le tau de Kendall mesure la différence entre la probabilité de concordance et celle de discordance pour deux variables aléatoires indépendantes (X_1, Y_1) (X_2, Y_2) . Alors on peut écrire

$$\tau(X, Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$$

Proposition 4.1 (Tau de Kendall). Soit (X_1, X_2) deux variables aléatoires de distribution elliptique de paramètres de localisation (μ_1, μ_2) et de coefficient de corrélation linéaire ρ .

$$si P(X_1 = \mu_1) = P(X_2 = \mu_2) = 0, alors \tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho)$$

Définition 4.8 (Rho de Spearman). Soit X et Y deux variables aléatoires de copule C , et soit $F_X(u)^{-1}, F_Y(v)^{-1}$ les fonctions quantiles, avec u et v les quantiles définies sur $[0, 1]=I$. Le Rho de Spearman est défini comme suit :

$$\rho_s(X, Y) = 12 \int_I \int_I C(u, v) du dv - 3$$

On peut montrer que le Rho de Spearman est un multiple de la différence entre la probabilité de concordance et la probabilité de discordance pour les vecteurs (X_1, Y_1)

(X_2, Y_2) . Si (X_1, Y_1) (X_2, Y_2) sont iid de copule C , alors

$$\rho_s = 3P((X_1 - X_2)(Y_1 - Y_2) > 0) - 3P((X_1 - X_2)(Y_1 - Y_2) < 0)$$

V La modélisation des dépendances avec les copules

On sait maintenant que le rôle des copules en premier degré est de décrire la dépendance des distributions marginales associées à une distribution conjointe. On a vu dans la partie précédente comment définir une copule, et comment déterminer sa fonction de répartition ainsi que sa densité. On a aussi abordé les copules paramétriques et notamment les copules elliptiques et archimédiennes. La plupart de ces copules se caractérisent par des paramètres et de même pour leurs distributions marginales. D'où l'importance de présenter des méthodes qui vont permettre de calibrer ces copules et d'estimer leurs paramètres à partir d'une base de données. Après on verra comment peut on générer des nombres aléatoires à partir des copules estimées.

V.1 L'estimation et la calibration à partir des données

Les copules offrent une grande flexibilité dans la modélisation des variables aléatoires car on peut choisir séparément les paramètres des marginaux et les paramètres de la distribution conjointe. Pour estimer et calibrer ces paramètres, il faut d'abord extraire la distribution des marginaux puis extraire la structure de dépendance de la distribution conjointe.

L'estimation de ces paramètres est basée sur la méthode du maximum de vraisemblance. Dans ce propos il y'a trois méthodes qu'on peut utiliser afin de calibrer une copule :

- Méthode du maximum de vraisemblance exacte : Cette méthode estime les paramètres des marginaux et les paramètres (structure de dépendance) de la distribution conjointe simultanément ;
- Méthode des fonctions d'inférence des marginales (IFM) : Avec cette méthode, on estime d'abord les paramètres des marginaux, puis on estime les paramètres de la distribution conjointe ;
- Méthode de pseudo-maximum de vraisemblance (maximum de vraisemblance canonique CML) : Cette méthode est similaire à la méthode IFM sauf le fait qu'avec la méthode CML, il n'est pas nécessaire de spécifier les marginaux. Il utilise les distributions empiriques et toute hypothèse sur la distribution marginale.

V.2 Méthode du maximum de vraisemblance exacte

Cette méthode n'est pas souvent utilisée dans la pratique car il s'agit d'une méthode très compliquée au niveau du calcul. En réalité, les paramètres des distributions marginales

et les paramètres de la structure de dépendance à partir des copules sont estimés ensemble ce qui pourrait être très intense en calcul dans le cas d'une grande dimension.

Supposons qu'on a les intensité précipitations sur n provinces et p années, représentés dans la matrice $(x_{i,1}, x_{i,2}, \dots, x_{i,n})$ avec i de $1, \dots, p$. On a F_i la fonction de répartition de la distribution marginale associée à chaque province, f_i sa densité, c la densité de la copule. On pose β_i les paramètres de la distribution marginale du i -ème province, α les paramètres de la copule (càd de la densité de la copule)

On pose (β, α) le vecteur représentatif des paramètres à estimer

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^p \log c(F_1(x_{i,1}, \beta_1), \dots, F_n(x_{i,n}, \beta_n), \alpha) + \sum_{i=1}^p \sum_{j=1}^n \log f_i(x_{i,j}, \beta_j)$$

Et on calcule donc : $\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} l(\theta)$

V.3 Méthode des fonctions d'inférence des marginales (IFM)

La méthode IFM peut être divisée en trois étapes :

1. L'estimation des paramètres des marginales $\hat{\theta}_i^{\text{margin}}$: Les paramètres $\hat{\theta}_i^{\text{margin}}$ des distributions marginales peuvent être estimés avec la méthode du maximum de vraisemblance. Afin de choisir la fonction de distribution marginale appropriée, on peut utiliser les QQ-plots des quantiles paramétriques en fonction des estimations empiriques.
2. La transformation des données avec les fonctions de distribution estimées : A partir des paramètres estimés dans l'étape précédente, on peut désormais transformer les données pour les utiliser dans l'estimation des paramètres de la copule. Ensuite, les données observées sont transformées sous forme de nombre aléatoires uniformes (uniform variates) ; $u_{i,j} = F_i(x_{i,j}, \hat{\theta}_i^{\text{margin}})$
3. L'estimation des paramètres de la copule $\hat{\theta}^{\text{copule}}$: On utilise maintenant les paramètres des distributions marginales estimés $\hat{\theta}_i^{\text{margin}}$ afin d'estimer les paramètres de la copule avec la méthode du maximum de vraisemblance.

Les paramètres estimés $\hat{\theta}^{\text{margin}}$ et θ^{copule} sont les paramètres à utiliser pour créer la copule ajustée qui décrit le mieux la distribution des données observées. Cette méthode d'estimation est plus efficace que la méthode du maximum de vraisemblance exacte car elle est moins intensive en calcul. C'est elle qu'on va utiliser dans la suite.

V.4 Méthode de pseudo-maximum de vraisemblance (maximum de vraisemblance canonique CML)

Cette méthode consiste à transformer les données de l'échantillon en données uniformes et estimer les paramètres de la copule à partir de ces données transformées. Cela signifie que les paramètres de la copule peuvent être estimés sans choisir des marginaux spécifiés. Cette méthode peut être divisée en deux étapes :

1. L'estimation des marginaux en utilisant une distribution empirique sans spécifier la forme de chaque marginale. On aura donc $\hat{F}_i(x_{i,j})$ et non pas $F_i(x_{i,j})$

2. L'estimation des paramètres de la copule en utilisant la méthode du maximum de vraisemblance.

Nous savons maintenant estimer et ajuster une copule à partir de données. Dans la partie suivante, nous allons appliquer ces méthodes aux données afin de simuler des nombres aléatoires à partir d'une copule.

VI Application aux données

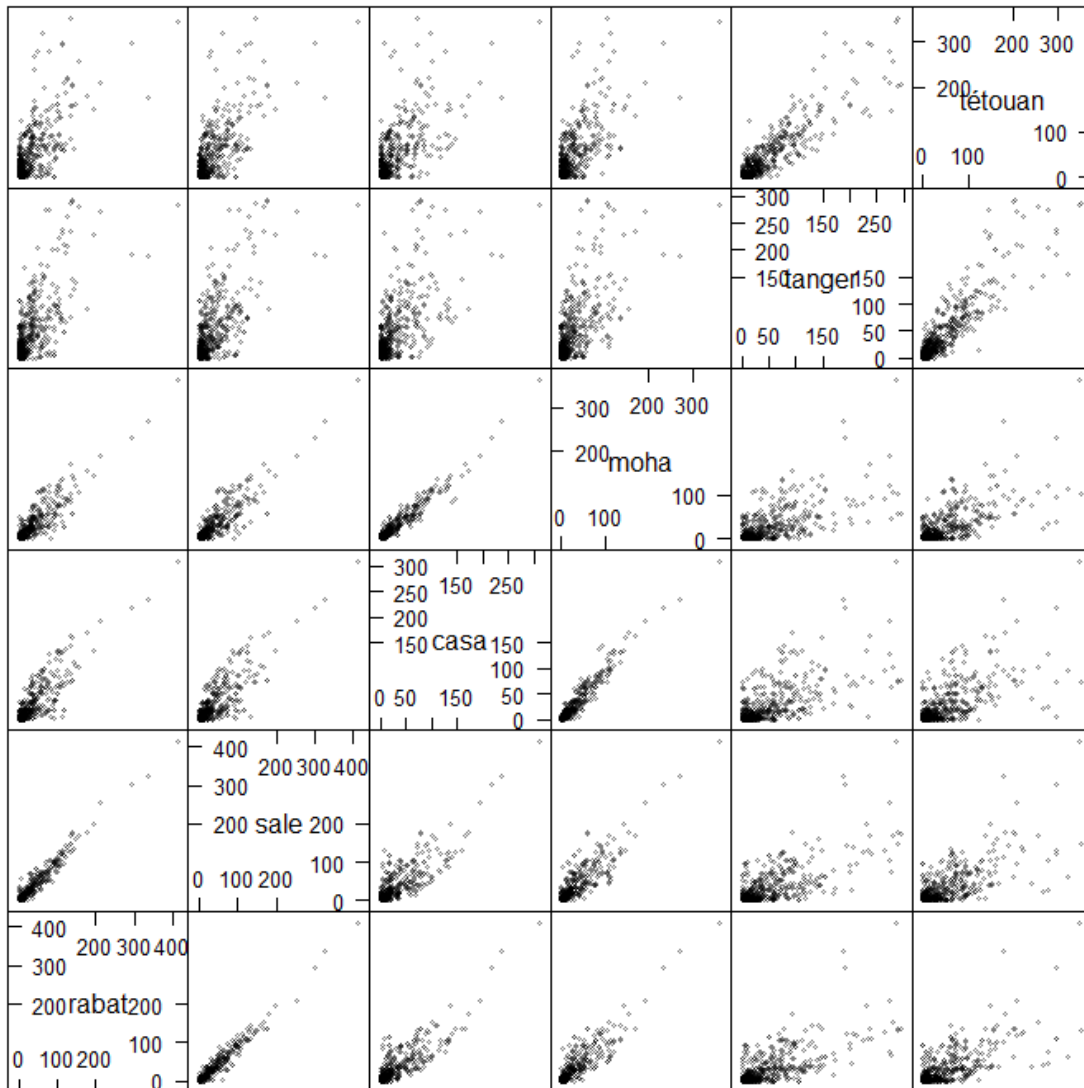


FIG. 4.6 : Nuage de points de cumules des précipitations

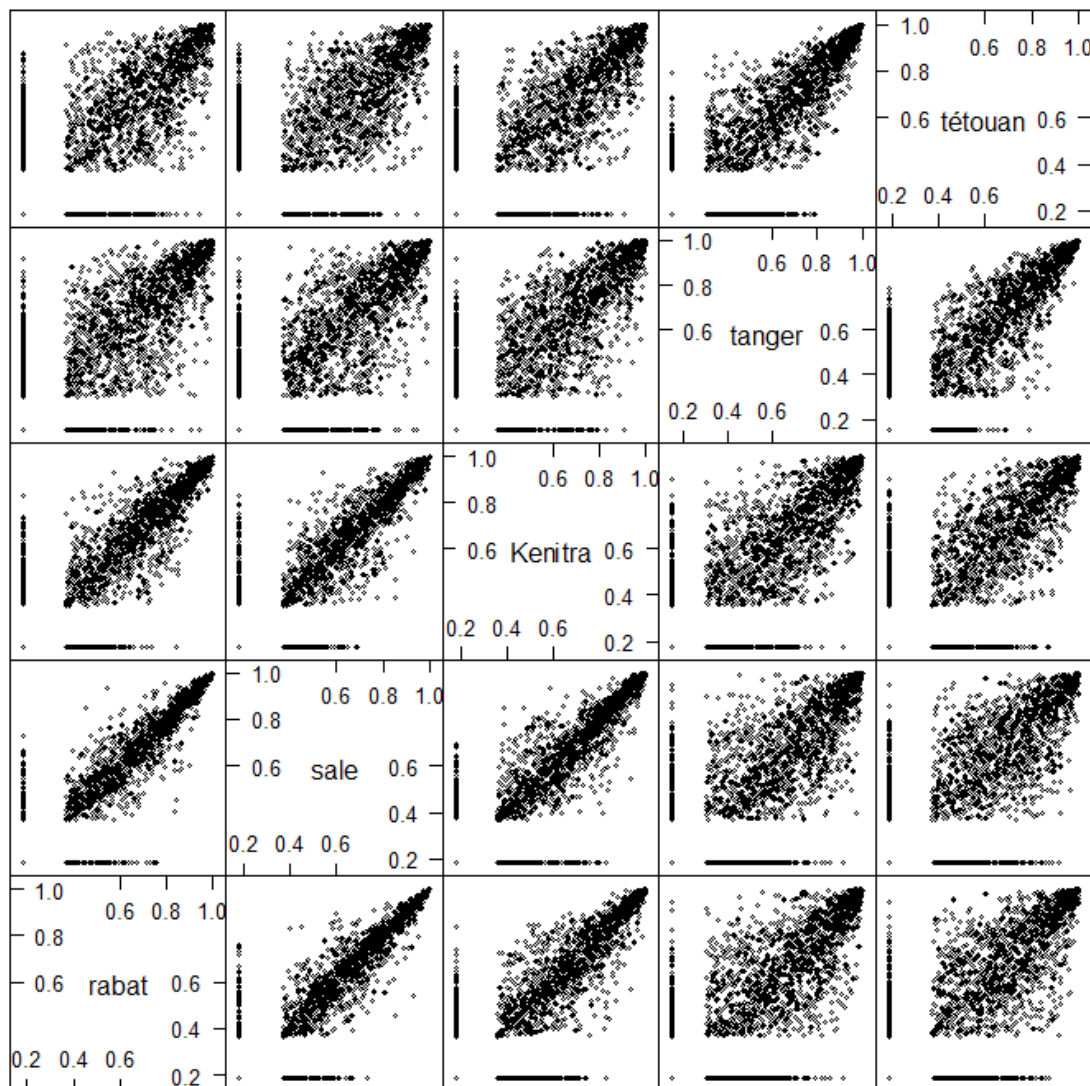


FIG. 4.7 : Nuage de points des données transformées

Ces deux graphiques exhibent une forte dépendance des précipitations entre les différentes provinces, notamment entre (Rabat, Sale, Kenitra), (Casablanca et Mohammadia), et (Tanger, Tétouan). Pour ajuster la copule, il faut tout d'abord choisir la copule qui permettrait de mieux modéliser la dépendance, puis estimer ces paramètres par une des méthodes présentées dans la partie précédente. Pour le choix de la copule, plusieurs candidates se présentent, notamment les copules elliptiques et les copules extrêmes. Les deux graphiques ci-dessous exposent les différences entre ces deux familles.

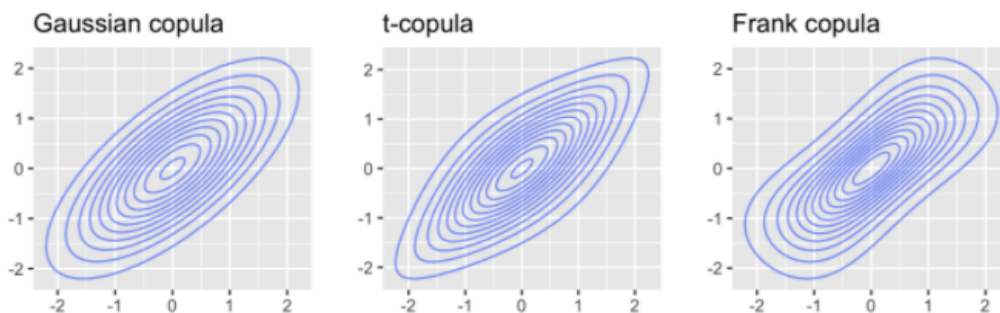


FIG. 4.8 : Copules elliptiques

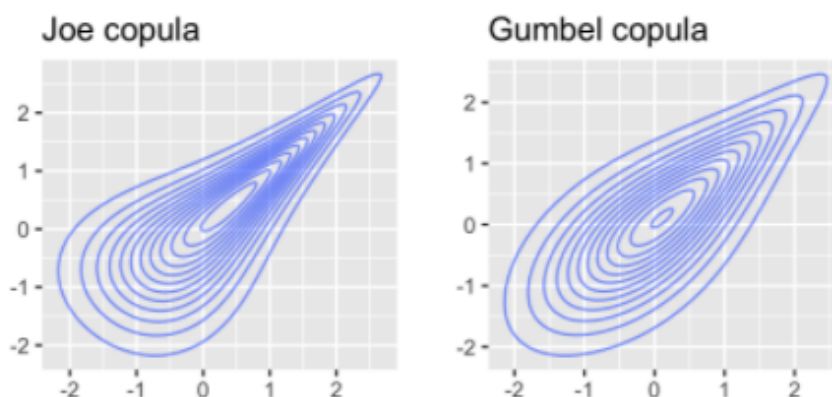


FIG. 4.9 : Copules extrêmes

Puisque on cherche à modéliser la dépendance des extrêmes, les copules de Gumbel et celle de Joe sont des candidats potentiels. Cependant, l'utilisation de la copule de Gumbel (ou la copule de Joe) multidimensionnelle n'est plus adaptée à notre problème lorsque la dimension est supérieure à 2. En effet, la copule de Gumbel (Copule de Joe) multidimensionnelle est toujours paramétrée par un seul paramètre α . Il est imprudent de vouloir modéliser la structure de dépendance mutuelle entre 7 provinces par un seul paramètre. Pour cette raison nous choisissons de privilégier l'utilisation de la copule gaussienne. Nous reviendrons sur le choix de la copule dans le dernier chapitre.

À partir de cumul des précipitations, les paramètres des marginales (GPD) et de la copule (Gaussienne) seront estimés via la Méthode IFM. Les paramètres estimés sont utilisés pour générer les variables aléatoires à partir de copule ajustée (voir la matrice de corrélation estimée en annexe VI). Ci-dessous les nuages de points des variables générées.

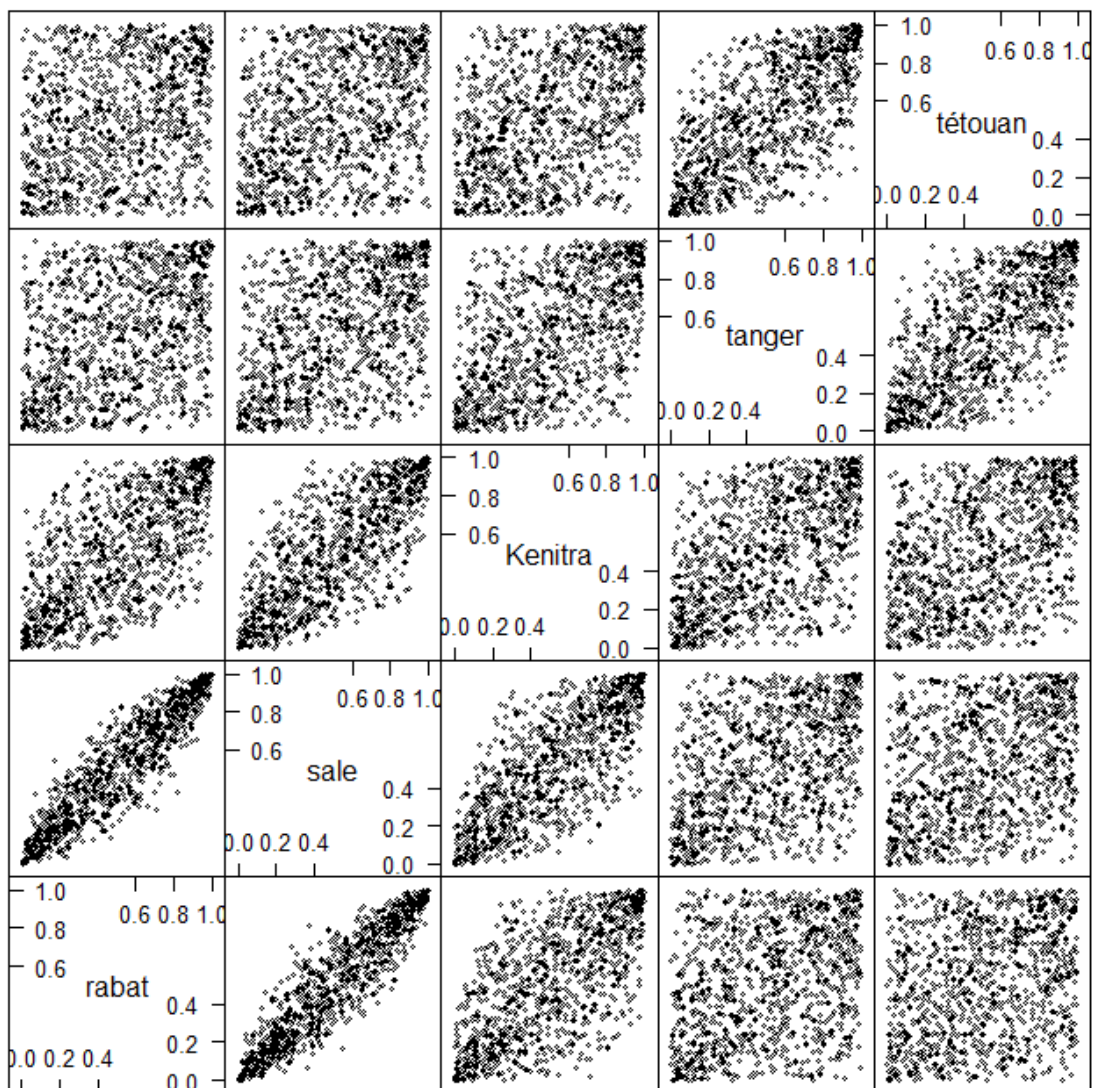


FIG. 4.10 : Nuage de points des données simulées

D'après le graphique, cette copule indique une forte dépendance au niveau de queue de distribution droite (upper tail). Ce qui est bien convenaient puisqu'on cherche à modéliser la dépendance des extrêmes.

VI.1 Tests d'adéquation

Les tests peuvent être utilisés pour affiner la recherche d'une famille de copules paramétrique adéquate. Il est alors généralement intéressant d'évaluer si l'inconnu (la copule C) appartient en fait à une classe \mathcal{C} , dans notre cas \mathcal{C} est la copule gaussienne. Cette question de qualité d'ajustement équivaut formellement tester

$$H_0 \ C \in \mathcal{C} \quad \text{vs} \quad H_1 \ C \notin \mathcal{C} \quad (4.6)$$

Un grand nombre de tests d'adéquation ont été proposés dans la littérature, parmi les procédures existantes, une approche intuitive basée sur la copule empirique C_n définie

ci-dessous. La copule empirique est un estimateur cohérent de la copule inconnue C que H_0 soit vrai ou non. Ainsi, un test naturel d'adéquation consiste à comparer C_n avec une estimation \hat{C}_n de C obtenue sous l'hypothèse H_0 (hypothèse que $C \in \mathcal{C}$ est vrai). Dans l'énoncé précédent, θ_n est un estimateur de paramètre de la copule C , la version la plus puissante de cette procédure est basée sur la statistique Cramér-von Mises

$$S_n^{gof} = \int_{[0,1]^d} n(C_n(u) - C_{\theta_n}(u))^2 = \sum_{i=1}^n (C_n(U_{i,n}) - C_n(U_{i,n}))^2. \quad (4.7)$$

avec

$$C_n(u) = \frac{1}{n} \sum_{i=1}^n 1(U_{i,n} < u) \text{ avec } u \in [0, 1]^d$$

Une valeur approximative de la p-value du test basé sur S^{gof} peut être obtenu au moyen à partir d'un bootstrap paramétrique. Le principal avantage de la procédure ci-dessus est sa simplicité conceptuelle. La capacité d'une copule C à modéliser la dépendance peut être évaluée tant que l'on sait comment estimer le vecteur des paramètres inconnu θ .

La fonction clé sous R pour effectuer le tests d'adéquation dans le package copula est appelé gofCopula(). Par défaut, elle est basée sur la statistique S^{gof} présentée ci-dessus. Nous pouvons alors évaluer si la famille des copules gaussiennes est un bon choix pour modéliser la dépendance de l'ensemble de nos données.

```

Multiplier bootstrap-based goodness-of-fit test of Normal copula, dim. d = 7, with 'method'="sn",
'estim.method'="mpl":

data: x
statistic = 0.033601, parameter1 = 0.92366, parameter2 = 0.71876, parameter3 = 0.67894, parameter4 =
0.71458, parameter5 = 0.37687, parameter6 = 0.32855, parameter7 = 0.79043, parameter8 = 0.67344,
parameter9 = 0.72440, parameter10 = 0.49503, parameter11 = 0.39787, parameter12 = 0.54368,
parameter13 = 0.62773, parameter14 = 0.56162, parameter15 = 0.43947, parameter16 = 0.83730,
parameter17 = 0.31225, parameter18 = 0.27723, parameter19 = 0.32412, parameter20 = 0.30342,
parameter21 = 0.71177, p-value = 0.6648
    
```

FIG. 4.11 : Goodness of fit test

L'estimation du paramètre inconnu de la copule gaussienne est effectuée par la méthode de pseudo-vraisemblance, qui est la méthode d'estimation par défaut. la statistique de test est $S^{gof} = 0.033601$ et p-value correspondante (calculée par le bootstrap) est 0.6648 donc on accepte l'hypothèse H_0 , la copule gaussienne est fortement retenue, on note que les paramètres de la copule sont également estimés.

Conclusion

Dans un premier temps nous avons exposé la théorie des valeurs extrêmes, nous avons modélisé les lois marginales des dépassements des pluies de 3 jours de façon indépendante. Cette modélisation ne prend pas en compte leur dépendance, pour remédier à ce problème nous avons utilisé la théorie des copules, nous avons choisi la copule gaussienne pour modéliser cette dépendance. Cette étape de modélisation est nécessaire à la tarification de notre produit que nous allons aborder dans la partie suivante.

Chapitre 5

Tarification

I Principe de calcul de la prime pure

Avant de passer au calcul de la prime pure, nous allons tout d'abord introduire les fondements théoriques de l'assurance non-vie, La tarification constitue l'un des cœurs de métier de l'actuaire. C'est une étape qui permet aux compagnies d'assurance d'évaluer les risques auxquelles elles doivent faire face. Elle se base sur l'estimation des flux futurs des prestations à verser afin de déterminer le montant des engagements probables au début d'un exercice. Dans le contexte du modèle collectif, on note la charge financière totale pour la période considérée par S tel que

$$S = \sum_{i=1}^N X_i \quad (5.1)$$

Avec :

- N est une variable aléatoire discrète représentant le nombre de sinistres.
- X_i une variable aléatoire continue indiquant le montant du $i^{\text{ème}}$ sinistre.

La variable aléatoire S suit une loi composée avec la convention $S=0$ lorsque $N=0$. La prime pure est donnée par l'espérance de la charge totale des sinistres auxquels l'assureur devra faire face, c'est un montant déterministe noté mathématiquement par $E[S]$. Elle s'agit d'une prime pure relatif aux coûts des sinistres payés par la compagnie et n'intègre pas les autres coûts d'administration ; frais de gestion et autres frais que la compagnie va déboursier afin d'assurer son activité. Pour simplifier le calcul de la prime pure, nous considérons que le nombre d'événements est indépendant des paiements. De même, nous supposons que les paiements par événement sont indépendants et identiquement distribués. Dans ce cas la prime pure s'écrit :

$$E(S) = E(N)E(X_1) \quad (5.2)$$

en effet

$$\begin{aligned} E[S] &= E[E[S|N]] \\ E[S] &= E[E[\sum_{i=1}^N X_i|N]] \end{aligned}$$

par indépendance :

$$E[S] = E[N]E[X_1]$$

Dans notre cas, N est le nombre d'événements d'inondations annuels, $E(N)$ peut être estimé à partir des données historiques, la loi de la variable aléatoire X n'étant pas explicitement connue, nous allons utiliser des simulations de Monte-Carlo pour calculer son espérance (loi forte des grands nombres).

Théorème - loi forte des grands nombres Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, intégrables. Alors

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E(X_1) \quad (5.3)$$

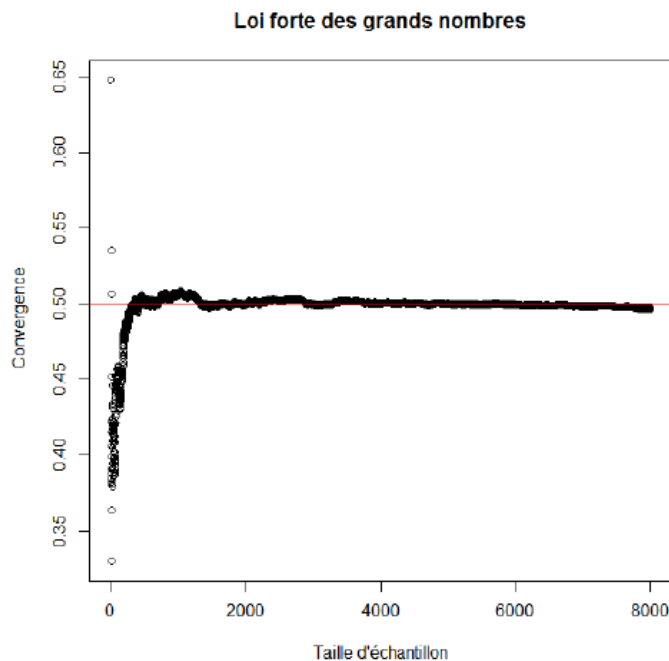


FIG. 5.1 : Convergence d'après la loi forte des grands nombres.

Le principe des techniques de simulation est d'approcher le résultat théorique en effectuant des tirages selon la loi du phénomène observé. Ainsi, il est donc nécessaire de savoir simuler des réalisations de lois de probabilité. Il existe un grand nombre de méthodes de simulation, nous nous limiterons à la méthode de simulation de **Monte Carlo par inversion de la fonction de répartition**. Il s'agit de l'une des méthodes les plus utilisées en simulation lorsque l'inversion de la fonction de répartition est possible. On rappelle que l'inverse généralisé de la fonction de répartition F d'une variable aléatoire réelle (v.a.r) X est donné par l'expression suivante :

$$F^{-1}(y) = \inf\{x | F(x) \geq y\}$$

Cette méthode repose sur le lemme suivant :

Lemme : Si U est v.a.r de loi uniforme sur $[0, 1]$ et F la fonction de répartition de la variable aléatoire X , alors $F^{-1}(U)$ suit la même loi que X .

Ainsi, pour simuler n réalisations i.i.d. d'une loi ayant pour fonction de répartition F , il suffit de simuler n réalisations indépendantes d'une v.a.r. de loi uniforme sur $[0, 1]$, puis d'appliquer l'inverse de la fonction de répartition à chacune de ces valeurs.

Pour une variable aléatoire de pareto généralisé, la fonction de distribution pour $\xi \neq 0$ est donnée par

$$G_{\xi,\sigma}(x) = (1 - (1 + \xi \frac{x}{\sigma})^{-1/\xi})$$

la fonction de distribution inverse est donnée par :

$$G_{\xi,\sigma}^{-1}(y) = \frac{\sigma}{\xi}((1-y)^{-\xi} - 1)$$

Or pour la loi des excès, $G_{\xi,\sigma}^{-1}(y) = X - u$ avec X correspond à la charge de sinistre grave et u représente le seuil choisi finalement pour une variable aléatoire v de loi uniforme sur $]0,1[$ on a :

$$X - u = \frac{\sigma}{\xi}((1-v)^{-\xi} - 1)$$

donc :

$$X = u + \frac{\sigma}{\xi}((1-v)^{-\xi} - 1)$$

cette dernière formule nous permettra de générer des variables aléatoires de loi GPD. On rappelle que la modélisation de la sinistralité grave se fait par le model fréquence sévérité :

$$S = \sum_{i=1}^N X_i$$

A partir de ce modèle de fréquence et de sévérité que nous avons ajusté on peut construire un modèle de sinistralité agrégée par le biais de la simulation Monte-Carlo.

II Application aux provinces

Dans ce paragraphe, nous considérons que la couverture se fait au niveau des provinces, et que chaque province est considérée indépendante des autres. La prime pure exigée pour une province est égale à l'espérance des indemnités reversées dans la province considérée. Nous avons à notre disposition une fonction de paiement qui relie l'indice local au montant reversé (voir figure 3.6).

La prime locale dépend essentiellement de la variable aléatoire $(X|X > 200)$ où X sont les précipitations agrégées sur la province considérée. C'est pour cette raison que nous allons utiliser la modélisation des dépassements de seuils mise en place dans le chapitre précédent. Dans un second temps, nous simulons les dépassements de seuil pour en déduire le montant de prime pour chaque province. Sur chaque province, la loi de $(X|X > 200)$ est estimée par modèle de dépassement de seuil. Nous pouvons dès lors simuler 100 000 fois la variable $(X|X > 200)$ qui suit une loi de Pareto de paramètre de forme ξ_0 et de paramètre d'échelle σ qui vérifie $\sigma = \sigma_0 + \xi_0 \times (200 - u_0)$. Ceci permet d'obtenir une estimation de $E[\text{paiement}(X|X > 200)]$ par province. $E(N)$ est estimée par la moyenne annuelle des événements historiques locaux de la province considérée. La prime pure d'une province est calculée par la formule (5.2).

Dans la partie qui a traité fonction de paiement (voir section Fonction de paiement), nous avons discuté de choix de la borne Exhaustion, cette borne sera prise comme un quantile de la distribution empirique des indices locaux simulés. Cela permet en particulier de définir la borne de la fonction de paiement selon les événements qu'on aurait voulu faire indemniser, donc chaque province aura une borne Exhaustion en fonction de son profil des précipitations. Les distributions des indices simulés (100 000 simulations) pour la province de Rabat en figure 5.2

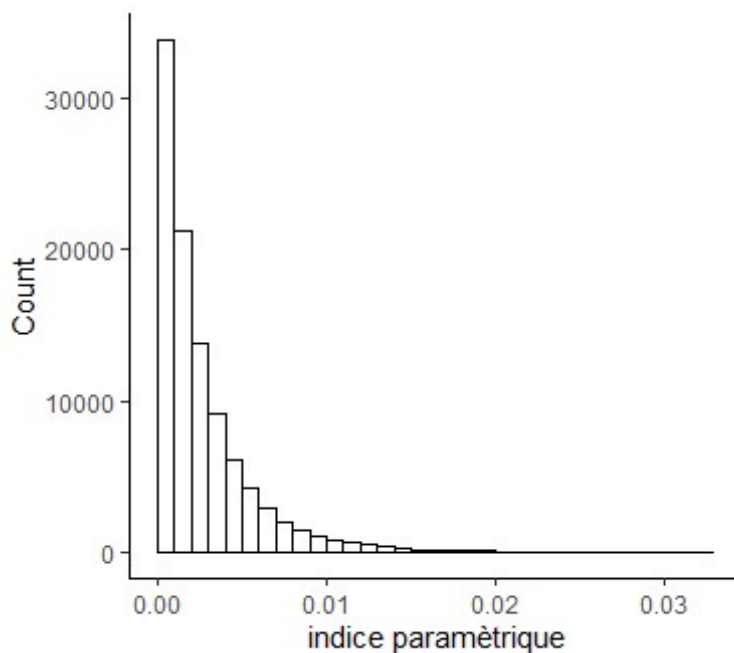


FIG. 5.2 : Histogramme des indices paramétriques de Rabat.

L'indice maximal simulé est de 3.3%, tandis que l'indice minimale simulé est de 2.04×10^{-9} , nous choisissons la borne Exhaustion comme le quantile 99% de la distribution de l'indice simulé, le coverage limit correspond au montant qu'encaissera la FSEC si l'indice dépasse la borne Exhaustion, idéalement il devrait être égale exactement à la perte de FSEC si l'indice dépasse la borne Exhaustion. Cette quantité est calculée à partir de la formule 2.3.

Notre modèle de tarification se base sur une approche par simulation comme souligné auparavant, nous calculons la prime pure grâce à la méthode Monte Carlo. La simulation Monte-Carlo nous permet d'obtenir des indemnités à partir desquelles nous déduisons simplement la prime pure. Le tableau ci-dessous résume les résultats de modélisation

Province	A= Coverage limit (MMAD)	B=Prime pure (MMAD)	B/A
RABAT	674,24	31,79	4,71%
SALE	685,21	27,92	4,07%
KENITRA	851,81	39,94	4,68%
CASABLANCA	1293,6	17,67	1,36%
MOHAMMEDIA	259,89	5,35	2,05%
TANGER ASSILAH	521,36	61,3	11,75%
TETOUAN	39,98	4,26	10,65%
Total	4326,09	188,23	4,35%

TAB. 5.1 : Les primes avec TD =5%

Les provinces avec les primes les plus élevées correspondent aux régions de nord, où les fortes précipitations sont les plus fréquentes. Ce qui est justifié par les données historiques, ci dessous les statistiques descriptives du cumul des précipitations sur 3 jours dans les provinces.

Province	1er quantile	Médiane	Moyenne	3 eme quantile	Max
RABAT	0.00	0.10	8.79	3.31	410.72
SALE	0.00	0.11	8.90	3.40	413.70
KENITRA	0.00	0.14	9.60	3.71	444.31
CASABLANCA	0.00	0.09	7.43	2.55	310.05
MOHAMMEDIA	0.00	0.08	7.73	2.57	363.22
TANGER ASSILAH	0.00	0.47	13.03	6.56	498.14
TETOUAN	0.00	0.10	12.60	5.28	477.49

TAB. 5.2 : Statistiques descriptives du cumul des précipitations sur trois jour (en mm)

Ces montants semblent être très élevés, l'augmentation de l'attachement baissera la prime, on prend un attachement qui correspond à un taux de dommage de 10% (qu'il faut multiplier par un taux d'exposition), le tableau ci-dessous résume les nouveaux résultats :

Province	A= Coverage limit (MMAD)	B= Prime (MMAD)	B/A
RABAT	674,24	22,92	3,39%
SALE	685,21	18,01	2,62%
KENITRA	851,81	27,63	3,24%
CASABLANCA	1293,6	7,701	0,59%
MOHAMMEDIA	259,89	3,72	1,43%
TANGER ASSILAH	521,36	37,27	7,14%
TETOUAN	39,98	2,63	6,57%
Total	4326,09	119,88	2,77%

TAB. 5.3 : Les primes avec TD= 10%

III Prime nationale

Pour calculer la prime pure nationale, nous avons besoin de connaître la distribution de l'indice national. Pour cela, nous allons simuler la distribution multivariée des indices locaux (I_1, \dots, I_7) en tenant compte de la dépendance qui serait modélisée par une copule gaussienne pour en déduire l'indice national défini :

$$I_{national} = \sum_{i=1}^7 I_i$$

Ci-dessous l'histogramme de l'indice national simulé (100 000 simulations)

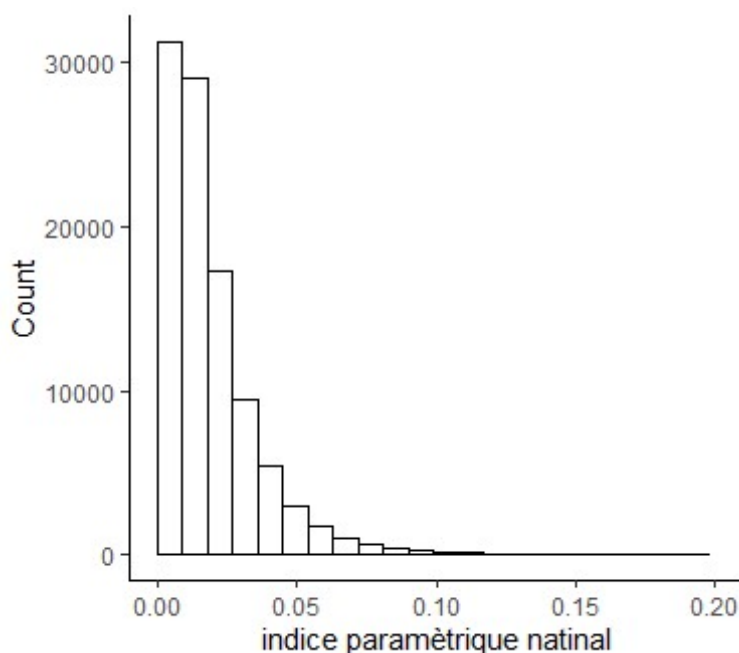


FIG. 5.3 : Histogramme de l'indice paramétrique national

Pour tarifier le produit, nous avons besoin de fixer la fonction de paiement. Pour se faire, nous fixons les points Attachement et Exhaustion. Attachement et Exhaustion sont également des indices nationaux, par définition un indice national est la somme des indices provinciaux, donc Attachement et Exhaustion seront fixé comme suit

- $Attachement_{national} = \sum_{i=1}^7 Attachement_i$
- $Exhaustion_{national} = \sum_{i=1}^7 Exhaustion_i$
- De même $Coverage\ limit_{national} = \sum_{i=1}^7 Coverage\ limit_i$

Cet $Exhaustion_{national}$ vaut à 0.11 ce qui correspond au quantile 99,8% de la distribution simulée. Pour rappel, la prime pure national est donnée par :

$$prime\ pure\ national = E(N) \times E(paiements(I_{national})) \tag{5.4}$$

$E(N)$ est estimée à partir l'historique, la prime pure annuelle correspondant est égale à 276,38 millions de DH.

Afin de baisser la prime pure annuelle, le FSEC peut décider d'augmenter la borne Attachment pour diminuer le nombre d'événements assuré. Nous gardons toujours les mêmes hypothèses (même borne Exhaustion, même Coverage limit); et faisons seulement varier les taux de dommages correspondants à la borne Attachment.

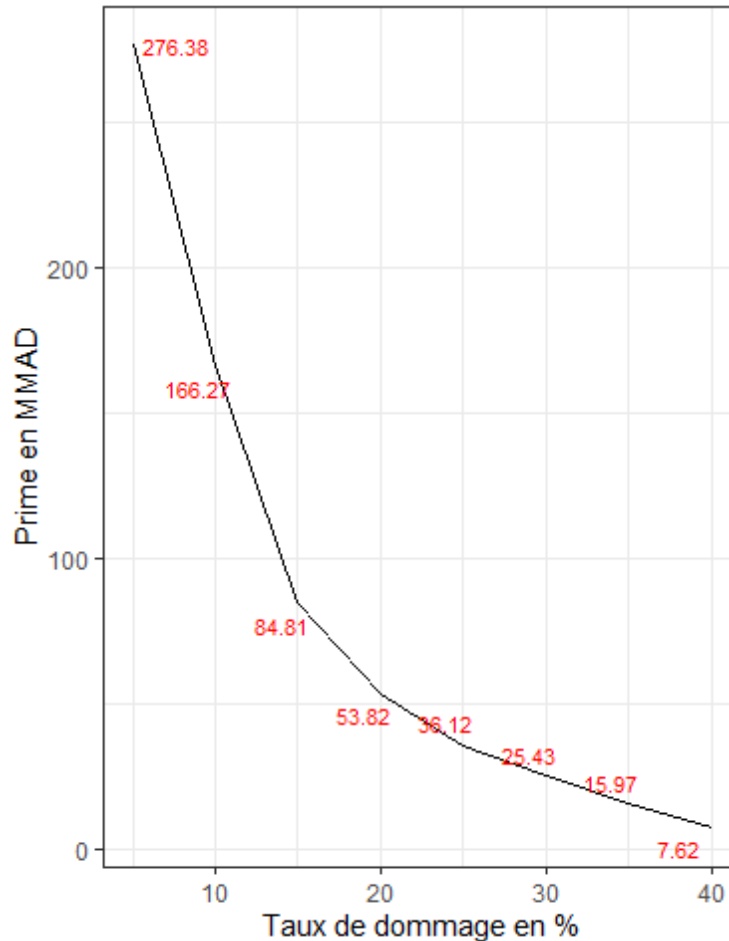


FIG. 5.4 : Sensibilité de la prime pure nationale au choix de l'attachement

Le graphique ci-dessus permet de voir l'impact de l'augmentation de la borne l'Attachment sur la prime pure nationale. Le FSEC a la possibilité de payer une prime plus au moins faible mais en contre partie, il y'aura moins d'événements assuré. La baisse de la prime est significative pour des valeurs de l'attachement correspondantes à des taux de dommages supérieures à 20%.

IV Comparaison des primes pures locales et de la prime nationale

Dans la partie précédent, nous avons calculé les primes locales et la prime nationale selon deux approches différentes :

- Pour la première approche, nous proposons 7 produits d'assurance paramétrique, le montant de couverture différent d'une province à une autre selon son exposition au risque, ce qui représente un avantage car il n'est pas raisonnable de répartir le montant de couverture national uniformément sur les provinces car les expositions au risque sont différentes. En effet, il est plus stratégique d'allouer un montant de couverture plus important dans les zones à haut risque. Les indices provinciaux associés à ces produits sont modélisés de façon indépendante selon la théorie des valeurs extrêmes.
- Pour la deuxième approche, nous proposons un produit d'assurance paramétrique qui couvre toutes les provinces, L'indice national associé dépend des indices extrêmes. La dépendance entre les provinces est modélisée selon la théorie des copules. La loi de l'indice national s'obtient par la sommation des indices extrêmes dont la dépendance est modélisée par la copule gaussienne.

La deuxième approche présente un avantage important. En effet cette approche permet de prendre en compte la dépendance des événements extrêmes extrêmes. En raison de la proximité géographique des provinces, une pluie extrême arrive en général simultanément dans plusieurs provinces . Le risque est donc plus important lors d'un événement national et se traduira par une prime pure plus importante que celle issue de l'agrégation d'événements extrêmes indépendants.

Cet écart important entre la modélisation univariée et multivariée peut être expliquer par plusieurs raisons :

- La raison la plus intuitive est que la modélisation des valeurs extrêmes univariées ne prend pas en compte la dépendance des pluies sur provinces, ce qui sous-estime le montant de prime de la couverture sur l'ensemble du royaume.
- Les fonctions de paiement ne sont pas les mêmes selon les deux approches. Exhaustion est défini comme un quantile (Value at Risk) de la distribution de l'indice. Le point Exhaustion dans la première approche correspond à un quantile de 0.99 de la distribution simulé, or pour la deuxième approche, le point Exhaustion correspond à un quantile de 0.998 de la distribution, c'est du au fait que la VaR est une mesure de risque non sous-additive, ce qui implique $VaR(\sum X_i) > \sum VaR(X_i)$

V Sensibilité au paramètre la fonction de vulnérabilité

Dans la section II, nous avons discuté le rôle de la fonction de vulnérabilité comme instrument fondamental dans la construction de notre indice paramétrique. Nous avons le choix de se baser sur la courbe construite à partir du produit MnHPRA (voir figure 2.2), cette fonction de dommages dus aux inondations est basée sur le niveau de l'eau à partir du niveau fini du rez-de-chaussée, pourtant la transformation de l'intensité des précipitations en hauteur d'eau est régie par des modèles physiques complexes qui sortent de cadre de notre formation. Nous avons choisi une fonction de vulnérabilité linéaire avec

un seuil maximal 1 000 mm. dans cette partie nous étudions la sensibilité de la prime nationale à la modification de ce seuil.

Après avoir modifié le seuil maximale de la fonction de vulnérabilité, nous recalculons la prime nationale avec les mêmes hypothèses.

seuil en mm	prime pure nationale en Millions MAD
800	291.02
900	286.60
1000	276,38

TAB. 5.4 : Sensibilité des primes au paramètre de la fonction de vulnérabilité

la prime reste peu sensible au seuil maximale, pour un seuil de 800 mm, la prime nationale est de 291.02 millions de DH, soit une augmentation de 5.3% par rapport au calcul original. Pour limiter l'impact de la sous-estimation de la prime, on pourra faire recours à la courbe de dommages construite à partir de MnHPRA, mais par contrainte de temps on a pas pu traiter cette partie.

VI Sensibilité au choix de la copule

Les copules se révèlent être un outil très puissant, plus particulièrement dans la modélisation de la dépendance entre les risques existants. Elles permettent dans un cadre multidimensionnel de modéliser la structure de dépendance du modèle. Toutes les informations sur la dépendance sont contenues dans la copule. Ainsi le choix de la copule qui va s'adapter aux données c'est très important. Habituellement, on prend une famille paramétrique (notamment la copule de student) de copules parmi beaucoup d'autres existantes et on l'adapte aux données en estimant les paramètres de la famille. Néanmoins, il n'existe pas de système rigoureux et de méthode systématique pour le choix des copules : il se peut que plusieurs copules permettent de modéliser la dépendance de la structure sous-jacente aux données. Cela peut fournir des résultats différents qui dépendent de la copule sélectionnée. Dans cette partie nous allons étudier la sensibilité de la prime pure nationale au choix de la copule.

Pour rappel, la copule gaussienne a été choisie pour modéliser la dépendance entre les provinces. La copule gaussienne est souvent utilisée pour modéliser la corrélation des risques en gestion des risques et en finance car elle est simple à paramétrer. On reproche à la copule gaussienne son incapacité à modéliser la dépendance des risques extrêmes. Par conséquent, nous testons d'autres copules pour quantifier la sensibilité de la prime pure au choix de copule. La copule de Gumbel est directement liée aux extensions multivariées de la théorie des valeurs extrêmes, elle a gagné en popularité dans la gestion des risques au cours de la dernière décennie mais cette copule reste inadaptée à notre problématique à cause au nombre limité de paramètres. La copule t de Student peut être considérée comme un compromis, car elle peut capturer la dépendance dans les queues sans abandonner la flexibilité pour modéliser la dépendance dans le centre.

```

Multiplier bootstrap-based goodness-of-fit test of t-copula, dim. d = 7, with 'method'='Sn',
'estim.method'='mpl':

data: x
statistic = 0.033586, parameter1 = 0.92417, parameter2 = 0.73708, parameter3 = 0.67589, parameter4 =
0.72437, parameter5 = 0.36668, parameter6 = 0.30257, parameter7 = 0.80325, parameter8 = 0.67261,
parameter9 = 0.73738, parameter10 = 0.47267, parameter11 = 0.36509, parameter12 = 0.54113,
parameter13 = 0.63508, parameter14 = 0.55250, parameter15 = 0.43139, parameter16 = 0.84082,
parameter17 = 0.29868, parameter18 = 0.25789, parameter19 = 0.31400, parameter20 = 0.27059,
parameter21 = 0.70367, p-value = 0.6998

```

FIG. 5.5 : Goodness of fit , copule t de Student .

Dans notre application, avec une p-value de 0.69 la copule t de Student est retenue. La prime pure annuelle correspondant est égale à 277,17 millions de DH, soit une augmentation de 0.28 %.

Pour pouvoir utiliser les copules archimédiennes, une autre approche peut être adoptée. On considère trois classes de provinces en se basant sur le critère de la proximité géographique et sur le coefficient de corrélation linéaire. Ainsi on construit les classes suivantes ;

- Classe A : {Rabat, Salé, Kénitra}
- Classe B : {Casablanca, Mohammadia}
- Classe C : {Tanger, Tétouan}

On considère que les indices des provinces dans une même classe sont dépendants, on suppose également l'indépendance entre les trois classes.

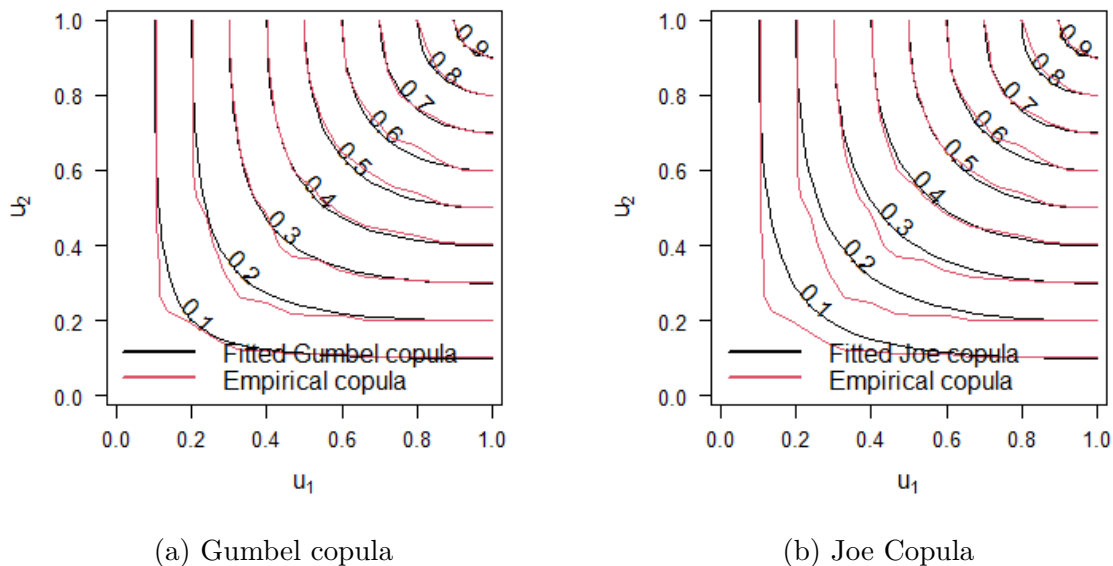


FIG. 5.6 : Contour plots de Gumbel copula et Joe copula (données Casablanca- Mohammadia)

Le graphique ci-dessous suggère de considérer la copule de gumbel comme outil pour modéliser la dépendance entre Casablanca et Mohammadia, sachant que ces deux copules présentent la dépendance de la queue supérieure. En effet le graphique nous permet de conclure que la copule de gumbel s'ajuste mieux aux données, chose prévue car le test

statistique accepte la copule de gumbel (p-value =0.1843) et rejette celle de Joe (p-value =0.0044) (voir figures ci-dessous).

```
Multiplier bootstrap-based goodness-of-fit test of Joe copula, dim. d = 2, with 'method'='sn',
'estim.method'='mpl':
data: x
statistic = 0.059352, parameter = 3.9342, p-value = 0.004496
```

FIG. 5.7 : Goodness of fit , copule de Joe

```
Multiplier bootstrap-based goodness-of-fit test of Gumbel copula, dim. d = 2, with 'method'='sn',
'estim.method'='mpl':
data: x
statistic = 0.019325, parameter = 2.8989, p-value = 0.1843
```

FIG. 5.8 : Goodness of fit , copule de Gumbel

On modélise la dépendance au sien de chaque classe par la copule de Gumbel (après vérification par les test statistiques). La prime pure nationale est la somme des primes des trois classes.

Classe	Prime en MAD
Classe A	153.85
Classe B	34.30
Classe C	102.688
Total	290,84

TAB. 5.5 : Primes pures, modélisation des dépendances au sien des régions

Conclusion

Les primes pures estimées pour les risques sont calculées selon des hypothèses simplistes en réponse au manque de données suffisamment fines. On considère les limitations et incertitudes sous-jacentes les calculs en interprétant les résultats. Les pertes autres que les pertes liées au résidentiel ne sont pas modélisées dans le mémoire ce qui pourrait mener à sous-estimer les pertes totales pour la population. En résumé, il existe une incertitude importante sous-jacente aux résultats en raison du manque de données et du manque d'hypothèses statistiquement prouvées pour évaluer l'estimation des coûts pour la population (fonction de vulnérabilité supposée linéaire ainsi que ces bornes ...), donc il faut considérer ces limitations et incertitudes sous-jacentes, hypothèses et méthodologie lors de l'utilisation de ces résultats à des fins de prises de décisions.

Conclusion et perspectives

Conclusion générale

La survenance de catastrophes naturelles toujours plus coûteuse et mondialement répartie crée une demande générale de couverture de plus en plus forte. Les polices d'assurance paramétrique sur « gros risques » sont devenues monnaie courante.

Avec pour objectif de gérer le risque d'inondation, nous avons proposé dans ce projet un modèle actuariel d'assurance indicielle visant à couvrir les pertes FSEC contre le risque d'inondations dans les provinces les plus exposées au risque dans le royaume. Un produit d'assurance dont les grands avantages se déclinent autour des deux points suivants : Suppression de l'aléa moral et de l'antisélection, rapidité des paiements et la réduction des coûts d'expertise. L'outil développé dans le cadre de ce mémoire permet maintenant de tarifer les catastrophes naturelles (inondations) en passant par une modélisation des précipitations par province et modéliser la dépendance via les copules. Cette idée tient en compte de la dépendance des provinces entre elles, ce qui est très envisageable puisque les provinces prises sont grandes et géographiquement proches. Les modèles proposés peuvent également être réutilisés dans le cadre d'autres phénomènes climatiques car les polices d'assurance paramétrique sur les risques catastrophiques sont des produits à fort rendement ce qui pourrait séduire de nombreux investisseurs par le biais de manque de risque d'asymétrie d'informations.

En conclusion, l'étude présentée dans ce mémoire constitue un premier support vers la couverture du risque d'inondations au Maroc par les polices d'assurance paramétrique. Les résultats obtenus sur les données des précipitations sont concluants et montrent qu'il est possible de mettre en place une modélisation du risque des inondations avec une source de données limitée sans passer par des modèles physiques externes tel que AIR, EQEcat ou RMS.

Bibliographie

- CHERUBINI, U. (2004). *Copula Methods in Finance*.
- DANIEL J. CLARKE, Ruth Vargas Hill (2013). “Cost-benefit analysis of the african risk capacity facility”. en. In :
- EMBRECHTS P. Kluppelberg C., Mikosch T. (2001). “Modelling Extremal Events for Insurance and Finance.” en. In : 3.
- FLORENT RITLENG, Cuong Nguyen (2014). “Etude d’un produit d’assurance paramétrique contre le risque de pluie torrentielle en Jamaïque”. fr. In : *Mémoire, ENSAE Paristech, 2013*.
- GUMBEL, E.J. (1958). “Statistics of Extremes”. en. In : *Columbia University Press*.
- JENKINSON, A. F. (1955). “The frequency distribution of the annual maximum (or minimum) values of meteorological elements”. en. In : *Quarterly Journal of the Royal Meteorological Society* 81, p. 158-171.
- SMITH, R. L. (1985). “Maximum likelihood estimation in a class of non-regular cases”. en. In : *Biometrika* 72, p. 67-90.
- VON MISES, R. (1954). “La distribution de la plus grande de n valeurs”. In : *American Mathematical Society, Providence, R.I.* 2, p. 237-256.

Annexes

Annexe I

Les principaux aspects de la loi 110-14

<... est considéré comme évènement catastrophique tout fait générateur des dommages directs survenus au Maroc, ayant pour origine déterminante l'action de l'intensité anormale d'un agent naturel ou l'action violente de l'Homme >

selon le Décret n° 2-18-785 du 23 chaabane 1440 (29 avril 2019) pris pour l'application de la loi n° 110-14 , Les agents naturels pouvant constituer un évènement catastrophique tel que défini par la loi n° 110-14 précitée sont :

- les crues ;
- les inondations y compris le ruissellement, le débordement des cours d'eau, la remontée de la nappe phréatique, la rupture de barrages causée par un phénomène naturel, les coulées de boue ;
- les tremblements de terre ;
- les tsunamis.

les dommages provoqués directement par les actions et mesures de secours, de sauvetage et de sécurisation sont assimilés à ceux résultant de l'évènement catastrophique lorsque les dites actions et mesures sont liées à cet évènement.

sont exclus du champ d'application de régime les dommages ou préjudices occasionnés par :

- L'utilisation d'agents ou d'armes chimiques, biologiques ,bactériologiques, radioactifs ou nucléaires ;
- La guerre civile, la guerre étrangère ou les actes d'hostilité assimilables, que la guerre soit ou non déclarée ;
- un acte de cybercriminilité.

La déclaration de la survenance de l'évènement catastrophique est établie par un acte administratif publié au bulletin officiel. ladite acte précise notamment, les zones

sinistrées, la datation et la durée de l'évènement catastrophique. la publication de l'acte administratif a pour effet :

- l'opération d'inscription des victimes sur le registre de recensement
- le processus d'octroi des indemnités par le fonds de solidarité contre les évènements catastrophique

le Fonds de solidarité a pour objet de :

- indemniser les victimes d'évènements catastrophiques ;
- Formuler des propositions et les communiquer à l'administration en vue d'améliorer le régime ;
- Etablir des données statistiques et financières relatives aux conséquences des évènements catastrophiques et les communiquer à l'administration à la demande de celle-ci.
- Réaliser ou faire réaliser toute étude qu'il juge nécessaire à l'exercice de ses missions.

Le budget du Fonds de solidarité comprend :

En recettes :

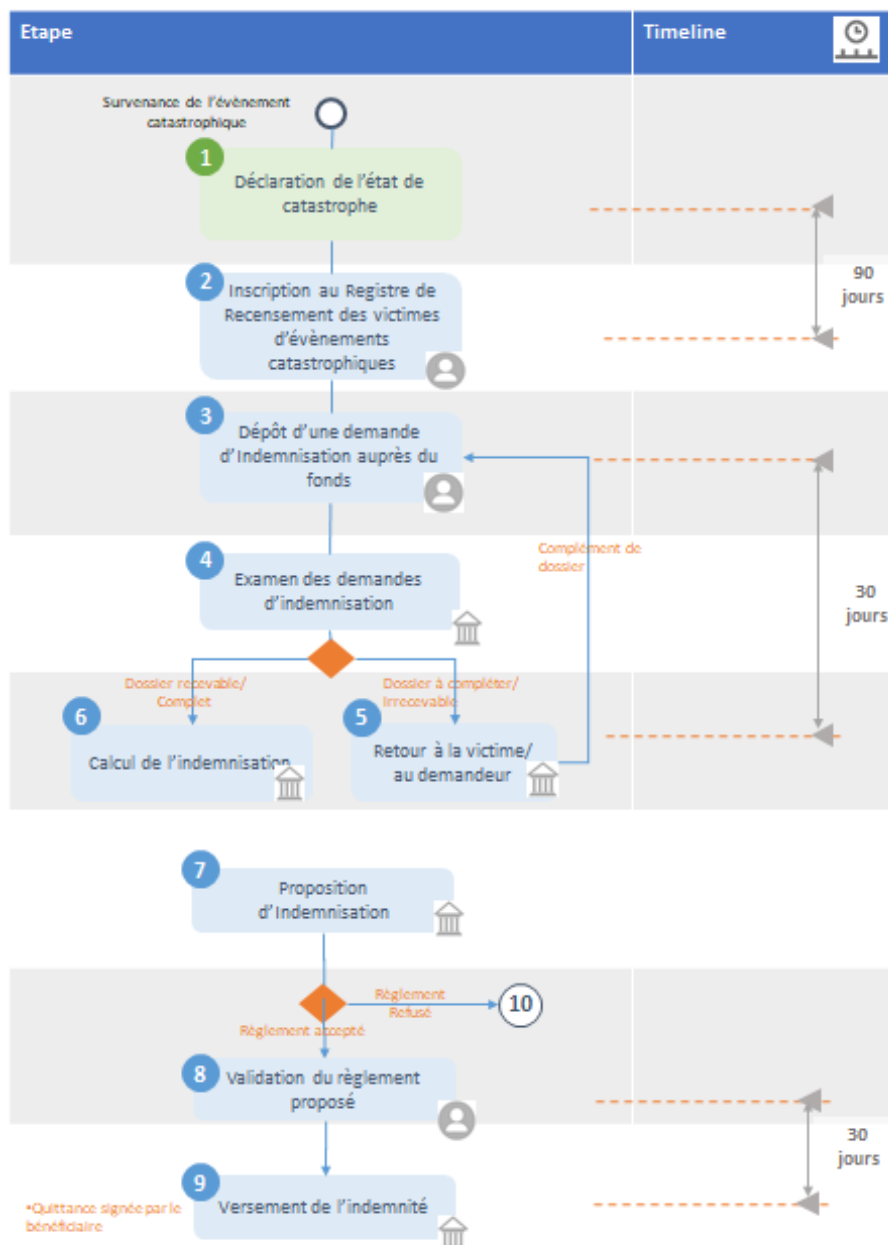
- une dotation initiale de l'Etat fixée par une loi de finance
- le produit des taxes parafiscales instituées à son profit par voie réglementaire ;
- la rémunération pour les services rendus ;
- les produits et intérêts de placement des fonds ;
- les subventions , les dons et les legs
- ...

En dépenses :

- les indemnités versées aux victimes d'évènements catastrophiques ;
- les dépenses de fonctionnement et d'investissement ;
- toutes autres dépenses en relation directe avec les missions du Fonds de solidarité.
- ...

Annexe II

Schéma d'indemnisation



Annexe II. Schéma d'indemnisation

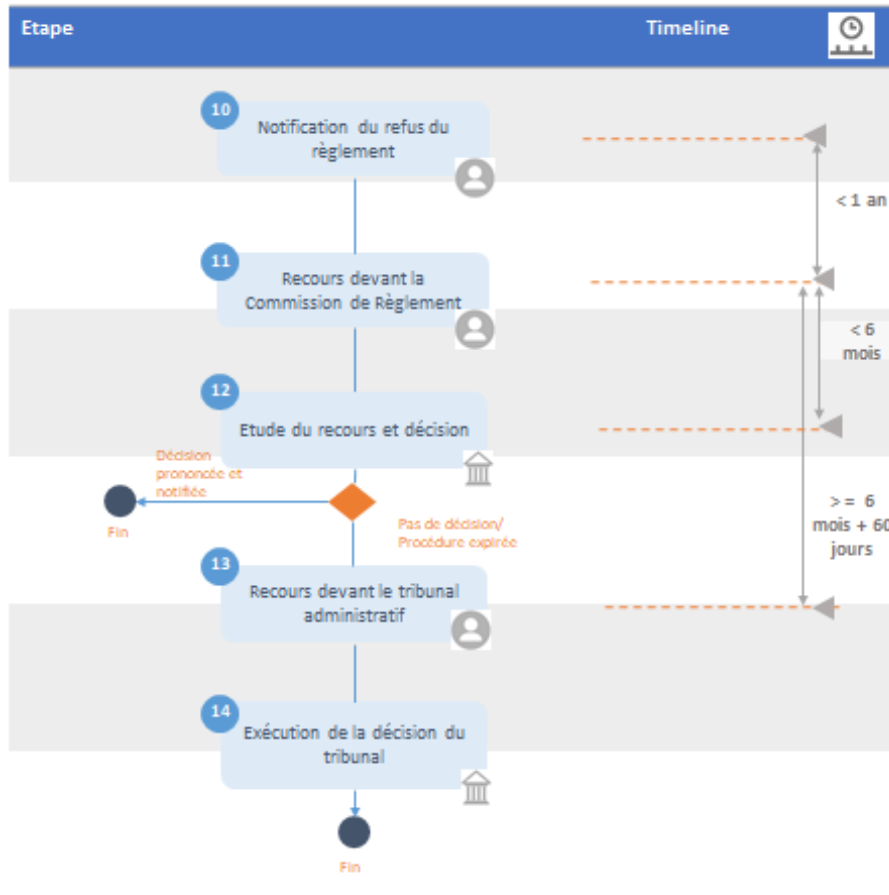


FIG. II.1 : Schéma d'indemnisation

Annexe III

Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov est un test d'adéquation non paramétrique permettant de tester l'hypothèse que nos données proviennent d'une loi de probabilité théorique connue. Ce test repose sur les propriétés asymptotiques de la fonction de répartition empirique :

$$F(\hat{x}) = \sum_{j=1}^n 1_{x_j \leq x}$$

La fonction de répartition empirique est un estimateur sans biais et fortement consistant pour la fonction de répartition F_X ,

nous souhaitons alors tester $H_0 : F_X = \hat{F}_X$ contre $H_1 : F_X \neq \hat{F}_X$.

Le test de Kolmogorov-Smirnov étudie la plus grande différence, en valeur absolue, entre la fonction de répartition empirique \hat{F} et la fonction de répartition F_X de la loi étudiée. On note D_n la statistique de ce test définie ainsi :

$$D_n = \sup_x |F(\hat{x}) - F_X(x)|$$

Plus la variable aléatoire D_n est proche de 0, plus la fonction de répartition empirique est proche de la fonction de répartition $F(x)$ postulée. Pour chaque test de Kolmogorov-Smirnov réalisée, nous avons indiqué la p-value renvoyé par la fonction `ks.test()` du logiciel R. Si D_n est la valeur obtenue de la statistique du test, la p-value représente la probabilité d'observer une réalisation pour une variable aléatoire D_n , sous l'hypothèse que l'échantillon (x_1, x_2, \dots, x_n) est issue de la loi de probabilité de fonction de répartition F_X postulée. Ainsi, plus cette *p-value* est proche de 0, plus l'hypothèse que les (x_1, x_2, \dots, x_n) sont issus de F_X sera à rejeter. Inversement, une *p-value* proche de 1 indique une bonne modélisation.

Annexe IV

Choix des seuils de modélisation

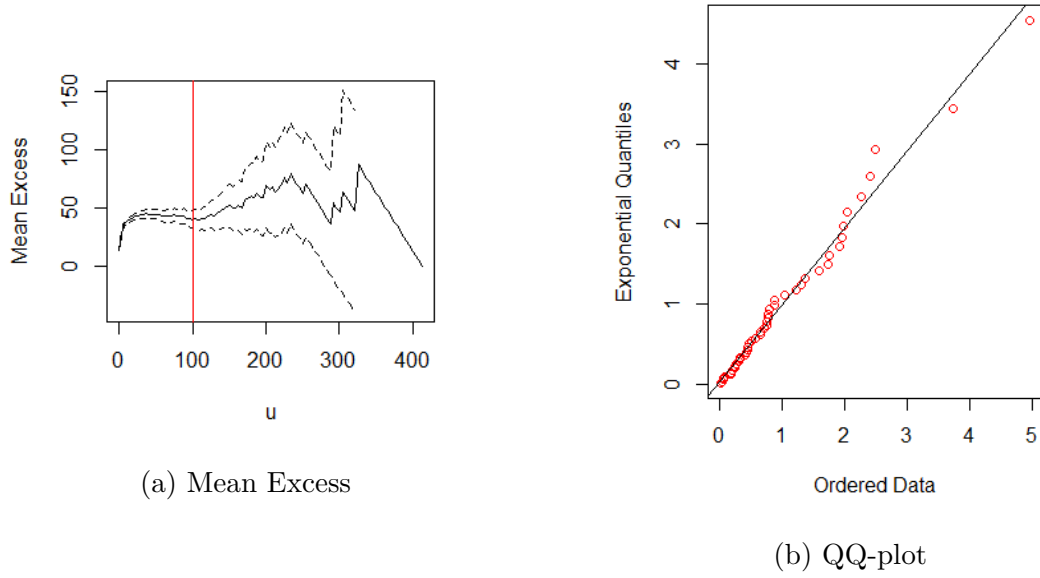


FIG. IV.1 : Modélisation- province Salé

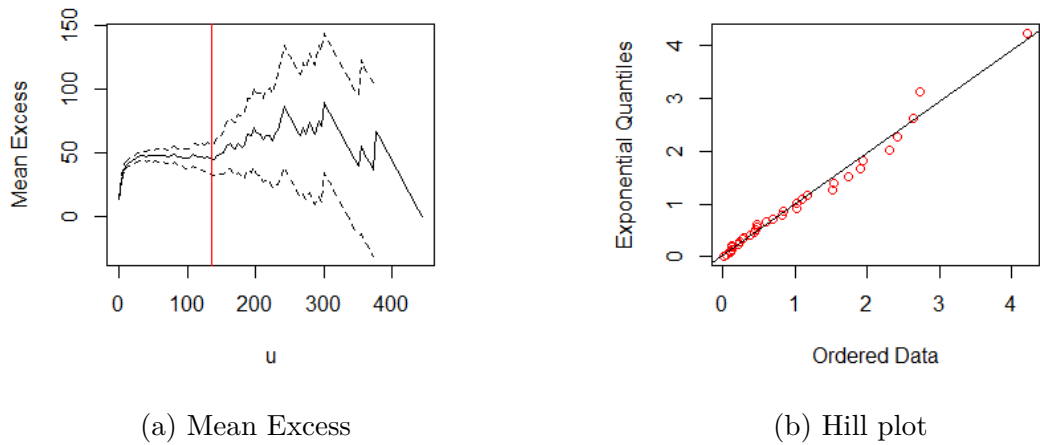
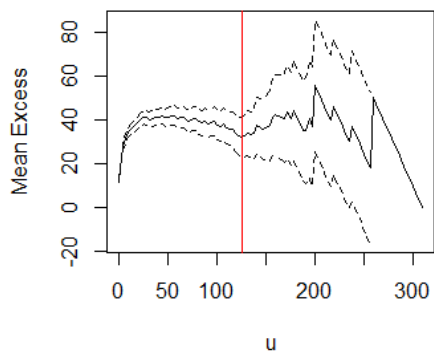
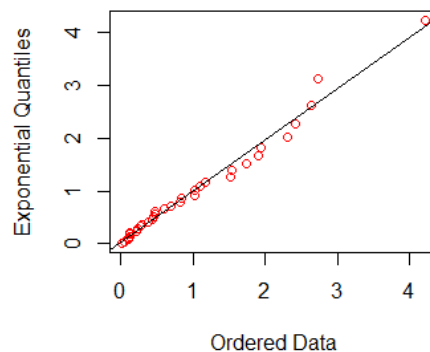


FIG. IV.2 : Modélisation- province Kenitra

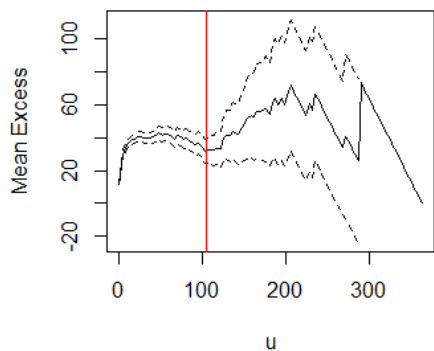


(a) Mean Excess

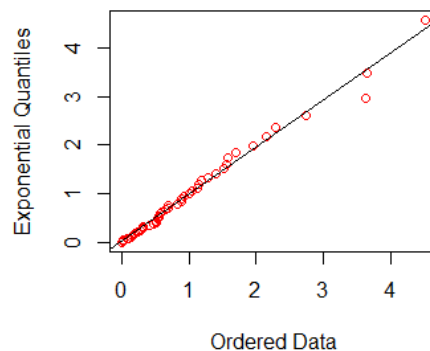


(b) QQ-plot

FIG. IV.3 : Modélisation- province Casablanca

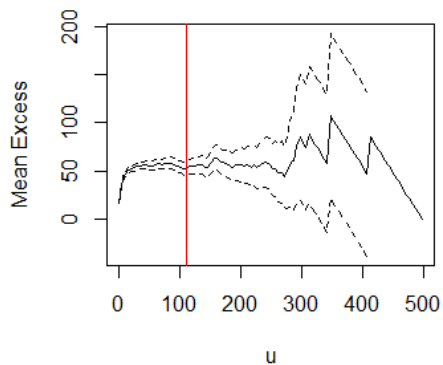


(a) Mean Excess

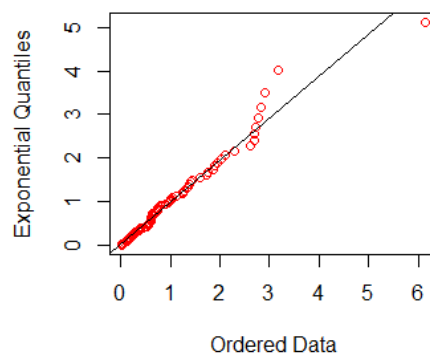


(b) QQ-plot

FIG. IV.4 : Modélisation- province Mohammedia

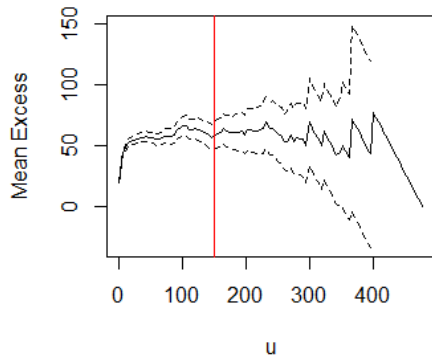


(a) Mean Excess

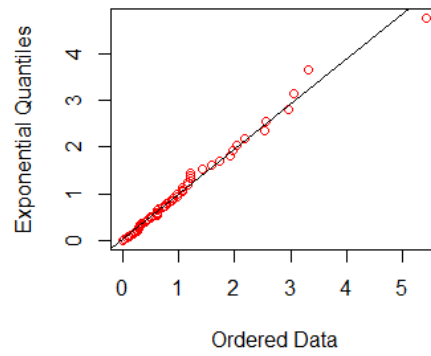


(b) QQ-plot

FIG. IV.5 : Modélisation province Tanger Assilah



(a) Mean Excess



(b) QQ-plot

FIG. IV.6 : Modélisation province Tétouane

Annexe V

Paramètres GPD Estimés & Expositions

Province	Seuil de modélisation en mm	ξ	σ	P-value KS
RABAT	110	0.287	31.525	0.5447
SALE	100	0.149	42.119	0.972
KENITRA	135	0.236	42.7175	0.965
CASABLANCA	125	-0.0344	44.980	0.999
MOHAMMEDIA	105	0.262	29.935	0.8122
TANGER ASSILAH	110	0.0016	62.846	0.7341
TETOUAN	150	-0.066	75.965	0.9871

TAB. V.1 : Estimation des paramètres GPD

Province	Nombre residences exposées	Exposition (MAD)	Taux d'exposition
RABAT	160 367	42 692 822 612	0,036
SALE	184 843	38 580 571 939	0,033
KENITRA	216 687	41 556 084 649	0,035
CASABLANCA	688 516	158 069 479 299	0,13
MOHAMMEDIA	70 227	13 499 324 935	0,011
TANGER ASSILAH	178 919	38 354 796 908	0,033
TETOUAN	22 584	3 441 665 135	0,0029
AGADIR	106 045	21 355 397 377	0,018
FES	219 044	49 749 134 747	0,042
MARRAKECH	232 516	50 679 193 080	0,043
OUARZAZATE	80 763	14 030 626 375	0,012

TAB. V.2 : Expositions de quelques provinces

Annexe VI

Matrice de corrélation estimée

	Rabat	Sale	Kenitra	Casablanca	Mohammadia	Tanger	Tétouan
Rabat	1.00	0.92	0.69	0.65	0.68	0.36	0.28
Sale	0.92	1.00	0.78	0.64	0.70	0.48	0.35
Kenitra	0.69	0.78	1.00	0.47	0.57	0.56	0.42
Casablanca	0.65	0.64	0.47	1.00	0.84	0.26	0.19
Mohammadia	0.68	0.70	0.57	0.84	1.00	0.29	0.23
Tanger	0.36	0.48	0.56	0.26	0.29	1.00	0.70
Tétouan	0.28	0.35	0.42	0.19	0.23	0.70	1.00

TAB. VI.1 : Matrice de corrélation estimée