

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuares
le 08/11/2023

Par : **Marie GANON**

Titre : **Prise en compte du risque climatique dans le
provisionnement agricole**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Entreprise : Milliman

Nom : Caroline HILLAIRET

Signature : 

*Membres présents du jury de l'Institut
des Actuares*

Directeur du mémoire en entreprise :

Nom : Eve Elisabeth TITON

Signature : 

**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents actuariels
(après expiration de l'éventuel délai de
confidentialité)**

Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat



Résumé

En raison du dérèglement climatique, les événements naturels, tels que la sécheresse ou la grêle, verront leur fréquence et leur intensité s'accroître dans les années à venir. Par conséquent, la prise en compte du climat est un enjeu majeur pour les acteurs du secteur de l'assurance agricole. Si la littérature actuarielle est vaste sur la prise en compte d'effets calendaires type inflation dans le provisionnement, elle est en revanche très pauvre en ce qui concerne la modélisation d'impacts climatiques dans ce domaine.

En qualité de « boîte à outils », ce mémoire apporte une démarche d'évaluation du risque climatique dans le cadre du provisionnement Non-Vie. Fortement exposé à ce type de risque, le secteur de l'assurance agricole est un bon exemple d'application, les événements climatiques détériorant la qualité des cultures et affectant les rendements. De plus, la réforme de l'assurance MRC risque d'accroître la souscription de contrats agricoles, et donc la problématique de provisionnement.

Dans un premier temps, une analyse qualitative et quantitative du risque climatique dans le provisionnement d'une branche agricole sera proposée : qualitative via une étude de la forme de la distribution des réserves, et quantitative par une extraction d'un facteur d'inflation climatique modélisé et projeté à partir de variables météorologiques. Par la suite, ce mémoire sera complété par des estimations de probabilités de retour de survenance de pertes de rendements extrêmes à partir d'indicateurs climatiques. Cette dernière analyse pourrait s'avérer pertinente afin d'aiguiller les actuaires cherchant à affiner leur provisionnement et / ou tarification.

Mots-clés : Provisionnement, Risque de réserve, Risque climatique, Assurance agricole, Dal Moro, Skewness, Kurtosis, Méthode de séparation, Théorie des valeurs extrêmes.

Abstract

As a result of climate change, natural events such as drought and hail are set to increase in frequency and intensity over the coming years. Consequently, taking climate into account is a major challenge for players in the insurance industry. While the actuarial literature is extensive on taking account of calendar effects such as inflation in reserving, it is very sparse on the modeling of climatic impacts in this field

As a “toolbox”, this thesis provides an approach for assessing climate risk in the context of Non-Life reserving. Highly exposed to this type of risk, the agricultural insurance sector is a good example of application, as climatic events deteriorate crop quality and affect yields. Besides, the reform of the multi-risk climate crop insurance is likely to increase the number of agricultural contracts underwritten, and hence the question of reserving.

Firstly, a qualitative and quantitative analysis of the climatic risk in the provisioning of an agricultural branch will be proposed: qualitative via a study of the shape of the distribution of reserves, and quantitative via the extraction of a modeled and projected climatic inflation factor from meteorological variables. Then, this thesis will be completed by estimates of return probabilities of extreme yield losses based on climatic indicators. This latter analysis could be relevant in guiding actuaries seeking to refine their reserving and/or pricing.

Key-words: Reserving, Reserving risk, Climate risk, Crop insurance, Dal Moro, Skewness, Kurtosis, Separation method, Extreme-value theory.

Remerciements

Je tiens tout d'abord à remercier Alexandre Boumezoued, *Principal* du département Recherche et Développement de Milliman, pour m'avoir offert l'opportunité de travailler sur ce sujet de mémoire. Son optimisme et sa faculté à se retourner vers d'autres solutions et d'autres approches ont beaucoup œuvré à la réalisation de ce mémoire.

Je remercie profondément ma tutrice Eve Elisabeth Titon pour sa qualité d'encadrement, sa bienveillance, sa disponibilité et ses précieux conseils. Elle m'a beaucoup apporté, professionnellement et humainement, et je lui en suis très reconnaissante.

Je souhaite également remercier l'ENSAE pour ses enseignements et ma tutrice pédagogique Caroline Hillairet pour ses relectures.

Je remercie de manière générale toute l'équipe R&D de Milliman et l'ensemble du cabinet pour leur accueil et leur aide sur certains points de ce mémoire.

J'ai une pensée particulière pour mes parents et mes grands-parents, qui m'ont donné le goût de l'école et qui m'ont accompagnée et soutenue durant tout mon parcours académique.

Je tiens enfin à remercier Yohan, pour son soutien sans faille depuis plus de 3 ans, et tous mes proches qui ont contribué à la relecture de ce mémoire.

Table des matières

Résumé	I
Abstract	II
Remerciements	III
Introduction	1
I Contexte : Risque climatique et assurance agricole	3
1 Risque climatique	4
1.1 Définition du risque climatique	4
1.2 Différentes projections envisagées	5
1.3 Enjeux pour les assureurs	9
1.4 Recommandations de l'ACPR	11
2 Assurance agricole et risque climatique	14
2.1 Assurance climatique des récoltes	14
2.2 Conséquences du risque climatique sur l'agriculture	17
2.3 Réforme de l'assurance MRC du 1 ^{er} janvier 2023	18
2.4 Problématiques de provisionnement soulevées par le climat	19
3 Panorama des aléas climatiques en agriculture	21
3.1 Sécheresse	21
3.2 Grêle	24
3.3 Gel	26
3.4 Autres aléas climatiques	29
II Provisionnement : Cadre théorique	31
4 Méthodes usuelles de provisionnement	32
4.1 Modèle de Mack	34
4.2 Modèle de Wüthrich, Merz et Lysenko	37
4.3 Modèles économétriques de provisionnement	43
4.4 Simulation des montants des réserves par bootstrap	44
5 Estimation du skewness et du kurtosis	48
5.1 Définitions et interprétations	48
5.2 Estimation du skewness et du kurtosis des paiements cumulés	49
5.3 Approche par formule fermée	52

5.4	Approche par simulation	55
5.5	Approche par bootstrap	58
6	Application des méthodes de provisionnement	61
6.1	Données de triangle	61
6.2	Estimation du montant des réserves et du risque de réserves	65
6.3	Evaluation du skewness et du kurtosis	68
6.4	Distribution finale des réserves	70
7	Intégration du climat dans le provisionnement	71
7.1	Traitement de l'inflation	71
7.2	Méthode de séparation climatique	72
III	Exemple d'application : Provisionnement climatique en assurance agricole	77
8	Présentation des données agricoles	78
8.1	Analyse d'ensemble des triangles	78
8.2	Validation des hypothèses de provisionnement	81
8.3	Déflation du triangle aléas climatiques	86
9	Etude de la distribution des réserves	89
9.1	Estimation des réserves et de l'erreur de prédiction	89
9.2	Considération de l'asymétrie et de l'aplatissement	91
9.3	Distributions obtenues	92
10	Application de la méthode de séparation climatique	94
10.1	Hypothèses de simulation	94
10.2	Résultats du modèle	95
10.3	Prise en compte de variables climatiques	98
10.4	Conclusions sur l'effet du climat sur les réserves	105
IV	Complément : Modélisation d'événements extrêmes	107
11	Eléments de théorie des valeurs extrêmes	108
11.1	Loi limite du maximum	108
11.2	Loi des dépassements de seuil	113
11.3	Calibrage de lois d'événements extrêmes	114
12	Données agricoles et données climatiques utilisées	121
12.1	Données agricoles	121
12.2	Données climatiques	127
13	Calibrage de lois de pertes de rendements extrêmes	131
13.1	Calibrage avec données historiques	131
13.2	Calibrage avec mise en <i>as if</i>	139
13.3	Conclusions sur l'effet du climat sur les calibrages	145
	Conclusion	147

Bibliographie	148
Table des figures	152
Liste des tableaux	155
Glossaire	157
Note de synthèse	158
Executive summary	165
Annexes	173
A Compléments sur l'assurance MRC	174
B Compléments sur la méthode de simulation de Dal Moro	175
B.1 Modèle GEV	175
B.2 Rappels sur les copules	176
C Application de la méthode de séparation climatique	178
C.1 Hypothèses formulées	178
C.2 Résultats obtenus	179
C.3 Distributions obtenues	180
D Validation des hypothèses de provisionnement	182
D.1 Triangle agricole hors tempête et aléas climatiques	182
D.2 Triangle agricole responsabilité civile	185
E Modèles de lien entre tendance climatique et SWI	188
E.1 Modèles simples	189
E.2 Modèles avec inclusion de retards	189
E.3 Modèles avec effet de seuil	191
F Modèles de lois de dépassements retenus	192
F.1 Modèle M2 pour les données historiques	192
F.2 Modèle M3 pour les données historiques	194
F.3 Modèle M2 pour les données mises en <i>as if</i>	196
F.4 Modèle M3 pour les données mises en <i>as if</i>	198

Introduction

En raison du dérèglement climatique, les événements naturels, tels que la sécheresse ou la grêle, verront leur fréquence et leur intensité s'accroître dans les années à venir. Dans l'agriculture, le changement climatique affecte de manière conséquente les rendements des exploitations. En effet, les hausses de température et les canicules favorisent notamment la sécheresse des terres agricoles, ce qui détériore leur fertilité et donc la qualité des cultures et leur prix de vente.

Par conséquent, la prise en compte du risque climatique est un enjeu majeur pour les acteurs du secteur de l'assurance, notamment dans le cadre de l'évaluation des réserves. Si la littérature actuarielle est vaste sur la prise en compte d'effets calendaires type inflation dans le provisionnement, elle est cependant très pauvre en ce qui concerne la modélisation d'impacts climatiques dans ce domaine.

Ce mémoire s'inscrit donc dans une optique de constitution d'une « boîte à outils » apportant une démarche d'évaluation du risque climatique dans les réserves d'une compagnie d'assurance Non-Vie. Il présente plusieurs méthodes répondant à diverses problématiques. Tout d'abord, il s'appuie sur des données émanant de triangles de provisionnement agricoles, produits à partir de garanties multirisque climatique des récoltes (MRC), en vue d'étudier la forme des distributions des réserves sous-jacentes aux branches considérées et d'identifier un éventuel effet dû au climat. Ensuite, l'objectif sera d'extraire une tendance liée à la survenance d'événements naturels. Enfin, cette étude de provisionnement s'accompagnera d'une modélisation des événements extrêmes, ici des pertes de rendements, que l'on cherchera à relier à des données météorologiques. Fortement exposé aux risques climatiques, le secteur agricole constitue un bon exemple d'application de notre démarche. De plus, la réforme sur l'assurance agricole du 1^{er} janvier 2023 visant à accroître la souscription de contrats MRC, la problématique de provisionnement pour cette branche devrait prendre de l'ampleur dans les années à venir.

Ce mémoire s'articule en quatre parties :

- La **première partie** fournit des éléments de contexte sur le risque climatique ainsi que l'assurance agricole. Elle dresse également un panorama des aléas climatiques qui menacent les cultures.
- La **deuxième partie** présente les différentes méthodes de provisionnement existantes, et se concentre en particulier sur le modèle de Mack - *chain-ladder*. Elle explicite également une approche de prise en compte des moments d'ordre 3 et 4, skewness et kurtosis, dans l'estimation des réserves à partir d'articles de Dal Moro. Une nouvelle méthodologie de bootstrap a notamment été développée en tenant compte de ses travaux. Ces différentes approches sont illustrées à travers un exemple de triangle issu d'un article de Mack [36]. Enfin, cette partie intègre un chapitre méthodologique, inspiré des méthodes de séparation utilisées pour l'inflation, permettant d'extraire une tendance climatique des réserves.
- La **troisième partie** applique les méthodologies décrites en amont dans le cadre de l'assurance agricole. Elle comporte une analyse qualitative et quantitative du risque climatique dans les réserves : qualitative via l'analyse des distributions des réserves pour des

triangles climatiques et non-climatiques, et quantitative à travers l'extraction d'un facteur d'inflation par année de survenance, corrélé à des variables climatiques. Par ailleurs, un point d'attention sera accordé à la validation des hypothèses des modèles utilisés, dans la mesure du possible. En effet, la particularité d'un triangle climatique pourrait venir compromettre ces dernières.

- La **dernière partie** complète le mémoire avec une analyse de la survenance d'événements extrêmes, ici des pertes importantes de rendements, et sur la manière dont ces dernières peuvent être reliées à des indicateurs climatiques. Après une exposition du cadre théorique relatif à la modélisation de ces événements extrêmes, elle s'articulera autour d'un exemple d'application sur des pertes de rendements agricoles historiques.

Première partie

Contexte : Risque climatique et assurance
agricole

Chapitre 1

Risque climatique

Ce chapitre introductif fait un état des lieux des connaissances actuelles à propos du risque climatique. Après une brève classification des différents risques, nous nous attarderons sur les scénarios de projections les plus utilisés pour les anticiper. En effet, le changement climatique est susceptible de fortement affecter les assureurs, c'est pourquoi l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) a émis quelques recommandations sur le traitement du risque climatique.

1.1 Définition du risque climatique

On appelle **risque climatique** l'ensemble des risques relatifs au changement climatique. Ceux-ci peuvent être classés en trois catégories, qui sont présentées dans une publication de l'ACPR [3] :

- le **risque physique**, qui mesure directement les conséquences du changement climatique sur les biens et les personnes. Ce risque impacte donc les assureurs sur leurs branches Vie et Non-Vie :
 - Pour l'assurance Non-Vie, le risque physique se traduit par une augmentation de la fréquence et de la sévérité des aléas climatiques extrêmes, tels que les sécheresses, les cyclones ou les inondations.
 - L'assurance Vie est également touchée par le risque climatique, notamment via l'accroissement des risques sanitaires, tels que la propagation de maladies ou la répétition de vagues de chaleur.

De plus, l'augmentation de la fréquence d'aléas climatiques extrêmes peut également fragiliser l'économie. Dès lors, les actifs détenus par des compagnies d'assurance verraient leur valeur se déprécier.

- le **risque de transition**, qui émane du changement de comportement des agents économiques et financiers devant se soumettre à des réglementations plus strictes, notamment en termes de politiques énergétiques. Ce risque affecte les entreprises dont les activités sont considérées incompatibles avec la transition écologique par le biais d'une perte de valeurs sur leurs actifs. Ainsi, les assureurs sont eux-mêmes exposés s'ils possèdent des actifs financiers dans ces firmes. Le risque de transition peut de surcroît conduire à une perte de contrats d'assurance couvrant les activités jugées polluantes.
- le **risque de responsabilité**, associé aux dommages et intérêts qui seraient à la charge d'une personne morale si elle était condamnée pour sa participation à l'aggravation du changement climatique. Un assureur est exposé de manière directe et/ou indirecte à ce risque :

- de manière directe, si ses actions ont contribué à accroître le changement climatique ;
- ou de manière indirecte, si ses contreparties sont elles-mêmes exposées au risque de réputation.

Il est à noter que ces différentes catégories ont des horizons temporels différents, ce qui complique leur anticipation.

1.2 Différentes projections envisagées

En vue de prévenir ces risques divers, plusieurs organismes se sont attelés à l'élaboration de scénarios climatiques. Ces derniers reposent sur des hypothèses de mise en œuvre de politiques d'adaptation et d'atténuation du changement climatique.

1.2.1 Scénarios du GIEC

Depuis 1988, le Groupe d'experts Intergouvernemental sur l'Evolution du Climat (GIEC) mesure le niveau de connaissance sur l'évolution du climat ainsi que ses causes et conséquences. Il a été fondé par deux organismes des Nations Unies : le Programme des Nations Unies pour l'Environnement et l'Organisation Météorologique Mondiale. Il est ouvert à tous les pays membres de ces institutions et rassemble aujourd'hui des experts de 195 nationalités. Son objectif est d'évaluer de manière méthodique et sans parti pris la qualité des informations scientifiques, socio-économiques et techniques relatives au changement climatique. Le GIEC cherche alors à identifier les éléments relevant d'un consensus de la communauté scientifique, mais également les limites, que ce soit en termes de connaissances ou d'interprétation des résultats. Son travail doit permettre de comprendre le rôle de l'Homme dans le changement climatique afin de mettre en œuvre des stratégies d'atténuation et d'adaptation [43].

Pour le cinquième rapport, paru en 2014, les experts du GIEC introduisent quatre scénarios de projections du climat, établis à partir d'hypothèses sur la quantité de gaz à effet de serre émise. Dès lors, pour chaque scénario, un niveau de **forçage radiatif**¹ peut être calculé. Dans ce cas, le forçage radiatif permet de déterminer dans quelle mesure le réchauffement climatique perturbe l'équilibre énergétique terrestre.

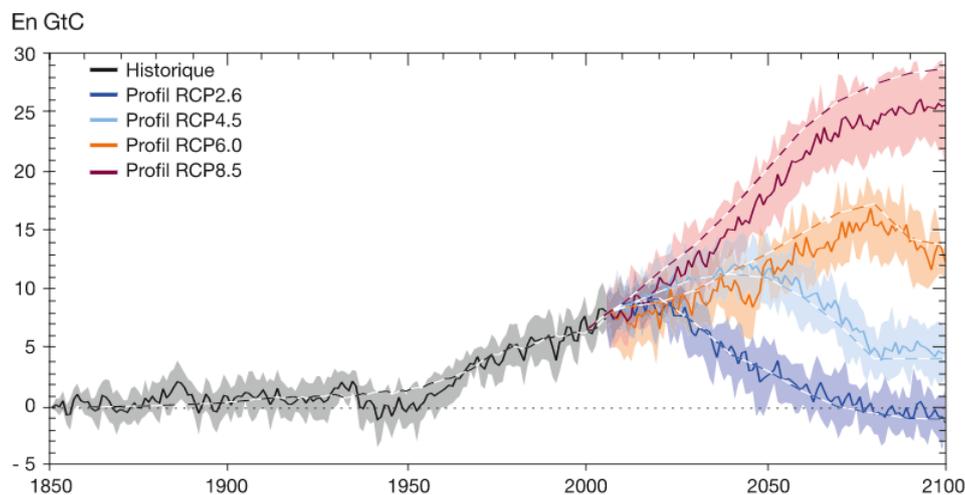


FIGURE 1.1 – Projection des émissions liées aux énergies fossiles selon les scénarios RCP, en gigatonnes de carbone [42]

1. défini comme la différence entre puissance de radiation reçue et puissance de radiation émise par un système climatique, ici la Terre

Ces scénarios sont nommés *Representative Concentration Pathways* (RCP). Le chiffre associé à leur nom quantifie le forçage radiatif résultant à l'horizon 2100 avec les hypothèses de trajectoires d'émissions de gaz à effet de serre choisies. En conséquence, le scénario **RCP 2.6** est le plus optimiste, tandis que le scénario **RCP 8.5** est le plus pessimiste, avec des concentrations de CO₂ très élevées, comme présenté en Figure 1.1.

A partir de ces trajectoires d'émissions, l'évolution des températures peut être modélisée sous différents scénarios. La Figure 1.2 retrace l'évolution de la température de surface moyenne selon les diverses projections.

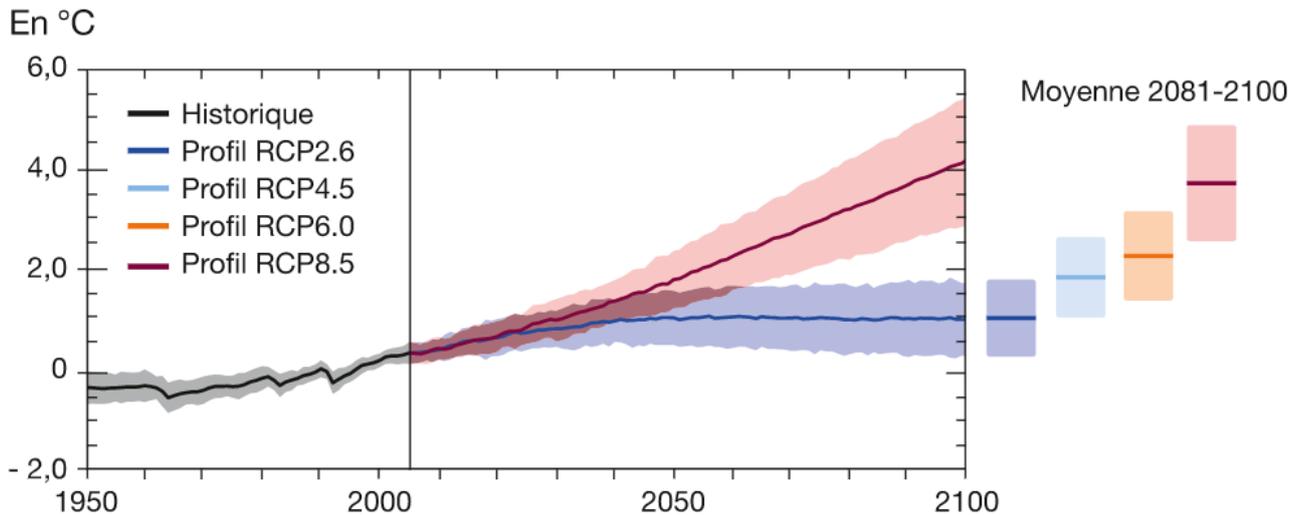


FIGURE 1.2 – Variations de la température de surface moyenne mondiale selon les scénarios RCP, relativement à la période 1986-2005 [42]

Afin de tenir compte des enjeux socio-économiques dans la stratégie d'adaptation au changement climatique, les experts du GIEC ont complété les scénarios RCP avec 5 trajectoires SSP (*Shared Socio-economic Pathways*). Celles-ci décrivent une évolution future du développement socio-économique en l'absence de politiques climatiques pour cinq thématiques et sont présentées dans le dernier rapport de 2022 :

- Les trajectoires **SSP1** et **SSP5** décrivent des scénarios de croissance économique rapide et d'investissements conséquents dans l'éducation et la santé. Si la trajectoire **SSP1** s'appuie pour cela sur les énergies durables, la **SSP5** repose quant à elle sur les énergies fossiles.
- Concernant les trajectoires **SSP3** et **SSP4**, peu d'investissements sont réalisés dans la santé et l'éducation. La croissance démographique est élevée et les inégalités s'accumulent. En particulier, la **SSP3** décrit une situation de rivalité entre les pays, tandis que la **SSP4** envisage de fortes inégalités intra et inter pays.
- Enfin, le scénario **SSP2** traduit une trajectoire intermédiaire, sans déviation des tendances actuelles.

Chacune d'entre elles est associée à un niveau de forçage radiatif y , dans l'optique d'établir un lien avec les scénarios climatiques. On obtient alors une trajectoire SSP_{x-y} , où x est le scénario socio-économique considéré. Les experts du GIEC utilisent notamment les trajectoires **SSP5-8.5** et **SSP3-7.0** pour modéliser des scénarios de fortes émissions de gaz à effet de serre. La trajectoire **SSP2-4.5** caractérise un taux d'émission plus modéré. Enfin, les scénarios **SSP1-2.6** et **SSP1-1.9** sont relatifs à de faibles voire très faibles émissions de gaz à effet de serre. En particulier, le scénario **SSP1-1.9** est conforme à l'objectif de réchauffement maximal

de 1,5°C déterminé au cours de l'accord de Paris de 2015 [34].

Les conséquences de chaque scénario en terme de réchauffement climatique sont illustrées sur la Figure 1.3.

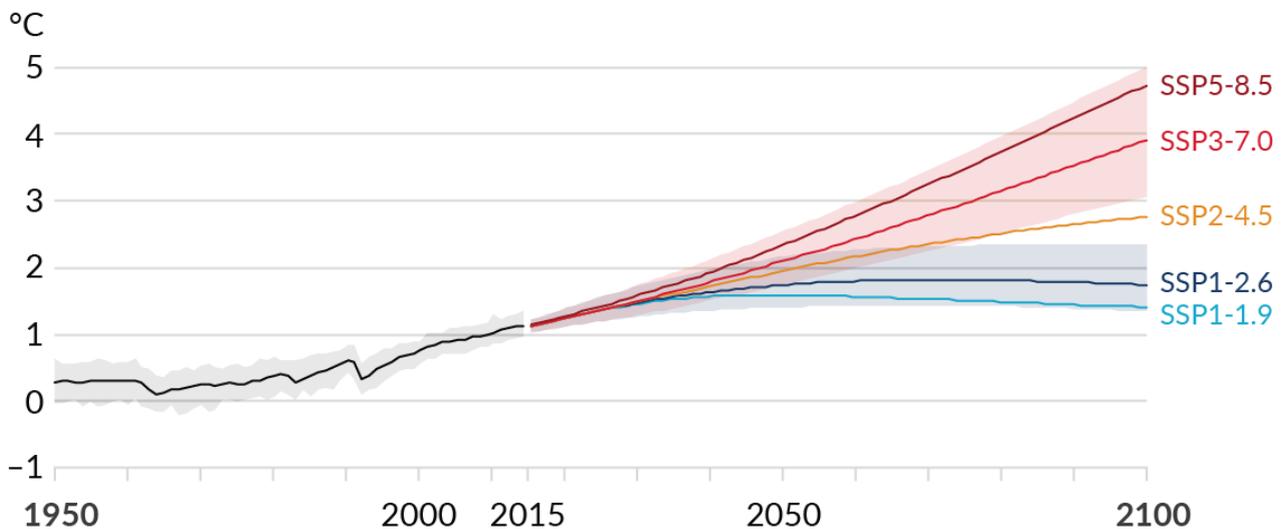


FIGURE 1.3 – Variations de la température de surface moyenne mondiale selon les scénarios SSP, relativement à la période 1850-1900 [24]

1.2.2 Scénarios du NGFS

Le *Network for Greening the Financial System* (NGFS) a été fondé par huit banques centrales et superviseurs lors du sommet *One Planet Summit* qui s'est tenu en Décembre 2017 à Paris [45]. Le Réseau a pour mission de contribuer au renforcement des politiques climatiques à instaurer en vue d'atteindre les objectifs fixés dans les Accords de Paris de 2015. Il vise également à accroître le rôle du système financier, que ce soit dans la gestion des risques ou dans la mobilisation de capitaux à investir dans des placements verts et à faible émission de carbone. Dans cette optique, le Réseau conduit ou réalise des travaux sur la finance verte et en parallèle développe et promeut les bonnes pratiques à adopter.

Les membres du NGFS présentent quatre grandes classes de scénarios [44], schématisées sur la Figure 1.4 :

1. **Transition ordonnée** (*Orderly*), qui suppose que les politiques sont rapidement mises en œuvre et deviennent de plus en plus contraignantes. Dès lors, les risques de transition et les risques physiques sont faibles.
2. **Transition désordonnée** (*Disorderly*), qui explore des scénarios dans lesquels le risque de transition est plus élevé, en raison de retards ou de divergences de politiques entre les secteurs et les pays.
3. **Monde sous serre** (*Hot house world*), qui caractérise un monde où certaines politiques climatiques sont mises en vigueur, mais où les efforts financiers sont insuffisants pour limiter le réchauffement climatique.
4. **Trop peu, trop tard** (*Too little, too late*), qui dépeint des situations dans lesquelles une transition tardive ne parvient pas à limiter les risques physiques.

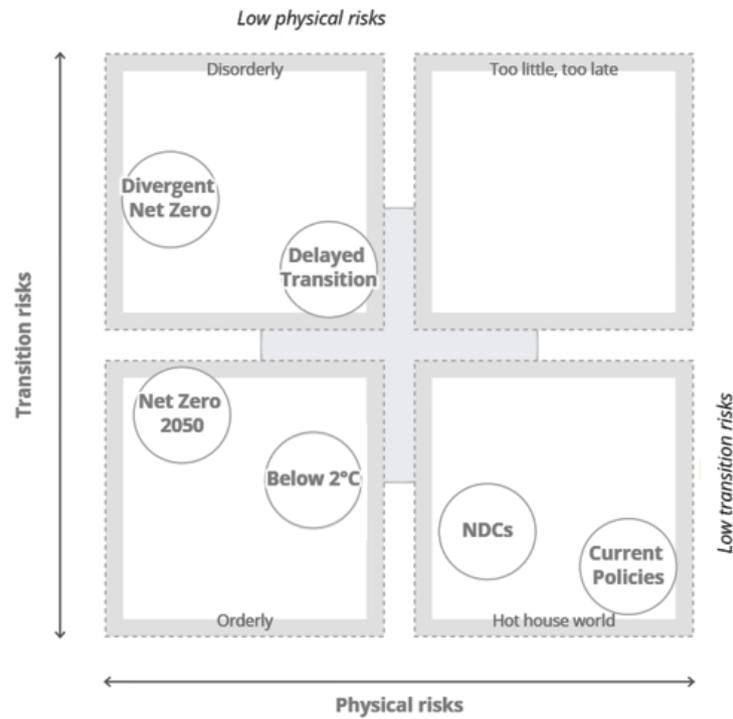


FIGURE 1.4 – Présentation des classes de scénarios du NGFS [44]

Parmi ces catégories, plusieurs scénarios sont envisagés [44] :

1. **Scénario « Net Zéro 2050 »** (*Net Zero 2050*), dans lequel le réchauffement climatique est limité à $1,5^{\circ}\text{C}$ grâce à des innovations et à des politiques climatiques contraignantes. Des émissions nettes zéro carbone sont atteintes en 2050, tandis que des technologies de captation de CO_2 sont massivement développées pour décarboner l'atmosphère.
2. **Scénario « Sous les 2°C »** (*Below 2°C*), dans lequel le renforcement croissant des politiques climatiques permet de limiter le réchauffement climatique en deçà de 2°C avec une probabilité de 67%. Des émissions nettes zéro carbone sont atteintes en 2070, tandis que les technologies de captation de CO_2 sont elles moins développées.
3. **Scénario « Divergent Net Zéro »** (*Divergent Net Zero*), pour lequel l'objectif zéro carbone est atteint en 2050 mais aura été plus coûteux en raison de politiques divergentes. Contrairement au scénario « Net Zéro 2050 », l'utilisation de technologies de décarbonation est plus faible, tandis que le risque de transition est plus élevé.
4. **Scénario « Transition retardée »** (*Delayed Transition*), qui émet l'hypothèse que les émissions annuelles mondiales ne chutent pas avant 2030. A compter de cette date, des politiques très strictes sont mises en œuvre pour contenir le réchauffement climatique sous la barre des 2°C . Les technologies de décarbonation sont là encore assez faibles. Ainsi, les niveaux de risques physique et de transition se voient supérieurs à ceux des scénarios « Net Zéro 2050 » et « Sous les 2°C », comme en témoigne la Figure 1.4.
5. **Scénario « Contributions déterminées au niveau national »** (*Nationally Determined Contributions, NDCs*), qui inclut toutes les politiques prévues, même si elles ne sont pas mises en œuvre. Ce scénario suppose que les politiques nationales demeurent modérées et hétérogènes. Bien que les émissions de carbone diminuent, le réchauffement climatique atteint les $2,6^{\circ}\text{C}$.
6. **Scénario « Politiques actuelles »**, qui suppose que les politiques en matière de changement climatique n'évoluent pas. De fait, les émissions carbone sont croissantes jusqu'en 2080, tandis que le réchauffement, atteignant 3°C , entraîne des changements irréversibles

tels qu'une forte montée des eaux. Ce scénario étant le plus pessimiste, il permet d'évaluer les conséquences en termes de risques physiques de l'inaction climatique.

Les caractéristiques propres à chaque scénario sont résumées dans la Table 1.1.

Scénario	Objectif	Réaction politique	Changement technologique	Utilisation de technologies de décarbonation	Variation régionale des politiques
1	1,4°C	Immédiate et en douceur	Rapide	Modéré à élevé	Modérée
2	1,6°C	Immédiate et en douceur	Modéré	Modéré à élevé	Faible
3	1,4°C	Immédiate mais divergente	Rapide	Faible à modéré	Modérée
4	1,6°C	Retardée	Lent ou rapide	Faible à modéré	Elevée
5	2,6°C	Dépendante du pays	Lent	Faible à modéré	Modérée
6	3°C	Aucune, politiques actuelles	Lent	Faible	Faible

TABLE 1.1 – Récapitulatif des différents scénarios

1.3 Enjeux pour les assureurs

1.3.1 Exposition aux risques climatiques

Courant 2018, l'Autorité de Contrôle Prudentiel et de Résolution a sondé des assureurs français sur l'identification et la gestion de risques climatiques [4]. Les 139 organismes ayant répondu au sondage enregistrent près de 2090 milliards d'euros d'actifs, soit 80% des placements des assureurs français, et comptabilisent 1758 milliards d'euros de provisions techniques, soit 83% des provisions du marché. Parmi les sondés, seuls 28% n'ont pas établi de processus d'identification des risques relatifs au changement climatique. Pour les autres, à partir des réponses des sociétés d'assurance, l'ACPR a pu élaborer le graphique suivant.

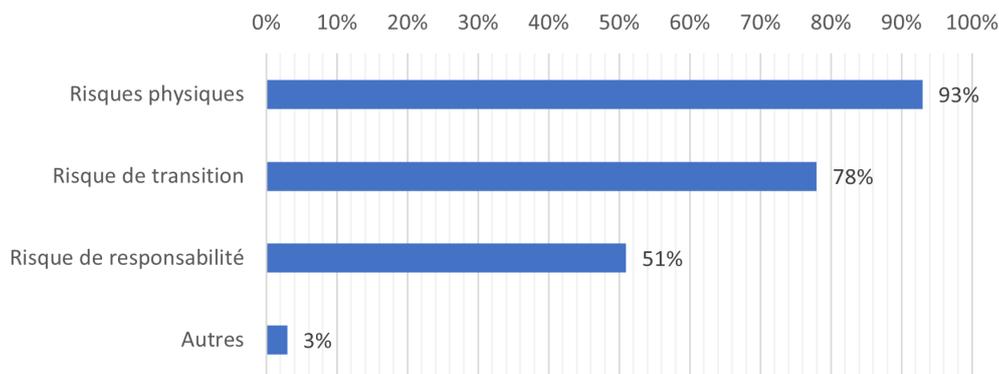


FIGURE 1.5 – Part de sociétés d'assurance interrogées exposée à chaque type de risque [4]

Le graphique ci-dessus démontre bien que tous les acteurs du secteur sont fortement exposés aux risques climatiques. Plus de neuf assureurs sur dix sont soumis aux risques physiques, tandis que près de huit assureurs sur dix doivent gérer le risque de transition. De fait, cette exposition majeure aux risques climatiques doit être mesurée plus précisément.

1.3.2 Projections des coûts pour les risques physiques

En 2021, France Assureurs a publié une étude visant à évaluer l'impact du changement climatique sur l'assurance en 2050 pour la France métropolitaine et les DOM [20]. Celle-ci se

fonde sur la projection de deux types de données :

- des données **socio-économiques**, telles que le nombre de logements, d’entreprises, leur concentration géographique ainsi que la croissance de l’activité ;
- et des données **climatiques**, pour lesquelles les projections reposent sur le scénario RCP 8.5 du GIEC.

Les auteurs de l’étude ont alors distingué différents effets expliquant les écarts entre projections et réalité constatée :

- des facteurs **socio-économiques** :
 - L’effet **richesse**, qui se traduit par une plus forte concentration en logements, en entreprises et en infrastructures territoriales. Ainsi, les conséquences d’un événement naturel futur seront accentuées en fonction du taux d’enrichissement de la zone touchée.
 - L’effet de **répartition** géographique, qui caractérise le niveau d’exposition aux risques naturels des richesses de la zone.
- et des facteurs **climatiques** :
 - Le **changement climatique**, qui, comme précédemment, se définit comme un accroissement des fréquences d’occurrence et d’intensité des événements naturels.
 - La **variabilité naturelle du climat**, indépendante du changement climatique, mais qui permet de tenir compte d’événements extrêmes non encore survenus dans la modélisation.

Les projections de coûts futurs ont ensuite été réalisées selon quatre périls : sécheresse, inondations, submersions marines et tempête. Dans une optique d’illustration des méthodes utilisées par France Assureurs, seules les études relatives aux deux premiers périls seront présentées.

Afin d’estimer le risque de sécheresse, l’étude utilise le zonage officiel du Bureau en Recherches Géologiques et Minières, lequel identifie les zones exposées au retrait-gonflement des sols argileux. Ensuite, les évolutions du nombre de logements par département et de la proportion de logements en zone à risque ont été croisées en vue de projeter la future exposition au risque. Les facteurs climatiques retenus pour effectuer ces projections sont les niveaux de précipitations moyens journaliers et les températures maximales.

Le détail des coûts futurs estimés est présenté en Figure 1.6. D’après les estimations de France Assureurs, en 2050, la charge annuelle allouée au péril sécheresse serait alors plus que triplée.

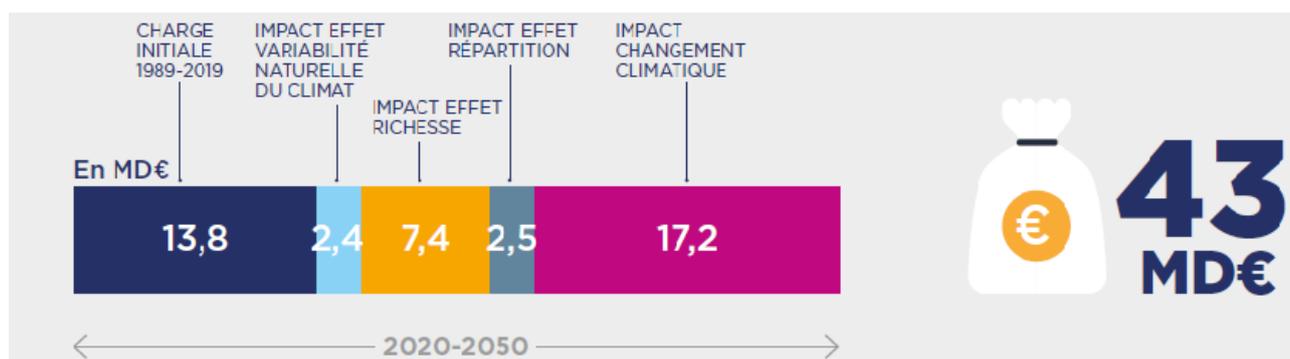


FIGURE 1.6 – Projection des coûts cumulés sécheresse entre 2020 et 2050 [20]

Concernant le risque inondation, des cartes présentant les zones inondables et les territoires à risque important ont été associées aux évolutions de la proportion de structures exposées et du nombre de logements par département. Pour cette partie de l'étude, l'indicateur climatique le plus pertinent correspond aux précipitations maximales par département.

Les résultats obtenus par ces projections sont détaillés sur la Figure 1.7. Ainsi, en 2050, la charge relative au risque d'inondation serait presque doublée.

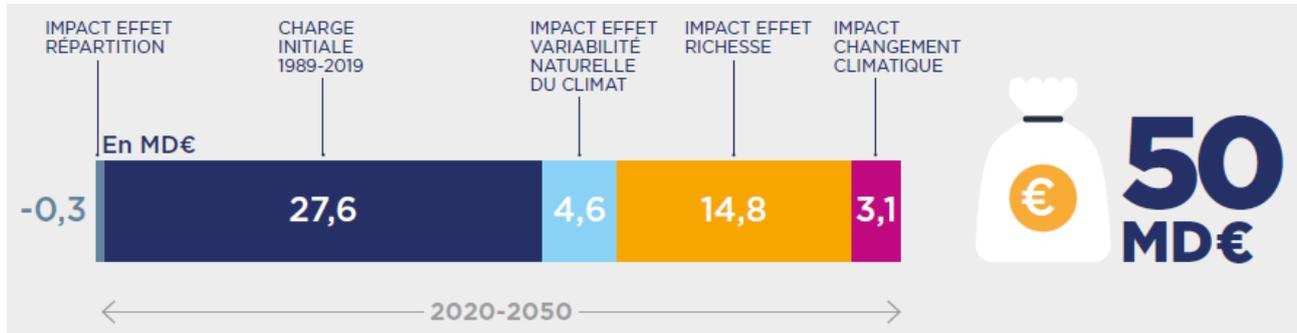


FIGURE 1.7 – Projection des coûts cumulés inondations entre 2020 et 2050 [20]

1.3.3 Impact sur le provisionnement non-vie

Dès lors, les risques climatiques ne peuvent être ignorés dans le processus d'estimation des réserves. Afin que les actuaires prennent en compte le changement climatique dans leurs travaux, Rothwell et al., membres de l'Institut des Actuaires britannique, publient en 2019 un guide de bonnes pratiques [52]. Ils s'intéressent notamment à la problématique de provisionnement et donnent des pistes de réflexion sur l'impact du changement climatique.

Pour les risques physiques, les actuaires sont invités à identifier les facteurs de sinistres liés au climat et à la météo, que ce soit en termes de fréquence ou de sévérité, qui seraient potentiellement affectés par le changement climatique. Les auteurs donnent ainsi l'exemple de la fréquence des sinistres liés au gel en assurance habitation et auto, qui pourrait éventuellement être réduite dans les années à venir en raison du réchauffement climatique.

En vue d'évaluer l'impact du changement climatique dans l'estimation des réserves, Rothwell et al. préconisent d'effectuer une analyse de sensibilité en fonction de divers scénarios, dans le but d'affiner l'intervalle de confiance autour du *Best Estimate*.

1.4 Recommandations de l'ACPR

Afin de tenir compte du risque climatique dans le secteur de l'assurance, l'ACPR développe un plan d'action sur cinq axes [3] :

- la **stratégie**,
- la **gestion des risques**,
- l'**organisation interne**,
- la **sensibilisation des parties prenantes**,
- et la **communication**.

L'expertise actuarielle étant clé dans les deux premiers axes, seuls ceux-ci seront développés.

1.4.1 Stratégie à mettre en œuvre

Pour mettre en œuvre une stratégie efficace, la première étape est d'identifier et détailler l'exposition de l'organisme d'assurance vis-à-vis des différentes composantes du risque climatique. Ce travail permet de mieux appréhender les risques ainsi que leurs impacts et leurs contraintes. L'assureur doit également garder à l'esprit que certaines politiques publiques peuvent modifier sa marge de manœuvre dans l'appréhension des risques.

Par ailleurs, il est nécessaire de tenir compte de la dimension long terme du risque climatique, notamment en fixant des objectifs à des horizons temporels plus courts. En effet, si les plans stratégiques des organismes d'assurance ont une vision à 5 ans maximum, les conséquences du changement climatique commencent déjà à être visibles et vont s'amplifier à l'avenir. Dès lors, la stratégie d'anticipation du risque climatique doit tenir compte du décalage temporel entre les plans stratégiques et les conséquences du changement climatique.

Pour ce faire, il est possible de se fixer des jalons à moyen terme afin de respecter les objectifs définis à long terme. Par exemple, pour respecter l'engagement de neutralité carbone du pacte vert pour l'Europe à l'horizon 2050, certains investisseurs, faisant partie de la *Net Zero Asset Owner Alliance*, cherchent déjà à réduire l'empreinte carbone de leurs portefeuilles d'ici 2025.

Afin de garantir la fiabilité de ces jalons, il est nécessaire de les combiner à des indicateurs qualitatifs ou quantitatifs régulièrement revus et analysés. Ainsi, pour revenir à l'engagement de réduction de l'empreinte carbone des portefeuilles d'investissement, celui-ci s'accompagne d'indicateurs basés sur des critères environnementaux, sociaux et de gouvernance, dits ESG.

Enfin, dans la mesure où le risque climatique est en constante évolution, il est essentiel de comparer les données ainsi que l'efficacité et la pertinence des méthodes employées dans le but d'améliorer la stratégie.

1.4.2 Gestion des risques climatiques

Afin d'établir une cartographie précise des risques climatiques, les assureurs sont encouragés à spécifier les sous-risques des différentes catégories de risques auxquels ils sont confrontés, ainsi que leurs horizons temporels. De là, ils doivent calculer la probabilité de réalisation de ces événements, accompagnée de l'impact financier associé. Dans la conduite de cette tâche, la cartographie définie par l'Autorité européenne des assurances et des pensions professionnelles (EIOPA), détaillant les divers sous-risques physiques et de transition, peut s'avérer particulièrement utile. Les assureurs doivent également envisager le fait que les risques climatiques agissent comme un amplificateur de risques déjà existants, et donc mesurer leurs dommages en conséquence.

L'un des principaux enjeux vis-à-vis du risque climatique concerne la fiabilité et la précision des données exploitées. De fait, la collecte de données fiables et homogènes est difficile. Certains organismes se voient donc dans l'obligation de mobiliser des ressources internes pour améliorer leur qualité, ou bien recourir à des prestataires externes, tandis que d'autres effectuent des croisements et des comparaisons entre différentes sources de données.

De manière générale, pour l'actif, les données sont collectées par pays ou zone géographique. Or, pour appréhender le risque de transition, il est préférable d'avoir une vision plus incrémentale du portefeuille afin d'affiner l'analyse. Néanmoins, un tel niveau de précision sur les actifs est difficile à obtenir.

Pour mesurer le risque physique, la principale difficulté concerne la géolocalisation des assurés avec un degré de précision suffisamment élevé. En effet, puisque l'efficacité de la géolocalisation dépend de son niveau de granularité, il est essentiel de pouvoir raisonner sur des données

fiables. Ce contrôle de fiabilité pourrait passer par une certification des géocodeurs employés.

En outre, les effets du changement climatique commencent déjà à se faire ressentir. Par conséquent, des tendances relatives à l'évolution de la sinistralité pour les risques physiques peuvent être identifiées et quantifiées.

Enfin, pour anticiper les effets du changement climatique, les données historiques collectées doivent être complétées par des scénarios climatiques et une prise en compte des divers horizons temporels. De nombreux organismes utilisent le processus ORSA pour suivre leurs risques climatiques, dans la mesure où ce dernier permet de déterminer les enjeux propres à chaque organisme, ainsi que les interactions avec les autres risques. Dans l'optique d'intégrer des scénarios climatiques au processus ORSA, la Commission européenne propose de réviser l'article 45 de Solvabilité II.

Par ailleurs, l'horizon temporel pris en compte dans le processus ORSA s'étale généralement de 3 à 5 ans. Ainsi, pour analyser le risque climatique, il serait pertinent d'incorporer des *stress tests* à un horizon plus lointain. Par exemple, des *stress tests* climatiques déterminés à partir de scénarios à long terme (après 2050) pourraient être intégrés au processus ORSA.

En parallèle, les assureurs sont invités à développer leurs propres scénarios, ou à utiliser des scénarios déjà disponibles, tels que ceux du NGFS par exemple, utilisés dans le cadre de l'exercice pilote mené par l'ACPR entre 2020 et 2021. Pour les organismes ayant des engagements assurantiels de court terme, l'EIOPA préconise l'utilisation d'un scénario de transition et d'un scénario sans transition. En outre, elle conseille aux organismes présentant des engagements à long terme d'analyser au moins deux scénarios de transition (ordonnée et désordonnée) et de les comparer au scénario sans transition.

Chapitre 2

Assurance agricole et risque climatique

Le 1^{er} janvier 2023, le régime des assurances multirisque climatique des récoltes a été réformé. Pour bien comprendre les enjeux de cette nouvelle réglementation, il faut tout d'abord connaître le fonctionnement de l'assurance agricole, avec les enjeux d'incitations à la souscription qu'elle entraîne. Les conséquences du risque climatique sur l'agriculture, qui ont été un moteur pour la réforme de ce régime, seront ensuite pleinement évoquées. Enfin, après avoir détaillé la nouvelle réglementation et ses impacts, la question du provisionnement dans le cadre des risques agricoles climatiques sera abordée.

2.1 Assurance climatique des récoltes

2.1.1 Types de contrats et couvertures

2.1.1.1 Le régime des calamités agricoles

Afin de protéger les exploitations agricoles contre le risque climatique, le Fonds National de Gestion des Risques en Agriculture (FNGRA) a mis un oeuvre le régime des calamités agricoles, visant à indemniser les pertes liées aux récoltes et aux fonds. Le FNGRA est majoritairement financé par l'intermédiaire de taxes payées par les exploitants agricoles, aquacoles et conchylicoles. L'Etat peut également participer en cas de nécessité.

Néanmoins, certains secteurs tels que les grandes cultures et les vignes ne sont plus éligibles au dispositif des calamités agricoles. De même, le risque lié à la grêle n'est plus couvert. Dans ces deux cas, seul un contrat d'assurance récoltes pourra compenser les pertes de l'exploitation [41].

2.1.1.2 Contrats d'assurance privés

Si un agriculteur souhaite recourir à une assurance privée pour se protéger contre le risque climatique, deux contrats peuvent être envisagés :

- un contrat d'**assurance grêle uniquement** ;
- ou un contrat d'**assurance multirisque climatique** (MRC).

Parmi les contrats d'assurance multirisque climatique, deux catégories de contrats existent :

- les contrats **à la culture**, pour lesquels une indemnisation est versée pour chaque nature de récolte assurée dès lors que la perte de production constatée sur cette dernière dépasse le seuil de déclenchement ;

— et les contrats **à l'exploitation**, pour lesquels une indemnisation est versée dès lors que le total des pertes constatées sur les natures de récoltes assurées dépasse le seuil de déclenchement. Dans ce type de contrat, la mutualisation s'opère au sein de l'exploitation, si bien que les gains effectués sur certaines récoltes peuvent compenser les pertes constatées sur d'autres. Par conséquent, ces contrats sont moins coûteux que les contrats à la culture.

Ces contrats ainsi que leurs garanties sont recensés en Annexe A.

2.1.2 Adoption et incitations

2.1.2.1 Adoption de contrats d'assurance multirisque climatique des récoltes

Grâce à une étude de TerrOïko [57] sur le suivi de l'aide à l'assurance multirisque climatique des récoltes mise en œuvre dans le cadre de la Politique Agricole Commune (PAC) entre 2010 et 2014, il est possible d'observer dans quelle mesure les contrats MRC ont été adoptés par les agriculteurs français.

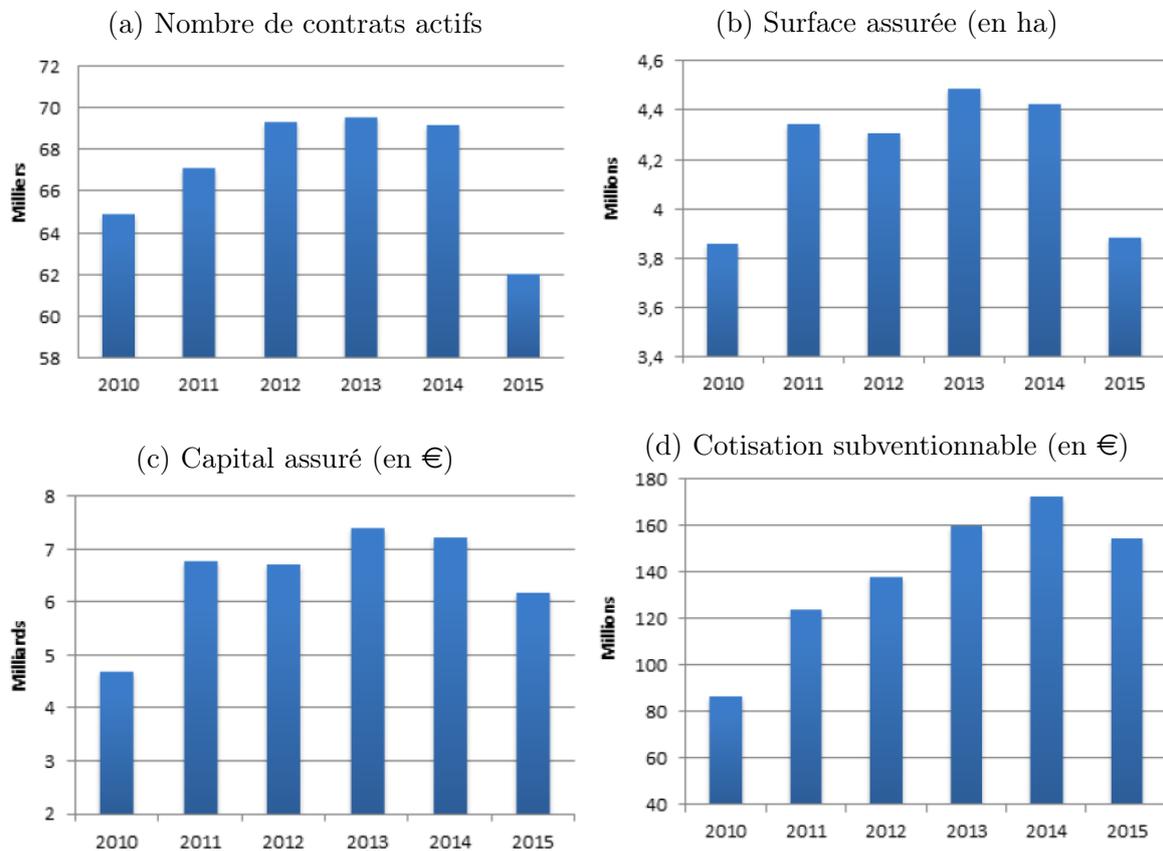


FIGURE 2.1 – Evolution du nombre de contrats souscrits, des surfaces et des capitaux assurés, ainsi que des cotisations subventionnables entre 2010 et 2015 [57]

Sur la Figure 2.1 précédente, nous constatons dans un premier temps une augmentation du nombre de contrats jusqu'en 2013. En parallèle, la surface et le capital assurés se sont soudainement accrus en 2011 avant d'atteindre leurs niveaux de 2013. Pour expliquer ce développement hétérogène de l'assurance multirisque climatique des récoltes, les auteurs proposent deux théories. La première serait que les agriculteurs assurés depuis 2010 ont revu à la hausse leur niveau de couverture, que ce soit en augmentant leurs surfaces ou capitaux assurés, ou bien en souscrivant à des garanties supplémentaires (températures basses, manque de rayonnement...). Pour

la seconde, ils exploitent les résultats de leur étude par suivi de culture. Celle-ci met en évidence une forte croissance du nombre de contrats actifs pour les vignobles. Or, cette activité se caractérisant par un capital assuré et un risque climatique élevés pour des petites surfaces, cela justifierait les tendances de la Figure 2.1.

A partir de 2014, une baisse des indicateurs se dessine, tandis qu'ils chutent considérablement en 2015. D'après les auteurs, ce phénomène pourrait s'expliquer par un dépassement des fonds de subvention prévue dans la PAC durant l'année 2013, qui a conduit à une diminution des taux de prise en charge des primes d'assurance pour certaines catégories de culture. De plus, le budget a encore été dépassé en 2014. La survenance de tous ces événements aurait alors pu conduire à une perte de confiance des agriculteurs sur la robustesse de ce dispositif.

Finalement, sur la période 2010-2015, on relève un taux de souscription moyen de 27,4% sur l'ensemble du territoire, avec notamment des écarts selon le type de culture allant jusqu'à 30%, comme le présente la Figure 2.2. Dans une telle situation, le principe de mutualisation peine à fonctionner, ce qui ne rend pas suffisamment attractif le produit d'assurance.

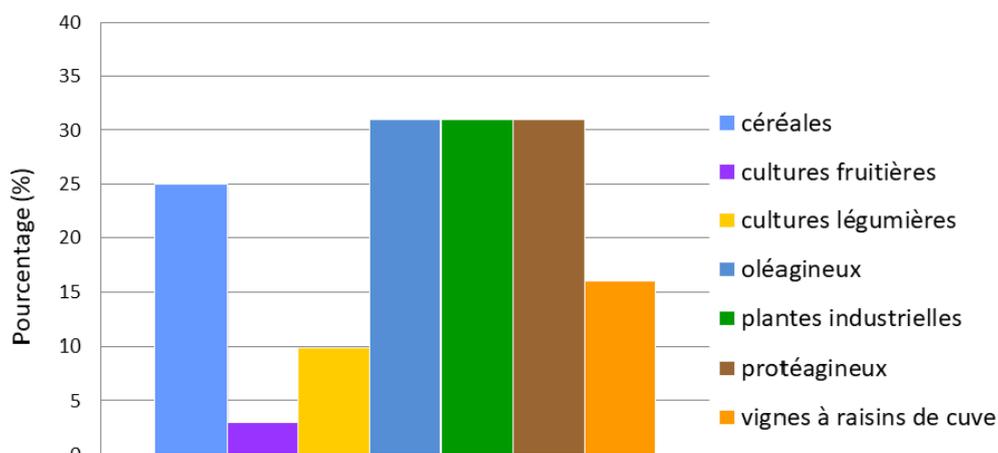


FIGURE 2.2 – Taux de diffusion par catégorie de culture - Moyenne 2010-2015 [57]

2.1.2.2 Incitations à la souscription

Dans l'optique d'inciter les agriculteurs à souscrire à une assurance pour protéger leurs récoltes, le Fonds Européen Agricole pour le Développement Rural (FEADER) a été mobilisé pour subventionner une partie de leurs primes d'assurance. En effet, le contrat multirisque climatique instauré en 2016 comporte trois niveaux de subvention [41] :

- un **niveau « socle »**, pour lequel le taux de subvention maximum est de 65%. Ce niveau permet de couvrir un capital dans la limite du barème prévu par le socle, en indemnisant essentiellement les pertes de quantité. Cette garantie se déclenche au seuil de 30%, tandis qu'elle prévoit une franchise minimale de 30% pour les contrats à la culture ou 20% pour les contrats à l'exploitation.
- des **garanties complémentaires optionnelles**, subventionnées à hauteur de 45% maximum. Elles permettent de couvrir le capital au-delà du barème défini par le socle, tout en indemnisant également les pertes de qualité. Là encore, la garantie se déclenche au seuil de 30% avec une franchise minimale de 25% pour les contrats à la culture.
- des **garanties optionnelles**, qui ne sont pas du tout subventionnées. Elles permettent d'ajouter des garanties supplémentaires au contrat d'assurance, telles que le rachat de rendement par exemple, et peuvent se déclencher en deçà du seuil de 30%. Les montants de franchises prévus sont inférieurs à 25% pour les contrats à la culture et 20% pour les contrats à l'exploitation.

Par ailleurs, l'Etat mène également des campagnes d'information sur l'assurance multirisque climatique des récoltes afin de sensibiliser les agriculteurs aux risques climatiques et les accompagner dans leur demande de subvention de primes par le FEADER. Ce fut par exemple le cas en 2021, lorsque le Ministère de l'agriculture et de l'alimentation publia un document expliquant l'intérêt de l'assurance multirisque climatique pour plusieurs types d'exploitants agricoles [41].

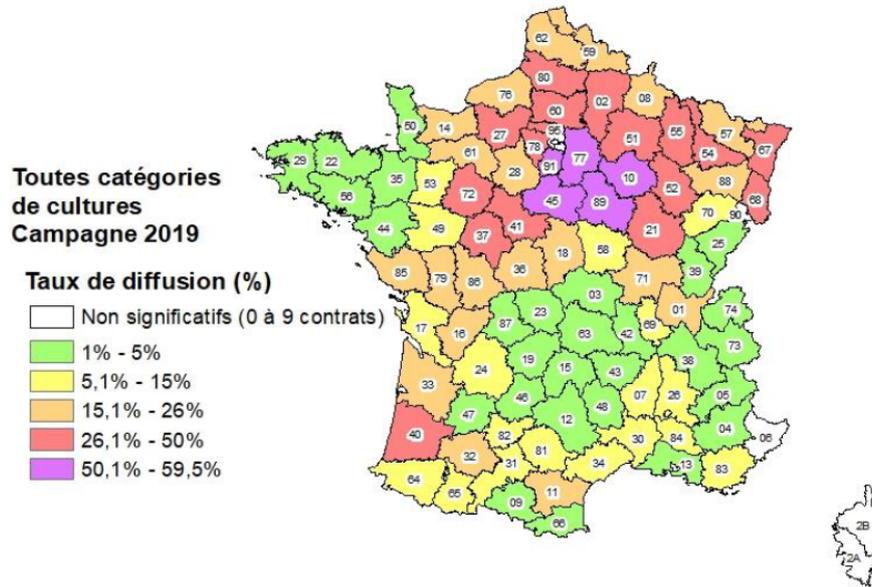


FIGURE 2.3 – Taux de diffusion de l'assurance multirisque climatique - Campagne de 2019 [8]

Malgré les différentes campagnes d'informations, l'assurance multirisque climatique peine à se développer, comme en témoigne la carte précédente. Dans la plupart des départements métropolitains, moins d'un quart des exploitations ont souscrit à ce type de contrat.

2.2 Conséquences du risque climatique sur l'agriculture

En France, le coût des sinistres liés à des événements climatiques a significativement augmenté depuis 2015 [12]. Cet accroissement de la sinistralité a deux conséquences pour les agriculteurs : non seulement ces derniers subissent des pertes de rendements dans leurs exploitations, mais ils doivent également payer davantage de primes pour être couverts par leur contrat d'assurance MRC.

2.2.1 Impact sur les rendements agricoles

Dans l'agriculture, le changement climatique affecte de manière conséquente les rendements des exploitations [47]. Les hausses de température et les canicules favorisent notamment la sécheresse des terres agricoles, ce qui détériore alors leur fertilité donc la qualité des cultures et leur prix de vente. L'augmentation de la fréquence et de l'intensité des épisodes de gel a également des conséquences dévastatrices sur les cultures, comme en France au printemps 2021 où les viticulteurs ont déploré une perte de production estimée à près de 30%. Ces gelées ont également détruit une partie des récoltes de pêches et d'abricots, faisant alors de l'année 2021 la saison pour laquelle la production est la plus faible depuis 46 ans.

A l'échelle mondiale, les rendements de certaines cultures, telles que le blé, la betterave et le maïs, pourraient chuter de 50%, voire 80% en 2100 si aucune mesure n'est mise en place [47]. A travers ces pertes de rentabilité, c'est directement l'économie de certains pays particulièrement agricoles qui pourrait s'effondrer.

2.2.2 Augmentation des primes d'assurance

Selon une étude de France Assureurs [21], en 2020 l'assurance MRC et grêle représentait près de 27% du total des cotisations agricoles, soit environ 544 millions d'euros. De plus, leur enquête auprès des divers assureurs agricoles révèle que le montant des primes encaissées depuis 2016 est en forte croissance. Si ce résultat peut être la conséquence des nombreuses campagnes d'incitations gouvernementales, il peut également s'interpréter par le fait que les assureurs se couvrent davantage contre les risques agricoles climatiques en diminuant leur rapport sinistres sur primes. En effet, d'après la Table 2.1, ce dernier s'élève à 159% pour l'année 2016, ce qui est particulièrement risqué pour un assureur.

	Marché national		
	Cotisations (en M€)	Sinistres (en M€)	Rapport S/P (en %)
2016	440	700	159
2017	468	411	88
2018	501	434	87
2019	520	517	99
2020	544	574	105
Δ 2020/2019	+4,6 %	+11,0 %	+6 pts

TABLE 2.1 – Evolution des primes, des sinistres et du rapport sinistres/primes pour les contrats MRC et grêle [21]

Cette hausse des primes a d'ailleurs été confirmée par les principaux assureurs du secteur. Selon les types de culture couverts, les cotisations auraient augmenté de 10 à 25% entre 2021 et 2022 [33].

2.3 Réforme de l'assurance MRC du 1^{er} janvier 2023

2.3.1 Détails de la réforme

Le 1^{er} janvier 2023, la réforme de l'assurance MRC est entrée en vigueur. Celle-ci vise à accroître le taux de souscription pour l'assurance récoltes. Les objectifs pour 2030 sont présentés dans la Table 2.2.

Type de culture	Nombre de cultures assurées en 2023 (Part des cultures assurables en 2023)	Nombre de cultures assurées en 2030 (cible) (Part des cultures assurables en 2030)	Évolution 2023-2030 de la part des cultures assurées
Grandes cultures	45 000 (30 %)	75 000 – 80 000 (60 %)	× 2
Viticulture	10 000 (30 %)	20 000 (60 %)	× 2
Prairies	1 700 (1,2 %)	26 700 – 31 700 (30 %)	× 25
Arboriculture	600 (3 %)	3 600 (30 %)	× 10

TABLE 2.2 – Objectifs de la réforme de l'assurance récoltes en taux de diffusion [46]

La réforme prévoit la mise en place d'un nouveau régime d'indemnisation à trois niveaux [49] :

1. Si les pertes de production sont inférieures à 20%, l'agriculteur ne sera pas couvert.
2. Pour couvrir des pertes plus élevées, l'exploitant peut choisir de souscrire à un contrat d'assurance avec une franchise de 20%, subventionné par la PAC. Selon le niveau d'offres du marché assurantiel, relatif au type de cultures, les pertes peuvent être indemnisées jusqu'à 30% voire 50%.
3. Au-delà de ce seuil, l'Etat agira, par l'intermédiaire du fonds de solidarité nationale, et ce pour toutes les cultures, assurées ou non. Ce fonds remplacera le fonds des calamités agricoles. Il sera financé par une contribution de l'Etat, qui atteindra les 680 millions par an d'ici à 2025, ainsi qu'un doublement du plafond de la taxe payée par les agriculteurs au FNGRA, qui passe de 5,5% à 11%.

Enfin, la réforme supprime les exclusions existantes pour bénéficier du fonds des calamités agricoles. De plus, elle permet une meilleure indemnisation des agriculteurs ayant souscrit à un contrat d'assurance.

2.3.2 Conséquences pour les assureurs

Déjà, puisqu'il s'agit de l'objectif principal de la réforme, les assureurs verront leurs portefeuilles d'assurés s'élargir. Par ailleurs, la loi prévoit la création d'un *pool* d'assureurs à adhésion obligatoire [46]. Ce dernier vise à la mutualisation des risques et au partage de données, en vue d'établir des contrats au tarif le plus adapté. Néanmoins, tous les assureurs n'ont pas le même intérêt à entrer dans ce *pool*, n'apportant pas le même niveau de connaissance du marché, ce qui retarde son instauration.

2.4 Problématiques de provisionnement soulevées par le climat

L'assurance récoltes est une branche à développement court, c'est-à-dire que les sinistres sont très rapidement indemnisés et que les déclarations tardives sont peu nombreuses. Bien que le dérèglement climatique amplifie les dommages sur les récoltes, les provisions restent encore facilement estimables avec les méthodes classiques.

Néanmoins, plusieurs axes de réflexion peuvent être envisagés afin d'évaluer l'impact climatique sur les réserves :

1. **Analyse de la distribution des réserves** de triangles issus de plusieurs sous-branches, en vue d'identifier un éventuel effet du risque climatique sur les paramètres d'asymétrie ou d'aplatissement.
2. **Lien entre provisions et variables météorologiques**, afin de corrélérer les réserves à des indicateurs climatiques et anticiper l'évolution de ces dernières et *a fortiori* de la sinistralité future à l'aide de projections calculées selon divers scénarios.

A l'aide du cadre théorique de la Partie II, les résultats de ces études seront présentés en Partie III, respectivement aux Chapitres 9 et 10.

Enfin, toujours en vue d'anticiper l'effet du climat sur les réserves futures, mais également sur la tarification, il peut être intéressant de se concentrer sur l'**étude des périodes de retour d'événements extrêmes**, tels que des pertes de rendements, et leur lien avec des variables climatiques.

Cette approche et ses résultats seront présentés dans la Partie IV.

Chapitre 3

Panorama des aléas climatiques en agriculture

Dans ce chapitre, nous dressons un panorama des divers risques auxquels sont exposées les cultures agricoles. Si certaines sont particulièrement menacées par le manque d'eau et la sécheresse, pour d'autres, les intempéries telles que la grêle ou le gel peuvent être beaucoup plus dévastatrices.

3.1 Sécheresse

La sécheresse est un phénomène complexe incluant diverses réalités. On dénombre quatre formes de sécheresse [22] :

- la **sécheresse météorologique**, correspondant à un déficit significatif et durable de précipitations ;
- la **sécheresse édaphique**, ou sécheresse agricole, se traduisant par un développement altéré de la végétation en raison d'un manque d'eau dans les sols ;
- la **sécheresse hydrologique**, pour laquelle le déficit de précipitation se répercute sur le niveau des nappes phréatiques et le débit des cours d'eau ;
- et la **sécheresse géologique**, qui se caractérise par une diminution de la teneur en eau du sous-sol.

3.1.1 Conséquences générales des sécheresses

Dans le monde, ces différents types de sécheresses induisent de forts dégâts, humains comme matériels [22]. En effet, entre 1970 et 2019, on dénombre 650 000 décès liés à ce phénomène. Par ailleurs, la sécheresse est considérée comme le danger le plus grave pour le bétail et les cultures, et ce dans presque toutes les régions du monde.

En France, la sécheresse géologique entretient le **phénomène de retrait-gonflement des argiles**. En raison des variations de la teneur en eau des sous-sols, certaines natures de sols telles que l'argile voient leur consistance s'altérer en fonction de leur humidité. De fait, un sol sec sera dur et cassant tandis qu'un sol humide sera davantage malléable. Lorsque la teneur en eau est très faible, le sol se rétracte alors drastiquement quand à l'inverse il se gonfle amplement lorsqu'il est bien plus humide. A travers l'amplification de ce phénomène naturel, la sécheresse peut ainsi engendrer des mouvements de terrain qui peuvent aller jusqu'à fragiliser les fondations des habitations, et en particulier les maisons individuelles.

Si le risque de retrait-gonflement majeur est l'un des enjeux principaux liés à la sécheresse pour les assureurs, ce péril recouvre d'autres conséquences loin d'être anodines.

Tout d'abord, la sécheresse peut amplifier les effets des feux de forêt, dans la mesure où leur intensité dépend entre autres du taux d'humidité de l'air et de l'état d'aridité de la végétation.

De plus, le manque d'eau peut conduire à une baisse de la production électrique. Par exemple, alors que 67% de la production d'électricité française repose sur le nucléaire, la sécheresse pourrait limiter le bon fonctionnement des réacteurs. La production des barrages se verrait également réduite.

Enfin, la sécheresse menace la biodiversité, notamment des milieux aquatiques. D'une part, la diminution des débits des cours d'eau altère la qualité de l'eau. D'autre part, la réduction de leurs niveaux d'eau peut conduire à empêcher la mobilité de certaines espèces voire provoquer leur disparition. Quant aux espèces non-aquatiques, celles-ci sont également victimes du manque d'eau.

3.1.2 Risque de sécheresse en France

3.1.2.1 2022, une année fortement marquée par les sécheresses

Pour quantifier les sécheresses, la Mission Risques Naturels utilise l'indicateur d'humidité des sols développé dans le cadre du projet ClimSec de Météo-France, le *Standardized Soil Wetness Index* (SWII) construit à partir du *Soil Wetness Index* (SWI). Utilisé par Météo-France pour suivre l'état hydrique des sols, il permet d'identifier les zones les plus touchées par des sécheresses extrêmes.

A partir de cet indice, il est possible de comprendre comment s'est matérialisé le phénomène de sécheresse en France en 2022. La Figure 3.1 retrace l'état des sols français au fil de ces mois.

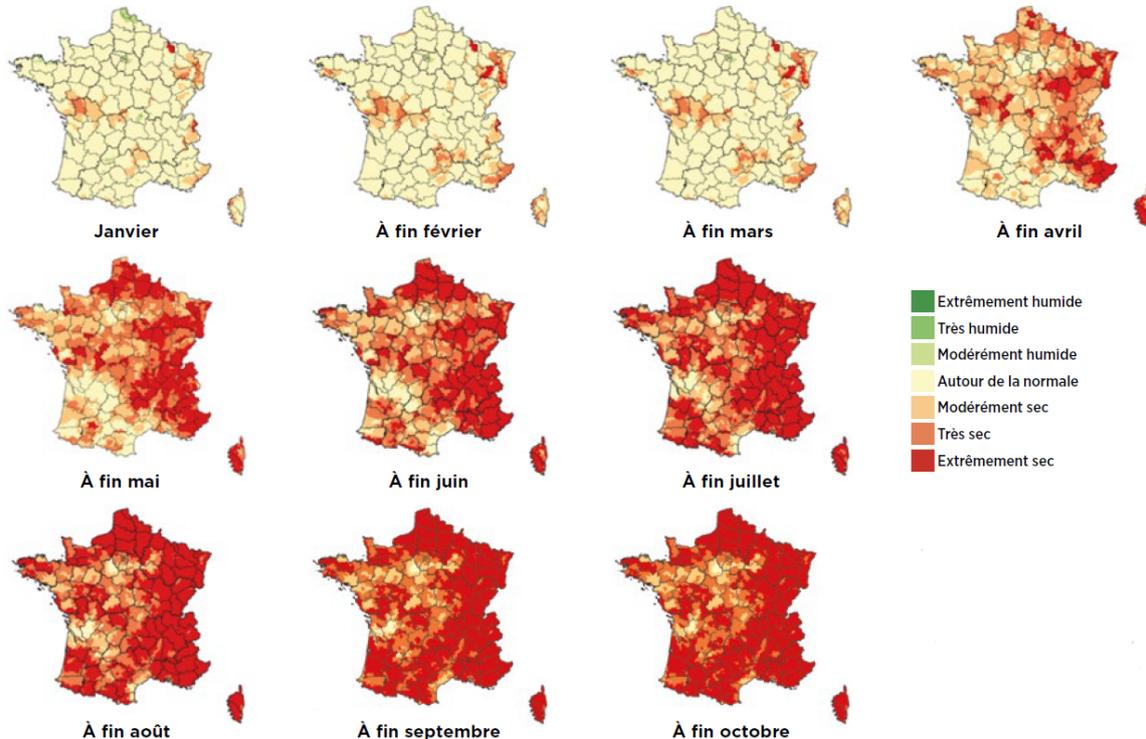


FIGURE 3.1 – Evolution de l'état des sols au fil des mois en 2022 [22]

En janvier 2022, le SWII est proche de la normale sur quasiment l'ensemble du territoire, à l'exception de certaines zones qui connaissent déjà une situation de stress hydrique. Durant les mois de mai et juillet, les plus fortes proportions de territoires en état de sécheresse extrême

sont atteintes : 30% en mai et 25% en juillet. Ainsi, l'année 2022 est marquée par une sécheresse printanière et estivale, qui aura touché presque tout le territoire à la fin octobre.

3.1.2.2 Projections météorologiques

Le livre blanc de Covéa, *Changement climatique et assurance : Quelles conséquences sur la sinistralité à horizon 2050 ?* [12] présente des projections météorologiques sur l'évolution des sécheresses en 2050. En particulier, les auteurs modélisent le climat futur à partir des projections du GIEC pour le scénario RCP 8.5, soit le scénario le plus pessimiste, et comparent ensuite les valeurs obtenues pour l'indicateur SWI aux valeurs actuelles. Les variations de l'indice d'humidité des sols sont cartographiées en Figure 3.2.

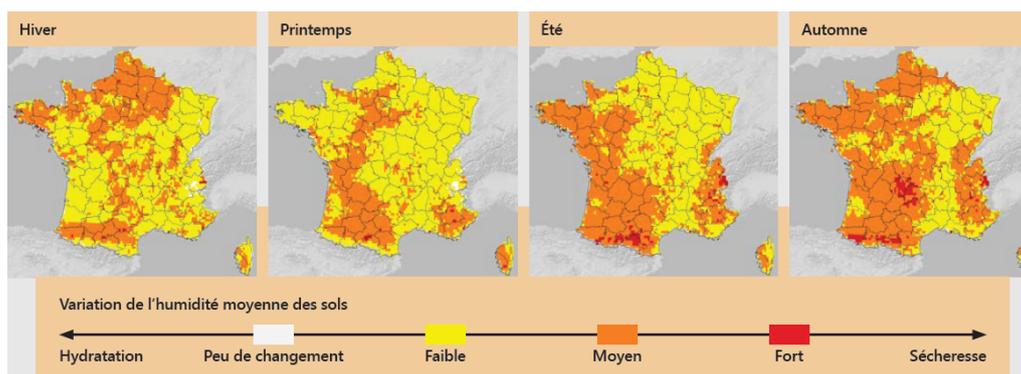


FIGURE 3.2 – Evolution de l'indicateur d'humidité des sols sous scénario RCP 8.5 sur la période 2035-2065 par rapport à la référence 1970-2005 [12]

De leurs recherches, il découle plusieurs constats :

- Quelle que soit la saison considérée, les sécheresses vont s'accroître sur l'ensemble du territoire.
- Des disparités géographiques sont observées : certains territoires seront plus ou moins touchés selon la saison.

En particulier, les zones qui ont actuellement des sols très humides telles que la Bretagne ou le Massif Central connaîtront une forte aggravation des sécheresses.

En outre, Météo France [40] présente quelques résultats sur le futur du cycle de l'eau en 2050, avec un scénario modéré d'émissions de gaz à effet de serre :

- Les températures en France pourront augmenter de 1°C à 1,5°C d'ici 2050.
- Les précipitations diminueront en été et amplifieront le phénomène d'évaporation dû à la chaleur.
- Le manteau neigeux pourra contenir jusqu'à deux fois moins d'eau dans les Alpes et fondra de manière plus précoce.
- Les sols vont s'assécher davantage : dans toutes les régions, le nombre de jours de sol sec augmentera de 10 à 25.
- Si les sécheresses sont déjà deux fois plus nombreuses qu'en 1960, elles vont encore se multiplier.

3.1.3 Impacts sur les cultures

De nombreuses cultures sont menacées par la sécheresse. Lorsque celle-ci survient au printemps, période de développement de nombreuses cultures céréalières d'hiver telles que le

blé ou l'orge, les rendements peuvent être très affectés [32]. A partir de la mi-mai, le blé entre en épiaison et les grains grossissent, comme en témoigne la Figure 3.3. Ainsi, si la plante manque d'eau, les grains ne pourront pas se développer correctement.

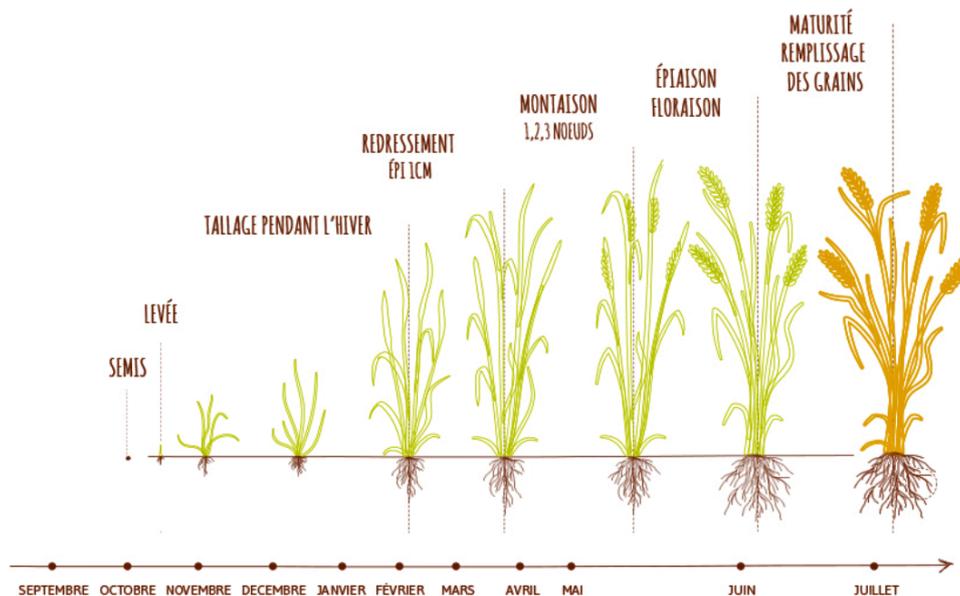


FIGURE 3.3 – Cycle de vie du blé [59]

Si le manque d'eau perdure, les cultures de printemps telles que les betteraves, le tournesol et le maïs peuvent également être impactées [32]. Par ailleurs, les restrictions d'eau imposées à cause des sécheresses peuvent amplifier la détérioration de la qualité des cultures. En effet, alors que 80% de la consommation d'eau en été en France sert à l'agriculture, certaines cultures de printemps, plus sensibles au manque d'eau, se verraient davantage affectées par un déficit d'arrosage [22].

3.2 Grêle

Selon Météo France [38], le phénomène de grêle est caractérisé par la chute de particules de glace de diamètre supérieur à 5 mm appelées grêlons. Généralement, la grêle survient lors d'orages violents au sein de cumulonimbus. Moins de 10% de ces nuages conduisent à l'arrivée de grêle au sol.

3.2.1 Risque de grêle en France

3.2.1.1 Saisonnalité de la grêle

Depuis 2006, Keraunos, l'observatoire français des tornades et orages violents, recense les chutes de grêles [30]. Leurs travaux ont permis d'établir une saisonnalité de la grêle dans le pays, représentée sur la Figure 3.4 ci-après :

- **De janvier à mars**, la plupart des chutes de grêle sont observées le long des côtes, plus rarement à l'intérieur du pays. Généralement, elles prennent la forme d'averses brèves, qui peuvent parfois être abondantes. En revanche, le diamètre des grêlons excède rarement plus d'1,5 cm.
- **D'avril à début mai**, les chutes de grêles sont de plus en plus intenses, avec des grêlons qui peuvent dépasser les 3 cm de diamètre. Les chutes sont de plus en plus fréquentes à l'intérieur des terres et se raréfient sur les côtes.

- C'est **entre mi-mai et fin juillet** que l'activité de grêle atteint son paroxysme. Les averses peuvent être violentes à l'intérieur des terres. Leur déclenchement est dû à des dégradations orageuses en provenance du sud-ouest.
- **D'août à septembre**, l'activité grêle décline progressivement. Les fortes chutes de grêle surviennent dans la moitié sud du pays, et en particulier dans les régions méditerranéennes.
- **En octobre**, les chutes de grêles perdent en intensité et sont moins fréquentes. Elles se concentrent en particulier autour de la Méditerranée. Il devient exceptionnel d'observer des grêlons de diamètre supérieur à 5 cm.
- **De novembre à décembre**, les chutes de grêle sont de moins en moins fréquentes et concernent presque exclusivement les côtes.

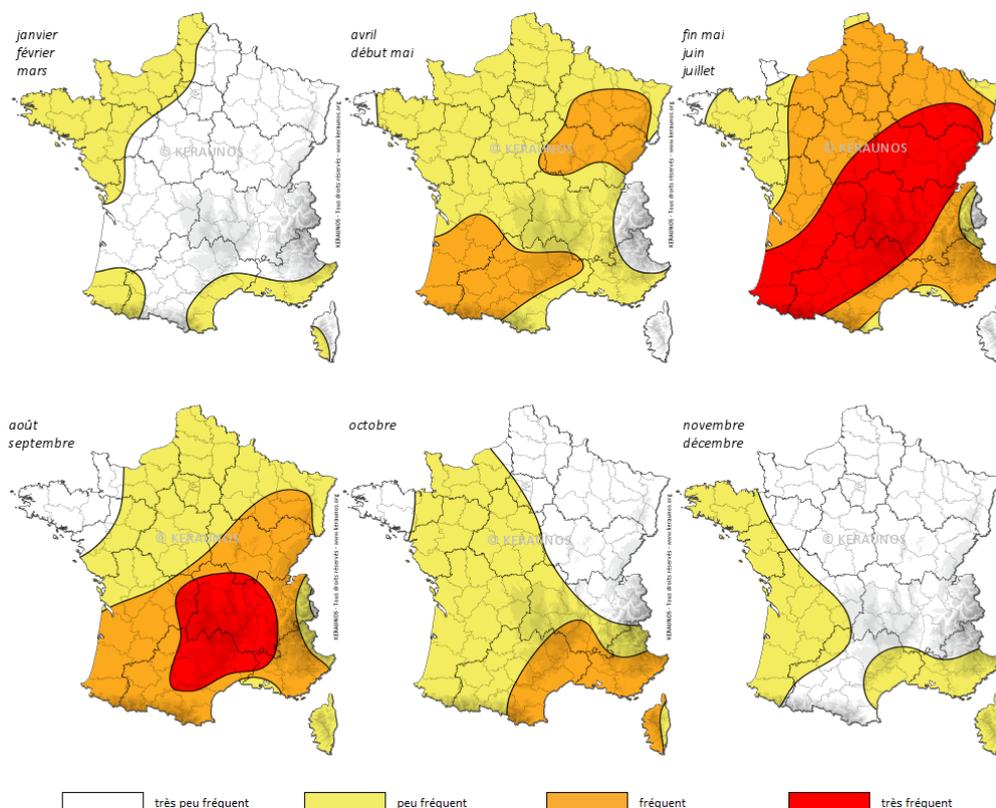


FIGURE 3.4 – Fréquence des chutes de grêle en France [30]

3.2.1.2 Projections météorologiques

Tout d'abord, suite au réchauffement climatique, la taille des grêlons serait amenée à grandir, leur vitesse de croissance dépendant du taux d'humidité de l'air [26]. En effet, des physiciens à l'*Environment and Climate Change Canada* ont démontré qu'un air plus chaud contient davantage de vapeur d'eau. Par ailleurs, la hausse des températures conduisant à une plus forte évaporation, les petits grêlons auront fondu avant d'arriver au sol. Cela n'étant pas le cas pour les plus gros grêlons, l'intensité des tempêtes serait donc plus importante dans les années à venir.

Concernant la fréquence des orages de grêle, le livre blanc de Covéa [12] présente des projections à horizon 2050.

Il ressort de la Figure 3.5 suivante que la fréquence des orages devrait sensiblement augmenter sur toute la France. Selon les modèles, cette augmentation grêligène est de l'ordre de

40% à l'échelle du territoire. En revanche, la localisation des orages de grêle ne devrait pas être grandement modifiée dans les années à venir.

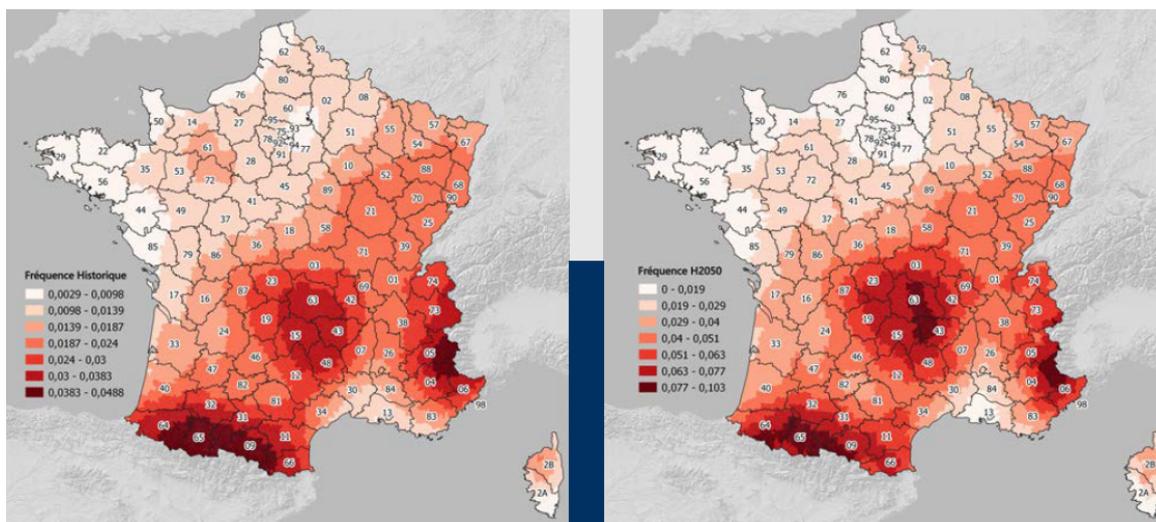


FIGURE 3.5 – Evolution du nombre d'occurrences annuel d'orages de grêle sur la période historique et à horizon 2050 [12]

3.2.2 Impacts sur les cultures

La grêle a des conséquences dévastatrices sur de nombreuses cultures, en particulier lorsque ces dernières sont proches de la récolte [1]. Au vu de la saisonnalité de la grêle, les cultures arrivant à maturité à la fin du printemps, telles que les céréales ou le colza, sont particulièrement menacées. Pour les céréales, les grêlons peuvent venir couper et égréner les épis, mais également plier des tiges, ce qui pourrait diminuer le remplissage des épis en grains. De la même manière, les plants de colza souffrent de la grêle à travers l'endommagement de leurs tiges.

La grêle détruit également d'autres types de cultures, telles que les vignes ou les arbres fruitiers [27]. Sur la vigne, les grêlons peuvent tout d'abord endommager les grappes, mais également conduire à une défoliation complète du pied. La souche du cep peut elle aussi être touchée, ce qui peut donc affecter les récoltes des années futures. Par ailleurs, la fragilisation des plants due à la grêle peut favoriser l'apparition de champignons comme le mildiou qui viennent détériorer le raisin. En arboriculture, les grêlons peuvent engendrer la chute ou des dommages sur les fruits, qui peuvent aller de la simple dégradation affectant leur prix de vente à la destruction complète, par fendaison ou perçage par exemple. La structure de l'arbre peut elle aussi être endommagée, conduisant à une baisse de rendements sur les années futures. Là encore, les orages de grêle peuvent favoriser l'apparition et la prolifération de champignons et maladies néfastes pour les arbres et les fruits.

3.3 Gel

Le phénomène de gel est caractérisé par une température de l'air au niveau du sol inférieure à 0°C. Il existe plusieurs types de gel [23] :

- Lorsqu'il y a une forte humidité dans l'air ambiant et un froid intense, des **gelées blanches** apparaissent. La surface des plantations est alors recouverte de givre.
- A l'inverse, les **gelées noires** surviennent dès lors que l'air est sec.

- Le **gel d'advection** se forme lorsque de grandes et épaisses masses d'air froid apparaissent brusquement.
- Suite à un transfert de chaleur du sol vers l'atmosphère, le **gel par radiation** peut survenir. Cependant, il ne se produit que dans des conditions météorologiques très spécifiques, plutôt satisfaites dans les fonds de vallées.
- Lors du processus d'évaporation, la température des plantes diminue drastiquement. En général, le **gel par évaporation** se produit lors de l'évaporation de l'eau à l'aube. Selon la période à laquelle les gelées surviennent, elles sont qualifiées d'« automnales », d'« hivernales » ou de « printemps ».

3.3.1 Risque de gel en France

3.3.1.1 Localisation des gelées

Météo France établit une carte du nombre moyen annuel de jours de gel, définis comme les jours où la température descend en-dessous de 0°C. Celle-ci est présentée en Figure 3.6.

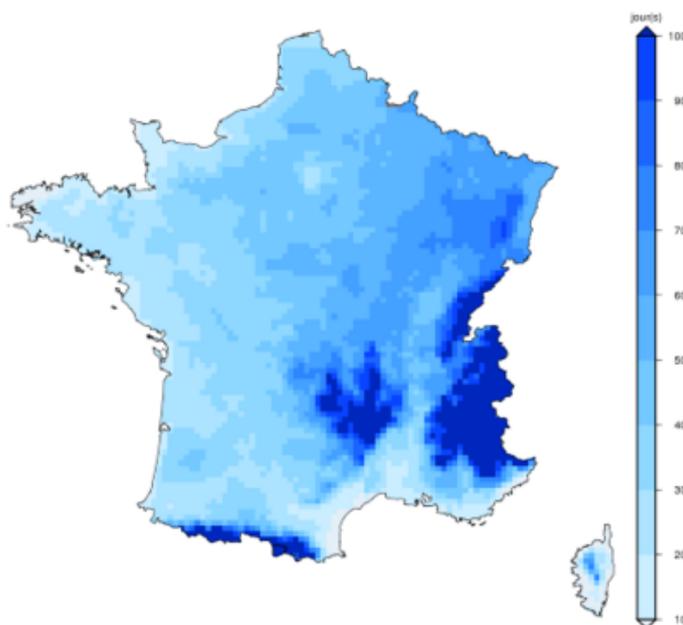


FIGURE 3.6 – Nombre de jours de grêle annuel moyen sur la période 1976-2005 [39]

Les zones montagneuses, telles que les Alpes, les Pyrénées et le Massif central dénombrent le plus de jours de gel. Le nord-est de la France est également très exposé. Enfin, les zones côtières, pour lesquelles les températures sont plus douces en hiver, expérimentent peu de gelées.

3.3.1.2 Projections météorologiques

En France, le nombre de jours de gel serait amené à diminuer drastiquement d'ici la seconde moitié du XXI^{ème} siècle. Météo France fournit l'outil Climadiag Commune permettant de projeter des indicateurs climatiques à l'échelle communale à horizon 2050 [37]. La Figure 3.7 suivante récapitule les projections obtenues pour le nombre de jours de gel dans plusieurs villes françaises. La valeur de référence est calculée sur la période 1976-2005. Pour les villes présentées, le scénario médian 2050 conduit à un déclin de l'ordre de 35% du nombre de jours de gel.

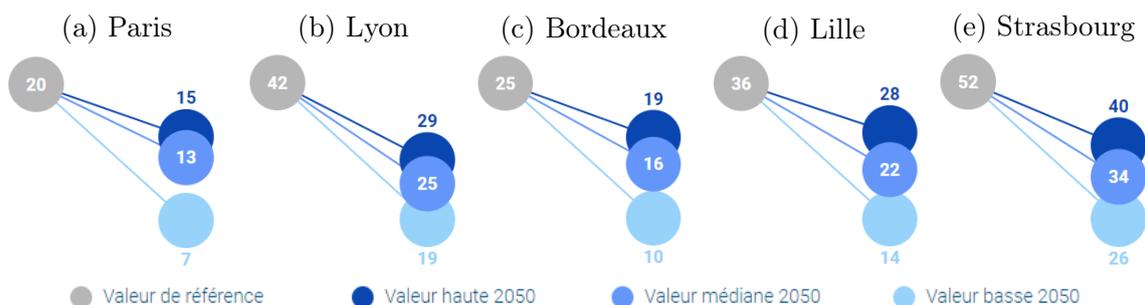


FIGURE 3.7 – Projections du nombre de jours de gel à horizon 2050 [37]

Sur un horizon temporel plus lointain, Météo France a projeté le nombre de jours de gel pour la période 2071-2100 à partir du scénario RCP 4.5 du GIEC [39]. La carte résultante est affichée sur la Figure 3.8.

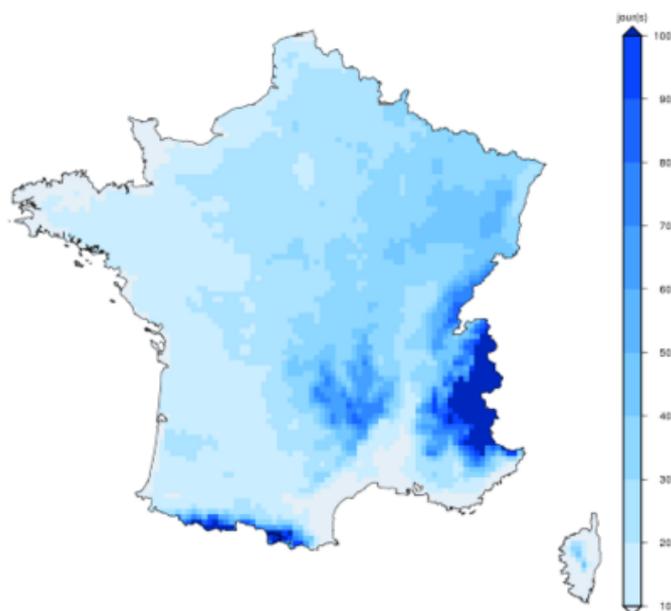


FIGURE 3.8 – Prévisions de nombre de jours de grêle à l'horizon 2071-2100 - Scénario RCP 4.5 [39]

Il ressort de la carte précédente que les zones pour lesquelles surviennent le plus de jours de grêle restent les zones montagneuses. Toutefois, le nombre de jours moyen a sensiblement diminué, par rapport à la période de référence. En plaine, cette diminution serait de l'ordre de 21 jours avec le scénario RCP 4.5, contre 33 avec le scénario RCP 8.5.

3.3.2 Impacts sur les cultures

Les différents types de gels ne sont pas tous aussi dommageables pour les cultures [23]. En raison du givre qu'elles engendrent, les gelées blanches permettent de protéger la plante, agissant comme une protection contre le froid. Les gelées noires sont quant à elles plus destructrices, le froid s'infiltrant directement dans les cultures. En outre, il est très difficile pour les agriculteurs de lutter contre le risque de gel d'advection, les masses d'air froid se mouvant en permanence sur de grandes étendues.

L'âge des cultures est un premier facteur qui impacte leur résistance au froid : les plants les plus jeunes seront les plus fragiles [48]. Ainsi, dans les zones exposées au risque de gel, les dates

de semis et les choix d'espèces doivent être anticipés pour permettre la meilleure résistance des cultures.

Toutes les espèces ne souffrent pas pareillement du froid [48]. Dans les cultures céréalières, le seigle et le triticale figurent parmi les cultures les plus résistantes au gel tandis que l'avoine et le blé dur sont davantage fragilisés par des vagues de froid. Pour les espèces protéagineuses, les pois et féveroles ont eux aussi une faible résistance au froid. Concernant les vignes, celles-ci sont exposées au gel de l'automne au printemps [27], bien que les gelées précoces soient surtout une menace pour les jeunes ceps. En hiver, le gel peut causer de sérieux dommages sur les bourgeons et la souche, allant parfois jusqu'à provoquer la mort des ceps.

3.4 Autres aléas climatiques

En plus de la sécheresse, de la grêle et du gel qui peuvent ravager les cultures, les agriculteurs peuvent faire face à d'autres aléas climatiques. Si le manque d'eau menace les récoltes, ces dernières peuvent également subir de plein fouet les conséquences d'un sur-arrosage. En effet, en 2016, les récoltes de blé se sont particulièrement amoindries en raison de pluies excessives. Combinées au froid et au manque de lumière, ces dernières ont alors engendré la prolifération de nuisibles et de maladies sur les épis. L'année 2016 a ainsi battu le record de la pire moisson de blé en France depuis 1976 [17].

De manière générale, les conséquences du dérèglement climatique risquent d'être coûteuses pour les agriculteurs et à plus forte raison leurs assureurs. En effet, le GIEC prévoit un accroissement de l'intensité des fortes pluies, des inondations et du risque de crue en fonction du niveau de réchauffement planétaire [25].

Deuxième partie

Provisionnement : Cadre théorique

Chapitre 4

Méthodes usuelles de provisionnement

Dans ce chapitre, nous présentons des méthodes de provisionnement usuelles, tels que Mack *chain-ladder*, Wüthrich, Merz et Lysenko, les approches par modèle généralisé ainsi que les méthodes simulatoires par bootstrap. Avant cela, nous expliquons le concept de provisionnement et nous introduisons les notations pertinentes.

Dans ce chapitre et les suivants, les modèles développés présenteront les mêmes spécifications et hypothèses que celles fournies par les auteurs dans leurs articles. La discussion de la validité des hypothèses vis-à-vis de triangles climatiques sera éludée, au profit d'une vérification de ces dernières au cours de la Partie III.

Introduction au provisionnement

Dans le monde de l'assurance, les assurés payent des primes afin d'obtenir une indemnisation en cas de sinistre. Lorsqu'un sinistre survient, quatre étapes se succèdent, schématisées sur la Figure 4.1 :

1. **Survenance** du sinistre durant la période de couverture du contrat d'assurance.
2. **Déclaration** du sinistre à l'assureur.
3. **Indemnisation** du sinistre par l'assureur, en un ou plusieurs versements.
4. **Clôture** du sinistre lorsque l'assureur estime que plus aucun paiement ne sera effectué.



FIGURE 4.1 – Etapes de la vie d'un sinistre

Ainsi, il existe un laps de temps entre le moment de survenance du sinistre, sa date de déclaration par l'assuré et la date à laquelle l'assureur effectuera le versement des indemnités. Par conséquent, les montants des sinistres non encore déclarés, communément appelés **IBNR**, *Incurred But Not Reported*, doivent être anticipés dans les réserves de la compagnie. Il existe deux sous-catégories d'IBNR :

- les **IBNeR**, *Incurred But Not enough Reported*, c'est-à-dire les sinistres pour lesquels les prestations à payer peuvent évoluer à la hausse ou à la baisse ;

— les **IBNyR**, *Incurred But Not yet Reported*, c'est-à-dire les sinistres qui n'ont pas encore été déclarés à l'assureur.

Les problématiques de provisionnement se posent à l'aide de triangles qui renseignent sur les paiements en fonction de l'année de survenance des sinistres et de l'année de développement de ces derniers, comme présenté sur la Table 4.1. Ces triangles sont communément appelés **triangles de provisionnement** ou **triangles *run-off***. Il est possible d'établir des triangles de deux manières : en y notant les incréments de paiements par année de développement, ou bien en les cumulant au fil des années. Dès lors, le but de l'actuaire est de compléter la partie inférieure du triangle grâce à des méthodes de provisionnement.

Année de survenance i	Année de développement j						
	1	...	$j-1$	j	$j+1$...	J
1	$C_{1,1}$						$C_{1,J}$
⋮							
$i-1$							
i				$C_{i,j}$			
$i+1$							
⋮							
I	$C_{I,1}$						

TABLE 4.1 – Exemple de triangle *run-off* en paiements cumulés

Notation 4.0.1. Les coefficients $C_{i,j}$ correspondent aux **paiements cumulés** versés dans l'année j pour les sinistres survenus dans l'année i .

Notation 4.0.2. Dans le cas d'un triangle écrit en incréments de paiements, on notera $X_{i,j}$ le **paiement incrémental** de l'année de développement j relatif aux sinistres survenus en année i . Il en ressort alors que

$$X_{i,j} = C_{i,j} - C_{i,j-1} \text{ d'où } C_{i,j} = \sum_{k=1}^j X_{i,k}.$$

Notation 4.0.3. Soit \mathcal{D}_I l'**information** sur les sinistres disponible au temps I . Alors

$$\mathcal{D}_I = \{C_{i,j}, 1 \leq i \leq I, 1 \leq j \leq J, i + j \leq I + 1\}.$$

Compte tenu des informations disponibles, l'objectif sera de calculer les réserves à faire en année J pour chacune des années de survenance i , via la formule suivante :

$$R_i = C_{i,J} - C_{i,J+1-i} \text{ pour } 1 \leq i \leq I.$$

Pour estimer les réserves, il sera donc nécessaire de compléter le triangle. La méthode de provisionnement la plus couramment utilisée est la méthode *chain-ladder*, complétée par Mack en 1993 [36] qui ajoutera une estimation de l'erreur de mesure des réserves.

4.1 Modèle de Mack

Dans son article *Distribution-free calculation of the standard error of chain ladder reserve estimates* [36], Mack propose d'estimer le montant des réserves en caractérisant la distribution des paiements cumulés $C_{i,j}$ par ses deux premiers moments, ainsi que l'incertitude liée à cette prédiction. Pour cela, il utilise la méthode *chain-ladder* sur un triangle *run-off* tel que celui de la Table 4.1. Cette dernière se base sur les cadences de paiements : autrement dit, elle requiert une certaine régularité entre les paiements selon l'année de développement. La méthode *chain-ladder* nécessite deux hypothèses principales.

Hypothèse 1 *Les paiements cumulés $C_{i,j}$ de différentes années de survenance sont **indépendants**, autrement dit :*

$$\{C_{i,1}, \dots, C_{i,I}\}, \{C_{k,1}, \dots, C_{k,I}\}, i \neq k \text{ sont indépendants.} \quad (H1)$$

Hypothèse 2 *Il existe des **facteurs de développement** $f_j > 0$, pour $1 \leq j \leq I - 1$ tels que*

$$\mathbb{E}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = f_j C_{i,j}, \text{ pour } 1 \leq i \leq I. \quad (H2)$$

Par ailleurs, Mack impose un cas de triangle de provisionnement bien précis dans lequel $I = J$.

4.1.1 Estimation du montant des réserves

A partir de là, il est possible de calculer les facteurs de développement \hat{f}_j :

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}} \text{ pour } 1 \leq j \leq I - 1. \quad (4.1)$$

En considérant cette formule, il est alors possible d'estimer les paiements cumulés à l'ultime, $\hat{C}_{i,I}$ par

$$\hat{C}_{i,I} = C_{i,I+1-i} \hat{f}_{I+1-i} \dots \hat{f}_{I-1} \text{ pour } 1 \leq i \leq I$$

De manière analogue, les réserves R_i de l'année de survenance i s'estiment par

$$\hat{R}_i = C_{i,I+1-i} (\hat{f}_{I+1-i} \dots \hat{f}_{I-1} - 1) \text{ pour } 1 \leq i \leq I.$$

Théorème 4.1.1. *Sous les hypothèses (H1) et (H2), il vient que*

$$\mathbb{E}[C_{i,I}|\mathcal{D}_I] = C_{i,I+1-i} f_{I+1-i} \dots f_{I-1} \text{ pour } 1 \leq i \leq I.$$

Ainsi, ce théorème démontre que $\hat{C}_{i,I}$ est de la même forme que $\mathbb{E}[C_{i,I}|\mathcal{D}_I]$, autrement dit de la même forme que la meilleure prédiction de $C_{i,I}$ connaissant l'information \mathcal{D}_I .

Il est également possible de justifier l'estimation des paramètres f_j par les \hat{f}_j calculés précédemment.

Théorème 4.1.2. *Sous les hypothèses (H1) et (H2), les estimateurs \hat{f}_j sont non-biaisés et non-corrélés.*

De ce théorème découle donc le fait que l'estimateur $\hat{C}_{i,I}$ de $C_{i,I}$ est non-biaisé et non-corrélé. A plus forte raison, l'estimateur du montant des réserves $\hat{R}_i = \hat{C}_{i,I} - C_{i,I+1-i}$ est également un estimateur non-biaisé et non-corrélé de $R_i = C_{i,I} - C_{i,I+1-i}$.

4.1.2 Estimation des erreurs de prédiction du modèle

A présent, nous disposons d'une estimation du montant des réserves. Nous pouvons désormais nous intéresser aux erreurs générées par la méthode *chain-ladder*. Ici Mack s'intéresse à l'**erreur de provisionnement à l'ultime**, autrement dit l'écart entre les montants de paiements estimés en ultime année de développement et les derniers paiements réels constatés.

De manière générale, l'**erreur quadratique moyenne**, ou *mean square error* (MSE) d'un modèle est la différence moyenne entre la valeur prédite et la valeur théorique, élevée au carré.

Définition 4.1.3. *L'erreur quadratique moyenne associée à l'estimateur $\hat{C}_{i,I}$ de $C_{i,I}$ est définie par*

$$\text{MSE}(\hat{C}_{i,I}) = \mathbb{E} \left[\left(\hat{C}_{i,I} - C_{i,I} \right)^2 \middle| \mathcal{D}_I \right]$$

En développant le carré dans l'espérance, il vient que

$$\begin{aligned} \text{MSE}(\hat{C}_{i,I}) &= \mathbb{E} \left[(C_{i,I} - \mathbb{E}[C_{i,I} | \mathcal{D}_I])^2 \middle| \mathcal{D}_I \right] + 2\mathbb{E} \left[(C_{i,I} - \mathbb{E}[C_{i,I} | \mathcal{D}_I]) \left(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I} \right) \middle| \mathcal{D}_I \right] \\ &\quad + \mathbb{E} \left[\left(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I} \right)^2 \middle| \mathcal{D}_I \right]. \end{aligned}$$

Puisque $\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I}$ est \mathcal{D}_I -mesurable, on obtient finalement

$$\text{MSE}(\hat{C}_{i,I}) = \text{Var}(C_{i,I} | \mathcal{D}_I) + \left(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I} \right)^2.$$

Ainsi, l'erreur de prédiction du modèle se décompose en :

- une **erreur d'estimation** (*variance error*), quantifiée par $\text{Var}(C_{i,I} | \mathcal{D}_I)$, liée à l'incertitude sur les paramètres ;
- et une **erreur de modèle** (*process error*), quantifiée par $(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^2$, relative à l'aléa sur les trajectoires de paiements futurs.

Remarque 4.1.4. *On peut également écrire que :*

$$\begin{aligned} \text{MSE}(\hat{R}_i) &= \mathbb{E} \left[\left(\hat{R}_i - R_i \right)^2 \middle| \mathcal{D}_I \right] \\ &= \mathbb{E} \left[\left((\hat{C}_{i,I} - C_{i,I-i+1}) - (C_{i,I} - C_{i,I-i+1}) \right)^2 \middle| \mathcal{D}_I \right] \\ &= \mathbb{E} \left[\left(\hat{C}_{i,I} - C_{i,I} \right)^2 \middle| \mathcal{D}_I \right] = \text{MSE}(\hat{C}_{i,I}). \end{aligned}$$

Par conséquent, l'erreur obtenue pour l'estimateur des réserves de l'année de survenance i correspond exactement à l'erreur faite sur le montant du paiement cumulé ultime des sinistres survenus en année i . *A fortiori*, celle-ci se décomposera également en un terme d'erreur de modèle et un terme d'erreur d'estimation.

Afin d'estimer l'erreur quadratique moyenne, Mack introduit une dernière hypothèse sur la variance des paiements cumulés.

Hypothèse 3 *Il existe des coefficients $\sigma_j > 0$ pour $1 \leq j \leq I - 1$ tels que :*

$$\text{Var}(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = \sigma_j^2 C_{i,j}, \quad \text{pour } 1 \leq i \leq I. \quad (\text{H3})$$

Par conséquent, il est nécessaire de trouver un estimateur de σ_j .

Théorème 4.1.5. *L'estimateur $\hat{\sigma}_j^2$ défini ci-après est un estimateur non-biaisé de σ_j^2 .*

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \text{ pour } 1 \leq j \leq I-2. \quad (4.2)$$

Remarque 4.1.6. *Pour déterminer $\hat{\sigma}_{I-1}^2$, trois possibilités s'offrent à nous :*

- Dans le cas où $f_{I-1} = 1$, on peut supposer que les sinistres ont fini de se développer d'où $\hat{\sigma}_{I-1}^2 = 0$.
- Dans le cas contraire, on peut extrapoler $\hat{\sigma}_{I-1}^2$ en régressant les valeurs des $\hat{\sigma}_j$.
- Dans le cas contraire, on peut également poser $\hat{\sigma}_{I-1}^2 = \min \left(\frac{\hat{\sigma}_{I-2}^4}{\hat{\sigma}_{I-3}^2}, \min(\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2) \right)$.

Désormais, nous pouvons déterminer l'erreur quadratique associée aux réserves.

Théorème 4.1.7. *Sous les hypothèses (H1), (H2) et (H3), l'erreur quadratique associée aux réserves peut être estimée par la formule suivante :*

$$\widehat{\text{MSE}}(\hat{R}_i) = \hat{C}_{i,I}^2 \sum_{j=I-i+1}^{I-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{I-j} C_{k,j}} \right). \quad (4.3)$$

Enfin, pour déterminer l'erreur quadratique moyenne associée aux réserves totales estimées $\hat{R} = \sum_{i=2}^I \hat{R}_i$, il n'est pas possible de simplement sommer les erreurs quadratiques moyennes obtenues pour chaque année, dans la mesure où ces dernières sont corrélées : il faudra également considérer un terme de covariance.

Corollaire 4.1.8. *Sous les hypothèses (H1), (H2) et (H3), l'erreur quadratique associée aux réserves totales peut être estimée par la formule suivante :*

$$\widehat{\text{MSE}}(\hat{R}) = \sum_{i=2}^I \left\{ \widehat{\text{MSE}}(\hat{R}_i) + \hat{C}_{i,I} \left(\sum_{k=i+1}^I \hat{C}_{k,I} \right) \sum_{j=I+1-i}^{I-1} \frac{2\hat{\sigma}_j^2/\hat{f}_j^2}{\sum_{k=1}^{I-j} C_{k,j}} \right\}. \quad (4.4)$$

L'erreur quadratique associée aux réserves totales peut elle aussi se décomposer en un terme d'erreur d'estimation et un terme d'erreur de modèle :

$$\begin{aligned} \text{MSE}(\hat{R}) &= \mathbb{E} \left[\left(\sum_{i=2}^I \hat{R}_i - \sum_{i=2}^I R_i \right)^2 \middle| \mathcal{D}_I \right] = \mathbb{E} \left[\left(\sum_{i=2}^I \hat{C}_{i,I} - \sum_{i=2}^I C_{i,I} \right)^2 \middle| \mathcal{D}_I \right] \\ &= \underbrace{\text{Var} \left(\sum_{i=2}^I C_{i,I} \middle| \mathcal{D}_I \right)}_{\text{erreur d'estimation}} + \underbrace{\left(\mathbb{E} \left(\sum_{i=2}^I C_{i,I} \middle| \mathcal{D}_I \right) - \sum_{i=2}^I \hat{C}_{i,I} \right)^2}_{\text{erreur de modèle}}. \end{aligned}$$

Notation 4.1.9. *Le terme d'erreur de modèle est parfois noté $\text{Var} \left(\sum_{i=2}^I \hat{C}_{i,I} \right)$.*

4.2 Modèle de Wüthrich, Merz et Lysenko

Dans *Uncertainty of the Claims Development Result in the Chain Ladder Method* [64], Wüthrich, Merz et Lysenko présentent une méthode pour évaluer les **erreurs de provisionnement à horizon un an**, c'est-à-dire entre la variabilité des montants réservés entre le début et la fin de l'année d'exercice. Cette approche est particulièrement intéressante dans le cadre de la norme Solvabilité II. En effet, la réglementation impose de quantifier non pas l'incertitude à horizon ultime, mais bien à horizon un an. Pour ce faire, les auteurs utilisent les triangles de provisionnement de la Table 4.2, calculés aux années I et $I + 1$.

Année de survenance i	Année de développement j							Année de survenance i	Année de développement j						
	0	...	$j-1$	j	$j+1$...	J		1	...	$j-1$	j	$j+1$...	J
0								0							
⋮								⋮							
$i-1$								$i-1$							
i								i							
$i+1$								$i+1$							
⋮								⋮							
I								I							

TABLE 4.2 – Triangles de provisionnement aux temps I (à gauche) et $I + 1$ (à droite) [64]

Remarque 4.2.1. *Comme nous pouvons le constater sur la Table 4.2, à la différence de Mack, les auteurs démarrent la numérotation des années à 0. Par conséquent, une certaine vigilance sera nécessaire lors de l'implémentation des formules, notamment dans un langage de programmation qui commence sa numérotation indicielle à 1. En particulier,*

$$\mathcal{D}_I = \{C_{i,j}, 0 \leq i \leq I, 0 \leq j \leq J, i + j \leq I\}.$$

Deux montants de réserves seront alors calculés pour chaque année de survenance i , $0 \leq i \leq I$:

- $R_i^I = C_{i,J} - C_{i,I-i}$, montant de réserves calculé avec l'information disponible en année I , qui correspond au cas de Mack (en décalant les indices).
- $R_i^{I+1} = C_{i,J} - C_{i,I-i+1}$, montant de réserves calculé avec l'information disponible en année $I + 1$.

Tout d'abord, les auteurs imposent deux hypothèses :

1. indépendance des $C_{i,j}$ pour différentes années de survenance $0 \leq i \leq I$;
2. existence de constantes $f_k > 0, \sigma_k > 0, 0 \leq k \leq J - 1$ telles que :

$$\forall 1 \leq j \leq J, \forall 0 \leq i \leq I, C_{i,j} = f_{j-1}C_{i,j-1} + \sigma_{j-1}\sqrt{C_{i,j-1}}\epsilon_{i,j}$$

où, en considérant $\mathcal{B}_0 = \{C_{0,0}, \dots, C_{I,0}\}$ avec $C_{i,0} > 0$, les variables aléatoires $\epsilon_{i,j}$ sont indépendantes et vérifient :

$$\forall 1 \leq j \leq J, \forall 0 \leq i \leq I, \mathbb{E}[\epsilon_{i,j}|\mathcal{B}_0] = 0, \mathbb{E}[\epsilon_{i,j}^2|\mathcal{B}_0] = 1 \text{ et } \mathbb{P}(C_{i,j} > 0|\mathcal{B}_0) = 1.$$

Remarque 4.2.2. Dès lors, il est possible de retrouver les hypothèses (H1) et (H3) données par Mack. Par ailleurs, les hypothèses fournies par Wüthrich, Merz et Lysenko impliquent que les coefficients $(C_{i,j})_{j \geq 0}$ sont des chaînes de Markov sous la probabilité \mathbb{P} . Les hypothèses de Mack se simplifient donc de la manière suivante :

$$\begin{aligned}\mathbb{E}[C_{i,j}|C_{i,j-1}] &= f_{j-1}C_{i,j-1} \\ \text{Var}(C_{i,j}|C_{i,j-1}) &= \sigma_{j-1}^2 C_{i,j-1}\end{aligned}$$

4.2.1 Estimation du montant des réserves

4.2.1.1 Estimation des réserves par *chain ladder*

Une fois ces hypothèses posées, il est alors possible de passer au calcul des facteurs de développement f_j . Là encore, une méthode type *chain-ladder* est appliquée :

— Au temps I , les facteurs de développement sont estimés par

$$\hat{f}_j^I = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{S_j^I} \text{ avec } S_j^I = \sum_{i=0}^{I-j-1} C_{i,j}.$$

— Au temps $I + 1$, les facteurs de développement sont estimés par

$$\hat{f}_j^{I+1} = \frac{\sum_{i=0}^{I-j} C_{i,j+1}}{S_j^{I+1}} \text{ avec } S_j^{I+1} = \sum_{i=0}^{I-j} C_{i,j}.$$

Remarque 4.2.3. L'estimateur \hat{f}_j^I est exactement l'estimateur \hat{f}_j de l'article de Mack. Par conséquent, pour $0 \leq j \leq J - 1$, les \hat{f}_j^I sont non-biaisés et non-corrélés. Le même raisonnement peut donc s'appliquer aux estimateurs \hat{f}_j^{I+1} .

A partir de ces estimateurs, les montants de paiements cumulés s'écrivent :

— sachant $C_{i,I-i}$,

$$\hat{C}_{i,j}^I = C_{i,I-i} \hat{f}_{I-i}^I \cdots \hat{f}_{j-1}^I$$

— sachant $C_{i,I-i+1}$,

$$\hat{C}_{i,j}^{I+1} = C_{i,I-i+1} \hat{f}_{I-i+1}^{I+1} \cdots \hat{f}_{j-1}^{I+1}$$

Ces deux montants sont donc respectivement des estimateurs de $\mathbb{E}[C_{i,j}|\mathcal{D}_I]$ pour $j \geq I - i$ et de $\mathbb{E}[C_{i,j}|\mathcal{D}_{I+1}]$ pour $j \geq I - i + 1$. De là, les montants des réserves estimés $\hat{R}_i^{\mathcal{D}_I}$ et $\hat{R}_i^{\mathcal{D}_{I+1}}$, obtenus respectivement avec les informations disponibles au temps I et $I + 1$ valent :

$$\begin{aligned}\hat{R}_i^I &= \hat{C}_{i,J}^I - C_{i,I-i} \\ \hat{R}_i^{I+1} &= \hat{C}_{i,J}^{I+1} - C_{i,I-i+1}\end{aligned}$$

Remarque 4.2.4. Suite à la Remarque 4.2.3 :

- $\hat{R}_i^{\mathcal{D}_I}$ est un estimateur non-biaisé de $\mathbb{E}[R_i^I|\mathcal{D}_I]$.
- $\hat{R}_i^{\mathcal{D}_{I+1}}$ est un estimateur non-biaisé de $\mathbb{E}[R_i^{I+1}|\mathcal{D}_{I+1}]$.

4.2.1.2 Calcul du *Claim Development Result*

Dans la suite de leur article, les auteurs vont supposer que les réserves sont exactement les engagements attendus pris par l'assureur, conditionnellement à l'information disponible. Cette hypothèse permet de définir le ***Claim Development Result*** (CDR) de l'année d'exercice comme étant la différence entre l'estimation des réserves avec l'information I et l'estimation des réserves effectuée avec l'information $I + 1$. Néanmoins, dans la mesure où, au temps $I + 1$, les réserves sont estimées avec des paiements cumulés plus récents que ceux employés dans le calcul des réserves au temps I , il faudra également ôter du calcul du CDR la différence des paiements cumulés entre les deux années considérées.

Définition 4.2.5. *Le CDR théorique de l'année de survenance i , $0 \leq i \leq I$, de l'année d'exercice $(I, I + 1]$ est donné par*

$$\text{CDR}_i(I + 1) = \mathbb{E}[R_i^I | \mathcal{D}_I] - (X_{i, I-i+1} + \mathbb{E}[R_i^{I+1} | \mathcal{D}_{I+1}])$$

où $X_{i, I-i+1} = C_{i, I-i+1} - C_{i, I-i}$ désigne le paiement incrémental de l'année i .
De plus, le ***CDR théorique agrégé*** est défini par :

$$\sum_{i=1}^I \text{CDR}_i(I + 1).$$

Remarque 4.2.6. *En réécrivant les montants des réserves en fonction des paiements cumulés, on trouve que le CDR théorique vaut*

$$\text{CDR}_i(I + 1) = \mathbb{E}[C_{i, J} | \mathcal{D}_I] - \mathbb{E}[C_{i, J} | \mathcal{D}_{I+1}]$$

Par ailleurs, puisque $\mathcal{D}_I \subset \mathcal{D}_{I+1}$, il vient

$$\mathbb{E}[\text{CDR}_i(I + 1) | \mathcal{D}_I] = 0$$

Définition 4.2.7. *Le CDR observable de l'année de survenance i , $0 \leq i \leq I$, de l'année d'exercice $(I, I + 1]$ est donné par*

$$\widehat{\text{CDR}}_i(I + 1) = \hat{R}_i^I - (X_{i, I-i+1} + \hat{R}_i^{I+1}).$$

De plus, le ***CDR observable agrégé*** est défini par :

$$\sum_{i=1}^I \widehat{\text{CDR}}_i(I + 1).$$

Remarque 4.2.8. *En considérant l'expression du CDR observable, nous remarquons que*

$$\widehat{\text{CDR}}_i(I + 1) = \hat{C}_{i, J}^I - \hat{C}_{i, J}^{I+1}.$$

Ainsi, le calcul du CDR de l'année i se résume à effectuer la différence entre les coefficients à l'ultime, estimés avec les informations \mathcal{D}_I et \mathcal{D}_{I+1} et calculés pour l'année de survenance i .

4.2.2 Estimation de la dispersion du *Claim Development Result*

Dans cette sous-partie, nous allons nous intéresser au calcul des variances conditionnelles des CDR agrégés théorique et observable. Avant toute chose, il est nécessaire de calculer le coefficient σ_j . Dans leur article, les auteurs appliquent la même formule que Mack pour déterminer un estimateur de σ_j^2 :

$$\hat{\sigma}_{j-1}^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} C_{i,j-1} \left(\frac{C_{i,j}}{C_{i,j-1}} - \hat{f}_{j-1}^I \right)^2 \text{ pour } 1 \leq j \leq J.$$

Remarque 4.2.9. Dans le cas $I = J$, Wüthrich, Merz et Lysenko utilisent l'approximation donnée dans l'article de Mack [36] à leurs données :

$$\hat{\sigma}_{I-1}^2 = \min \left(\frac{\hat{\sigma}_{I-2}^4}{\hat{\sigma}_{I-3}^2}, \min(\hat{\sigma}_{I-3}^2, \hat{\sigma}_{I-2}^2) \right).$$

4.2.2.1 Estimation de la variance conditionnelle du CDR théorique agrégé

La formule théorique de calcul de la variance du CDR théorique agrégé est donnée par le théorème suivant :

Théorème 4.2.10. $\text{Var} \left(\sum_{i=1}^I \text{CDR}_i(I+1) \middle| \mathcal{D}_I \right) = \sum_{i=1}^I \mathbb{E}[C_{i,j} | \mathcal{D}_I]^2 \cdot \frac{\sigma_{I-i}^2 / f_{I-i}^2}{C_{i,I-i}}$

Au temps I , la variance du CDR agrégé théorique peut alors être estimée par :

$$\widehat{\text{Var}} \left(\sum_{i=1}^I \text{CDR}_i(I+1) \middle| \mathcal{D}_I \right) = \sum_{i=1}^I (\hat{C}_{i,j}^I)^2 \frac{(\hat{\sigma}_{I-i}^I)^2 / (\hat{f}_{I-i}^I)^2}{C_{i,I-i}}. \quad (4.5)$$

4.2.2.2 Estimation de la variance conditionnelle du CDR observable agrégé

Tout d'abord, calculer la variance d'une somme revient à sommer les variances de chacun des termes et ajouter les termes de covariance croisés. Ainsi,

$$\text{Var} \left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \middle| \mathcal{D}_I \right) = \sum_{i=1}^I \text{Var}(\widehat{\text{CDR}}_i(I+1)) + 2 \sum_{0 < k < i} \text{Cov}((\widehat{\text{CDR}}_i(I+1)), \widehat{\text{CDR}}_k(I+1) | \mathcal{D}_I)$$

Théorème 4.2.11. La variance conditionnelle du CDR observable agrégé est estimée par la formule suivante :

$$\widehat{\text{Var}} \left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \middle| \mathcal{D}_I \right) = \sum_{i=1}^I \hat{\Gamma}_{i,J}^I + 2 \sum_{0 < k < i} \hat{\Upsilon}_{i,k}^I \quad (4.6)$$

avec, pour $i \geq 1$,

$$\hat{\Gamma}_{i,J}^I = (\hat{C}_{i,J}^I)^2 \left\{ \left(\left[1 + \frac{(\hat{\sigma}_{I-i}^I)^2 / (\hat{f}_{I-i}^I)^2}{C_{i,I-i}} \right] \prod_{j=I-i+1}^{J-1} \left(1 + \frac{(\hat{\sigma}_j^I)^2 / (\hat{f}_j^I)^2}{(S_j^{I+1})^2} \cdot C_{I-j,j} \right) \right) - 1 \right\}$$

et pour $i > k > 0$,

$$\hat{\Upsilon}_{i,k}^I = \hat{C}_{i,J}^I \hat{C}_{k,J}^I \left\{ \left(\left[1 + \frac{(\hat{\sigma}_{I-k}^I)^2 / (\hat{f}_{I-k}^I)^2}{S_{I-k}^{I+1}} \right] \prod_{j=I-k+1}^{J-1} \left(1 + \frac{(\hat{\sigma}_j^I)^2 / (\hat{f}_j^I)^2}{(S_j^{I+1})^2} \cdot C_{I-j,j} \right) \right) - 1 \right\}.$$

4.2.3 Estimation des erreurs de prédiction du modèle

Désormais, considérons les erreurs de prédiction du modèle. Wüthrich, Merz et Lysenko en dénombrent deux :

- l'erreur de prédiction du CDR théorique par le CDR observable au temps I ,
- et le biais généré par la prédiction du CDR observable.

4.2.3.1 Estimation de l'erreur quadratique moyenne entre le CDR théorique agrégé et le CDR observable agrégé

Avant de considérer l'erreur quadratique moyenne entre le CDR théorique agrégé et le CDR observable agrégé, concentrons-nous sur l'erreur quadratique moyenne obtenue pour une seule année de survénance.

Définition 4.2.12. *Pour une année de survénance i , $1 \leq i \leq I$, l'erreur quadratique moyenne conditionnelle sachant l'information \mathcal{D}_I entre le CDR théorique et le CDR observable est définie par :*

$$\text{MSE}_{\mathcal{D}_I}(\widehat{\text{CDR}}_i(I+1)) = \mathbb{E} \left[\left(\text{CDR}_i(I+1) - \widehat{\text{CDR}}_i(I+1) \right)^2 \middle| \mathcal{D}_I \right].$$

Grâce à la Remarque 4.2.6, il vient que

$$\text{MSE}_{\mathcal{D}_I}(\widehat{\text{CDR}}_i(I+1)) = \Phi_{i,J}^I + \left(\mathbb{E} \left[\widehat{\text{CDR}}_i(I+1) \middle| \mathcal{D}_I \right] \right)^2$$

avec $\Phi_{i,J}^I = \text{Var}(\text{CDR}_i(I+1) - \widehat{\text{CDR}}_i(I+1) | \mathcal{D}_I)$

$$= \text{Var}(\text{CDR}_i(I+1) | \mathcal{D}_I) + \text{Var}(\widehat{\text{CDR}}_i(I+1) | \mathcal{D}_I) - 2\text{Cov}(\text{CDR}_i(I+1), \widehat{\text{CDR}}_i(I+1) | \mathcal{D}_I).$$

A présent, il faut estimer cette erreur.

Théorème 4.2.13. *Pour une année de survénance i , $1 \leq i \leq I$, l'erreur quadratique moyenne conditionnelle sachant l'information \mathcal{D}_I entre le CDR théorique et le CDR observable est estimée par la formule suivante :*

$$\text{MSE}_{\mathcal{D}_I}(\widehat{\text{CDR}}_i(I+1)) = \hat{\Phi}_{i,J}^I + (\hat{C}_{i,J}^I)^2 \hat{\Delta}_{i,J}^I \quad (4.7)$$

avec $\hat{\Phi}_{1,J}^I = 0$

$$\hat{\Phi}_{i,J}^I = (\hat{C}_{i,J}^I)^2 \left[1 + \frac{(\hat{\sigma}_{I-i}^I)^2 / (\hat{f}_{I-i}^I)^2}{C_{i,I-i}} \right] \left(\prod_{j=I-i+1}^{J-1} \left(1 + \frac{(\hat{\sigma}_j^I)^2 / (\hat{f}_j^I)^2}{(S_j^{I+1})^2} C_{I-j,j} \right) - 1 \right) \text{ pour } i > 1$$

$$\hat{\Delta}_{i,J}^I = \frac{(\hat{\sigma}_{I-i}^I)^2 / (\hat{f}_{I-i}^I)^2}{S_{I-i}^I} + \sum_{j=I-i+1}^{J-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \frac{(\hat{\sigma}_j^I)^2 / (\hat{f}_j^I)^2}{S_j^I} \text{ pour } 1 \leq i \leq I.$$

Dès lors, nous pouvons définir et estimer l'erreur quadratique moyenne associée au CDR observable agrégé.

Définition 4.2.14. *L'erreur quadratique moyenne conditionnelle sachant l'information \mathcal{D}_I entre le CDR théorique agrégé et le CDR observable agrégé est définie par :*

$$\text{MSE}_{\mathcal{D}_I} \left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \right) = \mathbb{E} \left[\left(\sum_{i=1}^I \text{CDR}_i(I+1) - \sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \right)^2 \middle| \mathcal{D}_I \right].$$

Théorème 4.2.15. *Conditionnellement à l'information disponible au temps I , l'erreur quadratique moyenne du CDR observable agrégé s'estime par la formule suivante :*

$$\widehat{\text{MSE}}_{\mathcal{D}_I} \left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \right) = \sum_{i=1}^I \widehat{\text{MSE}}(\widehat{\text{CDR}}_i(I+1)) + 2 \sum_{i>k>0} (\hat{\Psi}_{i,k}^I + \hat{C}_{i,J}^I \hat{C}_{k,J}^I \hat{\Lambda}_{k,J}^I) \quad (4.8)$$

avec $\hat{\Psi}_{i,1}^I = 0$ pour $i > 1$

$$\hat{\Psi}_{i,k}^I = \frac{\hat{C}_{i,J}^I}{\hat{C}_{k,J}^I} \cdot \left(1 + \frac{(\hat{\sigma}_{I-k}^I)^2 / (\hat{f}_{I-k}^I)^2}{S_{I-k}^{I+1}} \right) \left(1 + \frac{(\hat{\sigma}_{I-k}^I)^2 / (\hat{f}_{I-k}^I)^2}{C_{k,I-k}} \right)^{-1} \hat{\Phi}_{k,J}^I \text{ pour } i > k > 1$$

$$\hat{\Lambda}_{k,J}^I = \frac{C_{k,I-k}}{S_{I-k}^{I+1}} \frac{(\hat{\sigma}_{I-k}^I)^2 / (\hat{f}_{I-k}^I)^2}{S_{I-k}^I} + \sum_{j=I-k+1}^{J-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \frac{(\hat{\sigma}_j^I)^2 / (\hat{f}_j^I)^2}{S_j^I} \text{ pour } i > k > 1.$$

4.2.3.2 Estimation du biais du modèle

Enfin, concluons cette partie avec l'estimation du potentiel biais engendré par le modèle. Pour ce faire, rappelons-nous de la formule de décomposition de la variance.

Théorème 4.2.16. *Soient X , Y , deux variables aléatoires sur le même espace de probabilité, avec $\text{Var}(Y) < \infty$. Alors*

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]).$$

En appliquant cette formule au CDR observable agrégé et compte tenu de la Remarque 4.2.6, en conditionnant l'équation précédente par rapport à \mathcal{D}_I , il ressort que :

$$\mathbb{E} \left[\left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \right)^2 \middle| \mathcal{D}_I \right] = \underbrace{\hat{\mathbb{E}}_{\mathcal{D}_I} \left[\mathbb{E} \left(\sum_{i=1}^I \left[\widehat{\text{CDR}}_i(I+1) \right] \middle| \mathcal{D}_I \right)^2 \right]}_{(u\text{-bias})^2} + \widehat{\text{Var}} \left(\sum_{i=1}^I \widehat{\text{CDR}}_i(I+1) \middle| \mathcal{D}_I \right). \quad (4.9)$$

Si la variance conditionnelle du CDR observable agrégé a été donnée dans le Théorème 4.2.11, il reste à calculer le terme de biais.

Théorème 4.2.17. *Le terme de biais est donné par la formule suivante :*

$$(u\text{-bias})^2 = \sum_{i=1}^I (\hat{C}_{i,J}^I)^2 \hat{\Delta}_{i,J}^I + 2 \sum_{i>k>0} \hat{C}_{i,J}^I \hat{C}_{k,J}^I \hat{\Lambda}_{k,J}^I. \quad (4.10)$$

Dès lors, il est possible de calculer l'erreur directement induite par le modèle. Par conséquent, le terme d'espérance de l'Equation (4.9) pourra être comparé au MSE déterminé par la méthode de Mack.

En effet, si dans l'article de Mack, on considère l'écart quadratique moyen entre $\hat{C}_{i,J}^I$ et $C_{i,J}^I$, pour toutes les années de survenance i , ici les auteurs considèrent l'erreur quadratique moyenne entre $\hat{C}_{i,J}^I$ et $\hat{C}_{i,J}^{I+1}$. Or, $\hat{C}_{i,J}^{I+1}$ peut s'apparenter en quelque sorte à la valeur théorique $C_{i,J}^I$, dans la mesure où elle en fournit une meilleure estimation que $\hat{C}_{i,J}^I$.

Remarque 4.2.18. *Cette comparaison a également été faite dans l'article de Merz et Wüthrich Modelling The Claims Development Result For Solvency Purposes [63]. En particulier, les auteurs constatent que, à la différence de l'erreur de prévision donnée par Mack, celle sur le CDR pondère l'erreur d'estimation des paiements au-delà de la diagonale suivante (i.e $j \geq I - i + 1$) par un facteur d'échelle $\frac{C_{i,I-i}}{S_j^{I+1}} \leq 1$.*

4.3 Modèles économétriques de provisionnement

4.3.1 Régressions linéaires et provisionnement

Une autre approche possible lorsque l'on s'intéresse au problème de provisionnement est d'adopter un modèle économétrique. Dans *Estimation of IBNR claims by least squares*, De Vylder [60] présente une régression linéaire classique appliquée aux paiements incrémentaux $X_{i,j}$, dans laquelle $X_{i,j}$ se décompose comme un produit d'un facteur de ligne α_i et d'un facteur de colonne β_j .

$$X_{i,j} = \alpha_i \beta_j + \epsilon_{i,j} \text{ avec } \epsilon_{i,j} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

L'hypothèse de normalité des résidus, ie $\epsilon_{i,j} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ est centrale dans un problème de régression linéaire. Sous la condition $X_{i,j} \sim \mathcal{N}(\alpha_i \beta_j, \sigma^2)$, les coefficients $\hat{\alpha}$ et $\hat{\beta}$ estimés par les moindres carrés sont :

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmin} \left\{ \sum_{i,j} (X_{i,j} - \alpha_i \cdot \beta_j)^2 \right\}.$$

Dès lors, les conditions d'ordre 1 pour α_i et β_j donnent :

$$\hat{\alpha}_i = \frac{\sum_j \hat{\beta}_j X_{i,j}}{\sum_j \hat{\beta}_j^2} \text{ et } \hat{\beta}_j = \frac{\sum_i \hat{\alpha}_i X_{i,j}}{\sum_i \hat{\alpha}_i^2}.$$

Néanmoins, ces deux équations sont dépendantes, donc on ne peut pas résoudre explicitement le problème d'optimisation. Pour palier cela, Christofides propose dans *Regression models based on log-incremental payments* [11] de réécrire le modèle précédent sous une forme log-linéaire, c'est-à-dire

$$\ln X_{i,j} \sim \mathcal{N}(a_i + b_j, \sigma^2).$$

L'estimation des paramètres est désormais possible car ces derniers ne sont plus dépendants l'un de l'autre. En effet, il vient que :

$$\hat{a}_i = \sum_j \ln X_{i,j} \text{ et } \hat{b}_j = \sum_i \ln X_{i,j}.$$

Toutefois, le modèle log-linéaire impose que X suive une loi log-normale. Par conséquent, $\mathbb{E}[X_{i,j}] = \exp(a_i + b_j + \frac{\sigma^2}{2})$. Ainsi, l'estimateur $\hat{X}_{i,j} = \exp(\hat{a}_i + \hat{b}_j)$ est biaisé : il faut plutôt estimer $X_{i,j}$ par $\tilde{X}_{i,j} = \exp(\hat{a}_i + \hat{b}_j + \frac{\hat{\sigma}^2}{2})$.

4.3.2 Modèle linéaire généralisé et lien avec la méthode *chain-ladder*

4.3.2.1 Modèle linéaire généralisé

Soient X les variables explicatives, Y la variable expliquée. Dans un **modèle linéaire généralisé**, un modèle de régression linéaire classique est relié à l'espérance $\mathbb{E}[Y|X]$ grâce à une **fonction de lien** g inversible. On écrit alors le modèle de régression linéaire classique $g(Y) = X\beta$. Puis on obtient $\mathbb{E}[Y|X] = g^{-1}(X\beta)$.

Ce type de modèle permet d'étendre la méthode de régression linéaire à des variables pour lesquelles la condition de normalité des résidus ne serait pas respectée.

4.3.2.2 Lien avec la méthode *chain-ladder*

Plusieurs auteurs, tels que Kremer [31] et Mack [35], ont démontré qu'il était possible de retrouver l'estimateur des réserves de *chain-ladder* à travers les résultats d'un GLM. Pour ce faire, ils choisissent un modèle log-Poisson sur les paiements incrémentaux, puis somment les paiements futurs.

Cependant, Dutang et Charpentier [9] rappellent que les paiements sont sujets à un phénomène de sur-dispersion. Par conséquent, l'utilisation d'une loi de Poisson, dont le paramètre de dispersion $\frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]}$ est de 1, est discutable. Ils proposent alors d'appliquer une loi dite quasi-Poisson ou Poisson surdispersée, dont le paramètre de dispersion est supérieur à 1.

4.3.3 Estimation de l'erreur de prédiction du modèle

Dans tous les types de modèles linéaires généralisés, il serait possible de déterminer les erreurs de prédiction à l'aide d'une formule fermée.

En particulier, dans *A stochastic model underlying the chain-ladder technique*, Renshaw et Verall [50] estiment l'incertitude sur le montant des réserves à l'ultime pour une régression Poisson surdispersée. Néanmoins, leur approximation n'est valable que dans un cas asymptotique, ce qui est problématique dans le cadre de provisionnement, où les données disponibles ne sont pas très nombreuses.

Théorème 4.3.1. *Considérons un modèle GLM Poisson surdispersé de paramètre ϕ et de fonction lien $g = \ln$, c'est-à-dire $\hat{X}_{i,j} = \hat{\mu}_{i,j} = \exp(\hat{\eta}_{i,j})$. Alors*

$$\mathbb{E}[(R - \hat{R})^2] \approx \sum_{i+j>I} \hat{\phi} \cdot \hat{\mu}_{i,j} + \hat{\mu}'_F \cdot \text{Var}(\hat{\eta}_F) \cdot \hat{\mu}_F$$

avec $\hat{\mu}_F$ et $\hat{\eta}_F$ les restrictions respectives des vecteurs $\hat{\mu}$ et $\hat{\eta}$ aux indices $i + j > I$.

4.4 Simulation des montants des réserves par bootstrap

4.4.1 Présentation de la technique de bootstrap

La technique de bootstrap est une méthode d'inférence statistique permettant, à travers l'étude d'échantillons, de déterminer la loi d'une distribution. Elle repose sur des rééchantillonnages multiples de données. Concrètement, des tirages avec remise vont être effectués parmi les données originelles afin de former de nouveaux échantillons. Dès lors, ces derniers vont contribuer à caractériser la distribution des données initiales, F , grâce à des estimations d'espérance ou de variance [16]. Cette méthode est très pertinente lorsque nous sommes face à un nombre de données insuffisant pour extrapoler une loi de distribution.

Il existe deux sortes de bootstrap :

- Le **bootstrap paramétrique**, pour lequel il est supposé que la loi de la distribution soit paramétrique. Autrement dit, il est alors nécessaire de déterminer les paramètres de cette dernière pour caractériser F . Par exemple, on peut faire l'hypothèse que F suive une loi normale de paramètres μ et σ^2 . Dans ce cas, le bootstrap devra permettre d'estimer μ et σ^2 , notamment via la méthode des moments ou la recherche du maximum de vraisemblance.

- Le **bootstrap non-paramétrique**, pour lequel aucune hypothèse n'est faite sur la forme de la loi de F . Le rééchantillonnage des données va alors permettre d'estimer la fonction de répartition empirique de F grâce au calcul de ses premiers moments.

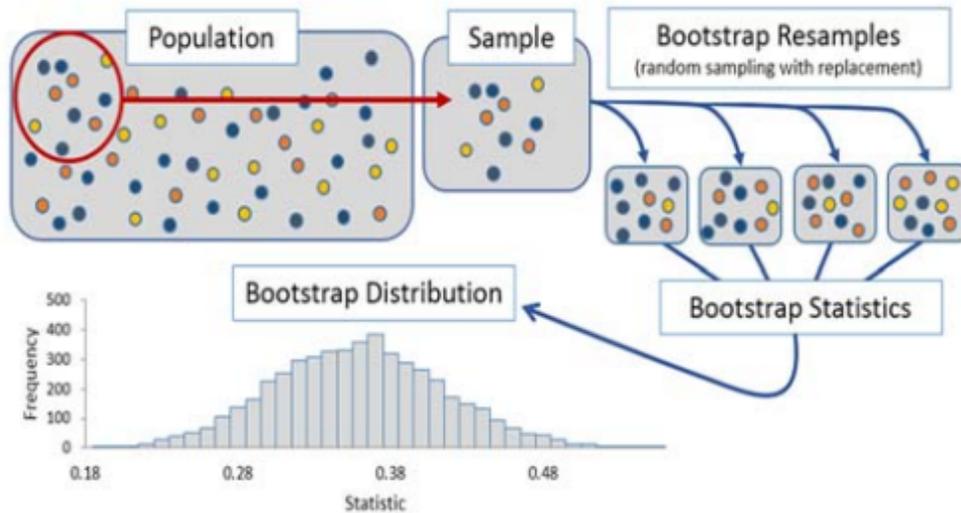


FIGURE 4.2 – Schéma du processus de bootstrap [5]

Il faut néanmoins être conscient des limites que peut présenter cette technique :

- Dans le cadre de bootstrap paramétrique, une erreur statistique est possible si la forme de la loi a été mal choisie.
- Si on ne simule pas assez d'échantillons, la loi de distribution peut s'éloigner de la loi théorique. Pour remédier à cela, on peut cependant augmenter le nombre de simulations, tout en ayant conscience des contraintes possibles (temps de calcul plus long, atteinte de la capacité maximale de calcul de la machine...).

4.4.2 Application au modèle de Mack (1993) - Cas général

Comme nous l'avons constaté, l'erreur de prédiction dans le modèle de Mack [36] présente deux sources d'incertitude : l'erreur d'estimation et l'erreur de modèle. Pour quantifier ces deux erreurs, il va falloir appliquer deux algorithmes de bootstrap, présentés par Charpentier et Dutang dans *L'actuariat avec R* [9]. Néanmoins, leurs structures sont très proches : seules quelques étapes vont différer.

L'idée générale derrière ces deux algorithmes est de régénérer la partie supérieure du triangle. Pour ce faire, il faut re-simuler des résidus. Ces derniers seront supposés indépendants et identiquement distribués. Pour les obtenir, on peut déterminer les résidus de Pearson relatifs à un modèle GLM. Pour rappel, pour une variable prédite $Y_{i,j}$, ces derniers se définissent ainsi :

$$r_{i,j} = \frac{Y_{i,j} - \hat{Y}_{i,j}}{\sqrt{\text{Var}(\hat{Y}_{i,j})}}.$$

Afin d'avoir une variance unitaire, les résidus sont ajustés grâce au nombre k de paramètres estimés de la façon suivante :

$$r_{i,j}^A = \sqrt{\frac{I}{I-k}} r_{i,j}.$$

Une fois ces résidus calculés, la boucle de simulations par bootstrap peut démarrer. Pour chaque itération k , un tirage avec remise des résidus $r_{i,j}^A$ centrés est effectué. Notons ces nouveaux résidus $r_{i,j}^k$. Le triangle supérieur peut alors être régénéré en inversant la formule des résidus de Pearson :

$$X_{i,j}^k = \hat{X}_{i,j} + \sqrt{\hat{X}_{i,j}} r_{i,j}^k$$

A ce stade, nous pouvons désormais distinguer les deux algorithmes. Concernant l'erreur d'estimation, cette dernière pourra être mesurée en recalculant les paiements futurs obtenus pour ce nouveau triangle, grâce aux coefficients de *chain ladder*, puis les réserves \hat{R}^k . La variance des montants de réserves quantifiera l'erreur d'estimation.

Pour estimer l'erreur de modèle, il est nécessaire de simuler les trajectoires de paiements possibles pour les années futures. Pour cela, deux méthodes s'offrent à nous :

1. La première consiste à effectuer un modèle linéaire généralisé (GLM) sur la partie inférieure du triangle à partir du triangle généré grâce aux résidus $r_{i,j}^k$. Se pose alors la question du choix de la loi de distribution. Si l'utilisation d'une loi de Poisson semble évidente, il peut toutefois être pertinent d'incorporer des distributions ayant davantage de variance [9], telles que la loi Gamma ou la loi binomiale négative.
2. La seconde consiste à modéliser la trajectoire des paiements cumulés grâce au modèle de *chain-ladder*. Pour chaque itération k ,

$$\hat{C}_{i,j}^k | \hat{C}_{i,j-1}^k \sim \mathcal{N}(\hat{f}_{j-1} \hat{C}_{i,j-1}^k, \hat{\sigma}_{j-1}^2 \hat{C}_{i,j-1}^k).$$

Il est alors possible d'obtenir une distribution du montant de réserves théorique.

Le point de départ pour ces deux algorithmes était l'application d'un modèle GLM afin de calculer des résidus permettant de re-simuler le triangle *run-off*. Il existe également une autre approche, qui se concentre essentiellement sur le modèle de Mack, sans faire intervenir de régression linéaire. Celle-ci fait l'objet de la section suivante.

4.4.3 Application au modèle de Mack (1993) - Algorithme avec bootstrap récursif

Dans cette partie, nous allons détailler un algorithme de bootstrap qui permet de mesurer simultanément les erreurs de modèle et les erreurs d'estimation dans le modèle de Mack [36]. Il est présenté dans le *Cours d'assurance dommage* de Baradel [6] et s'articule en deux temps :

1. Préliminaires au bootstrap.
2. Boucle bootstrap.

4.4.3.1 Préliminaires au bootstrap

Les étapes qui suivent s'effectuent en amont de la boucle de bootstrap :

1. Calcul des paramètres du modèle *chain ladder* \hat{f}_j et $\hat{\sigma}_j^2$, pour $0 \leq j \leq I - 1$.
2. Calcul des résidus obtenus à partir des estimations des paiements cumulés de la partie supérieure du triangle :

$$r_{i,j} = \sqrt{\frac{C_{i,j}}{\hat{\sigma}_j^2}} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right) \text{ pour } i, j \text{ tels que } i + j \leq n - 1$$

4.4.3.2 Boucle bootstrap

Pour chaque itération k de la boucle de bootstrap, on procède de la manière suivante :

1. Tirage avec remise des résidus $r_{i,j}$ du triangle supérieur. On obtient alors les résidus $r_{i,j}^k$.
2. Régénération de la partie supérieure du triangle :

$$C_{i,1}^k = C_{i,1} \text{ pour } i \text{ tel que } 1 \leq i \leq I$$

$$C_{i,j+1}^k = \hat{f}_j^k C_{i,j}^k + r_{i,j}^k \hat{\sigma}_j^k \sqrt{C_{i,j}^k} \text{ pour } i, j \text{ tels que } i + j \leq I - 1$$

3. Réestimation des paramètres \hat{f}_j^k et $\hat{\sigma}_j^k$ sur ce nouveau triangle de paiements cumulés.
4. Simulation des paiements cumulés futurs. On procède de manière analogue à l'étape 4, en utilisant les nouveaux paramètres \hat{f}_j^k et $\hat{\sigma}_j^k$:

$$C_{i,I-i+1}^k = C_{i,I-i+1} \text{ pour } i \text{ tel que } 1 \leq i \leq I$$

$$C_{i,j+1}^k = \hat{f}_j^k C_{i,j}^k + r_{i,j}^k \hat{\sigma}_j^k \sqrt{C_{i,j}^k} \text{ pour } i, j \text{ tels que } 2 \leq i \leq I \text{ et } I - i + 2 \leq j \leq I$$

5. Calcul des réserves R_i^k :

$$R_i^k = C_{i,I}^k - C_{i,I-i+1}, \quad 1 \leq i \leq I$$

4.4.3.3 Analyse des résultats du bootstrap

Si cet algorithme a permis de prendre en compte les deux sources d'incertitude de l'erreur quadratique moyenne, c'est parce qu'il s'est décomposé en deux phases :

- Estimation de nouveaux paramètres *chain-ladder* grâce au rééchantillonnage du triangle supérieur, agissant sur l'erreur d'estimation ;
- Re-simulation du triangle inférieur, agissant sur l'erreur de modèle.

Dès lors, la variance des réserves estimées totales générées durant les N itérations de la boucle bootstrap quantifie l'erreur quadratique moyenne. On peut alors la calculer de manière empirique :

$$\hat{s}^2 = \frac{1}{N} \sum_{k=1}^N \left(\left(\sum_{i=2}^I R_i^k \right) - \bar{R} \right)^2 \text{ avec } \bar{R} \text{ moyenne empirique telle que } \bar{R} = \frac{1}{I} \sum_{k=1}^N \left(\sum_{i=2}^I R_i^k \right).$$

Chapitre 5

Estimation du skewness et du kurtosis dans le modèle de Mack

Dans ses articles *A closed-form formula for the skewness estimation of non-life reserve risk distribution* (2012) [13] et *Kurtosis and skewness estimation for non-life reserve risk distribution* (2013) [15], Dal Moro propose une extension du modèle de Mack en donnant des estimateurs des moments d'ordre 3 et 4 du montant des réserves. Avant de présenter sa méthode, nous allons brièvement revenir sur la définition de ces deux moments.

5.1 Définitions et interprétations

5.1.1 Skewness

Le **skewness** ou **coefficient d'asymétrie** correspond au moment d'ordre 3 d'une variable aléatoire réelle centrée réduite. Il permet de mesurer l'asymétrie de sa distribution [61].

Définition 5.1.1. Soit X une variable aléatoire réelle, de moyenne μ et d'écart-type σ . On définit le skewness γ_1 de X par

$$\gamma_1 = \text{Skew}(X) = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right].$$

Ce coefficient s'interprète de la manière suivante :

- Si $\gamma_1 < 0$, la distribution se décale à droite de la médiane. Par conséquent, la queue de gauche est plus longue et épaisse (Figure 5.1a).
- Si $\gamma_1 = 0$, cela n'implique pas que la distribution est symétrique. Néanmoins, toutes les distributions symétriques sont telles que γ_1 est nul (Figure 5.1b).
- Si $\gamma_1 > 0$, la distribution se décale à gauche de la médiane. Par conséquent, la queue de droite est plus longue et épaisse (Figure 5.1c).

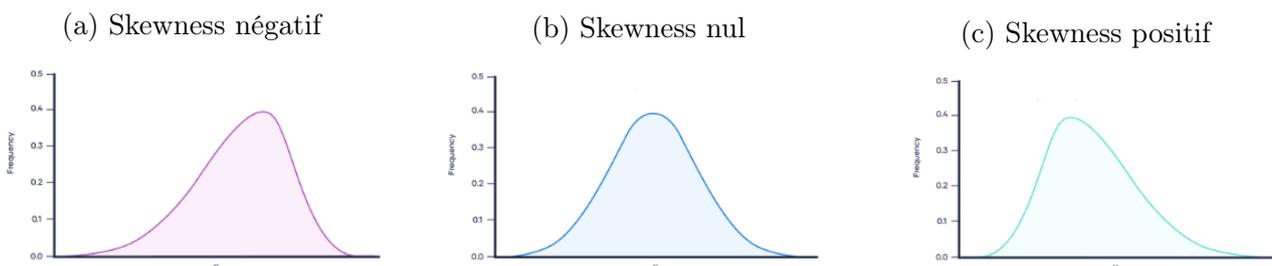


FIGURE 5.1 – Asymétries possibles [53]

5.1.2 Kurtosis

Le **kurtosis** ou **coefficient d'aplatissement**, **degré d'acuité** ou **degré de voussure** est le moment d'ordre 4 d'une variable aléatoire réelle centrée réduite. Il permet de mesurer l'aplatissement de sa distribution [62].

Définition 5.1.2. Soit X une variable aléatoire réelle, de moyenne μ et d'écart-type σ . On définit le kurtosis β_2 de X par

$$\beta_2 = \text{Kurt}(X) = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right].$$

Ce coefficient s'interprète de la manière suivante :

- Si $\beta_2 < 3$, il s'agit d'une distribution **platykurtique**. Dès lors, le centre et les queues de la distribution vont être écrasés, tandis que les flancs seront bien plus larges. Les outliers sont donc peu fréquents (Figure 5.2a).
- Si $\beta_2 = 3$, il s'agit d'une distribution **mésokurtique**. On peut par exemple retrouver les lois normales dans cette catégorie (Figure 5.2b).
- Si $\beta_2 > 3$, il s'agit d'une distribution **leptokurtique**. Les queues sont épaisses et les outliers fréquents (Figure 5.2c).

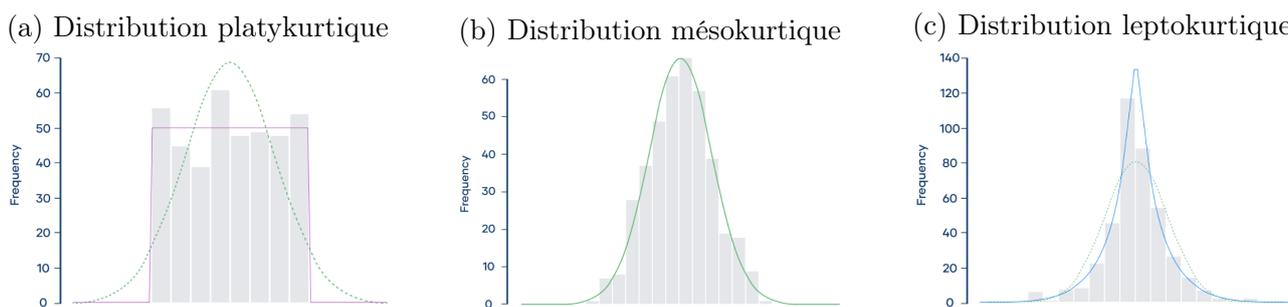


FIGURE 5.2 – Aplatissements possibles¹ [54]

Définition 5.1.3. On utilise parfois le **kurtosis normalisé**, défini par

$$\gamma_2 = \frac{\mathbb{E} [(X - \mu)^4] - 3(\sigma^2)^2}{(\sigma^2)^2} = \beta_2 - 3.$$

Dès lors, ce coefficient s'interprète comme le degré d'aplatissement de la distribution par rapport à la distribution gaussienne, de kurtosis égal à 3.

5.2 Estimation du skewness et du kurtosis des paiements cumulés

5.2.1 Skewness du paiement cumulé

Dans son article *Application of skewness in non-life reserving* [14], Dal Moro intègre le concept de skewness dans le cadre du modèle présenté par Mack en 1993 [36]. Pour ce faire, il introduit une notation, présentée ci-après.

1. Les distributions sont comparées à la distribution gaussienne (en vert).

Notation 5.2.1. Soit $SK(A|B)$ le moment centré d'ordre 3 de A sachant B . Autrement dit,

$$SK(A|B) = \mathbb{E}[(A - \mathbb{E}[A|B])^3|B].$$

$$\text{Ainsi } Skew(A|B) = \frac{SK(A|B)}{\text{Var}(A|B)^{3/2}}.$$

Dal Moro présente et démontre deux hypothèses déterminantes pour l'estimation du skewness du paiement cumulé.

Hypothèse 4 *Le skewness du paiement cumulé $C_{i,j+1}$ conditionnel aux paiements passés ne dépend pas de l'année de survenance i mais de l'année de développement j , soit*

$$Skew(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \frac{SK(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))}{\text{Var}(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))^{3/2}} \text{ dépend de } j \text{ mais pas de } i. \quad (\text{H4})$$

De fait, cette hypothèse impose que, pour une année de développement donnée, l'asymétrie des distributions de paiements cumulés est identique pour chaque année de survenance. Pour la justifier, il est nécessaire de tracer les droites $\frac{(C_{i,j+1} - \hat{f}_j C_{i,j})^3}{C_{i,j}^{3/2}}$ (car $\text{Var}(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \sigma_j^2 C_{i,j}$ d'après l'hypothèse (H1)) pour chaque année de développement j et de vérifier que l'on n'observe pas de tendance en fonction de l'année de survenance i .

Dès lors, on peut écrire que $Skew(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \gamma_j$, ce qui justifie l'hypothèse suivante.

Hypothèse 5 *Il existe des coefficients $Sk_j > 0$ pour $1 \leq j \leq I - 2$ tels que*

$$SK(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = Sk_j^3 C_{i,j}^{3/2}. \quad (\text{H5})$$

On aura en effet $Sk_j^3 = \gamma_j \sigma_j^3$.

Dal Moro donne un estimateur non-biaisé \hat{Sk}_j^3 de Sk_j^3 , dont la formule est présentée dans le théorème suivant.

Théorème 5.2.2. *Pour $1 \leq j \leq I - 2$,*

$$\hat{Sk}_j^3 = \frac{1}{\left(\sum_{i=1}^{I-j} C_{i,j}^{3/2} \right)^2} \sum_{i=1}^{I-j} C_{i,j}^{3/2} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^3. \quad (5.1)$$

Pour $j = I - 1$, on posera $\hat{Sk}_j^3 = 0$.

5.2.2 Kurtosis du paiement cumulé

Dans son article *Kurtosis and skewness estimation for non-life reserve risk distribution* [15], Dal Moro intègre le concept de kurtosis dans le cadre du modèle présenté par Mack en 1993 [36]. Comme ce qui a été fait pour le skewness, on introduit la notation suivante.

Notation 5.2.3. Soit $KT(A|B)$ le moment d'ordre 4 centré de A sachant B . Autrement dit,

$$KT(A|B) = \mathbb{E}[(A - \mathbb{E}[A|B])^4|B].$$

$$\text{Ainsi } Kurt(A|B) = \frac{SK(A|B)}{\text{Var}(A|B)^2}.$$

Dal Moro présente deux hypothèses déterminantes pour l'estimation du kurtosis du paiement cumulé, analogues à celles concernant le skewness.

Hypothèse 6 Le kurtosis du paiement cumulé $C_{i,j+1}$ conditionnel aux paiements passés ne dépend pas de l'année de survenance i mais de l'année de développement j , soit

$$Kurt(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \frac{KT(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))}{\text{Var}(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))^2} \text{ dépend de } j \text{ mais pas de } i. \quad (\text{H6})$$

De fait, cette hypothèse impose que, pour une année de développement donnée, le kurtosis des distributions de paiements cumulés est identique pour chaque année de survenance. Cette condition s'applique dans le cas où les risques du portefeuille présentent les mêmes caractéristiques pour chaque année de survenance.

Dès lors, on peut écrire que $Kurt(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \lambda_j$, ce qui justifie l'hypothèse suivante.

Hypothèse 7 Il existe des coefficients $Kt_j > 0$ pour $1 \leq j \leq I - 1$ tels que

$$KT(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = Kt_j^4 C_{i,j}^2. \quad (\text{H7})$$

On aura en effet $Kt_j^4 = \lambda_j \sigma_j^4$. Dal Moro donne un estimateur non-biaisé $\hat{K}t_j^4$ de Kt_j^4 , dont la formule est présentée dans le théorème suivant.

Théorème 5.2.4. Pour $1 \leq j \leq I - 4$,

$$\hat{K}t_j^4 = \frac{\left[\sum_{i=1}^{I-j} C_{i,j}^2 \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^4 - 3 (\hat{\sigma}_j^2)^2 \left(2 - 6 \frac{\sum_{i=1}^{I-j} C_{i,j}^2}{\left(\sum_{i=1}^{I-j} C_{i,j} \right)^2} + 4 \frac{\sum_{i=1}^{I-j} C_{i,j}^3}{\left(\sum_{i=1}^{I-j} C_{i,j} \right)^3} \right) \right]}{\left(\sum_{i=1}^{I-j} \left(1 - \frac{C_{i,j}}{\sum_{i=1}^{I-j} C_{i,j}} \right)^4 + \frac{\left(\sum_{i=1}^{I-j} C_{i,j}^2 \right)^2 - \sum_{i=1}^{I-j} C_{i,j}^4}{\left(\sum_{i=1}^{I-j} C_{i,j} \right)^4} \right)}. \quad (5.2)$$

Pour $I - 3 \leq j \leq I - 1$, on posera $\hat{K}t_j^4 = 3 (\hat{\sigma}_j^2)^2$. Cela revient à supposer que les distributions des paiements cumulés des dernières années de développement sont gaussiennes.

A partir de ces hypothèses sur le skewness et le kurtosis, deux approches sont possibles pour estimer l'asymétrie et l'aplatissement de la distribution des réserves :

- une approche par **formule fermée**,
- et une approche par **simulation**.

Si la première fait intervenir des calculs très lourds, la seconde a l'avantage d'être plus facilement applicable. Nous allons néanmoins détailler les deux méthodes. Enfin, nous présenterons une méthode de bootstrap que nous avons adaptée afin de tenir compte des hypothèses de skewness et de kurtosis données par Dal Moro.

5.3 Approche par formule fermée

5.3.1 Estimation du skewness d'une année de survenance

Définition 5.3.1. *Le skewness de l'année de survenance i de l'estimateur du paiement cumulé ultime $\hat{C}_{i,I}$ de $C_{i,I}$ s'écrit*

$$SK(\hat{C}_{i,I}) = \mathbb{E}[(C_{i,I} - \hat{C}_{i,I})^3 | \mathcal{D}_I]$$

De la même manière que le calcul de l'erreur de prédiction moyenne des réserves a nécessité un calcul d'erreur par année de survenance, qui ont ensuite été sommées, et l'ajout d'un terme de covariance, Dal Moro [13] applique la même méthode au calcul du skewness de la distribution des réserves.

Remarque 5.3.2. *Une démonstration analogue à celle de la Remarque 1.1.4. pourrait être faite pour comprendre que l'erreur faite sur le paiement cumulé ultime $C_{i,I}$ correspond à l'erreur commise sur le montant des réserves R_i .*

Dès lors, en développant le cube², cette formule peut se décomposer de la manière suivante :

$$\begin{aligned} SK(\hat{C}_{i,I}) &= \mathbb{E}[(C_{i,I} - \mathbb{E}[C_{i,I} | \mathcal{D}_I])^3 | \mathcal{D}_I] + 3\mathbb{E}[(C_{i,I} - \mathbb{E}[C_{i,I} | \mathcal{D}_I])^2 (\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I}) | \mathcal{D}_I] \\ &\quad + 3\mathbb{E}[(C_{i,I} - \mathbb{E}[C_{i,I} | \mathcal{D}_I]) (\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^2 | \mathcal{D}_I] + \mathbb{E}[(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^3 | \mathcal{D}_I] \\ &= SK(C_{i,I} | \mathcal{D}_I) + 3\text{Var}(C_{i,I} | \mathcal{D}_I) (\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I}) + (\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^3 \end{aligned}$$

car $\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I}$ est \mathcal{D}_I -mesurable. De plus, $SK(\hat{C}_{i,I})$ et $\text{Var}(C_{i,I} | \mathcal{D}_I)$ étant également \mathcal{D}_I -mesurables et $\hat{C}_{i,I}$ sans biais, on a

$$\begin{aligned} \mathbb{E}[SK(\hat{C}_{i,I}) | \mathcal{D}_I] &= \mathbb{E}[SK(C_{i,I} | \mathcal{D}_I) | \mathcal{D}_I] + 3\mathbb{E}[\text{Var}(C_{i,I} | \mathcal{D}_I) (\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I}) | \mathcal{D}_I] \\ &\quad + \mathbb{E}[(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^3 | \mathcal{D}_I] \\ \implies SK(\hat{C}_{i,I}) &= \underbrace{SK(C_{i,I} | \mathcal{D}_I)}_{(1)} + \underbrace{(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I})^3}_{(2)}. \end{aligned} \quad (5.3)$$

Remarque 5.3.3. *Nous pouvons noter que le skewness d'une année de survenance se décompose, comme l'erreur quadratique moyenne dans le modèle de Mack [36], en une erreur d'estimation (1) et une erreur de modèle (2).*

5.3.1.1 Calcul du terme (1)

En appliquant la loi du skewness total³ et les hypothèses (H2), (H3) et (H4), il vient que

$$\begin{aligned} SK(C_{i,I} | \mathcal{D}_I) &= SK(C_{i,I} | C_{i,1}, \dots, C_{i,I+1-i}) \\ &= \mathbb{E}[SK(C_{i,I} | C_{i,1}, \dots, C_{i,I-1}) | C_{i,1}, \dots, C_{i,I+1-i}] + SK(\mathbb{E}[C_{i,I} | C_{i,1}, \dots, C_{i,I-1}] | C_{i,1}, \dots, C_{i,I+1-i}) \\ &\quad + 3\text{Cov}(\mathbb{E}[C_{i,I} | C_{i,1}, \dots, C_{i,I-1}], \text{Var}(C_{i,I} | C_{i,1}, \dots, C_{i,I-1}) | C_{i,1}, \dots, C_{i,I+1-i}) \\ &= Sk_{I-1}^3 \mathbb{E}[C_{i,I-1}^{3/2} | C_{i,1}, \dots, C_{i,I+1-i}] + f_{I-1}^3 SK(C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}) \\ &\quad + 3f_{I-1} \sigma_{I-1}^2 \text{Var}(C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}). \end{aligned} \quad (5.4)$$

2. $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

3. Pour A et B variables aléatoires, $SK(A) = \mathbb{E}[SK(A|B)] + SK(\mathbb{E}[A|B]) + 3\text{Cov}[\mathbb{E}[A|B], \text{Var}(A|B)]$

Dans cette équation, seul le terme $\mathbb{E} \left[C_{i,I-1}^{3/2} | C_{i,1}, \dots, C_{i,I+1-i} \right]$ n'est pas directement estimable avec les formules de Mack. Dans son article, Dal Moro suppose que les paiements cumulés suivent une loi log-normale.

Définition 5.3.4. Soit X une variable aléatoire. On appelle **coefficient de variation** de X le ratio :

$$\text{CoV}[X] = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]}.$$

Remarque 5.3.5. Dans le cas où $X \sim \text{LogN}(\mu, \sigma^2)$,

$$E[X^{3/2}] = \exp\left(\frac{3}{2}\mu + \frac{9}{8}\sigma^2\right) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)^{3/2} \exp(\sigma^2)^{3/8}$$

Or, $\exp\left(\mu + \frac{1}{2}\sigma^2\right) = \mathbb{E}[X]$ et $\exp(\sigma^2) = 1 + \text{CoV}[X]^2$, d'où

$$E[X^{3/2}] = (1 + \text{CoV}[X]^2)^{3/8} \mathbb{E}[X]^{3/2}$$

Grâce à la remarque précédente, on peut écrire que

$$\mathbb{E} \left[C_{i,I-1}^{3/2} | C_{i,1}, \dots, C_{i,I+1-i} \right] = (1 + \text{CoV} [C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}]^2)^{3/8} \mathbb{E} [C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}]^{3/2}. \quad (5.5)$$

En combinant les Equations (5.4) et (5.5), il vient que

$$\begin{aligned} SK(C_{i,I} | \mathcal{D}_I) &= S k_{I-1}^3 (1 + \text{CoV} [C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}]^2)^{3/8} \mathbb{E} [C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}]^{3/2} \\ &\quad + f_{I-1}^3 SK(C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}) + 3 f_{I-1} \sigma_{I-1}^2 \text{Var}(C_{i,I-1} | C_{i,1}, \dots, C_{i,I+1-i}). \end{aligned} \quad (5.6)$$

Le skewness peut se calculer de manière récurrente via la formule suivante :

$$SK(C_{i,k+1}) = \hat{S} k_k^3 \left(1 + \frac{\text{Var}(\hat{C}_{i,k})}{\hat{C}_{i,k}^2} \right)^{3/8} \hat{C}_{i,k}^{3/2} + \hat{f}_k^3 SK(C_{i,k}) + 3 \hat{f}_k \sigma_k^2 \text{Var}(\hat{C}_{i,k})$$

où

$$\text{Var}(\hat{C}_{i,k}) = \hat{C}_{i,k-1} \hat{\sigma}_{k-1}^2 \left(1 + \frac{\hat{C}_{i,k-1}}{I-k+1} \sum_{j=1} C_{j,k-1} \right).$$

5.3.1.2 Calcul du terme (2)

On peut réécrire le terme (2) comme

$$\left(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I} \right)^3 = C_{i,I+1-i}^3 \left(f_{I+1-i} \dots f_{I-1} - \hat{f}_{I+1-i} \dots \hat{f}_{I-1} \right)^3.$$

Comme dans certaines démonstrations de Mack [36], on pose

$$F = f_{I+1-i} \dots f_{I-1} - \hat{f}_{I+1-i} \dots \hat{f}_{I-1} = S_{I+1-i} + \dots + S_{I-1} \text{ avec } S_k = \hat{f}_{I+1-i} \dots \hat{f}_{k-1} (f_k - \hat{f}_k) f_{k+1} \dots f_{I-1}.$$

Il vient alors que $F^3 = \sum_{k=I+1-i}^{I-1} S_k^3 + 3 \sum_k \sum_{j < k} S_j^2 S_k + 6 \sum_k \sum_l \sum_{j < k < l} S_j S_k S_l$.

Or, conditionnellement à $B_k = \{C_{i,j}, j \leq k, i + j \leq I + 1\}$, \hat{f}_k est un estimateur non-biaisé de f_k . D'où

$$\mathbb{E}[S_j^2 S_k | B_k] = 0 \text{ et } \mathbb{E}[S_j S_k S_l | B_k] = 0.$$

Par ailleurs, d'après (H1), les paiements cumulés d'années de survénance différentes sont indépendants. Donc

$$\mathbb{E} \left[\left(f_k - \hat{f}_k \right)^3 \middle| B_k \right] = SK(\hat{f}_k | B_k) = \frac{\sum_{j=1}^{I-k} SK(C_{j,k+1} | B_k)}{\left(\sum_{j=1}^{I-k} C_{j,k} \right)^3} = SK_k^3 \frac{\sum_{j=1}^{I-k} C_{j,k}^{3/2}}{\left(\sum_{j=1}^{I-k} C_{j,k} \right)^3}.$$

Puis

$$\begin{aligned} \mathbb{E} [S_k^3 | B_k] &= \hat{f}_{I+1-i}^3 \cdots \hat{f}_{k-1}^3 f_{k+1}^3 \cdots f_{I-1}^3 \mathbb{E} \left[\left(f_k - \hat{f}_k \right)^3 \middle| B_k \right] \\ &= \hat{f}_{I+1-i}^3 \cdots \hat{f}_{k-1}^3 \frac{\hat{f}_k^3}{\hat{f}_k^3} f_{k+1}^3 \cdots f_{I-1}^3 \mathbb{E} \left[\left(f_k - \hat{f}_k \right)^3 \middle| B_k \right]. \end{aligned}$$

Ainsi, on obtient finalement que

$$\left(\mathbb{E}[C_{i,I} | \mathcal{D}_I] - \hat{C}_{i,I} \right)^3 = C_{i,I}^3 \sum_{k=I+1-i}^{I-1} \left[\frac{\hat{S}k_k^3 \sum_{j=1}^{I-k} C_{j,k}^{3/2}}{\hat{f}_k^3 \left(\sum_{j=1}^{I-k} C_{j,k} \right)^3} \right]. \quad (5.7)$$

5.3.2 Estimation du skewness total

Pour simuler des corrélations entre les montants de réserves estimés \hat{R}_i , il est possible d'appliquer une structure de copule gaussienne⁴. Par cohérence avec les hypothèses émises par Dal Moro dans ses calculs de termes d'erreur, nous nous intéressons en particulier aux propriétés de la copule gaussienne pour des variables log-normales.

Lemme 5.3.6. Soit $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$.

On pose $X_1 = \exp(Y_1)$ et $X_2 = \exp(Y_2)$. Alors

$$\text{Cov}(X_1^2, X_2) = \mathbb{E}[X_1] \left(1 + \frac{\text{Var}(X_1)}{\mathbb{E}[X_1]^2} \right) \text{Cov}(X_1, X_2) \left(2 + \frac{\text{Cov}(X_1, X_2)}{\mathbb{E}[X_1]\mathbb{E}[X_2]} \right).$$

Lemme 5.3.7. Soit $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \right)$.

4. Ce point sera plus détaillé dans l'approche simulateur.

On pose $X_1 = \exp(Y_1)$, $X_2 = \exp(Y_2)$ et $X_3 = \exp(Y_3)$. Alors

$$\mathbb{E}[X_1 X_2 X_3] = \mathbb{E}[X_1] \mathbb{E}[X_2] \mathbb{E}[X_3] \left(1 + \frac{\text{Cov}(X_1, X_2)}{\mathbb{E}[X_1] \mathbb{E}[X_2]}\right) \left(1 + \frac{\text{Cov}(X_1, X_3)}{\mathbb{E}[X_1] \mathbb{E}[X_3]}\right) \left(1 + \frac{\text{Cov}(X_2, X_3)}{\mathbb{E}[X_2] \mathbb{E}[X_3]}\right).$$

Les deux lemmes précédents permettent d'établir une formule de skewness total en tenant compte de 3 années de survénance. Celle-ci peut être généralisée pour prendre en compte plus de trois années. Le résultat obtenu est le suivant.

$$\begin{aligned} SK \left(\sum_{i=2}^I \hat{C}_{i,I} \right) &= \sum_{i=2}^I SK(\hat{C}_{i,I}) + 3 \sum_{i=2}^{I-1} \sum_{j=i+1}^I \rho_{ij}^2 \text{Var}(\hat{C}_{i,I}) \text{Var}(\hat{C}_{j,I}) \left[\frac{\mathbb{E}[\hat{C}_{i,I}] + \mathbb{E}[\hat{C}_{j,I}]}{\mathbb{E}[\hat{C}_{i,I}] \mathbb{E}[\hat{C}_{j,I}]} \right] \\ &+ 3 \sum_{i=2}^{I-1} \sum_{j=i+1}^I \rho_{ij} \sqrt{\text{Var}(\hat{C}_{i,I}) \text{Var}(\hat{C}_{j,I})} \left[\frac{\text{Var}(\hat{C}_{i,I}) + \text{Var}(\hat{C}_{j,I})}{\mathbb{E}[\hat{C}_{i,I}] + \mathbb{E}[\hat{C}_{j,I}]} \right] \left[2 + \rho_{i,j} \frac{\sqrt{\text{Var}(\hat{C}_{i,I}) \text{Var}(\hat{C}_{j,I})}}{\mathbb{E}[\hat{C}_{i,I}] \mathbb{E}[\hat{C}_{j,I}]} \right] \\ &+ 6 \sum_{i=2}^{I-2} \sum_{j=i+1}^{I-1} \sum_{k=j+1}^I \rho_{ij} \rho_{ik} \rho_{jk} \frac{\text{Var}(\hat{C}_{i,I}) \text{Var}(\hat{C}_{j,I}) \text{Var}(\hat{C}_{k,I})}{\mathbb{E}[\hat{C}_{i,I}] \mathbb{E}[\hat{C}_{j,I}] \mathbb{E}[\hat{C}_{k,I}]} \\ &+ 6 \sum_{i=2}^{I-2} \sum_{j=i+1}^{I-1} \sum_{k=j+1}^I \rho_{ij} \rho_{ik} \rho_{jk} \sqrt{\text{Var}(\hat{C}_{i,I}) \text{Var}(\hat{C}_{j,I}) \text{Var}(\hat{C}_{k,I})} \left(\frac{\sqrt{\text{Var}(\hat{C}_{i,I})}}{\rho_{jk} \mathbb{E}[\hat{C}_{i,I}]} + \frac{\sqrt{\text{Var}(\hat{C}_{j,I})}}{\rho_{ik} \mathbb{E}[\hat{C}_{j,I}]} + \frac{\sqrt{\text{Var}(\hat{C}_{k,I})}}{\rho_{ij} \mathbb{E}[\hat{C}_{k,I}]} \right) \end{aligned} \quad (5.8)$$

où $\mathbb{E}[\hat{C}_{i,I}] = C_{i,I+1-i} \hat{f}_{I+1-i} \dots \hat{f}_{I-1}$, $\text{Var}(\hat{C}_{i,I}) := \widehat{\text{MSE}}(\hat{R}_i)$ et $\rho_{i,j}$ est tel que défini dans l'Equation (5.9). Rappelons toutefois qu'il nous faut enfin diviser les valeurs de $SK(\hat{C}_{i,I})$ obtenues par $\text{Var}(\hat{C}_{i,I})^{3/2}$ pour obtenir le skewness $Skew(\hat{C}_{i,I})$.

5.3.3 Estimation du kurtosis d'une année de survénance et du kurtosis total

Définition 5.3.8. *Le kurtosis de l'année de survénance i de l'estimateur du paiement cumulé ultime $\hat{C}_{i,I}$ de $C_{i,I}$ s'écrit*

$$KT(\hat{C}_{i,I}) = \mathbb{E}[(C_{i,I} - \hat{C}_{i,I})^4 | \mathcal{D}_I].$$

L'article de Dal Moro [15] ne propose pas de formule fermée ni pour le kurtosis d'une année de survénance, ni pour le kurtosis total. Ces estimations n'étant pas au cœur de l'objet de ce mémoire, nous nous contenterons des résultats obtenus par méthode de simulation. Toutefois, le mémoire d'actuariat d'Amoussa [2] présente un calcul de kurtosis par année de survénance.

Remarque 5.3.9. *Une démonstration analogue à celle de la Remarque 1.1.4. pourrait être faite pour comprendre que l'erreur faite sur le paiement cumulé ultime $C_{i,I}$ correspond à l'erreur commise sur le montant des réserves R_i .*

5.4 Approche par simulation

Afin de déterminer le skewness et kurtosis pour chaque année de survénance, Dal Moro [14] [15] propose deux lois de distribution pertinentes pour estimer les paiements cumulés $C_{i,k}$:

- la loi de Pareto généralisée (GPD),
- et la loi d'extremum généralisée (GEV).

L'idée derrière la méthode de simulation est de générer les paiements cumulés de proche en proche, puis d'estimer à chaque étape les paramètres de la loi choisie sous-jacente. La Figure 5.3 schématise le procédé utilisé dans l'algorithme.

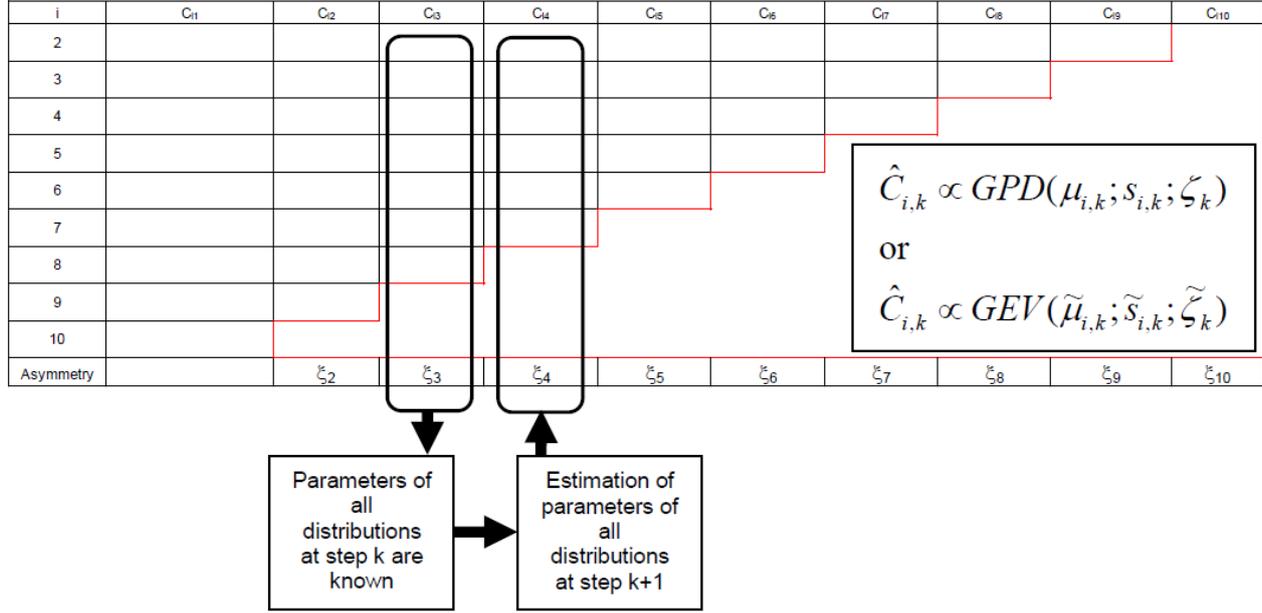


FIGURE 5.3 – Schéma de l'approche simulateur présentée par Dal Moro [14]

De proche en proche, ces modèles permettent de générer des paiements cumulés à l'ultime et par conséquent d'obtenir une distribution du montant des réserves. Les quatre premiers moments de cette dernière sont calculés de manière empirique. Le processus de simulation étant le même pour les deux types de distribution, nous détaillerons uniquement les différentes étapes pour le modèle GPD. Le processus relatif à la loi GEV est présenté en annexe B.1.

5.4.1 Modèle GPD

Définition 5.4.1. Soit X une variable aléatoire. X suit une **loi de Pareto généralisée** de paramètres $\mu \in \mathbb{R}$, $s > 0$ et $\xi \in \mathbb{R}$, i.e $X \sim GPD(\mu, s, \xi)$, si sa densité est telle que

$$f_X(x) = \begin{cases} \frac{1}{s} \left(1 + \xi \frac{x - \mu}{s} \right)^{-\frac{1}{\xi} - 1} & \text{si } \xi \neq 0 \\ \frac{1}{s} \exp\left(-\frac{x - \mu}{s}\right) & \text{si } \xi = 0 \end{cases}$$

où μ est le paramètre de position, s le paramètre d'échelle et ξ le paramètre de forme.

En considérant $\hat{C}_{i,k} \sim GPD(\mu_{i,k}, s_{i,k}, \xi_k)$, on peut exprimer les moments de $\hat{C}_{i,k}$ en fonction des paramètres :

$$\begin{aligned} \mathbb{E}[\hat{C}_{i,k}] &= \mu_{i,k} + \frac{s_{i,k}}{1 - \xi_k} \\ \text{Var}(\hat{C}_{i,k}) &= \frac{s_{i,k}^2}{(1 - \xi_k)^2(1 - 2\xi_k)} \\ \text{Skew}(\hat{C}_{i,k}) &= \frac{SK(\hat{C}_{i,k})}{\text{Var}(\hat{C}_{i,k})^{3/2}} = \frac{2(1 + \xi_k)\sqrt{1 - 2\xi_k}}{1 - 3\xi_k} \\ \text{Kurt}(\hat{C}_{i,k}) &= \frac{KT(\hat{C}_{i,k})}{\text{Var}(\hat{C}_{i,k})^2} = \frac{3(3 - 5\xi_k - 4\xi_k^3)}{1 - 7\xi_k + 12\xi_k^2}. \end{aligned}$$

Pour chaque année de développement k , on procède comme ceci :

Etape 1 Génération du paiement cumulé $\hat{C}_{i,k}$.

- i. On suppose les paramètres $(\mu_{i,k}, s_{i,k}, \xi_k)$ connus.
- ii. On génère une variable aléatoire $U_{i,k}$ de loi uniforme sur $(0, 1]$.
- iii. On détermine $\hat{C}_{i,k} = \mu_{i,k} + \frac{s_{i,k} (U_{i,k}^{-\xi_k} - 1)}{\xi_k}$.

Etape 2 Obtention des paramètres $(\mu_{i,k}, s_{i,k}, \xi_k)$

- i. On détermine le paramètre ξ_{k+1} :
 - Soit via la formule du skewness, en se rappelant que les hypothèses (H3) et (H5) donnent

$$\frac{\hat{S}k_{k+1}^3}{\hat{\sigma}_{k+1}^3} = \frac{2(1 + \xi_{k+1})\sqrt{1 - 2\xi_{k+1}}}{1 - 3\xi_{k+1}}.$$

- Soit via la formule du kurtosis, en se rappelant que les hypothèses (H3) et (H7) donnent

$$\frac{\hat{K}t_{k+1}^4}{\hat{\sigma}_{k+1}^4} = \frac{3(3 - 5\xi_{k+1} - 4\xi_{k+1}^3)}{1 - 7\xi_{k+1} + 12\xi_{k+1}^2}.$$

- ii. Puisqu'il résulte de la formule d'erreur quadratique moyenne donnée par Mack (Corollaire 1.1.8.) que

$$\text{Var}(\hat{C}_{i,k+1}) = \hat{C}_{i,k} \hat{\sigma}_k^2 \left(1 + \frac{\hat{C}_{i,k}}{I-k} \sum_{j=1}^{I-k} C_{j,k} \right),$$

la résolution de l'équation $\text{Var}(\hat{C}_{i,k+1}) = \frac{s_{i,k+1}^2}{(1 - \xi_{k+1})^2(1 - 2\xi_{k+1})}$ permet d'obtenir le paramètre $s_{i,k+1}$.

- iii. Enfin, pour obtenir le paramètre $\mu_{i,k+1}$ il faut résoudre l'équation

$$\mathbb{E}[\hat{C}_{i,k+1}] = \hat{f}_k \hat{C}_{i,k} = \mu_{i,k+1} + \frac{s_{i,k+1}}{1 - \xi_{k+1}}.$$

5.4.2 Prise en compte des corrélations

Afin de tenir compte des corrélations entre les paiements cumulés, Dal Moro choisit de donner une structure de **copule gaussienne** aux lois uniformes qu'il simule⁵. Cette dernière est indépendante de la loi choisie pour simuler les paiements cumulés. Dans son article [14], l'auteur préconise l'utilisation des coefficients de corrélations provenant de l'expression de l'erreur quadratique moyenne donnée par Mack [36] :

$$\widehat{\text{MSE}}(\hat{R}) = \sum_{i=2}^I \left\{ \widehat{\text{MSE}}(\hat{R}_i) + 2 \sum_{i < j} \rho_{i,j} \sqrt{\widehat{\text{MSE}}(\hat{R}_i) \widehat{\text{MSE}}(\hat{R}_j)} \right\}$$

5. Des rappels sur la théorie des copules sont présentés en annexe B.2.

$$\hat{C}_{i,I} \hat{C}_{j,I} \cdot \frac{\sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2 / \hat{f}_k^2}{I-k}}{\sum_{n=1}^{I-1} C_{n,k}}$$

avec $\rho_{i,j} = \frac{\hat{C}_{i,I} \hat{C}_{j,I} \cdot \frac{\sum_{k=I+1-i}^{I-1} \frac{\hat{\sigma}_k^2 / \hat{f}_k^2}{I-k}}{\sum_{n=1}^{I-1} C_{n,k}}}{\sqrt{\text{MSE}(\hat{R}_i) \text{MSE}(\hat{R}_j)}}$ (5.9)

Or, ces coefficients de corrélation sont valables pour la dernière année de survénance. Pour modéliser le fait que, dans l'algorithme, la vision n'est pas à l'ultime, Dal Moro corrige les coefficients de corrélation et utilise :

$$\tilde{\rho}_{i,j} = \rho_{i,j} \prod_{k=I-i+1}^{I-1} \hat{f}_k^2 \text{ pour } i < j$$

A partir des coefficients $\tilde{\rho}_{i,j}$, on peut construire la matrice de corrélation Σ qui permettra de simuler un échantillon X de loi $\mathcal{N}(0_m, \Sigma)$ et de générer les lois uniformes $U = \Phi(X)$.

5.5 Approche par bootstrap

Cette méthode, développée par nos soins, vise à étendre la théorie relative au bootstrap pour le modèle de Mack *chain-ladder* afin de tenir compte d'hypothèses supplémentaires sur la forme de la distribution.

Nous cherchons alors à exprimer les résidus en fonction des valeurs de skewness et de kurtosis par année de survénance. Pour ce faire, on utilise des polynômes de Fleishman [19].

5.5.1 Polynômes de Fleishman

Soit Y une variable aléatoire centrée réduite. L'objectif est de décomposer Y sous la forme d'un polynôme en X de degré 3, où X suit une loi normale centrée réduite :

$$Y = a + bX + cX^2 + dX^3$$

Il s'agit de ce que Fleishman appelle la *power method*.

Les coefficients (a, b, c, d) vont ensuite être déterminés par résolution d'un système d'équations en fonction des moments de Y . Puisque $X \sim \mathcal{N}(0, 1)$, il est utile de rappeler que, pour $k \in \mathbb{N}$:

$$\begin{cases} \mathbb{E}[X^{2k+1}] = 0 \\ \mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k k!} \end{cases}$$

5.5.1.1 Equation de $\mathbb{E}[Y]$

Puisque Y est centrée, $\mathbb{E}[Y] = 0$.

Or $\mathbb{E}[Y] = a + b \cdot \mathbb{E}[X] + c \cdot \mathbb{E}[X^2] + d \cdot \mathbb{E}[X^3]$, d'où l'équation

$$a + c = 0.$$

5.5.1.2 Equation de $\text{Var}(Y)$

Puisque Y est centrée réduite, $\text{Var}(Y) = \mathbb{E}[Y^2] = 1$. On a de plus :

$$\begin{aligned}\mathbb{E}[Y^2] &= a^2 + 2ab \cdot \mathbb{E}[X] + (2ab + b^2) \cdot \mathbb{E}[X^2] + 2(ad + bc) \cdot \mathbb{E}[X^3] \\ &\quad + (2bd + c^2) \cdot \mathbb{E}[X^4] + 2cd \cdot \mathbb{E}[X^5] + d^2 \cdot \mathbb{E}[X^6] \\ &= a^2 + 2ac + b^2 + 6bd + 3c^2 + 15d^2 \text{ en remplaçant } \mathbb{E}[X^k] \text{ par sa valeur, pour } 1 \leq k \leq 6.\end{aligned}$$

Or $a = -c$. On obtient donc l'équation

$$b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0.$$

5.5.1.3 Equation de $\text{Skew}(Y)$

Pour rappel,

$$\gamma_1 = \text{Skew}(Y) = \frac{\mathbb{E}[(Y - \mathbb{E}[Y])^3]}{\text{Var}(Y)^{3/2}}.$$

Puisque Y est centrée-réduite, calculer le skewness de Y revient à calculer $\mathbb{E}[Y^3]$. Après calculs, on a finalement, puisque $a = -c$ que

$$2c(b^2 + 24bd + 105d^2 + 2) - \gamma_1 = 0.$$

5.5.1.4 Equation de $\text{Kurt}(Y)$

Enfin, on rappelle que

$$\beta_2 = \text{Kurt}(Y) = \frac{\mathbb{E}[(Y - \mathbb{E}[Y])^4]}{\text{Var}(Y)^2} \text{ et que } \gamma_2 = \text{Kurt}(X) - 3.$$

Là encore, déterminer le kurtosis de Y revient à calculer $\mathbb{E}[Y^4]$. Finalement, comme $a = -c$, il vient que

$$24(bd + c^2[1 + b^2 + 28bd] + d^2[12 + 48bd + 141c^2 + 225d^2]) - \gamma_2 = 0.$$

5.5.2 Application à l'algorithme de bootstrap

Revenons à notre méthode de bootstrap récursive. On sait que, de manière générale, pour une année de survenance i et une année de développement j données, les résidus $r_{i,j}$ supposés gaussiens centrés réduits sont donnés par la formule suivante :

$$r_{i,j} = \sqrt{\frac{C_{i,j}}{\hat{\sigma}_j^2}} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)$$

Grâce aux polynômes de Fleishman que nous avons introduits précédemment, nous pouvons définir de nouveaux résidus $\tilde{r}_{i,j}$ vérifiant

$$\sqrt{\frac{C_{i,j}}{\hat{\sigma}_j^2}} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right) = a + b\tilde{r}_{i,j} + c\tilde{r}_{i,j}^2 + d\tilde{r}_{i,j}^3. \quad (5.10)$$

Puisque \hat{f}_j $\hat{\sigma}_j^2$ sont non-biaisés, la variable $Y_{i,j+1} := \sqrt{\frac{C_{i,j}}{\hat{\sigma}_j^2}} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right) = \frac{C_{i,j+1} - \hat{f}_j C_{i,j}}{\sqrt{C_{i,j} \hat{\sigma}_j^2}}$

vérifie :

$$\left\{ \begin{array}{l} \mathbb{E}[Y_{i,j+1}|C_{i,j}] = \frac{1}{\sqrt{C_{i,j} \hat{\sigma}_j^2}} \mathbb{E} [C_{i,j+1} - \hat{f}_j C_{i,j} | C_{i,j}] = 0 \\ \mathbb{E}[Y_{i,j+1}^2|C_{i,j}] = \frac{1}{C_{i,j} \hat{\sigma}_j^2} \mathbb{E} [(C_{i,j+1} - \mathbb{E}[C_{i,j+1}|C_{i,j}])^2 | C_{i,j}] = \frac{\text{Var}(C_{i,j+1}|C_{i,j})}{C_{i,j} \hat{\sigma}_j^2} = 1 \\ \mathbb{E}[Y_{i,j+1}^3|C_{i,j}] = \frac{1}{(C_{i,j} \hat{\sigma}_j^2)^{3/2}} \mathbb{E} [(C_{i,j+1} - \mathbb{E}[C_{i,j+1}|C_{i,j}])^3 | C_{i,j}] = \frac{SK(C_{i,j+1}|C_{i,j})}{(C_{i,j} \hat{\sigma}_j^2)^{3/2}} = \frac{\hat{S}k_j^3}{\hat{\sigma}_j^{3/2}} \\ \mathbb{E}[Y_{i,j+1}^4|C_{i,j}] = \frac{1}{(C_{i,j} \hat{\sigma}_j^2)^2} \mathbb{E} [(C_{i,j+1} - \mathbb{E}[C_{i,j+1}|C_{i,j}])^4 | C_{i,j}] = \frac{KT(C_{i,j+1}|C_{i,j})}{(C_{i,j} \hat{\sigma}_j^2)^2} = \frac{\hat{K}t_j^4}{\hat{\sigma}_j^4} \end{array} \right.$$

Par conséquent, on peut résoudre les équations relatives au polynôme de Fleishman. La connaissance des coefficients (a, b, c, d) nous permet de tirer avec remise de nouveaux résidus

$$r_{i,j} = a + b\tilde{r}_{i,j} + c\tilde{r}_{i,j}^2 + d\tilde{r}_{i,j}^3 ,$$

tenant compte de l'asymétrie et de l'aplatissement de la distribution des réserves. Pour cela, il suffit de résoudre l'Equation (5.10) pour déterminer les valeurs possibles de $\tilde{r}_{i,j}$.

Chapitre 6

Application des méthodes de provisionnement précédentes

Dans ce chapitre, nous appliquons les méthodes de provisionnement présentées au préalable sur un exemple. Ce procédé nous permettra de comparer les résultats des différents modèles avant de les appliquer sur nos données agricoles.

6.1 Données de triangle

6.1.1 Présentation des données

Nous souhaitons appliquer les différentes méthodes sur l'un des triangles de provisionnement présenté dans l'article de 1993 de Mack [36] et étudié dans un papier de Taylor et Ashe [56]. La branche à laquelle se rapporte ce triangle n'est en revanche pas connue. Ce choix nous permettra de comparer les éventuels écarts entre les résultats présentés dans les articles de recherche et ceux produits à partir de nos propres implémentations.

i	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$	$C_{i,4}$	$C_{i,5}$	$C_{i,6}$	$C_{i,7}$	$C_{i,8}$	$C_{i,9}$	$C_{i,10}$
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315		
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268			
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311				
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712					
7	440 832	1 288 463	2 419 861	3 483 130						
8	359 480	1 421 128	2 864 498							
9	376 686	1 363 294								
10	344 014									

TABLE 6.1 – Exemple de triangle de liquidation [36]

La Figure 6.1 ci-après représente, pour chaque année de survenance, l'évolution des paiements cumulés en fonction des années de développement.

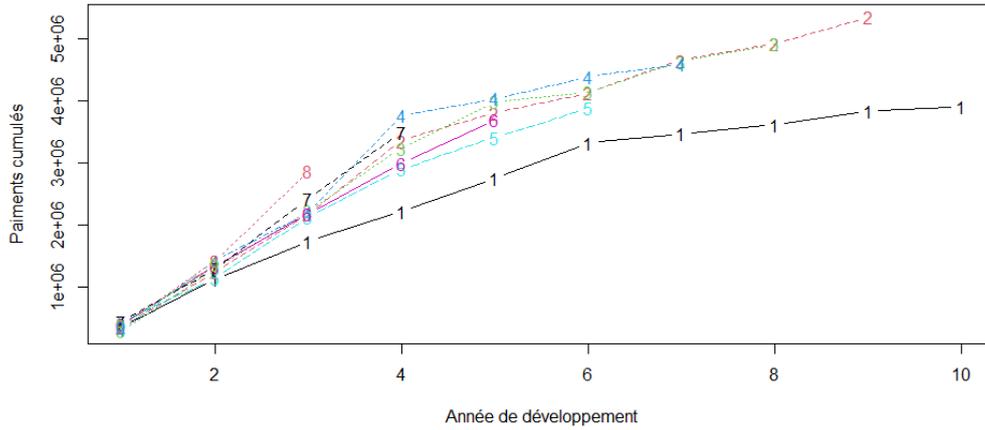


FIGURE 6.1 – Représentation graphique du triangle de Mack

Sans pour autant connaître la branche considérée, nous constatons que cette dernière possède un développement des sinistres plutôt long. En effet, les paiements cumulés ne se stabilisent pas lors des premières années de développement. En outre, ces derniers suivent une tendance strictement croissante, aucun incrément de paiement négatif n'est à noter.

6.1.2 Vérification des hypothèses de Mack *chain-ladder*

Bien que ce triangle soit souvent utilisé dans les exemples d'articles de recherche de provisionnement, nous vérifions tout de même les hypothèses relatives au modèle de Mack *chain-ladder*.

Hypothèse 1 Les paiements cumulés $C_{i,j}$ de différentes années de survenance sont *indépendants*, autrement dit

$$\{C_{i,1}, \dots, C_{i,I}\}, \{C_{k,1}, \dots, C_{k,I}\}, i \neq k \text{ sont indépendants.} \quad (H1)$$

Afin de vérifier l'indépendance des années de survenance, il suffit de représenter graphiquement l'évolution des facteurs de développement individuels $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ en fonction des années de développement j .

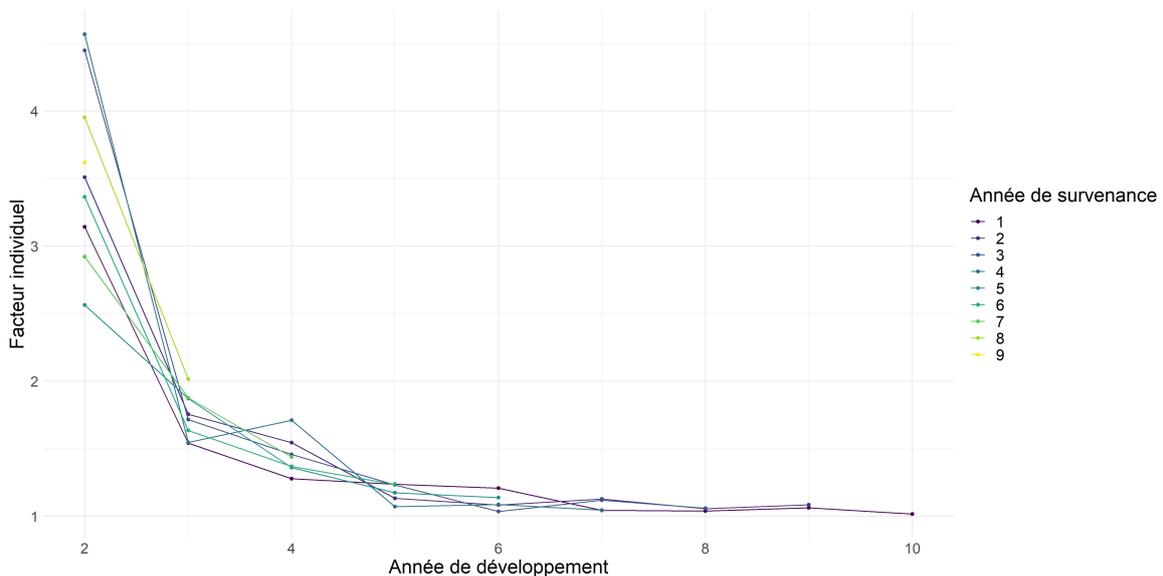


FIGURE 6.2 – Vérification de l'Hypothèse 1

Quelle que soit l'année de survenance, les courbes des facteurs de développement individuels de la Figure 6.2 présentent une tendance commune. Par conséquent, l'**Hypothèse 1** est vérifiée.

Hypothèse 2 Il existe des **facteurs de développement** $f_j > 0$, pour $1 \leq j \leq I - 1$ tels que

$$\mathbb{E}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = f_j C_{i,j}, \text{ pour } 1 \leq i \leq I. \quad (\text{H2})$$

Cette hypothèse se teste en traçant, pour chaque année de développement j , $1 \leq j \leq I - 1$ les points de coordonnées $(C_{i,j}, C_{i,j+1})$. Le graphique résultant mettra en lumière l'existence ou non d'un facteur de proportionnalité.

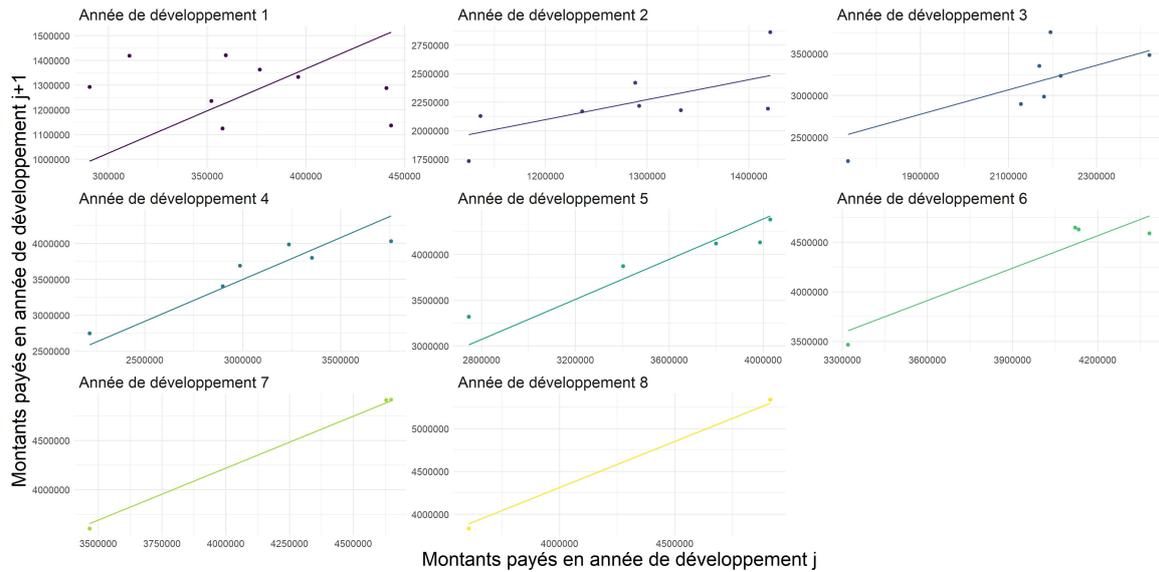


FIGURE 6.3 – Vérification de l'**Hypothèse 2**

L'existence d'un facteur de proportionnalité ne semble pas claire pour les trois premières années de développement mais se précise pour les années suivantes. Nous validons donc malgré tout l'**Hypothèse 2**.

Hypothèse 3 Il existe des coefficients $\sigma_k > 0$ pour $1 \leq j \leq I - 1$ tels que

$$\text{Var}(C_{i,j+1}|C_{i,1}, \dots, C_{i,j}) = \sigma_j^2 C_{i,j}, \text{ pour } 1 \leq i \leq I. \quad (\text{H3})$$

Dans l'article de Wüthrich, Merz et Lysenko de 2008 [64], cette hypothèse peut se réécrire comme

$$\forall 1 \leq j \leq I - 1, \forall 1 \leq i \leq I, C_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}} \epsilon_{i,j}$$

où, en considérant $\mathcal{B}_1 = \{C_{1,1}, \dots, C_{I,1}\}$ avec $C_{i,1} > 0$, les variables aléatoires $\epsilon_{i,j}$ sont indépendantes et vérifient

$$\forall 1 \leq j \leq I - 1, \forall 1 \leq i \leq I, \mathbb{E}[\epsilon_{i,j}|\mathcal{B}_1] = 0, \mathbb{E}[\epsilon_{i,j}^2|\mathcal{B}_1] = 1 \text{ et } \mathbb{P}(C_{i,j} > 0|\mathcal{B}_1) = 1.$$

Nous allons en particulier nous intéresser à la vérification de l'indépendance des résidus $r_{i,j}$. Pour ce faire, nous calculons les $r_{i,j}$ en inversant la formule précédente puis nous les représentons graphiquement.

Une structure aléatoire des résidus semble se dessiner sur la Figure 6.4 suivante. Ainsi, leur non-corrélation satisfait l'hypothèse d'indépendance sous-jacente à l'**Hypothèse 3**.

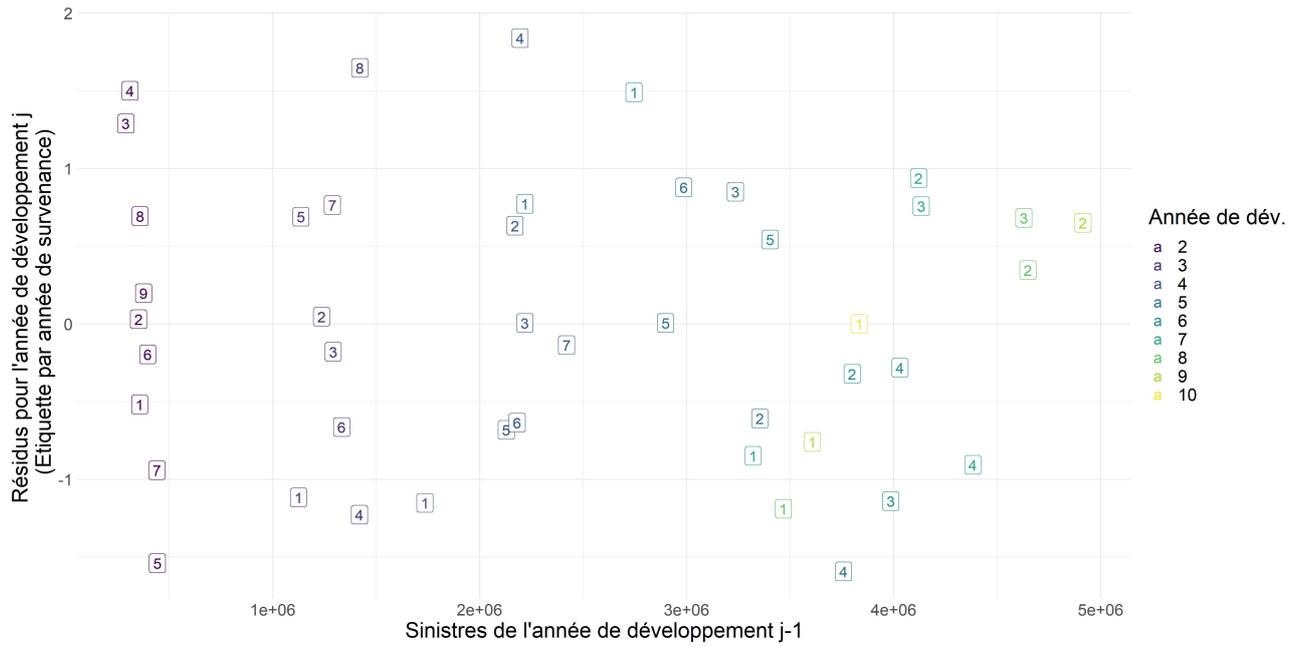


FIGURE 6.4 – Vérification de l'indépendance des résidus dans l'Hypothèse 3

6.1.3 Vérification des hypothèses de Dal Moro

D'après le chapitre précédent, les hypothèses (H5) et (H7), relatives à l'existence des facteurs Sk_j^3 et Kt_j^4 résultent de la validité des hypothèses (H3), (H4) et (H6). L'hypothèse (H3) ayant été vérifiée précédemment, il suffit donc de procéder à la vérification de ces deux dernières conditions uniquement.

Hypothèse 4 *Le skewness du paiement cumulé $C_{i,j+1}$ conditionnel aux paiements passés ne dépend pas de l'année de survénance i mais de l'année de développement j , soit*

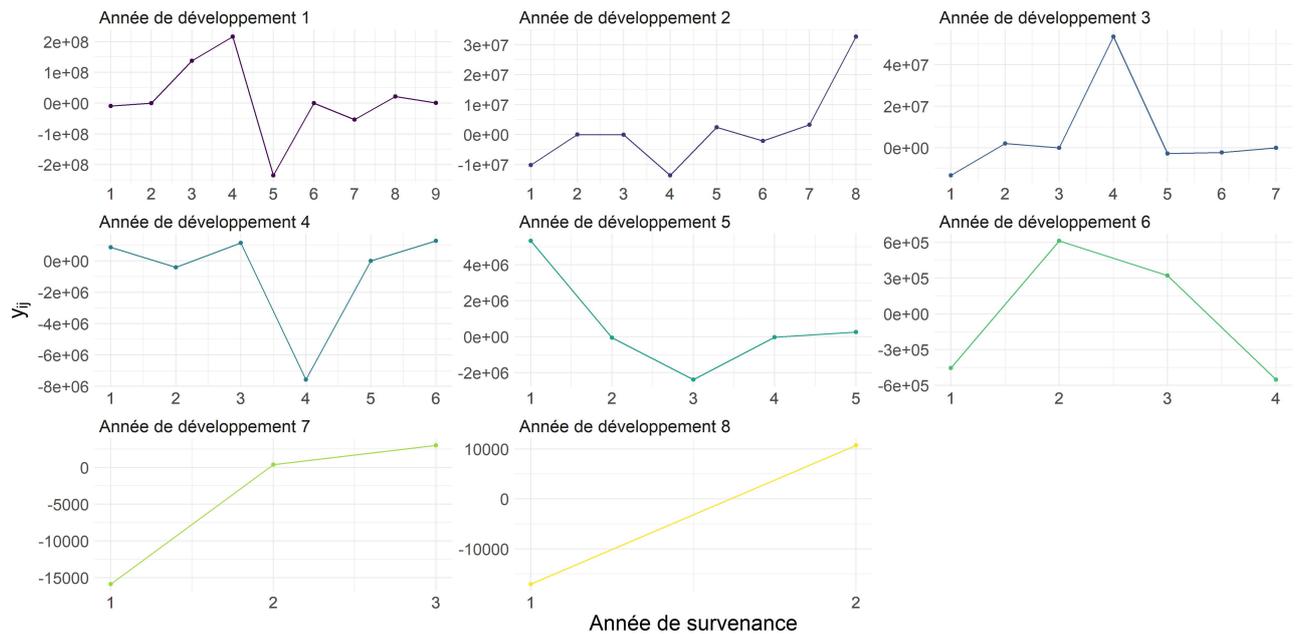
$$Skew(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \frac{SK(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))}{\text{Var}(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))^{3/2}} \text{ dépend de } j \text{ mais pas de } i. \quad (H4)$$


FIGURE 6.5 – Vérification de l'Hypothèse 4

Pour la justifier, il est nécessaire de tracer les courbes $y_{i,j} = \frac{(C_{i,j+1} - \hat{f}_j C_{i,j})^3}{C_{i,j}^{3/2}}$ pour chaque année de développement j en fonction de l'année de survenance i , comme sur la Figure 6.5 précédente.

Aucune tendance en fonction de l'année de survenance i n'étant observée, l'**Hypothèse 4** est valide.

Hypothèse 6 *Le kurtosis du paiement cumulé $C_{i,j+1}$ conditionnel aux paiements passés ne dépend pas de l'année de survenance i mais de l'année de développement j , soit*

$$Kurt(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j})) = \frac{KT(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))}{\text{Var}(C_{i,j+1}|(C_{i,1}, \dots, C_{i,j}))^2} \text{ dépend de } j \text{ mais pas de } i. \quad (\text{H6})$$

De manière analogue à la vérification de l'**Hypothèse 4**, il est nécessaire de tracer les coefficients $y_{i,j} = \frac{(C_{i,j+1} - \hat{f}_j C_{i,j})^4}{C_{i,j}^2}$ pour chaque année de développement j en fonction de l'année de survenance i .

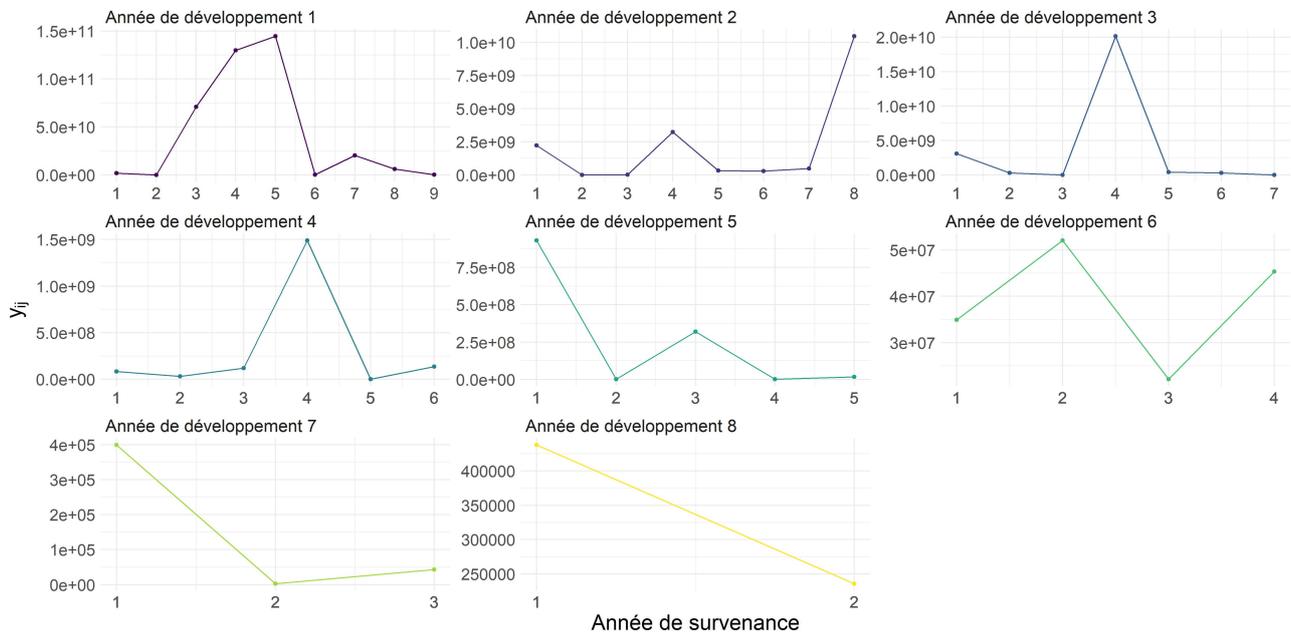


FIGURE 6.6 – Vérification de l'**Hypothèse 6**

Puisqu'aucune tendance en fonction de l'année de survenance i n'est observée sur la Figure 6.6 précédente, donc l'**Hypothèse 6** est valide.

6.2 Estimation du montant des réserves et du risque de réserves

Dans cette partie, nous nous intéressons aux deux premiers moments de la distribution des réserves. Pour évaluer leur montant et estimer le risque qui y est associé, nous disposons de plusieurs méthodes :

- des méthodes classiques, telles que les modèles de Mack *chain-ladder* et de Wüthrich, Merz et Lysenko, ainsi que le bootstrap récursif ;
- et des méthodes tenant compte des hypothèses de Dal Moro, telles que le bootstrap incluant des paramètres de skewness et de kurtosis et l'approche simulateur de Dal Moro,

que nous appliquons avec une loi de Pareto généralisée (GPD).

6.2.1 Application des méthodes usuelles

Rappelons tout d'abord que le modèle développé par Wüthrich, Merz et Lysenko s'appuie sur la méthode de Mack *chain-ladder* pour déterminer le montant des réserves. La Table 6.2 suivante présente alors les résultats d'estimations de montants de réserves à partir des modèles de Mack *chain-ladder*, notées \hat{R}_i^M , et de bootstrap récursif, lancé avec 1000 simulations, notées \hat{R}_i^B .

i	2	3	4	5	6	7	8	9	10	Total
\hat{R}_i^M	94 634	469 511	709 638	984 889	1 419 459	2 177 641	3 920 301	4 278 972	4 625 811	18 680 856
\hat{R}_i^B	94 640	468 699	712 665	998 635	1 438 910	2 207 862	3 950 875	4 335 591	4 736 241	18 944 118
Ecart	-0,01%	0,17%	-0,43%	-1,40%	-1,37%	-1,39%	-0,78%	-1,32%	-2,39%	-1,41%

TABLE 6.2 – Montants des réserves estimés

Nous constatons que les résultats des deux méthodes sont très proches, que ce soit pour les réserves par année de survenance ou totales. Le bootstrap permet de diminuer très légèrement le montant des réserves estimé.

Calculons désormais l'erreur associée à l'estimation des réserves pour les modèles de Mack *chain-ladder*, de Wüthrich, Merz et Lysenko et le bootstrap récursif.

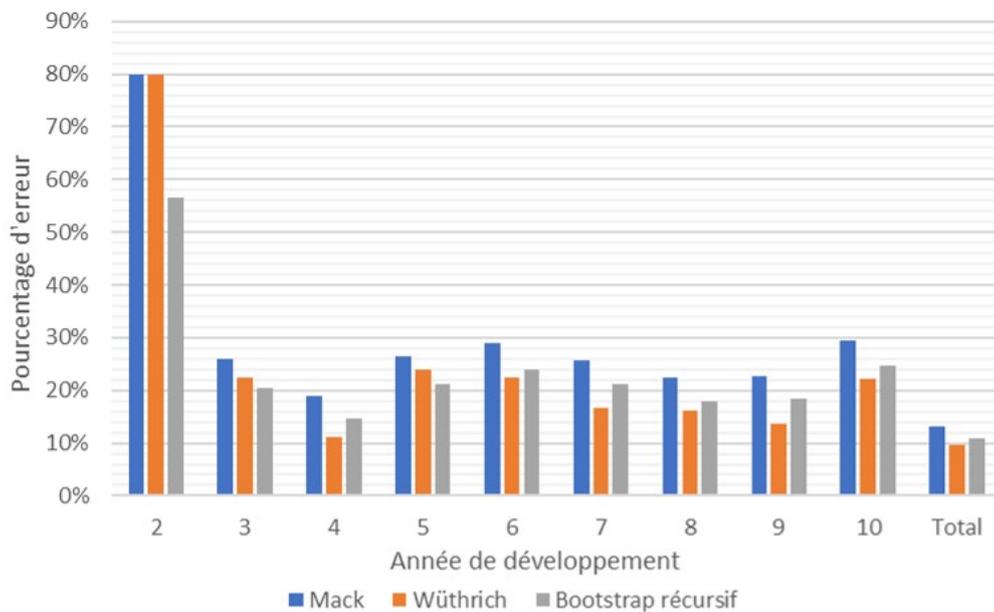


FIGURE 6.7 – Estimations du risque de réserve $\frac{\sqrt{MSE(\hat{R}_i)}}{\hat{R}_i^M}$ par année de survenance pour chaque modèle

La Figure 6.7 précédente représente le ratio $\frac{\sqrt{MSE(\hat{R}_i)}}{\hat{R}_i^M}$, où \hat{R}_i^M correspond au montant de réserves estimé avec le modèle de Mack *chain-ladder*. Nous pouvons noter que les simulations par bootstrap permettent de diminuer drastiquement l'erreur commise sur les montants des réserves

par année de survenance, tout comme les calculs d'erreur quadratique moyenne développés dans l'article de Wüthrich, Merz et Lysenko [64]. Leur approche conduit à une diminution de l'erreur moyenne totale de plus de 3%.

6.2.2 Prise en compte des hypothèses de Dal Moro

Nous allons maintenant appliquer les modèles qui tiennent compte des hypothèses de Dal Moro sur le skewness et le kurtosis des paiements cumulés. Nous comparons sur la Figure suivante les performances des différents modèles relativement à la méthode de Mack *chain-ladder*. Les estimations relatives à l'approche par loi de Pareto généralisée ont été obtenues après 1000 simulations. De plus, la méthode des moments a été appliquée sur les trois premiers moments de la distribution.

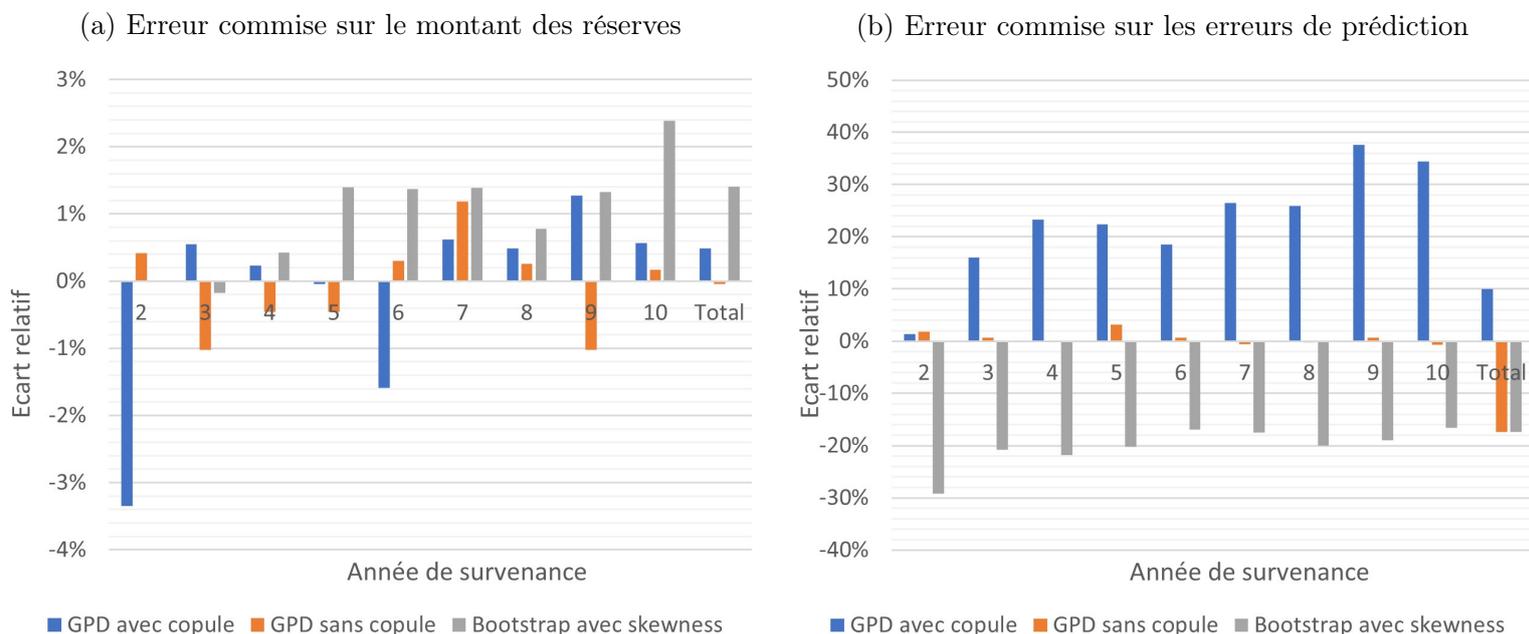


FIGURE 6.8 – Erreurs commises par les modèles tenant compte des hypothèses de Dal Moro

Tout d'abord, nous avons constaté des divergences de performances pour la simulation GPD en fonction de l'instauration ou non d'une structure de dépendance. Ainsi, bien que Dal Moro opte pour des copules gaussiennes dans son article [14], nous notons que nos résultats sont meilleurs sans leur utilisation. En effet, la Figure 6.8a présente des estimations de montants de réserves très éloignées de celles obtenues par la méthode de Mack *chain-ladder*. Or, ces dernières devraient être assez proches. Par ailleurs, l'intégration de copules perturbe l'évaluation du risque de réserve et élargit les erreurs quadratiques moyennes, comme en témoigne la Figure 6.8b.

Pour l'approche de bootstrap avec un paramètre de skewness, nous avons effectué 1000 simulations. Cette méthode génère davantage d'erreur sur les montants des réserves : une surévaluation se note sur la Figure 6.8a. Cependant, d'après la Figure 6.8b elle permet d'affiner les prédictions du modèle en diminuant les erreurs quadratiques moyennes.

Pour conclure cette analyse sur le risque de réserve, nous présentons les erreurs de prédictions des différents modèles proportionnellement aux montants estimés par le modèle de Mack *chain-ladder*, \hat{R}_i^M .

i	Mack	Simulation GPD avec copule	Simulation GPD sans copule	Bootstrap avec skewness
2	79,8%	80,9%	81,3%	56,5%
3	25,9%	30,1%	26,1%	20,5%
4	18,8%	23,2%	18,8%	14,7%
5	26,5%	32,5%	27,4%	21,2%
6	29,0%	34,3%	29,1%	24,1%
7	25,6%	32,4%	25,5%	21,1%
8	22,3%	28,1%	22,3%	17,9%
9	22,7%	31,2%	22,9%	18,4%
10	29,5%	39,6%	29,3%	24,6%
Total	13,1%	14,4%	10,8%	10,8%

 TABLE 6.3 – Pourcentage d’erreur de prédiction $\frac{\sqrt{MSE(\hat{R}_i)}}{\hat{R}_i^M}$ des différents modèles

Il ressort de la Table 6.3 que la méthode de simulation GPD est plus précise sans l’utilisation de copules. L’approche de bootstrap adaptée en tenant compte d’un paramètre d’asymétrie présente également de meilleures performances que la méthode de Mack *chain-ladder*. En réalité, dans cet exemple d’application, ce bootstrap se confond avec la méthodologie de bootstrap récursif. Ce phénomène est certainement dû à un manque d’asymétrie de la distribution du triangle, qui n’influence donc pas suffisamment la valeur des résidus tirés à chaque simulation.

6.3 Evaluation du skewness et du kurtosis

Nous cherchons désormais à quantifier l’asymétrie et l’aplatissement de la distribution des réserves. Pour estimer le skewness total et par année de survenance, nous pouvons appliquer plusieurs méthodes :

- le bootstrap récursif,
- le bootstrap incluant des paramètres de skewness et de kurtosis,
- l’approche simulatoire de Dal Moro, que nous appliquons avec une loi GPD,
- et enfin l’approche formule fermée du même auteur.

Pour déterminer le skewness via la formule fermée de Dal Moro, nous utilisons la fonction `quantile` du package R `ChainLadder`. Ses résultats nous serviront de référence afin de comparer les estimations produites par les autres méthodes.

Nous notons, sur la Table 6.4 ci-après, que la méthode de simulation GPD est plus efficace sans l’introduction de copules pour estimer le skewness total de la distribution. En effet, bien que les skewness par année de survenance obtenus par simulation GPD sans copule soient plus éloignés du résultat théorique, la non-considération d’une structure de dépendance fournit un résultat de skewness total plus proche du résultat théorique.

Ici encore, l’ajout d’un paramètre de skewness ne change pas les estimations de skewness obtenues avec bootstrap récursif classique.

Enfin, avec un skewness final positif, nous savons que la distribution des réserves est caractérisée par une queue de droite plus longue et plus épaisse.

i	Formule fermée	GPD avec copule	GPD sans copule	Bootstrap récursif	Bootstrap avec skewness
2	0,000	0,101	-0,001	0,030	0,030
3	-0,031	-0,067	0,058	0,117	0,117
4	-0,046	-0,072	-0,067	0,168	0,168
5	-0,001	-0,073	0,017	0,093	0,093
6	0,180	0,322	0,285	0,239	0,239
7	0,055	0,164	-0,005	0,183	0,183
8	0,268	0,341	0,377	0,260	0,260
9	0,286	0,395	0,382	0,259	0,259
10	0,314	0,584	0,402	0,314	0,314
Total	0,215	0,333	0,192	0,035	0,035

TABLE 6.4 – Résultats d'estimation de skewness pour le triangle de Mack

Pour quantifier le kurtosis de la distribution des réserves, nous disposons des mêmes méthodes que précédemment, à l'exception de la formule fermée. Rappelons par ailleurs que la loi de Pareto généralisée relative à la méthode de simulation a été calibrée ici à l'aide des coefficients d'asymétrie de la distribution. Cependant, il est tout à fait possible de considérer le quatrième moment au lieu du troisième en vue de se concentrer sur l'obtention du kurtosis de la distribution des réserves. Les résultats des différentes méthodes sont présentés sur la Table 6.5.

i	GPD avec copule	GPD sans copule	Bootstrap récursif	Bootstrap avec skewness
2	1,77	1,76	2,71	2,71
3	2,10	2,29	3,46	3,46
4	2,29	2,53	2,86	2,86
5	2,08	2,28	2,77	2,77
6	2,59	2,78	2,90	2,90
7	2,45	2,56	3,10	3,10
8	2,76	2,91	3,03	3,03
9	2,71	2,96	3,22	3,22
10	2,88	3,00	2,92	2,92
Total	3,07	2,74	2,89	2,89

TABLE 6.5 – Résultats d'estimation de kurtosis pour le triangle de Mack

Les diverses approches donnent des résultats de kurtosis finaux assez proches. En considérant que la méthode de simulation par GPD sans structure de dépendance est plus fiable vis-à-vis du skewness et de l'erreur de prédiction, nous pouvons dire que la distribution des réserves est légèrement platykurtique. En effet, les coefficients d'aplatissement étant inférieurs à 3, la distribution aura des flancs plus larges et des queues plus fines qu'une loi normale.

6.4 Distribution finale des réserves

Désormais, nous affichons la distribution des réserves obtenues à partir des méthodes de simulations GPD et bootstrap. Puisque la distribution des réserves résultant de la méthode de bootstrap est inchangée par ajout de paramètre de skewness, nous affichons uniquement le graphique relatif au bootstrap récursif classique.

(a) Simulation GPD avec structure de copule - 1000 simulations (b) Simulation GPD sans structure de copule - 1000 simulations

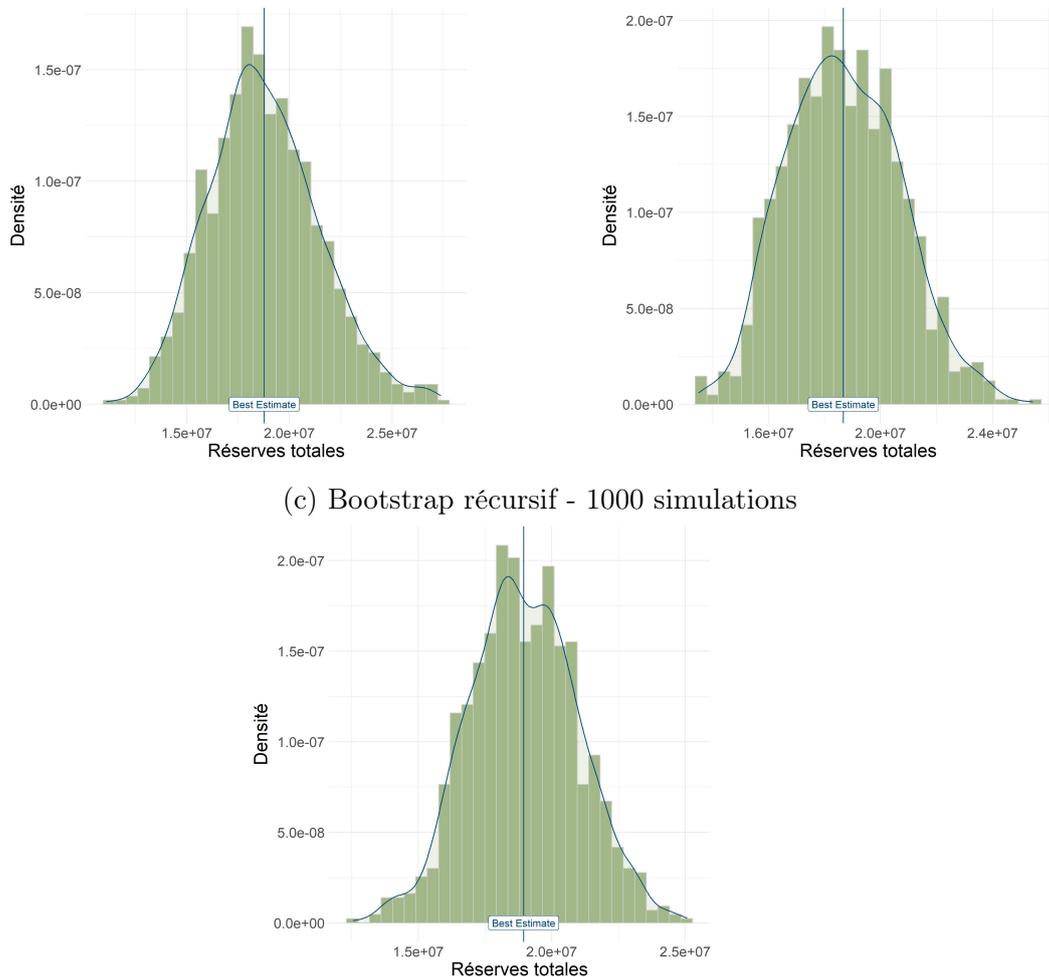


FIGURE 6.9 – Distribution des réserves pour le triangle de Mack

L'asymétrie positive de la distribution constatée lors des estimations de skewness est visible sur chaque graphique de la Figure 6.9. En effet, le pic de la distribution se trouve à gauche de la valeur moyenne *Best estimate*. Par ailleurs, comme prévu, les distributions des Figures 6.9b et 6.9c sont bien platykurtiques. Avec un kurtosis estimé supérieur à 3, la Figure 6.9a fait apparaître une distribution leptokurtique.

Chapitre 7

Intégration du climat dans le provisionnement

Ce chapitre méthodologique présente l'approche à suivre pour extraire une tendance climatique dans les réserves. Pour ce faire, nous allons détourner la méthode de séparation introduite par Verbeek en 1972 et appliquée par Taylor en 1977 [55] afin de capter un facteur de tendance climatique dans les réserves. Bien que ce mémoire ne s'intéresse pas spécifiquement à la prise en compte de l'inflation, nous devons néanmoins déflater les sinistres au préalable, dans l'optique de capter uniquement une tendance liée au climat. Nous procédons donc en plusieurs étapes :

1. **Traitement de l'inflation**
2. **Application de la méthode de séparation climatique**

Suite à cette deuxième étape, une modélisation de la tendance climatique extraite selon des variables météorologiques peut être menée, en vue de mieux comprendre l'évolution future des réserves.

7.1 Traitement de l'inflation

Pour éliminer tout effet inflationniste qui pourrait perturber l'extraction de la tendance climatique, nous pouvons appliquer la méthodologie de retraitement présentée dans le cours d'assurance non-vie de Baradel [6]. Rappelons que les sinistres sont réglés à différents moments, et que le coût des sinistres est impacté par l'inflation. Par conséquent, pour travailler en euros constants, il est nécessaire de déflater les sinistres.

Pour cela, on introduit $(I_m)_{m \geq 1}$ un indice des prix associé au secteur du triangle. Ces données sont généralement disponibles sur le site de l'INSEE. En notant n le nombre d'années considérées dans le triangle, on peut corriger les paiements incrémentaux $X_{i,j}$ de la manière suivante

$$X'_{i,j} = X_{i,j} \frac{I_{n+1}}{I_{i+j}}. \quad (7.1)$$

A partir de cette relation peuvent s'obtenir les paiements cumulés $C'_{i,j}$. Pour compléter la partie inférieure du triangle, la méthode *chain-ladder* est toujours valide.

Pour obtenir les paiements futurs corrigés de l'inflation, il est nécessaire de réinflater les paiements incrémentaux résultant du processus de complétion du triangle en les multipliant par le facteur $\frac{I_{i+j}}{I_{n+1}}$, avec $i + j \geq n + 1$. Dans ce cas, les indices des prix futurs font au préalable l'objet d'une étude de prévision temporelle. De là, il est possible d'estimer le coût futur

des sinistres manquants, ainsi que l'erreur d'estimation faite sur ces réserves en recumulant le triangle. Toutefois, ce travail sur l'inflation ne sera pas mené dans le cadre de ce mémoire, qui se concentre sur le risque climatique.

Le cheminement de cette méthode appliquée au modèle de Mack *chain-ladder* est résumé sur la Figure 7.1.

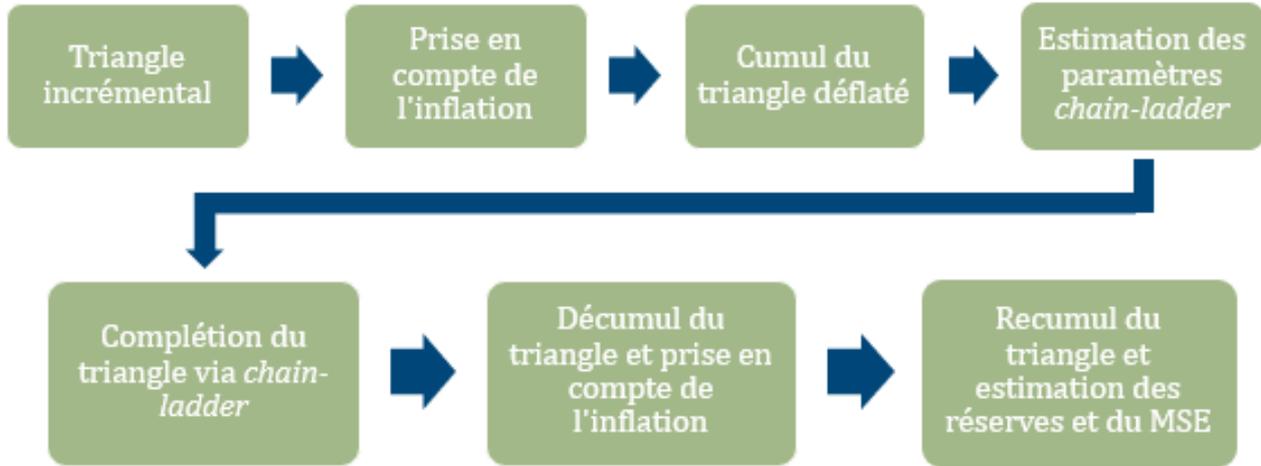


FIGURE 7.1 – Processus de prise en compte de l'inflation dans le modèle de Mack *chain-ladder*

Remarque 7.1.1. Si l'auteur utilise le modèle de Mack *chain-ladder* pour illustrer sa méthode, la réinflation finale des sinistres estimés peut s'effectuer après l'application de n'importe quel modèle de provisionnement. Dans le cadre de ce mémoire, la déflation des sinistres permettra d'ôter tout effet lié à l'inflation avant d'appliquer une méthode de séparation climatique.

7.2 Méthode de séparation climatique

La méthode de séparation, introduite par Verbeek en 1972 [58] et développée par Taylor en 1977 [55], permet de tenir compte de facteurs inflationnistes en utilisant deux triangles de liquidation incrémentaux, l'un pour les paiements, noté $(X_{i,j})$, et l'autre pour le nombre de sinistres, noté $(N_{i,j})$. Nous proposons ici une adaptation afin de capter l'effet relatif au climat dans les réserves.

Notation 7.2.1. On note ∇ (resp. Δ) l'ensemble des données du triangle supérieur (resp. inférieur).

$$\begin{aligned}\nabla &= \{0 \leq i \leq I, 0 \leq j \leq I - i\} \\ \Delta &= \{1 \leq i \leq I, I - i + 1 \leq j \leq I\}\end{aligned}$$

Notation 7.2.2. ∇_i (resp. Δ_i) caractérise, pour une année de survenance i , l'ensemble des années de développement associées au triangle supérieur ∇ (resp. inférieur Δ).

Le nombre de sinistres relatif à une année de survenance i est alors :

$$N_i = \sum_{j \in \nabla_i} N_{i,j} + \sum_{j \in \Delta_i} N_{i,j}. \quad (7.2)$$

La méthode de séparation de Taylor repose sur une hypothèse principale : le coût moyen des sinistres survenus en année i et déclarés en année j est proportionnel à un facteur $r_j \lambda_{i+j}$.

Autrement dit,

$$\mathbb{E} \left[\frac{X_{i,j}}{N_i} \middle| N_i \right] = r_j \lambda_{i+j}.$$

r_j correspond à la proportion de sinistres payés en année de développement j . Si l'on considère que les sinistres sont intégralement payés après I années de développement, on a alors la contrainte $\sum_{j=0}^I r_j = 1$. λ_{i+j} est quant à lui un facteur inflationniste.

A la différence de l'inflation, le climat intervient uniquement via les années de survenance des sinistres. Par conséquent, nous adaptons l'hypothèse précédente pour mettre en œuvre notre modèle climatique :

$$\mathbb{E} \left[\frac{X_{i,j}}{N_i} \middle| N_i \right] = r_j \lambda_i. \quad (\text{H})$$

λ_i décrira donc la tendance climatique dans le triangle. Nous tenterons ensuite de corrélérer son évolution avec celles de variables climatiques. Ici encore, on suppose que $\sum_{j=0}^I r_j = 1$.

7.2.1 Estimation du montant des réserves

Pour estimer de manière précise les coefficients r_j et λ_i , il est nécessaire d'avoir dans un premier temps un bon estimateur de N_i . Pour cela, on applique la méthode *chain-ladder* au cumul du triangle ($N_{i,j}$). Après décumul, on obtient l'estimateur \hat{N}_i . A partir du coût moyen $s_{i,j} = \frac{X_{i,j}}{\hat{N}_i}$, les estimateurs \hat{r}_j et $\hat{\lambda}_i$ peuvent être calculés de manière récursive :

$$\hat{\lambda}_i = \frac{\sum_{j=0}^{I-i} s_{i,j}}{1 - \sum_{j=I-i+1}^I \hat{r}_j} \quad \text{pour } 0 \leq i \leq I \quad \text{et} \quad \hat{r}_j = \frac{\sum_{i=0}^{I-j} s_{i,j}}{\sum_{i=0}^{I-j} \hat{\lambda}_i} \quad \text{pour } 0 \leq j \leq I.$$

La récurrence débute en $i = 0$ et permet de déterminer $\hat{\lambda}_0$. De là, \hat{r}_I peut être obtenu. De proche en proche, l'estimation de \hat{r}_{I-i} est déduite de l'estimation de $\hat{\lambda}_i$.

Suite à ces calculs, il est désormais possible de déterminer les paiements incrémentaux futurs

$$\hat{X}_{i,j} = \hat{N}_i \hat{r}_j \hat{\lambda}_i. \quad (7.3)$$

De là, on déduit le montant des réserves relatif à chaque année de survenance, puis leur total :

$$\hat{R}_i = \sum_{j \in \Delta_i} \hat{X}_{i,j} \quad \text{et} \quad \hat{R} = \sum_{\Delta} \hat{X}_{i,j}.$$

7.2.2 Obtention d'une distribution des réserves

7.2.2.1 Hypothèses

Toutefois, la méthode de séparation ne conduit pas à l'obtention d'une distribution des réserves. Björkwall et al. [7] proposent alors une méthode de bootstrap paramétrique pour

pallier cela. Ainsi, les auteurs font deux hypothèses de distribution sur le nombre de sinistres et les paiements incrémentaux :

$$N_{i,j} \sim \mathcal{P}(n_{i,j}) \text{ et } X_{i,j}|N_i \sim \Gamma\left(\frac{1}{\phi}, r_j \lambda_{i+j} \phi\right) \text{ avec } n_{i,j} = \mathbb{E}[N_{i,j}] \text{ et } \phi > 0.$$

Ces dernières peuvent s'adapter au modèle climatique que nous avons établi précédemment :

$$N_{i,j} \sim \mathcal{P}(n_{i,j}) \text{ et } X_{i,j}|N_i \sim \Gamma\left(\frac{N_i}{\phi}, r_j \lambda_i \phi\right) \text{ avec } n_{i,j} = \mathbb{E}[N_{i,j}] \text{ et } \phi > 0.$$

Avec cette paramétrisation, $\mathbb{E}[X_{i,j}|N_i] = \frac{N_i}{\phi} r_j \lambda_i \phi = N_i r_j \lambda_i$, ce qui est cohérent avec l'Hypothèse (H). Pour procéder aux simulations, il reste à déterminer la valeur de ϕ . Tout d'abord,

$$\text{Var}(X_{i,j}|N_i) = \frac{N_i}{\phi} (r_j \lambda_i \phi)^2 = \phi N_i (r_j \lambda_i)^2 = \phi \frac{\mathbb{E}[X_{i,j}|N_i]^2}{N_i}.$$

Les auteurs proposent alors d'utiliser un estimateur de Pearson pour ϕ :

$$\hat{\phi} = \frac{1}{|\nabla| - q} \sum_{\nabla} \hat{N}_i \frac{\left(X_{i,j} - \hat{\mathbb{E}}[X_{i,j}|N_i]\right)^2}{\hat{\mathbb{E}}[X_{i,j}|N_i]^2} = \frac{1}{|\nabla| - q} \sum_{\nabla} \hat{N}_i \frac{\left(X_{i,j} - \hat{N}_i \hat{r}_j \hat{\lambda}_i\right)^2}{\left(\hat{N}_i \hat{r}_j \hat{\lambda}_i\right)^2} \quad (7.4)$$

avec $|\nabla| = \frac{(I+1)(I+2)}{2}$, nombre d'observations dans le triangle de paiements supérieur

$q = 2I + 1$, nombre de paramètres à estimer par la méthode de séparation climatique

7.2.2.2 Estimation de la variance des réserves

Grâce à l'indépendance conditionnelle de $X_{i,j}$ pour tous i et j sachant N_i , il résulte que

$$\begin{aligned} \text{Var}(R_i) &= \mathbb{E}[\text{Var}(R_i|N_i)] + \text{Var}(\mathbb{E}(R_i|N_i)) \\ &= \mathbb{E}\left[\sum_{j \in \Delta_i} \phi N_i (r_j \lambda_i)^2\right] + \text{Var}\left(\sum_{j \in \Delta_i} N_i r_j \lambda_i\right) \\ &= \phi \mathbb{E}[N_i] \sum_{j \in \Delta_i} (r_j \lambda_i)^2 + \left(\sum_{j \in \Delta_i} r_j \lambda_i\right)^2 \text{Var}(N_i). \end{aligned}$$

Or, $N_{i,j} \sim \mathcal{P}(n_{i,j})$. Donc $\mathbb{E}[N_i] = \text{Var}(N_i) = \sum_{j \in \nabla_i \cup \Delta_i} n_{i,j}$. Par conséquent,

$$\text{Var}(R_i) = \left(\phi \sum_{j \in \Delta_i} (r_j \lambda_i)^2 + \left(\sum_{j \in \Delta_i} r_j \lambda_i\right)^2\right) \left(\sum_{j \in \nabla_i \cup \Delta_i} n_{i,j}\right).$$

Cette variance s'estime alors par

$$\widehat{\text{Var}}(R_i) = \left(\hat{\phi} \sum_{j \in \Delta_i} (\hat{r}_j \hat{\lambda}_i)^2 + \left(\sum_{j \in \Delta_i} \hat{r}_j \hat{\lambda}_i\right)^2\right) \left(\sum_{j \in \nabla_i \cup \Delta_i} \hat{n}_{i,j}\right). \quad (7.5)$$

Puis

$$\widehat{\text{Var}}(R) = \sum_i \widehat{\text{Var}}(R_i). \quad (7.6)$$

7.2.2.3 Algorithme de bootstrap

Etape 1 - Monde réel : Cette étape consiste en l'application de la méthode de séparation dans l'optique de déterminer les montant de réserves estimées \hat{R}_i et \hat{R} .

Etape 1.1 - Triangle de nombre de sinistres

1. Cumul du triangle de nombre de sinistres pour appliquer le modèle *chain-ladder*.
2. Récupération des projections $\hat{N}_{i,j}$ du triangle de nombre de sinistres.
3. Calcul de \hat{N}_i grâce à l'Equation (7.2).

Etape 1.2 - Triangle de paiements

1. Application de la méthode de séparation pour obtenir $\hat{X}_{i,j}$.
2. Calcul du montant des réserves par année \hat{R}_i , puis du montant total \hat{R} .
3. Calcul des variances des vrais montants de réserves par année de survenance R_i puis du montant total R .
4. Estimation de $\hat{\phi}$ en vue de procéder au bootstrap via l'Equation (7.4).

Etape 2 - Bootstrap : Cette étape vise à être répétée B fois. Par souci de clarté, les variables aléatoires $Z_{i,j}$ simulées sont notées :

- $Z_{i,j}^*$ si elles sont associées à des observations réelles, c'est-à-dire qu'elles se situent dans le triangle supérieur.
- $Z_{i,j}^{**}$ si elles ne sont pas observées, donc situées dans le triangle inférieur.

Les estimateurs calculés tout au long de la procédure de bootstrap seront également notés de cette manière.

Le processus de bootstrap s'articule en deux sous-étapes. La première se résume à l'évaluation de l'erreur d'estimation sur les paramètres r_j et λ_i du modèle. Dans la seconde, seule l'erreur de modèle est considérée : le montant des réserves dépendra alors uniquement de l'aléa présent dans le triangle.

Etape 2.1 - Captation de l'erreur d'estimation

1. Simulation du nombre de sinistres $N_{i,j}^*$ dans le triangle supérieur.
2. Cumul du triangle de nombre de sinistre en vue d'appliquer le modèle *chain-ladder*.
3. Estimation des projections $\hat{n}_{i,j}^*$.
4. Calcul du nombre de sinistres ultime par année de survenance i :

$$\hat{N}_i^* = \sum_{j \in \nabla_i} N_{i,j}^* + \sum_{j \in \Delta_i} \hat{n}_{i,j}^*$$

5. Simulation du nombre de sinistres $N_{i,j}^{**}$ dans le triangle inférieur.
6. Calcul du nombre de sinistres ultime par année de survenance i :

$$N_i^{**} = \sum_{j \in \nabla_i} N_{i,j}^* + \sum_{j \in \Delta_i} N_{i,j}^{**}$$

7. Simulation des paiements incrémentaux $X_{i,j}^*$ du triangle supérieur, conditionnellement à N_i^{**} .

8. Estimation des coefficients $\hat{\lambda}_i^*$ et \hat{r}_j^* à partir des coûts moyens $\frac{X_{i,j}^*}{\hat{N}_i^*}$.
9. Application de la méthode de séparation climatique avec $\hat{N}_{i,j}$ pour obtenir les paiements incrémentaux $\hat{X}_{i,j}^*$ du triangle inférieur :

$$\hat{X}_{i,j}^* = \hat{N}_i^* \hat{r}_j^* \hat{\lambda}_i^*$$

10. Calcul des montants des réserves par année $\hat{R}_i^* = \sum_{j \in \Delta_i} \hat{X}_{i,j}^*$, puis du montant total $\hat{R}^* = \sum_{j \in \Delta} \hat{X}_{i,j}^*$.

Etape 2.2 - Captation de l'erreur de modèle

1. Simulation des paiements incrémentaux $X_{i,j}^{**}$ du triangle inférieur, conditionnellement à N_i^{**} .
2. Estimation du paramètre de dispersion simulé $\hat{\phi}^*$:

$$\hat{\phi}^* = \frac{1}{|\nabla| - q} \sum_{\nabla} \hat{N}_i^* \frac{\left(X_{i,j}^* - \hat{N}_i^* \hat{r}_j^* \hat{\lambda}_i^* \right)^2}{\left(\hat{N}_i^* \hat{r}_j^* \hat{\lambda}_i^* \right)^2}$$

3. Calcul du montant des réserves par année $R_i^{**} = \sum_{j \in \Delta_i} X_{i,j}^{**}$, puis du montant total $R^{**} = \sum_{j \in \Delta} X_{i,j}^{**}$.
4. Estimation de la variance des réserves R_i^{**} et R^{**} .

$$\widehat{\text{Var}}(R_i^{**}) = \left(\hat{\phi}^* \sum_{j \in \Delta_i} \left(\hat{r}_j^* \hat{\lambda}_i^* \right)^2 + \left(\sum_{j \in \Delta_i} \hat{r}_j^* \hat{\lambda}_i^* \right)^2 \right) \left(\sum_{j \in \nabla_i \cup \Delta_i} \hat{n}_{i,j}^* \right)$$

$$\widehat{\text{Var}}(R^{**}) = \sum_i \widehat{\text{Var}}(R_i^{**})$$

5. Calcul des erreurs de prédiction sur les réserves par année de survivance pe_i^{**} et sur les réserves totales pe^{**} :

$$pe_i^{**} = \frac{R_i^{**} - \hat{R}_i^*}{\sqrt{\widehat{\text{Var}}(R_i^{**})}} \quad \text{et} \quad pe^{**} = \frac{R^{**} - \hat{R}^*}{\sqrt{\widehat{\text{Var}}(R^{**})}}$$

Etape 3 - Analyse des simulations

1. Calcul des variances des montants de réserves théoriques par année de survivance R_i puis du montant total R grâce aux Equations (7.5) et (7.6).
2. Calcul des prédictions des montants de réserves \tilde{R}_i^{**} et \tilde{R}^{**} , relatives à chaque itération de la boucle de bootstrap :

$$\tilde{R}_i^{**} = \hat{R}_i + pe_i^{**} \sqrt{\widehat{\text{Var}}(R_i)} \quad \text{et} \quad \tilde{R}^{**} = \hat{R} + pe^{**} \sqrt{\widehat{\text{Var}}(R)}$$

3. Affichage des montants des réserves \tilde{R}_i^{**} et \tilde{R}^{**} relatifs à chaque itération de la boucle de bootstrap.

Si les données clients que nous traitons dans ce mémoire n'ont pas rendu possible l'application de la technique de bootstrap, cette dernière a pu être testée sur des exemples simulés, dont les résultats sont disponibles en Annexe C.

Troisième partie

Exemple d'application : Provisionnement climatique en assurance agricole

Chapitre 8

Présentation des données agricoles

Dans ce chapitre sont présentées les données clients dont nous disposons pour appliquer les méthodologies de la Partie II. Il s'agit de triangles agricoles relatifs à la période 2002-2019 pour les trois sous-branches suivantes :

- **alés climatiques,**
- **hors alés climatiques et tempête,**
- et **responsabilité civile.**

Nous avons également accès au montant de primes acquises annuellement, sans pour autant connaître le nombre de contrats présents dans le portefeuille. Dans un souci de confidentialité, toutes les données ont été bruitées.

Nous allons dans un premier temps analyser les données, avant de procéder à une validation des hypothèses de Mack et Dal Moro, requise pour la suite de l'étude.

8.1 Analyse d'ensemble des triangles

La Figure 8.1 suivante présente l'évolution des primes acquises pour la période 2002-2019 pour les trois sous-branches considérées.

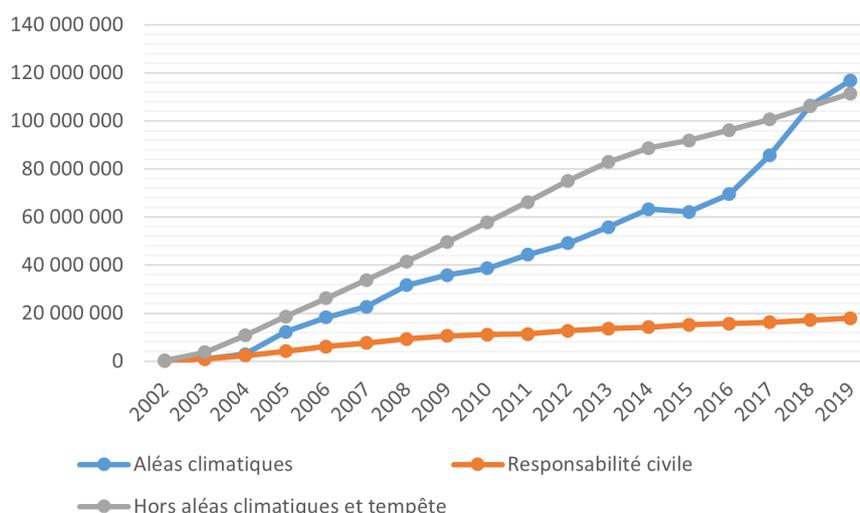


FIGURE 8.1 – Evolution du montant des primes acquises

Nous notons une très forte croissance des primes pour les sous-branches alés climatiques et hors alés climatiques et tempêtes. Cette augmentation est plus modérée pour la responsabilité

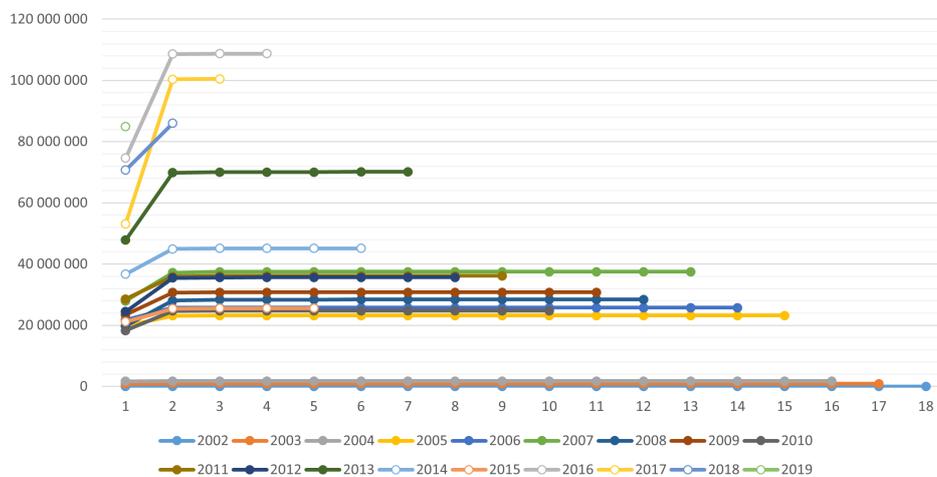
civile. N'ayant pas à notre disposition des informations sur le nombre de contrats par année, il n'est pas possible de déterminer si ces hausses sont dues à une augmentation des primes, à un grossissement du portefeuille agricole ou bien à un effet combiné des deux. Néanmoins, sur les dernières années, les primes acquises pour les aléas climatiques ont drastiquement augmenté, jusqu'à dépasser le niveau des primes excluant les tempêtes et les aléas climatiques, celles-ci observant une croissance plus lente. Cela pourrait donc marquer une récente importance des risques climatiques dans la sinistralité agricole.

Regardons de plus près les comportements des triangles de liquidation, représentés sur la Figure 8.2.

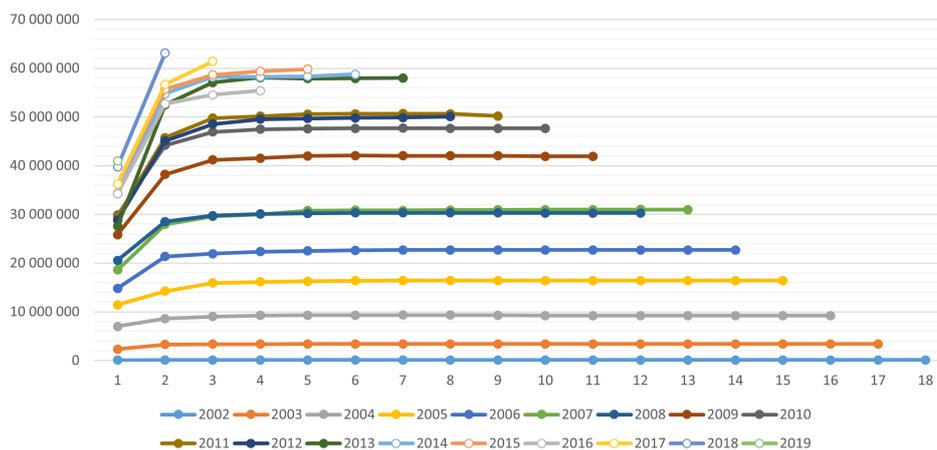
Les triangles des Figures 8.2a et 8.2b sont typiques de branches à développement court. En effet, les sinistres sont rapidement déclarés et évoluent peu. Ainsi, pour la plupart des années de survenance, les sinistres sont connus en totalité à la fin de la troisième année de développement.

Par ailleurs, nous observons un montant de sinistres en hausse au fil des années. Là encore, ne disposant pas du nombre de contrats par année, nous ne pouvons pas conclure quant à un accroissement du portefeuille, potentiellement lié à des campagnes d'information auprès des agriculteurs, ou à une hausse du coût des sinistres. Néanmoins, nous remarquons sur la Figure 8.2a que les dernières années, fortement marquées par des événements climatiques, ont été les plus coûteuses pour l'assureur. En particulier, l'année 2016, qui a enregistré la pire récolte de blé depuis 1976 [17], est celle qui a connu le montant de sinistres le plus élevé (voir Figure 8.2a).

(a) Aléas climatiques



(b) Hors aléas climatiques et tempête



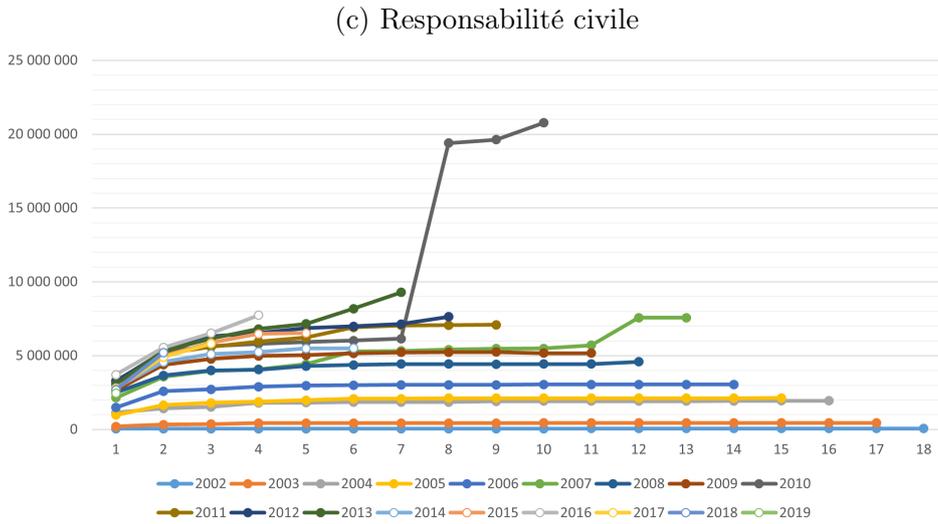


FIGURE 8.2 – Evolution des règlements nets par année de développement

Le triangle de responsabilité civile est quant à lui plus susceptible de connaître un nombre important d'IBNR, comme en témoigne la Figure 8.2c. Le développement des sinistres est parfois très tardif, comme pour les années 2007 et 2010. Cependant, la hausse des sinistres selon l'année de survenance est moins claire que pour les triangles précédents.

Intéressons-nous désormais aux ratios sinistres sur primes pour les trois branches considérées.

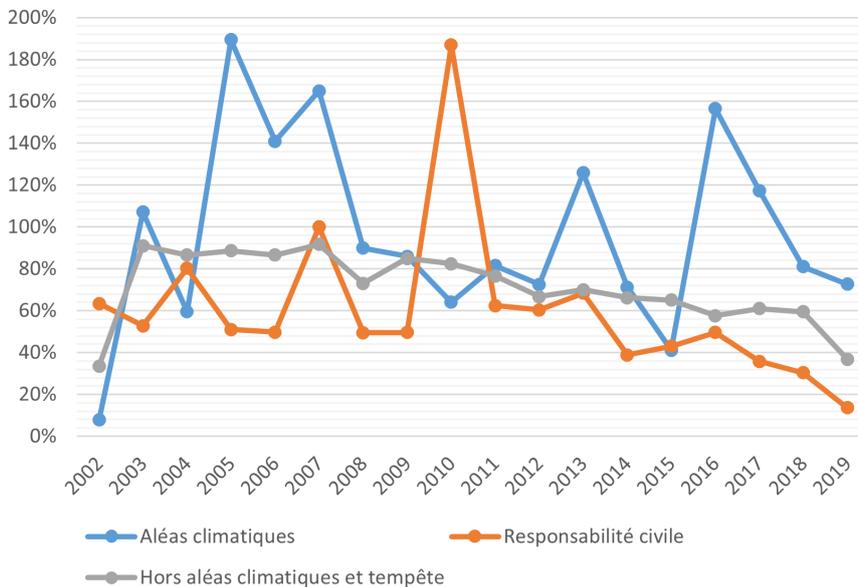


FIGURE 8.3 – Evolution du ratio sinistres sur primes

Nous constatons sur la Figure 8.3 de fréquents dépassements du seuil de 100% pour la sous-branche climatique. Dès lors, pour les sept années concernées par ces dépassements, les primes acquises n'ont pas été suffisantes pour couvrir les sinistres survenus. Cela a donc pu avoir comme effet d'augmenter le montant des primes enregistrées les années suivantes, ce qui expliquerait la forte croissance des primes aléas climatiques perçue sur la Figure 8.1. Cette hypothèse semble se vérifier graphiquement. En effet, l'assureur a répertorié un ratio sinistre

sur primes de 157% en 2016 et a massivement augmenté ses primes enregistrées sur la période 2017-2019 comme le montre la Figure 8.1.

8.2 Validation des hypothèses de provisionnement

Dans cette section, nous présentons uniquement les résultats obtenus pour le triangle agricole aléas climatiques. La validation des hypothèses pour les deux autres triangles est disponible en Annexe D.1 et D.2.

8.2.1 Validation de l'Hypothèse 1

Afin de vérifier l'indépendance des années de survenance, nous représentons graphiquement l'évolution des facteurs de développement individuels $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ en fonction des années de développement j .

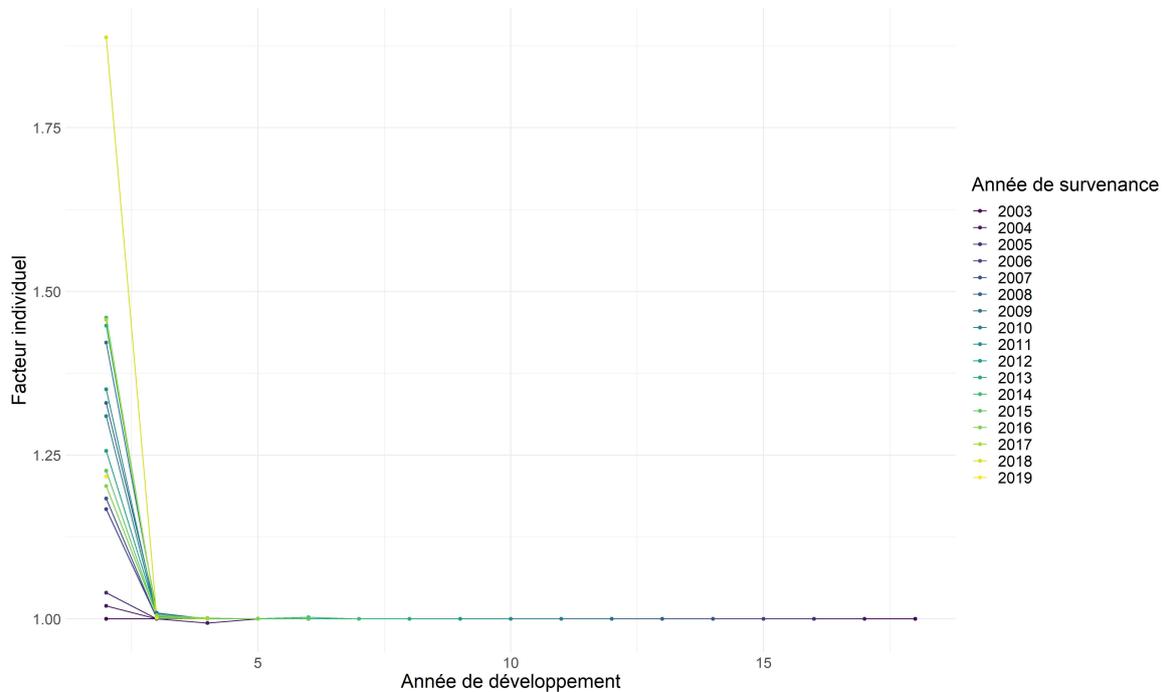


FIGURE 8.4 – Vérification de l'**Hypothèse 1**

Quelle que soit l'année de survenance, les courbes des facteurs de développement individuels de la Figure 8.4 présentent une tendance commune. Par conséquent, l'**Hypothèse 1** est vérifiée. Nous pouvons toutefois noter la présence d'effets climatiques : les valeurs initiales des facteurs de développement individuels sont croissantes en fonction du temps. De fait, les sinistres sont de plus en plus coûteux par rapport à l'année de survenance précédente.

8.2.2 Validation de l'Hypothèse 2

L'**Hypothèse 2** se teste en traçant, pour chaque année de développement j , $1 \leq j \leq I - 1$ les points de coordonnées $(C_{i,j}, C_{i,j+1})$. Le graphique résultant mettra en lumière l'existence ou non d'un facteur de proportionnalité.

Sur la Figure 8.5 ci-après, les points de coordonnées $(C_{i,j}, C_{i,j+1})$ sont alignés, quelle que soit l'année de développement considérée. L'**Hypothèse 2** est donc clairement validée.

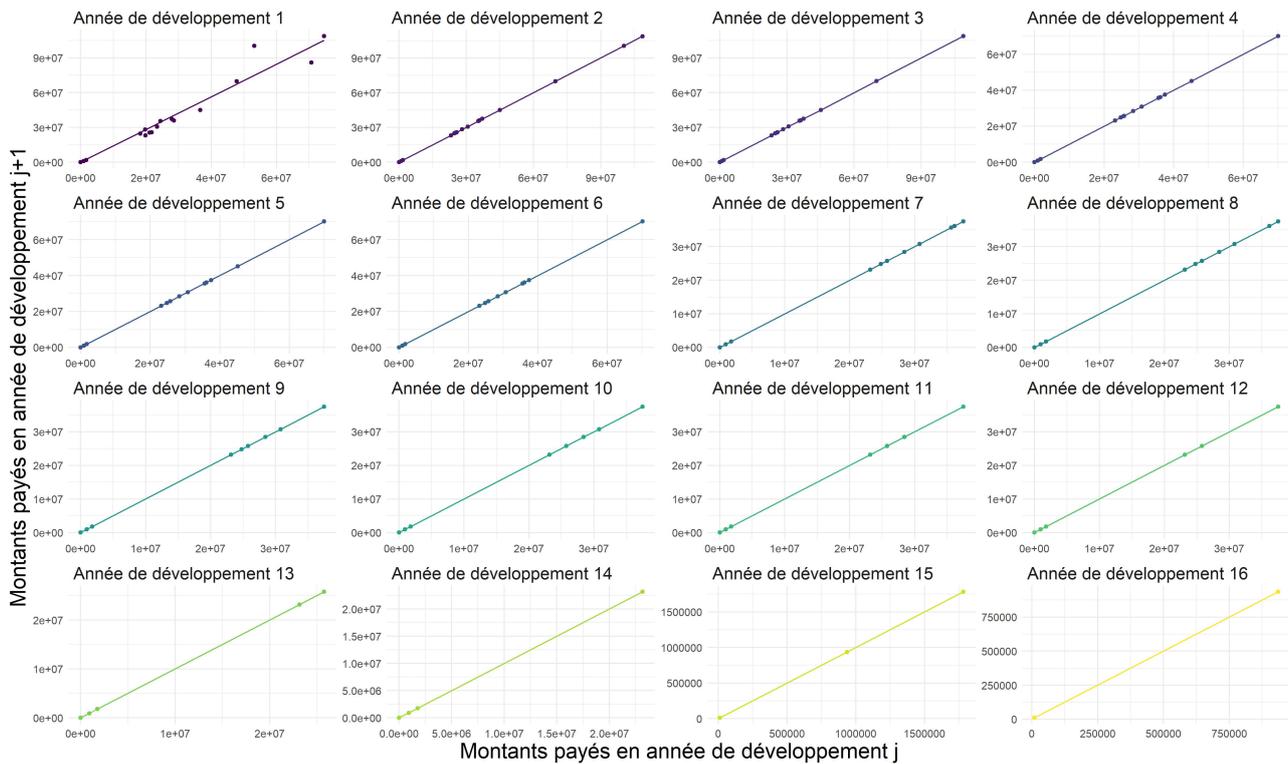


FIGURE 8.5 – Vérification de l’Hypothèse 2

8.2.3 Validation de l’Hypothèse 3

Comme dans le Chapitre 6, nous allons plus particulièrement vérifier l’hypothèse d’indépendance des résidus relative à l’Hypothèse 3.

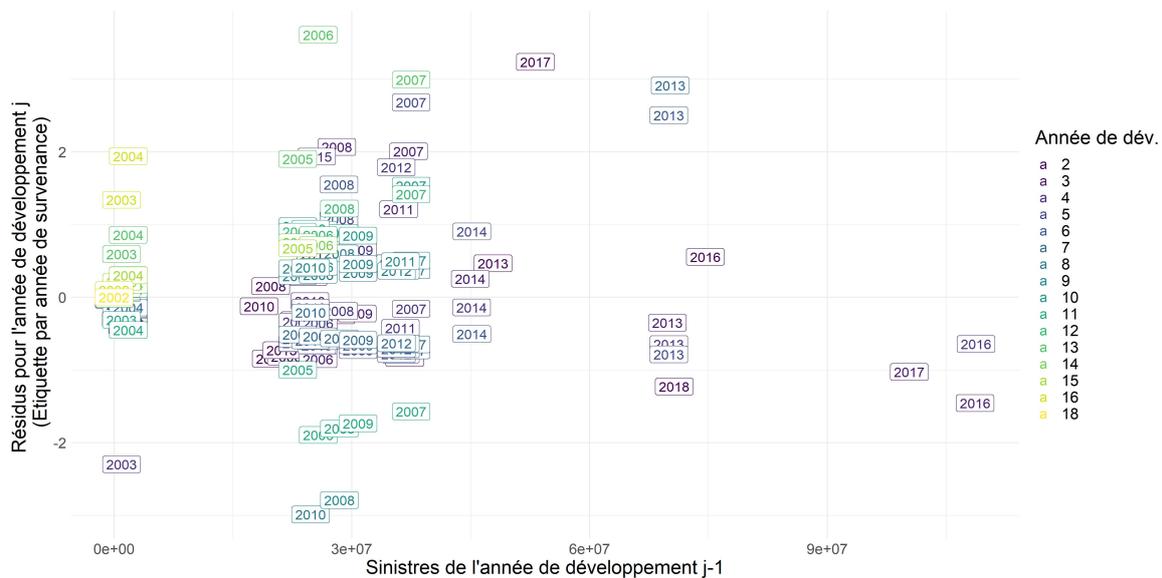
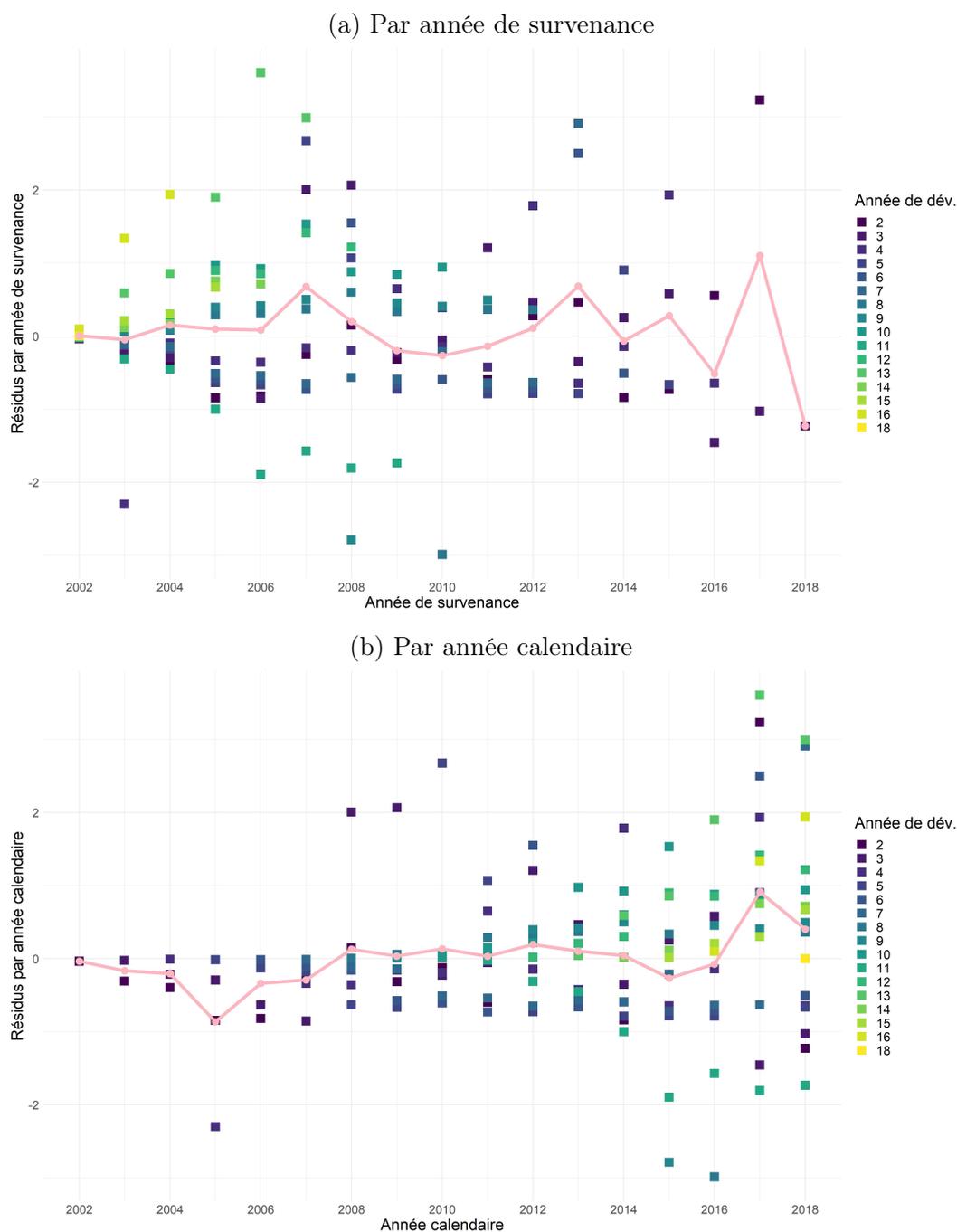


FIGURE 8.6 – Vérification de l’Hypothèse 3

Sur la Figure 8.6 ci-dessus, l’indépendance des résidus semble peu claire. En effet, les effets d’accroissement des montants des règlements nets au fil des années sont identifiables, via des groupes de résidus distincts.

Pour s'assurer de l'existence d'une structure aléatoire, nous allons donc procéder à des tests supplémentaires, menés dans le mémoire d'actuariat de Jaziri [29]. Nous traçons l'évolution de la moyenne des résidus par année de survenance, année calendaire et année de développement et vérifions que ceux-ci ne suivent aucune tendance temporelle. Pour rappel, l'inflation dans les réserves se traduit par des tendances calendaires ou par année de développement. A l'inverse, l'impact du climat est visible à travers des tendances sur les années de survenance.

Sur la Figure 8.7 suivante, nous observons que les résidus ne montrent pas de structure aléatoire. Tout d'abord, il semble se dessiner sur la Figure 8.7a une tendance à la baisse sur les dernières années de survenance, potentiellement due à des effets climatiques. Par ailleurs, les Figures 8.7b et 8.7c présentent une tendance de croissance au cours des dernières années calendaires et de développement. Ces hausses pourraient se justifier par des effets inflationnistes, qui altèrent le caractère indépendant des résidus.



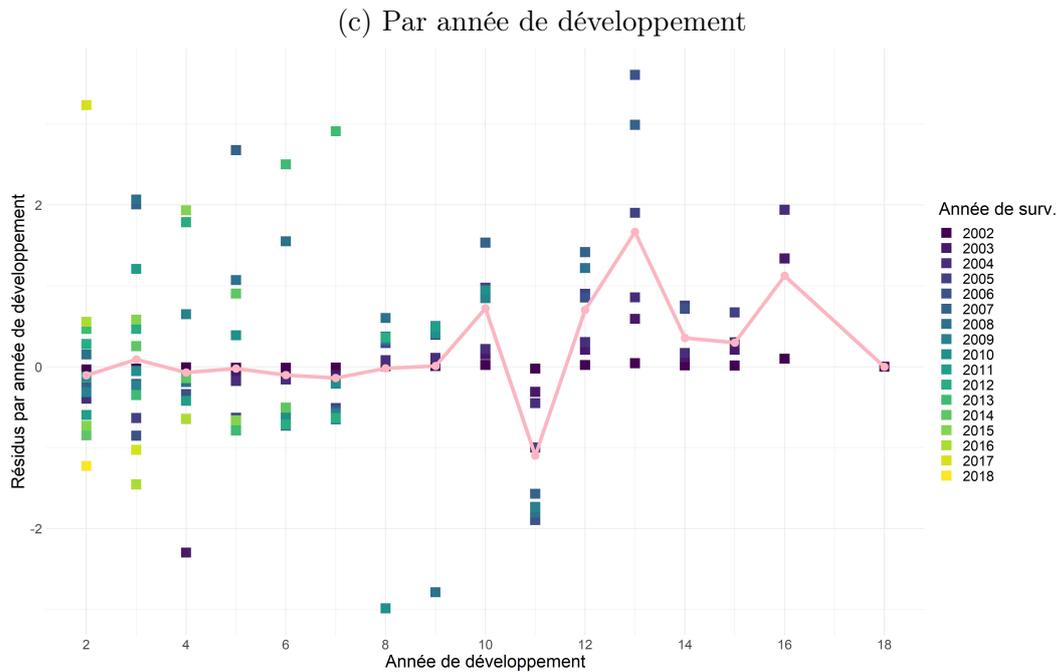


FIGURE 8.7 – Tests de stabilité des résidus

Pour plus de précisions, nous effectuons un test de blancheur sur les moyennes des résidus par année de survénance, année calendaire et année de développement. Les résultats des tests de Box-Pierce sont présentés en Table 8.1.

Tendance étudiée	p-value
Année de survénance	0,0555
Année calendaire	0,0838
Année de développement	0,9203

TABLE 8.1 – Tests de Box-Pierce sur le triangle aléas climatiques

Les p-values des séries de résidus moyenne par année de survénance et année calendaires sont inférieures à 10%. En conséquence, nous rejetons l’hypothèse de blancheur des résidus, et donc *a fortiori* l’**Hypothèse 3**. Nous pouvons toutefois tenter de recourir à une déflation du triangle afin d’ôter les effets inflationnistes qui entravent l’indépendance des résidus. Cette étude fera l’objet de la sous-section suivante. Avant cela, vérifions les hypothèses de Dal Moro sur le triangle aléas climatiques.

8.2.4 Validation de l’Hypothèse 4

Pour justifier cette hypothèse, il est nécessaire de tracer les courbes $y_{i,j} = \frac{(C_{i,j+1} - \hat{f}_j C_{i,j})^3}{C_{i,j}^{3/2}}$ pour chaque année de développement j en fonction de l’année de survénance i .

Aucune tendance en fonction de l’année de survénance i n’est observée sur la Figure 8.8 ci-après, donc l’**Hypothèse 4** est valide.

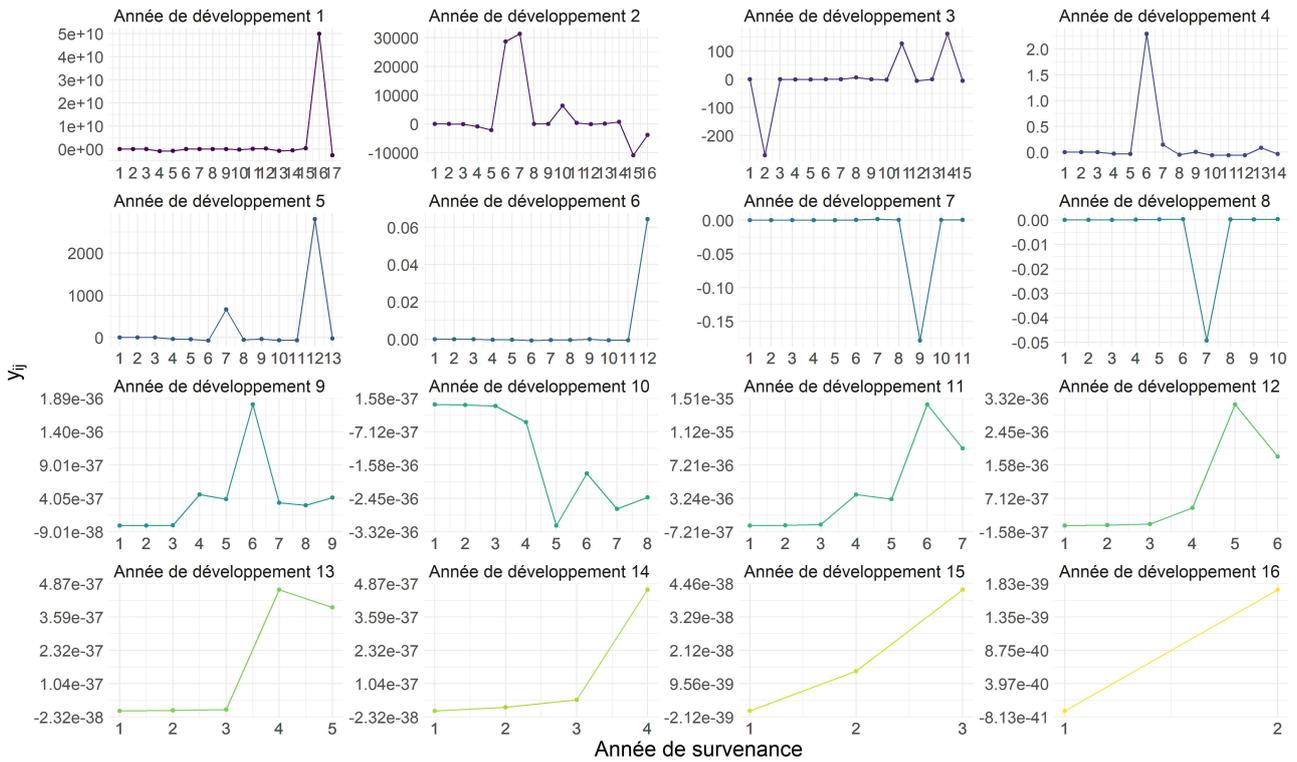


FIGURE 8.8 – Vérification de l’Hypothèse 4

8.2.5 Validation de l’Hypothèse 6

De manière analogue, cette hypothèse se vérifie en traçant les coefficients $y_{i,j} = \frac{(C_{i,j+1} - \hat{f}_j C_{i,j})^4}{C_{i,j}^2}$ pour chaque année de développement j en fonction de l’année de survivance i .

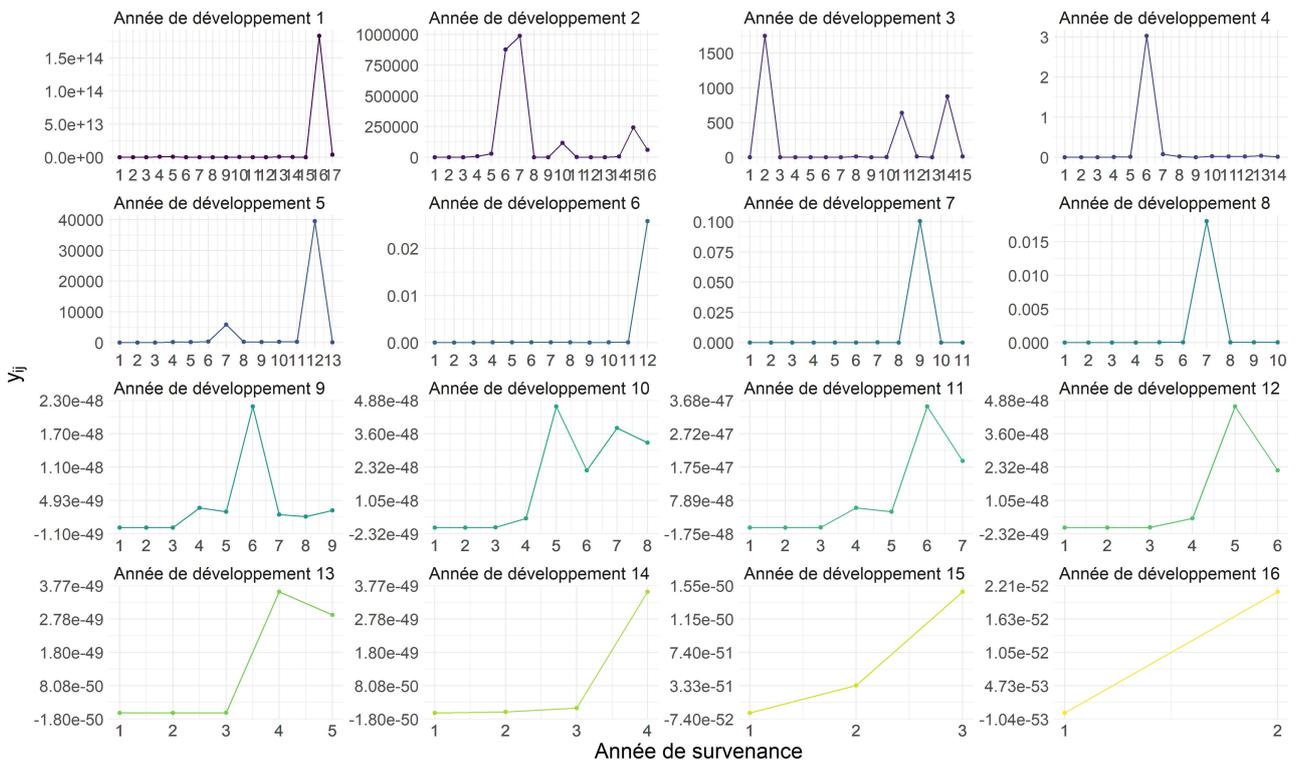


FIGURE 8.9 – Vérification de l’Hypothèse 6

Aucune tendance en fonction de l'année de survenance i n'est observée sur la Figure 8.9 précédente, donc l'**Hypothèse 6** est valide.

Comme indiqué précédemment, la validité des hypothèses (H5) et (H7) repose sur la validité de (H3), (H4) et (H6). Par conséquent, sous réserve de la validité de (H3), toutes les hypothèses sont vérifiées pour le triangle aléas climatiques. Pour s'en assurer, nous allons retraiter l'inflation sur ce triangle puis revérifier l'**Hypothèse 3** sur le triangle déflaté.

Compte tenu des résultats fournis en annexe D.1 et D.2, l'inflation ne semble pas avoir d'influence sur l'indépendance des résidus pour la branche responsabilité civile et celle excluant les aléas climatiques et tempêtes. Par conséquent, nous ne déflatons pas les règlements nets de ces deux triangles.

8.3 Déflation du triangle aléas climatiques

Afin d'ôter tous les effets relatifs à l'inflation, nous corrigeons les règlements nets du triangle agricole aléas climatiques via la formule de l'Equation (7.1). Ce mémoire ne portant pas sur la projection de l'inflation, nous nous contentons de raisonner par rapport aux valeurs de l'année suivant la dernière année de survenance du triangle, soit 2020.

8.3.1 Données d'inflation

Pour déflater le triangle climatique, nous avons choisi de tenir compte de l'Indice brut des Prix des Produits Agricoles à la Production, l'IPPAP. Dans la mesure où les assurances agricoles protègent contre une perte de rendement, tenir compte d'un indicateur de prix à la production global nous semble pertinent. Les données d'IPPAP sont disponibles sur le site de l'Agreste¹, organisme responsable des publications statistiques pour le Ministère de l'Agriculture et de la Souveraineté alimentaire.

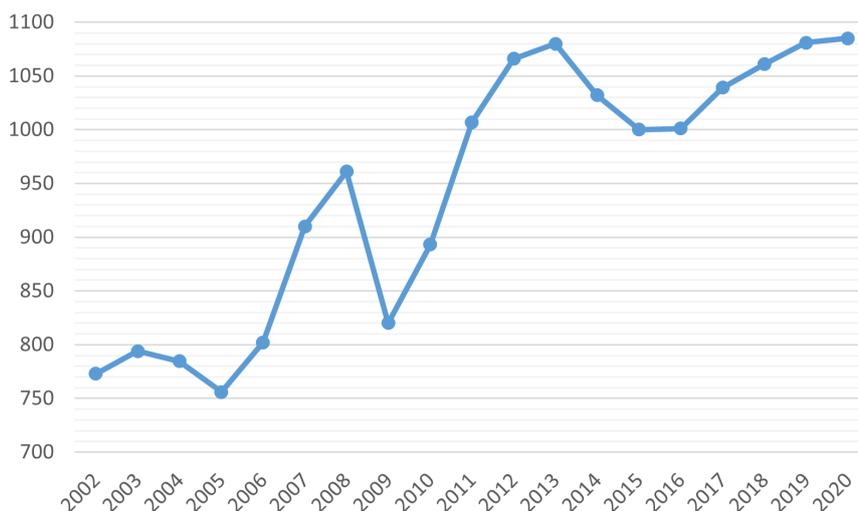


FIGURE 8.10 – Evolution de l'Indice des Prix des Produits Agricoles à la Production

D'après la Figure 8.10, la période d'étude est marquée par une forte croissance de l'IPPAP. Ce constat justifie la déflation du triangle aléas climatiques pour valider l'**Hypothèse 3**.

1. https://agreste.agriculture.gouv.fr/agreste-web/disaron/D_0037/detail/

8.3.2 Validation de l'hypothèse d'indépendance des résidus

Comme dans la section précédente, nous effectuons des tests de stabilité des résidus dans le temps. Leurs résultats sont présentés sur la Figure 8.11 ci-après.

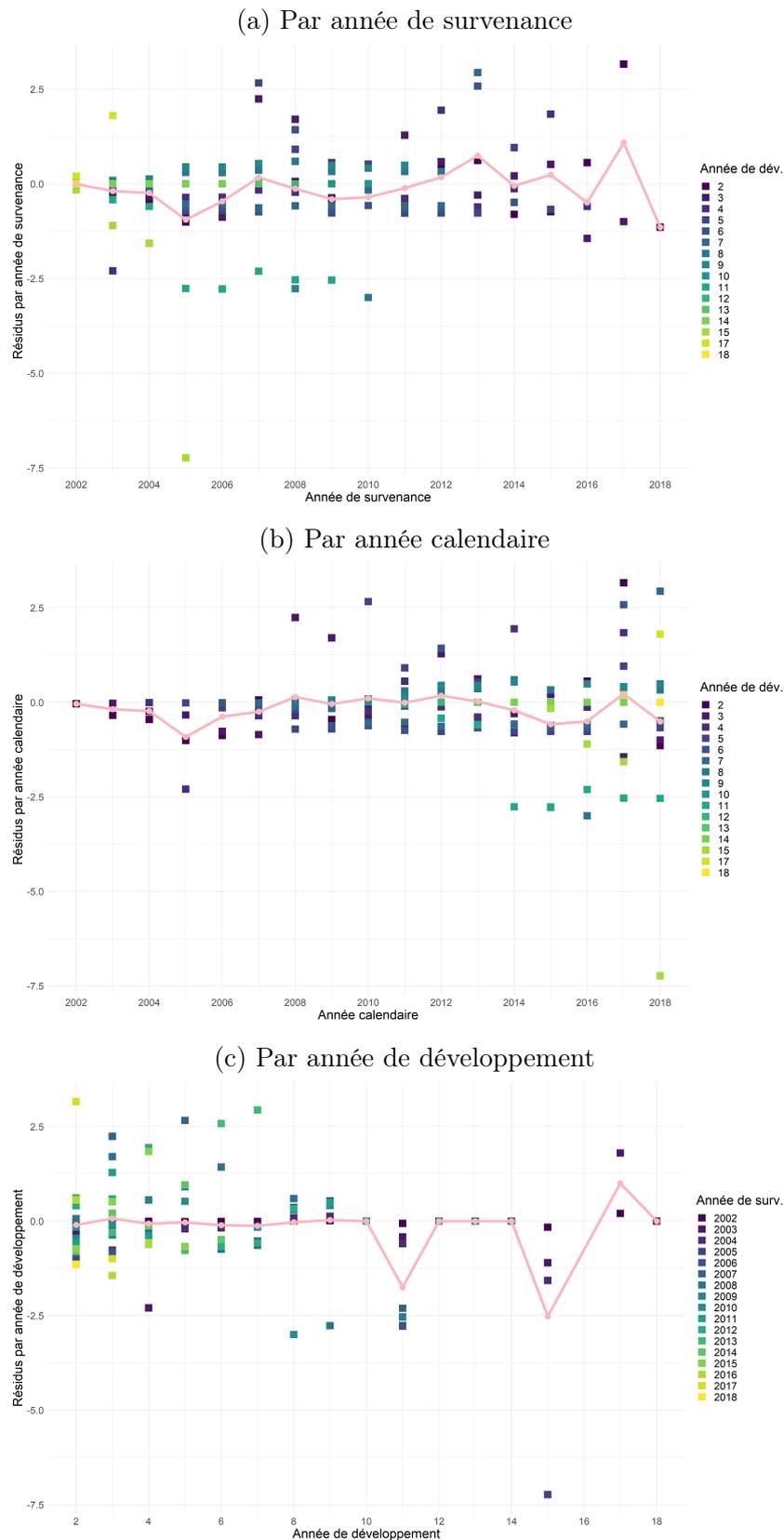


FIGURE 8.11 – Tests de stabilité des résidus pour le triangle déflaté

Il ressort des Figures 8.11b et 8.11c que le traitement de l'inflation a permis d'ôter les tendances en année calendaire et en année de développement pour les résidus. Toutefois, des effets d'année de survenance persistent comme le montre la Figure 8.11a. Finalement, cette dernière justifie l'application de techniques de séparation adaptées pour tenir compte d'effets en année de survenance, ici probablement liés au climat. Les résultats de cette étude seront présentés au Chapitre 10.

Pour avoir un indicateur plus quantitatif de l'indépendance des résidus, nous procédons à des tests de blancheur sur leurs moyennes par année de survenance, année calendaire et année de développement, dont les résultats sont présentés dans la Table 8.2.

Tendance étudiée	p-value
Année de survenance	0,3197
Année calendaire	0,3972
Année de développement	0,1620

TABLE 8.2 – Tests de Box-Pierce sur le triangle aléas climatiques déflaté

Au vu des p-values obtenues dans chaque cas considéré, nous ne pouvons rejeter l'hypothèse de blancheur des résidus. Cela nous conduit donc à valider l'**Hypothèse 3**, ce qui permettra d'appliquer les modèles de Mack *chain-ladder* et de Dal Moro afin de mener à bien notre étude sur les paramètres de la distribution des réserves.

Chapitre 9

Etude de la distribution des réserves

Dans ce chapitre, nous cherchons à identifier un éventuel effet du risque climatique sur les paramètres de forme de la distribution des réserves. Pour ce faire, nous appliquons les différentes méthodes de provisionnement présentées dans les Chapitres 4 et 5. Compte tenu de contraintes liées aux données, notre démarche est présentée à titre illustratif avec un triangle climatique et deux autres triangles non-climatiques. Néanmoins, l'analyse pourrait se faire de manière plus précise avec davantage de triangles climatiques. Par exemple, en restant sur la branche agricole, une étude de la déformation de la distribution pourrait s'effectuer par culture et péril considérés.

9.1 Estimation des réserves et de l'erreur de prédiction

Pour estimer le montant des réserves et l'erreur de prédiction, nous disposons de plusieurs méthodes : Mack *chain-ladder*, Wüthrich, le Bootstrap classique ou avec prise en compte du paramètre de skewness et la simulation par loi GPD. Pour les approches simulatoires, nous générons 1000 échantillons.

9.1.1 Estimation des réserves

Les résultats d'estimation des réserves relatifs à chacune des méthodes sont présentés sur la Figure 9.1 suivante.

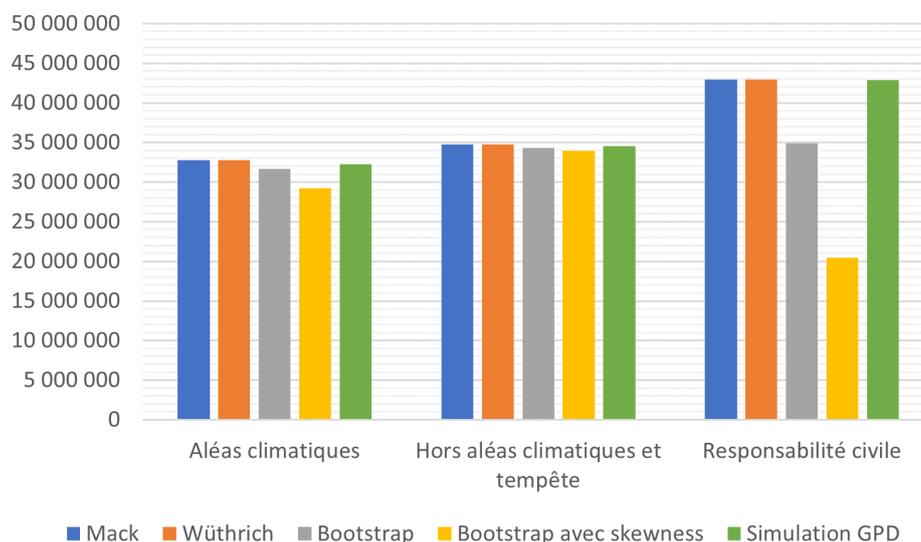


FIGURE 9.1 – *Best Estimates* obtenus

Nous constatons tout d'abord un montant de réserves plus important pour la branche de responsabilité civile. Cela est cohérent avec le fait que les deux autres triangles soient associés à des branches à développement plus court.

Sur ces exemples, nous remarquons que la méthode de bootstrap avec skewness donne des résultats bien différents de ceux obtenus par bootstrap classique. En particulier, ses estimations pour le triangle de responsabilité civile semblent aberrantes. A ce stade, nous n'avons pas d'explications pour justifier les écarts observés avec les autres méthodes. En conséquence, nous ne devrions pas accorder trop de valeur aux résultats suivants relatifs à cette approche.

9.1.2 Estimation de l'erreur de prédiction des réserves

Nous nous intéressons ensuite au risque de réserve induit dans chacun des modèles implémentés. Pour ce faire, nous rapportons l'erreur moyenne de prédiction commise au montant total de réserve estimé par la méthode de Mack *chain-ladder*.

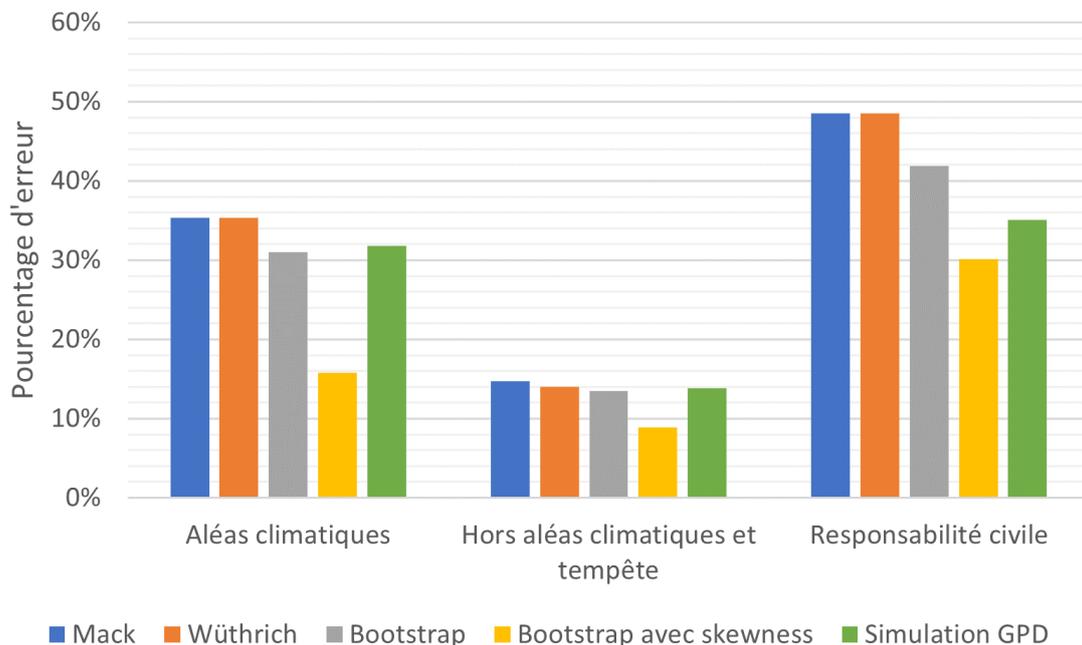


FIGURE 9.2 – Estimations du risque de réserve en pourcentage du montant des réserves obtenus par Mack *chain-ladder*

Selon la Figure 9.2, le triangle hors aléas climatiques et tempête est celui pour lequel l'erreur de prédiction est la plus reserrée autour du *Best Estimate*, se situant aux alentours de 14%. Quant au triangle aléas climatiques, les méthodes usuelles de Mack et Wüthrich engendrent une erreur de prédiction d'environ 35% du montant des réserves. Sur ce triangle, l'application de méthodes simulatoires de type bootstrap classique ou GPD permet une diminution notable de l'erreur de prédiction des réserves.

Concernant l'impact du climat sur l'estimation des réserves, il paraît cohérent que le triangle intégrant les aléas climatiques présente une plus forte variabilité que celui les excluant ainsi que les tempêtes. En effet, les événements naturels peuvent générer des pertes de rendements plus ou moins conséquentes selon la sévérité du phénomène subi. Toutefois, le triangle de responsabilité civile présente également une forte variabilité dans l'estimation des réserves. Par conséquent, le caractère climatique du triangle n'est pas le seul facteur affectant le moment d'ordre 2 de la distribution des réserves.

9.2 Considération de l'asymétrie et de l'aplatissement

Pour estimer le skewness et le kurtosis, nous disposons de méthodes de bootstrap, classique et avec prise en compte de paramètre de skewness, ainsi que la méthode de simulation GPD. Pour ces approches, nous effectuons 1000 simulations.

9.2.1 Estimation du skewness

L'approche par formule fermée de Dal Moro se rajoute aux méthodes d'estimations du skewness citées plus haut. Nous confrontons leurs résultats sur la Figure 9.3 suivante.

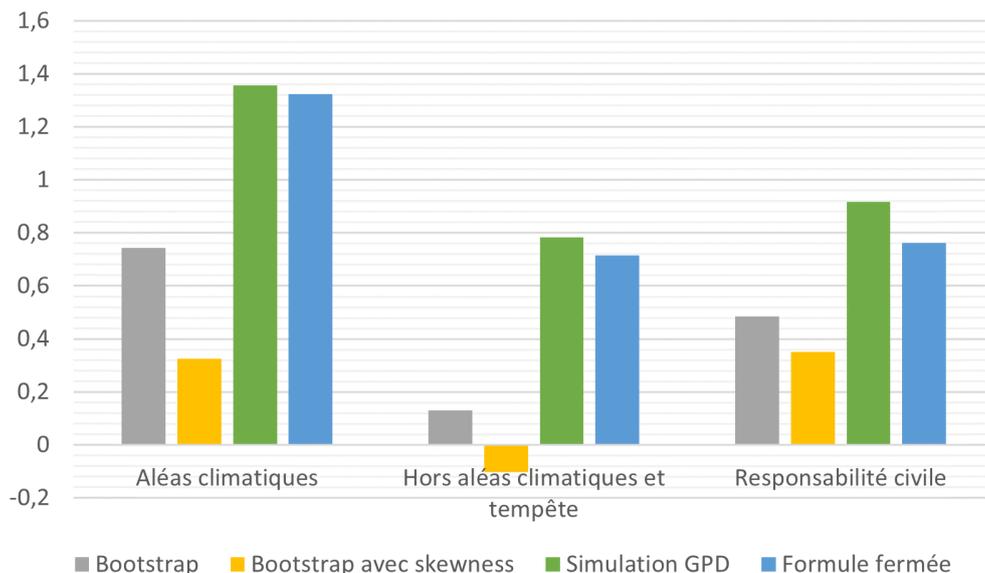


FIGURE 9.3 – Résultats de skewness obtenus

La Figure 9.3 nous apporte la confirmation que les méthodes de bootstrap ne sont pas adaptées pour évaluer le skewness de la distribution des réserves. En effet, seule la simulation par loi de Pareto généralisée procure des estimations proches des calculs par formule fermée. La méthode de bootstrap avec prise en compte du paramètre de skewness fournit quant à elle les résultats les plus éloignés, quel que soit le triangle considéré.

Concernant la valeur des paramètres en tant que telle, les trois distributions se décalent à gauche de leur médiane, dans la mesure où elles sont caractérisées par un skewness positif. Nous remarquons de plus que le triangle relatif aux aléas climatiques présente une asymétrie plus forte que les deux autres triangles : seul ce dernier possède un skewness supérieur à 1. Ce constat pourrait être lié au caractère climatique de ce triangle, mais il serait nécessaire de poursuivre l'étude sur des données émanant d'autres branches et d'autres assureurs pour confirmer cette conjecture.

9.2.2 Estimation du kurtosis

Pour le kurtosis, nous ne pouvons pas confirmer les estimations provenant des méthodes simulatoires avec une formule fermée. Néanmoins, les résultats du Chapitre 6 nous démontrent que l'approche par simulation GPD présentée par Dal Moro s'avère plus fiable que les bootstraps lorsque l'on considère les moments d'ordre 3 et 4 de la distribution des réserves.

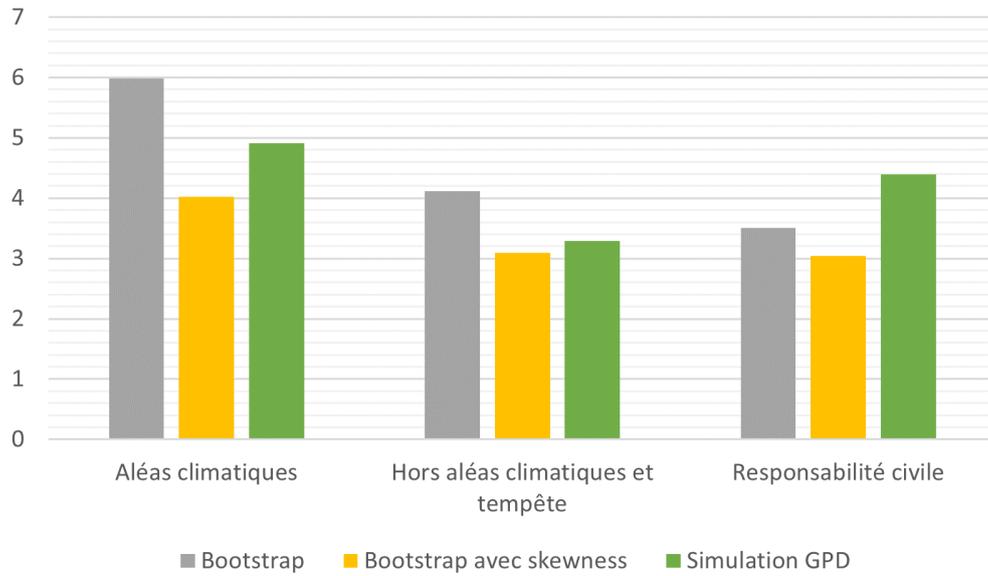


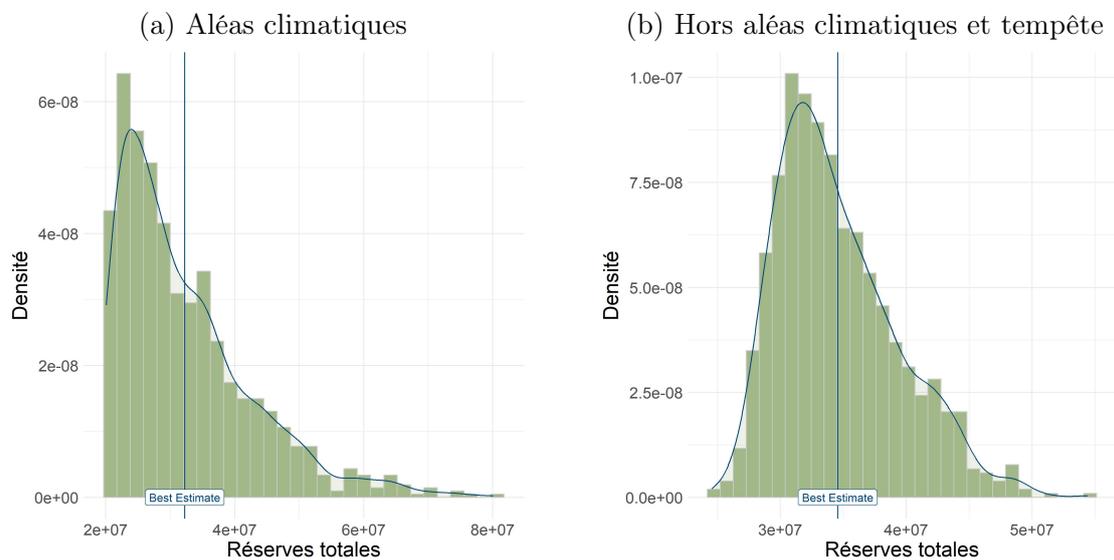
FIGURE 9.4 – Résultats de kurtosis obtenus

Considérant les résultats fournis par la simulation GPD présentés sur la Figure 9.4, la distribution triangle aléas climatiques est celle qui est la plus leptokurtique, avec un paramètre de kurtosis proche de 5 et donc supérieur à 3. En ce sens, la distribution est caractérisée par des queues très épaisses et de fréquents outliers. Nous pourrions vérifier cette forme sur la distribution des réserves générée par la simulation GPD et affichée dans la section suivante.

Nous pouvons donc conjecturer un impact positif du climat sur le paramètre d’aplatissement de la distribution des réserves : les événements climatiques conduiraient à un épaississement des queues de la distribution. Néanmoins, ce constat devrait être vérifié sur d’autres triangles émanant d’autres assureurs et relatifs à d’autres branches.

9.3 Distributions obtenues

Nous affichons ensuite les distributions générées par la simulation GPD pour chacun des triangles considérés.



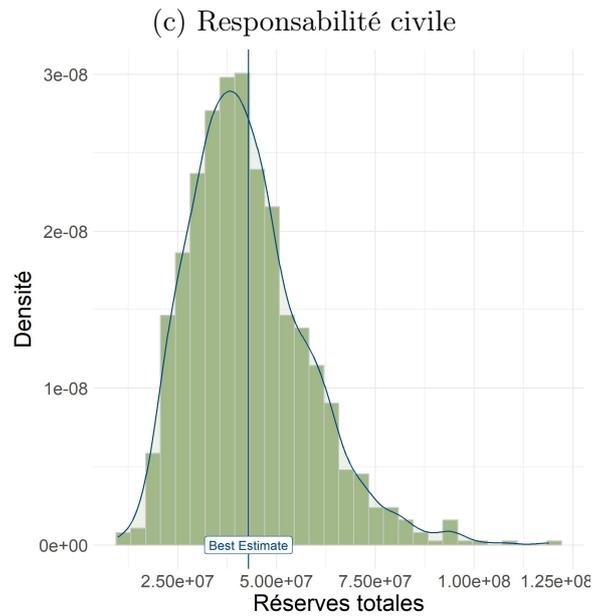


FIGURE 9.5 – Distributions des réserves obtenues par simulation GPD - 1000 simulations

D'après la Figure 9.5 précédente, les trois distributions sont caractérisées par une asymétrie positive, dans la mesure où celles-ci sont décalées vers la gauche. Par conséquent, ces observations sont cohérentes avec les estimations de skewness présentées à la section précédente. En outre, comme prévu par les estimations de kurtosis, tous les triangles possèdent des distributions de réserves à queues épaisses.

Chapitre 10

Application de la méthode de séparation climatique

Dans ce chapitre, nous appliquons la technique de séparation climatique présentée au Chapitre 7. En vue de ne capter qu'une tendance par année de survenance, nous utilisons le triangle aléas climatiques déflaté de l'IPPAP. Nous allons ensuite extraire une tendance d'inflation par année de survenance, que nous tenterons de corrélérer à des variables climatiques. Néanmoins, compte tenu de l'absence de données de triangle de nombre de sinistres, nous avons dû formuler des hypothèses afin de simuler ce dernier. Rappelons également que l'un des enjeux de ce mémoire est d'utiliser de nouvelles méthodologies et que cet exemple d'application est donné à titre d'illustration.

10.1 Hypothèses de simulation

10.1.1 Hypothèses formulées

Pour simuler le triangle de nombre de sinistres, nous avons supposé un nombre de contrats M et une proportion de sinistres α constants. De plus, nous avons fixé les proportions de sinistres par année de développement r_j . De là, nous avons formulé deux hypothèses possibles sur le triangle de nombre de sinistres $(N_{i,j})$:

(H1) Le nombre de sinistres moyen **dépend de la sinistralité annuelle** par la relation suivante :

$$\mathbb{E}[N_{i,j}] = r_j \times SP_i \times \alpha M$$

avec SP_i ratio sinistres sur primes de l'année de survenance i .

A travers cette hypothèse, on suppose alors que le coût unitaire des sinistres est constant sur toute la période d'étude et que les surcoûts sont liés à une sur-fréquence de sinistres. Les nombres de sinistres $N_{i,j}$ sont ensuite simulés par des lois de Poisson de paramètre $\mathbb{E}[N_{i,j}]$.

(H2) Le nombre de sinistres moyen est **indépendant de la sinistralité annuelle** et s'écrit :

$$\mathbb{E}[N_{i,j}] = r_j \times \alpha M.$$

Autrement dit, le nombre de sinistres moyen est constant sur la période d'étude. Les surcoûts s'expliquent par des sinistres de plus forte sévérité. Les nombres de sinistres $N_{i,j}$ sont ensuite simulés par des lois de Poisson de paramètre $\mathbb{E}[N_{i,j}]$.

Le choix des paramètres r_j a été effectué de sorte à ce que le caractère court du développement de la branche considérée soit respecté. Par conséquent, nous avons calculé les r_j de la manière suivante :

$$r_j = \begin{cases} 0,85 & \text{pour } j = 1 \\ \frac{1-r_1}{2} & \text{pour } j = 2 \\ \frac{r_{j-1}}{2} & \text{pour } 3 \leq j \leq I \end{cases}$$

10.1.2 Limites

Nous avons conscience que les hypothèses formulées sont très simplistes et ne peuvent refléter la réalité du portefeuille dont nous disposons. Néanmoins, compte tenu du peu d'informations que nous possédons, il était particulièrement complexe d'affiner ces hypothèses. En particulier, nous avons souhaité prendre en compte l'accroissement de la taille du portefeuille dans la modélisation. Nous avons donc formulé deux hypothèses supplémentaires, découlant de (H1) et (H2), avec un nombre de contrats M_i dépendant du temps :

(H3) Le nombre de sinistres moyen **dépend de la sinistralité annuelle** par la relation suivante :

$$\mathbb{E}[N_{i,j}] = r_j \times SP_i \times \alpha M_i$$

avec SP_i ratio sinistres sur primes de l'année de survenance i .

Les nombres de sinistres $N_{i,j}$ sont ensuite simulés par des lois de Poisson de paramètre $\mathbb{E}[N_{i,j}]$.

(H4) Le nombre de sinistres moyen est **indépendant de la sinistralité annuelle** et s'écrit :

$$\mathbb{E}[N_{i,j}] = r_j \times \alpha M_i.$$

Les nombres de sinistres $N_{i,j}$ sont ensuite simulés par des lois de Poisson de paramètre $\mathbb{E}[N_{i,j}]$.

Toutefois, l'utilisation de ces hypothèses dans la méthode de séparation climatique nous a fourni des résultats très différents de ceux obtenus avec (H1) et (H2). En particulier, les montants de réserves estimés étaient très éloignés de ceux de Mack *chain-ladder*. De plus, le facteur d'inflation « climatique » présentait une forme très différente : il n'y avait plus aucune tendance croissante, ce qui faussait nos conjonctures sur l'impact du changement climatique sur les réserves. Or, en ayant testé les quatre hypothèses sur des triangles totalement fictifs, dont les résultats sont disponibles en Annexe C, nous sommes parvenus à obtenir la même forme pour le facteur d'inflation « climatique ». Ainsi, ce n'était pas tant le modèle qui posait problème, mais bien les hypothèses qui n'étaient pas adaptées aux données.

10.2 Résultats du modèle

Le modèle de séparation climatique présenté au Chapitre 7 fournit trois résultats :

- le montant des réserves par année de survenance et total,
- la proportion de sinistre par année de développement r_j ,
- et la tendance d'inflation par année de survenance, nommée tendance climatique λ_i .

En adaptant les hypothèses du papier de Björkwall et al. [7], nous pouvions également obtenir une estimation de l'erreur de prédiction ainsi qu'une distribution des réserves. Or, nos hypothèses de modélisation du nombre de sinistres étant simplistes, nous n'avons pas obtenu de résultats concluants sur ces deux aspects. Par conséquent, nous présentons uniquement les trois points évoqués plus haut.

10.2.1 Estimation des réserves

Tout d’abord, intéressons-nous à l’estimation des réserves via le modèle de séparation climatique pour chacune des hypothèses, et comparons ces résultats à ceux du modèle de Mack *chain-ladder*.

Année i	Réserves Mack	Réserves H1	Réserves H2	Ecart H1 - Mack	Ecart H2 - Mack
1 à 10	0	0	0	0	0
11	-350	-399	-350	+14,00%	+0,138%
12	-1 473	-2 000	-1 474	+35,78%	+0,096%
13	-334	-777	-336	+132,47%	+0,518%
14	16 526	14 325	16 514	-13,32%	-0,078%
15	74 298	64 564	74 232	-13,10%	-0,089%
16	85 865	89 002	85 834	+3,65%	-0,036%
17	358 419	388 008	358 277	+8,26%	-0,040%
18	32 228 920	29 954 716	32 240 983	-7,06%	+0,037%
Total	32 761 872	30 507 438	32 773 681	-6,88%	+0,036%

TABLE 10.1 – Montant de réserves obtenu par chaque année de survenance et au total pour chaque hypothèse et écarts avec les estimations de Mack

Nous constatons sur la Table 10.1 que les sinistres sont si vite développés qu’il n’y a pas de provisions nécessaires à prévoir pour les dix premières années de survenance. De plus, les montants de réserves obtenus pour l’hypothèse (H2) sont bien plus proches de ceux de Mack. En effet, cette hypothèse conduit à surestimer les provisions totales de seulement 12 000 €, là où la première mène à une sous-estimation de plus de deux millions d’euros. De manière générale, le choix de l’hypothèse (H2) restreint à moins de 1% les écarts avec la méthode de Mack *chain-ladder*.

10.2.2 Proportion de sinistres par année de développement

La Table 10.2 ci-après confronte les proportions de sinistres estimées par année de survenance pour chaque hypothèse avec celles pré-établies lors de l’étape de simulation du triangle de nombre de sinistres. Nous affichons uniquement les cinq premiers r_j , dans la mesure où les autres sont très faibles, du fait du court développement du triangle.

Force est de constater que si les coefficients estimés pour chaque hypothèse sont proches, ils sont toutefois éloignés des proportions imposées lors de la simulation. Cependant, ces résultats devraient s’améliorer lors d’une application sur des données réelles. En outre, l’enjeu de la modélisation consiste surtout à extraire la tendance d’inflation « climatique » pour la corrélérer à des variables météorologiques. L’étude précise de la proportion de sinistres par année de survenance ne fait quant à elle pas l’objet de ce mémoire.

Année j	r_j théorique	r_j H1	r_j H2
1	0,850000	0,739216	0,724790
2	0,075000	0,256307	0,271075
3	0,037500	0,003604	0,003293
4	0,018750	0,000301	0,000184
5	0,009375	0,000052	0,000058

TABLE 10.2 – Comparaison des proportions de sinistres obtenues pour les cinq premières année de développement

10.2.3 Extraction de la tendance climatique

La Figure 10.1 ci-dessous présente l'évolution du facteur d'inflation « climatique » obtenu par année de survenance.

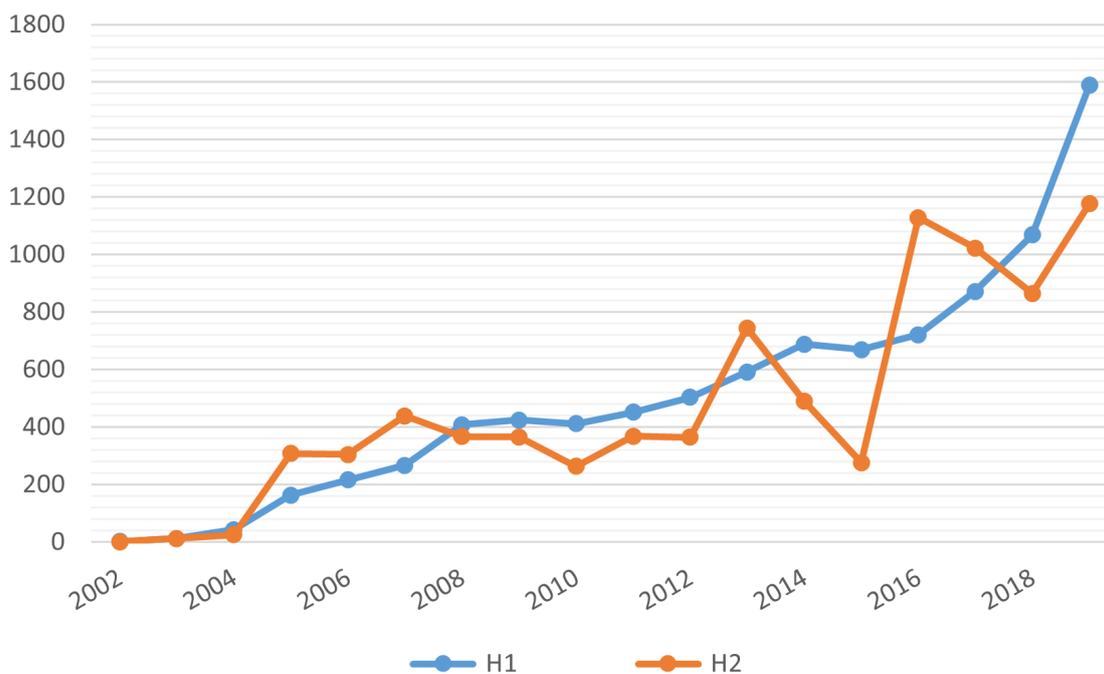


FIGURE 10.1 – Evolution du facteur d'inflation « climatique » λ_i

L'une des premières remarques qui peut être faite concerne la monotonie des deux courbes. Dans chaque cas, nous sommes bien face à un facteur d'inflation qui croît avec l'année de survenance. Cela est donc cohérent avec la conjecture que nous souhaitons vérifier : les événements climatiques ont bien une importance de plus en plus forte chaque année.

Néanmoins, les deux courbes ne sont pas strictement identiques. Si la tendance obtenue avec l'hypothèse (H1) observe une croissance sans à-coups, l'augmentation du facteur capté avec l'hypothèse (H2) est beaucoup plus saccadée. En fait, dans ce dernier cas, le facteur suit les évolutions du ratio sinistres sur primes. Ainsi, les points de 2013 et 2016 sont justifiés par des ratios supérieurs à 120%, quand le pic à la baisse observé en 2015 est lié à un ratio d'environ 40%. A l'inverse, le choix de tenir compte des ratios de sinistres sur prime dans l'hypothèse (H1)

conduit finalement à décorrélérer le facteur d'inflation des variations climatiques. En conséquence, pour mieux capter ces événements, il serait préférable d'utiliser la tendance extraite dans le cadre de l'hypothèse (H2).

10.3 Prise en compte de variables climatiques

10.3.1 Variables climatiques retenues

10.3.1.1 Difficultés rencontrées

Afin de mieux comprendre l'évolution du facteur d'inflation « climatique », il était crucial de choisir une ou plusieurs variables climatiques à relier à ce facteur. Toutefois, les données clients agricoles que nous avons utilisées nous ont contraint à réduire nos exigences sur les variables climatiques pour trois raisons :

- la **granularité spatiale**. Tout d'abord, le triangle étant fourni à la maille de la France entière, il est nécessaire de choisir des variables climatiques qui garderaient du sens lorsqu'elles sont moyennées à l'échelle nationale.
- la **granularité temporelle**. De plus, le triangle de provisionnement est calculé de manière annuelle. Dès lors, il est délicat de choisir des variables qui connaissent de fortes fluctuations dans l'année, sauf si l'on émet des hypothèses afin de ne considérer qu'une période de l'année en particulier.
- le **manque d'information sur les sinistres**. Enfin, lorsque les sinistres sont enregistrés dans la base pour fournir le triangle, l'événement climatique déclencheur n'est lui pas connu. En effet, seules les pertes de rendements sont déclarées, sans que la cause du sinistre ne soit spécifiée. Ainsi, notre triangle de provisionnement est certainement composé de sinistres liés au gel, à la grêle ou à la sécheresse sans que nous en connaissions l'importance. Par conséquent, au vu des données, il paraît complexe d'implémenter un modèle de prévisions tenant compte de plusieurs événements climatiques.

A ces problèmes liés au triangle client se rajoutent des difficultés relatives à la recherche de données météorologiques :

- D'une part, l'**accès aux données** météorologiques est souvent complexe : peu de bases sont disponibles gratuitement et certaines sont parfois très incomplètes. En particulier, lors de nos recherches, les bases de grêle téléchargeables en libre accès que nous avons étudiées étaient peu détaillées, en manquant par exemple d'informations sur la taille des grêlons.
- D'autre part, l'**existence de scénarios de projection** des variables considérées est un critère de choix important dans le cadre de notre étude. En effet, si ce mémoire porte sur le lien entre tendance d'inflation climatique capté dans les réserves et variables météorologiques, l'objectif est également de compléter l'analyse avec des projections de cette tendance selon plusieurs scénarios. Ainsi, il n'était pas possible d'exploiter des données de gel, par manque de scénarios de projection.

Au vu de ces différents éléments, nous nous sommes concentrés sur les pertes de rendements liées à l'état du sol, qu'il connaisse un surplus d'eau ou une sécheresse. Nous avons donc fait le choix d'intégrer une unique variable, un indicateur d'humidité des sols, dont les données sont disponibles en libre accès sur le site de Météo France, combinées à des prévisions fournies sur le portail de génération de scénarios du NGFS.

10.3.1.2 Données historiques

Définition de l'indicateur

Nous avons donc à notre disposition des données d'humidité des sols, issues de Météo France¹. L'indicateur en question est le *Soil Wetness Index* (SWI). Il se définit de la manière suivante :

$$SWI = \frac{W - W_{wilt}}{W_{wilt} - W_{fc}}$$

avec W contenu en eau du sol, W_{wilt} contenu en eau du sol au **point de flétrissement** et W_{fc} contenu en eau du sol à la **capacité du champ**. Si le point de flétrissement correspond au seuil à partir duquel la plante ne peut plus capter l'eau du sol, la capacité du champ exprime quant à elle le seuil au-dessus duquel il y a ruissellement, et non plus de drainage gravitationnel dans le sol.

De par sa définition, cet indicateur permet d'évaluer l'état de la réserve en eau du sol par rapport à la réserve utile $W_{wilt} - W_{fc}$:

- Si le SWI est proche voire supérieur à 1, le sol est humide et peut aller jusqu'à la saturation en eau.
- Si le SWI tend vers 0 voire devient négatif, le sol est en état de stress hydrique et peut s'assécher fortement.

Limite d'utilisation de la base

Météo France calcule cet indicateur via le modèle SIM pour le dispositif Catastrophes Naturelles et évalue l'état de la réserve en eau sur la profondeur de 2 mètres. Il est précisé que ces données ne peuvent pas être utilisées dans le cadre d'études relatives à l'humidité des sols superficiels. Bien que la sécheresse agricole soit caractérisé par un déficit hydrique sur une profondeur de 1 à 2 mètres, nous décidons malgré tout d'avoir recours à cette base. En effet, les deux autres jeux de données que nous avons trouvés, provenant de Copernicus, avaient également certaines limites :

- La première base² se concentre sur l'humidité des couches de sols de surface pour une épaisseur ne dépassant pas les 5 centimètres. Par conséquent, ces informations n'auraient pas été pertinentes pour notre analyse.
- La seconde, issue du modèle ERA5³, était bien plus complète, en fournissant plusieurs variables relatives à la volumétrie en eau du sol selon plusieurs profondeurs. Néanmoins, cette base ne renseigne pas sur l'indicateur d'humidité SWI en tant que tel. Dès lors, il n'aurait pas été possible de projeter ses valeurs à l'aide de scénarios climatiques. En effet, la seule base offrant des projections de volumétries⁴, issue de Copernicus, ne distingue pas les prévisions par profondeur de sol considérée. De plus, seuls les scénarios RCP 2.6 et RCP 8.5 y sont présentés.

Traitement des données

La base Météo France fournit des valeurs mensuelles de l'indicateur SWI pour la période 1969-2021. Chaque valeur tient compte du mois en cours ainsi que des deux mois précédents, en moyennant les moyennes mensuelles des valeurs quotidiennes du SWI. Cette base offre également une très grande granularité spatiale : la France métropolitaine est découpée en 8 981 mailles de 8 kilomètres de côté.

1. https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=301&id_rubrique=40

2. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture>

3. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means>

4. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-water-hydrological-change>

Pour obtenir une série de données de SWI à l'échelle nationale, nous avons moyenné les données de mailles relatives au mois de **mai**. Pour justifier ce choix, il faut tout d'abord revenir au risque de pertes de rendements agricoles. Les cultures sont particulièrement exposées aux aléas climatiques lors de leur croissance. Or, la phase de développement des cultures que nous avons mentionnées dans le Chapitre 3, telles que le blé ou l'orge, se déroule au printemps. Ainsi, utiliser l'indicateur relatif au mois de mai permet de tenir également compte des données d'humidité des sols des mois de mars et d'avril, qui pourraient expliquer l'observation de pertes de rendements.

Analyse des données

La Figure 10.2 ci-après présente l'évolution de l'indicateur national d'humidité des sols entre 2000 et 2020.

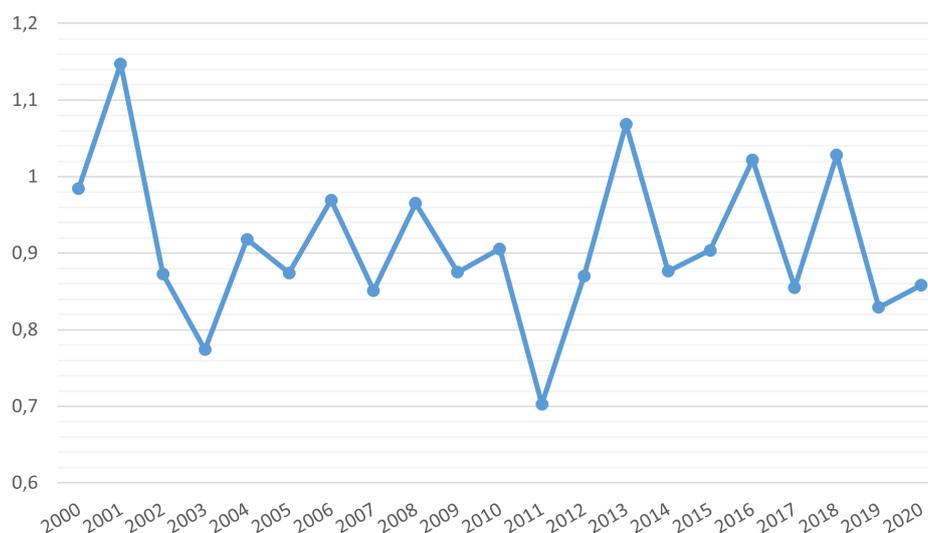


FIGURE 10.2 – Evolution de l'indice moyen d'humidité des sols entre 2000 et 2020 pour le mois de mai

A l'échelle du territoire, l'état d'humidité des sols semble correct. Certaines années sont toutefois marquées par une saturation en eau, comme l'année 2016 par exemple. En effet, en raison d'un printemps très pluvieux qui entraîna la prolifération de mauvaises herbes et de nuisibles qui détériorent les cultures, les récoltes de cette année furent particulièrement mauvaises [17]. Les stress hydriques subis dans certaines localités ne sont pas visibles lorsque les valeurs du SWI sont moyennées. Néanmoins, la localisation géographique du portefeuille client étant inconnue, il paraît complexe d'affiner les données météorologiques pour mieux refléter la situation de ce dernier.

10.3.1.3 Données de projection

Si les données historiques d'indice d'humidité des sols proviennent de Météo France, leur site web ne disposait pas de projections selon des scénarios climatiques. Par conséquent, nous avons complété ces données historiques par des données émanant du *Climate Impact Explorer*⁵, une banque de données de projections développée par *Climate Analytics* en collaboration avec le NGFS. De nombreux indicateurs climatiques y sont disponibles pour une multitude de pays, de manière bidécennale à partir de 2015, et modélisés avec les scénarios suivants :

- les scénarios du GIEC **RCP 2.6**, **RCP 4.5**, **RCP 6.0** et **RCP 8.5**,

5. <https://climate-impact-explorer.climateanalytics.org/>

- les scénarios du NGFS **Politiques actuelles**, **Transition retardée** et **Net Zéro 2050**,
- et enfin un autre scénario de **Politiques actuelles**, développé par l'organisme indépendant *Climate Action Tracker*, qui suit les actions climatiques des gouvernements des 39 pays du monde qui représentent 85% des émissions totales de gaz à effets de serre.

La granularité des données est toutefois limitée : du point de vue temporel, la plus grande précision est associée à des prévisions trimestrielles (mars à mai, juin à août, septembre à novembre et décembre à février), tandis que la plus petite maille spatiale est régionale.

Dans le cas de cette étude, pour rester en cohérence avec nos choix temporels pour les données historiques, nous choisissons les prévisions relatives aux mois de mars à mai à l'échelle nationale. En vue d'utiliser ces données dans la prédiction du facteur d'inflation, nous interpolons linéairement les projections afin d'obtenir un indicateur à horizon un an plutôt qu'à horizon cinq ans. Nous présentons l'évolution de l'indice d'humidité des sols médian selon les divers scénarios avec leurs intervalles de confiance à 95% sur la Figure 10.3 suivante.

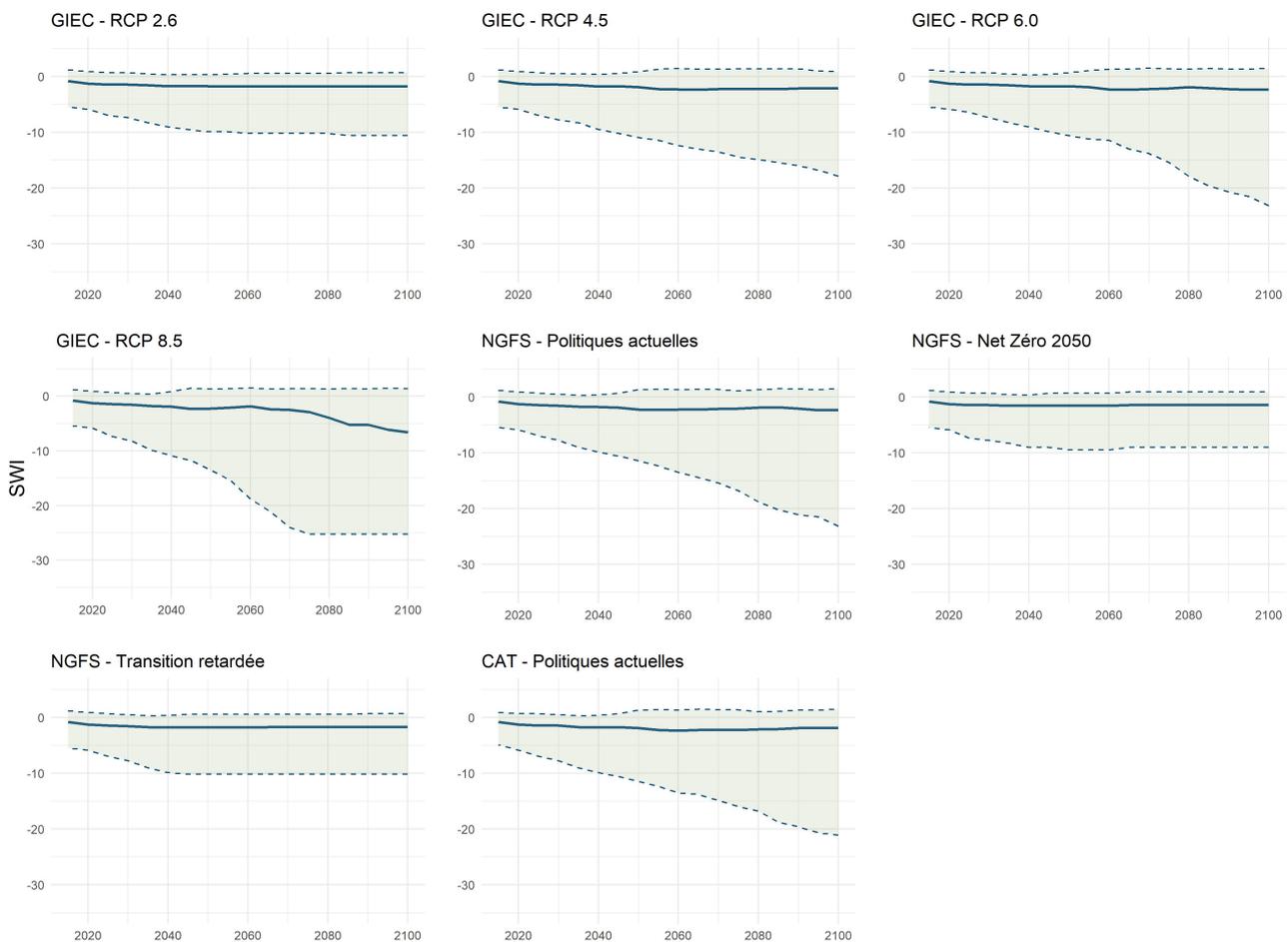


FIGURE 10.3 – Evolution de l'indice d'humidité des sols selon divers scénarios

Tout d'abord, nous pouvons noter que les scénarios supposant un réchauffement de la température de surface plus élevé, tels que les **Politiques actuelles** ou les **RCP 6.0** et **RCP 8.5**, sont marqués par une plus grande incertitude sur la valeur de l'indice d'humidité des sols.

De plus, dans tous les scénarios, la médiane de l'indicateur est négative, y compris pour les années antérieures à 2020, dont les données historiques sont fournies par Météo France. Si en moyenne, à l'échelle nationale, nous n'avons pas observé historiquement de SWI négatif, ce fut néanmoins le cas pour certaines mailles. En outre, selon le *Climate Impact explorer*,

l'indicateur quantifie le contenu en eau sur une profondeur d'environ un mètre, ce qui pourrait donc expliquer cette différence de niveau d'humidité. Ce problème de congruence entre données historiques et données de projection est très courant et dans notre cas, il n'a pas été possible de le contourner. Par ailleurs, l'objectif est de projeter le facteur d'inflation « climatique » en vue d'obtenir un ordre de grandeur. En conséquence, c'est surtout de la tendance d'évolution globale de l'indicateur plutôt que ses valeurs absolues qui nous intéresse. Enfin, les modèles fournissent des intervalles de confiance sur des valeurs futures. Pour les années les plus proches, il restera toujours plus fiables de se fier aux données réelles historiques si elles sont disponibles plutôt qu'aux valeurs modélisées.

10.3.2 Lien avec la tendance climatique

Nous avons ensuite cherché à corréliser les séries climatiques avec le facteur d'inflation climatique extrait. Les résultats obtenus pour les deux premières hypothèses de simulation sont présentés sur la Table 10.3, accompagné des résultats obtenus sans déflation du triangle.

Hypothèse	Triangle non-déflaté		Triangle déflaté	
	H1	H2	H1	H2
Corrélation SWI_i	0,129	0,2890	0,150	0,315
Corrélation SWI_{i-1}	0,163	0,178	0,162	0,178
Corrélation SWI_{i-2}	-0,122	-0,313	-0,134	-0,345

TABLE 10.3 – Corrélation de la tendance climatique avec la variable retenue pour les triangles déflaté et non-déflaté

Tout d'abord, nous pouvons remarquer que le choix de l'hypothèse (H2) conduit à de plus fortes corrélations entre le facteur de tendance climatique et l'indice d'humidité des sols. Toujours sous l'hypothèse (H2), les corrélations augmentent légèrement si l'on déflate le triangle de liquidation au préalable.

Concernant les signes des corrélations, il semblerait que le facteur de tendance climatique croisse avec l'indice d'humidité des sols et sa série retardée d'une année. Cela paraît contre-intuitif, dans la mesure où un indice proche de 0 voire négatif est synonyme de sécheresse. Néanmoins, rappelons que le triangle aléas climatiques inclut également des dommages relatifs aux intempéries, telles que la grêle, le gel ou les inondations, qui pourraient être liés positivement à l'indice d'humidité des sols. De plus, nos estimations du SWI étant calculées à l'échelle de la France entière, n'ayant pas de maille plus précise sur la sinistralité du triangle, nous risquons de manquer des effets régionaux qui pourraient altérer la corrélation entre l'indicateur climatique et la tendance inflationniste.

A l'inverse, le facteur de tendance climatique est négativement corrélé avec SWI_{i-2} . C'est d'ailleurs avec la variable doublement retardée que nous observons la plus forte dépendance entre les deux séries. Par conséquent, cette série est la plus à même de permettre l'obtention de résultats significatifs pour un modèle linéaire.

Afin de mieux capter le changement climatique dans l'estimation du modèle linéaire, nous décidons d'intégrer l'année de survenance dans la régression. Nous avons également tenté d'intégrer l'indice d'humidité des sols ainsi que ses séries retardées, mais les coefficients obtenus étaient trop peu significatifs et les prédictions trop peu adéquates. D'autres modèles, visant à

distinguer les effets d'un SWI supérieur à 1 d'un SWI en-deçà de cette valeur seuil ont également été lancés, sans résultat. Les résultats de ces régressions sont disponibles en Annexe E. Finalement, nous avons retenu le modèle linéaire suivant :

$$\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma \times \text{SWI}_{i-2}.$$

Les résultats de la régression linéaire sont présentés sur la Table 10.4.

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,125e+05	1,660e+04	-6,776	6,25e-06
Année _i	56,55	8,227	6,873	5,29e-06
SWI _{i-2}	-836,9	426,4	-1,963	0,0685

R² ajusté : 0,7595

TABLE 10.4 – Résultats du modèle de régression linéaire retenu

Tous les coefficients sont significatifs au seuil de 10%. Le R² ajusté de la régression linéaire étant de 0,76, la prédiction est de bonne qualité. En outre, les signes des coefficients relatifs à l'année et à l'indice d'humidité des sols sont cohérents. En effet, le facteur d'inflation « climatique » a tendance à croître avec le temps, dans la mesure où les événements naturels prennent davantage d'ampleur en raison du changement climatique. Par ailleurs, un indice d'humidité faible est propice à la survenance de sinistres, dans la mesure où des sols secs offrent de moins bons rendements, d'où le signe négatif devant le coefficient du SWI dans la régression linéaire. Toutes choses égales par ailleurs, une diminution d'une unité du SWI conduit à une augmentation de la tendance climatique d'environ 840. D'une année sur l'autre, sous réserve de la stabilité de l'indice d'humidité des sols, cette tendance va augmenter de 56,55.

Pour illustrer les performances du modèle linéaire, nous confrontons sur la Figure 10.4 ses prédictions ainsi que leurs intervalles de confiance à 95% avec l'estimation de la tendance extraite de la méthode de séparation.

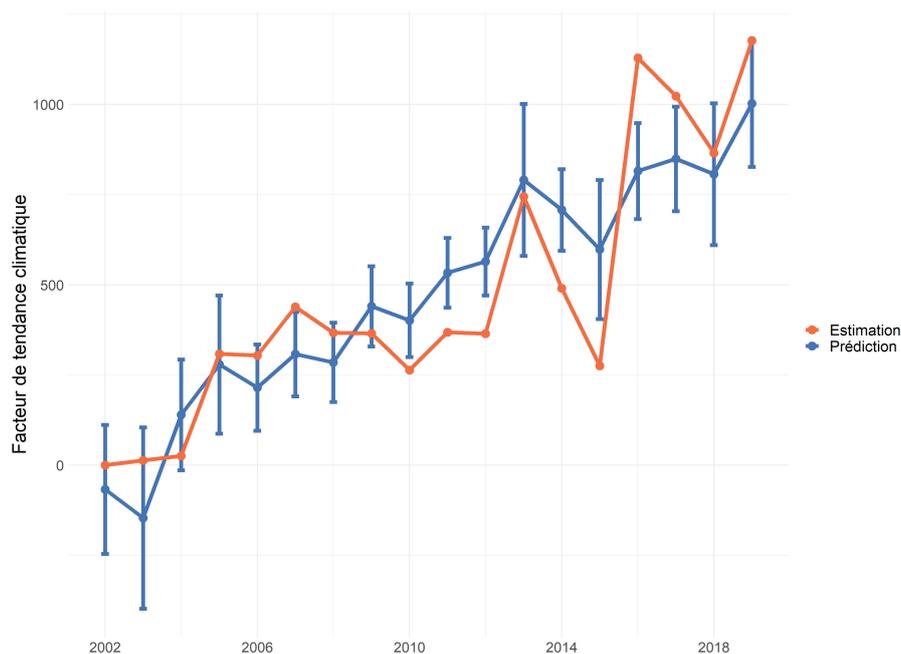


FIGURE 10.4 – Prédications du modèle linéaire

Certains points estimés par la méthode de séparation climatique ne rentrent pas dans les intervalles de confiance des coefficients prédits par la régression linéaire. De plus, cette dernière peine à capter les points issus de sauts dans la courbe de ratios sinistres sur primes, tels que 2015 et 2016. Toutefois, dans une optique de projection, nous ne cherchons pas à déterminer précisément la valeur de la tendance climatique mais un ordre de grandeur sur lequel les assureurs pourraient se fier pour affiner leurs estimations de réserves. Sous cet angle, les résultats de la régression linéaire sont finalement assez fiables.

10.3.3 Projection de la tendance climatique

Une fois nos données météorologiques futures traitées, nous pouvons désormais projeter la tendance climatique grâce au modèle linéaire retenu. La Figure 10.5 récapitule les résultats obtenus avec les différents scénarios de projection, que ce soit avec la valeur médiane ou les bornes de l'intervalle de confiance à 95% de l'indice d'humidité des sols.

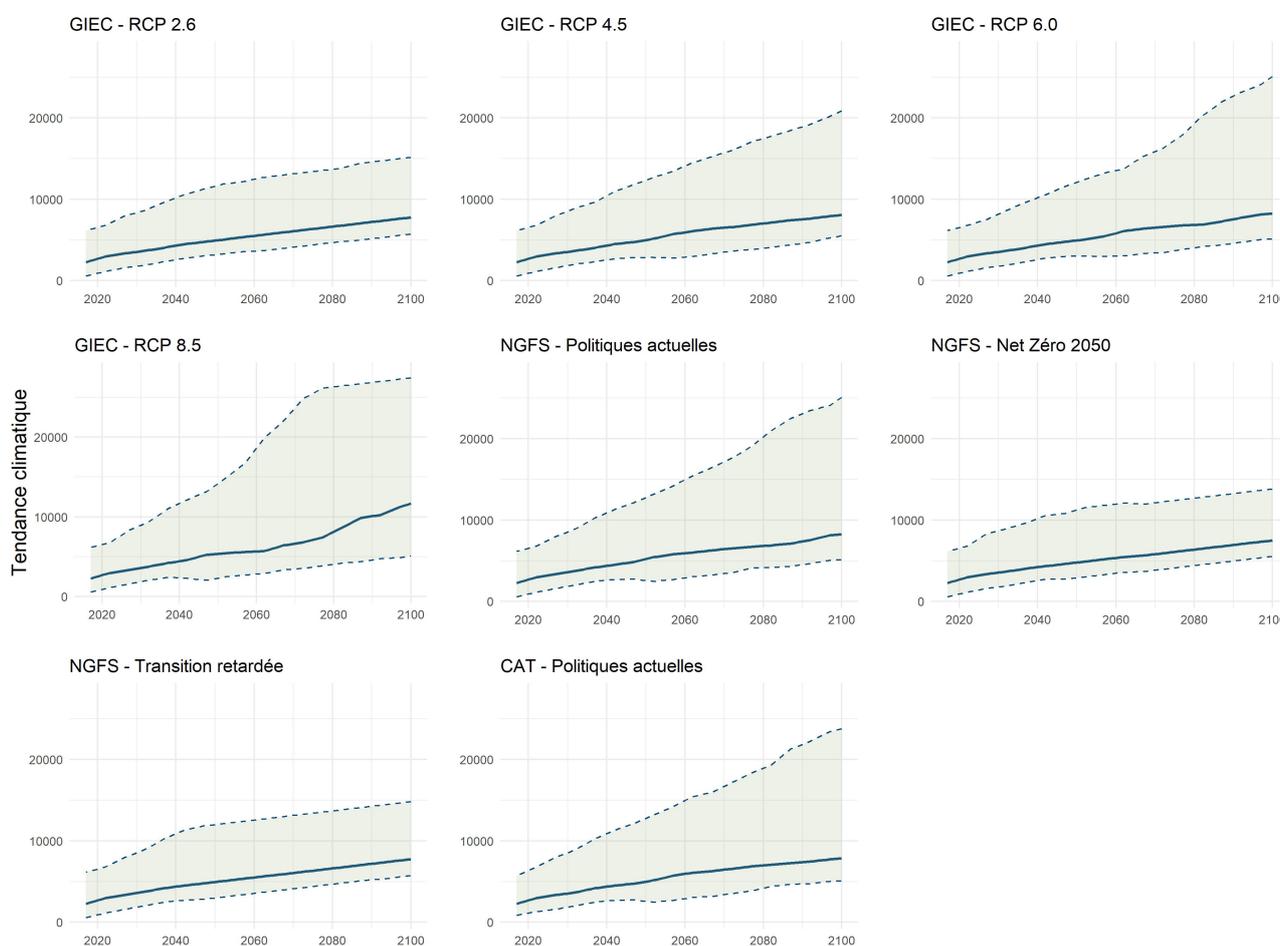


FIGURE 10.5 – Evolution de la tendance climatique prédite selon divers scénarios

Pour tous les scénarios, le facteur d'inflation « climatique » est croissant, signe que l'indicateur d'humidité des sols est trop faible pour compenser la tendance de croissance due à l'année considérée. Pour rappel, la tendance climatique prédite s'élevait à environ 850 en 2017. Ici, quel que soit le scénario considéré, elle grimpe à près de 2300. Cet effet de seuil est intégralement dû aux écarts entre les valeurs du SWI historiques et celles modélisées dans des projections climatiques. En effet, la tendance climatique est particulièrement sensible aux variations de cet indicateur : rappelons que toutes choses égales par ailleurs, une diminution de 1 du SWI

conduit à lui ajouter presque 840.

A titre d'illustration, la Table 10.5 suivante donne les estimations du facteur de tendance climatique pour quelques années.

Année	GIEC	GIEC	GIEC	GIEC	NGFS	NGFS	NGFS	CAT
	RCP 2.6	RCP 4.5	RCP 6.0	RCP 8.5	Pol. actuelles	Net Zéro 2050	Transition retardée	Pol. actuelles
2025	3201,044	3201,044	3201,044	3201,044	3213,056	3213,056	3213,056	3201,044
2050	4928,357	5002,517	4938,623	5365,502	5217,100	4788,984	4950,631	5002,517
2075	6364,341	6749,623	6709,817	7164,329	6634,370	6097,280	6342,067	6751,683

TABLE 10.5 – Prévisions de la tendance climatiques pour quelques années

Si tous les scénarios offrent des projections très similaires pour les années à venir, les écarts sont plus marqués pour les années 2050 et 2075. La tendance d'inflation des réserves pourrait même plus que doubler d'ici à 2075, signe que les événements climatiques affecteront de plus en plus les agriculteurs et donc de fait les assureurs qui protègent leurs rendements.

10.4 Conclusions sur l'effet du climat sur les réserves

Ce chapitre clôt l'analyse du risque climatique dans le cadre du provisionnement. Au cours de la partie précédente, nous avons recensé des méthodes nous permettant d'évaluer, de manière qualitative et quantitative, l'importance du risque climatique. Ces approches ont été appliquées sur des triangles agricoles dans cette dernière partie.

Les résultats présentés au chapitre précédent nous démontrent que le triangle agricole relatif aux aléas climatiques est caractérisé par des paramètres de distribution d'ordre 3 et 4 bien différents des triangles des autres sous-branches. Rappelons toutefois que cette analyse n'a pu être effectuée que sur une branche et pour un seul assureur. Par conséquent ses résultats devraient être confirmés par d'autres travaux mais elle fournit d'ores et déjà une méthodologie qualitative pour observer le risque climatique dans la distribution des réserves.

Lors de ce dernier chapitre, nous avons réussi à lier, à travers un modèle linéaire, un indicateur climatique d'humidité des sols à une tendance d'inflation par année de survenance, extraite grâce à une méthode de séparation. Suite à cela, des prévisions ont été effectuées via différents scénarios climatiques, en vue d'observer l'évolution temporelle du facteur de tendance. Néanmoins, pour arriver à de tels résultats, des hypothèses simplificatrices ont été nécessaires afin de pallier à l'absence de données. Dès lors, puisque ce mémoire se place en tant que « boîte à outils », il serait pertinent de relancer cette démarche avec des données plus complètes. Cela permettrait ainsi d'une part d'assurer la stabilité de la méthode, mais également de complexifier le modèle de lien entre variables météorologiques et facteur d'inflation « climatique » en vue d'acquérir une plus grande précision dans les estimations et les projections.

Quatrième partie

Complément : Modélisation d'événements extrêmes

Chapitre 11

Eléments de théorie des valeurs extrêmes

Ce chapitre vise à présenter les résultats fondamentaux de la théorie des valeurs extrêmes, que ce soit sur la loi limite du maximum ou sur la loi des dépassements de seuil. Il reprend pour cela le cours d'*Extreme Value Theory* de Robert donné à l'ENSAE Paris [51]. Ce chapitre donne enfin des méthodes de calibrage des lois d'événements extrêmes.

Objectifs de la théorie des valeurs extrêmes

Les valeurs extrêmes sont définies comme des événements rares qui peuvent conduire à de fortes pertes financières. Pour les assureurs, il peut s'agir de catastrophes naturelles ou de krack boursier. Dès lors, il est nécessaire de modéliser correctement ces événements extrêmes en vue d'anticiper leurs conséquences. La théorie des valeurs extrêmes donne alors un cadre permettant de calibrer :

- la distribution du maximum d'une distribution ou de ses plus grandes statistiques d'ordre,
- la distribution des dépassements de seuil,
- ou encore la distribution de la survenance des dépassements de seuil.

En conséquence, la théorie des valeurs extrêmes permet de répondre à plusieurs questions :

- Quelle est la période de retour d'un événement de sévérité x ?
- Quel est le niveau de retour associé à un temps d'attente de T années ?
- Quel est le montant moyen des sinistres dépassant un certain seuil u ?

Les deux sections suivantes fournissent des éléments théoriques permettant de répondre à ces questions, à travers des estimations de la loi limite du maximum et de la loi des dépassements de seuil. La dernière section traitera des méthodes à appliquer pour calibrer de telles lois.

Notation 11.0.1. Dans ce chapitre, (X_1, \dots, X_n) est une suite de variables indépendantes et identiquement distribuées, de fonction de répartition F et de maximum $M_n = \max(X_1, \dots, X_n)$.

Définition 11.0.2. La fonction de survie des (X_i) est définie par :

$$\bar{F} = 1 - F.$$

11.1 Loi limite du maximum

Les hypothèses précédentes conduisent à ce que

$$\mathbb{P}(M_n \leq x) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \leq x) = F(x)^n.$$

Néanmoins, si la distribution des X_i est inconnue, cette formule n'est pas très utile. De plus, il serait pertinent d'avoir des approximations asymptotiques quand la taille de l'échantillon est très grande.

Définition 11.1.1. Le *point extrême* de F , noté x^F est défini par :

$$x^F = \sup\{x, F(x) < 1\}.$$

Remarque 11.1.2. Si $F(x) < 1$, alors $\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq x) = 0$. Ainsi, $M_n \rightarrow x^F$ en probabilité lorsque $n \rightarrow \infty$.

En conséquence, la distribution de M_n est dégénérée. Il faudra donc passer par une normalisation, comme ce qui est fait pour la somme via le théorème central limite, pour obtenir une loi limite du maximum.

11.1.1 Existence de la loi limite

Avant de caractériser la loi limite du maximum, il faut tout d'abord s'assurer de son existence. Pour cela, il faut trouver une suite (u_n) telle que la limite de $\mathbb{P}(M_n \leq u_n)$ existe et soit non-nulle. La proposition et le théorème suivants donnent des conditions sur F assurant une telle convergence.

Proposition 11.1.3. Soient $\tau > 0$, (u_n) une suite de réels. Les conditions suivantes sont équivalentes :

1. $\lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq u_n) = e^{-\tau}$,
2. $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$.

Théorème 11.1.4. Soient $\tau > 0$ et F fonction de répartition de point extrême $x^F \leq \infty$. Il existe une suite (u_n) telle que $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$ si et seulement si

$$\lim_{x \nearrow x^F} \frac{\bar{F}(x)}{\bar{F}(x^-)} = 1.$$

Le choix de la suite (u_n) peut également avoir des conséquences sur le nombre de dépassement de ce seuil u_n , comme en témoigne la proposition suivante.

Proposition 11.1.5. Supposons qu'il existe une suite (u_n) de nombres réels telle que

$$\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau.$$

Soit $N_n = \sum_{i=1}^n \mathbb{1}_{\{X_i > u_n\}}$ le nombre de dépassements au-delà du seuil u_n . Alors $N_n \xrightarrow[n \rightarrow \infty]{d} N$, où N est une variable aléatoire de distribution de Poisson de paramètre τ .

11.1.2 Caractérisation de la loi limite du maximum

Le théorème suivant, dit de Fisher-Typett, fournit une caractérisation des distributions limites possibles pour la loi du maximum.

Théorème 11.1.6. Supposons qu'il existe deux suites (a_n) et (b_n) de nombres réels, avec $c_n > 0$ pour tout n , telles que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x)$$

où G est une distribution non-dégénérée. Alors G est de la forme de l'une des trois distributions suivantes :

- **Fréchet** : Pour $\alpha > 0$, $\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases}$
- **Gumbel** : $\Lambda(x) = \exp(-e^{-x})$ pour $x \in \mathbb{R}$
- **Weibull** : Pour $\alpha > 0$, $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$

Les densités de ces trois distributions sont représentées sur la Figure 11.1.

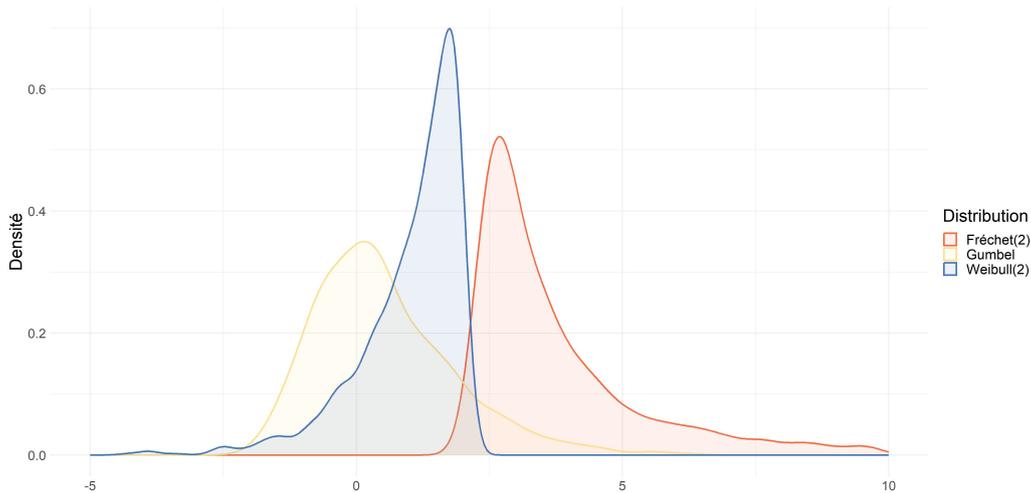


FIGURE 11.1 – Densité des distributions de Fréchet, Weibull et Gumbel

Il apparaît alors que les trois lois n'ont pas la même « dangerosité ». En effet, la distribution de Fréchet possède une queue de distribution droite épaisse et peut ainsi prendre des valeurs très élevées avec une probabilité non-négligeable. Dans le cadre de l'assurance, le calibrage d'une telle loi peut conduire à devoir provisionner des montants conséquents pour anticiper la survenance d'événements extrêmes d'une telle intensité. A l'inverse, la loi de Weibull est caractérisée par une queue droite de distribution très fine, ce qui fait d'elle une distribution non-dangereuse. La loi de Gumbel est quant à elle dans l'entre deux, possédant une queue de distribution droite moins épaisse qu'une loi de Fréchet.

Outre l'interprétation graphique, le caractère « dangereux » de ces lois limites se justifie par leur propriété de **max-stabilité**, énoncée ci-après.

Proposition 11.1.7.

- Soient X, X_1, \dots, X_n des variables indépendantes et identiquement distribuées de loi de Fréchet Φ_α . Alors

$$M_n \xrightarrow[n \rightarrow \infty]{d} n^{1/\alpha} X.$$

- Soient X, X_1, \dots, X_n des variables indépendantes et identiquement distribuées de loi de Gumbel Λ . Alors

$$M_n \xrightarrow[n \rightarrow \infty]{d} X + \ln n.$$

- Soient X, X_1, \dots, X_n des variables indépendantes et identiquement distribuées de loi de Weibull Ψ_α . Alors

$$M_n \xrightarrow[n \rightarrow \infty]{d} n^{-1/\alpha} X.$$

Dès lors, la propriété précédente implique que, à partir d'un certain nombre d'observations n , le maximum d'un échantillon suivant une loi de Fréchet de paramètre α suit la même loi, normalisée par un facteur $n^{1/\alpha}$ qui tend vers l'infini. Cela justifie donc le caractère risqué de ce maximum.

11.1.3 Généralisation par la loi GEV

Il est parfois préférable d'uniformiser les trois distributions précédentes à l'aide d'une seule distribution. Pour cela, la loi d'extremum généralisée (GEV) est utilisée.

Définition 11.1.8. Soit X une variable aléatoire. X suit une **loi d'extremum généralisé** de paramètres $\mu \in \mathbb{R}$, $\sigma > 0$ et $\xi \in \mathbb{R}$, i.e $X \sim GEV(\mu, \sigma, \xi)$, si sa fonction de répartition est telle que :

$$G(x) = \begin{cases} \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)_+^{-\frac{1}{\xi}} \right] & \text{pour } \xi \neq 0 \\ \exp \left[- \exp \left(- \frac{x - \mu}{\sigma} \right) \right] & \text{pour } \xi = 0 \end{cases}$$

où $x_+ = \max(x, 0)$, μ est le paramètre de position, σ le paramètre d'échelle et ξ le paramètre de forme.

Remarque 11.1.9. Le cas $\xi = 0$ correspond exactement à la limite de $G(x)$ lorsque $\xi \rightarrow 0$.

Notation 11.1.10. Pour $\mu = 0$ et $\sigma = 1$, il est noté : $G_\xi(x) = \exp \left[- (1 + \xi x)_+^{-\frac{1}{\xi}} \right]$.

Remarque 11.1.11.

- Si $\xi > 0$, la distribution associée est celle de Fréchet : $GEV(1, 1/\alpha, 1/\alpha) = \text{Fréchet}(\alpha)$.
- Si $\xi = 0$, la distribution associée est celle de Gumbel : $GEV(0, 1, 0) = \text{Gumbel}$.
- Si $\xi < 0$, la distribution associée est celle de Weibull : $GEV(-1, 1/\alpha, -1/\alpha) = \text{Weibull}(\alpha)$.

Tout comme les distributions de Fréchet, Gumbel et Weibull, les lois GEV sont également max-stables, via la relation suivante.

Proposition 11.1.12. Soient X, X_1, \dots, X_n des variables indépendantes et identiquement distribuées de loi $GEV(0, 1, \xi)$. Alors

$$M_n \xrightarrow[n \rightarrow \infty]{d} n^\xi X + \frac{n^\xi - 1}{\xi}.$$

Il est alors possible de reformuler le théorème de Fisher-Typett à partir de cette nouvelle distribution.

Théorème 11.1.13. Supposons qu'il existe deux suites (a_n) et (b_n) de nombres réels, avec $c_n > 0$ pour tout n , telles que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = G(x)$$

où G est une distribution non-dégénérée. Alors G est de la forme suivante :

$$G_\xi(x) = \exp \left[- (1 + \xi x)_+^{-\frac{1}{\xi}} \right].$$

11.1.4 Notion de domaine d'attraction

Définition 11.1.14. Soit G une distribution GEV. F appartient au **domaine d'attraction** de G s'il existe deux suites (a_n) et (b_n) de nombres réels, avec $c_n > 0$ pour tout n , telles que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = G(x).$$

La Proposition suivante fournit une caractérisation d'appartenance au domaine d'attraction de la loi $GEV(0, 1, \xi)$.

Proposition 11.1.15. Soit x un réel appartenant au support de la loi de X . Soit h , l'inverse du taux de risque, défini comme $h(x) = \frac{\bar{F}(x)}{f(x)}$. Supposons f dérivable et $\lim_{x \rightarrow x^F} h'(x) = \xi$. Alors F appartient au domaine d'attraction de la loi $GEV(0, 1, \xi)$ et les suites (a_n) et (b_n) peuvent être choisies de la manière suivante

$$1 - F(b_n) = 1/n \text{ et } a_n = h(b_n).$$

L'appartenance à un domaine d'attraction est également une question d'équivalence en queue, dont la notion est définie ci-après.

Définition 11.1.16. Soient deux distributions F_1 et F_2 . F_1 et F_2 sont **équivalentes en queues** si elles possèdent le même point extrême $x^{F_1} = x^{F_2}$ et que

$$\lim_{x \rightarrow x^{F_1}} \frac{\bar{F}_1(x)}{\bar{F}_2(x)} = c \in \mathbb{R}_+ \cup \{\infty\}.$$

Proposition 11.1.17. Deux distributions équivalentes en queues appartiennent au même domaine d'attraction.

La notion de domaine d'attraction est très large. Les trois distributions limites citées dans le Théorème de Fisher-Typett ont leur propre caractérisation de domaine d'attraction. Ces dernières ne sont pas présentées dans le cadre de ce mémoire, mais sont disponibles dans le cours d'*Extreme Value Theory* de Robert donné à l'ENSAE Paris [51].

11.1.5 Loi limite des statistiques d'ordre

Soient $M_n = X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)} = \min(X_1, \dots, X_n)$. Le théorème suivant caractérise la distribution des statistiques d'ordre de l'échantillon (X_1, \dots, X_n) .

Théorème 11.1.18. Soit une suite (u_n) de nombres réels telle que $\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau$. Alors

$$\lim_{n \rightarrow \infty} \mathbb{P} (X_{(k)} \leq u_n) = \lim_{n \rightarrow \infty} \mathbb{P} (N_n < k) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}.$$

Si u_n est de la forme $u_n = a_n x + b_n$ avec $\tau = -\ln G(x)$ et G distribution GEV, alors

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_{(k)} - b_n}{a_n} \leq x \right) = G(x) \sum_{i=0}^{k-1} \frac{(-\ln G(x))^i}{i!}.$$

11.2 Loi des dépassements de seuil

Plutôt que de considérer le maximum M_n d'un échantillon, il peut être utile de s'intéresser aux excédents de seuil u_n , à savoir les observations $Y_i := (X_i - u_n)_+$, schématisées sur la Figure 11.2. Par définition, ces observations sont positives. Elles sont caractérisées par une loi de Pareto généralisée (GPD).

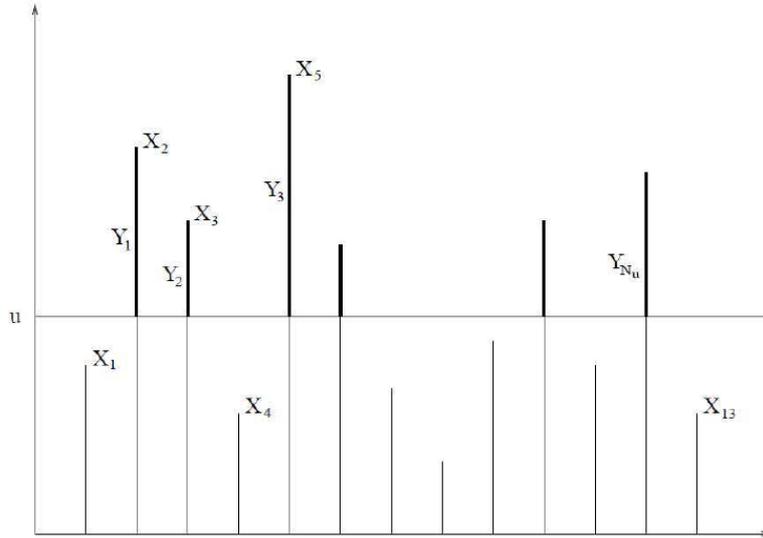


FIGURE 11.2 – Echantillon de dépassements de seuil [51]

11.2.1 Loi de Pareto généralisée

Définition 11.2.1. Soit X une variable aléatoire. X suit une **loi de Pareto généralisée** de paramètres $\beta > 0$ et $\xi \in \mathbb{R}$, i.e $X \sim GPD(\beta, \xi)$, si sa fonction de répartition est telle que :

$$G_{\xi, \beta}^p(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)_+^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{si } \xi = 0 \end{cases} \quad \text{pour } \begin{cases} x \geq 0 & \text{si } \xi \leq 0 \\ 0 \leq x \leq -\beta/\xi & \text{si } \xi < 0 \end{cases}$$

où β est le paramètre d'échelle et ξ le paramètre de forme.

Les distributions de Pareto généralisée possèdent quelques propriétés, énoncées dans la proposition suivante.

Proposition 11.2.2.

- Si $\xi = -1$, $G_{-1, \beta}^p$ est la loi uniforme sur $[0, \beta]$.
- Si $\xi = 0$, $G_{0, \beta}^p$ est la distribution exponentielle de paramètre $1/\beta$.
- Si $U \sim \mathcal{U}([0, 1])$, alors

$$\beta \left(\frac{U^{-\xi} - 1}{\xi} \right) \sim GPD(\beta, \xi).$$

- Si $X \sim GPD(\beta, \xi)$ avec $\xi < 1$, alors

$$\mathbb{E}[X] = \frac{\beta}{1 - \xi}.$$

- Si $X \sim GPD(\beta, \xi)$, alors $X - u | X > u \sim GPD(\beta + \xi u, \xi)$. En particulier, si $\xi < 1$, alors

$$\mathbb{E}[X - u | X > u] = \frac{\beta + \xi u}{1 - \xi}.$$

11.2.2 Lien avec la loi limite du maximum

Le Théorème 11.2.6 établit un lien entre la distribution des dépassements de seuil et la loi limite du maximum. Avant cela, quelques notations et définition doivent être introduites.

Notation 11.2.3. Soient x, u dans le domaine de définition de F . Alors

$$F_u(x) := \mathbb{P}(X - u \leq x | X > u) = \frac{F(u+x) - F(u)}{1 - F(u)}.$$

Définition 11.2.4. L'inverse généralisé de la fonction de répartition F , noté F^{\leftarrow} est défini par

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\} \quad \forall t \in [0, 1].$$

Notation 11.2.5. Soit $t > 0$. Alors

$$U(t) = F^{\leftarrow}(1 - t^{-1}).$$

Théorème 11.2.6. Soit $\xi \in \mathbb{R}$. Les propositions suivantes sont équivalentes :

1. Il existe deux suites (a_n) et (b_n) , $a_n > 0$, telles que $\lim_{n \rightarrow \infty} F(a_n x + b_n)^n = G_\xi(x)$.
2. Il existe une fonction $a(\cdot)$ telle que

$$\lim_{u \rightarrow x^F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)_+^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0 \end{cases}$$

3. Pour $x, y > 0$ et $y \neq 1$

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} (x^\xi - 1) / (y^\xi - 1) & \text{si } \xi \neq 0 \\ \ln x / \ln y & \text{si } \xi = 0 \end{cases}$$

4. Il existe une fonction $b(\cdot)$ positive telle que

$$\lim_{u \rightarrow x^F} \sup_{0 < x < x^F - u} |F_u(x) - G_{\xi, \beta}^p(x)| = 0.$$

11.3 Calibrage de lois d'événements extrêmes

11.3.1 Outils d'analyse de distribution

Avant de calibrer des lois de distribution, quelques outils permettent d'obtenir davantage d'informations sur la loi des observations recherchée. Nous présentons dans cette section deux d'entre eux :

- le graphique quantile-quantile
- et la *mean excess function*.

11.3.1.1 Graphique quantile-quantile

Définition 11.3.1. *Le graphique quantile-quantile, ou QQ-plot, est le graphe suivant*

$$\{X_{(i)}, F^{-1}(1 - i/n)\}.$$

Les graphiques quantile-quantile permettent de comparer les quantiles de la loi empirique avec ceux d'une loi de référence. Ils ont deux utilités :

- **Etude des queues de distribution** : en fonction du comportement de la courbe, il est possible de déterminer si la distribution empirique a des queues plus fines ou plus épaisses que la loi théorique comparative. Ainsi, une convexité de la courbe indique la présence de queues plus épaisses que la loi de référence, tandis que sa concavité est signe de présence de queues plus fines. A partir de cette analyse, le choix de la distribution théorique sur laquelle effectuer le calibrage peut donc se préciser.
- **Validation d'une distribution théorique** : si les données sont en adéquation avec la loi de distribution théorique choisie, alors le QQ-plot doit être approximativement linéaire. En particulier, après calibrage des paramètres, le QQ-plot doit suivre la première bissectrice.

La Table 11.1 présente quelques exemples de QQ-plots selon la distribution considérée. Φ représente la fonction de répartition de la loi normale centrée-réduite.

Distribution	QQ-plot associé
Normale	$\{X_{(i)}, \Phi^{-1}(1 - i/n)\}$
Log-Normale	$\{\ln X_{(i)}, \Phi^{-1}(1 - i/n)\}$
Exponentielle	$\{X_{(i)}, -\ln(i/n)\}$
Pareto	$\{\ln X_{(i)}, -\ln(i/n)\}$

TABLE 11.1 – Exemples de QQ-plots [51]

11.3.1.2 Mean excess function

Définition 11.3.2. *Soient X une variable aléatoire et u un réel. La mean excess function associée à X est définie comme*

$$e(u) = \mathbb{E}[X - u | X > u].$$

En assurance, cette fonction peut s'interpréter comme le coût moyen des sinistres dépassant une franchise u , ou bien l'espérance de vie résiduelle d'un individu ayant passé l'âge u .

En pratique, on utilise la version empirique de la mean excess function, définie ci-dessous.

Définition 11.3.3. *Soient X une variable aléatoire et u un réel. Notons*

$$\Delta_n(u) = \{i, i \in \llbracket 1, n \rrbracket, X_i > u\}.$$

La mean excess function empirique se définit par

$$e_n(u) = \frac{\sum_{i \in \Delta_n(u)} (X_i - u)}{|\Delta_n(u)|} \text{ avec } |\Delta_n(u)| \text{ nombre d'observations de l'ensemble } \Delta_n(u).$$

Cette fonction est représentée dans le graphe $(X_{(i)}, e_n(X_{(i)}), i \in \llbracket 1, n \rrbracket)$, appelé mean excess plot.

A partir de ce graphique, et notamment de la monotonie de la courbe résultante, il est possible d'identifier la nature de la distribution sous-jacente aux observations considérées. La Figure 11.3 associe certains comportements de la *mean excess function* à des lois de distribution.

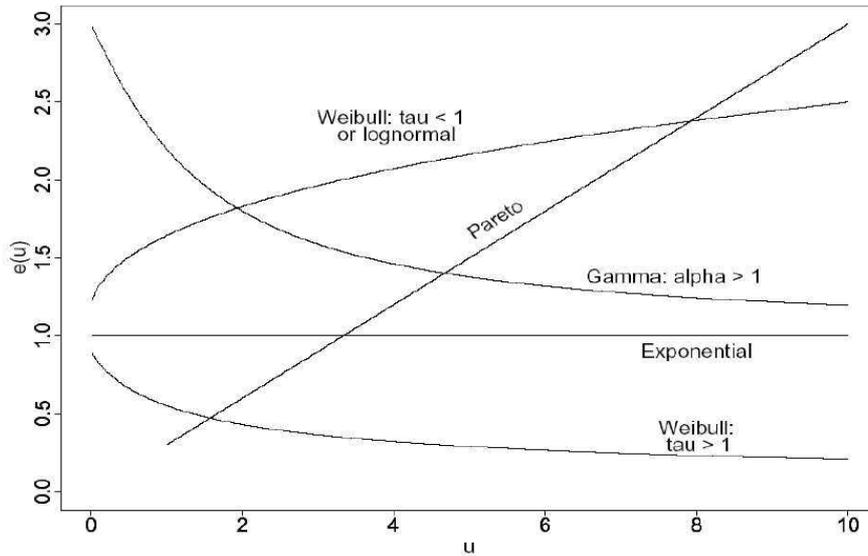


FIGURE 11.3 – Lien entre *mean excess function* et distributions de probabilité [51]

11.3.2 Calibrage de la loi limite du maximum

Pour rappel, les suites a_n et b_n dépendent de la distribution F sous-jacente. Par conséquent, à n fixé, en posant $y = a_n x + b_n$, $\sigma = a_n > 0$ et $\mu = b_n$, il vient que

$$\mathbb{P}(M_n \leq y) \approx \exp \left(- \left(1 + \xi \frac{y - \mu}{\sigma} \right)_+^{-1/\xi} \right).$$

La validité de la distribution limite, pour n fixé, repose sur deux conditions. Tout d'abord, le **choix de n** , dans la mesure où les blocs d'observations utilisés pour calculer les maxima doivent être de taille suffisamment large. Par ailleurs, il est important que h' **tende vers une constante**, pour assurer la convergence vers la loi limite.

Pour calibrer les paramètres de la loi GEV, l'approche la plus communément utilisée est l'estimation par maximum de vraisemblance. L'estimateur doit être calculé à partir de méthodes de maximisation numériques, aucune formule fermée ne pouvant être déterminée analytiquement. Par ailleurs, certaines contraintes peuvent être imposées, dans la mesure où l'ensemble de définition de la loi peut dépendre des paramètres (μ, σ, ξ) . De plus, selon la valeur du paramètre de forme, l'estimateur du maximum de vraisemblance $\hat{\theta}$ n'a pas les mêmes propriétés. En particulier, la normalité asymptotique peut être non-vérifiée.

D'autres méthodes existent pour déterminer le domaine d'attraction auquel appartient la loi des observations, via des estimations du paramètre de forme ξ :

— **Estimateur de Pickands :**

$$\xi_{k,n}^{\text{Pickands}} = \left[\ln \frac{X_{(k)} - X_{(2k)}}{X_{(2k)} - X_{(4k)}} \right] / \ln 2$$

— **Estimateur de Dekkers Einmahl et de Haan :**

$$\xi_{k,n}^{\text{DEdH}} = \xi_{k,n}^{H(1)} + 1 - \frac{1}{2} \left[1 - \frac{\left(\xi_{(k)}^{H(1)} \right)^2}{\xi_{(k)}^{H(2)}} \right]^{-1}$$

avec pour $r \in \{1, 2\}$, $\xi_{k,n}^{H(r)} = \frac{1}{k} \sum_{i=1}^{k-1} [\ln X_{(i)} - \ln X_{(k)}]^r$.

— **Estimateur de Hill :**

$$\xi_{k,n}^{\text{Hill}} = \frac{1}{k} \sum_{i=1}^k \ln X_{(i)} - \ln X_{(k+1)}$$

Ces estimateurs sont faiblement consistants. Sous réserve d'hypothèses plus forte sur k et F , ils sont également asymptotiquement gaussiens :

$$k^{1/2} \left(\xi_{k,n}^{\text{Pickands}} - \xi \right) \xrightarrow[k, n \rightarrow \infty]{d} \mathcal{N} \left(0, \xi^2 \frac{2^{2\xi+1} + 1}{(2(2^\xi - 1) \ln 2)^2} \right)$$

$$k^{1/2} \left(\xi_{k,n}^{\text{DEdH}} - \xi \right) \xrightarrow[k, n \rightarrow \infty]{d} \mathcal{N} \left(0, 1 + \xi^2 \right), \xi \geq 0$$

$$k^{1/2} \left(\xi_{k,n}^{\text{H}} - \xi \right) \xrightarrow[k, n \rightarrow \infty]{d} \mathcal{N} \left(0, \xi^2 \right).$$

11.3.3 Calibrage de la loi des dépassements de seuil

La méthode la plus classique d'estimation des paramètres des lois de dépassements de seuil repose sur la mise en œuvre d'un maximum de vraisemblance. Cette approche est couramment utilisée lors de calibrages de lois de distribution, qu'elles soient extrêmes ou non. Pour accroître la précision des estimations, Chavez-Demoulin, Embrechts et Hofert (2016) [10] étendent la méthode de maximum de vraisemblance en autorisant les paramètres de la GPD à dépendre de covariables. Leur approche sera détaillée dans la Section 11.3.3.2.

Or, les méthodes de calibrage de lois extrêmes sont conditionnelles au choix du bon seuil. Quelques outils existent afin de déterminer un seuil suffisamment grand pour quantifier des événements extrêmes, sans pour autant être trop large au risque de manquer de données pour estimer correctement la loi. Ces derniers vont faire l'objet de la Section 11.3.3.1 suivante.

11.3.3.1 Choix du seuil

La propriété de stabilité par seuil de la loi de Pareto généralisée est essentielle pour déterminer le bon seuil. Pour rappel, soit Y_u une variable aléatoire caractérisant les dépassements du seuil u telle que $Y_u | Y_u > 0 \sim GPD(\sigma_u, \xi)$. Si Y_v caractérise les dépassements au-delà du seuil $v \geq u$, alors il vient que

$$Y_v | Y_v > 0 \sim GPD(\sigma_v, \xi), \text{ où } \sigma_v = \sigma_u + \xi(v - u).$$

A l'aide de cette propriété, deux outils peuvent servir à choisir un seuil raisonnable pour calibrer la distribution GPD :

- la *mean excess function*
- et le **graphique de stabilité des paramètres**.

Mean excess function

Soit $X = Y_u + u$, avec $Y_u | Y_u > 0 \sim GPD(\sigma_u, \xi)$, avec $\xi < 1$. Pour rappel, la *mean excess function* de X au-delà du seuil $v > u$ s'écrit

$$\mathbb{E}[X - v | X > v] = \frac{\sigma_v}{1 - \xi} = \frac{\sigma_u + \xi(v - u)}{1 - \xi}.$$

Cette fonction est linéaire en v , de pente $\frac{\xi}{1-\xi}$. Ainsi, il suffit de représenter sa version empirique avec le graphe $(X_{(i)}, e_n(X_{(i)}), i \in \llbracket 1, n \rrbracket)$ pour différents seuils et d'observer à partir duquel la *mean excess function* devient approximativement linéaire. Cela signifiera la présence d'une loi de Pareto généralisée sur les données excédant ce seuil.

Au-delà de la question du choix du seuil, la monotonie de la *mean excess function* permet également de déterminer le signe du paramètre de forme ξ . Mathématiquement :

- Si $\xi > 0$, alors $\mathbb{E}[X - v | X > v]$ est croissante à partir du seuil u .
- Si $\xi = 0$, alors $\mathbb{E}[X - v | X > v]$ est constante à partir du seuil u .
- Si $\xi < 0$, alors $\mathbb{E}[X - v | X > v]$ est décroissante à partir du seuil u .

L'exemple de la Figure 11.4 illustre ce propos, en simulant 1000 observations de lois GPD de seuil nul. Si la monotonie des courbes des Figures 11.4a et 11.4c confirme le signe du paramètre de forme, dans le cas $\xi = 0$, la *mean excess function* n'est pas vraiment constante mais oscillante.

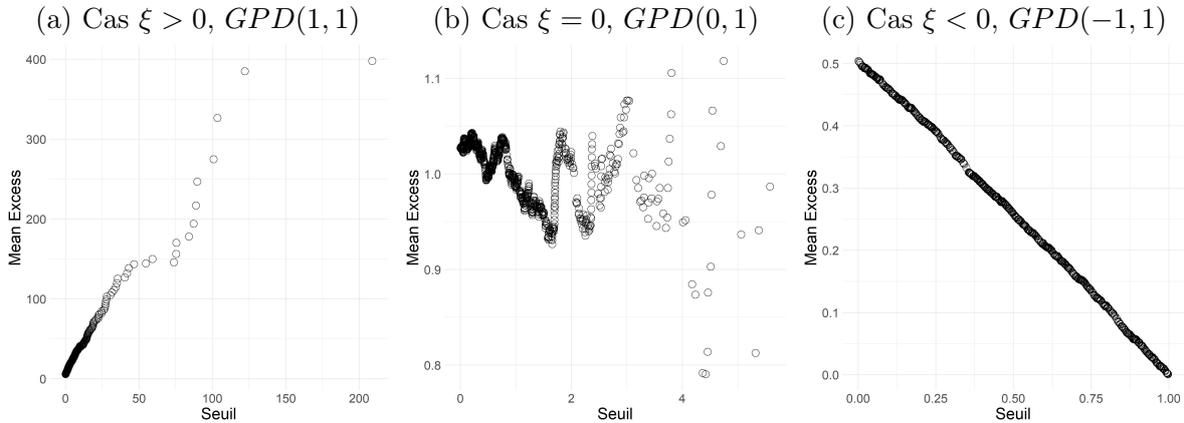


FIGURE 11.4 – *Mean excess function* de trois types de lois de Pareto généralisée

Graphique de stabilité des paramètres

Notons $\sigma^* = \sigma_v - \xi v$. Si la variable aléatoire considérée a une distribution de Pareto généralisée à partir du seuil $u \leq v$, alors σ^* ne dépend plus de v . Il faut alors estimer les paramètres σ^* et ξ en fonction de plusieurs seuils u et déterminer graphiquement à partir duquel les estimateurs semblent approximativement constants.

11.3.3.2 Estimation dynamique des paramètres de la GPD

Dans *An extreme value approach for modeling operational risk losses depending on covariates*, Chavez-Demoulin, Embrechts et Hofert [10] présente une méthode de calibrage de GPD dite dynamique permettant d'intégrer l'effet de covariables pour prédire les paramètres. Dans l'article, l'approche est illustrée à partir de données de pertes liées au risque opérationnel dans

plusieurs secteurs. Les auteurs considèrent deux covariables dans leur exemple : le secteur d'activité x et l'année de survenance de la perte t . Néanmoins, il est possible d'intégrer davantage de covariables à leurs modèles.

Spécification du modèle

Pour des questions de convergence, les auteurs reparamétrisent la GPD (ξ, β) en (ξ, ν) , où ν est tel que

$$\nu = \ln((1 + \xi)\beta) \text{ d'où } \beta = \frac{\exp(\nu)}{1 + \xi}.$$

Cette reparamétrisation est toutefois valable uniquement si $\xi > -1$. Pour inclure les covariables dans les modèles, les auteurs font les deux hypothèses suivantes :

$$\xi = \xi(x, t) = f_\xi(x) + h_\xi(t)$$

$$\nu = \nu(x, t) = f_\nu(x) + h_\nu(t)$$

où les fonctions f_ξ et f_ν indiquent l'appartenance au secteur x et sont de la forme

$$f(x) = \sum_i \alpha_i \times \mathbb{1}_{\{x=x_i\}}$$

et les fonctions h_ξ et h_ν peuvent être des fonctions linéaires, voire des modèles additifs généralisés de t .

Estimation des paramètres

L'estimation des paramètres s'effectue via un maximum de vraisemblance pénalisée. Soient les observations $\mathbf{y} = (y_{t_1}, \dots, y_{t_n})$ et $z_i = (t_i, x_i, y_{t_i})$. Soit ℓ la log-vraisemblance associée aux paramètres (ξ, β) . Soient $\gamma_\xi, \gamma_\nu \geq 0$ deux paramètres de lissage. La log-vraisemblance pénalisée ℓ^p s'écrit alors :

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu, \mathbf{z}) = \sum_{i=1}^n \ell \left(\xi_i, \frac{\exp(\nu_i)}{1 + \xi_i}, y_{t_i} \right) - \gamma_\xi \int_0^T h_\xi''(t)^2 dt - \gamma_\nu \int_0^T h_\nu''(t)^2 dt.$$

L'estimation s'effectue ensuite à partir d'un algorithme appliquant la méthode de Newton qui réévalue à chaque étape les modèles sur ξ et ν . Sur R, les auteurs ont développé la fonction `gamGPDfit` du package `QRM` qui permet d'obtenir les résultats des calibrages.

Remarque 11.3.4. *L'article de Chavez-Demoulin, Embrechts et Hofert [10] présente également une méthodologie pour calibrer la fréquence des événements extrêmes à partir des covariables précédentes. Ainsi, le nombre de dépassements λ est modélisé par un processus de Poisson et estimé par la régression logistique suivante :*

$$\ln \lambda(x, t) = f_\lambda(x) + h_\lambda(t).$$

Cette nouvelle hypothèse permet de caractériser totalement la loi des dépassements de seuil à partir de sa fréquence (suivant une loi de Poisson) et sa sévérité (suivant une loi GPD).

11.3.4 Estimation de périodes et niveaux de retour

Une fois la loi des événements extrêmes calibrée, il peut être intéressant d'évaluer des **périodes** et **niveaux de retour**. Dans le cadre assurantiel, ces derniers peuvent permettre d'évaluer la fréquence de survenance de certains événements, tels que des catastrophes naturelles, et donc d'adapter les provisions et ratios de solvabilité en conséquence.

Définition 11.3.5.

- La **période de retour** associée à un niveau x se définit comme le temps d'attente moyen avant un nouveau dépassement du seuil x .
- Le **niveau de retour** associé à une période de T années correspond au seuil atteint après un temps d'attente moyen de T années.

Dans le cas de données indépendantes et identiquement distribuées, ces deux indicateurs sont facilement calculables. Si la distribution \hat{F} a été calibrée sur les observations, période de retour et niveau de retour sont reliés par l'équation

$$1 - \hat{F}(x_p) = p.$$

En effet, le niveau de retour x_p aura une période de retour de $T = p^{-1}$ observations.

Chapitre 12

Données agricoles et données climatiques utilisées

Ce chapitre présente les données, agricoles et climatiques, utilisées pour le calibrage des pertes de rendements extrêmes.

12.1 Données agricoles

12.1.1 Présentation des données

Les données agricoles requises pour l'étude sont issues de la Statistique Agricole Annuelle (SAA) pour les cultures développées de l'organisme Agreste, le service statistique ministériel de l'agriculture. L'extraction des données s'effectue par le biais d'un tableau interactif¹, pour lequel plusieurs variables peuvent être intégrées. Trois indicateurs sont disponibles pour ces données :

- la **production**, exprimée en quintal,
- la **superficie développée**, en hectare,
- et le **rendement**, en quintal par hectare.

Nous retenons pour l'étude les indicateurs de production et de rendement. Ces derniers sont calculés annuellement sur la période 2000-2020, pour plusieurs mailles géographiques :

- France entière,
- Métropole ou Départements d'Outre-Mer (DOM),
- région
- et département.

Dans le cadre de l'étude, nous extrayons les données de toute la période de référence à la maille régionale. Ce choix de granularité spatiale nous permettra de faire le lien avec les données climatiques retenues. Du fait de la grande différence d'espèces cultivées en Outre-Mer, nous excluons les DOM de l'étude pour nous concentrer sur les régions métropolitaines.

Six niveaux de précision peuvent être pris en compte dans le choix des cultures. Les deux premiers niveaux correspondent à de grandes catégories, comme en témoigne la Figure 12.1 suivante. Si le troisième niveau est relatif au type de culture, les derniers détaillent précisément les différentes variétés.

1. https://agreste.agriculture.gouv.fr/agreste-saiku/?plugin=true&query=query/open/SAANR_DEVELOPPE_2#query/open/SAANR_DEVELOPPE_2

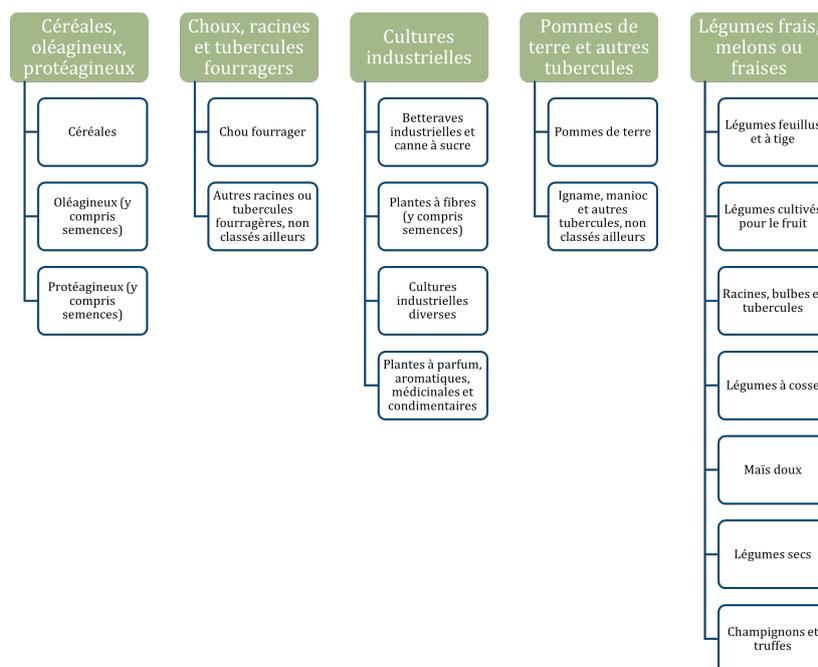


FIGURE 12.1 – Cultures développées de niveaux 1 et 2

Pour cette étude, nous nous concentrons sur des cultures de la catégorie « Céréales, oléagineux, protéagineux », listées dans la Table 12.1, dont la croissance s’effectue au cours du printemps, dans l’optique d’être en cohérence avec les données climatiques utilisées.

Avoine	Féveroles et fèves	Pois protéagineux
Blé	Maïs grain et maïs semence	Seigle et méteil
Colza grain et navette	Orge et escourgeon	Triticale

TABLE 12.1 – Cultures considérées pour l’étude

12.1.2 Analyse des données

La Figure 12.2 suivante effectue un état des lieux de la production agricole en 2020, pour les cultures citées en Table 12.1.

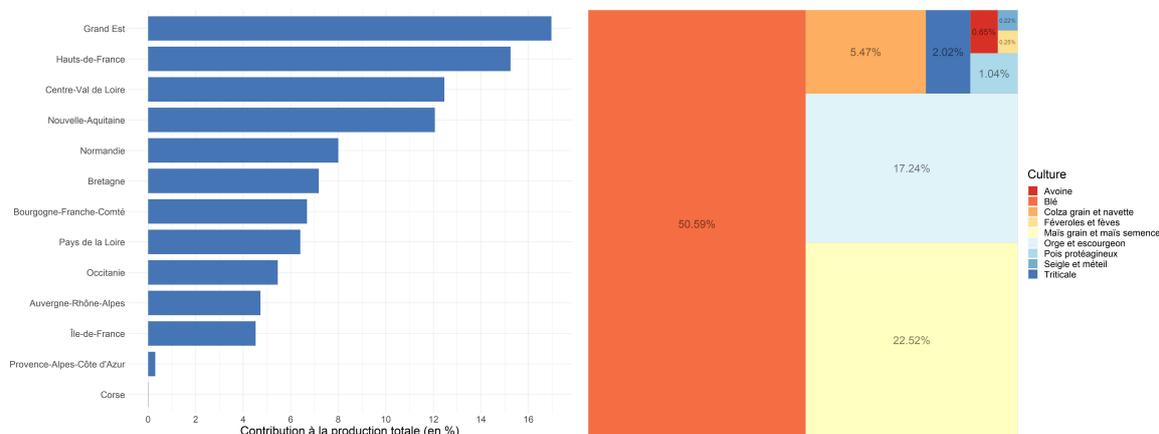


FIGURE 12.2 – Etat des lieux de la production agricole en 2020 pour les cultures considérées

Il ressort de la Figure 12.2a que les trois régions les plus productrices sont le Grand Est, les Hauts-de-France ainsi que le Centre-Val de Loire, suivie de près par la Nouvelle-Aquitaine. A elles seules, ces quatre régions contribuent à hauteur de 56% à la production totale métropolitaine pour les cultures considérées. A l'inverse, la Corse et la Provence-Alpes-Côte d'Azur sont les régions les moins productrices, contribuant à moins d'1% à la production totale.

En outre, selon la Figure 12.2b, le blé occupe une place majeure dans l'agriculture française, représentant plus de la moitié de la production totale. D'autres céréales sont également très cultivées, telles que le maïs et l'orge. Au total, ces deux espèces représentent près de 40% de la production totale métropolitaine. D'autres cultures sont peu exploitées en France, telles que l'avoine, les féveroles et fèves et les pois protéagineux : moins d'1% de la production totale leur est due.

En croisant les données régionales avec les données par type de culture, nous obtenons la Figure 12.3 suivante, sur laquelle sont présentés les niveaux de production en tonnes pour l'année 2020.

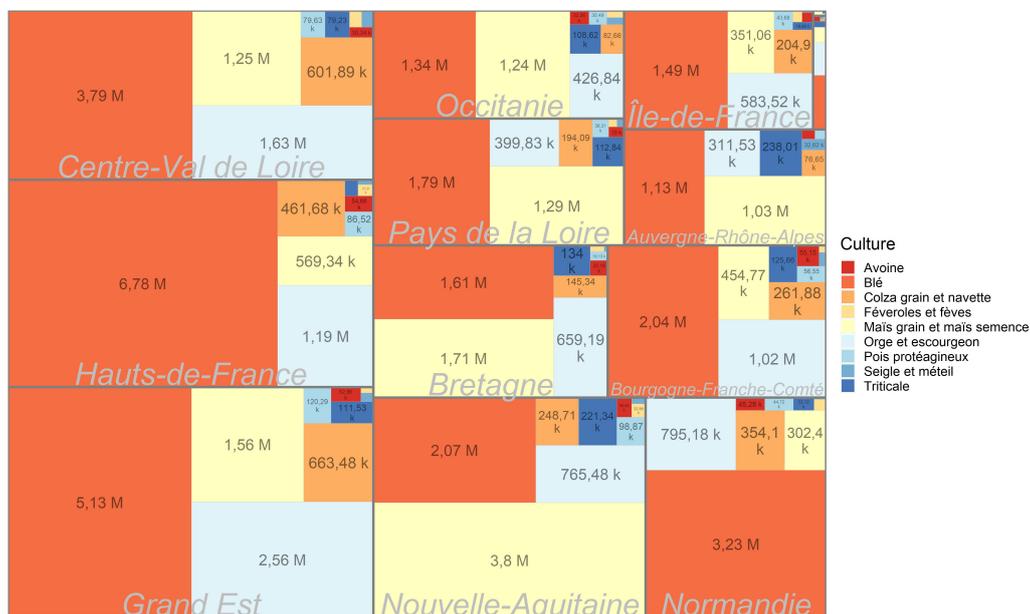


FIGURE 12.3 – Production agricole régionale en 2020 (en tonnes) par culture considérée

Si le Grand Est est la première région agricole métropolitaine d'après la Figure 12.2a, les Hauts-de-France produisent davantage de blé, 6,78 millions de tonnes en 2020 selon la Figure 12.3. Le Grand Est est toutefois le premier producteur d'orge et d'escourgeon, avec 2,56 millions de tonnes récoltées en 2020. Enfin, la Nouvelle-Aquitaine est caractérisée par la plus grande production de maïs grain et maïs semence, représentant près de 4 millions de tonnes en 2020.

La Figure 12.4 ci-après permet de visualiser l'évolution temporelle des rendements agricoles à la maille régionale.

Dans toutes les régions métropolitaines, à l'exception de la Corse, l'année 2020 est marquée par une baisse des rendements, par rapport à la moyenne des décennies 2000 et 2010. Si le Grand Est est la région la plus productrice, les rendements agricoles sont plus élevés dans les Hauts-de-France, quelle que soit la période considérée. Cela implique donc que les superficies exploitées sont plus grandes dans le Grand Est. Par ailleurs, les effets du changement climatique ne sont pas visibles sur ce graphique : les rendements de la décennie 2010 sont assez proches de ceux des années 2000, voire plus élevés pour certaines régions.

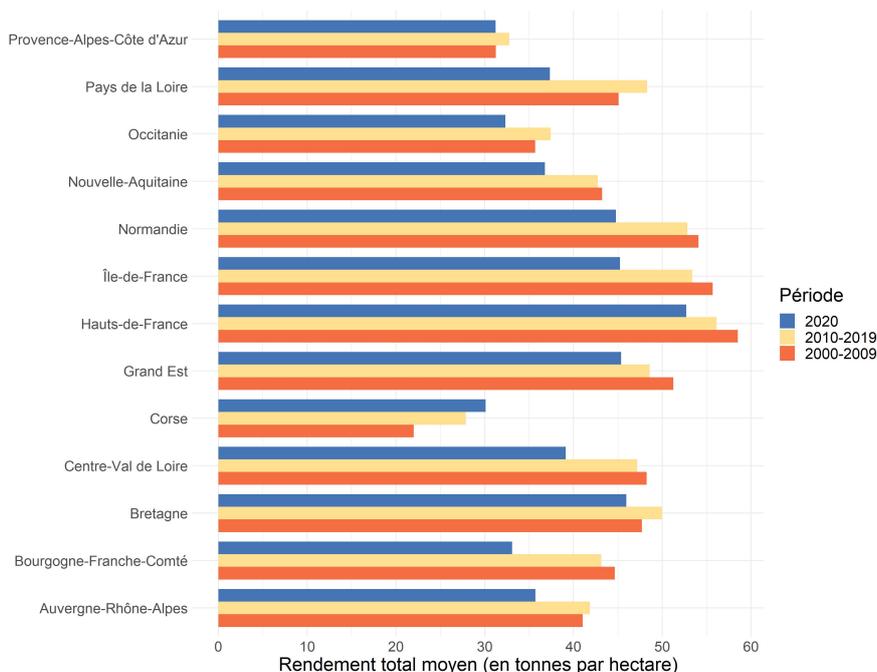


FIGURE 12.4 – Rendements moyens par région - Toutes cultures considérées

Intéressons-nous plus en détail à l'évolution des rendements sur la période d'étude pour une culture en particulier : le blé. La Figure 12.5 suivante représente cette évolution pour les cinq régions les plus productrices.

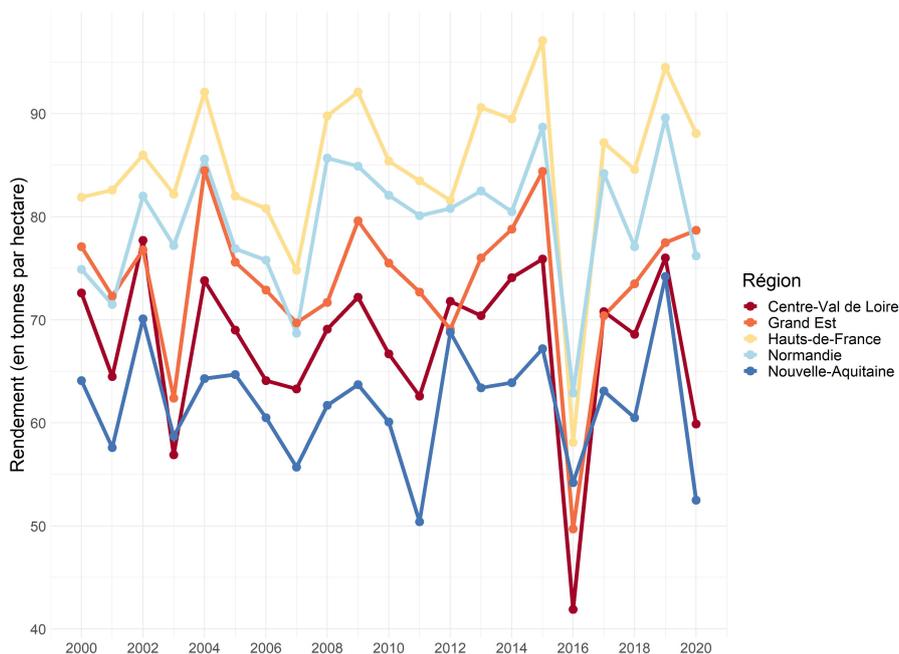


FIGURE 12.5 – Evolution des rendements de blé sur la période 2000-2020 pour les cinq régions les plus productrices

Sur la Figure 12.5, l'impact des conditions météorologiques sur les cultures est bien visible. En effet, pour toutes les régions, les rendements de blé sont marqués par une forte baisse en 2016, soit l'année de la pire récolte depuis 1976 [17]. Une baisse plus modérée est également à noter en 2003, année de canicule et de sécheresse qui n'a pas épargné les récoltes.

12.1.3 Traitement des données

12.1.3.1 Méthodes de traitement appliquées

Pour obtenir des données sous la forme de pertes de rendements, il suffit de calculer l'écart relatif des rendements annuels vis-à-vis d'une moyenne de référence. Cette dernière peut être déterminée de deux manières.

Pour la première approche, dite **historique**, les données de rendements sont rapportées à la moyenne de rendements des cinq années précédentes. Cette méthode présente toutefois plusieurs limites. Tout d'abord, elle induit une réduction de la période d'étude des pertes de rendements, ces dernières ne pouvant être calculées pour les cinq premières années de l'échantillon. De plus, le référentiel évoluant au fil des années, une perte de rendement extrême relative aux premières années de la période étudiée ne pourrait plus être considérée comme extrême, du fait d'une tendance à la baisse dans l'évolution des rendements.

Pour pallier ce problème, une **mise en *as if*** peut être une solution. De manière générale, cette méthode consiste à tenir compte des sinistres comme s'ils survenaient aujourd'hui. Dans le cadre de cette étude, la moyenne de référence peut donc être fixée à la moyenne de rendements sur les dernières années d'historique, 2016-2020. Si cette approche ne conduit pas à une exclusion des cinq premières années de la période d'étude, elle possède en revanche une autre limite : si les années considérées ont été marquées par de faibles rendements, la moyenne de référence risque d'être assez basse. Ainsi, les pertes de rendements calculées pourraient être moins nombreuses et d'intensité plus faible. En particulier, dans ce cas d'étude précis, l'année 2016 risque d'influencer fortement la moyenne de référence, comme observé sur la Figure 12.5 pour le blé.

12.1.3.2 Analyse des pertes de rendements extrêmes

Sur les graphiques suivants, nous comparons le nombre de données obtenu selon chaque méthode de calcul considérée, en tenant compte des pertes de rendements supérieures à 15%, et ce par année, culture et région.

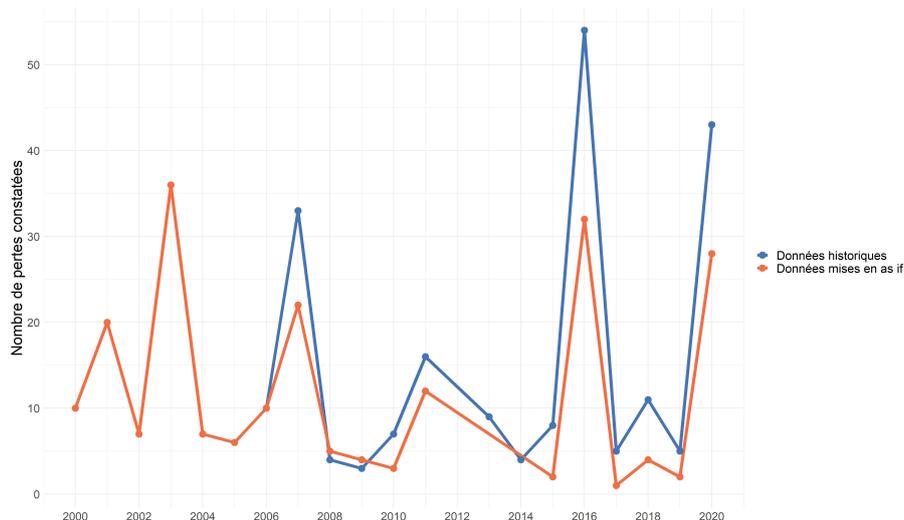


FIGURE 12.6 – Nombre de pertes de rendements supérieures à 15% constatées entre 2000 et 2020

Il ressort de la Figure 12.6 que la mise en *as if* atténue la valeur des pertes de rendements et donc *a fortiori* le nombre de pertes supérieures à 15%. Ce phénomène est particulièrement

visible pour l'année 2016, pour laquelle plus de 20 événements ne sont pas recensés comme des pertes de rendements supérieures à 15%. Par ailleurs, la Figure 12.6 ne présente pas spécialement de tendance calendaire liée au changement climatique.

Pour plus de cohérence, nous excluons des deux Figures 12.7 et 12.8 suivantes les pertes de rendements obtenues par la méthode de mise en *as if* pour les années antérieures à 2005. Commençons avec une analyse par culture.

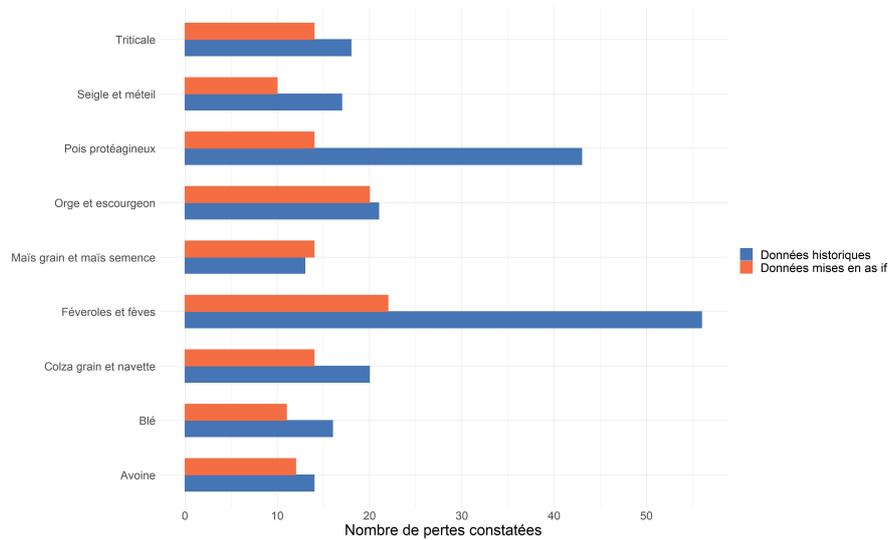


FIGURE 12.7 – Nombre de pertes de rendements supérieures à 15% constatées par culture

Une fois encore, la mise en *as if* réduit le nombre de pertes de rendements extrêmes constatées sur la période 2005-2020, quelle que soit la culture considérée. En outre, la considération de données historiques induit de très nombreuses pertes de rendements pour les cultures de pois protéagineux et de féveroles et fèves.

Enfin, nous comparons le nombre de données obtenu par région selon les deux approches de traitement sur la Figure 12.8.

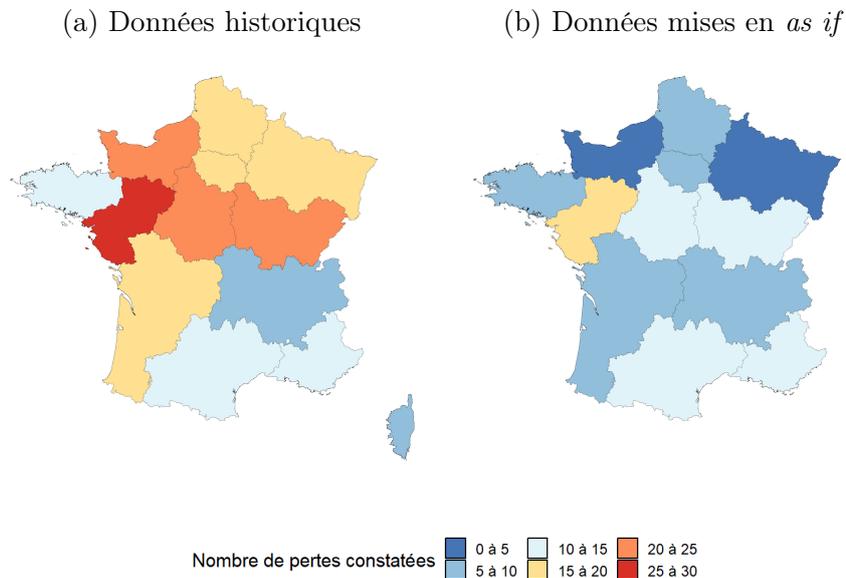


FIGURE 12.8 – Nombre de pertes de rendements supérieures à 15% constatées par région

A l'exception de la Corse, la mise en *as if* conduit à une diminution du nombre de pertes de rendements observées par région. De plus, l'impact des régions Normandie et Grand Est risque d'être mal capté dans le calibrage des paramètres de la GPD, du fait d'un trop faible nombre d'observations. A l'inverse, l'approche historique permet l'obtention de plus de 5 pertes de rendements par région, ce qui pourra favoriser la significativité de la variable régionale dans les calibrages dynamiques de la GPD. Quelle que soit la méthode considérée, le Pays de la Loire est la région qui dénombre le plus de pertes de rendements extrêmes.

Pour conclure cette analyse des pertes de rendements, nous comptons le nombre de données observées par région et par année, puis nous recensons le pourcentage d'observations manquantes par période sur la Table 12.2. Cette étude nous permet de contrôler la qualité de l'échantillon considéré avant d'effectuer des calibrages dynamiques.

	Données historiques	Données mises en <i>as if</i>
Période 2000-2004	Aucune donnée disponible	52%
Période 2005-2009	60%	61%
Période 2010-2014	60%	89%
Période 2015-2020	46%	68%

TABLE 12.2 – Pourcentage de données manquantes par période considérée

Pour les années postérieures à 2010, la mise en *as if* réduit drastiquement la taille de l'échantillon à considérer pour le calibrage des paramètres de la loi de Pareto généralisée. S'il est tout à fait normal d'observer peu de données par période considérée, dans la mesure où nous analysons des événements extrêmes donc rares, il faudra toutefois avoir conscience que le calibrage risque de dépendre fortement de la structure de l'échantillon et donc pourra s'avérer biaisé.

12.2 Données climatiques

La modélisation d'événements extrêmes conduit à une réduction drastique de la base de données. Dès lors, le calibrage de paramètres risque de ne pas être significatif si l'on multiplie le nombre de variables considérées. Ainsi, comme dans la Partie III, nous ne tiendrons compte que d'un seul indicateur, le *Soil Wetness Index*. Si les bases utilisées sont les mêmes que dans la partie précédente et ont déjà été décrites au préalable dans le Chapitre 10, cette section met l'accent sur les données régionales extraites de ces deux bases.

12.2.1 Données historiques

12.2.1.1 Source et traitements

Pour rappel, la base Météo France² fournit des valeurs mensuelles de l'indicateur SWI pour la période 1969-2021. Chaque valeur tient compte du mois en cours ainsi que des deux mois précédents, en moyennant les moyennes mensuelles des valeurs quotidiennes du SWI. Cette base offre également une très grande granularité spatiale : la France métropolitaine est découpée en 8 981 mailles de 8 kilomètres de côté.

2. https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=301&id_rubrique=40

Pour construire une base de données de SWI à l'échelle régionale, nous avons tout d'abord associé les mailles à leur région grâce à un logiciel SIG (Système d'Information Géographique), QGIS. Puis les indicateurs relatifs au mois de **mai** ont été moyennés pour chaque région. Cela permet de tenir également compte des données d'humidité des sols des mois de mars et d'avril, qui pourraient expliquer l'observation de pertes de rendements dues à une croissance altérée des cultures considérées.

12.2.1.2 Analyse des données

La Figure 12.9 suivante représente le niveau moyen de l'indice d'humidité des sols par région pour la période 2000-2020.

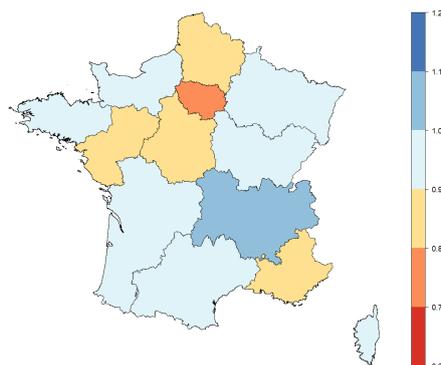


FIGURE 12.9 – Indice d'humidité des sols moyen sur la période 2000-2020

Il ressort de la Figure 12.9 que la région Île-de-France est celle qui, en moyenne, est la plus sèche. De plus, puisque le SWI des régions Grand Est et Bourgogne-Franche-Comté est très proche voire supérieure à 1, celles-ci sont caractérisées par la plus forte humidité des sols. Néanmoins, il est à noter que les traitements effectués sur les données Météo France fournissent uniquement une vision agrégée à l'échelle régionale. Certains effets locaux peuvent donc être masqués par cette agrégation.

Sur la Figure 12.10 suivante sont comparés les niveaux de l'indice d'humidité des sols pour deux années marquées par de fortes pertes de rendements agricoles.

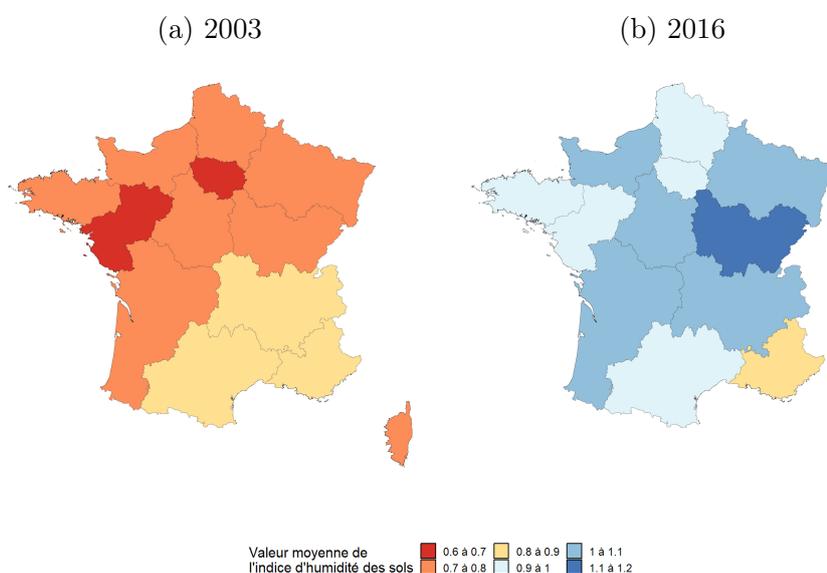


FIGURE 12.10 – Indice d'humidité des sols moyen constaté par région en 2003 et 2016

Si la sécheresse subie en 2003 est bien visible sur la Figure 12.10a, le printemps pluvieux enregistré en France en 2016 est également observable sur la Figure 12.10b. En particulier, les régions Grand Est, Centre-Val de Loire et Nouvelle-Aquitaine, qui produisent plus de la moitié de la récolte de blé français, ont connu une saturation en eau, le SWI dépassant la valeur seuil de 1. En raison de ce surplus d'eau, les mauvaises herbes et nuisibles se sont développés, conduisant à une détérioration des cultures [17].

12.2.2 Données de projection

12.2.2.1 Source et traitements

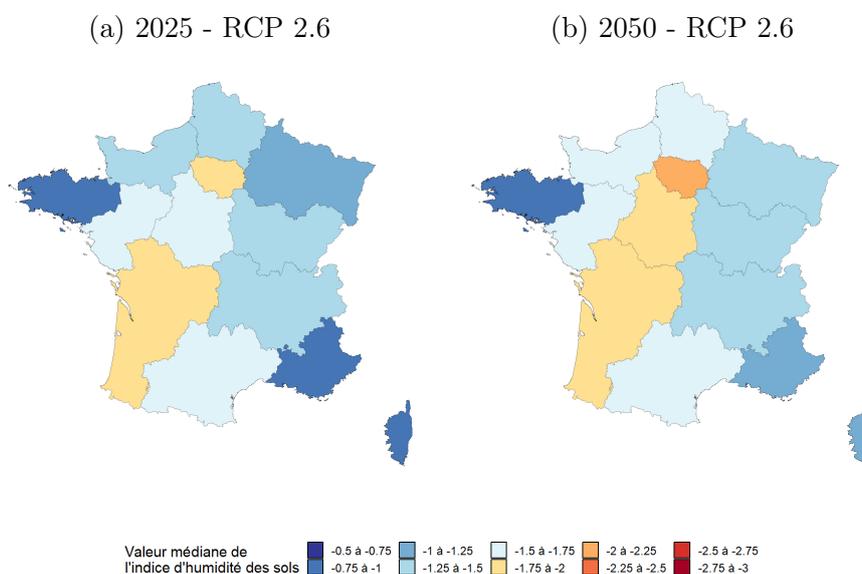
Une nouvelle fois, les données historiques de Météo France ont été complétées par des données émanant du *Climate Impact Explorer*³, une banque de données de projections développée par *Climate Analytics* en collaboration avec le NGFS.

Pour rappel, les projections sont disponibles à partir de 2015 de manière bidécennale par saison et par année, pour de nombreux pays et de nombreuses provinces. Pour la France, le niveau de granularité spatiale maximal est régional. Ainsi, pour rester en cohérence avec nos choix temporels pour les données historiques, nous choisissons les prévisions relatives aux mois de mars à mai et à l'échelle régionale.

12.2.2.2 Analyse des données

La Figure 12.11 suivante compare les projections de l'indice d'humidité des sols obtenues selon les deux scénarios du GIEC les plus adverses.

Sur la période 2015-2025, tous les scénarios du GIEC offrent les mêmes projections, comme en témoignent les Figures 12.11a et 12.11c. A partir de 2030, l'effet des hypothèses de modélisation se fait ressentir. En particulier, en 2050, là où le SWI se situe entre -1,75 et -3 sur toute la France avec le scénario RCP 8.5, les trajectoires du scénario RCP 2.6 le placent plutôt sur une échelle de -0,5 à -2,25. Plus localement, si la Bretagne, la Corse et Provence-Alpes-Côte d'Azur expérimentent un indicateur de sécheresse inférieur à -1,25 dans le scénario RCP 2.6, le choix du scénario RCP 8.5 induit pour ces régions une chute d'environ 1 point du SWI.



3. <https://climate-impact-explorer.climateanalytics.org/>

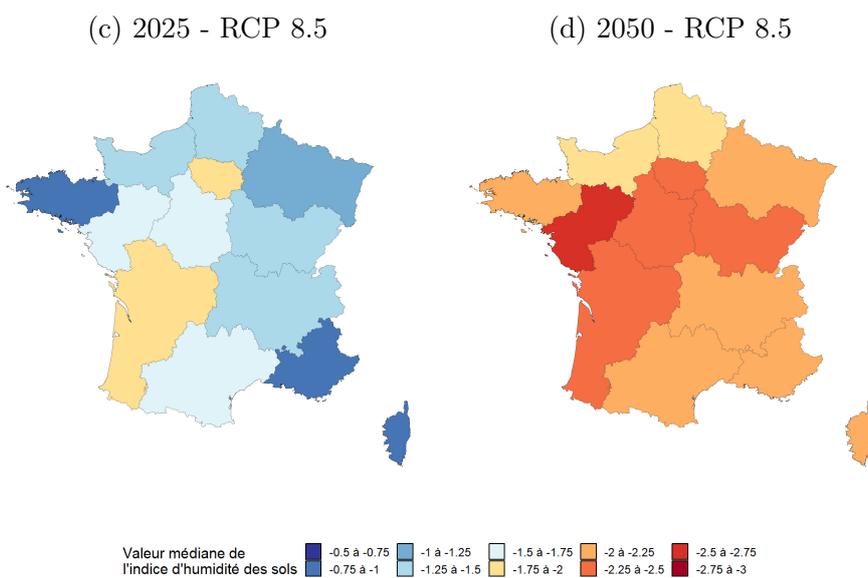


FIGURE 12.11 – Projections de l'indice d'humidité des sols selon les scénarios RCP 2.6 et 8.5 du GIEC

Chapitre 13

Calibrage de lois de pertes de rendements extrêmes

Ce dernier chapitre vise à évaluer l'intérêt de la prise en compte d'effets fixes et de variables climatiques dans le calibrage de lois de dépassements de seuil. Dans cet exemple, ces lois seront associées à des pertes de rendements extrêmes. En particulier, la pertinence de ce type de paramétrisation sera étudiée à travers l'analyse de probabilités de retour projetées pour des années futures. Pour ce faire, trois modèles seront considérés :

- le modèle **M1**, qui correspond à un calibrage statique des paramètres de la GPD, soit une estimation classique par maximum de vraisemblance,
- le modèle **M2**, qui se rapporte à un calibrage dynamique des paramètres de la GPD, pour lequel des variables d'effets fixes (année, région, culture) pourront être considérées,
- et le modèle **M3**, qui consiste en un approfondissement de la méthode dynamique, pour lequel les variables climatiques seront intégrées au modèle de calibrage des paramètres de la GPD.

L'accent sera également mis sur l'étude de l'influence du calcul des pertes de rendements sur les résultats obtenus. Ainsi, chacun des modèles sera calibré deux fois : avec des données de pertes de rendements historiques et avec des données mises en *as if*. Ainsi, pour chaque approche de traitement de données, trois étapes seront effectuées :

1. Choix du seuil à partir de l'étude de la *mean excess function*.
2. Calibrage des modèles **M1**, **M2** et **M3**.
3. Calcul des probabilités de retour pour le niveau de perte de 30% pour les années 2025 et 2050.

Dans le cadre de cette étude, seule la sévérité de la loi des dépassements de seuil sera modélisée. En effet, l'accent est mis sur l'analyse de l'intérêt de l'ajout de covariables aux calibrages des lois GPD. La modélisation de la loi de fréquence étant éludée, les probabilités de retour mentionnées seront conditionnelles à la survenance d'un événement extrême.

13.1 Calibrage avec données historiques

13.1.1 Choix du seuil

La première étape du processus de calibrage consiste à choisir le seuil à partir duquel est calibrée la loi des dépassements de seuil. Pour ce faire, il est nécessaire d'étudier la *mean excess function* à partir de la Figure 13.1 ci-après.

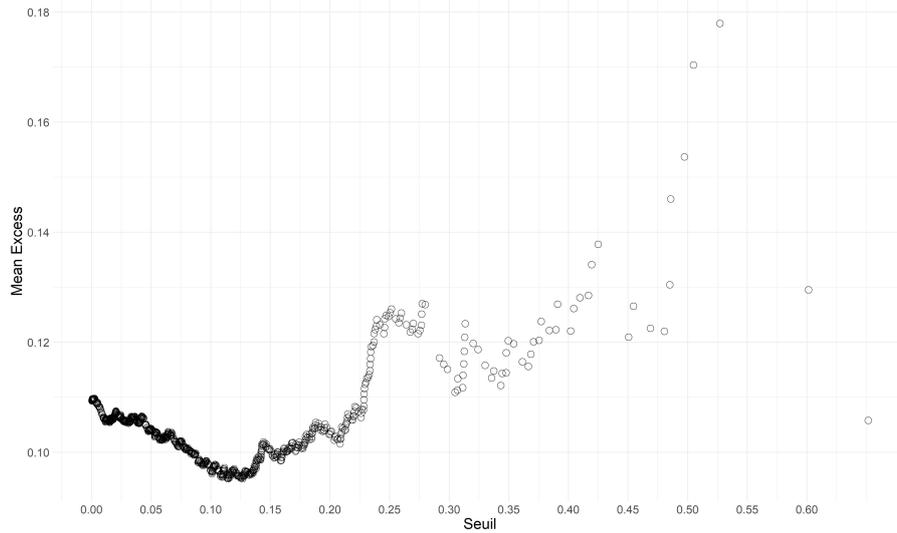


FIGURE 13.1 – *Mean excess plot* relatif aux données historiques

Pour choisir un bon seuil, il faut déterminer à partir de quelle valeur la *mean excess function* est linéaire. En parallèle, le seuil ne doit pas être trop élevé, sous peine de ne pas bien calibrer la distribution des valeurs extrêmes. Or, la *mean excess function* présentée sur la Figure 13.1 n’a pas un comportement spécialement linéaire, même en choisissant un seuil très élevé. Par conséquent, nous choisissons grossièrement le seuil de **16%**, guidés par la faible tendance linéaire existant jusqu’au seuil de 23% de pertes.

Pour confirmer le choix du seuil de 16%, nous traçons les graphiques de stabilité des paramètres σ^* et ξ .

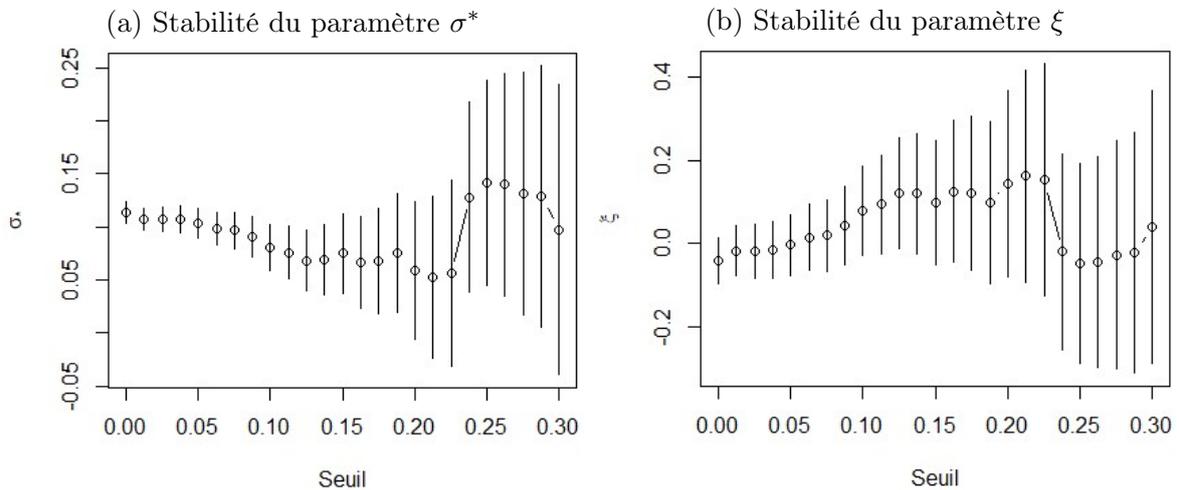


FIGURE 13.2 – Graphiques de stabilité des paramètres de la loi GPD pour les données historiques

Les graphiques précédents ne laissent pas apparaître une stabilité des paramètres σ^* et ξ vis-à-vis du choix du seuil. Par conséquent, dans l’optique d’avoir un nombre suffisant de données pour calibrer des modèles dynamiques dépendants de covariables, nous décidons de conserver le seuil à **16%**. En effet, pour une meilleure significativité des covariables utilisées, il sera nécessaire d’estimer les modèles sur des échantillons de taille conséquente.

13.1.2 Modèles calibrés

13.1.2.1 Modèle M1

Pour calibrer les paramètres de la loi GPD de manière statique, il suffit d'employer la fonction `gpd` de R, qui effectue une estimation par maximum de vraisemblance. Les valeurs des paramètres ainsi que leurs écarts-type sont présentés sur la Table 13.1.

Paramètre ξ	Paramètre β
0,1216 (0,0848)	0,0870 (0,0096)

TABLE 13.1 – Paramètres du modèle **M1** - Données historiques

Le paramètre de forme ξ obtenu est positif, ce qui est cohérent avec la tendance croissante de la *mean excess function* observée sur la Figure 13.1.

La Figure 13.3 suivante fournit des éléments graphiques permettant de vérifier le modèle calibré précédemment.

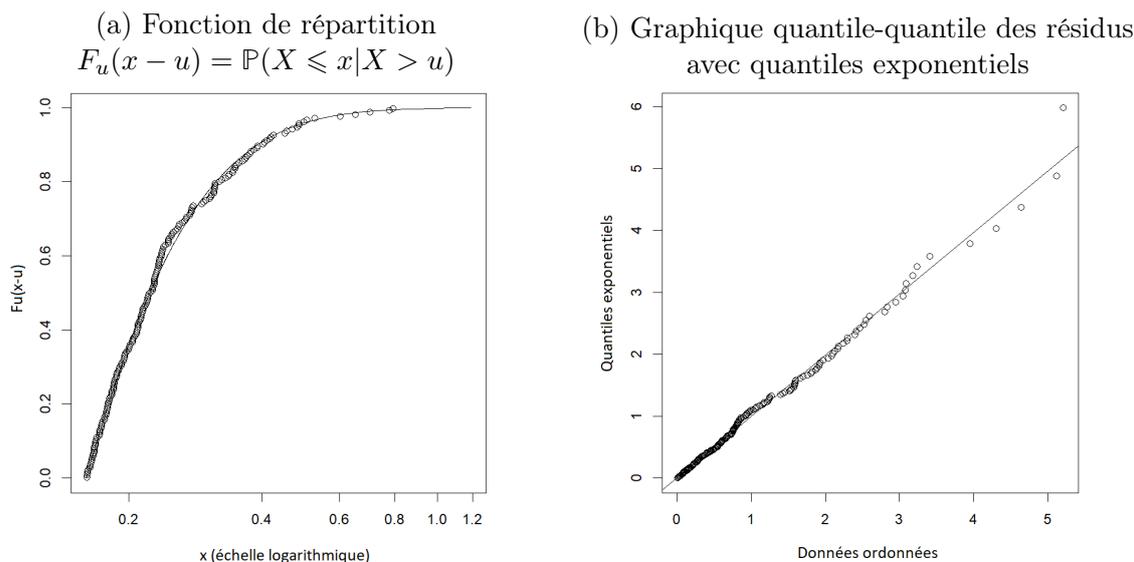


FIGURE 13.3 – Détails du modèle **M1** calibré sur les données historiques

Le modèle calibré semble bien adapté aux données. En effet, sur la Figure 13.3a, la fonction de répartition calibrée est proche de la fonction de répartition empirique. De plus, les quantiles empiriques des résidus de la Figure 13.3b sont alignés sur la bissectrice, signalant ainsi la cohérence du calibrage.

13.1.2.2 Modèle M2

Pour obtenir le modèle **M2**, de nombreuses spécifications des paramètres ont été étudiées. Selon le modèle de Chavez-Demoulin, Embrechts et Hofert [10], les paramètres ξ et ν sont tous deux estimés à partir de modèles linéaires, voire additifs généralisés. Pour cette étude, seuls les modèles linéaires ont été appliqués. Les covariables possibles sont :

- la culture c ;

- la région k ;
- et l'année t associées à la perte de rendement.

Il faut alors formuler deux modèles linéaires à partir de ces covariables. Il est ressorti de nos analyses que, quel que soit le modèle choisi pour ξ , le modèle le plus significatif pour ν n'utilise que la région comme covariable, soit :

$$\nu_{c,k,t} = \alpha_k \times \mathbb{1}_{\text{Région}_k}.$$

Concernant la prédiction du paramètre ξ en fonction des covariables précédentes, de nombreux modèles ont été testés. Néanmoins, l'algorithme appliquant la méthode de maximum de vraisemblance pénalisée fournissait une vraisemblance finale infinie pour la plupart des modèles. En conséquence, le calcul de critères d'information tels que l'AIC (*Akaike Information Criterion*) n'aboutissant pas, le choix du modèle final s'est effectué selon la qualité de la prédiction, donc la valeur du R^2 , et la significativité des paramètres. La Table 13.2 fait l'inventaire des modèles qui ont été testés et fournit leurs résultats, notamment le nombre de coefficients non-significatifs au seuil de 10%.

Modèle pour ξ	$R^2 \xi$	$R^2 \nu$	Déviante expliquée ξ	Déviante expliquée ν	Nb. coeffs non-significatifs ξ	Nb. coeffs non-significatifs ν
$f(\text{Région})$	0,0611	-0,0455	19,0%	63,30%	9	1
$f(\text{Culture})$	0,14	0,0293	24,5%	72,2%	4	0
$f(\text{Année})$	-5,65e-06	-0,0152	1,00%	58,1%	1	0
$f(\text{Région, Année})$	0,893	0,0827	97,9%	81,2%	0	0
$f(\text{Culture, Région})$	0,96	0,261	99,30%	88,5%	2	0
$f(\text{Culture, Année})$	0,661	0,115	91,30%	85,4%	0	0
$f(\text{Culture, Région, Année})$	0,973	0,23	99,5%	89,5%	0	0

TABLE 13.2 – Spécifications de modèles testées pour le choix de **M2** pour les données historiques

Parmi les sept modèles initiaux, il ne reste que trois candidats qui offrent une significativité satisfaisante. En termes de qualité de prédiction, le modèle $f(\text{Culture, Région, Année})$ offre de très bons résultats pour la régression de ξ , et l'un des meilleurs R^2 obtenus pour la régression de ν . Un point essentiel à noter est la difficulté pour obtenir deux régressions de qualité correcte. En effet, si la prédiction est très bonne pour l'un des paramètres, le second peut pâtir du modèle choisi pour le premier, dans la mesure où les deux paramètres dépendent l'un de l'autre.

Bien que le modèle $f(\text{Culture, Région})$ soit moins significatif que les trois autres, il offre la meilleure prédiction du paramètre ν ainsi qu'une excellente qualité de prédiction pour ξ . Toutefois, en plus de son manque de significativité, ce modèle ne permet pas de capter des effets temporels, qui pourraient être intéressants à étudier lors de la phase de projection. Par conséquent, le modèle final retenu pour ξ reste :

$$\xi_{c,k,t} = \alpha_c \times \mathbb{1}_{\text{Culture}_c} + \beta_k \times \mathbb{1}_{\text{Région}_k} + \gamma_t \times \text{Année}_t, \text{ en jaune dans la Table 13.2.}$$

Pour des soucis de lisibilité, la Table 13.3 suivante résume les impacts des covariables sur les deux paramètres de la loi GPD. Les résultats détaillés des régressions linéaires sont disponibles en Annexe F.1.

	Impact sur ξ	Impact sur ν
Culture	De -0,158 à +0,756	Nul
Région	De +29,23 à +30,22	De -3,933 à -1,805
Année	-0,015 \times Année	Nul

TABLE 13.3 – Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle **M2** - Données historiques

Pour la régression sur ξ , la modalité de référence est l’avoine : les coefficients des autres cultures correspondent donc aux variations du paramètre ξ initialement calculé pour l’avoine. Le cas où le paramètre de forme est minimal correspond à la combinaison (Blé, Grand Est, 2020). ξ vaut alors -0,938. Puisque $\xi > -1$, la reparamétrisation du modèle (ξ, β) en (ξ, ν) est donc valide.

13.1.2.3 Modèle M3

Pour le modèle **M3**, l’indicateur SWI se rajoute à la liste des covariables. Comme pour le modèle précédent, le meilleur choix de régression linéaire sur ν est le suivant :

$$\nu_{c,k,t} = \alpha_k \times \mathbb{1}_{\text{Région}_k}.$$

Pour inclure la variable climatique dans les modélisations de nombreuses spécifications de régressions linéaires sur ξ ont été testées :

- Région et SWI ainsi que ses retards ;
- Culture, Région et SWI accompagné de ses retards ;
- Culture, Région, Année et SWI ou ses retards ;
- et enfin Culture, Région, SWI ou ses retards ; en tenant compte d’effets distincts selon la valeur prise par le SWI (inférieure ou supérieure à 1).

Les résultats des modèles testés sont résumés en Table 13.4. Une fois encore, la vraisemblance retournée par l’algorithme étant presque toujours infinie, la sélection du modèle final s’effectuera à partir de critères de qualité de prédiction et de significativité des coefficients, au moins au seuil de 10%.

Presque tous les modèles conduisent à une significativité idéale sur les coefficients de la régression de ν . Parmi les dix-neuf modèles testés, neuf d’entre eux peuvent être éliminés en raison de la faiblesse des prédictions fournies sur ν . Sur les dix restants, 5 autres ne donnent pas de coefficients significatifs au seuil de 10% pour la variable climatique. Ces modèles ne pouvant confirmer l’effet de l’indice d’humidité des sols sur les pertes de rendements, ils ne sont pas retenus. Il reste alors les cinq modèles suivants, classés par R^2 obtenu sur ν décroissant :

- $f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t)$, de R^2 0,270, qui a fournit des résultats incohérents lors de la phase de projections : les probabilités de pertes étant plus faibles dans le cas RCP 2.6 que RCP 8.5, tandis que ce dernier scénario est plus pessimiste ;

- $f(\text{Région}, \text{Culture}, \text{SWI}_t)$, de R^2 0,204 et pour lequel la régression sur ξ donne des coefficients non-significatifs sur une région et une culture au seuil de 10% ;
- $f(\text{Région}, \text{Culture}, \text{SWI}_t, \text{SWI}_{t-2})$, de R^2 0,192 ;
- $f(\text{Région}, \text{Culture}, \text{Année}; \text{SWI}_{t-2})$, de R^2 0,114 ;
- et $f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-1} \times \mathbb{1}_{\{\text{SWI}_{t-1} > 1\}})$, de R^2 0,109.

Au vu des difficultés rencontrées avec les deux premiers modèles, nous retenons le troisième, en jaune dans la Table 13.4 et d'équation

$$\xi_{c,k,t} = \alpha_c \times \mathbb{1}_{\text{Culture}_c} + \beta_k \times \mathbb{1}_{\text{Région}_k} + \gamma_{k,t}^{(1)} \times \text{SWI}_{k,t} + \gamma_{k,t}^{(2)} \times \text{SWI}_{k,t-2}.$$

En particulier, nous avons préféré ce modèle à la spécification $f(\text{Région}, \text{Culture}, \text{Année}, \text{SWI}_{t-2})$ pour s'assurer de la significativité des effets de chacune des régions sur la valeur du paramètre de forme ξ , bien que cela conduise à une diminution de la qualité de la prédiction de ν .

Modèle pour ξ	$R^2 \xi$	$R^2 \nu$	Déviante expliquée ξ	Déviante expliquée ν	Nb. coeffs non-significatifs ξ	Nb. coeffs non-significatifs ν
$f(\text{Région}, \text{SWI}_t)$	0,959	0,042	99,3%	82,5%	0	1
$f(\text{Région}, \text{SWI}_{t-1})$	0,057	-0,050	19,4%	62,5%	14	0
$f(\text{Région}, \text{SWI}_{t-2})$	-0,070	-0,040	2,9%	62,9%	14	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-1})$	0,952	0,035	99,1%	85,4%	0	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-2})$	0,950	0,037	99,1%	82,2%	1 (SWI)	1
$f(\text{Région}, \text{SWI}_{t-1} + \text{SWI}_{t-2})$	0,077	-0,044	22,3%	61,9%	15	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-1}, \text{SWI}_{t-2})$	0,913	0,040	98,3%	84,4%	1 (SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t)$	0,981	0,204	99,6%	88,0%	2	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-1})$	0,945	0,235	99,0%	86,8%	6 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-2})$	0,891	0,100	98,1%	86,0%	9 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t, \text{SWI}_{t-1})$	0,879	0,207	97,6%	87,7%	8 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t, \text{SWI}_{t-2})$	0,983	0,192	99,7%	90,5%	0	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-1}, \text{SWI}_{t-2})$	0,919	0,083	98,5%	84,9%	10 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t)$	0,969	0,270	99,5%	90,0%	1	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-1})$	0,969	0,212	99,4%	89,8%	2 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-2})$	0,974	0,114	99,5%	88,5%	0	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t \times \mathbb{1}_{\{\text{SWI}_t > 1\}})$	1,000	-0,027	100%	31,60%	1	7
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-1} \times \mathbb{1}_{\{\text{SWI}_{t-1} > 1\}})$	0,979	0,109	99,60%	88%	0	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-2} \times \mathbb{1}_{\{\text{SWI}_{t-2} > 1\}})$	0,991	0,395	99,80%	91,10%	2 (dont SWI)	0

TABLE 13.4 – Spécifications de modèles testées pour le choix de **M3** pour les données historiques

Les résultats des deux régressions sont disponibles en Annexe F.2. La Table 13.5 ci-après résume les effets des différentes covariables sur les paramètres de la GPD pour le modèle retenu.

	Impact sur ξ	Impact sur ν
Culture	De -0,156 à +0,382	Nul
Région	De +0,623 à +1,869	De -4,108 à -1,788
SWI_t	$-1,098 \times SWI_t$	Nul
SWI_{t-2}	$-0,458 \times SWI_{t-2}$	Nul

TABLE 13.5 – Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle **M3** - Données historiques

Le coefficient de l'avoine, modalité de référence, est nul. Les valeurs des coefficients des autres cultures permettent donc de modifier les résultats fournis pour l'avoine, toutes choses égales par ailleurs. Pour cette régression, les coefficients relatifs à l'indice d'humidité des sols et son retard sont de même signe. De plus,, l'impact de la région sur ν est du même ordre de grandeur que dans le modèle **M2** précédent. Par ailleurs, la combinaison (Pois protéagineux, Île-de-France) pourrait invalider la reparamétrisation du modèle (ξ, β) en (ξ, ν) , conduisant au paramètre de forme minimal sans tenir compte du SWI. Or, sur les données observées, ses évolutions sont telles que tous les ξ résultent comme supérieurs à -1.

13.1.3 Projection de pertes de rendements futures

Sur les Figures 13.4 et 13.5 suivantes sont présentées les prédictions de probabilité de perte de 30%¹, par région en 2025 et 2050 pour le blé, par l'intermédiaire des modèles **M2** et **M3**, ce dernier étant couplé aux projections de l'indice d'humidité des sols du NGFS.

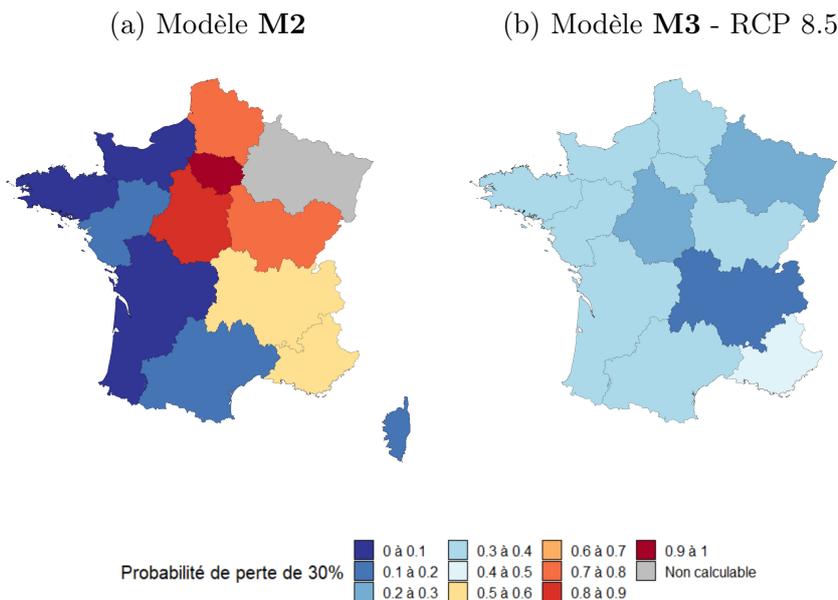


FIGURE 13.4 – Probabilité de perte à 30% projetée par région en 2025 selon les modèles **M2** et **M3** pour la culture de blé - Données historiques

1. conditionnelles à la survenance d'une perte supérieure à 16%

Les régions grisées sont celles pour lequel le calcul de la probabilité de retour n'était pas possible, que ce soit en raison d'un paramètre de forme inférieur à -1 ou d'un ensemble de définition de la distribution trop limité. A horizon 2025, toutes les projections du GIEC étant identiques, indépendamment du niveau de forçage radiatif, nous affichons uniquement les résultats pour le scénario RCP 8.5.

En 2025, le choix du modèle **M2** conduit à l'obtention de probabilités de retour assez hétérogènes à l'échelle régionale. En effet, pour six des treize régions métropolitaines, la probabilité de perte de 30% sachant la perte de rendements de 16% est supérieure à 50%. En particulier, en Île-de-France, l'occurrence de cette perte est quasi-certaine.

Avec le modèle **M3**, les prévisions du GIEC donnent une probabilité de perte plus raisonnable, allant de 30% à 40% pour la plupart des régions. Seule la Provence-Alpes Côte d'Azur est caractérisée par une probabilité de perte supérieure à 40%.

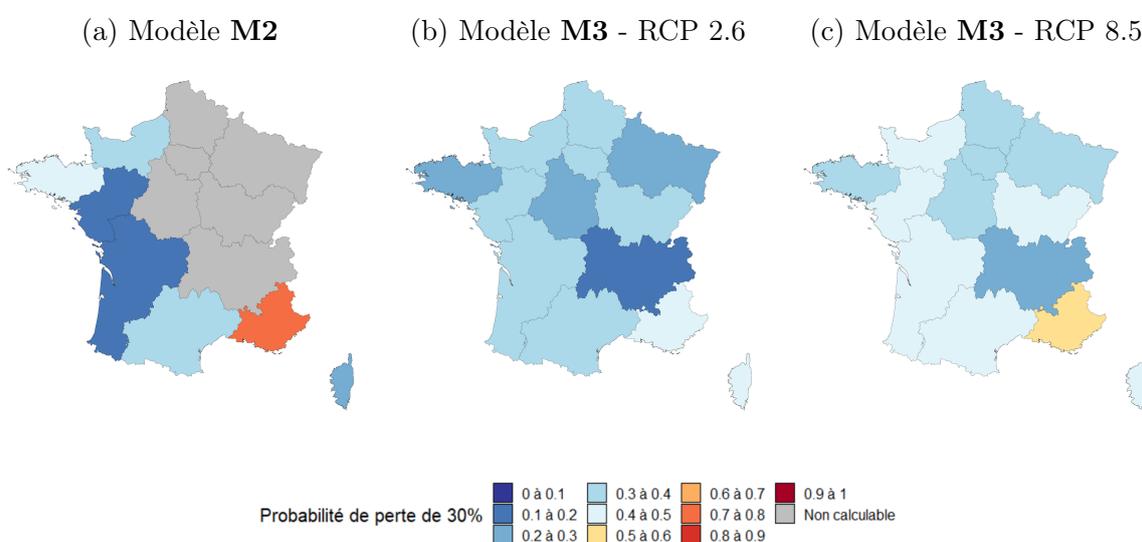


FIGURE 13.5 – Probabilité de perte à 30% projetée par région en 2050 selon les modèles **M2** et **M3** pour la culture de blé - Données historiques

Le modèle **M2** atteint ses limites en 2050, comme en témoigne la Figure 13.5a sur laquelle six régions apparaissent grisées. Seule la Corse connaît une diminution de la probabilité de perte de rendements. Pour les six régions restantes, cette dernière s'est accrue de 4% à 24%.

Les projections obtenues à partir du scénario RCP 8.5 sont plus pessimistes que celles du scénario RCP 2.6, ce qui est cohérent avec les hypothèses sous-jacentes de ces deux trajectoires. En effet, le RCP 8.5 modélise une plus grande sécheresse des sols, qui pourrait donc détériorer davantage les cultures. Par ailleurs, à l'exception de la Bretagne, les ordres de grandeur obtenus pour les probabilités de retour de la Figure 12.11b sont très proches de ceux calculés en 2025 avec le modèle **M3**.

A titre de comparaison, le modèle **M1** fournit une probabilité de perte de 23%, indépendamment de covariables spatio-temporelles. Des graphiques précédents, il ressort que le modèle **M2** ainsi spécifié n'est pas très pertinent pour deux raisons : tout d'abord car il ne permet pas toujours le calcul des probabilités de retour, mais également car celles fournies sont souvent bien plus élevées. De plus, considérer l'année de survenance comme effet linéaire dans le modèle possède certaines limites lorsque les projections sont effectuées sur un horizon lointain. A l'inverse, le modèle **M3** semble plus raisonnable.

13.2 Calibrage avec mise en *as if*

13.2.1 Choix du seuil

Une fois encore, la première étape du processus de calibrage consiste à choisir le seuil à partir duquel est calibrée la loi des dépassements de seuil. Pour ce faire, il est nécessaire d'étudier la *mean excess function* à partir de la Figure 13.1 suivante.

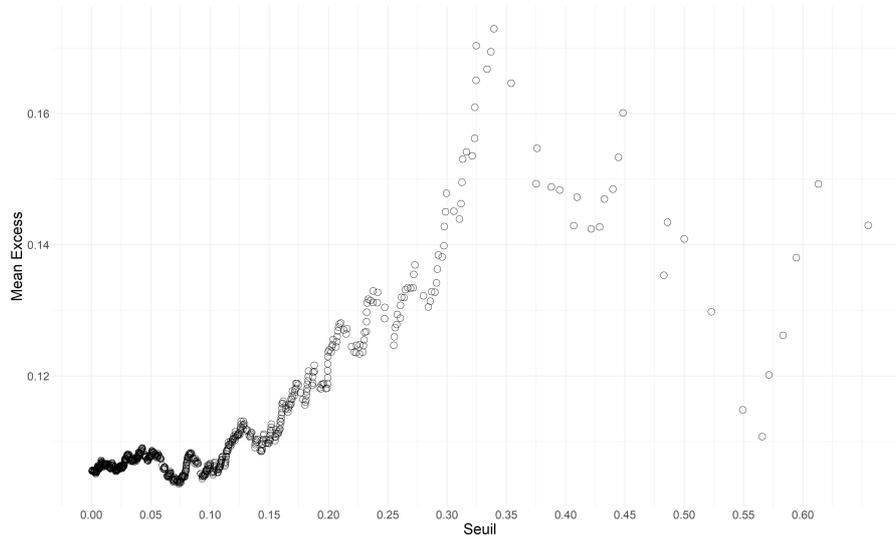


FIGURE 13.6 – *Mean excess plot* pour les données mises en *as if*

Sur les données mises en *as if*, la *mean excess function* présentée sur la Figure 13.6 semble caractérisée par une tendance linéaire à partir du seuil de **14%**. En conséquence, ce seuil permettra de calibrer les trois modèles de distributions des dépassements de seuil.

Le choix du seuil peut également être confirmé ou infirmé par les graphiques de stabilité des paramètres σ^* et ξ , présentés ci-après.

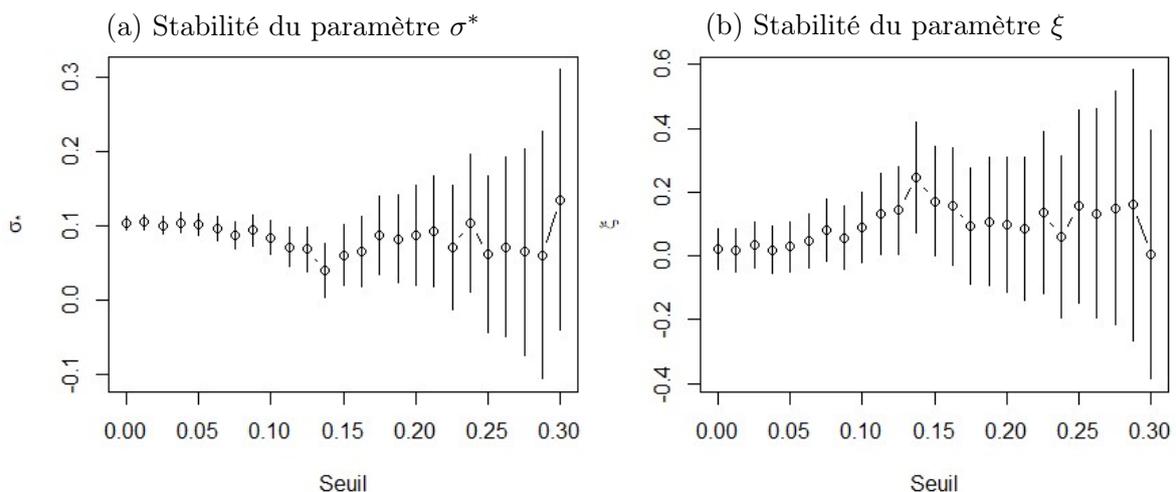


FIGURE 13.7 – Graphiques de stabilité des paramètres de la loi GPD pour les données mises en *as if*

Avec les données mises en *as if*, les paramètres σ^* et ξ sont moins dépendants du seuil que lorsqu'ils étaient calculés avec les données historiques. En effet, une plus grande stabilité est

observée sur les Figures 13.7a et 13.7b. Concernant le choix du seuil, tenir compte d'environ 20% de perte semble cohérent au vu des graphiques. Toutefois, accroître ce seuil peut poser des problèmes de significativité lors des calibrages des modèles **M2** et **M3** qui dépendent de covariables. Par conséquent, le seuil de 14% précédemment convenu est conservé pour la suite de l'étude.

13.2.2 Modèles calibrés

13.2.2.1 Modèle M1

De manière analogue à la section précédente, il suffit d'employer la fonction `gpd` de R pour calibrer les paramètres de la loi du modèle **M1**. Les valeurs de ces derniers ainsi que leurs écarts-type sont présentés sur la Table 13.6.

Paramètre ξ	Paramètre β
0,1883 (0,0830)	0,0825 (0,0086)

TABLE 13.6 – Paramètres du modèle **M1** - Données mises en *as if*

Ici encore, le paramètre de forme ξ obtenu est positif, ce qui est cohérent avec la tendance croissante de la *mean excess function* observée sur la Figure 13.6.

La Figure 13.8 suivante fournit des éléments graphiques permettant de vérifier le modèle calibré précédemment.

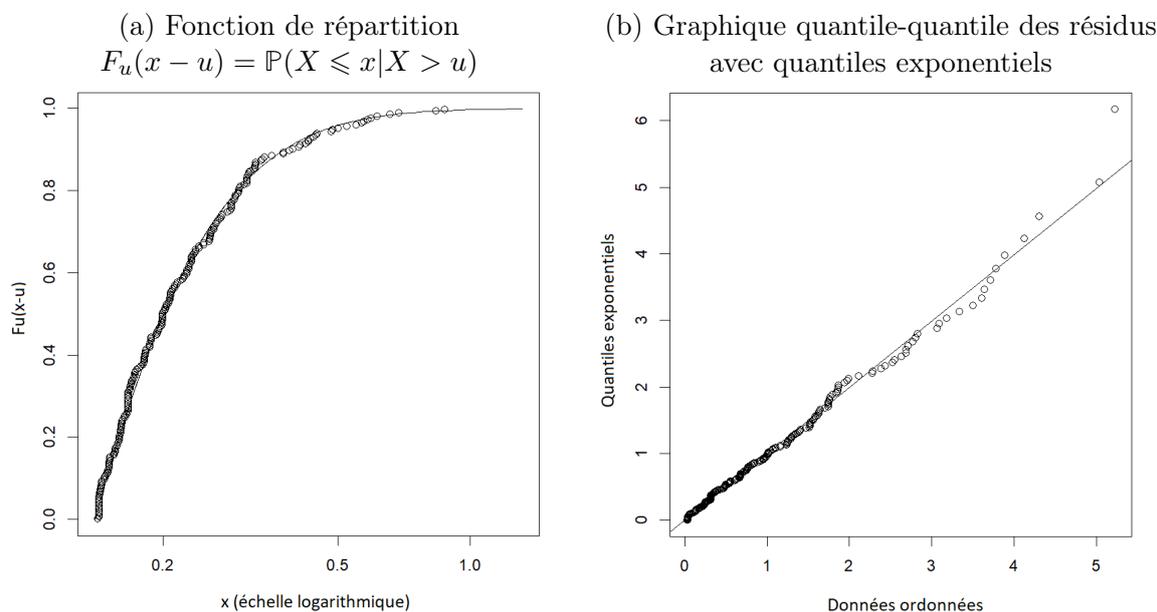


FIGURE 13.8 – Détails du modèle **M1** calibré sur les données mises *as if*

Le modèle calibré semble bien adapté aux données. En effet, sur la Figure 13.8a, la fonction de répartition calibrée est proche de la fonction de répartition empirique. De plus, les quantiles empiriques des résidus de la Figure 13.8b sont alignés sur la bissectrice, signalant ainsi la cohérence du calibrage.

13.2.2.2 Modèle M2

De manière analogue aux modèles présentés pour les données historiques, le meilleur choix de régression linéaire sur ν est le suivant :

$$\nu_{c,k,t} = \alpha_k \times \mathbb{1}_{\text{Région}_k}.$$

Là encore, la sélection du modèle final retenu pour ξ s'est effectué selon des critères de qualité des prédictions et de significativité des coefficients, au moins au seuil de 10%. Le panel des spécifications testées est résumé en Table 13.7.

Modèle pour ξ	R ² ξ	R ² ν	Déviance expliquée ξ	Déviance expliquée ν	Nb. coeffs non-significatifs ξ	Nb. coeffs non-significatifs ν
$f(\text{Région})$	0,765	0,1	89,8%	82,5%	2	0
$f(\text{Culture})$	0,193	-0,005	32,5%	64,5%	4	0
$f(\text{Année})$	4,45e-05	-0,014	0,00%	58,1%	1	0
$f(\text{Région, Année})$	0,685	0,05	87,6%	78,9%	0	0
$f(\text{Culture, Région})$	0,989	0,308	99,7%	89,0%	2	0
$f(\text{Culture, Année})$	0,39	-0,06	68,8%	71,3%	0	0
$f(\text{Culture, Région, Année})$	0,978	0,208	99,2%	86,8%	1	0

TABLE 13.7 – Spécifications de modèles testées pour le choix de **M2** avec les données mises en *as if*

Si la significativité des coefficients est essentielle, la faiblesse des prédictions fournies par les modèles $f(\text{Culture, Année})$ et $f(\text{Région, Année})$ pour les deux régressions nous obligent à les rejeter.

Seuls deux modèles candidats ressortent alors comme suffisamment prédictifs et significatifs : $f(\text{Culture, Région})$ et $f(\text{Culture, Région, Année})$. Le premier souffre d'une absence de captation d'effets temporels, qui peut s'avérer regrettable pour produire des projections de probabilités de retour. Le second fournit certes des prédictions de moins bonne qualité pour ν , mais possède une meilleure significativité : si moins de coefficients ne sont pas significatifs, pour les autres le seuil de significativité est également réduit. Nous retenons donc le modèle $f(\text{Culture, Région, Année})$, en jaune dans la Table 13.7 d'équation :

$$\xi_{c,k,t} = \alpha_c \times \mathbb{1}_{\text{Culture}_c} + \beta_k \times \mathbb{1}_{\text{Région}_k} + \gamma_t \times \text{Année}_t.$$

Les résultats des deux régressions sont disponibles en Annexe F.3. Notons que le seul coefficient significatif correspond aux fèves et féveroles. Si nous procédons aux prédictions sur d'autres cultures, alors cette absence de significativité ne sera plus problématique. Pour plus de clarté, les impacts de chacune des covariables sont résumés dans la Table 13.8.

	Impact sur ξ	Impact sur ν
Culture	De -0,218 à +0,277	Nul
Région	De -13,88 à -12,52	De -3,780 à -1,809
Année	+0,007 \times Année	Nul

TABLE 13.8 – Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle **M2** - Données mises en *as if*

A la différence du modèle **M2** obtenu avec les données historiques, l'année de survenance agit désormais de manière positive sur le paramètre de forme ξ . En outre, si la région influence moins ce dernier que dans le modèle historique, ses répercussions sur la valeur de ν sont assez similaires. Pour la combinaison (Pois protéagineux, Île-de-France, 2000), le paramètre de forme atteint la valeur minimale de -1,004. Dans ce cas, la reparamétrisation du modèle (ξ, β) en (ξ, ν) pourrait poser problème. Or, cette combinaison n'existant pas dans les données servant à l'estimation des paramètres, la reparamétrisation du modèle reste valable.

13.2.2.3 Modèle M3

Une fois encore, le meilleur choix de régression linéaire sur ν est le suivant :

$$\nu_{c,k,t} = \alpha_k \times \mathbb{1}_{\text{Région}_k}.$$

Pour prédire la valeur du paramètre de forme en fonction de covariables, les mêmes spécifications que pour le modèle **M3** avec données historiques ont été testées. Les résultats de chaque modèle sont présentés dans la Table 13.9 ci-après. La qualité de la prédiction de ξ est excellente pour tous les modèles. De plus, la régression sur ν ne pose aucun problème de significativité. Il faudra donc arbitrer entre la qualité de cette dernière et le nombre de coefficients non-significatifs obtenus pour le modèle sur ξ . En ne tenant compte que des régressions menant à moins de 3 coefficients non-significatifs, seules cinq spécifications conduisent à un R^2 sur ν supérieur à 0,20 :

- $f(\text{Culture}, \text{Région}, \text{SWI}_{t-2})$, de R^2 0,269 mais qui prédit de nombreux paramètres de forme inférieurs à -1 lors de la phase de projection ;
- $f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t \times \mathbb{1}_{\{\text{SWI}_t > 1\}})$, de R^2 0,238, pour laquelle l'indice d'humidité des sols a un effet du même signe et du même ordre de grandeur avant et après le seuil. Par conséquent, le modèle semble trop spécifique sans raison.
- $f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t)$, de R^2 0,216 ;
- $f(\text{Région}, \text{SWI}_{t-1})$, de R^2 0,212 ;
- et $f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-2})$, de R^2 0,208.

Compte tenu des remarques faites sur les deux premiers modèles, nous retenons le troisième pour des questions de qualité de prédiction. Ce dernier est encadré en jaune dans la Table 13.9 et a pour équation :

$$\xi_{c,k,t} = \alpha_c \times \mathbb{1}_{\text{Culture}_c} + \beta_k \times \mathbb{1}_{\text{Région}_k} + \gamma_t \times \text{Année}_t + \delta_{k,t} \times \text{SWI}_{k,t}.$$

Modèle pour ξ	$R^2 \xi$	$R^2 \nu$	Déviante expliquée ξ	Déviante expliquée ν	Nb. coeffs non-significatifs ξ	Nb. coeffs non-significatifs ν
$f(\text{Région}, \text{SWI}_t)$	0,939	0,090	98,7%	83,7%	4	0
$f(\text{Région}, \text{SWI}_{t-1})$	0,959	0,212	98,7%	84,8%	1	0
$f(\text{Région}, \text{SWI}_{t-2})$	0,941	0,094	98,2%	83,7%	1	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-1})$	0,959	0,129	98,8%	85,4%	3	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-2})$	0,956	0,161	98,7%	84,8%	5	0
$f(\text{Région}, \text{SWI}_{t-1} + \text{SWI}_{t-2})$	0,932	0,102	97,9%	81,5%	1	0
$f(\text{Région}, \text{SWI}_t, \text{SWI}_{t-1}, \text{SWI}_{t-2})$	0,927	0,150	97,7%	85,7%	5 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t)$	0,988	0,285	99,7%	89,7%	3	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-1})$	0,952	0,359	98,6%	89,5%	10	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-2})$	0,991	0,269	99,7%	90,9%	1	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t, \text{SWI}_{t-1})$	0,95	0,272	98,7%	88,6%	10 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_t, \text{SWI}_{t-2})$	0,986	0,254	99,6%	90,7%	4 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{SWI}_{t-1}, \text{SWI}_{t-2})$	0,991	0,269	99,7%	90,7%	2 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t)$	0,979	0,216	99,4%	86,8%	1	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-1})$	0,938	0,149	98,7%	85,7%	15 (dont SWI)	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-2})$	0,976	0,208	99,2%	89,4%	0	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_t \times \mathbb{1}_{\{\text{SWI}_t > 1\}})$	0,980	0,238	99,4%	87,4%	2	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-1} \times \mathbb{1}_{\{\text{SWI}_{t-1} > 1\}})$	0,672	0,053	84,7%	78,9%	6	0
$f(\text{Culture}, \text{Région}, \text{Année}, \text{SWI}_{t-2} \times \mathbb{1}_{\{\text{SWI}_{t-2} > 1\}})$	0,957	0,189	98,6%	86,8%	1	0

TABLE 13.9 – Spécifications de modèles testées pour le choix de **M3** avec les données mises en *as if*

Les résultats détaillés des régressions du modèle **M3** sont disponibles en Annexe F.4. La Table 13.10 suivante résume les impacts des différentes covariables sur les paramètres de la GPD.

	Impact sur ξ	Impact sur ν
Culture	De -25,28 à -24,78	Nul
Région	De -0,30 à +1,168	De -3,900 à -1,832
Année	+0,013 \times Année	Nul
SWI _t	-1,061 \times SWI _t	Nul

TABLE 13.10 – Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle **M3** - Données mises en *as if*

L'effet de l'indice d'humidité des sols sur le paramètre de forme est négatif. Ainsi, une diminution de ce dernier conduit à accroître la dangerosité de la loi des dépassements de seuil. Comme pour le modèle **M2**, l'année a un effet positif sur ξ . Sans tenir compte du SWI, la combinaison (Pois protéagineux, Île-de-France, 2000) donne le paramètre de forme minimal. En l'occurrence, ce dernier est toujours supérieur à -1 pour les données de la base agricole, ce qui n'affecte donc pas la validité de la reparamétrisation (ξ, β) en (ξ, ν) .

13.2.3 Projection de pertes de rendements futures

Sur les Figures 13.9 et 13.10 suivantes sont présentées les prédictions de probabilité de perte de 30%² par région en 2025 et 2050, par l'intermédiaire des modèles **M2** et **M3**, ce dernier étant couplé aux projections de l'indice d'humidité des sols du NGFS. Comme pour les projections historiques, les régions grisées sont celles pour lequel le calcul de la probabilité de retour n'était pas possible.

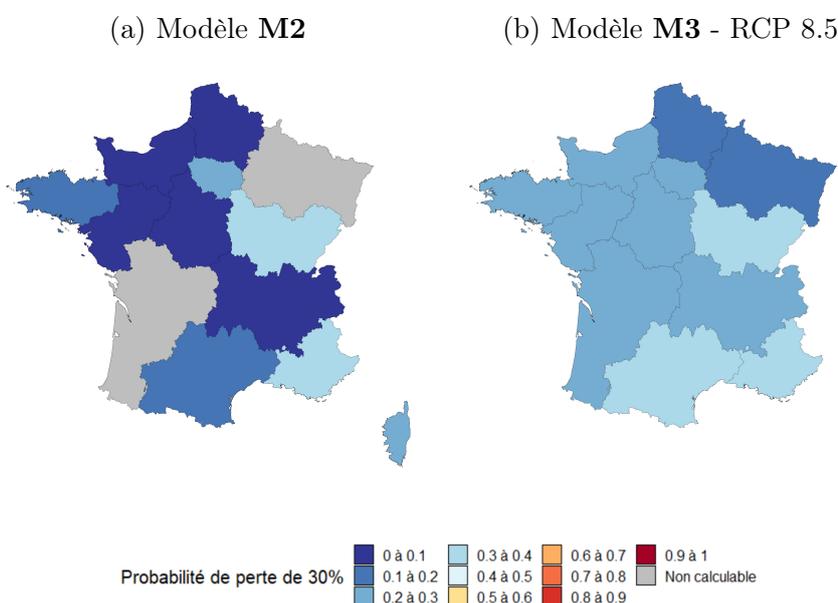


FIGURE 13.9 – Probabilité de perte à 30% projetée par région en 2025 selon les modèles **M2** et **M3** pour la culture de blé - Données mises en *as if*

Par rapport au résultat obtenu avec les données historiques, le modèle **M2** avec les données mises en *as if* conduit à l'obtention de probabilité de retour plus modérées. La probabilité de perte de rendements de 30% est toujours inférieure à 40% et atteint son maximum en Bourgogne-Franche-Comté avec une valeur de 0,37%.

Les prévisions du modèle **M3** sont également plus mesurées que les projections fournies à partir des données historiques. Par rapport au modèle **M2**, les probabilités de retour prédites sont toutefois légèrement plus élevées selon les régions considérées.

Peu de changements sont observés entre 2025 et 2050 pour le modèle **M2**, comme en témoigne la Figure 13.10 ci-après. En effet, à l'exception de l'Île-de-France, pour laquelle la probabilité de retour diminue, et des Hauts-de-France, pour laquelle cette dernière n'est plus calculable, la Figure 13.10 présente les mêmes ordres de grandeur que la carte de la Figure 13.9.

Là encore, les projections sont cohérentes avec les scénarios du GIEC, la Figure 13.10c donnant des probabilités de retour plus élevées que la Figure 13.10b, construite à partir du

2. conditionnelles à la survenance d'une perte supérieure à 14%

scénario RCP 2.6, plus optimiste. De plus, indépendamment du scénario retenu, le modèle **M2** conduit à l'obtention de probabilités de perte plus faibles.

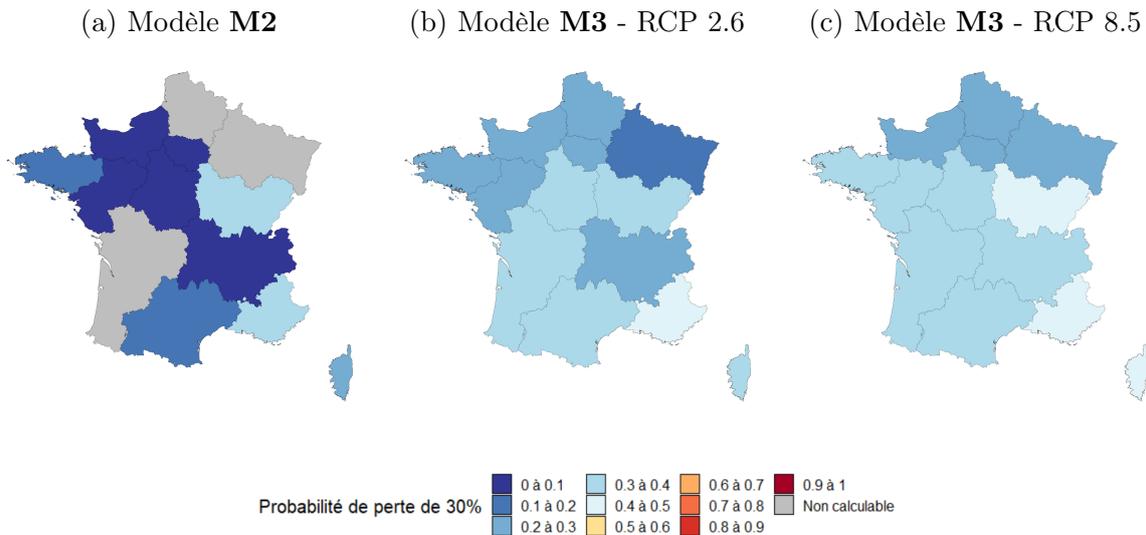


FIGURE 13.10 – Probabilité de perte à 30% projetée par région en 2050 selon les modèles **M2** et **M3** pour la culture de blé - Données mises en *as if*

A titre indicatif, le modèle **M1** fournit une probabilité de perte globale de 19%. Au vu des résultats présentés, les deux modèles semblent satisfaisants pour calibrer et prédire les paramètres de la GPD selon les covariables retenues.

13.3 Conclusions sur l'effet du climat sur les calibrages

Intégrer des variables climatiques aux calibrages des paramètres d'une loi de dépassement de seuil n'est pas aisé. De manière générale, le choix des covariables pour les deux paramètres reste assez complexe : si l'une des régressions peut être de très bonne qualité et avec une bonne significativité des coefficients, il peut être difficile de conserver de bons résultats sur la seconde.

Par ailleurs, tenir compte de plusieurs covariables peut tendre vers une individualisation du calibrage, conduisant donc à un surentraînement du modèle qui perdrait ses qualités de prédiction sur d'autres échantillons. En particulier, la structure de données initiale pourrait générer des corrélations entre deux covariables, par exemple si une espèce est cultivée dans une seule région uniquement.

En outre, la question de la méthode de calcul des pertes de rendements doit faire l'objet d'une analyse poussée. Si l'approche de mise en *as if* ne capte pas certaines pertes de rendements puisque la référence de rendements moyens est indépendante de l'année considérée, elle donne également des résultats plus cohérents que la méthode historique, notamment via le modèle **M2**. Cette dernière souffre également d'une réduction de la période d'analyse des pertes de rendements, due intrinsèquement à leur calcul.

Enfin, pour poursuivre cette analyse, une étude complète de la loi des dépassements de seuil intégrant une modélisation de la fréquence de survenance de ces événements pourrait être réalisée. Nous avons observé que la prise en compte de covariables peut significativement affecter le calibrage des paramètres de la loi de sévérité. Il serait intéressant de modéliser la loi de fréquence à partir de ces mêmes covariables afin de tester l'intérêt de la méthodologie vis-à-vis d'un calibrage classique. Cela permettrait également d'estimer des probabilités de retour absolues et non plus conditionnelles à la survenance de pertes extrêmes, comme présenté dans l'article de Chavez-Demoulin, Embrechts et Hofert [10].

Conclusion

Dans ce mémoire, le changement climatique a été pris en compte dans le provisionnement d'une branche agricole de deux manières : qualitative et quantitative.

Sur le plan qualitatif, l'étude s'est portée sur la comparaison des paramètres de forme de distribution des réserves selon le caractère climatique ou non de la sous-branche considérée. Pour cela, différentes méthodes de provisionnement ont été appliquées, notamment celles de Dal Moro qui propose une extension au modèle de Mack en intégrant des hypothèses sur le skewness et le kurtosis dans le triangle de liquidation. Il s'est avéré que le triangle relatif aux aléas climatiques présentait des paramètres de distribution d'ordre 3 et 4 bien différents des triangles des autres sous-branches. Bien que cette analyse ait été conduite sur une seule branche et pour un seul assureur, elle fournit néanmoins une méthodologie qualitative permettant d'observer le risque climatique dans la distribution des réserves.

Cette approche pourrait donc être étendue à d'autres branches émanant d'autres assureurs pour vérifier sa cohérence. En restant sur une analyse de portefeuilles agricoles, elle pourrait être appliquée en distinguant les triangles par type de culture assuré afin d'estimer si les effets climatiques impactent de la même manière le provisionnement.

Sur le plan quantitatif, l'étude s'est portée sur l'application d'une méthode de séparation climatique. Cette approche, initialement utilisée pour mesurer les effets inflationnistes, a été adaptée dans le but d'extraire une tendance calendaire liée au climat. Cette dernière a ensuite fait l'objet d'une modélisation à partir d'un indicateur d'humidité des sols, le *Soil Wetness Index*. Suite à cela, des projections ont été effectuées via différents scénarios climatiques, en vue d'observer l'évolution temporelle de la tendance extraite. Néanmoins, pour arriver à de tels résultats, des hypothèses simplificatrices ont été nécessaires afin de pallier à l'absence de données.

Dès lors, puisque ce mémoire se place en tant que « boîte à outils », il serait pertinent de relancer cette démarche avec des données plus complètes, fournissant par exemple la localisation géographique du portefeuille, le triangle de nombres de sinistres et l'événement climatique cause du sinistre. Cela permettrait ainsi d'une part d'assurer la stabilité de la méthode, mais également de complexifier le modèle de lien entre variables météorologiques et facteur d'inflation « climatique » en vue d'acquérir une plus grande précision dans les estimations et les projections. En particulier, cette méthode pourrait avoir de l'intérêt sur des branches à développement long, telles que la sécheresse en assurance habitation, pour lesquelles les montants de provisions à constituer sont difficilement estimables par les méthodes classiques type *chain-ladder*.

En complément, ce mémoire propose une méthodologie visant à intégrer des variables climatiques dans l'estimation de probabilités de retour de pertes de rendements agricoles. Cela pourrait être pertinent pour aiguiller les actuaires qui cherchent à affiner leur provisionnement et/ou tarification.

Ainsi, des lois de sévérité de dépassements de seuil sur les pertes de rendements ont été modélisées, en calibrant les paramètres en fonction d'effets fixes uniquement (culture, année, région) ou en ajoutant également un indicateur climatique. Les modèles calibrés à partir de l'indice d'humidité des sols conduisent à l'obtention de probabilités de retour de pertes, conditionnelles à la survenance d'événement extrême, assez proches de celles résultant d'un calibrage classique. Cependant, cette méthode permet d'affiner les résultats par région et type de culture, en intégrant des spécifications météorologiques propre à chaque localité.

De manière générale, le choix des covariables pour les deux paramètres reste toutefois assez complexe : si l'une des régressions peut être de très bonne qualité et avec une bonne significativité des coefficients, il peut être difficile de conserver de bons résultats sur la seconde. Par ailleurs, tenir compte de plusieurs covariables peut tendre vers une individualisation du calibrage, conduisant donc à un surentraînement du modèle qui perdrait ses qualités de prédiction sur d'autres échantillons. L'approche présentée dans le mémoire pourrait donc certes être facilement étendue sur d'autres risques et avec d'autres indicateurs climatiques, mais ces limites opérationnelles resteraient toujours présentes. Néanmoins, l'étude pourrait être complétée par une modélisation de la loi de fréquence des dépassements de seuil en tenant compte de covariables. Dès lors, les probabilités de retour d'événements extrêmes pourraient être estimées de manière absolue à partir du modèle fréquence - sévérité.

In fine, de par son caractère émergent et sa complexité, le risque climatique devra faire l'objet de davantage de travaux de recherche, que ce soit vis-à-vis de la problématique de provisionnement ou pour d'autres thématiques actuarielles.

Bibliographie

- [1] Agriculteurs-85. *Agronomie : conséquences de la grêle sur les cultures*. <https://www.agriculteurs-85.fr/25-06-2013/240/>. Consulté le 01/08/23.
- [2] M. AMOUSSA. « Méthodes de conversion d'un montant d'ajustement pour risque en quantile IFRS 17 ». Mém. de mast. 2021.
- [3] Autorité de Contrôle Prudentiel et de Résolution. *La gouvernance des risques liés au changement climatique dans le secteur de l'assurance*. 2022.
- [4] Autorité de Contrôle Prudentiel et de Résolution. *Les assureurs français face au risque de changement climatique*. 2019.
- [5] E. S. BANJANOVIC et J. W. OSBORNE. « Confidence Intervals for Effect Sizes : Applying Bootstrap Resampling ». In : *Practical Assessment, Research, and Evaluation : Vol. 21, Article 5* (2016). DOI : 10.7275/dz3r-8n08.
- [6] N. BARADEL. *Cours d'assurance dommage*. Juin 2021.
- [7] S. BJÖRKWALL, O. HÖSSJER et E. OHLSSON. « Bootstrapping the Separation Method in Claims Reserving ». In : *ASTIN Bulletin : The Journal of the IAA* 40.2 (2010), p. 845-869. DOI : 10.2143/AST.40.2.2061138.
- [8] Caisse Centrale de Réassurance. *Note de synthèse sur le suivi de l'aide à l'assurance multirisque climatique pour l'exercice 2019*.
- [9] A. CHARPENTIER et C. DUTANG. *L'actuariat avec R*. Déc. 2012.
- [10] V. CHAVEZ-DEMOULIN, P. EMBRECHTS et M. HOFERT. « An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates ». In : *The Journal of Risk and Insurance* 83.3 (2016), p. 735-776. ISSN : 00224367, 15396975. URL : <http://www.jstor.org/stable/43998282> (visité le 01/08/2023).
- [11] S. CHRISTOFIDES. « Regression models based on log-incremental payments ». In : *Claim Reserving Manual* (1989).
- [12] Covéa. *Changement climatique et assurance : Quelles conséquences sur la sinistralité à horizon 2050 ?* Jan. 2022.
- [13] E. DAL MORO. « A Closed-Form Formula for the Skewness Estimation of Non-Life Reserve Risk Distribution ». In : *UNSW Business School Research Paper Series* (2013).
- [14] E. DAL MORO. « Application of Skewness to Non-Life Reserving ». In : *Microeconomics : Decision-Making under Risk & Uncertainty eJournal* (2012).
- [15] E. DAL MORO. « Kurtosis and Skewness Estimation for Non-Life Reserve Risk Distribution ». In : *Risk Management & Analysis in Financial Institutions eJournal* (2013).
- [16] A. DAVISON. *Bootstrap Methods and their Application*. 2006.
- [17] C. ECHCHIHAB. « 2016, la pire récolte de blé en France depuis 40 ans ». In : *Les Echos* (juin 2016).

- [18] J.-D. FERMANIAN. *Copulas and financial applications*. ENSAE Paris. 2022.
- [19] A. I. FLEISHMAN. « A method for simulating non-normal distributions ». In : *Psychometrika* 43 (déc. 1978), p. 521-532. DOI : 10.1007/BF02293811.
- [20] France Assureurs. *Impact du changement climatique sur l'assurance à l'horizon 2050*.
- [21] France Assureurs. *L'assurance agricole en 2020*. 2021.
- [22] France Assureurs. *Le risque sécheresse et son impact sur les habitations*. 2022.
- [23] M. FRANQUESA. « Les types de gel dans l'agriculture ». In : Agroptima Blog (août 2019).
- [24] Groupe d'experts Intergouvernemental sur l'Evolution du Climat. *Climate Change 2021 : The Physical Science Basis*. 2021.
- [25] Groupe d'experts Intergouvernemental sur l'Evolution du Climat. *Rapport spécial du GIEC sur les conséquences d'un réchauffement planétaire de 1,5°C*. 2019.
- [26] D. HAMBLING. « How climate change is leading to bigger hailstones ». In : *BBC* (mars 2022).
- [27] M. HIRSCHY et al. *Gel et grêle en viticulture et arboriculture - Etat des lieux des dispositifs de protection contre les aléas climatiques*. Rapport technique. ACTA - Association de Coordination Technique Agricole, juin 2020.
- [28] Institut français de la vigne et du vin. *Aléas climatiques de la vigne - La grêle et le gel de printemps : comment s'en protéger ?* 2019.
- [29] S. JAZIRI. « Méthodes de provisionnement non-vie et risque de réserve à un an ». Mém. de mast. 2011.
- [30] Keraunos. *La grêle en France*. <https://www.keraunos.org/climatologie/les-orages-en-france/grele-climatologie-france-records-saisonnalite-risque-degats>. Consulté le 23/06/23.
- [31] E. KREMER. « IBNR-claims and the two-way model of ANOVA ». In : *Scandinavian Actuarial Journal* 1982.1 (1982), p. 47-55. DOI : 10.1080/03461238.1982.10405432.
- [32] « La sécheresse menace les récoltes de céréales en France ». In : L'Info durable (mai 2022).
- [33] La terre de chez nous. *Vers une hausse des primes de 10 à 25% en 2022*. http://laterredecheznous.com/news/archivestory.php/aid/7985/Vers_une_hausse_des_primes_de_10__E0_25__25_en_2022_.html. Consulté le 27/01/2023.
- [34] V. LEPOUSEZ et M. ABOUKRAT. « Les scénarios SSP ». In : *Carbone4* (juill. 2022).
- [35] T. MACK. « A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves ». In : *ASTIN Bulletin* 21.1 (1991), p. 93-109. DOI : 10.2143/AST.21.1.2005403.
- [36] T. MACK. « Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates ». In : *ASTIN Bulletin* 23.2 (1993), p. 213-225. DOI : 10.2143/AST.23.2.2005092.
- [37] Météo France. *Climadiag Commune*. <https://meteofrance.com/climadiag-commune>. Consulté le 02/08/23.
- [38] Météo France. *La grêle*. <https://meteofrance.com/comprendre-la-meteo/orages/la-grele>. Consulté le 24/08/2023.
- [39] Météo France. *Les gelées seront-elles de plus en plus rares dans le futur ?* <https://meteofrance.com/actualites-et-dossiers/magazine/les-gelees-seront-elles-de-plus-en-plus-rares-dans-le-futur>. Consulté le 02/08/23. Avr. 2023.

- [40] Météo France. *Ressource en eau, sécheresses et changement climatique*. Mars 2023.
- [41] Ministère de l'agriculture et de l'alimentation. *Campagne 2021 - L'assurance multirisque climatique des récoltes*. 2021.
- [42] Ministère de la Transition Ecologique et de la Cohésion des territoires. « Scénarios et projections climatiques ». In : *Chiffres clés du climat France, Europe et Monde* (2021).
- [43] Ministère de la Transition Ecologique et de la Cohésion des territoires et Ministère de la Transition énergétique. *Comprendre le GIEC*. <https://www.ecologie.gouv.fr/comprendre-giec>. Consulté le 13/02/2023. 2022.
- [44] Network for Greening the Financial System. *NGFS Scenarios Portal*. <https://www.ngfs.net/ngfs-scenarios-portal/>. Consulté le 02/01/2023.
- [45] Network for Greening the Financial System. *Origin and Purpose*. <https://www.ngfs.net/en/about-us/governance/origin-and-purpose>. Consulté le 05/01/2023.
- [46] Optimind. *Réforme de l'assurance récolte*. Nov. 2022.
- [47] Oxfam France. *Les événements climatiques extrêmes : quand la planète s'emballe*. Consulté le 16/01/2022. URL : <https://www.oxfamfrance.org/climat-et-energie/les-evenements-climatiques-extremes-quand-la-planete-semballe/#:~:text=Les%5C%20hausses%5C%20de%5C%20temp%5C%3%5C%A9ratures%5C%20et,sur%5C%20la%5C%20production%5C%20agricole%5C%20fran%5C%3%5C%A7aise>.
- [48] Plein Champ. *Comment réagissent les céréales face au gel ?* <https://www.pleinchamp.com/actualite/comment-reagissent-les-cereales-face-au-gel>. Consulté le 24/08/2023.
- [49] Plein Champ. *Gestion des risques climatiques : le nouveau dispositif en 15 questions*. <https://www.pleinchamp.com/actualite/gestion-des-risques-climatiques-le-nouveau-dispositif-en-15-questions>. Consulté le 12/01/2023.
- [50] A. E. RENSHAW et R. J. VERRALL. « A stochastic model underlying the chain-ladder technique ». In : *British Actuarial Journal* 4.4 (1998), p. 903-923. ISSN : 13573217, 20440456. (Visité le 10/11/2022).
- [51] C. Y. ROBERT. *Extreme Value Theory*. ENSAE Paris. 2023.
- [52] M. ROTHWELL et al. « Practical guide to climate change for general insurance practitioners ». In : *Institute and Faculty of Actuaries* (sept. 2019).
- [53] Scribbr. *Skewness | Definition, Examples & Formula*. <https://www.scribbr.com/statistics/skewness/>. Consulté le 07/11/2022.
- [54] Scribbr. *What Is Kurtosis ? | Definition, Examples & Formula*. <https://www.scribbr.com/statistics/kurtosis/>. Consulté le 07/11/2022.
- [55] G. C. TAYLOR. « Separation of Inflation and other Effects from the Distribution of Non-Life Insurance Claim Delays ». In : *ASTIN Bulletin : The Journal of the IAA* 9.1-2 (1977), p. 219-230. DOI : 10.1017/S0515036100011533.
- [56] G. TAYLOR et F. ASHE. « Second moments of estimates of outstanding claims ». In : *Journal of Econometrics* 23.1 (1983), p. 37-61. ISSN : 0304-4076. DOI : [https://doi.org/10.1016/0304-4076\(83\)90074-X](https://doi.org/10.1016/0304-4076(83)90074-X).
- [57] TerrOïko. *Suivi de l'aide à l'assurance multirisque climatique des récoltes - Période 2010-2015*.
- [58] H. G. VERBEEK. « An approach to the analysis of claims experience in motor liability excess of loss reinsurance ». In : *ASTIN Bulletin : The Journal of the IAA* 6.3 (1972), p. 195-202. DOI : 10.1017/S0515036100010989.

-
- [59] VIVESCIA. *Céréale : quel est le cycle du blé?* <https://www.vivescia.com/grand-angle/tous/cereale-quel-est-le-cycle-du-ble>. Consulté le 01/08/23.
- [60] F. E. D. VYLDER. *Estimation of IBNR claims by least squares*. Jan. 1978.
- [61] Wikipédia. *Asymétrie (statistiques)*. [https://fr.wikipedia.org/wiki/Asym%C3%A9trie_\(statistiques\)](https://fr.wikipedia.org/wiki/Asym%C3%A9trie_(statistiques)). Consulté le 07/11/2022.
- [62] Wikipédia. *Kurtosis*. <https://fr.wikipedia.org/wiki/Kurtosis>. Consulté le 07/11/2022.
- [63] M. V. WÜTHRICH et M. MERZ. « Modelling the Claims Development Result for Solvency Purposes ». In : *Casualty Actuarial Society E-Forum* (jan. 2008).
- [64] M. V. WÜTHRICH, M. MERZ et N. LYSENKO. « Uncertainty of the claims development result in the chain ladder method ». In : *Scandinavian Actuarial Journal* 2009.1 (2009), p. 63-84. DOI : 10.1080/03461230801979732.

Table des figures

1.1	Projection des émissions liées aux énergies fossiles selon les scénarios RCP, en gigatonnes de carbone [42]	5
1.2	Variations de la température de surface moyenne mondiale selon les scénarios RCP, relativement à la période 1986-2005 [42]	6
1.3	Variations de la température de surface moyenne mondiale selon les scénarios SSP, relativement à la période 1850-1900 [24]	7
1.4	Présentation des classes de scénarios du NGFS [44]	8
1.5	Part de sociétés d'assurance interrogées exposée à chaque type de risque [4] . . .	9
1.6	Projection des coûts cumulés sécheresse entre 2020 et 2050 [20]	10
1.7	Projection des coûts cumulés inondations entre 2020 et 2050 [20]	11
2.1	Evolution du nombre de contrats souscrits, des surfaces et des capitaux assurés, ainsi que des cotisations subventionnables entre 2010 et 2015 [57]	15
2.2	Taux de diffusion par catégorie de culture - Moyenne 2010-2015 [57]	16
2.3	Taux de diffusion de l'assurance multirisque climatique - Campagne de 2019 [8] .	17
3.1	Evolution de l'état des sols au fil des mois en 2022 [22]	22
3.2	Evolution de l'indicateur d'humidité des sols sous scénario RCP 8.5 sur la période 2035-2065 par rapport à la référence 1970-2005 [12]	23
3.3	Cycle de vie du blé [59]	24
3.4	Fréquence des chutes de grêle en France [30]	25
3.5	Evolution du nombre d'occurrences annuel d'orages de grêle sur la période historique et à horizon 2050 [12]	26
3.6	Nombre de jours de grêle annuel moyen sur la période 1976-2005 [39]	27
3.7	Projections du nombre de jours de gel à horizon 2050 [37]	28
3.8	Prévisions de nombre de jours de grêle à l'horizon 2071-2100 - Scénario RCP 4.5 [39]	28
4.1	Etapes de la vie d'un sinistre	32
4.2	Schéma du processus de bootstrap [5]	45
5.1	Asymétries possibles [53]	48
5.2	Aplatissements possibles ³ [54]	49
5.3	Schéma de l'approche simulateur présentée par Dal Moro [14]	56
6.1	Représentation graphique du triangle de Mack	62
6.2	Vérification de l' Hypothèse 1	62
6.3	Vérification de l' Hypothèse 2	63
6.4	Vérification de l'indépendance des résidus dans l' Hypothèse 3	64
6.5	Vérification de l' Hypothèse 4	64
6.6	Vérification de l' Hypothèse 6	65

6.7	Estimations du risque de réserve $\frac{\sqrt{MSE(\hat{R}_i)}}{\hat{R}_i^M}$ par année de survenance pour chaque modèle	66
6.8	Erreurs commises par les modèles tenant compte des hypothèses de Dal Moro . .	67
6.9	Distribution des réserves pour le triangle de Mack	70
7.1	Processus de prise en compte de l'inflation dans le modèle de Mack <i>chain-ladder</i>	72
8.1	Evolution du montant des primes acquises	78
8.2	Evolution des règlements nets par année de développement	80
8.3	Evolution du ratio sinistres sur primes	80
8.4	Vérification de l' Hypothèse 1	81
8.5	Vérification de l' Hypothèse 2	82
8.6	Vérification de l' Hypothèse 3	82
8.7	Tests de stabilité des résidus	84
8.8	Vérification de l' Hypothèse 4	85
8.9	Vérification de l' Hypothèse 6	85
8.10	Evolution de l'Indice des Prix des Produits Agricoles à la Production	86
8.11	Tests de stabilité des résidus pour le triangle déflaté	87
9.1	<i>Best Estimates</i> obtenus	89
9.2	Estimations du risque de réserve en pourcentage du montant des réserves obtenus par Mack <i>chain-ladder</i>	90
9.3	Résultats de skewness obtenus	91
9.4	Résultats de kurtosis obtenus	92
9.5	Distributions des réserves obtenues par simulation GPD - 1000 simulations . . .	93
10.1	Evolution du facteur d'inflation « climatique » λ_i	97
10.2	Evolution de l'indice moyen d'humidité des sols entre 2000 et 2020 pour le mois de mai	100
10.3	Evolution de l'indice d'humidité des sols selon divers scénarios	101
10.4	Prédictions du modèle linéaire	103
10.5	Evolution de la tendance climatique prédite selon divers scénarios	104
11.1	Densité des distributions de Fréchet, Weibull et Gumbel	110
11.2	Echantillon de dépassements de seuil [51]	113
11.3	Lien entre <i>mean excess function</i> et distributions de probabilité [51]	116
11.4	<i>Mean excess function</i> de trois types de lois de Pareto généralisée	118
12.1	Cultures développées de niveaux 1 et 2	122
12.2	Etat des lieux de la production agricole en 2020 pour les cultures considérées . .	122
12.3	Production agricole régionale en 2020 (en tonnes) par culture considérée	123
12.4	Rendements moyens par région - Toutes cultures considérées	124
12.5	Evolution des rendements de blé sur la période 2000-2020 pour les cinq régions les plus productrices	124
12.6	Nombre de pertes de rendements supérieures à 15% constatées entre 2000 et 2020	125
12.7	Nombre de pertes de rendements supérieures à 15% constatées par culture . . .	126
12.8	Nombre de pertes de rendements supérieures à 15% constatées par région	126
12.9	Indice d'humidité des sols moyen sur la période 2000-2020	128
12.10	Indice d'humidité des sols moyen constaté par région en 2003 et 2016	128
12.11	Projections de l'indice d'humidité des sols selon les scénarios RCP 2.6 et 8.5 du GIEC	130

13.1	<i>Mean excess plot</i> relatif aux données historiques	132
13.2	Graphiques de stabilité des paramètres de la loi GPD pour les données historiques	132
13.3	Détails du modèle M1 calibré sur les données historiques	133
13.4	Probabilité de perte à 30% projetée par région en 2025 selon les modèles M2 et M3 pour la culture de blé - Données historiques	137
13.5	Probabilité de perte à 30% projetée par région en 2050 selon les modèles M2 et M3 pour la culture de blé - Données historiques	138
13.6	<i>Mean excess plot</i> pour les données mises en <i>as if</i>	139
13.7	Graphiques de stabilité des paramètres de la loi GPD pour les données mises en <i>as if</i>	139
13.8	Détails du modèle M1 calibré sur les données mises <i>as if</i>	140
13.9	Probabilité de perte à 30% projetée par région en 2025 selon les modèles M2 et M3 pour la culture de blé - Données mises en <i>as if</i>	144
13.10	Probabilité de perte à 30% projetée par région en 2050 selon les modèles M2 et M3 pour la culture de blé - Données mises en <i>as if</i>	145

Liste des tableaux

1.1	Récapitulatif des différents scénarios	9
2.1	Evolution des primes, des sinistres et du rapport sinistres/primes pour les contrats MRC et grêle [21]	18
2.2	Objectifs de la réforme de l'assurance récoltes en taux de diffusion [46]	18
4.1	Exemple de triangle <i>run-off</i> en paiements cumulés	33
4.2	Triangles de provisionnement aux temps I (à gauche) et $I + 1$ (à droite) [64]	37
6.1	Exemple de triangle de liquidation [36]	61
6.2	Montants des réserves estimés	66
6.3	Pourcentage d'erreur de prédiction $\frac{\sqrt{MSE(\hat{R}_i)}}{\hat{R}_i^M}$ des différents modèles	68
6.4	Résultats d'estimation de skewness pour le triangle de Mack	69
6.5	Résultats d'estimation de kurtosis pour le triangle de Mack	69
8.1	Tests de Box-Pierce sur le triangle aléas climatiques	84
8.2	Tests de Box-Pierce sur le triangle aléas climatiques déflaté	88
10.1	Montant de réserves obtenu par chaque année de survenance et au total pour chaque hypothèse et écarts avec les estimations de Mack	96
10.2	Comparaison des proportions de sinistres obtenues pour les cinq premières année de développement	97
10.3	Corrélation de la tendance climatique avec la variable retenue pour les triangles déflaté et non-déflaté	102
10.4	Résultats du modèle de régression linéaire retenu	103
10.5	Prévisions de la tendance climatiques pour quelques années	105
11.1	Exemples de QQ-plots [51]	115
12.1	Cultures considérées pour l'étude	122
12.2	Pourcentage de données manquantes par période considérée	127
13.1	Paramètres du modèle M1 - Données historiques	133
13.2	Spécifications de modèles testées pour le choix de M2 pour les données historiques	134
13.3	Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle M2 - Données historiques	135
13.4	Spécifications de modèles testées pour le choix de M3 pour les données historiques	136
13.5	Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle M3 - Données historiques	137
13.6	Paramètres du modèle M1 - Données mises en <i>as if</i>	140
13.7	Spécifications de modèles testées pour le choix de M2 avec les données mises en <i>as if</i>	141

13.8	Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle M2	
	- Données mises en <i>as if</i>	142
13.9	Spécifications de modèles testées pour le choix de M3 avec les données mises en	
	<i>as if</i>	143
13.10	Impacts des covariables sélectionnées sur les paramètres de la GPD - Modèle	
	M3 - Données mises en <i>as if</i>	143

Glossaire

Acronyme	Signification	Acronyme	Signification
ACPR	Autorité de Contrôle Prudentiel et de Résolution	IBNR	<i>Incurring But Not Reported</i>
CDR	<i>Claim Development Result</i>	INSEE	Institut National de la Statistique et des Etudes Economiques
DOM	Département d’Outre-Mer	IPPAP	Indice des Prix des Produits Agricoles à la Production
EIOPA	<i>European Insurance and Occupational Pensions Authority</i>	MRC	Assurance MultiRisque Climatique
ESG	Environnementaux, Sociaux et de Gouvernance	MSE	<i>Mean Square Error</i>
FEADER	Fonds Européen Agricole pour le Développement Rural	NGFS	<i>Network for Greening the Financial System</i>
FNGRA	Fonds National de Gestion des Risques en Agriculture	ORSA	<i>Own Risk and Solvency Assessment</i>
GEV	<i>Generalized Extreme Value</i>	PAC	Politique Agricole Commune
GIEC	Groupe d’experts Intergouvernemental sur l’Evolution du Climat	RCP	<i>Representative Concentration Pathways</i>
GLM	<i>Generalized Linear Model</i>	SAA	Statistique Agricole Annuelle
GPD	<i>Generalized Pareto Distribution</i>	SIG	Système d’Information Géographique
IBNe/yR	<i>Incurring But Not enough/yet Reported</i>	SSP	<i>Shared Socio-economic Pathways</i>

TABLE – Liste des acronymes utilisés

Note de synthèse

Introduction

En raison du dérèglement climatique, les événements naturels, tels que la sécheresse ou la grêle, verront leur fréquence et leur intensité s'accroître dans les années à venir. Dans l'agriculture, ces changements climatiques affectent de manière conséquente les rendements des exploitations. En effet, les hausses de température et les canicules favorisent notamment la sécheresse des terres agricoles, ce qui détériore leur fertilité et donc la qualité des cultures et leur prix de vente.

Par conséquent, la prise en compte du risque climatique est un enjeu majeur pour les acteurs du secteur de l'assurance, notamment dans le cadre de l'évaluation des réserves. Ce mémoire s'inscrit donc dans une optique de constitution d'une « boîte à outils » apportant une démarche d'évaluation de ce risque émergent dans les réserves d'une compagnie d'assurance Non-Vie. Il présente plusieurs méthodes répondant à diverses problématiques :

- une **analyse qualitative du risque climatique** dans le provisionnement, via une étude détaillée de la forme de la distribution des réserves ;
- une **analyse quantitative du risque climatique**, visant à extraire une tendance inflationniste des réserves liée à des variables météorologiques et à la projeter selon des scénarios climatiques ;
- et une **modélisation des pertes de rendements extrêmes** en fonction de variables climatiques.

Fortement exposé aux risques climatiques, le secteur agricole constitue un bon exemple d'application de notre démarche.

Données de l'étude

Dans l'ensemble du mémoire, la variable météorologique utilisée est un indicateur d'humidité des sols, le *Soil Wetness Index*. Ces données sont issues de deux bases, provenant de Météo France¹ pour les valeurs historiques et du *Climate Impact Explorer*² pour les données futures suivant des scénarios climatiques.

Par ailleurs, les triangles de liquidation requis pour répondre aux deux premières problématiques émanent de données clients, anonymisées avant étude. Trois sous-branches sont disponibles : **aléas climatiques**, **hors aléas climatiques** et **tempête** et **responsabilité civile**. Ces triangles sont fournis pour la période 2002-2019.

Enfin, les données agricoles de production et de rendements nécessaires pour la dernière partie de l'étude sont issues de la Statistique Agricole Annuelle (SAA) pour les cultures déve-

1. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture>

2. <https://climate-impact-explorer.climateanalytics.org/>

loppées de l'organisme Agreste³, le service statistique ministériel de l'agriculture. La période d'étude est 2000-2020 et les indicateurs sont recueillis à la maille régionale pour la France métropolitaine. L'analyse s'articule autour de plusieurs cultures, incluses dans la catégorie « Céréales, oléagineux, protéagineux ».

Analyse qualitative du risque climatique

Modèle de provisionnement de Dal Moro

Dans son article [13], Dal Moro propose une extension du modèle de Mack *chain-ladder* [36] en le complétant avec deux hypothèses sur le moment d'ordre 3, nommé *skewness*, des paiements cumulés conditionnel aux paiements passés :

1. Le skewness du paiement cumulé $C_{i,j+1}$ conditionnel aux paiements passés ne dépend pas de l'année de survenance i mais de l'année de développement j .
2. Il existe des coefficients $Sk_j > 0$ pour $1 \leq j \leq I - 2$ tels que

$$\mathbb{E} [(C_{i,j+1} - \mathbb{E}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}])^3 | C_{i,1}, \dots, C_{i,j}] = Sk_j^3 C_{i,j}^{3/2}.$$

A partir de ces deux hypothèses, l'auteur propose deux méthodologies pour calculer le skewness de la distribution des réserves :

- une **approche par formule fermée**, dont les résultats sont donnés par la fonction quantile du package `ChainLadder` de R ;
- et une **approche simulatoire** à partir d'une loi de Pareto Généralisée (GPD). Cette dernière propose de générer les paiements cumulés du triangle inférieur de proche en proche, en calibrant les paramètres de la loi GPD à partir des hypothèses de Mack et Dal Moro via une méthode des moments.

Résultats obtenus

Une vérification des hypothèses de Mack et Dal Moro a conduit à la déflation du triangle aléas climatiques à partir de l'indice des Prix des Produits Agricoles à la Production, l'IPPAP, fourni par l'Agreste⁴. Ce traitement n'a pas été nécessaire pour les deux autres triangles. L'approche par simulation GPD a ensuite été lancée sur les triangles clients. La Table 1 récapitule les résultats obtenus pour le skewness et le kurtosis. Pour la simulation GPD, les valeurs de skewness et de kurtosis ont été obtenues de manière empirique à partir des distributions finales, schématisées en Figure 1.

Triangle / Méthode	SKEWNESS		KURTOSIS
	Formule fermée	Simulation GPD	Simulation GPD
Aléas climatiques déflaté	1,32	1,36	4,91
Hors aléas climatiques et tempête	0,72	0,78	3,29
Responsabilité civile	0,76	0,92	4,40

TABLE 1 – Résultats de skewness et de kurtosis des distributions des réserves

3. https://agreste.agriculture.gouv.fr/agreste-saiku/?plugin=true&query=query/open/SAANR_DEVELOPPE_2#query/open/SAANR_DEVELOPPE_2

4. https://agreste.agriculture.gouv.fr/agreste-web/disaron/D_0037/detail/

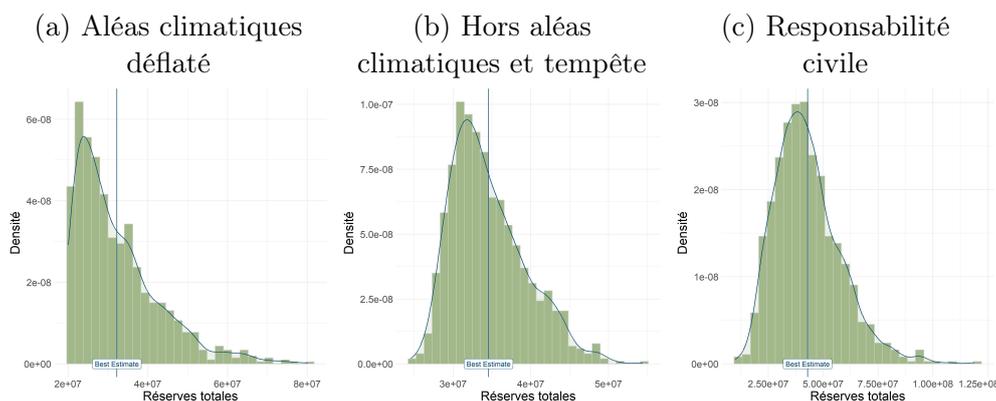


FIGURE 1 – Distributions des réserves obtenues par simulation GPD - 1000 simulations

Il ressort de la Table 1 et de la Figure 1 que la sous-branche aléas climatiques présente une asymétrie (donnée par le skewness) et un aplatissement (donné par le kurtosis) bien plus élevés que les deux autres sous-branches. Cela pourrait être un effet potentiel de la présence du risque climatique dans cette sous-branche. La validation de cette conjecture nécessiterait l'étude de davantage de triangles émanant d'autres assureurs. En restant sur le secteur agricole, l'analyse de la forme des distributions des réserves par type de culture pourrait également être envisagée afin de vérifier si le risque climatique affecte pareillement leurs provisionnements.

Analyse quantitative du risque climatique

Adaptation de la méthode de séparation de Verbeek

La littérature actuarielle fournit plusieurs méthodologies afin de tenir compte d'effets calendaires type inflation dans le provisionnement. L'une de ces approches, la méthode de séparation introduite par Verbeek [58] et développée par Taylor [55] a été adaptée dans ce mémoire pour extraire des réserves non pas une tendance inflationniste mais une tendance climatique. La Figure 2 présente l'hypothèse principale sous-jacente à chacun des modèles.

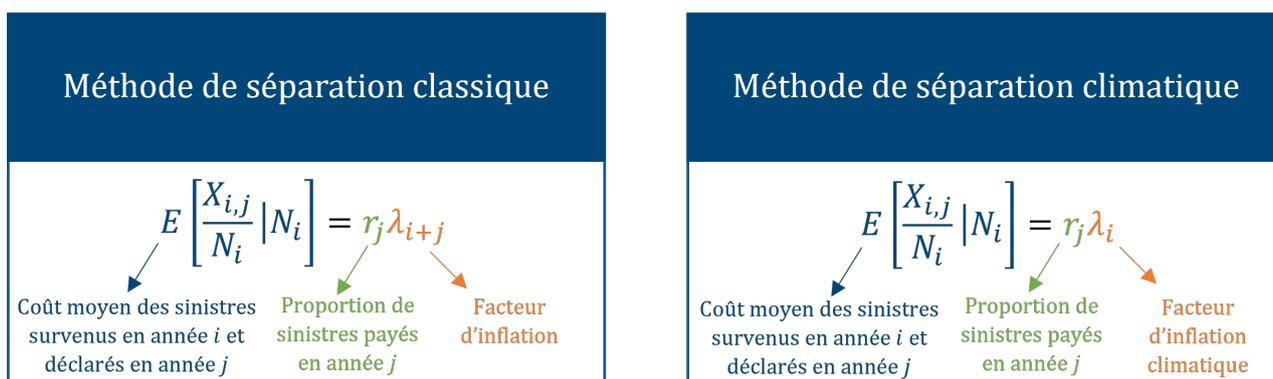


FIGURE 2 – Adaptation de l'hypothèse principale de la méthode de séparation de Verbeek [58]

Pour appliquer cette méthode au triangle aléas climatique, deux étapes préliminaires se sont succédées :

1. Déflation du triangle permettant d'ôter les effets parasites liés à l'inflation.
2. En raison d'une absence de données, simulation d'un triangle de nombre de sinistres, selon deux hypothèses :

(H1) Le nombre de sinistres moyen **dépend de la sinistralité annuelle** :

$$\mathbb{E}[N_{i,j}] = r_j \times SP_i \times \alpha M$$

avec SP_i ratio sinistres sur primes de l'année de survenance i .

(H2) Le nombre de sinistres moyen est **indépendant de la sinistralité annuelle** :

$$\mathbb{E}[N_{i,j}] = r_j \times \alpha M.$$

Dans chacune d'entre elles, les nombres de sinistres $N_{i,j}$ sont ensuite simulés à partir de lois de Poisson de moyenne $\mathbb{E}[N_{i,j}]$. α représente une proportion de sinistres constatés parmi un nombre de contrats M fixé car non-fourni dans les données. r_j , proportion de sinistres payés par année de développement a quant à elle été fixée en respectant le caractère court du développement des sinistres dans le portefeuille.

Suite à ces deux étapes, le triangle de nombres de sinistres a été complété par le modèle de Mack *chain-ladder* et la méthode de séparation climatique a pu être appliquée au triangle de paiements.

Résultats obtenus

Les montants de réserves estimés ainsi que la tendance climatique extraite grâce à la méthode de séparation sont représentés sur la Figure 3 suivante.

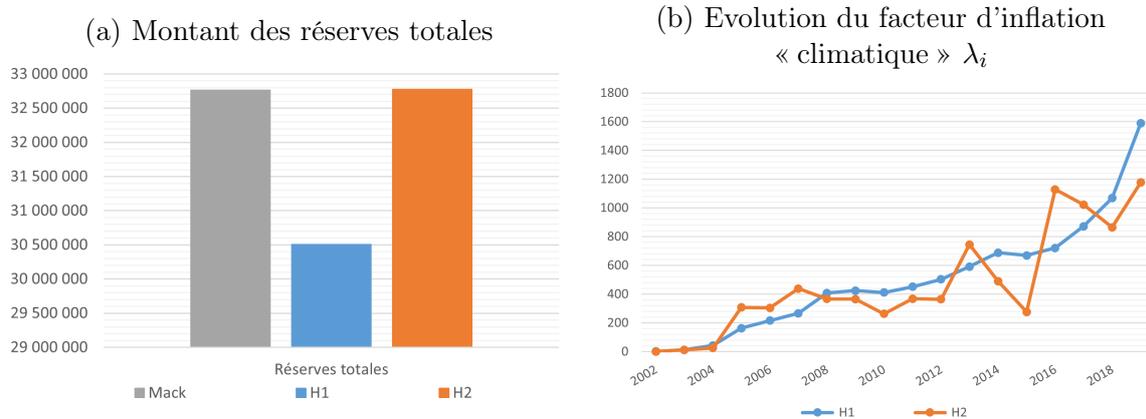


FIGURE 3 – Résultats de la méthode de séparation climatique

L'hypothèse (H2) semble plus adaptée aux données : elle fournit un *Best Estimate* proche de celui calculé avec le modèle de Mack *chain-ladder*. De plus, le facteur d'inflation « climatique », bien qu'observant une tendance croissante, suit également les variations du ratio sinistres sur primes. Dès lors, il peut retranscrire plus fidèlement le comportement d'une variable météorologique que le ferait le facteur extrait avec (H1).

Une phase de recherche du meilleur modèle linéaire afin de relier la tendance climatique aux données d'humidité des sols s'est ensuite opérée. La spécification retenue est la suivante :

$$\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma \times \text{SWI}_{i-2}.$$

Les résultats de la régression linéaire sont présentés sur la Table 2. Une diminution de l'indice d'humidité des sols conduit à un accroissement de la tendance climatique. En outre, la croissance

du facteur d'inflation « climatique » en fonction de l'année de survenance est bien captée dans ce modèle.

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,125e+05	1,660e+04	-6,776	6,25e-06
Année _{<i>i</i>}	56,55	8,227	6,873	5,29e-06
SWI _{<i>i-2</i>}	-836,9	426,4	-1,963	0,0685

R² ajusté : 0,7595

TABLE 2 – Résultats du modèle de régression linéaire retenu

Enfin, le modèle de lien entre facteur d'inflation « climatique » et SWI a été projeté à partir de scénarios d'évolution émanant entre autres du GIEC.

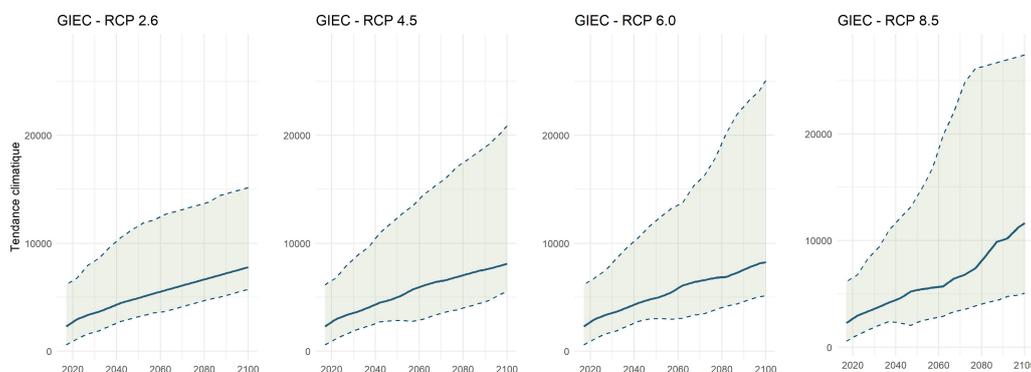


FIGURE 4 – Evolution de la tendance climatique selon les scénarios du GIEC

Sur la Figure 4, tous les scénarios conduisent à une forte augmentation de la tendance climatique dans les années à venir. Bien que le modèle retenu ait peu de sens physique, en considérant un indicateur retardé de deux ans, il permet toutefois d'établir un ordre de grandeur permettant de guider le provisionnement de branches agricoles. Ces premiers résultats sont ainsi prometteurs et démontrent la pertinence de la méthodologie appliquée, bien que cette dernière puisse être améliorée, par exemple en complexifiant la spécification du modèle linéaire avec plusieurs indicateurs climatiques.

Modélisation des pertes de rendements extrêmes

Traitement des données agricoles

La base des pertes de rendements est constituée à partir des données de rendements issues de la SAA. Pour ce faire, un écart relatif entre les rendements annuels et une moyenne de référence est calculé. Cette dernière peut être déterminée de deux manières :

- à partir de **données historiques**, en tenant compte de la moyenne de rendements des cinq années précédents la donnée de rendements ;
- à partir de **données mises en *as if***, en tenant compte de la même moyenne de rendements pour toutes les années de la période d'étude.

Méthodologie

L'objectif de cette partie était de calibrer la loi des dépassements de seuil pour modéliser les probabilités de retour de pertes de rendements extrêmes. Pour cela, la distribution retenue est une loi de Pareto Généralisée (GPD), de fonction de répartition :

$$G_{\xi,\beta}^p(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)_+^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{si } \xi = 0 \end{cases} \quad \text{pour } \begin{cases} x \geq 0 & \text{si } \xi \leq 0 \\ 0 \leq x \leq -\beta/\xi & \text{si } \xi < 0 \end{cases}$$

où β est le paramètre d'échelle et ξ le paramètre de forme.

L'analyse s'est effectuée selon chaque méthode de traitement des données présentée ci-dessus. Dans chaque cas se sont enchaînées les étapes suivantes :

1. Choix du seuil de la GPD.
2. Calibrage d'une loi GPD globale (Modèle **M1**) selon une méthode de maximum de vraisemblance.
3. Calibrage de deux lois GPD dynamiques à partir de la méthodologie de Chavez-Demoulin, Embrechts et Hofert [10] intégrant un maximum de vraisemblance pénalisé. Une première loi (Modèle **M2**) intègre des effets fixes type année, culture et région dans le calibrage des paramètres de la GPD. La seconde (Modèle **M3**) tient également compte de ces covariables en plus de données climatiques.
4. Calcul de probabilités de retour de pertes de rendements supérieures à 30%.

Résultats obtenus

La Table 3 présente les résultats de calibrages obtenus avec le modèle **M1**. Le seuil requis est légèrement plus faible pour les données mises en *as if*. Ce modèle conduit également à une probabilité de perte plus basse qu'avec un calibrage à partir de données historiques.

	Données historiques	Données mises en <i>as if</i>
Seuil retenu	16%	14%
Probabilité de retour de perte supérieure à 30%	23%	19%

TABLE 3 – Résultats des calibrages du modèle **M1**

La Table 4 résume les spécifications retenues pour les modèles **M2** et **M3**.

	Données historiques	Données mises en <i>as if</i>
Modèle M2	Modèle pour ξ	$f(\text{Culture, Région, Année})$
	Modèle pour ν	$f(\text{Région})$
Modèle M3	Modèle pour ξ	$f(\text{Culture, Région, SWI}_t, \text{SWI}_{t-2})$
	Modèle pour ν	$f(\text{Région})$

TABLE 4 – Récapitulatif des modèles retenus

Les covariables du premier modèle sont identiques, quelle que soit l'approche de traitement des données utilisée. Elles diffèrent davantage pour le second modèle. En outre, le calibrage et le choix des covariables furent particulièrement délicat. En effet, en raison de la rareté des données de pertes de rendements extrêmes, les prédictions des différents modèles testés n'étaient pas toujours de bonne qualité. Des arbitrages ont donc dû être effectués entre qualité de la prédiction, significativité du modèle et pertinence des covariables.

La Figure 5 fournit les cartes résultant du calcul de la probabilité de retour de perte supérieure à 30%, conditionnelle à la survenance d'un événement dépassant le seuil prédéfini, pour les modèles **M2** et **M3** pour la culture de blé en 2025. Dans ce dernier modèle, le SWI est projeté à partir du scénario RCP 8.5 du GIEC.

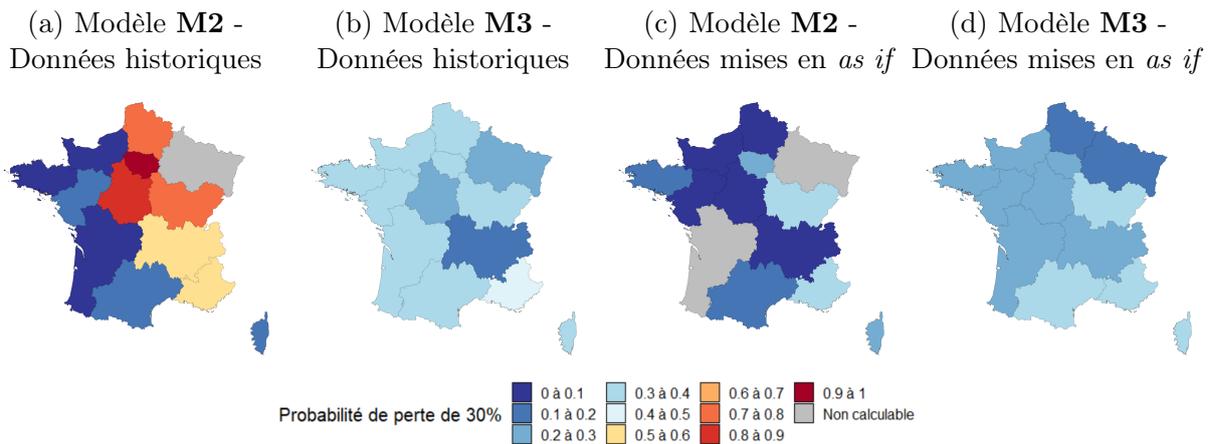


FIGURE 5 – Probabilité de perte à 30% projetée par région en 2025 pour la culture de blé

Par rapport au résultat obtenu avec les données historiques, le modèle **M2** avec les données mises en *as if* conduit à l'obtention de probabilité de retour plus modérées. Les prévisions du modèle **M3** sont également plus mesurées que les projections fournies à partir des données historiques. Par rapport au modèle **M2**, les probabilités de retour prédites sont toutefois légèrement plus élevées selon les régions considérées.

Bien que la méthode de Chavez-Demoulin, Embrechts et Hofert [10] permette d'affiner les lois de dépassements de seuil à partir de covariables, la faible qualité des prédictions peut conduire à une surestimation des probabilités de retour, notamment dans les modèles intégrant l'année de survenance de la perte. Considérer un indicateur climatique peut permettre d'obtenir des résultats plus cohérents vis-à-vis d'un calibrage classique des paramètres de la GPD. Enfin, compléter la modélisation par un calibrage de la loi de fréquence des dépassements de seuil en fonction de covariables pourrait être pertinent.

Conclusion

En qualité de « boîte à outils », ce mémoire propose plusieurs méthodes d'évaluation du risque climatique dans le provisionnement. Chacune d'elle a un objectif différent, et leurs limites ont déjà été évoquées plus haut dans cette note de synthèse. Les résultats obtenus avec ces premières applications sont encourageants, bien que les approches présentées restent perfectibles. Ces dernières pourraient de plus être employées sur d'autres données, afin de confirmer les conjectures émises mais également affiner les modèles calibrés. En particulier, si le secteur agricole est le fil conducteur de ce mémoire, ce dernier fournit des pistes d'analyse du risque climatique pouvant servir pour d'autres branches Non-Vie.

Executive summary

Introduction

As a result of climate change, natural events such as drought and hail are set to increase in frequency and intensity over the coming years. In agriculture, these climatic changes will have a significant impact on yields. Particularly, rising temperatures and heatwaves tend to dry out farmland, reducing its fertility and thus the quality and selling price of crops.

Consequently, taking climate risk into account is a major challenge for players in the insurance sector, especially when it comes to assessing reserves. The purpose of this thesis is to provide a toolbox for assessing this emerging risk in the reserves of a Non-Life insurance company. It presents a number of methods to address different issues:

- a **qualitative analysis of climate risk** in reserving, via a detailed study of the form of reserve distribution;
- a **quantitative analysis of climate risk**, aiming to extract an inflationary trend in reserves linked to meteorological variables and project it according to climate scenarios;
- and an **extreme yield losses modeling** based on climate data.

Highly exposed to climatic risks, the agricultural sector is a good example of the application of our approach.

Data

Throughout the thesis, the meteorological variable used is a soil moisture indicator, the Soil Wetness Index. These data come from two databases, originating from *Météo France*¹ for historical values and from the Climate Impact Explorer² for future data based on climate scenarios.

In addition, the liquidation triangles required to address the first two issues are derived from customer data, anonymized prior to study. Three sub-lines of business are available: **climatic hazards, excluding climatic hazards and storms** and **third-party liability**. These triangles are provided for the period 2002-2019.

Lastly, the agricultural production and yield data required for the final part of the study are taken from the Annual Agricultural Statistics (SAA) for crops grown by the French organism Agreste³, the ministerial statistical service for agriculture. The study period is 2000-2020, and indicators are collected at the regional level for metropolitan France. The analysis is organized around several crops, included in the category “Cereals, oilseeds, protein plants”.

1. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture>

2. <https://climate-impact-explorer.climateanalytics.org/>

3. https://agreste.agriculture.gouv.fr/agreste-saiku/?plugin=true&query=query/open/SAANR_DEVELOPPE_2#query/open/SAANR_DEVELOPPE_2

Qualitative analysis of climate risk

Dal Moro's reserving model

In his article [13], Dal Moro proposes an extension of Mack chain-ladder's model [36] by supplementing it with two assumptions about the third-order moment, named skewness, of cumulative payments conditional on past payments:

1. The skewness of the cumulative payment $C_{i,j+1}$ conditional on past payments does not depend on the year of occurrence i but on the year of development j .
2. There are coefficients $Sk_j > 0$ for $1 \leq j \leq I - 2$ such that

$$\mathbb{E} [(C_{i,j+1} - \mathbb{E}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}])^3 | C_{i,1}, \dots, C_{i,j}] = Sk_j^3 C_{i,j}^{3/2}.$$

Based on these two hypotheses, the author proposes two methodologies for calculating the skewness of the distribution of reserves:

- a **closed-formula approach**, whose results are given by the **quantile** function in the **ChainLadder** package in R;
- and a **simulatory approach** based on a Generalized Pareto Distribution (GPD). The latter proposes to generate the cumulative payments of the lower triangle from one step to the next, by calibrating the parameters of the GPD law on the basis of Mack and Dal Moro's hypotheses via a method of moments.

Results

A verification of Mack and Dal Moro's hypotheses led to the deflation of the climatic hazards triangle using the Producer Price Index for Agricultural Products (IPPAP) provided by the Agreste⁴. This treatment was not necessary for the other two triangles. The GPD simulation approach was then run on the customer triangles. Table 1 summarizes the results obtained for skewness and kurtosis. For the GPD simulation, skewness and kurtosis values were empirically obtained from the final distributions, shown in Figure 1.

Triangle / Method	SKEWNESS		KURTOSIS
	Closed formula	GPD Simulation	GPD Simulation
Deflated climtic hazards	1,32	1,36	4,91
Excluding climatic hazards and storms	0,72	0,78	3,29
Third-party liability	0,76	0,92	4,40

TABLE 1 – Skewness and kurtosis results for reserve distributions obtained for the three triangles

4. https://agreste.agriculture.gouv.fr/agreste-web/disaron/D_0037/detail/

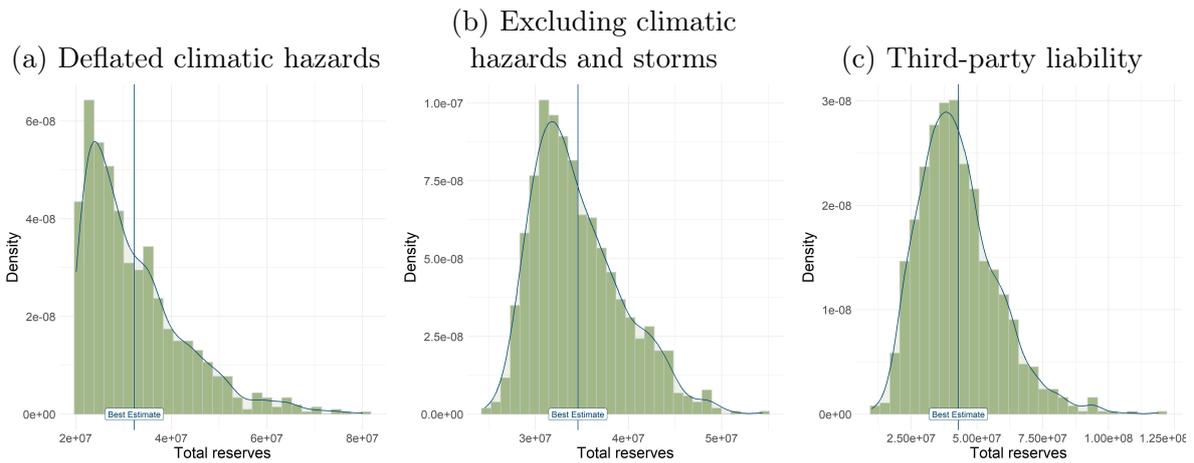


FIGURE 1 – Reserve distributions obtained per GPD simulation - 1000 simulations

It can be seen from Table 1 and Figure 1 that the climate hazards sub-branch has a much higher asymmetry (given by skewness) and kurtosis (given by kurtosis) than the other two sub-branches. This could be a potential effect of the presence of climate risk in this sub-branch. To validate this conjecture, we would need to study more triangles from other insurers. Staying with the agricultural sector, an analysis of the shape of reserve distributions by type of crop could also be considered, in order to verify whether climate risk similarly affects their reserving.

Quantitative analysis of climate risk

Adaptation of Verbeek’s separation method

The actuarial literature provides several methodologies for taking inflation-type calendar effects into account in reserving. One of these approaches, the separation method introduced by Verbeek [58] and developed by Taylor [55], has been adapted in this thesis to extract not an inflationary trend but a climatic trend from reserves. Figure 2 shows the main hypothesis underlying each model.

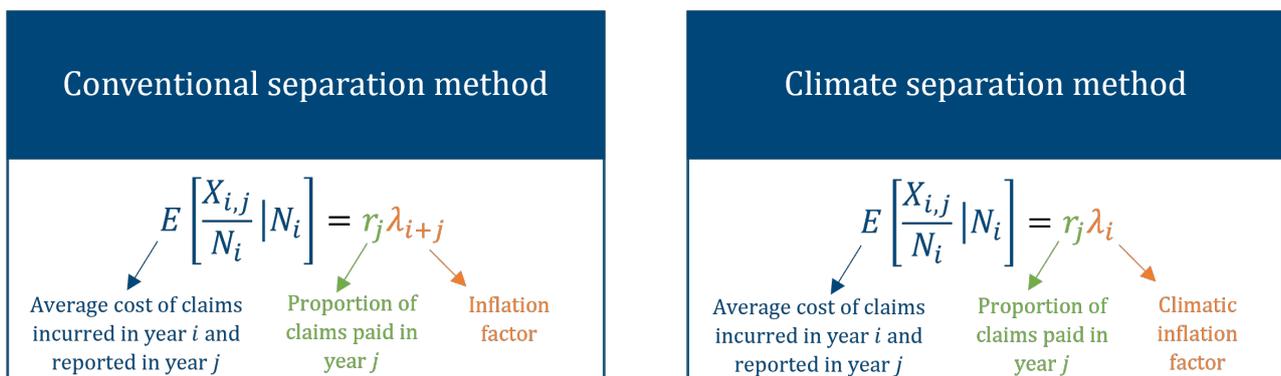


FIGURE 2 – Adaptation de l’hypothèse principale de la méthode de séparation de Verbeek [58]

To apply this method to the climatic hazard triangle, two preliminary steps were followed:

1. Deflation of the triangle to remove the spurious effects of inflation.
2. Due to a lack of data, simulation of a triangle of claim numbers, based on two assumptions:

(H1) The average number of claims depends on the annual claims experience:

$$\mathbb{E}[N_{i,j}] = r_j \times SP_i \times \alpha M$$

with SP_i loss ratio for the year of occurrence i .

(H2) The average number of claims is **independent of the annual loss ratio** :

$$\mathbb{E}[N_{i,j}] = r_j \times \alpha M.$$

In each case, the numbers of claims $N_{i,j}$ are then simulated using Poisson distributions with mean $\mathbb{E}[N_{i,j}]$. α represents the proportion of claims recorded among a fixed number of contracts M . This number is fixed since it is not provided in the data. r_j , the proportion of claims paid per year of development, has been set to reflect the short development of claims in the portfolio.

Following these two steps, the triangle of claim numbers was completed with the Mack chain-ladder's model, and the climate separation method was applied to the run-off triangle.

Results

The estimated reserve amounts and the climate trend extracted using the separation method are shown in Figure 3 below.

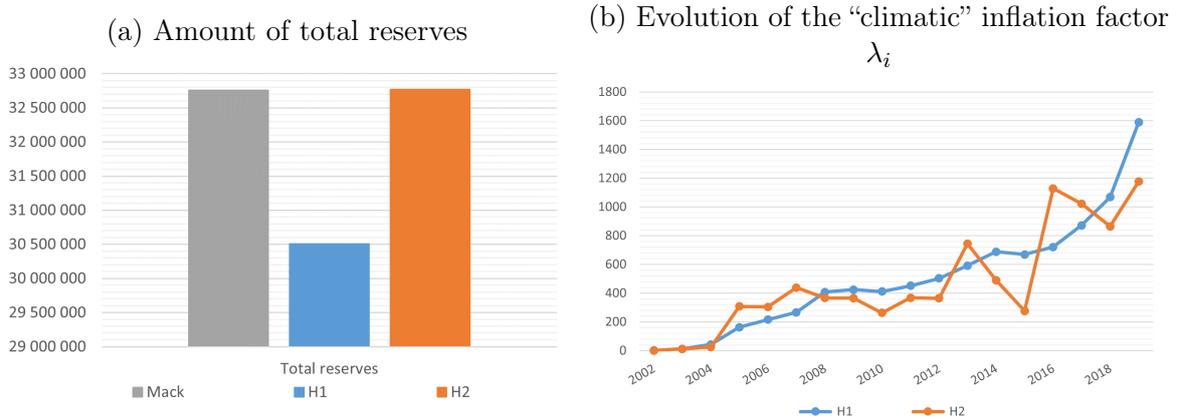


FIGURE 3 – Results of the climate separation method

Assumption (H2) seems more appropriate to the data: it provides a Best Estimate close to that calculated with Mack chain-ladder's model. Moreover, the "climatic" inflation factor, although observing an increasing trend, also follows variations of the loss ratio. Therefore, it can more accurately reflect the behavior of a meteorological variable than the factor extracted with (H1).

This was followed by a search for the best linear model to link the climatic trend to the soil moisture data. The selected specification is as follows:

$$\lambda_i = \alpha + \beta \times \text{Year}_i + \gamma \times \text{SWI}_{i-2}.$$

The results of the linear regression are presented in Table 2. A decrease in the soil moisture index leads to an increase in the climatic trend. Besides, the growth in "climatic" inflation factor as a function of the year of occurrence is well captured in this model.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-1,125e+05	1,660e+04	-6,776	6,25e-06
Year _i	56,55	8,227	6,873	5,29e-06
SWI _{i-2}	-836,9	426,4	-1,963	0,0685

Adjusted R²: 0,7595

TABLE 2 – Results of the selected linear regression model

Last but not least, the model linking the “climatic” inflation factor and SWI was projected on the basis of evolution scenarios from the IPCC and others.

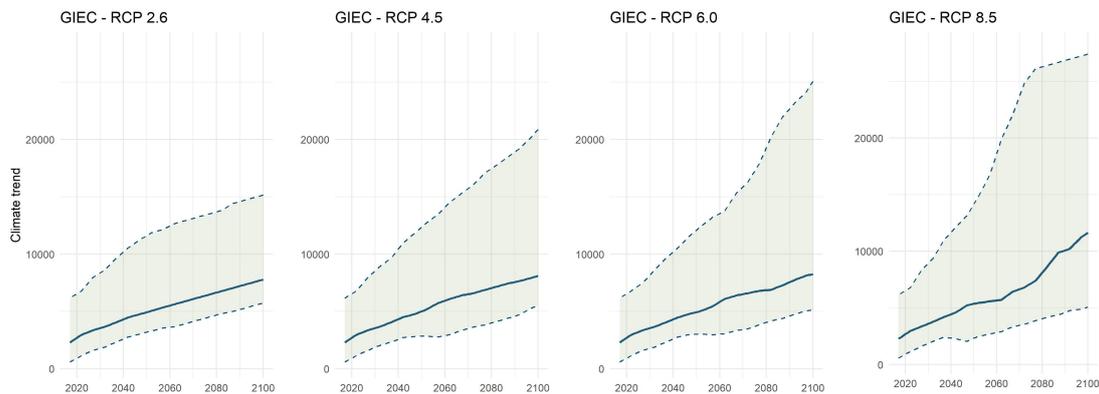


FIGURE 4 – Evolution of climate trend according to IPCC scenarios

In Figure 4, all scenarios point to a sharp increase in climate trends over the coming years. Eventhough the model selected makes little physical sense, considering an indicator delayed by two years, it nevertheless enables us to establish an order of magnitude to guide the reserving of agricultural lines of business. These initial results are therefore promising and demonstrate the relevance of the methodology applied, although it could be improved, for example by making the specification of the linear model more complex with several climate indicators.

Extreme yield losses modeling

Processing agricultural data

The database for yield losses is built with yield data extracted from the SAA. To do so, a relative difference between annual yields and a reference average is calculated. The latter can be determined in two ways:

- from **historical data**, taking into account the average yield for the five years preceding the yield data;
- from **data put into as if**, taking into account the same average yield for all years of the study period.

Methodology

The aim of this section was to calibrate the threshold exceedance distribution to model the probability of return of extreme yield losses. For this purpose, a Generalized Pareto Distribution

(GPD) was chosen:

$$G_{\xi,\beta}^p(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)_+^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{if } \xi = 0 \end{cases} \quad \text{for } \begin{cases} x \geq 0 & \text{if } \xi \leq 0 \\ 0 \leq x \leq -\beta/\xi & \text{if } \xi < 0 \end{cases}$$

with β scale parameter and ξ shape parameter.

The analysis was carried out according to each of the data processing methods described above. In each case, the following steps were followed:

1. Choice of GPD threshold.
2. Calibration of a global GPD law (**M1** Model) using a maximum likelihood method.
3. Calibration of two dynamic GPD laws using the Chavez-Demoulin, Embrechts and Hofert methodology [10] involving penalized maximum likelihood. A first law (**M2** Model) integrates fixed effects such as year, crop and region in the calibration of GPD parameters. The second (**M3** Model) also takes these covariates into account, in addition to climatic data.
4. Calculation of return probabilities of yield losses greater than 30%.

Results

Table 3 shows the calibration results obtained with the **M1** model. The required threshold is slightly lower for the data set in as if. This model also leads to a lower probability of loss than calibration using historical data.

	Historical data	Data put into as if
Selected threshold	16%	14%
Return probabilities of yield losses greater than 30%	23%	19%

TABLE 3 – Calibration results of **M1** Model

Table 4 summarizes the specifications retained for the **M2** and **M3** models.

	Historical data	Data put into as if
M2 Model	Model for ξ	$f(\text{Crop, Region, Year})$
	Model for ν	$f(\text{Region})$
M3 Model	Model for ξ	$f(\text{Crop, Region, SWI}_t, \text{SWI}_{t-2})$ $f(\text{Crop, Region, Year, SWI}_t)$
	Model for ν	$f(\text{Region})$

TABLE 4 – Overview of selected models

The covariates in the first model are identical, regardless of the data processing approach used. They differ more for the second model. Moreover, the calibration and choice of covariates

were particularly challenging. Indeed, given the scarcity of extreme yield loss data, the predictions of the various models tested were not always of good quality. Consequently, trade-offs had to be made between prediction quality, model significance and covariate relevance.

Figure 5 provides the maps resulting from the calculation of the probability of return of losses greater than 30%, conditional on the occurrence of an event exceeding the predefined threshold, for the **M2** and **M3** models and wheat cultivation in 2025. In the latter model, the SWI is projected from the IPCC RCP 8.5 scenario.

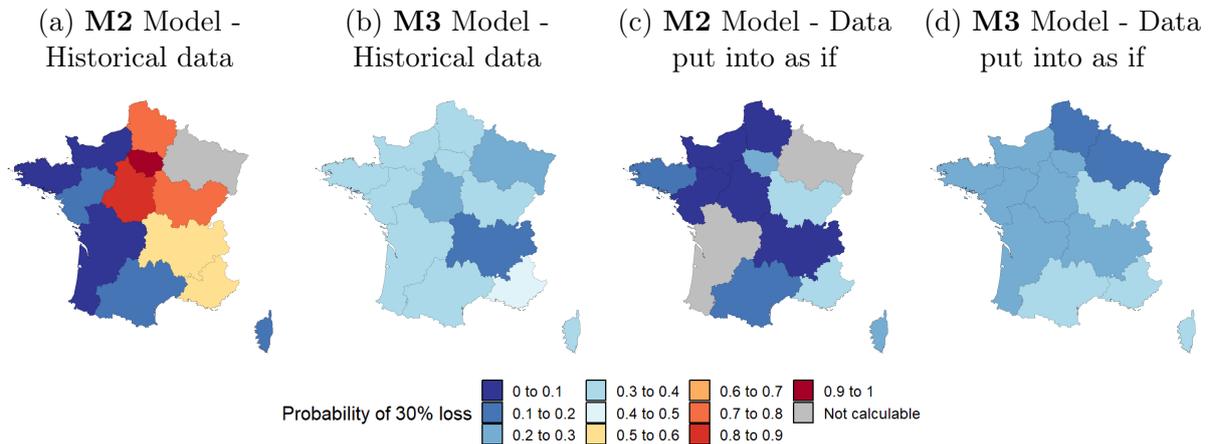


FIGURE 5 – Return probabilities of yield losses greater than 30% projected by region in 2025 for wheat cultivation

Compared with the result obtained with historical data, the **M2** model implemented with the data put into as if leads to more moderate return probabilities. The **M3** model's forecasts are also more balanced than the projections based on historical data. Compared with the **M2** model, however, the predicted return probabilities are slightly higher, depending on the regions considered.

Although the Chavez-Demoulin, Embrechts and Hofert [10] method can be used to refine threshold exceedance laws based on covariates, the low quality of predictions can lead to an overestimation of return probabilities, particularly in models incorporating the year of loss occurrence. Considering a climatic indicator can help to obtain results that are more consistent with a conventional calibration of GPD parameters. Finally, it might be useful to enhance the modeling by calibrating the frequency distribution of threshold exceedances as a function of covariates.

Conclusion

As a “toolbox”, this thesis proposes several methods for assessing climate risk in reserving. Each of them has a different objective, and their limitations have already been discussed earlier in this summary. The results obtained with these initial applications are encouraging, although there is still room for improvement. They could also be tested with other data, in order to confirm the conjectures made and refine the calibrated models. In particular, while the agricultural sector is the main focus of this thesis, it also provides a number of avenues for analyzing climate risk that could be applied to other Non-Life sectors.

Annexes

Annexe A

Compléments sur l'assurance MRC

	Assurance grêle	Assurance multirisque climatique	
		Contrat à la culture	Contrat à l'exploitation
Evènements garantis	Grêle (+ tempête obligatoire)	17 aléas dont le gel, la tempête, la grêle, l'inondation et la sécheresse. Possibilité de couvrir la perte de qualité par extension	
Obligation de couverture	Couvrir l'ensemble de la sole de la culture assurée	Assurer 70 % des cultures de vente et 100 % pour la vigne et les prairies	Assurer 80 % des cultures de vente et au moins deux natures de récoltes
Base indemnitaire	Décompte des pertes sur terrain	Décompte des pertes sur terrain	
Choix des rendements	Libre	Rendements historiques moyens (+15 % sur option)	
Déclaration de sinistre	OUI	OUI	
Expertise terrain	OUI	OUI pour tous les aléas	
Seuil de déclenchement	Fonction du choix de la franchise	30% (rachat possible)	
Franchise	Au choix (10 à 30 %) et souvent à la parcelle	30 % (rachat possible dont grêle à 10 % à la parcelle)	20 %
Gestion	Simple déclaration d'assolement	Lourdeur et obligations administratives	
Subvention	NON	OUI (deux niveaux de subvention selon les choix de garanties : 65 % et 45 %)	
Prime (faux)	1 à 3 % (en fonction de la culture et de la franchise)	1,5 à 12 % (avant subvention) et 1,5 à 8 % net de subvention selon les options choisies	1,5 à 10 % (avant subvention) et 1,5 à 5 % net de subvention selon les options choisies

TABLE A.1 – Récapitulatif des contrats d'assurance récoltes climatiques existants [28]

Annexe B

Compléments sur la méthode de simulation de Dal Moro

B.1 Modèle GEV

Définition B.1.1. Soit X une variable aléatoire. X suit une **loi d'extremum généralisée** de paramètres $\mu \in \mathbb{R}$, $s > 0$ et $\xi \in \mathbb{R}$, i.e $X \sim GEV(\mu, s, \xi)$, si sa densité est telle que :

$$f_X(x) = \begin{cases} \frac{1}{s} \left(1 + \xi \frac{x - \mu}{s}\right)^{-\frac{1}{\xi}-1} \exp \left[-\left(1 + \xi \frac{x - \mu}{s}\right)^{-\frac{1}{\xi}} \right] & \text{pour } \xi \neq 0 \\ \frac{1}{s} \exp \left(-\frac{x - \mu}{s} \right) \exp \left[-\exp \left(-\frac{x - \mu}{s} \right) \right] & \text{pour } \xi = 0 \end{cases}$$

où μ est le paramètre de position, s le paramètre d'échelle et ξ le paramètre de forme.

En considérant $\hat{C}_{i,k} \sim GEV(\mu_{i,k}, s_{i,k}, \xi_k)$, on peut exprimer les moments de $\hat{C}_{i,k}$ en fonction des paramètres :

$$\mathbb{E}[\hat{C}_{i,k}] = \mu_{i,k} + s_{i,k} \frac{\Gamma(1 + \xi_k) - 1}{\xi_k}$$

$$\text{Var}(\hat{C}_{i,k}) = s_{i,k}^2 \frac{\Gamma(1 - 2\xi_k) - \Gamma(1 - \xi_k)^2}{\xi_k^2}$$

$$\text{Skew}(\hat{C}_{i,k}) = \frac{\Gamma(1 - 3\xi_k) - 3\Gamma(1 - \xi_k)\Gamma(1 - 2\xi_k) + 2\Gamma(1 - \xi_k)^3}{(\Gamma(1 - 2\xi_k) - \Gamma(1 - \xi_k)^2)^{3/2}}$$

$$\text{Kurt}(\hat{C}_{i,k}) = \frac{\Gamma(1 - 4\xi_k) - 4\Gamma(1 - \xi_k)\Gamma(1 - 3\xi_k) + 6\Gamma(1 - 2\xi_k)\Gamma(1 - \xi_k)^2 - 3\Gamma(1 - \xi_k)^4}{(\Gamma(1 - 2\xi_k) - \Gamma(1 - \xi_k)^2)^2}$$

où Γ est la fonction Gamma d'Euler définie par $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ pour $x \in \mathbb{R}_+^*$.

Les étapes de l'algorithme suivant sont analogues à celles présentées pour le modèle GPD. Pour chaque année de développement k :

Etape 1 Génération du paiement cumulé $\hat{C}_{i,k}$

- i. On suppose les paramètres $(\mu_{i,k}, s_{i,k}, \xi_k)$ connus.
- ii. On génère une variable aléatoire $U_{i,k}$ de loi uniforme sur $(0, 1]$.
- iii. On détermine $\hat{C}_{i,k} = \mu_{i,k} + \frac{s_{i,k} ((-\ln(U_{i,k}))^{-\xi_k} - 1)}{\xi_k}$.

Etape 2 Obtention des paramètres $(\mu_{i,k}, s_{i,k}, \xi_k)$

i. On détermine le paramètre ξ_{k+1} :

— Soit via la formule du skewness, en se rappelant que les hypothèses (H3) et (H5) donnent

$$\frac{\hat{S}_{k+1}^3}{\hat{\sigma}_{k+1}^3} = \frac{\Gamma(1 - 3\xi_{k+1}) - 3\Gamma(1 - \xi_{k+1})\Gamma(1 - 2\xi_{k+1}) + 2\Gamma(1 - \xi_{k+1})^3}{(\Gamma(1 - 2\xi_{k+1}) - \Gamma(1 - \xi_{k+1})^2)^{3/2}}.$$

— Soit via la formule du kurtosis, en se rappelant que les hypothèses (H3) et (H7) donnent

$$\frac{\hat{K}_{k+1}^4}{\hat{\sigma}_{k+1}^4} = \frac{\Gamma(1 - 4\xi_{k+1}) - 4\Gamma(1 - \xi_{k+1})\Gamma(1 - 3\xi_{k+1}) + 6\Gamma(1 - 2\xi_{k+1})\Gamma(1 - \xi_{k+1})^2 - 3\Gamma(1 - \xi_{k+1})^4}{(\Gamma(1 - 2\xi_{k+1}) - \Gamma(1 - \xi_{k+1})^2)^2}.$$

ii. On obtient le paramètre $s_{i,k+1}$ en résolvant l'équation suivante :

$$\text{Var}(\hat{C}_{i,k+1}) = \hat{C}_{i,k} \hat{\sigma}_k^2 \left(1 + \frac{\hat{C}_{i,k}}{\sum_{j=1}^{I-k} C_{j,k}} \right) = s_{i,k+1}^2 \frac{\Gamma(1 - 2\xi_{k+1}) - \Gamma(1 - \xi_{k+1})^2}{\xi_{k+1}^2}.$$

iii. Enfin, pour obtenir le paramètre $\mu_{i,k+1}$ il faut résoudre l'équation

$$\mathbb{E}[\hat{C}_{i,k+1}] = \hat{f}_k \hat{C}_{i,k} = \mu_{i,k+1} + s_{i,k+1} \frac{\Gamma(1 + \xi_{k+1}) - 1}{\xi_{k+1}}.$$

B.2 Rappels sur les copules

Dans cette partie, nous rappelons quelques définitions et propriétés relatives aux copules, présentées dans le cours de Fermanian [18].

Définition B.2.1. Soit une application $G : \overline{\mathbf{R}}^m \rightarrow [0, 1]$. Soient $\mathbf{x} \in \mathbf{R}^m, a, b \in \mathbf{R}$. Pour $1 \leq k \leq m$, on note :

$$(\mathbf{x}_{-k}, a) = (x_1, \dots, x_{k-1}, a, x_{k+1}, \dots, x_n)$$

$$\Delta_{k,a}^b G(\mathbf{x}) = G(\mathbf{x}_{-k}, b) - G(\mathbf{x}_{-k}, a)$$

On définit le *volume d'une boîte* $\mathbf{B}_{\mathbf{a},\mathbf{b}} := \prod_{k=1}^m (a_k, b_k]$ par :

$$V_G(\mathbf{B}_{\mathbf{a},\mathbf{b}}) = \Delta_{1,a_1}^{b_1} \dots \Delta_{1,a_m}^{b_m} G(\mathbf{x})$$

Définition B.2.2. Soit une application $G : \overline{\mathbf{R}}^m \rightarrow [0, 1]$. G est **m-croissante** si pour tous $\mathbf{a}, \mathbf{b} \in \mathbf{R}^m$ tels que $a_k \leq b_k$ pour tous $1 \leq k \leq m$,

$$V_G(\mathbf{B}_{\mathbf{a},\mathbf{b}}) \geq 0$$

Définition B.2.3. Soit une application $G : \overline{\mathbf{R}}^m \rightarrow [0, 1]$. G est une **fonction de répartition m-dimensionnelle** si :

1. G est m-croissante.
2. $G(\mathbf{x}) = 0$ lorsqu'une des composantes de \mathbf{x} est $-\infty$.
3. $G(+\infty, \dots, +\infty) = 1$.

Lemme B.2.4. Soient X une variable aléatoire, F une fonction de répartition univariée. Si F est continue, alors

$$F(X) \sim \mathcal{U}([0, 1]) \iff X \sim F.$$

Définition B.2.5. Une **copule m-dimensionnelle** est une fonction de répartition $C : [0, 1]^m \rightarrow [0, 1]$ dont les marges sont uniformes.

Théorème B.2.6. (Théorème de Sklar) Pour toute fonction de répartition H , de marges $F_k, 1 \leq k \leq m$, il existe une copule m-dimensionnelle C telle que

$$\forall \mathbf{x} \in \mathbb{R}, H(\mathbf{x}) = C(F_1(x_1), \dots, F_m(x_m))$$

et on peut imposer la continuité de C .

Réciproquement, considérons une copule C et des marges $F_k, 1 \leq k \leq m$ respectant la relation précédente. Alors celle-ci définit une fonction de répartition H de \mathbb{R}^m de marges F_k .

Définition B.2.7. Soient $\mathbf{u} = (u_1, \dots, u_m) \in [0, 1]^m$ et Σ matrice de corrélation de taille m . On note Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$ et Φ_Σ la fonction de répartition de la loi $\mathcal{N}(0_m, \Sigma)$. La **copule gaussienne** m-dimensionnelle de paramètre Σ est définie par

$$C_\Sigma(\mathbf{u}) = \Phi_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m)).$$

De manière générale, pour générer une loi ayant une structure de copule gaussienne, on procède en deux étapes :

1. Générer un échantillon X de loi $\mathcal{N}(0_m, \Sigma)$.
2. Calculer $U = \Phi(X)$. Puisque Φ est une fonction continue, on sait alors que

$$X \sim \Phi \iff U \sim \mathcal{U}([0, 1]).$$

Annexe C

Application de la méthode de séparation climatique sur un exemple simulé

Dans ce chapitre annexe, nous appliquons la méthode de séparation climatique à un exemple simulé en vue de vérifier son bon fonctionnement.

C.1 Hypothèses formulées

Nous reprenons dans cet exemple les quatre hypothèses évoquées au Chapitre 10. Pour capter une tendance climatique, nous décidons de modéliser les montants incrémentaux en tenant compte de l'année de survenance. Toutes les hypothèses formulées sur le nombre de sinistres ou leurs montants incrémentaux sont résumées dans la Table C.1 suivante.

	Nombre de contrats M constant		Nombre de contrats M_i en hausse	
	Hypothèse 1	Hypothèse 2	Hypothèse 3	Hypothèse 4
$\mathbb{E}[N_{i,j}]$	$\alpha \times M \times r_j \times SP_i$	$\alpha \times M \times r_j$	$\alpha \times M_i \times r_j \times SP_i$	$\alpha \times M_i \times r_j$
Loi du nombre	$N_{i,j} \sim \text{Poisson}(\mathbb{E}[N_{i,j}])$			
$X_{i,j}$	$N_{i,j} \times \text{Année}_i$			

TABLE C.1 – Récapitulatif des hypothèses

Une nouvelle fois, α est la proportion de sinistre par rapport au nombre de contrats, supposée constante, et SP_i correspond au ratio sinistres sur primes de l'année de survenance i , ici simulé selon une loi uniforme sur l'intervalle $[0, 7; 1, 3]$.

Année	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
SP_i	73%	71%	117%	74%	80%	114%	110%	80%	85%	94%

TABLE C.2 – Ratios sinistres sur primes simulés

Dans cet exemple, 2012, 2015 et 2016 sont des années à forte sinistralité, qui pourraient correspondre à des années de mauvaises récoltes dues à de mauvaises conditions météorologiques.

Concernant la proportion de sinistres par année, nous choisissons un développement selon la relation suivante :

$$r_j = \begin{cases} 0,5 & \text{pour } j = 1 \\ \frac{1-r_1}{2} & \text{pour } j = 2 \\ \frac{r_{j-1}}{2} & \text{pour } 3 \leq j \leq I = 10 \end{cases}$$

C.2 Résultats obtenus

Nous avons simulé des triangles de nombre de sinistres et de paiements incrémentaux, relatifs à chacune des hypothèses énoncées dans la Table C.1. Les Figures C.1 et C.2 présentent les résultats obtenus en appliquant la méthode de séparation climatique, ainsi que l'approche de bootstrap sous-jacente.

Quelques constats peuvent être effectués à partir des Figures C.1 et C.2 et de la Table C.3 suivantes :

- La méthode de séparation climatique permet d'obtenir des montants de réserves proches de ceux estimés via Mack *chain-ladder*. Néanmoins elle induit un risque de réserve plus élevé.
- Sur chacun des triangles étudiés, l'approche par bootstrap conduit à une diminution du risque de réserve par rapport à la méthode de séparation climatique, bien que les réserves soient légèrement sur-évaluées.
- Tenir compte de l'évolution du nombre de contrats dans la modélisation influence peu le résultat obtenu pour la tendance climatique.
- Les hypothèses sur la tendance climatique et la proportion de sinistres par année de développement se retrouvent bien dans les résultats du modèle, quelle que soit l'hypothèse de simulation initiale considérée.

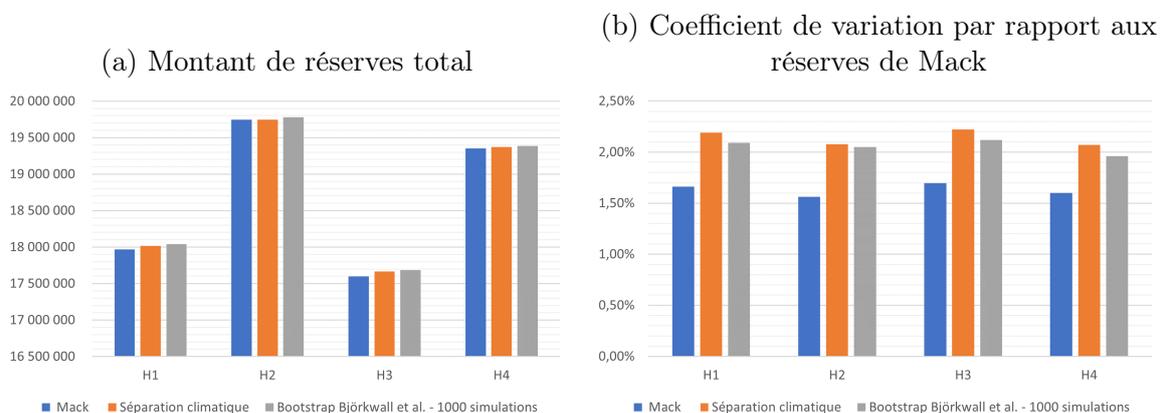


FIGURE C.1 – Comparaison avec les résultats de Mack *chain-ladder*

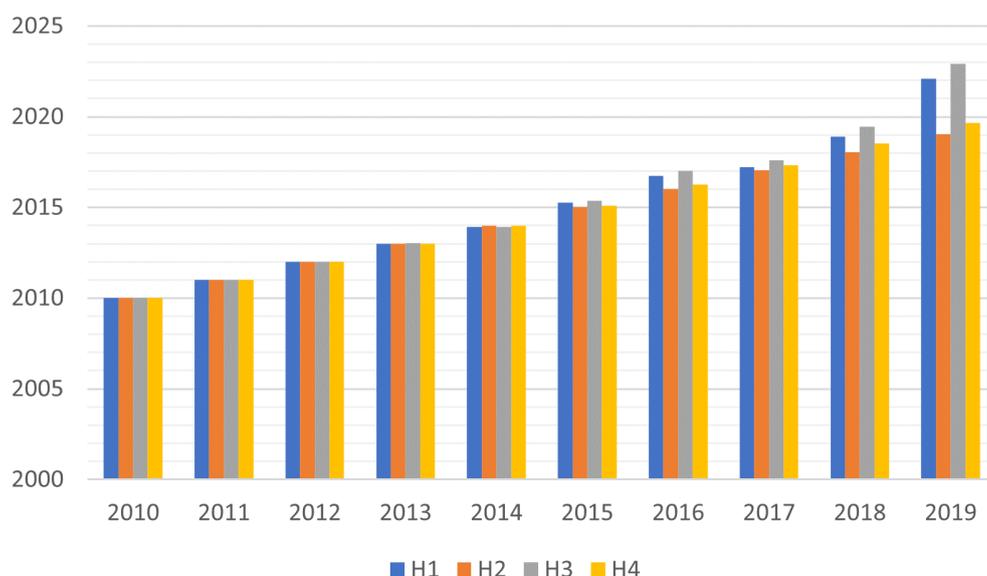


FIGURE C.2 – Evolution du facteur d’inflation « climatique » pour chaque hypothèse

Année j	r_j théorique	r_j H1	r_j H2	r_j H3	r_j H4
1	0,5	0,50028	0,50060	0,50026	0,50055
2	0,25	0,25056	0,25030	0,25060	0,25030
3	0,125	0,12561	0,12546	0,12567	0,12547
4	0,0625	0,06253	0,06261	0,06251	0,06266
5	0,03125	0,03167	0,03154	0,03171	0,03158
6	0,015625	0,01593	0,01582	0,01595	0,01587
7	0,0078125	0,00742	0,00748	0,00739	0,00745
8	0,00390625	0,00356	0,00362	0,00353	0,00355
9	0,001953125	0,00175	0,00177	0,00169	0,00183
10	0,000976563	0,00069	0,00081	0,00068	0,00074

TABLE C.3 – Comparaison des proportions de sinistres obtenues par année de développement

C.3 Distributions obtenues

L’approche par bootstrap conduit à une obtention de la distribution des réserves, sur laquelle peuvent être calculés des indicateurs de risque de réserve, que ce soit des quantiles particuliers ou l’écart-type empirique.

Les distributions relatives à chacun des triangle de paiements sont représentées sur la Figure C.3 ci-après.

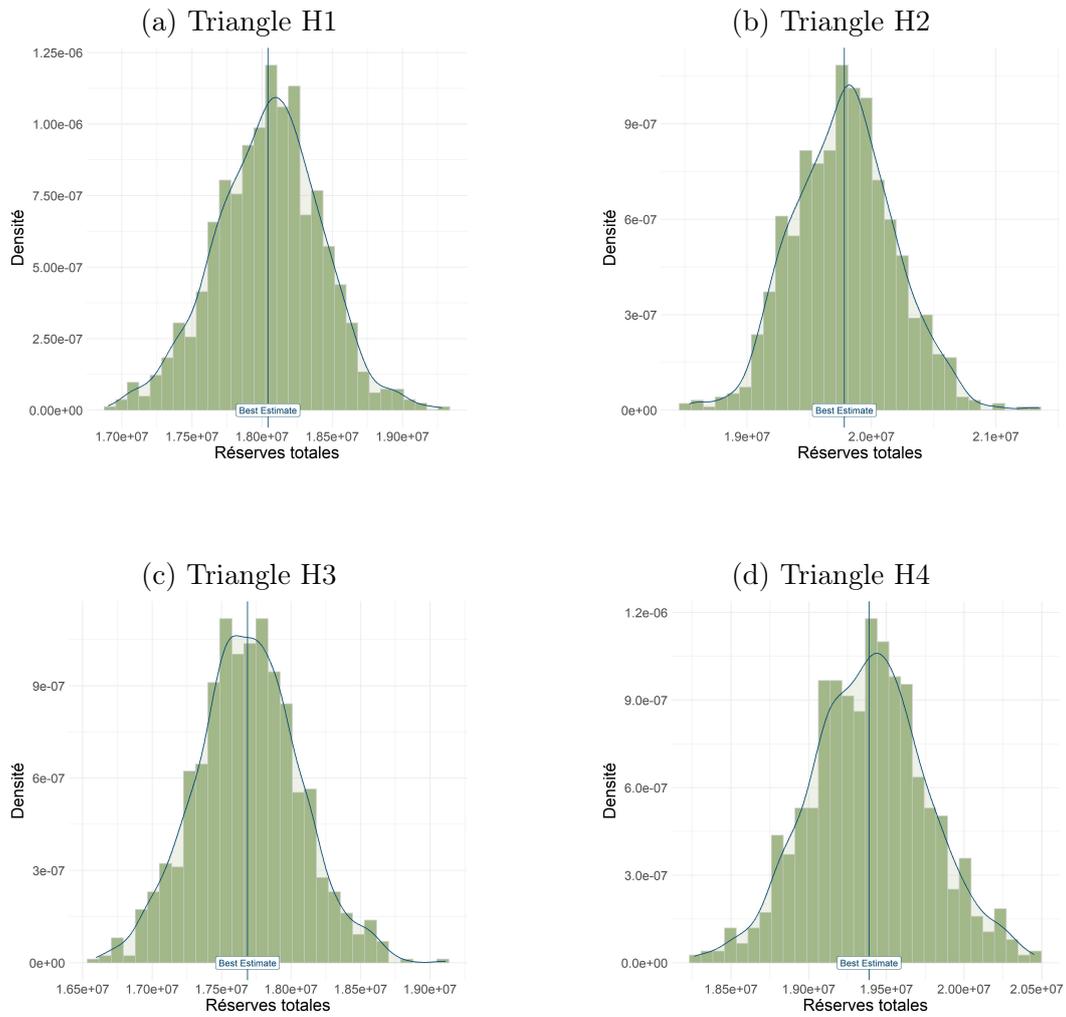


FIGURE C.3 – Distributions obtenues par bootstrap de Björkwall et al. [7] climatique avec 1000 simulations

Annexe D

Validation des hypothèses de provisionnement pour les autres triangles

D.1 Triangle agricole hors tempête et aléas climatiques

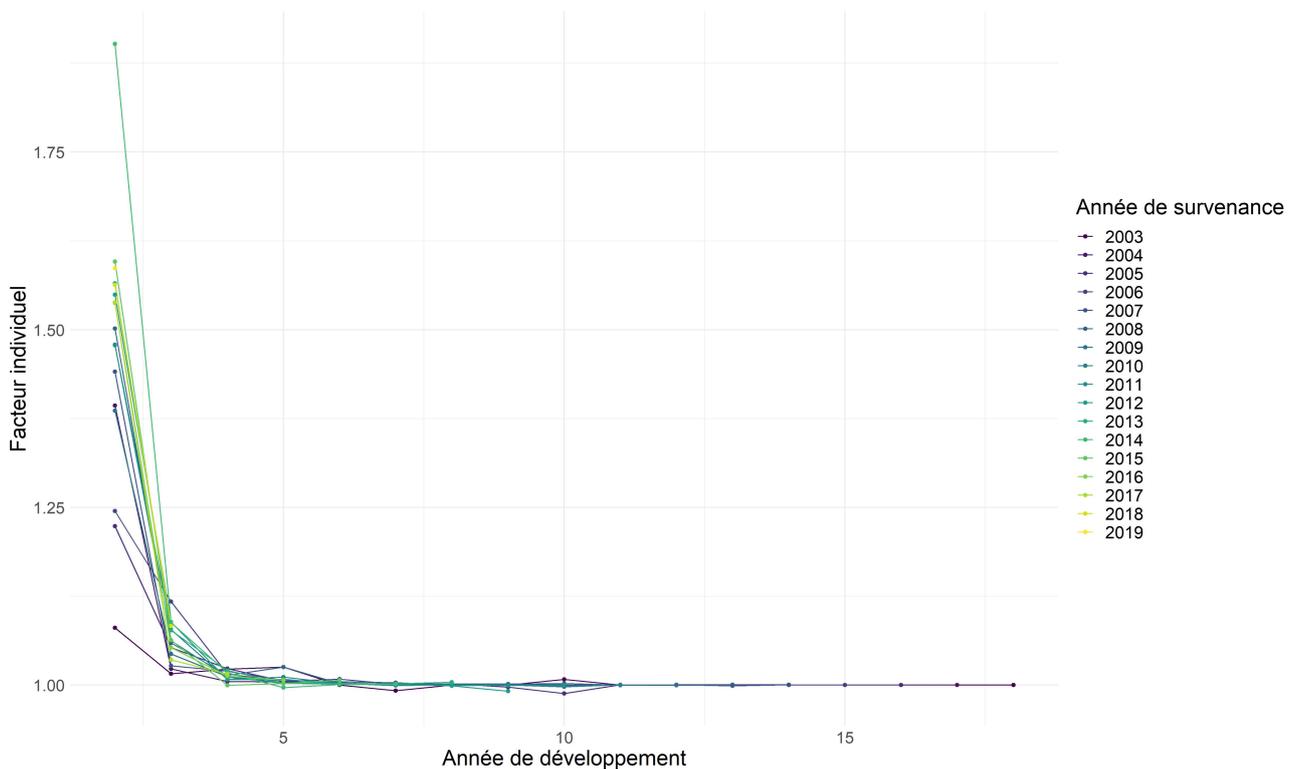


FIGURE D.1 – Vérification de l’Hypothèse 1

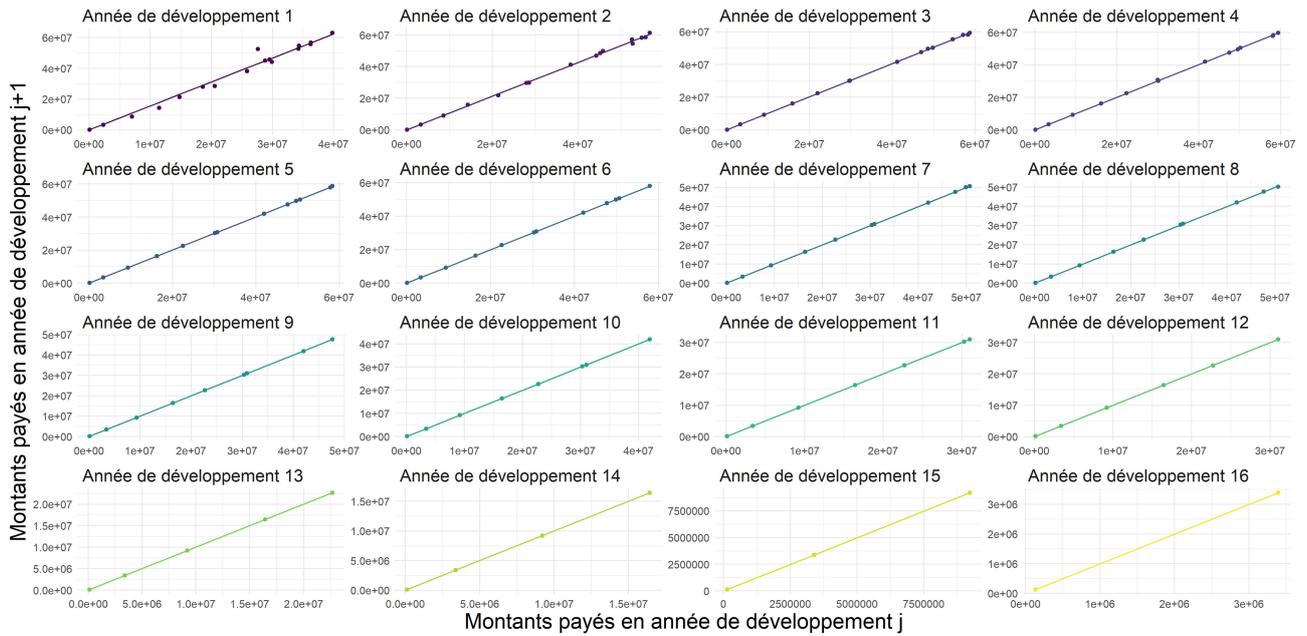


FIGURE D.2 – Vérification de l’Hypothèse 2

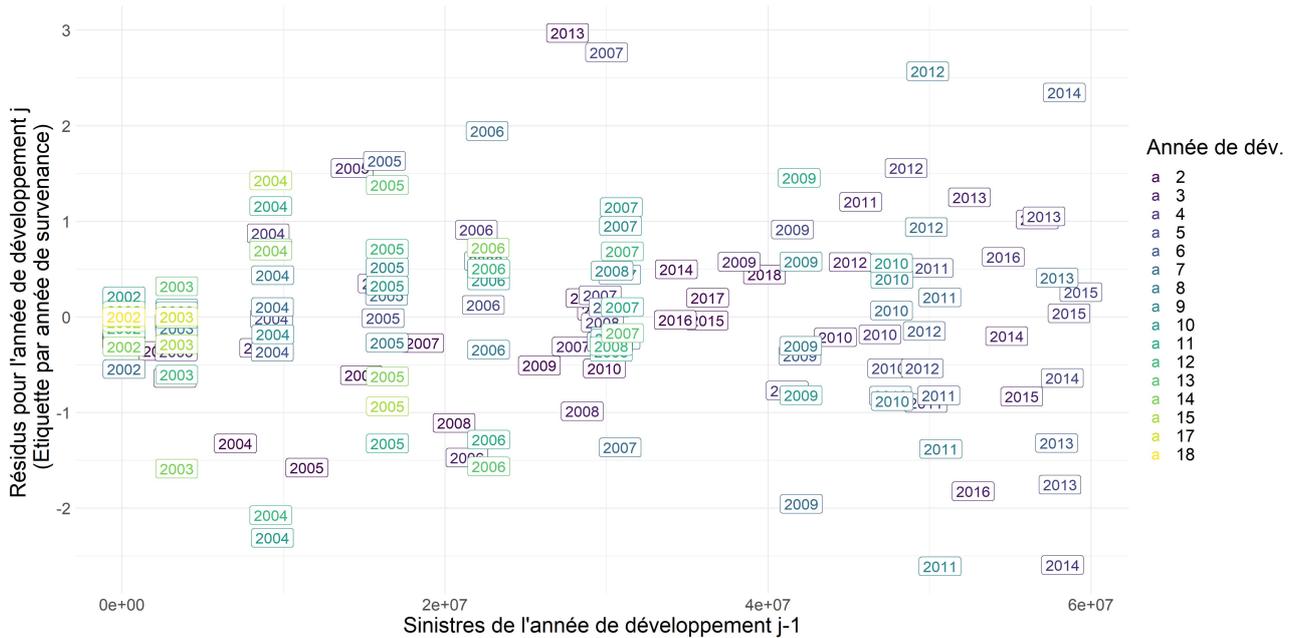


FIGURE D.3 – Vérification de l’Hypothèse 3

Tendance étudiée	p-value
Année de survénance	0,787
Année calendaire	0,803
Année de développement	0,121

TABLE D.1 – Tests de Box-Pierce

La Figure D.3 précédente et les tests de Box-Pierce nous permettent d'accepter l'hypothèse de blancheur des résidus, et donc d'indépendance.

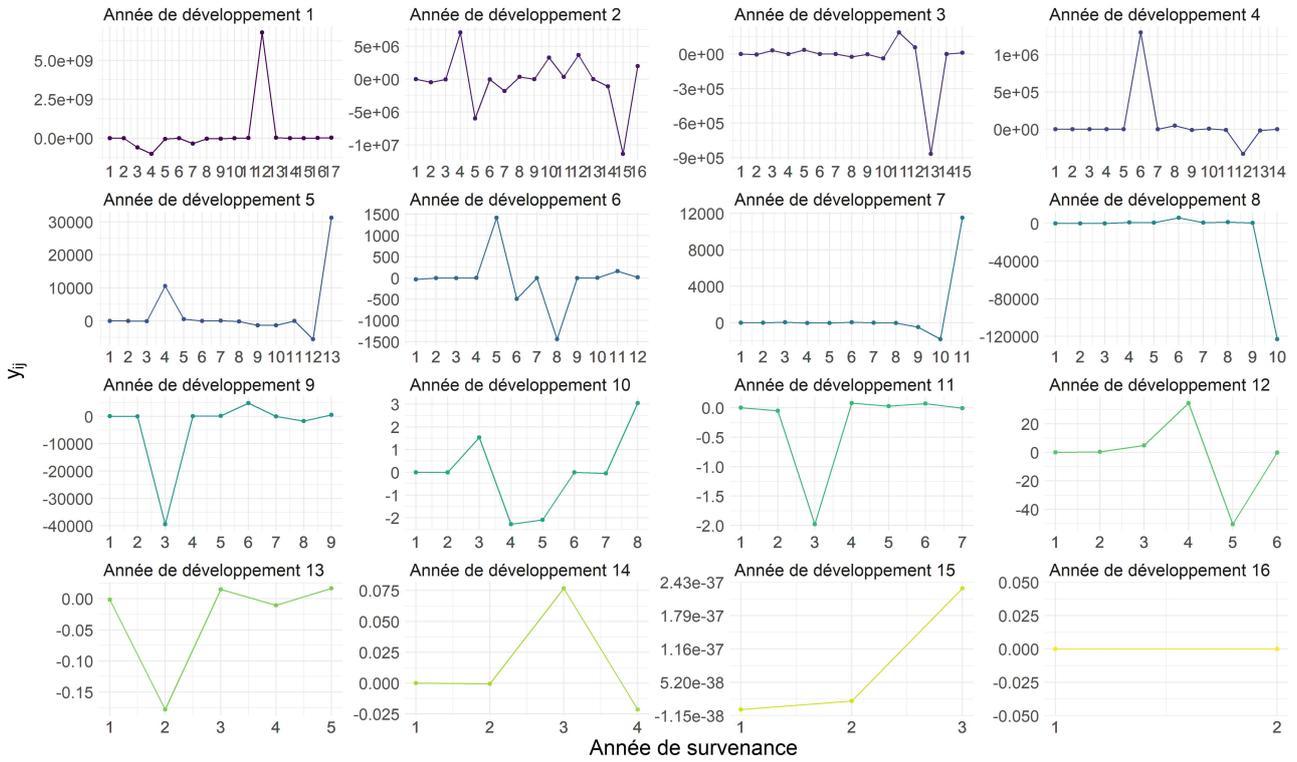


FIGURE D.4 – Vérification de l'Hypothèse 4

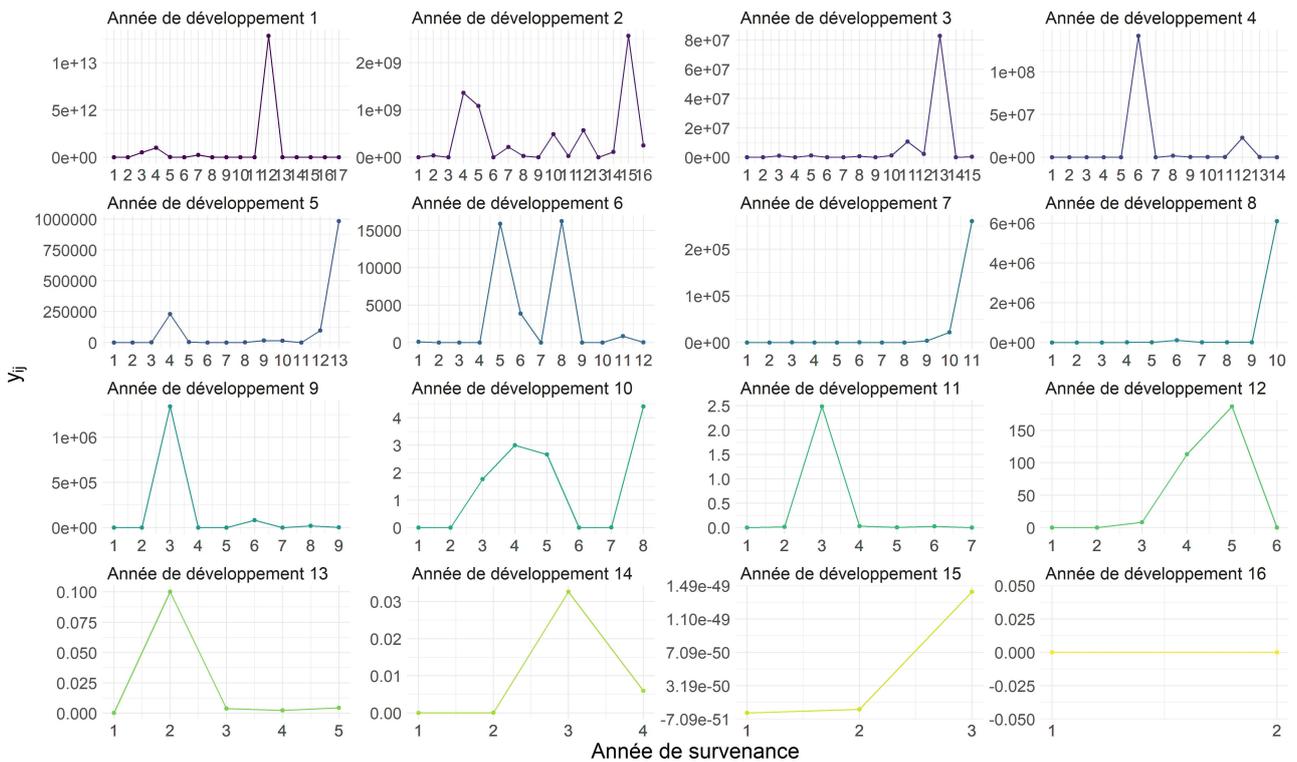


FIGURE D.5 – Vérification de l'Hypothèse 6

D.2 Triangle agricole responsabilité civile

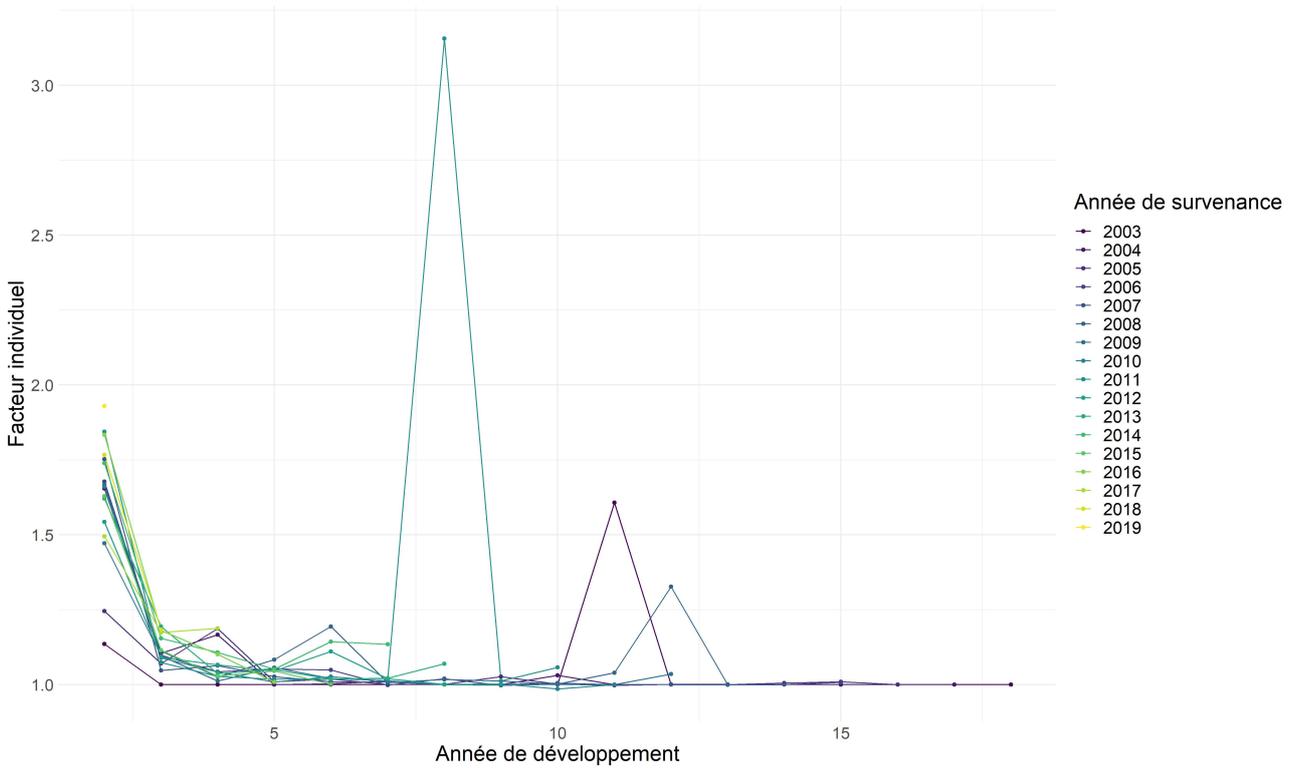


FIGURE D.6 – Vérification de l’Hypothèse 1

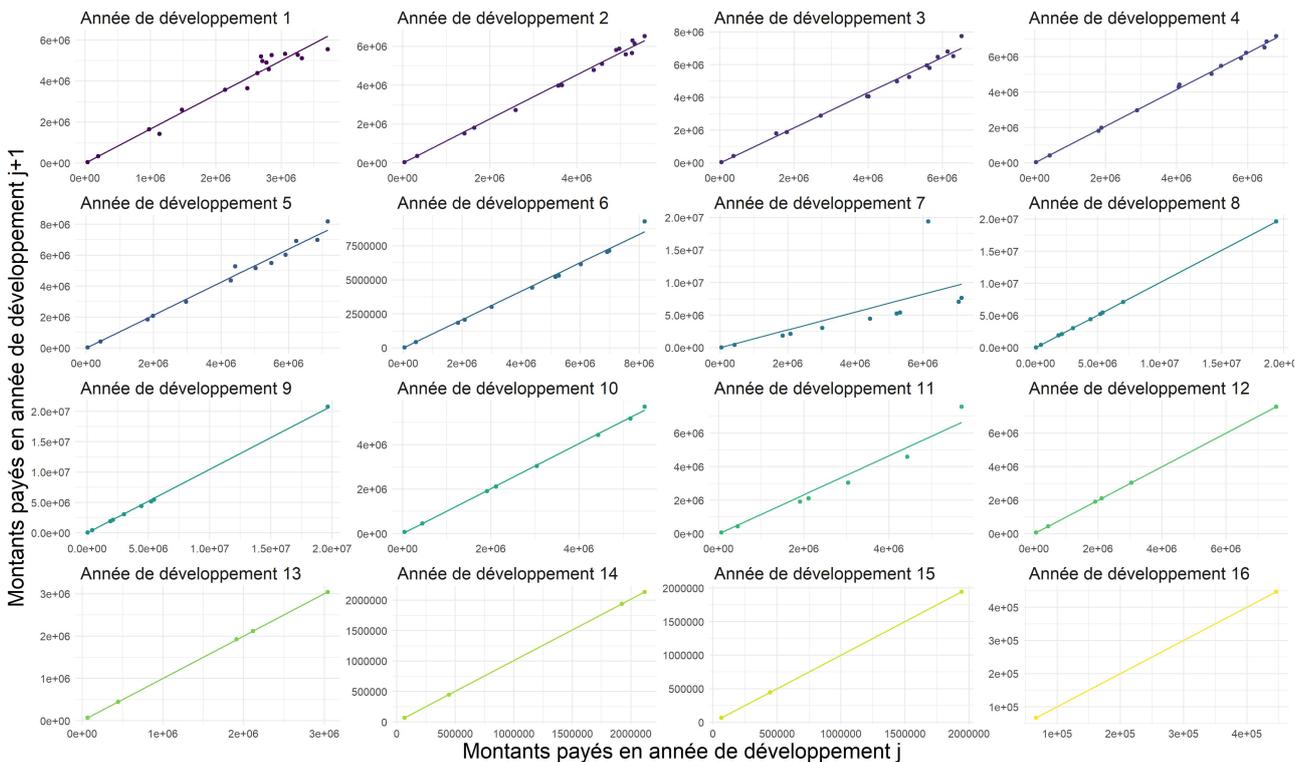


FIGURE D.7 – Vérification de l’Hypothèse 2

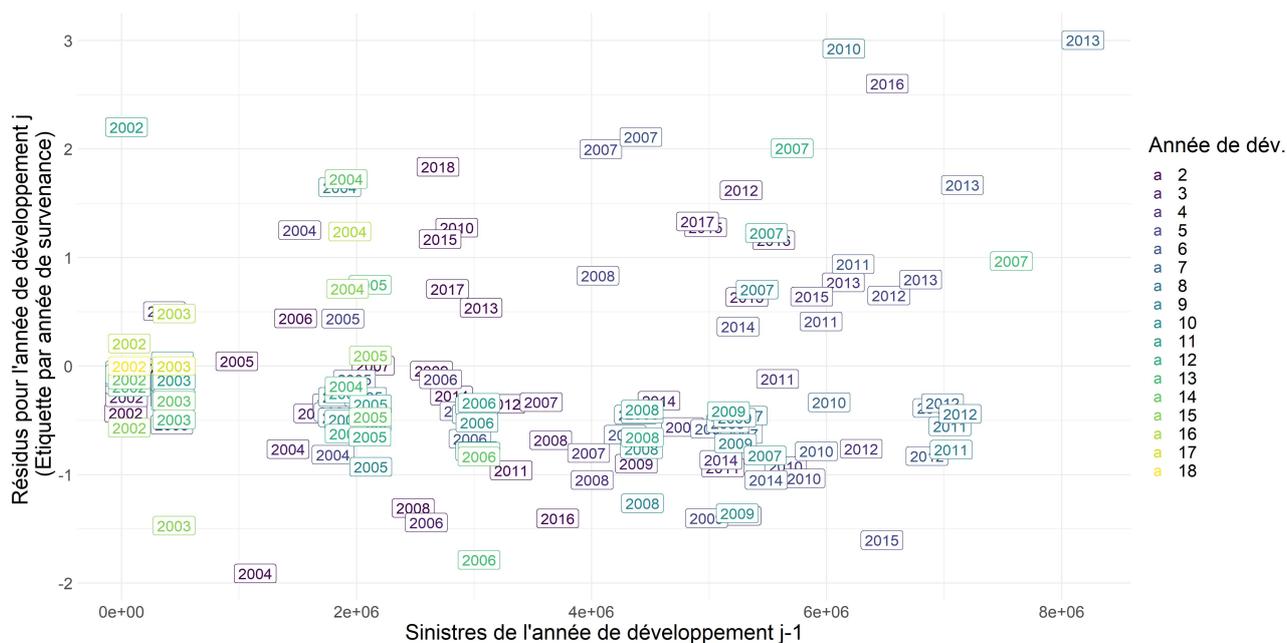


FIGURE D.8 – Vérification de l’Hypothèse 3

Tendance étudiée	p-value
Année de survenance	0,414
Année calendaire	0,734
Année de développement	0,206

TABLE D.2 – Tests de Box-Pierce

La Figure D.8 précédente et les tests de Box-Pierce nous permettent d’accepter l’hypothèse de blancheur des résidus, et donc d’indépendance.

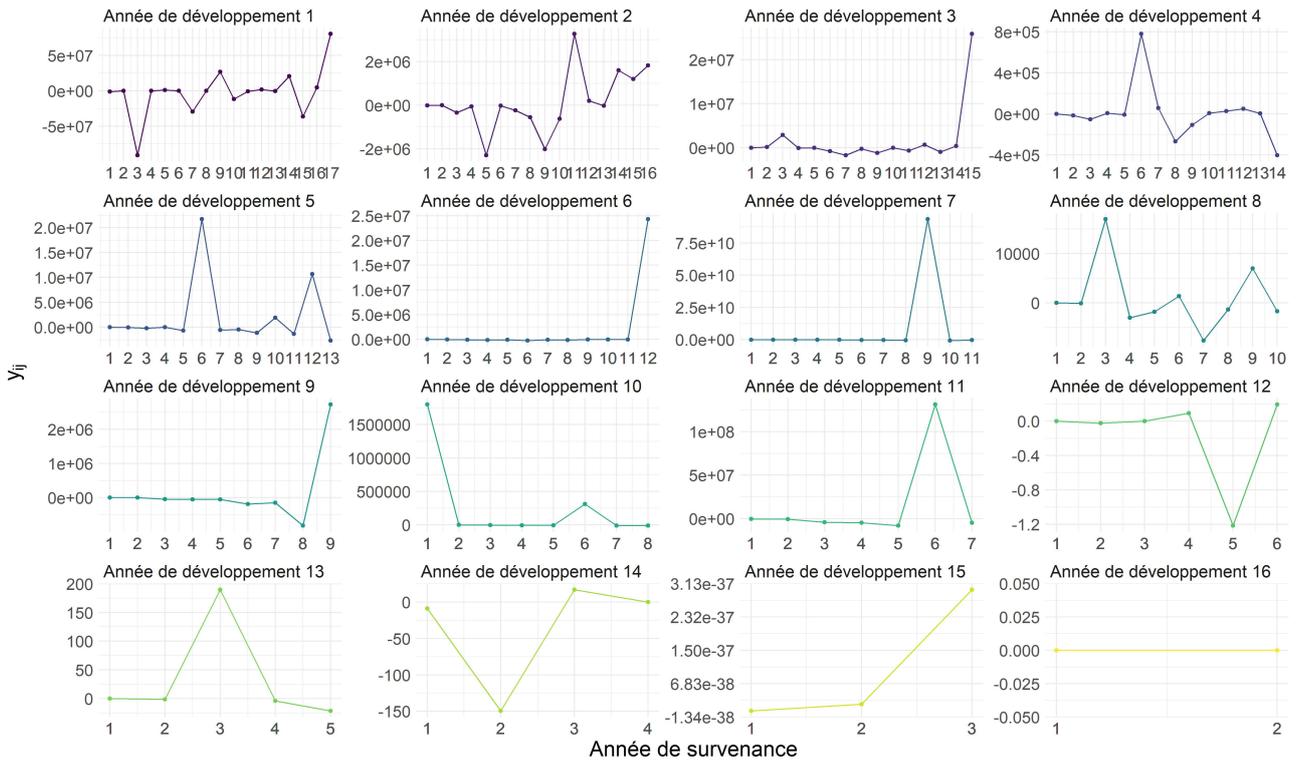


FIGURE D.9 – Vérification de l'Hypothèse 4

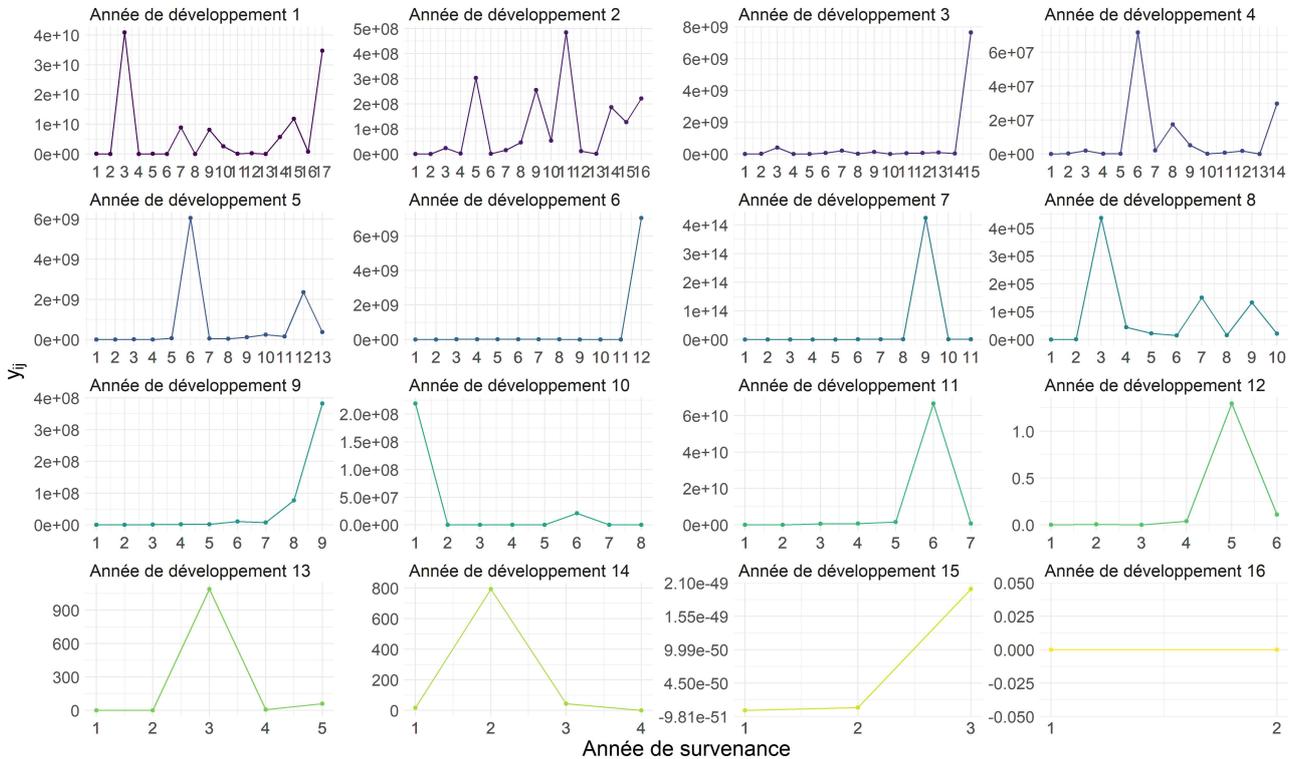


FIGURE D.10 – Vérification de l'Hypothèse 6

Annexe E

Modèles de lien entre tendance climatique et indice d'humidité des sols

E.0.1 Modèles simples sans année

	Estimate	Std. Error	t value	Pr(> t)
Constante	-680,7	873,9	-0,779	0,447
SWI _i	1285	968,7	1,327	0,203

R² ajusté : 0,0428

TABLE E.1 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{SWI}_i$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-92,11	787,8	-0,117	0,908
SWI _{i-1}	617,8	855,2	0,722	0,480

R² ajusté : -0,0289

TABLE E.2 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{SWI}_{i-1}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	1593	765,6	2,082	0,0538
SWI _{i-2}	-1227	833,5	-1,472	0,1603

R² ajusté : 0,0643

TABLE E.3 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{SWI}_{i-2}$

E.1 Modèles simples

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,141e+05	1,826e+04	-6.246	1,57e-05
Année _i	56,74	9,136	6,211	1,67e-05
SWI _i	505,6	544,1	0,929	0,367

R² ajusté : 0,7141

TABLE E.4 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma \times \text{SWI}_i$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,172e+05	1,768e+04	-6,625	8,08e-06
Année _i	58,28	8,802	6,621	8,13e-06
SWI _{i-1}	490,7	446,3	1,099	0,289

R² ajusté : 0,7202

TABLE E.5 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma \times \text{SWI}_{i-1}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,125e+05	1,660e+04	-6,776	6,25e-06
Année _i	56,55	8,227	6,873	5,29e-06
SWI _{i-2}	-836,9	426,4	-1,963	0,0685

R² ajusté : 0,7595

TABLE E.6 – Résultats du modèle de régression linéaire $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma \times \text{SWI}_{i-2}$

E.2 Modèles avec inclusion de retards

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,121e+05	1,771e+04	-6,332	1,85e-05
Année _i	59,39	8,883	6,235	2,18e-05
SWI _i	712,2	545,5	1,306	0,213
SWI _{i-1}	647,1	452,4	1,430	0,175

R² ajusté : 0,7328

TABLE E.7 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_i + \gamma_2 \times \text{SWI}_{i-1}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,121e+05	1,771e+04	-6,332	1,85e-05
Année _i	55,39	8,581	6,454	1,51e-05
SWI _i	335,3	518,1	0,647	0,528
SWI _{i-2}	-783,8	442,6	-1,771	0,098

R² ajusté : 0,7498

TABLE E.8 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_i + \gamma_2 \times \text{SWI}_{i-2}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,123e+05	1,654e+04	-6,791	8,73e-06
Année _i	56,26	8,202	6,859	7,81e-06
SWI _{i-1}	435,3	413,4	1,053	0,310
SWI _{i-2}	-805,2	425,9	-1,891	0,0796

R² ajusté : 0,7612

TABLE E.9 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_{i-1} + \gamma_2 \times \text{SWI}_{i-2}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,091e+05	1,634e+04	-6,482	2,06e-05
Année _i	54,35	8,42	6,454	2,15e-05
SWI _i	528,87	527,66	1,002	0,334
SWI _{i-1}	557,80	431,02	1,294	0,218
SWI _{i-2}	-712,32	435,78	-1,635	0,126

R² ajusté : 0,7613

TABLE E.10 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_0 \times \text{SWI}_i + \gamma_1 \times \text{SWI}_{i-1} + \gamma_2 \times \text{SWI}_{i-2}$

E.3 Modèles avec effet de seuil

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,040e+05	2,002e+04	-5,194	1,36e-04
Année _i	52,02	9,91	5,249	1,23e-04
SWI _i × 1 _{SWI_i>1}	80,75	651,1	0,124	0,903
SWI _i × 1 _{SWI_i≤1}	-137,7	773,1	-0,178	0,861

R² ajusté : 0,7205

TABLE E.11 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_i \times \mathbb{1}_{\{\text{SWI}_i > 1\}} + \gamma_2 \times \text{SWI}_i \times \mathbb{1}_{\{\text{SWI}_i \leq 1\}}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,205e+05	1,932e+04	-6,235	2,18e-05
Année _i	59,80	9,529	6,276	2,04e-05
SWI _{i-1} × 1 _{{SWI_{i-1}>1}}	716,1	642,2	1,115	0,284
SWI _{i-1} × 1 _{{SWI_{i-1}≤1}}	808,7	783,2	1,033	0,319

R² ajusté : 0,7055

TABLE E.12 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_{i-1} \times \mathbb{1}_{\{\text{SWI}_{i-1} > 1\}} + \gamma_2 \times \text{SWI}_{i-1} \times \mathbb{1}_{\{\text{SWI}_{i-1} \leq 1\}}$

	Estimate	Std. Error	t value	Pr(> t)
Constante	-1,155e+05	1,805e+04	-6,403	1,65e-05
Année _i	57,97	8,875	6,532	1,33e-05
SWI _{i-2} × 1 _{{SWI_{i-2}>1}}	647,9	570,7	-1,135	0,275
SWI _{i-2} × 1 _{{SWI_{i-2}≤1}}	-559,3	693,8	-0,806	0,434

R² ajusté : 0,7471

TABLE E.13 – Résultats du modèle de régression linéaire
 $\lambda_i = \alpha + \beta \times \text{Année}_i + \gamma_1 \times \text{SWI}_{i-2} \times \mathbb{1}_{\{\text{SWI}_{i-2} > 1\}} + \gamma_2 \times \text{SWI}_{i-2} \times \mathbb{1}_{\{\text{SWI}_{i-2} \leq 1\}}$

Annexe F

Modèles de lois de dépassements retenus

F.1 Modèle M2 pour les données historiques

	Estimate	Std. Error	t value	Pr(> t)
Culture - Blé	-0,158	0,036	-4,42	1,75e-05
Culture - Colza grain et navette	0,401	0,040	10,01	<2e-16
Culture - Féveroles et fèves	-0,142	0,032	-4,463	1,44e-05
Culture - Maïs grain et maïs semence	0,756	0,046	16,55	<2e-16
Culture - Orge et escourgeon	0,215	0,056	3,847	1,67e-04
Culture - Pois protéagineux	-0,132	0,032	-4,136	5,46e-05
Culture - Seigle et méteil	-0,147	0,032	-4,555	9,77e-06
Culture - Triticale	-0,119	0,035	-3,399	8,36e-04
Région - Auvergne-Rhône-Alpes	29,31	3,540	8,280	2,99e-14
Région - Bourgogne-Franche-Comté	29,57	3,535	8,366	1,78e-14
Région - Bretagne	29,85	3,525	8,468	9,58e-15
Région - Centre-Val de Loire	29,32	3,535	8,294	2,75e-14
Région - Corse	30,22	3,524	8,573	5,02e-15
Région - Grand Est	29,23	3,529	8,283	2,94e-14
Région - Hauts-de-France	29,32	3,532	8,303	2,61e-14
Région - Île-de-France	29,25	3,533	8,280	3,00e-14
Région - Normandie	29,91	3,528	8,479	8,95e-15
Région - Nouvelle-Aquitaine	30,11	3,527	8,537	6,29e-15
Région - Occitanie	30,01	3,533	8,494	8,16e-15

	Estimate	Std. Error	t value	Pr(> t)
Région - Pays de la Loire	30,19	3,533	8,544	6,00e-15
Région - Provence-Alpes-Côte d'Azur	29,93	3,518	8,506	7,57e-15
Année	-0,015	0,002	-8,469	9,49e-15

R^2 ajusté : 0,973

TABLE F.1 – Modèle **M2** retenu pour ξ avec les données historiques

	Estimate	Std. Error	t value	Pr(> t)
Région - Auvergne-Rhône-Alpes	-3,933	0,460	-8,549	4,59e-15
Région - Bourgogne-Franche-Comté	-1,893	0,224	-8,463	7,82e-15
Région - Bretagne	-2,923	0,826	-3,539	5,07e-04
Région - Centre-Val de Loire	-2,808	0,122	-22,99	<2e-16
Région - Corse	-2,579	1,063	-2,426	0,016
Région - Grand Est	-3,809	0,193	-19,76	<2e-16
Région - Hauts-de-France	-3,268	0,254	-12,85	<2e-16
Région - Île-de-France	-2,995	0,212	-14,14	<2e-16
Région - Normandie	-2,873	0,455	-6,311	2,00e-09
Région - Nouvelle-Aquitaine	-2,960	0,659	-4,492	1,24e-05
Région - Occitanie	-2,533	0,611	-4,147	5,13e-05
Région - Pays de la Loire	-2,729	0,557	-4,897	2,12e-06
Région - Provence-Alpes-Côte d'Azur	-1,805	0,605	-2,984	0,003

R^2 ajusté : 0,230

TABLE F.2 – Modèle **M2** retenu pour ν avec les données historiques

F.2 Modèle M3 pour les données historiques

	Estimate	Std. Error	t value	Pr(> t)
Culture - Blé	-0,111	0,030	-3,751	2,39e-04
Culture - Colza grain et navette	0,382	0,034	11,32	< 2e-16
Culture - Féveroles et fèves	-0,148	0,028	-5,335	2,93e-07
Culture - Maïs grain et maïs semence	0,351	0,042	8,370	1,78e-14
Culture - Orge et escourgeon	0,116	0,043	2,678	8,12e-03
Culture - Pois protéagineux	-0,156	0,028	-5,512	1,25e-07
Culture - Seigle et méteil	-0,098	0,029	-3,350	9,90e-04
Culture - Triticale	-0,129	0,029	-4,405	1,84e-05
Région - Auvergne-Rhône-Alpes	0,709	0,103	6,858	1,16e-10
Région - Bourgogne-Franche-Comté	1,027	0,108	9,518	< 2e-16
Région - Bretagne	1,437	0,090	15,90	< 2e-16
Région - Centre-Val de Loire	0,724	0,098	7,401	5,44e-12
Région - Corse	1,869	0,107	17,54	< 2e-16
Région - Grand Est	0,854	0,099	8,624	3,80e-15
Région - Hauts-de-France	0,849	0,094	9,001	3,69e-16
Région - Île-de-France	0,623	0,088	7,051	3,94e-11
Région - Normandie	1,547	0,102	15,23	< 2e-16
Région - Nouvelle-Aquitaine	1,475	0,097	15,20	< 2e-16
Région - Occitanie	1,420	0,108	13,12	< 2e-16
Région - Pays de la Loire	1,547	0,090	17,20	< 2e-16
Région - Provence-Alpes-Côte d'Azur	1,474	0,093	15,84	< 2e-16
SWI _t	-1,098	0,062	-17,64	< 2e-16
SWI _{t₂}	-0,458	0,071	-6,469	9,56e-10

R² ajusté : 0,983

TABLE F.3 – Modèle **M3** retenu pour ξ avec les données historiques

	Estimate	Std. Error	t value	Pr(> t)
Région - Auvergne-Rhône-Alpes	-4,108	0,330	-12,43	< 2e-16
Région - Bourgogne-Franche-Comté	-2,311	0,200	-11,56	< 2e-16
Région - Bretagne	-2,904	0,814	-3,566	4,61e-04
Région - Centre-Val de Loire	-3,570	0,196	-18,26	< 2e-16
Région - Corse	-2,447	0,945	-2,589	0,010
Région - Grand Est	-3,136	0,131	-23,99	< 2e-16
Région - Hauts-de-France	-2,883	0,229	-12,58	< 2e-16
Région - Île-de-France	-3,030	0,196	-15,49	< 2e-16
Région - Normandie	-2,875	0,477	-6,024	9,00e-09
Région - Nouvelle-Aquitaine	-3,136	0,585	-5,360	2,45e-07
Région - Occitanie	-2,517	0,403	-6,248	2,79e-09
Région - Pays de la Loire	-2,750	0,516	-5,331	2,83e-07
Région - Provence-Alpes-Côte d'Azur	-1,788	0,581	-3,077	0,002

R² ajusté : 0,192

TABLE F.4 – Modèle **M3** retenu pour ν avec les données historiques

F.3 Modèle M2 pour les données mises en *as if*

	Estimate	Std. Error	t value	Pr(> t)
Culture - Blé	-0,146	0,038	-3,839	1,62e-04
Culture - Colza grain et navette	0,277	0,045	6,181	3,10e-09
Culture - Féveroles et fèves	-0,034	0,038	-0,896	0,371
Culture - Maïs grain et maïs semence	0,223	0,037	6,070	5,62e-09
Culture - Orge et escourgeon	0,182	0,035	5,253	3,55e-07
Culture - Pois protéagineux	-0,218	0,039	-5,629	5,56e-08
Culture - Seigle et méteil	-0,148	0,037	-4,017	8,12e-05
Culture - Triticale	-0,186	0,035	-5,292	2,94e-07
Région - Auvergne-Rhône-Alpes	-13,58	3,588	-3,786	1,98e-04
Région - Bourgogne-Franche-Comté	-13,31	3,593	-3,703	2,70e-04
Région - Bretagne	-13,43	3,588	-3,743	2,32e-04
Région - Centre-Val de Loire	-13,63	3,593	-3,793	1,92e-04
Région - Corse	-12,52	3,575	-3,502	5,60e-04
Région - Grand Est	-13,80	3,593	-3,841	1,61e-04
Région - Hauts-de-France	-13,82	3,599	-3,840	1,61e-04
Région - Île-de-France	-13,88	3,599	-3,857	1,51e-04
Région - Normandie	-12,96	3,593	-3,608	3,82e-04
Région - Nouvelle-Aquitaine	-13,81	3,585	-3,851	1,54e-04
Région - Occitanie	-13,08	3,580	-3,652	3,25e-04
Région - Pays de la Loire	-13,32	3,585	-3,714	2,59e-04
Région - Provence-Alpes-Côte d'Azur	-12,89	3,573	-3,607	3,84e-04
Année	0,007	0,002	3,669	3,07e-04

R² ajusté : 0,977

TABLE F.5 – Modèle M2 retenu pour ξ avec les données mises en *as if*

	Estimate	Std. Error	t value	Pr(> t)
Région - Auvergne-Rhône-Alpes	-2,828	0,274	-10,324	<2e-16
Région - Bourgogne-Franche-Comté	-1,960	0,336	-5,834	1,85e-08
Région - Bretagne	-2,484	0,466	-5,335	2,30e-07
Région - Centre-Val de Loire	-2,878	0,222	-12,97	<2e-16
Région - Corse	-2,334	0,383	-6,096	4,63e-09
Région - Grand Est	-3,380	0,286	-11,82	<2e-16
Région - Hauts-de-France	-3,273	0,183	-17,86	<2e-16
Région - Île-de-France	-3,086	0,198	-15,61	<2e-16
Région - Normandie	-3,253	0,903	-3,601	3,89e-04
Région - Nouvelle-Aquitaine	-3,780	0,235	-16,09	<2e-16
Région - Occitanie	-2,637	0,390	-6,755	1,18e-10
Région - Pays de la Loire	-2,752	0,295	-9,323	<2e-16
Région - Provence-Alpes-Côte d'Azur	-1,809	0,456	-3,967	9,74e-05

R² ajusté : 0,208

TABLE F.6 – Modèle **M2** retenu pour ν avec les données mises en *as if*

F.4 Modèle M3 pour les données mises en *as if*

	Estimate	Std. Error	t value	Pr(> t)
Culture - Avoine	-25,15	3,658	-6,876	6,47e-11
Culture - Blé	-25,24	3,658	-6,899	5,65e-11
Culture - Colza grain et navette	-24,78	3,641	-6,806	9,68e-11
Culture - Féveroles et fèves	-25,09	3,666	-6,843	7,80e-11
Culture - Maïs grain et maïs semence	-25,01	3,655	-6,843	7,79e-11
Culture - Orge et escourgeon	-25,10	3,662	-6,855	7,28e-11
Culture - Pois protéagineux	-25,28	3,660	-6,907	5,39e-11
Culture - Seigle et méteil	-25,23	3,657	-6,900	5,62e-11
Culture - Triticale	-25,27	3,657	-6,910	5,30e-11
Région - Bourgogne-Franche-Comté	0,241	0,048	5,022	1,07e-06
Région - Bretagne	0,265	0,047	5,588	6,84e-08
Région - Centre-Val de Loire	0,064	0,054	1,174	0,242
Région - Corse	1,168	0,039	29,82	< 2e-16
Région - Grand Est	-0,179	0,042	-4,275	2,87e-05
Région - Hauts-de-France	-0,239	0,039	-6,111	4,54e-09
Région - Île-de-France	-0,304	0,038	-7,965	9,23e-14
Région - Normandie	0,507	0,056	9,004	< 2e-16
Région - Nouvelle-Aquitaine	0,589	0,058	10,17	< 2e-16
Région - Occitanie	0,523	0,041	12,80	< 2e-16
Région - Pays de la Loire	0,187	0,042	4,482	1,20e-05
Région - Provence-Alpes-Côte d'Azur	0,689	0,054	12,72	< 2e-16
Année	0,013	0,002	7,009	2,99e-11
SWI _t	-1,061	0,087	-12,23	< 2e-16

R² ajusté : 0,979TABLE F.7 – Modèle M3 retenu pour ξ avec les données mises en *as if*

	Estimate	Std. Error	t value	Pr(> t)
Région - Auvergne-Rhône-Alpes	-2,855	0,291	-9,829	< 2e-16
Région - Bourgogne-Franche-Comté	-2,044	0,259	-7,898	1,21e-13
Région - Bretagne	-2,440	0,448	-5,451	1,30e-07
Région - Centre-Val de Loire	-2,825	0,229	-12,31	< 2e-16
Région - Corse	-2,388	0,367	-6,506	4,88e-10
Région - Grand Est	-3,900	0,242	-16,12	< 2e-16
Région - Hauts-de-France	-3,491	0,224	-15,56	< 2e-16
Région - Île-de-France	-3,184	0,171	-18,65	< 2e-16
Région - Normandie	-3,344	0,809	-4,134	5,01e-05
Région - Nouvelle-Aquitaine	-3,147	0,632	-4,979	1,26e-06
Région - Occitanie	-2,652	0,375	-7,066	1,94e-11
Région - Pays de la Loire	-2,909	0,238	-12,22	< 2e-16
Région - Provence-Alpes-Côte d'Azur	-1,832	0,446	-4,113	5,46e-05

R² ajusté : 0,216

TABLE F.8 – Modèle **M3** retenu pour ν avec les données mises en *as if*