

Mémoire présenté le : 04/07/2023

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : EMERY Quentin

Etude de la déviation d'un portefeuille MRH face au risque climatique d'inondation.

Confidentialité : NON (Durée : 1 an 2 ans)

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

*Membres présents du jury de Signature
l'Institut des Actuaires*

C. Vigouroux
.....

M. Juillard
.....

*Membres présents du jury de
l'ISFA*

A. Eyraud
.....

Entreprise :

Nom : SeaBird Conseil

*Directeur de mémoire en entre-
prise :*

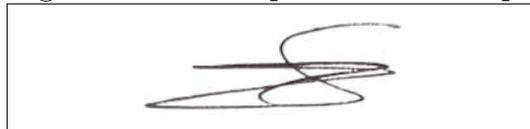
Nom : SILVA Benjamin

Signature :

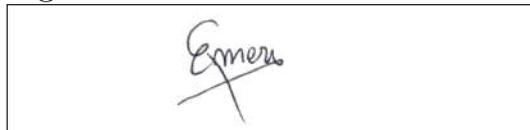


***Autorisation de publication et
de mise en ligne sur un site de
diffusion de documents actua-
riels (après expiration de l'éventuel
délai de confidentialité)***

Signature du responsable entreprise



Signature du candidat



Résumé

La hausse récente du nombre d'événements climatiques d'inondation affectant les bilans des assureurs au passif couplée à l'émergence de normes et réglementations au sujet de l'évolution du risque de catastrophes naturelles poussent les assureurs à vouloir anticiper l'évolution du risque inondation liée au changement climatique.

Néanmoins, la validité des modèles basés sur les seuls historiques de pertes est mise à mal, comment mesurer non plus un risque bien connu mais son évolution ? Comment mesurer la rupture entre un environnement maîtrisé passé et une période de doutes future ? C'est dans ce contexte là que le mémoire présentera les problématiques de modélisation de ce que l'on peut appeler le risque climatique dans le but de donner via une approche exploratoire des méthodes de valorisation du risque climatique d'inondation tout en essayant de donner des clés d'analyses utiles à la gestion du risque et à la réponse aux enjeux réglementaires à venir.

De manière concrète, les méthodes de valorisation envisagées s'appuient sur l'utilisation de classification, afin de rendre compte d'une évolution du risque de catastrophe naturelle éparse sur les territoires, sur la prédiction grâce à la mise en contact de variables climatiques open-source avec les pertes historiques via des algorithmes d'apprentissage supervisé d'un côté et via des statistiques semi-paramétriques de l'autre et enfin sur la construction d'indices propres à l'appréhension du risque climatique pouvant servir dans la construction de scénarios économique-climatiques.

Mots-clés : risque climatique physique, modèle semi-paramétrique, clustering, apprentissage supervisé, scénario climatique, métriques climatiques, inondation.

Abstract

The recent increase in the number of flooding events affecting insurers' balance sheets on the liabilities side, coupled with the emergence of standards and regulations concerning the evolution of natural catastrophe risk, are pushing insurers to try to anticipate the evolution of flooding risk linked to climate change.

Nevertheless, the validity of models based on historical losses alone is being challenged: how to measure not a well-known risk but its evolution? How to measure the break between a past controlled environment and a future period of doubt? It is in this context that the thesis presented will address the problems of modelling what can be called climate risk with the aim of providing, via an exploratory approach, methods of valuing climate risk while trying to provide useful analysis keys for risk management and for responding to future regulatory issues.

In concrete terms, the valuation methods envisaged are based on the use of classification, in order to account for an evolution of the risk of natural disasters scattered over the territories, on prediction thanks to the contact of open-source climatic variables with historical losses and supervised learning algorithms on the one hand, and semi-parametric statistics on the other, and finally on the construction of indices specific to the understanding of climatic risk that can be used in the construction of economic climatic scenarios.

Keywords : climate risk, clustering, supervised machine learning semi-parametric model, climate scenario, climate metrics

Note de synthèse

Quand nous parlons de risque climatique, nous parlons en réalité d'une diversité de risques. En témoigne le graphique construit ci-après.

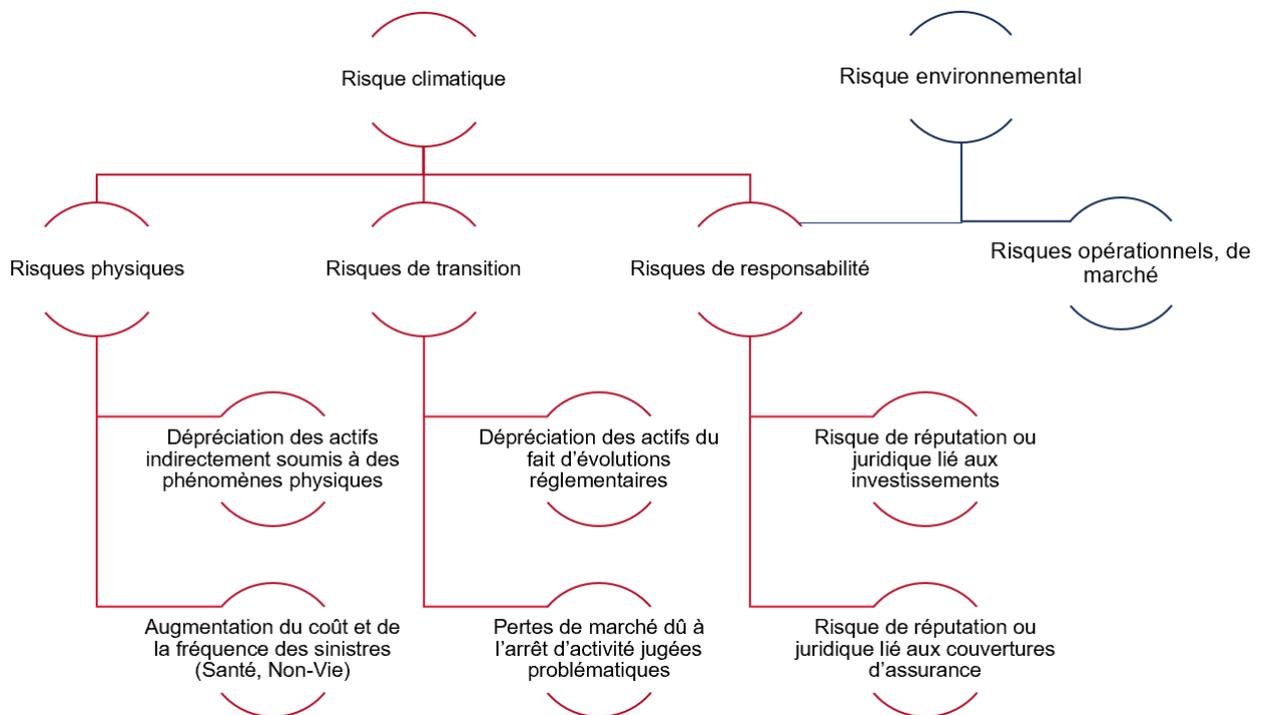


Figure 0-1- Organigramme définitionnel du risque climatique et environnemental

La pluralité des risques affichées nous contraint de nous concentrer sur un risque plus précis, ainsi le risque étudié est le risque d'augmentation en fréquence et en sévérité du risque inondation au regard du changement climatique.

Dès lors, il s'agit de contextualiser l'étude menée. L'évaluation du risque climatique de catastrophes naturelles se place en effet dans un contexte de fortes évolutions normatives et de pilotage du risque. On peut notamment citer les évolutions attendues ou effectives suivantes :

- L'inclusion du risque climatique de catastrophes naturelles dans la norme européenne Solvabilité II à travers :
 - La potentielle évolution du calcul standard du SCR
 - L'inclusion de scénarios climatiques dans le processus ORSA
- La mise en place de stress-tests climatiques sur le bilan des assureurs organisée par le régulateur (l'ACPR)
- L'augmentation des primes d'assurances et de réassurance lié à la déviation des portefeuilles lié au risque climatique

Ainsi, l'étude menée a pour but d'éclairer l'évolution du risque d'inondation sur un portefeuille MRH et de donner des pistes de réponses aux défis normatifs et de pilotage de l'activité. Pour cela, on se fixe d'estimer les charges annuelles à une échelle départementale et un pas annuel sur la période 2020-2050 à comparer avec la période de référence 1998-2020.

Pour ce faire, nous utiliserons les données de sinistralités relatives au portefeuille MRH étudié et les données climatiques produites par le GIEC fournies sous la forme de 3 scénarios :

- Scénario le plus extrême : RCP8.5
- Scénario à « +2° » moyen : RCP4.5
- Scénario positif : RCP2.6

Le traitement des bases de données et l'analyse des données nous permet d'identifier les variables climatiques les plus explicatives :

- Le pourcentage de pluies intenses pour la fréquence
- Le nombre de jours de pluies à plus d'1MM et le nombre de jours de sécheresses pour la sévérité

En outre, une forte dispersion et un poids important de fréquence nulles présentes dans les données nous incite à utiliser un modèle bimodal. Il prend la forme d'une Random Forest pour la prédiction des départements sinistrés à l'avenir selon les données climatiques et d'un modèle GLM pénalisé de réponse binomiale négative pour gérer la surdispersion.

Notez que les deux pans du modèle de fréquence sont par nature adaptés à la forte dimensionalité des variables climatiques, la sélection des variables effectuées par ces modèles est d'ailleurs en adéquation avec la corrélation étudiée. Par suite, on établit des résultats de validation satisfaisants.

	Département prédit sinistré	Département prédit sauf
Département sinistré	358	4
Département sauf	60	19

F1 - Score :	91,8%
---------------------	--------------

Figure 0-2- Validation du modèle de Random Forest

On parvient à retrouver une bonne proportion de départements sinistrés, en revanche, l'erreur est relativement importante quant au nombre de départements prédit sinistrés alors qu'ils ne le sont pas.

Ceci, combiné à un modèle GLM peu dispersée (RMSE relatif fort, presque 60%), nous donne un résultat moyen de 16,2 sinistres par départements par année contre un historique à 14,1 sinistres. Ainsi, on établit une marge d'erreur native au modèle de 20% mais conservons le modèle. Par ailleurs, nous ne notons pas de différence majeure de qualité du modèle entre les différents scénarios climatiques.

Du côté de la sévérité, nous adoptons également un modèle bimodal. Nous traitons les attritionnels par régression non-paramétrique et les sinistres graves par régression paramétrique. La source de ce split provient du fait que le modèle non-paramétrique indépendant génère moins d'erreur mais ne sera pas robuste face à un changement d'échelle des données climatiques puisque puisant directement dans l'historique (13% des enregistrements climatiques dépassent leurs extremum sur la période de référence dans la période 2020-2050).

On aboutit à un modèle raisonnable en termes de prédiction moyenne :

Rapports de moyennes	Rapports d'écart-type	RMSE	MAE	ME
93%	68%	11 002 €	6 134 €	584 563 €

Figure 0-3- Validation du modèle de sévérité

Il faudra garder en tête une marge d'erreur de 7% minimum pour ce modèle.

Ainsi le modèle sera utilisé pour prédire des charges de sinistralité sur la période 2020-2050.

On procède ensuite à un bootstrap pour combiner les modèles de fréquences et de sévérité car la régression non-paramétrique ne nous permet pas la création de formules fermées.

Finalement, en gardant en tête les réserves que l'on peut avoir sur les modèles créés, on aboutit à la création d'une décomposition de l'évolution du risque global sur le portefeuille étudié pour le risque inondation.

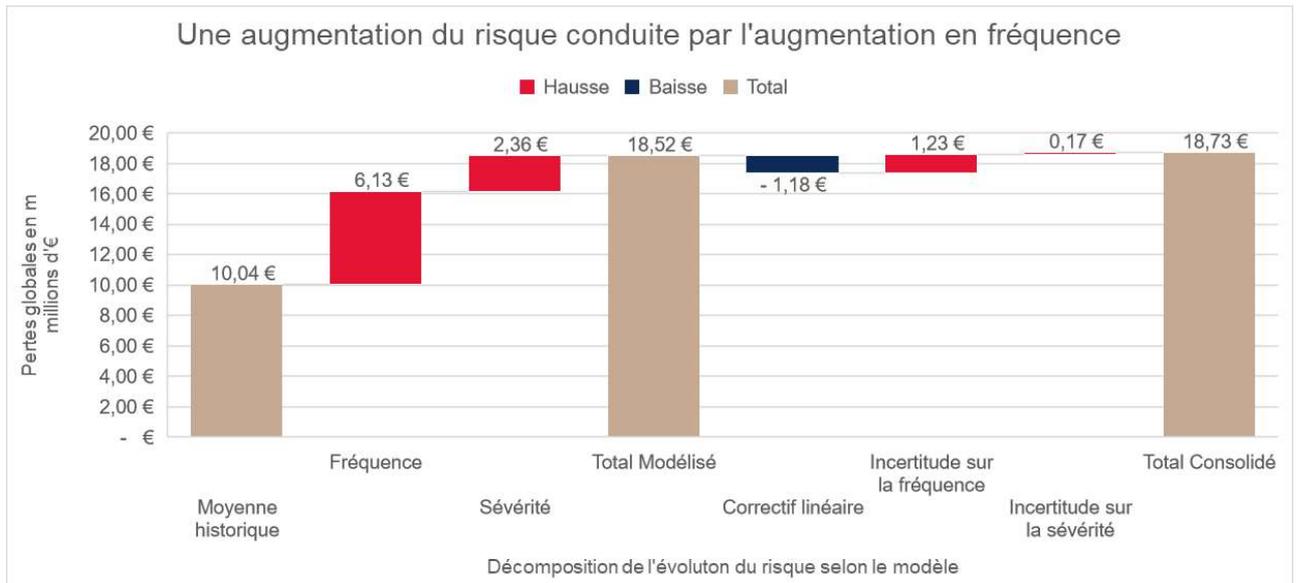


Figure 0-4- Décomposition de l'estimation d'augmentation du risque inondation

La première conclusion à tirer de cela est l'importance de la fréquence dans l'évolution du risque : la fréquence dans le scénario RCP8.5 est responsable pour 72% de l'augmentation de la charge totale du portefeuille en inondation. La seconde est l'importance de l'augmentation du risque avec un atterrissage à 18 millions d'€ soit une charge augmentant de 88%.

Par un calcul simple d'annualisation de taux, on estime par conséquent que cette augmentation estimée pourrait être à l'origine d'une augmentation annuelle de 2% des primes d'assurances et de réassurance hors effets exogènes.

En outre, nous n'estimons pas de changements majeurs dans le bassin géographique du risque inondation, la densité semble juste plus étendue mais concentrés sur les mêmes points.

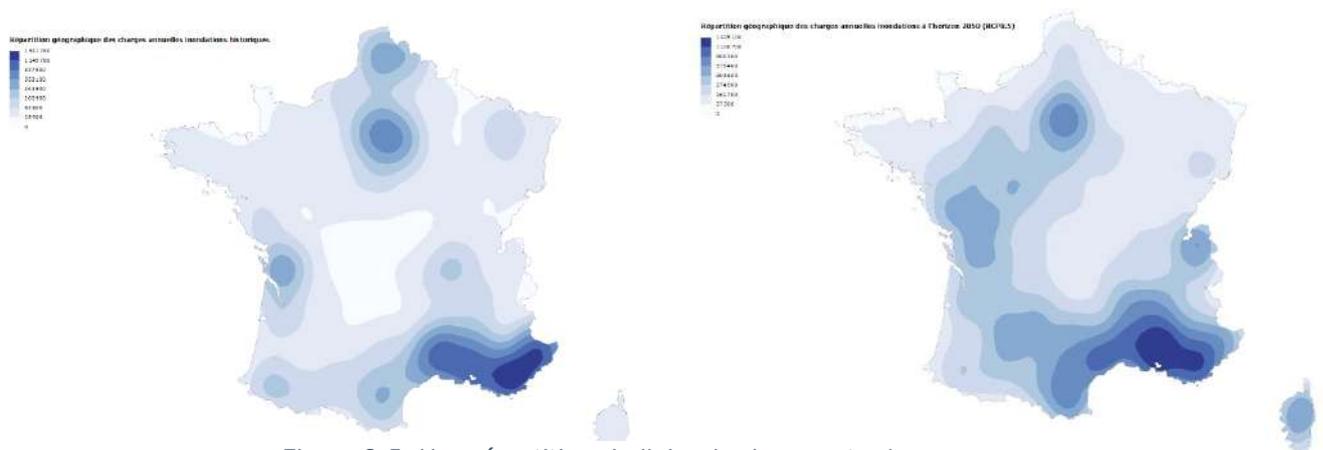


Figure 0-5- Une répartition similaire du risque entre les deux périodes

Finalement, on s'intéresse aux différences entre les différents scénarios climatiques au sujet de la vitesse de déviation du risque. Pour cela, une métrique qui pondère le risque évalué par les années d'estimation est créé. On trouve alors que le scénario le plus extrême implique un changement de risque 20% plus rapide que dans les autres scénarios.

Avant de conclure, il est bon de souligner la cohérence du modèle avec les modèles de marché en termes de résultats.

Acteur	↗ Inondation
FFA	11%
Covéa	120%
CCR x Météo France	38%
Moyenne	56%
Mémoire (brut)	84%
Mémoire (incidence)	57%

Figure 0-6-Quelques résultats de modèles

Pour conclure, ce qu'il faut retenir de ce mémoire est

- La difficile modélisation du risque climatique étant donné son aspect multifactorielle, la présence d'enregistrements de variables climatiques extrêmes jamais atteints et la source de données relativement faible sur les catastrophes naturelles.
- Une augmentation très forte du risque d'inondation à l'horizon 2050 avec des primes ajustées à hauteur de 2% par an.
- Le rôle d'un changement de scénario climatique pour la gestion de risque : passer d'un scénario climatique RCP8.5 à RCP4.5 permet par exemple de réduire de 20% la vitesse d'augmentation du risque.

Synthesis note

When we talk about climate risk, we are really talking about a diversity of risks. This is illustrated by the graph below.

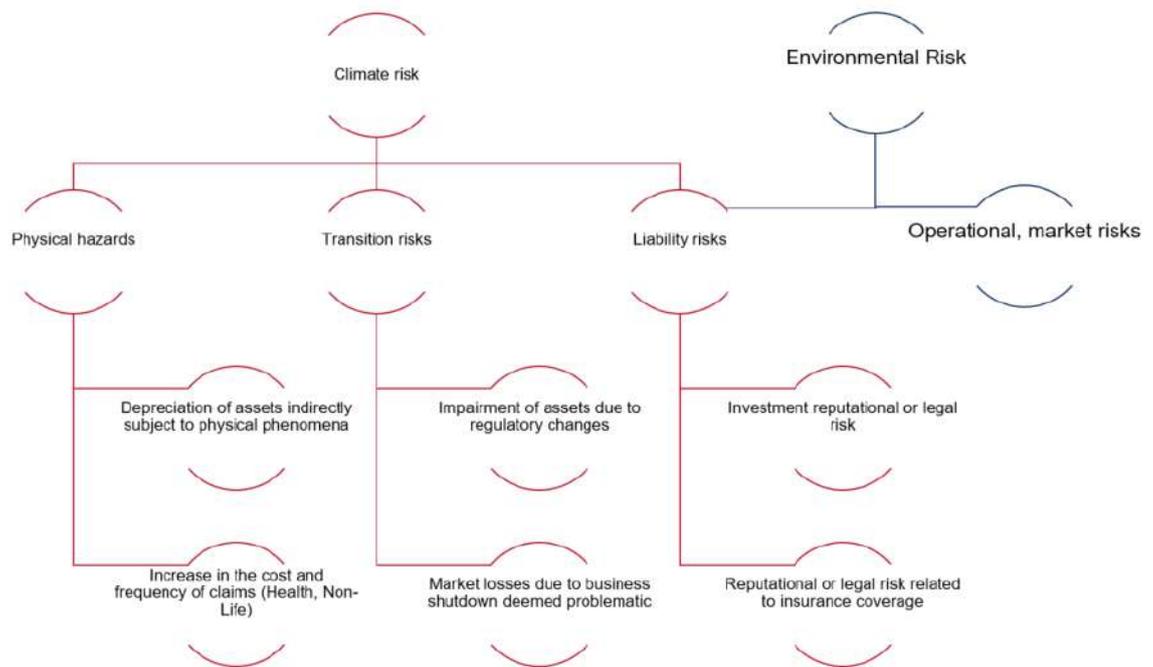


Figure 0-1- Definitional diagram of climate and environmental risk

The plurality of risks displayed forces us to focus on a more precise risk, so the risk studied is the risk of increasing the frequency and severity of the flood risk with regard to climate change.

Therefore, it is a question of contextualizing the study conducted. The assessment of the climate risk of natural disasters is indeed placed in a context of strong normative changes and risk management. These include the following expected or actual developments:

- The inclusion of the climate risk of natural disasters in the European Solvency II standard through:
 1. The potential evolution of the standard calculation of the SCR
 2. The inclusion of climate scenarios in the ORSA process
- The implementation of climate stress tests on insurers' balance sheets organised by the regulator (ACPR)
- The increase in insurance and reinsurance premiums linked to the deviation of portfolios linked to climate risk

Thus, the study conducted aims to shed light on the evolution of flood risk on an HRM portfolio and to provide answers to normative and business management challenges. To do this, we set ourselves to estimate the annual expenses at a departmental level and an annual step over the period 2020-2050 to be compared with the reference period 1998-2020.

To do this, we will use the claims data relating to the HRM portfolio studied and the climate data produced by the IPCC provided in the form of 3 scenarios:

- Most extreme scenario: RCP8.5
- Medium +2° scenario: RCP4.5
- Positive scenario: RCP2.6

Database processing and data analysis allows us to identify the most explanatory climate variables:

- The percentage of intense rainfall for frequency
- The number of rainy days at more than 1MM and the number of drought days for severity

In addition, a high dispersion and a high weight of zero frequencies present in the data encourages us to use a bimodal model. It takes the form of a Random Forest for predicting the affected departments in the future according to climate data and a penalized GLM model of negative binomial response to manage superdispersion.

Note that the two parts of the frequency model are by nature adapted to the high dimensionality of climatic variables, the selection of variables made by these models is also in line with the correlation studied. As a result, satisfactory validation results are established.

	Predicted as damaged	Predicted as not damaged
Actual damaged	358	4
Actual not damaged	60	19

F1 - Score :	91,8%
---------------------	--------------

Figure 0-2- Validation of Random Forest model

We manage to find a good proportion of departments affected, on the other hand, the error is relatively large as to the number of departments predicted to be affected when they are not.

This, combined with a poorly dispersed GLM model (high relative RMSE, almost 60%), gives us an average result of 16.2 claims per department per year against a history of 14.1 claims. Thus, we establish a margin of error native to the model of 20% but keep the model. In addition, we do not note any major difference in model quality between the different climate scenarios.

On the severity side, we are also adopting a bimodal model. We treat attritional by non-parametric regression and severe claims by parametric regression. The source of this split comes from the fact that the independent non-parametric model generates less error but will not be robust in the face of a change in scale of climate data since it draws directly from history (13% of climate records exceed their extremum over the reference period in the period 2020-2050).

We arrive at a reasonable model in terms of average prediction:

Mean relation	Standard deviation relation	RMSE	MAE	ME
93%	68%	11 002 €	6 134 €	584 200 €

Figure 0-3- Vulnerability model validation

It will be necessary to keep in mind a margin of error of 7% minimum for this model.

Thus, the model will be used to predict loss costs over the period 2020-2050.

We then perform a bootstrap to combine the frequency and severity models because non-parametric regression does not allow us to create closed formulas.

Finally, keeping in mind the reservations that may be had on the models created, we end up with the creation of a breakdown of the evolution of the overall risk on the portfolio studied for flood risk.

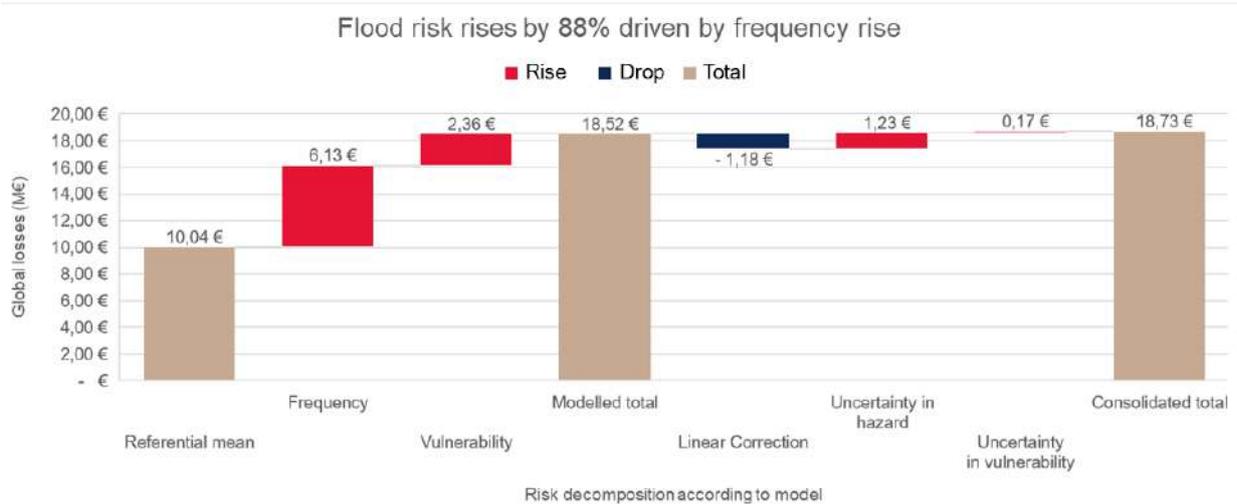


Figure 0-4- Flood risk rise decomposition

The first conclusion to be drawn from this is the importance of frequency in the evolution of risk: frequency in the RCP8.5 scenario is responsible for 72% of the increase in the total flood portfolio load. The second is the importance of the increase in risk with a landing at € 18 million, an increase of 88% charge.

By a simple calculation of annualization of rates, it is therefore estimated that this estimated increase could be the cause of an annual increase of 2% in insurance and reinsurance premiums excluding exogenous effects.

Besides, we do not estimate major changes in the geographical basin of flood risk, the density seems just more extensive but concentrated on the same points.

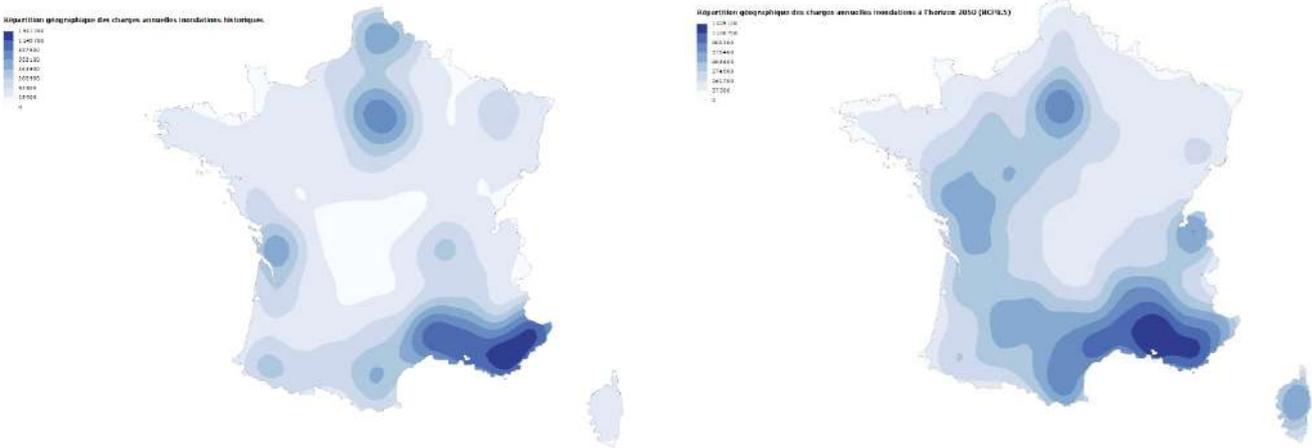


Figure 0-5- Similar hazard basins on the referential period and the forecasted period

Finally, we are interested in the differences between different climate scenarios regarding the rate of risk deviation. For this, a metric that weights the risk assessed by the years of estimation is created. We then find that the most extreme scenario involves a change in risk 20% faster than in other scenarios.

Before concluding, it is worth highlighting the consistency of the model with market models in terms of results.

Actor	↗ Flooding
FFA	11%
Covéa	120%
CCR x Météo France	38%
Moyenne	56%
Essay (gross)	84%
Essay (offset)	57%

Figure 0-6- Market figures estimating flood risk increase

To conclude, what must be remembered from this brief is

- The difficulty of modelling climate risk given its multifactorial aspect, the presence of records of extreme climate variables never reached and the relatively weak source of data on natural disasters.
- A very sharp increase in the risk of flooding by 2050 with premiums adjusted at 2% per year.
- The role of a change in climate scenario for risk management: moving from a RCP8.5 climate scenario to RCP4.5 allows, for example, to reduce the rate of increase in risk by 20%.

Remerciements

Je tiens tout d'abord à remercier mon tuteur de mémoire, Benjamin Silva, pour son soutien et ses conseils tout au long de ce travail. Ton expertise et ta confiance en mon autonomie ont été très importantes pour moi, et je suis reconnaissant de t'avoir eu comme référent tout au long de ce projet. Un grand merci également à tous mes collègues de SeaBird pour l'atmosphère de travail conviviale et pour tout le soutien que vous m'avez apporté durant l'écriture de ce mémoire.

Je remercie également mes amis rencontrés à l'ISFA pour les moments mémorables que nous avons partagés tout au long de notre formation et particulièrement lors de l'expérience de la colocation. Les discussions animées, les projets en équipe et les moments de détente ont rendu ces années d'études et de premier emploi particulièrement enrichissantes.

Plus largement je voudrais ici remercier du fond du cœur tous mes proches pour avoir été une source importante de motivation et d'encouragement dans l'écriture de ce mémoire et la poursuite de mes études.

Quentin Emery

Etude de la déviation d'un
portefeuille multirisque habitation
face au risque climatique
d'inondation

Table des matières

Note de synthèse.....	3
Synthesis note.....	8
Remerciements	13
I. Introduction.....	17
II. Contexte de l'étude : un cadre réglementaire en pleine évolution et une couverture du risque à surveiller	19
A. Définitions et périmètre de l'étude	19
1. Définition du risque climatique.....	19
2. Périmètre de l'étude.....	22
3. Scénarios climatiques.....	23
B. Enjeux réglementaires et de pilotage	26
1. Adaptation et nouveautés dans le cadre normatif actuel.....	26
2. Enjeux autour du pilotage de l'activité.....	31
C. Présentation du cas traité et des données associées	36
1. Données climatiques.....	36
2. Données de sinistralité.....	43
3. Analyse des données jointes.....	57
4. Conclusion générale sur l'analyse des données	59
III. Modélisation de l'évolution du risque : modèle semi-paramétrique	60
A. Projection de la fréquence.....	65
1. Méthodologie.....	65
2. Adéquation et paramétrage.....	68
3. Validation.....	72
4. Sensibilité aux autres scénarios climatiques.....	73
B. Projection de la sévérité.....	75
1. Méthodologie.....	75
2. Adéquation et paramétrage.....	79
3. Validation.....	82
4. Sensibilité aux autres scénarios climatiques.....	82
C. Agrégation entre fréquence et sévérité.....	84
1. Méthodologie.....	84
2. Adéquation.....	84

3.	Validation.....	85
4.	Sensibilité.....	85
IV.	Conclusions quantitatives.....	Erreur ! Signet non défini.
A.	Evolution des métriques sur le portefeuille et décomposition de l'évolution du risque.....	87
1.	Décomposition et répartition des évolutions moyennes globales et départementales	87
2.	Taux d'augmentation de primes espérables et avancement des limites de réassurance.....	90
B.	Proposition d'une mesure de déviation du risque.....	92
1.	Vitesse de déviation du portefeuille.....	92
2.	Conclusions quant à la métrique.....	93
C.	Axes d'amélioration du modèle et comparaisons avec quelques modèles de marché.....	94
1.	Axes d'améliorations	94
2.	Comparaison avec quelques modèles de marché	95
V.	Conclusion	96
VI.	Bibliographie	97

I. Introduction

Le présent mémoire a pour but de répondre à la question suivante : **comment quantifier l'évolution du risque d'inondation sur un portefeuille MRH du fait du changement climatique et quelles déviations peut-on prévoir en termes de charges globales futures ?**

Pour répondre à cette question, dont la formulation n'est pas laissée au hasard et est le fruit d'un cheminement sur la notion de risque climatique que l'on peut retrouver en première partie, deux parties sont proposées.

D'abord, il s'agit de créer une méthode d'évaluation du risque climatique au sens du risque d'augmentation en fréquence et en sévérité des catastrophes naturelles du fait du changement climatique. De ce côté-ci, l'opportunité de travailler à temps plein sur une méthode d'évaluation du risque climatique au sein de l'entreprise a permis d'aboutir à une méthode fonctionnelle et dont les aboutissants sont en adéquation avec les modèles que l'on peut observer sur le marché. Ensuite il s'agit d'appliquer la méthode développée pour déterminer les conséquences du changement climatique (s'il en est), les analyser et répercuter ces analyses sur des questions concrètes.

En termes de structure, la première partie du mémoire est ainsi une partie définitionnelle et contextuelle. Elle traite des enjeux autour du risque climatique et doit permettre au lecteur de comprendre les différents pans du risque climatique notamment celui traité. D'autre part, cette partie doit permettre de saisir la forme globale des données utilisées dans le traitement du problème de régression climatique. On présentera ainsi une analyse sommaire des données climatiques utilisées et une analyse plus importante de la base de données de sinistralité utilisée ainsi que les traitements qui ont été nécessaires sur ces bases.

La seconde partie du mémoire décrit la méthode mise en place pour projeter les charges de sinistralité futures liées aux changements climatiques : cette partie est exhaustive sur la méthode employée en axant d'abord l'explication sur le modèle théorique développé en développant puis en rentrant dans les détails sur la méthode employée pour estimer les fréquences de sinistres futures ainsi que les charges. Par suite, la calibration du modèle sera interrogée sur la base du portefeuille Multi-Risque Habitation (MRH) dont nous disposons en termes d'adéquation, validation et sensibilité notamment aux différents scénarios climatiques.

Enfin, la dernière partie du mémoire concerne l'exploitation des résultats de modélisation et leurs analyses ainsi que les conclusions à tirer en termes d'évolution du risque. De manière concrète, il s'agira d'étudier l'évolution du portefeuille MRH concernant le péril d'inondation et d'utiliser la projection des charges fournie par le modèle créé pour étudier les changements en termes de répartition du risque et

d'augmentations globales ainsi que de proposer une métrique cohérente avec l'horizon de temps d'étude. Enfin, les axes d'améliorations du modèle seront présentés afin de nuancer les résultats.

II. Contexte de l'étude : un cadre réglementaire en pleine évolution et une couverture du risque à surveiller

A. Définitions et périmètre de l'étude

1. Définition du risque climatique

Débutons par donner une définition simple et sans détour du risque climatique : l'ACPR (Autorité de Contrôle Prudentiel et de Résolution) définit dans le rapport de gouvernance des risques climatiques [2] les risques climatiques comme les risques liés au changement climatique et précise que par soucis de simplicité l'on peut prendre pour égal le risque climatique et les risques climatiques.

Bien que cette définition simpliste soit juste, elle ne permet pas de comprendre la pluralité du risque auquel sont confrontés les institutions financières et assureurs. Ainsi, pour aller plus loin, l'ACPR après une étude des définitions du risque climatique internes aux entreprises sur la place française donne une définition plus détaillée reposant sur une division du risque climatique en plusieurs sous-risques [3]. Par soucis de traçabilité, disons maintenant que la classification que l'on s'apprête à énoncer est héritée d'un discours prononcé en septembre 2015 à la Lloyds de Londres par Mark Carney alors gouverneur de la Banque d'Angleterre et fait office de consensus sur la place.

La définition est la suivante :

Le risque climatique est le risque lié au changement climatique qui peut affecter le bilan des assureurs par la présence de :

- risques physiques, qui résultent des dommages directement causés par les phénomènes météorologiques et climatiques, comme par exemple :
 - la perte de valeur des placements détenus par les assureurs et émis par des entités touchées par ces événements climatiques ;
 - l'augmentation de la fréquence et du coût des sinistres à régler par les assureurs aux assurés ;
- risques de transition, qui résultent des ajustements effectués en vue d'une transition vers une économie bas-carbone, en particulier lorsque ceux-ci sont mal anticipés ou interviennent brutalement. Ces risques sont liés par exemple à :

- une dépréciation des actifs, faisant suite à des évolutions réglementaires qui viendraient pénaliser, voire interdire, certaines activités jugées trop émettrices en gaz à effets de serre (GES) ;
- des pertes de contrats d'assurance faisant suite à la fin de certaines activités assurées jugées trop polluantes ou émettrices de GES ;
- risques de responsabilité induits (risques juridiques et de réputation), liés aux impacts financiers des demandes de compensation de la part de ceux qui subissent des dommages dus au changement climatique, comme par exemple :
 - des investissements finançant le développement d'industries et d'activités polluantes ou fortement émettrices de GES ;
 - des assurances professionnelles, responsabilité civile exploitant, construction d'infrastructures.

Notons que malgré le consensus autour de cette définition parmi les assureurs, le NGFS (Network of Central Banks and Supervisors), organisme chargé de l'organisation du verdissement de la finance notamment par la création d'indicateurs financiers prenant en compte le risque climatique et environnemental, ne donne pas la même définition [4].

Pour le NGFS, une différenciation est nécessaire entre le risque climatique et le risque environnemental. Pour l'organisme, le risque climatique prend en compte le risque physique et de transition lié au changement climatique alors que le risque environnemental prend en compte le risque de responsabilité sans toutefois s'y limiter : il s'agit des risques liés à une dégradation de l'environnement qui ont un effet direct et durable sur l'activité (et les sous-jacents de l'activité comme les contrats d'assurance ou les actifs financiers). Par exemple, participerait au risque environnemental une entreprise qui viendrait à polluer un cours d'eau et provoquer la genèse de maladies parmi les habitants autour de ce même cours entraînant la dévaluation durable de sa valeur et des poursuites judiciaires pour sa responsabilité (risque de responsabilité) tout en demandant une intervention pour « réparer » les dégâts (c'est également un risque opérationnel).

De là, on peut quand même noter une forte correspondance des définitions du risque climatique, le risque de responsabilité étant incorporé à un plus grand champ définitionnel qu'est le risque environnemental.

Schématiquement, les superviseurs et banques nationales européennes composent le risque climatique comme suit :

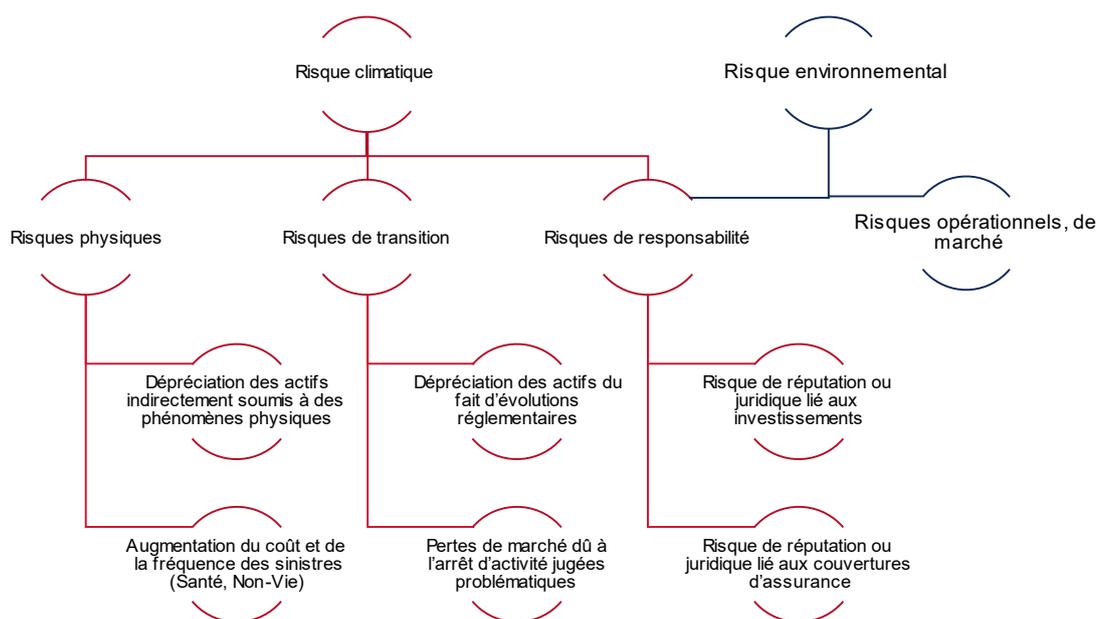


Figure II-1- Organigramme définitionnel du risque climatique et environnemental

En bref, les régulateurs (id est les organismes en charge du renforcement de la réglementation et de la prévention des risques ainsi que de la sécurité des dépôts des épargnants cf. la description que fait d'elle même la Banque de France) ont des définitions qui sont recomposables et adoptées par les assureurs.

Il reste à dire que du point de vue du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) qui fait office de référence en expertise climatique, le risque climatique n'est pas défini autrement que par la potentialité de conséquences néfastes induites par les changements climatiques eux-mêmes ou aux réponses apportées au changement climatique [5]. Finalement, on reconnaît sans qu'il existe un développement définitionnel important les notions de risques physiques et risques de transition.

Ainsi, nous prendrons pour définition primaire la définition du risque climatique donnée ci-avant dans ce mémoire d'autant qu'elle semble rassembler tous les aspects, pourtant nombreux, du risque climatique.

Bien que la définition soit satisfaisante, complétons là en ajoutant que le risque climatique est un risque de longue durée. Cette caractéristique induit deux obligations. La première est de se fixer une étendue temporelle pour évaluer le risque. En effet, si l'effet du changement climatique n'est pas directement perceptible sur un portefeuille d'assurance, il faut néanmoins le prendre en compte pour assurer une adaptation des politiques de primes et de réassurance à moyen terme. La seconde est de se fixer un

pas de temps raisonnable pour évaluer le risque, un pas de temps trop petit impliquant une mauvaise évaluation du risque.

Pour tout cela -l'envergure du risque climatique et sa temporalité- il est nécessaire de se circonscrire à un sous-risque particulier afin de mener à bien une étude a minima pertinente a maxima robuste.

Ainsi, sera exposé dans la suite le risque spécifique auquel s'intéresse ce mémoire.

2. Périmètre de l'étude

De manière brève, l'étude portée par ce mémoire est l'étude du risque climatique physique associé aux catastrophes naturelles (que l'on pourra dénommer CAT-NAT au cours du mémoire) liées au phénomène d'inondation sur la zone géographique de la France métropolitaine à une maille départementale, à un horizon de temps 2050 au maximum et un pas annuel.

Le choix de ce périmètre n'est pas laissé au hasard.

D'abord, la maille et la zone géographique est choisi conformément à la disponibilité des données même si le modèle construit ci-après dans le mémoire pourrait éventuellement s'adapter à une maille et une zone différente.

De même, le pas de temps choisi est annuel car il représente un bon compromis entre volume de données et précision de la modélisation, cependant on pourrait imaginer se placer à un pas de temps trimestriel ou saisonnier afin de prendre en compte la saisonnalité des événements CAT-NAT.

Ensuite, l'horizon temporelle de 30 ans et moins est un horizon classique dans le monde de l'expertise climatique et en particulier est en adéquation avec les attentes des régulateurs dans les essais de stress- tests climatiques menés auparavant. Il a l'avantage de permettre la prise en compte d'un effet durable et futur mais de laisser également la place à des actions opérationnelles.

Enfin, le choix de ne s'intéresser qu'au risque physique associé aux catastrophes naturelles est la conséquence de plusieurs éléments. Avant toute chose, le traitement du risque climatique dans son ensemble est trop complexe et il ne serait pas pertinent de tenter d'évaluer tous les aspects de celui-ci : il est nécessaire de se concentrer sur un unique aspect du risque climatique. Par choix, et du fait que la problématique est moins explorée que le risque de transition et de responsabilité (au travers par exemple des articles 173 de la loi relative à la transition énergétique pour la croissance verte et articles 128 du code de l'environnement), ce mémoire se concentre sur le risque physique « direct ». Ensuite, il convient de dire que le risque physique « direct » c'est à dire afférent à l'augmentation de la fréquence et de la sévérité des sinistres, au vue de l'exercice pilote climatique de l'ACPR, est surtout le fait de risques : le risque CAT-

NAT et le risque Santé. Ainsi, étant donné la différence entre ces risques et toujours dans l'idée de s'attacher à un aspect unique du risque climatique puis par appétence aux sujets relatifs aux CAT-NAT et enjeux réassuranciers, ce mémoire traite de l'évaluation du risque CAT-NAT au regard du changement climatique et plus précisément au risque d'inondation.

La problématique est donc la suivante, comment quantifier l'évolution du risque d'inondation sur un portefeuille MRH du fait du changement climatique et quelles déviations peut-on prévoir en termes de charges globales futures ? Par souci de simplification, au sein du mémoire et en l'absence de contre-indication, sera désigné par « risque climatique » le risque décrit et défini en amont.

3. Scénarios climatiques

Pour évaluer et quantifier monétairement le risque climatique, il est nécessaire de projeter en amont l'évolution du climat par l'ensemble des variables atmosphériques et météorologiques (humidité, pressions, températures. . .) caractérisant ce qu'on appelle le climat. Depuis 1988, le GIEC s'attache à prévoir les circonstances atmosphériques et météorologiques futures notamment par la production de scénarios optimistes ou pessimistes quant aux actions humaines limitant ou accélérant le processus de changement climatique. Les dernières générations de scénarios qui projettent à un pas de temps journalier les variables climatiques sont les SSP (Shared Socioeconomic Pathways), ils font suite aux scénarios RCP (Representative Concentration Pathways), et aux scénarios SRES et IS92 avant eux. Les scénarios sont ensuite régionalisés, par exemple, pour la France Métropolitaine, c'est le Drias (acronyme de Donner accès aux scénarios climatiques régionalisés français pour l'impact et l'adaptation de nos sociétés et environnements) qui a pour vocation de mettre à disposition des projections climatiques régionalisées réalisées dans les laboratoires français de modélisation du climat (IPSL, CERFACS, CNRM).

Les scénarios RCP sont de la forme RCPx où x correspond au forçage radiatif (perturbation du bilan radiatif du système climatique de la Terre) employé pour projeter les variables climatiques. Plus le forçage radiatif est élevé plus les changements sont drastiques et négatifs pour la vie sur Terre. Le pire scénario est ainsi le RCP8.5.

De manière plus aboutie le 6e rapport du GIEC paru en 2022 crée les SSP qui incluent des hypothèses supplémentaires sur la gestion du changement climatique par l'humanité, ils sont de la forme SSPx-y où x est le numéro du scénario social choisi et y le forçage radiatif adopté. Ci-dessous, la liste des narratifs associés aux scénarios SSPx :

Scénario	Narratif	Défi pour l'atténuation	Défi pour l'adaptation
SSP1 Développement durable	Fort coopération internationale, priorité donnée au développement durable, amélioration des conditions de vie et préférences des consommateurs pour des biens et services respectueux de l'environnement, peu intensifs en ressources et en énergie.	Faible	Faible
SSP2 Poursuite des tendances	Les tendances sociales, économiques et technologiques actuelles se poursuivent, le développement et la croissance progressent de manière inégale selon les pays et les régions. Les institutions nationales et internationales œuvrent à la réalisation des objectifs de développement durable qui progresse lentement. L'environnement se dégrade malgré un développement moins intensif en ressources et en énergie.	Moyen	Moyen
SSP3 Rivalités régionales	Résurgence des nationalismes, développement économique lent, persistance des inégalités et des conflits régionaux. Les pays sont guidés par des préoccupations en matière de sécurité et de compétitivité. Ils se concentrent sur les problèmes nationaux voire régionaux et sur les enjeux de sécurité alimentaire et énergétique. Faible priorité internationale pour la protection de l'environnement, qui se dégrade fortement dans certaines régions.	Elevé	Elevé
SSP4 Inégalités	Développement marqué par de fortes inégalités entre et à l'intérieur des pays. Dégradation de la cohésion sociale et multiplication des conflits. Fossé croissant entre une élite connectée et mondialisée, responsable de la majorité des émissions de GES, et un ensemble fragmenté de populations à faible revenu, peu éduqué et vulnérable au changement climatique. Le secteur énergétique se diversifie entre sources d'énergie fortement carbonées et décarbonées. Les politiques environnementales se concentrent sur les enjeux locaux.	Faible	Elevé
SSP5 Développement conventionnel	Développement adossé à l'exploitation forte des énergies fossiles et marqué par des investissements élevés dans la santé, l'éducation et les nouvelles technologies. Adoption de modes de vie intensifs en ressources et en énergie à travers le monde. La croissance économique et le progrès technologique sont élevés. Les problèmes de pollution locale sont bien gérés et l'adaptation est facilitée notamment grâce au recul de la pauvreté.	Elevé	Faible

Sources : Ffahí et al. (2017), O'Neill et al. (2015), Bauer et al. (2017)

Figure II-2- Les SSP et leurs narratifs respectifs

A titre d'exemple, un graphique de l'augmentation de température globale selon les scénarios adoptés :

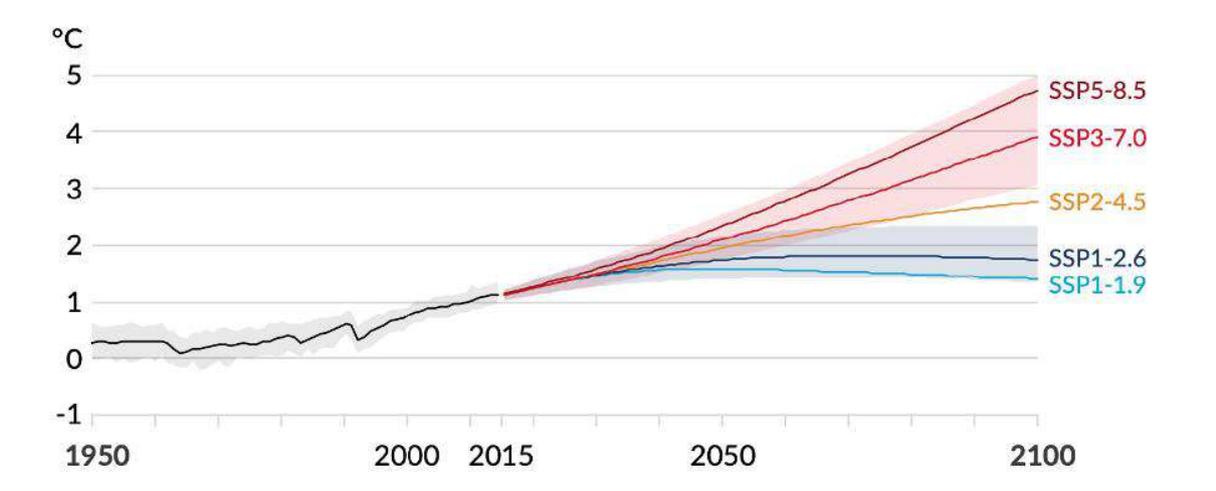


Figure II-3- Evolution des températures selon les différents scénarios climatiques

Les scénarios les plus dommageables et qui sont corrélés à un risque climatique très important sont ainsi le SSP5 – 8.5 (ancien RCP8.5). On prendra en priorité dans notre modèle d'évaluation du risque climatique ces scénarios en compte, les plus pessimistes, pour deux raisons : ils sont les scénarios les plus réalistes pour le moment et il est prudent de les considérer au premier chef.

Maintenant que le risque climatique a été défini ainsi que le périmètre auquel s'intéresse ce mémoire, nous allons expliquer quels sont les enjeux réglementaires et de pilotage auxquels sont exposés les assureurs au regard de ce risque.

B. Enjeux réglementaires et de pilotage

1. Adaptation et nouveautés dans le cadre normatif actuel

Les premiers enjeux associés au risque climatique sont des enjeux réglementaires. En effet, étant donné l'importance de ce risque émergent (classé 1er risque émergent dans le Future Risk Report d'AXA par exemple [1]) les superviseurs veulent progressivement inclure le risque dans les processus prudentiels voire créer de nouvelles obligations prudentielles [6] [7] [8]. D'abord, est envisagé par l'ACPR l'élaboration d'un stress-test climatique régulier pour choquer les bilans des périmètres risques physiques [6], ensuite l'EIOPA (Autorité Européenne des Assurances et des Pensions) réfléchit à l'inclusion du risque climatique dans le SCR (Solvency Capital Requirement) [7], Value-At-Risk à 99, 5%, de Solvabilité II ainsi que de dans l'ORSA des entreprises [8] ce qui pose plus largement des questions de gouvernance réglementaire du risque climatique.

a. Les stress-tests climatiques de l'ACPR

Afin de mesurer la solidité financière des établissements financiers face au changement climatique à l'horizon 2050, l'ACPR a mis en place un exercice pilote qui consiste en l'application de « stress-tests » sur les expositions risquées de l'organisme considéré. L'objectif général est double : garantir la sécurité financière des établissements et permettre une prise de conscience amenant à une finance durable. Si l'exercice n'est pour le moment pas obligatoire et ne mène pas à des exigences quantitatives, l'ACPR pourrait avec la demande de divers ONG et d'autres acteurs le rendre obligatoire [9].

Quels risques sont pris en compte dans les stress tests ? Divers risques climatiques affectant le passif et l'actif du bilan sont pris en compte. Seulement, en ce qui concerne le passif d'assurance et le risque physique climatique, seuls les risques liés à l'augmentation de la fréquence de sinistres de type catastrophe naturelle et santé sont nécessairement pris en compte. Les risques chroniques, trop difficiles à évaluer à date ne sont pas obligatoirement pris en compte. Il est aussi important de noter que le stress-test, dont la restitution se fait via un tableau de charges pour le passif laisse la liberté de considérer toutes les branches que l'on souhaite pour projeter son passif d'assurance au regard climatique même si les branches privilégiées sont l'automobile, le dommage aux biens, la santé et les catastrophes naturelles (placées hors dommages aux biens et automobile).

Quel scénario climatique est utilisé ? Le scénario utilisé pour faire la projection de charges de sinistralité est le scénario RCP8.5. En revanche, 3 scénarios sociaux viennent compléter le scénario purement climatique. On peut commodément les résumer en scénario « conséquences attendues », scénario « conséquences plus

graves que prévu», scénario « conséquences moins graves que prévu». Pour chacun de ces scénarios de stress-test, l'ACPR fournit des indicateurs macroéconomiques pour compléter la projection climatique comme l'indice de chômage, l'inflation prévue jusqu'en 2050 par exemple.

Quelles sont les attentes en termes de reporting ? Les charges de sinistralité doivent être estimées à plusieurs dates butoirs : 2025, 2035, 2040 et 2050 grâce aux scénarios données par l'ACPR. En outre, jusqu'en 2025 le bilan doit être statique i.e. il ne doit pas prendre en compte des réactions de gestion aptes à réduire l'effet du risque climatique sur le bilan, puis à partir de 2025 l'hypothèse d'un bilan dynamique est prise. Cela permet d'avoir une vision d'impact à court terme sur les institutions ainsi qu'à long terme tout en observant les choix de gestion qui sont faits. Par ailleurs, d'un point de vue géographique, la granularité du rapport est départementale pour la France et continentale pour le reste du monde. Pour l'exercice pilote, il était attendu des assureurs de faire parvenir au régulateur, sur le périmètre CAT-NAT, un tableau d'exposition de la forme :

Expositions géographiques pour les catastrophes naturelles (tous périmètres confondus, y compris non inclus dans le régime CAT NAT)

	Nombre de risques assurés				Valeurs assurées				Primes émises CatNat				Sinistres CatNat			
	2010	2015	2040	2050	2010	2015	2040	2050	2010	2015	2040	2050	2010	2015	2040	2050
France métropolitaine	2100	2300	2450	2550	3000	3300	3500	3700	1000	1100	1200	1300	1400	1500	1600	1700
FR01																
FR02																
FR03																
FR04																
FR05																
FR06																
FR07																
FR08																

Figure II-4- Reporting attendu d'un stress-test climatique

Quelles conclusions tirer de l'exercice pilote ? L'exercice pilote a permis de mettre en lumière une augmentation globale significative des primes du fait de l'augmentation de la fréquence et de la sévérité des événements catastrophes : entre 2, 8% et 3, 7% d'augmentation par an résultant d'une augmentation entre 130% et 200% sur 30 ans pour conserver un ratio combiné constant [9]. Un autre élément que l'on doit retenir est la prise en compte d'un bilan dynamique : le pilotage du risque climatique a un intérêt double. D'abord, il est intrinsèquement intéressant de maîtriser ce risque pour le pilotage de l'activité, mais en outre un pilotage argumenté et solide revêtirait un intérêt réglementaire si toutefois les stress-tests devenaient obligatoires.

Nous nous attacherons à comparer les résultats globaux du premier stress-test (le seul dont on a les résultats) avec l'évolution prédite sur notre portefeuille. Passons maintenant aux implications que pourrait avoir le risque climatique sur le SCR de Solvabilité II (S2) (dans le pilier 1 de la norme).

b. L'inclusion du risque climatique dans Solvabilité II

i. Pilier I : évolution du Solvency Capital Requirement

En juillet 2021, l'EIOPA a publié un papier sur la méthodologie qui pourrait être appliquée pour l'inclusion du risque climatique -comme on l'a défini- dans le SCR catastrophes naturelles du module SCR CAT Non-Vie. [7]

Rappels sur le SCR

Le SCR (pour Solvency Capital Requirement) représente l'exigence de fonds propres à respecter par un organisme assurantiel sous peine de faire l'objet d'une surveillance renforcée et d'une intervention de l'ACPR. Il correspond au niveau de fonds propres nécessaire pour limiter la probabilité de défaillance de l'organisme à 0,5% ou, autrement dit, pour faire face à un choc extrême bicentenaire. Son calcul est basé sur l'exposition aux risques liés à l'activité des organismes d'assurance.

Deux méthodes sont possibles pour le calcul dudit SCR :

- l'utilisation de la formule standard fournie par le règlement délégué
- l'utilisation d'un modèle interne à l'entreprise

Dans la formule standard, le SCR est divisé en plusieurs sous SCRs qui sont ensuite agrégés par des pseudo normes euclidiennes (pour permettre la sous-additivité de la mesure et donc les bénéfices de diversifications). Pour ce qui est de l'utilisation d'un modèle interne, il convient de différencier les modèles internes complets dans lesquels une Value-At-Risk à 99,5% est calculée sans l'aide de la structure standard et les modèles internes partiels dans lesquels la structure standard est toute ou partie adoptée sans que les SCR de catégories infra soient tous calculés de manière standard. Bien évidemment, les modèles internes partiels ou complets doivent être validés par l'ACPR avant d'être utilisés. Maintenant que le cadre global du calcul du SCR est globalement rappelé (une multitude de sources permettent l'approfondissement de ce point-ci), détaillons le calcul en formule standard du SCR auquel nous nous intéressons.

Nous nous intéressons ici au SCR de risque de catastrophes naturelles présent au sein du sous- module SCR CAT du module SCR Non-Vie. La formule standard permet d'obtenir le capital requis de solvabilité par des calculs relativement simples, elle est à 3 niveaux pour le SCR catastrophes naturelles du SCR CAT Non-Vie : pour chaque péril considéré par Solvabilité II (tempête, séisme, inondation, grêle, affaissement de terrain) un SCR est calculé, ce SCR est calculé comme une norme euclidienne pondérée par des facteurs de corrélations de SCR régionaux pour chaque région dans laquelle l'organisme a des affaires. Quant à eux, les SCR régionaux sont calculés à partir d'une pondération par des facteurs de risques fournis par l'ACPR des sommes assurées

pondérées par zones de la région considérée, dans certaines régions, par manque d'étude, on utilise une formule différente basée sur les volumes de primes. De manière synthétique, en formule standard :

$$SCR_{natcat} = \sqrt{\sum_{peril} SCR_{peril}}$$

Avec :

$$SCR_{peril} = \sqrt{\left(\sum_{(r,s)} CorrPeril_{(r,s)} SCR_{(peril,r)} SCR_{(peril,s)} \right) + SCR_{(peril,autres)}^2}$$

avec (r, s) des régions dont les facteurs de risques sont donnés par le régulateur et dont les coefficients de corrélations $CorrPeril_{(r,s)}$ sont également donnés et *autres* les régions dont les SCR sont calculés avec les volumes de primes. En particulier :

$$SCR_{(peril,r)} = Q_{(peril,r)} \sqrt{\sum_{(i,j)} CorrPeril_{(r,i,j)} W_{(peril,r,i)} SI_{(peril,r,i)} W_{(peril,r,j)} SI_{(peril,r,j)}}$$

avec Q le facteur de risque associé au péril et à la région d'intérêt, W la pondération des sommes assurées (SI) de la zone i de la région r pour le péril donné et $CorrPeril$ les coefficients de corrélations entre zones de la région.

$$SCR_{(peril,autres)} = c_{(peril,autres)}(0,5DIV_{autre} + 0,5)P_{peril}$$

avec DIV le facteur de diversification des zones hors scope qui se calcule comme le quotient de la variance intra et de la variance totale du risque sur ces zones et P le volume de primes récolté dans les zones hors scope.

De manière synthétique, les paramètres correspondent :

- Pour les facteurs de risques à des taux de destructions pour les événements à période de retour de 200 ans (donc à un risque 0,5%) i.e. des ratios sur sommes assurées des PML (Probable Maximum Loss) sur 200 ans.
- Pour les coefficients de corrélations à des coefficients de corrélations sur les distributions des événements à période de retour 200 ans par zone.

Malgré la simplicité des calculs de la formule standard, nous voyons qu'il existe une réelle complexité quant à la détermination des entrants : valoriser les sommes assurées propres à chaque zone par exemple est difficile. D'autre part, les paramètres fournis par le régulateur sont discutables, ainsi, sous couvert de justification, un organisme peut utiliser ses propres paramètres avec la formule standard. Il s'agit des USP (pour undertaking specific parameters) qui sont un moyen d'estimer au mieux le SCR de catastrophes naturelles sans pour autant adopter un modèle interne qui peut être coûteux et difficile à mettre en place.

Risque climatique et SCR

Plusieurs éléments sont donnés par l'EIOPA pour l'inclusion du risque climatique au sein du SCR. Il est à noter qu'au moment où est écrit ce mémoire, il ne s'agit que d'une note méthodologique qui n'a pas vocation de loi, cependant elle pourrait être à la genèse de mouvements réglementaires prochainement. Dans ce cadre là, le présent mémoire a vocation à donner quelques pistes de réflexions sur le sujet.

D'abord, selon la note, dans le cas où l'organisme utilise un modèle interne pour traiter le risque catastrophe naturelle au sein de son SCR, il deviendra nécessaire d'y inclure le risque climatique [7, p.22]. Le SCR étant une vision à court terme (il s'agit d'une Value-At-Risk de niveau 99,5% à 1 an), les modèles internes sont censés prendre indirectement en compte le risque climatique du fait de leurs mises-à-jour récurrentes néanmoins il faudra pouvoir le justifier auprès du régulateur. Ensuite, si l'on s'intéresse à la formule standard, la problématique principale est l'évolution des paramètres (USP ou fournis). En effet, la note prévoit l'évolution des paramètres fournis par le régulateur tous les 5 ans avec une recalibration 2 ans [7, p.4] avant de telle sorte que l'on est la chronologie suivante :

Ainsi, le risque climatique sera pris en compte sur une durée de 5 ans en avant à chaque nouvelle recalibration : il faudra alors prendre en compte l'évolution sur 5 ans du risque catastrophe naturelle liée au risque climatique. La problématique est surtout importante pour les organismes utilisant des USP, si la recalibration des USP au regard des évolutions climatiques n'est pas faite, il est probable que les USP ne soient pas validés par le régulateur et donc qu'ils soient inutilisables.

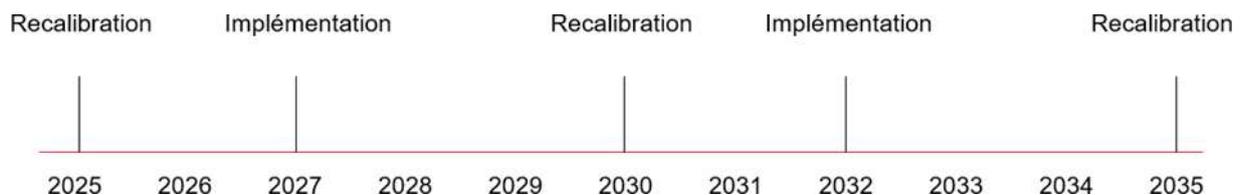


Figure II-5- Chronologie d'évolution des paramètres du SCR CAT

Nous voyons donc que l'enjeu, dans Solvabilité II est important : la recalibration des USP ou des modèles internes s'ils sont trop statiques devra potentiellement être faite. D'autre part, d'un point de vue opérationnel, l'utilisation des paramètres standards nécessitera probablement une refonte des modèles si les paramètres ne sont pas des entrants des modèles mais des paramètres statiques. Si nous n'étudions pas directement l'évolution du SCR, nous regarderons l'évolution sur la queue de distribution du risque inondation de notre portefeuille et tirerons des conclusions en matière de solvabilité.

ii. Pilier II : scénarios climatiques dans l'ORSA

Une autre obligation réglementaire insufflée par Solvabilité II (cette fois dans son pilier 2) est l'obligation de construire un ORSA (Own Risk and Solvency Assessment) c'est à dire un processus d'évaluation de la structure des risques et de la solvabilité de l'entreprise matérialisé par un rapport. Cet ORSA permet de mettre en lien l'activité avec les exigences de Solvabilité II, notamment en s'intéressant à une échéance à plus long terme des exigences quantitatives du pilier 1 de Solvabilité II. Dans ce cadre-là, selon le rapport sur la gouvernance du risque climatique produit par l'ACPR, l'ORSA permet une clarification des rôles et des méthodes employées par l'organisme afin d'évaluer et de gérer le risque climatique. Cela fait en réalité suite à une consultation de l'EIOPA menée en 2021 sur l'inclusion de scénarios climatiques dans l'ORSA [8]. La consultation parvient à la conclusions que les entreprises devront inclure a minima deux scénarios climatiques dans leur ORSA à horizon de leur plan stratégique (c'est à dire à un horizon de 5 ans en général) :

- scénario climatique dans lequel la température moyenne mondiale n'augmente pas de plus de 2°C
- scénario climatique dans lequel la température moyenne mondiale dépasse l'augmentation de 2°C

En outre, l'ACPR [2] conseille aux organismes de commencer à travailler sur la création d'indices de sensibilité au risque climatique qui pourront être inclus à la discrétion des organismes dans leur ORSA.

Finalement, nous voyons que la gouvernance du risque climatique au sein de l'entreprise et en particulier au sein de l'ORSA est réglementé, néanmoins, elle se rapproche de problématique concernant l'activité de l'entreprise et le pilotage du risque climatique plus que la simple solvabilité demandée par le régulateur. Nous allons donc maintenant nous intéresser aux enjeux de pilotage qu'induisent le risque climatique.

2. Enjeux autour du pilotage de l'activité

a. *Question d'assurabilité et de déviation des portefeuilles*

L'évolution des primes d'assurance liées au risque d'inondation face au changement climatique est un sujet complexe et crucial pour la société, les assureurs et les gouvernements du monde entier. Pour comprendre pleinement cette question, il est essentiel d'examiner en détail les facteurs sous-jacents, les défis posés par l'assurabilité et les réponses potentielles à ce défi grandissant.

L'impact du changement climatique sur les inondations

Le changement climatique est une réalité indéniable qui s'est traduit par des phénomènes météorologiques de plus en plus extrêmes, notamment des

précipitations intenses, des tempêtes plus fréquentes et une montée du niveau de la mer. Ces changements entraînent des répercussions directes sur les inondations, les rendant plus fréquentes, plus graves et plus imprévisibles. Les événements climatiques extrêmes, tels que les ouragans, les tempêtes tropicales et les pluies torrentielles, peuvent également déclencher des inondations soudaines, mettant en danger les biens et les vies humaines.

L'assurabilité des risques d'inondation

L'assurabilité des risques d'inondation est un concept clé qui fait référence à la possibilité pour les assureurs de fournir une couverture pour ces risques. Elle est influencée par plusieurs facteurs, notamment la fréquence et la gravité des inondations, les caractéristiques géographiques et climatiques des zones à risque, la capacité des assureurs à évaluer et à gérer ces risques, ainsi que la disponibilité de données fiables et de modèles de risque précis.

La montée des coûts d'indemnisation

L'augmentation des sinistres liés aux inondations en raison du changement climatique a entraîné une montée significative des coûts d'indemnisation pour les compagnies d'assurance. Les inondations peuvent causer des dégâts matériels considérables, notamment des dommages aux habitations, aux entreprises, aux infrastructures publiques et aux biens personnels. Ces coûts sont exacerbés par la hausse des valeurs immobilières et des biens assurés, ainsi que par l'urbanisation croissante dans les zones à risque d'inondation.

Les ajustements tarifaires

Face à cette réalité économique, de nombreuses compagnies d'assurance ont dû réajuster leurs tarifs pour tenir compte de l'augmentation des risques d'inondation. Ces ajustements se traduisent souvent par des primes d'assurance plus élevées pour les propriétaires et les entreprises situés dans des zones à risque. Les assureurs ont également commencé à adopter des approches plus sophistiquées de modélisation des risques pour évaluer de manière plus précise les probabilités d'inondation et les coûts potentiels associés.

L'incertitude et les défis de la modélisation des risques

La modélisation des risques liés aux inondations est un domaine complexe qui exige une compréhension approfondie de multiples facteurs, tels que les données historiques, les caractéristiques géographiques, les systèmes de drainage, les mesures d'atténuation des risques, et bien sûr, les données climatiques. Cependant, le changement climatique introduit une grande incertitude dans cette modélisation, car il est difficile de prédire avec précision comment les conditions météorologiques vont évoluer à l'avenir. Cette incertitude pose un défi majeur aux assureurs qui doivent prendre des décisions financières à long terme.

La question de l'assurabilité en zone à risque

Dans certaines régions, en particulier celles fortement exposées aux inondations, l'assurabilité devient une préoccupation majeure. Les primes d'assurance peuvent devenir prohibitives pour de nombreux résidents et entreprises, ce qui les laisse vulnérables aux pertes financières en cas d'inondation. Cette situation soulève des questions sur l'accès à une couverture d'assurance adéquate et sur la manière de protéger les populations vivant dans des zones à risque.

Les réponses possibles

Face à ces défis, il est impératif de trouver des solutions pour garantir l'assurabilité des risques d'inondation tout en encourageant la résilience face au changement climatique. Voici quelques-unes des réponses possibles à cette problématique :

- Investir dans la prévention et l'atténuation des risques : Les gouvernements et les collectivités locales peuvent investir dans des mesures telles que la construction de digues, le renforcement des infrastructures de drainage, et la gestion des zones inondables pour réduire les risques.
- Promouvoir la réassurance : Les assureurs peuvent se tourner vers la réassurance pour partager les risques d'inondation à l'échelle mondiale. Cela peut aider à stabiliser les coûts d'indemnisation et à maintenir des primes plus abordables.
- Encourager la recherche et l'innovation : La recherche continue sur la modélisation des risques d'inondation et les méthodes d'atténuation est essentielle pour mieux comprendre et gérer ces risques.
- Sensibilisation et éducation : Informer le public sur les risques d'inondation et les mesures à prendre pour se protéger peut contribuer à réduire les pertes et les coûts d'indemnisation.
- Politiques publiques adaptées : Les gouvernements peuvent mettre en place des politiques qui encouragent les résidents à prendre des mesures d'atténuation des risques, comme l'élévation des maisons au-dessus du niveau de l'eau ou l'achat de dispositifs de protection contre les inondations.

L'évolution des primes d'assurance liées au risque d'inondation face au changement climatique est un défi de taille qui nécessite une approche multidimensionnelle. Dans notre cas, loin d'avoir des clés de résilience face au risque d'inondation d'ordre physique, nous montrerons simplement une évolution possible des primes lié au changement climatique.

b. Durcissement du marché de la réassurance

Le durcissement du marché de la réassurance en contexte d'augmentation des catastrophes liées au risque climatique peut être expliqué par plusieurs facteurs interdépendants. Voici les principales raisons qui contribuent à ce durcissement :

Augmentation des pertes liées au climat : L'augmentation des phénomènes météorologiques extrêmes et des catastrophes naturelles causées par le changement climatique a entraîné une hausse significative des pertes financières pour les réassureurs. Les coûts d'indemnisation après des événements tels que les inondations ont augmenté, ce qui a mis sous pression les réserves financières des réassureurs.

Épuisement des capacités de réassurance : L'accumulation de pertes dues à des catastrophes naturelles a conduit à l'épuisement des capacités de réassurance. Les réassureurs ont atteint leurs limites en termes de montant total qu'ils peuvent assurer. Cette situation a conduit à une raréfaction de la réassurance disponible sur le marché, ce qui a entraîné une hausse des prix et des conditions plus strictes.

Réévaluation des modèles de risque : Le changement climatique a forcé les réassureurs à réévaluer leurs modèles de risque. Ils doivent prendre en compte les données climatiques en constante évolution pour évaluer plus précisément les probabilités d'événements extrêmes. Cette réévaluation peut conduire à une augmentation des primes pour refléter le risque accru.

Pression sur les marges bénéficiaires : Les réassureurs ont été confrontés à une pression accrue sur leurs marges bénéficiaires en raison de l'augmentation des pertes. Pour maintenir leur rentabilité, ils ont dû augmenter les primes et durcir les conditions de souscription.

Incertitude climatique croissante : L'incertitude entourant les tendances climatiques futures rend la modélisation des risques plus complexe. Les réassureurs doivent prendre en compte la variabilité climatique et les changements imprévisibles, ce qui accroît l'incertitude quant à l'évaluation des risques.

Évolution des réglementations et des normes de solvabilité : Si les régulateurs financiers renforcent les exigences de solvabilité pour les réassureurs afin de garantir leur stabilité financière face aux risques climatiques. Cela pourrait pousser les réassureurs à mobiliser davantage de capital pour respecter ces normes, ce qui a également exercé une pression sur les prix.

En résumé, le durcissement du marché de la réassurance dans un contexte d'augmentation des catastrophes liées au risque climatique est le résultat de multiples facteurs, notamment des pertes accrues, des capacités limitées, des modèles de risque révisés et une concurrence croissante. Cette évolution souligne la

nécessité d'une approche pour comprendre l'évolution du risque, notamment la quantifier et tirer des conclusions en milieu et fin de distribution.

C. Présentation du cas traité et des données associées

Dans cette partie, les données utilisées seront présentées et analysées.

1. Données climatiques

a. Présentation de la base de données brute

Les données climatiques utilisées sont composées de deux base de données open-source provenant du DRIAS, institut climatologique sous tutelle du ministère de la transition écologique. Les base de données contiennent symétriquement les indicateurs suivant :

Nom	Description
TAV	Température moyenne journalière (°C)
TNAV	Température minimale journalière (°C)
TXAV	Température maximale journalière (°C)
TRAV	Amplitude thermique journalière (°C)
TXQ90	Extrême chaud de la température maximale journalière (°C)
TXQ10	Extrême froid de la température maximale journalière (°C)
TNQ90	Extrême chaud de la température minimale journalière (°C)
TNQ10	Extrême froid de la température minimale journalière (°C)
HDD	degré-jours de chauffage (°C)
CDD	degré-jours de climatisation (°C)
SD	Nombre de journées d'été (jour(s))
TXND	Nombre de jours anormalement chauds (jour(s))
TNHT	Nombre de nuits anormalement chaudes (jour(s))
TR	Nombre de nuits tropicales (jour(s))
TNFD	Nombre de jours de gel (jour(s))
TXFD	Nombre de jours sans dégel (jour(s))
TNND	Nombre de jours anormalement froids (jour(s))
TNCWD	Nombre de jours d'une vague de froid (jour(s))
TXHWD	Nombre de jours d'une vague de chaleur (jour(s))
PAV	Précipitations journalières moyennes (mm)
RR	Cumul de précipitation (mm)
PINT	Précipitation moyenne les jours pluvieux (mm)
PFL90	Fraction des précipitations journalières intenses (%)
RR1MM	Nombre de jours de pluie (jour(s))
PN20MM	Nombre de jours de fortes précipitations (jour(s))
PXCWD	Nombre maximum de jours pluvieux consécutifs (jour(s))
PXCDD	Période de sécheresse (jour(s))

Tableau II-1- Indices climatiques de la base de données climatique

Les deux bases sont :

- Une base référentielle contenant des valeurs numériques pour les indicateurs listées ci-dessus par station météorologique française à un pas de temps annuel sur la période 1978 – 2005.
- Une base prospective contenant des valeurs numériques pour les indicateurs listées ci-dessus par station météorologique française à un pas de temps annuel sur la période 2006-2100 selon les différents scénarios climatiques RCP (2.6, 4.5 et 8.5).

b. Traitements réalisés sur la base

Dans le cadre de notre étude, nous avons choisi de nous intéresser à l'évolution du risque à une échelle départementale et à un pas de temps annuel. Dès lors, nous devons procéder à un traitement des données pour adapter la granularité de celles-ci. En somme, il faut que les données climatiques à la maille de la station météorologique soient à la maille du département de France métropolitaine pour chaque scénario climatique.

Etape 1 : jointure de la table référentielle et prospective

La première étape de la démarche consiste à la jointure des deux base de données, cette jointure est simple car les bases sont à la même granularité et l'une n'est qu'un prolongement prospectif de l'autre.

Etape 2 : Association d'une couverture spatiale aux données ponctuelles des stations météorologiques

La seconde étape consiste à convertir les données ponctuelles associées aux stations météorologiques en données surfaciques. Le but ici est de trouver la zone maximale dans laquelle les données climatiques d'une station peuvent être utilisées, c'est-à-dire de trouver en quelque sorte la zone d'influence de la station météorologique.

Pour cela, on utilise le logiciel QGIS (logiciel de traitement de données géographique open-source) et nous traçons des polygones de Voronoï autour de chaque point correspondant à une station météorologique. Les polygones de Voronoï sont les plus grand polygones autour d'un point qui maximisent les polygones autour des points voisins. Schématiquement :

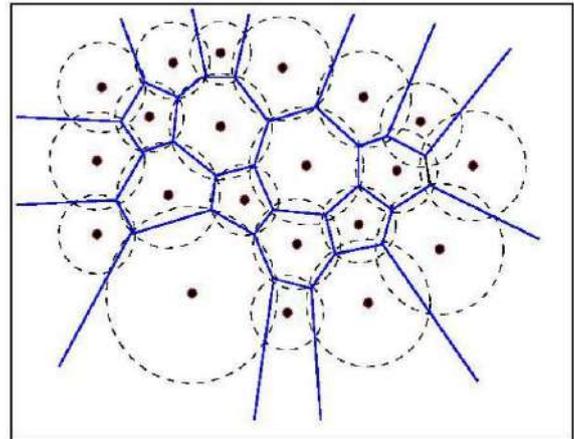
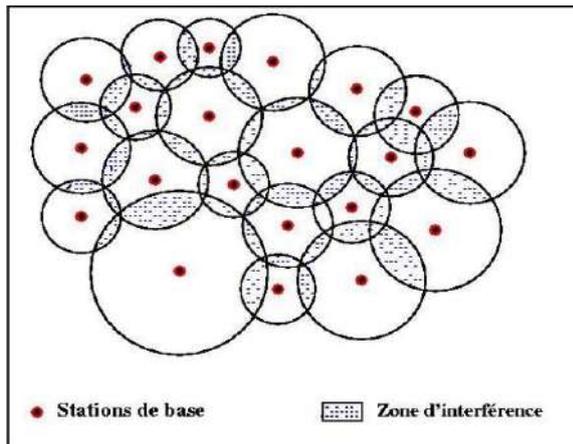


Tableau II-2- Construction schématique de polygones de Voronoï autour de données ponctuelles [27]

Dès lors, est associé à chaque station météorologique contenant un ensemble de relevés annuels des surfaces d'influence sur lesquelles on considérera que l'on peut prendre pour indicateurs climatiques ceux de la station.

Etape 3 : Association de poids pour chaque station météorologiques dans chaque département

Maintenant que nous disposons de la surface d'influence de chaque station météorologique, il s'agit de croiser cette information avec les zones administratives dont on veut extraire les données, ici, le département. Ainsi, en croisant les polygones de Voronoï hérités des stations météorologiques avec le découpage de la France métropolitaine des départements nous sommes capable d'avoir le poids de chaque station météorologique dans un département. Ce poids est en effet défini égal à la surface d'influence de la station dans le département sur la surface totale du département.

Du fait de la disponibilité des données surfaciques des communes sur QGIS (échelle du code INSEE), nous avons établi le calcul des surfaces de croisement à l'échelle communale puis sommer les surfaces par département afin d'obtenir le résultat souhaité.

Pour rendre compte du procédé, ci-dessous un schéma dans lequel se trouve une commune (en jaune) dont l'empreinte se croise avec 3 stations météorologiques (la 7633, la 7489 et la 7490) dont les polygones sont rectangulaires.

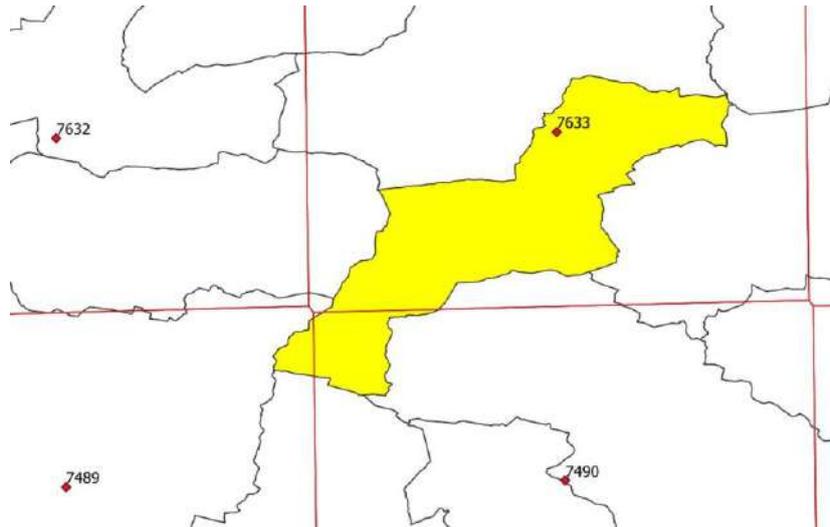


Figure II-6- Exemple de croisement des surfaces d'influence des stations météorologique et d'une zone administrative

On récupère alors la surface de croisement entre la commune et chaque station.

On a alors pour chaque code INSEE (i.e. commune) la surface de croisement entre la commune et chaque station météorologique.

Dès lors, il est possible de sommer sur les 2 premiers chiffres INSEE (qui indiquent le département) et on obtient pour chaque département la surface de croisement entre le département et chaque station météorologique. Il suffit ensuite de diviser par la surface totale du département pour obtenir le poids de chaque station météorologique dans chacun des départements.

Etape 4 : Recomposer les indices climatiques par département

Ayant obtenu les poids de chaque station météorologique pour chaque département, il suffit de procéder au calcul suivant pour tous les indices climatiques de la base (Tableau II-1) :

$$\text{Indice}_{\text{Département}} = \sum_{\text{Stations}} \text{Poids}_{\text{Station} \times \text{Département}} \text{Indice}_{\text{Station}}$$

A ce stade, on a obtenu pour chaque département des relevés météorologiques annuels.

On a donc une base pour chaque scénario climatique de la forme :

departement	annee	TAV	NAV	---	PXCDD
01	1977	9.094	5.367	---	15.515
01	1978	9.733	5.775	---	20.839
01	1979	9.975	6.365	---	13.478
01	1980	9.136	5.456	---	13.821
01	1981	9.960	5.746	---	18.659
---	---	---	---	---	---
95	2096	14.96	11.01	---	28.254
95	2097	15.23	11.47	---	26.191
95	2098	15.94	11.77	---	35.179
95	2099	14.81	11.08	---	21.055
95	2100	15.74	11.68	---	35.825

c. Premières analyses de données et traitements consécutifs

Maintenant que nos données ont été départementalisées, nous pouvons nous intéresser plus en profondeur aux valeurs qu'elles prennent et aux échelles de valeurs prises.

Nous pouvons déjà remarquer que nos données peuvent être fractionnées en 4 groupes.

Nom	Description
TAV	Température moyenne journalière (°C)
TNAV	Température minimale journalière (°C)
TXAV	Température maximale journalière (°C)
TRAV	Amplitude thermique journalière (°C)
TXQ90	Extrême chaud de la température maximale journalière (°C)
TXQ10	Extrême froid de la température maximale journalière (°C)
TNQ90	Extrême chaud de la température minimale journalière (°C)
TNQ10	Extrême froid de la température minimale journalière (°C)
HDD	degré-jours de chauffage (°C)
CDD	degré-jours de climatisation (°C)
SD	Nombre de journées d'été (jour(s))
TXND	Nombre de jours anormalement chauds (jour(s))
TNHT	Nombre de nuits anormalement chaudes (jour(s))
TR	Nombre de nuits tropicales (jour(s))
TNFD	Nombre de jours de gel (jour(s))
TXFD	Nombre de jours sans dégel (jour(s))
TNND	Nombre de jours anormalement froids (jour(s))
TNCWD	Nombre de jours d'une vague de froid (jour(s))
TXHWD	Nombre de jours d'une vague de chaleur (jour(s))
PAV	Précipitations journalières moyennes (mm)
RR	Cumul de précipitation (mm)
PINT	Précipitation moyenne les jours pluvieux (mm)
PFL90	Fraction des précipitations journalières intenses (%)
RR1MM	Nombre de jours de pluie (jour(s))
PN20MM	Nombre de jours de fortes précipitations (jour(s))
PXCWD	Nombre maximum de jours pluvieux consécutifs (jour(s))
PXCDD	Période de sécheresse (jour(s))

Figure II-7- Regroupements naturels des variables climatiques explicatives

En effet, les deux premiers groupes d'indices nous donnent des mesures sur les températures pour un département donné, l'un en degrés donc en valeur, l'autre en jour donc temporellement. On peut s'attendre à une très forte corrélation entre ces groupes et à l'intérieur de ces groupes. D'autre part, nous avons des données qui

concernent les précipitations, et de la même manière que pour les températures, des informations en termes de quantités (en mm ou fraction de mm de pluie) comme de temporalités (en jours) sont disponibles.

Centrage et réduction des données

La première remarque que l'on peut faire en observant les données est l'absence d'unité de mesure semblable que ce soit la mesure effective -°C contre mm par exemple- ou bien dans la façon de compter les valeurs -relevé de quantile contre de moyenne journalière de la température par exemple. Ainsi, pour pallier les problèmes que l'on pourrait avoir dans l'analyse des données (notamment s'agissant de la variance intra-groupe de ces données), nous centrons-réduisons les données.

Maintenant que le travail de normalisation a été fait, on peut procéder à une analyse en composantes principales dans le but d'étudier la colinéarité des variables climatiques ainsi que les groupements éventuels des départements que l'on pourrait faire. Pour une justification technique de la nécessité à centrer-réduire voir [28].

Analyse en Composantes Principales (ACP)

La première étape consistant à réduire et centrer les variables afin de ne pas brouiller les variances intra et inter-groupe (l'inertie dans le vocabulaire de l'ACP) étant faite, nous pouvons procéder à une ACP (on utilisera pour cela le package FactoMineR sur le logiciel R).

Affichons les pourcentages d'inertie expliqués par les différentes composantes principales obtenues selon les différents scénarios climatiques :

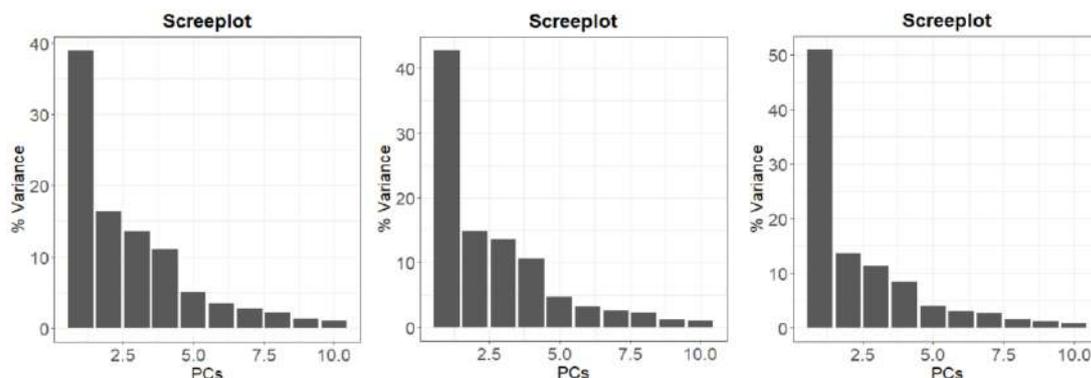


Figure II-8- Inerties des composantes pour les bases de données des scénarios RCP2.6, 4.5 et 8.5

Les 3 scénarios présentent des « screeplot » ou histogrammes d'inerties semblables, selon le critère du coude, les 4 premières composantes sont à conserver et les autres n'apportent que peu d'information.

De plus, la première composante est très informative, ainsi le graphique des variables sur les 2 premières dimensions sera très informatif quant aux relations entre les variables. Nous l'affichons

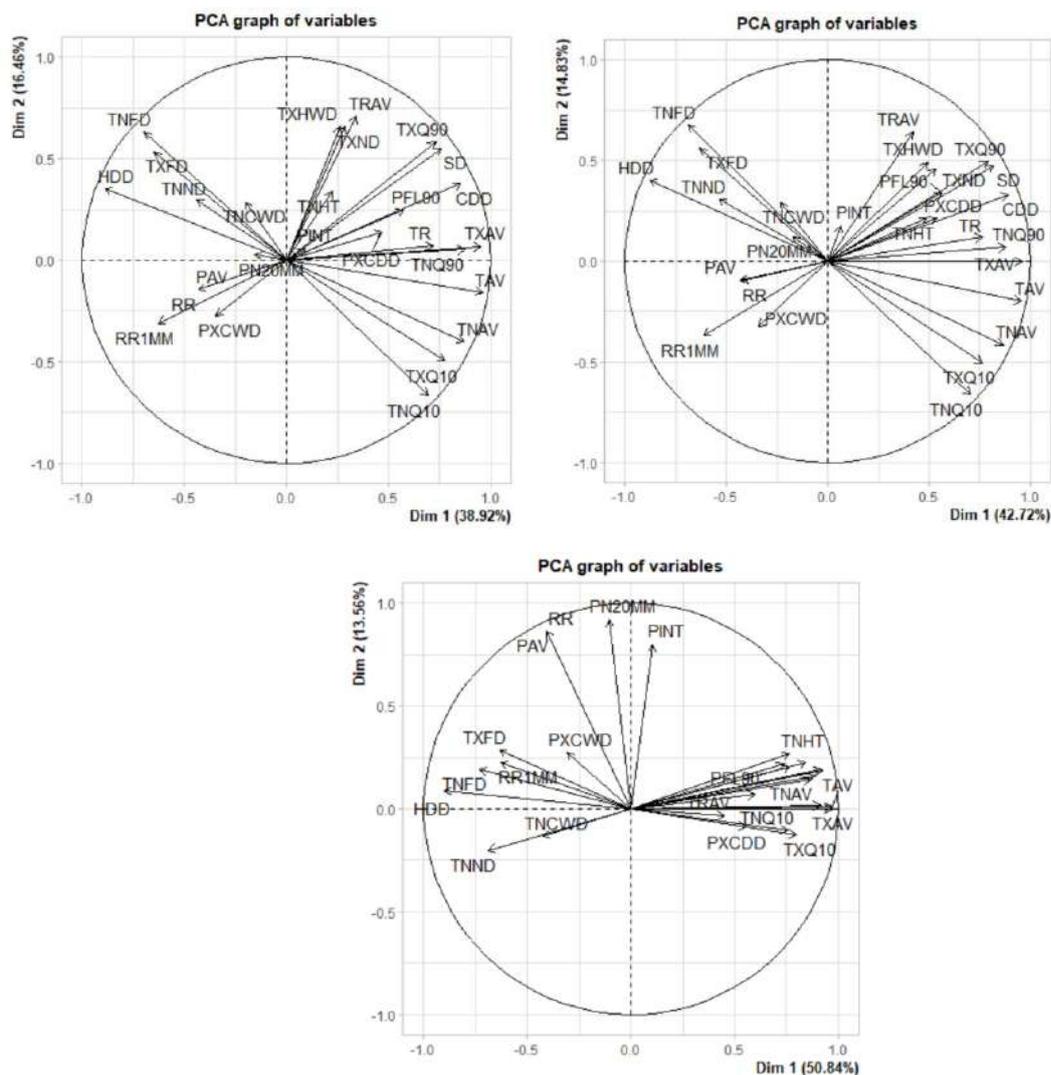


Figure II-9- Graphique des variables de chaque scénario climatique sur les 2 premiers axes principaux

Les 3 graphiques présentent à des degrés différents les mêmes faisceaux de variables. Ainsi, on peut faire l'analyse commune que les variables qui ont trait aux précipitations sont peu corrélées avec les autres variables mais assez fortement corrélées entre elles, surtout dans le cas du scénario RCP8.5.

De manière logique, les variables de température sont fortement corrélées entre elles également avec deux groupes s'opposant, les mesures de températures chaudes (jours anormalement chaud par exemple) et les températures faibles (jours de gels par exemple). Pour nuancer, seules quelques variables sont purement colinéaires (opposées dans le cercle) ce qui nous contraindra à devoir explorer toutes les possibilités de variable et d'utiliser des moyens de sélectionner les variables dans nos régressions de manière automatique.

Shift entre la période de référence et la période future

Nous observons un décalage important entre les valeurs des variables sur la période de référence (2000-2020) et la période prospective (2020-2050).

En effet, en moyenne 13% des valeurs dépassent les extrêmes climatiques (minimum ou maximum) de la période de référence sur la période prospective.

Variable	Nombre de dépassements d'extrêmes
TAV	1067
TNAV	512
TXAV	946
TRAV	3
TXQ90	1088
TXQ10	493
TNQ90	902
TNQ10	34
HDD	947
CDD	1272
SD	636
TXND	3109
TNHT	3639
TR	1007
TNFD	117
TXFD	1060
TNND	978
TNCWD	5322
TXHWD	2867
PAV	2
RR	2
PINT	7
PFL90	0
RR1MM	0
PN20MM	3
PXCWD	10
PXCDD	7
Pourcentage de valeur "hors-champ"	
13%	

De plus, 26% des enregistrements de la période prospective sont situés hors des centiles 5 et 95 de la période de référence. Ainsi, nous observons un décalage fort entre les données « passées » et « futures » estimées, ce qui nous obligera à opter au moins pour partie à une régression paramétrique pour gérer le cas de valeurs hors champs.

2. Données de sinistralité

a. Présentation de la base de données brute

La base de données de sinistralité dont nous disposons est une base de données composée de 541 361 enregistrements de sinistres survenus entre 1999 et 2020 provenant d'un acteur assurantiel implanté de manière homogène sur l'ensemble du territoire français avant 1998.

Chaque enregistrement est doté de 11 informations contenues dans les variables suivantes.

Variable	Description	Type	Valeurs possibles
année	Année de survenance du sinistre	Integer	1998-2020
réseau	Réseau par lequel le contrat lié au sinistre a été souscrit	String	'agent', 'courtier', 'salarié'
péril	Péril associé au sinistre	String	'tempête', 'rga', 'inondation', 'autre hydrolique', 'autre climatique'
lob	Line Of Business du contrat lié au sinistre	String	'MRC', 'MRH', 'RI', 'autre'
insee	Code INSEE du lieu du sinistre	String	Codes INSEE sur la France métropolitaine
risque_inondation	Risque d'inondation du site du sinistre échelonné	Integer	Niveau de risque (avis d'expert) : 1-6
risque_rga	Risque de sécheresse du site du sinistre échelonné	Integer	Niveau de risque (avis d'expert) : 1-5
charge	Charge en vision finale 2020 du sinistre	Double	Charge monétaire du sinistre actualisée en euros 2020 en vision finale (net de recours et provisionnements)
qualité_geocodage	Qualité du géocodage du site du sinistre	String	'Code postale', 'numero', 'voie', "
x	Si qualité géocodage suffisante, latitude du site du sinistre	Double	Coordonnées GPS
y	Si qualité géocodage suffisante, longitude du site du sinistre	Double	Coordonnées GPS

Tableau II-3- Variables de la base de sinistralité brute

La base de données bien que très complète va être modifiée pour se conformer au problème de régression que l'on souhaite résoudre, c'est-à-dire une régression au niveau du département sur le périmètre MRH en inondation.

b. Traitements réalisés sur la base

Etape 1 : Suppression des variables inutilisées

Le premier traitement que l'on fait sur la base est la suppression des variables dont on ne se servira pas dans la régression.

L'utilisation de données d'exposition étant trop spécifique au site des sinistres et non adaptée à la projection de sinistralité à l'échelle départementale, nous supprimons les variables *risque_inondation* et *risque_rga* de notre base.

Par suite et dans la même logique, nous supprimons les données relatives au géocodage précis des sinistres (*qualité_geocodage*, *x*, *y*).

Enfin, du fait que nous nous intéressons à la déviation globale du portefeuille sans étudier la source de souscription du sinistre, nous supprimons la variable *réseau*.

Etape 2 : Filtrage sur le périmètre d'intérêt

Du fait que nous nous intéressons à la déviation possible du portefeuille sur le périmètre Multi-Risque Habitation (MRH) et pour le péril inondation nous filtrons les données grâce à la variable *lob* en sélectionnant les enregistrements qui ont attiré au MRH. En outre, est filtré la base grâce à la variable *péril* en sélectionnant les

enregistrement dont la valeur est « inondation ». Après cette étape, nous pouvons supprimer les variables *lob* et *péril* devenues inutiles puisque de valeur unique.

Etape 3 : Suppression des enregistrements dont le code INSEE n'est pas renseigné
Comme nous souhaitons obtenir des résultats à l'échelle départementale, nous avons besoin de la géolocalisation des sinistres en code INSEE, ainsi les enregistrements dont le code INSEE est manquant sont supprimés.

Notons qu'il s'agit de l'unique champ conservé qui souffre d'un manque de quelques données, le problème n'étant pas grave car seulement 83 code INSEE ne sont pas renseignés sur un total de 25 309 enregistrements après filtrage soit 0,32% de la base. De plus, cela représente seulement 0,28% de la valeur totale de charge de la base.

Etape 4 : Transformation du code INSEE en code département

On extrait les deux premiers caractères de la chaîne de caractère de la variable *INSEE* donnant le code INSEE du lieu du sinistre afin d'obtenir les départements désignés par le code INSEE départementale que l'on assigne à la variable *departement*.

Etape 5 : Suppression des enregistrements de France d'Outre-Mer

Etant donné que nous ne disposons pas de données climatiques pour les outre-mers les enregistrements relatifs aux Outre-Mer, i.e. les enregistrements pour lesquels la variable nouvellement formée *departement est égale à 97*. Cela ne constitue que 14 enregistrements sur les 25 227 restants, soit 0,055% de la base.

A ce stade, la base a été nettoyée, filtrée et mise à échelle et nous obtenons une base de données relativement simple qui constituera la variable réponse de notre problème de régression que ce soit pour la prédiction de la fréquence ou de la sévérité. La base contient ainsi 25 213 enregistrements et est de la forme :

departement	année	charge
01	1999	0
01	1999	1924.95
01	2001	25876.32
01	2001	3151.83
---	---	---
95	2019	1878.94
95	2019	9304.2
95	2019	1788.5
95	2019	230
95	2020	4190

Tableau II-4- Base de données réponse nettoyée et filtrée pour la régression de la sévérité

Notons que les charges peuvent être nulles puisque nettes de recours et provisionnements, un sinistre ouvert peut n'avoir aucune charge associée dans le cas de provisionnements incorrects ou de recours compensant la perte. Cela étant dit, aucunes valeurs négatives ne sont présentes dans la base. Nous décidons de garder

ces sinistres nulles dans notre modèle en tant qu'ils sont informatifs quant à la fréquence de survenance des sinistres et sont peu nombreux (3,02% de la base) donc peu susceptible d'être vecteur d'un bruit important dans la projection de la sévérité.

Etape 6 : Création d'une base de fréquence et d'une base de sévérité

Pour la projection du coût des sinistres, nous utiliserons la base ainsi obtenue car elle associe **des** charges à un département et une année c'est-à-dire à un contexte climatique au sens de la base climatique construite.

Pour la projection de la fréquence des sinistres, nous allons procéder à un comptage d'enregistrements par département et par année. On obtient une base de la forme :

departement	année	freq
01	1998	0
01	1999	2
01	2000	0
01	2001	8
01	2002	5
---	---	---
95	2016	14
95	2017	3
95	2018	16
95	2019	5
95	2020	1

Tableau II-5- Base de données réponse nettoyée et filtrée pour la régression de la fréquence

c. Premières analyses de données et traitements

i. Fréquence

Résumé de données : moyenne, variance annuelle départementale

Sur la période de référence (1998-2020), nous avons les informations suivantes :

Moyenne	Ecart-Type	Quantile 75%	Quantile 95%
11.41	40.28	9	38.56

Tableau II-6- Récapitulatif de la distribution des fréquences annuelles

Il semble que la répartition des fréquences annuelles soit très complexe : on observe une grande dispersion des données et l'existence de valeurs extrêmes. Tentons de visualiser les données à l'échelle départementale.

Hétérogénéité de la fréquence selon les départements

Le premier élément que l'on peut remarquer est la normale hétérogénéité des fréquences de survenances de sinistres annuelles au sein des départements de la France Métropolitaine. Nous traçons les moyennes annuelles du nombre de sinistres sur la carte de France Métropolitaine.

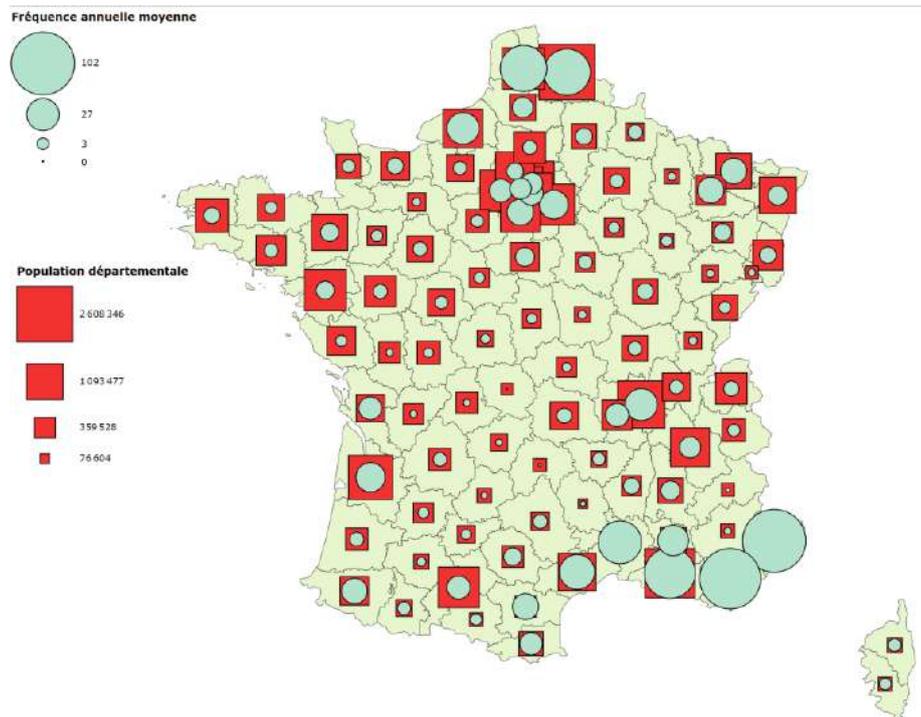


Figure II-10- Carte des fréquences annuelles moyennes en France métropolitaine contre la population pour chaque département

Deux éléments viennent expliquer l'hétérogénéité des fréquences parmi les départements. Le premier est la différente exposition liée aux populations assurées : les départements les plus habités souffrent naturellement d'un nombre de sinistres plus important. On le visualise sur la carte précédente sur les départements dans lesquelles les cercles et carrés se confondent. Le second élément d'explication est la différente exposition liée aux aléas naturels, ce qui se visualise par des cercles largement plus grand ou petit que les carrés sur la carte : il n'y a pas de correspondance normale entre nombre de sinistres et population.

Cette analyse induit deux choses, la première est que nous devons utiliser le département comme variable explicative de la fréquence car il contient intrinsèquement des informations sur la population et la seconde que l'aléa climatique dont nous avons une base représentative de données est également explicative et permettra certainement de rajouter de l'information au département quant à l'exposition naturelle au risque.

Un poids important en 0

Affichons un histogramme des valeurs prises par la variable de fréquence annuelle pour se rendre compte du poids de la valeur 0.

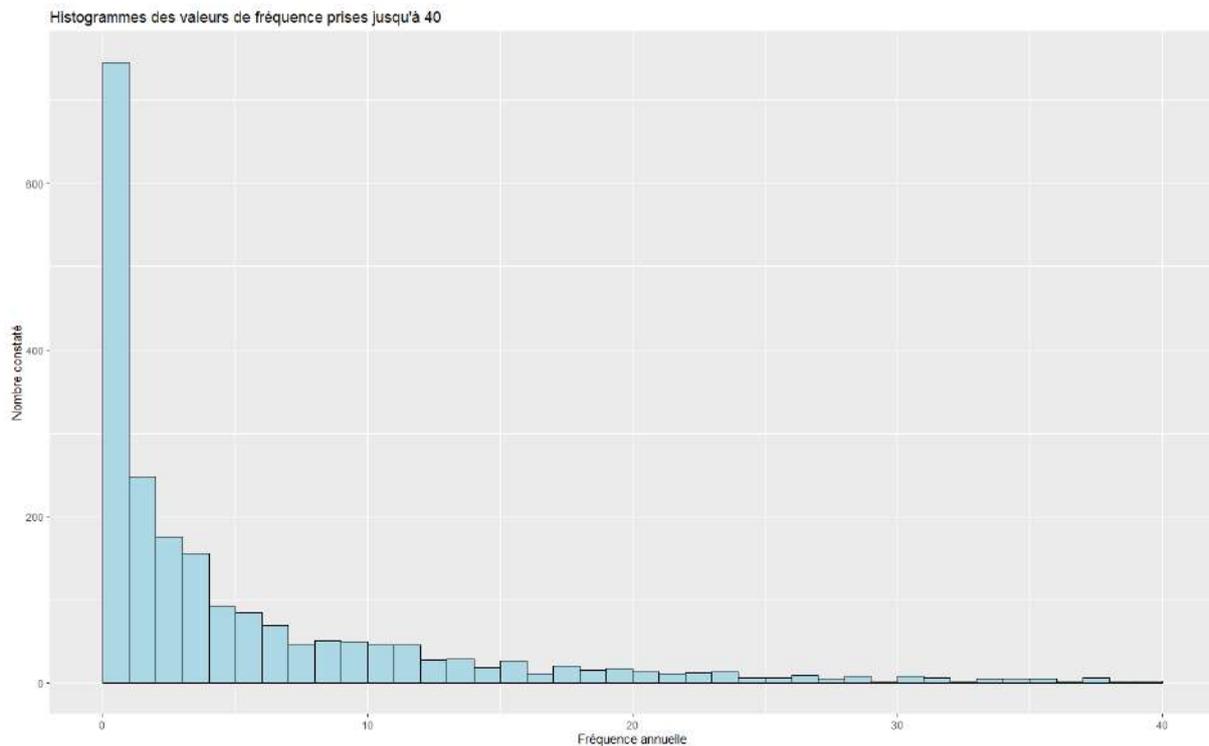


Figure II-11- Histogramme de répartition des fréquences annuelles départementales de sinistres

Ce graphique montre un important nombre d'années sans survenance de sinistre, ce poids en 0 devra être traité spécifiquement afin d'éviter le bruit dans un modèle de régression simple. Ainsi, on pourra voir la régression de fréquence comme un double problème : déterminer les années futures pour lesquelles aucun sinistre n'est déclaré et déterminer le nombre de sinistres en cas d'année sinistré.

Surdispersion de la variable de fréquence

Afin de savoir si les fréquences annuelles sont effectivement surdispersées, nous affichons un tableau récapitulatif des ratios par département entre les moyennes de fréquences annuelles et les variances de fréquences annuelles par département.

Min.	1er Qu.	Moyenne	3ème Qu.	Max.
1.000	4.513	32.439	29.286	498.271

Tableau II-7-Distributions des ratios de Variance / Moyenne des fréquences annuelles par département

Ce qu'on observe aisément c'est l'importance du ratio *Variance/Moyenne* ce qui implique une forte surdispersion.

On peut donc conclure que les données de comptage dont nous disposons ne sont pas Poissoniennes et que la régression que l'on sera amené à faire ne doit pas se faire dans une optique poissonnienne.

Années non informatives

Les graphiques suivants affichent la relation linéaire qu'il y a entre l'année et le nombre de sinistre. Pour mieux visualiser, on a mis un graphique avec et un graphiques sans valeurs « extrêmes » avec une droite de régression.

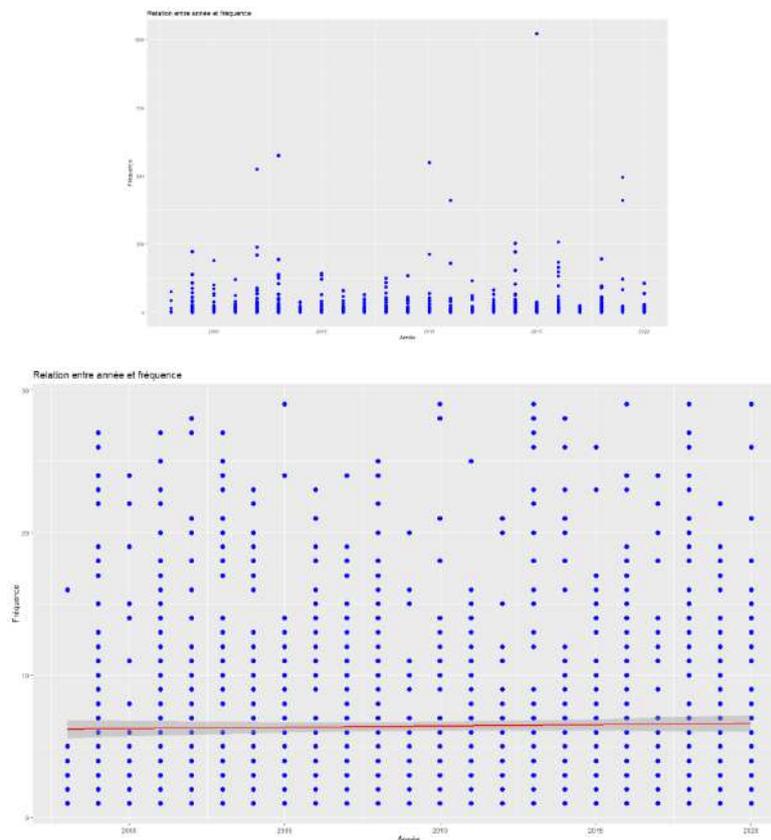


Figure II-12- Fréquences annuelles et années en nuage de point : pas de relations linéaires entre les deux variables

Très peu d'information ne semble être contenu dans les années, ce qui justifie l'enrichissement à un niveau local (départemental) de variables climatiques pour la prédiction de fréquences futures.

ii. Coût

Quelques macro-données

La base de données extraite nous permet d'avoir les premiers indicateurs suivant :

Min.	1er Qu.	Mediane	Moyenne	3ème Qu.	Max.
0.0	656.3	1855.0	7235.5	5500.1	972515.2

Tableau II-8- Répartition des charges historiques contenues dans la base de données

Comme pour la fréquence, ils nous indiquent l'existence de valeurs extrêmes ainsi que d'une répartition qui semble être importante sur la droite de la distribution.

On peut le vérifier en affichant l'histogramme des charges historiques :

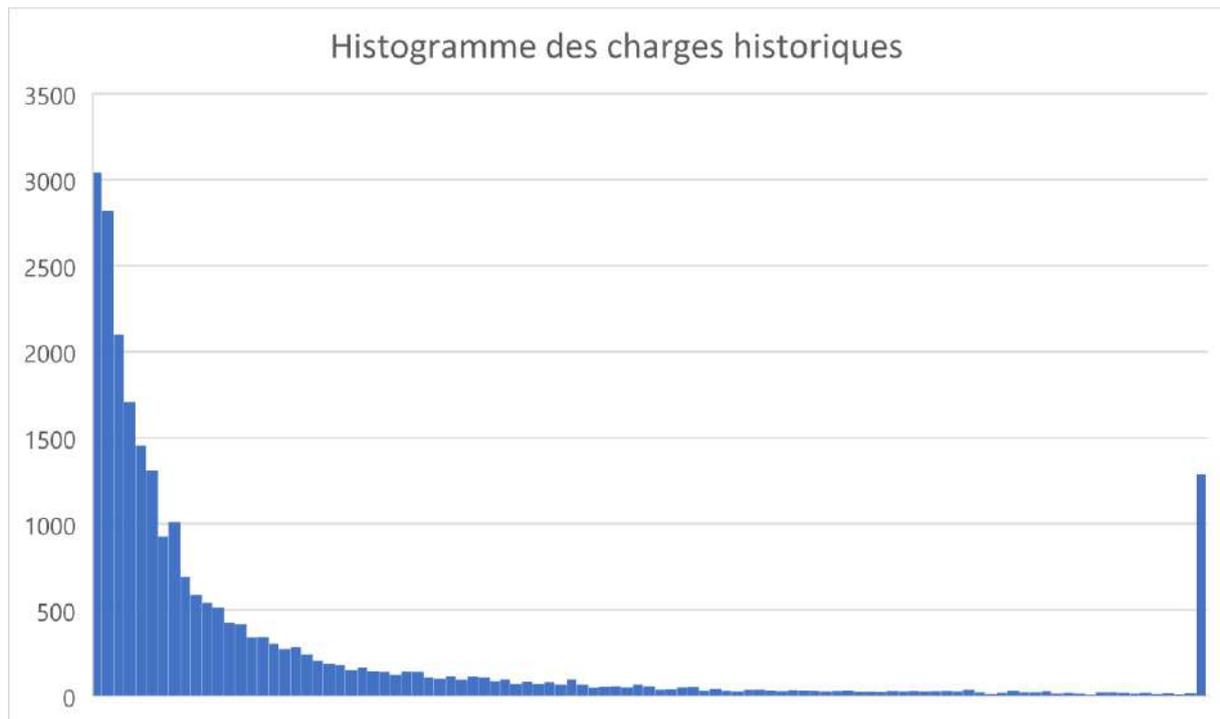


Figure II-13- Queue de répartition très lourde dans l'historique des charges

L'histogramme est un histogramme à largeur de classe fixe égal à 300€ et une classe de dépassement qui concentre toutes les valeurs au-delà de 30 000€.

On remarque d'abord sur ce graphique que la distribution des charges est dotée d'une queue extrêmement lourde, i.e. la distribution de charge appartient probablement à une loi à queue lourde comme les lois log-normales ou log-gammas.

Il s'agit maintenant de réduire la maille d'étude et de regarder si l'on peut établir des distributions au niveau du département ou de regroupements de départements.

Clustering

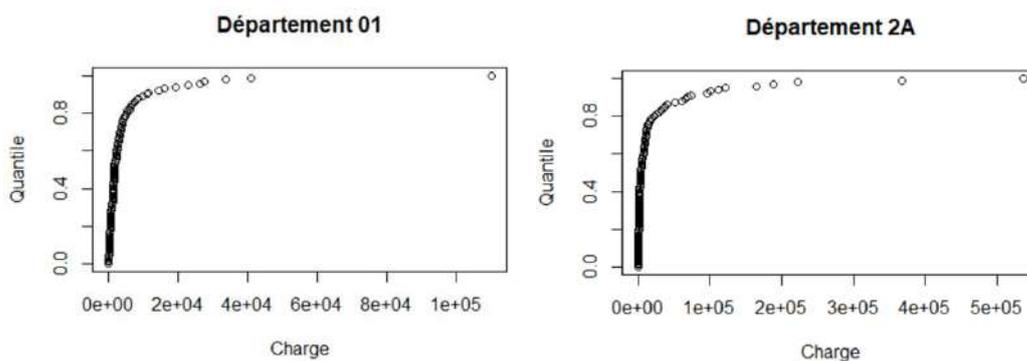
Avant de modéliser les charges futures de sinistralité en fonction du département et pour pallier le faible volume de données disponible pour certains départements peu soumis aux catastrophes naturelles il est bon de procéder à un « clustering » [11] c'est-à-dire un regroupement des départements considérés en fonction de l'historique de charge qui leur est lié. En somme, nous voulons regrouper les départements avec les mêmes dynamiques de sinistralité. Pour ce faire, nous procédons à une méthode en 3 étapes :

- Classification de sinistres graves et attritionnels
- Lissage des séries temporelles de charges attritionnels
- Clustering par méthode de Fréchet

En premier chef, nous commençons par séparer les attritionnels, c'est-à-dire les sinistres plus fréquents mais moins massifs en termes de charge, des sinistres qu'on appellera graves tout au long du mémoire, moins fréquents, moins prédictibles et surtout associés à des charges très élevées.

La raison pour laquelle nous faisons cette séparation est le manque de régularité des sinistres graves opposé au principe de clustering. L'hypothèse est forte car il serait également pertinent de regrouper les départements en fonction de leur dynamique d'augmentation de fréquence de sinistres graves, mais cela impliquerait une régularité parmi les graves pour pouvoir opérer au clustering alors qu'elle est difficilement trouvable et sans compter que cette variabilité pourrait augmenter avec le changement climatique. En bref, nous nous concentrons sur la distribution des sinistres au ventre de la distribution.

Plus concrètement, la séparation est faite en observant le centile de la distribution des charges après lequel on assiste à une explosion de variance pour chaque département, puis nous prenons le centile moyen. Par exemple les deux distributions suivantes nous amèneraient à prendre pour attritionnels les sinistres en dessous du 85e centile :



C'est en effet au-delà du 85e centile que l'on observe une forme de coude et que la variance augmente. De manière plus robuste, nous devons prendre le centile à partir duquel la courbe s'aplanit le plus. Pour cela il s'agit de calculer la dérivée discrète entre chaque centile. La même classification sera mise en place pour ajuster la distribution de sévérité future.

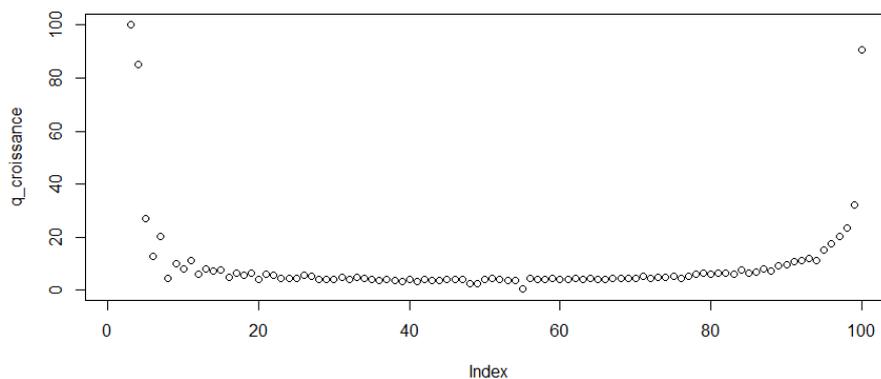


Figure II-14- q-croissance des charges historiques

Pour notre portefeuille le graphique de q-croissance ci-dessus nous indique de choisir comme seuil de graves le 85^e centile, centile pour lequel la courbe commence à se lever.

De là, une fois notre choix de centile fait, nous pouvons passer à l'étape suivante.

Le clustering que nous souhaitons faire se base sur des séries temporelles, ainsi, nous créons les séries temporelles des charges annuelles moyennes par département. Ensuite pour éviter le bruit lors du clustering, il est bon de lisser les courbes obtenues via une méthode de lissage.

Dans le cas présent, l'interpolation par polynôme par segment [12] (ou LOESS) permet de lisser les séries temporelles et d'éviter les données trop hétérogènes. Le paramètre à ajuster lors de l'utilisation de cette méthode est le « span » (ou fenêtre de lissage), plus le span est élevé plus la courbe est lissée, du fait de l'importante variance liée aux sinistres de catastrophes naturelles, le choix a été fait de lisser avec un span d'une valeur de 0.75 de manière arbitraire.

Une fois cette étape faite, il ne reste plus qu'à opérer au clustering.

De manière globale, le clustering, méthode d'apprentissage non-supervisé, consiste au regroupement en un nombre de groupes choisi d'éléments fonctionnels. Le partitionnement automatisé qui est fait dans une méthode de clustering se base sur une mesure de proximité ou distance : le but est de créer des groupes les plus éloignés possibles avec des éléments à l'intérieur de ses groupes les plus proches possibles. On parle de minimisation de l'inertie intra-classe et de maximisation de l'inertie inter-classe. Le point le plus important est donc de choisir une distance qui soit en adéquation avec l'objectif du clustering. Dans notre cas, il faut déterminer une distance entre des séries temporelles, i.e. entre des courbes (après lissage).

Plusieurs mesures de proximités entre séries temporelles existent, ici, nous utiliserons la distance de Fréchet [13]. La distance de Fréchet mesure la plus petite corde entre deux courbes qui permet de relier à tout moment les deux courbes, elle prend ainsi en compte l'emplacement et l'ordre des points le long des courbes ce qui nous intéresse dans la mesure où nous cherchons à relier des dynamiques.

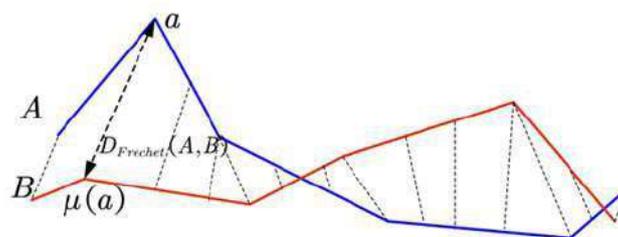


Figure II-15- Schéma explicatif de la distance de Fréchet

A l'issu de la méthode nous pouvons assigner à chaque département un cluster (i.e. un groupe), ce qui nous sera utile dans la modélisation des charges de sinistralité futures.

Notons pour finir que pour notre historique si l'on s'abstient à l'inondation sur les contrats MRH, le clustering le plus pertinent semble être celui en 5 classes car il permet de rendre compte de dynamiques de sinistralité fines tout en écartant les groupements inutiles.

Les courbes brutes de clustering sont les suivantes :

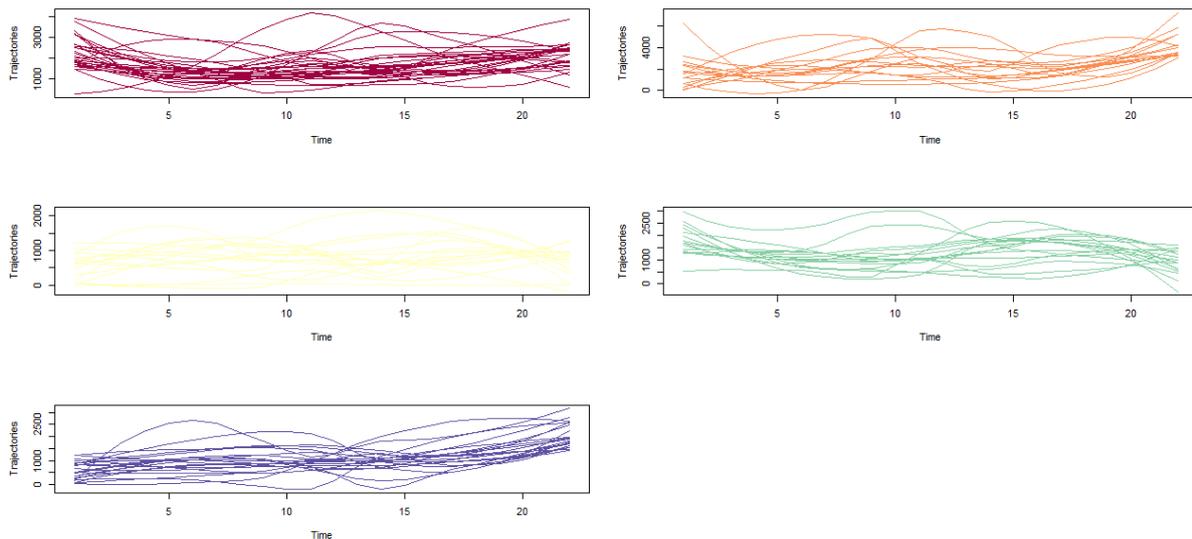


Figure II-16- Clusters de départements à tendances de sinistralité différentes

Les classes présentées via le graphique ci avant nous indiquent 5 classes homogènes de risque en regardant les bornes et les tendances des courbes :

- Le cluster rouge regroupe les départements à la sinistralité élevée stable
- Le cluster orange regroupe les départements à la sinistralité élevée croissante
- Le cluster jaune regroupe les départements à faible sinistralité stable
- Le cluster vert regroupe les départements à sinistralité moyenne stable
- Le cluster bleu regroupe les départements à sinistralité moyenne croissante

Cette validation graphique n'étant pas suffisante, on regarde si les appartenances des départements dans les différents cluster est attendu par rapport à la carte de sinistralité historique.

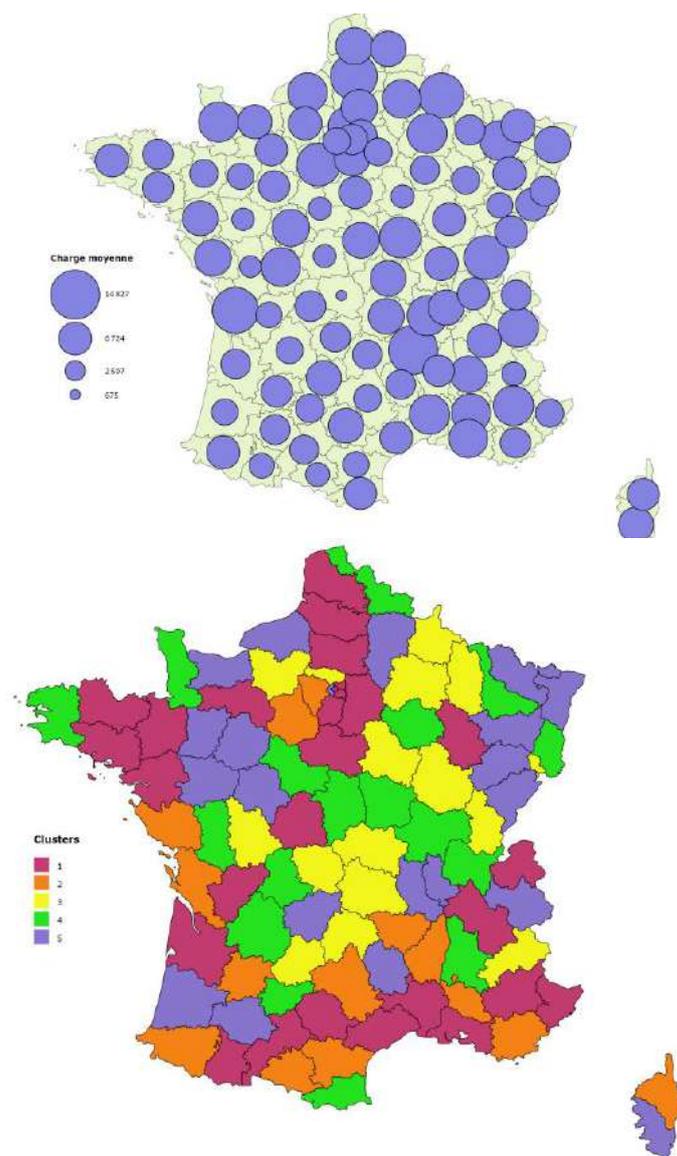


Figure II-17- Carte des clusters de département selon le risque inondation

Géographiquement, nous retrouvons relativement fidèlement les départements concentrant le risque dans la première carte : l'arc sud-sud ouest est bien représentée dans les clusters à hautes pertes et le milieu de la France est bien dans les clusters associant moins de pertes.

Ainsi, les clusters ainsi produits semblent cohérents.

Vérification de la cohérence du risque au sein des clusters et affiliation à la loi gamma

La première représentation que l'on peut se donner est la suivante :

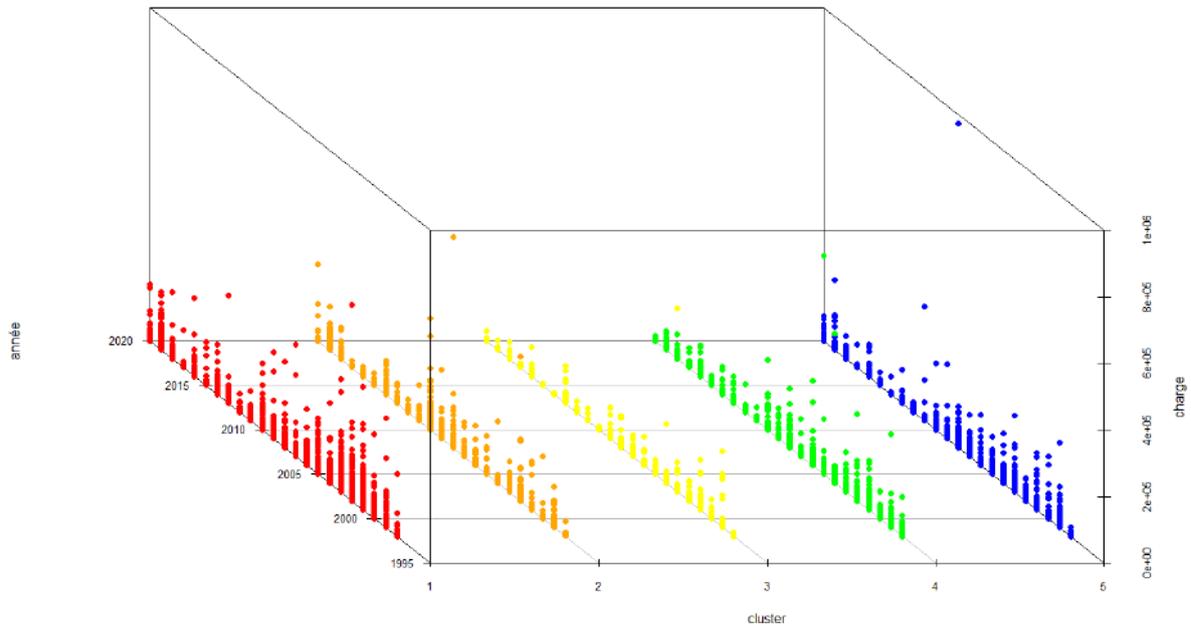


Figure II-18-Pertes par clusters par années

Si le graphique n'est pas complètement informatif, on peut néanmoins distinguer les différents clusters via la hauteur des pics qu'ils produisent, ainsi, le risque semble homogène au sein des différents clusters.

Pour continuer l'analyse, nous allons tenter d'ajuster des lois log-gammas pour chaque cluster afin de déterminer si le risque est bien homogène.

Pour déterminer le meilleur ajustement, nous réalisons des tests de Kolmogorov-Smirnov après avoir ajusté via différentes méthodes. Ci-dessous un tableau récapitulatif.

Ajustement	MLE	MME	MGE
Descriptif	Maximum de vraisemblance	Méthode des moments	Maximisation de la GOF
Cluster	Tests de KS		
1			
2			
3			
4			
5			

Tableau II-9- Tests d'ajustement réalisés sur les charges par cluster

Ainsi, chaque cluster peut être assimilé à une loi log-gamma spécifique, ce qui induit l'homogénéité du risque dans les clusters et nous permet ainsi de les utiliser dans le cadre d'une régression non-paramétrique.

Nous allons utiliser la maximisation de la Goodness-Of-Fit afin de réaliser nos ajustements.

Afin de vérifier que les distributions sont bien ajustées nous regardons les différents QQ-plot que les distributions gammas sur les logs-charges par cluster ont produit.

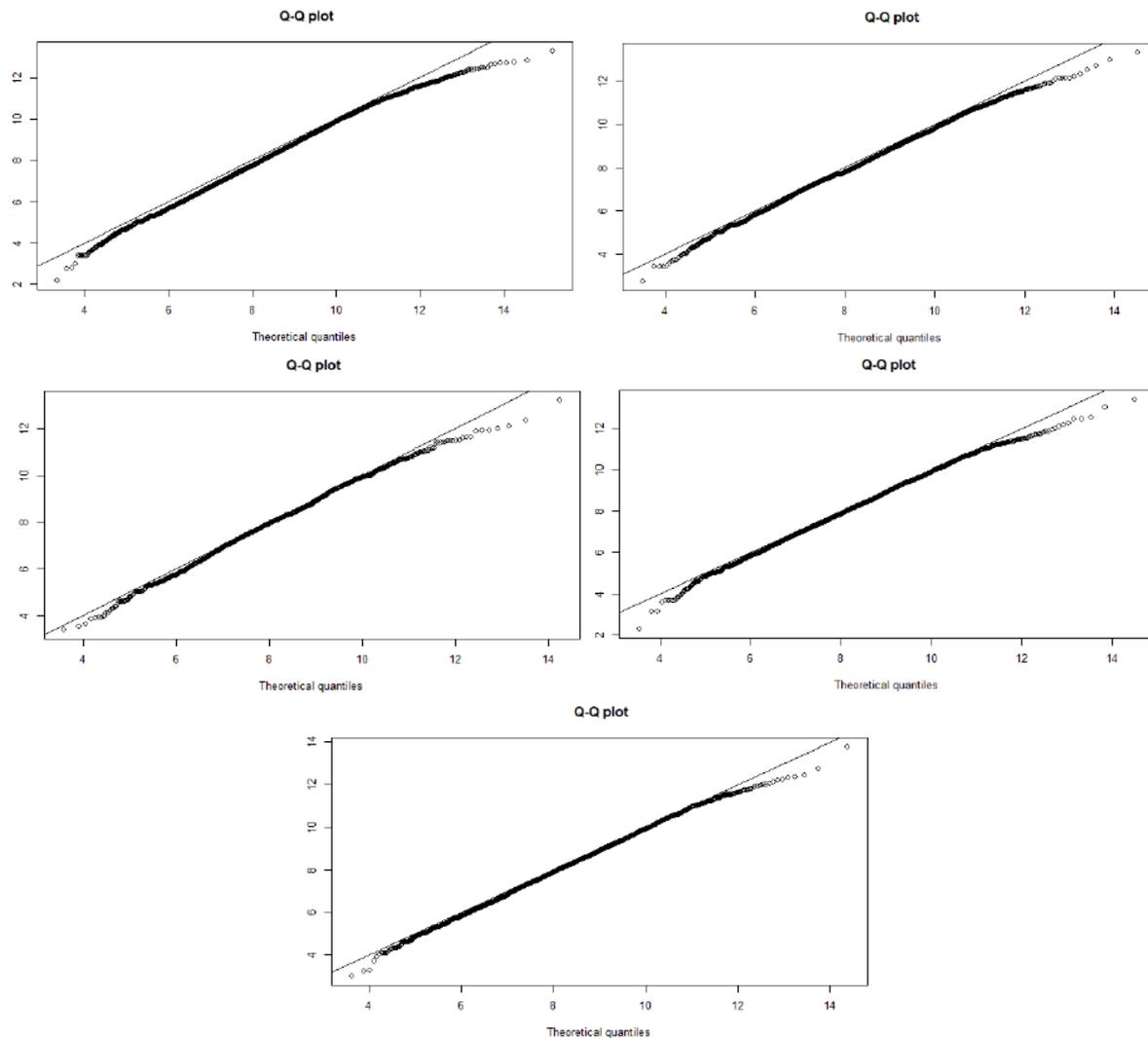


Figure II-19- QQ-plot des ajustements de lois gammas sur chacun des clusters (dans l'ordre)

On remarque sur les graphique du QQ-plot les mêmes réactions aux ajustements pour chacun des clusters. La droite de référence et les points sont bien alignés de manière globale ce qui rassure sur la qualité de l'ajustement et ainsi sur l'homogénéité du risque dans chaque cluster.

Néanmoins, on remarque également une surévaluation du risque légère sur chacun des extrêmes de la distribution et le début des distributions des cluster 1, 2 et 3. Dès lors, il pourra être pertinent de penser à « caper » la distribution modélisée dans le

cadre de pertes trop importantes générées par un modèle de régression s'appuyant sur les lois log-gammas. D'autre part, on pourra utiliser une régression non-paramétrique par cluster pour les débuts de distribution qui sera potentiellement plus fidèle à la réalité.

Pour finir, les valeurs des paramètres trouvées sont les suivants :

Cluster	Shape	Rate
1	23,9882	3,135582
2	27,96985	3,644241
3	22,31823	2,881665
4	25,36545	3,279816
5	26,3105	3,392212

Tableau II-10- Paramètres des lois gammas associées à chaque cluster

Notez qu'il s'agit ici des paramètres de la loi gamma sous-jacente. On remarque que les formes sont bien différentes pour chaque cluster, ce qui montre l'inhomogénéité inter-classes des clusters et justifie l'utilisation de cluster plutôt que d'une étude de la distribution globale.

Année informative quant à la charge annuelle

Le taux de corrélation linéaire entre les charges annuelles par département et la variable année est de **-14,13%**, ainsi l'année concentre une certaine quantité d'information. On peut attribuer cette corrélation négative à des politiques de gestion de risques plus importantes au fur et à mesure des années. Quoiqu'il en soit, il s'agit d'une variable qui pourra être utile à une régression mais pourrait être source d'erreur importante sur des horizons lointains pour lesquels la variable année prend des grandes valeurs, ainsi, on l'écarte de nos variables candidates.

3. Analyse des données jointes

Commençons par afficher les corrélations linéaires entre les fréquences annuelles par département et les variables climatiques.

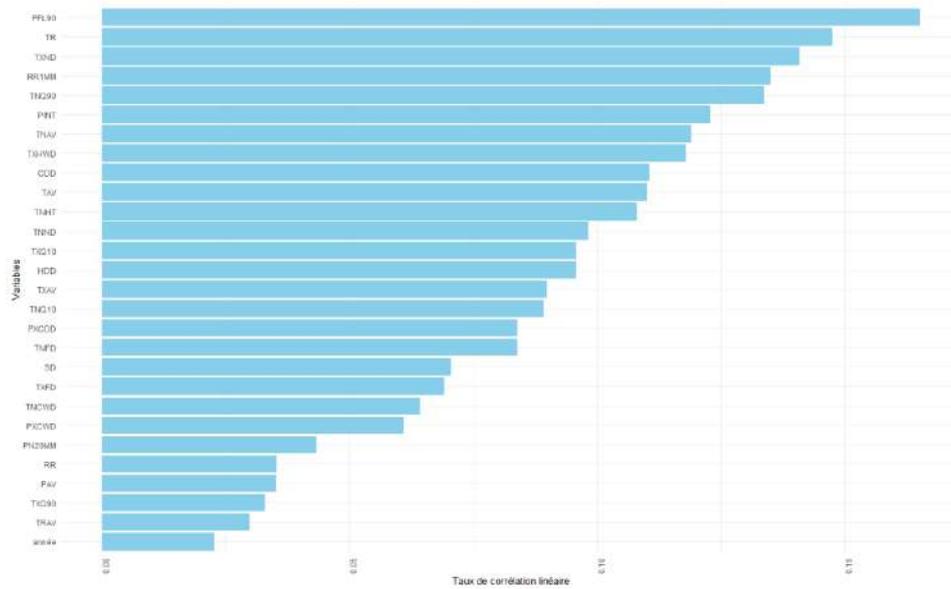


Figure II-20- Corrélations absolues disparates parmi les indicateurs climatiques face à la fréquence

On observe que les variables les plus informatives proviennent logiquement du groupes des indicateurs climatiques sur les précipitations. Notamment la variable la plus corrélée qui est la fraction de pluie intense de l'année définie par le Drias comme la fraction de pluie produisant plus de 90mm soit une pluie extrême.

Néanmoins, est aussi présenté sur ce graphique qu'aucune variable est très informative, et que, de même, on ne peut pas exclure les variables les moins corrélées à ce stade.

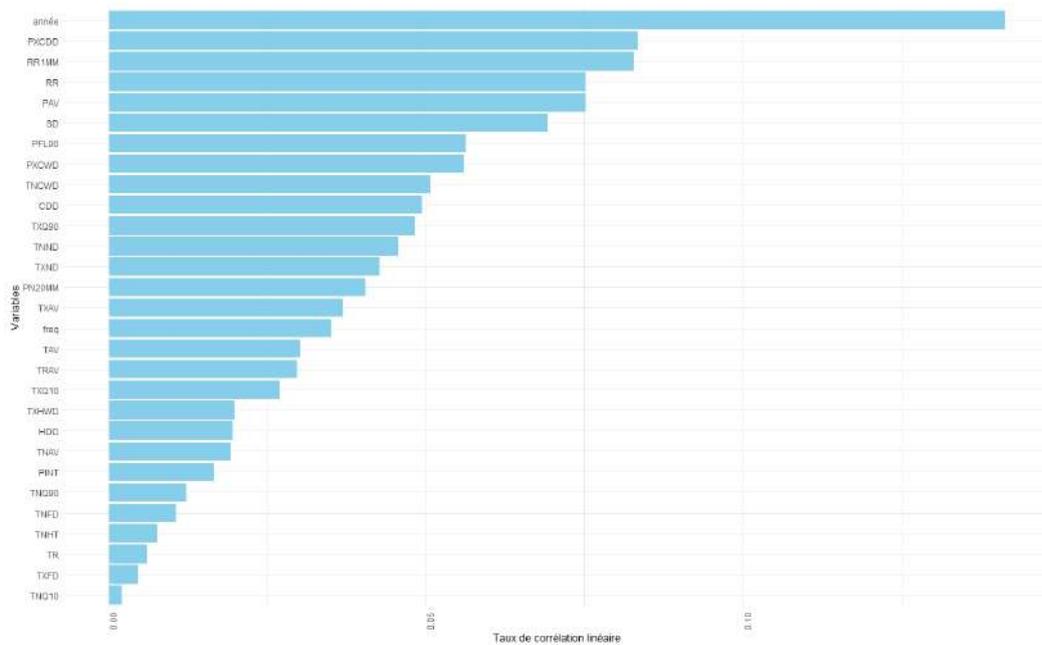


Figure II-21- Corrélations absolues des indicateurs climatiques avec les charges annuelles

De même que pour la fréquence, les données de précipitation sont en tête de classement des corrélations. En revanche, on peut remarquer que les variables sont globalement moins informatives, notamment les variables relatives à la température journalière (TN...). Ainsi, on pourra considérer de ne pas utiliser ces variables comme pivots dans nos régressions.

4. Conclusion générale sur l'analyse des données

Brièvement, après analyse de données, nous décidons de mettre en place un modèle double pour la fréquence : un modèle de classification pour gérer le problème du poids important de la distribution en 0, puis un GLM pénalisé de type binomiale négative pour gérer les données surdispersées ainsi que de procéder à un choix automatisé des variables de régression étant donné leurs faibles taux d'information.

De même, étant donnée la relative cohérence des charges au sein des clusters nous allons utiliser les clusters comme source de régression : chaque département appartenant à un cluster aura pour courbe de sévérité celle du cluster. Dans cette optique, étant donné la forte queue de distribution observée, nous séparerons la régression en deux : une régression non-paramétrique pour l'ensemble gauche de la distribution et une régression paramétrique pour la queue de distribution utilisant les lois log-gammas ajustées.

III. Modélisation de l'évolution du risque inondation

Le modèle que nous allons proposer a pour but d'évaluer le risque climatique, comme on l'a abondamment énoncé en première partie, le modèle doit donc être en mesure de matérialiser la distribution future des charges de sinistralité, du ventre de la distribution (car la moyenne est un indicateur absolument nécessaire à l'évaluation du risque) à la queue à droite de la distribution dont la connaissance met en lumière la solvabilité de l'entreprise.

Pour pouvoir avoir des estimations de la distribution des charges futures par année, le modèle créé procède ainsi à la simulation des charges futures (en se conformant à plusieurs hypothèses statistiques). Pour ce qui est de la structure du modèle, elle est composée de quatre modules opérationnels :

- Le module de fréquence des sinistres futurs
- Le module de sévérité de sinistres attritionnels futurs
- Le module de sévérité de sinistres graves futurs
- Le module d'assemblage et simulation des données

Le modèle que nous appliquerons aux données disponibles est présenté ici d'un point de vue théorique. Nous tenterons de souligner les différentes hypothèses faites et leurs implications.

Commençons par introduire la variable aléatoire réelle X_d qui représente la charge associée à un sinistre unitaire pour la maille géographique d avec $d \in [[1; D]]$ et D un entier représentant le nombre total de zones géographiques. On a $X_d : \Omega \rightarrow R$ et nous observons la variable sur t années de telle sorte que nous observons en réalité le processus $(X_{d,1}, X_{d,2}, \dots, X_{d,t})$ pour chaque maille d .

Par analyse de séries temporelles des processus $(X_{1,1}, \dots, X_{1,t}), \dots, (X_{D,1}, \dots, X_{D,t})$ sont rassemblé les séries de même tendances moyennes (il s'agit ici d'une hypothèse forte que de s'intéresser à la moyenne que l'on a justifié dans la partie clustering) de telle sorte que l'on forme C clusters rassemblant c_1, c_2, \dots, c_C séries temporelles chacun.

On fait ensuite l'hypothèse d'assimilation des séries au cluster, ainsi, au sein du cluster $c \in [[1; C]]$ doté de c_c éléments on pose :

$$(X_1^c, \dots, X_t^c) \approx \begin{cases} (X_{1c,1}, \dots, X_{1c,t}) \\ \vdots \\ (X_{c_c,1}, \dots, X_{c_c,t}) \end{cases}$$

Cette forte hypothèse a pour objectif de faire un compromis entre robustesse des estimations statistiques (en augmentant le nombre d'observations en rassemblant les observations de même tendance) et conservation d'une granularité intéressante.

Dès lors, on a à disposition, au sein de chaque cluster c , c_c observations du processus (X_1^c, \dots, X_t^c)

Introduisons maintenant la variable $Y : \Omega \rightarrow R^v$ où $v \in N^*$ qui représente la ou les variables climatiques (selon si $v \geq 1$). On fait maintenant l'hypothèse que pour chaque cluster $c \in [[1; C]]$ le processus (X_1^c, \dots, X_t^c) est assimilable à une unique variable $X^c | Y$ dont on dispose de $c_c \times t$ observations :

$$\overbrace{X^c | Y = y_1, \dots, X^c | Y = y_1}^{c_c \text{ fois}}, \overbrace{X^c | Y = y_2, \dots, X^c | Y = y_2}^{c_c \text{ fois}}, \dots, \overbrace{X^c | Y = y_t, \dots, X^c | Y = y_t}^{c_c \text{ fois}}$$

De là, connaissant les valeurs que prendra Y en $t + 1, \dots, t + k$ avec $k \in N^*$ du fait qu'elles sont des entrants pour le modèle, on peut estimer la distribution (et les indicateurs statistiques associés) à la variable $X^c | Y = y_{t+j}$ (pour tout j dans $[[1; k]]$).

Pour une meilleure estimation, on procède à la décomposition suivante (au seuil $Q_c \in R$) :

$$(X^c | Y = y) = (X^c | Y = y) \mathbb{1}_{X^c | Y = y \leq Q_y^c} + (X^c | Y = y) \mathbb{1}_{X^c | Y = y \geq Q_y^c}$$

$$(X^c | Y = y) = {}^{np}X_y^c + {}^pX_y^c$$

Ceci nous permet de séparer le comportement de la variable en deux variables, l'une représentant la distribution d'attrition des montants de sinistre, l'autre représentant la distribution grave des montants de sinistre. Il est important de procéder à cette décomposition dans la mesure où l'on veut proposer un modèle semi-paramétrique pour les raisons qui suivent. Pour la partie attritionnels la distribution dépendra directement de l'historique du fait qu'on utilisera une méthode d'inversion par quantile conditionnel pour se conformer à la fois au climat variant et un historique donné. Cette manière d'évaluer la distribution de sévérité future ne sera donc pas en mesure de capter des montants de sinistres pour des valeurs y_{t+j} hors champ par rapport à l'historique, c'est pourquoi pour la partie grave, le modèle devra être capable de gérer des nouvelles valeurs extrêmes de y_{t+j} et sera ainsi paramétrique.

Il reste à dire quelle est la valeur du seuil Q_c que nous nous donnons comme seuil de séparation entre le comportement d'attrition et le comportement grave de la variable et de préciser ainsi la forme des variables présentes dans la décomposition.

Pour établir Q_c , on fait deux hypothèses, la première est l'hypothèse selon laquelle le seuil historique de séparation entre montants de sinistres graves et d'attrition est conservée (i.e. le seuil Q_c déterminé par l'historique reste le même dans les périodes futures), la seconde est que le seuil historique est déterminé en observant la « q-croissance ». On définit ici la « q-croissance » comme suit :

Soit $Q_y^c(q) = F_{(X^c|Y=y)}^{-1}(q)$ la fonction quantile associée à la distribution $(X_c|Y=y)$, et $p \in [0; 1]$ un pas d'évaluation de la fonction tel qu'il existe M qui vérifie que $Mp = 1$. On détermine alors les points $\omega_m = (Q_y^c(pm) - Q_y^c(p(m-1))) / Q_y^c(pm)$ pour m dans $[1; M]$, on appellera cet ensemble de points la q-croissance de la distribution $(X_c|Y=y)$ du fait qu'ils décrivent l'évolution de la fonction quantile rapportée à elle-même, en effet, si un point est important en valeur cela signifie ainsi que la progression du quantile sur la tranche précédente a été non linéaire et donc forte. Ainsi, on regarde à partir de quel point w_m^* on observe un soulèvement de la valeur par rapport au point précédent et décidons ainsi du seuil que nous appliquerons entre les montants de sinistres graves et d'attrition en prenant $Q_y^c = Q_y^c(pm^*)$. On notera également $q_c = pm^*$.

On peut maintenant décrire plus précisément les variables issues de la décomposition de la distribution $(X_c|Y=y)$.

$$\begin{aligned} {}^{np}X_y^c &= F_{(X^c|Y=y)}^{-1}(U) \mathbb{1}_{F_{(X^c|Y=y)}^{-1}(U) \leq Q_y^c} \\ &= q_y^c F_{(X^c|Y=y)}^{-1}(U) | F_{(X^c|Y=y)}^{-1}(U) \leq Q_y^c \\ {}^pX_y^c &= (1 - q_y^c) G_y^c | G_y^c \geq Q_y^c \end{aligned}$$

avec $U \rightarrow U[0; 1]$ une loi uniforme sur le segment unitaire et $G \rightarrow \exp(G(\alpha_c, \beta_c))$ une loi log-gamma dont les paramètres dépendent de c et y .

On voit ici que la distribution de sévérité est donc celle d'un modèle composite avec un point d'attachement (ou slicing point) à Q_y^c et des poids de q_y^c pour la variable d'attrition et symétriquement $(1 - q_y^c)$ pour la variable de grave.

Ici, l'aspect semi-paramétrique du modèle est ainsi bien visible. Nous rentrerons plus dans les détails des estimations des paramètres et statistiques d'intérêt par la suite, par contre, on peut déjà dire ici que l'on a complètement caractérisé le modèle théorique du point de vue de la sévérité i.e. des montants de sinistre du fait que la distribution est maintenant décrite. Il reste maintenant à introduire la notion de fréquence des sinistres i.e. le nombre de sinistres dans le modèle.

D'un point de vue pratique les méthodes utilisées pour déterminer les fréquences de sinistres futures ne permettent pas d'obtenir la distribution des fréquences futures de sinistre mais seulement une valeur espérée. Ainsi, on a à disposition d'un point de vue théorique, pour tout d dans $[[1; D]]$, c le cluster associé et j dans $[[1; k]]$ l'entier $n_{d,t+j}$ égal au nombre de sinistres attendus de sévérité $X_{d,t+j} = X^c | Y = y_{t+j}$. Dès lors, le

modèle ainsi fourni permet de pouvoir s'intéresser à la variable de charge totale sur une année dans une zone donnée :

$$\begin{aligned}
 Z_{d,t+j} &= \sum_{n=0}^{n_{d,t+j}} X_{d,t+j} \\
 &= \sum_{n=0}^{n_{d,t+j}} {}^{np}X_{y_{t+j}}^c + {}^pX_{y_{t+j}}^c \\
 Z_{d,t+j} &= \sum_{n=0}^{n_{d,t+j}} q_{y_{t+j}}^c F_{(X^c|Y=y_{t+j})}^{-1}(U) | F_{(X^c|Y=y_{t+j})}^{-1}(U) \leq Q_{y_{t+j}}^c + (1 - q_{y_{t+j}}^c) G_{y_{t+j}}^c | G_{y_{t+j}}^c \geq Q_{y_{t+j}}^c
 \end{aligned}$$

A ce stade, on obtient aisément dans le cadre classique du modèle collectif et par indépendance des lois tronquées symétriquement à gauche et à droite :

$$\begin{aligned}
 \mathbb{E}(Z_{d,t+j}) &= n_{d,t+j} \left(\mathbb{E}({}^{np}X_{y_{t+j}}^c) + \mathbb{E}({}^pX_{y_{t+j}}^c) \right) \\
 \mathbb{V}(Z_{d,t+j}) &= n_{d,t+j} \left(\mathbb{V}({}^{np}X_{y_{t+j}}^c) + \mathbb{V}({}^pX_{y_{t+j}}^c) \right)
 \end{aligned}$$

L'espérance et la variance de ${}^{np}X_{y_{t+j}}^c$ ne pouvant être obtenue que par simulation grâce aux estimateurs classiques de l'espérance et la variance sur ces simulations (par exemple la moyenne pour l'espérance et la somme des carrés des écarts à la moyenne divisé par le nombre de simulations moins un pour la variance) et l'espérance et la variance de ${}^pX_{y_{t+j}}^c$ étant donné par le calcul que nous expliciterons par la suite.

D'autre part, le modèle composite étant connu, on peut notamment trouver la formulation dans le papier [14], on a la distribution de $X_{d,t+j}$ caractérisée par sa fonction de répartition $F_{X_{d,t+j}}$:

$$F_{X_{d,t+j}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ F_{(X^c|Y=y_{t+j})}^{-1}(x) & \text{si } 0 < x < Q_{y_{t+j}}^c \\ q_{y_{t+j}}^c + (1 - q_{y_{t+j}}^c) \frac{F_{G_{y_{t+j}}^c}(x) - F_{G_{y_{t+j}}^c}(Q_{y_{t+j}}^c)}{1 - F_{G_{y_{t+j}}^c}(Q_{y_{t+j}}^c)} & \text{si } x > Q_{y_{t+j}}^c \end{cases}$$

La connaissance de cette distribution sera par ailleurs fondamentale dans l'établissement par méthode de rééchantillonnage de la distribution $Z_{d,t+j}$.

Disons avant de terminer que d'un point de vue pratique, pour éviter des écueils liés notamment à une mauvaise projection climatique (sur laquelle on n'a pas la main) influant sur les paramètres de la loi G_c ou à une mauvaise projection de ces paramètres qui aurait pour conséquence une explosion des coûts des sinistres graves, la variable G_c est « capée » à une valeur K . En somme, on fait en sorte

qu'un sinistre unitaire ne puisse pas exprimer un montant supérieur à K , ce qui équivaut à faire valoir en pratique ${}^pX_y^c = \max\left(G_y^{cl}, G_y^c \geq Q_y^c, K\right)$.

En outre, toujours d'un point de vue pratique, pour permettre de meilleures interprétations plus robustes et ne pas tomber dans la trop grande ambition de vouloir s'intéresser à des périodes de temps trop faibles d'un futur très lointain et incertain, on pourra assimiler les distributions de telle sorte que l'on pourra s'intéresser à la distribution $Z_{d,t+j \leftrightarrow t+j'}$ (par exemple la distribution de charge annuelle du département de l'Aube entre 2030 et 2035 plutôt que 2030) égale à $\frac{1}{j'-j+1} \sum_{l=j}^{j'} Z_{d,t+l}$.

Pour finir, en considérant que la corrélation entre les zones géographiques est intrinsèquement prise en compte d'une part du fait de l'historique « corrélé » et d'autre part par l'utilisation de variables climatiques extérieures rendant compte de la corrélation voire de la causalité des sinistres entre les zones géographiques, on peut, pour obtenir des distributions globales ou moins granulaires procéder à la sommation sur d des distributions locales.

Maintenant que le modèle théorique d'obtention des distributions de charges futures a été décrit, intéressons-nous plus en détail au modèle sous-jacent d'obtention du nombre de sinistres futures et au modèle sous-jacent d'obtention de la distribution du montant des sinistres unitaires futurs (non- paramétrique et paramétrique).

A. Projection de la fréquence

1. Méthodologie

La modélisation des fréquences futures d'occurrence de catastrophes naturelles à une granularité départementale est difficile dans la mesure où il existe un grand nombre de valeurs nulles que l'on observe une sur-dispersion des valeurs sur plusieurs années et enfin que la dimension du problème est importante du fait du grand nombre de données climatiques dont on dispose. La stratégie que l'on adopte est donc de réduire le problème à deux problèmes : la prédiction des années durant lesquelles les départements seront sinistrés et la fréquence, en cas d'occurrence, des sinistres dans les départements touchés par année.

Pour le premier pan de notre modèle, nous utiliserons une forêt aléatoire : il s'agit d'un modèle particulièrement adapté à la grande dimension car il ne souffre pas de biais de dimensionnalité. En effet la forêt aléatoire est dans notre cas, où la variable à prédire est booléenne donc catégorielle, un ensemble d'arbres de décisions façonnés sur des variables piochées aléatoirement pour chaque arbre.

Rentrons dans le détail.

Commençons par définir ce qu'est un arbre de décision ou un arbre CART[15]. Un arbre CART est un arbre de la forme :

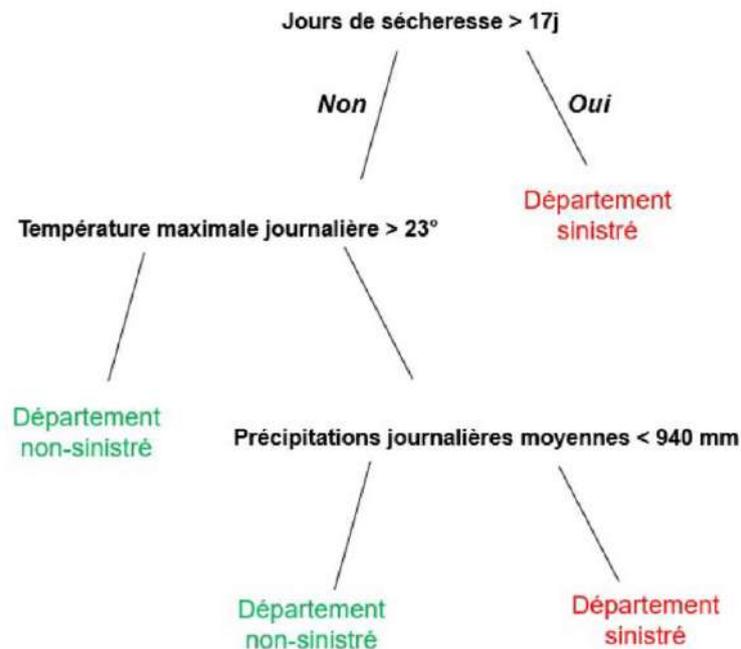


Figure III-1- Schéma explicatif d'un arbre CART

Schématiquement, un arbre CART de classification est un arbre de décision qui permet avec un ensemble de variables qualitatives et quantitatives explicatives X_j d'assigner à un individu Y une classe. Sur le schéma basique ci-dessus, les classes sont « Département sinistré » et « Département non-sinistré » et correspondent à l'occurrence d'événement d'inondation ou non dans les départements de la France conditionnellement aux valeurs de variables climatiques.

Dans les faits, l'algorithme CART choisi l'arbre optimal en 3 étapes. D'abord, est construit l'arbre maximal en isolant tous les individus de la base de données historique via des branches de décision successives, ensuite on élague l'arbre sur un critère de pénalité -on prend le niveau pour lequel l'erreur quadratique est la plus basse par exemple- afin de faire un compromis entre la taille de l'arbre et l'information retenue, et enfin on segmente notre échantillon en plusieurs sous-échantillons afin de déterminer le meilleur arbre par validation croisée. A noter que la dernière étape n'est pas nécessaire dans le cadre d'une Random Forest du fait que l'on cherche à avoir un ensemble d'arbres.

La question à laquelle il faut s'attacher à répondre pour une explication un peu plus exhaustive est la suivante : comment sont déterminées les variables et les seuils de séparation en deux groupes à chaque branche ? La manière de faire est de sélectionner la variable à chaque nœud qui va le mieux segmenter le groupe restant en deux groupes, pour cela, plusieurs mesures de dissimilarités existent mais c'est l'indice de diversité de Gini qui est le plus fréquemment utilisé. C'est celui qu'on utilisera.

Le problème de l'arbre CART est la grande variance et le faible biais qu'il observe, ainsi, pour palier cela, nous avons besoin d'utiliser une méthode de « Bagging », i.e. une méthode ensembliste statistique, qui permet de combiner la maniabilité de l'arbre CART avec une plus grande robustesse dans les sorties de modélisation : la Random Forest [16].

Comme son nom l'indique, la Random Forest est un ensemble d'arbres CART. Pour éviter le biais et neutraliser la variance, l'algorithme de Random Forest va piocher aléatoirement (comprendre ici uniformément) des variables explicatives et construire un arbre avec celles-ci et ce autant de fois que la paramétrisation l'ordonne. Ensuite, un vote parmi les réponses des arbres est fait et l'attribution à la classe de l'individu cible est faite par vote majoritaire. Ceci permet de neutraliser la variance et également de se prémunir contre une forte corrélation des variables explicatives puisque l'on n'en pioche qu'un nombre restreint parmi la totalité durant la construction de chaque arbre.

Schématiquement une Random Forest de classification est de la forme :

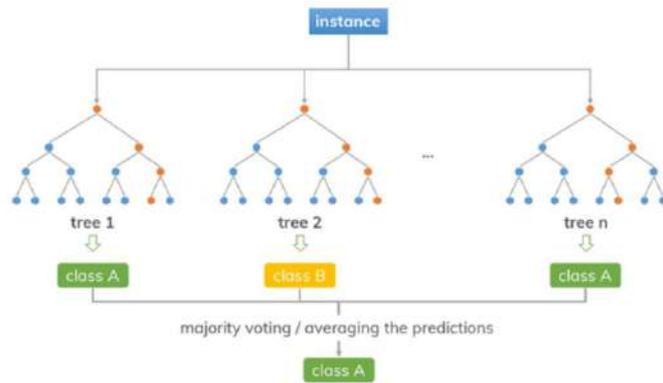


Figure III-2- Schéma explicatif d'une Random Forest

Une fois que les départements dans lesquels des sinistres peuvent avoir lieu sur une année sont établis via la Random Forest, nous utiliserons un algorithme Elastic Net [17] pour prédire le nombre effectif de sinistres qu'enregistreraient les départements sinistrés.

Pourquoi utilisons-nous une méthode Elastic Net ? Pour deux raisons principales, d'abord, la proximité de la méthode Elastic Net avec celle du GLM lui confère une interprétabilité utile à la qualification du modèle, ensuite et surtout l'Elastic Net permet la sélection automatique de variables et le groupement de variables trop colinéaires pour l'utilisation d'un GLM.

De plus, étant donné la surdispersion importante observée dans le portefeuille, même après l'extraction des départements non sinistrés dans l'historique (puisque déjà prédits par la Random Forest), il sera bon d'utiliser la famille Négative Binomiale.

Dans le cadre du GLM, on cherche à minimiser l'erreur quadratique d'une fonction de lien associée à la famille choisie évaluée en l'espérance attendue associée également à la famille, ici, on cherche ainsi les prédicateurs linéaires associés au logarithme de l'espérance d'une loi binomiale négative. Pour permettre l'ajout de variables corrélées comme explicatives nous mettons en plus dans la fonction à minimiser des éléments de pénalité si bien que la fonction à minimiser est désormais la suivante :

$$\min_{\beta_1, \dots, \beta_p} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j z_{ij} \right)^2 + \lambda \left[\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right]$$

$R(\beta)$ = Fonction de pénalité de la régression elasticnet. ($\alpha = 0$) : Ridge ; ($\alpha = 1$) Lasso.

Pour rentrer dans le détail, le premier terme de pénalité est la norme des coefficients prédicateurs, ce terme permet de traiter les coefficients trop faibles et vecteurs de

bruits en les annulant, de même le deuxième terme permet d'empêcher une trop forte variance en réduisant la force donnée à des coefficients prédicateurs trop variant.

Une fois l'Elastic Net mis en place, il est nécessaire de choisir par validation croisée dans l'idéal les meilleurs coefficients de régularisation : faut-il préférer minimiser la norme L1 ou L2 ?

2. Adéquation et paramétrage

a. *Random Forest*

Les deux macro-paramètres du modèle sont le nombre d'arbre utilisé et le nombre de variables à piocher pour la création de chaque arbre.

Afin de vérifier que le modèle est adéquat pour les données utilisées, nous entraînons une forêt aléatoire sur une séquence de nombre d'arbres et un nombre de variables changeant. Nous regardons alors la capacité du modèle à retrouver les départements sinistrés dans l'historique. Pour cela, on regarde le F1-score de prédiction, il s'agit de la moyenne harmonique entre la précision et le rappel du modèle en termes de prédiction.

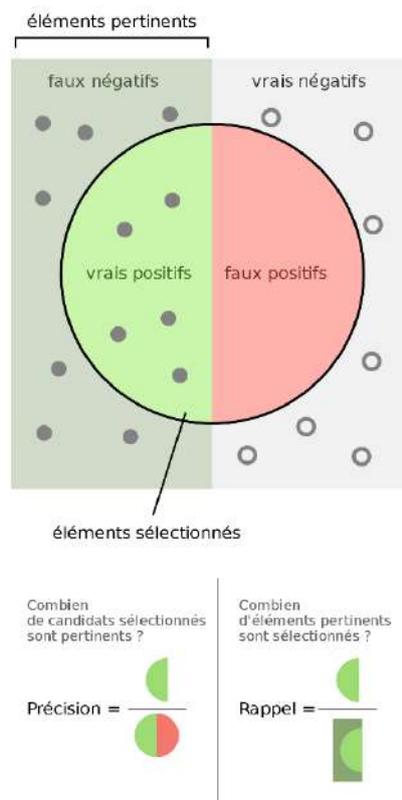


Figure III-3- Schéma explicatif des notions de précision et rappel

On obtient le graphique de convergence suivant :

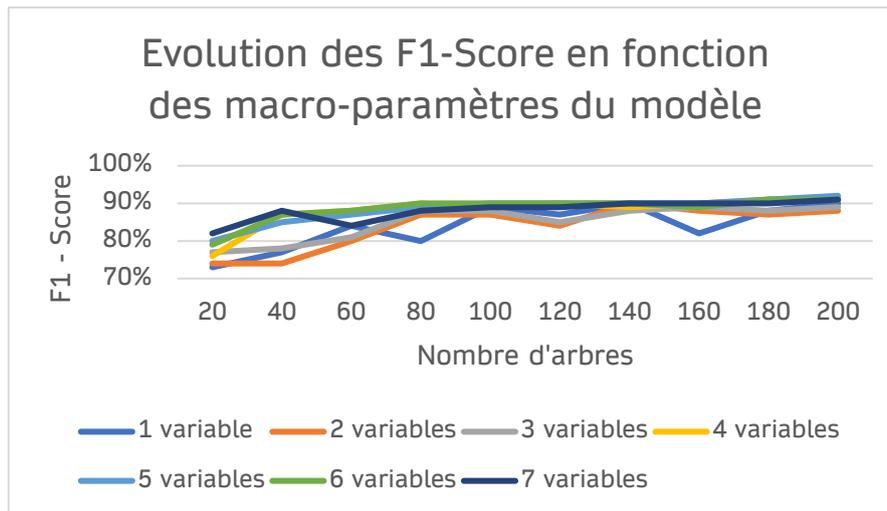


Figure III-4- Graphique de convergence de la forêt aléatoire en fonction du F1-Score

Ceci nous permet de conclure que le modèle est convergent avec 5 variables par arbre et 200 arbres. De plus, il nous indique que le modèle est bien adapté aux données puisque le F1-Score obtenu dans ce cadre est égal à **91,8%**.

On peut donner l'interprétation suivante, les départements sinistrés sont bien identifiés avec moins de 10% d'erreur (faux positifs et faux négatifs compris). En revanche, il faut nuancer le résultat étant donné la présence plus importante de départements sinistrés que de départements sains dans l'historique.

b. GLM binomial négatif pénalisé

Pour vérifier la bonne adéquation du modèle de GLM binomial négatif pénalisé aux données, nous vérifions d'abord que la variable de réponse est approximable par une variable binomiale négative. Pour cela, on réalise un test du χ^2 entre la variable observée et la variable binomiale négative dont on a estimé les paramètres par méthode des moments ($r = 0.6$, $p = 0,007$).

Test du Khi 2 d'adéquation	
Valeur du Khi 2 :	20161,33
Degrès de libertés :	1020
P-value	0

Figure III-5- Résultat du test du khi 2

Pour rappel, nous choisissons d'utiliser une régression elastic-net

Ensuite, nous allons procéder au choix optimal du facteur de régularisation λ par validation croisée.

- Séparation de la base de données en K sous-échantillons
- Estimation du modèle sur $K - 1$ sous-échantillons et validation sur le dernier sous-échantillon

- On répète le procédé K fois en changeant l'échantillon de validation
- Chaque point aura servi à l'apprentissage et à la validation

Nous prenons pour coefficient de validation l'erreur quadratique : plus celle-ci est élevée moins le lambda est optimal. Au final, on retient deux valeurs possibles de lambda au niveau des traits pointillés sur le graphique.

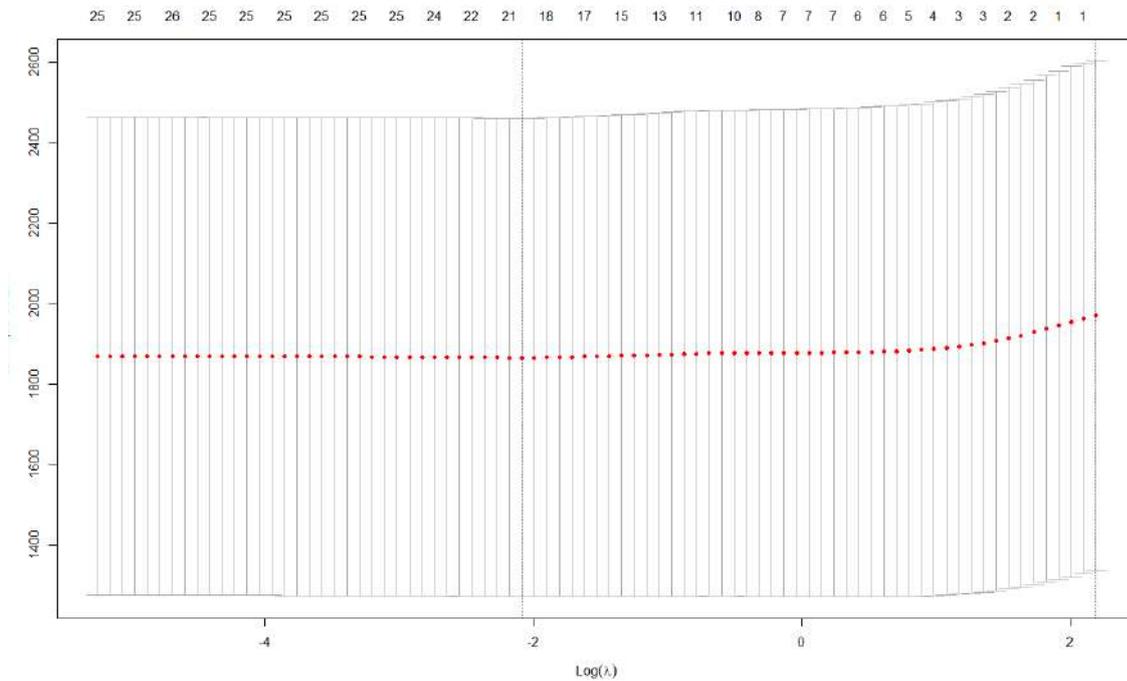


Figure III-6- Choix du lambda de pénalisation général

Le plus à droite est le premier qui associe une erreur quadratique faible et stable avec 1960 d'erreur quadratique globale, le second est le lambda minimum optimale avec une erreur de 1865.

On choisira le lambda minimum, il est d'une valeur de 0,12 et indique une pénalisation faible de notre glm.

Enfin, nous itérons sur les valeurs possibles d'alpha, dont la valeur choisit le type de pénalité afin d'avoir une erreur quadratique minimum. Nous vérifions aussi qu'alpha est suffisamment grand pour annuler des variables en annulant les coefficients du fait de la norme L1.

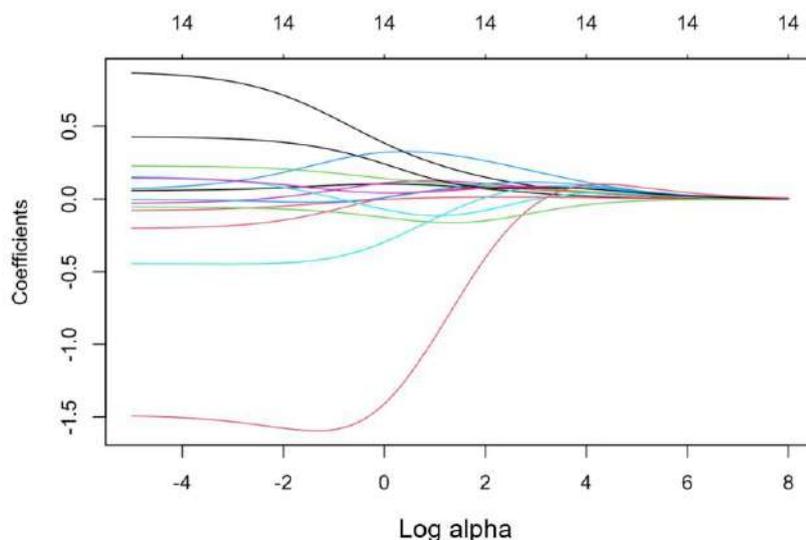


Figure III-7- Aide au choix du coefficient de balancement entre normes

Ici, le graphique montre qu'un alpha optimal se trouvera autour de 0.37 (ou -1 en log) car la sélection de variable n'est pas abusive et que l'aplatissement des coefficients et regroupement des variables peu importantes semble être réalisé. On choisit donc alpha égale à 0.40. Ainsi nous mettons un poids équivalent aux deux normes : nous devrions voir une sélection de variable automatique et de la parcimonie dans les coefficients.

Pour terminer, nous affichons les coefficients du GLM pénalisé ainsi réalisé ainsi que le ratio de déviance du modèle i.e. son taux d'explication de la variance théorique.

Variables conservées	Coefficients associés	Significativité
PINT	0.0085672217	***
PFL90	0.1877771355	***
RR1MM	-0.0311631816	***
année	0.0013133644	***
TRAV	-0.0003256346	**
Variance expliquée par le modèle :		91,21%

Tableau III-1- Récapitulatif de l'Elastic-Net

En tant que la variance semble être bien expliquée par le modèle, nous pouvons penser que le modèle est bien en adéquation avec les données. D'autre part, on retrouve les variables les plus corrélées linéairement avec la fréquence annuelle ce qui est rassurant et nous indique une bonne adéquation avec les données.

Toute chose égale par ailleurs, on peut d'ailleurs dire qu'une augmentation de la fraction de pluie intense (PFL90) de 100% induit une augmentation du risque d'inondation de 18% en fréquence.

3. Validation

Afin de valider le pouvoir prédictif du modèle nous avons effectué un split train-test dans nos données historiques.

Notre split est un split 80-20, nous avons conservé 80% de la base pour construire le modèle et 20% de la base a été utilisé comme test.

Nous obtenons la matrice de confusion suivante pour la prédiction des départements sinistrés :

	Département prédit sinistré	Département prédit sauf
Département sinistré	358	4
Département sauf	60	19

F1 - Score :	91,8%
---------------------	--------------

Tableau III-2- Résultat du test sur le modèle de forêt aléatoire

On conclut de ces chiffres que le premier pan du modèle semble adapté à la prédiction car le F1-Score est élevé. On remarquera cependant que le modèle surévalue le risque bien qu'il prédise bien les départements sinistrés (faux positifs à 60). Cela aura un impact sur les résultats globaux.

RMSE*	AME**	ME***
21,8	2,8	177

Tableau III-3- Résultat du test sur le modèle Elastic net

*RMSE : Root Mean Square Error, il s'agit de l'erreur de prédiction selon la norme euclidienne, elle induit la distance euclidienne

**AME : Absolute Mean Error, il s'agit de l'erreur de prédiction moyenne

***ME : Maximum Error, il s'agit de l'erreur de prédiction en norme infini, soit l'erreur absolue maximum de l'échantillon, elle induit la distance de Tchebychev

Les résultats que nous obtenons par régression elastic net sont plus mitigés.

Comme nous pouvons le voir grâce au RMSE et au ME, les prédictions sont trop dispersées par rapport au risque. On a donc une surestimation d'en moyenne 2,8 sinistres par département, ce qui en relatif nous donne une marge d'erreur de 26%. Bien que l'erreur soit importante, dans le contexte de travail dans lequel nous nous trouvons nous allons valider le modèle et garder en tête la marge d'erreur à appliquer.

En appliquant les deux modèle on obtient les résultats suivant :

RMSE*	AME**	ME***
22,1	2,1	177

Tableau III-4- Résultat global du test sur le module fréquence

Ainsi, on a réduit la marge d'erreur en moyenne, qui ne représente plus que 20% de la moyenne.

Ainsi, nous trouvons en moyenne 16,2 sinistres par départements par année contre un historique à 14,1 sinistres.

Enfin, nous validons le modèle du fait qu'il semble de toute manière difficile de créer un modèle parfait tant les inputs explicatifs sont réduits.

4. Sensibilité aux autres scénarios climatiques

Nous avons calibré et validé le modèle de projection de fréquence en s'appuyant sur le scénario RCP8.5 du Drias. Ainsi, nous allons tester le modèle sur les scénarios 4.5 et 2.6 (moins extrêmes) et voir si le modèle est sensible à cela.

RCP4.5

Premièrement, il n'y a pas de différence majeure sur la prédiction des départements sinistrés par la Random Forest. En témoigne la matrice de confusion.

	Département prédit sinistré	Département prédit sauf
Département sinistré	353	9
Département sauf	59	20

F1 - Score :	91,2%
---------------------	--------------

Tableau III-5- Sensibilité du modèle RF à l'utilisation du scénario RCP4.5

En outre, les résultats globaux sont moins bons mais la déviation entre les deux modèles n'est pas significative :

RMSE*	AME**	ME***
22,3	2,8	178

Tableau III-6- Sensibilité du modèle global de fréquence à l'utilisation du scénario RCP4.5

Ainsi, on ne constate pas de forte sensibilité au changement de scénario en ce qui concerne le scénario RCP8.5.

RCP2.6

Comme pour le scénario précédent, pas de différence majeure ni concernant le modèle de Random Forest ni au global.

On a les résultats suivant :

	Département prédit sinistré	Département prédit sauf
Département sinistré	355	7
Département sauf	58	21

F1 - Score : **91,6%**

Tableau III-7 Sensibilité du modèle RF à l'utilisation du scénario RCP2.6

RMSE*	AME**	ME***
22,7	2,6	177

Tableau III-8 Sensibilité du modèle global de fréquence à l'utilisation du scénario RCP4.5

Ces résultats nous confortent dans l'idée de pouvoir comparer l'évolution du risque à travers plusieurs scénarios.

B. Projection de la sévérité

1. Méthodologie

En reprenant les notations que l'on a déjà explicité en début de chapitre III, on utilisera deux méthodes afin de déterminer la distribution des montants de sinistres $X_{d,t}$ pour un département donné d appartenant au cluster c l'année t et considérant que, en reprenant les écritures précédentes, la distribution d'un sinistre unitaire est issue de la décomposition suivante :

$$X_{d,t} = {}^{np}X_{y_t}^c + {}^pX_{y_t}^c$$

avec ${}^{np}X_{y_t}^c$ une distribution estimée par une méthode non paramétrique correspondant à la charge de sinistralité attritionnels et ${}^pX_{y_t}^c$ une distribution estimée par méthode paramétrique.

Projection des charges de sinistralité attritionnels

Pour les sinistres attritionnels, le regroupement par cluster permet de disposer d'un grand nombre de données, ainsi, nous allons projeter les charges de sinistralité futures par transformée inverse grâce à la classique méthode d'inversion [18]. Cependant, afin de prendre en compte les variables climatiques dans notre démarche, il s'agira non pas d'une transformation par inversion « classique » du quantile généralisé de la distribution, mais d'une transformation par inversion du quantile conditionnel de la distribution (comme dans [19]).

Schématiquement nous souhaitons piocher des pertes conformes à un cadre climatique donné.

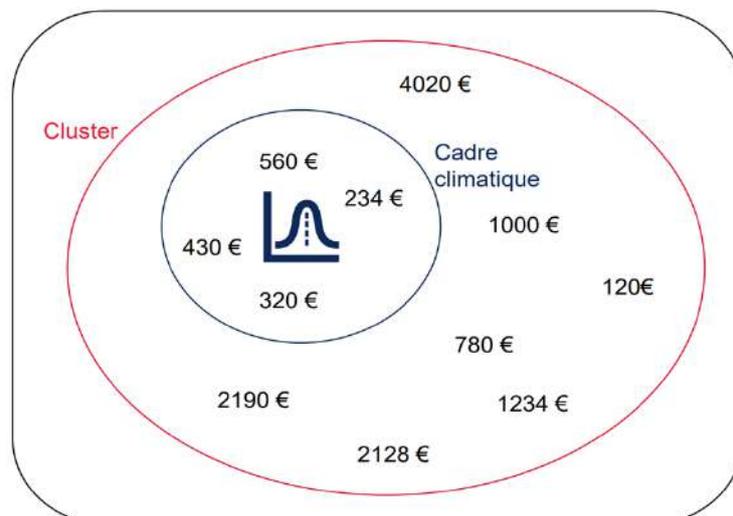


Figure III-8- Schéma explicatif de la démarche de régression non-paramétrique

Pour cela, il suffit de filtrer la base de données sur le climat futur à un voisinage près puis d'appliquer une méthode de quantilification empirique au niveau u où u est une réalisation d'une loi uniforme sur le segment unitaire.

Mathématiquement et de manière concise, l'inversion précédemment introduite s'appuie sur le fait que $F_V^{-1}(U) = V$ pour toute variable V .

Pour rappel, $V = X^c | Y = y_t$ où X^c représente la charge d'un sinistre unitaire au sein du cluster c associé à la zone d , Y est la variable climatique (ou l'ensemble de variables climatiques) choisie comme pivot de l'inversion et y_t sa valeur dans le département considéré l'année considérée.

Pour chaque département et année future, on détermine donc la fonction de répartition empirique sur la base des données disponibles au sein du cluster et de la valeur de la variable climatique de la zone au moment t :

$$F_{X^c|Y=y_t}(x) = \frac{P(X^c \leq x \cap Y=y_t)}{P(Y=y_t)}$$

En approximant comme suit l'événement $[Y = y_t]$:

$$[Y = y_t] \approx \left[(y_t - V(y_t))_+ \leq Y \leq y_t + V(y_t) \right]$$

où $V(y_t)$ est un voisinage de y_t qui sert à neutraliser la forme continue des variables climatiques.

On utilise par suite la méthode d'inversion sur autant de points du segment unitaire que l'on souhaite de précision :

$$F_{X^c|Y=y_t}^{-1}(u) = \inf(x \in \mathbb{R}; F_{X^c|Y=y_t}(x) \geq u)$$

Pour que le modèle soit robuste, il est nécessaire d'avoir un nombre suffisant de points ainsi que d'adopter potentiellement une méthode d'approximation du quantile conditionnel comme celui de la fonction quantile du logiciel R qui interpole les points manquants linéairement. De plus, afin d'avoir une vision globale de la distribution ainsi créée, il est nécessaire de choisir un nombre de points u suffisant sur lesquels sont calculés les quantiles conditionnels, par exemple 1000 points à intervalles réguliers donnent une distribution avec une précision de 0,001.

Avant de continuer sur l'estimation paramétrique des montants de sinistre grave, notons que la distribution ci-avant présentée est tronquée à droite (en dessous de la valeur Q_c) comme cela a déjà été exposé avant.

Projection des charges de sinistralité graves

La première méthode utilisée capterait mal les graves et leurs augmentations, en effet, étant non paramétrique, la distribution n'est pas en mesure d'atteindre des montants

de sinistre jusqu'alors non atteints par l'historique liés à des conditions climatiques inobservées dans le passé. De ce fait, il est obligatoire d'inclure une manière de projeter les sinistres graves au sein du modèle.

La méthode utilisée se base ainsi sur l'estimation de paramètres d'une loi en cohérence avec l'historique et l'hypothèse que la forme paramétrique des distributions de sinistres graves ne change pas au cours du temps.

Notre point de départ est la paramétrisation des lois de montants générales par cluster faite durant la partie I. Après plusieurs tests sur notre historique, le meilleur ajustement de loi sur l'historique complet de montant que l'on a par cluster est la loi log-gamma. Ceci peut s'expliquer par la présence d'une queue lourde en fin de distribution associée à des montants élevées comme il est possible d'en imaginer lorsque l'on travaille sur les catastrophes naturelles. On procède ainsi à l'ajustement par MGE des log-charges à des lois gammas par cluster (sous la paramétrisation (α, β)), l'utilisation de cette méthode plutôt que d'utiliser la méthode de maximum de vraisemblance se justifie par le fait que l'on s'intéresse davantage à une fonction de répartition qui plus est à queue lourde plutôt qu'au centre de la distribution pour le cas des graves.

Après cette étape d'ajustement nous projetons les paramètres des loi gammas contre des variables climatiques (non corrélées entre elles pour éviter le bruit) grâce à une simple régression linéaire. Nous vérifions le bien fondé du modèle par une vérification du R2 du modèle. La méthode n'est pas la plus heureuse mais elle permet de prendre en compte les changements climatiques sans retravailler sur la paramétrisation déjà effectuée.

En effet, de manière plus théorique, on obtient des paramètres globaux hérités des observations pour chaque cluster après ajustement par MGE, c'est-à-dire pour C clusters : $((\alpha_1, \beta_1), \dots, (\alpha_C, \beta_C))$. Ensuite, ces paramètres sont granularisés à la maille d en assimilant de la même manière que globalement avec les distributions globales des montants unitaires la variable Y , en effet, on dit alors que pour toute zone d , il existe des paramètres (α_d, β_d) qui sont égaux aux paramètres α_c, β_c où c est le cluster de la zone d en prenant en compte le climat d'estimation de α [resp. β] et on fait l'hypothèse naturelle que α_d [resp. β_d] = $\alpha_c(y^*)$ [resp. $\beta_c(y^*)$] avec y^* la moyenne d'une variable climatique sur l'historique et $\alpha_c(\cdot)$ [resp. $\beta_c(\cdot)$] : $y \rightarrow R$ des fonctions qui associent à y des paramètres (α, β) .

Dès lors, il est possible de projeter les paramètres de ces lois par un modèle choisi qui estimera les valeurs des paramètres pour une valeur climatique future y_t . De manière simple, est choisie ici une régression linéaire mais l'on pourrait imaginer des modèles plus complexes permettant de réduire l'erreur d'estimation et par conséquent réduire les cas limites où la fonction est capée en pratique pour éviter une explosion des charges.

Nous avons maintenant obtenu la distribution paramétrique des sinistres selon le climat, on souhaite à présent appliquer la troncature à gauche issue de la décomposition grave/attritionnels exposée en amont.

Troncature de la variable à gauche

Étant donné un quantile donné par la distribution conditionnelle empirique (loi d'attrition) Q_c , on tronque à gauche la variable associée à la sinistralité grave.

Ainsi, il est possible de calculer des indicateurs propres à cette distribution tronquée.

Pour faciliter l'écriture des calculs, considérons V une loi gamma de paramètre (α, β) i.e $V \rightarrow G(\alpha, \beta)$. Et notons Q la valeur de troncature à gauche.

En préliminaire, notons que l'on a que $e^{V\mathbb{1}_{e^V \geq Q}} = e^{V\mathbb{1}_{V \geq \ln(Q)}}$. Notons également que si l'on note f_V [resp. S_V] la densité [resp. la fonction de survie] de la loi V , on sait que pour tout v dans \mathbb{R} :

$$f_V(v) = \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v} \mathbb{1}_{v>0}$$

$$S_V(v) = \frac{\Gamma(\alpha, \beta v)}{\Gamma(\alpha)}$$

avec $\Gamma(\cdot)$ la fonction Gamma d'Euler et $\Gamma(\cdot, \cdot)$ la fonction Gamma incomplète.

De là, on peut aisément calculer l'espérance d'une loi log-gamma tronquée à gauche par théorème de transfert :

$$\begin{aligned} \mathbb{E}(e^V \mathbb{1}_{V \geq \ln(Q)}) &= \int_{\ln(Q)}^{+\infty} e^V f_V(v) dv \\ &= \int_{\ln(Q)}^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v + v} dv \\ &= \int_{\ln(Q)}^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} e^{-(\beta-1)v} dv \\ &= \int_{\ln(Q)}^{+\infty} \frac{(\beta-1)^\alpha}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\beta-1)^\alpha} v^{\alpha-1} e^{-(\beta-1)v} dv \\ &= \left(\frac{\beta}{\beta-1} \right)^\alpha \mathbb{P}(G(\alpha, \beta-1) \geq \ln(Q)) \\ \mathbb{E}(e^V \mathbb{1}_{V \geq \ln(Q)}) &= \left(\frac{\beta}{\beta-1} \right)^\alpha \frac{\Gamma(\alpha, (\beta-1)\ln(Q))}{\Gamma(\alpha)} \end{aligned}$$

On voit bien ici qu'il est possible en pratique d'observer une explosion de charge grave si le coefficient β est projeté de telle sorte que l'espérance ne converge plus. Cela justifie l'utilisation d'un maximum en capital en pratique.

Il suffit maintenant pour avoir l'espérance de notre loi log-gamma tronquée d'utiliser la relation suivante :

$$\mathbb{E}(e^V | V \geq \ln(Q)) = \frac{\mathbb{E}(e^V \mathbb{1}_{V \geq \ln(Q)})}{S_V(\ln(Q))}$$

En réutilisant les lignes antérieures on a alors par simplification :

$$\mathbb{E}(e^V | V \geq \ln(Q)) = \left(\frac{\beta}{\beta - 1} \right)^\alpha \frac{\Gamma(\alpha, (\beta - 1)\ln(Q))}{\Gamma(\alpha, \beta \ln(Q))}$$

Et de même on a de manière équivalente :

$$\mathbb{E}(e^V | V \geq \ln(Q)) = \left(\frac{\beta}{\beta - 1} \right)^\alpha \frac{\mathbb{P}(\mathcal{G}(\alpha, \beta - 1) \geq \ln(Q))}{\mathbb{P}(\mathcal{G}(\alpha, \beta) \geq \ln(Q))}$$

Pour la variance, on utilise de la même manière le théorème de transfert sur la variable $e^{2V} \mathbb{1}_{V \geq \ln(Q)}$ pour obtenir le moment d'ordre 2 de la variable cible, puis par un calcul très similaire (mettant en jeu du $(\beta - 2)$ et non du $(\beta - 1)$) et la formule de Koëning-Huygens :

$$\mathbb{V}(e^V | V \geq \ln(Q)) = \left(\frac{\beta}{\beta - 2} \right)^\alpha \frac{\Gamma(\alpha, (\beta - 2)\ln(Q))}{\Gamma(\alpha, \beta \ln(Q))} - \left(\frac{\beta}{\beta - 1} \right)^{2\alpha} \left(\frac{\Gamma(\alpha, (\beta - 1)\ln(Q))}{\Gamma(\alpha, \beta \ln(Q))} \right)^2$$

En outre, on peut également calculer le quantile théorique de cette variable (cf. A.2), mais il faudra nécessairement obtenir le quantile de la variable globale de sévérité (grave et attritionnels) par simulation étant donné qu'il n'existe pas de formule fermée pour son obtention à l'inverse de l'espérance et la variance.

2. Adéquation et paramétrage

Macro-paramètres

Pour le modèle non-paramétriques, deux éléments sont à définir en priorité la variable pivot sur laquelle on va piocher les réalisations de sinistres ainsi que le voisinage autour de celle-ci.

Pour choisir et vérifier l'adéquation des données au modèle nous avons regardé l'erreur relative moyenne pour plusieurs variables et plusieurs voisinages définis en % de l'écart-type à la variable pivot. Ici, pas comme nous travaillons en moyenne, pas besoin de préciser un degrés de précision de la distribution extraite. Le tableau qui suit présente les variables les plus prometteuses, il s'agit par ailleurs des variables les plus corrélées linéairement que l'on avait vu dans l'analyse de données ce qui rassure quant à l'adéquation du modèle.

Variable	Voisinage en % de l'écart-type	Erreur relative moyenne
PXCDD	25%	29%
PXCDD	50%	28%
PXCDD	75%	28%
PXCDD	100%	29%
RR1MM	25%	24%
RR1MM	50%	16%
RR1MM	75%	20%
RR1MM	100%	26%
TAV	25%	27%
TAV	50%	26%
TAV	75%	27%
TAV	100%	27%
PFL90	25%	29%
PFL90	50%	28%
PFL90	75%	27%
PFL90	100%	27%
RR	25%	26%
RR	50%	26%
RR	75%	26%
RR	100%	26%
PXCWD	25%	27%
PXCWD	50%	28%
PXCWD	75%	28%
PXCWD	100%	44%

Tableau III-9- Choix de la variable pivot et du voisinage à utiliser

Dès lors, on pourra choisir le nombre de jours de pluie à plus d'1MM (RR1MM) comme variable de régression non-paramétrique.

Min.	1er Qu.	Median	Mean	3ème Qu.	Max.
-224337.36	8.54	4250.43	1581.17	6474.45	12805.36

Comme en témoigne le tableau précédent, il faut noter que la dispersion des résidus est toute de même élevée ce qui implique un modèle peu régulier et source d'erreur marginales. Une marge d'erreur devra être prise en compte lors de l'utilisation du modèle.

Du côté paramétrique, nous allons procéder à une régression linéaire simple entre les paramètres de loi log-gamma que nous avons obtenu par cluster et une combinaison de variables climatiques.

En procédant à une routine d'inclusion/exclusion des variables nous trouvons que le meilleur modèle linéaire en terme d'R2 ajusté est le même pour les deux paramètres (forme et taux) :

Variable	Coefficient (shape)	Coefficient (rate)	Significativité (shape & rate)
TRAV	-7,4	-0,9616	***
RR1MM	3,134	0,4022	***
TNND	13,356	1,7316	***

	Modèle Shape	Modèle Rate
R2-Ajusté	39,30%	39,20%

Tableau III-10- Résultat du modèle linéaire sur paramètres de lois gammas

Le modèle linéaire semble être peu en adéquation avec les données, toutefois, faute de mieux car on ne souhaite pas se lancer dans une modélisation plus complexe, sa qualité reste raisonnable. Testons l'erreur générée.

Nous appliquons la limite légale du contrat d'assurance du portefeuille, 600 000 €, et obtenons que l'erreur absolue du modèle est de **33%**.

Min.	1er Qu.	Median	Mean	3ème Qu.	Max.
1,5	1220,4	2235,9	5624,6	5099	600 000

Tableau III-11- Dispersion de l'erreur absolue du modèle paramétrique seul

De plus, les résidus absolus nous indiquent une forte disparité dans les prédictions.

Combinaison des deux modèles

S'agissant du modèle mis en place, nous avons deux distributions à assembler au point de quantile 85%, c'est en effet le point établi comme point de séparation entre sinistres graves et attritionnels dans l'historique.

On peut donc établir la distribution totale en pondérant nos résultats des deux précédents pans de modèle. On obtient alors.

	Estimée	Réel
Moyenne	5 465 €	5 792 €
Ecart-Type	4 200 €	11 300 €
Erreur Absolue Moyenne	6 047 €	

Tableau III-12- Résultats globaux de la modélisation de sévérité

Notre modèle est ainsi capable de générer des pertes en moyenne en accord avec celles de l'historique néanmoins la dispersion du modèle n'est pas suffisante et cela explique l'erreur absolue importante observée. Notez que le modèle non-paramétrique semble meilleur mais que la présence d'un shift dans les données climatiques nous contraint d'adopter une méthode paramétrique pour la queue de courbe.

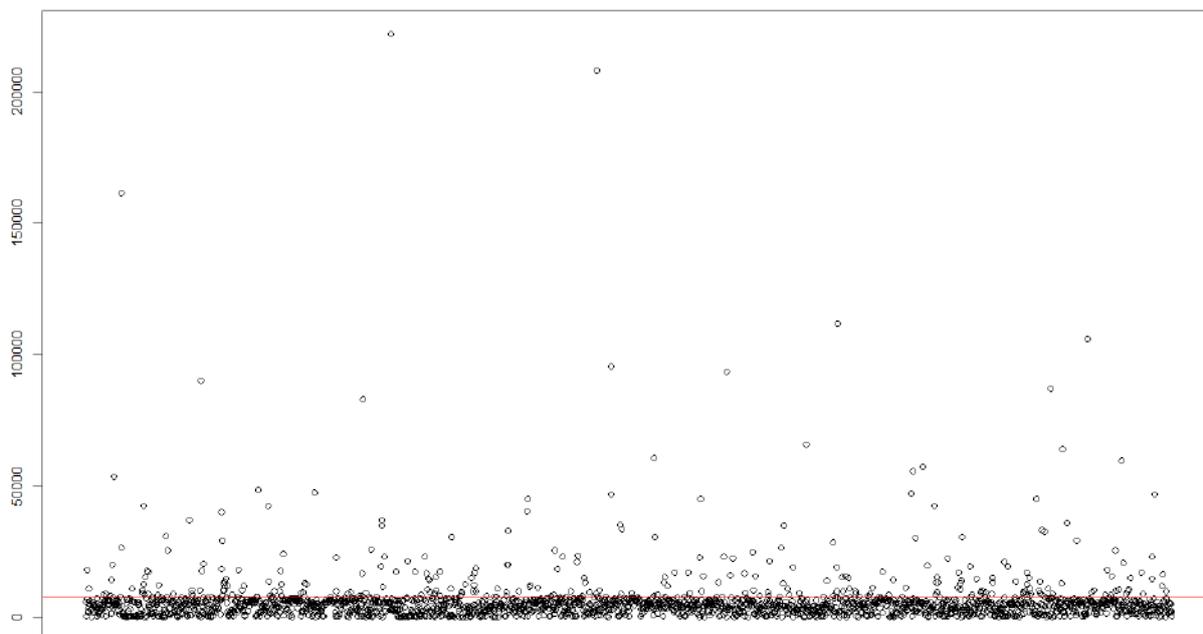


Figure III-9- Niveaux des résidus absolus du modèle de sévérité

Le graphique des résidus nous indique par ailleurs que les résidus sont pour la plupart compris entre 0 et 8000 euros, ainsi cela confirme que le modèle est erroné par les valeurs extrêmes mal appréhendées malgré un effort de modélisation sur ce point-là.

Le modèle est acceptable mais il faudra garder en tête d'appliquer une marge de risque de 20% en cohérence avec les différents résultats ci-avant présentés.

3. Validation

Comme pour la fréquence nous adoptons un split en train et test à hauteur de 80% - 20%.

Nous trouvons sur le set de test les résultats suivant :

Rapports de moyennes	Rapports d'écart-type	RMSE	MAE	ME
93%	68%	11 002 €	6 134 €	584 563 €

Tableau III-10- Résultat de validation du modèle

Les résultats sont cohérents avec les résultats d'adéquation du modèle.

On peut donc, en gardant en tête la relative erreur générée par le modèle l'utiliser dans un modèle prédictif.

4. Sensibilité aux autres scénarios climatiques

Nous avons utilisé le modèle établi sur les scénarios climatiques RCP4.5 et RCP2.6. Nous obtenons le tableau suivant.

	Rapports de moyennes	Rapports d'écart-type	RMSE	MAE	ME	R2 - Modèle paramétrique
RCP4.5	95%	63%	11 876 €	6 166 €	600000	23%
RCP2.6	91%	56%	11 567 €	5 988 €	600000	16%

Tableau III-13- Tests de sensibilité du modèle aux autres scénarios RCP

Nous pouvons constater que le modèle reste correct bien que le taux d'explication du modèle linéaire paramétrique soit bas. Notons qu'il est logique que dans un scénario climatique moins extrêmes les paramètres de lois à queue lourdes soit moins bien prédites.

En conclusion, le modèle construit pourra servir à étudier la déviation en moyenne du risque mais il faudra être très prudent sur les quantiles les plus élevées.

C. Agrégation entre fréquence et sévérité

1. Méthodologie

Une fois que les distributions de sévérités et de fréquences sont obtenues, il est nécessaire de recomposer la distribution par application du modèle collectif (la variable $Z_{d,t+j}$ si l'on reprend la nomenclature du début de chapitre).

Pour cela, étant donné que la distribution de sévérité construite possède un nombre de k points, il faudrait pour recomposer la distribution complète un nombre total de distributions théoriques de $k \times n$ (par calcul combinatoire simple).

Ainsi, du fait de l'impossibilité d'un tel calcul du point de vue des ressources informatiques dont on dispose, on procédera à un rééchantillonnage pour établir la distribution $Z_{d,t+j}$.

Pour éviter trop de temps de calcul nous prendrons des distributions de sévérité de 100 points et iront piocher un nombre d'événement correspondant à la fréquence n_{sim} fois.

2. Adéquation

Il faut procéder à un nombre suffisant de rééchantillonnage pour que le modèle soit robuste ou au moins stable. Ainsi, nous allons choisir le nombre de simulation n_{sim} nécessaire à stabiliser la moyenne de pertes annuelle sur la période de référence.

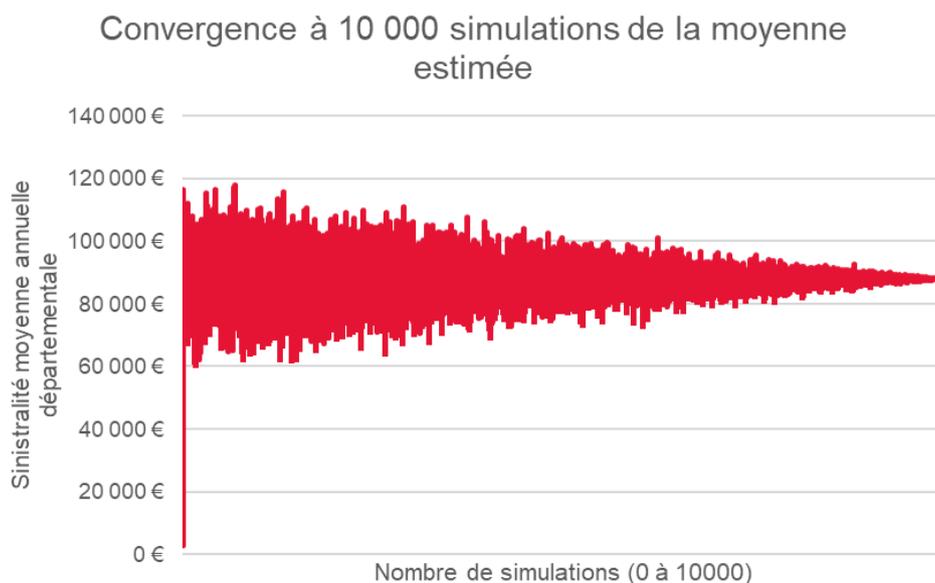


Figure III-11- Convergence de l'estimation de la moyenne avec le nombre de simulations

3. Validation

Afin de valider le modèle nous comparons la distribution de pertes annuelles globales sur la période de référence produite par le modèle et recomposée pour l'historique.

Nous obtenons les fonctions de répartitions suivantes :

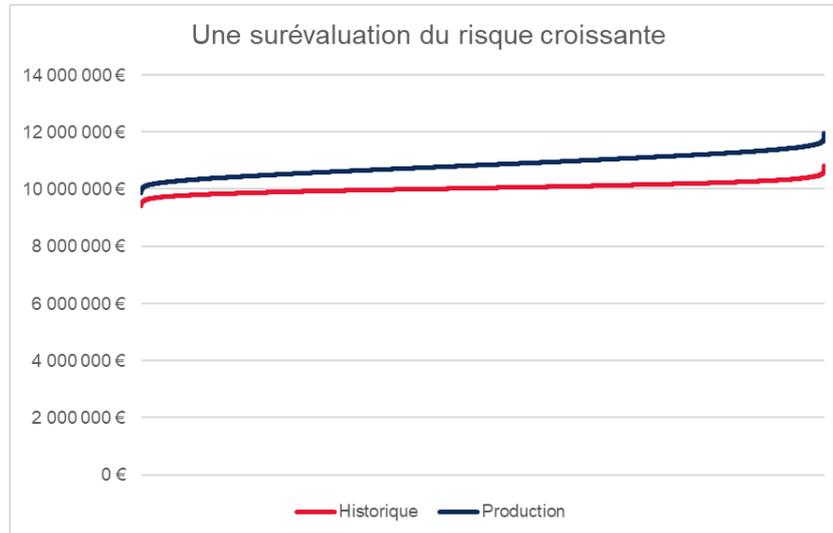


Figure III-12- Erreur linéaire entre prédiction des charges globales et historique

Nous observons un écart entre les distributions qui suit approximativement la fonction linéaire $f(x)=105\%*x + 300\ 030$. (obtenue avec un R2 de 96%).

Ainsi, nous validons le modèle et nous servons de la fonction affine pour créer une marge d'erreur sur la période future que l'on cherche in fine à estimer.

4. Sensibilité

La convergence est semblable pour les autres scénarios climatiques. Par contre, les charges globales sont sensibles au scénario RCP à des niveaux différents.

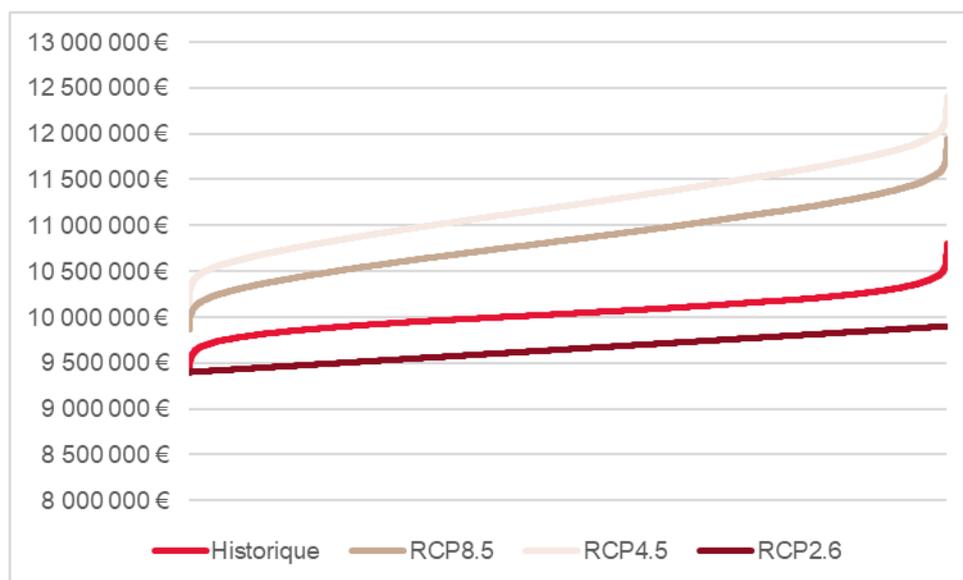


Figure III-13- Différentes prédictions de courbe selon les scénarios RCP

Ce qui se traduit par des coefficient linéaires différents i.e. des incertitudes différentes dans les prédictions.

Incertitude face à l'historique			
	Pente	Intercepte	R2
RCP8.5	105%	300 030,00	96%
RCP4.5	103%	50 010,00	95%
RCP2.6	100%	251 020,00	97%

Tableau III-14-Correctifs linéaires à appliquer sur les prédictions du modèle de sévérité

Nous utiliserons ces coefficients affines pour appliquer un correctif à la décomposition du risque que nous proposerons dans la partie suivante.

IV. Evolution du risque d'inondation sur le portefeuille étudié

A. Evolution des métriques sur le portefeuille et décomposition de l'évolution du risque

1. Décomposition et répartition des évolutions moyennes globales et départementales

Commençons par afficher la décomposition du risque global sur le portefeuille que l'on peut faire sur les deux scénarios les plus pessimistes (RCP8.5 et RCP4.5).

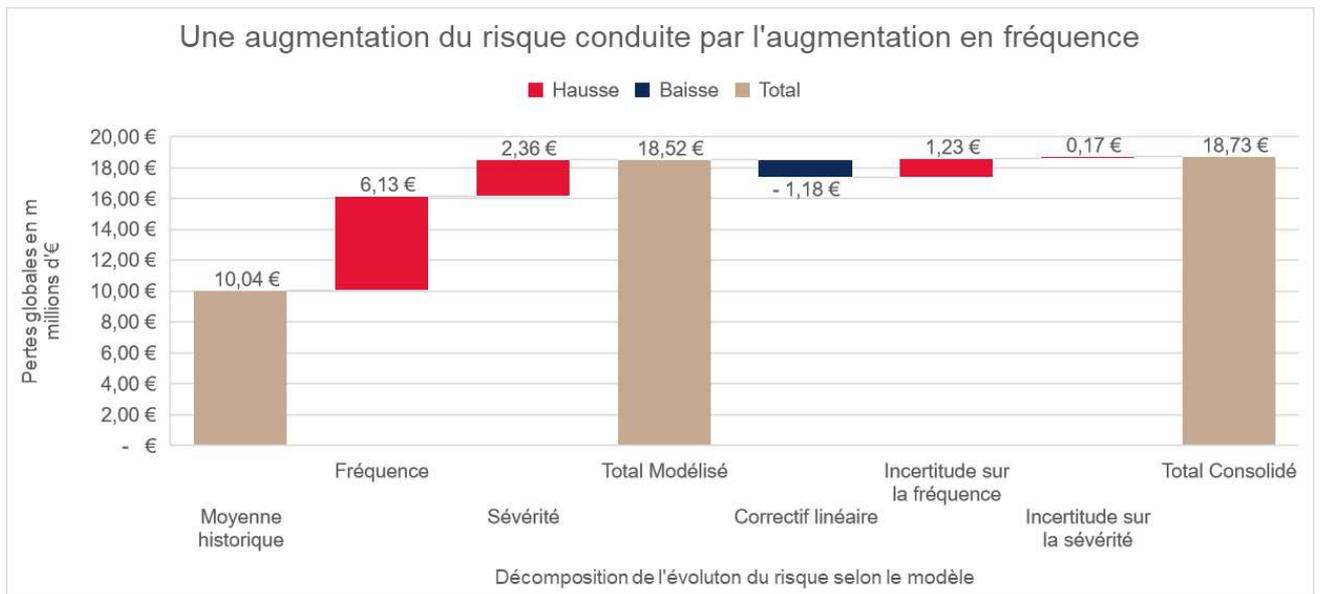


Figure IV-1-Augmentation de la moyenne selon le scénario RCP8.5 sur la période future 2021-2050

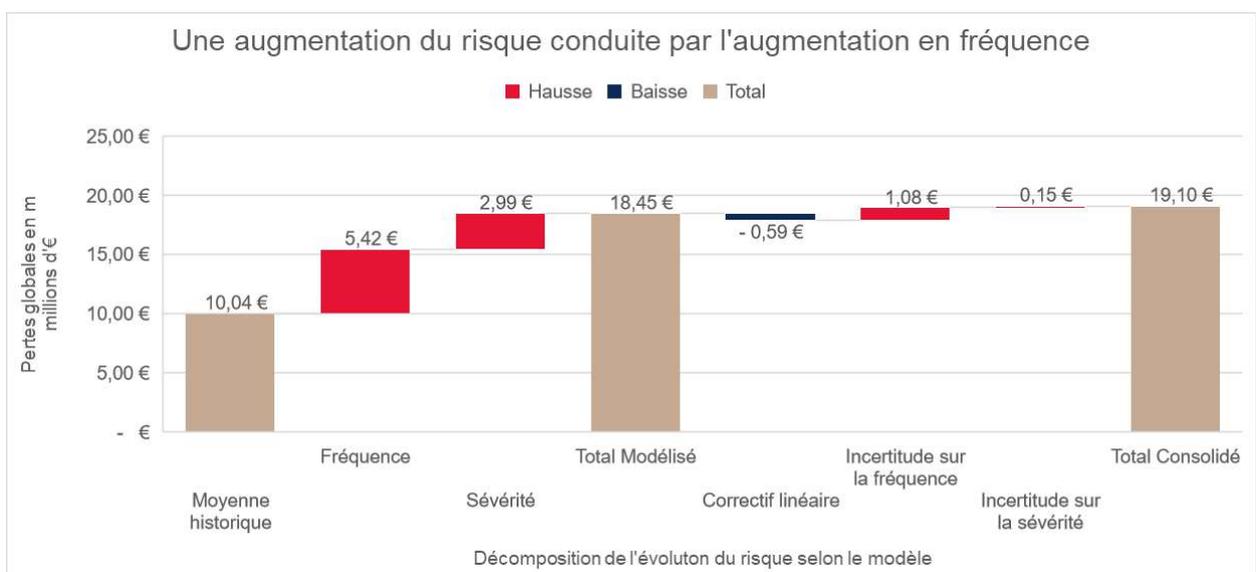


Figure IV-2- Augmentation de la moyenne selon le scénario RCP4.5 sur la période 2021-2050

Pour rentrer dans le détail, a été affiché ci-dessus une décomposition hérité des sorties du modèle : l'augmentation de la moyenne de fréquence est mise en avant en premier chef puis l'augmentation de la moyenne (le reste pour le modèle) est assimilé à la sévérité. Ensuite, nous appliquons le correctif linéaire hérité des écarts observés entre prédiction et historique en fin de partie III. Puis, des taux d'incertitudes liées aux validations de modèles effectués en partie III sont appliqués : 20% de marges d'erreur pour la fréquence correspondante à l'erreur de prédiction observée et 5 à 7% pour la sévérité pure aussi selon les données de validation (l'écart entre la moyenne prédite et la moyenne réelle).

La première conclusion que l'on peut tirer est que dans le cas de scénarios pessimistes c'est la fréquence d'événements qui est responsable à plus 67% de l'augmentation du risque. Combiné avec une augmentation en sévérité de 25% par rapport à la base en moyenne et avec les correctifs d'incertitude que l'on a fait on obtient que le risque inondation sur le portefeuille étudié augmente semblablement entre les deux scénarios.

Pour finir, on anticipe un niveau de risque, moyennant les deux scénarios, qui atteindrait sur la période prospective 188% de la base soit 19 millions d'euros.

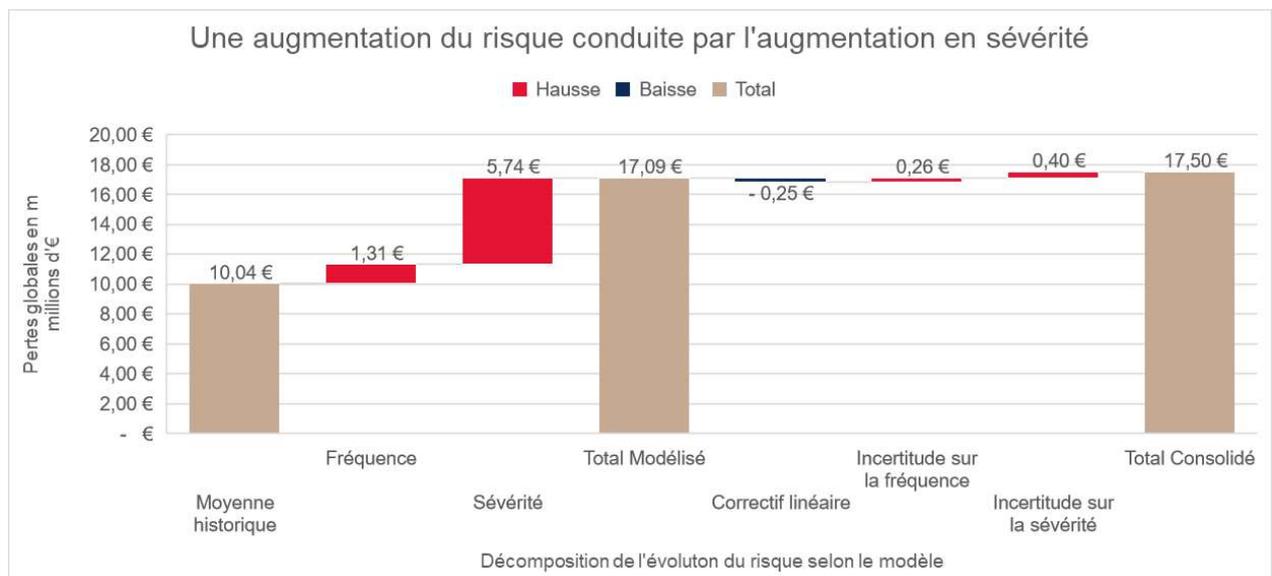


Figure IV-3- Augmentation de la moyenne selon le scénario RCP2.6 sur la période 2021-2050

Dans le cas du scénario le plus optimiste, c'est la sévérité des événements qui tirent les pertes vers le haut. En effet, dans ce cas, la sévérité est responsable à 81% de l'augmentation des pertes annuelles globales. Il faut nuancer ce résultat en considérant que le modèle de sévérité n'était pas très robuste concernant le scénario RCP2.6. On peut quand même imaginer une augmentation du risque bien que plus raisonnable. Ici, 173% d'augmentation en gardant en tête la forte incertitude de modélisation sur la sévérité.

Si le risque augmente globalement, sa répartition reste peu changée : les cartes de densités affichées ci-après nous montre un bassin de risque inchangé mais plus dense.

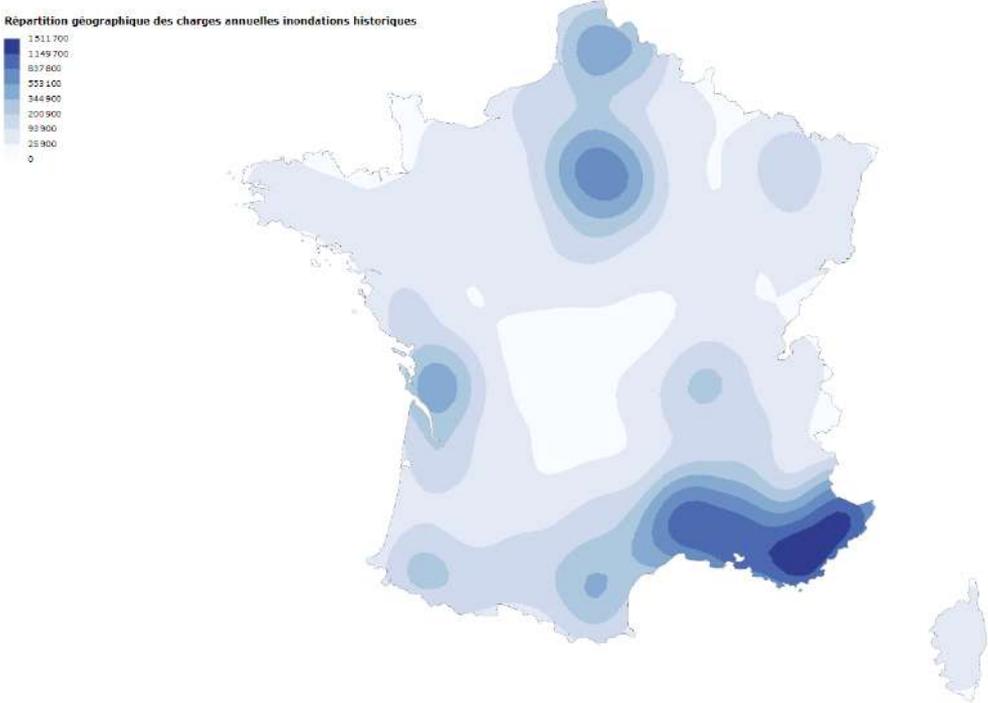


Figure IV-4- Carte de densité du risque inondation du portefeuille (fréquence et sévérité)

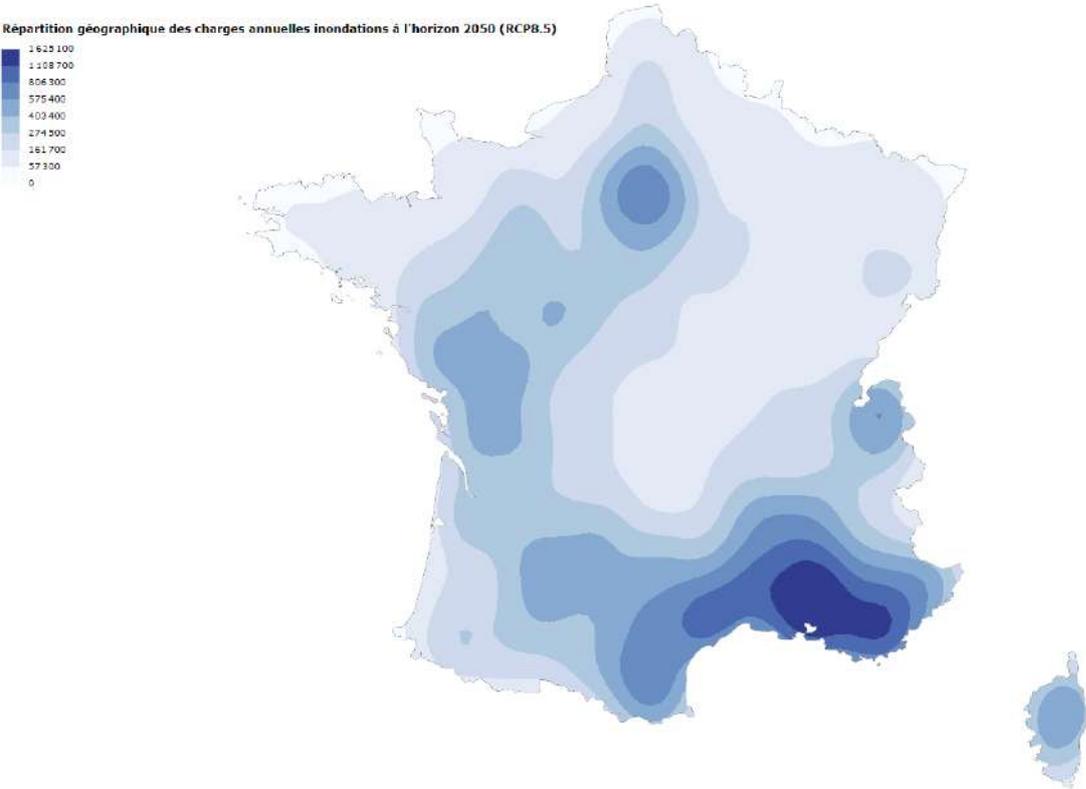


Figure IV-5- Carte prospective de densité du risque inondation du portefeuille (fréquence et sévérité)

Enfin, s'agissant des questions relatives au SCR et moyenne, voici un graphique récapitulatif des valeurs estimées et de l'incertitude associée à la mesure.

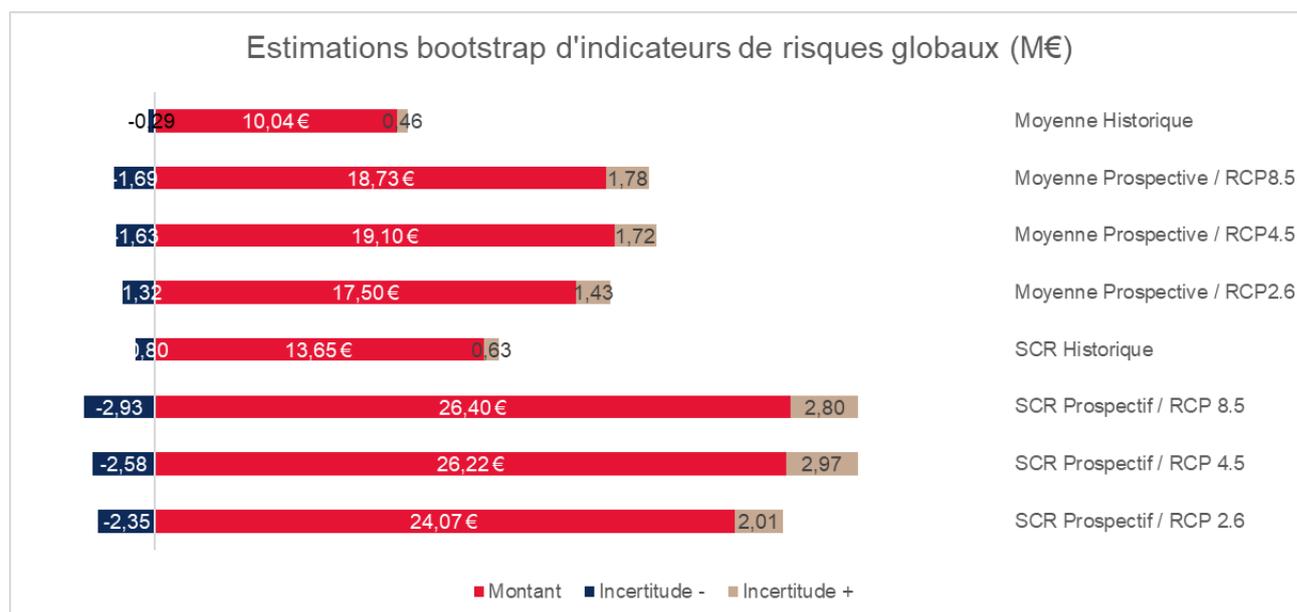


Figure IV-6- Forte augmentation en moyenne et en SCR du risque d'inondation

Les distributions bootstrap obtenues après rééchantillonnage permettent d'obtenir des estimations des indicateurs clés pour les différents scénarios.

Nous pouvons tirer quelques conclusions de ce graphique :

- La forte incertitude de mesure (amplitude de 20%)
- Une augmentation semblable en SCR et en moyenne :
 - En moyenne, l'augmentation en quantile extrême (SCR ici) est de 188% contre 183% pour l'espérance (ici la moyenne)
- Un seul scénario plus favorable en termes de risque, le scénario RCP2.6
- Une forte augmentation du risque sur la période prospective quel que soit le scénario

2. Taux d'augmentation de primes espérables et avancement des limites de réassurance

Après avoir établi des estimations d'indicateurs à l'échelle du portefeuille pour le risque inondation. Nous annualisons les résultats pour trouver un taux d'augmentation annuel. On utilise précisément le calcul de taux suivant (avec π le montant estimé de charge annuelle à horizon prospectif et actuel).

$$\Pi_{2020}(1+r)^{31} = \Pi_{2050} \iff r = \left(\frac{\Pi_{2050}}{\Pi_{2020}}\right)^{1/31} - 1$$

De là, on obtient le tableau suivant :

Indicateur / Scénario	Taux annuel
Espérance / RCP 8.5	2,032%
Espérance / RCP 4.5	2,095%
Espérance / RCP 2.6	1,808%
SCR / RCP 8.5	2,150%
SCR / RCP 4.5	2,128%
SCR / RCP 2.6	1,846%

Tableau IV-1- Taux annualisés d'augmentation du risque

Pour donner une interprétation face à ce tableau.

Un assureur doté du portefeuille dont nous disposons devrait augmenter ses montants de primes non chargées de 2,032% annuellement (hors frais liés à l'inflation ou autre frais financiers). Ceci couvrirait ainsi l'augmentation modélisée sur la période 2020-2050.

Ce taux est relativement faible en considération des résultats du premier stress-test de l'ACPR. Néanmoins, notre périmètre d'étude n'est pas le même : nous nous sommes concentrés sur le risque inondation là où le risque sécheresse (attendu comme fortement augmentant) est aussi pris en compte.

De plus, il est intéressant de noter que dans le cadre d'une production de scénario ORSA, nous pouvons voir le décalage entre les deux scénarios +2° (RCP8.5 et RCP4.5) et le scénario adverse (RCP2.6). Ainsi, un gap faible de 0,23% est modélisé entre les deux scénarios : aucun des deux scénarios ORSA n'est neutre en termes de risque.

Finalement, si l'on suppose que le besoin en réassurance est à un niveau SCR, c'est-à-dire que le bilan est réassuré à une hauteur d'un scénario bicentenaire, l'augmentation en primes de réassurance est ici estimée à un niveau de 2,150%. Ceci induit une augmentation plus forte en primes de réassurance qu'en primes. Ainsi, les acteurs de l'assurance peuvent s'attendre à une augmentation du coût naturelle des primes de réassurance, et, par mécanisme de disponibilité du capital (côté réassureur), une pression des prix. Ceci pourrait finalement avoir un impact sur les primes.

Finalement, en considération de l'augmentation de SCR modélisé, nous pouvons anticiper que la formule standard devra évoluer rapidement pour suivre l'augmentation pratique des besoins en capitaux de solvabilité.

B. Proposition d'une mesure de déviation du risque

1. Vitesse de déviation du portefeuille

Nous souhaitons proposer une métrique rendant compte de l'évolution du risque d'inondation en prenant en compte la temporalité. L'idée est de présenter la vitesse de déviation du portefeuille, concrètement, sur l'augmentation de 188% du risque en espérance estimée sur la période prospective, à quelle vitesse le portefeuille atteindra-t-il ce niveau de risque ?

La métrique ci-introduite sera nommée l'incidence climatique du portefeuille et définie comme le taux de croissance moyen de l'espérance (ou la SCR) annuelle pondérée par l'écart temporel à l'espérance (ou SCR) actuelle. En effet, dans une logique de pilotage en transition, les dynamiques les plus proches sont privilégiées et on affecte ainsi un poids correspondant à la distance temporelle à la date 0 considérée normalisée.

Mathématiquement, on écrira cette quantité τ et la calculerons comme suit :

$$\tau^{\mathbb{E}} = \sum_{i=1}^n \frac{d_i}{\sum d_i} \tau_i^{\mathbb{E}}$$

avec

$$d_i = t_{n+1-i} - t_0$$

$$\tau_i^{\mathbb{E}} = \frac{\mathbb{E}_{t_i} - \mathbb{E}_{t_0}}{\mathbb{E}_{t_0}} \frac{1}{t_i - t_0}$$

Qualitativement parlant, il s'agit du taux de croissance de la sinistralité moyenne annuelle pondérée par l'écart temporel à la sinistralité moyenne historique.

Schématiquement :

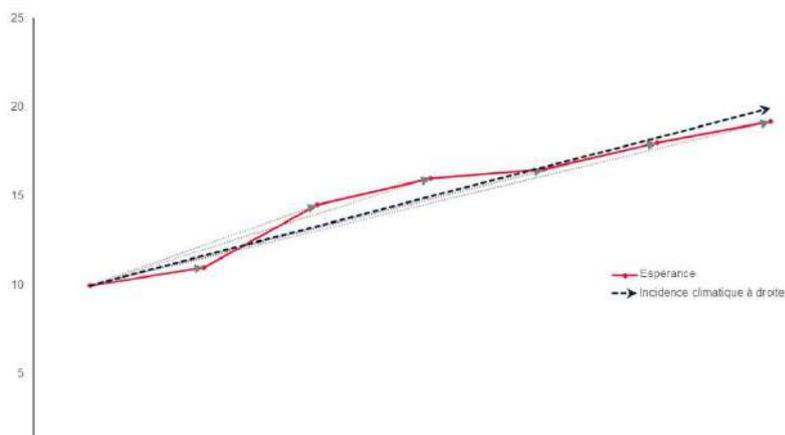


Figure IV-7- Illustration du taux d'incidence climatique construit

L'indicateur construit a plusieurs visées : comparaisons des portefeuilles, utilisation au sein d'un scénario climatique type ORSA ou interprétation de l'augmentation du risque en vitesse, ce qui donne une indication de l'augmentation du risque sur un temps court.

2. Conclusions quant à la métrique

Après calcul sur un pas de 5 ans de la métrique présenté ci-avant, nous obtenons les vitesses d'augmentation du risque en espérance et SCR pour notre portefeuille selon les différents scénarios climatiques.

Scénario	Espérance	SCR	Moyenne harmonique
RCP 8.5	57,43%	43,73%	49,65%
RCP 4.5	48,52%	36,52%	41,67%
RCP 2.6	47,37%	34,63%	40,01%

Figure IV-8- Coefficient de vitesse d'augmentation du risque inondation

La conclusion que l'on peut faire de cette métrique simple est la différence que l'on voit entre les scénarios en termes de vitesse. De manière notable, le premier scénario (le plus extrême) forme une déviation plus rapide du risque que les deux autres.

Ainsi, si le scénario RCP4.5 atteint un niveau de risque semblable au RCP8.5 dans notre modélisation, nous estimons que la vitesse d'augmentation du risque est nettement plus modérée. En effet, si l'on considère la moyenne harmonique des deux vitesses (ce qui reste donc une vitesse), nous pouvons dire que la déviation du portefeuille est 20% plus rapide sur le scénario le plus extrême que sur le second scénario.

Cela signifie que dans l'hypothèse où l'évolution climatique suit la tendance du premier scénario, les primes devront augmenter de manière plus soutenue que dans le cas d'un scénario plus modéré.

C. Axes d'amélioration du modèle et comparaisons avec quelques modèles de marché

Afin de terminer notre étude sur le portefeuille multi-risque habitation dont nous disposons, précisons les limites du modèle prédictif mis en place : les hypothèses fortes et leurs implications, les manquements en termes de données. Puis, nous comparerons les résultats obtenus aux résultats d'autres acteurs du marché.

1. Axes d'améliorations

Données

Les améliorations suivantes sont envisageables en termes de données :

- Se baser sur une base de pertes par événements et non par sinistre pour profiter d'une causalité plus directe entre données climatiques et données de sinistralité.
- Utilisation des scénarios SSP plutôt que RCP dans la modélisation
- Augmentation du nombre de simulation pour mieux estimer les valeurs extrêmes.

Hypothèses fortes dans la construction du modèle

Maintenant que les manquements en termes de données ont été listés, les différentes hypothèses fortes du modèles vont être exposées du fait qu'elles représentent des limites du modèle.

- Hypothèse d'indépendance entre fréquence et sévérité
- Clustering des données créé à partir de la moyenne malgré son utilisation dans la prédiction de valeurs extrêmes.
- Pas d'apports de données exogènes sur l'exposition au risque (population, carte de risque naturelle) bien qu'elle soit intrinsèquement considérée via la structure de nos données.

- Hypothèse de conservation des lois (log-gammas en particulier) et des seuils de séparation grave-attributionnels
- Choix de conserver une régression linéaire pour les paramètres de loi gammas

2. Comparaison avec quelques modèles de marché

En s'appuyant notamment sur le rapport du collectif « SinonVirgule » [22] traitant de l'effondrement du monde de l'assurance et de la note de la CCR à propos du risque climatique, on dispose d'éléments de comparaison du modèle construit face aux modèles que d'autres acteurs du marché ont construits.

Le tableau ci-dessous présente les résultats du modèle face à ceux de grands acteurs du monde de l'assurance qui ont été extraits du comparatif de SinonVirgule :

Acteur	↗ Inondation
FFA	11%
Covéa	120%
CCR x Météo France	38%
Moyenne	56%
Mémoire (brut)	84%
Mémoire (incidence)	57%

Figure IV-9-Tableau comparatif de l'augmentation moyenne du risque inondation sur la période 2020-2050

La première conclusion que l'on peut tirer de ce tableau est l'extrême hétérogénéité du tableau malgré un cadre de travail très similaire : Covéa et la FFA affichent des résultats presque miroirs concernant les évolutions des deux périls. Ceci souligne la difficulté d'estimation à un horizon très lointain.

En outre, on remarque que le modèle présente une tendance à une surestimation du risque brut par rapport au marché mais il démontre une correspondance plus précise avec la moyenne des quelques acteurs répertoriés une fois que des pondérations temporelles sont appliquées.

V. Conclusion

Le changement climatique est l'une des préoccupations les plus pressantes de notre époque, avec des conséquences potentiellement dévastatrices pour les entreprises et les investisseurs du monde entier. Parmi les nombreux défis posés par le changement climatique, l'augmentation du risque d'inondation occupe une place importante en raison de ses implications financières et de ses répercussions sur les portefeuilles d'assurance. Ce mémoire a tenté d'examiner en détail l'impact du changement climatique sur le risque d'inondation d'un portefeuille MRH et les implications financières qui en découlent.

L'une des conclusions de cette étude est la préoccupation quant à l'imprécision des résultats obtenus en raison de la complexité inhérente à la modélisation du risque d'inondation lié au changement climatique. Le caractère multifactoriel des phénomènes climatiques, la variabilité des données historiques et la diversité de scénarios climatiques rendent cette tâche particulièrement ardue. Cependant, malgré cette imprécision inhérente, des tendances alarmantes émergent.

L'analyse a révélé une augmentation significative du risque d'inondation sur la période 2020-2050. Cette augmentation est estimée à 88 % sur le portefeuille étudié, ce qui démontre l'ampleur du défi auquel sont confrontés les assureurs. Cette hausse substantielle du risque s'explique principalement par une augmentation en fréquence des événements.

Une autre estimation clé qui mérite d'être soulignée est l'augmentation annuelle en prime du risque d'inondation sur le portefeuille. Selon nos résultats, cette augmentation est de 2 % par an, ce qui indique une trajectoire croissante du risque au fil du temps. Cette tendance est particulièrement inquiétante pour les assurés, car elle suggère que les coûts augmenteront régulièrement et pourraient devenir prohibitifs si des mesures appropriées ne sont pas prises.

L'incertitude entourant ces estimations souligne la nécessité d'une approche prudente en matière de gestion du risque. Les gestionnaires devront prendre en compte l'incertitude inhérente aux modèles de changement climatique et aux données disponibles. Il est essentiel de maintenir une certaine flexibilité dans la gestion du risque et d'adopter une approche adaptative, notamment en scénario, pour faire face à l'évolution des conditions météorologiques et climatiques.

En conclusion, l'augmentation du risque d'inondation doit être une préoccupation majeure. Bien que les estimations puissent être entachées d'incertitude, les tendances alarmantes et les chiffres clés présentés dans cette étude soulignent la nécessité d'une gestion proactive du risque d'inondation.

VI. Bibliographie

- [1] AXA. AXA Future Risk Report 2021, 2021.
- [2] ACPR. La gouvernance des risques liés au changement climatique dans le secteur de l'assurance. Technical report, Février 2022.
- [3] ACPR. Les assureurs français face au risque de changement climatique. Technical report, 2019.
- [4] NGFS. Le changement climatique comme source de risque financier. Technical report, Avril 2019.
- [5] IPCC. Climate change 2022 Impacts, Adaptation and Vulnerability. Technical report, 2022.
- [6] ACPR. Principales hypothèses de l'exercice pilote. Technical report, 2020.
- [7] EIOPA. Methodological paper on potential inclusion of climate change in the Nat Cat standard formula, Juin 2021.
- [8] EIOPA. Consultation paper on Application guidance on running climate change materiality assessment and using climate change scenarios in the ORSA. Technical report, Décembre 2021.
- [9] ACPR. Principaux résultats de l'exercice pilote. Technical report, 2021.
- [10] Unidata. NetCDF FactSheet.
- [11] Denis Clot. Analyse de données.
- [12] R Core Team. R Stats Package, 2022.
- [13] Anne Driemel, Amer Krivošija, and Christian Sohler. Clustering time series under the fréchet distance, 2015.
- [14] Jan Beirlant Katrien Antonio Tom Reynkens, Roel Verbelen. Modelling censored losses using splicing : a global fit strategy with mixed erlang and extreme value distributions, 2017.
- [15] Christophe Chesneau. Introduction aux arbres de décision (de type CART). page 67.
- [16] Tin Kam Ho. Random Decision Forests.
- [17] J. Kenneth Tay, Balasubramanian Narasimhan, and Trevor Hastie. Elastic Net Regularization Paths for All Generalized Linear Models, 2021.
- [18] Luc Devroye. Non-uniform random variate generation. Springer, New York Heidelberg, 1986.

- [19] Xavier D'Haultfœuille Pauline Givord. La régression quantile en pratique, 2013.
- [20] Galen R. Shorack and Jon A. Wellner. Empirical Processes with Applications to Statistics. 2009.
- [21] Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. 1943.
- [22] SinonVirgule. Peut-on assurer un monde qui s'effondre ?, 2023.
- [23] IPCC. Guidance for IPCC authors. page 15, Septembre 2020.
- [24] Guan Schellekens and Joris van Toor. Values at risk ? Sustainability risks and goals in the Dutch financial sector. Technical report, 2019.
- [25] EIOPA. Factsheet on the pilot exercise application guidance in the ORSA, 2021.
- [26] EIOPA. Orientations relatives à l'évaluation interne des risques et de la solvabilité, 2016.
- [27] Réseaux cellulaires de Voronoï, A-E Baert, 2002
- [28] [Analyse de données](#), D. Clot