

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

Par : GREVILLOT Jeanne

Titre : Construction et projection d'une prime pure pour le risque inondation.

Confidentialité : NON OUI (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*

.....
.....
.....

*Membres présents du jury de
l'ISFA*

.....
.....
.....

*Entreprise :
Nom : ADDACTIS France*

Signature :

*Directeur de mémoire en entreprise :
Nom : MORIAH Mulah*

Signature : Morah Mulah

*Invité :
Nom :*

Signature :

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

Signature du responsable entreprise



Signature du candidat



Résumé

Face à l'intensification des événements climatiques extrêmes, l'adaptation des modèles assurantiels et l'étude des impacts du changement climatique sur la sinistralité future deviennent des enjeux majeurs. Parmi ces phénomènes, les inondations, qui représentent plus de la moitié des arrêtés Cat Nat en France, sont particulièrement préoccupantes. Leur fréquence et leur intensité devraient continuer d'augmenter sous l'effet du renforcement des précipitations extrêmes et d'autres facteurs environnementaux.

Ce mémoire s'inscrit dans cette dynamique en proposant une approche innovante pour affiner la tarification du risque inondation en assurance multirisques habitation (MRH). Pour ce faire, le modèle s'appuie sur un ensemble de données variées. Celles-ci comprennent, non seulement les historiques de sinistralité provenant d'un portefeuille assureur, mais également des données météorologiques disponibles en open data et des données complémentaires à l'adresse (namR). En combinant ces différentes sources, cette étude développe un modèle de prime pure plus précis, capable de mieux refléter la vulnérabilité de chaque logement face au risque inondation.

Au-delà de cette première étape d'ajustement tarifaire, l'étude cherche à anticiper l'évolution du risque inondation à horizon 2030. En s'appuyant sur les projections climatiques du scénario SSP2-4.5 établies par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), elle vise à analyser les effets du changement climatique sur la sinistralité en France métropolitaine, mettant en lumière les tendances à venir et les zones qui seront les plus exposées.

Mots clés : *Tarification, Inondation, Projection, Assurance Multirisques Habitation, Changement climatique*

Abstract

In the face of the intensification of extreme climate events, the adaptation of insurance models and the study of the impacts of climate change on future claims are becoming major challenges. Among these phenomena, floods, which account for more than half of the Cat Nat decrees in France, are particularly concerning. Their frequency and intensity are expected to continue increasing due to the strengthening of extreme precipitation and other environmental factors.

This thesis is part of this dynamic, proposing an innovative approach to refine the pricing of flood risk in multi-risk home insurance (MRH). To do this, the model relies on a set of varied data. These include not only the historical claims data from an insurer's portfolio but also meteorological data available in open data and complementary building database (namR). By combining these different sources, this study develops a more precise pure premium model, able to better reflect the vulnerability of each property to flood risk.

Beyond this first step of pricing adjustment, the study aims to anticipate the evolution of flood risk by 2030. Relying on the climate projections of the SSP2-4.5 scenario established by the Intergovernmental Panel on Climate Change (IPCC), it aims to analyze the effects of climate change on claims in mainland France, highlighting future trends and the areas that will be most exposed.

Key words : *Pricing, Flood, Projection, Multi-risk Home insurance, Climate Change*

Note de Synthèse

Contexte et problématique

Les catastrophes naturelles en France, bien que relativement rares, entraînent des coûts de sinistres considérables. Parmi ces événements, les inondations occupent une place centrale dans les préoccupations des assureurs, représentant plus de la moitié des arrêtés Catastrophes Naturelles (Cat Nat) et environ 50% du montant global de la sinistralité, qui s'élève à 49,9 milliards d'euros au titre de la garantie Cat Nat entre 1982 et 2022. Ces dernières années, la fréquence et l'intensité des événements climatiques ont significativement augmenté, un phénomène confirmé par les projections climatiques à venir. En effet, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) prévoit une intensification de ces risques, avec des précipitations extrêmes qui deviennent de plus en plus fréquentes. La CCR (Caisse Centrale de Réassurance) prévoit une hausse attendue de la sinistralité climatique comprise entre 47% et 85% d'ici 2050, en raison de l'intensification des événements climatiques. Dans ce contexte, les assureurs se trouvent contraints d'adapter leurs modèles de tarification pour mieux appréhender l'évolution des risques et ajuster leurs stratégies sur le long terme. C'est dans ce contexte que les travaux s'inscrivent en proposant d'une part d'affiner la tarification du risque en intégrant dans la modélisation des données externes et, d'autre part, de projeter les primes pures obtenues à horizon 2030 afin de mieux anticiper les impacts futurs du changement climatique.

Description des données utilisées

Afin d'établir une prime pure pour le risque inondation en France métropolitaine (hors Corse), nous disposons de trois bases de données complémentaires. La première est une base de sinistralité fournie par un assureur, contenant des informations sur l'assuré, son domicile, le contrat souscrit ainsi que l'historique des sinistres. Elle permet de calculer les bases de "Fréquence" et de "Coût moyen", utilisées en tarification non-vie.

Grâce à la géolocalisation des contrats, cette base est enrichie par des données complémentaires issues d'une base développée par Addactis en partenariat avec NamR. Ces nouvelles données concernent à la fois les caractéristiques du bâtiment (surface, nombre de pièces, type de toiture, présence de panneaux solaires, etc.) et son environnement (distance aux points d'eau, proximité des services d'urgence, densité de population, occupation des sols, etc.). Elles proviennent de sources officielles telles que l'INSEE, le Ministère de la Transition Écologique et Solidaire (MTES) ou l'ADEME (Agence de la transition écologique).

Sont ajoutées ensuite, des données météorologiques provenant de la base ERA5-Land, un jeu de données de réanalyse produit par le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT) dans le cadre du programme Copernicus. Ces données sont issues d'observations passées et de modèles climatiques. Avec une résolution de 9 km et une mise à jour quotidienne des valeurs horaires, ce jeu de données permet notamment d'ajouter une variable de cumul de précipitations journalières. La figure 1 récapitule les informations principales.



FIGURE 1 – Schéma récapitulatif des 3 bases de données utilisées

Construction d'un indice de précipitation

La première étape de la construction de cette prime pure, est de comprendre le phénomène inondation d'un point de vue physique et d'identifier les paramètres à l'origine de ces événements. D'après la rubrique «Inondation » de GEORISQUES s. d.(a) , l'inondation se définit comme une submersion temporaire, rapide ou lente, d'une zone habituellement hors de l'eau. Ce phénomène est principalement causé par des précipitations soutenues et durables. Cependant, contrairement aux tempêtes par exemple, pour lesquelles l'impact est direct entre la force du vent et le dommage au bâtiment, pour les inondations, les précipitations ne sont qu'un facteur parmi d'autres car leur effet dépend du comportement de l'eau une fois arrivée sur le sol.

Afin de mieux prendre en compte cette complexité, nous avons élaboré un indice de précipitation adapté. La construction de cet indice s'appuie sur plusieurs étapes :

Dans un premier temps, nous identifions les bâtiments sinistrés et leur associons les quatre points de grille ERA5-Land les plus proches. Pour ces points, nous collectons les cumuls de précipitations sur différentes périodes précédant l'inondation (de 1 à 30 jours) afin d'analyser les effets des épisodes pluvieux. L'objectif est de déterminer si l'inondation résulte plutôt d'un épisode intense et soudain ou d'un événement prolongé et soutenu. Pour chaque période étudiée, nous retenons le maximum des précipitations observées parmi les quatre points sélectionnés.

Ensuite, nous établissons la distribution historique (de 1985 à 2015) des cumuls de précipitations pour chaque période considérée. Cela nous permet d'attribuer à chaque cumul une valeur de quantile par rapport à l'historique. Par exemple, si le cumul sur une journée correspond à un quantile modéré tandis que le cumul sur 30 jours atteint le quantile 99, on peut supposer que l'inondation est liée à des précipitations prolongées. À l'inverse, un cumul très élevé sur une courte période indique un épisode pluvieux soudain et intense. La figure 2 illustre la distribution historique des précipitations sur une période de trois jours pour un bâtiment sinistré. Trois jours avant l'inondation, le cumul des précipitations correspondait au quantile 99 par rapport aux cumuls historiques.

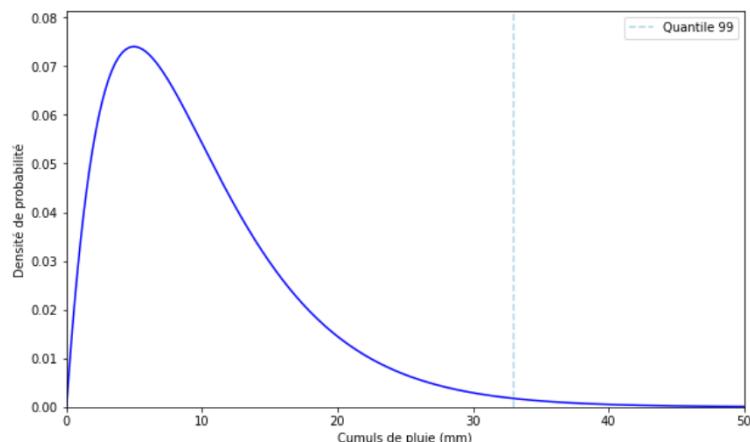


FIGURE 2 – Distribution des cumuls de précipitations historiques sur 3 jours

Enfin, pour chaque bâtiment sinistré, le maximum des quantiles est sélectionné. Cette valeur devient donc, pour chaque bâtiment de la base, la variable "cumul_precip", représentant l'intensité des précipitations, en cohérence avec les spécificités du risque d'inondation.

Méthode de calcul de la prime pure au bâtiment

La méthodologie de calcul de la prime pure climatique suit une approche similaire à celle couramment utilisée en tarification non-vie, tout en intégrant un conditionnement basé sur l'intensité des inondations. La prime pure pour le risque inondation est définie par l'équation suivante :

$$PP_{\text{climatique}} = \sum_i P(\text{intensité}_i) \times \mathbb{E}(\text{fréquence}_{\text{sinistres}}) \times \mathbb{E}(\text{coût}_{\text{sinistres}} \mid \text{inondation}_{\text{intensité}_i}) \quad (1)$$

Cette formule découle du principe des probabilités totales, et décompose le coût moyen des sinistres en fonction des différentes intensités d'inondation i . L'objectif est d'estimer la probabilité d'occurrence d'une inondation d'intensité donnée afin d'intégrer cette information dans le calcul de la prime pure climatique. L'intensité des inondations est évaluée à l'aide de la variable *cumul_precip*. Cette variable est discrétisée en sept classes d'intensité, chacune associée à un intervalle de quantiles et à une probabilité d'occurrence.

Le tableau 1 résume cette discrétisation :

intensité	<i>cumul_precip</i> discrétisée	Probabilité associée
1	[0,0,75[0.75
2	[0.75,0.85[0.10
3	[0.85,0.95[0.10
4	[0.95,0.97[0.02
5	[0.97,0.98[0.01
6	[0.98,0.999[0.019
7	[0.999,1[0.001

TABLE 1 – Description des différentes intensités

Ensuite, le modèle de coût moyen des sinistres définis dans la formule 1, est construit conditionnellement à l'intensité de l'inondation i .

En revanche, la formule utilisée ne lie pas la fréquence des sinistres à l'intensité des inondations, car une base de contrats ne contient pas de dates précises d'inondation, mais seulement une indication d'impact durant la période de validité du contrat. D'autres indicateurs météorologiques complémentaires disponibles en version historique et projetée ont donc été intégrés dans la modélisation. Cet ajout permet la projection de la fréquence en utilisant la valeur projetée de ces indicateurs pour mettre à jour le modèle.

Les modèles utilisés sont des GLM (modèles linéaires généralisés) intégrant des variables explicatives provenant de l'assureur, ainsi que les variables météorologiques, climatiques et caractéristiques du bâtiment. Enfin, la prime pure climatique est obtenue en multipliant la fréquence prédictive par le coût moyen global, ce dernier étant calculé comme une moyenne pondérée des coûts prédicts pour chacune des sept intensités. Cette approche permet d'évaluer de manière plus précise le risque d'inondation impactant les maisons en France métropolitaine.

Une fois la prime pure modélisée, elle a été comparée à d'autres modèles et à une tarification basée uniquement sur les données de sinistralité de l'assureur. Les résultats se révèlent satisfaisants et montrent que l'intégration de données complémentaires permet d'affiner l'évaluation du risque.

Projection de la prime pure

Données et hypothèses de projection

Par la suite, nous souhaitons projeter la fréquence et le coût moyen des sinistres liés aux inondations, dans un contexte de changement climatique, à horizon 2030. Le but est d'élaborer une projection de la prime pure climatique déterminée précédemment, en prenant en compte l'évolution de la sinistralité.

Les projections faites dans cette étude reposent sur l'utilisation de modèles climatiques sophistiqués, tels que les modèles de circulation générale (MCG) et les modèles climatiques régionaux (RCM), qui permettent de simuler l'évolution du climat à l'échelle mondiale et régionale. Elles reposent également sur des scénarios de forçage radiatif (RCP et SSP) qui sont utilisés pour simuler les trajectoires futures du climat, en prenant en compte les activités humaines et les variations naturelles. Le scénario SSP2-4.5 a été retenu comme base d'analyse pour notre étude. Il est considéré comme une trajectoire intermédiaire plausible, avec une évolution socio-économique modérée et une stabilisation des émissions de gaz à effet de serre. Ce scénario se positionne d'après le sixième rapport du GIEC comme un compromis pertinent pour évaluer les impacts à moyen terme, pour son caractère réaliste, évitant les extrêmes des scénarios plus optimistes ou pessimistes, et s'aligne bien avec les tendances observées depuis 2015 (figure 3).

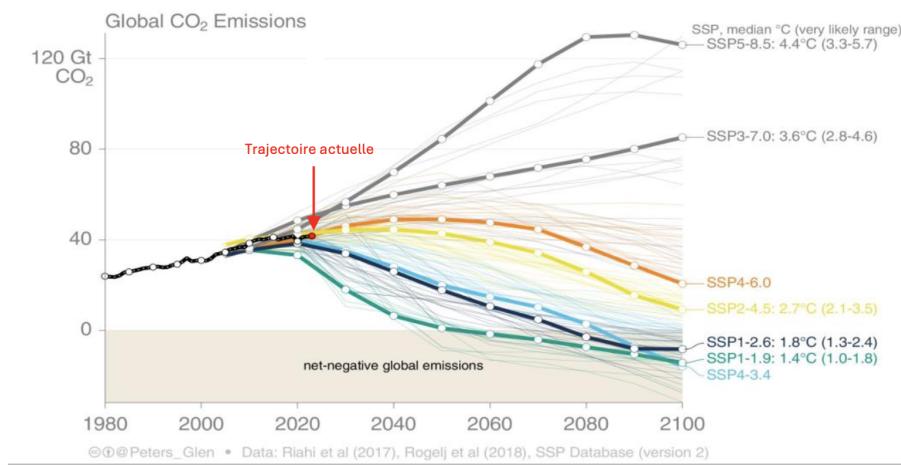


FIGURE 3 – Trajectoires des différents scénarios

Source :IPCC-AR6-TS

Les données utilisées lors de la projection ont été produites dans le cadre de l'exercice international CMIP6 (Coupled Model Inter-comparison Project Phase 6) et sont fournies par le service Copernicus sur le changement climatique. Celles-ci incluent des projections de précipitations totales pour la période 2015-2100, disponibles à l'échelle mondiale et selon divers scénarios, notamment SSP2-4.5. En complément, nous intégrons également les valeurs projetées des indicateurs météorologiques utilisés dans le modèle de fréquence.

Méthode de projection

La projection de la prime pure climatique repose ensuite sur la formule 1, en projetant séparément ses composantes principales : la fréquence des sinistres et les probabilités d'intensité.

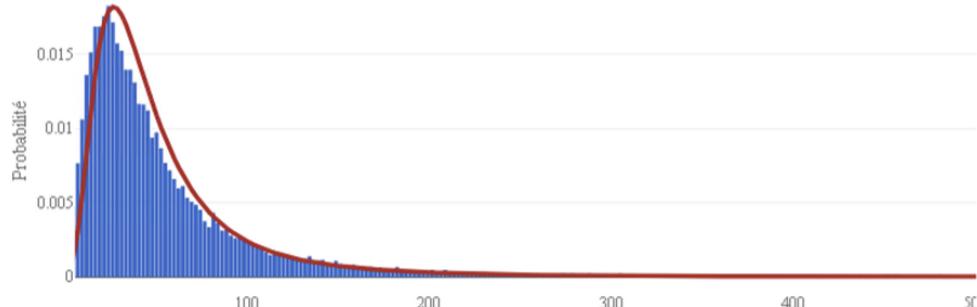
Le modèle de fréquence, précédemment élaboré à partir d'indicateurs météorologiques historiques, est appliqué aux valeurs projetées de ces mêmes indicateurs issues du scénario SSP2-4.5 (2015-2045). Cette approche permet d'estimer, pour chaque bâtiment, une fréquence projetée des sinistres (*Freq_proj*).

La projection des probabilités d'intensité repose sur l'analyse de la variable *cumul_precip*. En comparant la distribution de cette variable sur la période passée (1985-2015) et sa projection future (2015-2045) selon le scénario SSP2-4.5, on observe un décalage des quantiles de précipitations, indiquant une augmentation attendue des événements extrêmes (figure 4). Par exemple, une précipitation autrefois associée à une période de retour de 7 ans pourrait désormais survenir tous les 5 ans.

Afin d'intégrer ces évolutions, les probabilités d'intensité sont ajustées pour chaque bâtiment en tenant compte du déplacement des distributions. Concrètement, les seuils de précipitation correspondant aux différentes classes d'intensité sont recalibrés, ce qui se traduit par une augmentation des probabilités associées aux événements les plus extrêmes et une diminution de celles des événements plus modérés. À l'issue de ce processus, chaque bâtiment se voit attribuer une nouvelle série de probabilités d'intensité projetées, reflétant l'impact du changement climatique sur les risques d'inondation.



(a) Passé



(b) Projeté

FIGURE 4 – Comparaison des répartitions de la variable *cumul_precip* passée et future.

Enfin, en appliquant les nouvelles probabilités d'intensité aux coûts moyens associés à chaque classe, on obtient un coût moyen global projeté (CM_proj). La prime pure projetée pour chaque bâtiment est ensuite déterminée en multipliant ce coût moyen par la fréquence projetée :

$$Prime\ pure\ climatique\ projetée = CM_proj \times Freq_proj$$

Résultats

Ce travail permet alors de comparer la prime pure historique et la prime pure projetée pour identifier les bâtiments les plus vulnérables et quantifier les variations prévues d'ici 2030. L'évolution de la prime pure est calculée selon le ratio suivant :

$$Evolution\ Prime\ Pure\ inondation = Evol_PP = \frac{PP_{projete} - PP_{historique}}{PP_{historique}}$$

Afin d'obtenir une vision globale des évolutions sur l'ensemble du territoire français, y compris dans les zones où aucun bâtiment n'est assuré par l'assureur, une approche complémentaire est adoptée. Un bâtiment type, aux caractéristiques uniformes pour toutes les communes, est défini afin d'isoler l'impact du changement climatique sur l'évolution des primes. En appliquant les méthodes de calcul des primes pures historiques et projetées à cette base, il devient possible de cartographier les variations attendues des primes pures à l'horizon 2030 pour l'ensemble des communes de France métropolitaine.

Finalement, l'analyse de l'évolution générale des primes pures en France métropolitaine permet de conclure à une augmentation légère. En moyenne, les primes pures augmentent de 0,7% à horizon 2030, une hausse modérée mais qui révèle tout de même les impacts croissants du réchauffement climatique. Cependant, cette évolution varie selon les régions.

Si l'on fait un focus sur la région Grand Est, qui se démarque par sa diversité géographique mais aussi par son important réseau fluvial, on constate que cette région est particulièrement vulnérable aux inondations.

Nettement supérieur à la tendance nationale, l'étude projette une hausse des primes pures pour le risque inondation dans cette région de 6%. A l'image de la figure 5, tous les départements de la région affichent une évolution positive des primes, variant entre 0 et 10%.

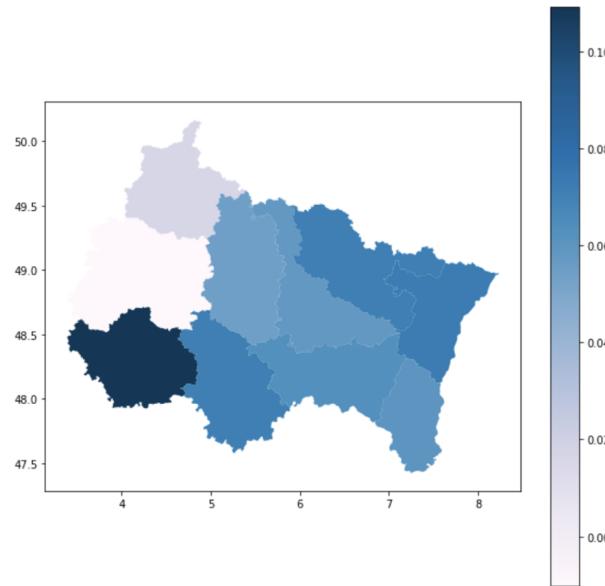


FIGURE 5 – Cartographies des évolutions moyennes de prime pure pour le risque inondation par département du Grand Est

Limites et perspectives

Bien que les résultats obtenus soient encourageants, l'étude présente cependant plusieurs limites.

Il est à noter que ce travail constitue une représentation partielle des travaux en cours au sein de la pratique. En effet, la méthode de projection a permis d'obtenir une première approximation de l'évolution du risque d'inondation en France, mais elle reste une simplification de la réalité, qui pourra être affinée au fil des travaux futurs.

De plus, nous pouvons discuter de la fiabilité des données météorologiques, provenant de la base ERA5-Land, qui peuvent présenter des écarts par rapport aux observations réelles et ne reflètent pas toujours les valeurs extrêmes nécessaires à une évaluation précise du risque. De même, les données relatives aux bâtiments sont observées à un instant donné, ce qui peut limiter leur pertinence pour une projection à long terme.

Enfin, d'autres éléments, tels que l'évolution du nombre de bâtiment ou de la répartition des populations sur le territoire pourraient être pris en compte pour affiner l'évaluation.

Toutefois, des ajustements futurs, enrichis par de nouvelles données et méthodes, permettront de renforcer la précision des projections et d'offrir une vision encore plus claire des impacts du changement climatique sur le risque d'inondation.

Conclusion

L'actualité sur les inondations en France souligne l'importance d'étudier l'impact du changement climatique sur la sinistralité future et met en avant la nécessité, pour les assureurs, d'avoir une bonne compréhension de ce risque. Ce mémoire s'inscrit dans cette optique en proposant d'intégrer des données innovantes à la modélisation afin d'affiner la tarification mais également d'avoir une meilleure connaissance du risque. Par la suite, l'étude a eu pour objectif de projeter les primes pures obtenues permettant d'avoir une vision complète de leur évolution à horizon 2030 sur le territoire français. Finalement, ces premiers résultats offrent des perspectives intéressantes pour une meilleure gestion du risque inondation.

Synthesis Note

Context and Problem

Natural disasters in France, although relatively rare, result in significant claims costs. Among these events, floods hold a central position for insurers, accounting for more than half of the Natural Disaster (Cat Nat) declarations and approximately 50% of the total claims cost, which amounts to €49.9 billion under the Cat Nat coverage from 1982 to 2022.

In recent years, the frequency and intensity of climate-related events have increased significantly, a trend confirmed by future climate projections. The Intergovernmental Panel on Climate Change (IPCC) predicts an intensification of these risks, with extreme precipitation events becoming increasingly frequent. Similarly, the CCR (Caisse Centrale de Réassurance) anticipates a rise in climate-related claims costs ranging from 47% to 85% by 2050 due to the intensification of extreme weather events.

In this context, insurers are compelled to adapt their pricing models to better assess risk trends and adjust their long-term strategies. This study aims, on the one hand, to refine risk pricing by incorporating external data into modeling and, on the other hand, to project the obtained pure premiums to 2030 to better anticipate the future impacts of climate change.

Description of the Data Used

To establish a pure premium for flood risk in mainland France (excluding Corsica), we rely on three complementary databases. The first is a claims database provided by an insurer, containing information about the policyholder, their residence, the subscribed contract, and the claims history. This database enables the calculation of "Frequency" and "Average Cost" bases, which are used in non-life insurance pricing.

Thanks to the geolocation of contracts, this database is enriched with additional data from a dataset developed by Addactis in partnership with NamR. These new data cover both building characteristics (surface area, number of rooms, roof type, presence of solar panels, etc.) and its environment (distance to water bodies, proximity to emergency services, population density, land use, etc.). They originate from official sources such as INSEE, the Ministry of Ecological Transition and Solidarity (MTES), or ADEME (Agency for Ecological Transition).

Additionally, meteorological data from the ERA5-Land database, a reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus program, are incorporated. These data are derived from past observations and climate models. With a resolution of 9 km and daily updates of hourly values, this dataset enables the addition of a daily cumulative precipitation variable. Figure 6 summarizes the key information.



FIGURE 6 – Summary Diagram of the Three Databases Used

Construction of a Precipitation Index

The first step in building this pure premium is to understand the flood phenomenon from a physical perspective and identify the key parameters behind these events. According to the "Flood" section of GEORISQUES s. d.(a), a flood is defined as a temporary, rapid or slow submersion of an area that is usually dry. This phenomenon is primarily caused by sustained and prolonged precipitation. However, unlike storms, for example, where there is a direct impact between wind strength and building damage, in the case of floods, precipitation is just one factor among others, as its effect depends on how water behaves once it reaches the ground.

To better account for this complexity, we have developed a tailored precipitation index. The construction of this index involves several steps :

First, we identify the affected buildings and associate them with the four closest ERA5-Land grid points. For these points, we collect precipitation totals over various periods preceding the flood (ranging from 1 to 30 days) to analyze the effects of rainfall episodes. The goal is to determine whether the flood results from an intense and sudden event or a prolonged and sustained one. For each studied period, we retain the maximum observed precipitation among the four selected points.

Next, we establish the historical distribution (from 1985 to 2015) of precipitation totals for each considered period. This allows us to assign each total a quantile value relative to the historical record. For example, if the one-day total corresponds to a moderate quantile while the 30-day total reaches the 99th quantile, we can assume that the flood is linked to prolonged precipitation. Conversely, a very high total over a short period indicates a sudden and intense rainfall episode. Figure 7 illustrates the historical distribution of three-day precipitation totals for a flood-affected building. Three days before the flood, the precipitation total corresponded to the 99th quantile compared to historical totals.

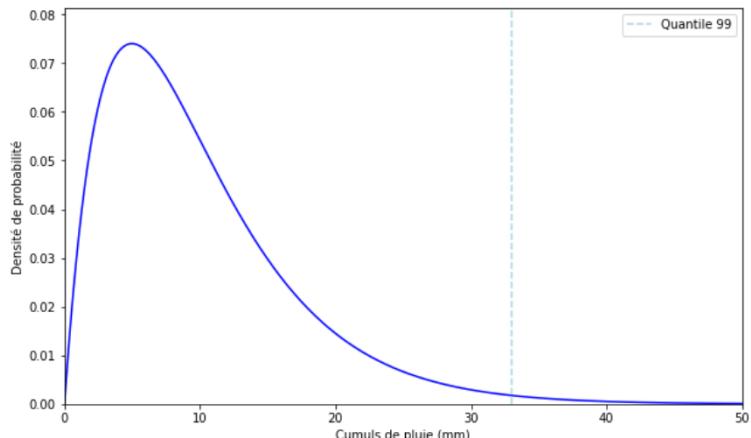


FIGURE 7 – Distribution of Historical 3-Day Precipitation Totals

Finally, for each affected building, the maximum of the quantiles is selected. This value then becomes, for each building in the database, the variable "cumul_precip", representing precipitation intensity in accordance with the specific characteristics of flood risk.

Method for Calculating the Pure Flood Risk Premium

The methodology for calculating the pure climatic premium follows a similar approach to the one commonly used in non-life insurance pricing, while incorporating a conditioning based on the intensity of floods. The pure climatic premium for the flood risk is defined by the following equation :

$$PP_{\text{climatic}} = \sum_i P(\text{intensity}_i) \times \mathbb{E}(\text{frequency}_{\text{claims}}) \times \mathbb{E}(\text{cost}_{\text{claims}} \mid \text{flood}_{\text{intensity}_i}) \quad (2)$$

This formula derives from the principle of total probability and breaks down the average cost of claims based on the different flood intensities i . The goal is to estimate the probability of the occurrence of a flood of a given intensity in order to incorporate this information into the calculation of the pure climatic premium. The intensity of the floods is evaluated using the variable *cumul_precip*. This variable is discretized into seven intensity classes, each associated with a quantile interval and an occurrence probability.

The table 2 summarizes this discretization :

Intensity	<i>cumul_{precip}</i> discretized	Associated Probability
1	[0,0.75[0.75
2	[0.75,0.85[0.10
3	[0.85,0.95[0.10
4	[0.95,0.97[0.02
5	[0.97,0.98[0.01
6	[0.98,0.999[0.019
7	[0.999,1[0.001

TABLE 2 – escription of the different intensities

Then, the average claims cost model defined in equation 2 is constructed according to the flood intensity i . However, the formula used does not link the frequency of claims to the intensity of floods, as a contract database does not contain specific flood dates, but only an indication of impact during the contract's validity period. Other complementary meteorological indicators, available in both historical and projected versions, have therefore been integrated into the modeling. This addition allows for the projection of frequency by using the projected values of these indicators to update the model.

The models used are GLMs (Generalized Linear Models) incorporating explanatory variables from the insurer, as well as meteorological, climatic, and building characteristics. Finally, the pure climatic premium is obtained by multiplying the predicted frequency by the overall average cost, which is calculated as a weighted average of the predicted costs for each of the seven intensities. This approach allows for a more precise evaluation of the flood risk impacting houses in mainland France.

Once the pure premium was modeled, it was compared to other models and to a pricing method based solely on the insurer's claims data. The results are satisfactory and show that the integration of complementary data helps refine the risk assessment.

Projection of the Pure Premium

Data and Projection Assumptions

Subsequently, we aim to project the frequency and average cost of flood-related claims in the context of climate change, with a 2030 horizon. The goal is to develop a projection of the previously determined pure climatic premium, taking into account the evolution of claims frequency.

The projections made in this study are based on the use of sophisticated climate models, such as General Circulation Models (GCMs) and Regional Climate Models (RCMs), which allow for simulating climate evolution at both global and regional scales. They are also based on radiative forcing scenarios (RCP and SSP) that are used to simulate future climate trajectories, taking into account human activities and natural variations. The SSP2-4.5 scenario has been chosen as the baseline for our analysis. It is considered a plausible intermediate trajectory, with moderate socio-economic development and stabilization of greenhouse gas emissions. According to the sixth IPCC report, this scenario is viewed as a relevant compromise for assessing medium-term impacts, due to its realistic nature, avoiding the extremes of more optimistic or pessimistic scenarios, and aligning well with trends observed since 2015 (figure 8).

The data used in the projections were produced as part of the international CMIP6 (Coupled Model Inter-comparison Project Phase 6) exercise and are provided by the Copernicus Climate Change Service. These include projections of total precipitation for the period 2015-2100, available globally and according to

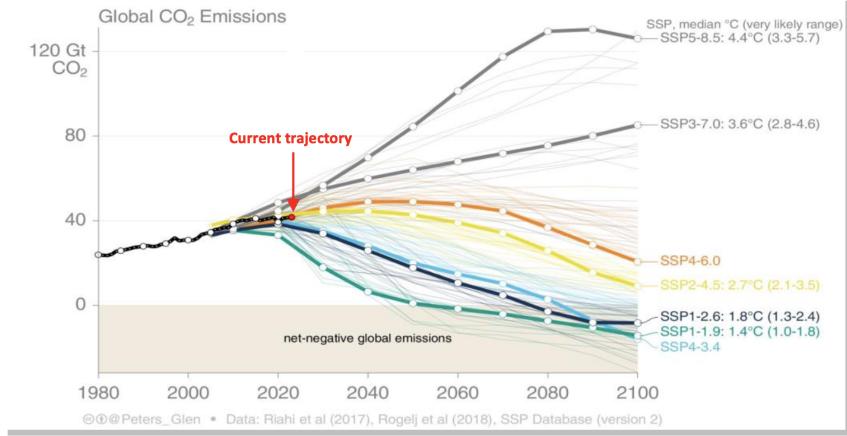


FIGURE 8 – Trajectories of the different scenarios

Source :IPCC-AR6-TS

various scenarios, including SSP2-4.5. Additionally, we also integrate the projected values of the meteorological indicators used in the frequency model.

Projection Method

The projection of the pure climatic premium is then based on equation 2, by separately projecting its main components : the frequency of claims and the intensity probabilities.

The frequency model, previously developed from historical meteorological indicators, is applied to the projected values of these same indicators from the SSP2-4.5 scenario (2015-2045). This approach allows for estimating, for each building, a projected claims frequency ($Freq_{proj}$).

The projection of intensity probabilities is based on the analysis of the variable *cumul_precip*. By comparing the distribution of this variable over the past period (1985-2015) and its future projection (2015-2045) under the SSP2-4.5 scenario, a shift in precipitation quantiles is observed, indicating an expected increase in extreme events (figure 9). For example, a precipitation event previously associated with a return period of 7 years could now occur every 5 years.

In order to integrate these changes, the intensity probabilities are adjusted for each building by taking into account the shift in the distributions. Specifically, the precipitation thresholds corresponding to the different intensity classes are recalibrated, which results in an increase in the probabilities associated with the most extreme events and a decrease in those of more moderate events. At the end of this process, each building is assigned a new set of projected intensity probabilities, reflecting the impact of climate change on flood risks.

Finally, by applying the new intensity probabilities to the average costs associated with each class, a projected overall average cost (CM_{proj}) is obtained. The projected pure premium for each building is then determined by multiplying this average cost by the projected frequency :

$$\text{Projected pure climatic premium} = CM_{proj} \times Freq_{proj}$$

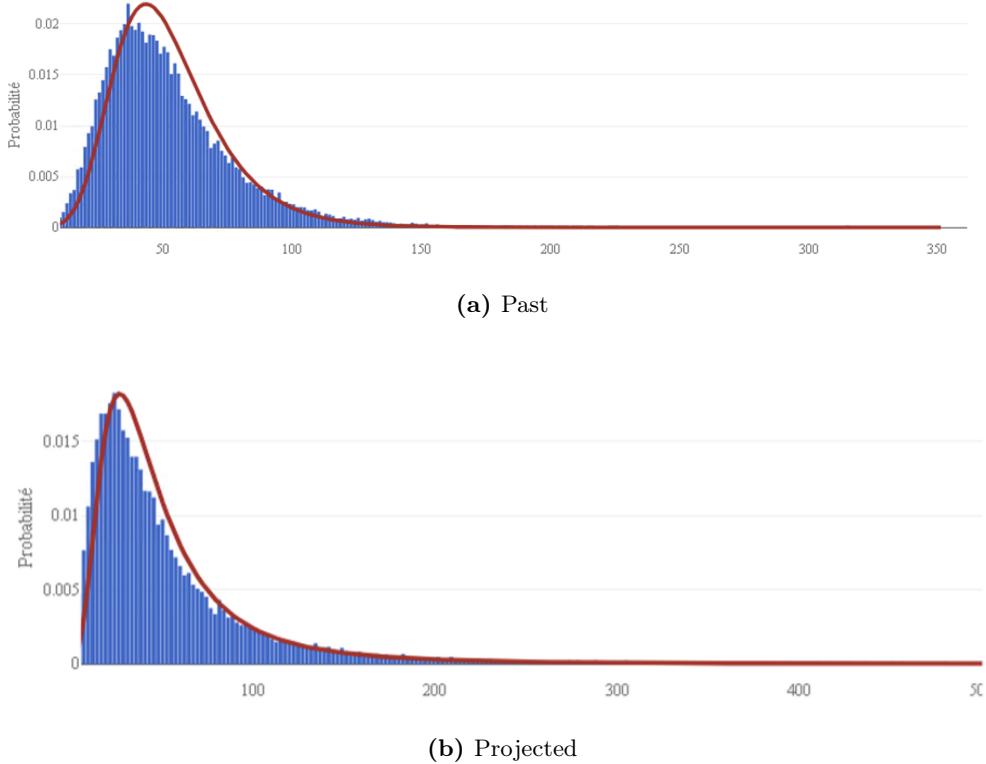


FIGURE 9 – Comparison of the past and future distributions of the variable *cumul_precip*.

Results

This work allows for comparing the historical pure premium and the projected pure premium to identify the most vulnerable buildings and quantify the expected variations by 2030. The evolution of the pure premium is calculated according to the following ratio :

$$\text{Evolution of Pure Premium flood} = \text{Evol_PP} = \frac{PP_{\text{projected}} - PP_{\text{historical}}}{PP_{\text{historical}}}$$

In order to obtain a comprehensive view of the changes across the entire French territory, including areas where no buildings are insured by the insurer, a complementary approach is adopted. A typical building, with uniform characteristics for all municipalities, is defined in order to isolate the impact of climate change on the evolution of premiums. By applying the methods for calculating both historical and projected pure premiums to this base, it becomes possible to map the expected variations in pure premiums by 2030 for all municipalities in mainland France.

Finally, the analysis of the general evolution of pure premiums in mainland France concludes in a slight increase. On average, pure premiums increase by 0.7% by 2030, a moderate rise, but one that still reveals the growing impacts of climate change. However, this evolution varies across regions.

Focusing on the Grand Est region, which stands out for its geographical diversity as well as its extensive river network, we observe that this region is particularly vulnerable to flooding. Significantly higher than the national trend, the study projects an increase of 6% in pure premiums for flood risk in this region. As shown in figure 10, all departments in the region show a positive evolution of premiums, ranging between 0% and 10%.

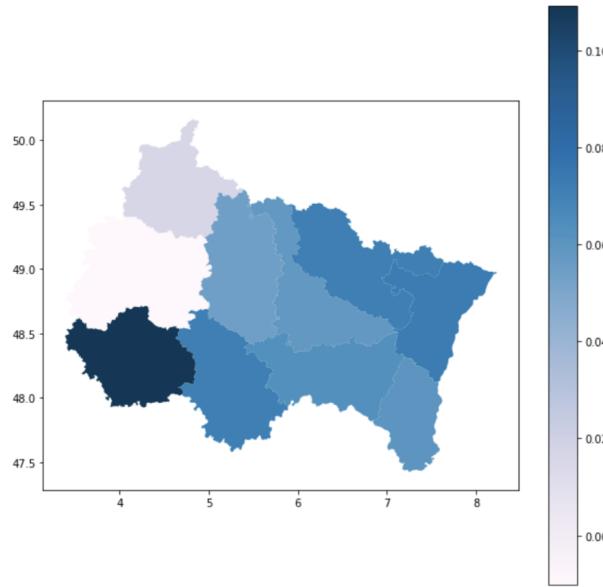


FIGURE 10 – Mapping of the average changes in pure premiums for flood risk by department in the Grand Est region.

Limitations and Perspectives

Although the results obtained are encouraging, the study does have several limitations.

It should be noted that this work represents a partial snapshot of ongoing efforts within the practice. Indeed, the projection method has provided a first approximation of the evolution of flood risk in France, but it remains a simplification of reality, which can be refined as future work progresses.

Additionally, we can discuss the reliability of the meteorological data from the ERA5-Land database, which may exhibit discrepancies compared to actual observations and may not always reflect the extreme values required for an accurate risk assessment. In the same way, building data is observed at a specific point in time, which may limit its relevance for long-term projections.

Finally, other factors, such as the evolution of the number of buildings or the distribution of populations across the territory, could be considered to refine the assessment.

However, future adjustments, enriched by new data and methods, will help enhance the accuracy of the projections and provide an even clearer view of the impacts of global warming on flood risk.

Conclusion

The current situation regarding floods in France highlights the importance of studying the impact of climate change on future claims and emphasizes the need for insurers to have a good understanding of this risk. This thesis was developed with this goal in mind, proposing the integration of innovative data into modeling to refine pricing as well as gain a better understanding of the risk. Subsequently, the study aimed to project the pure premiums obtained, providing a comprehensive view of their evolution by 2030 across French territory. Ultimately, these initial results offer promising perspectives for better flood risk management.

Table des matières

Résumé	3
Abstract	4
Note de Synthèse	5
Synthesis Note	11
Remerciements	19
Introduction	20
1 Généralités	21
1.1 Généralités et contexte climatique	21
1.1.1 L'assurance habitation	21
1.1.2 La sinistralité climatique en France	23
1.1.3 Le phénomène inondation	27
1.1.4 Le risque inondation	29
1.2 Description des éléments	31
1.2.1 Base de données assureur	31
1.2.2 Base de données complémentaire au bâtiment	32
1.2.3 Base de données météorologique et expertises climatiques	33
2 Mise en place des modèles	38
2.1 Méthodologie	38
2.1.1 La tarification en assurance non-vie	38
2.1.2 Calcul de la prime pure climatique	39
2.1.3 Modèle linéaire généralisé	40
2.1.4 Méthodes de validation	45
2.2 Modèle de fréquence	47
2.2.1 Analyse exploratoire des données	47
2.2.2 Sélection de variables	48
2.2.3 Modélisation	49
2.2.4 Validation du modèle	58
2.3 Modèle de coût moyen	61
2.3.1 Analyse exploratoire des données	61
2.3.2 Sélection de variables	63
2.3.3 Modélisation	63
2.3.4 Résultats et validation	69
2.4 Obtention de la prime pure	71
3 Projection	73
3.1 Éléments de projection	73
3.1.1 Contexte	73
3.1.2 Définition et choix du scénario	75
3.1.3 Présentation des données de projection	79
3.2 Méthode de projection	79
3.3 Analyse et interprétation des résultats	84
3.3.1 Analyse des résultats à l'échelle nationale	84

3.3.2 Analyse des résultats à l'échelle régionale - Focus sur le Grand Est	84
3.4 Limites et perspectives	86
Conclusion	88
Bibliographie	90

Remerciements

Je remercie toute l'équipe d'Addactis France pour son accueil et son encadrement. Je remercie particulièrement Mullah Moriah pour son suivi, son expertise et ses nombreux conseils qui ont guider ma réflexion durant la réalisation de ce mémoire. Je tiens également à remercier Mederick Besarabov, Bilal Sadou, Franck Baton, Mathilde Rochelle et tous mes collègues de la practice Pricing & Analytics P&C avec qui j'ai eu la chance de travailler et d'échanger durant mon alternance.

J'aimerais ensuite remercier l'ensemble de mes professeurs et intervenants de l'ISFA, notamment Frédéric Planchet, mon tuteur académique, pour ses conseils.

Je remercie chaleureusement Yohann, Marie et mes parents, pour leur accompagnement, leur relecture et leur soutien.

Introduction

Ces dernières années, la préoccupation des français à l'égard des catastrophes naturelles s'est fortement accrue. Cette augmentation a pu être observée à la suite d'évènements météorologiques extrêmes tels que la tempête Xynthia (2010), les inondations dans l'Aude (2018) et dans les Alpes-Maritimes (2020) (MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE ET DE LA COHÉSION DES TERRITOIRES 2023), indiquant un tournant dans la sinistralité climatique, qui impacte la vie de la population. L'année 2023, par exemple, devient l'année la plus sinistrée depuis 1989 et enregistre des records de chaleur et de jours consécutifs sans pluie selon Meteo-France. La CCR prévoit une hausse de la sinistralité climatique comprise entre 47% et 85% d'ici 2050 en raison de l'intensification des événements climatiques (CCR 2023a).

Ces prévisions posent de nombreux défis pour les assureurs, qui devront se préparer à une hausse significative des dommages couverts liés aux risques climatiques.

Parmi eux, le risque inondation suscite particulièrement l'attention ; il est perçu comme l'un des plus préoccupants en France (MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE ET DE LA COHÉSION DES TERRITOIRES 2023). Entre 1982 et 2022, 54% des arrêtés CAT NAT correspondaient à des inondations. Placé au cœur de nombreuses études, ce péril occasionne de nombreux dégâts, dont un coût moyen annuel pour les assureurs de 600 millions d'euros, qui devrait connaître, d'après les projections climatiques de la CCR (CCR 2023a), une augmentation comprise entre 6% et 20% selon les scénarios mis en place par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), d'ici 2050.

Ces phénomènes à forte intensité sont parfois à l'origine de lourds bilans comme cela a été le cas lors des inondations de mai-juin 2016 qui ont fait 4 morts, 24 blessés et un montant cumulé de dommages de l'ordre d'un milliard d'euros dans le Loiret, le Loir-et-Cher, l'Essonne, la Seine-et-Marne et l'Yonne. Cependant, l'impact matériel d'une inondation ne se mesure pas uniquement à l'intensité du phénomène lui-même ; il dépend également des caractéristiques et de l'environnement de chaque bâtiment, dont la plupart sont couverts par la garantie Catastrophes naturelles de l'assurance multirisques habitation.

Face à cette menace croissante, le marché de l'assurance dommage se voit contraint d'adapter ses offres et ses politiques. Ainsi l'étude de ces phénomènes et de leurs conséquences actuelles et futures sur les bâtiments devient essentiel afin de pouvoir mieux les appréhender. De plus en plus de sociétés d'assurance ont leur propre modèle de catastrophes naturelles afin d'accéder à une meilleure connaissance de leur risque (CHICHE 2013). C'est dans cette optique qu'Addatis France travaille actuellement sur le développement d'une nouvelle approche pour le service de souscription MRH, en utilisant des données innovantes (météorologiques, environnementales, économiques ainsi que les caractères physiques du bâtiment) disponibles à la maille adresse. Ce mémoire a donc pour objectif d'établir une prime pure, propre à chaque bâtiment situé en France métropolitaine hors Corse, à l'appui de données de sinistralité historique, de données complémentaires à la maille adresse et de données météorologiques pour le risque inondation. L'intérêt est ensuite de tenter de comprendre son évolution en projetant cette prime à horizon 2030 en se basant sur les projections climatiques des météorologues en lien avec les scénarios du GIEC.

La première partie (1) sera consacrée à l'introduction du contexte de l'étude, ainsi qu'à la présentation des bases de données (météorologique, sinistralité et complémentaire à l'adresse) utilisées pour la modélisation. La seconde partie (2) détaillera la méthodologie employée pour la modélisation de la prime pure. La mise en place de la projection fera l'objet du Chapitre 3.

Chapitre 1

Généralités

1.1 Généralités et contexte climatique

Cette première partie a pour objectif d'introduire le contexte de l'étude et d'en préciser les enjeux ainsi que les objectifs.

1.1.1 L'assurance habitation

L'assurance habitation est une couverture essentielle dans la vie quotidienne, qui offre à l'assuré une protection pour son patrimoine familial, contre les conséquences d'incidents touchant son domicile ou impliquant sa responsabilité ou la responsabilité de sa famille (MINISTÈRE DE L'ÉCONOMIE DES FINANCES ET DE LA SOUVERAINETÉ INDUSTRIELLE ET NUMÉRIQUE 2023).

En France métropolitaine, en 2023, 82% des logements sont des résidences principales, leur nombre s'élève à 31 millions. Les autres logements sont des résidences secondaires ou des logements vacants. En ce qui concerne les biens de particuliers, ils sont composés à 57% de maisons et à 43% d'appartements, ces proportions restent stables au cours du temps. Dans toute la suite de ce mémoire, nous prendrons en compte uniquement les logements individuels désignés comme "maisons", que ce soit des résidences principales ou secondaires. Selon la publication 2023 des chiffres clés du logement (INSEE 2023), les logements individuels représentent 54,8% du parc en France.

La majorité des assureurs proposent, afin de protéger ce parc de logements, des assurances multirisques habitation (MRH), couvrant principalement les dommages aux biens et la responsabilité civile, y compris la responsabilité civile "vie privée".

- La garantie dommages aux biens permet de se couvrir contre les dommages qui pourraient affecter les biens de l'assuré tels que l'incendie, les dégâts des eaux, le vol, les tempêtes, les catastrophes naturelles, etc.
- La garantie responsabilité civile de l'assuré indemnise les tierces victimes de dommages dont il serait responsable.
- La garantie responsabilité civile "vie privée" couvre les membres de son foyer en cas de dommages causés à un tiers, incluant les dommages corporels, matériels et immatériels.

L'assurance Multirisque Habitation est la deuxième plus importante sur le marché des assurances de dommages aux biens et de responsabilité, juste derrière l'assurance automobile. En effet, le marché de l'assurance habitation maintient sa dynamique de croissance avec un nombre de contrats MRH qui s'élève à 45,4 millions, en parallèle avec la hausse du nombre de logements. La cotisation totale en 2022 pour ces contrats représentait 12,2 milliards d'euros (FRANCE ASSUREURS 2023c).

Cependant, la sinistralité a été amplifiée par des épisodes climatiques extrêmes, notamment de fortes grêles, entraînant une hausse du nombre de sinistres de 35% en 2022 par rapport à l'année 2021, et une augmentation significative des coûts associés. En conséquence, le ratio sinistres/primes s'est détérioré. Malgré une augmentation des primes de 2,4% par rapport à 2021, les charges de prestations sont toujours en forte hausse, exerçant ainsi une pression financière croissante sur les assureurs. Cette situation est en grande partie due aux catastrophes naturelles, qui attirent désormais une attention particulière.

Concernant la sinistralité en assurance Habitation, la fréquence de la garantie Catastrophes naturelles est très faible par rapport aux autres garanties multirisques habitation. Comme le montre la figure 1.1, elle est égale, pour l'année 2022 à 0,06 % pour une fréquence globale de sinistres MRH de 91,9 %.

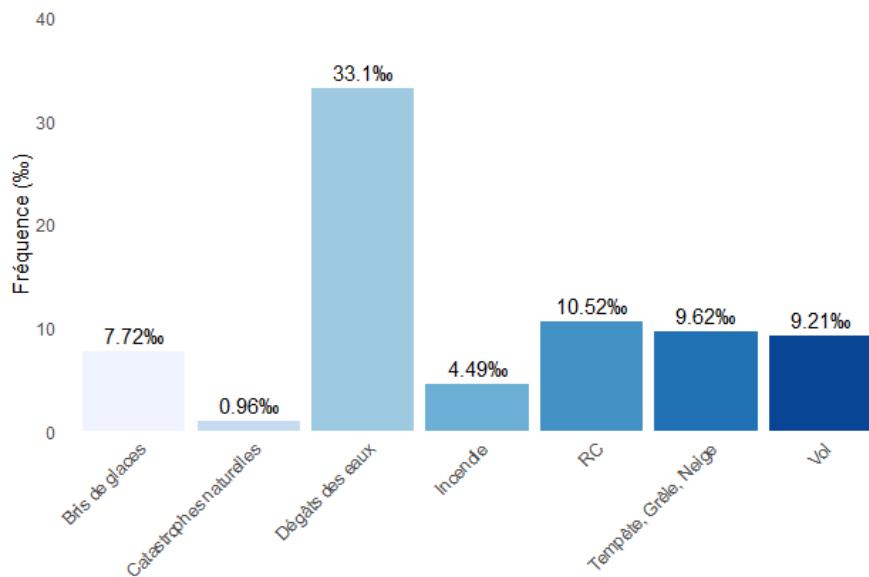


FIGURE 1.1 – Fréquence de sinistres entre 2013 et 2022 en assurance MRH selon les garanties

Source : « *L'assurance Habitation en 2022 - Assurances de Biens et de Responsabilité* » (2023), « *L'assurance française - Données clés 2018* » (2019), « *L'assurance française - Données clés 2020* » (2021)

S'ils sont marqués par une fréquence faible en assurance habitation, les sinistres catastrophes naturelles représentent en revanche un coût moyen très important. En effet, comme l'illustre la figure 1.2, le coût moyen des sinistres survenus entre 2013 et 2022 pour la garantie catastrophe naturelle, dépasse largement le coût des autres garanties. Il atteint en moyenne 15090 €.

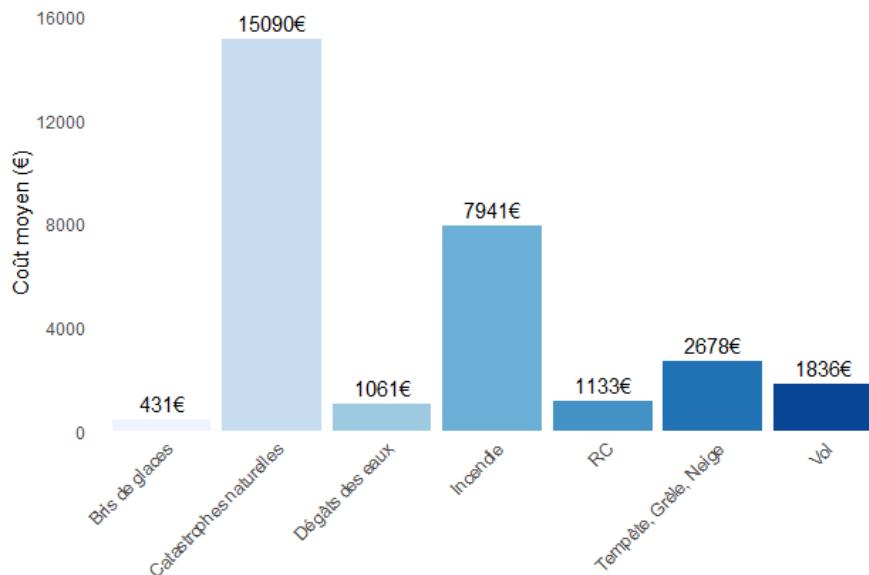


FIGURE 1.2 – Coût moyen des sinistres entre 2013 et 2022 en assurance MRH selon les garanties

Source : « *L'assurance Habitation en 2022 - Assurances de Biens et de Responsabilité* » (2023), « *L'assurance française - Données clés 2018* » (2019), « *L'assurance française - Données clés 2020* » (2021)

Ces coûts élevés reflètent l'intensité des événements récents, mais ils s'inscrivent également dans un contexte plus large de réchauffement climatique.

1.1.2 La sinistralité climatique en France

Le climat a effectivement évolué au cours des dernières années et la sinistralité due aux événements naturels a significativement augmenté. Celle-ci est complexe puisqu'elle résulte non seulement de l'intensité et de la fréquence des événements extrêmes, mais également de l'exposition et de la vulnérabilité des territoires et des assurés. De plus, l'efficacité des mesures de prévention et de sensibilisation ont également un impact important sur l'atténuation de cette sinistralité. Cette partie vise donc, afin de mieux comprendre l'enjeu de l'étude, à examiner le contexte climatique actuel et futur.

Tous les contrats d'assurance de biens (multirisques habitation, multirisques entreprise, agricoles, etc.) incluent obligatoirement une garantie contre les catastrophes naturelles. Le montant de cette garantie est un taux de surprime fixé par décret. En effet, depuis le 13 juillet 1982, il existe en France, un régime d'indemnisation des catastrophes naturelles obligatoire. Ce régime a permis de combler un manque de couverture face aux risques naturels, après des années de pertes dues aux événements naturels et à la réticence des assureurs à couvrir ces risques (FRANCE ASSUREURS 2024c).

Définition d'une catastrophe naturelle

Une catastrophe naturelle est caractérisée par un événement imprévisible lié à un phénomène non généré par l'homme (INSEE 2016).

Parmi elles, figurent les tempêtes, les inondations, les cyclones, les glissements de terrain, les incendies, etc. Comme vu précédemment, les catastrophes naturelles se caractérisent par une fréquence relativement faible mais génèrent souvent des pertes importantes tant en vies humaines qu'en biens. Cette distinction est notable par rapport aux sinistres plus fréquents, mais de moindre coût.

Selon l'article L125-1 du code des assurances, la garantie Cat-Nat prend en charge les « dommages matériels directs non assurables ayant eu pour cause déterminante l'intensité anormale d'un agent naturel, lorsque les mesures habituelles à prendre pour prévenir ces dommages n'ont pu empêcher leur survenance ou n'ont pu être prises. » (GEORISQUES s. d.(b)). On peut établir une liste non exhaustive des périls couverts par le régime :

- Les inondations
- Les phénomènes liés à l'action de la mer
- Les mouvements de terrains
- Les avalanches
- Les séismes
- Le volcanisme
- Les vents cycloniques

Les dommages provoqués par des vents qui ne sont pas caractérisés de cyclonique, la grêle ou la neige, sont couvert par d'autres garanties des contrats dommages aux biens.

Procédure d'indemnisation

Afin de pouvoir débuter la procédure d'indemnisation des catastrophes naturelles, deux conditions doivent être remplies :

- Les biens endommagés sont assurés par une assurance de dommages aux biens (multirisques habitation, multirisques automobile, local professionnel, etc.). Les biens couverts uniquement par une assurance responsabilité civile ne bénéficient pas de cette garantie.
- Un arrêté ministériel publié dans le journal officiel de la République française (JO) doit reconnaître la catastrophe naturelle comme telle, en précisant les zones touchées, les périodes auxquelles les faits se sont produits et la nature des dommages causés par l'événement.

Lorsqu'un événement de type catastrophe naturelle se produit, les maires des communes concernées, doivent adresser une demande de reconnaissance de l'état de catastrophe naturelle auprès des préfets de départements dans un délai maximum de 24 mois après l'événement. Les préfets transmettent l'ensemble des demandes au ministère de l'intérieur. Une commission ministérielle sur l'événement étudie l'état de catastrophe naturelle ou non, et s'il est avéré, un arrêté est signé et publié au Journal Officiel.

En contre partie de l'obligation de proposer une garantie catastrophe naturelle pour tous les contrats de dommages aux biens et responsabilités, l'Etat intervient comme réassureur auprès des compagnies d'assurance via la CCR (Caisse Centrale de Réassurance). La CCR est une Société Anonyme détenue à 100% par l'Etat français. Ses objectifs sont de :

- Maintenir l'équilibre financier du régime d'indemnisation des catastrophes naturelles.
- Evaluer les conséquences financières des catastrophes naturelles
- Mesurer l'exposition des territoires à des évènements possibles.
- Contribuer à la prévention des risques naturels.

Pour les contrats relatifs aux charges couvertes par le régime CAT-NAT, l'Etat offre une garantie illimitée, c'est à dire qu'il n'y a pas de limite maximale à la prise en charge des sinistres par la CCR.

La figure 1.3 retrace les différentes étapes de ce mécanisme d'indemnisation des catastrophes naturelles.

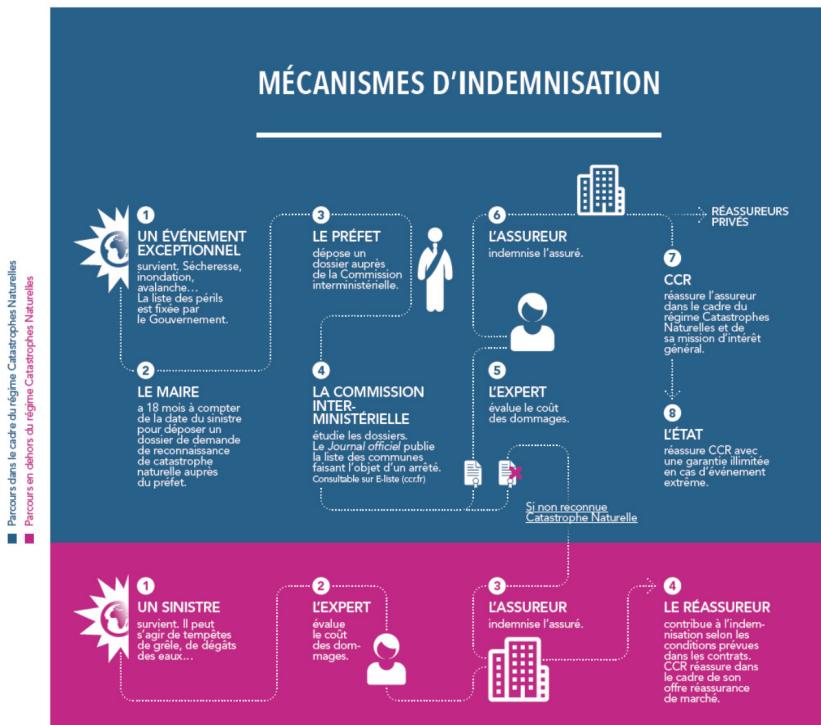


FIGURE 1.3 – Mécanismes d'indemnisation des sinistres dépendant de la garantie Catastrophes naturelles et dépendant de la garantie TGN

Source : *Caisse Centrale de Réassurance*

Mesures de prévention et de sensibilisation

L'augmentation des risques, mentionnée précédemment, soulève la question des conditions d'assurabilité, où la prévention joue un rôle essentiel.

En 1982, la loi sur l'indemnisation des victimes de catastrophes naturelles a instauré le plan d'exposition aux risques (PER) pour encourager la prévention parmi les assurés. En 1995, les **plans de prévention des risques naturels prévisibles (PPRN)** ont remplacé ces PER. Ils ont pour objectif de réduire l'exposition au risque et la vulnérabilité des biens et des personnes face aux phénomènes dommageables. La même année, le fonds Barnier a été créé pour financer les mesures visant à réduire la vulnérabilité et à promouvoir la prévention, conformément à l'article L561-3 du code de l'environnement. Ce fonds, géré par la CCR, est partiellement financé par les cotisations des contrats multirisques habitation (MRH), avec environ 12% des primes des contrats MRH revenant au fonds Barnier. La mise en œuvre de cette politique de prévention est une compétence partagée, impliquant les services déconcentrés de l'État, les collectivités territoriales, plusieurs ministères, ainsi que les citoyens.

Le PPRN est composé d'un rapport de présentation, d'un plan de zonage et d'un règlement décrivant les contraintes constructives. Il permet d'identifier les zones exposées aux événements naturels et d'évaluer leur intensité potentielle. A l'intérieur de ces zones, ils réglementent les futures constructions comme les existantes dans le but de gérer et limiter leur vulnérabilité. La figure 1.4 représente les effets potentiels de la mise en place d'un PPR pour le risque inondation.

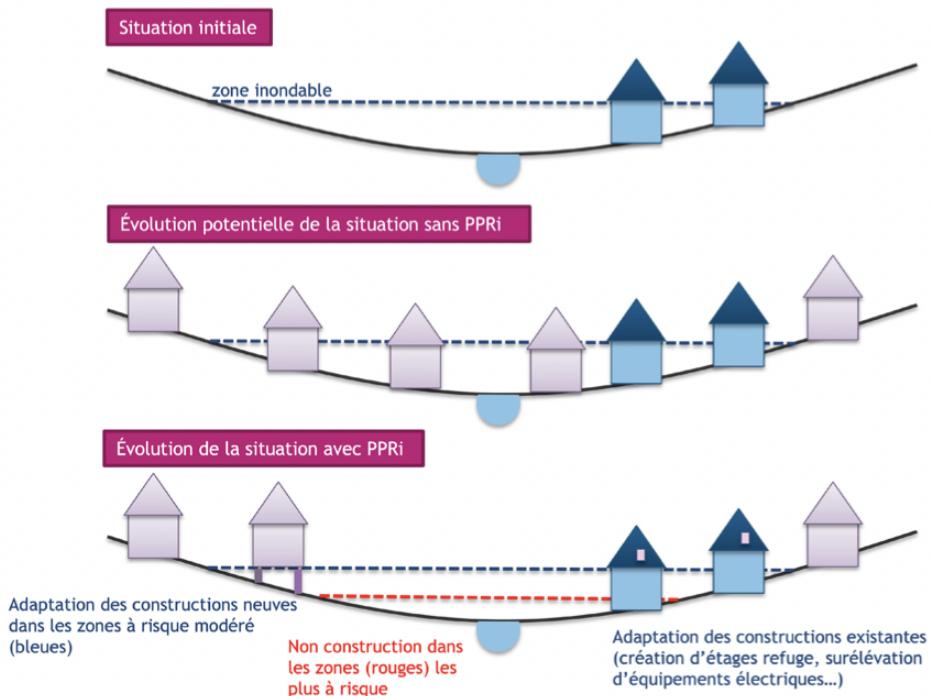


FIGURE 1.4 – Schéma théorique simplifié de l'impact du PPRI sur la sinistralité

Source : *Caisse Centrale de Réassurance (2023)*

Une partie de la section 1.2.3 sera consacrée à l'ajout des PPRI (Plans de Préventions des Risques Inondations) dans la base de modélisation.

Ce système de garantie Cat Nat français a ainsi montré une efficacité notable pour réduire les risques d'absence d'assurance ou de nécessité d'intervention étatique post-catastrophe. Il permet également une mutualisation partielle des primes, facilitant ainsi l'accès à l'assurance dans les zones les plus exposées aux risques. Toutefois, il nécessite un renforcement et un financement accru pour maintenir son efficacité à long terme au vu de la hausse de la sinistralité.

Sinistralité et impact sur les biens immobiliers

Quelle que soit l'exposition aux événements naturels, la cotisation catastrophes naturelles est fixe et correspond à un taux de surprime de :

- 20% sur l'assurance dommages aux biens des particuliers et professionnels
- 9% sur les garanties vol et incendie de l'assurance automobile.

Au 1er janvier 2025, ces taux, qui n'avaient pas été modifiés depuis plus de 25 ans, sont passés de 12 à 20% sur les contrats d'assurance des biens d'habitation et professionnels, et de 6 à 9% sur les contrats automobiles. La hausse des primes d'assurance de dommages aux biens a conduit à une augmentation des cotisations émises au titre des catastrophes naturelles de 5,7% et atteint 1 907 M € dont 50% environ pour les particuliers hors automobile. Cette révision des taux est notamment justifiée par l'augmentation des sinistres liés aux catastrophes naturelles, conséquence directe du dérèglement climatique. Entre 1989 et 2022, CCR a recensé 199 événements climatiques couverts par le régime Cat Nat. La figure 1.5 présente l'historique des principaux événements

survenus entre 1989 et 2019. La figure 1.6 présente certains évènements survenus entre 2020 et 2022.

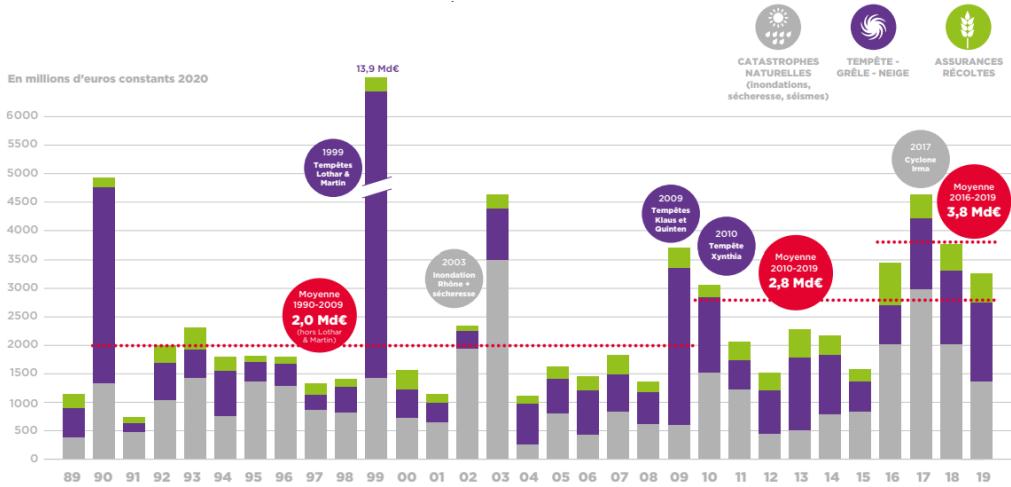


FIGURE 1.5 – L'historique des indemnisations versées par les assureurs à la suite d'aléas naturels

Source : «Impact du changement climatique sur l'assurance à horizon 2050» (2020)

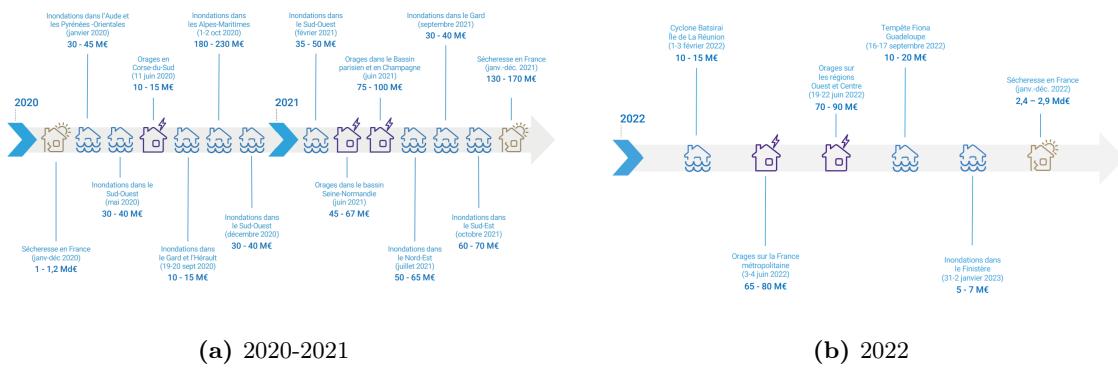


FIGURE 1.6 – Les principaux évènements naturels survenus entre 2020 et 2022

Source : « Les catastrophes naturelles en France - Bilan 1982-2022 » (2023) & « Les catastrophes naturelles en France - Bilan 1982-2021 » (2022)

En 2023, la sinistralité climatique prend une nouvelle dimension tant en fréquence qu'en sévérité. Elle obtient le record de l'année la plus sinistrée entre 1989 et 2023, enregistrant 37 sinistres graves et la troisième année la plus coûteuse derrière 2022 et 1999, avec un coût de sinistres liés au climat de 6,5 milliards d'euros. Elle enregistre par exemple, le record du nombre de jours sans pluie (déjà battu en 2020), et est la deuxième année la plus chaude selon Météo-France.

Ainsi, l'année 2023 reflète la tendance significative à la hausse de la sinistralité climatique en France.

En effet, sur les 4 dernières années, les périls naturels atteignent en moyenne 6 milliards d'euros alors qu'ils s'élevaient seulement à 3,7 milliards d'euros en moyenne pour les années 2010 et 2,7 milliards d'euros pour les années 2000. Cette situation met en évidence la nécessité d'adapter les modèles d'assurance en permanence pour faire face à l'augmentation des risques et des coûts engendrés par les événements climatiques qui seront de plus en plus fréquents et sévères d'après les prévisions.

Impact du changement climatique à horizon 2050

Les travaux réalisés par la CCR en septembre 2023 sur les conséquences du changement climatique sur le coût des catastrophes naturelles en France à l'horizon 2050 (CCR 2023a) concluent à une hausse attendue de la sinistralité climatique comprise entre 47% et 85%, en prenant en compte l'impact du climat et l'accroissement des valeurs assurées.

Indépendamment du scénario RCP du GIEC retenu (défini en partie 3.1.1), les montants des dommages dus aux aléas climatiques augmenteront d'ici 2050, en raison de l'intensification des événements climatiques. En particulier, selon le scénario RCP 8.5, l'intensité de la sécheresse géotechnique est plus prononcée, avec une augmentation estimée des dommages de 162% en raison de l'aléa, et de 190% en intégrant l'évolution des biens assurés en 2050.

Par ailleurs, on anticipe une augmentation significative des dommages causés par les inondations, particulièrement marquée sous le scénario RCP 4.5, en raison de la hausse de la fréquence des crues rapides. La montée du niveau marin, liée à l'augmentation des températures qui pourrait atteindre +2,4°C en 2050, entraînera également une augmentation de la fréquence et de l'intensité des submersions marines. Globalement, on estime une hausse des dommages comprise entre 27% et 62% à l'horizon 2050, due uniquement au changement climatique.

En parallèle, les projections climatiques et hydro-climatiques à l'horizon 2050 réalisées par le groupe mutualiste Covéa (CovÉA 2022), portant sur quatre aléas (tempêtes, grêle, inondations et sécheresse) selon le scénario d'émissions de gaz à effet de serre RCP 8.5, indiquent une augmentation de la sinistralité de +60% dans les années à venir, corroborant les projections de la CCR.

Ces prévisions posent de nombreux défis pour le régime d'indemnisation des catastrophes naturelles, qui devra faire face à l'augmentation des dommages assurés. Elles soulignent également l'importance de l'étude de ces phénomènes afin de mieux appréhender les risques futurs.

1.1.3 Le phénomène inondation

Entre 1982 et 2022, 54% des arrêtés Cat Nat publiés correspondaient à des inondations, comme le montre la figure 1.7. Elles ont contribué à elles seules pour 24,8 milliards d'euros de dommages assurés, soit 50% du montant global de la sinistralité au titre de la garantie Catastrophes Naturelles. Le coût moyen annuel correspond à 604 millions d'euros (FRANCE ASSUREURS 2023a).

Considérant le fort impact du risque inondation sur la garantie Cat Nat, et l'intérêt particulier des assureurs pour cette garantie, il apparaît essentiel de définir le phénomène.

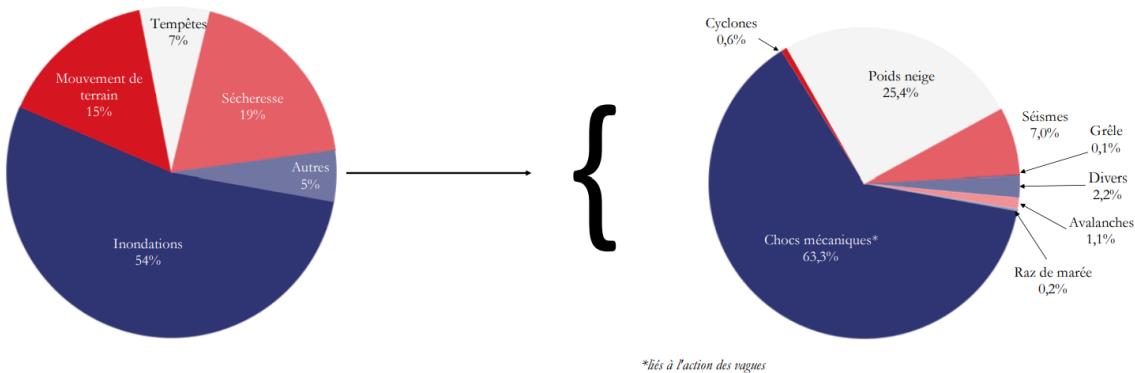


FIGURE 1.7 – Répartition des arrêtés par nature de périls entre 1982 et 2022

Source : « *L'assurance des événements naturels en 2022 – Assurance de biens et de responsabilité* » (2023)

Définition du phénomène

D'après la rubrique «Inondation » de GEORISQUES s. d.(a) , l'inondation se définit comme une submersion temporaire, rapide ou lente, d'une zone habituellement hors de l'eau. Ce phénomène trouve principalement sa source dans les précipitations soutenues et durables, ou intenses et soudaines, mais il peut également provenir de la mer ou des eaux souterraines. C'est un phénomène fortement aggravé par l'activité humaine et les aménagements.

On distingue différents types d'inondations en fonction de leur origine :

Débordement d'un cours d'eau

Les inondations par débordement de cours d'eau sont la conséquence de fortes pluies répétées ou de fonte de neige dans les cas montagneux. Le cours d'eau déborde de son lit habituel et vient occuper son lit majeur en submergeant les plaines alentours. Cette crue (ou remontée des eaux), peut être soit lente et avoir surtout un impact sur le fonctionnement des infrastructures et des réseaux au quotidien, soit "éclair" et se caractériser par une montée extrêmement rapide des eaux avec un débit très puissant et provoquer d'importants dégâts.

Le débordement des cours d'eau représente la majorité des cas d'inondations. La figure 1.8 illustre le phénomène. Ce type d'inondation occasionne des dommages parmi les plus coûteux de l'histoire du régime des catastrophes naturelles en France. Par exemple, les inondations du Gard en 2002 ont coûté plus de 830 millions d'euros, de même pour les inondations du Rhône en 2003 qui ont fait plus de 1,5 milliard d'euros de dégâts. Plus récemment, la catastrophe du Var, en juin 2010 a fait, elle, plus d'un milliard d'euros de dégâts.

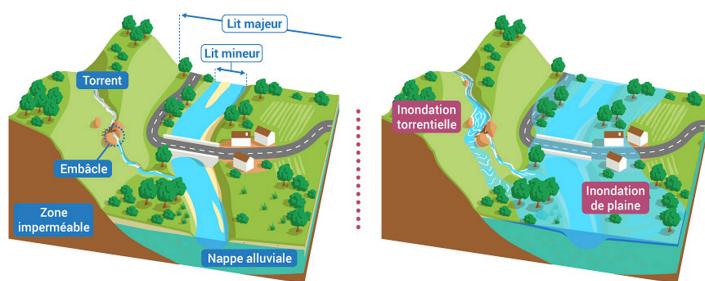


FIGURE 1.8 – L'inondation par débordement de cours d'eau

Source : « *Les inondations et les submersion marines* » – Eaufrance (2018)

Ruisseaulement

En cas d'évènements climatiques importants, par exemple une forte pluie ou un cumul important de pluie sur plusieurs jours, si les eaux ne peuvent plus s'infiltrer dans les sols, elles s'écoulent en dehors des réseaux d'évacuations des eaux pluviales et dans les zones habituellement sèches provoquant donc des inondations par ruissellement. Le ruissellement est aussi la cause d'inondation sur des territoires éloignés de tout cours d'eau. Certains facteurs augmentent le risque de survenue d'inondation par ruissellement. Parmi eux on peut noter, l'essor de l'urbanisation qui entraîne un accroissement des surfaces imperméabilisées, la topographie, l'état du sol et les caractéristiques des sous-sols.

Ce phénomène s'est produit lors des inondations meurtrières et dévastatrices d'octobre 2018 dans le département de l'Aude (METEO FRANCE 2020).

Remontée de nappe

L'inondation par remontée de nappes est due à l'élévation du niveau de la nappe phréatique au-dessus de la surface du sol. Les nappes phréatiques sont approvisionnées par l'eau de pluie, si des évènements extrêmes de pluie surviennent, alors le niveau des nappes peut atteindre une hauteur supérieure à la surface du sol et provoquer une inondation. Le déclenchement de ce type d'inondation est dû à plusieurs éléments :

- Une suite d'années à forte pluviométrie conduisant à des niveaux moyens de plus en plus élevés.
- Des variations importantes du niveau de la nappe au cours de l'année.

- Un volume d'eau conséquent dans les nappes phréatiques.
- Les caractéristiques de l'aquifère (formation géologique souterraine qui retient l'eau) dans lequel est contenu l'eau.

Ces facteurs ont notamment déclenché les inondations par remontée de nappes de Poitier en 2004 et dans l'Aude en novembre 2015.

Submersion marine

Une submersion marine désigne une inondation temporaire de la zone côtière par la mer dans des conditions météorologiques extrêmes. Elle survient généralement au cours d'une pleine mer, lorsque le coefficient de marée est élevé, que la pression atmosphérique est faible et le vent fort, ce qui implique une forte houle et un niveau de surcote élevé. Ces conditions sont souvent déclenchées par le passage d'une tempête proche du littoral. On peut noter à titre d'exemple, la tempête Xynthia en février 2010 et les tempêtes Lothar et Martin en 1999 qui sont à l'origine de submersion marine sur les côtes atlantiques majoritairement et qui ont causé de nombreux dommages pour tous les habitants du littoral.

1.1.4 Le risque inondation

A chaque phénomène d'inondation est associé un risque. Afin de modéliser et projeter une prime pure liée à ce risque, il est désormais essentiel de s'approprier les notions climatiques et techniques qui lui sont propres. Cette compréhension permettra d'identifier les paramètres clés qui interviendront dans notre modèle.

Le risque inondation correspond à la confrontation en un même lieu géographique d'un aléa (une inondation potentiellement dangereuse) avec des enjeux (humains, économiques, ou environnementaux) susceptibles de subir des dommages ou des préjudices. (GEORISQUES s. d.(a))

Qualification de l'aléa et enjeux

Afin de pouvoir comparer deux inondations différentes qui se sont produites à des endroits différents et des périodes différentes, il est important de pouvoir leur attribuer une caractérisation globale. Quatre paramètres principaux sont nécessaires pour qualifier l'aléa inondation (MINISTÈRE DE L'ÉCONOMIE ET DU DÉVELOPPEMENT DURABLE 2004) :

- La hauteur et la durée de submersion : La hauteur de submersion est une des caractéristiques qui permet de qualifier l'inondation puisqu'elle peut avoir un impact important sur les structures bâties, à travers la poussée qu'elle génère. Pour l'homme, des hauteurs d'eau supérieures à 50 cm sont considérées comme dangereuses. La durée de submersion a également un impact significatif : plus une inondation dure longtemps, plus elle contribue à propager l'humidité dans les murs et à dégrader les matériaux par gonflement. Lorsqu'elle dure plus de 24 à 48h, des problèmes sanitaires peuvent également apparaître présentant un risque pour la santé publique.
- La vitesse du courant : La vitesse du courant dépend de la pente et de la rugosité du lit du cours d'eau. Elle atteint parfois plusieurs mètres par seconde. La vitesse, associée à la hauteur, sont les principales caractéristiques qui déterminent la gravité de l'écoulement. On estime qu'à partir de 0,5m/s, il y a un risque pour l'homme d'être emporté par le courant ou blessé par des objets transportés par le courant.
- Le volume de matière transportée : Les matériaux de type argile, sable, gravier, galet etc, qui se trouvent dans les cours d'eau, appelés « transport solide » peuvent également faire des dégâts. Ils sont transportés par suspension dans l'eau ou par la force du courant sur le fond du lit.
- La période de retour : Il est nécessaire de caractériser une inondation par son importance. Les débits ou les hauteurs d'eau n'ont une signification que sur un point donné d'un cours d'eau. Il est donc impossible de comparer deux événements différents, s'ils se sont produits à deux endroits distants sur le cours d'eau et encore moins pour deux cours d'eau différents. Les événements sont souvent représentés par leur probabilité d'apparition. C'est la valeur du temps de retour qui est le plus souvent utilisée. Elle correspond à l'inverse de la probabilité d'occurrence du phénomène. Les événements les plus souvent représentés sur les cartes d'aléa sont les crues décennales et centennales. (RÉGION CENTRE VAL DE LOIRE 2016)

Chacun des paramètres sera plus ou moins important en fonction du type d'inondation.

Le risque d'inondation concerne aujourd'hui plus de 18 millions d'habitants en France soit plus de 10 millions de logements. Cela signifie que plus de 27% de la population française réside en zone inondable

(MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE ET DE LA COHÉSION DES TERRITOIRES 2023). Ce risque menace également l'environnement puisque les phénomènes d'érosion, de charriage, de suspension de matériaux et d'alluvionnement peuvent altérer le milieu naturel. En ce qui concerne les zones industrielles, il existe également un risque de pollution et d'accident technologique.

Vulnérabilité des bâtiment face à l'inondation

L'objectif étant de calculer une prime pure pour les bâtiments du portefeuille de notre assureur (limité au périmètre "maison" comme mentionné en section 1.1.1), il est également essentiel d'examiner la vulnérabilité des bâtiments et leur comportement face aux inondations. Cela permettra de déterminer les variables spécifiques au bâtiment qui influenceront notre modélisation.

D'après le CENTRE EUROPÉEN DE PRÉVENTION DES RISQUES D'INONDATION (CEPRI) 2010, il existe trois dimensions de la vulnérabilité d'un bâtiment :

- **La vulnérabilité liée au risque des personnes.** Cela concerne tous les éléments qui sont susceptibles de porter une atteinte physique aux personnes se trouvant dans le bâtiment ou devant intervenir pendant ou après l'inondation. Cela inclut la capacité du bâtiment à résister aux forces exercées par la montée des eaux, la présence d'une zone hors d'eau pour abriter les occupants, et l'accès des secours à cette zone. Il faut également considérer les risques liés aux équipements techniques, tels que l'électrocution, les explosions de gaz, et l'intoxication par des substances dangereuses. Enfin, les modifications de l'environnement, comme l'effondrement de la voirie ou la présence de zones dangereuses masquées par l'eau, doivent être prises en compte. A titre d'exemple, la présence de plancher à l'intérieur du bâtiment le rend plus vulnérable par rapport à la sécurité des personnes, il peut se déformer, gonfler voire se rompre au contact de l'eau. De même, si les parties électriques des pompes, des brûleurs, des tableaux de commande, etc, sont dans la partie basse de la maison, il y a un risque d'atteinte par l'eau qui peut avoir une incidence sur la sécurité des personnes dans la mesure où les appareils de protection, les disjoncteurs, sont eux mêmes endommagés.
- **La vulnérabilité liée au retour à la normale** se mesure par le temps nécessaire entre une inondation et la reprise effective des activités dans un bâtiment, qu'elles soient publiques ou privées. Ce délai inclut le nettoyage, le séchage des locaux, du mobilier, et des équipements récupérables, ainsi que les travaux de réparation si besoin. Le retour à la normale dépendra donc de l'ampleur des dégâts, du temps nécessaire pour remettre les biens en état, et de la restauration des services essentiels. Les principaux risques concernent les cloisons, doublages et planchers, dont la réparation nécessite souvent le remplacement complet des éléments endommagés par l'eau, ce qui peut être très coûteux. Il en va de même pour les installations électriques et les ouvrages annexes, où le nettoyage et le raccordement peuvent s'avérer longs et onéreux.
- **La vulnérabilité liée aux effets domino.** En effet, les dommages d'un bâtiment du à l'inondation peuvent entraîner une succession de dégâts et avoir un impact important sur des sites alentours, qui sont à prendre en compte pour le propriétaire. Par exemple, le renversement d'une cuve peut provoquer une pollution de l'environnement, ou une citerne mal arrimée peut percer les bâtiments aux alentours.

Certaines stratégies peuvent cependant être adoptées afin de réduire la vulnérabilité du bâtiment, c'est à dire minimiser les risques pour les occupants, les délais de retour à la normale, et les effets en chaîne. Il est possible, soit de résister à l'eau en scellant toutes les voies d'entrée possibles, comme les ouvertures, le sous-sol, et le réseau d'évacuation des eaux usées. Cette approche aide à maintenir le bâtiment relativement sec et à réduire le temps nécessaire pour le remettre en état, bien qu'elle requière une mise en œuvre prudente. On peut également accepter que l'eau pénètre dans le bâtiment et prendre des mesures pour limiter les dégâts et accélérer le retour à la normale, telles que la surélevation des équipements et l'utilisation de matériaux résistants. Cette stratégie est parfois la seule option lorsque la prévention est impossible.

Ces éléments permettront ainsi d'identifier les bâtiments assurés les plus vulnérables à l'inondation, ou au contraire, les moins susceptibles d'être endommagés. Toutefois, face à des événements extrêmes aggravés par le dérèglement climatique, ces mesures peuvent s'avérer insuffisantes, rendant impossible toute garantie de protection absolue.

Impact du changement climatique sur le phénomène inondation

De nombreuses études réalisées ces dernières années (CCR 2023a ou COVÉA 2022 par exemple) montrent que le changement climatique est à l'origine de phénomènes de précipitations extrêmes et de modifications saisonnières en lien avec des variations d'enneigement. Par exemple, l'augmentation des précipitations intenses dans le climat actuel s'explique par la capacité accrue d'une atmosphère plus chaude à transporter de la vapeur d'eau. Pour chaque degré de réchauffement, les couches basses de l'atmosphère peuvent transporter 7% de vapeur d'eau supplémentaire, ce qui favorise les inondations par ruissellement. Le changement climatique impacte également les inondations par débordement non seulement en modifiant les régimes de précipitations, mais aussi en influençant les régimes d'enneigement, les cumuls de précipitations saisonnières et l'augmentation de l'évapotranspiration (l'évapotranspiration correspond à la quantité d'eau qui depuis le sol passe dans l'air à l'état de vapeur en raison de l'effet conjoint de la transpiration des plantes et de l'évaporation au niveau du sol). Ainsi, ces variations dues au changement climatique, combinées à l'action humaine influencent de manière significative le risque d'inondation et aggravent les dommages.

A l'échelle de l'Europe, les projections indiquent que les tendances observées concernant les pluies intenses vont se poursuivre. En effet, COVEA estime que les précipitations qui surviennent aujourd'hui tous les 10 ans, se produiront tous les 4,5 ans en 2050, de même, celles qui surviennent tous les 20 ans aujourd'hui, se produiront tous les 7 ans et les précipitations qui se produisent aujourd'hui tous les 50 ans, se produiront tous les 11 ans. Cela signifie que les inondations par ruissellement vont continuer à s'aggraver et occasionneront des difficultés d'adaptation pour les communes qui n'auront pas les moyens de redimensionner leurs systèmes d'évacuation des eaux.

En revanche, il est difficile d'identifier une tendance uniforme pour l'évolution future des débits des grands cours d'eau en France métropolitaine. Par exemple, on prévoit une augmentation des inondations par débordement dans la région des Cévennes et le Nord-Est de la France, tandis qu'une diminution de ce phénomène est anticipée dans les Alpes, les Pyrénées et le Jura.

La CCR en déduit une hausse du montant annuel moyen des dommages due à l'aléa comprise entre 6% et 19% selon les scénarios du GIEC. L'augmentation des biens assurés pourrait entraîner une hausse supplémentaire de 20% des dommages.

En ce qui concerne les submersions marines, le phénomène le plus préoccupant en France métropolitaine est l'élévation du niveau marin et la variation climatique en est une des principales causes. Les niveaux marins extrêmes semblent augmenter de manières cohérentes avec l'augmentation du niveau marin moyen. Bien que la modélisation de ce phénomène soit complexe, on estime une hausse significative des dommages annuels moyens, entre 75% et 91% d'après la CCR. Ces montants de dommages pourraient également être intensifiés davantage avec l'augmentation des biens assurés sur les littoraux.

Au regard de ces prévisions climatiques, l'objectif de l'étude est de s'interroger sur l'impact de ces changements sur l'évolution des primes pures d'assurance pour le risque inondation.

1.2 Description des éléments

Cette partie présente les différentes bases de données utilisées pour la modélisation d'une prime pure concernant le péril inondation. Ainsi la base de sinistralité, la base de complément à la maille adresse, provenant d'un prestataire externe, et la base de données météorologiques y seront décrites.

1.2.1 Base de données assureur

Pour établir une prime pure pour le risque d'inondation, nous disposons d'une base de données de sinistralité qui sera utilisée pour entraîner les modèles que nous allons développer. Cette base de données, fournie par un assureur, contient des informations sur plusieurs millions de contrats. Elle regroupe un vaste ensemble de données relatives à l'assuré, à sa couverture, à son domicile, ainsi qu'à son historique de sinistralité.

En ce qui concerne l'assuré, la base de données inclut des informations telles que son âge et sa catégorie socioprofessionnelle. Les données sur le domicile de l'assuré comprennent des éléments comme l'adresse, le nombre de pièces et d'étages, ainsi que la présence d'éléments spécifiques tels qu'une piscine, une cheminée, ou

une dépendance, dont la surface est également précisée. Il est également indiqué si le bâtiment constitue une résidence principale ou secondaire.

Pour ce qui est du contrat d'assurance lui-même, la base de données recense des informations comme les dates de début et de fin de couverture, le niveau de franchise choisi, et la formule souscrite. En cas de sinistre, nous possédons des informations sur la date de survenance, le coûts du sinistre et son type, ce qui va permettre de segmenter la base pour ne s'intéresser qu'au risque inondation.

Ainsi, à partir de ces données initiales, nous créerons deux bases :

- Une base "contrat", qui sera une copie de la base initiale.
- Une base "sinistre", qui ne contiendra que les informations sur les contrats ayant subi au moins un sinistre d'inondation. (Il ne sera pas possible de distinguer le type d'inondation)

Par la suite, ces bases subiront plusieurs transformations afin d'obtenir les bases de Fréquence et de Coût moyen utilisées en tarification non-vie.

Comme évoqué dans la partie 1.1.1, la sinistralité du risque inondation sera étudiée uniquement sur le périmètre "maison". Cela représente 64% des sinistres inondation du portefeuille assureur.

La collaboration de Addactis avec son partenaire externe namR va permettre de géolocaliser chaque contrat et de lui rattacher des données complémentaires à l'adresse.

1.2.2 Base de données complémentaire au bâtiment

La prise en compte des spécificités du logement de l'assuré dans l'évaluation de la sinistralité est un atout important pour l'assureur. Cela permet d'être plus précis et de pouvoir comparer deux bâtiments dans une même zone. A titre d'exemple, il est plus probable qu'un assuré possédant un logement de plein pied (sans zone surélevée) soit victime d'un sinistre inondation que son riverain qui lui, possède dans son logement une zone de refuge en hauteur.

L'ajout de données complémentaires à la maille adresse a été rendue possible grâce à la collaboration entre les entreprises Addactis France et NamR, premier hub français de données géolocalisées à l'adresse.

Construction des données par NamR

Les données doivent d'abord être géolocalisées. Le géocodage est le processus par lequel une description d'emplacement, comme une adresse postale, est convertie en coordonnées géographiques. Ainsi, il est nécessaire de relier les adresses postales présentes dans la base de données "assureur" à leurs coordonnées géographiques correspondantes.

Pour ce faire, le référentiel d'adresses BAN (Base Adresse Nationale) est utilisé. Cette base est l'une des neuf bases de référence du service public et contient la correspondance entre les adresses non nominatives et leurs positions géographiques pour plus de 25 millions d'adresses à travers la France. La base de l'Institut National de l'Information Géographique et Forestière (IGN), optimisée par namR, est également sollicitée pour compléter le géocodage.

La correspondance entre les bâtiments et les adresses est établie grâce à un système de "string matching". Cette méthode permet de trouver des chaînes de caractères qui correspondent approximativement à un motif, plutôt qu'à une correspondance exacte.

Présentation des variables complémentaires

Ainsi, de nombreuses informations complémentaires ont été ajoutées à la base initiale "assureur". Au total, plus de 600 variables ont été intégrées, enrichissant les données sur la sinistralité et les renseignements concernant les assurés. Parmi ces variables, on trouve des attributs liés directement au bâtiment mais également des attributs liés aux alentours du bâtiment, ainsi que leur niveau de confiance.

Pour le bâtiment, les nouvelles variables englobent des données sur les surfaces et caractéristiques du bâtiment, comme la surface au sol, la surface habitable, le nombre de pièces, et la valeur au mètre carré, ainsi que des informations sur les murs et la toiture, telles que la période de construction, la surface vitrée, le type de

toit, et la présence de panneaux solaires. Des éléments concernant le terrain et les bâtiments annexes, tels que la surface de végétation, la présence d'une piscine, et la surface des annexes, viennent également enrichir la base de données.

Quant aux alentours du bâtiment, les variables comprennent des mesures de distance et d'altitude, telles que la distance au cours d'eau le plus proche, la distance à la mer, la distance à une caserne de pompiers, etc. Elles incluent aussi le nombre de bâtiments dans un rayon de 100 mètres et la différence d'altitude avec le point d'eau le plus proche par exemple. Des données complémentaires sur la démographie locale, comme le nombre d'écoles dans le quartier, la densité de population, et le nombre de campings complètent également la base.

Ces variables peuvent être regroupées en différentes échelles : celles correspondant à la maille « Ilôts Regroupés pour l'Information Statistique » (IRIS), celles au niveau de la maille INSEE, ainsi que celles au niveau de l'adresse (comprenant des variables géographiques, spécifiques aux bâtiments).

Ces données fournis par namR proviennent de différentes sources dont par exemple l'INSEE (bases « SIRENE », « logements ordinaires », etc.), le Ministère de la Transition Écologique et Solidaire (MTES), l'Agence De l'Environnement et de la Maîtrise de l'Énergie (ADEME), le répertoire des logements locatifs des bailleurs sociaux (RPLS), la Direction Générale des Finances Publiques (DGFIP), le Bureau de Recherches Géologiques et Minières (BRGM), entre autres.

En intégrant ces variables, il est possible de créer un modèle plus précis et complet. Certains attributs pourront contribuer à une meilleure compréhension de la vulnérabilité du bâtiment et de son environnement, permettant ainsi une évaluation plus fine du risque.

1.2.3 Base de données météorologique et expertises climatiques

Compte tenu de l'évolution de la sinistralité climatique en France, l'intégration de l'historique météorologique dans la tarification des contrats de Multi-Risques Habitation (MRH) représente un avantage considérable. Elle permet d'améliorer les modèles d'évaluation des risques et de prendre en compte des données spécifiques à chaque bâtiment.

Dans l'objectif de caractériser le péril inondation, nous utiliserons une base de données météorologiques de cumuls de précipitations.

Les données météorologiques

Les données météorologiques les plus pertinentes pour cette étude sont les précipitations totales sur le même périmètre que la base de sinistralité, la France métropolitaine hors Corse. Pour cela, nous utilisons des données en libre accès.

Diverses sources de données en open data ont été étudiées afin de s'orienter vers la plus adaptée à notre sujet. En prenant en compte la finesse des mailles géographiques et temporelles ainsi que la disponibilité des historiques, nous décidons d'utiliser les données de réanalyse ERA5-Land pour notre modèle. Ces données sont construites à partir des données de réanalyses atmosphériques ERA5. Il s'agit de données produites par le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT), qui est une organisation intergouvernementale, fondée en 1975, entre 22 pays européens (France, Allemagne, Espagne, Norvège, Suisse, etc.) et 12 pays associés ou en voie de devenir membre (Bulgarie, République Tchèque, Maroc, Hongrie, etc.) réunis pour mettre en commun leurs ressources et bénéficier collectivement d'une meilleure prévision numérique des conditions météorologiques. (CEPMMT 2021).

Ces dernières années, le CEPMMT a également joué un rôle clé dans le programme *Copernicus* de l'Union européenne, offrant des informations à l'échelle mondiale sur le changement climatique, la composition atmosphérique, les inondations et les risques d'incendie à ses utilisateurs. Le produit de réanalyse ERA5 a été récemment conçu dans le cadre de ce programme et devient l'outil de surveillance du climat mondial le plus performant de CEPMMT. Son objectif est de suivre l'évolution des perturbations climatiques en fournissant des informations fiables sur l'état passé et présent du climat. C'est via le site de Copernicus, qui offre un accès gratuit et illimité aux observations météorologiques historiques, que nous téléchargeons les données relatives aux précipitations. La figure 1.9 donne un aperçu du portail d'accès au données de ce site, celles-ci sont téléchargées au format NETCDF.

FIGURE 1.9 – Portail d'accès aux données de ERA5-Land

Source : *Copernicus (2024)*

Les données de réanalyse de ERA5 proviennent de la combinaison entre des observations passées réelles issues de stations météorologiques, des données satellitaires et des modèles climatiques (partie 3.1.1) créés à partir des lois physiques, pour avoir une vision complète du climat actuel et passé. Elles sont disponibles à des résolutions horaires ou mensuelles et couvrent l'ensemble de la planète avec une grille régulière de 0,25 degré en latitude et en longitude (un point tous les 31 kilomètres environ) (Figure 1.10a). Ces données, dont l'historique remonte à 1950, sont mises à jour quotidiennement avec un décalage de cinq jours par rapport au temps réel.

Comme mentionné précédemment, notre étude nécessite exclusivement l'utilisation de données terrestres. C'est pourquoi nous avons opté pour les données ERA5-Land, plus adaptées que le produit ERA5. En effet, ERA5-Land se concentre uniquement sur la surface terrestre, avec une résolution de 9 km ($0,1^\circ \times 0,1^\circ$) (Figure 1.10b), offrant ainsi une meilleure granularité des données. De la même façon qu'ERA5, ERA5-Land fournit des variables horaires (CLIMATE CHANGE SERVICE 2024).

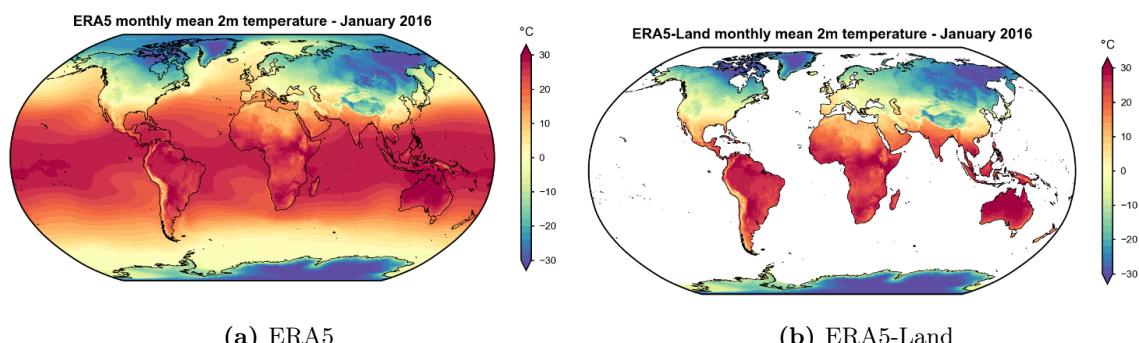


FIGURE 1.10 – Comparaison ERA5 et ERA5-Land

Source : *Copernicus (2024)*

Récupération des cumuls de précipitations

Afin de caractériser l'inondation, nous avons extrait la variable "*Total precipitation*" des données ERA5-Land, couvrant la période de 1950 à aujourd'hui. Ce sont en effet, les précipitations qui sont à l'origine de la plupart des inondations (BROUSSE 2021). Cette variable mesure la profondeur (en m) de l'accumulation

d'eau, tant liquide que solide, incluant la pluie et la neige, qui atteint la surface terrestre. Le brouillard, la rosée, et les précipitations qui s'évaporent dans l'atmosphère avant d'atterrir à la surface de la Terre ne sont pas pris en compte. Cependant, ces éléments n'interviennent pas dans l'étude des inondations. Cette variable est disponible à l'échelle horaire, mais en téléchargeant les données pour la dernière heure de la journée, nous obtenons le cumul des précipitations depuis le début de la journée jusqu'à cette dernière heure, ce qui correspond à un cumul journalier.

La figure 1.11 représente les données de précipitation recueillies pour la journée du 30 mai 2016. Ce jour-là, de fortes pluies se sont abattues en France notamment dans le Sud de l'Ile-de-France ainsi que dans l'Yonne et le Loiret, provoquant des inondations qui ont fait l'objet de plusieurs arrêtés catastrophes naturelles.

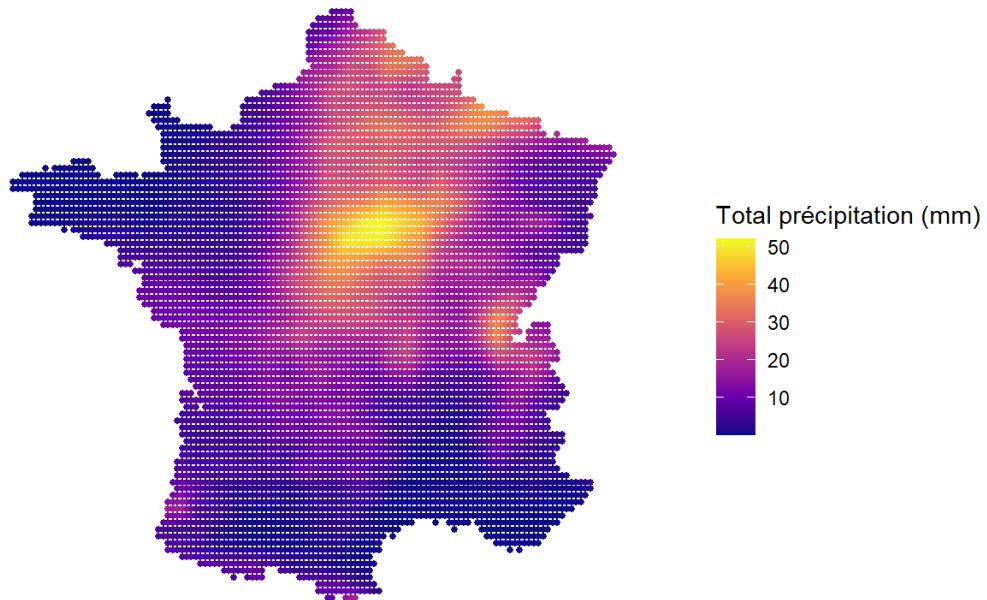


FIGURE 1.11 – Données de précipitation pour la journée du 30 mai 2016 (ERA5-Land Hourly data)

Source : *Climate Change Service (2024)*

Complexité du risque inondation

Le lien entre les inondations et la variable météorologique "*cumul de précipitations*" est beaucoup plus complexe que pour d'autres périls, comme les tempêtes par exemple, qui n'impliquent pas de processus intermédiaires. En effet, pour les tempêtes, la relation entre la force des rafales de vent et les dommages aux bâtiments est directe : si des rafales de vent de 150 km/h sont observées à un moment donné, ce sont ces rafales qui viendront toucher le bâtiment et affecter directement les structures.

En revanche, pour les inondations, le processus est plus compliqué. L'effet des conditions météorologiques est modifié par des facteurs intermédiaires tels que les processus hydrologiques et hydrauliques, ainsi que par l'état du sol et la végétation. Les inondations ne sont pas seulement causées par la pluie elle-même, mais aussi par la façon dont l'eau se comporte après être tombée. Si elle ne s'infiltre pas dans le sol, elle peut s'écouler sur la surface terrestre ou faire déborder des cours d'eau, provoquant des inondations. Ce processus de déplacement de l'eau est une conséquence indirecte des conditions météorologiques, comme les fortes pluies ou la fonte des neiges, plutôt qu'une conséquence directe de la météo elle-même. Il faut également prendre en compte la notion de bassin versant. Certaines zones peuvent subir des inondations sans être directement affectées par les conditions météorologiques locales, comme l'exemple des inondations de la Seine à Paris en 2016, qui étaient causées par des pluies intenses survenues en amont du bassin versant parisien. De plus, il est important de considérer la temporalité : une faible pluviométrie persistante sur plusieurs jours peut provoquer des inondations, alors que de fortes précipitations en peu de temps peuvent ne pas suffire à elles seules.

Il semble donc important, de construire un indicateur à partir des données météorologiques prenant en compte cette spécificité du péril inondation.

Construction d'un indice de précipitation cohérent

Nous détaillons ici, la méthode de construction d'un indice de précipitation pertinent pour la modélisation du risque inondation.

Il est à noter que ce mémoire est réalisé en parallèle de travaux de R&D au sein de la practice tarification non-vie d'Addactis. La méthodologie de construction de cet indice est donc fortement inspirée d'un article rédigé par M.Moriah qui sera publié par la suite.

Nous disposons dans la base de sinistre, pour chaque bâtiment sinistré, de ses coordonnées géographiques, ainsi que de la date de survenance de l'inondation.

La première étape pour construire l'indice, consiste à associer à chaque bâtiment de la base de sinistres, les quatre points géographiques de la grille ERA5-Land les plus proches. Pour ces quatre points, on récupère les cumuls de précipitations correspondant à différentes périodes avant l'inondation : 1 jour, 2 jours, 3 jours, 5 jours, 7 jours, 10 jours, 15 jours et 30 jours. Ces différentes périodes permettent d'étudier les effets de la pluie sur différents horizons temporels, afin de saisir à la fois les épisodes pluvieux intenses de courte durée et les cumuls plus modérés mais prolongés qui peuvent également causer des inondations. Pour chaque période, on sélectionne le maximum des précipitations observées parmi ces quatre points géographiques. Cette approche permet de capter le plus fort épisode de pluie survenu à proximité du bâtiment sinistré. Ce maximum sera retenu pour caractériser la période considérée, assurant ainsi que l'indice reflète fidèlement l'intensité des précipitations à l'échelle locale.

Finalement, comme l'illustre le tableau 1.1, nous attribuons à chaque bâtiment sinistré un cumul de précipitation pour les n jours précédents l'inondation.

id bâtiment	date sinistre	precip_1j	precip_2j	precip_3j	precip_5j	precip_7j	precip_10j	precip_15j	precip_30j
1	02-06	7.7919	25.6674	33.9298	45.5124	52.6754	68.3111	140.2076	128.1979
2	04-01	11.7589	17.7647	26.4047	52.0243	65.1834	107.5563	143.9834	186.9007
3	16-05	2.5661	21.9884	31.7807	42.9677	60.1544	74.4079	117.2245	180.2469

TABLE 1.1 – Exemple : attribution des cumuls de précipitation n avant l'inondation pour chaque bâtiment

Pour chacune des différentes fenêtres de temps, les distributions historiques des précipitations (depuis 1985 jusqu'à aujourd'hui) au point géographique considéré sont alors tracées, afin de déterminer à quel quantile de la distribution correspond la valeur maximale sélectionnée précédemment.

En prenant l'exemple du **bâtiment 1** du tableau 1.1. Nous cherchons à savoir à quel quantile correspond la valeur **33.9229** qui représente le cumul de précipitation sur les 3 jours avant l'inondation. Pour cela, la distribution historique des cumuls de précipitations au point sélectionné sur 3 jours est tracée(figure 1.12). Ensuite, on regarde à quel quantile cela correspond, ici le quantile 99.023.

Huit nouvelles colonnes représentant les quantiles sont alors ajoutées à la base sinistre à l'image du tableau 1.2.

id bâtiment	date sinistre	...	precip_3j	quantile_precip_3j	...	precip_10j	quantile_precip_10j	...	precip_30j	quantile_precip_30j	var_precip	jours
1	02-06	...	33.929805	99.023163	...	68.31104	98.393723	...	128.197945	96.422090	99.023163	3j
2	04-01	...	26.404721	97.032487	...	107.556337	99.885266	...	186.900727	99.729620	99.885266	10j
3	16-05	...	31.780711	97.694812	...	74.407934	97.057626	...	180.246935	97.833253	97.833253	30j

TABLE 1.2 – Exemple : ajout des quantiles n avant l'inondation pour chaque bâtiment

Enfin, le **maximum** de ces différents quantiles est sélectionné, et cette valeur devient la variable **cumul_precip**, représentant le cumul des précipitations, en cohérence avec les spécificités du risque d'inondation.

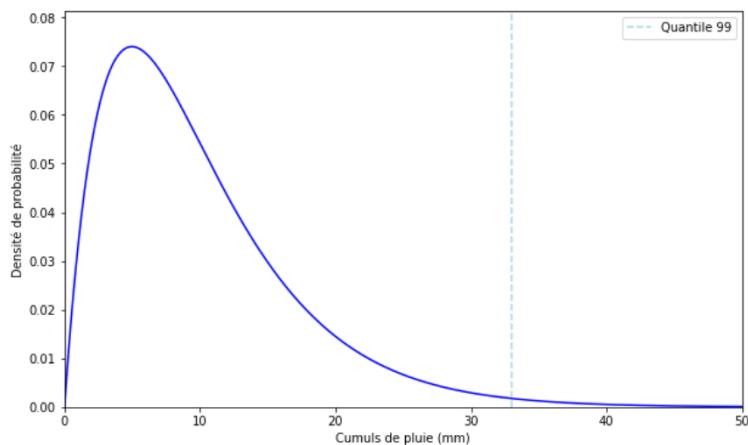


FIGURE 1.12 – Distribution des cumuls de précipitations historiques sur 3 jours

Ajout d'une expertise climatique

En complément de cette variable de précipitation, la modélisation inclura également d'autres variables météorologiques. Parmi ces variables supplémentaires figurent, par exemple, les maximums de précipitations journalières ou encore le nombre de jours consécutifs de précipitations sur une année.

Ces variables météorologiques complémentaires sont issues d'une base de données interne à Addactis, qui compile une variété d'informations climatologiques spécifiques. La description détaillée de ces variables sera abordée dans la partie 2.2.

La modélisation sera également renforcée par l'intégration de données provenant de sources open data liées aux expertises climatiques. Parmi ces données figurent notamment des informations sur les Territoires à Risque d'Inondation (TRI) et les Plans de Prévention des Risques Naturels (PPRN). Ces éléments fournissent des informations sur les risques naturels associés à chaque zone géographique, notamment la fréquence et la gravité des inondations passées. Ces bases de données open data sont déjà intégrées dans des bases complémentaires, accessibles à l'adresse fournie par namR.

Nous prendrons également en compte les zonages climatiques et hydrographiques de la France. Ces zones divisent le territoire en fonction de ses particularités climatiques (par exemple, zones tempérées, méditerranéennes...) et hydrographiques (bassins versants, régions sujettes aux crues). Cette approche régionale assure que les risques d'inondation soient évalués en tenant compte des spécificités locales.

Grâce à cette combinaison de données météorologiques, d'expertises climatiques et de zonages géographiques adaptés, la modélisation pourra mieux capturer la complexité du phénomène d'inondation. Nous pourrons donc obtenir une prime pure au bâtiment qui tient compte à la fois des particularités locales et des tendances climatiques historiques.

Chapitre 2

Mise en place des modèles

2.1 Méthodologie

Dans cette partie, sont présentées les différentes méthodes et outils utilisés qui permettront d'obtenir une prime pure pour chaque bâtiment assuré.

2.1.1 La tarification en assurance non-vie

La tarification vise à déterminer les primes commerciales complètes, permettant à l'assureur de faire face aux prestations garanties ainsi qu'aux coûts de gestion et de commercialisation. Cette prime commerciale complète est constituée de la prime pure, à laquelle s'ajoutent des chargements de gestion pour financer les coûts d'acquisition et d'administration supportés par l'assureur, ainsi que des chargements de sécurité pour couvrir le risque de mauvaise tarification.

La prime pure se rapporte uniquement à la tarification du risque, elle représente l'espérance des sinistres. Si on note la **charge totale** des sinistres S telle que :

$$S = \begin{cases} 0 & \text{si } N = 0 \\ \sum_{i=1}^N Y_i & \text{si } N \neq 0 \end{cases}$$

Où N est la variable aléatoire qui représente le nombre de sinistres survenus sur une période de temps définie, et Y_i la variable aléatoire qui représente le coût du $i^{\text{ème}}$ sinistre durant la période d'exposition au risque.

Alors la **prime pure** peut être donnée par la formule :

$$\text{prime pure} = \mathbb{E}[S] = \mathbb{E}[Y] \times \mathbb{E}[N]$$

Ainsi, la modélisation de cette prime pour un risque donné, revient à modéliser séparément le nombre moyen de sinistres et leur coût moyen. Sans cette décomposition, la modélisation directe de la prime pure est compliquée.

Pour cela, la première approche est de séparer les différents contrats et individus de la base de données en plusieurs classes ou catégories afin que les risques soient considérés homogènes au sein de chaque groupe. En effet, en l'absence de segmentation tarifaire, l'assureur est exposé au phénomène d'antisélection. Si l'assureur ne propose qu'une tarification unique, les « bons risques » partiront vers des concurrents offrant des tarifs plus attractifs en fonction de leur profil. Les « mauvais risques » resteront dans le portefeuille, car leurs tarifs seraient moins avantageux ailleurs. Pour contrer cette antisélection, les assureurs ont tout intérêt à segmenter leur portefeuille en fonction des profils de leurs assurés.

Dans le cas où les variables Y_i sont indépendantes, identiquement distribuées et indépendantes de N , on peut donc réécrire la relation précédentes comme :

$$\mathbb{E}[S|X] = \mathbb{E}[Y|X] \times \mathbb{E}[N|X]$$

Les variables $X = (X_1, \dots, X_p)$ qui permettent la segmentation, sont appelées **Variables de tarification**. Elles peuvent être quantitatives ou qualitatives, en général à modalités finies car les variables continues sont regroupées en classes. Elles représentent les caractéristiques des individus (age, contrat, localisation,...).

En ajoutant des informations complémentaires en plus des variables de tarification classiques, il sera possible de mieux refléter les impacts potentiels du changement climatique sur les sinistres. L'objectif de la suite de ce mémoire est donc d'intégrer ces éléments clés pour l'inondation tout en gardant une structure cohérente.

2.1.2 Calcul de la prime pure climatique

La méthodologie utilisée pour calculer la prime pure climatique suit donc une approche similaire à celle évoquée précédemment, tout en y ajoutant un conditionnement spécifique basé sur l'intensité des événements climatiques.

La prime pure climatique pour le risque inondation dans notre étude, est définie par l'équation suivante :

$$PP_{\text{climatique}} = \sum_i P(\text{intensité}_i) \times \mathbb{E}(\text{fréquence}_{\text{sinistres}}) \times \mathbb{E}(\text{coût}_{\text{sinistres}} \mid \text{inondation}_{\text{intensité}_i}) \quad (2.1)$$

Cette formule découle du principe des probabilités totales. Elle consiste à décomposer le coût moyen des sinistres en fonction de différents scénarios d'intensité, chaque scénario étant caractérisé par une intensité spécifique de précipitation. Nous utilisons donc une modélisation du coût moyen en cas d'inondation complétée par une probabilité d'observer une inondation d'**intensité i** . Cela permet d'intégrer la variabilité climatique et ses effets potentiels dans le calcul des risques.

Dans cette formule :

- $P(\text{intensité}_i)$, représente la probabilité qu'une inondation d'intensité i se produise. La méthode de construction de cette probabilité sera expliquée plus en détail par la suite.
- $\mathbb{E}(\text{fréquence}_{\text{sinistres}})$, est la fréquence attendue des sinistres, c'est-à-dire le nombre moyen d'événements de sinistres sur une période donnée.
- $\mathbb{E}(\text{coût}_{\text{sinistres}} \mid \text{inondation}_{\text{intensité}_i})$, est le coût moyen attendu des sinistres, sous la condition qu'une inondation d'intensité i ait lieu.

Le modèle du coût des sinistres est conditionné à la survenance d'une inondation d'une intensité i . En d'autres termes, il estime le coût moyen des dommages subis par les bâtiments en France lors d'une inondation avec une intensité donnée.

Les différents coûts sont ensuite pondérés en fonction de leur probabilité d'intensité associée, qui est modélisée à partir de la variable *cumul_precip*. Cette variable, décrite plus en détail dans la partie 1.2.3, représente la quantité totale de précipitations observée sur une période de temps spécifique et permet de quantifier l'intensité des inondations.

Contrairement au coût moyen, la fréquence des sinistres n'est pas dépendante de l'intensité des inondations dans la formule utilisée. En effet, la base de contrat qui servira lors de la modélisation de la fréquence, ne contient pas de date précise d'inondation mais uniquement une indication d'impact durant la période de validité du contrat. Il était donc impossible d'associer une intensité de précipitation spécifique à un contrat couvrant une période donnée.

Afin de contourner cette contrainte, l'approche la plus pertinente a été d'intégrer à la modélisation de la fréquence d'autres indicateurs météorologiques disponibles aussi bien en version historique qu'en version projetée dans une base de données interne à Addactis. Ces indicateurs supplémentaires seront détaillés par la suite. Cet ajout permettra d'une part d'apporter des informations supplémentaires mais également de projeter la fréquence en utilisant la valeur projetée de ces indicateurs pour mettre à jour le modèle.

Calcul de la probabilité d'intensité

Les probabilités de survenance des différentes intensités d'inondation sont calculées de manière empirique à partir de la variable *cumul_precip*. On rappelle que cette variable, dont la méthode de détermination est expliquée en détail dans la section 1.2.3, représente, pour chaque bâtiment ayant subi un sinistre, le quantile maximal correspondant aux précipitations observées à la date du sinistre, calculé à partir de la distribution historique de ces précipitations.

Afin d'utiliser cette variable dans la modélisation, nous procérons à une discréétisation en sept classes d'intensité distinctes. Dans l'équation 2.1, l'indice i correspond à l'une de ces sept classes, avec i allant de 1 à 7. La discréétisation est effectuée en fonction des quantiles associés à la variable *cumul_precip*, permettant de définir des intervalles d'intensité pour chaque classe.

Le tableau 2.1 récapitule cette discréétisation, indiquant pour chaque classe d'intensité l'intervalle de quantiles correspondant. À chaque intervalle de quantiles est ensuite associée une probabilité de survenance. Par exemple, une inondation d'intensité 1 correspond à un cumul de précipitations se situant entre le quantile 0 et 0.75 des distributions historiques. Cela signifie que l'inondation a lieu avec une intensité de précipitation qui se produit dans 75% du temps. De même, une inondation d'intensité 2, correspondant à un cumul de précipitations compris entre le quantile 0.75 et 0.85, et donc une intensité de précipitations qui se produit dans 10% des cas.

Cette approche permet d'associer une probabilité à chaque tranche d'intensité, facilitant ainsi l'intégration des précipitations dans le calcul de la prime pure climatique.

intensité	<i>cumul_precip</i> discréétisée	Probabilité associée
1	[0,0.75[0.75
2	[0.75,0.85[0.10
3	[0.85,0.95[0.10
4	[0.95,0.97[0.02
5	[0.97,0.98[0.01
6	[0.98,0.999[0.019
7	[0.999,1[0.001

TABLE 2.1 – Description des différentes intensités

2.1.3 Modèle linéaire généralisé

Enfin, l'estimation des modèles sous-jacents (fréquence et coût moyen) se fait par le biais de modèles statistiques. Les modèles de référence utilisés dans la majorité des travaux d'évaluation de risque en tarification non-vie sont les modèles de régression qui permettent de quantifier les relations entre la variables dite *réponse* Y et les variables explicatives $X = (X_1, \dots, X_p)$. Ces modèles ont gagné en sophistication depuis plusieurs années. Dans les années 1980, les modèles linéaires généralisés (MLG), dont on détaillera le principe par la suite, ont remplacés les modèles de régression linéaire simples. Ces GLM ont amélioré la précision des prédictions de risque et sont désormais largement utilisés par les compagnies d'assurance.

Modèle de régression linéaire

Dans un premier temps, nous rappelons les bases du modèle linéaire (*Linear Model (LM)*) et ses conditions d'application. Ce modèle est défini tel que :

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{j,i} + \epsilon_i, \quad i \in \{1, \dots, n\} \quad (1.1)$$

Il s'appuie sur les hypothèses suivantes :

1. Les erreurs ϵ_i sont indépendantes et identiquement distribuées, de loi normale centrée et de variance constante, $\mathbb{E}[\epsilon_i] = 0$ et $\text{Var}(\epsilon_i) = \sigma^2$.
2. Les variables explicatives X_1, \dots, X_p sont déterministes et connues.

3. Les coefficients β_1, \dots, β_p sont constants.

Le modèle peut également être donné sous forme matricielle par :

$$Y = \beta X + \epsilon \quad (1.2)$$

Grâce à la méthode de minimisation des **Moindres Carrés Ordinaires (MCO)**, on obtient une estimation du vecteur β :

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y \quad (2.2)$$

A partir de là, il est possible de définir le vecteur des **valeurs ajustées** :

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y \quad (2.3)$$

Ainsi que le vecteur des **résidus** :

$$\hat{\epsilon} = Y - \hat{Y} = Y - X(X'X)^{-1}X'Y \quad (2.4)$$

Un indicateur classique qui permet de juger de la qualité d'un modèle linéaire est le **coefficient de détermination**, qu'on note :

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \in [0, 1] \quad (2.5)$$

Où \hat{Y}_i est la valeur prédictive de Y_i et \bar{Y} est la moyenne de tous les termes Y_i pour $i = 1, \dots, n$. Il correspond au rapport entre la variance expliquée par le modèle et la variance totale. Plus ce coefficient est proche de 1, meilleure est l'estimation. Cependant, le R^2 augmente de façon injustifiée avec l'ajout de variables dans le modèle. Dans ce cas il est utile de se référer au coefficient de détermination ajusté, R^2_{ajuste} , qui lui, tient compte du nombre de variables dans le modèle.

Modèle linéaire généralisé

Développés dans les années 1970 par Nelder et Wedderburn, les Modèles Linéaires Généralisés (*Generalized Linear Model, GLM*), permettent de regrouper les modèles linéaires, logistiques ou de Poisson en modifiant la structure de dépendance linéaire entre la variable à expliquer et les prédicteurs à l'aide d'une fonction de lien adaptée.

Ils présentent trois composantes :

— La composante aléatoire du modèle

Ce sont tout simplement les variables à expliquer Y_1, \dots, Y_n dont les densités appartiennent à la famille exponentielle.

Une loi de probabilité sur \mathbb{R} , discrète ou continue, appartient à la famille exponentielle si elle possède une densité de la forme :

$$f(y | \theta, \phi) = \exp \left(\frac{c(y, \phi) + y\theta - b(\theta)}{a(\phi)} \right) \quad (2.6)$$

Où $a(\cdot)$, $b(\cdot)$ et $c(\cdot)$ sont des fonctions connues, $\theta \in \mathbb{R}$ est le « paramètre naturel » et $\phi \in \mathbb{R}$ est le « paramètre de dispersion » dans le cas où $a(\phi) = \phi$.

Cette formulation inclut la plupart des lois usuelles telles que les lois normales, Poisson, Gamma, Binomiale mais aussi des lois moins classiques comme les lois Tweedie, Binomiale négative ou Gaussienne inverse.

On peut montrer que si Y appartient à la famille exponentielle, alors

$$\mu = \mathbb{E}(Y) = b'(\theta)$$

$$\text{Var}(Y) = a(\phi)b''(\theta) = a(\phi)V(\mu)$$

où $V(\mu)$ est appelée fonction variance.

Les éléments μ , $a(\phi)$ et $V(\cdot)$ caractérisent exhaustivement la loi de Y .

Loi Gamma

Pour modéliser le coût moyen des sinistres liés aux inondations, nous utiliserons la loi Gamma, couramment employée en actuariat pour représenter les coûts de sinistres. Cette distribution est particulièrement adaptée, car elle permet de modéliser des variables aléatoires strictement positives, ce qui correspond aux coûts des sinistres sur les bâtiments. De plus, elle prend en compte l'asymétrie de la distribution des coûts et la présence d'une queue à droite, correspondant au faible nombre de sinistres avec un coût très élevé.

La densité de la loi Gamma $\Gamma(k, \alpha)$ s'écrit :

$$f_Y(y; k, \alpha) = \frac{y^{k-1} e^{-\alpha y}}{\Gamma(k) \alpha^k}$$

En prenant le logarithme :

$$\ln f_Y(y; k, \alpha) = (k-1) \ln(y) - \ln(\Gamma(k)) - k \ln(\alpha) - \alpha y$$

On peut également l'écrire sous la forme :

$$\ln f_Y(y; k, \alpha) = (k-1) \ln(y) - \ln(\Gamma(k)) + k \ln(k) - k \ln \alpha - k \frac{y}{\alpha} + \ln k - \ln \alpha$$

Par identification il est possible de retrouver les éléments de l'équation 2.6 et ainsi prouver l'appartenance de la loi Gamma $\Gamma(k, \alpha)$ à la famille exponentielle :

$$\begin{cases} \theta = -\frac{1}{k\alpha} \\ b(\theta) = -\ln(-\theta) = -\ln(k\alpha) \\ a(\phi) = \frac{1}{k} \\ c(y, \phi) = (k-1) \ln(y) - \ln(\Gamma(k)) + k \ln(k) \end{cases}$$

On peut ainsi déterminer la valeur de l'espérance de Y :

$$E(Y) = \mu = b'(\theta) = -\frac{1}{\theta} = k\alpha$$

et celle de sa variance :

$$\text{Var}(Y) = b''(\theta) \cdot a(\phi) = \frac{(k\alpha)^2}{k} = k\alpha^2$$

Ainsi,

$$v(\mu) = (k\alpha)^2.$$

Loi de Poisson

La loi de poisson en revanche est souvent utilisée en assurance pour modéliser la survenue d'événements rares sur une période donnée. Cette loi est définie par un unique paramètre λ , représentant l'espérance du nombre d'événements. Elle repose sur des principes en cohérence avec la sinistralité climatique : les sinistres sont indépendants les uns des autres et la moyenne des sinistres est supposée stable dans le temps.

Ainsi, si l'on note λ le nombre moyen d'occurrences de sinistres, alors pour $k \in \mathbb{N}$:

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!} = \exp(k \ln(\lambda) - \lambda - \ln(k!)).$$

Par identification, il est possible de retrouver les éléments de l'équation 2.6 tels que :

$$\begin{cases} \theta = \ln(\lambda), \\ b(\theta) = e^\theta = \lambda, \\ a(\phi) = 1, \\ c(y, \phi) = -\ln(y!). \end{cases}$$

L'espérance et la variance de Y sont donnés, par identification, par :

$$\mu = b'(\theta) = b''(\theta) \cdot a(\phi) = v(\mu) = \lambda.$$

— La composante déterministe

La composante systématique est le prédicteur linéaire de taille $n \times 1$ défini tel que :

$$E(Y) = \mu = \beta X.$$

Avec β un vecteur de p paramètres et X la matrice des observations planifiées des variables explicatives constituées des vecteurs :

$$X_1 = \begin{pmatrix} X_{11} \\ \vdots \\ X_{1n} \end{pmatrix}, \dots, X_p = \begin{pmatrix} X_{p1} \\ \vdots \\ X_{pn} \end{pmatrix}$$

— La fonction de lien

La fonction de lien est une fonction monotone, différentiable et inversible qui permet de lier la composante déterministe à la composante aléatoire du modèle : $\mu = \mathbb{E}(Y)$ est liée au prédicteur linéaire $\eta = \beta_0 + \sum_{j=1}^p \beta_j X_j$ par une fonction de lien (notée g dans la suite).

$$\begin{aligned} g(\mu) &= g(\mathbb{E}(Y \mid X = x)) = \beta_0 + \sum_{j=1}^p \beta_j x_j = \eta \\ &\Rightarrow \mu = g^{-1}(\eta) \end{aligned} \tag{2.7}$$

La fonction de lien canonique est la fonction lien qui associe la moyenne μ_i au paramètre canonique θ_i . Elle est définie telle que

$$g(\mu_i) = \theta_i.$$

Or, $\mu_i = b'(\theta_i)$ d'où $g^{-1} = b'$.

Estimation des paramètres et qualité du modèle

L'estimation des paramètre β_0, \dots, β_p , se fait par la méthode du maximum de vraisemblance. La fonction de log-vraisemblance s'écrit :

$$\log L(Y, \beta) = \sum_{i=1}^n \ln(f(y_i \mid \theta_i, \phi))$$

L'estimateur du Maximum de Vraisemblance (EMV) est défini comme

$$\hat{\beta}_{\text{EMV}} \in \arg \max_{\beta} L = \arg \max_{\beta} \log L.$$

De plus, lorsque g est la fonction de lien canonique, $\theta_i = \beta x_i$, ce qui donne :

$$\log L(Y, \beta) = \sum_{i=1}^n (c(y_i, \phi) + y_i \beta x_i - b(\beta x_i)) / a(\phi)$$

Les équations de vraisemblance obtenues sont :

$$\forall j = 0, \dots, p, \sum_{i=1}^n \frac{x_{i,j}}{a(\phi)} (y_i - b'(\beta x_i)) = 0$$

Ces équations n'ont généralement pas de solutions explicites. Dans ce cas, il est possible d'utiliser des méthodes numériques pour obtenir une estimation des paramètres.

La qualité des modèles sera évaluée selon différents axes (détaillés par la suite) :

- La qualité d'ajustement du modèle sera évaluée à l'aide de plusieurs indicateurs mesurant les écarts entre les observations et les prédictions. Ces analyses permettront également d'examiner la significativité des variables.
- L'analyse des résidus aidera à déterminer si le modèle nécessite des améliorations, en vérifiant leur distribution aléatoire et en identifiant d'éventuelles valeurs aberrantes.
- La représentation graphique des *multipliers* permettra de visualiser l'effet marginal des variables explicatives sur la variable réponse, toutes choses étant égales par ailleurs.

Utilisation du machine learning comme outil de sélection et de comparaison

Pour mieux comprendre le phénomène, les modèles GLM seront comparés à des modèles issus de méthodes de machine learning, qui offrent plusieurs avantages. Parmi ces méthodes, les arbres de classification et de régression (CART) se démarquent en tant qu'outils non paramétriques particulièrement adaptés à la modélisation des relations non linéaires et complexes entre variables. Contrairement aux modèles linéaires généralisés (GLM), qui supposent une distribution spécifique des données et peuvent conduire à des prévisions erronées, les arbres CART gèrent efficacement les interactions multiples entre variables et hiérarchisent automatiquement les variables explicatives en fonction de leur importance (BELLINA 2014).

Ainsi, certains modèles dérivés des arbres CART, comme la *forêt aléatoire* (*Random Forest*) et le modèle *CatBoost*, seront utilisés comme outils de comparaison et d'analyse.

Forêt aléatoire

Une forêt aléatoire ou *random forest*, est un modèle d'apprentissage automatique basé en partie sur une méthode appelée "bagging". Introduite par Breiman en 1996, cette méthode, contraction de "bootstrap aggregating" (que l'on peut traduire par "rééchantillonnage et agrégation"), consiste à tirer aléatoirement et avec remise k observations parmi un ensemble de n observations. Ce processus est répété m fois pour générer m échantillons, chacun étant utilisé pour entraîner un modèle, tel qu'un arbre de décision. Les prédictions obtenues sont ensuite agrégées, soit par une moyenne pour la régression, soit par un vote majoritaire pour la classification. La méthode de Random Forest ajoute à la méthode de bagging une part d'aléatoire dans la construction des arbres.

Comme expliqué précédemment, cette méthode est efficace pour la prédiction mais elle permet également de mesurer pour chaque variables explicatives leur importance dans la prédiction de la variable à expliquer. Cette approche sera donc exploitée lors des étapes de sélection des variables pour les modèles de fréquence et de coût moyen (parties 2.2.2 et 2.3.2).

CatBoost

Le modèle CatBoost en revanche, est un modèle basé sur le renforcement de gradient (Gradient Boosting Machines), une technique d'agrégation d'arbres fondée sur le boosting. Son principe général est similaire à celui du bagging mais diffère dans la manière de construire les modèles. En effet, contrairement au bagging qui construit les arbres séparément, le boosting les construit les uns après les autres, afin que chaque arbre corrige les erreurs du précédent.

Ces méthodes de machine learning sont détaillées dans les ouvrages de BELLINA 2014, BREIMAN 1996.

2.1.4 Méthodes de validation

Nous proposons dans cette partie de définir les éléments utilisés pour la validation des modèles.

Lors de la construction de nos modèles, nous avons employé diverses métriques d'évaluation, telles que l'Explained Deviance Ratio (EDR), l'Erreur Quadratique Moyenne (MSE), l'Erreur Absolue Moyenne (MAE) et le coefficient de Gini. Ces indicateurs permettent d'analyser la performance des modèles sous différents angles, en prenant en compte la précision des prédictions, la variabilité des erreurs et l'équité de la répartition des sinistres prévus. En complément de ces métriques, des analyses graphiques ont également été réalisées pour visualiser les écarts entre les prédictions et les observations, et pour identifier d'éventuelles tendances ou biais dans les résultats.

L'ensemble de ces analyses, combinant métriques et visualisations, servira à valider les choix de variables ainsi que la performance des modèles retenus. Nous comparons principalement deux modèles :

- un modèle dit «**assureur**», qui ne prendra en compte que les variables classiques fournies par l'assureur (formule du contrat, nature de la résidence, etc...);
- un modèle appelé «**complet**», qui sera le modèle retenu. Il inclura non seulement les variables assureur, mais également les variables externes relatives à la météo, aux risques environnementaux et au bâtiment.

Les métriques, associées aux analyses graphiques, permettent d'une part de tester la pertinence de la sélection des variables, et d'autre part de vérifier que l'écart entre les observations et les prédictions du modèle reste acceptables, tout en s'assurant que les ajustements apportés à chaque version des modèles améliorent leur qualité d'ajustement et leur pouvoir prédictif.

Métriques de performance

1. L'**EDR** a été développé pour surmonter les limitations du coefficient de détermination, noté R^2 , dans le cadre des modèles GLM. Il correspond à la proportion de déviance résiduelle apprise. C'est à dire qu'il évalue la proportion de déviance du modèle nul M_0 , qui ne contient que l'intercept, expliquée par le modèle M qui inclut les prédicteurs :

$$\text{EDR} = 1 - \frac{D(M)}{D(M_0)} \in [0, 1] \quad (2.8)$$

Avec la déviance, notée D , telle que $D = -2(\log L - \log L_{\text{sat}})$. $\log L$ correspond à la log-vraisemblance du modèle estimé et $\log L_{\text{sat}}$ à celle du modèle dit "saturé", possédant autant de paramètres que d'observations (ROCHELLE 2023).

2. L'indice de **Gini** est un indicateur qui mesure la capacité de segmentation du modèle. Il permet de comparer deux modèles ou de tester l'ajout de nouvelles variables.

Il est calculé à partir de la fonction représentée par la courbe de Lorenz. Cette courbe a été développée en 1905 par Marx O. Lorenz, et permettait de représenter les inégalités de revenus au sein d'une population, en comparant la part des revenus avec la part des détenteurs.

Dans notre cas d'étude, pour le modèle de fréquence, la courbe de Lorenz représente en abscisses la part cumulée des contrats et en ordonnées, la part cumulée du nombre de sinistres observés. Dans le cas du modèle de coût moyen, on représenterait en ordonnées, la part cumulée du coût moyen de sinistres observés.

L'indice de **Gini normalisé**, correspond donc au ratio entre les aires A et B présentes sur la figure 2.1.

La courbe verte sur la figure 2.1 illustre la courbe de "Gini", qui correspond à une courbe de Lorenz appliquée aux prédictions du modèle ordonnancées de façon décroissante.

La courbe bleue, quant à elle, représente la courbe de Lorenz appliquée aux observations réelles ordonnancées de manière décroissante.

L'**aire B** correspond à l'aire située entre ces deux courbes.

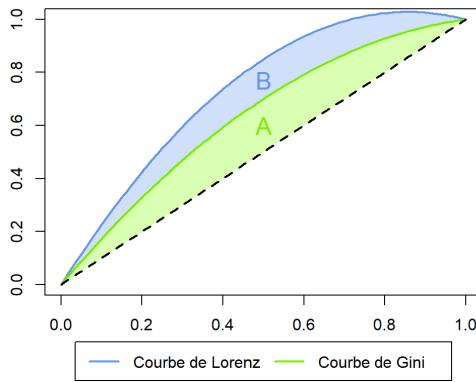


FIGURE 2.1 – Calcul de l'indice de Gini normalisé à l'aide des aires A et B

Si le nombre de sinistres observés (ou coût moyen des sinistres observés) était également répartis par rapport au cumul des contrats, la courbe de Lorenz serait confondue avec la courbe de la fonction identité représentée en pointillé sur la figure 2.1. L'**aire A** est comprise entre la courbe de "Gini" et la courbe de la fonction identité.

Ainsi l'indice de Gini normalisé est calculé tel que :

$$\text{Gini} = \frac{A}{A + B} \quad (2.9)$$

Dans le cas où les observations réelles conservent un ordre décroissant après ordonnancement selon les prédictions du modèles, les courbes de Gini et de Lorenz sont alors confondues et le Gini vaut alors 1, sa valeur maximale. C'est pourquoi, il doit être utilisé conjointement à d'autres mesures comme le Root Mean Square Error (RMSE)(détailé en partie 2.2.4) ou l'EDR car l'indice Gini ne mesure pas l'écart entre les prédictions et les observations (PARENTE 2017).

3. Le **Mean Squared Error (MSE ou Erreur Quadratique Moyenne)** est une métrique qui mesure la moyenne des carrés des écarts entre les valeurs prédictes par le modèle et les valeurs observées. La formule est donnée par (en reprenant les notations précédentes) :

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y - \hat{Y})^2$$

4. L'indicateur **Mean Absolute Error (MAE ou erreur absolue moyenne)**, correspond à la moyenne arithmétique des valeurs absolues des écarts entre prévisions et observations :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |Y - \hat{Y}|$$

Études graphiques

Les graphiques sont également des outils permettant de mesurer l'importance d'une variable dans un modèle. Le premier graphique utile pour la sélection est le graphique des «**Observés vs prédicts**» (Figure 2.2). Il va permettre d'évaluer l'effet de l'ajout d'une variables X_n dans un modèle à $(n - 1)$ variables. Ce graphique est établi en traçant, sur la base d'entraînement, les observations de la variables à expliquer et les prédictions du modèle avant ajout de la variable, noté M_{n-1} , et du modèle (M_n) après ajout de la variable candidate X_n . Si X_n a un réel effet explicatif sur la variable réponse qui n'est pas encore capté par le modèle existant, alors le tracé des observations et celui des prédictions du modèle M_N , seront approximativement semblables.

Sur la figure 2.2, les observations sont représentées par la courbe rouge, tandis que les prédictions du modèle M_n apparaissent en bleu et celles du modèle M_{n-1} en vert. Ces données correspondent à la fréquence des

sinistres observés et prédits selon les modalités de la variable candidate. L'exposition en bleu clair, correspond au nombre de lignes dans la base. Il est clair que les prédictions du modèle M_{n-1} s'écartent considérablement des observations, même dans les zones où les expositions sont élevées. En revanche, les prédictions du modèle M_n s'ajustent bien mieux aux observations, reflétant une amélioration notable dans la capture des tendances de la fréquence des sinistres. Cela montre que l'inclusion de la variable candidate X_n dans le modèle M_n permet de mieux expliquer les sinistres observés. Ainsi, la variable X_n semble apporter une contribution significative à la précision du modèle, réduisant les écarts et améliorant sa performance par rapport au modèle précédent M_{n-1} . Ce graphique peut donc être exploité afin de valider l'ajout d'une variable au modèle.

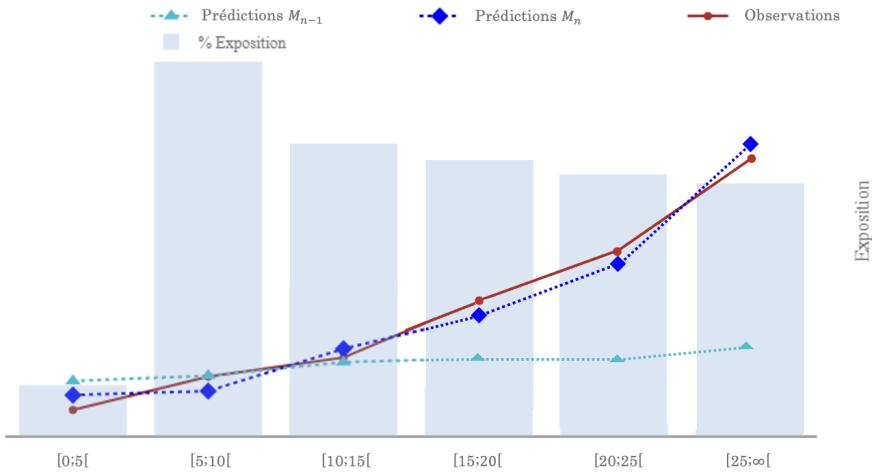


FIGURE 2.2 – Graphique Observés vs prédits

Dans les analyses qui suivront, les graphiques des *Observés vs prédits* présenteront une légère adaptation. Les prédictions issues des modèles « **assureur** » et « **complet** » seront superposées sur un même graphique. Cette approche permettra de comparer directement les deux modèles, facilitant ainsi l'analyse de leurs performances respectives.

Ensuite, nous procéderons à une analyse graphique des *multipliers* du modèle «complet». Cette analyse vise à comparer les effets marginaux des effets croisés. Observer ces *multipliers* offrira une vision plus précise de la manière dont les variables modifient les prédictions. Il est à noter que les axes des ordonnées seront absents sur l'ensemble des graphiques par souci de confidentialité.

2.2 Modèle de fréquence

Dans cette partie, nous présenterons les différentes étapes de construction du modèle de fréquence. Nous y détaillerons les différentes analyses exploratoires réalisées, la sélection des variables et la modélisation de la fréquence au moyen d'un GLM.

L'objectif principal est d'explorer des données innovantes, d'évaluer leur pertinence et de comprendre leur contribution à l'amélioration du modèle. Dans cette optique, nous consacrerons une analyse approfondie de certaines variables clés, en détaillant leur intégration dans le modèle et leur impact sur la fréquence des sinistres, sans pour autant examiner l'ensemble des variables utilisées de façon exhaustive pour des raisons de confidentialité.

2.2.1 Analyse exploratoire des données

Présentation de la variable cible

La variable cible ici est la **fréquence des sinistres** liés aux inondations. Celle-ci est calculée en divisant le nombre de sinistres déclarés pour un bâtiment, enregistré dans la base de données de l'assureur, par son exposition. L'exposition correspond à la proportion de la période de couverture effective d'un bâtiment par

rapport à la période totale d'observation historique disponible. Cette fréquence est une variable continue qui mesure le taux d'occurrence des sinistres par unité d'exposition.

L'analyse de la distribution de cette variable montre qu'une majorité de bâtiments n'ont enregistré aucun sinistre sur la période étudiée, ce qui se traduit par une forte concentration de valeurs égales à zéro. Toutefois, le nombre d'observations non nulles reste suffisant pour mener une analyse pertinente des sinistres subis et de leurs caractéristiques.

De plus, la cartographie de la fréquence des sinistres sur le territoire français met aussi en évidence des disparités marquées. Les fréquences varient considérablement d'une région à l'autre, en fonction des caractéristiques locales telles que la topographie ou la proximité de cours d'eau ou de zones inondables, ainsi que la vulnérabilité des infrastructures. Cette hétérogénéité géographique justifie l'intégration de variables spatiales dans le modèle.

Analyse bivariée

Pour mieux comprendre les relations entre la fréquence des sinistres et les autres variables explicatives, nous avons procédé à une analyse bivariée. Cette approche consiste à étudier la relation entre la fréquence des sinistres avec une autre variable, en visualisant par exemple la fréquence en fonction de la variable explicative étudiée sur un graphique. En complément, nous avons également cartographié certaines variables explicatives potentielles sur le territoire français afin d'évaluer leur influence géographique.

Cela nous a permis d'identifier les premières tendances de la fréquence des sinistres en fonction des différentes modalités des variables explicatives. Par exemple, nous avons observé, que la fréquence des sinistres pour un bâtiment varie bien en fonction de la distance au cours d'eau le plus proche de celui-ci.

Nous avons aussi pu identifier les variables nécessitant un retraitement ou des ajustements pour améliorer leur pertinence dans le modèle.

Enfin, cette première étape d'analyse a permis de guider nos choix méthodologiques, notamment en ce qui concerne la sélection des variables à inclure dans le modèle de prévision.

2.2.2 Sélection de variables

La sélection de variables consiste à sélectionner de manière manuelle ou automatique les caractéristiques les plus pertinentes pour la construction du modèle. Elle permet de réduire le nombre de variables potentielles en éliminant les variables trop corrélées et celles qui n'apportent pas suffisamment d'informations au modèle pour la prédiction. Elle permet également d'avoir une meilleure compréhension de la relation de cause à effet entre les variables explicatives et la variable à expliquer.

Comme expliqué en partie 2.1.3, d'autres approches que les GLM peuvent être employées pour effectuer cette sélection, notamment des méthodes de machine learning qui vont s'appuyer sur l'importance de chaque variable dans le modèle.

Dans notre cas d'étude, le nombre de variables candidates est très élevé, plus d'une centaine de variables issues de la base de fréquence pourraient potentiellement être incluses dans la modélisation. Afin d'aboutir à une sélection optimale, nous avons appliquée plusieurs méthodes permettant de réduire le nombre de variables tout en conservant celles ayant un impact significatif sur le modèle.

La première phase d'analyse exploratoire des données (partie 2.2.1), a permis, avant même toute analyse statistique, d'éliminer les variables n'ayant aucune pertinence métier ou ne présentant aucun lien évident avec la fréquence ou le coût des sinistres liés aux inondations. Cette étape a, par exemple, conduit à l'exclusion de variables telles que la distance du bâtiment au centre commercial le plus proche ou encore le type de toiture du bâtiment, qui ne sont pas censées influencer significativement la fréquence d'un sinistre d'inondation. Grâce à cette première sélection, nous avons réduit la base de données de plus d'une centaine de variables à quelques dizaines, rendant ainsi l'analyse plus efficace.

Dans un second temps, nous avons étudié les corrélations entre les variables explicatives afin de détecter d'éventuelles redondances et de limiter la multicolinéarité. Pour cela, une matrice de corrélation est construite, permettant d'identifier les variables fortement corrélées et de sélectionner celles ayant le lien le plus pertinent

avec la variable cible. La figure 2.3 illustre cette matrice appliquée à un sous-ensemble de variables sélectionnées.

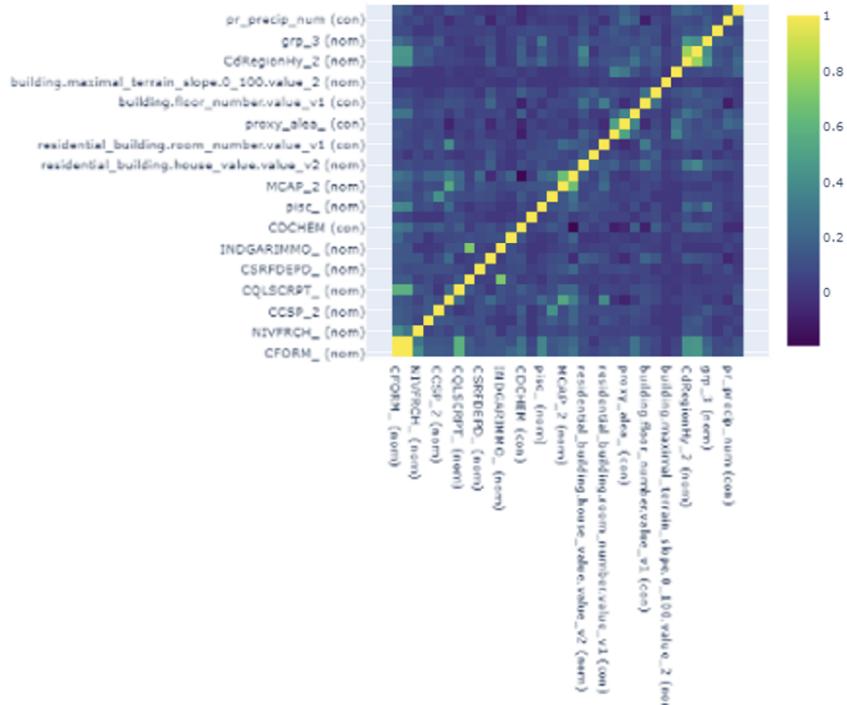


FIGURE 2.3 – Matrice de corrélation

Pour compléter cette analyse, nous avons utilisé un modèle de type Random Forest afin d'estimer l'importance relative de chaque variable (partie 2.1.3). Ce type de modèle étant capable de capturer des relations non linéaires et des interactions complexes, a permis de mettre en évidence les variables influentes qui auraient pu être sous-estimées dans l'analyse de corrélation.

Enfin, nous avons utilisé le modèle GLM pour étudier l'effet marginal de chaque variable sur la variable cible. Les variables présentant des coefficients non significatifs ou un effet incohérent ont été éliminées afin d'assurer la robustesse du modèle final. Cette dernière sélection garantit que seules les variables ayant un impact réel sur la variable cible sont retenues. Des analyses de métriques ainsi que des études graphiques viendront valider la sélection.

2.2.3 Modélisation

Comme détaillé en section 2.1.3, la modélisation du coût moyen et de la fréquence repose sur des modèles linéaires généralisés (GLM). Dans les deux cas, nous avons choisi une fonction de lien logarithmique afin de préserver l'effet multiplicatif entre les variables explicatives. Pour modéliser la fréquence des sinistres, nous avons retenu la loi de Poisson (voir section 2.1.3). Cependant, comme mentionné précédemment, d'autres approches de modélisation ont également été envisagées, notamment les modèles Random Forest et CatBoost. Les résultats de ces trois modèles (GLM, Random Forest et CatBoost) seront présentés dans la section 2.2.4, permettant ainsi une comparaison des performances des modèles.

Pour ce faire, nous séparons notre base de données de fréquence en une base d'entraînement et une base de test. Ce découpage est réalisé aléatoirement de manière à ce que la répartition soit homogène pour les sinistres. Les bases d'apprentissage et de test représentent respectivement 70% et 30% des contrats de la base initiale.

Comme expliqué précédemment, l'objectif de cette partie est d'évaluer la pertinence des variables utilisées, en détaillant leur intégration dans le modèle et leur impact sur la fréquence des sinistres.

Pour cela, nous nous appuierons sur les graphiques *observés vs prédicts* ainsi que sur les graphiques des *multipliers* afin de mieux visualiser ces effets. Nous expliciterons également le retraitement des variables. Enfin,

pour chaque variable incluse dans le modèle complet, nous avons vérifié sa pertinence dans le cadre de l'analyse des inondations, tout en nous assurant que l'ajout de nouvelles variables ne dégradait pas les métriques de performance décrites précédemment.

Variables assureurs

Les variables issues de la base de sinistralité de l'assureur, intégrées au modèle, apportent des informations sur le profil du souscripteur ainsi que sur la nature de la résidence. Elles permettent notamment de distinguer si l'assuré est locataire ou propriétaire, ainsi que de déterminer si le logement constitue une résidence principale ou secondaire. Le statut d'occupation semble être une variable intéressante pour la modélisation de la fréquence de sinistres. En effet, en cas de sinistre inondation, un propriétaire déclarera l'ensemble des dommages à son assureur, tandis qu'un locataire se limitera à déclarer les dégâts subis par son mobilier et ses biens personnels à l'intérieur du logement, le reste des dommages étant à la charge du bailleur. On s'attend donc à ce que la fréquence des sinistres liés aux inondations soit plus élevée pour un propriétaire que pour un locataire.

Une seconde variable fournie par la base assureur et qui entre en compte dans la modélisation, concerne la formule de couverture. La formule de couverture choisie a effectivement un impact direct sur la fréquence de sinistres pour diverses raisons. D'une part, un assuré possédant une couverture étendue, aura tendance à déclarer davantage de sinistres. D'autre part, une couverture plus généreuse peut aussi influencer le comportement de l'assuré, le rendant moins prudent ou plus enclin à déclarer des sinistres mineurs, sachant que ces derniers seront pris en charge par l'assureur.

Présence de piscine

Cette variable a également été exploitée car la présence de piscine peut avoir une incidence sur le risque inondation. En effet la construction d'une piscine modifie souvent le terrain environnant, en ajoutant des surfaces imperméables comme une terrasse par exemple, ce qui diminue la capacité naturelle du sol à absorber l'eau et augmente les risques de ruissellement. De plus, en cas de fortes pluies ou de précipitations abondantes, la piscine peut rapidement déborder si elle n'est pas correctement vidangée ou équipée de dispositifs de drainage efficaces, créant ainsi une accumulation d'eau autour de la propriété.

Valeur du bâtiment

Cette variable représente le prix de l'immobilier divisé par la surface habitable du bâtiment. Cela permet de comparer des bâtiments de tailles différentes en tenant compte de la valeur marchande au mètre carré, plutôt que du prix total, qui pourrait varier en fonction de la superficie.

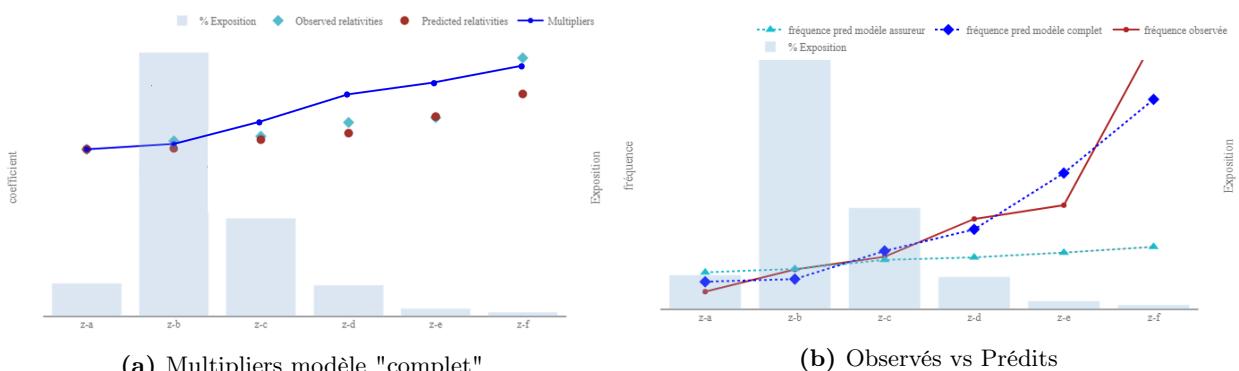


FIGURE 2.4 – Observés vs Prédits et multipliers - Variable "Valeur du bâtiment"

La variable valeur du bâtiment a été discrétisée en plusieurs classes de valeurs croissantes. Le graphe des *Observés vs prédits* (figure 2.4b) indique que la fréquence croît avec le prix du mètre carré. On retrouve également cette tendance avec le graphique des *multipliers* du modèle "complet" (figure 2.4a). Cela peut s'expliquer par le fait que des maisons chères sont souvent accompagnées d'équipements extérieurs beaucoup plus exposés aux inondations comme des jardins, piscines, vérandas, etc.

L'augmentation de la fréquence liée à la valeur des biens immobiliers peut aussi être due à la localisation géographique des propriétés les plus onéreuses. Comme illustré dans la figure 2.5, les biens à forte valeur immobilière sont souvent situés dans des zones attractives, telles que les quartiers situés en bord de mer ou à proximité des cours d'eau. Ces emplacements sont particulièrement prisés en raison de leur cadre de vie agréable, cependant ils sont également plus vulnérables aux inondations en raison de leur exposition directe aux risques liés à la montée des eaux, aux crues fluviales ou à l'érosion côtière.

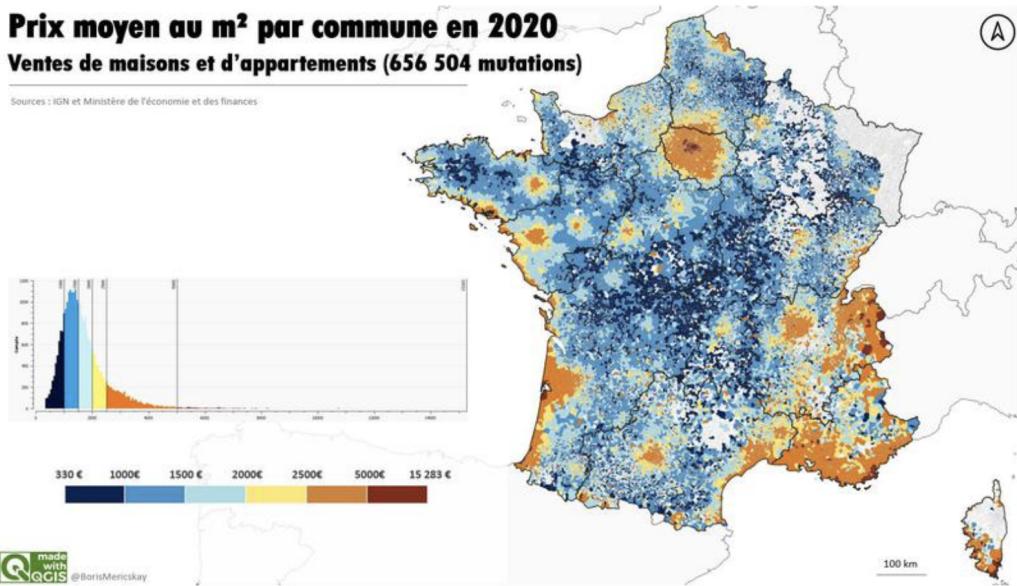


FIGURE 2.5 – Carte des prix moyens au m² par commune des ventes de maisons et d'appartements en 2020

Source : *Data.gouv* (2021)

En complément de la variable valeur du bâtiment, nous avons intégré la variable **période de construction**. En effet, aux alentours des années 1990, des réglementations liées à la construction des bâtiments ont été mises en place pour faire face aux risques majeurs comprenant le risque d'inondation. Le décret n° 95-1089 du 5 octobre 1995, par exemple, a officialisé la mise en place des Plans de Prévention des Risques d'Inondation (PPRI). Ces plans dont la préparation avait débuté à la fin des années 1980, imposent des restrictions aux constructions dans les zones inondables (LÉGIFRANCE 2024).

Après 1995, des mesures spécifiques, comme la surélevation des infrastructures lors de la construction, ont gagné en popularité et sont devenues parfois obligatoires, assurant ainsi une meilleure résilience face aux inondations. L'identification des bâtiments qui ont été construits avant cette date permet de mesurer l'impact des évolutions législatives sur la sinistralité et d'affiner l'estimation du risque en tenant compte des normes de constructions.

TRI et interactions

La variable *TRI et interactions* correspond à une valeur, attribuée à chaque bâtiment, construite à partir de la variable TRI que nous avons mis en interaction avec d'autres paramètres afin d'améliorer son pouvoir explicatif.

Elle a donc été élaborée en prenant en considération :

- Le niveau de risque d'inondation publié dans les TRI de la commune. Ces Territoires à Risques importants d'inondation (TRI) sont des zones où les enjeux potentiellement exposés aux inondations sont les plus importants (DREAL GRAND EST 2022). Ils incluent les risques liés aux risques de débordement des rivières, de débordement de nappe phréatique, de submersion et de ruissellement. Leur cartographie a été imposée par la directive européenne 2007/60/CE, dite "Directive inondation". Son objectif était de cartographier l'ensemble des zones exposées au risque inondation afin de prioriser les investissements de l'Etat en matière de prévention (construction de digues,...). A l'image de la figure

2.6, les cartes des TRI délimitent les zones inondables et indiquent les hauteurs d'eau potentielles en cas d'inondation.

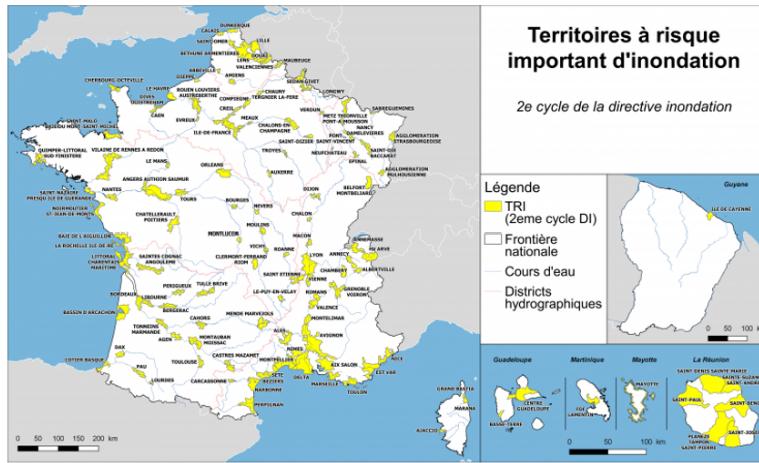


FIGURE 2.6 – TRI

Source : *Ministère de la transition écologique*

- La distance au cours d'eau le plus proche.
- La différence d'altitude entre le bâtiment et ce cours d'eau.

La combinaison de la variable TRI avec les autres paramètres offre ainsi une meilleure représentation du niveau de risque que ne le ferait la variable TRI seule.

Comme escompté, nous avons pu observer sur le graphique des *multipliers* que, toutes choses étant égales par ailleurs, la fréquence des sinistres augmente avec la valeur du proxy associée. Le graphique des *observés vs prédis*, indiquait également une augmentation de la fréquence observée en fonction de la valeur du proxy, particulièrement pour les niveaux les plus élevés. Cette croissance du risque est également nettement mieux capturée par le modèle "complet" que par le modèle "assureur".

PPRI

Les Plan de Prévention des Risques d'Inondation (PPRI) sont des outils de gestion des risques d'inondation mis en place par les services de l'état. Leur objectif principal est de réguler l'urbanisation dans les zones à risques afin de limiter l'exposition des populations et des biens. Comme les autres plans de prévention des risques naturels, les PPRI sont issus de la loi du 2 février 1995, connue sous le nom de Loi Barnier, et sont encadrés par les articles L.562-1 à L.562-12 du Code de l'environnement (DRIEAT ILE-DE-FRANCE 2024).

Il définissent dans la commune, les zones exposées au risque d'inondation, dites zones réglementaires, et les mesures de prévention et de protection à mettre en place pour réduire l'exposition. La carte de zonage réglementaire des PPRI, qui représente les zones à risques, distingue en général deux zones :

- Une zone rouge exposée à des aléas forts ou très forts, pour lesquelles la construction est impossible.
- Une zone bleue exposée à des aléas faibles ou modérés, pour laquelle la construction est possible sous conditions (construction sans sous-sol etc...)

La figure 2.7 est un exemple de PPRI pour la ville de Clamecy.

La variable *PPRI* dont nous disposons, représente le niveau de risque inondation du bâtiment selon le PPRI. Elle se décline en plusieurs catégories : «Élevé +», «Élevé», «Moyen +», «Moyen», «Faible» et «Aucun». Pour simplifier l'analyse et tenir compte des différentes expositions au risque, cette variable a été regroupée en deux catégories : «possède un PPRI» et «ne possède pas de PPRI» («moyen» et «aucun» sur la figure 2.8).

On observe sur le graphique 2.8b, la fréquence de sinistre augmente avec la présence de PPRI dans la zone du bâtiment. Les *multipliers* présentés sur la figure 2.8a confirment cette tendance : la présence d'un PPRI est associée à une fréquence de sinistres plus importante, ce qui souligne l'impact des zones réglementées sur la

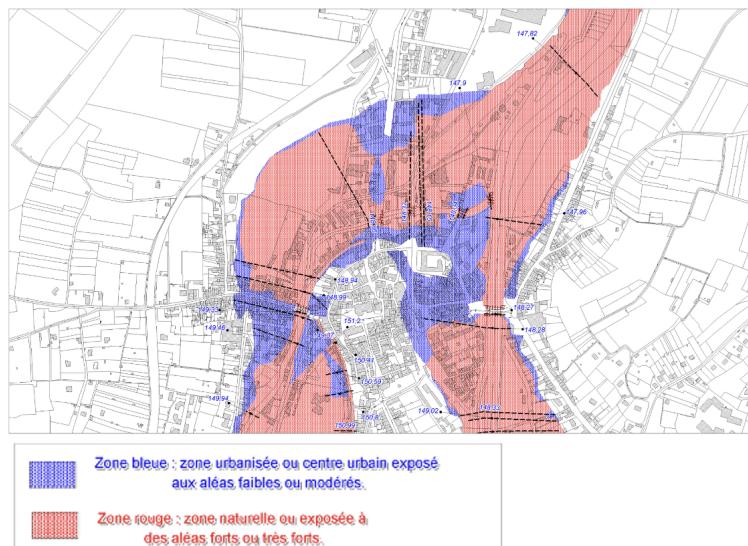


FIGURE 2.7 – Exemple de PPRI pour la ville de Clamecy

Source : *Ville de Clamecy*

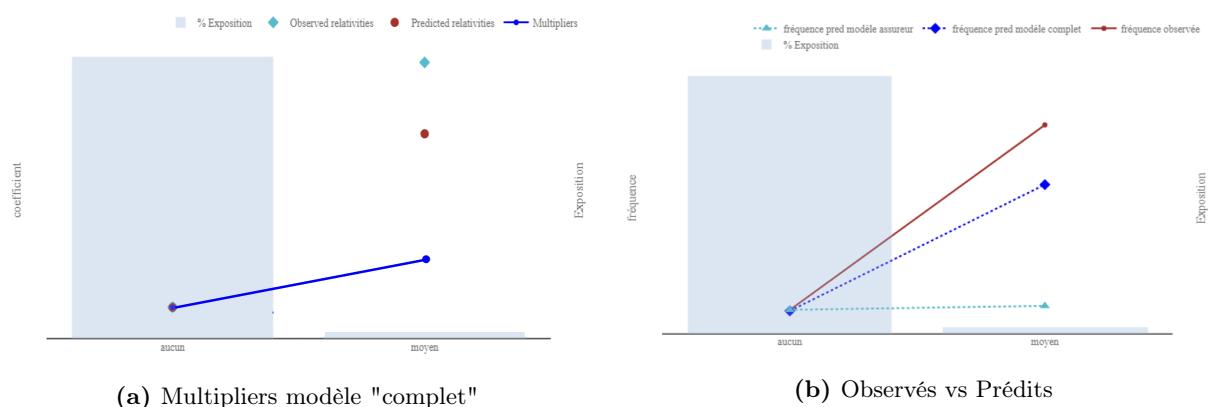


FIGURE 2.8 – Observés vs Prédits et multipliers - Variable «PPRI»

vulnérabilité des bâtiments face aux inondations. Cependant, la faible exposition de la modalité «moyen» limite l'interprétation de son effet. Cette faible représentation peut entraîner une variabilité accrue des estimations et nécessiter une prudence particulière dans l'analyse des résultats.

Nous avons également ajouté à la modélisation, la variable **Zone sujette aux remontées de nappes phréatiques**. Cette variable va permettre de compléter l'information apportée par la variable «PPRI» en capturant un risque sous estimé dans les PPRI mais potentiellement dommageable : celui des remontées de nappes phréatiques. L'ajout de cette variable permet d'avoir une vision plus complète du risque en tenant compte non seulement des inondations classiques mais aussi des phénomènes souterrains liés à la nappe phréatique.

Les graphiques des *multipliers* et des *observés vs prédits* pour la variable **Zone sujette aux remontées de nappes phréatiques** ont été analysés, et leurs résultats sont cohérents avec ceux obtenus pour la variable «PPRI». Ces graphiques montrent une dynamique similaire, confirmant que les zones exposées à des remontées de nappes phréatiques présentent également un risque plus important de sinistres.

Région hydrographique

Avant de définir la notion de région hydrographique, il est important de bien comprendre le concept de **bassin versant**. Un bassin versant désigne le territoire sur lequel tous les écoulements des eaux de surfaces et de ruissellement convergent vers le même point, comme dans un entonnoir. C'est le point de plus basse altitude, appelé exutoire. Les limites séparant deux bassins versants sont appelées lignes de partage des eaux et sont constituées des sommets et des crêtes qui séparent les directions d'écoulement des eaux de ruissellement. Le schéma 2.9 permet d'illustrer cette notion.

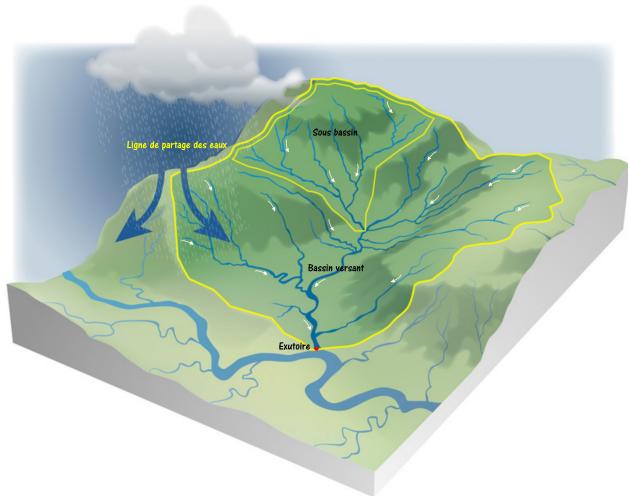


FIGURE 2.9 – Schéma illustratif d'un bassin versant

Source : *SIBBCA*

Ainsi, chaque goutte qui tombe ou arrive sur le bassin versant et destinée à rejoindre l'exutoire de ce bassin, après avoir parcouru la surface du sol. Effectivement, le bassin hydrographique ne concerne que les sols imperméables et les écoulements de surface et ne prend pas en compte les écoulements souterrains et leurs échanges avec les cours d'eau.

De ce fait, tous les cours d'eau finissent par se jeter dans des cours d'eau plus grands, de sorte qu'un bassin versant est constitué de sous-bassins imbriqués. Ce découpage peut donc être affiné à l'intérieur des bassins versants pour identifier des sous-unités, appelées régions hydrographiques. Ces régions permettent une gestion plus locale des ressources en eau et des risques associés, comme les inondations ou la gestion des crues.

Initialement, nous disposions d'une variable «*région_hydro*», associant une région hydrographique spécifique à chaque bâtiment. Cependant, nous avons choisi de regrouper ces différentes régions en 4 grandes catégories hydrographiques homogènes en terme de risque et d'écoulement, afin d'éviter la complexité d'un trop grand nombre de catégories pour un même niveau de risque.

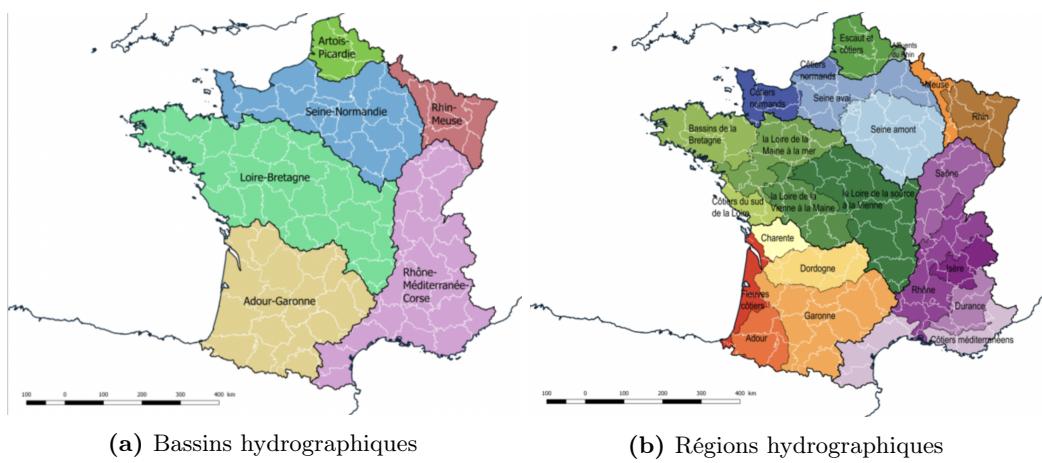


FIGURE 2.10 – Bassins et régions hydrographiques de France métropolitaine

Source : Wiklimate (2020)

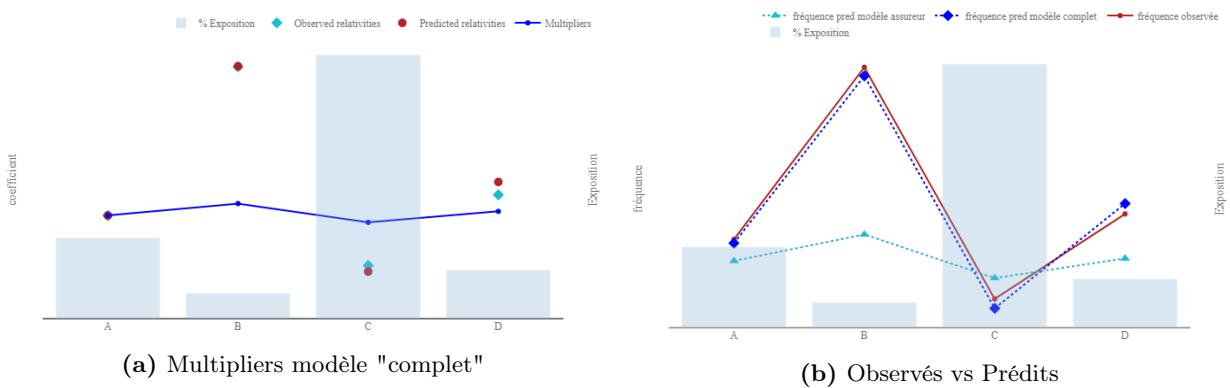


FIGURE 2.11 – Observés vs Prédits et multipliers - Variable "Région hydrographique"

D'après le graphique 2.11b, on constate que, comparé au modèle «assureur», le modèle «complet» permet de mieux capter les variations de la fréquence des sinistres, pour les différentes régions hydrographiques. Ces variations reflètent correctement la réalité observée, et sont en cohérence avec les tendances montrées dans le graphique des *multipliers*. Par exemple, les régions hydrographiques appartenant à la *zone B* sur la figure 2.11, qui affichent les fréquences de sinistres les plus élevées, incluent des zones particulièrement sujettes aux inondations en raison de leur exposition à des conditions climatiques spécifiques, telles que des pluies intenses et des crues soudaines, mais également leur proximité aux cours d'eau majeurs et à la mer.

Inclinaison du terrain

Afin de compléter l'analyse des régions hydrographiques, il est intéressant de prendre en compte la topographie des terrains alentour afin d'avoir une meilleure estimation des risques d'accumulation d'eau ou de débordement.

La variable **Inclinaison du terrain** correspond à la pente du terrain à proximité du bâtiment, calculée sur un rayon de 50 mètres. Concrètement, on mesure la différence d'altitude maximum dans ce périmètre, puis la variable est obtenue en divisant la différence d'altitude par la distance (50 mètres), cela permet d'obtenir un pourcentage représentant l'inclinaison du terrain. Par la suite, nous discrétons la variable en plusieurs catégories.

Après avoir examiner les deux graphiques, celui des *multipliers* et celui des *observés vs prédicts*, on observe, contrairement à ce qu'on pourrait attendre, que la fréquence des sinistres augmente avec l'inclinaison du terrain, toutes choses étant égales par ailleurs. Cela pourrait s'expliquer par le fait que les terrains en pente favorisent un ruissellement plus rapide et plus concentré des eaux de pluie. Au lieu de s'infiltrer lentement dans le sol, l'eau tend à s'écouler rapidement sur des pentes plus raides, provoquant des crues soudaines ou des accumulations

d'eau temporaires. De plus, sur des terrains fortement inclinés, des glissements de terrain ou des coulées de boue peuvent aggraver l'impact des précipitations extrêmes, contribuant à une sinistralité plus élevée. Cette dynamique est d'ailleurs mieux captée par le modèle «complet», qui intègre ces interactions de manière plus précise.

Cette variable est une illustration des variables pertinentes sur le bâtiment et l'environnement qui l'entourent qui peuvent servir à mieux comprendre le risque.

Nombre d'arrêtés Cat Nat

Afin de compléter les indicateurs météorologiques et géographiques tels que les PPRI, la distance aux cours d'eau ou la pente du terrain, nous apportons une dimension temporelle au modèle à travers la variable **Nombre d'arrêtés Cat nat**. Cette variable est issue des données publiées par le journal officiel et mises à disposition par la CCR (Caisse Centrale de Réassurance). Elle couvre la période de 1982 à 2022 et retrace tous les arrêtés officiels de reconnaissance de catastrophes naturelles. Cette variable reflète pour chaque bâtiment son niveau d'exposition historique au risque. La variable a été catégorisée en 4 modalités croissantes.

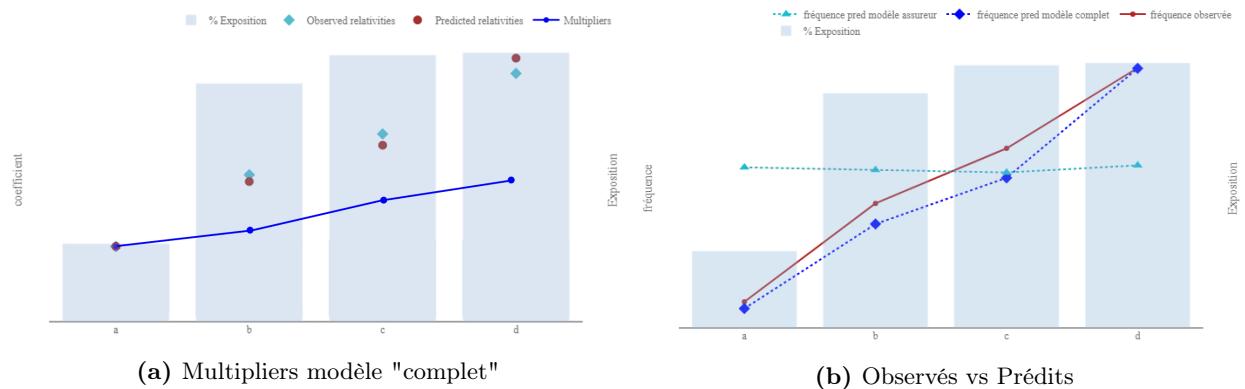


FIGURE 2.12 – Observés vs Prédits et multipliers - Variable "Nombre d'arrêtés Cat Nat"

Les *multipliers* (figure 2.12a) sont croissants avec un nombre plus élevé d'arrêtés, cela signifie que les bâtiments situés dans des zones touchées à plusieurs reprises par des catastrophes naturelles sont beaucoup plus exposés aux sinistres futurs. Ainsi, on peut accepter l'idée qu'un bâtiment fréquemment touchés dans le passé est plus susceptible de subir de nouveaux événements climatiques extrêmes, car ces zones sont identifiées comme récurrentes ou particulièrement vulnérables. On observe également (figure 2.12b), que les prédictions du modèle «complet» sont plus proches des valeurs observées que les prédictions du modèle «assureur». La fréquence de sinistres est plus élevée dans les zones avec un grand nombre d'arrêtés. La variable contribue donc efficacement à expliquer la fréquence des sinistres.

Variables météorologiques

Les variables météorologiques qui vont suivre, utilisées pour la modélisation de la fréquence des sinistres inondation sont issues d'une base de données Addactis. Ce sont des données d'analyse et de réanalyse construite à partir de plusieurs sources de données.

Nombre de jours par an avec des précipitations intenses

Cette variable représente le nombre de jours dans l'année durant lequel le bâtiment a été exposé à des précipitations intenses. Ces précipitations intenses sont caractérisées par un volume d'eau important tombant sur une courte période. L'ajout de cette variable a pour objectif de quantifier la fréquence des épisodes pluvieux extrêmes.

Les graphiques associés montrent une corrélation entre le nombre de jours de précipitations intenses et la fréquence des sinistres observés. Plus le nombre de jours de précipitations intenses est élevé, plus la fréquence des sinistres tend à augmenter. Le graphique des *multipliers* illustre cette tendance. De plus, le graphique des *Observés vs Prédits* montre une bonne adéquation entre les sinistres prévus par le modèle «complet» et ceux réellement observés, validant ainsi la pertinence de cette variable pour expliquer l'exposition au risque d'inondation.

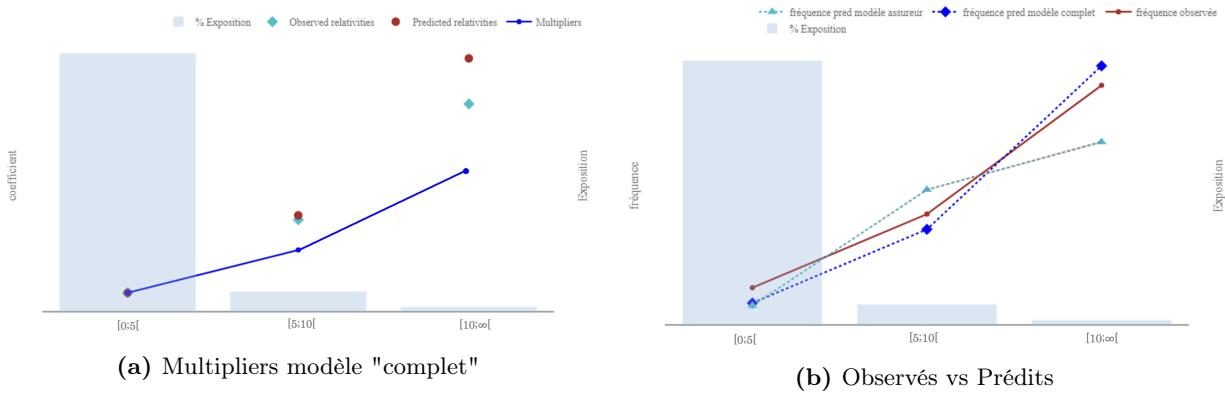


FIGURE 2.13 – Observés vs Prédits et multipliers - Variable "Nb jours avec precip intenses"

Nombre max de jours consécutifs par an avec des précipitations quotidiennes d'au moins 1mm

Certaines régions en France sont naturellement plus exposées aux précipitations que d'autres. On pourrait donc supposer que ces zones, étant plus pluvieuses, présentent un risque d'inondation plus important. Cependant, l'étude de la variable «*Nombre max de jours consécutifs par an avec des précipitations quotidiennes d'au moins 1mm*» montre que les bâtiments situés dans des zones régulièrement exposées à la pluie sont en réalité moins fréquemment inondés. Cela peut s'expliquer par le fait que dans ces régions, les bâtiments sont souvent mieux adaptés et renforcés pour faire face aux précipitations. Ainsi, on a vu que les épisodes de pluies extrêmes augmentent le risque d'inondation mais l'expérience et l'exposition prolongée à de faibles précipitations ont conduit à des constructions plus résilientes.

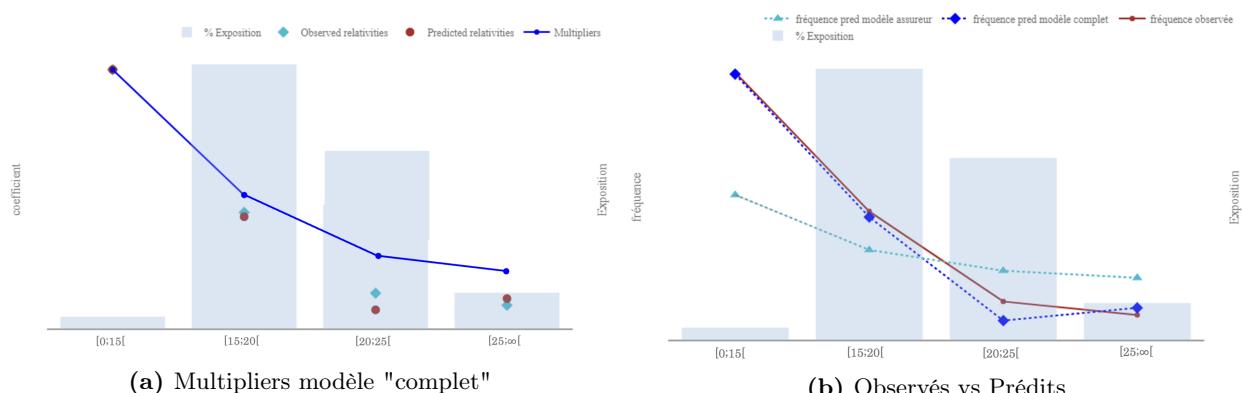


FIGURE 2.14 – Observés vs Prédits et multipliers - Variable "jours consécutifs avec des precip >1mm"

Les prédictions du modèle «complet» s'ajustent aux observations qui reflètent cette dynamique, les bâtiments les plus fréquemment exposés à la pluie sont aussi les moins touchés par des inondations. Quant aux multiplicateurs, figure 2.14a, ils suivent la même tendance : ils diminuent pour les bâtiments situés dans des zones fréquemment touchées par des pluies modérées, cela confirme que ces bâtiments, mieux adaptés aux conditions météorologiques locales, sont moins susceptibles de subir des sinistres liés aux inondations.

Précipitations max sur un jour

La variable **Précipitation maximale sur un jour** représente le volume le plus important de précipitations tombées en une seule journée au cours de l'année. Elle mesure l'intensité maximale d'un épisode pluvieux enregistré sur le bâtiment concerné. Cette variable permet d'évaluer dans quelle zone on a le plus de précipitations intenses en un court laps de temps. Ces conditions augmentent la probabilité de débordement ou de saturation des sols et des systèmes d'évacuation d'eau, provoquant des inondations.

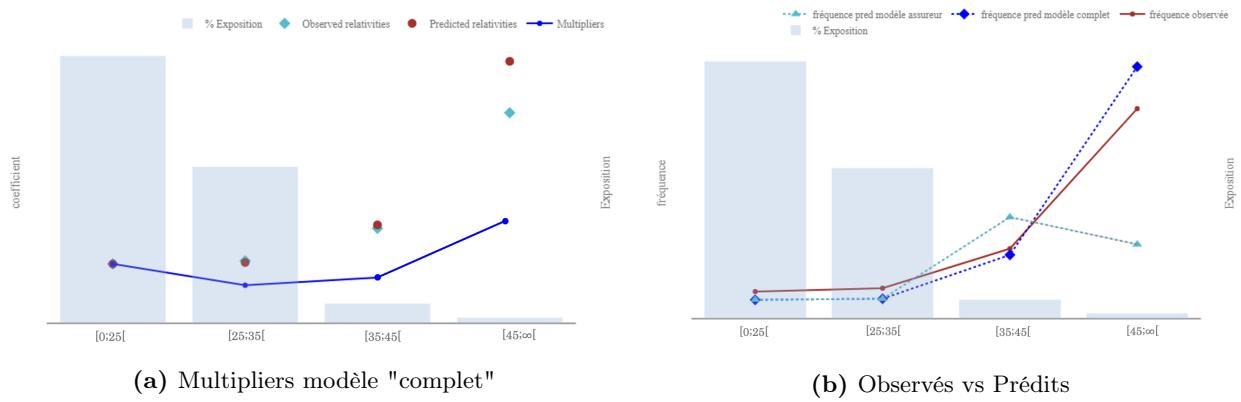


FIGURE 2.15 – Observés vs Prédits et multipliers - Variable "Precip max sur un jour"

Les graphes associés à cette variable (figure 2.15) montrent, tout comme pour les autres variables liées aux précipitations extrêmes, qu'une augmentation des précipitations maximales historiques enregistrées sur un jour est fortement corrélée à une fréquence plus élevée de sinistres inondation.

Cette analyse des variables descriptives a permis de démontrer que le modèle « complet », intégrant non seulement les variables fournies par l'assureur mais également des variables météorologiques, climatiques et spécifiques aux bâtiments, permettait de capter plus d'effets variés et de mieux ajuster le modèle aux observations que le modèle « assureur ». Ces données supplémentaires permettent une amélioration de la précision des prévisions. Elles enrichissent également la compréhension globale du phénomène étudié, en apportant des informations sur les interactions entre les sinistres et les conditions environnementales.

2.2.4 Validation du modèle

Comme mentionné précédemment, lors de l'élaboration du modèle, nous avons évalué le gain en termes de métriques pour chaque variable intégrée, ainsi que la pertinence des représentations graphiques, notamment celles des *observés vs prédits* et des *multipliers* (partie 2.2.3) pour le modèle GLM. Nous avons également appliqué des algorithmes plus complexes, tels que le random forest et le catboost, aux mêmes variables sélectionnées pour évaluer leur efficacité. Ces algorithmes ont été choisis pour vérifier si leur capacité de modélisation non linéaire offrait un avantage supplémentaire par rapport au GLM. À présent, nous allons présenter les valeurs des métriques pour les différents modèles, afin de démontrer que l'utilisation du modèle GLM est justifiée et que les résultats des métriques confirment que le modèle «complet» est nettement plus pertinent. L'objectif global est de minimiser les erreurs, telles que le MSE et le MAE, tout en maximisant les indicateurs de performance comme le Gini et l'EDR.

Le tableau 2.2 récapitule donc les valeurs des métriques pour chacun des modèles examinés.

Les deux premières métriques présentées permettent d'évaluer la précision du modèle. En ce qui concerne le MSE, il mesure l'écart quadratique moyen entre les prédictions et les observations. On constate que pour chaque modèle, il diminue lorsqu'on passe du modèle assureur au modèle complet, ce qui signifie que l'ajout de variables explicatives améliore la qualité des prédictions. Le Random Forest est le modèle avec une amélioration la plus marquée, passant d'un MSE de 0.0504 à 0.0487.

Le MAE en revanche, mesure l'écart absolu moyen entre les prédictions et les observations, sans donner de poids supplémentaire aux erreurs extrêmes. Contrairement au MSE, cette métrique met davantage en évidence les erreurs systématiques. L'amélioration la plus marquée entre modèle assureur et complet, se retrouve encore dans le modèle Random Forest, dont le MAE diminue significativement de 0.0023 à 0.0018. Pour le GLM et le modèle CatBoost, la baisse est moins importante.

Globalement, la précision des prédictions s'améliore avec le modèle complet, en particulier pour les modèles non linéaires. Toutefois, l'amélioration observée lors du passage du modèle assureur au modèle complet du GLM reste significative, même si elle est moins marquée.

	MSE	MAE	Gini	EDR
GLM				
Modèle Assureur	0.0502	0.0023	0.19	0.35
Modèle Complet	0.0501	0.0023	0.24	0.42
Random forest				
Modèle Assureur	0.0504	0.0023	0.22	0.40
Modèle Complet	0.0487	0.0018	0.31	0.43
Catboost				
Modèle Assureur	0.0505	0.0025	0.22	0.36
Modèle Complet	0.0502	0.0023	0.29	0.42

TABLE 2.2 – Valeurs des métriques pour les différents modèles

Le Gini et l'EDR sont des métriques qui permettent d'évaluer la capacité de segmentation des modèles. Pour le GLM, l'ajout de nouvelles variables explicatives améliore l'indice de Gini. Le Random Forest bénéficie d'une amélioration encore plus marquée, avec un Gini qui passe de 0.22 à 0.31. Bien que le gain pour le modèle CatBoost soit légèrement inférieur, il suit la même tendance avec une progression du Gini de 0.22 à 0.29. Cela confirme que l'ajout de variables explicatives améliore systématiquement la capacité de segmentation des modèles. Les modèles de machine learning sont cependant plus performants que le GLM sur cette métrique, car par nature, ils capturent mieux les relations non linéaires et complexes.

L'EDR suit une tendance similaire et augmente pour tous les modèles avec l'ajout de variables. Pour le GLM, l'EDR passe de 0.35 à 0.42, ce qui traduit une amélioration significative de la capacité explicative. Ces résultats confirment que, quel que soit le modèle de prédiction choisi, le modèle complet est nettement plus performant que le modèle assureur en termes d'indice de segmentation.

Enfin, bien que les modèles de machine learning affichent des meilleures performances en termes de Gini et d'EDR, le GLM reste un choix pertinent pour plusieurs raisons. Contrairement aux modèles d'apprentissage automatique, le GLM offre une interprétabilité des coefficients et permet d'analyser directement l'impact de chaque variable sur la fréquence des sinistres. Dans un cadre assurantiel, cette clarté est essentielle pour expliquer les résultats et justifier les tarifications auprès des clients. Il permet également une meilleure compréhension des facteurs qui influencent le risque. De plus, bien que les modèles Random Forest et CatBoost affichent des meilleurs scores de Gini et d'EDR, ils peuvent capturer des non-linéarités inutiles, augmentant le risque de surajustement, que le GLM ne prendra pas en compte.

Le modèle GLM est donc un choix fiable et justifié dans notre cas d'étude. Ses performances sont validées par les différentes métriques et il offre un bon compromis entre précision, robustesse et interprétabilité.

Analyse des résidus

Pour évaluer la qualité de l'ajustement du modèle en tenant compte de la distribution utilisée, nous avons également choisi d'examiner les résidus du modèle.

Les résidus sont la différence entre les observations réelles y_i et les prédictions \hat{y}_i . Cependant, ces résidus ne prennent pas en compte l'asymétrie de la distribution des erreurs dans un modèle de Poisson. Ils ne sont pas non plus normalisés, ce qui signifie qu'ils n'intègrent pas le fait que, pour une loi de Poisson, la variance augmente avec la valeur prédite. C'est pourquoi, nous utiliserons plutôt les **résidus de déviance**. Ces résidus mesurent la contribution de chaque observation à la déviance du modèle par rapport au modèle saturé (WIKISTAT 2016).

Ces résidus sont définis à partir de la déviance. Le résidu associé à la i ème observation est donné par la formule suivante :

$$rD_i = \sqrt{D_i} \times \text{sgn}(y_i - \hat{y}_i) = \sqrt{2y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i)} \times \text{sgn}(y_i - \hat{y}_i) \quad (2.1)$$

où D_i représente la contribution à la déviance pour l'observation i et $\text{sgn}(y_i - \hat{y}_i)$ est le signe de la différence entre l'observation et la prédiction.

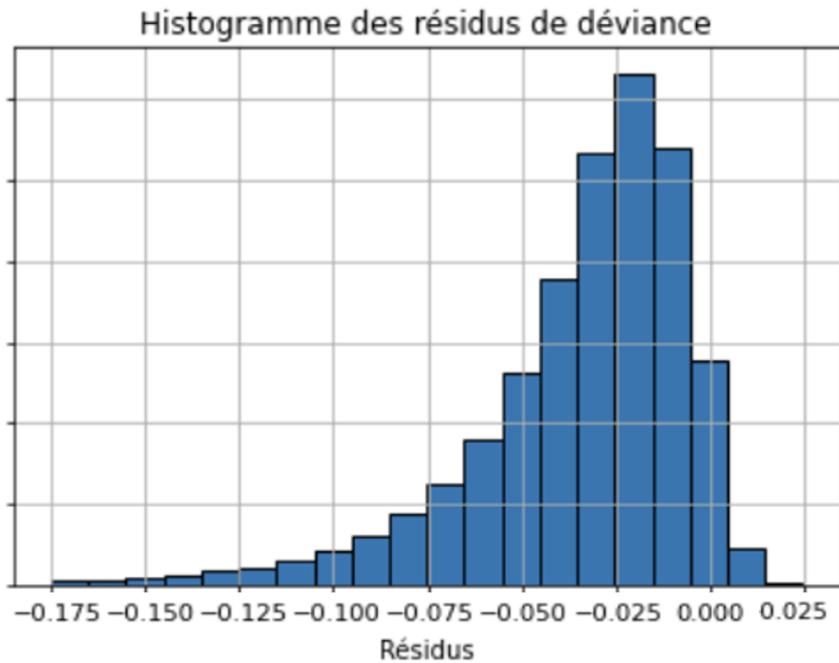


FIGURE 2.16 – Résidus de déviance

Les résidus de déviance pour le modèle de fréquence sont présentés sur la figure 2.16. On remarque que la distribution est majoritairement négative. En effet, dans le cas de la modélisation d'une fréquence, et en particulier avec un modèle GLM utilisant une loi de Poisson, les prédictions du modèle sont toujours positives ou nulles. Ainsi, il est rare que le modèle prédise une fréquence égale à zéro exactement. Le terme $sgn(y_i - \hat{y}_i)$ est donc toujours négatif pour une fréquence observée nulle.

Cependant, les résidus sont concentrés autour de -0.025 , ce qui signifie que pour beaucoup d'observations, l'erreur du modèle reste faible et qu'il est globalement bien ajusté. Ainsi, même si les résidus sont majoritairement négatifs, cela n'affecte pas nécessairement les performances du modèle, car celui-ci reste capable de capturer la tendance générale et de fournir des prédictions cohérentes.

Lift Curve

En complément, une analyse par *Lift curve* sera également réalisée.

La Lift Curve est un outil graphique utilisé pour évaluer la performance d'un modèle prédictif. Elle compare les valeurs moyennes des observations avec les valeurs moyennes des prédictions du modèle pour différents segments qui possèdent une exposition similaire (nombre de lignes équivalent). Cette représentation graphique permet donc de visualiser si le modèle capture correctement les tendances dans l'ensemble des sous-groupes notamment les extrêmes.

Nous avons donc ordonné les données selon les valeurs prédictives afin de produire la courbe (figure 2.17).

On observe que la courbe des prédictions (en rouge) suit de près, en moyenne, la courbe des observations réelles (en bleu), même dans les segments les plus extrêmes. Cette adéquation entre les prédictions et les observations témoigne de la capacité du modèle à bien ajuster les données, y compris dans les cas où les fréquences de sinistres sont particulièrement élevées ou faibles.

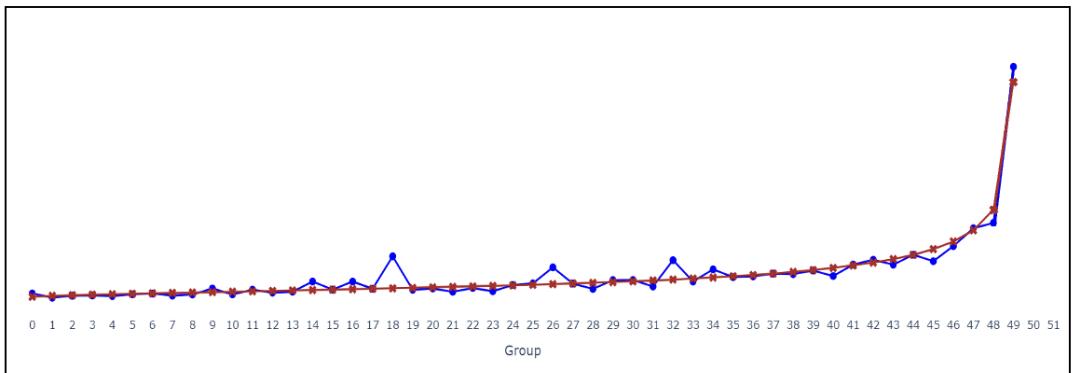


FIGURE 2.17 – Lift curve

2.3 Modèle de coût moyen

Cette partie récapitule les étapes de construction du modèle de coût moyen conditionnellement à la survenance d'une inondation d'intensité i . Elle englobe les phases de préparation des données, de sélection des variables, de modélisation, ainsi que de validation du modèle.

Tout comme pour le modèle de fréquence, les variables présentées ci-après ne constituent qu'un sous-ensemble des variables réellement utilisées, par souci de confidentialité.

2.3.1 Analyse exploratoire des données

Présentation de la variable cible

Nous utilisons ici la base de sinistralité, qui est extraite de la base de contrats d'assurance, mais ne retient que les lignes associées à des sinistres déclarés. Par conséquent, cette base est beaucoup plus restreinte en taille comparée à la base de contrats complète. En effet, le faible historique disponible limite le nombre total de sinistres observés, ce qui réduit la quantité de données exploitables pour l'analyse. Le coût moyen des sinistres liés aux inondations est directement disponible dans cette base de données. Il représente la dépense associée à un sinistre et est exprimé en euros.

La distribution de cette variable présente une large dispersion, allant de quelques dizaines d'euros à plus de 100 000 euros. Cependant, la majorité des sinistres se situent dans une fourchette comprise entre 1 000 et 20 000 euros. Par la suite, un retraitement sera effectué afin d'écrêter les valeurs trop importantes. Par ailleurs, une cartographie des coûts moyens révèle des variations importantes à travers le territoire français. Ces disparités géographiques suggèrent que certains territoires sont plus touchés par des sinistres coûteux que d'autres. Ces écarts peuvent être liés à plusieurs facteurs, tels que la vulnérabilité locale aux inondations, la nature des biens assurés, ou encore les pratiques de gestion des sinistres dans différentes régions.

Cette analyse a également permis de déceler la présence de coûts d'ouverture dans la base de sinistre. Ceux-ci correspondent à un montant arbitraire que fixe l'assureur lorsqu'il ne connaît pas encore la vraie valeur du sinistre. Ils sont donc retirer de la base car ils n'apportent aucune information sur la véritable valeur du sinistre.

Prise en compte de l'inflation

Pour tenir compte de l'inflation sur la période de couverture observée du portefeuille, un ajustement du coût moyen a été effectué en utilisant l'indice FFB (Fédération Française du Bâtiment) du coût de la construction. Cet indice, exprimé en base 1 au 1^{er} janvier 1941, est publié trimestriellement. Il se base sur le prix de revient d'un immeuble moyen à Paris et enregistre les variations de coût des différents éléments qui rentrent dans le calcul du coût de la construction. Il prend en compte le coût de la main d'œuvre par exemple, des matériaux de construction, etc. L'objectif initial de cet indice est d'assurer l'indexation des polices d'assurance

(FÉDÉRATION FRANÇAISE DU BATIMENT 2024).

La moyenne annuelle de cet indice pour chaque année a été utilisée pour calculer les coefficients d'ajustement au niveau de l'année de référence. La table 2.3 présente les coefficients d'ajustement pour une vision 2024 des coûts.

Année	Indice FFB	Coefficient d'ajustement
2019	994,13	1,178
2020	996,9	1,175
2021	1044,3	1,122
2022	1129,1	1,037
2023	1157,7	1,012
2024	1172	1

TABLE 2.3 – Tableau des indices FFB et coefficients d'ajustement de 2019 à 2024

Écrêttement des valeurs extrêmes

Ensuite, nous avons écrêté les sinistres afin de limiter l'impact des valeurs extrêmes sur notre modèle. Concrètement, cela signifie que tous les sinistres initialement supérieurs à un seuil prédéfini seront ajustés à ce seuil. Cette démarche permet de réduire la variance induite par des sinistres très élevés, qui pourraient fausser les résultats de l'analyse et nuire à la précision des prévisions.

Le seuil a été déterminé à l'aide d'outils statistiques, tout en tenant compte de l'expertise métier. Nous avons notamment tracé la courbe du mean excess, qui illustre l'excédent moyen au-dessus d'un certain seuil, et la courbe de Hill, qui permet d'évaluer la queue de distribution des sinistres (des informations supplémentaires sur ces méthodes sont fournies en annexe). Ces deux représentations graphiques nous ont orientés vers un seuil de valeurs extrêmes fixé à 75 000 euros.

Pour valider notre choix de seuil, nous avons examiné le coût des sinistres en fonction de la hauteur d'eau lors d'une inondation d'une durée supérieure à 48 heures, en considérant des logements de 115 m² et de 200 m² (voir figure 2.18).

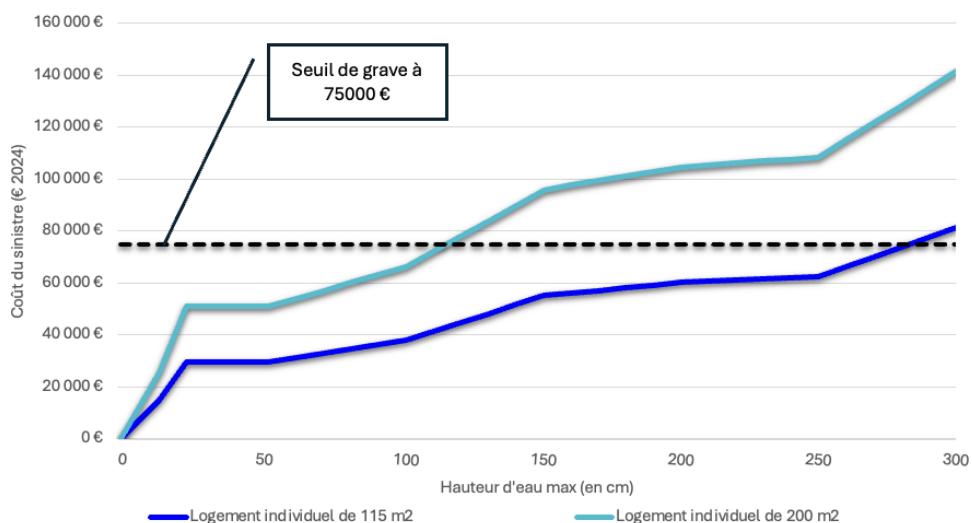


FIGURE 2.18 – Dommages au bâti et mobilier pour une durée de submersion >48h

Source : Ministère de la transition écologique « Analyse multicritère des projets de prévention des inondations » (2018)

Selon une étude de l'INSEE menée en 2020, la surface moyenne des maisons individuelles en France métropolitaine est de 113,8 m², très proche des 115 m² pris en compte dans l'étude. L'analyse révèle que la courbe représentant les dommages pour un logement de 115 m² atteint le seuil de 75 000 € uniquement dans le cas d'une inondation dont la hauteur d'eau dépasse 280 cm, une situation considérée comme extrêmement grave et relativement rare. Pour un logement individuel plus spacieux de 200 m², ce même seuil est atteint dès qu'une hauteur d'eau dépasse 1 mètre pendant plus de 48 heures. Cette situation, bien que moins extrême, demeure très critique. Ces observations corroborent notre décision d'établir le seuil à 75 000 €.

Analyse bivariée

Tout comme pour le modèle de fréquence, les relations entre le coût moyen des sinistres liés aux inondations et les différentes variables explicatives ont été analysées. Une cartographie de ces variables a également été réalisée, permettant ainsi d'identifier les retraitements nécessaires et de faire ressortir les premières tendances concernant le coût moyen.

2.3.2 Sélection de variables

La sélection des variables pour le coût moyen est similaire à celle décrite pour l'obtention d'un modèle de fréquence (partie 2.2.2).

2.3.3 Modélisation

Conformément au modèle de fréquence, le modèle de coût moyen conditionnel sera également basé sur un GLM (modèle linéaire généralisé). Dans ce cas, nous utiliserons la loi Gamma, couramment employée en assurance IARD pour modéliser le coût moyen des sinistres, comme expliqué en partie 2.1.3.

Cette section, à l'image de la précédente, détaillera l'intégration de quelques variables utilisées dans le modèle, ainsi que leur impact sur le risque lié aux inondations. Pour chaque variable incluse dans le modèle «complet», nous avons également vérifié que son inclusion était pertinente, tant au niveau des métriques que par des analyses graphiques. Enfin, nous avons également veillé à ce que l'interprétation des coefficients soit cohérente avec les connaissances métiers. Cela nous a permis de nous assurer que l'ajout de chaque variable contribuait à améliorer l'ajustement du modèle, ainsi que son pouvoir explicatif et prédictif.

Cumul de Précipitation

La variable **cumul de précipitations**, décrite en détail en partie 1.2.3, reflète l'intensité des inondations. Elle constitue un indicateur clé pour évaluer les coûts associés aux inondations. En intégrant cette variable, le modèle gagne en précision. Elle permet également de segmenter les coûts en fonction des différentes intensités d'inondation et de projeter le coût moyen en fonction de scénarios climatiques futurs.

Le graphique des *multipliers* (figure 2.19a) illustre une relation positive : à mesure que l'indice de cumul des précipitations augmente, les coefficients liés au coût moyen des sinistres s'élèvent, traduisant des dommages de plus en plus importants. Quant au graphique des *observés vs prédis* (figure 2.19b), il montre que le modèle "complet" reproduit correctement cette tendance. Ce bon ajustement souligne l'importance de cette variable tant dans la modélisation que dans l'estimation du risque futur.

Variables assureur

Certaines variables initialement présentes dans la base de données de l'assureur ont été conservées dans notre modèle. Ces variables ont été sélectionnées en raison de leur pertinence pour expliquer et prédire le coût moyen des sinistres liés aux inondations.

Par exemple, la qualité du souscripteur se révèle être une variable intéressante dans l'étude du coût moyen. Les dommages subis par les locataires et les propriétaires non occupants semblent généralement de moindre ampleur, notamment en raison d'une quantité réduite de biens mobiliers à risque. Cela conduit potentiellement à des coûts de sinistres plus faibles pour ces catégories.

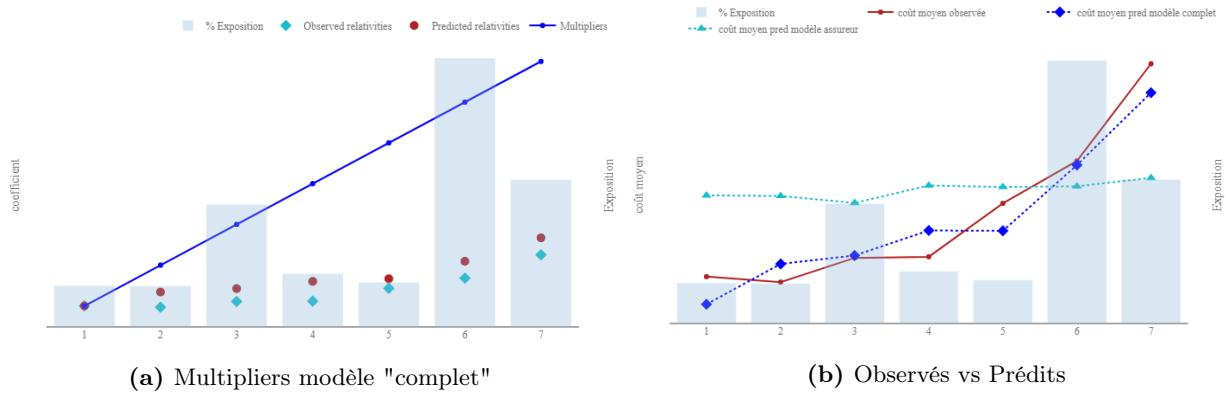


FIGURE 2.19 – Observés vs Prédits et multipliers - Variable "Cumul de précipitation"

Période de construction

La période de construction a également été intégrée dans la modélisation. En effet, cette variable peut influencer considérablement le coût moyen des sinistres inondation. Les bâtiments plus anciens, en particulier ceux construits avant la mise en place de réglementations spécifiques (autour des années 1990) peuvent être plus susceptibles d'être endommagés par les inondations en raison de techniques et matériaux de constructions obsolètes (FASTERCAPITAL 2024). Par exemple, les propriétés plus anciennes peuvent avoir des systèmes de drainage inadéquats ou des matériaux de construction moins résistants. De plus les coûts de remise de reconstruction pour ces logements anciens peuvent être très élevés en raison de remise aux normes.

Nombre de pièces

Le nombre de pièces, fortement corrélé à la surface habitable (tableau 2.4), constitue un indicateur clé du coût moyen des dommages en cas d'inondation. A priori, une surface plus grande implique un volume plus important de matériaux et de biens mobiliers susceptibles d'être endommagés, augmentant ainsi le coût potentiel des sinistres.

Nombre de pièces (p)	Surface moyenne en m^2
1 pièce	29
2 pièces	48
3 pièces	70
4 pièces	90
5 pièces	109
6 pièces et plus	147

TABLE 2.4 – Surface moyenne des logements français selon le nombre de pièces

Source : INSEE 2002

Le graphique des *multipliers* montre que les coefficients augmentent presque proportionnellement au nombre de pièces dans un logement. Cela signifie que, toutes choses étant égales par ailleurs, le coût moyen des sinistres croît avec le nombre de pièces. Bien que cette tendance soit moins marquée sur le graphique des *observés vs prédits*, on y observe tout de même que les prévisions de coûts s'ajustent en fonction de la taille du logement.

Cette analyse démontre l'importance d'intégrer la surface du logement dans la modélisation des coûts de sinistres dus aux inondations.

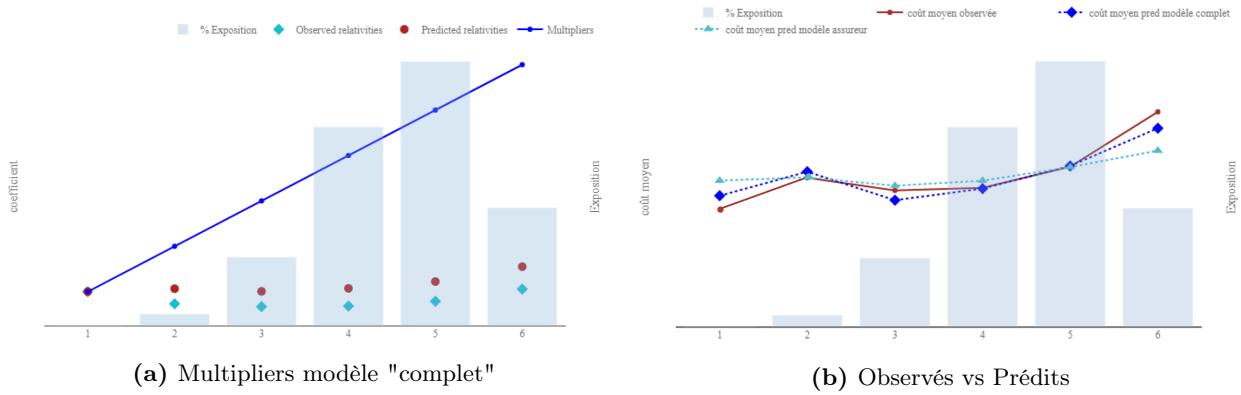


FIGURE 2.20 – Observés vs Prédits et multipliers - Variable "Nombre de pièces"

Nombre d'étages

En plus de prendre en compte la taille du logement, il est essentiel d'intégrer l'aménagement intérieur, et en particulier la présence d'un étage. Comme expliqué en partie 1.1.4, disposer d'un ou plusieurs étages réduit les dommages causés par les inondations. En effet, les biens mobiliers ainsi que les installations électriques et énergétiques, peuvent être déplacés à l'étage, à l'abri de l'eau, ce qui limite significativement le coût des sinistres.

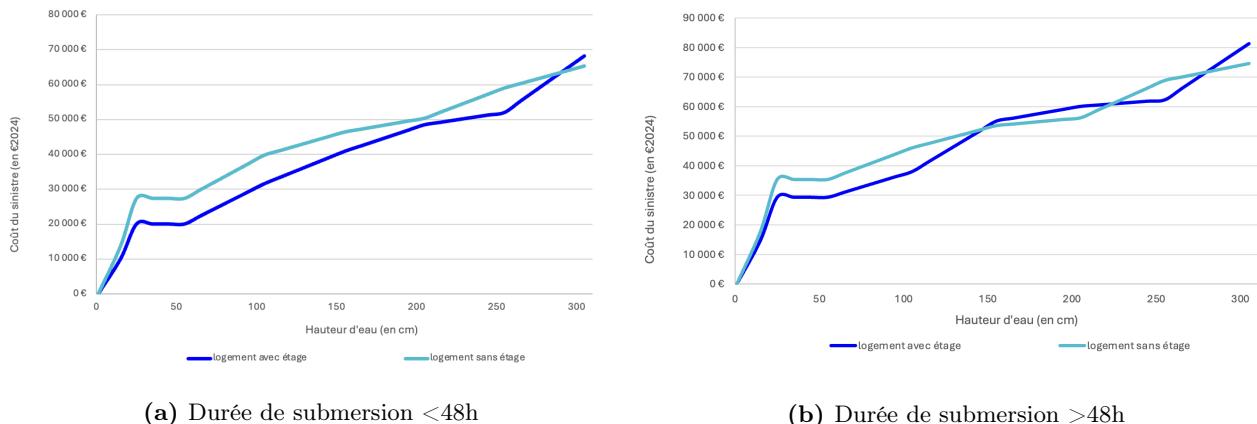


FIGURE 2.21 – Dommages au bâti et mobilier pour un logement de 115 m²

Source : Ministère de la transition écologique « Analyse multicritère des projets de prévention des inondations » (2018)

La figure 2.21 illustre l'évolution des coûts des dommages dus aux inondations (en euros constants de 2024) en fonction de la hauteur d'eau et de la durée de submersion pour des logements de 115 m², avec ou sans étage (pour référence, la surface moyenne des maisons individuelles en France est de 113 m²). Il ressort que, pour une durée de submersion inférieure à 48 heures, le coût des sinistres est bien plus faible pour les maisons avec étage, les biens ayant pu être protégés à l'étage. Les deux courbes se rejoignent aux alentours de 300 cm de submersion, hauteur à laquelle l'eau atteint généralement les étages, annulant cet avantage.

Pour des durées de submersion supérieures à 48 heures, l'effet protecteur de l'étage persiste jusqu'à environ 150 cm de submersion, au-delà de laquelle les deux courbes se rejoignent et se croisent, indiquant que l'inondation prolongée pourrait saturer les structures et causer des dégâts dans toute la maison, étage inclus.

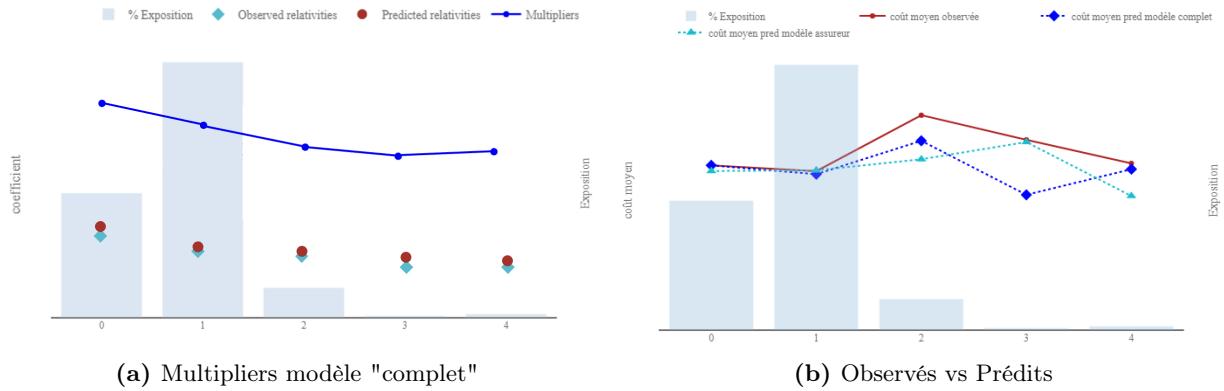


FIGURE 2.22 – Observés vs Prédits et multipliers - Variable "Nombre d'étages"

Le graphique des *multipliers* (figure 2.22a) confirme cet effet de manière claire : toute chose étant égale par ailleurs, la présence d'un ou plusieurs étages réduit significativement le coût moyen des sinistres liés aux inondations. Cependant, le graphique des valeurs *observés vs prédicts* n'est pas aussi concluant. Cette différence peut s'expliquer par le fait qu'un logement avec étage tend à avoir une plus grande surface habitable et donc potentiellement plus de biens mobilier susceptibles d'être endommagés lors d'une inondation. Autrement dit, bien que la présence d'un étage permette de protéger certains biens en les élevant hors de portée de l'eau, l'augmentation de la surface habitable et du volume de mobilier entraîne également une hausse du coût potentiel de sinistre. Il est donc nécessaire d'intégrer à la fois le nombre d'étages et le nombre de pièces dans le modèle.

TRI et interactions

Cette variable est la même qu'en partie 2.2.3, elle prend en compte le niveau de risque d'inondation publié dans les TRI de la commune, qui inclut les risques liés aux risques de débordement des rivières, de débordement de nappes phréatiques, de submersion et de ruissellement. En plus, nous avons associé à cet indicateur, la distance au cours d'eau le plus proche et la différence d'altitude entre le bâtiment et ce cours d'eau. Cela a permis de construire une variable offrant une meilleure représentation du niveau de risque associé au bâtiment.

Les deux graphiques (qui ne seront pas présentés pour des raisons de confidentialité) indiquent que les coûts des dommages subis par les bâtiments situés dans les zones à risque sont plus élevés. En effet, les *multipliers* du modèle augmentent avec la valeur du niveau de risque. On peut l'expliquer par le fait que le coût des dommages peut être plus important dans les zones de TRI à haut risques car ils peuvent nécessiter des mesures de protection et de reconstruction plus coûteuses, comme des fondations spéciales, ou la reconstruction de certaines parties du bâtiment pour répondre aux normes plus strictes de sécurité contre les inondations. La répétition des sinistres dans ces zones peut également augmenter le coût global des sinistres au fil du temps. Le graphique des *observés vs prédicts*, permet également de constater que le modèle "complet" capture nettement mieux la tendance de la fréquence à croître avec le niveau de risque associé au TRI, que le modèle "assureur".

PPRI

De même, les bâtiments situés dans les zones couvertes par un Plan de Prévention des Risques d'Inondation (PPRI) présentent des coûts de dommages significativement plus élevés, pour des raisons similaires à celles observées pour les bâtiments situés dans les zones TRI.

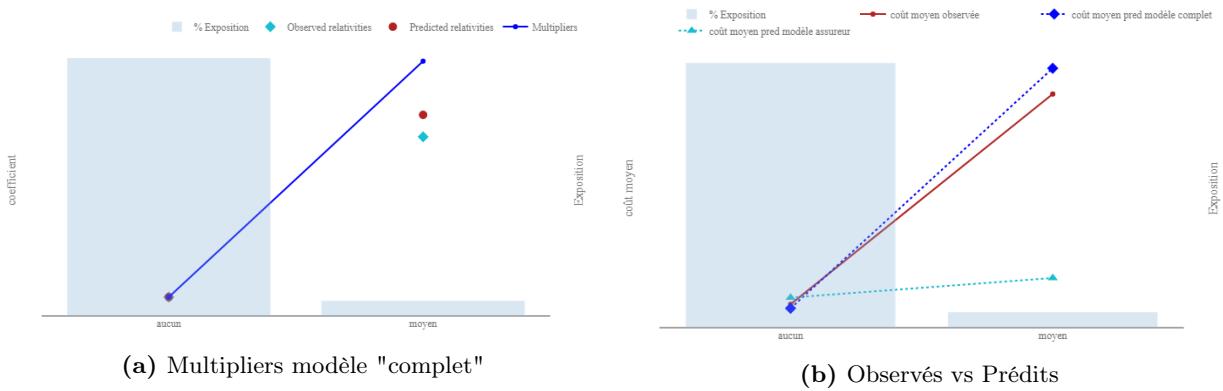


FIGURE 2.23 – Observés vs Prédits et multipliers - Variable "PPRI"

Région hydrographique

La variable **région hydrographique**, décrite en partie 2.2.3, a été intégrée au modèle de coût moyen pour capturer l'impact des bassins versants sur le coût des inondations. Le graphique des *multipliers* (figure 2.24a) montre une nette distinction entre les régions, certaines affichant des coefficients plus élevés, témoignant d'un coût des dommages significativement supérieur.

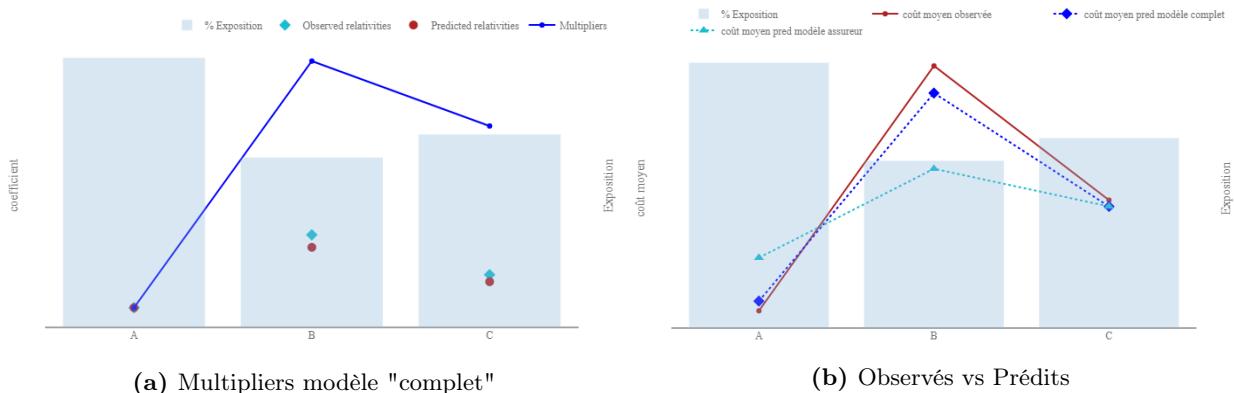


FIGURE 2.24 – Observés vs Prédits et multipliers - Variable "Région hydrographique"

Région climatique

En plus des régions hydrographiques, notre modèle prendra en compte les différences de climat entre les régions de France métropolitaine. En effet, le climat français est globalement tempéré, avec des précipitations réparties tout au long de l'année. Cependant, certains paramètres comme l'altitude, la proximité de la mer ou la distance aux massifs montagneux font varier le climat selon les régions (MÉTÉOFRANCE 2023).

Météo-France identifie cinq grands types de climats en métropole (figure 2.25) :

- Océanique,
- Océanique altéré,
- Semi continental,
- De montagne
- Méditerranéen.

Dans le cadre de notre modélisation, les bâtiments ont été regroupés en deux grandes catégories de régions climatiques.



FIGURE 2.25 – Cinq grands types de climats en France métropolitaine

Source : *Météo France (2023)*

Certaines régions climatiques peuvent plus ou moins influencer le coût des sinistres inondation, car certaines sont plus sujettes à des épisodes de pluie intenses et violentes. Par exemple, les épisodes méditerranéens (MÉTÉOFRANCE 2024), fréquents dans les zones proches de la mer Méditerranée, se caractérisent par des pluies torrentielles en un temps très court. Ces phénomènes sont causés par des remontées d'air chaud, humide et instable provenant de la Méditerranée. En raison de leur intensité et de leur rapidité, ils sont difficiles à anticiper, rendant la protection des bâtiments et des biens mobiliers complexe. Cette montée des eaux brutale augmente les coûts de réparation.

Un exemple marquant est l'épisode méditerranéen catastrophique du 3 octobre 2015 dans les Alpes-Maritimes (L'ARGUS DE L'ASSURANCE 2015). A Mandelieu, Cannes et Nice on compte plus de 130 mm d'eau en seulement 2h. Selon Le Monde, la CCR a évalué les dégâts entre 550 et 650 millions d'euros. À l'opposé, dans les régions climatiques océaniques ou semi-continentales, les inondations sont généralement plus fréquentes, mais moins violentes.

Les deux graphiques présentés (figure 2.26) confirment cet effet. La zone B présente des coûts plus importants, mieux capté par le modèle "complet".

Les analyses graphiques, notamment des *multipliers* et des *observés vs prédis*, confirment la pertinence des choix effectués et la capacité du modèle à refléter fidèlement les variations des coûts moyens. En complément, d'autres variables concernant le bâtiment ont été ajoutées à la modélisation mais ne seront pas présentées.

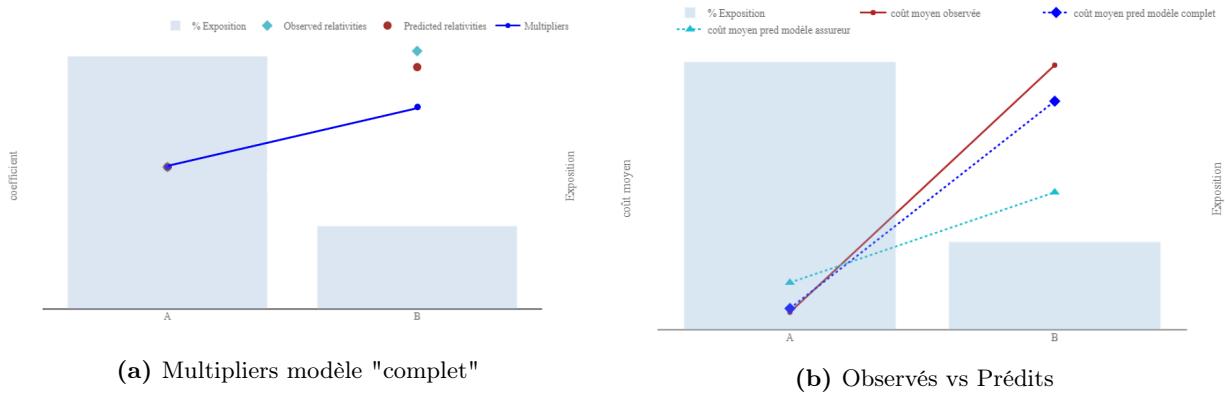


FIGURE 2.26 – Observés vs Prédits et multipliers - Variable "Région climatique"

2.3.4 Résultats et validation

Comme pour le modèle de fréquence des sinistres, l'objectif de cette partie est d'évaluer et de valider la pertinence du modèle choisi par rapport à des alternatives non linéaire. De plus il va permettre de confirmer que le modèle dit "complet" offre une meilleure performance que le modèle constitué uniquement des variables issues de la base assureur.

Le tableau 2.5 présente une comparaison des résultats obtenus sur différents tests de modèles. L'objectif principal reste de minimiser les erreurs, notamment le MSE et le MAE, tels que décrits précédemment, tout en maximisant les indicateurs de performance comme le Gini et l'EDR. Les résultats présentés concernent la base de test, mais les performances pour la base d'entraînement ont aussi été étudiés puis le modèle a été appliqué à l'ensemble de la base.

	MSE	MAE	Gini	EDR
<hr/>				
GLM				
Modèle Assureur	9827	4842	0.4	0,42
Modèle Complet	7407	3589	0.51	0.55
<hr/>				
Random forest				
Modèle Assureur	7225	3472	0.53	0.56
Modèle Complet	5106	2577	0.57	0.62
<hr/>				
Catboost				
Modèle Assureur	7830	3962	0.52	0,58
Modèle Complet	4906	2563	0.60	0.69

TABLE 2.5 – Valeurs des métriques pour les différents modèles

Le MSE et le MAE diminuent systématiquement pour tous les modèles lorsque l'on passe du modèle assureur au modèle complet, confirmant ainsi que l'ajout de nouvelles variables explicatives améliore la précision des prédictions, y compris pour le coût moyen des sinistres. Bien que le GLM affiche un MSE plus élevé que les modèles de machine learning, cet écart reste modéré. De plus, la baisse du MAE, qui mesure l'erreur absolue moyenne, est significative pour tous les modèles, ce qui démontre une réduction globale des écarts entre les prédictions et les valeurs réelles lors du passage au modèle "complet".

L'indice de Gini s'améliore également nettement avec le modèle complet, notamment pour le GLM, indiquant une meilleure capacité à discriminer les individus à risque. L'EDR suit la même tendance, confirmant que l'intégration de nouvelles variables permet une segmentation plus efficace des sinistres.

Malgré des performances parfois légèrement inférieures aux modèles de Machine Learning en termes de précision brute, le GLM reste le modèle privilégié pour des raisons similaires à celles évoquées pour la fréquence des sinistres. En effet, il offre une meilleure interprétabilité, permettant d'évaluer précisément l'impact de chaque variable sur le coût moyen, ce qui est essentiel en tarification. De plus, les valeurs des métriques du GLM restent cohérentes par rapport aux autres modèles, confirmant qu'il constitue une approche fiable.

Analyse des résidus

Les résidus de déviance du modèle ont ensuite été analysés. Nous reprenons pour cela la formule 2.1 où D_i représente la déviance de la loi Gamma. Nous avons donc :

$$r_{D_i} = \sqrt{2} \cdot \left[-\log \left(\frac{s}{s_{\text{obs}}} \right) + \left(\frac{\bar{y}_{\text{obs}_i} - \bar{y}_{\text{pred}_i}}{\bar{y}_{\text{obs}}} \right) \right] \cdot \text{sgn} (\bar{y}_{\text{obs}_i} - \bar{y}_{\text{pred}_i}) \quad (2.1)$$

La distribution des résidus de déviance est également observée en figure 2.27. On peut donc constater que ces résidus sont relativement centrés et tendent à valider le modèle choisi.

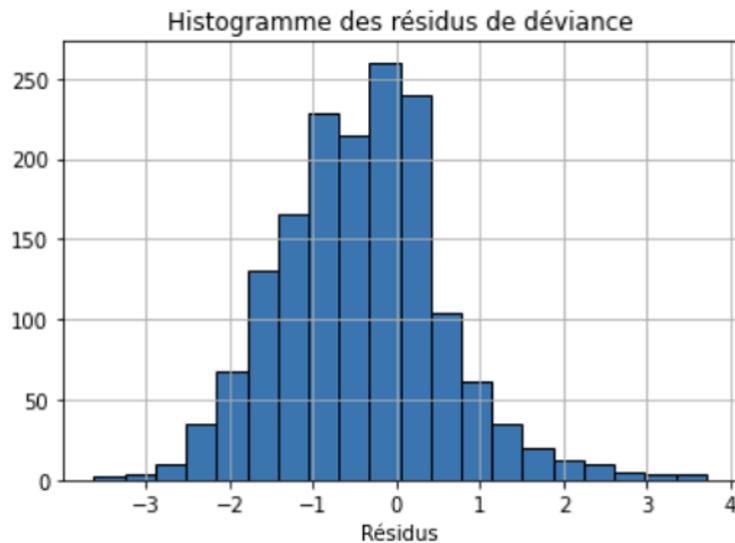


FIGURE 2.27 – Résidus de déviance

Lift Curve

La Lift curve présentée en figure 2.28 représente les coûts moyens observés et prédits selon des groupes équivalents en terme de nombre de sinistres. Elle permet de s'assurer qu'en moyenne les coûts moyens sont satisfaisants, tant pour les faibles valeurs que pour les valeurs extrêmes.

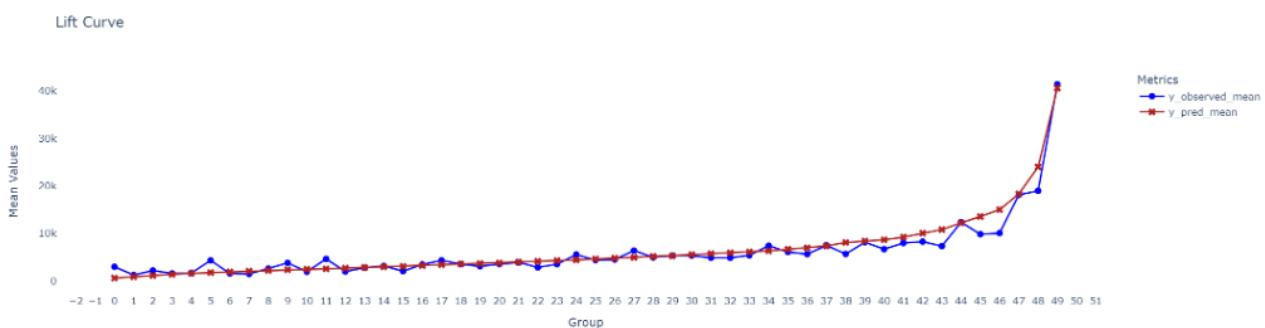


FIGURE 2.28 – Lift curve

2.4 Obtention de la prime pure

Cette section détaille la méthodologie utilisée pour calculer la prime pure à partir des modèles de coût moyen et de fréquence. La prime pure est obtenue à l'aide de l'équation 2.1, qui agrège les éléments prédits par ces modèles.

Prédiction de la fréquence

La base de modélisation utilisée pour prédire la fréquence reprend les variables explicatives nécessaires à l'application de ce modèle.

Les variables explicatives sont ensuite retraitées conformément à la procédure décrite dans la partie 2.2, afin d'assurer leur compatibilité avec le modèle de fréquence. Les prédictions obtenues à partir de ce modèle seront désignées par *pred_freq* dans les étapes suivantes.

Prédiction du coût moyen

Les variables explicatives nécessaires au modèle de coût moyen sont également issues de la base de modélisation, sauf pour la variable cumul de précipitations. Comme expliqué en partie 1.2.3, la construction de cette variable repose sur une date d'inondation et des données de précipitations antérieures. Dans le cadre des prédictions prospectives sans date d'événement, cette variable est recalculée en attribuant des probabilités aux différentes intensités possibles.

Pour chaque bâtiment, le coût moyen est prédit pour chacune des sept intensités de précipitations possibles (voir tableau 2.6). Par exemple, pour le bâtiment 1, le coût moyen prédit pour un cumul de précipitation d'intensité 1 (*pred_cm_1*) est obtenu en appliquant le modèle de coût moyen et en fixant la variable cumul de précipitations à la valeur correspondant à l'intensité 1 ([0,0.75]). De manière similaire, le coût moyen prédit pour un cumul de précipitation d'intensité 7 (*pred_cm_7*) pour ce même bâtiment est calculé en fixant la variable cumul de précipitations à l'intensité 7 ([0.999,1]). Ces prédictions seront notées *pred_cm_i* où *i* correspond à la valeur de l'intensité. La prédiction globale du coût moyen est ensuite obtenue en effectuant une moyenne pondérée des coûts prédits par les probabilités associées à chaque intensité :

$$\text{pred_cm_global} = \sum_{i=1}^7 \text{pred_cm}_i \times \text{proba}_i$$

Les probabilités associées à chaque intensité sont rappelées dans le tableau 2.6.

Modalité	<i>cumul_precip</i> discrétisée	Probabilité associée
1	[0,0.75[0.75
2	[0.75,0.85[0.10
3	[0.85,0.95[0.10
4	[0.95,0.97[0.02
5	[0.97,0.98[0.01
6	[0.98,0.999[0.019
7	[0.999,1[0.001

TABLE 2.6 – Probabilité associée à chaque modalité de la variable *cumul_precip*

Calcul de la prime pure

Enfin, pour chaque bâtiment assuré, la prime pure climatique pour le risque inondation est obtenue en multipliant la fréquence prédite par le coût moyen global :

$$\text{Prime pure climatique} = \sum_{i=1}^7 \text{pred_cm}_i \times \text{proba}_i \times \text{pred_freq} = \text{pred_cm_global} \times \text{pred_freq}$$

Le tableau 2.7 illustre les différentes étapes du calcul de la prime pure pour quelques exemples de bâtiments assurés.

Id building	<i>pred_freq</i>	<i>pred_cm_1</i>	...	<i>pred_cm_7</i>	<i>pred_cm_global</i>	Prime Pure
1	0.0085	1246.97	...	2204.38	1370.56	11.28
2	0.0004	2005.67	...	3545.20	2204.10	0.98
3	0.0041	5169.10	...	8311.02	5100.10	20.65
4	0.0017	780.96	...	1254.51	857.72	1.43
...

TABLE 2.7 – Calcul de la prime pure climatique au bâtiment

Une représentation visuelle sous forme de carte des primes pures peut être générée pour la France métropolitaine, en utilisant le code INSEE associé à chaque bâtiment. Cette carte permet de visualiser, à l'échelle de la commune, le montant des primes pures attribuées. Par souci de confidentialité, celle-ci ne sera pas présentée dans ce mémoire. Toutefois, elle permet une analyse géographique de la distribution des primes en identifiant les zones où les primes sont particulièrement élevées ou au contraire beaucoup plus faibles.

Les résultats obtenus vont désormais permettre la projection des primes pures.

Chapitre 3

Projection

Dans cette section, nous détaillons notre démarche de projection de la fréquence et du coût moyen des sinistres pour le risque d'inondation, en nous appuyant sur les modèles créés précédemment. Ces éléments seront agrégés ensuite pour élaborer une projection de la prime pure à l'horizon 2030. L'objectif de cette projection est d'évaluer l'évolution de la sinistralité liée aux inondations dans un contexte de changement climatique.

3.1 Éléments de projection

3.1.1 Contexte

Dans la partie 1, nous avons établi que le secteur de l'assurance est particulièrement vulnérable aux impacts du changement climatique en raison de son rôle central dans la protection contre les risques et l'indemnisation des dommages causés par les aléas climatiques. Des études récentes, comme le Livre blanc de Covéa (COVÉA 2022) ou le Rapport Langreney (LANGRENEY 2023), confirment une tendance à la hausse de la fréquence des événements extrêmes, notamment les inondations. La CCR projette, par exemple, une augmentation de 50% du coût des dommages liés aux inondations pour les biens des particuliers d'ici 2050.

La projection d'une prime pure individuelle pour chaque bâtiment à l'horizon 2030 dans le cadre du risque d'inondation permettrait de dresser une cartographie précise des évolutions de primes à différentes échelles géographiques sur le territoire français. L'objectif est d'avoir une vision détaillée des taux d'évolution des primes et d'identifier les zones où les sinistres devraient s'aggraver.

Cette démarche aiderait les assureurs à anticiper les zones émergentes en termes de risque. Les zones actuellement les plus exposées pourraient ne plus l'être en 2030, en raison des évolutions climatiques ou des politiques d'aménagement du territoire. Une meilleure connaissance de ces régions permettra non seulement d'ajuster la tarification, mais également d'adapter les produits et services proposés. Par exemple, les assureurs pourraient inclure des garanties optionnelles couvrant les coûts d'évacuation temporaire, de relogement ou de remise en état rapide des bâtiments après une inondation. Ils pourraient également offrir des incitations, telles que des réductions de primes, aux propriétaires qui adoptent des mesures de prévention, comme le rehaussement des entrées ou l'utilisation de matériaux résistants aux inondations. Une étude a été réalisée en septembre 2023 par la CCR sur l'efficacité des plans de prévention des risques d'inondation sur le coût des dommages assurés (CCR 2023b). Depuis leur mise en place progressive en 1995, les PPRI auraient permis de réduire le coût des dommages assurés pour les biens des particuliers de 2,2 milliards d'euros entre 1995 et 2018. Grâce à ces dispositifs, la tendance des coûts liés aux inondations a pu être contenue. A l'image de la figure 3.1, cette étude montre qu'en France en 2050, le coût des dommages assurés dus aux inondations pourrait être réduit de 38% grâce au maintien et au développement de ces dispositifs. À l'inverse, leur absence entraînerait une augmentation estimée à 145% de ces coûts. L'enjeu, pour les assureurs, d'inciter à la prévention est donc extrêmement important.

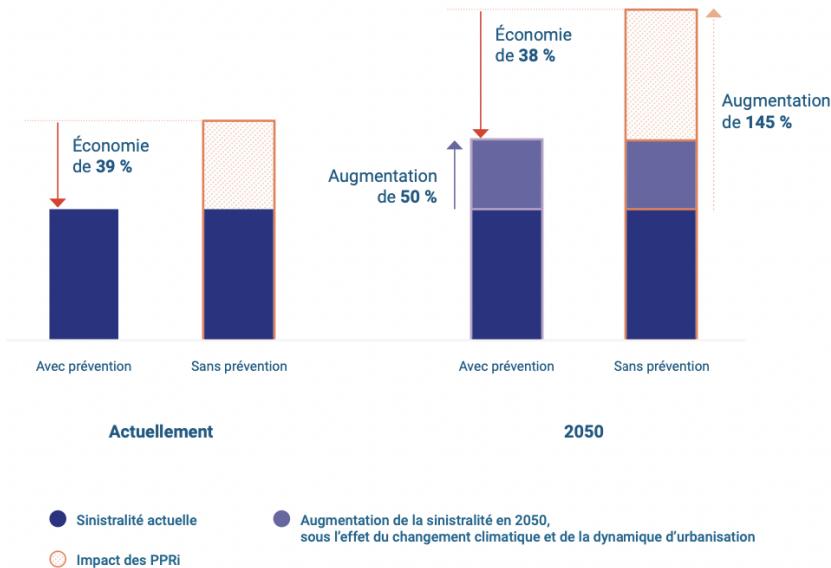


FIGURE 3.1 – Effet des PPRI sur le montant des dommages assurés, à climat actuel et en 2050.

Source :CCR (2023)

Cette cartographie des risques permettrait également aux compagnies d'assurance d'évaluer la répartition géographique de leurs portefeuilles à l'horizon 2030. Les assureurs pourraient ainsi réorienter leurs stratégies commerciales ou cibler des segments de clientèle moins exposés, afin d'optimiser la gestion globale des risques.

Modèles climatiques

Afin d'établir ces projections, la méthode nécessite l'utilisation de données issues de modèles climatiques.

D'après le GIEC (Groupe d'Experts Intergouvernemental sur l'Évolution du Climat), les modèles climatiques sont "des programmes informatiques extrêmement sophistiqués qui englobent notre compréhension du système climatique et simulent, avec autant de fidélité que possible, les interactions complexes entre l'atmosphère, l'océan, la surface terrestre, la neige et la glace, l'écosystème mondial et divers processus chimiques et biologiques". Ce sont des supercalculateurs, des ordinateurs extrêmement puissants conçus pour exécuter des calculs complexes à une vitesse très élevée, qui vont permettre la résolution des équations qui découlent des modèles. Pour ce faire, le globe terrestre est divisé en mailles tridimensionnelles (latitude, longitude, altitude) et le temps est segmenté en intervalles réguliers (figure 3.2). L'objectif est d'avoir à chaque point de cette grille, l'évolution des variables d'état des différentes composantes du système climatique telles que l'atmosphère, la surface terrestre, les océans et les glaces.

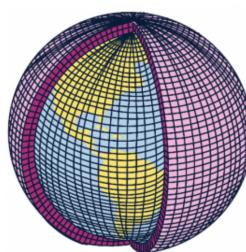


FIGURE 3.2 – Schéma simplifié du maillage du globe terrestre

Source :CNRS (2021)

Les premiers modèles informatiques capables de coupler l'évolution de l'atmosphère avec celle des océans ont vu le jour dans les années 1960 (DRIAS 2023). Ces avancées faisaient suite aux premières tentatives de modélisation météorologique dans les années 1950. Au fil des décennies, les modèles climatiques ont progressivement intégré des éléments supplémentaires, comme les sols, le cycle du carbone ou encore les interactions entre les écosystèmes, devenant ainsi de plus en plus complexes et sophistiqués.

Aujourd'hui, les simulations climatiques globales reposent principalement sur des **Modèles de Circulation Générale** (MCG, ou GCM en anglais pour *General Circulation Model*). Ces modèles numériques décrivent le climat à l'échelle mondiale avec une résolution horizontale qui varie généralement entre 100 et 500 km et une fréquence temporelle typique de 6 heures. Cependant, cette résolution globale peut s'avérer insuffisante pour étudier les phénomènes locaux ou régionaux, qui nécessitent une plus grande précision. De plus, la complexité des calculs et le grand nombre de points à traiter exigent des ressources informatiques considérables.

Pour étudier les impacts du changement climatique à des échelles plus fines, on utilise des **Modèles Climatiques Régionaux** (RCM, pour *Regional Climate Model*). Contrairement aux MCG, ces modèles couvrent uniquement une partie du globe, comme l'Europe, et offrent une résolution spatiale plus précise, variant entre 10 et 20 km. Cela permet une meilleure représentation des particularités locales, telles que le relief ou les contrastes terre-mer. Grâce à ces modèles régionaux, il devient possible d'affiner les prévisions et de mieux comprendre les conséquences locales du changement climatique.

Les modèles climatiques, ne sont pas exempts de limites. Ils sont notamment contraints par la puissance de calcul disponible, qui reste insuffisante pour représenter simultanément toutes les échelles spatiales et temporelles avec une précision optimale. Par ailleurs, ces modèles sont également limités par notre propre compréhension des processus et de leurs interactions. Cependant, leur fiabilité est rigoureusement testée et validée par des équipes scientifiques internationales au terme de nombreuses années de recherche. De plus, ils ont prouvé leur efficacité en reconstituant avec succès les climats anciens et récents. Cette capacité à reproduire des conditions passées renforce la confiance dans la fiabilité de leurs simulations pour le futur. Ces outils sont continuellement améliorés grâce aux efforts des centres de recherche à travers le monde. Parmi ces institutions, le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT) participe à l'amélioration des modèles, comme mentionné dans la section 1.2.3.

Ces modèles climatiques partagent de nombreuses similitudes avec les modèles de prévisions météorologiques : leurs formulations mathématiques et méthodes de calcul sont comparables, et ils utilisent souvent des outils logiciels communs. Cependant, il est essentiel de bien différencier ces deux approches. Les modèles météorologiques s'appuient sur les conditions initiales de l'atmosphère, telles que la température de l'air, la pression ou l'humidité à un instant donné (t), pour prédire l'évolution du temps sur les jours suivants. En revanche, les modèles climatiques se fondent sur des données météorologiques passées et actuelles couvrant l'ensemble des composantes du système climatique (atmosphère, océans, glaces, surfaces terrestres) afin de simuler des tendances à long terme. Ainsi, les concepts de météo et de climat renvoient non seulement à des processus de modélisation distincts, mais aussi à des échelles temporelles différentes. La météorologie se concentre sur des périodes courtes, de l'ordre de quelques jours ou semaines, tandis que les études climatiques se basent sur des moyennes établies sur des périodes d'au moins 30 ans.

3.1.2 Définition et choix du scénario

Pour produire des projections climatiques, les modèles climatiques utilisent en entrée des scénarios de **forçages radiatifs** prévus sur le siècle à venir.

Le forçage radiatif, également appelé forçage climatique, désigne la perturbation du bilan radiatif du système climatique (différence entre le rayonnement entrant et sortant) au sommet de la troposphère, causée par des facteurs externes. Il s'exprime en watts par mètre carré (W/m^2) (DRIAS 2013). On distingue classiquement deux types de forçages : naturels et anthropiques.

- **Les forçages naturels** incluent le forçage solaire, lié aux variations de l'activité solaire et aux fluctuations astronomiques de l'orbite terrestre, ainsi que le forçage volcanique, provoqué par les émissions de particules au cours des éruptions volcaniques. Ces particules rendent l'atmosphère moins transparente au rayonnement solaire, entraînant ainsi un refroidissement temporaire du système

climatique.

- Les **forçages anthropiques** résultent principalement des activités humaines, qui modifient la composition de l'atmosphère et l'utilisation des sols. Parmi ces forçages, les gaz à effet de serre (GES) occupent une place centrale. Essentiels au fonctionnement naturel du système climatique, ces gaz, comme le dioxyde de carbone ou le méthane, piègent une partie du rayonnement infrarouge émis et réémis par la surface terrestre. Cependant, les émissions dues aux activités humaines ont considérablement augmenté leur concentration dans l'atmosphère, générant un effet de serre additionnel responsable d'une grande partie du réchauffement climatique actuel. En parallèle, d'autres facteurs anthropiques influencent l'atmosphère. Par exemple, les émissions d'aérosols modifient sa transparence et affectent la formation des nuages. De plus, les changements dans l'utilisation des sols, comme la déforestation ou les pratiques agricoles, perturbent les échanges énergétiques entre la surface terrestre et l'atmosphère. Ces différentes contributions anthropiques, prises en compte dans les modèles climatiques, illustrent l'impact des activités humaines sur le bilan énergétique de la planète.

Un forçage radiatif positif contribue au réchauffement du système climatique, tandis qu'un forçage négatif engendre un refroidissement.

Les modèles climatiques, alimentés par des données de forçages produisent ensuite les projections climatiques selon les scénarios de forçage.

Les scénarios RCP

Le forçage radiatif passé est connu. Afin d'imaginer le futur du climat, des courbes d'évolution du forçage radiatif ont été caractérisées par les experts du climat et reprises par les rapports du GIEC.

Dans le 5eme rapport du GIEC, ce sont les scénarios RCP (Representative Concentration Pathway) qui ont été utilisés. Ce sont quatre scénarios de référence de l'évolution du forçage radiatif (tableau 3.1) sur la période 2006-2300.

Nom	Forçage radiatif
RCP8.5	$> 8.5 \text{ W.m}^{-2}$ en 2100
RCP6.0	$\sim 6 \text{ W.m}^{-2}$ au niveau de stabilisation après 2100
RCP4.5	$\sim 4.5 \text{ W.m}^{-2}$ au niveau de stabilisation après 2100
RCP2.6	Pic à $\sim 3 \text{ W.m}^{-2}$ avant 2100 puis déclin

TABLE 3.1 – Résumé des scénarios RCP selon le forçage radiatif et leur trajectoire.

Source :DRIAS (2021)

Les scénarios SSP

En complément des scénarios RCP, le sixième rapport du GIEC a introduit un nouvel ensemble de scénarios, appelés SSP (Shared Socioeconomic Pathways). Ces scénarios, conçus pour enrichir l'analyse des trajectoires climatiques, décrivent différents contextes socio-économiques susceptibles d'influencer les émissions de gaz à effet de serre et l'évolution du climat. Contrairement aux RCP, qui se concentrent principalement sur les trajectoires de concentration des gaz à effet de serre, les SSP mettent en lumière les interactions entre les dynamiques sociales, économiques et environnementales.

Ce sont 5 scénarios principaux qui ont été utilisés par le GIEC (figure 3.3) :

- Le **SSP1** et le **SSP5** représentent des trajectoires optimistes, marquées par une amélioration significative des conditions de vie, une croissance économique rapide et des institutions efficaces. Toutefois, le SSP1 favorise des pratiques durables et une transition vers des énergies propres, tandis que le SSP5 repose sur une exploitation intensive des énergies fossiles, entraînant des impacts environnementaux importants malgré les progrès économiques.
- Le **SSP2**, qualifié de scénario intermédiaire, reflète une évolution socio-économique inégale, avec des disparités marquées entre les pays. La démographie mondiale se stabilise, mais les systèmes environnementaux subissent une dégradation modérée.
- Les **SSP3** et **SSP4** illustrent des trajectoires plus pessimistes, marquées par un faible investissement dans la santé et l'éducation ainsi qu'une vulnérabilité accrue au changement climatique. Le SSP3

se distingue par une priorité donnée à la sécurité régionale, limitant la coopération internationale, tandis que le SSP4 se caractérise par des inégalités sociales et économiques importantes, amplifiant les disparités tant entre, qu'au sein des pays.

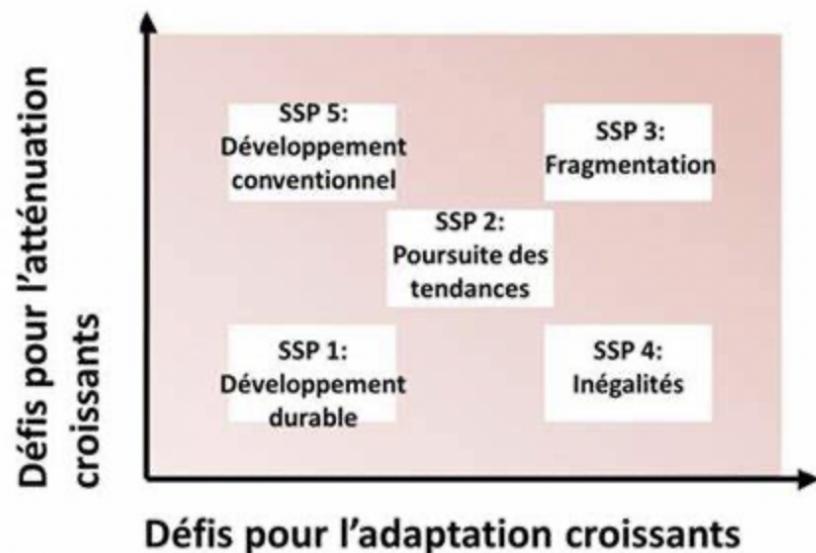


FIGURE 3.3 – Les 5 grandes trajectoires socio-économiques des SSP

Source : *DRIAS (2021)*

Les projections climatiques sont exprimées en fonction de ces scénarios, sous la forme « SSPx-y », où *x* correspond au numéro du SSP et *y* indique le niveau approximatif de forçage radiatif prévu en 2100.

Ainsi dans son sixième rapport, le GIEC évalue la réponse du climat à 5 scénarios combinant l'activité humaine et les facteurs externes. Ces cinq scénarios retenus par le GIEC sont (figure 3.4) :

- SSP1-1.9
- SSP1-2.6
- SSP2-4.5
- SSP3-7.0
- SSP5-8.5

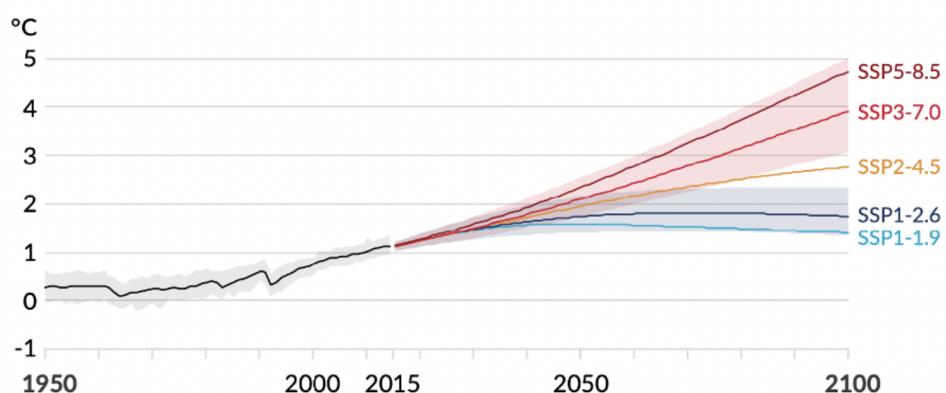


FIGURE 3.4 – Trajectoires de réchauffement planétaire selon les cinq scénarios SSPx-y retenus dans le résumé pour décideur du GIEC

Source : *Carbone4 (2022)*

Les scénarios socio-économiques (SSP) et climatiques (RCP) sont conçus séparément, ce qui permet de les combiner de manière flexible. Une approche en matrice relie les trajectoires socio-économiques, comme le développement durable ou l'exploitation intensive des énergies fossiles, aux niveaux de forçage radiatif définis par les RCP. Ainsi, pour un même niveau de réchauffement climatique, les efforts ou coûts nécessaires pour réduire les émissions peuvent varier considérablement en fonction du contexte socio-économique. A l'image de la figure 3.5, chaque scénario RCP peut donc être associé à plusieurs SSP, car un niveau donné d'émission de gaz à effet de serre peut résulter de différents types de développement socio-économiques.

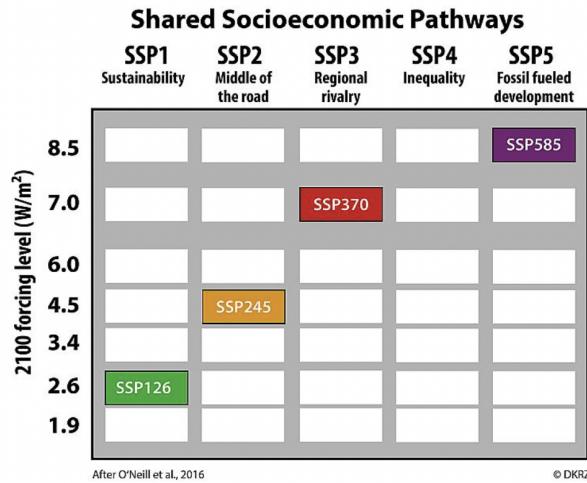


FIGURE 3.5 – Matrice des scénarios RCP-SSP

Source :DRIAS (2021)

Pour cette étude, le scénario SSP2-4.5 a été retenu comme base d'analyse. Ce scénario, souvent qualifié de « trajectoire intermédiaire », suppose une évolution socio-économique modérée, sans changements drastiques ni vers une durabilité accrue ni vers un développement fortement inégal ou orienté sur les énergies fossiles. Ce scénario reflète une trajectoire plausible qui évite les extrêmes, comme les hypothèses irréalistes de décarbonation rapide (SSP1-1.9), ou les scénarios de forte intensité carbone (SSP5-8.5), souvent critiqués pour leur caractère alarmant. Par ailleurs, ce scénario semble refléter la trajectoire globale actuelle observée depuis 2015 (figure 3.6).

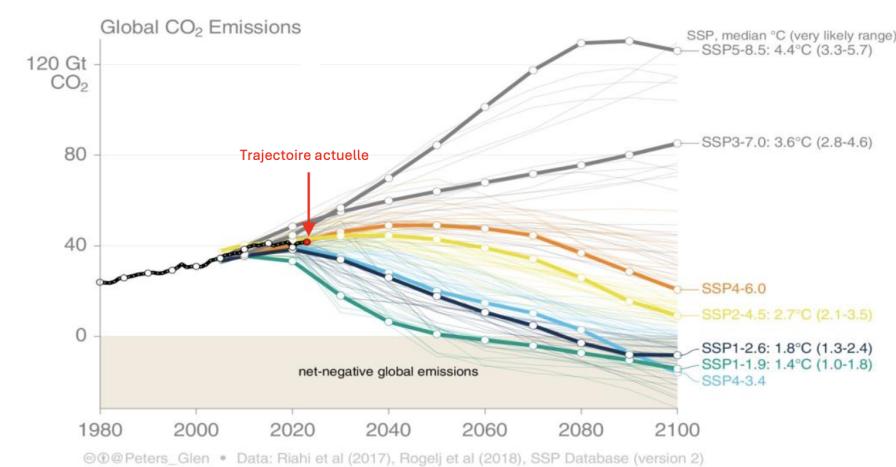


FIGURE 3.6 – Trajectoires des différents scénarios

Source :IPCC-AR6-TS

Des études, notamment celles issues du sixième rapport du GIEC, positionnent le SSP2-4.5 comme un compromis pertinent pour évaluer les impacts climatiques à moyen terme. Étant donné que l'horizon temporel

de nos projections est fixé à 2030, ce scénario s'impose comme une base crédible et réaliste pour analyser les évolutions climatiques et leurs implications.

3.1.3 Présentation des données de projection

Données issues du projet CMIP6

Comme décrit en partie 1.2.3, le service Copernicus sur le changement climatique (mis en œuvre par le Centre européen pour les prévisions météorologiques à moyen terme au nom de la Commission européenne) met à disposition un vaste catalogue de données projetées en open data.

Parmi les variables disponibles, celle qui nous intéresse est la variable "Total précipitation", qui représente la profondeur cumulée d'eau au sol (en mètres) sur une période donnée. Ces données sont disponibles au format NetCDF4 pour la période de 2015 à 2100, selon les différents scénarios SSP définis dans le cadre du CMIP6. Les séries peuvent être téléchargées à des échelles temporelles quotidienne ou mensuelle. Dans le cadre de cette étude, nous téléchargeons spécifiquement les séries quotidiennes pour le scénario SSP2-4.5.

Ces données s'appuie sur plus d'une centaines de modèles provenant de plus de 50 centres de modélisations climatiques, produits dans le cadre de l'exercice international CMIP6 (*Coupled Model Inter-comparison Project Phase 6*). Le Projet d'inter-comparaison de modèles couplés (*CMIP*) est une initiative coordonnée par le Programme mondial de recherche sur le climat des Nations Unies. Les données du CMIP6 sont les plus récentes données de modèles climatiques mondiaux disponibles. Reconnues pour leur robustesse scientifique, elles constituent la base des rapports du GIEC (CNRS, CEA ET MÉTÉOFRANCE 2019).

Cependant, une contrainte majeure des données issues de CMIP6 est leur échelle spatiale globale. Aucune régionalisation n'est directement disponible dans ces produits, ce qui rend nécessaire un travail supplémentaire de descente d'échelle. Cette étape consiste à adapter les projections globales à des résolutions plus fines, mieux adaptées aux spécificités régionales. Pour cette étude, les données régionalisées étaient disponibles dans la base de données interne de Addactis.

Le choix d'utiliser cette source de données s'inscrit dans une démarche de continuité méthodologique. Bien que ERA5-Land et CMIP6 reposent sur des approches différentes (réanalyse vs. projection), l'objectif était d'avoir la même variable entre le passé et le projeté pour une équivalence entre les analyses passées et projetées.

Données complémentaires

Les données climatiques additionnelles intégrées aux modèles historiques de fréquence sont également disponibles sous forme projetée. Ces données proviennent d'une base interne développée par Addactis, construite à partir d'une combinaison d'analyses et de réanalyses issues de multiples sources et modèles climatiques. Leur valeur projetée a donc été ajouté à la base de modélisation pour permettre la mise à jour des modèles.

3.2 Méthode de projection

Cette section vise à détailler les différentes étapes mises en place pour obtenir une prime pure projetée pour le risque inondation en France par bâtiment.

Rappel de la formule

Dans un premier temps, nous rappelons la formule 2.1, utilisée pour calculer les primes pures climatiques historiques pour chaque bâtiment :

$$PP_{\text{climatique}} = \sum_i P(\text{intensité}_i) \times \mathbb{E}(\text{fréquence}_{\text{sinistres}}) \times \mathbb{E}(\text{coût}_{\text{sinistres}} \mid \text{inondation}_{\text{intensité}_i}) \quad (3.1)$$

Afin de déterminer une prime pure projetée, les différents éléments de cette formule seront projetés séparément, puis agrégés à l'aide de la même structure. Nous allons donc projeter la fréquence des sinistres

($\mathbb{E}(\text{fréquence}_{\text{sinistres}})$) à l'aide des données complémentaires, puis déterminer la probabilité future d'intensité ($P(\text{intensité}_i)$).

Projection de la fréquence

On rappelle que, contrairement au coût moyen, la fréquence des sinistres n'est pas conditionnée selon l'intensité des événements, car il est impossible d'associer une intensité de précipitation historique à un contrat qui s'étend sur toute une période. La projection ne peut donc pas se faire au niveau des probabilités d'intensité.

Pour palier cette contrainte, nous avons intégré dans le modèle de fréquence construit en partie 1, trois variables climatiques disponibles aussi bien en version historique qu'en version projetée dans la base de données interne à Addactis. En utilisant leur valeur projetée, nous pourrons mettre à jour le modèle et ainsi obtenir une projection. Ces variables, essentielles à la projection de la fréquence des sinistres, sont les suivantes :

- Le nombre annuel de jours avec des précipitations intenses.
- Le nombre maximal de jours consécutifs par an avec des précipitations quotidiennes d'au moins 1 mm.
- Les précipitations maximales sur une journée.

Elles ont été extraites pour la période 2015 à 2045, conformément au scénario climatique SSP2-4.5. Les mêmes traitements appliqués aux données historiques ont été répliqués sur les nouvelles variables projetées afin de garantir la cohérence des analyses.

Une fois ces ajustements réalisés, le modèle de fréquence est appliqué à la nouvelle base de modélisation. Cette base intègre les caractéristiques spécifiques de chaque bâtiment et remplace les données climatiques historiques par leur version projetée. À l'issue de cette étape, pour chaque bâtiment, une variable notée *Freq_proj* est obtenue, qui représente la projection de la fréquence du risque d'inondation associée au bâtiment.

Projection des probabilités d'intensités

Dans le chapitre 1, nous avons construit une variable *cumul_precip*, qui sert de base pour définir les probabilités d'intensités. La méthodologie permettant d'obtenir cette variable et les probabilités associées est détaillée dans les parties 1.2.3 et 2.1.2.

Précédemment, des probabilités d'intensité ont été attribuées en fonction des cumuls de précipitations historiques. Par exemple, pour un bâtiment donné, une intensité 1 correspond à un cumul de précipitations compris entre le quantile 0 et le quantile 75 de la répartition historique des cumuls de précipitations enregistrés sur la période 1985-2015 à l'endroit de ce bâtiment. La correspondance entre les intensités, les quantiles de *cumul_precip* et les probabilités associées est donnée par la table suivante (Table 3.2) :

intensité	<i>cumul_precip</i> discrétisée	Probabilité associée
1	[0,0.75[0.75
2	[0.75,0.85[0.10
3	[0.85,0.95[0.10
4	[0.95,0.97[0.02
5	[0.97,0.98[0.01
6	[0.98,0.999[0.019
7	[0.999,1[0.001

TABLE 3.2 – Description des différentes intensités passées

Pour projeter les probabilités d'intensité dans le futur, nous analysons la distribution de la variable *cumul_precip* basées sur le scénario SSP2-4.5. L'objectif est d'observer comment les périodes de retour associées aux intensités évoluent dans le futur et d'ajuster les probabilités en conséquence.

La Figure 3.7 compare les répartitions de la variable *cumul_precip* pour les périodes passées (1985-2015) et futures (2015-2045), pour un bâtiment donné. On observe clairement un décalage de la distribution projetée vers des valeurs plus élevées, indiquant une augmentation des cumuls de précipitations attendus. Par exemple, un événement correspondant dans le passé à une période de retour de 7 ans est projeté pour se produire environ tous les 5 ans dans le futur.

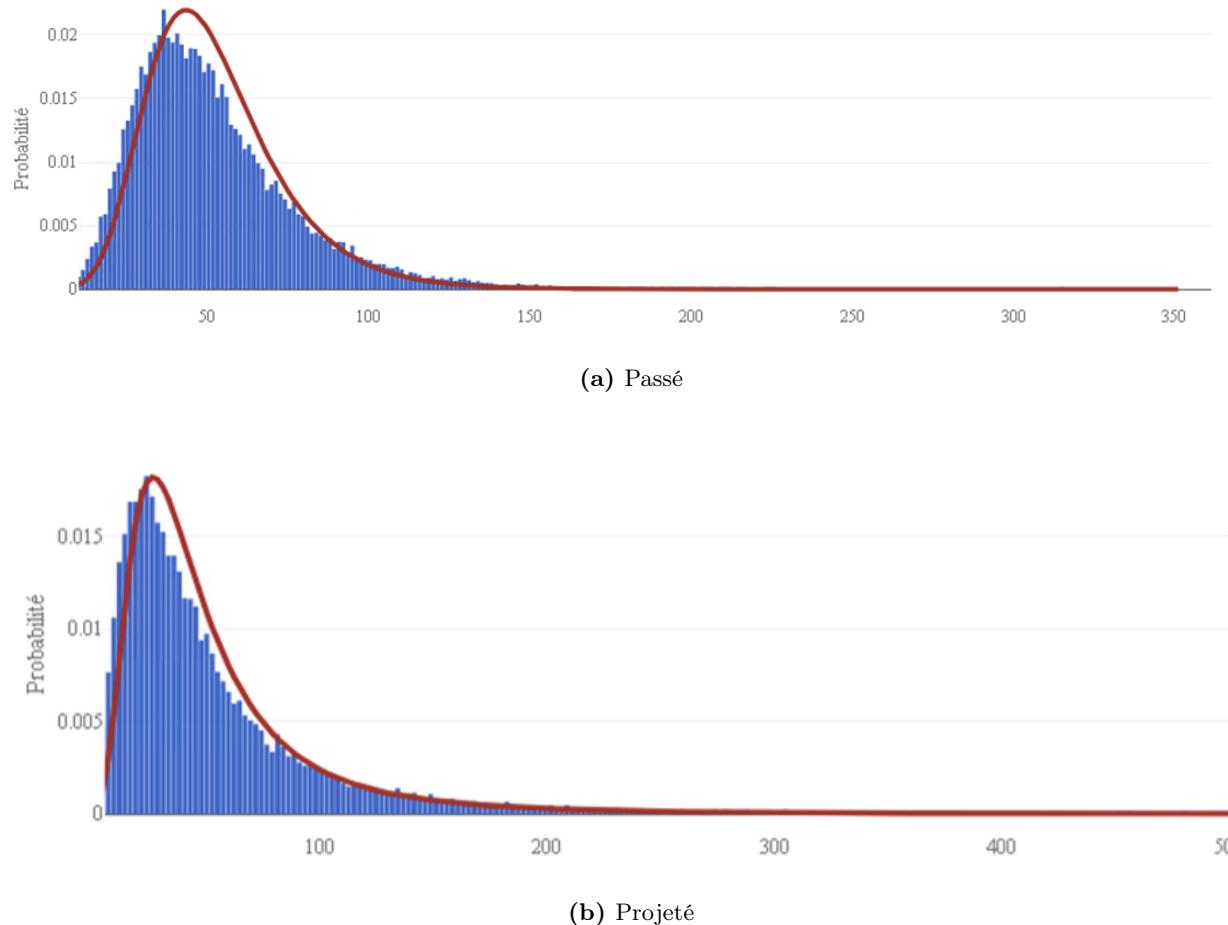


FIGURE 3.7 – Comparaison des répartitions de la variable *cumul_precip* passée et future.

Nous procédons ensuite à une comparaison des distributions passées et projetées de *cumul_precip* pour chaque bâtiment. Par exemple, dans le passé, un quantile 75 correspondant à un cumul de 100 mm de précipitations peut désormais, dans la distribution projetée, correspondre à un quantile 73. Cela signifie que la probabilité associée à la première intensité (*cumul_precip* entre le quantile 0 et le quantile 75) devra être ajustée à 0,73 pour ce bâtiment.

Une fois les ajustements effectués, nous attribuons pour chaque bâtiment une nouvelle série de probabilités d'intensité, réparties sur les 7 classes d'intensité définies. Ces probabilités sont stockées sous forme de 7 colonnes dans la base de données, où chaque ligne représente un bâtiment (Tableau 3.3). Ces nouvelles probabilités reflètent l'évolution des risques liés aux précipitations.

Bâtiment	...	P0_1	P0_2	P0_3	P0_4	P0_5	P0_6	P0_7
1	...	0.7369	0.1070	0.1052	0.0195	0.0102	0.0196	0.0013
2	...	0.7335	0.1056	0.077	0.0198	0.0104	0.0197	0.0011
3	...	0.7337	0.1024	0.1083	0.0202	0.0108	0.0196	0.0012
...

TABLE 3.3 – Exemple des probabilités projetées pour chaque bâtiment, réparties sur les 7 classes d'intensité.

Les probabilités projetées, regroupées par classe d'intensité, révèlent des tendances claires quant à l'évolution des risques liés aux précipitations. On observe une diminution des probabilités associées à la plus faible intensité (*P0_1*) et une augmentation des probabilités correspondant aux intensités élevées.

Ces ajustements reflètent le déplacement des distributions climatiques futures vers des événements plus extrêmes, comme observé dans les analyses précédentes. Par exemple, pour le bâtiment 1, la probabilité d'observer un cumul de précipitation d'intensité 1 est passée à 0.7369, alors que les probabilités d'avoir des intensités plus élevées ($P0_6$ et $P0_7$) augmentent légèrement, atteignant respectivement 0.0196 et 0.0013.

Cela signifie qu'à l'échelle des bâtiments, les risques d'événements climatiques rares et destructeurs sont en hausse, ce qui s'accorde avec les études menées par la CCR et le livre blanc Covéa cité précédemment.

Obtention de la prime pure projetée

De manière analogue à la méthodologie décrite en partie 2.4 pour le calcul de la prime pure historique, la prime pure projetée pour chaque bâtiment est calculée en intégrant les coûts moyens et les nouvelles probabilités d'intensité obtenues.

Pour chaque bâtiment, les coûts moyens calculés pour chaque intensités, $pred_cm_i$, (où i représente les intensités de 1 à 7) et les probabilités projetées associées $P0_i$ sont utilisés. Le coût moyen global projeté est alors obtenu de la façon suivante :

$$\text{Coût moyen projeté (CM_proj)} = \sum_{i=1}^7 pred_cm_i \times P0_i$$

Enfin la prime pure projetée pour le risque inondation pour un bâtiment est donnée par :

$$\text{Prime pure climatique projetée} = CM_proj \times Freq_proj$$

À l'issue de ce calcul, chaque bâtiment se voit attribuer une prime pure climatique projetée, reflétant les risques d'inondation prévus (Tableau 3.4).

Id Bâtiment	...	<i>Freq_proj</i>	<i>cm_pred_1</i>	<i>P0_1</i>	...	<i>cm_pred_7</i>	<i>P0_7</i>	<i>CM_proj</i>	<i>PP_proj</i>
1	...	0.0085	1246.97	0.7369	...	2204.38	0.0013	1370.56	11.65
2	...	0.0004	2005.67	0.7335	...	3545.20	0.0011	2204.10	0.88
3	...	0.0041	5169.10	0.7337	...	8311.02	0.0012	5100.10	20.91
4	...	0.0017	780.96	0.7300	...	1254.51	0.0017	857.72	1.46
...

TABLE 3.4 – Synthèse des données pour le calcul de la prime pure climatique projetée

Calcul de l'évolution par bâtiment

Pour chaque bâtiment du portefeuille, nous disposons d'une prime pure historique et d'une prime pure projetée pour le risque d'inondation.

En analysant les primes pures projetées brutes, l'assureur pourra identifier les bâtiments présentant les primes les plus élevées à l'horizon 2030. Cela permettra de repérer les bâtiments les plus vulnérables et les zones les plus exposées au risque d'inondation. Cependant, il est aussi important d'avoir un regard sur les variations entre passé et futur. L'objectif de cette étape est donc d'évaluer l'augmentation ou la diminution prévue de la prime pure pour chaque bâtiment à horizon 2030. Cette analyse permettra de quantifier l'ampleur des hausses ou des baisses prévues pour chaque bâtiment en fonction des résultats des projections climatiques.

L'évolution de la prime pure est calculée selon la formule suivante :

$$\text{Evolution Prime Pure inondation} = Evol_PP = \frac{PP_{projete} - PP_{historique}}{PP_{historique}}$$

Calcul de l'évolution par commune de France métropolitaine

Afin de compléter l'analyse, nous souhaitions obtenir une évolution à l'échelle communale pour l'ensemble de la France métropolitaine. Pour cela, nous constituons une nouvelle base, comprenant une ligne par commune. En effet, cette nouvelle manipulation permettra d'avoir une vision des évolutions sur tout le territoire en incluant les zones dans lesquelles il n'y a pas de bâtiments assurés par l'assureur, offrant ainsi une couverture exhaustive.

Pour chaque commune, un bâtiment type unique est défini, c'est à dire que pour toutes les communes de la base, les caractéristiques des bâtiments sont strictement identiques. Ces caractéristiques sont choisies de manière cohérente afin de refléter une situation moyenne par commune. Les données climatiques utilisées, en revanche, varient en fonction de la position géographique de chaque commune.

En appliquant les méthodes de calcul des primes pures historiques et projetées à cette base, il est possible d'estimer les évolutions dues exclusivement au changement climatique. Les résultats obtenus sont ensuite cartographiés, fournissant une représentation des variations des primes pures à l'horizon 2030 pour toutes les communes de France métropolitaine.

Évolution du coût des sinistres

Dans cette étude, l'analyse des projections a été réalisée sans prise en compte de l'évolution des coûts des sinistres. Toutefois, il serait possible d'intégrer l'inflation en ajustant les coûts moyens obtenus grâce aux projections de l'indice des prix à la consommation réalisées par la Banque de France. Cela permettra de refléter plus fidèlement leur évolution dans le temps. Nous avions utilisé l'indice FFB du coût de la construction afin d'homogénéiser les coûts des sinistres sur la période historique. Cet indice est fortement corrélé à l'inflation annuelle, comme l'illustre la figure 3.8.

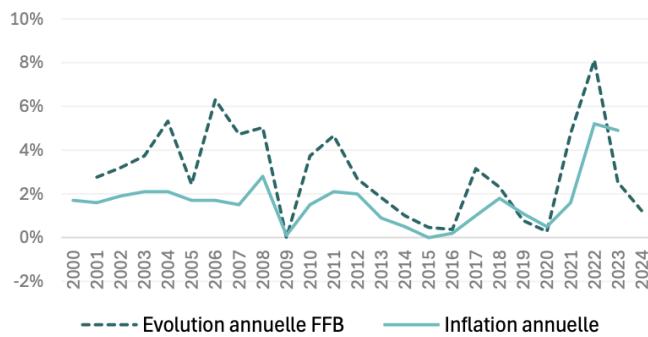


FIGURE 3.8 – Historique des indices FFB du coût de la construction et de l'inflation annuelle

Source :*Banque de France et Fédération Française du bâtiment*

Sur la base de ces éléments, un ajustement des coûts des sinistres peut être appliqué pour tenir compte de l'évolution des prix (tableau 3.5).

Année	Inflation annuelle
2024	2,5%
2025	1,7%
2026	1,7%
2027	1,7%
2028	1,7%
2029	1,7%
2030	1,7%
Ajustement 2030	+13,4%

TABLE 3.5 – Taux d'inflation annuelle et ajustement 2030

Source :*Banque de France : Projections macroéconomiques – Juin 2024*

3.3 Analyse et interprétation des résultats

3.3.1 Analyse des résultats à l'échelle nationale

Sur l'ensemble de la France métropolitaine, l'évolution globale des primes pures est marquée par une légère augmentation moyenne, cela témoigne de l'impact du réchauffement climatique sur le risque inondation. Cette hausse reflète la projection d'une intensification des événements extrêmes liés aux précipitations.

La cartographie des évolutions globales à l'échelle communale permet de distinguer certaines tendances territoriales.

Les résultats indiquent que les évolutions ne sont pas uniformes et sont fortement influencées par les facteurs géographiques. Certaines zones hydrographiques spécifiques, par exemple, présentent des augmentations plus marquées. Par ailleurs, à travers la cartographie des évolutions, on distingue également la démarcation de certaines régions climatiques, ce qui reflète leur rôle dans la répartition des variations sur le territoire français. Ce sont les régions où les températures hivernales sont historiquement basses, ainsi que les zones du pourtour méditerranéen qui enregistrent les augmentations les plus fortes des primes pures pour le risque inondation. Cela pourrait s'expliquer par une augmentation prévue des précipitations extrêmes en hiver et des épisodes méditerranéens plus fréquents et intenses. En revanche, certaines zones, bien que minoritaires, voient une diminution projetée de leurs primes pures. Ces baisses sont particulièrement concentrées dans le nord de la France, où les conditions climatiques futures prévoient une possible diminution des précipitations intenses.

Au global, on constate une évolution moyenne des primes pures en France métropolitaine pour le risque inondation à horizon 2030 de **+0,7%**. Une augmentation qui peut paraître modérée, mais qui témoigne néanmoins de l'impact progressif du réchauffement climatique et qui dissimule des effets très variés sur le territoire. De plus, cette évolution moyenne est projetée à l'horizon 2030, mais pourrait s'accentuer au-delà en fonction de l'intensité du réchauffement climatique et des mesures d'adaptation mises en place.

3.3.2 Analyse des résultats à l'échelle régionale - Focus sur le Grand Est

Afin d'affiner l'analyse des évolutions des primes pures, nous nous concentrons sur une région spécifique : le Grand Est. Cette région se distingue par sa diversité climatique et géographique, mais aussi par son conséquent réseau de cours d'eau, qui la rend particulièrement vulnérable aux risques d'inondation.

Le Grand Est regroupe 10 départements aux conditions climatiques variées et est traversé par plusieurs rivières ainsi que par cinq grands fleuves (figure 3.9) : le Rhin, la Meuse, la Moselle, la Marne et la Seine.



FIGURE 3.9 – Grands Est - départements et grands cours d'eau

Source :Régions et départements français

En raison de sa configuration, le Grand Est est souvent confronté à des crues majeures, en particulier pour les bassins fluviaux à forte densité urbaine. Toutes ces caractéristiques font donc du Grand Est, un terrain intéressant pour analyser la façon dont ces facteurs influencent l'évolution des primes pures.

Au global, la région Grand Est suit la tendance nationale avec une hausse projetée des risques inondations. Cependant, les résultats obtenus prévoient une hausse moyenne de **6%** pour cette région, nettement supérieure à la moyenne nationale. En effet, tous les départements présentent une évolution projetée positive comprise entre 0 et 10%. Cette région fait donc partie des régions de France métropolitaine qui seront les plus influencées par le réchauffement climatique en terme d'inondation.

La cartographie des évolutions des primes pures pour le risque inondation dans la région Grand Est est présentée en figure 3.10 :

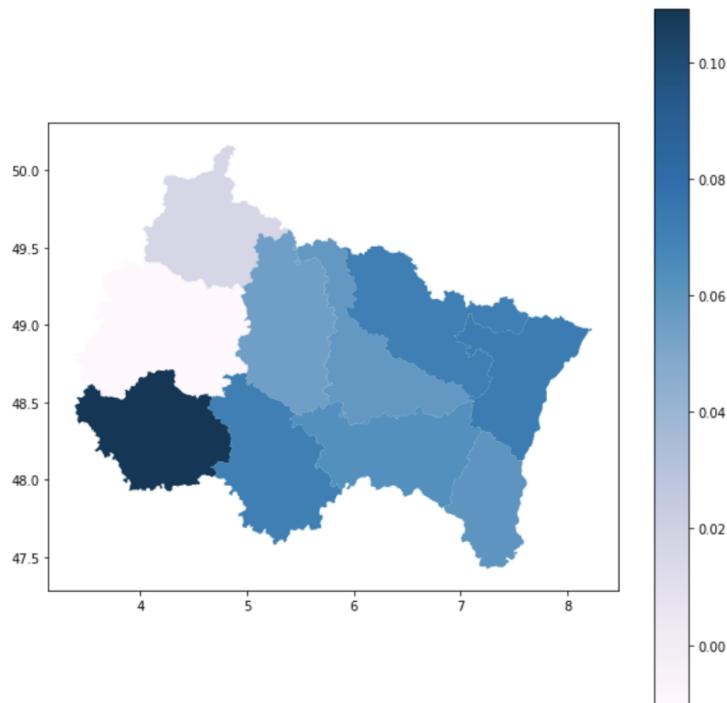


FIGURE 3.10 – Cartographies des évolutions moyennes de prime pure pour le risque inondation par département du Grand Est

Le département de l'Aube apparaît comme le plus impacté par le risque d'inondation à l'horizon 2030, avec une augmentation estimée d'environ 10% des primes pures. Ce constat peut s'expliquer d'une part, par le fait que ce département a historiquement été plutôt épargné par les événements majeurs malgré son relief plat et ses nombreux épisodes de pluies intenses. En effet, comme illustré par la figure 3.11, l'Aube est le département du Grand Est qui enregistre le moins d'arrêts catastrophes naturelles inondation depuis 1992. Ainsi cette faible fréquence d'événement majeur conduit paradoxalement à une forte augmentation de la prime pure en 2030. D'autre part, cette forte hausse peut aussi être expliquée, par la présence de la Seine au centre du département dont les crues pourraient être fortement impactées par le réchauffement climatique.

La Marne, en revanche apparaît comme le département le moins influencé par le changement climatique à horizon 2030. Avec une évolution proche de 0%, le département des Ardennes semble également très peu concerné par un potentiel changement de prime. Une première explication peut être liée à la mise en place, dès les années 2000, de nombreuses réglementations en faveur de la prévention des inondations dans ces zones (PPRi, normes de construction...). Par ailleurs, on peut supposer une évolution stable des conditions climatiques dans ces départements également.

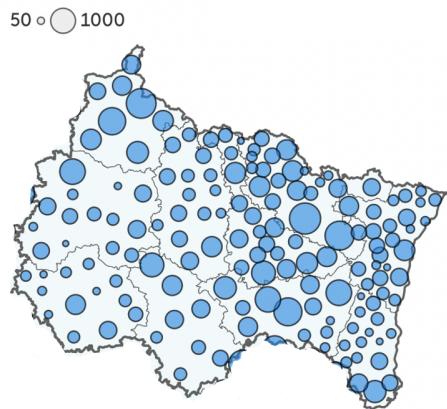


FIGURE 3.11 – Nombre d'arrêtés catastrophe naturelle pour l'inondation dans le Grand Est

Source :*FranceInfo - Grand Est (2024)*

3.4 Limites et perspectives

Les analyses menées ont permis d'obtenir une estimation fiable de la prime pure pour le risque inondation à l'horizon 2030. Toutefois, certaines limites méritent d'être soulignées, et leur prise en compte permettrait d'affiner davantage l'évaluation.

Données météorologiques et au bâtiment

Les données météorologiques utilisées dans la mise en place des modèles proviennent de la base de réanalyse ERA5-Land. Elles peuvent présenter des écarts par rapport aux observations, notamment en raison des incertitudes inhérentes aux modèles de réanalyse. De plus, ces modèles sont susceptibles de lisser ou de corriger à tort certaines valeurs extrêmes, alors que ces dernières seraient utiles pour une évaluation précise du risque inondation.

De la même manière, les données relatives aux bâtiments restent perfectibles. Elles sont observées à un instant donné, ce qui peut limiter leur pertinence pour une projection. Toutefois, à horizon court terme, l'évolution du bâti reste marginale et n'affecte que faiblement l'analyse. Par ailleurs, cette base de données fait l'objet d'améliorations continues et est régulièrement enrichie par de nouvelles informations fournies par l'entité NamR.

Méthode de projection

Bien que la méthode de projection mise en place ait permis d'obtenir une première approximation de l'évolution du risque en France métropolitaine, elle reste une simplification de la réalité. Elle constitue une représentation partielle des travaux en cours au sein de la pratique, visant à affiner l'estimation du risque futur d'inondation.

Choix du modèle climatique

Le choix du modèle climatique utilisé dans les projections constitue une source supplémentaire d'incertitude dans l'estimation du risque inondation. Chaque modèle repose sur des hypothèses et des paramétrages spécifiques qui influencent la représentation des phénomènes climatiques extrêmes. Ainsi, le couple GCM/RCM sélectionné peut impacter les résultats obtenus (PLATEL 2022).

Dans cette étude, un modèle a été utilisé pour réaliser les projections. Cependant, le fait que l'horizon de projection soit court, permet de diminuer la variabilité des résultats entre les modèles. Bien que cette approche permette d'obtenir des résultats cohérents et exploitables, une alternative aurait été de travailler avec un ensemble de modèles afin d'évaluer la dispersion des projections. Cela pourrait offrir une vision plus complète des impacts du changement climatique.

Dans une logique de continuité, il serait intéressant de réaliser ce même travail de projection en intégrant plusieurs scénarios climatiques afin d'évaluer l'impact des différentes trajectoires d'émissions de gaz à effet de serre sur le risque inondation. Cela permettrait de comparer les résultats selon des hypothèses plus ou moins optimistes quant à l'évolution du climat.

Evolution du nombre de bâtiments et de la répartition des populations sur le territoire

Afin d'affiner l'étude, il serait intéressant de prendre en compte l'évolution du nombre de bâtiments ainsi que leur répartition spatiale sur le territoire. En effet, l'urbanisation croissante influence directement la localisation des nouvelles constructions, entraînant une concentration accrue dans certaines zones potentiellement exposées aux aléas climatiques.

À court terme, l'impact de l'évolution démographique reste relativement limité. Cependant, dans une perspective à plus long terme, notamment à l'horizon 2100, ces changements pourraient avoir des répercussions plus significatives.

Conclusion

Avec l'intensification des phénomènes climatiques extrêmes, il devient essentiel d'adapter les modèles assurantiels et d'étudier l'impact du changement climatique sur les sinistres futurs. Les inondations, représentant plus de la moitié des arrêtés Cat Nat en France, suscitent une inquiétude particulière.

Traditionnellement, la tarification en multirisques habitation, repose uniquement sur les données de sinistralité détenues par les assureurs. Dans ce mémoire, nous avons enrichi cette tarification en intégrant des données innovantes supplémentaires. La méthodologie adoptée repose ainsi sur l'exploitation des données de sinistralité fournies par l'assureur mais également sur l'exploitation de données complémentaires au bâtiment et de données météorologiques. Les données complémentaires au bâtiment, ont permis d'affiner l'estimation de la prime pure en prenant en compte les spécificités structurelles et environnementales de chaque bâtiment. Par ailleurs, les données météorologiques ont permis d'une part de projeter le risque inondation et d'autre part de mieux appréhender le phénomène. En effet, l'une des principales difficultés liées au risque inondation, réside dans la complexité du comportement de l'eau après son arrivée au sol et dans le décalage temporel entre les précipitations et la survenue de l'inondation. Pour répondre à cette problématique, un indice de précipitations adapté a été élaboré, basé sur les cumuls de précipitations journaliers. Cet indice a rendu possible la capture des différentes intensités de précipitations.

Une fois la prime pure modélisée, celle-ci a été comparée à d'autres modèles ainsi qu'à une tarification fondée uniquement sur les données de sinistralité détenues par l'assureur. Les résultats obtenus se révèlent satisfaisants et montrent que l'intégration de données complémentaires permet d'affiner significativement l'estimation du risque.

L'objectif final de cette étude était de projeter les primes pures obtenues à horizon 2030 afin d'anticiper leur évolution sous l'effet du changement climatique. Cette projection, rendue possible grâce à l'intégration de données climatiques, est basée sur le scénario SSP2-4.5, qui semble être la trajectoire actuellement suivie. Les résultats mettent en évidence une hausse moyenne de +0,8% en France métropolitaine, avec des disparités régionales marquées.

Toutefois, ces résultats doivent être interprétés avec prudence en raison des limites inhérentes aux données et aux modèles utilisés. De nombreuses pistes d'amélioration restent à étudier pour affiner davantage les projections.

Finalement, cette approche pourrait permettre aux assureurs MRH de mieux appréhender le risque inondation en intégrant de nouvelles informations à la modélisation des primes pures, et offre des perspectives intéressantes pour adapter la tarification face aux défis du changement climatique.

Annexes

Mean Excess Plot

La mean excess function est un outil statistique permettant d'analyser le comportement des valeurs extrêmes d'une distribution. Elle est particulièrement utile pour identifier un seuil pertinent d'écrêtage des sinistres en analysant la distribution des valeurs extrêmes. Elle se définit comme suit :

Soit X une variable aléatoire de fonction de répartition F et de point terminal x_F . Pour tout $u < x_F$, on définit la fonction

$$F_u(x) = P[X - u \leq x \mid X > u], \quad x \geq 0$$

appelée *fonction de répartition des excès au-dessus du seuil u* (BORCHANI 2010).

Le *Mean Excess Plot* est le graphe des points $(u, e(u))$, où $e(u)$ représente la moyenne des excès au-delà du seuil u , définie par :

$$e(u) = \mathbb{E}[X - u \mid X > u]$$

Pour estimer cette fonction empiriquement, on utilise l'estimateur $\hat{e}(u)$, défini par :

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

où $(X_i - u)^+$ représente l'excès positif de X_i au-dessus du seuil u et $\mathbb{1}_{\{X_i > u\}}$ est une fonction indicatrice valant 1 si $X_i > u$, et 0 sinon.

Cette courbe permet d'identifier le comportement des valeurs extrêmes :

- Si $e(u)$ est constante, la distribution des sinistres suit une loi exponentielle.
- Si $e(u)$ est croissante de manière linéaire, la distribution des sinistres suit une loi de Pareto généralisée (GPD), caractérisant une queue lourde (ALEXANDRE GODZINSKI 2007).

Le seuil optimal u_0 est choisi en identifiant le point où $e(u)$ présente un changement de tendance, passant d'une phase constante à une croissance linéaire.

En traçant la courbe de $e(u)$ pour notre ensemble de sinistres, nous avons observé une croissance marquée à partir de 75 000 €, suggérant que les sinistres supérieurs à ce seuil ont un impact significatif sur la variance du modèle.

Hill Plot

La courbe de Hill est un outil graphique utilisé en théorie des valeurs extrêmes. Elle permet d'estimer l'indice de queue ξ et de déterminer un seuil optimal en exploitant la propriété de stabilité de la loi de Pareto généralisée (GPD). En effet, si la variable $[X - u \mid X > u] \sim \text{GPD}(\xi, \sigma)$, alors pour tout $u' > u$, on a également $[X - u' \mid X > u'] \sim \text{GPD}(\xi, \sigma')$. Cela signifie que le paramètre de queue ξ reste constant au-delà d'un certain seuil (ALEXANDRE GODZINSKI 2007).

La méthode consiste ainsi à représenter les estimateurs de Hill $\hat{\xi}_k$ en fonction de k , qui correspond au nombre des plus grandes valeurs retenues pour l'estimation.

L'estimateur de Hill est défini par :

$$\hat{\xi}_k = \frac{1}{k} \sum_{i=1}^k (\log X_{(i)} - \log X_{(k+1)})$$

où :

- $X_{(i)}$ représente la i -ème plus grande valeur de l'échantillon (statistiques d'ordre).
- k est le nombre d'excès pris en compte pour l'estimation.

Lorsque k est trop petit, le nombre de valeurs extrêmes sélectionnées est insuffisant, entraînant une forte volatilité de $\hat{\xi}_k$. À l'inverse, si k est trop grand, des valeurs non extrêmes sont incluses, ce qui biaise l'estimation de ξ .

Ainsi, le seuil optimal est déterminé en repérant la zone où la courbe de Hill se stabilise. Le nombre d'excès k , correspondant au seuil de valeur extrême, est choisi comme la plus petite valeur à partir de laquelle $\hat{\xi}_k$ devient approximativement constant.

Dans notre étude, la courbe de Hill confirme notre choix d'un seuil de 75 000 € pour l'écrêtement des sinistres et montre un stabilisation de l'estimateur de Hill au même niveau.

Bibliographie

- [1] Emilie Deprey et ALEXANDRE GODZINSKI. "Problématique de seuil dans la modélisation de la sinistralité en Réassurance Non Vie" (2007). URL : [https://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/d512ad5b22d73cc1c1257052003f1aed/0628c7d53c231edac1257570003b67ca/\\$FILE/AGodzinski_Actuariat2007.pdf](https://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/d512ad5b22d73cc1c1257052003f1aed/0628c7d53c231edac1257570003b67ca/$FILE/AGodzinski_Actuariat2007.pdf).
- [2] Rémi BELLINA. "Méthodes d'apprentissage appliquées à la tarification non-vie". Mém. de mast. ISFA, 2014.
- [3] Anis BORCHANI. "Statistiques des valeurs extrêmes dans le cas de lois discrètes" (2010). URL : <https://essec.hal.science/hal-00572559/document>.
- [4] Leo BREIMAN. "Bagging Predictors". *Machine Learning*, 24(2), 123-140 (1996).
- [5] Claire BROUSSE. "Modélisation de l'impact du changement climatique sur le risque de catastrophes naturelles causées par les inondations" (2021).
- [6] CCR. "Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050" (2023).
- [7] CCR. "EFFICACITÉ DES PLANS DE PRÉVENTION DES RISQUES D'INONDATION SUR LE COÛT DES DOMMAGES ASSURÉS : ESTIMATION DES IMPACTS PASSÉS, PRÉSENTS ET FUTURS" (2023).
- [8] CENTRE EUROPÉEN DE PRÉVENTION DES RISQUES D'INONDATION (CEPRI). "Le bâtiment face à l'inondation : Vulnérabilité des ouvrages" (2010).
- [9] CEPMMT. "Feuille de route du CEPMMT vers 2030" (2021).
- [10] Robin CHICHE. "Construction d'un modèle catastrophe stochastique d'inondation" (2013).
- [11] CLIMATE CHANGE SERVICE. "ERA5-Land hourly data from 1950 to present" (2024). URL : <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>.
- [12] CNRS, CEA ET MÉTÉOFRANCE. "Changement climatique : Les résultats des nouvelles simulations frnacaises" (2019).
- [13] COVÉA. "Changement climatique Assurance : Quelles conséquences sur la sinistralité à horizon 2050 ?" *Livre blanc Covéa* (2022).
- [14] DREAL GRAND EST. "Qu'est ce qu'un territoire à risque important d'inondation (TRI)" (2022).
- [15] DRIAS. "La modélisations du climat" (2023).
- [16] DRIAS. "Les scénarios RCP" (2013).
- [17] DRIEAT ILE-DE-FRANCE. "Les Plans de Prévention des Risques d'Inondation (PPRI) en Île-de-France" (2024).
- [18] FASTERCAPITAL. "le cout de l'assurance contre les inondations facteurs qui influencent vos primes" (2024). URL : <https://fastercapital.com/fr/contenu/Le-cout-de-1-assurance-contre-les-inondations---facteurs-qui-influencent-vos-primes.html#-valuer-1--ge-et-la-construction-de-votre-propri-t->.
- [19] FÉDÉRATION FRANÇAISE DU BATIMENT. "Indice FFB du coût de la construction" (2024). URL : https://www.outils.ffa batiment.fr/federation-francaise-du-batiment/le-batiment-et-vous/en-chiffres/indices-index/Chiffres_Index_FFB_Construction.html.
- [20] FRANCE ASSUREURS. "Inondations : l'indemnisation des dommages en cas de catastrophe naturelle" (2024). URL : <https://www.franceassureurs.fr/lassurance-protege-finance-et-emploie/lassurance-protege/lassurance-en-pratique-pour-les-particuliers/inondations-lindemnisation-des-dommages-en-cas-de-catastrophe-naturelle/>.

- [21] FRANCE ASSUREURS. "Inondations : l'indemnisation des dommages en cas de catastrophes naturelles" (2024).
- [22] FRANCE ASSUREURS. "L'assurance des catastrophes naturelles" (2024). URL : <https://www.franceassureurs.fr/lassurance-protege-finance-et-emploie/lassurance-protege/lassurance-en-pratique-pour-les-particuliers/assurance-des-catastrophes-naturelles/>.
- [23] FRANCE ASSUREURS. "L'assurance des évènements naturels en 2022 : Assurance de biens et de responsabilité" (2023).
- [24] FRANCE ASSUREURS. "L'assurance française Données clés 2022" (2023).
- [25] FRANCE ASSUREURS. "L'Assurance habitation en 2022 - Assurance de biens et de responsabilité" (2023).
- [26] GEORISQUES. "Inondation : Le premier risque naturel en France" (). URL : <https://www.georisques.gouv.fr/minformer-sur-un-risque/inondation>.
- [27] GEORISQUES. "La garantie Cat-Nat" (). URL : <https://www.georisques.gouv.fr/minformer-sur-la-prevention-des-risques/la-garantie-cat-nat>.
- [28] INSEE. "37,8 millions de logements en France au 1er Janvier 2023" (2023).
- [29] INSEE. "Catastrophe naturelle" (2016).
- [30] L'ARGUS DE L'ASSURANCE. "C'était en 2025 : Après les inondations des Alpes Maritimes, la mobilisation et les propositions de l'assurance" (2015).
- [31] Thierry LANGRENEY. "Adapter le système assurantiel français face à l'évolution des risques climatiques" (2023).
- [32] LÉGIFRANCE. "Décret n°95-1089 du 5 octobre 1995 relatif aux plans de prévention des risques naturels prévisibles" (2024). URL : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000556069>.
- [33] Esterina MASIELLO. "Modèles Linaires Généralisés (Support de cours)" (2022).
- [34] METEO FRANCE. "14-15 octobre 2018 : inondations catastrophiques sur l'Aude" (2020).
- [35] MÉTÉOFRANCE. "Le climat en France métropolitaine" (2023). URL : <https://www.meteofrance.com/comprendre-climat/france/le-climat-en-france-metropolitaine>.
- [36] MÉTÉOFRANCE. "Qu'est ce qu'un épisode méditerranéen" (2024). URL : <https://www.meteofrance.com/comprendre-la-meteo/precipitations/quest-ce-quun-episode-mediterraneen>.
- [37] MINISTÈRE DE L'ÉCONOMIE DES FINANCES ET DE LA SOUVERAINETÉ INDUSTRIELLE ET NUMÉRIQUE. "ce qu'il faut savoir sur l'assurance habitation" (2023). URL : <https://www.economie.gouv.fr/particuliers/assurance-habitation#>.
- [38] MINISTÈRE DE L'ÉCONOMIE ET DU DÉVELOPPEMENT DURABLE. "Les inondations" (2004).
- [39] MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE ET DE LA COHÉSION DES TERRITOIRES. "Chiffres clés des risques naturels" (2023).
- [40] Jennifer PARIENTE. "Modélisation du risque géographique en assurance habitation". Mém. de mast. Dauphine, 2017.
- [41] Juliette PLATEL. "Risque tempête : refléter la densité d' événements dommageables à l'aide d'un indice construit sur de l'open data" (2022).
- [42] Antoine QUANTIN. "Modélisation du péril inondation : passage d'une approche déterministe à une approche probabiliste en vue d'une tarification à l'exposition des couvertures de réassurance dans le cadre du régime des Catastrophes naturelles". Mém. de mast. CEA, 2011.
- [43] RÉGION CENTRE VAL DE LOIRE. "Importance des crues, notion de temps de retour" (2016).
- [44] Mathilde ROCHELLE. "Construction d'un score de risque Tempêtes au bâtiment en assurance". Mém. de mast. Duas, 2023.
- [45] WIKISTAT. "Introduction au modèle linéaire général" (2016). URL : <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-modlin-mlg.pdf>.

Table des figures

1	Schéma récapitulatif des 3 bases de données utilisées	5
2	Distribution des cumuls de précipitations historiques sur 3 jours	6
3	Trajectoires des différents scénarios	8
4	Comparaison des répartitions de la variable <i>cumul_precip</i> passée et future.	9
5	Cartographies des évolutions moyennes de prime pure pour le risque inondation par département du Grand Est	10
6	Summary Diagram of the Three Databases Used	11
7	Distribution of Historical 3-Day Precipitation Totals	12
8	Trajectories of the different scenarios	14
9	Comparison of the past and future distributions of the variable <i>cumul_precip</i>	15
10	Mapping of the average changes in pure premiums for flood risk by department in the Grand Est region.	16
1.1	Fréquence de sinistres entre 2013 et 2022 en assurance MRH selon les garanties	22
1.2	Coût moyen des sinistres entre 2013 et 2022 en assurance MRH selon les garanties	22
1.3	Mécanismes d'indemnisation des sinistres dépendant de la garantie Catastrophes naturelles et dépendant de la garantie TGN	24
1.4	Schéma théorique simplifié de l'impact du PPRI sur la sinistralité	25
1.5	L'historique des indemnisations versées par les assureurs à la suite d'aléas naturels	26
1.6	Les principaux événements naturels survenus entre 2020 et 2022	26
1.7	Répartition des arrêtés par nature de périls entre 1982 et 2022	27
1.8	L'inondation par débordement de cours d'eau	28
1.9	Portail d'accès aux données de ERA5-Land	34
1.10	Comparaison ERA5 et ERA5-Land	34
1.11	Données de précipitation pour la journée du 30 mai 2016 (ERA5-Land Hourly data)	35
1.12	Distribution des cumuls de précipitations historiques sur 3 jours	37
2.1	Calcul de l'indice de Gini normalisé à l'aide des aires A et B	46
2.2	Graphique Observés vs prédicts	47
2.3	Matrice de corrélation	49
2.4	Observés vs Prédits et multiplicateurs - Variable "Valeur du bâtiment"	50
2.5	Carte des prix moyens au m ² par commune des ventes de maisons et d'appartements en 2020	51
2.6	TRI	52
2.7	Exemple de PPRI pour la ville de Clamecy	53
2.8	Observés vs Prédits et multiplicateurs - Variable «PPRI»	53
2.9	Schema illustratif d'un bassin versant	54
2.10	Bassins et régions hydrographiques de France métropolitaine	55
2.11	Observés vs Prédits et multiplicateurs - Variable "Région hydrographique"	55
2.12	Observés vs Prédits et multiplicateurs - Variable "Nombre d'arrêtés Cat Nat"	56
2.13	Observés vs Prédits et multiplicateurs - Variable "Nb jours avec precip intenses"	57
2.14	Observés vs Prédits et multiplicateurs - Variable "jours consécutifs avec des precip >1mm"	57
2.15	Observés vs Prédits et multiplicateurs - Variable "Precip max sur un jour"	58
2.16	Résidus de déviance	60
2.17	Lift curve	61
2.18	Dommages au bâti et mobilier pour une durée de submersion >48h	62
2.19	Observés vs Prédits et multiplicateurs - Variable "Cumul de précipitation"	64
2.20	Observés vs Prédits et multiplicateurs - Variable "Nombre de pièces"	65
2.21	Dommages au bâti et mobilier pour un logement de 115 m ²	65
2.22	Observés vs Prédits et multiplicateurs - Variable "Nombre d'étages"	66

2.23 Observés vs Prédits et multiplicateurs - Variable "PPRI"	67
2.24 Observés vs Prédits et multiplicateurs - Variable "Région hydrographique"	67
2.25 Cinq grands types de climats en France métropolitaine	68
2.26 Observés vs Prédits et multiplicateurs - Variable "Région climatique"	69
2.27 Résidus de déviance	70
2.28 Lift curve	70
3.1 Effet des PPRi sur le montant des dommages assurés, à climat actuel et en 2050.	74
3.2 Schéma simplifié du maillage du globe terrestre	74
3.3 Les 5 grandes trajectoire socio-économiques des SSP	77
3.4 Trajectoires de réchauffement planétaire selon les cinq scénarios SSPx-y retenus dans le résumé pour décideur du GIEC	77
3.5 Matrice des scénarios RCP-SSP	78
3.6 Trajectoires des différents scénarios	78
3.7 Comparaison des répartitions de la variable <i>cumul_precip</i> passée et future.	81
3.8 Historique des indices FFB du coût de la construction et de l'inflation annuelle	83
3.9 Grands Est - départements et grands cours d'eau	84
3.10 Cartographies des évolutions moyennes de prime pure pour le risque inondation par département du Grand Est	85
3.11 Nombre d'arrêtés catastrophe naturelle pour l'inondation dans le Grand Est	86

Liste des tableaux

1 Description des différentes intensités	7
2 escription of the different intensities	13
1.1 Exemple : attribution des cumuls de précipitation n_j avant l'inondation pour chaque bâtiment	36
1.2 Exemple : ajout des quantiles n_j avant l'inondation pour chaque bâtiment	36
2.1 Description des différentes intensités	40
2.2 Valeurs des métriques pour les différents modèles	59
2.3 Tableau des indices FFB et coefficients d'ajustement de 2019 à 2024	62
2.4 Surface moyenne des logements français selon le nombre de pièces	64
2.5 Valeurs des métriques pour les différents modèles	69
2.6 Probabilité associée à chaque modalité de la variable <i>cumul_precip</i>	71
2.7 Calcul de la prime pure climatique au bâtiment	72
3.1 Résumé des scénarios RCP selon le forçage radiatif et leur trajectoire.	76
3.2 Description des différentes intensités passés	80
3.3 Exemple des probabilités projetées pour chaque bâtiment, réparties sur les 7 classes d'intensité.	81
3.4 Synthèse des données pour le calcul de la prime pure climatique projetée	82
3.5 Taux d'inflation annuelle et ajustement 2030	83