

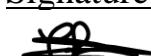
**Mémoire présenté le :
 pour l'obtention du diplôme
 de Statisticien Mention Actuarial
 et l'admission à l'Institut des Actuaires**

Par : Madame Laurène Choquer

Titre du mémoire : Perspectives d'évolution du régime CatNat pour le risque sécheresse

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

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

<u>Membres présents du jury de la filière :</u>	Signature : <u>Entreprise : WTW</u> Nom : Nicolas Clément Signature : 
<u>Membres présents du jury de l'Institut des Actuaires :</u>	Signature : <u>Directeur de mémoire en entreprise</u> Nom : Nicolas Clément Signature : 
<u>Invité :</u>	Nom : Signature :
<u>Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels</u> (après expiration de l'éventuel délai de confidentialité)	
<u>Signature du responsable entreprise :</u> 	
<u>Signature du candidat :</u> 	

0.1 Résumé

Dans un contexte de changement climatique, ce mémoire se penche sur les probables mutations de la sécheresse en France et, plus précisément, sur leurs répercussions dans le secteur de l'assurance, notamment en ce qui concerne l'assurance MRH (Multirisques Habitation). La sécheresse, phénomène complexe et multifacette, est tout d'abord exposée en profondeur pour en saisir toute la subtilité.

Pour mieux anticiper les impacts de la sécheresse, le domaine de l'assurance utilise divers indices. Présentés en détail dans cette étude, ils permettent d'évaluer l'intensité de la sécheresse, d'identifier les communes pouvant prétendre à la reconnaissance CatNat (catastrophes naturelles), ainsi que d'estimer leur sinistralité à venir et le coût associé pour les assureurs. Cette étude offre également un aperçu approfondi des indices les plus fréquemment employés, enrichi d'analyses et de prévisions fondées sur ces indicateurs climatiques.

S'appuyant sur des données modélisées, disponibles via le portail DRIAS, une exploration rigoureuse des modifications climatiques est effectuée pour le territoire français. D'ores et déjà, les résultats montrent une tendance inquiétante vers un assèchement croissant de certaines régions françaises.

Au-delà des observations évoquées, ce mémoire se penche spécifiquement sur la dimension assurantielle de cette sécheresse croissante. Nous proposons de modéliser la fréquence des sinistres dans les communes, de déterminer si une commune va être reconnue en état de catastrophe naturelle (CatNat) et d'évaluer le pourcentage de maisons affectées par la sécheresse. Pour cela, nous utilisons des techniques de machine learning, telles que le modèle linéaire généralisé, les forêts aléatoires (Random Forests) et le Gradient Boosting. Notre objectif est de comprendre l'évolution du nombre de sinistres liés à ce phénomène et, en y associant une estimation des coûts, d'illustrer l'impact financier sur les compagnies d'assurance. Cette analyse fournit une base solide pour intégrer ce risque et sa tendance future dans les modèles d'évaluation du risque et de solvabilité propres à l'entreprise (ORSA), qui sont essentiels à la gestion des risques dans le secteur de l'assurance.

Après cette étape de modélisation, des analyses de sensibilité seront menées. L'objectif est de déterminer quels facteurs intrinsèques au portefeuille ont le plus d'impact sur la sinistralité. Cette démarche offre ainsi aux assureurs des informations précieuses pour élaborer une stratégie de pilotage plus adaptée et efficiente.

0.2 Abstract

In the context of climate change, this dissertation delves into the probable changes in drought in France and, more specifically, its repercussions in the insurance sector, particularly with regards to home protection insurance. Drought, a complex and multifaceted phenomenon, is first thoroughly presented to grasp its intricacies.

To better anticipate the impacts of drought, the insurance field utilizes various indices. Detailed in this study, they are used to assess the intensity of drought, identify municipalities eligible for CatNat (natural disasters) recognition, and estimate their future claims frequency and the associated cost for insurers. This study also provides a comprehensive overview of the most commonly used indices, augmented by analyses and forecasts based on these climatic indicators.

Relying on modeled data available through the DRIAS portal, a rigorous exploration of climatic changes is conducted for the French territory. The findings already indicate a concerning trend towards increased aridity in certain French regions.

Beyond the mentioned observations, this dissertation specifically addresses the insurance dimension of this increasing drought. It introduces a modeling of the frequency of drought-related claims using machine learning techniques such as the generalized linear model, random forests, and gradient boosting. The aim is to understand the evolving trend in the number of claims associated with this phenomenon. Coupled with a cost projection, this modeling seeks to highlight the financial impact on insurance companies. Thus, it offers an opportunity to incorporate this risk and its future trajectory more insightfully into ORSA models (Own Risk and Solvency Assessment), vital tools for risk management in insurance.

Following this modeling phase, sensitivity analyses will be carried out. The goal is to identify which intrinsic factors in the portfolio have the most significant impact on the claims frequency. This approach provides insurers with valuable insights for crafting a more appropriate and efficient management strategy.

0.3 Note de synthèse

Face à l'intensification des changements climatiques, le risque de sécheresse occupe une place prépondérante pour le secteur de l'assurance, en particulier concernant les contrats Multirisques Habitation (MRH). Ces contrats visent à couvrir les propriétaires contre divers sinistres, tant au niveau des dégâts matériels de l'habitation que des biens à l'intérieur. La prise en charge des dommages liés à la sécheresse nécessite une reconnaissance en tant que "CatNat" (catastrophe naturelle). Cette reconnaissance est soumise à des critères spécifiques : la présence d'au moins 3% d'argile dans le sol de la commune, une moyenne du Soil Water Index sur 3 mois inférieure à celle des 3 mêmes mois sur les 25 dernières années, et enfin, une demande officielle du maire de la commune qui doit être acceptée. Malgré cette reconnaissance, les assurés sont souvent soumis à une franchise spécifique en plus de la surprime fixée à 12% de la prime dans le cadre de l'assurance MRH.

Le GIEC, dans son rapport du 9 août 2021, met en évidence les enjeux liés au changement climatique, en accentuant l'impact des activités humaines. Le GIEC utilise les scénarios RCP (Representative Concentration Pathways) pour évaluer l'évolution des émissions. Ces scénarios, basés sur différentes hypothèses d'évolution technologique, politique et économique, vont du plus optimiste, envisageant des mesures draconiennes de réduction des émissions, au plus pessimiste, avec une augmentation continue de ces émissions. Complétant les RCP, les SSP (Shared Socioeconomic Pathways) envisagent différentes trajectoires socio-économiques possibles, donnant ainsi une vision complète des futures trajectoires climatiques. L'étude se concentre sur trois de quatre scénarios (le 2.6, le 4.5 et le 8.5) pour observer au mieux les effets du dérèglement climatique.

L'ACPR, consciente de ces réalités, incite les assureurs à une adaptation stratégique. Elle les encourage notamment à intégrer, dans leurs ORSA (Own Risk and Solvency Assessment), les risques liés à la sécheresse. Cette évaluation des risques et de la solvabilité est primordiale pour assurer la pérennité des compagnies d'assurance face à l'augmentation des risques climatiques. En se projetant sur ces défis futurs, le monde de l'assurance doit adapter ses offres et garanties pour protéger efficacement les assurés des conséquences du changement climatique.

Face à l'ampleur croissante des défis climatiques et à la nécessité d'anticiper les risques futurs pour le secteur de l'assurance, nous avons rassemblé un ensemble de données exhaustives couvrant les communes françaises. Ces informations contiennent des détails sur la hauteur des bâtiments, leur âge, ainsi que des données précieuses sur la composition argileuse des sols, fournies par le BRGM. Le BRGM a identifié quatre zones distinctes en fonction de la répartition de l'argile dans les sols et du risque associé au phénomène de Retrait-Gonflement des Argiles (RGA) :

- Zone d'exposition forte : Dans cette zone, les sols de toutes les communes sont très exposés au RGA. Il existe une forte probabilité de survenance de sinistres liés à la sécheresse, avec des phénomènes d'une intensité élevée.
- Zone d'exposition moyenne : Bien que le sol des communes soit exposé au RGA, la probabilité et l'intensité des sinistres liés à la sécheresse sont moins prononcées qu'en zone d'exposition forte.
- Zone d'exposition faible : Le risque de sinistres liés à la sécheresse est plus faible dans cette zone, tout comme l'intensité des phénomènes attendus.

- Zone d'exposition nulle : Cette zone est caractérisée par des sols non argileux, donc sans risque de RGA. Les risques liés à la sécheresse y sont absents.

Outre ces données géotechniques, nous intégrons également des informations climatiques fournies par le DRIAS, qui incluent des éléments essentiels comme la pluviométrie, les températures et l'évapotranspiration, reflétant l'eau évaporée du sol et celle transpirée par les plantes.

Pour affiner notre compréhension du phénomène de retrait-gonflement des sols, l'indice SWI (Soil Water Index) est central. La moyenne du SWI sur trois mois, nommée SWI Moy, a été introduite pour mieux appréhender les variations de cet indice, offrant une perspective enrichie des impacts de la sécheresse sur le RGA.

Grace à ces informations détaillées, nous développons des modèles prédictifs qui nous aideront à anticiper quelles communes seront reconnues comme zones de catastrophe naturelle (CatNat) ainsi que le nombre de sinistres associés dans un avenir proche, ainsi que pour des horizons plus éloignés tels que 2050, 2075 et 2100. Pour ce faire, nous employons des techniques de machine learning, notamment le modèle linéaire généralisé, les forêts aléatoires (Random Forest) et le Gradient Boosting. Notre but est de façonner nos stratégies d'assurance pour les adapter à l'évolution du climat. L'efficacité de ces modèles est évaluée à travers la courbe ROC.

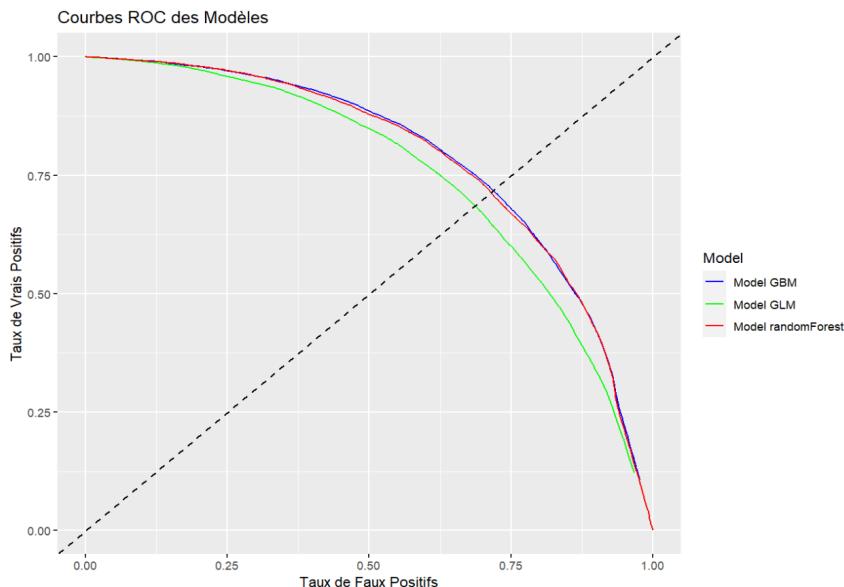


FIGURE 1 – Comparaison des courbes ROC des modèles

On constate que la méthode la plus efficace est le Gradient boosting c'est donc cette méthode qui est conservée pour obtenir les fréquences futures probables.

Suite à la mise en place de notre modèle prédictif, nous avons procédé à des analyses de sensibilité pour identifier les variables ayant le plus d'impact sur la sinistralité. Parmi les techniques employées, la méthode de Morris s'est avérée particulièrement éclairante. Nos analyses ont ré-

vélé que le pourcentage d'argile de type 3, la hauteur des bâtiments, ainsi que la fréquence des bâtiments construits avant 1970 et après 1990 sont les variables les plus influentes dans la prédition de notre sinistralité. Face à ces constatations, il est devenu primordial de comprendre les interactions entre ces variables.

Afin d'offrir une visualisation claire et synthétique de ces interactions, nous avons représenté ces variables sur un graphique PDD (Partial Dependence Diagram). L'analyse de ce graphique PDD a révélé que le pourcentage d'argile de type 3, étonnamment, interagit assez peu avec les autres variables du portefeuille. Plus ce pourcentage augmente, plus la sinistralité s'accroît de manière linéaire, quelles que soient les variations des autres paramètres. En revanche, les fréquences de construction présentent des interactions notables entre elles, particulièrement lorsqu'elles sont représentées au même pourcentage. De plus, elles montrent des interactions significatives avec les variables d'argile de types 1 et 2. Ces indicateurs nous guident dans l'affinement de nos stratégies d'assurance et la prise de décisions éclairées.

Dans le cadre de notre étude sur les impacts climatiques, nous avons procédé à une analyse préliminaire des variables climatiques clés : les précipitations, l'évapotranspiration et le SWI. Cette analyse est basée sur divers scénarios du GIEC et se penche sur trois horizons temporels : 2050, 2075 et 2100.

Précipitations : Quel que soit le scénario ou l'horizon envisagé, nos résultats montrent une constance notable. Bien que nous observions une tendance à la diminution des précipitations, le nombre de jours de précipitations reste constant par rapport à la période de référence.

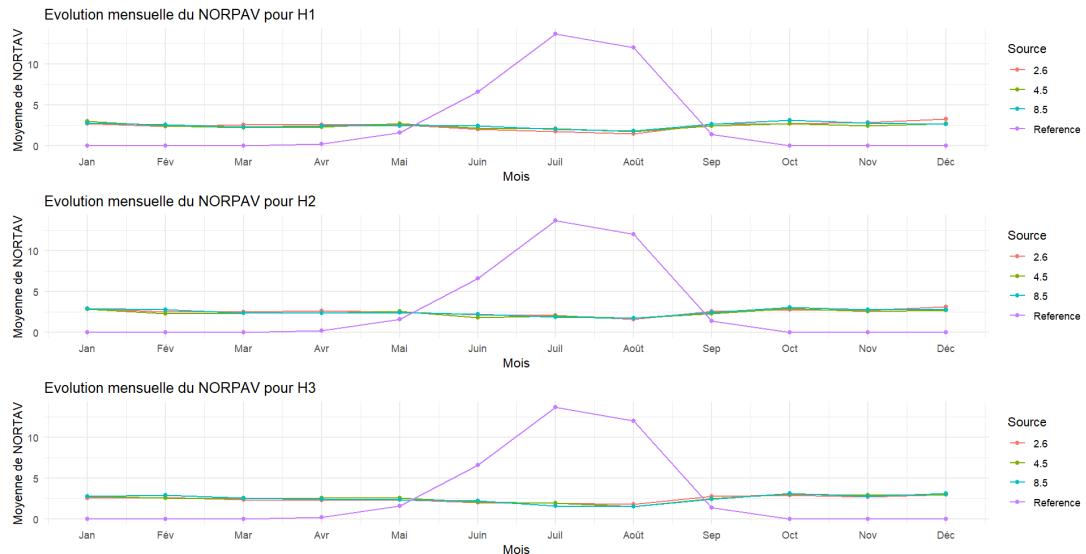


FIGURE 2 – Évolution mensuelle de la quantité de pluie par jour par période.

Évapotranspiration : De manière similaire, l'évapotranspiration, sur tous les scénarios et horizons, ne présente pas de variation significative. Elle demeure stable par rapport à la période de référence, soulignant une régularité dans les prévisions à travers les scénarios et les horizons temporels.

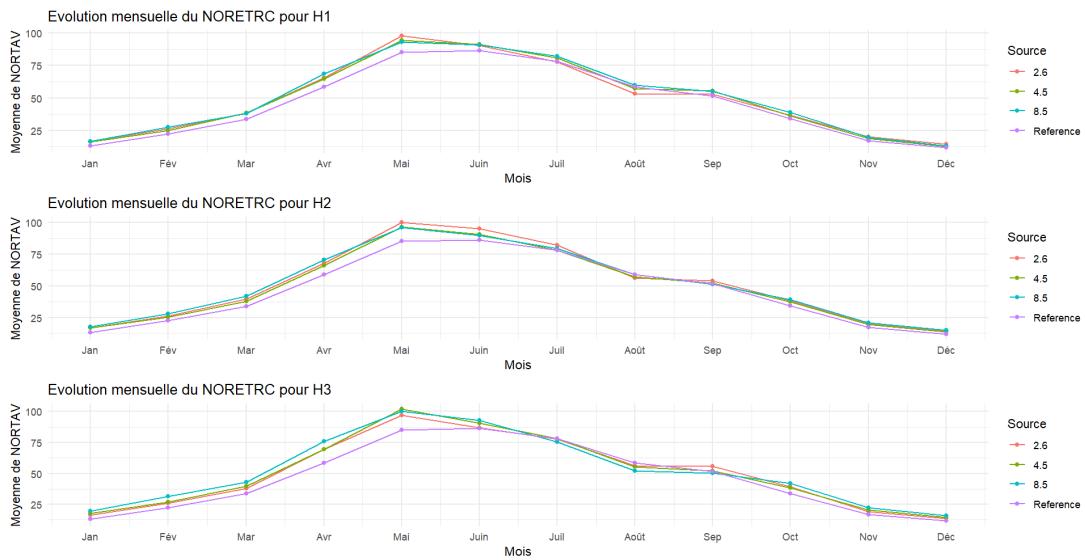


FIGURE 3 – Évolution mensuelle de l'évapotranspiration moyenne par mois par période

SWI : Les analyses du SWI montrent, quant à elles, une divergence marquée. Bien que les résultats soient constants à travers les différents scénarios et horizons, une augmentation du nombre de jours de sécheresse est notable, spécialement pour les scénarios les plus pessimistes. Le mois d'août se distingue par des pics de sécheresse. Toutefois, le scénario 8.5, souvent considéré comme le plus alarmant, prévoit une augmentation considérable du nombre de jours de sécheresse, y compris pendant les mois d'hiver.

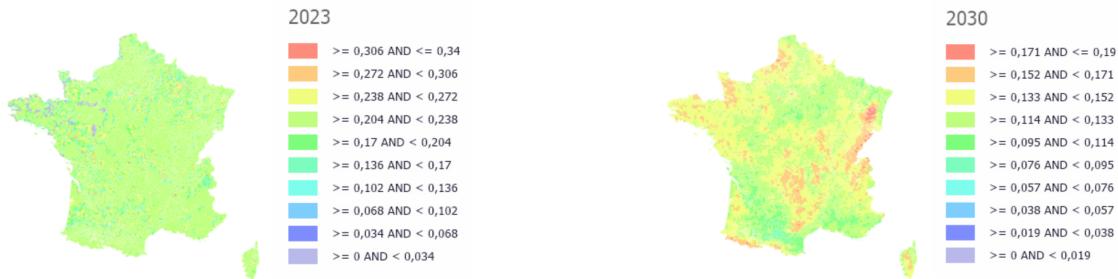


FIGURE 4 – Critère SWI actuel

FIGURE 5 – Critère SWI à l'année 2050

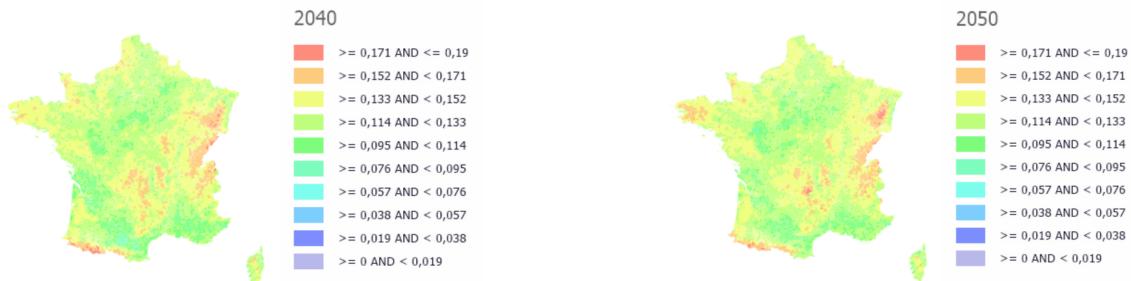


FIGURE 6 – Critère SWI à l'année 2075

FIGURE 7 – Critère SWI à l'année 2100

Le graphique dévoile l'évolution du score SWI, qui évalue la fréquence d'éligibilité d'une commune au critère de SWI sur une période donnée. De 2022 à 2050, une augmentation notable du SWI est observée. Cependant, dans les années qui suivent, ce score semble se stabiliser avec des variations légères. Cette tendance pourrait s'expliquer par le fait que les sols, après de longues périodes de sécheresse, ne remplissent plus aussi souvent les critères de faible humidité comparativement aux données SWI des 25 années précédentes.

Désormais, nous illustrons les fréquences de sinistres observées.

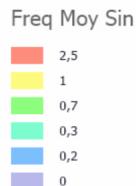
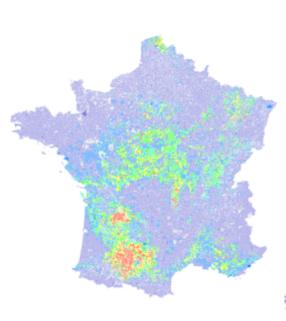


FIGURE 8 – Fréquence de sinistres pour la période de référence

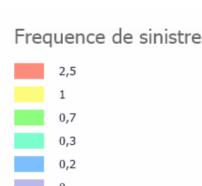
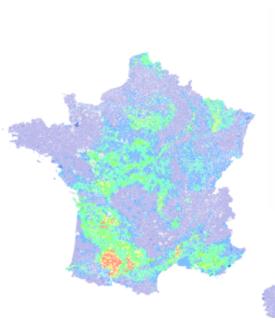


FIGURE 9 – Fréquence de sinistres pour la période de 2050 et le scénario 4.5

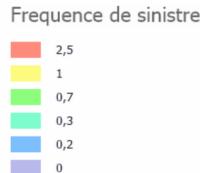


FIGURE 10 – Fréquence de sinistres pour la période de 2075 et le scénario 4.5

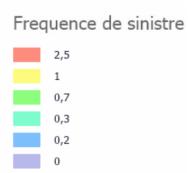
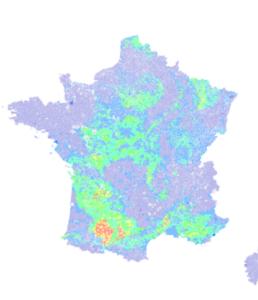


FIGURE 11 – Fréquence de sinistres pour la période de 2100 et le scénario 4.5

Surprenamment, entre 2075 et 2100, moins de communes semblent touchées par la sécheresse, notamment à l'est de Paris où la fréquence des sinistres diminue. Tous les scénarios montrent, après 2050, une stabilisation des fréquences de sinistres, possiblement due à une baisse de la pluviométrie et des conditions d'évapotranspiration constantes. Ces facteurs limitent le retrait-gonflement des sols. La région de la Bretagne est la moins touchée à cause de la rareté des sols argileux sensibles à ce phénomène. En revanche, le "croissant argileux", une zone s'étendant du nord-est de Paris à la Bourgogne et au centre de la France, est fortement impacté. Le bassin méditerranéen, autrefois peu touché, risque de voir une augmentation des sinistres à l'avenir à cause des changements climatiques amplifiant les chaleurs et réduisant les précipitations. Si la Bretagne reste protégée, d'autres zones, comme le bassin méditerranéen, pourraient devenir plus vulnérables.

Nous utilisons l'indice du coût à la construction (ICC) comme principal outil pour évaluer l'infla-

tion annuelle liée à la construction. Cet indice reflète les variations des coûts et permet d'ajuster les valeurs financières en fonction des évolutions du marché. Grâce à l'ICC, nous sommes en mesure d'estimer avec précision le coût futur des sinistres, en prenant en compte les fluctuations économiques ce qui donne pour un département :

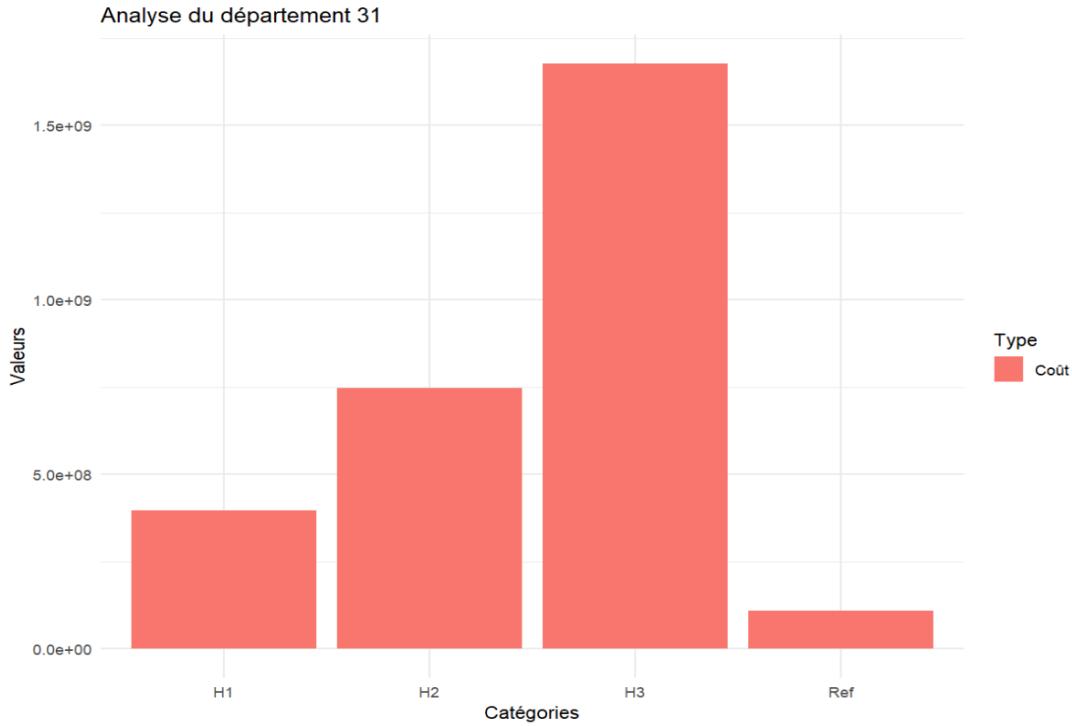


FIGURE 12 – Evolution des coûts des sinistres pour le département de Haute-Garonne

L'accroissement du coût des sinistres entre 2022 et 2050 est notable, avant de connaître une hausse plus modérée par la suite. Toutefois, il apparaît que la majeure partie de cette augmentation peut être attribuée à l'inflation.

Après avoir estimé le coût global et local, nous pouvons désormais déterminer un pourcentage de variation. Celui-ci quantifie l'augmentation du coût pour les assureurs. Cette variation peut être intégrée au bilan ORSA, et plus spécifiquement à l'évolution du SCR de l'assureur, afin d'anticiper au mieux les défis futurs qui l'attendent.

Néanmoins, l'interprétation des résultats doit se faire avec circonspection en raison des évolutions possibles des règles de reconnaissance CatNat. Suite à la sécheresse sans précédent de 2022, générant des dommages de 2,9 milliards d'euros, principalement sur les sols argileux, l'ordonnance n° 2023-78 a été introduite pour repenser le régime d'indemnisation des catastrophes naturelles. Ces discussions envisagent notamment des changements majeurs tels que la réduction de la période de retour et la modification des critères d'éligibilité à l'indemnisation. De plus, des obligations plus strictes sont imposées aux experts évaluant les dommages, avec des sanctions en cas de manquement. Il est également à noter que les interprétations doivent être considérées avec prudence, car elles ne prennent pas en compte les évolutions législatives récentes concernant les nouvelles constructions. Notamment, les lois visant à préserver les espaces ruraux. Ainsi, à l'avenir, il est probable que nous observerons une tendance accrue aux rénovations plutôt qu'à

la construction de nouveaux bâtiments.

En conclusion, cette étude détaille les impacts alarmants du changement climatique, en soulignant la corrélation entre hausse des températures, diminution des précipitations et assèchement des sols, en particulier dans le sud de la France. Ces modifications climatiques ont des répercussions économiques, pouvant influencer le montant des primes d'assurance, mettant ainsi en question le modèle actuel du régime Cat Nat. Des mesures préventives, telles que l'application de la loi ELAN, sont envisageables, mais une adaptation continue des critères d'indemnisation semble essentielle face à la dégradation climatique croissante. Cette situation nous incite à reconsidérer notre relation avec l'environnement et les systèmes d'assurance qui nous couvrent.

0.4 Synthesis

In the face of intensifying climate change, the risk of drought occupies a prominent place in the insurance sector, particularly concerning Multi-Risk Home (MRH) contracts. These contracts aim to cover homeowners against various damages, both in terms of property damage and interior belongings. Addressing damages related to drought requires recognition as a "CatNat" (natural catastrophe). This recognition is subject to specific criteria : the presence of at least 3% of clay in the soil of the municipality, an average Soil Water Index (SWI) over 3 months lower than the average of the same 3 months over the past 25 years, and finally, an official request from the mayor of the municipality that must be accepted. Despite this recognition, policyholders often face a specific deductible in addition to a surcharge set at 12% of the premium in the context of MRH insurance.

The IPCC, in its report of August 9, 2021, highlights the challenges related to climate change, emphasizing the impact of human activities. The IPCC employs Representative Concentration Pathways (RCP) scenarios to assess emission trends. These scenarios, based on various assumptions about technological, political, and economic developments, range from the most optimistic, envisioning drastic emission reduction measures, to the most pessimistic, featuring a continuous increase in emissions. Complementing the RCPs, Shared Socio-economic Pathways (SSP) consider various possible socio-economic trajectories, providing a comprehensive view of future climate trajectories. This study focuses on three out of four scenarios (2.6, 4.5, and 8.5) to better observe the effects of climate change.

Recognizing these realities, the ACPR encourages insurers to engage in strategic adaptation. This includes urging insurers to integrate drought-related risks into their Own Risk and Solvency Assessment (ORSA). This risk and solvency evaluation is crucial for ensuring the resilience of insurance companies in the face of increasing climate risks. In anticipation of these future challenges, the insurance industry must adjust its offerings and coverage to effectively protect policyholders from the consequences of climate change.

In response to the growing magnitude of climate challenges and the need to anticipate future risks for the insurance sector, we have compiled a comprehensive dataset covering French municipalities. These data include details on building heights, their ages, as well as valuable information about soil clay composition provided by BRGM. BRGM has identified four distinct zones based on clay distribution in soils and associated risk of Shrink-Swell Clays (RGA) :

- High Exposure Zone : In this zone, soils in all municipalities are highly exposed to RGA. There is a high likelihood of drought-related damages, with intense occurrences.
- Medium Exposure Zone : While the soil in municipalities is exposed to RGA, the probability and intensity of drought-related damages are less pronounced than in the high exposure zone.
- Low Exposure Zone : The risk of drought-related damages is lower in this zone, as is the intensity of expected phenomena.
- No Exposure Zone : This zone is characterized by non-clay soils, thus devoid of RGA risk. Drought-related risks are absent in these areas.

Beyond these geotechnical data, we also incorporate climatic information provided by DRIAS, including essential elements such as precipitation, temperatures, and evapotranspiration, reflecting soil-evaporated and plant-transpired water.

To refine our understanding of the soil shrink-swell phenomenon, the Soil Water Index (SWI) is pivotal. The three-month average SWI, referred to as SWI Moy, has been introduced to better grasp the variations of this index, offering an enriched perspective on drought impacts on RGA.

Armed with these detailed pieces of information, we construct predictive models for the horizons of 2050, 2075, and 2100, utilizing machine learning methods such as Generalized Linear Models, Random Forest, and Gradient Boosting. Our aim is to tailor our insurance strategies to this changing climate reality. The performance of these models is depicted in ROC curves.

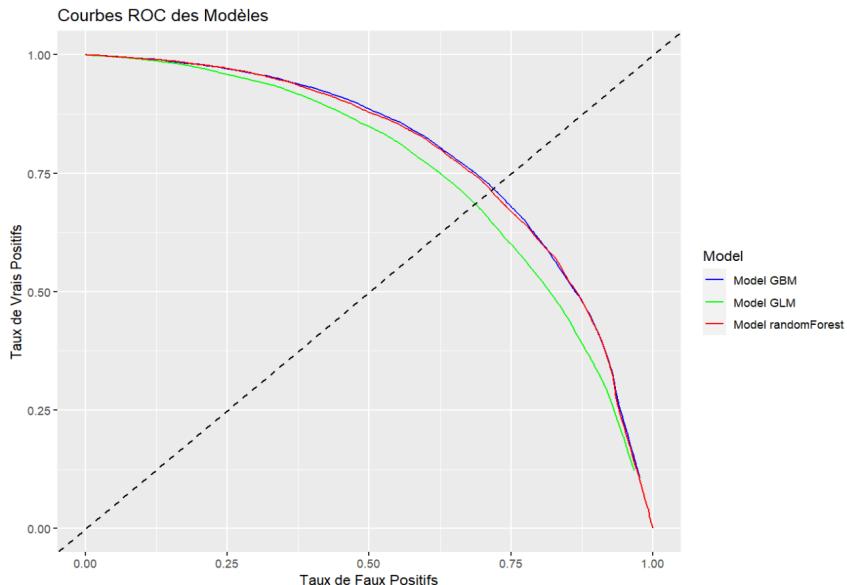


FIGURE 13 – Comparison of ROC curves of models

It is evident that the most effective method is Gradient Boosting, which is retained for obtaining probable future frequencies.

Following the establishment of our predictive model, we conducted sensitivity analyses to identify the variables with the most impact on claim frequency. Among the techniques employed, the Morris method proved to be particularly illuminating. Our analyses revealed that the percentage of type 3 clay, building height, as well as the frequency of buildings constructed before 1970 and after 1990, are the most influential variables in predicting our claim frequency. Faced with these findings, understanding the interactions between these variables became paramount.

To offer a clear and concise visualization of these interactions, we depicted these variables in a Partial Dependence Diagram (PDD). Analyzing this PDD graph revealed that the percentage of type 3 clay surprisingly interacts relatively little with other portfolio variables. As this per-

tage increases, the claim frequency linearly grows, regardless of variations in other parameters. In contrast, construction frequencies exhibit notable interactions among themselves, particularly when represented at the same percentage. Moreover, they show significant interactions with type 1 and type 2 clay variables. These insights guide us in refining our insurance strategies and making informed decisions.

As part of our study on climate impacts, we conducted a preliminary analysis of key climatic variables : precipitation, evapotranspiration, and SWI. This analysis is based on various IPCC scenarios and focuses on three time horizons : 2050, 2075, and 2100.

Precipitation : Regardless of the scenario or horizon considered, our results exhibit a notable consistency. Although we observe a tendency for decreased precipitation, the number of precipitation days remains constant compared to the reference period.

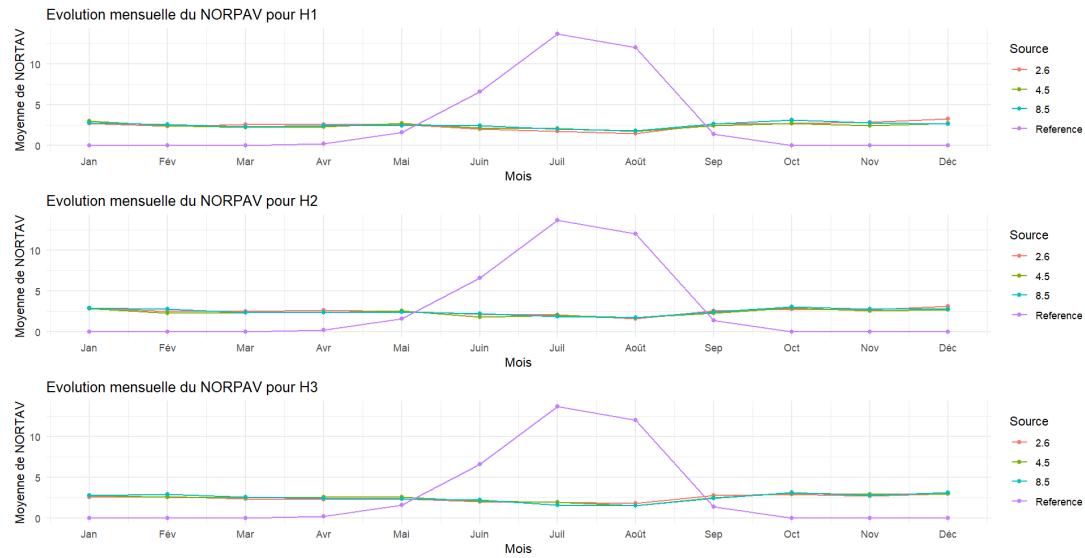


FIGURE 14 – Monthly change in the number of rain per day per period

Evapotranspiration : Similarly, evapotranspiration across all scenarios and horizons does not display significant variation. It remains stable relative to the reference period, underscoring consistency in forecasts across scenarios and time horizons.

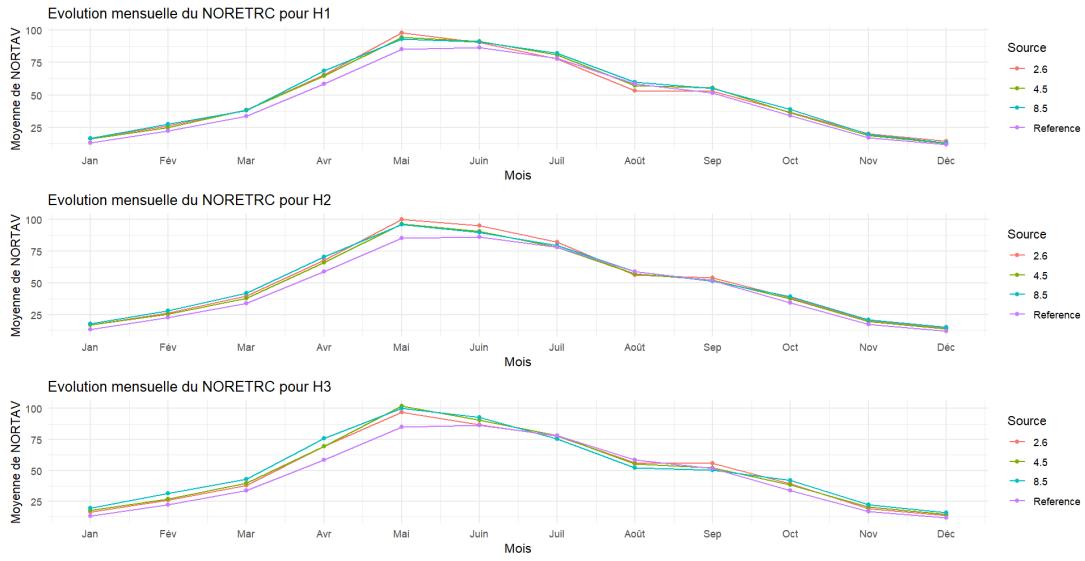


FIGURE 15 – Monthly evolution of average evapotraspiration per month per period

SWI : Analyses of the SWI show a pronounced divergence. While results remain consistent across different scenarios and horizons, an increase in the number of drought days is notable, especially for the more pessimistic scenarios. August stands out with drought peaks. However, scenario 8.5, often considered the most alarming, predicts a substantial increase in drought days, including during the winter months.

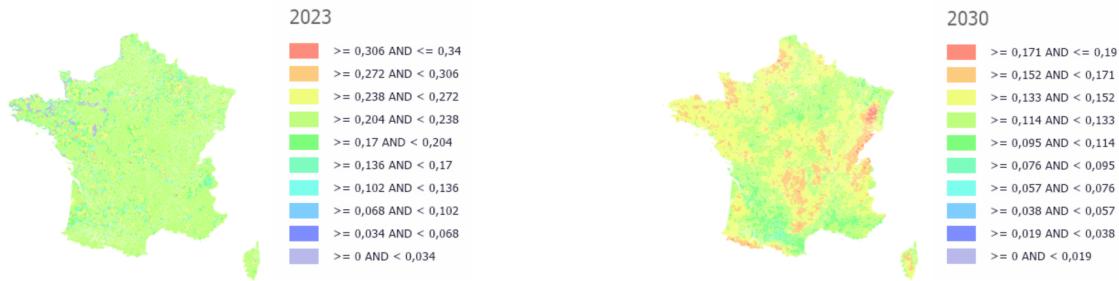


FIGURE 16 – Current SWI criterion

FIGURE 17 – SWI criterion to year 2050

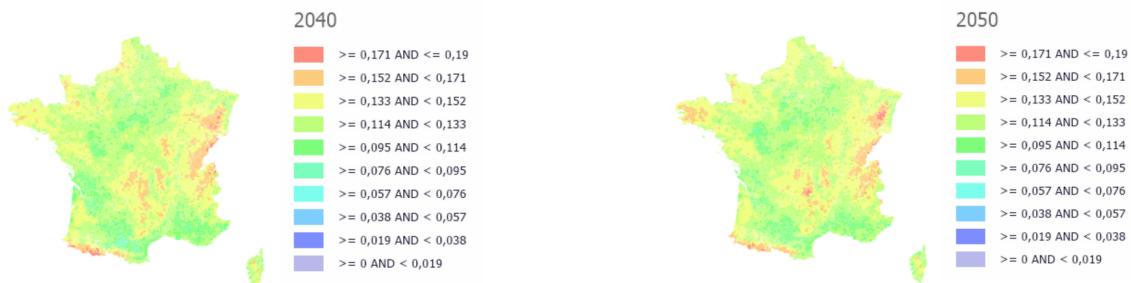
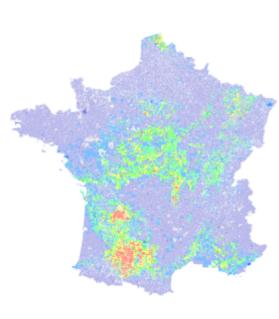


FIGURE 18 – SWI criterion to year 2075

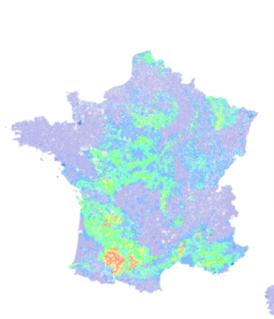
FIGURE 19 – SWI criterion to year 2100

The graph reveals the evolution of the SWI score, assessing the frequency of a municipality's eligibility for the SWI criterion over a given period. From 2022 to 2050, a notable increase in SWI is observed. However, in the subsequent years, this score appears to stabilize with minor fluctuations. This trend could be attributed to the fact that soils, after extended periods of drought, no longer meet the low humidity criteria as frequently compared to the SWI data from the previous 25 years.

Now, we illustrate the observed claim frequencies



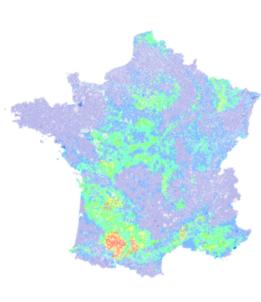
Freq Moy Sin
2,5
1
0,7
0,3
0,2
0



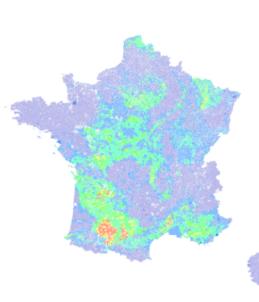
Frequence de sinistre
2,5
1
0,7
0,3
0,2
0

FIGURE 20 – Frequency of loss for the reference period

FIGURE 21 – Disaster frequency for the period 2050 and scenario 4.5



Frequence de sinistre
2,5
1
0,7
0,3
0,2
0



Frequence de sinistre
2,5
1
0,7
0,3
0,2
0

FIGURE 22 – Claim Frequency for 2075 and Scenario 4.5

FIGURE 23 – Disaster frequency for the period 2100 and scenario 4.5

Surprisingly, between 2075 and 2100, fewer municipalities seem affected by drought, particularly east of Paris where claim frequency decreases. All scenarios show, after 2050, a stabilization of claim frequencies, possibly due to reduced precipitation and constant evapotranspiration conditions. These factors limit soil shrink-swell phenomena. The Brittany region is the least affected due to the scarcity of clay-sensitive soils. In contrast, the "clay crescent," an area extending from northeast Paris to Burgundy and central France, is heavily impacted. The Mediterranean basin, once minimally affected, could experience increased claims in the future due to climate changes amplifying heat and reducing precipitation. While Brittany remains safeguarded, other areas like the Mediterranean basin could become more vulnerable.

We utilize the Construction Cost Index (CCI) as the primary tool to evaluate annual construction-related inflation. This index reflects cost fluctuations and allows for adjusting financial values based on market developments. Through the CCI, we are able to accurately estimate the fu-

ture cost of claims, considering economic fluctuations. This information allows us to calculate a percentage variation, indicating the increase in cost for insurers. This variation can be integrated into the ORSA balance sheet, specifically within the insurer's Solvency Capital Requirement (SCR) evolution, to better anticipate the forthcoming challenges. Nevertheless, the interpretation

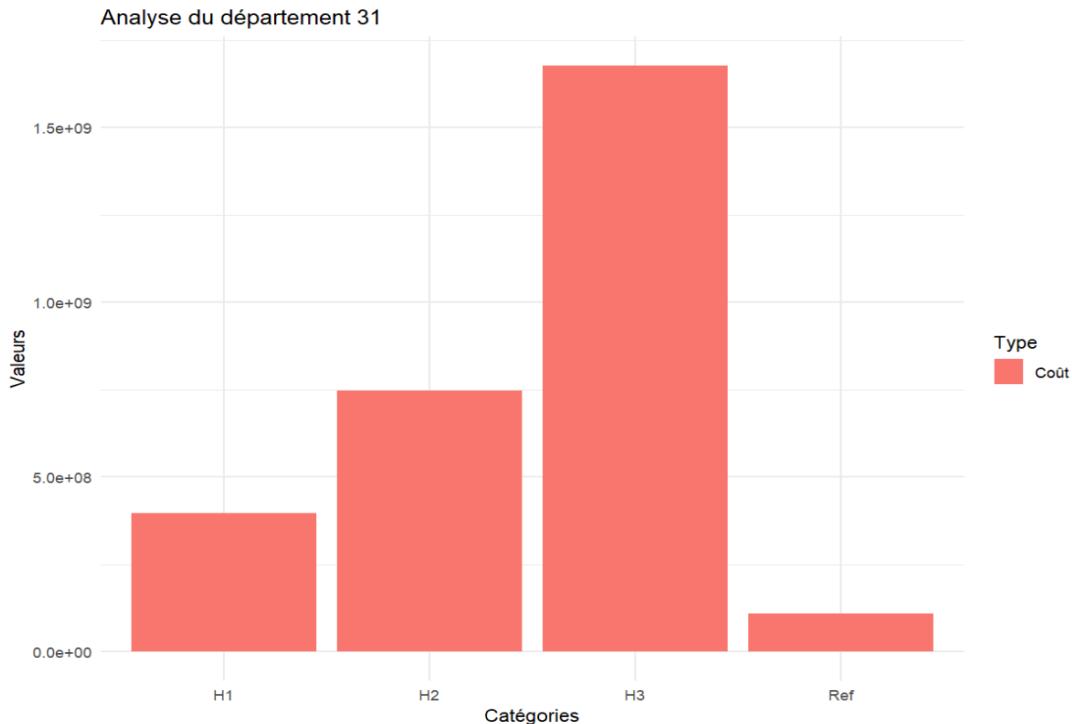


FIGURE 24 – Evolution of claims costs for the department of Haute-Garonne

of results must be approached cautiously due to potential changes in CatNat recognition rules. Following the unprecedented drought of 2022, resulting in €2.9 billion in damages, primarily to clayey soils, Ordinance No. 2023-78 was introduced to reform the regime of natural disaster compensation. These discussions contemplate major changes such as reducing the return period and modifying eligibility criteria for compensation. Furthermore, stricter obligations are imposed on experts assessing damages, with sanctions in case of non-compliance.

In conclusion, this study delineates the alarming impacts of climate change, highlighting the correlation between rising temperatures, decreasing precipitation, and soil desiccation, particularly in southern France. These climatic changes have economic repercussions that may influence insurance premium amounts, thus questioning the current model of the Cat Nat regime. Preventive measures, such as the implementation of the ELAN law, are feasible, but an ongoing adaptation of compensation criteria seems essential in the face of escalating climate degradation. This situation urges us to reconsider our relationship with the environment and the insurance systems that provide coverage.

0.5 Remerciements

Je tiens à exprimer ma gratitude envers toutes les personnes qui ont contribué à la réalisation de ce mémoire. Tout d'abord, je souhaite remercier tous les membres de l'équipe ICT pour m'avoir offert l'opportunité d'effectuer ce travail au sein de l'entreprise et pour la confiance qu'ils m'ont témoignée tout au long de sa rédaction.

Je tiens particulièrement à exprimer ma reconnaissance envers Eric Gettler et Nicolas Clement pour leur encadrement et leurs précieux conseils tout au long de ce processus.

Enfin, je tiens à dédier une pensée spéciale à mes proches qui ont participé à la relecture de ce mémoire et m'ont apporté leur soutien tout au long de sa rédaction.

Table des matières

0.1 Résumé	1
0.2 Abstract	2
0.3 Note de synthèse	3
0.4 Synthesis	10
0.5 Remerciements	16
1 Contextualisation	20
1.1 Contextualisation	20
1.2 Solvabilité II	22
Pilier 1 : Exigence quantitative	22
Pilier 2 : Exigence qualitative	23
Pilier 3 : Communication et transparence	24
1.3 Les risques de durabilité	24
Les risques climatiques	25
1.4 Contexte légal	29
1.5 L'assurance MRH	31
1.6 Le changement climatique et les rapports du GIEC	32
2 Présentation de la base de données	37
2.1 Présentation de la Base de données	37
Construction d'une sinistralité par commune	37
Données sur les communes françaises	38
Indicateur Climatique	41
3 Élaboration d'un modèle prédictif de la sinistralité	52
3.1 Contexte	52
3.2 Modèles linéaires généralisés	54
Théorie	54
Choix des variables du modèle	54
Résultats	57
3.3 Random Forest	64
Théorie	64
Choix des paramètres du modèle	65
Résultats	66
3.4 Gradient Boosting	67
Théorie	67
Choix des paramètres du modèle	69
Résultats	70
3.5 Comparaison des modèles	71

4 Analyse de sensibilité du modèle	75
4.1 La méthode de Morris	75
Théorie : Le plan OAT	75
Mesure de sensibilité	77
5 Projection de la sinistralité	86
5.1 Construction de la composition du portefeuille	86
5.2 Analyse préliminaire	87
Analyse du SWI	87
Analyse des précipitations	90
Analyse de l'évapotranspiration	91
Projection de la sinistralité	92
6 Projection de la sévérité	98
6.1 Indice de construction en bâtiment	98
6.2 Analyse de l'évolution des coûts en fonction des scénarios	99
7 Conclusion	103
8 Annexe	106
8.1 Annexe 1 :La correction statistique des simulations	106
8.2 Annexe 2 : Calcul de l'FTP	109

Chapitre 1

Contextualisation

1.1 Contextualisation

Le risque climatique est devenu un enjeu crucial pour l'avenir des assureurs, car il représente désormais un risque majeur et en constante augmentation. Les impacts des événements climatiques extrêmes tels que les tempêtes, les inondations, les incendies de forêt et les sécheresses sont de plus en plus fréquents et intenses, ce qui met en danger les biens assurés et les populations. Les assureurs doivent donc prendre en compte ce risque majeur dans leur activité et adapter leurs stratégies en conséquence pour garantir leur pérennité à long terme.

En effet, la température globale de la Terre a augmenté d'environ 1 degré Celsius depuis l'ère préindustrielle. Selon les chiffres de Météo-France, cette tendance à la hausse est clairement visible, avec une augmentation constante des températures au fil des ans. De plus, il est important de noter que sur les 10 années les plus chaudes jamais enregistrées, 8 appartiennent à la dernière décennie.^[3]

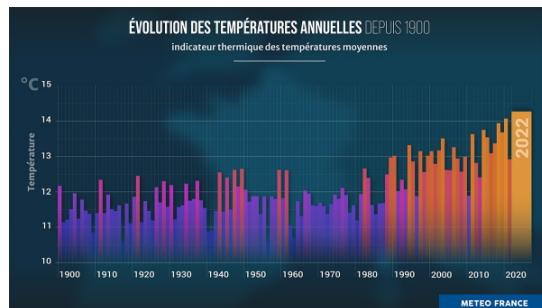


FIGURE 1.1 – Évolution de la température moyenne par ans

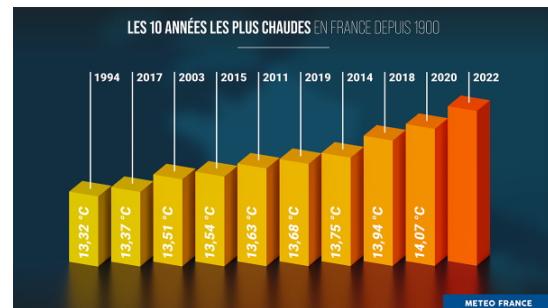


FIGURE 1.2 – 10 années ayant enregistré les températures moyennes les plus élevées

Il y a une corrélation directe entre le risque climatique et l'augmentation des températures. En effet, l'augmentation des températures est un symptôme du changement climatique qui est causé par les émissions de gaz à effet de serre, principalement le dioxyde de carbone, dans l'atmosphère.^[13]

Plus les températures augmentent, plus les effets du changement climatique sont susceptibles de s'aggraver, avec des conséquences telles que la fonte des glaciers et des calottes glaciaires, la montée du niveau de la mer, l'acidification des océans, l'augmentation de la fréquence et de l'intensité des événements climatiques extrêmes tels que les tempêtes, les inondations, les sécheresses, etc.

En 2022, on a d'ailleurs pu noter que de nombreux événements climatiques graves ont frappé le monde. Ainsi l'ouragan Ian a touché la Floride en faisant de lourds dégâts, des inondations ont eu lieu en Afrique du Sud. En France et plus généralement en Europe, l'année 2022 a aussi été marquée par une période de sécheresse longue, qui s'est avérée être la 3e plus longue de son histoire. Cette situation s'explique par un déficit persistant de précipitations depuis la fin de l'hiver, aggravé par des températures extrêmement élevées qui ont encore accentué le phénomène.

D'après le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) si les émissions de gaz à effet de serre se poursuivent, il est estimé que le réchauffement planétaire atteindra 1,5°C à court terme, selon les scénarios envisagés et les modèles de trajectoire. Cette augmentation du réchauffement planétaire entraînera une intensification des événements multiples et simultanés, avec un niveau de confiance élevé. Pour éviter cela, une réduction soutenue des émissions de gaz à effet de serre doit être mise en place dans les deux prochaines décennies, ainsi que des changements perceptibles dans la composition atmosphérique dans les années à venir.

Dans le cadre de Solvabilité 2, les compagnies d'assurance et de réassurance sont tenues de prendre en compte tous les risques auxquels elles pourraient être confrontées, à court et à long terme, y compris ceux qui ne sont pas ou peu pris en compte dans le calcul du SCR. Ceci doit être intégré dans leur système de gouvernance, leur système de gestion des risques ainsi que leur propre évaluation des risques et de la solvabilité (ORSA). Et c'est donc dans cette situation économique actuelle que l'Autorité de contrôle prudentiel et de résolution (ACPR) incite les actuaires à intégrer les risques de durabilité dans l'ORSA [1] cette notion sera définie à cette section [1.3].

Bien que l'accord de Paris (COP21, 2015) prévoit de limiter l'augmentation de la température « bien en-dessous de 2°C », l'ACPR recommande de réaliser des simulations prenant en compte deux scénarios différents : une augmentation de la température de 1,5 degrés Celsius, ainsi qu'une augmentation de la température au-delà de 2 degrés Celsius. Dans le cadre de ce mémoire nous nous référerons au scénario envisagé par le GIEC (qui sera défini dans cette section [1.6]) pour prévoir l'évolution future des températures et nous nous focaliserons sur le péril sécheresse.

La conception d'un modèle de prédiction de sinistres de sécheresse est cruciale pour les compagnies d'assurance, car cela leur permettrait d'évaluer les risques et de fixer des primes appropriées pour les polices d'assurance. Cependant, cela nécessite des données de qualité sur les conditions météorologiques et les niveaux de précipitations, ce qui peut être difficile à obtenir dans certaines régions. L'impact économique des sinistres de sécheresse sur le portefeuille des compagnies d'assurance est également important à considérer, car cela peut avoir un impact significatif sur les résultats financiers. En outre, une étude de sensibilité sur la construction d'un portefeuille avec des risques modérés est essentielle pour aider les compagnies d'assurance à gérer les risques de sinistres de sécheresse tout en garantissant la pérennité de leur entreprise. Cette étude de

sensibilisation peut aider les compagnies d'assurance à mieux comprendre les risques liés à la sécheresse et à élaborer des stratégies pour gérer efficacement ces risques, en prenant en compte les facteurs économiques, environnementaux et sociaux.

1.2 Solvabilité II

Entrée en vigueur le 1er janvier 2016, la directive européenne Solvabilité II a pour but de mieux appréhender les risques effectifs auxquels les assureurs sont exposés et d'adapter les fonds propres nécessaires à leur activité. Elle vise également à harmoniser les pratiques à l'échelle européenne. Solvabilité II est articulée autour de trois piliers distincts :

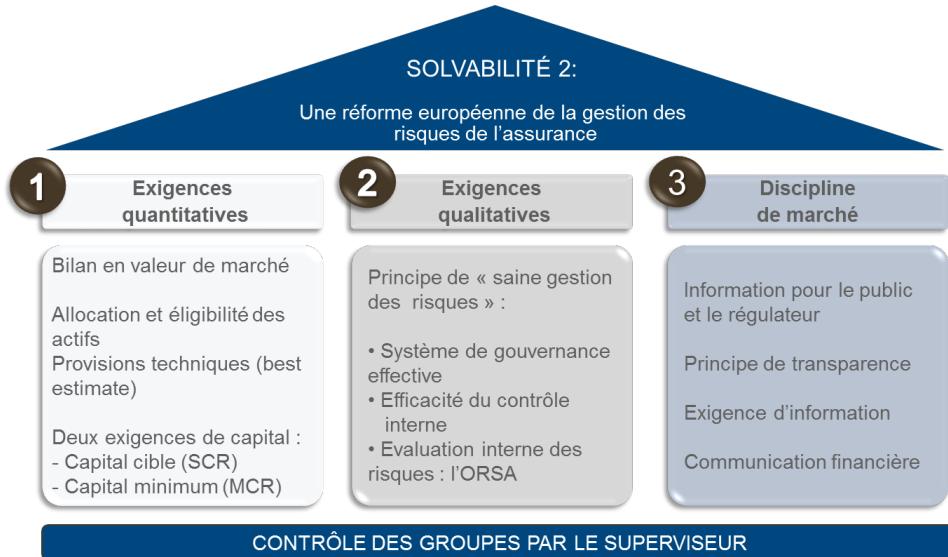


FIGURE 1.3 – Présentation rapide des 3 piliers de solvabilité II

Pilier 1 : Exigence quantitative

Le pilier 1 de Solvabilité II a pour objectif d'établir les normes de calcul des différentes composantes du bilan économique, à savoir la valorisation de l'actif, du passif et le calcul des fonds propres, dans le but d'harmoniser ces normes à l'échelle européenne.

Il est à noter que les créances clients (hors réassurance) à l'actif et les dettes au passif, considérées comme non prioritaires, sont simplifiées et ne sont pas incluses dans les formulations « actif » et « passif » du bilan économique.

Le bilan économique de Solvabilité II valorise l'actif en "fair value" afin de donner une image cohérente avec celle que les marchés financiers ont de l'entreprise. Contrairement au principe de prudence comptable en vigueur en France, la fair value permet de comptabiliser toutes les plus et moins-values, reflétant ainsi la vraie valeur de l'entreprise. En cas de hausse des actifs, les fonds propres économiques augmenteraient également, améliorant ainsi le ratio de couverture du SCR. En conséquence, l'entreprise serait moins pénalisée par son SCR en « valeur économique ».

Le passif quant à lui se compose des fonds propres, de la marge pour risque et de la meilleure estimation, qui en France sont appelés « provisions techniques prudentielles ». Ces provisions techniques sont évaluées soit en "fair value" pour les risques couvrables sur les marchés (risques hedgeables), soit en Best Estimate pour les autres risques. Cette dernière méthode repose sur des projections basées sur des hypothèses crédibles telles que le business plan de l'entreprise ou du secteur.

Suite à ces estimations le bilan comptable devient un bilan économique.

En outre, Solvabilité II impose des montants minimums de fonds propres :

- Le SCR (Solvency Capital Requirement) : Montant de fonds propres qu'un assureur se doit de détenir afin d'honorer ses engagements vis-à-vis des assurés dans 99,5% des cas sur un an. Le SCR peut être calculé soit à l'aide de la formule standard soit à l'aide d'un modèle interne développé par l'entreprise.
- Le MCR (Minimum Capital Requirement) : Montant minimal de fonds propres en-dessous duquel le superviseur intervient automatiquement afin d'éviter un défaut de paiement de l'assureur. L'agrément de l'assureur peut lui être retiré s'il ne détient pas ce niveau minimal de fonds propres. Le MCR se doit de respecter l'inégalité suivante :

$$25\% * SCR < MCR < 45\% * SCR$$

Pilier 2 : Exigence qualitative

L'objectif de Solvabilité II est d'assurer une gestion saine et prudente de l'activité des organismes d'assurance, en imposant la mise en place d'un système de gouvernance efficace.

Pour cela, il est nécessaire d'avoir une structure organisationnelle transparente, avec des responsabilités clairement réparties et séparées. Ce système de gouvernance doit être proportionné à la nature et à la complexité de l'organisme et est basé sur des politiques écrites, validées par le Conseil d'Administration/Conseil de Surveillance. Par ailleurs, le système de gouvernance doit comprendre au minimum deux dirigeants effectifs pour respecter le principe des "4 yeux", ainsi que quatre responsables de fonctions clés :

- la fonction actuarielle,
- la fonction gestion des risques,
- la fonction audit interne,
- la fonction conformité.

En outre il est également imposé d'avoir des mécanismes de contrôle interne adaptés à l'activité. L'ACPR (Autorité de Contrôle Prudentiel et de Résolution) a le pouvoir d'ajouter un capital add-on au SCR (Solvency Capital Requirement) de l'entreprise en cas d'insuffisance du contrôle interne. Autrement dit, si l'ACPR considère que le contrôle interne de l'entreprise n'est pas suffisant pour garantir une gestion saine et prudente de l'activité, elle peut exiger que l'entreprise augmente son niveau de capital pour couvrir les risques supplémentaires. Cela permet à l'ACPR de renforcer la supervision de l'entreprise et d'assurer la sécurité financière des assurés.

Cette évaluation interne des risques porte le nom de Own Risk and Solvency Assessment (ORSA). L'ORSA doit démontrer que l'organisme ou le groupe est capable d'identifier, de mesurer et de gérer les facteurs qui peuvent influencer sa solvabilité ou sa situation financière. Il s'agit d'un outil stratégique majeur qui doit être considéré comme un moyen de piloter l'activité de l'organisme en fonction des risques.

L'horizon de projection pour l'ORSA doit être d'au moins trois ans mais il est en général plutôt de 4 à 6 ans , selon les exigences de Solvabilité II. Cela signifie que les organismes d'assurance doivent projeter leurs risques et leur solvabilité pour une période de cinq ans, en tenant compte de différents scénarios de stress et en évaluant leur impact potentiel.

Pilier 3 : Communication et transparence

La directive Solvabilité II vise à instaurer une discipline sur les marchés de l'assurance en exigeant des entreprises du secteur une publication claire et transparente d'informations. L'objectif est d'harmoniser cette publication d'informations à l'échelle de l'Union Européenne (UE), afin qu'elle puisse être utilisée par le marché et les superviseurs. Cette publication est fondée sur trois principes clés :

- Les publications comptables et réglementaires pour le superviseur doivent être cohérentes
- La publication d'informations pour le régulateur européen doit être harmonisée à l'échelle de l'UE
- Les mêmes règles doivent s'appliquer à tous les assureurs

1.3 Les risques de durabilité

Les risques de durabilité, également connus sous le nom de risques ESG (environnementaux, sociaux et de gouvernance), sont des risques financiers ou non-financiers découlant de facteurs environnementaux, sociaux et de gouvernance qui peuvent avoir un impact sur la durabilité à long terme d'une entreprise, d'un secteur ou de l'économie dans son ensemble.

Le risque de durabilité peut être divisé en trois catégories distinctes : le risque de transition, le risque physique et le risque de responsabilité :

- Le risque physique regroupe l'ensemble des risques climatiques dont le nombre et l'intensité tendent à augmenter,
- Le risque de transition représente le risque de dépréciation des actifs dû à un changement de réglementation,
- le risque de responsabilité est un risque relatif au risque de poursuite judiciaire ou au risque de réputation dans le cas où une entreprise commet des actions contribuant au réchauffement climatique.

Du côté de l'actif, les risques de durabilité peuvent avoir un impact sur la performance financière à court et à long terme d'une entreprise, ainsi que sur sa réputation, sa capacité à attirer et à retenir des employés et des investisseurs, ainsi que sur les réglementations futures qui peuvent affecter l'entreprise. Par conséquent, les risques de durabilité sont de plus en plus pris en compte par les investisseurs et les régulateurs dans leurs décisions d'investissement et de surveillance.

Du côté du passif, les risques de durabilité, et notamment les risques climatiques, ont un impact significatif sur la sinistralité. En effet, les événements climatiques extrêmes tels que les inondations, les tempêtes et les sécheresses peuvent causer des dégâts importants aux biens assurés et augmenter le nombre de sinistres. Les changements climatiques peuvent également avoir des effets à plus long terme, tels que l'élévation du niveau de la mer et la montée des températures, qui peuvent entraîner des risques accrus d'inondation, d'incendie et de maladies liées à la chaleur. Enfin, les risques liés à la durabilité peuvent également affecter la solvabilité de l'entreprise d'assurance si elle ne parvient pas à adapter ses activités aux nouvelles exigences environnementales et sociales, ce qui peut entraîner une perte de clients et une baisse de la confiance des investisseurs.

Les risques climatiques

Parmi les risques de durabilité, le risque climatique fait partie des préoccupations majeures des assureurs et réassureurs. Celui-ci peut se décomposer en 5 sous risques principaux :

Inondations

L'inondation est une submersion, rapide ou lente, d'une zone habituellement hors de l'eau. Le risque d'inondation est le risque naturel le plus important en France en raison de l'ampleur des dommages qu'il peut causer, du nombre de communes touchées, de l'étendue des zones inondables et du nombre de personnes vivant dans ces zones.

Il existe de nombreux types d'inondation :

- Les inondations de plaine : La rivière sort de son lit lentement et peut inonder la plaine pendant une période relativement longue.

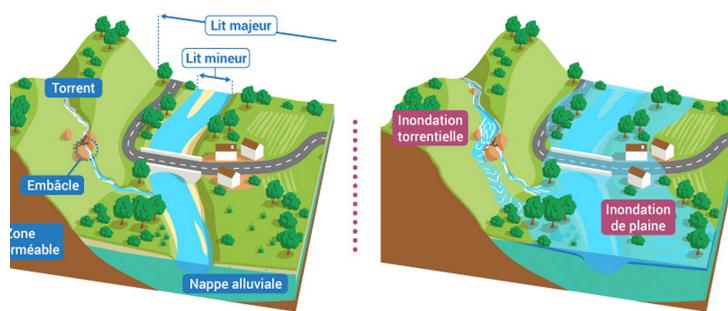


FIGURE 1.4 – Schéma de l'inondation des plaines

- Les inondations par remontée de nappe : Lorsque plusieurs années humides se succèdent, cela entraîne une montée du niveau de la nappe phréatique jusqu'à la surface du sol

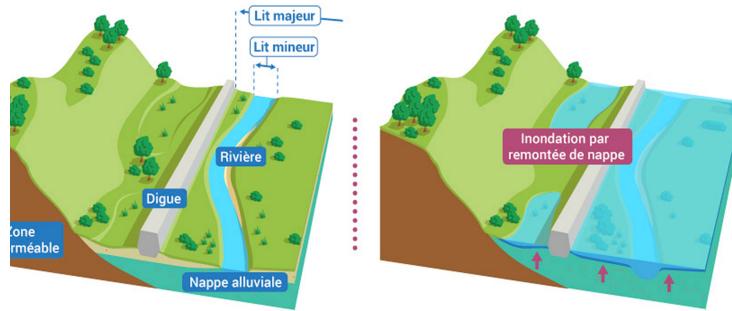


FIGURE 1.5 – Schéma de l'inondation par remontée des nappes phréatiques

- Les submersions marines sont des inondations temporaires de la zone côtière par la mer.

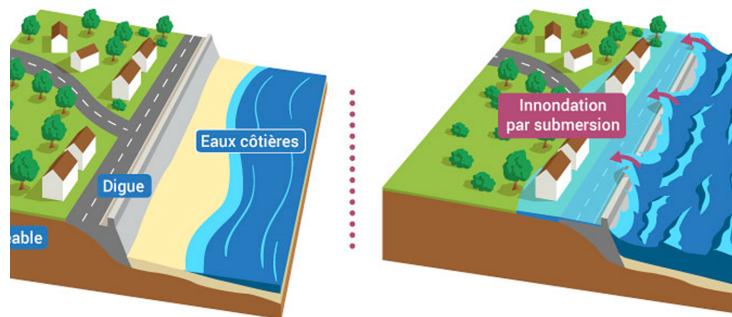


FIGURE 1.6 – Schéma de l'inondation par submersion marine

- Les crues des rivières torrentielles et des torrents.

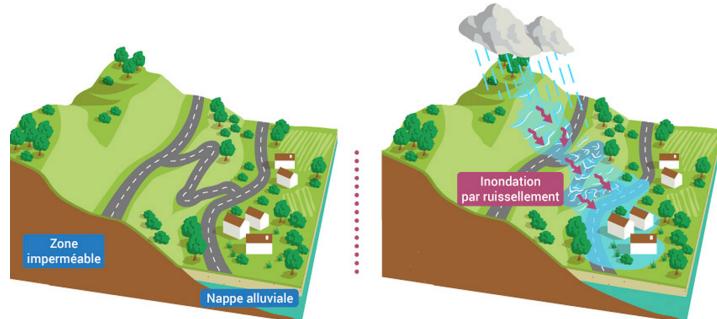


FIGURE 1.7 – Schéma de l'inondation par ruissellement de pluie

Tempêtes

Les tempêtes, telles que les ouragans, les typhons, les cyclones et les tempêtes hivernales, peuvent causer des dommages importants aux bâtiments et aux infrastructures, ainsi que des pertes de vies humaines. Les vents violents et les précipitations intenses sont des caractéristiques communes de ces phénomènes météorologiques extrêmes.

Les termes "ouragan", "cyclone" et "typhon" désignent tous le même phénomène météorologique, mais la principale différence entre eux est la région géographique où ils se produisent. Voici une explication :

- Ouragan : Utilisé dans l'océan Atlantique Nord, la mer des Caraïbes, le golfe du Mexique et l'océan Pacifique Nord-Est. Exemple : Ouragan Katrina qui a frappé les États-Unis en 2005.
- Cyclone : Utilisé dans l'océan Indien et le Pacifique Sud. Exemple : Cyclone Idai qui a touché le Mozambique en 2019.
- Typhon : Utilisé dans l'océan Pacifique Nord-Ouest, spécifiquement à l'est de l'Asie. Exemple : Typhon Haiyan (ou Yolanda) qui a frappé les Philippines en 2013.

Un ouragan se forme lorsque l'air chaud et humide au-dessus de l'océan commence à s'élever et à se refroidir, formant ainsi des nuages de tempête. À mesure que la chaleur latente est libérée pendant la condensation de l'humidité contenue dans l'air, elle génère une énorme quantité d'énergie qui alimente la tempête. L'air humide et chaud monte à travers une zone de basse pression atmosphérique, où il est renforcé par les vents convergents qui se déplacent vers le centre de la tempête. À mesure que la tempête se renforce, les vents deviennent de plus en plus violents, ce qui entraîne des conditions météorologiques extrêmes, telles que des pluies torrentielles, des vents forts et des vagues de tempête dévastatrices.

La température de l'eau chaude de l'océan est un facteur clé dans la formation d'un ouragan, car elle fournit l'énergie nécessaire à la tempête pour se renforcer. Lorsque les eaux sont suffisamment chaudes, cela peut déclencher la formation d'une tempête tropicale, qui peut évoluer en ouragan si les conditions sont favorables.

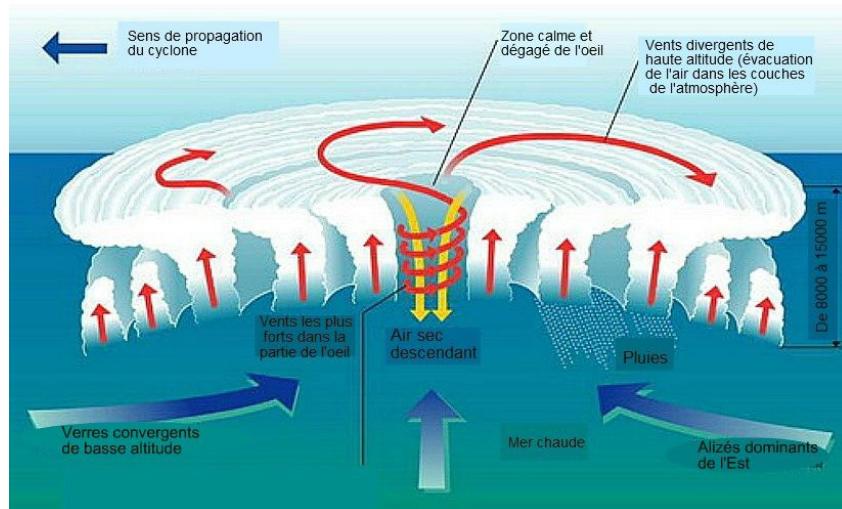


FIGURE 1.8 – Schéma de la formation d'un ouragan

Sécheresses

Les sécheresses peuvent avoir des conséquences sur les cultures, la production alimentaire, les ressources en eau, ainsi que sur les habitations. La sécheresse est un phénomène qui doit être pris en compte dans les polices d'assurance multi-risques habitation . Lorsqu'il y a un manque de précipitations et des températures élevées, cela peut entraîner une sécheresse.

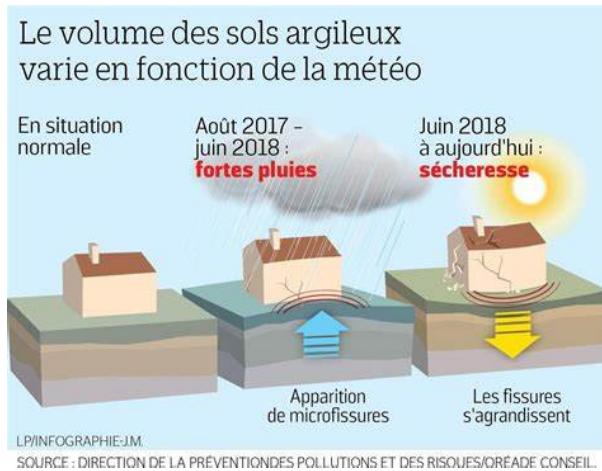


FIGURE 1.9 – Schéma de l'impact de la sécheresse sur une maison

La sécheresse est un terme global qui englobe trois sous-catégories de sécheresse :

- La sécheresse météorologique, qui se caractérise par un déficit prolongé de précipitations. Ce manque est mesuré en comparant les valeurs de précipitations aux normales, qui représentent des moyennes historiques sur une période généralement de 30 ans.
- La sécheresse édaphique, également appelée sécheresse agricole, qui fait référence au manque d'humidité des sols en surface. Dans ce contexte, les sols considérés sont ceux situés entre 1 et 2 mètres de profondeur, là où les plantes puisent l'eau nécessaire à leur développement.
- La sécheresse hydrologique, qui se manifeste par des niveaux insuffisants dans les cours d'eau, les lacs et les nappes phréatiques.

Ces différentes formes de sécheresse sont liées, mais l'une ne conduit pas nécessairement aux autres. Par exemple, de fortes précipitations peuvent ne pas suffire à nourrir adéquatement les sols.

La sécheresse des sols peut entraîner un phénomène appelé retrait-gonflement des sols argileux, également connu sous le nom de RGA. Les sols se composent de sable, d'argile ou de limons, voire d'une combinaison de ces éléments. Les argiles sont constituées de silicium Si, d'aluminium Al et d'hydroxyle OH organisés en couches qui s'empilent en feuillets, cela permet à l'eau de venir s'y loger. Lorsque les argiles absorbent l'eau, elles augmentent de volume, entraînant un gonflement du sol. À l'inverse, lorsque les argiles s'assèchent, elles se rétractent et on observe alors

un phénomène de subsidence, autrement dit le sol s'affaisse. Ce processus peut être exacerbé en cas de perturbation de la circulation de l'eau dans le sol, que ce soit par la sécheresse ou par des activités humaines.

De par la nature de ses sols, la France y est confrontée et compte une trentaine de départements ayant principalement des sols argileux. La sensibilité au phénomène de RGA varie en fonction de la composition minéralogique des argiles et des différences climatiques. Les sols argileux se rétractent en période de sécheresse et se dilatent lorsqu'ils sont hydratés. L'amplitude du changement de volume des sols dépend donc de la variation d'humidité, et pas seulement de la sécheresse elle-même. Cependant, il existe une limite au gonflement des sols, qui est leur saturation. À un certain niveau, les argiles ont absorbé la quantité maximale d'eau et le sol ne peut plus gonfler davantage. Or, la saturation des sols est fréquente en France en raison de son climat tempéré, où les sols sont généralement bien imbibés. C'est pourquoi la sécheresse est considérée comme le principal déterminant du risque de RGA. En effet, si l'on considère que les sols reviennent régulièrement à une saturation ou à un niveau proche de celle-ci, l'amplitude de l'humidité des sols est déterminée par son niveau minimum, c'est-à-dire par l'intensité de la sécheresse. La réhydratation des sols accentue les dommages, mais n'en est pas à l'origine. En France, la sécheresse est un risque majeur pour les compagnies d'assurance, car c'est le deuxième événement naturel le plus coûteux à indemniser, juste après les inondations.

Grêle

La grêle est un phénomène météorologique qui se produit lorsqu'il y a une forte convection dans les nuages, entraînant la formation de grêlons. Les grêlons sont des particules de glace qui peuvent mesurer de quelques millimètres à plusieurs centimètres de diamètre. Lorsqu'ils tombent sur le sol ou sur les véhicules, ils peuvent causer des dommages importants. En ce qui concerne l'assurance auto, les dommages causés par la grêle sont généralement couverts par la garantie "dommages tous accidents" ou "bris de glace", selon les contrats et les assureurs. [5]

1.4 Contexte légal

Le régime d'indemnisation des catastrophes naturelles (dénommé CatNat) a été créé par la loi du 13 juillet 1982. Ce nouveau régime a permis de palier une carence de couverture des risques naturels qui n'étaient que très peu assurés jusqu'alors.

Pour être éligible au régime CatNat, la catastrophe doit être considérée comme naturelle, c'est-à-dire qu'elle doit être due à un phénomène naturel violent, soudain et imprévisible. De plus, elle doit être constatée par un arrêté interministériel qui déclare l'état de catastrophe naturelle. Les biens et les personnes concernés doivent être assurés au titre de la garantie catastrophes naturelles dans leur contrat d'assurance habitation ou automobile.

Le régime CATNAT couvre les dommages matériels directs causés aux biens assurés, tels que les bâtiments, les meubles et les équipements, ainsi que les pertes d'exploitation pour les entreprises. Les dommages indirects, tels que les pertes de revenus ou les frais de relogement, ne sont pas pris en charge par le régime CATNAT. Pour être reconnue comme CATNAT il faut remplir les critères suivants :

- Que le maire ait effectué une demande et que l'état de catastrophe naturelle soit déclaré par une commission interministérielle,
- Que les biens sinistrés soient couverts par une assurance de dommage aux biens,
- Que le sinistre ait été provoqué par la catastrophe naturelle.

Ces étapes peuvent se récapituler par le schéma suivant :

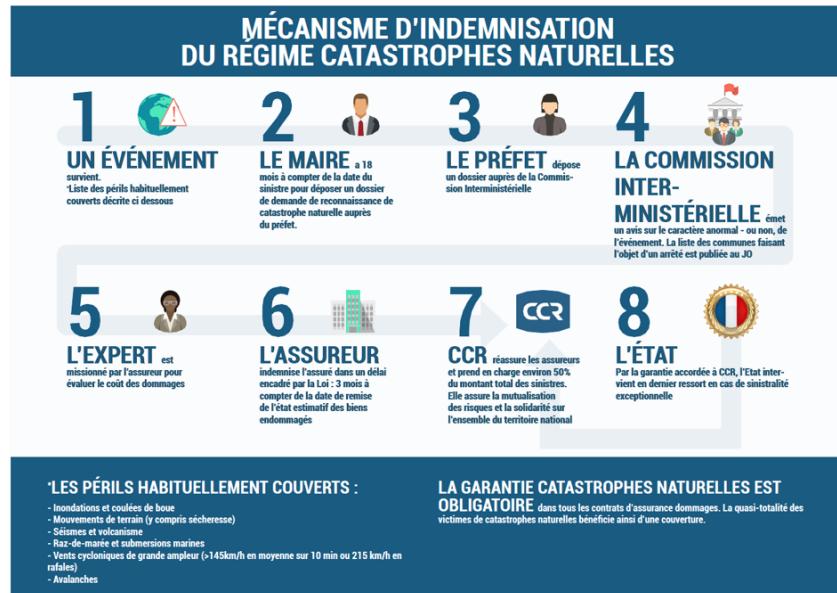


FIGURE 1.10 – Étape de la reconnaissance CatNat

Afin d'être indemnisés au titre d'un arrêté CatNat les particuliers se doivent d'être assurés. L'assurance catastrophe naturelle est une assurance qui permet d'être indemnisé pour les dommages causés par les catastrophes naturelles telles que les tremblements de terre, les inondations, les sécheresses, les glissements de terrain et l'action mécanique des vagues. Contrairement aux assurances obligatoires, elle n'est pas incluse dans les contrats d'assurance de base. Cependant, elle est incluse dans l'assurance multirisques habitation et l'assurance automobile. Il est important de noter qu'un assureur n'a pas le droit de refuser de souscrire la garantie "catastrophes naturelles".

Intéressons-nous plus particulièrement au cas de la sécheresse qui sera celui abordé au cours de ce mémoire. En France, les critères climatiques pour la reconnaissance d'une commune en tant que catastrophe naturelle liée à la sécheresse peuvent varier selon la législation en vigueur. Les critères d'éligibilité CatNat ont changé depuis 2018. Les principaux critères climatiques pris en compte actuellement sont les suivants :

- La présence de plus de 3% d'argile dans le sol de la commune,
- Que la moyenne du Soil Water Index sur 3 mois soit inférieure à la moyenne des 3 mêmes mois sur les 25 dernières années,
- Que le maire de la commune fasse la demande de reconnaissance CatNat et que celle-ci soit acceptée.

Ces critères climatiques peuvent être adaptés et actualisés au fil du temps pour refléter les réalités climatiques et les besoins de chaque région. Il est essentiel de se référer à la législation en vigueur, aux directives des organismes compétents et aux experts en gestion des risques naturels pour obtenir les informations les plus récentes et précises concernant les critères climatiques spécifiques pour la reconnaissance CatNat liée à la sécheresse en France.

Cependant il est à noter que la déclaration d'état de catastrophe naturelle est accordée par saison :

- La saison hivernale = du 1er janvier au 31 mars ,
- La saison printanière = Du 1er avril au 31 juin ,
- La saison estivale = du 1er au 31 septembre ,
- La saison automnale = du 1er octobre au 31 décembre.

Ainsi il suffit que l'un des mois de la période vérifie la condition pour que toute la saison soit reconnue CatNat. Une fois chaque zone affectée à chaque commune, on obtient notre base complète.

1.5 L'assurance MRH

Dans le cadre de ce mémoire nous nous intéresserons plus particulièrement à l'impact du risque de sécheresse dans le cadre de l'assurance MRH ou multirisques habitation.

La MRH est une assurance qui couvre les risques liés à un logement. Elle offre une protection contre les dommages causés par les incendies, les dégâts des eaux, les vols, les catastrophes naturelles, les actes de vandalisme et autres sinistres qui peuvent survenir dans une maison ou un appartement.

La MRH peut être souscrite par tout propriétaire ou locataire d'un logement. Elle permet de protéger l'habitation elle-même, ainsi que les biens qu'elle contient, tels que les meubles, les appareils électroménagers, les équipements électroniques, les vêtements, et autres objets de valeur.

En cas de sinistre, la MRH permet d'indemniser le propriétaire ou le locataire pour les pertes subies, et de prendre en charge les frais de réparation ou de remplacement des biens endommagés. Elle peut également couvrir les frais de relogement temporaire en cas d'impossibilité de résider dans le logement sinistré.

Les dommages causés par la sécheresse peuvent inclure des fissures dans les murs et les fondations, des dommages aux toits, des dommages aux structures en bois et en plâtre, ainsi que des dommages aux plantes et aux jardins.

Pour être couvert contre les dommages causés par la sécheresse, il est important de souscrire une assurance MRH. Cette garantie peut inclure une indemnisation en cas de fissures ou de dommages causés aux structures du logement, ainsi qu'une couverture pour les pertes subies dans le jardin et les plantations. Il est important de noter que les garanties spécifiques pour la sécheresse peuvent varier selon les compagnies d'assurance et les contrats proposés. La jouissance de cette garantie impose le paiement d'une surprime uniforme sur le territoire, et dont le

montant (exprimé comme un pourcentage de la prime de départ) est actuellement fixé par le gouvernement à 12% de la prime initiale. Le régime CatNat ne fait donc pas de distinction entre les individus, même s'ils ne sont pas exposés de la même manière aux catastrophes naturelles. Ce dispositif repose donc entièrement sur le principe de mutualisation.

Néanmoins l'assuré garde une partie des frais à sa charge. D'après l'article A125-1 du Code des Assurances la franchise peut être définie de la façon suivante :

Pour les biens à usage d'habitation et les autres biens à usage non professionnel, le montant de la franchise est fixé à 380 euros, sauf en ce qui concerne les dommages imputables aux mouvements de terrain différentiels consécutifs à la sécheresse et/ ou à la réhydratation des sols, pour lesquels le montant de la franchise est fixé à 1 520 euros.

1.6 Le changement climatique et les rapports du GIEC

Le 9 août 2021, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) a publié son sixième rapport d'évaluation sur l'état du changement climatique. Ce rapport représente l'étude la plus complète à ce jour sur les impacts du changement climatique, l'état actuel du climat et les projections futures, ainsi que les mesures nécessaires pour limiter le réchauffement climatique. Le rapport a mobilisé plus de 200 scientifiques de 66 pays, qui ont analysé des milliers d'études scientifiques et de données. Les conclusions sont alarmantes, confirmant que les activités humaines sont la principale cause du changement climatique et que les effets du réchauffement sont désormais inévitables, même si des mesures immédiates sont prises pour réduire les émissions de gaz à effet de serre. Le rapport souligne également l'urgence d'une action collective à l'échelle mondiale pour limiter le réchauffement à 1,5°C et pour adapter les sociétés aux impacts inévitables du changement climatique.

Le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) utilise des scénarios pour explorer différentes trajectoires d'émissions de gaz à effet de serre (GES) et de leurs impacts sur le climat. Le GIEC a utilisé deux ensembles de scénarios différents dans ses rapports d'évaluation. Le premier ensemble est le scénario Representative Concentration Pathways (RCP), qui a été utilisé dans les quatrième et cinquième rapports d'évaluation du GIEC. Les scénarios RCP décrivent des trajectoires d'émissions de GES futures qui sont basées sur des hypothèses quant à l'évolution des technologies, de la politique et de l'économie. Les scénarios RCP ont été choisis pour couvrir un large éventail de futurs possibles, allant d'un scénario où les émissions de GES augmentent considérablement jusqu'à un scénario où les émissions sont considérablement réduites.

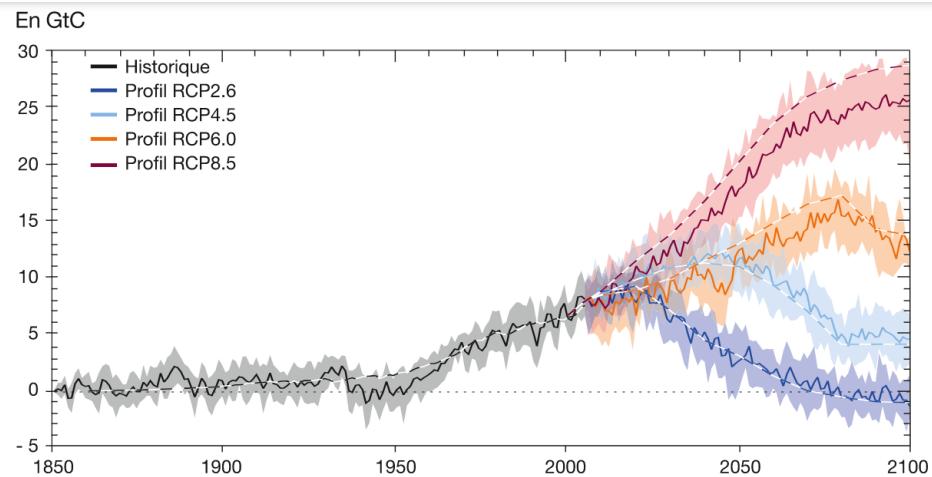


FIGURE 1.11 – Projection des émissions liées aux énergies fossiles selon les profils d'évolution des GES

Le deuxième ensemble de scénarios est le scénario Shared Socioeconomic Pathways (SSP), qui a été utilisé dans le sixième rapport d'évaluation du GIEC. Les scénarios SSP sont similaires aux scénarios RCP dans la mesure où ils décrivent des trajectoires futures possibles, mais ils ont une approche différente en ce qui concerne la façon dont ils intègrent les facteurs socioéconomiques. Les scénarios SSP décrivent des futurs possibles en fonction de différentes voies socioéconomiques, qui sont basées sur des hypothèses sur la croissance économique, la gouvernance, la technologie et la démographie. Ces scénarios incluent des trajectoires d'émissions de GES élevées sans atténuation du changement climatique, ainsi que de nouvelles trajectoires d'émissions de GES faibles. Ces 5 scénarios sont représentés ci-dessous :

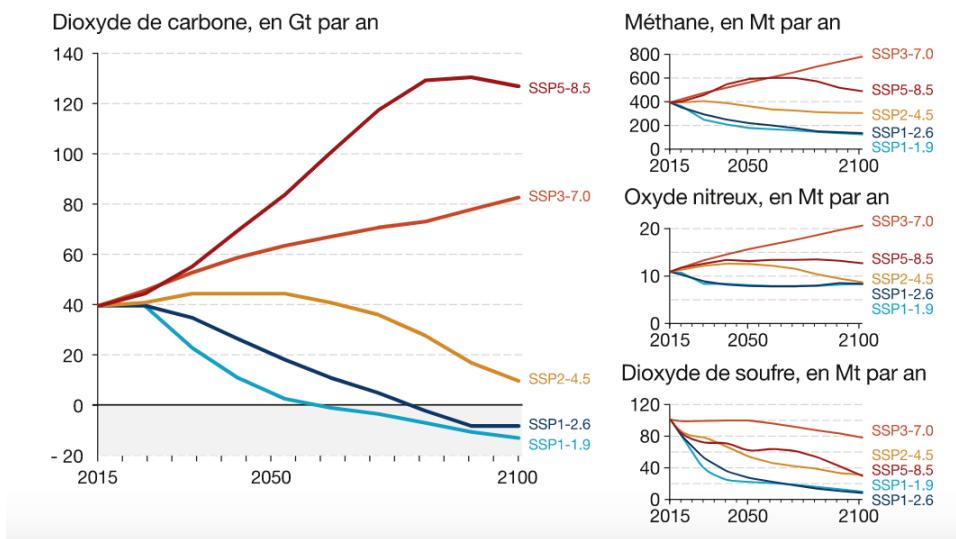


FIGURE 1.12 – Évolution des émissions de Gaz à effet de serre en fonction des différents scénarios du GIEC

Compte tenu de ces scénarios, on obtient des prévisions sur l'augmentation de la température.

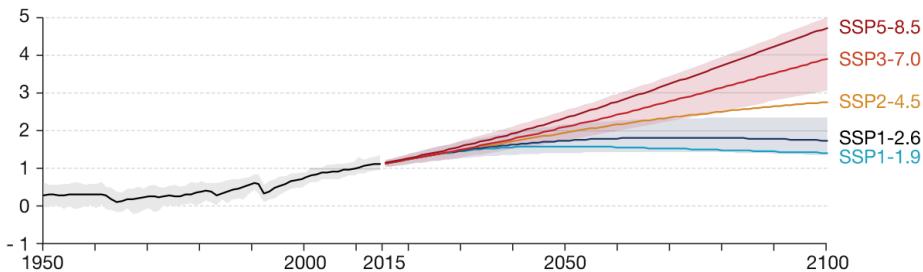


FIGURE 1.13 – Évolution de la température mondiale en fonction des différents scénarios du GIEC

- Le SSP1 (faible défi d'adaptation, faible défi d'atténuation) : correspond à un monde marqué par une forte coopération internationale, donnant la priorité au développement durable
- Le SSP2 (défi d'adaptation moyen, défi d'atténuation moyen), décrit un monde caractérisé par la poursuite des tendances actuelles.
- Le SSP3 (défi d'adaptation élevé, défi d'atténuation élevé) décrit un monde fragmenté où la compétition entre pays est prédominante. La croissance économique est lente, et les politiques sont axées sur la sécurité et la production industrielle, avec peu d'attention portée à l'environnement.
- Le SSP4 (défi d'adaptation élevé, faible défi d'atténuation) décrit un monde marqué par de grandes inégalités entre pays et en leur sein. Une minorité y serait responsable de l'essentiel des émissions de GES.
- Le SSP5 (faible défi d'adaptation, défi d'atténuation élevé) décrit un monde qui se concentre sur un développement traditionnel et rapide des pays en voie de développement, fondé sur une forte consommation d'énergie et des technologies émettrices de carbone. Cependant, la hausse du niveau de vie permettrait d'augmenter la capacité d'adaptation, notamment grâce au recul de l'extrême pauvreté.

Ces scénarios, qui prennent en compte les comportements sociaux, ne sont pas tous compatibles avec certains scénarios RCP qui ne reflètent que les émissions de gaz à effet de serre. Il est donc important de prendre conscience de ces incompatibilités pour choisir le scénario de projection le plus adapté.

	SSP1	SSP2	SSP3	SSP4	SSP5
RCP8.5	Red	Red	Green	Red	Red
RCP6.0	Red	Green	Green	Green	Green
RCP4.5	Green	Green	Green	Green	Green
RCP2.6	Green	Green	Red	Green	Red

FIGURE 1.14 – Relation entre les scénarios SSP et RCP

Considérant les caractéristiques intrinsèques des différents scénarios, le scénario SSP1, axé sur la durabilité, est conçu pour envisager un monde où les actions en faveur de l'environnement et de la durabilité sont prioritaires. Cela s'accompagne naturellement d'une réduction significative des émissions de gaz à effet de serre. À l'opposé, le scénario 8.5 est défini par une augmentation

continue et significative des émissions de GES, reflétant un monde où les préoccupations environnementales sont reléguées au second plan.

Dans ce contexte, associer le SSP1 au scénario 8.5 reviendrait à associer une stratégie de réduction maximale des émissions avec un contexte d'émissions en forte hausse. Les objectifs et les prémisses sous-jacentes de ces deux scénarios sont tout simplement incompatibles.[10]

Chapitre 2

Présentation de la base de données

2.1 Présentation de la Base de données

Construction d'une sinistralité par commune

Au cours de ce mémoire nous allons chercher à prédire le nombre de sinistres sécheresses ainsi que leurs coûts pour l'assureur.

Pour chaque commune la base est constituée de données concernant :

- La fréquence des sinistres
- Le coût moyen des sinistres
- Le Soil Wet Index (SWI)
- Des indicateurs permettant de faire état du type de bâtiment ainsi que de la hauteur du bâtiment, de la période de construction et du type de sol
- Des indicateurs climatiques

Pour cela, nous disposons de différentes bases de données publiques issues de l'Observatoire National des Risques Naturels (ONRN). Ces bases de données recensent une fourchette de la fréquence des sinistres dans une commune donnée ainsi que le coût moyen associé, sachant qu'un événement CatNat s'est produit. Ces données sont relatives à des données d'observation sur une période de 25 ans

Dans le cadre de cette étude et par mesure de précaution, nous considérerons la valeur maximale de la fourchette.

Grâce à ces bases de données et aux informations sur le nombre de maisons par commune fournies par l'INSEE, il est possible d'appliquer une formule très simple pour déterminer le nombre de maisons impactées par commune pour chaque sinistre CatNat :

Notons S le nombre de sinistres dans la commune.

Notons C la probabilité qu'une commune soit reconnue CatNat.

$$P(S|C) = \frac{FreqMoy * nbMaison}{nbSinistreCatNat} \quad (2.1)$$

On peut obtenir le nombre de logements de chaque commune sur le site de l'INSEE. [14]

Données sur les communes françaises

Pour compléter notre analyse, nous utilisons des données fournies par un partenaire. Pour une commune ces données concernent la composition en terme de hauteur des bâtiments, l'âge des bâtiments, ainsi que le pourcentage d'argile présent.

Âge des bâtiments

Concernant l'âge des bâtiments on définit des tranches d'âge de construction : la première représente les bâtiments construits avant 1970, la deuxième tranche les bâtiments construits entre 1970-1990 et la dernière tranche celle des bâtiments construits après 1990 ; les tranches sont choisies en fonction des normes de construction en vigueur durant les différentes époques. D'après l'historique, on constate que les bâtiments construits avant les années 70 ont 3 fois plus de chance de subir un sinistre en cas d'événement CatNat. On établit un score d'âge qui nous servira dans notre analyse :

$$Score = \frac{\text{Nb maison dans la tranche d'âge} * \text{freq de sinistre sur cette tranche}}{\text{Nombre de bâtiment}} \quad (2.2)$$

Pour plus de lisibilité on normalise ce score et on obtient la carte suivante :

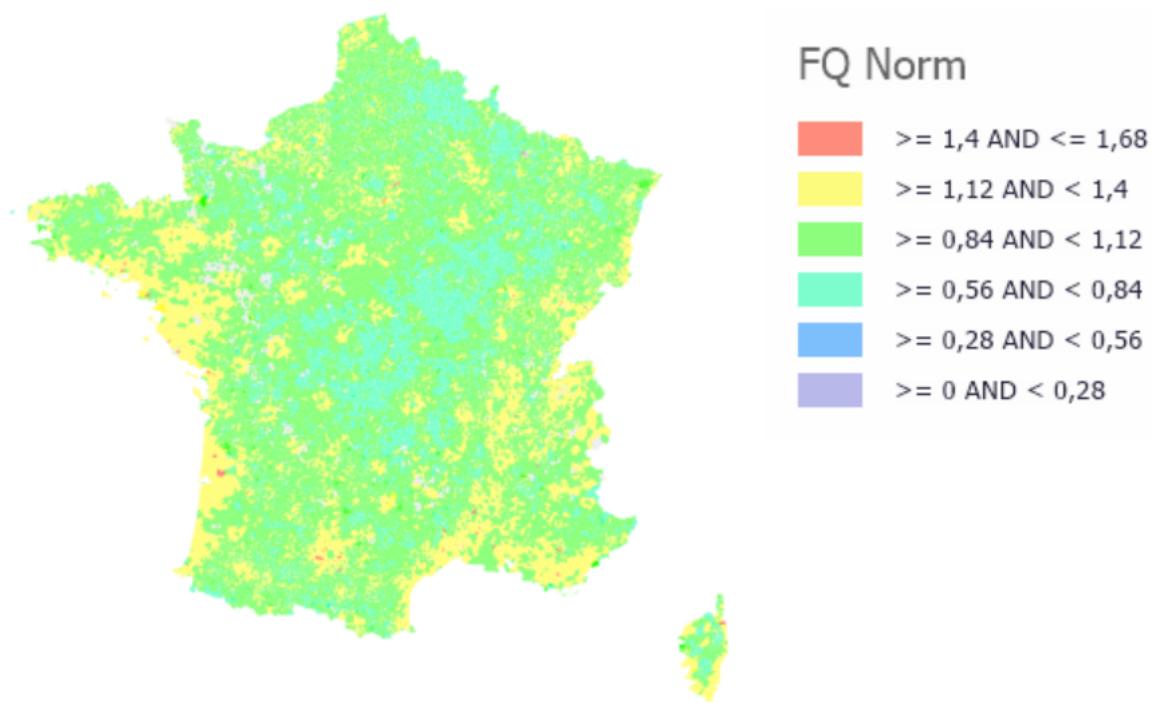


FIGURE 2.1 – Répartition du score d'âge des bâtiments sur la France

Par ailleurs, il peut s'avérer pertinent d'examiner la proportion moyenne de chaque catégorie

dans les communes. En effet, une surreprésentation d'une catégorie pourrait conduire à une analyse biaisée. Nous observons que les proportions sont presque identiquement réparties ; nous remarquons seulement une légère prédominance des bâtiments construits avant les années 70. Comme évoqué précédemment, les bâtiments construits avant les années 70 semblent être davantage affectés par les sinistres. Cependant, la légère prédominance de cette catégorie de bâtiments ne suffit pas à elle seule à expliquer cette observation.

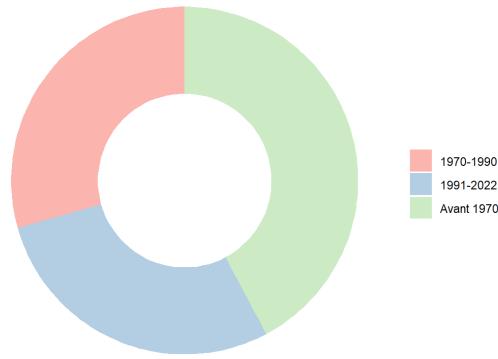


FIGURE 2.2 – Proportion de date de construction des bâtiments dans les communes

Pourcentage d'argile dans la commune

Le pourcentage d'argile par commune est issu des données du BRGM. Le Bureau de Recherches Géologiques et Minières (BRGM) fournit des données sur la présence d'argile dans différentes régions. L'argile est une composante importante des sols et joue un rôle crucial dans de nombreux aspects géologiques et géotechniques. Les données du BRGM concernant l'argile sont généralement fournies sous forme de cartes et de rapports.

Ces données mettent en évidence la répartition de l'argile dans les sols, en identifiant les zones où sa présence est significative. Le BRGM prend en compte quatre zones :

1. Zone d'exposition forte : Dans cette zone, le sol de toutes les communes sont grandement exposées au phénomène de Retrait-Gonflement des Argiles (RGA). Il y a une forte probabilité de survenance de sinistres liés à la sécheresse, et les phénomènes attendus sont d'une intensité élevée.
2. Zone d'exposition moyenne : Le sol des communes est exposé au phénomène de RGA, mais la probabilité et l'intensité des sinistres liés à la sécheresse sont moins élevées que dans la zone d'exposition forte.
3. Zone d'exposition faible : La probabilité de survenance de sinistres liés à la sécheresse est moins élevée et l'intensité des phénomènes attendus est également réduite.
4. Zone d'exposition nulle : Dans cette zone, le sol n'est pas argileux, ce qui fait qu'il n'est pas exposé au phénomène de RGA. Il s'agit d'une zone où les risques liés à la sécheresse sont absents.

Les données du BRGM sur l'argile permettent de mieux comprendre la distribution spatiale de ces différents types d'argile, ce qui peut avoir une incidence sur diverses activités humaines, telles que la construction, l'agriculture et la gestion des risques géologiques. Ces informations

sont précieuses afin de comprendre les événements de sécheresse.4

On représente sur la carte suivante l'aléa moyen présent dans chaque commune :

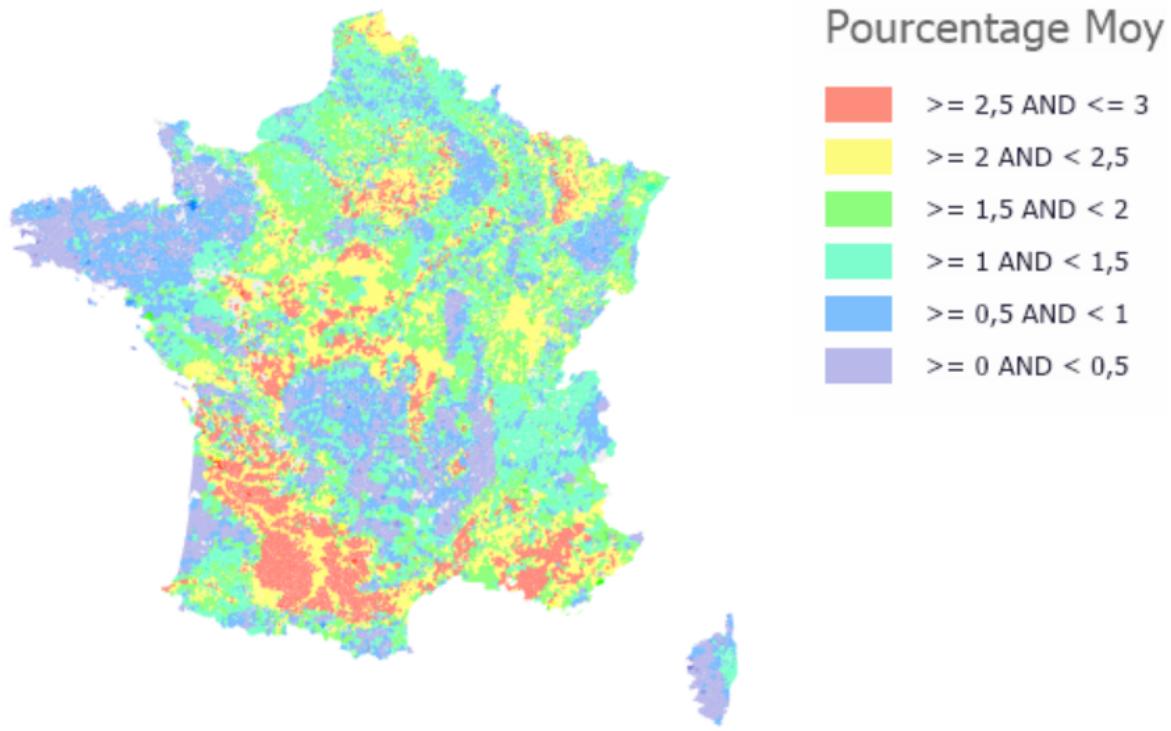


FIGURE 2.3 – Type d'argiles moyen en France

11

Le score d'une commune est déterminé par le type d'argile, multiplié par le pourcentage de surface couverte par cette argile. Par conséquent, une commune affichant un score de 2,5 est principalement composée du type d'argile 3, ce qui la place dans une catégorie à haut risque.

Hauteur

La dernière information à notre disposition concerne un score basé sur la hauteur des bâtiments. Pour calculer ce score, nous procédons à une moyenne des hauteurs des bâtiments, cette moyenne étant pondérée par le nombre de sinistres associés à chaque hauteur de bâtiment. L'analyse des sinistres révèle une tendance intéressante : les bâtiments de faible hauteur sont généralement plus susceptibles d'être affectés par des sinistres. Cette observation peut s'expliquer par le fait que, lors de la construction, plus un bâtiment est élevé, plus ses fondations sont conçues pour être profondes et solides, afin de supporter le poids supplémentaire. Ce phénomène se traduit par une meilleure résilience face à certaines catégories de sinistres. Ainsi, le score basé sur la hauteur des bâtiments offre une perspective enrichissante pour comprendre et évaluer le risque de sinistre.

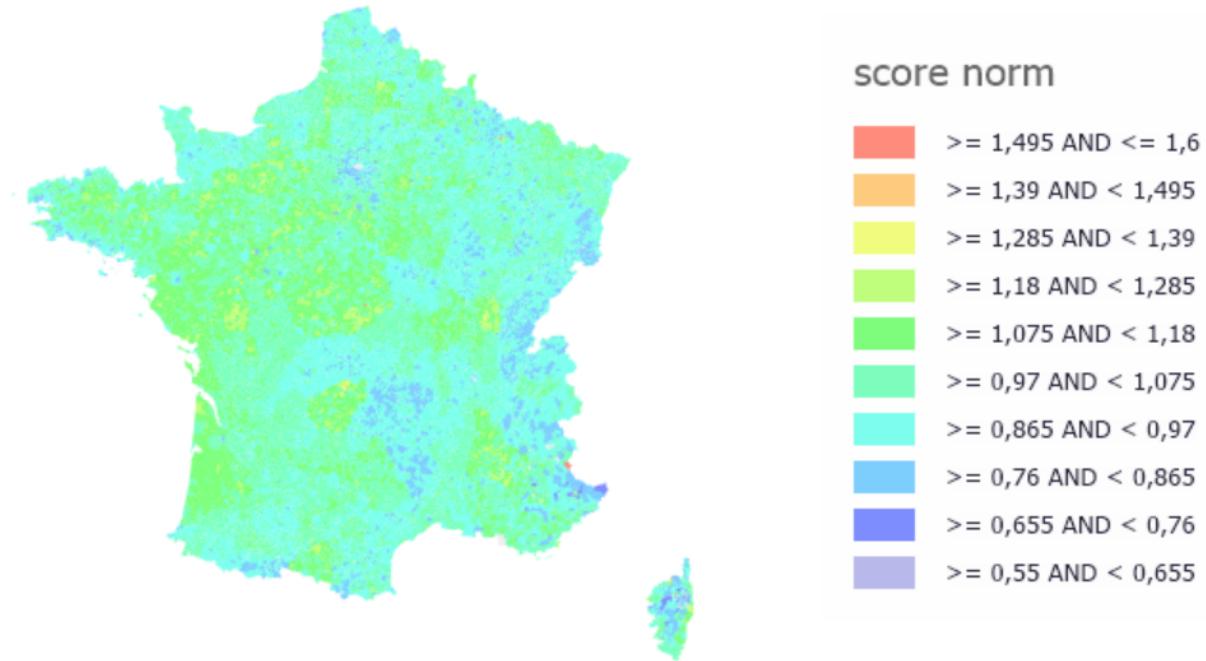


FIGURE 2.4 – Répartition du score de Hauteur des bâtiments sur la France

Indicateur Climatique

Simulation du DRIAS

Le projet "DRIAS les futurs du climat" (Donner accès aux scénarios climatiques Régionalisés français pour l'Impact et l'Adaptation de nos Sociétés et environnement) a été lancé en 2012 et a donné naissance au portail web "Drias les futurs du climat". Développé par Météo-France en collaboration avec des laboratoires français de modélisation du climat tels que le CERFACS, le CNRM et l'IPSL, ce projet est soutenu par le ministère de la Transition écologique. Son objectif principal est de faciliter l'accès et l'utilisation des modèles climatiques régionaux sur le territoire français.

Dans le cadre du projet DRIAS, les scientifiques sont amenés à remplir différentes missions dont les principales sont :

- Collecte des données : Les chercheurs rassemblent des données climatiques provenant de différentes sources, telles que les observations météorologiques, les satellites, les stations de mesure, etc. Ces données sont utilisées pour alimenter les modèles climatiques régionaux.
- Sélection des modèles climatiques mondiaux (GCM) parmi l'ensemble CMIP5 et des modèles climatiques régionaux (RCM) parmi l'ensemble Euro-CORDEX.
- Application d'une méthode de correction des biais des modèles à partir des données historiques, permettant d'ajuster les distributions simulées aux observations réelles.
- Calcul d'indicateurs à partir des données brutes, telles que le nombre moyen de jours de pluie par an, en utilisant les précipitations quotidiennes sur la période étudiée.
- Simplification du téléchargement des données.

Afin de prendre en compte les incertitudes liées aux modèles climatiques dans les projections

climatiques futures, il est essentiel de travailler avec un ensemble étendu de simulations de la meilleure qualité possible pour la zone d'intérêt. C'est pourquoi, en 2014, le portail DRIAS a inclus un jeu de données basé sur onze couples GCM/RCM provenant de l'ensemble Euro-CORDEX. Ces simulations constituaient alors les premières disponibles pour l'Europe à une résolution de 12 km.

Depuis 2014, le nombre de simulations disponibles dans l'initiative Euro-CORDEX a considérablement augmenté. De plus, des évaluations des premières simulations utilisées en 2014 ont révélé des anomalies pour certaines d'entre elles, qui ont été documentées dans l'espace d'accompagnement DRIAS.

Ainsi, la production du nouveau jeu de données, DRIAS 2020, vise à tirer parti des nouvelles simulations réalisées dans le cadre d'Euro-CORDEX au cours des dernières années. De plus, il s'agit de mettre à jour les simulations de 2014 en utilisant des versions corrigées qui ont été développées. Cette démarche permet d'améliorer la qualité et la fiabilité des projections climatiques disponibles sur le portail DRIAS, offrant ainsi une base plus solide pour l'étude des impacts futurs du changement climatique.

L'ensemble des simulations disponibles est résumé ci-dessous :

Nom de la simulation	Institution	GCM	RCM	Scénarios	Périodes disponibles	Variabil
CNRM-CERFACS-CNRM-CM5_CNRM-ALADIN63	CNRM	CNRM-CM5	ALADIN63	RCP8.5, RCP4.5, RCP2.6	1951-2100	9
MPI-M-MPI-ESM-LR_CLMcom-CCLM4-8-17	CLMcom	MPI-ESM	CCLM4-8-17	RCP8.5, RCP4.5, RCP2.6	1950-2100	7
MOHC-HadGEM2-ES_ICTP-RegCM4-6	ICTP	HadGEM2	RegCM4-6	RCP8.5, — , RCP2.6	1970-2099	7
ICHEC-EC-EARTH_SMHI-RCA4	SMHI	EC-EARTH	RCA4	RCP8.5, RCP4.5, RCP2.6	1970-2100	7
IPSL-IPSL-CM5A-MR_IPSL-WRF381P	IPSL	IPSL-CM5A	WRF381P	RCP8.5, RCP4.5, —	1951-2100	7
NCC-NorESM1-M_GERICS-REMO2015	GERICS	Nor-ESM1	REMO2015	RCP8.5, — , RCP2.6	1950-2100	7
MPI-M-MPI-ESM-LR_MPI-CSC-REMO2009	CSC	MPI-ESM	REMO2009	RCP8.5, RCP4.5, RCP2.6	1970-2100	7
MOHC-HadGEM2-ES_CLMcom-CCLM4-8-17	CLMcom	HadGEM2	CCLM4-8-17	RCP8.5, RCP4.5, —	1950-2099	7
ICHEC-EC-EARTH_KNMI-RACMO22E	KNMI	EC-EARTH	RACMO22E	RCP8.5, RCP4.5, RCP2.6	1950-2100	9
IPSL-IPSL-CM5A-MR_SMHI-RCA4	SMHI	IPSL-CM5A	RCA4	RCP8.5, RCP4.5, —	1970-2100	7
CNRM-CERFACS-CNRM-CM5_KNMI-RACMO22E	KNMI	CNRM-CM5	RACMO22E	RCP8.5, RCP4.5, RCP2.6	1950-2100	9
NCC-NorESM1-M_DMI-HIRHAM5	DMI	Nor-ESM1	HIRHAM5 v3	RCP8.5, RCP4.5, —	1951-2100	7

FIGURE 2.5 – Liste des projections climatiques composant l'ensemble DRIAS-2020

La mise à jour de l'ensemble Euro-CORDEX pour le portail DRIAS a été réalisée avec plus de trente couples GCM/RCM disponibles sur le domaine EUR-11 depuis le printemps 2019. Ces projections climatiques sont le fruit des expériences menées dans le cadre du projet d'inter-comparaison de modèles couplés (CMIP) pour étudier le climat mondial, ainsi que des nombreux modèles régionaux utilisés dans le cadre du projet CORDEX pour affiner les modèles mondiaux à l'échelle régionale.

Cependant, le nombre d'expériences réalisées avec chaque RCM dépend des financements disponibles et des ressources informatiques des centres régionaux de modélisation. Par conséquent, les GCM et les RCM ne sont pas également répartis dans l'ensemble, ce qui a un impact significatif sur les caractéristiques des données. De plus, les simulations sont publiées de manière progressive, ce qui signifie que le moment où l'analyse est effectuée influence les résultats.

En outre, certaines contraintes de calcul dans certains centres ou bureaux d'études limitent généralement l'utilisation de l'ensemble complet, ce qui justifie la réduction du nombre de simulations climatiques dans cet ensemble. Cependant, il est important de maintenir une dispersion équilibrée du signal climatique couvert par les modèles. C'est la démarche adoptée par le portail DRIAS, qui propose une nouvelle sous-sélection du jeu Euro-CORDEX.

C'est pourquoi il a fallu appliquer des critères pour sélectionner des scénarios. Ces critères sont les suivants :

1. Sélectionner environ 12 couples GCM / RCM
2. GCMs jugés réalistes sur l'Europe (Mc Sweeney et al. 2015)
3. Diversité des RCMs
4. Simulations disponibles pour au moins 2 scénarios d'émissions RCP (Representative Concentration Pathway)
5. Rejeter les couples GCM/RCM concernés par une erreur connue
6. Inclure les simulations réalisées par des centres de recherche ou instituts français
7. Optimiser la dispersion du changement climatique simulé par les couples sélectionnés
8. Favoriser des couples avec une cohérence de physique entre le GCM et le RCM

Suite à cette sélection on conserve uniquement les couples GCM/RCM suivants :

GCM	RCM	HISTO	RCP2.6	RCP4.5	RCP8.5
CNRM-CM5	Aladin63 V2				
CNRM-CM5	Racmo22E v2				
IPSL-CM5A-MR	WRF381P				
IPSL-CM5A-MR	RCA4				
HadGEM2-ES	RegCM4-6				
HadGEM2-ES	CCLM4-8-17				
EC-EARTH	Racmo22E v2				
EC-EARTH	RCA4				
MPI-ESM-LR	CCLM4-8-17				
MPI-ESM-LR	REMO*				
NorESM1-M	HIRHAM5 v3				
NorESM1-M	REMO**				

* REMO 2009 ; ** REMO 2015

FIGURE 2.6 – Liste des couples GCM/RCM sélectionnés

Ce qui nous permet d'avoir :

- 12 modèles simulés aux conditions historiques
- 8 modèles simulés selon le scénario RCP 2.6
- 10 modèles simulés selon le scénario RCP 4.5
- 12 modèles simulés selon le scénario RCP 8.5

Ces modèles permettent de simuler sept variables quotidiennes, à savoir les températures moyennes, la température minimale et maximale, l'humidité spécifique près de la surface, les précipitations totales et neigeuses, ainsi que la vitesse du vent.

Les données sont simulées selon 4 horizons :

- une période horizon proche : les simulations sont effectuées sur la période 2021-2050,
- une période horizon moyen : les simulations sont effectuées sur la période 2041-2070,
- une période horizon lointain : les simulations sont effectuées sur la période 2071-2100.
- Les simulations historiques : les simulations sont effectuées sur la période 1976-2005. Les données historiques disponibles sur le portail DRIAS ne sont pas basées sur des observations réelles. Au lieu de cela, elles proviennent de simulations du passé, où les modèles sont exécutés avec les émissions de gaz à effet de serre de l'époque correspondante. Cette approche est nécessaire pour éviter l'introduction de biais dans les analyses.

Le Drias poursuit également en effectuant une correction statistique des simulations cette correction sera détaillé en annexe

Indicateur Calculé dans le cadre du projet DRIAS

Suite à toutes ces corrections, nous disposons désormais d'un ensemble de données simulées complet comprenant les variables suivantes :

- TN_i : La température minimale quotidienne du jour i.
- TX_i : La température maximale quotidienne du jour i.
- $TM_i = \frac{TX_i + TN_i}{2}$: La température moyenne quotidienne du jour i.
- TNR_i : La température minimale quotidienne de référence du jour i (calculée sur la période de référence).
- TXR_i : La température maximale quotidienne de référence du jour i (calculée sur la période de référence).
- RR_i : Les précipitations quotidiennes du jour i.

A partir de ces simulations on peut définir les indicateurs qui seront utilisés dans notre modélisation :

- NORTAV : Température moyenne journalière ($^{\circ}\text{C}$) : $TM = \sum_{i=1}^N \frac{TM_i}{N} =$
- NORTNAV : Température minimale journalière ($^{\circ}\text{C}$) : $TN = \sum_{i=1}^N \frac{TN_i}{N} =$
- NORTXAV : Température maximale journalière ($^{\circ}\text{C}$) : $TX = \sum_{i=1}^N \frac{TX_i}{N} =$
- NORSD : Nombre de journées d'été (jour(s)) Nombre de jours pour lesquels : $TX_i > 25C$
- NORTX35 : Nombre de jours de forte chaleur $TX_i > 35C$ (jour(s))
- NORTXND : Nombre de jours anormalement chauds (jour(s)) : Nombre de jours pour lesquels $TX_i > TXR_i + 5C$

- NORTNHT : Nombre de nuits anormalement chaudes (jour(s)) : Nombre de jours pour lesquels $TN_i > TNR_i + 5C$
- NORTR : Nombre de nuits tropicales (jour(s)) : Nombre de jours pour lesquels $TN_i > 20C$
- NORTNFD : Nombre de jours de gel (jour(s)) : Nombre de jours pour lesquels $TN_i < 20C$
- NORTNND : Nombre de jours anormalement froids (jour(s)) : Nombre de jours pour lesquels $TN_i < TNR_i - 5C$
- NORTXFD : Nombre de jours sans dégel (jour(s)) : Nombre de jours pour lesquels $TX_i \leq 0C$
- NORTNCWD : Nombre de jours d'une vague de froid (jour(s)) : Nombre de jours où $TN_i < TNR_i - 5C$ dans une séquence de plus de 5 jours consécutifs
- NORTXHWD : Nombre de jours d'une vague de chaleur (jour(s)) : Nombre de jours où $TX_i > TXR_i - 5C$ dans une séquence de plus de 5 jours consécutifs
- NORTRAV : Amplitude thermique journalière ($^{\circ}C$) : Moyenne des amplitudes thermiques quotidiennes de la période $A = \sum_{i=1}^N \frac{TX_i - TN_i}{N}$
- NORTXQ90 : Extrême chaud de la température maximale journalière ($^{\circ}C$) : 90ème centile des TX_i
- NORTXQ10 : Extrême froid de la température maximale journalière ($^{\circ}C$) : 10ème centile des TX_i
- NORTNQ10 : Extrême froid de la température minimale journalière ($^{\circ}C$) : 10ème centile des TN_i
- NORTNQ90 : Extrême chaud de la température minimale journalière ($^{\circ}C$) : 90ème centile des TN_i
- NORHDD : degré-jours de chauffage (C) : Cumul sur la période des écarts négatifs au seuil de $17C$ par la température quotidienne moyenne : $DG17 = \sum_{i=1}^N (17 - TM_i) si TM_i \leq 17C$
- NORCDD : degré-jours de climatisation ($^{\circ}C$) : Cumul sur la période des dépassements du seuil de $18^{\circ}C$ par la température quotidienne moyenne $DG17 = \sum_{i=1}^N (TM_i - 18) si TM_i > 18C$
- NORPAV : Précipitations journalières moyennes (mm) : Moyenne des précipitations quotidiennes de la période $RRM = \sum_{i=1}^N \frac{RR_i}{N}$
- NORRR : Cumul de précipitations (mm) : Cumul sur la période des précipitations $RRC = \sum_{i=1}^N RR_i$
- NORPFL90 : Fraction des précipitations journalières intenses (%) : 90e centile des précipitations annuelles F_{90}
- NORTPSPI : Temps passé en sécheresse météorologique : Pour évaluer la fréquence des événements, nous utilisons différents indicateurs en fonction de la période considérée. Pour la fréquence mensuelle, nous utilisons le SPI-1 de chaque mois. Pour la fréquence saisonnière, nous utilisons le SPI-3 de février (pour l'hiver), de mai (pour le printemps), d'août (pour l'été) et de novembre (pour l'automne). Enfin, pour la fréquence annuelle, nous utilisons le SPI-12 de décembre.

Pour chaque année sur une période de 30 ans, nous évaluons si le SPI correspondant est inférieur au seuil de -1,28. Ce seuil représente une sécheresse avec une période de retour de 10 ans en utilisant le climat de la période historique comme référence.

- NORPINT : Précipitation moyenne les jours pluvieux (mm) : on désigne N_w le nombre de jours de pluie N_w il s'agit du nombre de jours où il y a eu plus d'un millimètre de pluie.

Et le nombre de précipitation moyenne les jours pluvieux $RRP = \frac{1}{N_w} \sum_{i=1}^N RR_i$ si $RRI > 1mm$

- NORHUSAV : Humidité spécifique moyenne (g/kg) : Moyenne des humidités spécifiques moyennes quotidiennes de la période

Cependant, bien que ces données proviennent de points de mesure précis, il a été nécessaire d'attribuer des valeurs spécifiques à chaque commune. Plus concrètement, la démarche consistait à identifier pour chaque commune le point de mesure le plus proche en se basant sur ses coordonnées de longitude et de latitude. Cette approche assure que chaque commune est représentée par une donnée basée sur la proximité géographique. L'idée sous-jacente était simple mais essentielle : garantir que chaque commune soit représentée de manière aussi précise que possible en se basant sur le point de mesure le plus proche.

Afin d'enrichir notre base de données et surtout de pouvoir prédire l'apparition d'événements CatNat il nous faut prendre en compte des indicateurs climatiques pour prendre en compte l'humidité des sols, dans notre cas on utilisera le Soil Water Index (SWI).

Le SWI, également appelé indice d'humidité des sols permet d'évaluer l'état de la réserve en eau du sol par rapport à sa réserve utile, sur une profondeur d'environ deux mètres. Il mesure donc l'humidité du sol superficiel, et non le niveau des nappes phréatiques. Un SWI de zéro indique un sol très sec, où les plantes ne peuvent plus puiser d'eau, tandis qu'un SWI d'une valeur de 1 indique un sol saturé d'eau, atteignant sa réserve utile.

Bien qu'il soit techniquement possible de mesurer le contenu en eau dans différentes couches de sol et d'estimer l'indice d'humidité des sols, cela nécessite des dispositifs d'observation complexes et coûteux. Seules quelques stations réalisent des observations en temps réel de l'indice d'humidité des sols sur le territoire métropolitain. De plus, ces observations locales sont difficilement extrapolables à de vastes territoires, car elles ne sont que peu ou pas représentatives de grandes surfaces.

C'est pourquoi Meteo-France estime le SWI en utilisant une méthode basée sur la modélisation numérique. La représentation de la météo et du climat nécessite une prise en compte réaliste d'une large gamme de processus physiques, y compris les échanges d'eau entre le sol et l'atmosphère tels que l'évaporation, l'infiltration, l'interception de l'eau par la végétation, le ruissellement en surface, le drainage vers les couches profondes du sol, etc. Les modèles numériques utilisés par Meteo-France pour les prévisions météorologiques et les simulations climatiques intègrent donc un module représentant ces échanges complexes.

Une simulation a dû être effectuée de façon quotidienne en utilisant la formule suivante :

$$SWI = \frac{W - W_{wilt}}{W_{fc} - W_{wilt}} \quad (2.3)$$

Avec W le contenu intégré en eau du sol, W_{wilt} le contenu en eau au point de flétrissement et W_{fc} le contenu en eau du sol à la capacité au champ.

Le SSWI, ou Indice Standardisé de Sécheresse du Sol, est un indicateur utilisé pour évaluer le degré de sécheresse des sols. Cet indice est basé sur l'indice d'humidité du sol (SWI) et exprime les probabilités de manière standardisée. Un SSWI de 0 indique des précipitations moyennes par rapport à une climatologie de référence établie sur une période de 30 ans. L'indice SSWI est négatif lorsqu'il indique une sécheresse et positif lorsqu'il indique des conditions humides.

A partir de ces estimations journalières on obtient un SWI mensuel en effectuant une moyenne des jours du mois :

$$\overline{SWI_m} = \sum_j^{31} SWI_{j,m} \quad (2.4)$$

Avec m le nombre de mois, m ∈ 1; 2.....; 12 et j le nombre de jours dans le mois.

Dans le contexte de la sécheresse, les données mensuelles ne suffisent pas à capturer pleinement le phénomène du retrait-gonflement des sols (RGA) car celui-ci est de nature assez long. Afin d'illustrer ce phénomène de manière plus précise, une approche alternative est adoptée par l'État. Elle consiste à définir la valeur pour un mois donné comme étant la moyenne du SWI de ce mois et des deux mois précédents. Cette approche permet de mieux appréhender les variations du SWI sur une période plus étendue et de mieux évaluer l'impact de la sécheresse sur le retrait-gonflement des sols. On appellera l'indice retenu le SWI_{Moy} .

En 2018, le SWI a été adopté pour remplacer le Standardized Precipitation Index (SPI) dans la reconnaissance de l'état de catastrophe naturelle (CatNat). Ce changement a été motivé par la valeur ajoutée apportée par le SWI par rapport au SPI. Le SWI présente des avantages significatifs par rapport au SPI. Il permet une évaluation plus complète de l'état hydrique des sols en prenant en compte non seulement les précipitations, mais aussi d'autres facteurs tels que l'évaporation, l'infiltration et le ruissellement. Cela offre une vision plus précise de l'impact des événements météorologiques sur les terres, en permettant notamment de détecter les zones touchées par une saturation excessive des sols ou par une sécheresse prolongée.

[6]
Actuellement, l'indice SWI est utilisé pour déterminer si une commune sera reconnue en tant que CatNat. En effet, l'un des critères requis dans les normes actuelles de reconnaissance des catastrophes est que la valeur moyenne du SWI sur une période de 3 mois soit inférieure à la valeur moyenne du SWI pour les mêmes 3 mois sur les 25 dernières années. Cependant, des discussions sont en cours pour modifier ce critère. Il est envisagé de prendre en compte la moyenne annuelle du SWI et de la comparer à la moyenne annuelle du SWI des 10 dernières années.

Le Standardized Soil Wetness Index (SSWI) est dérivé du SWI classique en utilisant une méthode similaire au Standardized Precipitation Index (SPI). Il projette le SWI classique sur une distribution normale centrée réduite. Le SSWI partage le même pas de temps que le SWI, ce qui permet de le calculer sur une base mensuelle. Il est utilisé pour évaluer la sécheresse du sol et représente un indice de probabilité basé sur l'indice d'humidité du sol (SWI). Les probabilités sont standardisées, de sorte qu'un SSWI de 0 indique une quantité de précipitation médiane par rapport à une climatologie moyenne de référence établie sur 30 ans. Un indice SSWI négatif indique des conditions de sécheresse, tandis qu'un indice positif indique des conditions humides (McKee et al., 1993). Ainsi si les critères de reconnaissance CatNat en terme de SWI sont vérifiés ils sont également vérifiés pour le SSWI. Cependant le SSWI comporte quelques avantages comparé au SWI qui sont les suivants :

- il est plus facilement interprétable,
- il requiert seulement la moyenne mensuelle du SWI,
- il permet l'intercomparaison de régions aux climats différents,
- la normalisation de l'indice permet de déterminer la rareté d'une sécheresse.

Les indicateurs relatifs au SWI simulés par le DRIAS sont les suivants :

- NORSWIAV : Indicateur moyen d'humidité des sols
- NORSWI04 : Nombre de jours avec $\text{SWI} < 0.4$ (jour)
- NORTPSSWI : Temps passé en sécheresse des sols (jour)

Le dernier indice que l'on utilisera est l'évapotranspiration. L'évapotranspiration regroupe l'évaporation de l'eau du sol et la transpiration des plantes, qui est le rejet d'une grande partie de l'eau absorbée par les racines pour refroidir la plante.

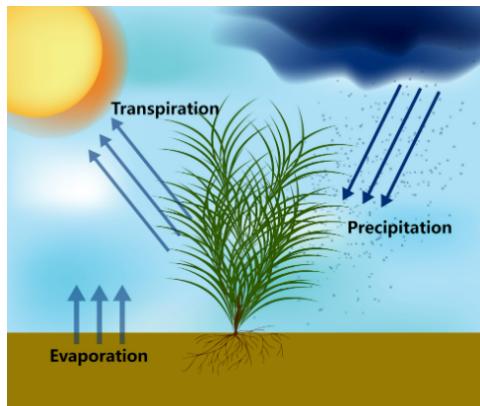


FIGURE 2.7 – Schéma du phénomène d'évapotranspiration

Combiner ces deux phénomènes est pratique car les mesurer indépendamment est complexe. Le rapport entre évaporation et transpiration varie considérablement en fonction des conditions climatiques et de la croissance végétale. L'évapotranspiration joue un rôle majeur dans la différence entre sécheresse météorologique et agricole, car environ deux tiers des précipitations retournent dans l'atmosphère de cette manière, tandis que le reste alimente les écoulements de surface et souterrains, qualifiés de précipitations efficaces.

La nature du sol et la végétation influencent fortement l'évapotranspiration. Les surfaces micro-relief du sol, comme un sol nu, peuvent entraîner une évaporation plus importante que des étendues d'eau de taille équivalente. De même, les forêts rejettent une grande quantité d'eau dans l'atmosphère, faisant d'elles de véritables "rivière volantes" qui peuvent influencer les précipitations à des milliers de kilomètres. Le couvert végétal (qui désigne l'ensemble des plantes : arbres, herbe) interagit également avec l'eau du sol par interception, où les végétaux retiennent une partie de l'eau de pluie sur leurs feuilles, ralentissant ainsi le ruissellement et favorisant l'absorption par le sol.

L'évapotranspiration est plus importante lorsque la température augmente, mais lors d'épisodes de sécheresse avec de fortes chaleurs, elle reste relativement faible en raison du manque d'eau dans le sol. Cela peut renforcer l'intensité et l'étendue géographique de la sécheresse, car l'évapotranspiration est restreinte. Il convient de distinguer l'évapotranspiration réelle (ETR) de l'évapotranspiration potentielle (ETP), qui représente l'évapotranspiration théorique d'un sol suffisamment humide avec un couvert végétal bas et homogène, et qui varie positivement avec les températures.

Il existe plusieurs équations pour calculer l'évapotranspiration potentielle (ETP) en fonction des

variables météorologiques disponibles, et leur efficacité a été évaluée en vue d'applications en modélisation hydrologique (Oudin, 2004). Actuellement, la formule préconisée à l'échelle internationale repose sur l'approche Penman-Monteith, avec des coefficients déterminés par la FAO. Celle-ci se calcule selon la formule suivante :

$$ET_0 = \frac{0,408 * \Delta * (R_n - G) + Y \frac{900}{T_{moy} + 273} ff_{2m} * \max(e_s - e_a; 0)}{\Delta + y(1 + 0,34ff_{2m})} \quad (2.5)$$

avec :

- ET0 référence évapotranspiration [mm/jour]
- R_n Rayonnement net de la surface du couvert [MJ m⁻² jour⁻¹]
- G densité de flux de conduction dans le sol [MJ m⁻² jour⁻¹]
- T température journalière moyenne à 2m [°C]
- ff_{2m} vitesse du vent à 2m [ms⁻¹]
- e_s pression saturante de vapeur [kPa] ea pression réelle de vapeur [kPa]
- e_a déficit hydrique de l'air [kPa]
- Δ pente de courbe de saturation [kPa °C⁻¹]
- y constante psychrométrique [kPa °C⁻¹]

Les étapes pour arriver à cette formule sont données en annexe. 

Les indicateurs considérés relatifs à l'évapotranspiration sont les suivants :

- NORETRC : ETR cumulée (mm)
- NORCWBC : Pluie efficace cumulée (mm)

La Caisse Centrale de Réassurance

En dernier lieu pour entraîner notre modèle à reconnaître si oui ou non une commune est reconnue CatNat, il nous a fallu l'historique des communes reconnues CatNat sur la période de référence de calcul de nos indices. Pour cela, nous faisons appel aux données de la Caisse Centrale de Réassurance.

La Caisse Centrale de Réassurance (CCR) joue un rôle crucial en matière de gestion des risques liés aux catastrophes naturelles telles que les sécheresses. En tant qu'institution publique française spécialisée dans la réassurance, la CCR intervient en tant que réassureur des assureurs opérant sur le territoire français. Lorsque des événements climatiques tels que les sécheresses surviennent, ils peuvent engendrer d'importants dégâts matériels et économiques, et mettre en péril la stabilité financière des compagnies d'assurance. C'est là que la CCR entre en jeu pour réduire le fardeau financier qui pèse sur ces assureurs. En se portant garant de leur portefeuille d'assurance, la CCR les aide, en échange d'une partie de la prime d'assurance à faire face aux coûts élevés des indemnisations et contribue à maintenir la confiance du public dans le système d'assurance.

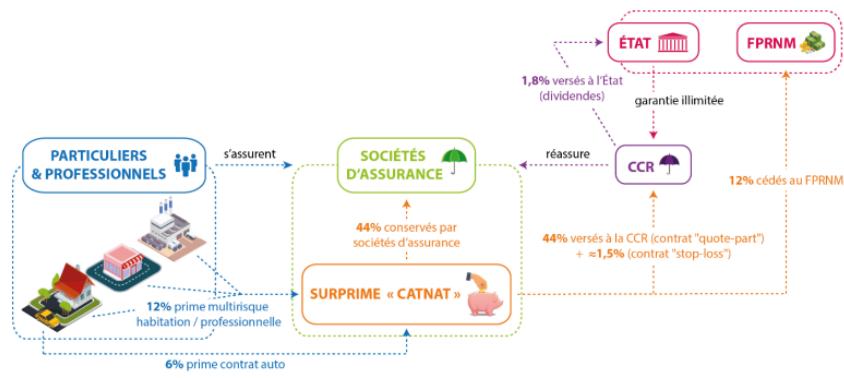


FIGURE 2.8 – Schéma du fonctionnement du régime CatNat et rôle de la CCR

Outre son rôle de réassureur, la Caisse Centrale de Réassurance joue également un rôle clé dans la diffusion de l'information concernant les arrêtés CatNat. En effet, suite à la survenue d'une catastrophe naturelle, le gouvernement français peut publier un arrêté déclarant l'état de catastrophe naturelle dans certaines zones touchées. Cet arrêté permet aux sinistrés de bénéficier de conditions d'indemnisation spécifiques, notamment en accélérant les procédures de dédommagement. La CCR collabore étroitement avec le gouvernement pour suivre de près l'évolution de ces arrêtés et les rendre accessibles au public. Cette transparence dans la publication des arrêtés CatNat est cruciale pour assurer une réponse rapide et appropriée aux sinistrés, contribuant ainsi à atténuer les conséquences des catastrophes naturelles sur les personnes et les biens. [2]

Chapitre 3

Élaboration d'un modèle prédictif de la sinistralité

3.1 Contexte

Le défi de prédire la reconnaissance en tant que CatNat réside non seulement dans la connaissance de la valeur du SWI à un moment donné, mais surtout dans la question de savoir si le maire fera une demande de reconnaissance CatNat. En réalité, si aucun sinistre n'est déclaré, aucune démarche n'est entreprise et la reconnaissance n'est pas accordée. Il est donc essentiel de commencer par déterminer le nombre de sinistres potentiels dans une commune. De nombreux facteurs peuvent influencer l'occurrence de sinistres dans une commune. Il convient de noter que le type d'argile présent dans la commune peut avoir une influence significative sur le nombre de sinistres, car il existe trois types d'argile qui varient en termes de susceptibilité au retrait-gonflement :

1. Argiles kaolinitiques : Ces argiles se caractérisent par leur faible plasticité et leur texture fine. Elles sont généralement présentes dans les régions où le climat est plutôt humide. Les argiles kaolinitiques sont utilisées dans diverses industries, notamment la céramique, la fabrication du papier et la construction.
2. Argiles illitiques : Les argiles illitiques ont une plus grande plasticité que les argiles kaolinitiques. Elles sont répandues dans les régions tempérées et sont souvent associées à des formations géologiques sédimentaires. Les argiles illitiques sont utilisées dans des applications telles que la fabrication de matériaux de construction et l'industrie pétrolière.
3. Argiles smectiques : Les argiles smectiques sont connues pour leur grande plasticité et leur capacité à absorber et à retenir l'eau. Elles se trouvent dans des régions caractérisées par des conditions climatiques variées. Les argiles smectiques sont utilisées dans divers domaines, notamment l'industrie du génie civil, l'agriculture et la construction de revêtements d'étanchéité.

Les autres facteurs influençant la survenance de sinistres sont principalement l'âge des bâtiments dans la commune. Les bâtiments datant d'avant les années 1980 sont souvent plus vulnérables aux sinistres en cas de sécheresse en raison de plusieurs facteurs liés à leurs méthodes de construction et aux matériaux utilisés à l'époque.

Tout d'abord, les normes de construction et de gestion de l'eau n'étaient pas aussi rigoureuses qu'aujourd'hui. Ces bâtiments plus anciens peuvent donc être moins bien équipés pour faire face aux effets néfastes de la sécheresse sur le sol. En période de sécheresse, le sol se rétracte et perd de sa cohésion, ce qui peut entraîner des mouvements du sol et des déformations des fondations. Les bâtiments plus anciens ont généralement des fondations moins résilientes, ce qui les rend plus sujets aux tassements différentiels et aux fissures.

De plus, les matériaux de construction utilisés avant les années 1980 peuvent être moins résistants aux contraintes de la sécheresse. Par exemple, certaines constructions antérieures à cette période peuvent avoir utilisé des matériaux tels que la brique non armée ou des bétons moins performants, qui sont plus sensibles aux variations de la teneur en eau du sol. Ces matériaux peuvent se dégrader plus rapidement en cas de sécheresse prolongée, augmentant ainsi le risque de dommages structurels.

En outre, les techniques de construction et les normes de qualité étaient moins avancées par rapport aux normes actuelles. Cela signifie que les bâtiments anciens peuvent avoir été construits avec des méthodes moins résilientes face aux changements climatiques, tels que la sécheresse, et peuvent donc présenter une plus grande vulnérabilité aux déformations et aux sinistres.

Les bâtiments les plus récents, c'est-à-dire ceux de moins de 10 ans, sont généralement couverts par la garantie décennale. Ainsi, en cas de sinistre causé par une sécheresse, cela est souvent considéré comme une malfaçon imputable au constructeur, et c'est cette garantie qui couvre les dégâts.

La hauteur des bâtiments est aussi à prendre en compte : en effet plus un bâtiment est élevé, plus les charges verticales et horizontales exercées sur les fondations sont importantes. Cela signifie que les fondations doivent être dimensionnées de manière adéquate pour supporter le poids du bâtiment ainsi que les forces potentiellement exercées par les séismes et les sécheresses. Un autre paramètre à prendre en considération est la norme de construction en vigueur dans la commune. En effet, l'application de normes de construction antisismiques réduit considérablement le risque de sinistres.

Avec l'instauration des normes antisismiques, l'importance des fondations dans la construction des bâtiments a été grandement renforcée. Ces normes ont été élaborées pour garantir une résistance accrue des structures face aux secousses sismiques, mais elles ont également eu un impact significatif sur leur comportement lors de périodes de sécheresse. En effet, les fondations sont devenues un élément essentiel pour maintenir la stabilité et l'intégrité des bâtiments dans des régions sujets à la sécheresse et aux risques sismiques.

Grâce à ces nouvelles normes, les concepteurs et les ingénieurs se sont tournés vers des techniques de fondation plus avancées, incluant l'utilisation de matériaux de construction hautement résistants et des méthodes de renforcement adaptées aux conditions locales. Les fondations sont désormais conçues de manière à mieux résister aux effets néfastes de la sécheresse sur le sol, tels que la rétraction et la perte de cohésion, ce qui les rend moins sensibles aux mouvements du sol pendant les séismes.

En cas de sécheresse, ces fondations robustes permettent aux bâtiments de conserver leur stabilité structurelle, même lorsque le sol devient plus instable. Les risques de fissuration ou d'effondrement des constructions sont ainsi réduits de manière significative. En assurant une meilleure

répartition des forces et en absorbant efficacement l'énergie sismique, les fondations conformes aux normes antisismiques apportent une protection cruciale aux bâtiments, contribuant ainsi à minimiser les dégâts matériels et à préserver la sécurité des occupants.

En somme, grâce aux normes antisismiques, les fondations sont devenues un maillon essentiel dans la chaîne de résistance des bâtiments face aux séismes, leur permettant de rester remarquablement stables et fonctionnels même en période de sécheresse. Ces avancées dans la conception et la construction garantissent une meilleure protection des communautés vivant dans des zones à risques multiples et renforcent la résilience des infrastructures face aux défis climatiques et sismiques.

3.2 Modèles linéaires généralisés

Théorie

Le Modèle Linéaire Généralisé (GLM) est devenu la méthode prédominante depuis plus de vingt ans, et son utilisation est aujourd'hui presque systématique dans le domaine de l'assurance non-vie. Les GLM sont une extension des modèles linéaires classiques, offrant la possibilité de modéliser des relations non linéaires entre la variable à expliquer et les variables explicatives. Le but du modèle linéaire généralisé est de modéliser la relation entre la variable réponse Y et les variables explicatives X :

$$g(E[Y|X]) = \beta^t X \quad (3.1)$$

Les GLM se caractérisent par 3 composantes principales :

La composante aléatoire La composante aléatoire se définit par la distribution de probabilité de la variable réponse Y.

La composante déterministe la composante déterministe dans un GLM représente la partie du modèle qui est prévisible et déterminée par les variables explicatives et les paramètres du modèle, sans tenir compte de la variation aléatoire. Elle permet de capturer la relation systématique entre les variables explicatives et la variable de réponse. Cela peut se formaliser par

$$g(E[Y_i|X_i]) = \beta^t X = \beta_1 X_1 + \dots + \beta_n X_n \quad (3.2)$$

La fonction de lien : la fonction de lien établit la relation entre l'espérance mathématique de la variable de réponse et les variables explicatives. Le choix de la fonction de lien dépend en grande partie de la nature du problème et des données étudiées.

Choix des variables du modèle

En premier lieu il convient de faire une analyse rapide de nos données afin de choisir quelles sont les variables appropriées ou non pour notre modèle.

Les matrices de corrélation sont des outils statistiques utilisés pour mesurer les relations linéaires entre les variables dans un ensemble de données. Elles sont particulièrement utiles pour explorer la structure des données, détecter des dépendances entre les variables et identifier les associations potentielles. Une matrice de corrélation est généralement présentée sous forme de tableau, où les valeurs sont comprises entre -1 et 1.

L'interprétation d'une matrice de corrélation se fait en analysant les coefficients de corrélation entre les différentes paires de variables. Un coefficient de corrélation de 1 indique une corrélation positive parfaite, c'est-à-dire que les variables augmentent proportionnellement l'une par rapport à l'autre. Un coefficient de corrélation de -1 indique une corrélation négative parfaite, où une variable diminue proportionnellement à l'augmentation de l'autre. Un coefficient de corrélation proche de 0 indique une absence de corrélation linéaire entre les variables.

Dans le cas de notre matrice nous allons utiliser la corrélation de Pearson :

La corrélation de Pearson, également connue sous le nom de coefficient de corrélation produit-moment de Pearson, est une mesure statistique de la relation linéaire entre deux variables continues. Elle varie entre 1 et -1 où :

- 1 indique une corrélation linéaire positive parfaite : lorsque l'une des variables augmente, l'autre fait de même de manière proportionnelle.
- 1 indique une corrélation linéaire négative parfaite : lorsque l'une des variables augmente, l'autre diminue de manière proportionnelle.
- 0 indique l'absence de corrélation linéaire : les variations de l'une des variables n'entraînent pas systématiquement des variations proportionnelles de l'autre.

Le coefficient de corrélation de Pearson r est défini par la formule :

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (3.3)$$

où :

- n est le nombre d'observations.
- $\sum xy$ est la somme des produits des paires de scores.
- $\sum x$ est la somme des scores de la première variable et
- $\sum y$ est la somme des scores de la deuxième variable.
- $\sum x^2$ est la somme des carrés des scores de la première variable et
- $\sum y^2$ est la somme des carrés des scores de la deuxième variable.

Il est à noter que la corrélation de Pearson ne détecte que les relations linéaires. Si une relation est curviligne ou de toute autre nature non linéaire, le coefficient de Pearson pourrait ne pas la capter, même si les variables sont fortement liées. Dans de tels cas, d'autres méthodes de corrélation, comme la corrélation de Spearman, pourraient être plus appropriées.

Lorsqu'il y a des variables fortement corrélées entre elles, cela peut poser des problèmes dans un modèle de régression logistique, notamment en raison du phénomène de multicollinéarité. La multicollinéarité se produit lorsque deux ou plusieurs variables sont fortement corrélées entre elles. Cela peut rendre difficile la distinction des effets individuels de ces variables sur la variable dépendante dans le modèle de régression.

Lorsque des variables sont fortement corrélées, le modèle de régression peut être instable, et les coefficients de régression peuvent être difficiles à interpréter de manière significative. En présence de multicollinéarité, les coefficients peuvent avoir des valeurs implausibles ou devenir sensibles aux petites variations des données, ce qui rend les prédictions moins fiables.

Pour remédier au problème de la multicollinéarité, il est important d'identifier les variables fortement corrélées et de prendre des mesures appropriées. Cela peut inclure l'exclusion d'une des variables corrélées du modèle, la combinaison de ces variables en une seule variable ou l'utilisation de techniques d'analyse de données avancées telles que la régression Ridge ou la régression Lasso pour atténuer l'effet de la multicollinéarité.

Ici on obtient les résultats suivants ;

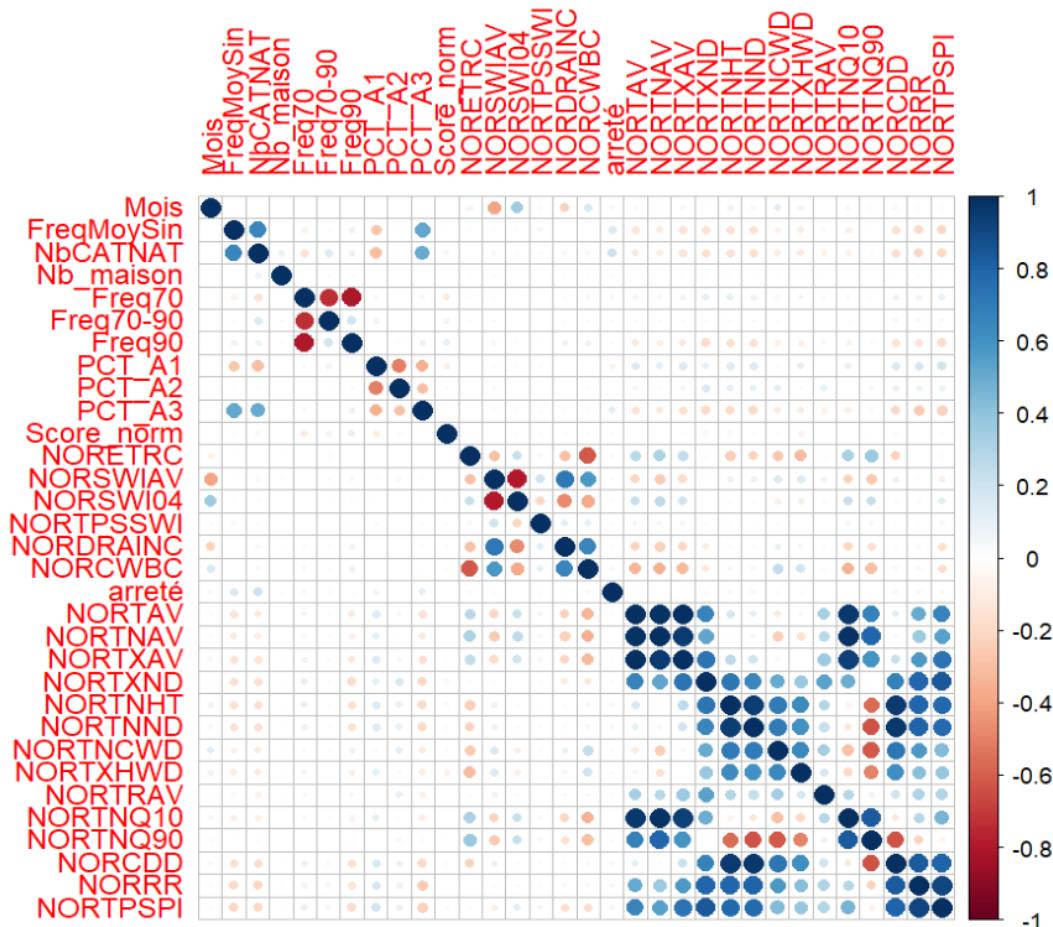


FIGURE 3.1 – Matrice de corrélation de Pearson des variables

On observe une forte corrélation entre les variables Freq70-90 et Freq90, ainsi qu'entre les variables NORSWI04 et NORTPSSWI. En conséquence, nous avons pris la décision d'exclure les variables Freq90 et NORSWI04 de notre modèle. La corrélation entre NORSWI04 et NORTPSSWI est aisément explicable, car ces deux indicateurs sont liés au SWI (Indice d'Eau de Surface), mais ils sont calculés de manière différente. En ce qui concerne les variables liées au climat, une corrélation très significative est observée entre toutes les variables relatives à la température. Dans la poursuite de notre étude, nous déciderons de conserver uniquement les variables qui mesurent les jours de chaleur extrême ainsi que ceux de froid extrême. En ce qui concerne les variables liées à la pluviométrie, nous maintiendrons exclusivement la variable qui enregistre les cumuls de pluie sur le mois.

Quant à la forte corrélation entre Freq90 et Freq70-90, elle peut être attribuée à une urbanisation similaire des villes durant ces deux périodes, entraînant un nombre approximativement équivalent de bâtiments construits. Cette corrélation reflète donc une certaine stabilité dans la croissance urbaine entre les deux périodes.

En prenant en compte ces corrélations et en excluant les variables appropriées, nous cherchons à améliorer la stabilité et la fiabilité de notre modèle de régression en éliminant les effets indésirables de la multicollinearité. Cela nous permettra de mieux comprendre les effets individuels des variables restantes sur notre variable dépendante et d'obtenir des résultats plus interprétables et pertinents.

Il nous faut également choisir une loi pour modéliser nos données pour cela on choisit la distribution de Poisson. Celle-ci est souvent considérée comme adaptée aux données de fréquence de sinistres pour plusieurs raisons, notamment le fait que ces données sont positives et définies dans un intervalle borné.

Tout d'abord, la distribution de Poisson est appropriée pour modéliser des événements rares et discrets qui se produisent dans un intervalle de temps ou d'espace donné. Dans le cas des fréquences de sinistres, elles correspondent au nombre de sinistres survenant dans une commune pendant une période de temps donnée. Comme les sinistres sont des événements relativement rares, la distribution de Poisson offre un cadre approprié pour modéliser ces occurrences discrètes.

Deuxièmement, la distribution de Poisson est définie pour des valeurs positives et est capable de modéliser des données comptées, c'est-à-dire des données qui prennent des valeurs entières non négatives. Dans le contexte des fréquences de sinistres, les valeurs observées sont généralement des nombres non négatifs représentant le nombre de sinistres dans une commune. La distribution de Poisson permet donc de représenter et de modéliser correctement ces valeurs discrètes positives.

Résultats

On obtient la sortie R avec la fonction summary du package glm suivante :

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.876e+00	4.747e-02	-60.597	< 2e-16	***
Mois	1.447e-02	1.438e-03	10.062	< 2e-16	***
Nb_maison	-3.249e-06	9.031e-07	-3.598	0.000321	***
Freq70	-9.278e-01	3.585e-02	-25.877	< 2e-16	***
Freq90	-7.098e-01	5.196e-02	-13.661	< 2e-16	***
PCT_A1	-5.059e-01	3.458e-02	-14.633	< 2e-16	***
PCT_A2	1.843e+00	2.471e-02	74.577	< 2e-16	***
PCT_A3	3.163e+00	2.384e-02	132.633	< 2e-16	***
Score_norm	-7.025e-02	1.872e-02	-3.753	0.000175	***
NORETRC	1.095e-03	1.851e-04	5.914	3.33e-09	***
NORSWIAV	5.913e-01	3.010e-02	19.647	< 2e-16	***
NORTPSSWI	1.516e+00	1.501e-01	10.100	< 2e-16	***
NORDRAINC	-2.780e-03	2.266e-04	-12.266	< 2e-16	***
NORCWBC	-1.480e-03	1.956e-04	-7.564	3.90e-14	***
NORTXND	8.900e-02	7.637e-03	11.654	< 2e-16	***
NORTNHT	-3.500e-02	1.378e-02	-2.540	0.011085	*
NORTNNND	1.284e-01	1.492e-02	8.603	< 2e-16	***
NORTNCWD	4.610e-02	4.792e-03	9.621	< 2e-16	***
NORTXHWD	-3.070e-01	1.846e-02	-16.633	< 2e-16	***
NORPAV	6.114e-03	6.059e-04	10.091	< 2e-16	***
NORRR	-7.263e-02	1.322e-02	-5.495	3.91e-08	***
NORTPSPI	-1.666e-02	7.341e-04	-22.687	< 2e-16	***

FIGURE 3.2 – Summary du modèle

La valeur de p (p-valeur) est un concept essentiel en statistiques, utilisée pour évaluer la significativité statistique d'un résultat observé dans le cadre d'une analyse. Elle est souvent utilisée en conjonction avec des tests d'hypothèses pour prendre des décisions objectives concernant les modèles statistiques et les expériences scientifiques.

Lorsqu'on effectue une analyse statistique, on formule une hypothèse nulle (H_0) selon laquelle il n'y a pas de différence ou d'effet significatif entre les groupes ou les variables étudiées. Ensuite, on recueille des données et on effectue des calculs pour obtenir une statistique de test, qui mesure l'écart entre les données observées et les valeurs attendues sous l'hypothèse nulle.

La p-valeur représente la probabilité d'obtenir une statistique de test aussi extrême (ou plus extrême) que celle observée, sous l'hypothèse nulle. Si la p-valeur est très petite (par exemple, $p < 0,05$), cela signifie qu'il est très peu probable d'obtenir une telle statistique de test en supposant que l'hypothèse nulle est vraie. Dans ce cas, on rejette l'hypothèse nulle au profit de l'hypothèse alternative (H_1) qui suggère qu'il y a un effet ou une différence significative.

En revanche, si la p-valeur est grande (par exemple, $p > 0,05$), cela signifie qu'il est assez probable d'obtenir une telle statistique de test sous l'hypothèse nulle. Dans ce cas, on ne dispose pas de suffisamment de preuves pour rejeter l'hypothèse nulle, et on ne peut pas conclure à l'existence d'un effet ou d'une différence significative.

L'utilisation appropriée de la p-valeur est essentielle pour interpréter correctement les résultats.

tats statistiques. Il est important de noter que la p-valeur ne mesure pas la taille de l'effet ou l'importance pratique, mais plutôt la probabilité d'observer les données sous l'hypothèse nulle. Par conséquent, une p-valeur faible ne signifie pas nécessairement que l'effet est important, et une p-valeur élevée ne signifie pas nécessairement que l'effet est faible.

En résumé, la p-valeur est un outil statistique fondamental qui permet d'évaluer la significativité d'un résultat. Cependant, elle doit être interprétée avec prudence et en combinaison avec d'autres informations contextuelles pour prendre des décisions éclairées dans une analyse statistique.

Ici si on analyse nos valeurs elles sont toutes inférieures à 0,001 ce qui révèle que nos variables sont toutes très significatives pour le modèle.

Nous allons également réaliser une analyse de la matrice de confusion. Une matrice de confusion est un tableau qui permet de mesurer les performances d'un modèle de classification. Elle croise les valeurs prédites par le modèle avec les valeurs réelles, permettant ainsi d'identifier les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs. Cette analyse nous permettra de vérifier si nous parvenons à obtenir des fréquences de sinistres cohérentes. Nous obtenons la matrice suivante :

Matrice de confusion									
Observé	Plus de 50%	0	0	0	0	0	1	0	0
	40% - 50%	5	8	2	0	0	1	2	22
	30% - 40%	2	0	0	0	99	18	0	3
	20% - 30%	0	142	198	981	10	87	1	61
	15% - 20%	0	2	0	20	198	7	0	0
	10% - 15%	12	411	490	25	216	0	50	0
	5% - 10%	2128	991	1614	192	8	817	0	0
	0,1% - 5%	10401	2896	1107	561	671	24	122	0
	0%	62117	8395	2056	398	100	0	0	0
	0%	0%	0,1% - 5%	5% - 10%	10% - 15%	15% - 20%	20% - 30%	30% - 40%	40% - 50%
Prédiction									

FIGURE 3.3 – Matrice de confusion pour le modèle GLM

Pour évaluer la qualité de notre modèle, nous nous penchons sur le taux de bonne prédiction. Dans le contexte de la classification, il s'agit de la proportion des prédictions correctes (vrais positifs et vrais négatifs) par rapport au nombre total de prédictions faites par un modèle. Ainsi, il reflète l'efficacité générale du modèle à classifier correctement les observations. Nous observons dans notre cas un taux de vrais positifs de 68%. Néanmoins, pour une évaluation complète, il est essentiel de comparer cette performance à celle d'autres modèles.

En complément de l'analyse de fréquence, nous entreprenons également une étude de classi-

fication binaire afin de déterminer si une commune sera reconnue comme zone de catastrophe naturelle (CatNat). Pour ce faire, nous considérons dans notre échantillon qu'une fréquence strictement supérieure à zéro correspond à la survenue d'un sinistre.

Nous obtenons une base de données où 94% des cas sont des zéros, correspondant à des non-reconnaissances en tant que zone de catastrophe naturelle (CatNat), contre seulement 6% de reconnaissances CatNat. Il est également important de souligner que cette proportion a augmenté depuis l'instauration du régime CatNat en 1982 ; en effet, chaque année, un nombre croissant de communes est reconnu comme zone CatNat.

Nous procéderons à l'analyse de la courbe ROC (Receiver Operating Characteristic) pour évaluer la performance de cette classification binaire.

La courbe ROC (Receiver Operating Characteristic), un outil puissant utilisé en statistiques et en apprentissage automatique pour évaluer la performance des modèles de classification binaire. Elle est largement utilisée dans des domaines tels que la médecine, l'ingénierie, la finance et bien d'autres. La courbe ROC présente graphiquement la relation entre le taux de vrais positifs (sensibilité) et le taux de faux positifs (1 - spécificité) en fonction du seuil de classification.

Lorsqu'on effectue une classification binaire, le modèle produit des scores ou des probabilités de classe pour chaque observation. Le seuil de classification est utilisé pour convertir ces scores en prédictions de classe (0 ou 1). En faisant varier le seuil, on obtient différents taux de vrais positifs et de faux positifs, ce qui donne lieu à une courbe ROC.

Interprétation de la courbe ROC :

- Plus la courbe ROC se rapproche du coin supérieur gauche du graphique, plus le modèle est performant, car il parvient à atteindre un taux élevé de vrais positifs tout en maintenant un faible taux de faux positifs. Une courbe ROC idéale atteindrait le coin supérieur gauche, formant un angle de 90 degrés avec les axes.
- La diagonale de la courbe ROC représente la performance d'un modèle aléatoire, où le taux de vrais positifs est égal au taux de faux positifs. Un modèle qui suit la diagonale n'a aucune capacité prédictive et est équivalent à une décision aléatoire.
- La distance entre la courbe ROC et la diagonale (aire sous la courbe ROC ou AUC) mesure la capacité discriminative du modèle. Une AUC de 0,5 indique une performance équivalente à un modèle aléatoire, tandis qu'une AUC de 1,0 indique une performance parfaite avec une séparation complète entre les classes.

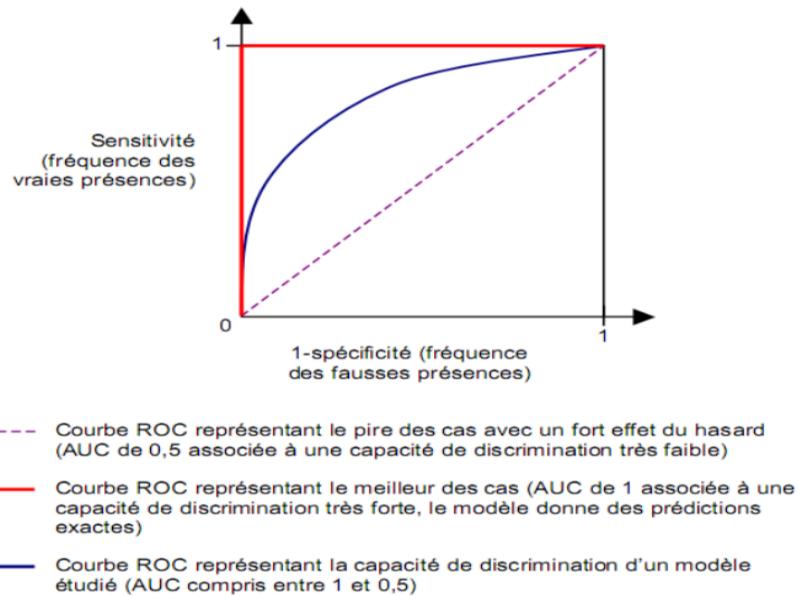


FIGURE 3.4 – Schéma explicatif de la courbe ROC

Le calcul de la courbe ROC implique également le calcul du critère AUC (Aire sous la courbe ROC) .

L'AUC est un critère essentiel pour évaluer la performance globale d'un modèle de classification binaire. Une AUC élevée (proche de 1) indique une meilleure capacité prédictive du modèle, avec une séparation claire entre les classes. Cela signifie que le modèle est capable de distinguer efficacement les observations des classes positives et négatives.

L'AUC est également utile pour comparer différents modèles. Un modèle avec une AUC plus élevée est généralement préféré, car il est plus performant pour discriminer entre les classes. Cependant, il est important de garder à l'esprit que l'AUC ne fournit pas d'informations sur les performances du modèle à un seuil spécifique, et qu'il est toujours nécessaire d'examiner d'autres métriques et d'ajuster le seuil en fonction des besoins spécifiques de l'application.

Ici nous disposons des résultats suivants : on considère l'événement où il y a eu une déclaration de catastrophe naturelle (catnat) si la fréquence de sinistre prédite est supérieure à 0,01%.

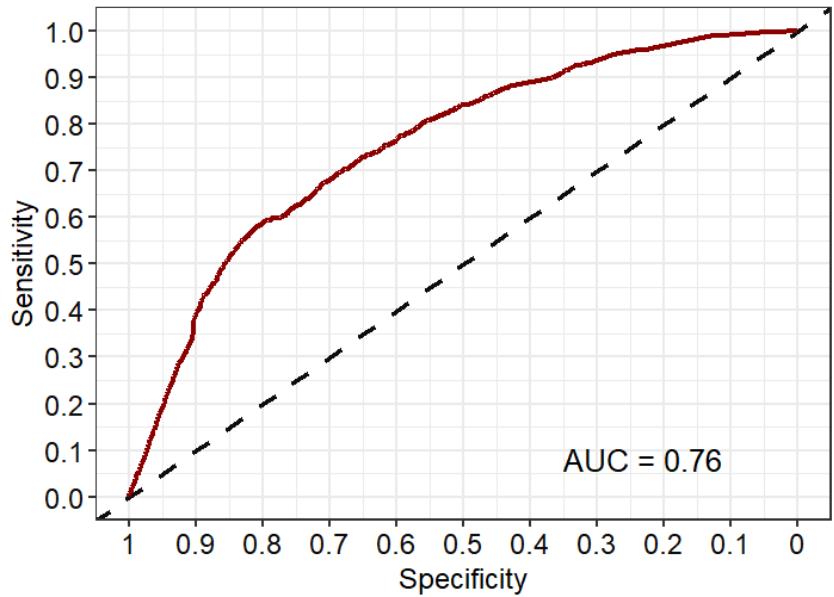


FIGURE 3.5 – Courbe ROC du modèle linéaire généralisé

On constate un critère AUC de 0.76 ce qui semble assez satisfaisant.

L'erreur quadratique moyenne (RMSE) est une métrique couramment utilisée en statistiques et en apprentissage automatique pour évaluer la performance d'un modèle de régression. Elle mesure l'écart moyen entre les valeurs prédites par le modèle et les valeurs réelles de la variable de réponse. Le RMSE est particulièrement adapté pour évaluer la précision d'un modèle de régression lorsque les valeurs de la variable de réponse sont continues.

La formule du RMSE est la suivante :

$$RMSe = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3.4)$$

où :

- y_i représente la valeur réelle de la variable de réponse pour l'observation
- \hat{y}_i représente la valeur prédite par le modèle pour l'observation
- n est le nombre total d'observations.

Interprétation du RMSE : Le RMSE mesure la différence moyenne entre les valeurs prédites et les valeurs réelles du modèle. Une valeur de RMSE faible indique que le modèle est capable de faire des prédictions très proches des vraies valeurs, ce qui est un bon signe de précision et d'ajustement aux données. En revanche, un RMSE élevé suggère que les prédictions du modèle sont éloignées des valeurs réelles, ce qui indique une faible précision du modèle.

Le RMSE a l'avantage de considérer les écarts sous forme quadratique, ce qui a pour effet de pénaliser davantage les grosses erreurs. Cela signifie que les valeurs aberrantes ou les erreurs importantes auront un impact plus significatif sur le RMSE, ce qui reflète bien leur influence sur

la performance globale du modèle.

Il est important de noter que le RMSE doit être interprété en contexte. Par exemple, si la variable de réponse a des valeurs très élevées ou très faibles, le RMSE peut également être grand en valeur absolue, mais cela ne signifie pas nécessairement que le modèle est mauvais. Il est donc souvent utile de comparer le RMSE avec la variation naturelle des données ou avec le RMSE d'autres modèles pour évaluer la qualité relative de différentes approches de modélisation.

En résumé, le RMSE est une mesure essentielle pour évaluer la précision d'un modèle de régression en comparant les valeurs prédites aux valeurs réelles. Il fournit une indication quantitative de la qualité des prédictions du modèle et aide les praticiens à prendre des décisions éclairées lors de la sélection du meilleur modèle pour leurs tâches de régression.

Dans le contexte de prédiction des fréquences de sinistres pour différentes communes, le modèle présenté a obtenu une valeur de RMSE de 0,38. Cette mesure indique que, en moyenne, les prédictions de fréquences de sinistres effectuées par le modèle diffèrent d'environ 0,38 unité par rapport aux fréquences réelles observées pour chaque commune. Étant donné que les fréquences réelles peuvent varier considérablement, avec des valeurs allant de 0,2 à 0,7, et que les zéros sont fréquents, cette RMSE de 0,38 peut être considérée comme relativement modérée et raisonnable. Cela montre que le modèle est capable de fournir des prédictions relativement proches des valeurs réelles malgré la complexité de la distribution des fréquences de sinistres.

La courbe ROC a également été utilisée pour évaluer la performance du modèle, et elle a donné un critère AUC de 0,76. Cette mesure suggère que le modèle a une capacité raisonnable à discriminer entre les communes touchées et non touchées par des sinistres, avec une probabilité d'environ 0,76. Cette capacité de discrimination est encourageante, car cela indique que le modèle est capable de bien classer les observations en fonction de leurs probabilités prédites. Cependant, malgré cette bonne capacité de discrimination, il est important de reconnaître que les zéros représentent une catégorie spéciale dans les données, et leur prédiction peut être délicate pour le modèle. Cela peut expliquer en partie la RMSE de 0,38, car les prédictions peuvent être plus éloignées des valeurs réelles pour les communes sans sinistre.

En conclusion, le modèle présente une performance globale satisfaisante, capable de bien discriminer entre les communes touchées et non touchées par des sinistres grâce à une AUC de 0,76. Cependant, la prédiction des zéros, qui sont fréquents dans les données, reste un défi pour le modèle, ce qui peut expliquer la valeur de RMSE de 0,38. Pour améliorer davantage la performance du modèle, il pourrait être intéressant de se concentrer sur des méthodes spécifiques pour mieux prédire les zéros, tout en continuant à considérer d'autres métriques d'évaluation et le contexte spécifique de l'application. L'objectif est de développer un modèle qui parvient à fournir des prédictions précises pour toutes les communes, qu'elles aient des fréquences de sinistres élevées ou nulles, afin d'obtenir une modélisation plus robuste et appropriée pour le problème de prédiction des fréquences de sinistres.

3.3 Random Forest

Théorie

Le Random Forest est une technique d'apprentissage automatique relativement récente, introduite pour la première fois par Leo Breiman en 2001. Cette technique est issue de l'agrégation d'arbres de décision, une méthode d'apprentissage populaire depuis les années 1980.

Le principe du Random Forest est de construire un grand nombre d'arbres de décision, chacun étant entraîné sur un sous-échantillon aléatoire de données et un sous-ensemble aléatoire de variables. En combinant les prédictions de ces arbres, le modèle final est plus stable et performant que les arbres de décision individuels. Cette technique a rapidement gagné en popularité en raison de sa capacité à traiter des données complexes et de grande dimensionnalité, tout en étant relativement simple à mettre en œuvre.

En premier lieu il est important de bien comprendre ce qu'est un arbre de décision.

Un arbre de décision est un modèle d'apprentissage automatique utilisé pour la classification et la prise de décision. Il consiste en une série de questions logiques basées sur les caractéristiques des données. Les noeuds représentent les questions, les branches relient les réponses possibles, et les feuilles représentent les décisions finales. Les arbres de décision sont faciles à interpréter et à utiliser, mais peuvent être sensibles aux variations de données et au surapprentissage.

L'aspect le plus délicat de l'algorithme Random Forest réside dans l'importance attribuée au choix du nombre d'arbres et à leur caractère aléatoire. De plus, étant donné que le résultat est une moyenne de plusieurs arbres, l'aspect visuel intuitif des arbres de décision individuels est perdu.

Cette méthode se formalise de la manière suivante :

1. On commence par créer B échantillons bootstrap à partir d'un échantillon d'apprentissage donné. Chaque échantillon est construit en effectuant un tirage aléatoire avec remise de n observations. On note l'échantillon d'apprentissage $Z = ((x_1, y_1), \dots, (x_n, y_n))$, où X est un vecteur de taille p des variables explicatives et Y la variable réponse. On tire B échantillons Z qu'on numérote Z^1, \dots, Z^B
2. Pour chaque échantillon B tiré : On choisit un sous ensemble de taille m des variables explicatives. Avec $m < \sqrt{p}$. On entraîne ensuite un arbre de classification sur ce jeu de données et on obtient une fonction $f(x)$ avec $x = (x^1, \dots, x^m)$
3. On stocke pour chaque échantillon les prédictions $\hat{f}_1(x), \dots, \hat{f}_B(x)$ obtenues
4. On dispose à présent de B arbres de décision qui nous donne B prédictions. Afin d'obtenir une prédiction globale en calculant l'estimation moyenne obtenue :

$$\hat{f}_B = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x)$$

Ces étapes peuvent se résumer avec le schéma suivant :

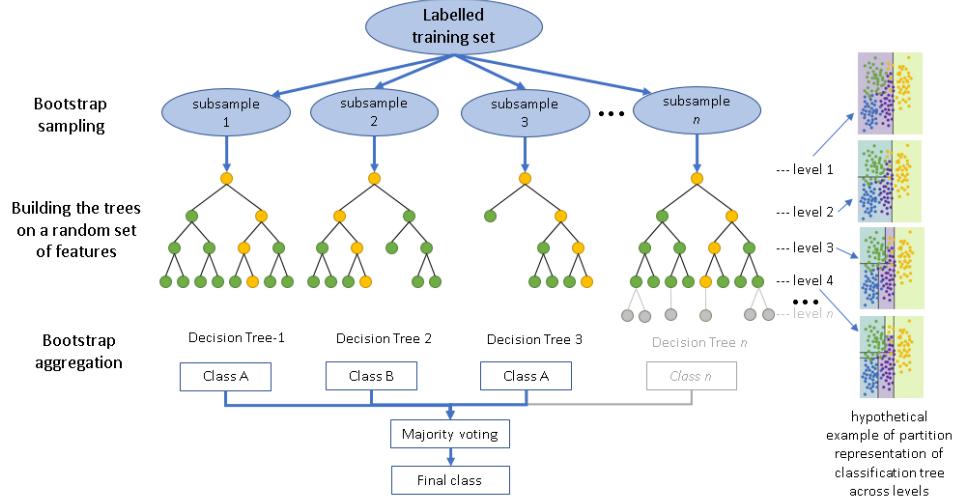


FIGURE 3.6 – Schéma illustrant l'algorithme du Random Forest

Choix des paramètres du modèle

Le paramètre le plus délicat à estimer dans l'algorithme Random Forest est le nombre d'arbres à choisir. Pour résoudre cette problématique, nous allons mettre en œuvre l'algorithme et comparer le taux d'erreur pour différentes valeurs du nombre d'arbres. Cela nous permettra de sélectionner le nombre d'arbres le plus approprié pour notre modèle.

L'idée ici est de tester plusieurs valeurs du nombre d'arbres (par exemple, de 1 à 100) en utilisant la validation croisée pour évaluer les performances du modèle. Nous calculerons le taux d'erreur pour chaque configuration et choisisrons celui qui donne le taux d'erreur le plus bas. En optant pour le nombre d'arbres correspondant au taux d'erreur minimal, nous cherchons à obtenir un modèle robuste et performant qui généralise bien sur de nouvelles données.

En somme, cette approche nous permettra de déterminer le nombre optimal d'arbres pour notre modèle Random Forest, ce qui est essentiel pour obtenir des résultats fiables et efficaces dans diverses applications de l'apprentissage automatique et de la statistique.

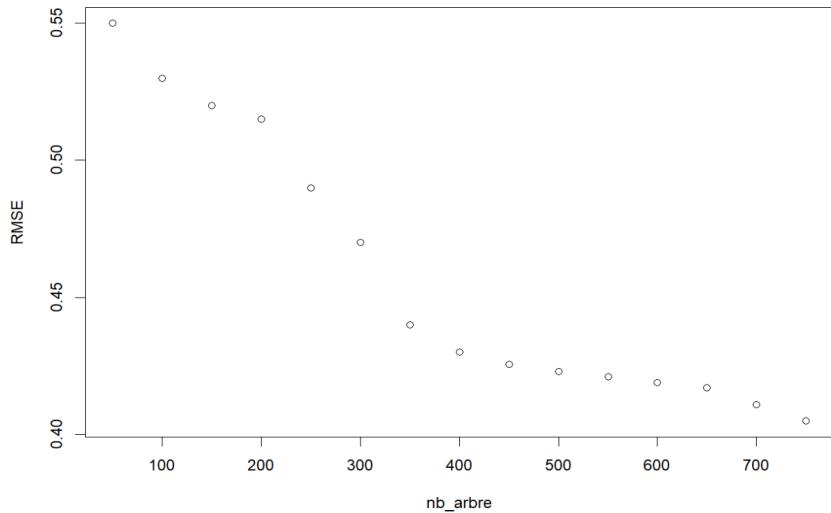


FIGURE 3.7 – Choix du nombre d’arbres de l’algorithme du Random Forest

On observe que la RMSE atteint un plateau à environ 600 arbres, ce qui signifie que l’erreur de prédiction ne diminue plus significativement au-delà de ce nombre d’arbres. Par conséquent, nous retiendrons 600 arbres pour la suite de l’étude.

Résultats

Observons la matrice de confusion de ce modèle :

Matrice de confusion									
Observé	Plus de 50%	40% - 50%	30% - 40%	20% - 30%	15% - 20%	10% - 15%	5% - 10%	0,1% - 5%	0%
	0	0	0	0	1	0	0	0	0
	1	1	2	4	0	0	2	17	1
	0	1	0	0	0	18	0	0	0
	0	121	143	981	10	64	1	0	1
	4	4	0	20	198	2	0	0	0
	5	312	453	25	216	3	50	0	0
	4313	1871	1624	492	8	217	0	0	3
	10964	2352	1331	478	678	32	7	0	0
	65241	4220	2898	5	90	4	0	0	0

FIGURE 3.8 – Matrice de confusion pour le modèle Random forest

On constate des résultat légèrement meilleur que dans le modèle précédent en effet on obtient un taux de bonne prédiction de 69%

Regardons la courbe ROC du nombre de bonnes prédictions d’un événement de catastrophe naturelle :

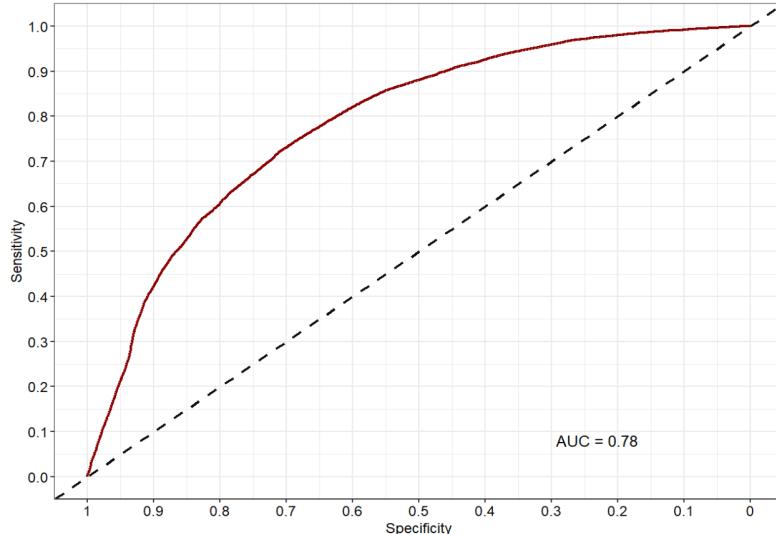


FIGURE 3.9 – Schéma illustrant l’algorithme du Random Forest

Le modèle démontre une capacité satisfaisante à discriminer entre les classes positives et négatives, comme indiqué par son score AUC solide de 0,78. Cependant, en raison de la prévalence significative de zéros dans les données, il est important de noter que le modèle peut être influencé par cette répartition.

Bien que le coefficient de détermination R^2 soit modéré à 0,43, il convient de rappeler que les modèles prédictifs pour les données comprenant de nombreux zéros peuvent avoir des ajustements plus complexes. Les prédictions peuvent être plus influencées par la majorité de zéros, ce qui peut se refléter dans l’ajustement et les performances globales du modèle.

3.4 Gradient Boosting

Théorie

Le Gradient Boosting (GBM) est un algorithme d’apprentissage automatique qui permet de construire des modèles de prédiction à partir de données. Il s’agit d’une méthode de boosting, qui consiste à construire des modèles simples et les combiner pour obtenir un modèle plus complexe et plus précis.

L’algorithme GBM commence par construire un premier modèle de prédiction, appelé modèle de base. Ce modèle est souvent un arbre de décision, qui permet de diviser les données en sous-groupes en fonction de certaines caractéristiques. Ensuite, GBM ajuste le modèle pour qu’il prédise plus précisément les données d’entraînement. Pour cela, il cherche les erreurs de prédiction du premier modèle et ajuste le modèle de base en ajoutant un deuxième modèle de prédiction pour corriger les erreurs du premier. Ce processus est répété plusieurs fois en utilisant de nouveaux modèles pour corriger les erreurs des modèles précédents jusqu’à ce que la précision soit satisfaisante.

Le GBM est capable de prendre en compte de nombreuses caractéristiques de données et de traiter des données de différentes natures (numériques, catégoriques, etc.). Il est également capable de traiter de grands ensembles de données avec une grande précision de prédiction. Cependant, l'algorithme peut être sensible aux valeurs aberrantes et nécessite souvent un temps de traitement plus long que d'autres méthodes de machine learning.

L'algorithme du Gradient Boosting (GBM) est un cas particulier de Boosting qui est basé sur la méthode d'optimisation de la descente du gradient.

Le Boosting est une méthode d'agrégation de modèles pour obtenir un résultat final unique. L'agrégation se fait de manière séquentielle, en commençant par un modèle initial qui évolue ensuite. Le premier modèle est basé sur un échantillon d'apprentissage de taille n , avec une pondération uniforme de chaque observation. Par la suite, chaque individu est re-pondéré en fonction de sa performance en prédiction : un poids plus élevé indique une prédiction moins précise, le but étant d'améliorer la prédiction dans le modèle suivant. Cette procédure est répétée selon le désir de l'utilisateur, et la prédiction finale est la somme pondérée des prédictions de chaque modèle.

Les étapes du Boosting sont :

Initialisation : Pondération uniforme des n observations de l'échantillon d'apprentissage avec des poids w_{1i} , où :

$w_{11} = w_{12} = \dots = \dots = w_{1n} = \frac{1}{n}$ Récurrence : L'utilisateur fixe un nombre B d'itérations. Pour $b \in \{1, \dots, B\}$

- Construction de l'arbre A_b à partir de l'échantillon.
- Calcul du taux d'erreur pondéré $\epsilon_b = \sum_{i=1}^n w_{bi} * \mathbb{1}_{y_i \neq y_{pred}^i}$
- Si $\epsilon_b < \epsilon$, où ϵ est un seuil d'erreur maximal prédéfini, l'algorithme s'arrête. Sinon :
 1. Calcul du poids α_b pour A_b
 2. Mise à jour des poids pour la prochaine itération : $w_{b+1,i} = f(w_{bi})$, où f est une fonction définie par l'utilisateur, puis normalisation des poids.

Le GBM est une forme spécifique de Boosting utilisant le gradient de la fonction de perte pour le calcul des poids lors de la construction de chaque nouvel arbre. Cette méthode itérative optimise une fonction, généralement par minimisation. Dans le GBM, le but est de minimiser la fonction de perte. Le processus de la descente du gradient est :

1. Définition de la fonction de perte $l(y_i, y_i^{pred0})$ et calcul de sa dérivée $\frac{\partial l(y_i, y_i^{pred0})}{\partial y_i^{pred0}}$
2. pour $b \in 1, \dots, B$:

Si $|\frac{\partial l(y_i, y_i^{predb})}{\partial y_i^{predb}}| < \epsilon$, où ϵ est fixé par l'utilisateur, l'algorithme s'arrête, le point minimal est trouvé.

Sinon, on fait le choix d'un pas η_b et mise à jour : $y_{predb+1}^1 = y_{predb}^1 - \eta_b \times \frac{\partial l(y_i, y_i^{predb})}{\partial y_i^{predb}}$, et recalcul du gradient pour $l(y_i, y_i^{pred+1})$

Le but du Gradient Boosting Machine est donc de trouver une combinaison linéaire d'arbres optimaux. Il va construire une nouvelle base corrélée avec le gradient de la fonction de perte.[\[17\]](#) Ce qui peut se traduire par le schéma suivant :

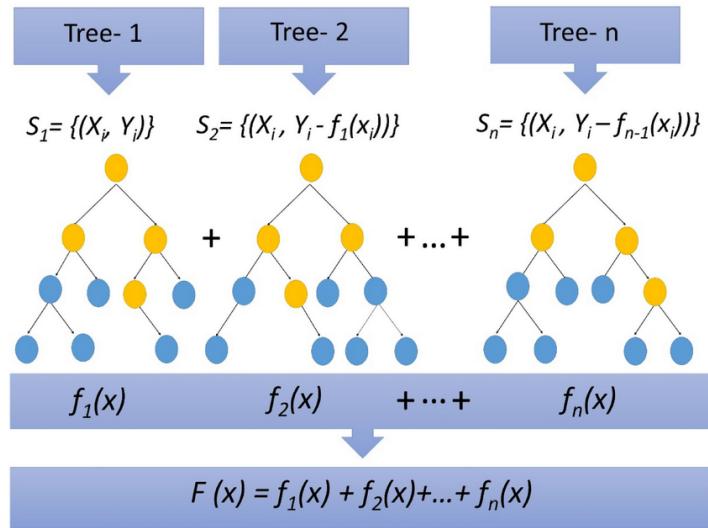


FIGURE 3.10 – Schéma illustrant l'algorithme du Gradient Boosting

Choix des paramètres du modèle

De même que pour l'algorithme du Random Forest, il est important de faire un choix sur le nombre d'arbres à considérer dans le modèle GBM (Gradient Boosting Machine). Pour ce faire, on peut représenter la RMSE (Root Mean Squared Error) en fonction du nombre d'arbres de décision considérés.

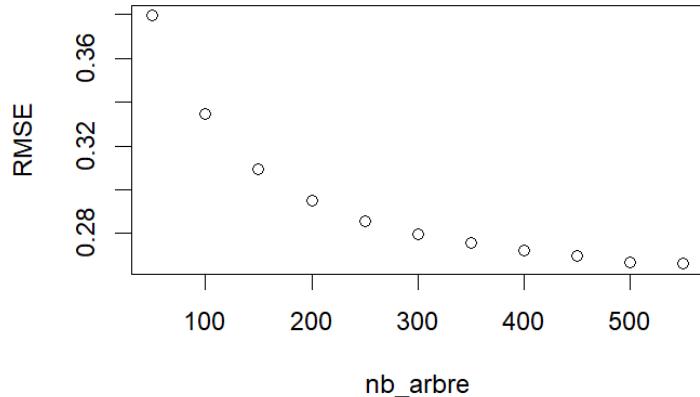


FIGURE 3.11 – Choix du nombre d'arbres pour l'algorithme du GBM

On observe que la RMSE atteint un plateau à environ 400 arbres, ce qui signifie que l'erreur de prédiction ne diminue plus significativement au-delà de ce nombre d'arbres. Par conséquent,

nous retiendrons 400 arbres pour la suite de l'étude.

En utilisant 400 arbres, nous pouvons être confiants que le modèle GBM a atteint un niveau de stabilité satisfaisant en termes de performance de prédiction. Cela permet également de réduire le risque de surajustement (overfitting), car un grand nombre d'arbres peut conduire à un modèle complexe qui s'adapte trop précisément aux données d'entraînement.

Nous pouvons maintenant utiliser ce modèle GBM entraîné avec 400 arbres pour faire des prédictions sur de nouvelles données et obtenir des résultats fiables pour notre étude.

Résultats

Observons la matrice de confusion de notre modèle

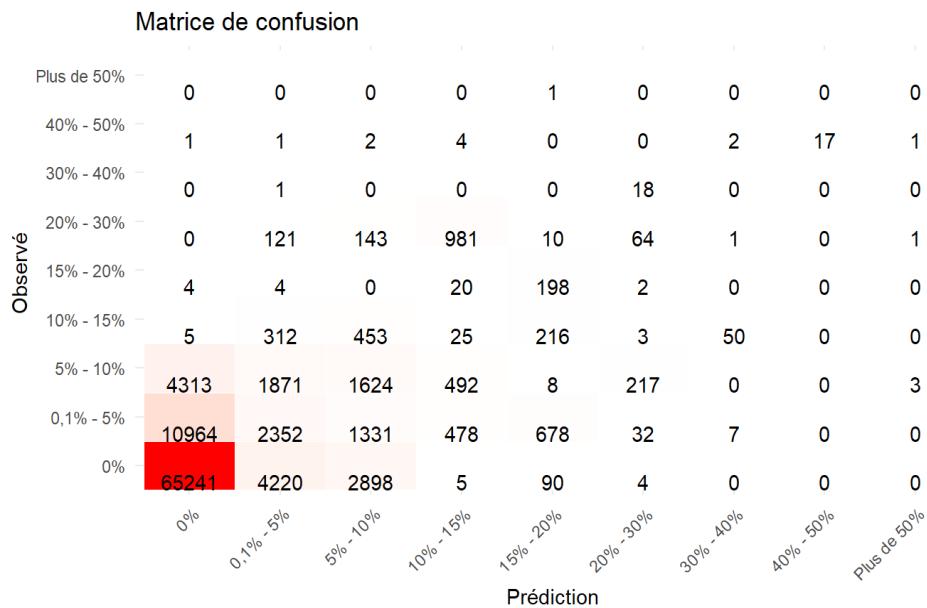


FIGURE 3.12 – Matrice de confusion pour le modèle Gradient boosting

Nous constatons un taux de bonne prédiction de 72%. De plus, le nombre de prédictions très mal classées, c'est-à-dire celles qui s'écartent significativement de la diagonale, a nettement diminué. Par conséquent, les prédictions des fréquences se rapprochent plus fidèlement de la réalité, grâce à une diminution notable des prédictions aberrantes. Cette amélioration contribue à la minimisation de l'erreur globale.

Regardons la Courbe ROC de notre modèle.

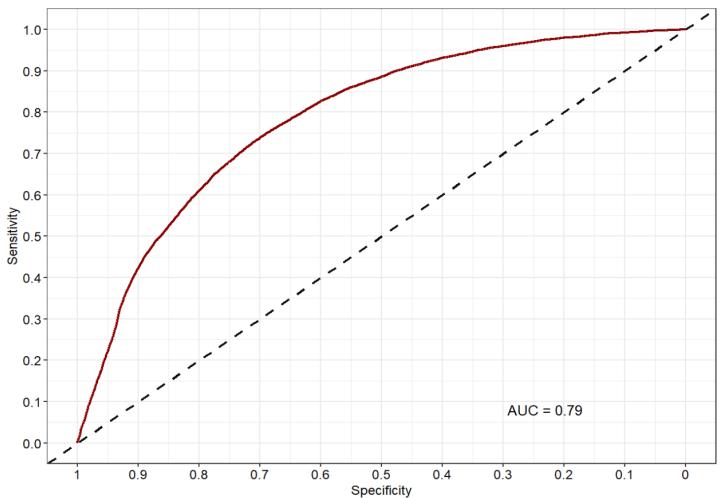


FIGURE 3.13 – Courbe ROC du modèle GBM retenu

Suite à l'analyse du modèle de prédiction des fréquences de sinistres, nous avons obtenu des résultats prometteurs. La Root Mean Squared Error (RMSE) du modèle est de 0,2722349, ce qui indique que, en moyenne, les prédictions du modèle diffèrent d'environ 0,27 unité par rapport aux fréquences réelles des sinistres pour chaque commune. Cette RMSE relativement faible suggère que le modèle parvient à bien approcher les fréquences observées.

De plus, le coefficient de détermination (R-squared, R^2) du modèle est de 0,6699698, ce qui signifie que le modèle explique environ 67% de la variation totale des fréquences de sinistres dans les données. Un R^2 élevé indique une bonne adéquation du modèle aux données, ce qui est encourageant pour la qualité des prédictions.

Enfin, le critère Area Under the Curve (AUC) de 0,79 issu de la courbe ROC est excellent. L'AUC mesure la capacité du modèle à bien discriminer entre les communes touchées et non touchées par des sinistres.

Ces résultats nous permettent de conclure que le modèle de prédiction des fréquences de sinistres se révèle solide et précis dans sa capacité à estimer les probabilités de sinistres pour différentes communes.

3.5 Comparaison des modèles

Pour déterminer le modèle le plus adapté à nos projections, nous évaluerons les performances de divers modèles. En commençant par analyser les matrices de confusion, nous avons constaté que la meilleure prédiction des fréquences de sinistres attendues est obtenue avec le modèle de gradient boosting. En effet, ce modèle se distingue non seulement par le meilleur taux de bonne prédiction, mais aussi par sa capacité à minimiser les erreurs aberrantes.

Lors de la comparaison des courbes ROC, nous obtenons les résultats suivants :

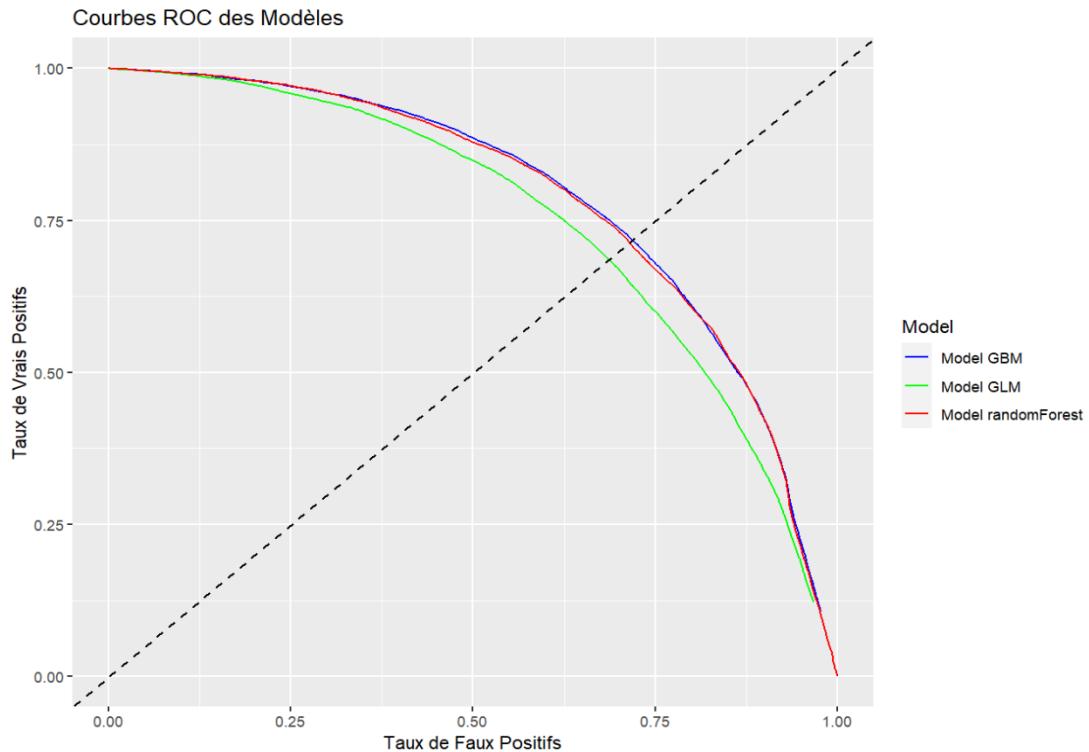


FIGURE 3.14 – Comparaison des courbes ROC des modèles

La courbe ROC du modèle GBM se rapproche davantage d'une AUC égale à 1, signifiant qu'il présente une performance supérieure en matière de distinction entre les classes positives et négatives. Cependant, la courbe ROC du modèle Random Forest demeure assez similaire à celle du GBM. Cette proximité suggère que, malgré la légère supériorité apparente du GBM. La courbe ROC du modèle linéaire est nettement moins performante, comparativement à celle du GBM, ce qui indique une capacité moindre de ce modèle à distinguer les classes.

Indicateur	Modèle GLM	Modèle Random Forest	Modèle GBM
RMSE	0,380737	0.340737	0,27223
R ²	0.4663157	0.452672	0,66997

FIGURE 3.15 – Récapitulatif des indicateurs de qualité des modèles

Les modèles de Random Forest et GBM, bien que basés sur des techniques d'apprentissage automatique avancées, présentent des différences notables en termes de performance. En analysant les métriques RMSE et R², on observe que le GBM semble surpasser le modèle Random Forest. Plus précisément, le GBM bénéficie d'une RMSE plus faible, ce qui indique une moindre dispersion des résidus autour de la ligne de meilleure adaptation. De plus, son R² est significativement plus proche de 1, suggérant une meilleure capacité d'explication de la variance par le modèle.

Cela dit, il est essentiel de contextualiser ces performances dans le cadre du problème spécifique à résoudre. Le modèle GBM offre des performances respectables. Toutefois, ces résultats doivent être pris avec prudence. En effet, la présence d'un grand nombre de zéros dans les données complique le processus de prédition. Cette situation conduit souvent le modèle GBM à prévoir des fréquences de sinistres supérieures à la réalité. En effet, prévoir un nombre plus élevé de sinistres permet aux assureurs de provisionner davantage, offrant ainsi une meilleure protection contre d'éventuels aléas. En somme, bien que le GBM se distingue par ses performances, il est crucial de comprendre les nuances et les spécificités des prédictions fournies pour une prise de décision éclairée.

Chapitre 4

Analyse de sensibilité du modèle

L'analyse de sensibilité globale est une approche utilisée pour évaluer l'importance des différentes variables d'entrée dans un modèle numérique et comprendre comment elles influencent ses résultats. Elle vise à quantifier l'effet des variations des variables d'entrée sur les sorties du modèle. L'analyse de sensibilité globale est réalisée en explorant systématiquement l'espace des variables d'entrée et en mesurant la variation de la sortie du modèle en réponse à ces variations.

Le fonctionnement global de l'analyse de sensibilité consiste généralement à appliquer une méthode statistique ou numérique pour échantillonner l'espace des variables d'entrée. Cela peut être fait en utilisant des techniques telles que les plans d'expérience, les méthodes de Monte Carlo ou les méthodes de variance réduite. Pour chaque combinaison d'échantillons, le modèle est exécuté et les sorties sont enregistrées. Les résultats sont ensuite analysés pour évaluer l'importance des variables d'entrées et les interactions éventuelles entre elles.

L'intérêt de l'analyse de sensibilité globale est multiple. Tout d'abord, elle permet d'identifier les variables d'entrées qui ont le plus d'influence sur les résultats du modèle, ce qui permet de concentrer les efforts de collecte de données ou d'améliorer la précision de ces variables. De plus, elle aide à comprendre la complexité du modèle en mettant en évidence les interactions entre les variables d'entrées. Cela permet de détecter les relations non linéaires ou les seuils critiques dans le modèle. Enfin, l'analyse de sensibilité globale fournit des informations précieuses pour prendre des décisions éclairées et évaluer les incertitudes associées aux résultats du modèle.

Dans le cadre de ce mémoire, il est essentiel d'effectuer une analyse de sensibilité de nos paramètres afin de déterminer quelles variables ont le plus d'influence sur notre sinistralité. Cette étude nous permettra ensuite de construire différents portefeuilles présentant des caractéristiques variées, afin de projeter notre sinistralité de manière plus précise et adaptée. En comprenant les variables clés qui influent sur nos résultats, nous pourrons prendre des décisions éclairées et améliorer la gestion des risques liés au portefeuille d'exposition.[\[16\]](#)

4.1 La méthode de Morris

Théorie : Le plan OAT

Pour mesurer l'importance d'une variable il semble intuitif de vouloir évaluer le modèle pour toutes les combinaisons de niveau d'entrée ce qui pourrait permettre l'estimation de tous les effets d'entrée et de leur interaction. Cependant, cela serait extrêmement coûteux en termes de

temps de calcul et donc pratiquement irréalisable.

Le plan One-At-a-Time (OAT) est une technique d'analyse de sensibilité qui évalue l'influence individuelle des variables d'entrée sur un modèle. Dans cette méthode, une seule variable est modifiée à la fois, tandis que les autres restent constantes. Mathématiquement, pour une variable X_i , cette influence est estimée par le gradient partiel $\frac{\partial f}{\partial X_i}$, approximé comme $\frac{f(X_1, \dots, X_i + \Delta X_i, \dots, X_n) - f(X_1, \dots, X_i, \dots, X_n)}{\Delta X_i}$. Où ΔX_i est une petite variation de X_i . Après avoir défini des valeurs de référence pour chaque variable, on modifie la valeur de X_i sur une plage prédéfinie et observe les variations de la sortie du modèle.

Le plan OAT présente des limites, notamment son incapacité à capturer les interactions entre variables, car il ne prend pas en compte les termes d'interaction tels que $\frac{\partial^2 f}{\partial X_i \partial X_j}$ pour $i \neq j$. Si les variables sont fortement corrélées, les variations individuelles ne reflètent pas pleinement l'effet global des variables sur la sortie du modèle.

Malgré ces limitations, le plan OAT est utile pour une première exploration de sensibilité et pour identifier les variables influentes. Pour une analyse plus approfondie, on peut utiliser la méthode de Morris.

La méthode de Morris, une approche d'analyse de sensibilité, évalue la sensibilité des paramètres un par un en utilisant des trajectoires pour calculer les effets élémentaires (EE). Les EE pour un paramètre X_i sont définis comme la variation moyenne de la sortie du modèle en raison d'un petit changement dans X_i , calculés sur toutes les trajectoires comme $\frac{1}{r} \sum_{j=1}^r \frac{f(X_{ij} + \Delta e_i) - f(X_{ij})}{\Delta}$.

Ici, r représente le nombre de trajectoires, et Δe_i est un vecteur unitaire dans la direction de X_i . Cette méthode combine la rapidité des approches OAT avec une exploration plus globale, mais elle dépend de la discréétisation des paramètres et du nombre de trajectoires.

Le nombre de simulations pour la méthode de Morris est déterminé par $N=r(K+1)$, où K est le nombre de paramètres et r le nombre de trajectoires. Cette méthode permet de classer les variables d'entrée en catégories basées sur les EE : variables à effets négligeables, à effets linéaires et sans interaction, et celles ayant des effets non-linéaires avec potentiellement des interactions.

En combinant ces approches, on obtient une analyse complète de la sensibilité des différents paramètres dans un modèle, permettant une compréhension approfondie de leur influence individuelle et collective.

Cette nouvelle méthode permet de passer outre les limites du modèle OAT et permet de classer les variables d'entrée en 3 catégories :

- Les entrées dont les effets sont négligeables
- Les entrées dont les effets sont linéaires et sans interaction
- Les entrées ayant des effets non-linéaires et qui peuvent en outre présenter des interactions avec d'autres entrées.

[12]

Mesure de sensibilité

Dans le cadre de la gestion de notre portefeuille assurantiel, nous mettons en œuvre des mesures d'importance afin d'identifier les données clés à gérer. Ces analyses visent principalement à comprendre la pertinence de nos variables, notamment celles relatives aux communes. Des éléments tels que les trois types d'argile, la hauteur des bâtiments et la taille des édifices sont au cœur de notre évaluation. La démarche permettra à l'assureur de mieux appréhender les risques potentiels et d'optimiser la sélection des risques à assurer, garantissant ainsi une gestion plus affinée et pertinente de son portefeuille.

Dans la gestion de notre portefeuille, nous nous sommes focalisés sur les variables que nous pouvons contrôler. C'est pourquoi nous avons choisi des données caractéristiques de la commune, en particulier :

- Le pourcentage d'argile : influençant la stabilité des fondations.
- La hauteur des bâtiments : reflétant la densité urbaine.
- L'âge des bâtiments : indiquant le patrimoine architectural.

La variable "Freq70" fait référence au pourcentage de bâtiments construits avant les années 1970. Elle prend des valeurs allant de 0 à 1, illustrant ainsi le taux de bâtiments anciens dans une région donnée. Dans notre analyse, pour mieux comprendre l'impact de cette variable tout en maintenant les autres variables constantes, nous avons choisi de la faire varier par incrément de 0,1.

Un "Partial Dependence Plot" (PDP) est un outil puissant pour visualiser l'effet d'une variable sur la prédiction d'un modèle, tout en contrôlant ou en "neutralisant" les effets des autres variables. En d'autres termes, il montre comment la variable d'intérêt (dans ce cas "Freq70") affecte la prédiction moyenne du modèle lorsque toutes les autres variables sont maintenues à leur moyenne (ou à une autre valeur fixe). C'est particulièrement utile pour comprendre les relations non linéaires entre une variable et la sortie prédite.

Dans le contexte de notre étude, en utilisant un PDP pour "Freq70", nous cherchons à visualiser comment le pourcentage de bâtiments construits avant 1970 influence la prédiction de notre modèle. Cela peut nous donner une idée de l'importance de la conservation architecturale ou de la rénovation par rapport aux prédictions du modèle.

Passons maintenant à la représentation graphique de ce PDP pour la variable "Freq70" :

PDP pour Freq70

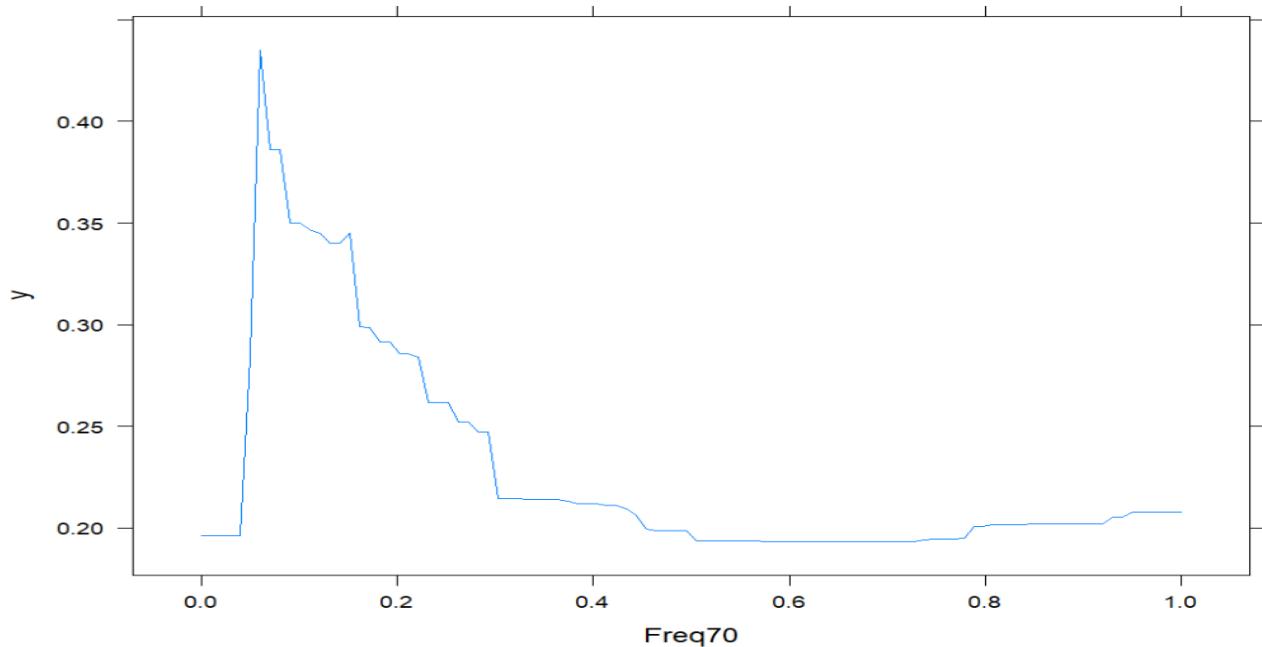


FIGURE 4.1 – PDP pour la variable Freq70

Lors de l'analyse de notre modèle, nous avons observé un pic significatif de la prédiction de sinistralité lorsque la variable "Freq70" est aux alentours de 10%. Autrement dit, il semble que lorsqu'environ 10% des bâtiments ont été construits avant les années 1970 (représenté par la valeur de 0,1 pour la variable Freq70), le risque de sinistralité atteint son apogée. Cependant, au-delà de ce seuil, nous constatons une diminution progressive de cette prédiction de sinistralité. Cette observation suggère que c'est précisément autour de cette valeur de 0,1 que la variable "Freq70" a un impact maximal sur notre modèle, influençant ainsi fortement nos prédictions de risques.

Les trois indicateurs clés pour cette analyse sont :

- μ^* : Il représente la moyenne des effets absolus. Plus sa valeur est élevée, plus la variable a tendance à avoir un impact significatif sur la sortie du modèle. C'est une mesure de l'importance globale d'une variable.
- μ : La moyenne des effets, sans prendre la valeur absolue. Elle donne une idée de la direction de l'effet (positif ou négatif).
- σ : L'écart-type des effets. Une valeur élevée indique que l'effet de la variable dépend des niveaux des autres variables (interaction ou non-linéarité).

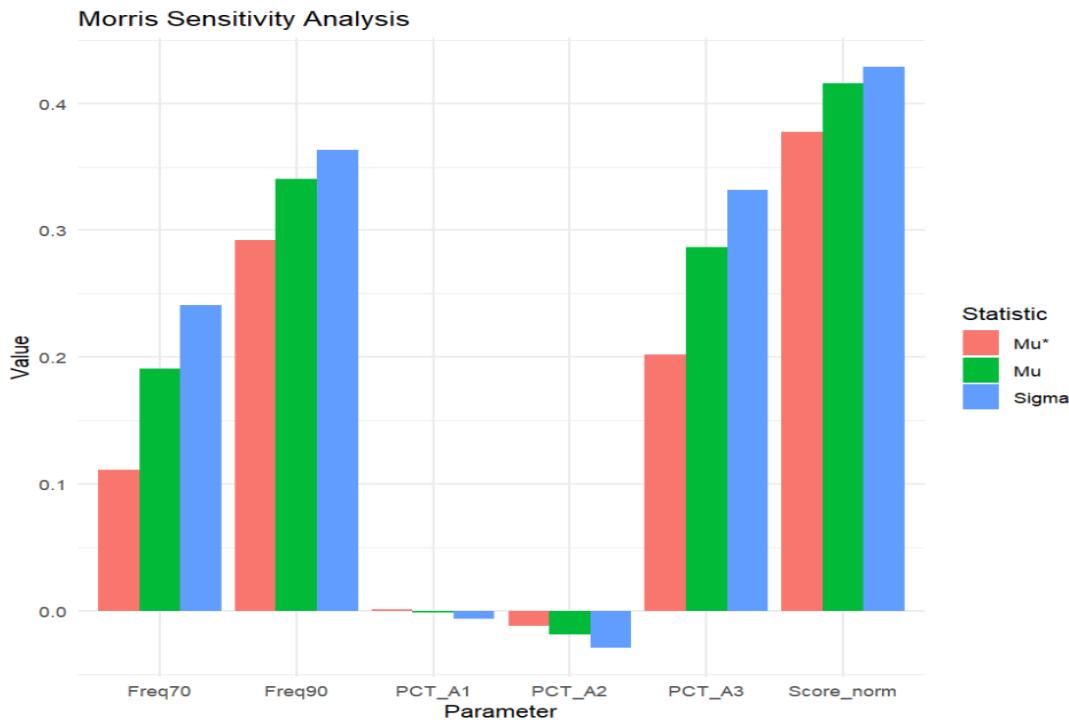


FIGURE 4.2 – Résultats de l’analyse de Morris

Concernant l’indicateur μ , il peut s’interpréter seul. Les variables avec des valeurs positives de μ ont tendance à augmenter la sortie du modèle lorsqu’elles augmentent, tout en gardant les autres variables fixes. A l’inverse les variables avec des valeurs négatives de μ ont tendance à diminuer la sortie du modèle lorsqu’elles augmentent.

De cette analyse, nous pouvons déduire que les valeurs de μ pour les variables Freq70, PCTA3 et Score_norm sont élevées. Cela suggère que ces variables ont une forte influence sur la fréquence des sinistres. En revanche, les variables PCTA2 et PCTA1 semblent avoir un impact réducteur sur la fréquence des sinistres. Cette observation peut être interprétée comme suit : plus la présence d’argile de types 1 et 2 est élevée, moins il y a d’argile de type 3. Or, l’argile de type 3 est nettement plus sensible au phénomène de retrait-gonflement, ce qui conduit à une réduction des sinistres.

Individuellement on peut interpréter les indicateurs μ^* et σ comme suit :

μ^* Élevé : Cela indique que la variable a un impact significatif sur la sortie du modèle en moyenne. Une valeur élevée de μ^* suggère que des variations dans cette variable entraînent des variations substantielles dans la sortie, indiquant son importance.

σ Élevé : Un sigma élevé indique que l’effet de cette variable sur la sortie du modèle n’est pas linéaire et/ou présente des interactions avec d’autres variables. En d’autres termes, l’effet de cette variable sur la sortie n’est pas constant mais dépend du niveau de la variable elle-même ou du niveau d’autres variables.

On peut également les interpréter ensemble ce qui les rend plus intéressants :

Lorsque μ^* et σ sont tous les deux élevés pour une variable donnée, cela suggère que la variable est non seulement importante pour le modèle, mais aussi que son effet est complexe. Elle peut avoir un comportement non linéaire, des seuils ou des interactions avec d'autres variables.

Les variables avec des valeurs élevées pour μ^* et faibles pour σ sont celles qui ont un effet important et constant, indépendamment des autres variables.

Si σ est plus grand que μ mais pas de manière significative, cela suggère que la variable concernée a un effet non linéaire ou qu'elle interagit avec d'autres variables dans le modèle.

On va se pencher sur l'analyse de la relation entre ces variables.

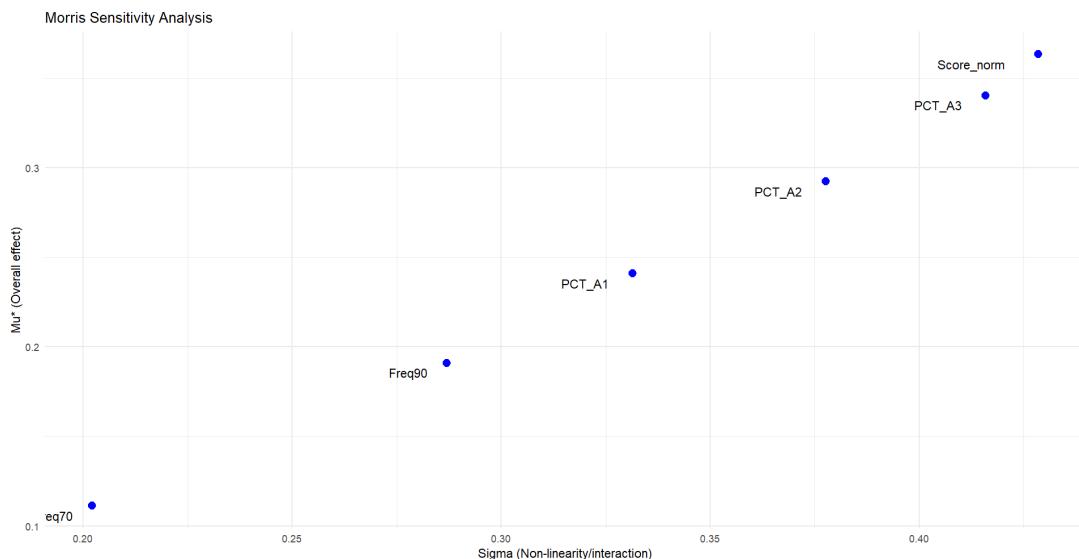


FIGURE 4.3 – Comparaison de μ^* par rapport à sigma

L'analyse révèle que les variables Freq90, Scorenorm, et PctA3 sont les plus influentes pour notre modèle. Toutefois, ces variables présentent des valeurs de σ élevées, suggérant des interactions complexes avec d'autres variables du modèle. Une valeur de σ élevée peut indiquer des effets non linéaires ou des interactions avec d'autres variables, rendant l'interprétation de ces facteurs moins directe. Néanmoins, malgré cette complexité, il est impératif pour les assureurs de prêter une attention particulière à ces données.

Nous allons maintenant nous pencher sur les interactions entre les variables. Le Partial Dependence Plot (PDP) en 2D s'avère particulièrement éclairant à cet égard. Il permet de visualiser comment une variable spécifique influence la prédiction, tout en intégrant les effets et interactions des autres variables du modèle. Ce qui rend le PDP si efficace pour mettre en lumière les interdépendances est sa capacité à isoler et à illustrer l'effet d'une variable tout en moyennant les effets des autres variables. Ainsi, plutôt que de voir l'effet d'une variable dans le contexte d'un échantillon de données fixes, le PDP montre cet effet sur l'ensemble du jeu de données, révélant les nuances et les tendances qui pourraient autrement être masquées.

Ce graphique s'interprète comme suit :

- Zones de forte interaction : Sur le graphique, lorsque certaines zones affichent une teinte chaude, cela indique qu'il y a une interaction marquée entre les variables à cet endroit précis. Une couleur intense à la croisée de certaines valeurs des variables signifie que leur combinaison a un impact notable sur le résultat ou la prédiction.
- Zones de faible interaction : Dans les zones où la teinte est uniforme ou peu variée, cela indique que l'interaction entre les variables est faible ou inexiste. Autrement dit, l'une des variables n'a pas ou peu d'influence sur le résultat lorsqu'elle est combinée à certaines valeurs de l'autre variable.
- Contours serrés : La présence de contours rapprochés sur le graphique est un indicateur d'une interaction forte et rapide entre les variables. Cela signifie que de petites variations dans l'une des variables peuvent entraîner des changements significatifs dans le résultat, surtout lorsque elles sont combinées à certaines valeurs de l'autre variable.

Commençons par les interactions entre les variables relatives à de l'argile :

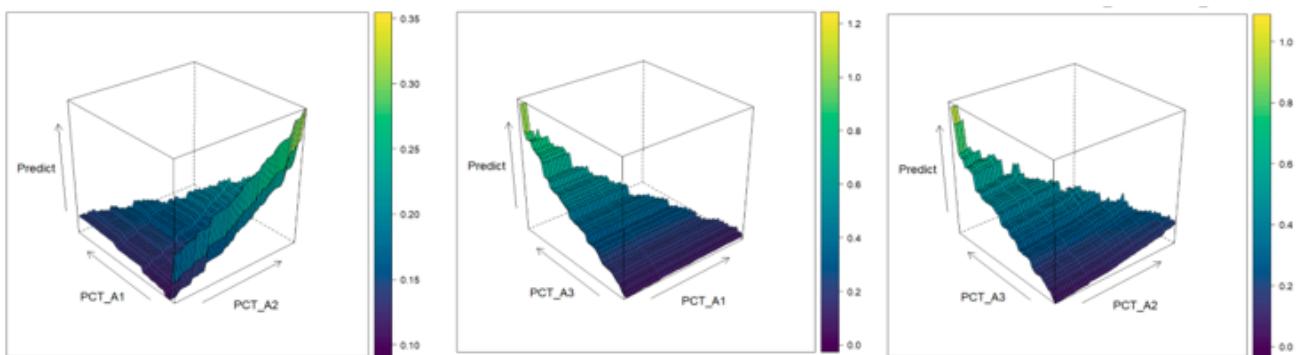


FIGURE 4.4 – Analyse des interactions entre les types d'argile

Il est cohérent de remarquer qu'à mesure que la proportion d'argile de type A3 augmente, en comparaison avec l'argile de type A1, la fréquence des sinistres monte également. Cette observation est confirmée par les graphiques d'interaction entre A3 et A2, ainsi qu'entre A3 et A1. Cela vient étayer la notion selon laquelle une concentration accrue d'argile favorisant les phénomènes de retrait-gonflement entraîne une augmentation du risque de sinistres.

Regardons des variables un peu moins intuitives, regardons les interactions entre les variables d'argile et de la hauteur des bâtiments :

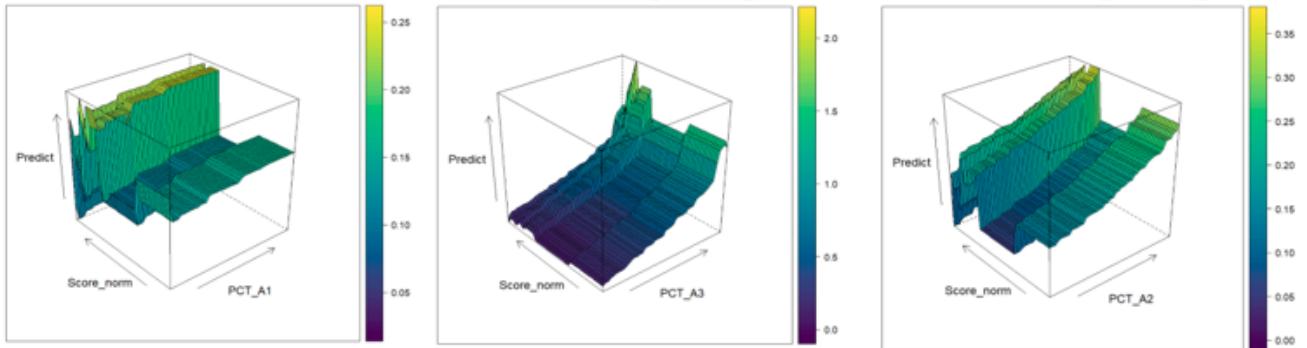


FIGURE 4.5 – Analyse des interactions entre la variable *Score_norm* et les types d’argile

L’analyse des données dévoile des constatations notables concernant la sinistralité en relation avec les spécificités des sols et des bâtiments. Premièrement, une nette augmentation de la sinistralité est observée sur les terrains à forte concentration d’argile. Cette tendance s’accentue de manière très prononcée pour les sols à forte présence d’argile de type 3 lorsque les bâtiments sont particulièrement hauts. En effet, lorsque ces deux variables - la hauteur des bâtiments et la présence d’argile de type 3 - atteignent leurs valeurs les plus élevées, l’effet combiné sur la sinistralité est très accru. Cette interaction entre ces deux facteurs renforce considérablement le niveau de risque.

En ce qui concerne les interactions entre les autres types d’argile et la hauteur des bâtiments, on relève également des tendances notables. L’interaction entre la hauteur des bâtiments et l’argile de type 1 et de type 2 est mise en évidence par l’apparition de zones aux couleurs chaudes et de contours serrés sur le graphique. Ces zones indiquent des niveaux élevés de sinistralité, ce qui implique que ces variables doivent également être considérées conjointement lors de la gestion du portefeuille de polices d’assurance. En somme, il est essentiel de prendre en compte ces interactions pour évaluer correctement les risques associés à chaque police et adapter en conséquence la stratégie d’assurance.

Intéressons-nous aux interactions entre les bâtiments construits avant 1970 et les différents types d’argile :

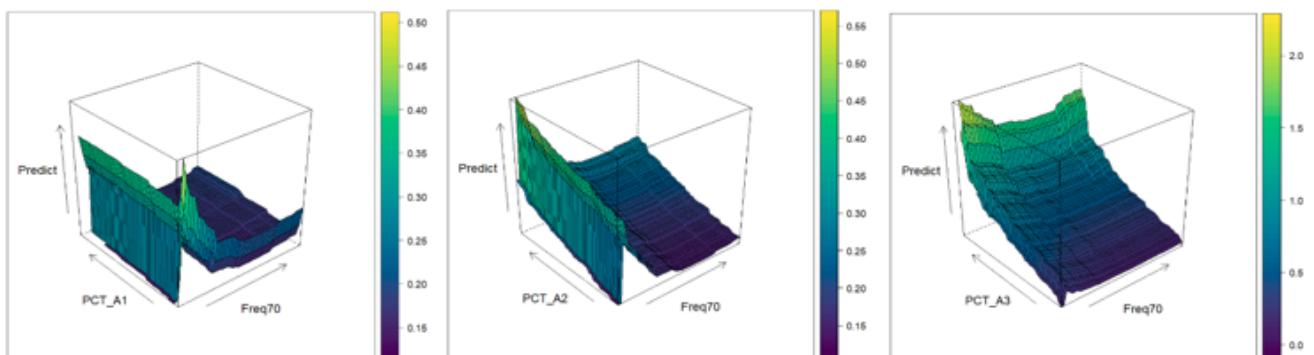


FIGURE 4.6 – Analyse des interactions entre la variable *Freq70* et les types d’argile

En étudiant les interactions entre la variable de la fréquence de construction avant les années 70 et les types d'argile, des tendances distinctes émergent. La variable A3 se démarque nettement par son influence prédominante. Lorsque le type d'argile 3 est en augmentation, la sévérité croît presque linéairement. Cependant, pour les deux autres types d'argile, les interactions sont plus complexes. Dans ces deux cas, on observe des interactions particulièrement marquées lorsque le nombre de bâtiments construits avant les années 70 est faible.

Comparons à présent les argiles et la fréquence des bâtiments construits après les années 1990.

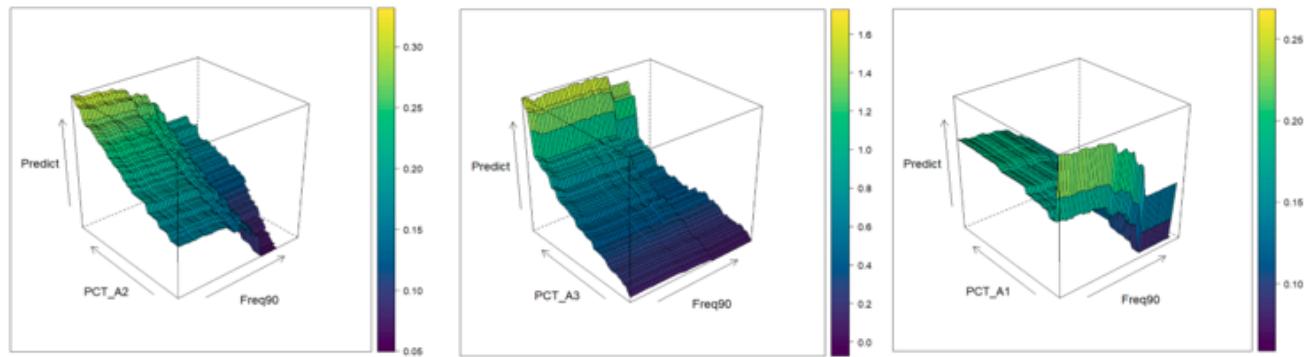


FIGURE 4.7 – Analyse des interactions entre la variable Freq90 et les types d'argile

Dans notre étude des interactions avec les différents types d'argile, nous observons des comportements contrastés selon le type d'argile en question. Pour l'argile A2, les interactions se montrent plus prononcées lorsque la présence d'argile est minimale, s'approchant des valeurs autour de 90. À l'opposé, en ce qui concerne l'argile A1, les interactions sont d'autant plus marquées que la quantité d'argile est faible. Cette distinction entre les deux types d'argile souligne la nécessité de les considérer différemment dans nos analyses et stratégies.

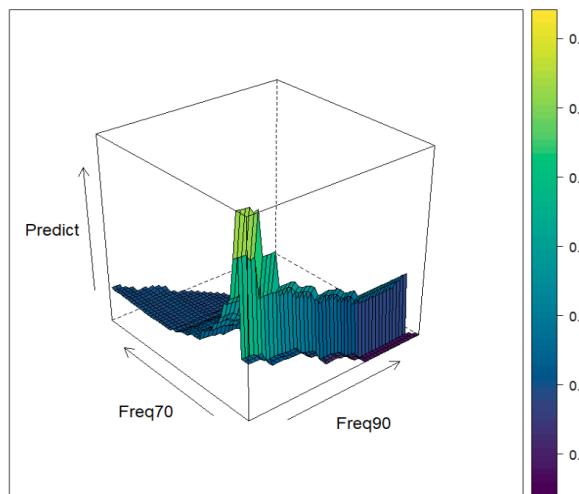


FIGURE 4.8 – Analyse des interactions entre la variable Freq90 et la variable Freq70

En examinant de près les deux fréquences de construction, une particularité se démarque. Lorsque ces deux fréquences sont représentées en proportions quasi équivalentes, nous constatons la présence de contours resserrés. Cette observation suggère que c'est dans ces conditions spécifiques que les interactions entre ces variables sont les plus intenses. Il semble que l'équilibre entre ces deux fréquences amplifie leur interdépendance, conduisant à des effets marqués dans notre étude.

Chapitre 5

Projection de la sinistralité

5.1 Construction de la composition du portefeuille

Nous avons déjà en notre possession les projections climatiques pour les périodes futures. La prochaine étape consiste à anticiper l'évolution démographique et architecturale de chaque commune. Pour ce faire, il est crucial d'estimer le nombre de maisons qui seront construites durant ces périodes. À cet effet, nous nous appuierons sur les données de construction fournies par l'INSEE pour chaque commune, réparties sur différentes années. Grâce à ces données, nous serons en mesure de déterminer un indice représentant le nombre de nouvelles constructions. Cet indice sera désigné sous le terme de *indice_construc* :

$$\text{indice_construc} = \frac{1}{N} \sum_{i=1}^N \frac{\text{nb_construc}_i - \text{nb_construc}_{i-1}}{\text{nb_construc}_{i-1}} \quad (5.1)$$

Avec :

- N l'année d'observation
- *nb_construc* le nombre de constructions présentes dans la commune

En ne prenant en compte que les nouvelles constructions, nous évitons de comptabiliser deux fois un bâtiment, notamment si celui-ci est construit en remplacement d'un précédent détruit.

Cependant, cette hypothèse reste limitée car elle ne tient pas compte des législations en vigueur sur la préservation des espaces ruraux, lesquelles encouragent davantage la rénovation que la construction de nouveaux bâtiments. En effet, des dispositifs comme la loi n° 2014-366 du 24 mars 2014, dite loi ALUR (Accès au Logement et un Urbanisme Rénové), mettent l'accent sur la protection des terres agricoles et des zones naturelles, limitant ainsi l'urbanisation excessive. Cette législation favorise la réhabilitation du bâti existant, en insistant sur la densification des zones déjà urbanisées et la lutte contre l'étalement urbain, afin de préserver les espaces ruraux et naturels. Ainsi, l'indice de construction pris en compte pourrait être surestimé. En effet, les réglementations visant à préserver les espaces ruraux et naturels, comme celles mentionnées dans la loi ALUR, imposent des contraintes sur l'étalement urbain et favorisent la rénovation plutôt que la construction de nouveaux bâtiments. De ce fait, l'augmentation de l'indice de construction pourrait ne pas refléter pleinement les réalités imposées par ces législations. Le potentiel de construction est donc limité par ces politiques publiques.

Il ne nous reste qu'à appliquer l'indice à nos données pour obtenir une estimation du nombre

de maisons dans la commune : $nb_maison * (indice_construc)^n$ où n est le nombre d'années écoulées (exemple entre 2022 et 2050 on prend n=28)

5.2 Analyse préliminaire

Face à la menace croissante du changement climatique, l'assurance multirisques habitation (MRH) pour la sécheresse gagne en importance. En effet, les sécheresses, amplifiées par les dérèglements climatiques, peuvent occasionner des dommages structurels considérables aux bâtiments, comme des fissures dues au retrait-gonflement des argiles. Les fondations des habitations sont particulièrement vulnérables à ces mouvements de terrain, engendrant des coûts de réparation souvent élevés pour les propriétaires. Dans ce contexte, il devient impératif pour les assureurs MRH de scrupuleusement étudier l'évolution des variables climatiques liées à la sécheresse. Ceci leur permettra d'ajuster leurs offres, de mieux tarifer les risques, et d'offrir une protection optimale à leurs assurés. Anticiper ces évolutions est essentiel pour la pérennité des compagnies d'assurance, tout en assurant la sérénité des propriétaires face aux caprices du climat.

Analyse du SWI

Une des variables climatiques majeures permettant d'expliquer la sécheresse est le SWI (Soil Water Index). Ainsi, nous nous concentrerons sur l'évolution temporelle de cet indice, en tenant compte des différents scénarios proposés par le GIEC.

On va donc comparer pour la période de 2050 les nombres de jours moyens passés en jours de sécheresse selon les différents scénarios :

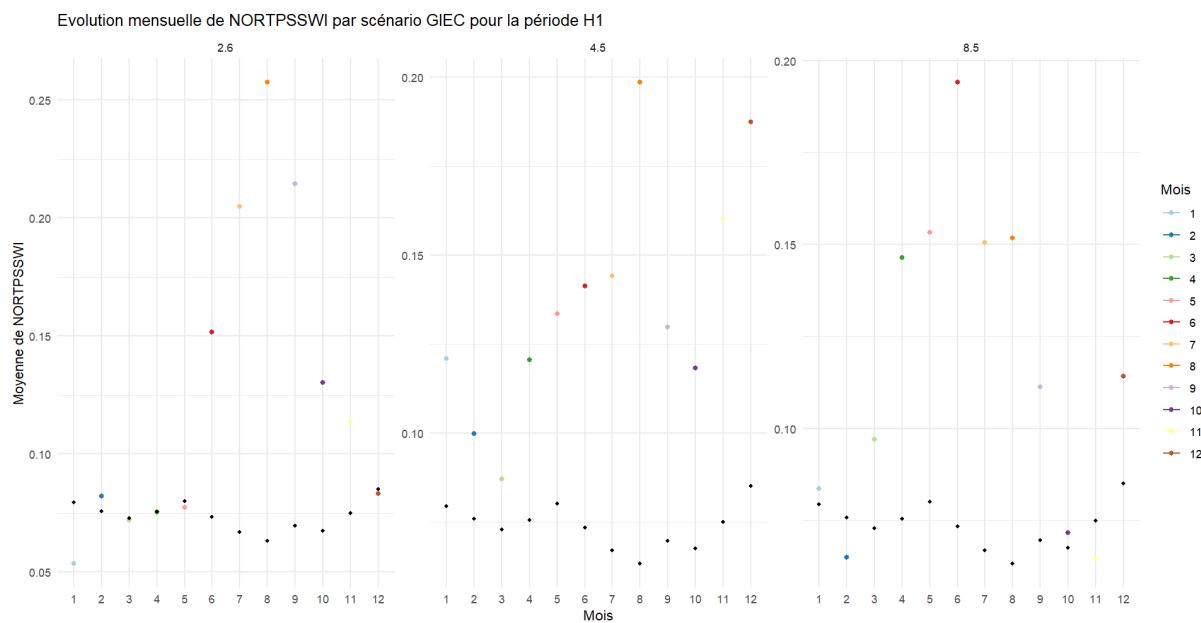


FIGURE 5.1 – Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2050

L'impact du changement climatique est devenu une préoccupation majeure pour notre planète et son avenir. À l'horizon 2050, l'urgence de cette situation est mise en évidence par des projections

qui montrent une augmentation alarmante du nombre moyen de jours de sécheresse, par rapport à une période de référence déjà confrontée à des défis climatiques.

En examinant ces projections plus en détail, on constate une intéressante nuance : lors de la comparaison entre différents scénarios envisagés par les climatologues, on remarque que les valeurs, en termes d'augmentation des jours de sécheresse, tendent à se regrouper, suggérant une certaine convergence dans leurs prédictions. Cependant, il est important de mettre en avant le scénario 8.5. Ce scénario, souvent qualifié de "scénario du pire cas" en raison de ses prévisions basées sur des taux élevés d'émissions de gaz à effet de serre, présente une moyenne de jours de sécheresse légèrement supérieure à celle des autres. Bien que cette différence puisse sembler minime à première vue, elle pourrait avoir des implications majeures, notamment en termes d'agriculture, de gestion de l'eau et d'écologie des écosystèmes.

Un autre élément d'importance cruciale à souligner est la présence de pics de sécheresse aigus pendant les mois d'été. Ces pics ne sont pas seulement l'expression d'un nombre accru de jours sans pluie, mais traduisent une intensité accrue des périodes de sécheresse. Les conséquences sont multiples : sols plus arides, réduction des réserves d'eau douce, stress sur la faune et la flore, ainsi qu'un risque accru pour les activités humaines, qu'il s'agisse de l'agriculture ou de la consommation d'eau.

Examinons donc à un horizon plus éloigné et observons les résultats obtenus. Regardons donc en 2075 :

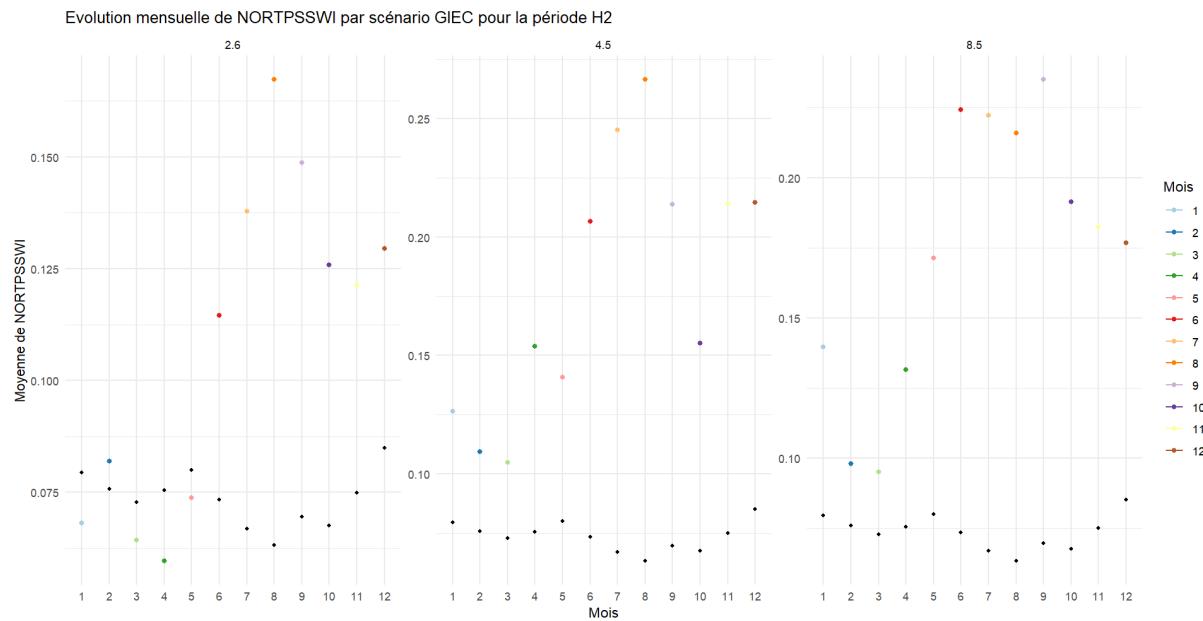


FIGURE 5.2 – Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2075

L'analyse des projections climatiques pour 2075 révèle des tendances inquiétantes en matière de sécheresse en France, avec des implications directes pour le secteur des assurances MRH.

Tout d'abord, l'augmentation du nombre moyen de jours de sécheresse touchant tous les mois

de l'année, même en dehors de la période estivale habituellement plus sèche, indique que les propriétés pourraient être exposées à des risques accrus tels que des fissures dans les fondations ou d'autres dommages structurels. Cela pourrait entraîner une hausse des demandes d'indemnisation liées à ces dommages.

Les différences entre les scénarios sont particulièrement frappantes. Sous le scénario 2.6, la France bénéficierait d'une certaine clémence, avec nettement moins de jours de sécheresse. Pour les compagnies d'assurance, cela pourrait signifier un risque moindre dans certaines régions, permettant potentiellement de maintenir des primes d'assurance plus compétitives.

Cependant, le scénario 4.5 présente un défi bien plus grand. En enregistrant le plus grand nombre moyen de jours de sécheresse sur une période étendue, il est probable que les propriétés soient soumises à des stress prolongés, augmentant ainsi le risque de dommages. Les assureurs pourraient alors avoir besoin d'ajuster leurs tarifs en conséquence.

Le scénario 8.5, le plus pessimiste des trois, montre un paysage où la sécheresse est une réalité presque constante. Pour les assureurs, cela représenterait un environnement de risque extrême, où les politiques traditionnelles de couverture pourraient devoir être repensées.

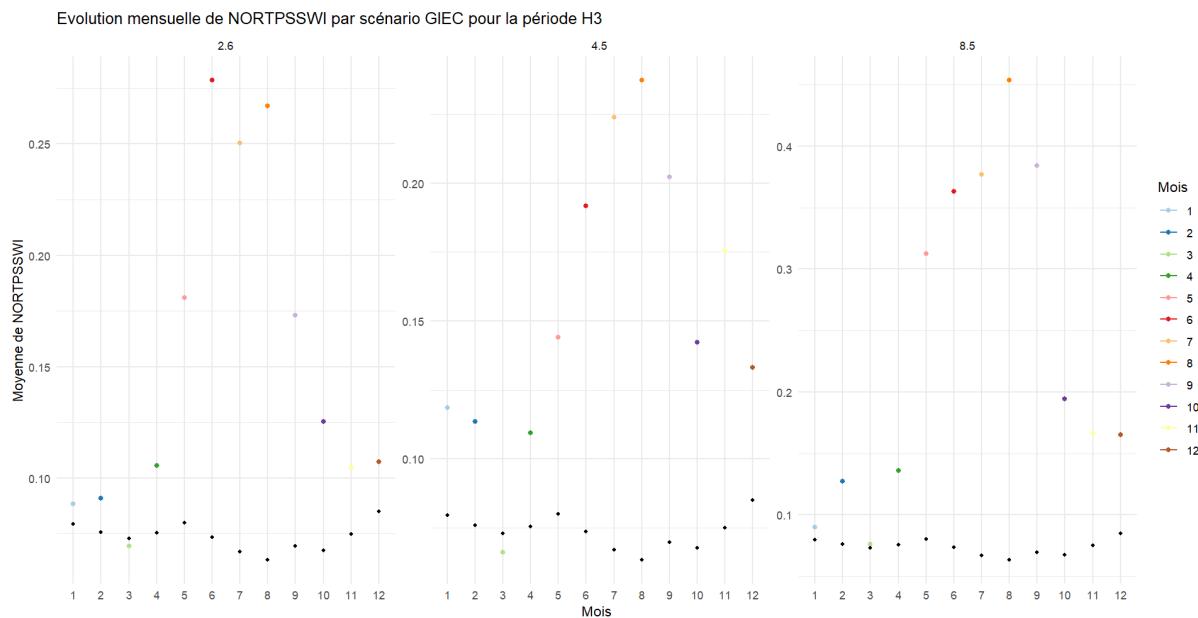


FIGURE 5.3 – Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2100

Pour l'année 2100, selon les données du scénario 8.5, nous observons une nette amplification du nombre moyen de jours subissant la sécheresse. Au-delà d'un simple constat, cette progression confirme un réchauffement climatique d'une magnitude préoccupante. Les effets de ces changements climatiques se manifestent avec une intensité qui pourrait avoir des conséquences sévères pour notre environnement. Pour affiner notre compréhension de ces impacts, il serait pertinent d'étudier ces phénomènes de manière régionale, car les effets peuvent varier considérablement d'une région à l'autre.

2023



FIGURE 5.4 – Critère SWI actuel

2030

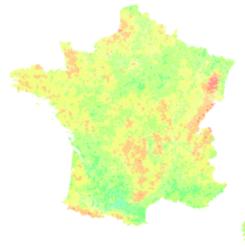


FIGURE 5.5 – Critère SWI à l'année 2050

2040

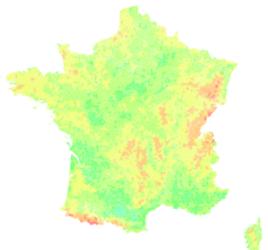


FIGURE 5.6 – Critère SWI à l'année 2075

2050

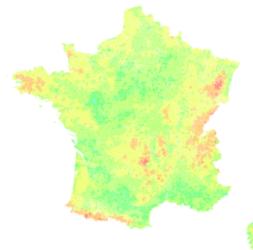


FIGURE 5.7 – Critère SWI à l'année 2100

Sur le graphique présenté, nous mettons en exergue le score SWI, calculé en fonction de sa période de retour. Cette métrique est directement indicative de la fréquence à laquelle une commune serait éligible au critère de SWI. Un examen minutieux révèle une évolution significative du SWI entre 2022 et 2050. Toutefois, pour les années subséquentes, ce score semble montrer une tendance à la stabilisation. En effet, les variations observées sont minimales, et l'on note même par moments un léger recul de la fréquence de reconnaissance en tant que CatNat. Une explication plausible à ce phénomène est que les terrains, ayant été soumis à des conditions de sécheresse pendant de nombreuses années, peinent à satisfaire la condition exigeant qu'ils soient moins humides que l'ensemble des SWI enregistrés pour le mois sur une période de 25 ans.

Analyse des précipitations

La France, avec sa diversité géographique, a toujours été influencée par une variété de régimes de précipitations. Historiquement, les régions du nord et de l'ouest ont été caractérisées par des précipitations plus régulières tout au long de l'année, tandis que le sud et l'est connaissent des étés plus secs suivis d'épisodes pluvieux intenses en automne. Cependant, avec l'avènement du changement climatique, les schémas traditionnels de précipitations ont commencé à évoluer. Des épisodes de pluies plus intenses alternant avec de longues périodes de sécheresse sont désormais fréquemment observés. Ces changements ont des implications notables sur les sols, la gestion de l'eau, et même sur la stabilité des infrastructures. Il est essentiel d'analyser en détail les variations des précipitations afin de comprendre leur impact global.

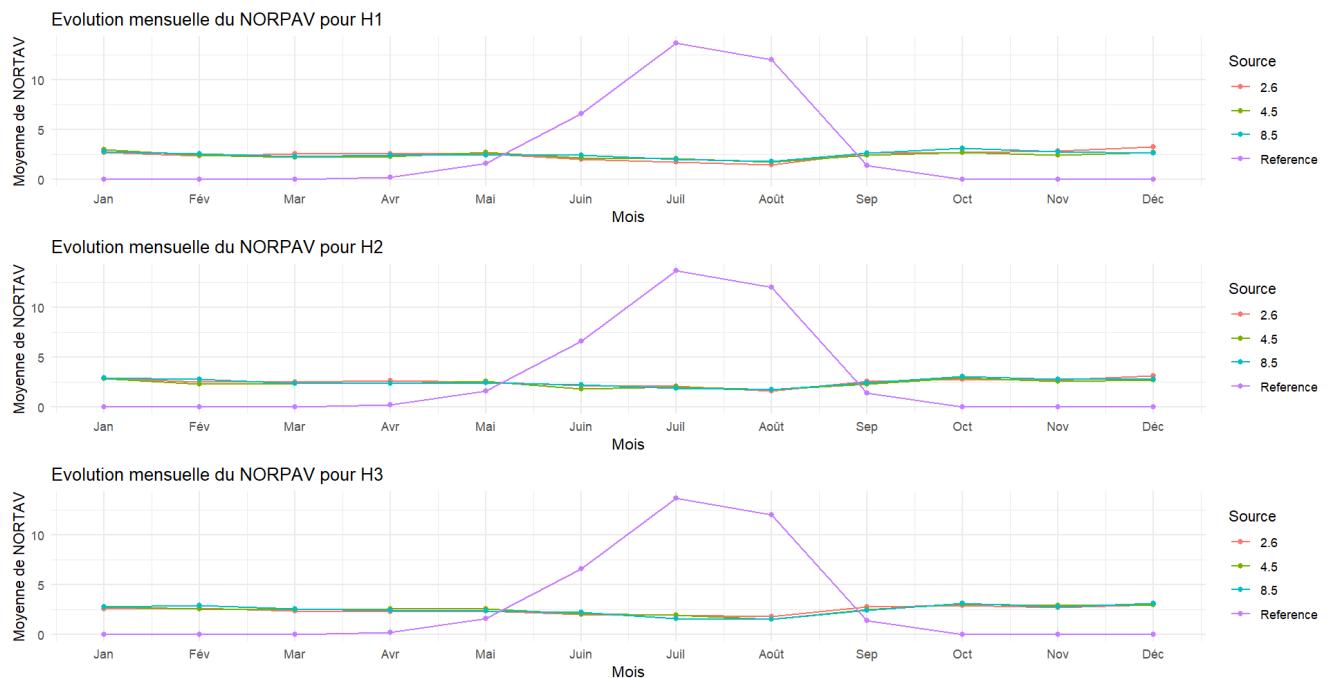


FIGURE 5.8 – Évolution mensuelle du nombre de pluie par jour par période

L'analyse du graphique illustre une tendance distincte à travers tous les scénarios envisagés. Durant la période estivale, une diminution marquée des précipitations se manifeste, potentiellement accentuée par les effets du réchauffement climatique. Par contre, les autres périodes de l'année montrent une légère augmentation des précipitations. Cette compensation des volumes d'eau en dehors de l'été pourrait avoir des conséquences directes sur le remplissage des nappes phréatiques. En effet, un apport plus constant de précipitations tout au long de l'année peut favoriser une recharge plus uniforme des réservoirs souterrains.

Analyse de l'évapotranspiration

L'évapotranspiration joue un rôle crucial dans le contexte de la sécheresse, particulièrement en ce qui concerne les assurances liées aux Maisons et Résidences Habitées (MRH). Ce processus combiné d'évaporation de l'eau à la surface du sol et de transpiration des plantes influe directement sur le volume d'eau disponible dans le sol. Dans le cadre de la sécheresse, une évapotranspiration accrue peut entraîner un assèchement rapide des sols, exacerbant le phénomène de retrait-gonflement des argiles.

Observons l'évolution des paramètres :

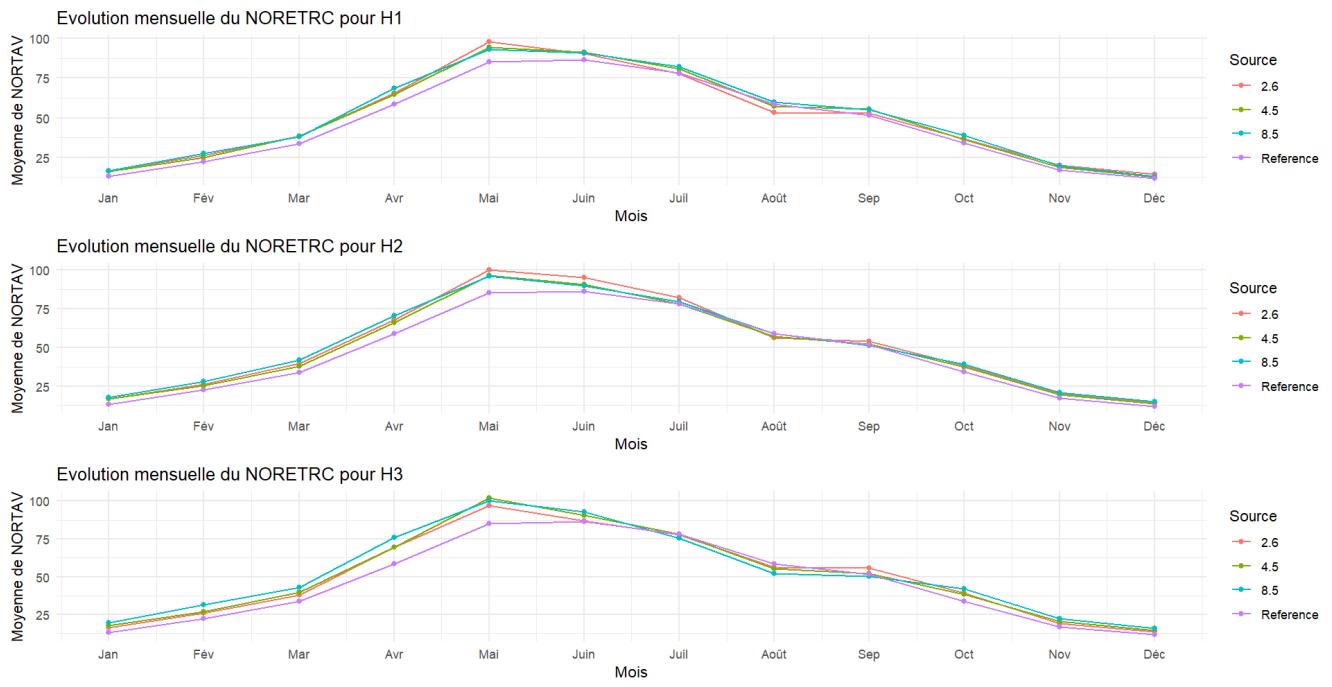


FIGURE 5.9 – Évolution mensuelle de l'évapotranspiration moyenne par mois par période

L'analyse des données montre que l'évapotranspiration, contrairement à certaines attentes, ne semble pas évoluer significativement avec le temps. Cette constance est quelque peu surprenante, étant donné les variations notables des températures et des précipitations que nous avons observées. Plusieurs interprétations sont possibles. Premièrement, il se pourrait que les augmentations des températures soient contrebalancées par d'autres facteurs environnementaux, limitant ainsi les variations de l'évapotranspiration. Deuxièmement, les plantes pourraient adapter leurs mécanismes de transpiration face à des conditions changeantes, ce qui maintiendrait un taux d'évapotranspiration stable. Enfin, cette stabilité pourrait aussi indiquer que, malgré les changements climatiques, certaines propriétés hydrologiques du sol restent inchangées. Toutefois, cette constance ne doit pas être interprétée comme une absence de risque.

Projection de la sinistralité

Après une analyse approfondie des projections climatiques et de leurs conséquences possibles en matière de sécheresse et d'évapotranspiration, Nous allons quantifier l'impact concret de ces changements sur nos populations et infrastructures. À cet effet, nous orientons notre attention vers l'estimation de la fréquence moyenne des sinistres pour les périodes 2024-2050, 2050-2075, et 2075-2100. Cette fréquence moyenne de sinistres a été calculée en établissant le rapport entre le nombre de sinistres et le volume de risques, (ce dernier étant extrapolé pour le marché dans son ensemble à partir des données collectées à l'échelle nationale). L'objectif de ces projections, basées sur les trois scénarios du GIEC, est de quantifier la fréquence attendue des sinistres durant ces périodes clés. Cette approche est destinée à fournir une compréhension précise des risques futurs et à préparer les stratégies nécessaires en termes de prévention, de tarification et de provisionnement.

Nous allons maintenant visualiser les fréquences de sinistres pour le mois d'août, mois durant lequel nous observons le plus grand nombre de sinistres.

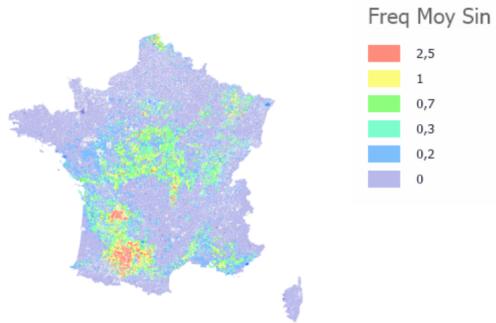


FIGURE 5.10 – Fréquence de sinistres pour la période de référence

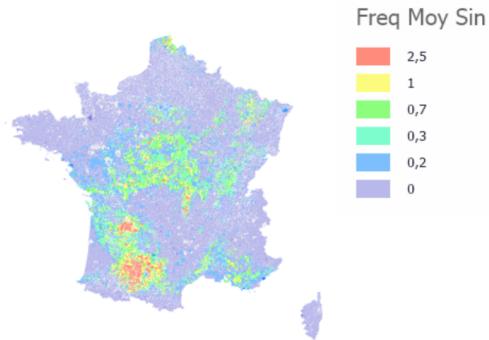


FIGURE 5.11 – Fréquence de sinistres pour la période de 2024-2050 et le scénario 2.6

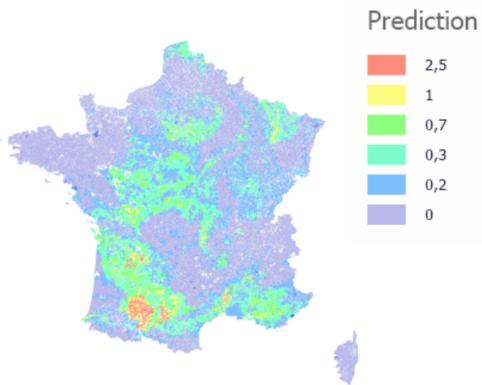


FIGURE 5.12 – Fréquence de sinistres pour la période de 2050-2075 et le scénario 2.6

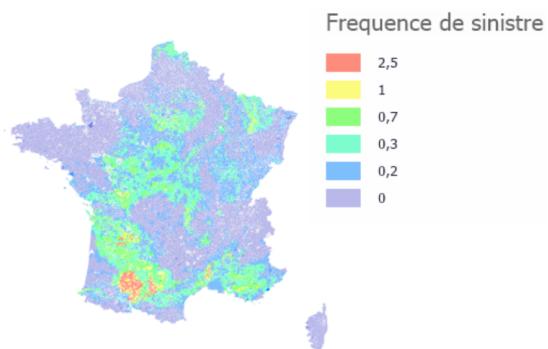


FIGURE 5.13 – Fréquence de sinistre pour la période de 2075-2100 et le scénario 2.6

On remarque que les fréquences de sinistres présentent globalement une tendance similaire au SWI qui a été présenté précédemment. Cependant, on note une diminution des fréquences les plus élevées. En revanche, des fréquences de sinistres moins élevées font leur apparition, bien qu'elles soient plus nombreuses.

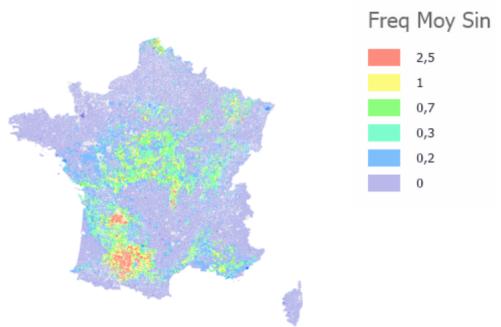


FIGURE 5.14 – Fréquence de sinistres pour la période de référence

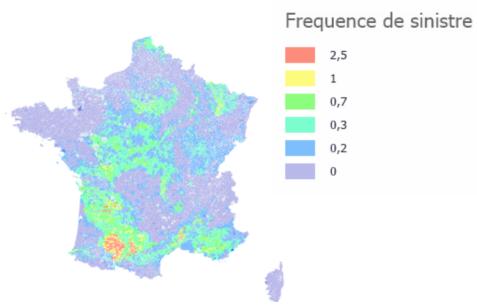


FIGURE 5.15 – Fréquence de sinistres pour la période de 2024-2050 et le scénario 4.5

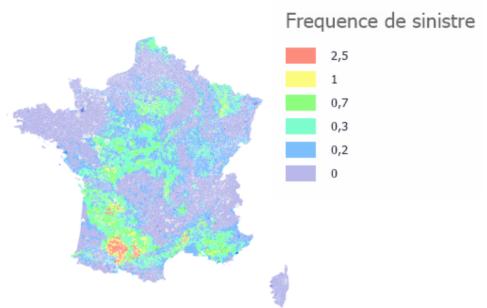


FIGURE 5.16 – Fréquence de sinistres pour la période de 2050-2075 et le scénario 4.5

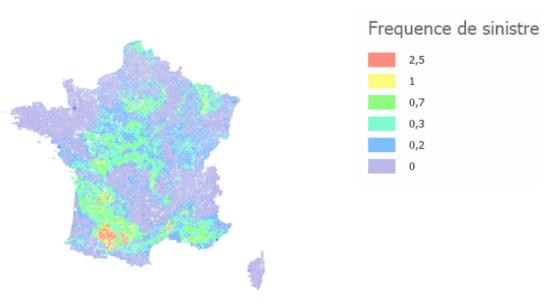


FIGURE 5.17 – Fréquence de sinistres pour la période de 2075-2100 et le scénario 4.5

Dans ce scénario, nous remarquons des tendances semblables à celles identifiées dans le scénario 2.6. Toutefois, un aspect distinctif de ce scénario se manifeste : une augmentation marquée du nombre de sinistres dans la région de l'Est. Cette observation est d'autant plus intrigante qu'entre les années 2075 et 2100, un nombre réduit de communes semble être impacté par la sécheresse. Plus précisément, dans la zone à l'est de Paris, on discerne une baisse notable de la fréquence des sinistres. Ce recul suggère une évolution potentiellement positive.

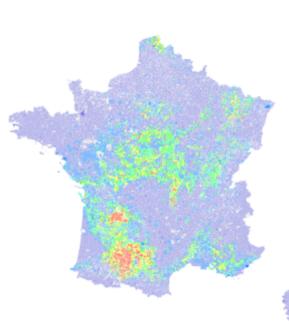


FIGURE 5.18 – Fréquence de sinistres pour la période de référence

Freq Moy Sin

2,5
1
0,7
0,3
0,2
0

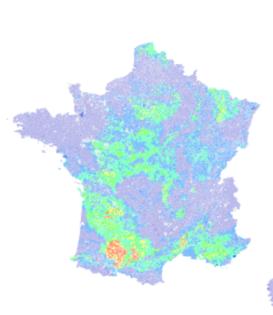


FIGURE 5.19 – Fréquence de sinistres pour la période de 2024-2050 et le scénario 8.5

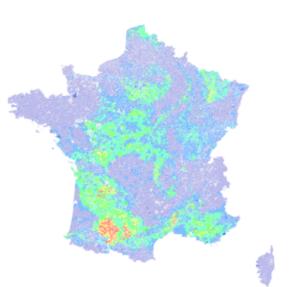


FIGURE 5.20 – Fréquence de sinistres pour la période de 2050-2075 et le scénario 8.5

Fréquence de sinistre

2,5
1
0,7
0,3
0,2
0

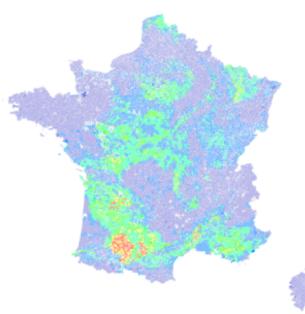


FIGURE 5.21 – Fréquence de sinistres pour la période de 2075-2100 et le scénario 8.5

Fréquence de sinistre

2,5
1
0,7
0,3
0,2
0

Pour le scénario 8.5, nous observons une tendance similaire à celle des deux autres scénarios. Toutefois, les fréquences y semblent légèrement supérieures et, en projetant à l'année 2100, les zones de sinistralité apparaissent plus étendues.

En conclusion, les trois scénarios présentent des similitudes notables. Comme on pouvait s'y attendre, le scénario 8.5 montre une sinistralité légèrement accrue par rapport au scénario 2.6. Toutefois, à partir de 2050, on note une stabilisation des fréquences sinistrées dans tous les scénarios. Cette évolution pourrait être attribuée aux observations préliminaires qui montrent une pluviométrie en nette diminution, combinée à des conditions d'évapotranspiration quasi-identiques. Cette combinaison réduit le phénomène de retrait-gonflement des sols. En effet, un sol déjà sec, s'il est peu réhydraté, minimise ce phénomène. De la même manière, le sol peine à être plus sec que son état observé au cours des 25 dernières années, comme le suggèrent les analyses avec le SWI. Pour toutes ces raisons, il semblerait que les effets de la sécheresse et les déclarations Catastrophes Naturelles (CatNat) tendent à se stabiliser. Certains scénarios indiquent même une légère régression de ce phénomène.[9]

La région de la Bretagne se distingue comme étant la moins affectée par les sinistres liés à la sécheresse. Cette particularité s'explique en grande partie par la faible présence d'argile sensible au phénomène de retrait-gonflement dans cette zone, réduisant ainsi la probabilité de sinistres. À l'opposé, le croissant argileux, souvent cité dans les études relatives aux sinistres liés à la sécheresse, est une zone géographique spécifique de la France où les sols argileux prédominent.

Cette région s'étend du nord-est de la région parisienne, traverse la Bourgogne, et se poursuit jusqu'au centre du pays, englobant des départements tels que le Val-d'Oise, l'Oise, la Seine-et-Marne, l'Yonne, la Nièvre, et le Cher, pour n'en nommer que quelques-uns, et demeure la plus impactée.

Dans tous les scénarios étudiés, nous observons une évolution notable concernant le bassin méditerranéen. Alors qu'il était auparavant relativement épargné, les projections suggèrent que cette région commencera à connaître davantage de sinistres à l'avenir. Cette tendance est probablement induite par les conditions climatiques de cette zone : des chaleurs intenses combinées à une réduction des précipitations due au changement climatique. Ainsi, bien que la Bretagne reste préservée, d'autres régions, comme le bassin méditerranéen, pourraient voir leur vulnérabilité augmenter face à ce phénomène.

Chapitre 6

Projection de la sévérité

6.1 Indice de construction en bâtiment

La projection des coût des sinistres est une démarche essentielle pour les acteurs de la construction, les assureurs, ainsi que pour les décideurs publics. L'un des paramètres à prendre en compte dans cette projection est l'inflation de l'indice de la construction en bâtiment. L'indice de la construction reflète l'évolution des coûts associés aux matériaux, à la main-d'œuvre, et à d'autres facteurs qui influencent directement le coût de la construction. En projetant les sinistres, si l'on ne tient pas compte de cette inflation, on risque de sous-estimer significativement les coûts futurs associés à la réparation ou à la reconstruction des bâtiments endommagés. De plus, l'inflation de cet indice peut varier en fonction de divers facteurs économiques, technologiques ou même sociopolitiques. Par conséquent, en intégrant l'inflation de l'indice de la construction dans la projection des sinistres, on s'assure d'une estimation plus réaliste et pertinente des coûts potentiels, ce qui permet à tous les intervenants de mieux planifier, budgetter et anticiper les risques à venir.

L'Institut National de la Statistique et des Études Économiques (INSEE) fournit un indice de la construction depuis 1990, qui est un outil essentiel pour comprendre l'évolution des coûts dans le secteur du bâtiment. Pour calculer le taux d'inflation annuel moyen à partir de cet indice, nous utilisons la formule suivante :

$$r = \left(\frac{I_{2022}}{I_{1990}} \right)^{\frac{1}{2022-1990}} - 1 \quad (6.1)$$

r est le taux d'inflation annuel moyen, I_{2022} est l'indice de construction en 2022, et I_{1990} est l'indice de construction en 1990. Si nous disposons d'une estimation du coût moyen des sinistres par commune en 2022, notée C_{2022} , alors pour estimer le coût moyen des sinistres par commune en 2050, C_{2050} , en tenant compte de l'inflation projetée, nous utilisons la formule de croissance composée :

$$C_{2050} = C_{2022} * (1 + r)^{2050-1990} \quad (6.2)$$

Ainsi, en combinant les données fournies par l'INSEE et nos estimations pour 2022, nous pouvons obtenir une projection éclairée du coût moyen des sinistres par commune pour l'année 2050.

Calculer le taux d'inflation annuel moyen de l'indice du coût de la construction en se basant sur des données remontant à 1990 n'est pas optimal. En effet, l'inflation cumulée sur une période aussi longue, entre 1990 et aujourd'hui, peut masquer les nuances et les tendances plus récentes. Cette approche risque de donner un aperçu biaisé, influencé par des événements économiques anciens qui ne sont plus nécessairement pertinents pour comprendre les dynamiques actuelles et

futures. Il est donc plus judicieux de se concentrer sur une période plus réduite, telle que les cinq dernières années. Cela permettrait de capturer une image plus actuelle de l'évolution de l'indice et de fournir une base plus solide pour les analyses et projections futures.
Sous ces conditions on obtient un taux d'inflation moyen $r = 0,0338$.

6.2 Analyse de l'évolution des coûts en fonction des scénarios

Suite aux observations précédentes, nous avons décidé de nous concentrer sur un scénario précis. Compte tenu des similitudes entre les différents scénarios, focalisons-nous sur l'évolution d'un département spécifique. Nous examinerons le département numéro 31, connu sous le nom de Haute-Garonne, car sa sinistralité est particulièrement significative. Le graphique illustre la somme totale des coûts des sinistres liés à la sécheresse pour le département.

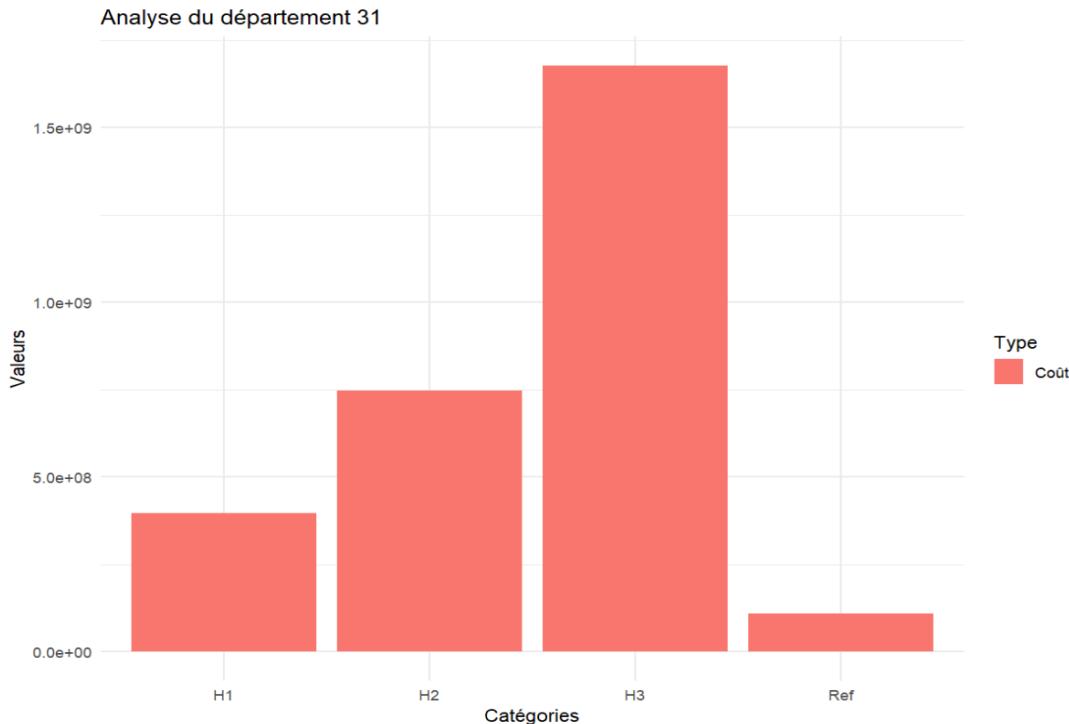


FIGURE 6.1 – Évolution des coûts des sinistres pour le département de Haute-Garonne

Nous allons à présent réaliser une treemap afin de visualiser quels sont les départements dont le coût est plus élevé. Une treemap, également appelée carte arborescente, est une visualisation graphique utilisée pour représenter une structure hiérarchique de données. Sous la forme de rectangles imbriqués, chaque segment de la hiérarchie est dépeint par un rectangle dont la taille est proportionnelle à une certaine quantité. Ainsi, plus le rectangle est grand, plus la valeur qu'il représente est importante. La couleur de ces rectangles peut également véhiculer des informations, en fonctionnant généralement sur un gradient : par exemple, un ton plus foncé pourrait signifier une valeur plus élevée et vice versa. Les étiquettes accompagnent ces rectangles pour faciliter leur identification. Cependant, si vous observez une profusion d'étiquettes ou des étiquettes trop petites pour être déchiffrées, cela peut suggérer une surcharge d'informations. En essence, les treemaps sont excellentes pour obtenir une vision globale des données, déceler des tendances

ou des schémas généraux, plutôt que pour déduire des valeurs exactes. Pour une précision numérique, d'autres méthodes de visualisation ou présentations sont recommandées. L'avantage, dans notre contexte, est que cela nous donne une perspective claire de la proportion que ce coût représente par rapport au coût total, indépendamment de l'effet d'inflation.

Coût Référence par département

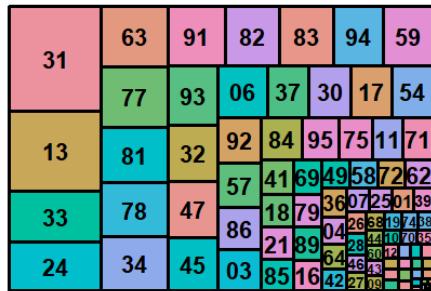


FIGURE 6.2 – Fréquence de sinistres pour la période de référence

Coût H1 par département

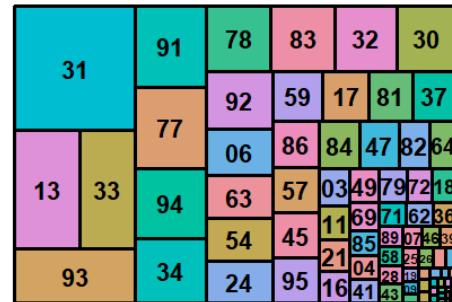


FIGURE 6.3 – Fréquence de sinistres pour la période de 2050 et le scénario 8.5

Coût H2 par département

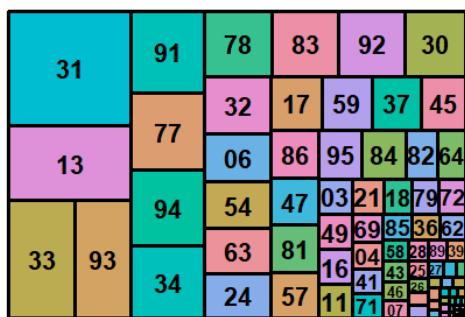


FIGURE 6.4 – Fréquence de sinistres pour la période de 2075 et le scénario 8.5

Coût H3 par département

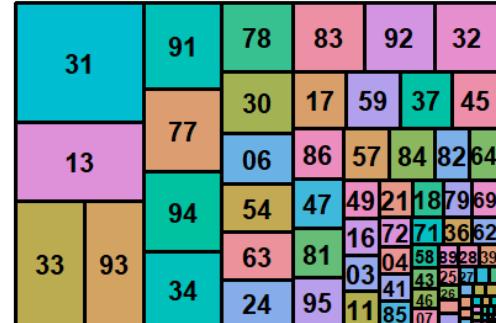


FIGURE 6.5 – Fréquence de sinistres pour la période de 2100 et le scénario 8.5

Nous observons que, conformément à ce que nous avions noté précédemment, la somme des sinistres tend à se stabiliser. L'augmentation constatée sur le graphique précédent est attribuable à l'inflation. Toutefois, l'augmentation due à l'accroissement du nombre de sinistres est principalement observée en conséquence de l'évolution de 2022 à 2050.

Dans le cadre d'une saine gestion des risques, l'évaluation de la solvabilité est d'une importance cruciale pour les assureurs. La prise en compte des variations du coût des sinistres joue un rôle prépondérant pour obtenir une évaluation précise du risque associé à des événements tels que la sécheresse. Dans le contexte actuel, où les changements climatiques rendent ces variations

plus volatiles et moins prévisibles, et où des changements de réglementation sont attendus, nécessitant un travail adapté pour les prendre en compte.

En effet, suite à la sécheresse exceptionnelle de 2022 ayant causé 2,9 milliards d'euros de dommages, principalement sur les sols argileux présents sous une grande partie des habitations en France, l'ordonnance n° 2023-78 a été présentée pour réformer le régime d'indemnisation des catastrophes naturelles (CatNat). Ce régime, qui a payé 16 milliards d'euros pour des dégâts liés à la sécheresse depuis 1989, voit ses critères élargis et assouplis. Des discussions sont actuellement en cours pour envisager une réduction de la période de retour de 25 ans à 10 ans, et remplacer la comparaison mensuelle des SWI (Standardized Water Index) par une comparaison annuelle.

Les principales modifications apportées par l'ordonnance au code des assurances sont les suivantes :

L'éligibilité à l'indemnisation ne se limite plus à l'intensité anormale d'un agent naturel, mais prend aussi en compte la succession d'événements de sécheresse d'ampleur notable. La garantie se limite désormais aux dommages qui menacent la solidité du bâtiment ou son usage normal. Des obligations sont imposées aux experts des assureurs pour évaluer les dommages. Les agents publics peuvent contrôler le respect de ces obligations. En cas de non-respect des obligations par un expert, des sanctions peuvent être appliquées, telles que l'invalidation du rapport d'expertise, une interdiction temporaire d'exercer, et des amendes. Certains bâtiments, notamment ceux construits sans permis ou ne respectant pas certaines dispositions du code de la construction post-2024, sont exclus des garanties pour les dommages liés aux mouvements de terrain dûs à la sécheresse et la réhydratation. Globalement, cette ordonnance vise à mieux prendre en compte les dégâts liés à la sécheresse en France, tout en mettant en place un cadre plus strict pour l'évaluation et l'indemnisation de ces dommages.[15]

Chapitre 7

Conclusion

L'étude approfondie présentée ici dévoile avec clarté les conséquences préoccupantes des changements climatiques, en mettant un accent particulier sur le lien causal entre l'augmentation des températures, la diminution des précipitations et la constance de l'évapotranspiration. L'interaction de ces éléments climatiques crée un cocktail inquiétant, menant inexorablement à l'assèchement des sols, notamment dans les régions méditerranéennes françaises.

En effet, la montée des températures, conjuguée à une moindre pluviométrie, intensifie le stress hydrique des sols. L'évapotranspiration constante, malgré ces conditions, témoigne d'une extraction continue de l'humidité du sol, exacerbant ainsi sa dessiccation. Ces constats mettent en lumière la fragilité de ces écosystèmes face aux bouleversements climatiques.

Dans l'évaluation des risques liés aux sinistres de sécheresse, certains paramètres se distinguent par leur importance. Le critère de SWI, joue un rôle significatif dans la prévision de ces événements. De plus, l'âge et la hauteur des bâtiments dans le portefeuille sont des facteurs importants à considérer. Ces éléments influencent la vulnérabilité des structures aux effets de la sécheresse. Toutefois, parmi tous ces paramètres, le type d'argile présent dans le sol est prépondérant. La nature argileuse du sol joue un rôle déterminant dans la susceptibilité du terrain au phénomène de retrait-gonflement, ce qui rend son étude essentielle dans l'évaluation des risques. Cependant, certains facteurs restent difficiles à appréhender, comme le facteur politique, par exemple. Il est également envisageable que l'accès à des données plus précises, à une échelle plus fine, ainsi qu'à davantage d'informations sur les bâtiments, pourrait nous permettre d'affiner nos prédictions.

Face à cette réalité environnementale, les implications économiques et sociales sont manifestes. Les coûts liés à la protection et à la prévention des risques climatiques sont amenés à s'accroître. Pour les assureurs, cela pourrait se traduire par une augmentation des primes d'assurance. Si le régime CatNat opère actuellement sur la base de la solidarité nationale, son modèle pourrait être mis à l'épreuve. Les habitants des zones non argileuses pourraient voir un surcoût sur leurs assurances, d'autant plus que ces régions sont déjà fortement impactées par la sécheresse. Par ailleurs, une hausse forfaitaire généralisée de ces primes, voire une tarification différenciée pour les régions à risques comme les zones argileuses, pourrait s'avérer nécessaire pour couvrir les coûts croissants des sinistres.

Cette étude comporte néanmoins certaines limites. En effet, certaines hypothèses ont été formulées, notamment concernant le nombre de constructions de bâtiments. Ces prévisions pourraient être surestimées, car elles ne tiennent pas suffisamment compte de la volonté de l'État de préser-

ver les espaces ruraux, comme le prévoient les législations en vigueur, qui favorisent la rénovation plutôt que la construction de nouveaux bâtiments. De plus, les données dont nous disposons sont à la maille communale, ce qui limite la précision de l'analyse. Ce manque de données détaillées par adresse ne permet pas une représentation exacte du risque futur.

Cependant, des solutions préventives, comme la mise en œuvre effective de la loi ELAN, peuvent offrir un certain répit. Toutefois, une adaptation constante des critères d'indemnisation est requise, en tenant compte de la détérioration rapide du climat. La conjugaison de ces défis exige de repenser notre rapport au climat, à la terre et aux mécanismes de solidarité et d'assurance qui nous protègent.

Chapitre 8

Annexe

8.1 Annexe 1 :La correction statistique des simulations

De manière générale, les projections climatiques ne peuvent être utilisées directement pour les études d'impact à l'échelle locale car elles présentent des biais par rapport aux observations. Afin de pouvoir comparer les résultats des projections climatiques avec le climat actuel ou passé et de calculer divers indicateurs climatiques basés sur des seuils absolus, il est essentiel d'utiliser une méthode de correction statistique des simulations climatiques. Cette correction permet d'ajuster les données simulées pour les rendre cohérentes avec les observations réelles, en prenant en compte les différences entre les deux. Grâce à cette correction, il devient possible de réaliser des analyses plus précises et fiables des impacts du changement climatique.

Il est donc important d'appliquer une méthode de correction. Dans le cas de nos simulations les modèles regroupés dans l'ensemble DRIAS-2020 utilisent principalement la méthode ADMONT, qui s'inspire largement de la méthode quantile-quantile. Cette approche permet de tenir compte des différentes variations météorologiques en analysant les centiles pour chaque saison et chaque régime de temps, plutôt que sur l'ensemble de la période.

La méthode quantile-quantile est une approche utilisée pour ajuster les simulations climatiques afin qu'elles correspondent davantage aux observations réelles. Elle repose sur la création d'une fonction de correction qui aligne la distribution statistique des données simulées sur celle des valeurs observées, en prenant en compte chaque variable, saison et période correspondante. Cette méthode non paramétrique considère que le modèle est capable de prédire la répartition des variables climatiques, mais pas les valeurs exactes de chaque quantile.

Pour chaque variable météorologique, des fonctions de densité de probabilités sont calculées à partir des séries temporelles de la simulation et des séries observées provenant d'une base de données de climat de référence telle que SAFRAN. La fonction de transfert consiste à associer chaque centile du modèle à un centile observé. Lorsque l'on souhaite projeter dans le futur, on détermine à quels centiles correspondent les valeurs du jour de la projection, on utilise les valeurs correspondantes dans le climat actuel, puis on effectue une interpolation linéaire entre les centiles. Pour les valeurs au-delà du dernier centile estimé ou en deçà du premier centile, on applique une correction constante basée sur le centile extrême correspondant. Dans certains cas, tels que les précipitations ou l'humidité relative, aucune extrapolation n'est réalisée, car il existe une limite absolue de la distribution (distribution bornée).

Grâce à cette méthode, il est possible de corriger les principales lacunes des modèles climatiques.

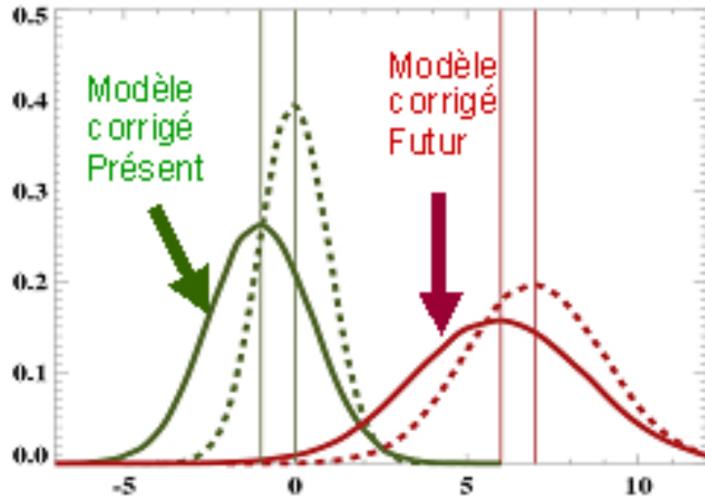


FIGURE 8.1 – Application de la méthode quantile-quantile aux données

tiques, notamment en ce qui concerne l'intensité des phénomènes extrêmes :

Pour mettre en œuvre cette méthode sur nos données, il est nécessaire de disposer de données de référence historique sur la période 1976-2005 considérée. On utilise donc la base de données Safran, elle est produite par Météo-France et est constituée de données horaires sur la France métropolitaine avec une résolution de 8 kilomètres. Cette résolution est comparable à la maille de 12 kilomètres utilisée dans les données Euro-CORDEX. On associe à chaque point de grille des modèles RCM, le point de grille de la base Safran le plus proche.

La correction s'effectue en quatre étapes :

1. Les variables météorologiques quotidiennes simulées sur la période sont interpolées sur la grille SAFRAN. L'interpolation consiste à prendre pour chaque point de la grille SAFRAN la valeur du quantile simulé au point le plus proche de la grille du modèle du DRIAS.
2. On calcule les quantiles, le plus souvent au pas de 1% mais avec une résolution de 0,5% ou même parfois 0,1% pour les queues de distribution, des variables météorologiques quotidiennes de la simulation considérée et de SAFRAN, en prenant en compte une période commune.
3. En chaque point i de la grille SAFRAN et pour chaque ordre k de quantile, on calcule alors le coefficient de correction $\text{Corr}(k)$ donné par la formule suivante :

$$\text{Corr}(k) = \begin{cases} \text{Corr}_k(i) = 0 \text{ si } Q_k(i) = 0 \\ \text{Corr}_k(i) = \frac{Q_k^{\text{Safran}}}{Q_k(i)} \text{ sinon.} \end{cases} \quad (8.1)$$

4. Pour chaque jour j de la période et en chaque point i de la grille SAFRAN, on cherche l'ordre k du quantile modèle directement inférieur à la valeur de la variable quotidienne du jour j au point i que l'on note $(\text{RR}(j,i))$. Ce qui se formalise par :

Si $RR(j, i) \geq Q_0,999(i)$ alors k=0,999

Sinon k est tel que : $Q_k(i) \leq RR(j, i) \leq Q_{k+1}(i)$

Ainsi les données corrigées sont obtenues en calculant :

$$RR_{corr}(j, i) = rr(j, i) * Corr_k(i) \quad (8.2)$$

[7]

8.2 Annexe 2 : Calcul de l'ETP

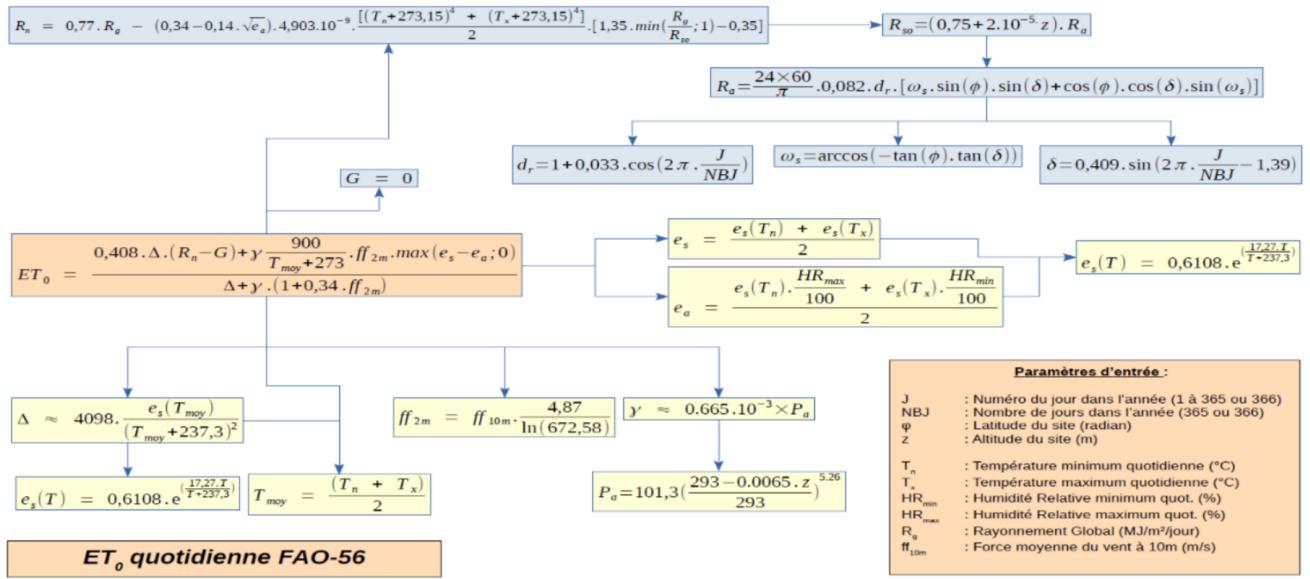


FIGURE 8.2 – Schéma du calcul de l'ETP

Table des figures

1	Comparaison des courbes ROC des modèles	4
2	Évolution mensuelle de la quantité de pluie par jour par période.	5
3	Évolution mensuelle de l'évapotranspiration moyenne par mois par période.	6
4	Critère SWI actuel	6
5	Critère SWI à l'année 2050	6
6	Critère SWI à l'année 2075	6
7	Critère SWI à l'année 2100	6
8	Fréquence de sinistres pour la période de référence	7
9	Fréquence de sinistres pour la période de 2050 et le scénario 4.5	7
10	Fréquence de sinistres pour la période de 2075 et le scénario 4.5	7
11	Fréquence de sinistres pour la période de 2100 et le scénario 4.5	7
12	Evolution des coûts des sinistres pour le département de Haute-Garonne	8
13	Comparison of ROC curves of models	11
14	Monthly change in the number of rain per day per period	12
15	Monthly evolution of average evapotraspiration per month per period	13
16	Current SWI criterion	13
17	SWI criterion to year 2050	13
18	SWI criterion to year 2075	13
19	SWI criterion to year 2100	13
20	Frequency of loss for the reference period	14
21	Disaster frequency for the period 2050 and scenario 4.5	14
22	Claim Frequency for 2075 and Scenario 4.5	14
23	Disaster frequency for the period 2100 and scenario 4.5	14
24	Evolution of claims costs for the department of Haute-Garonne	15
1.1	Évolution de la température moyenne par ans	20
1.2	10 années ayant enregistré les températures moyennes les plus élevées	20
1.3	Présentation rapide des 3 piliers de solvabilité II	22
1.4	Schéma de l'inondation des plaines	25
1.5	Schéma de l'inondation par remontée des nappes phréatiques	26
1.6	Schéma de l'inondation par submersion marine	26
1.7	Schéma de l'inondation par ruissellement de pluie	26
1.8	Schéma de la formation d'un ouragan	27
1.9	Schéma de l'impact de la sécheresse sur une maison	28
1.10	Étape de la reconnaissance CatNat	30
1.11	Projection des émissions liées aux énergies fossiles selon les profils d'évolution des GES	33

1.12 Évolution des émissions de Gaz à effet de serre en fonction des différents scénarios du GIEC	33
1.13 Évolution de la température mondiale en fonction des différents scénarios du GIEC	34
1.14 Relation entre les scénarios SSP et RCP	34
2.1 Répartition du score d'âge des bâtiments sur la France	38
2.2 Proportion de date de construction des bâtiments dans les communes	39
2.3 Type d'argiles moyen en France	40
2.4 Répartition du score de Hauteur des bâtiments sur la France	41
2.5 Liste des projections climatiques composant l'ensemble DRIAS-2020	42
2.6 Liste des couples GCM/RCM sélectionnés	43
2.7 Schéma du phénomène d'évapotranspiration	48
2.8 Schéma du fonctionnement du régime CatNat et rôle de la CCR	50
3.1 Matrice de corrélation de Pearson des variables	56
3.2 Summary du modèle	58
3.3 Matrice de confusion pour le modèle GLM	59
3.4 Schéma explicatif de la courbe ROC	61
3.5 Courbe ROC du modèle linéaire généralisé	62
3.6 Schéma illustrant l'algorithme du Random Forest	65
3.7 Choix du nombre d'arbres de l'algorithme du Random Forest	66
3.8 Matrice de confusion pour le modèle Random forest	66
3.9 Schéma illustrant l'algorithme du Random Forest	67
3.10 Schéma illustrant l'algorithme du Gradient Boosting	69
3.11 Choix du nombre d'arbres pour l'algorithme du GBM	69
3.12 Matrice de confusion pour le modèle Gradient boosting	70
3.13 Courbe ROC du modèle GBM retenu	71
3.14 Comparaison des courbes ROC des modèles	72
3.15 Récapitulatif des indicateurs de qualité des modèles	72
4.1 PDP pour la variable Freq70	78
4.2 Résultats de l'analyse de Morris	79
4.3 Comparaison de μ^* par rapport à sigma	80
4.4 Analyse des interactions entre les types d'argile	81
4.5 Analyse des interactions entre la variable Score_norm et les types d'argile	82
4.6 Analyse des interactions entre la variable Freq70 et les types d'argile	82
4.7 Analyse des interactions entre la variable Freq90 et les types d'argile	83
4.8 Analyse des interactions entre la variable Freq90 et la variable Freq70	83
5.1 Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2050	87
5.2 Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2075	88
5.3 Évolution mensuelle de NORTPSSWI par scénario GIEC pour la période 2100	89
5.4 Critère SWI actuel	90
5.5 Critère SWI à l'année 2050	90
5.6 Critère SWI à l'année 2075	90
5.7 Critère SWI à l'année 2100	90
5.8 Évolution mensuelle du nombre de pluie par jour par période	91
5.9 Évolution mensuelle de l'évapotranspiration moyenne par mois par période	92
5.10 Fréquence de sinistres pour la période de référence	93

5.11 Fréquence de sinistres pour la période de 2024-2050 et le scénario 2.6	93
5.12 Fréquence de sinistres pour la période de 2050-2075 et le scénario 2.6	93
5.13 Fréquence de sinistre pour la période de 2075-2100 et le scénario 2.6	93
5.14 Fréquence de sinistres pour la période de référence	94
5.15 Fréquence de sinistres pour la période de 2024-2050 et le scénario 4.5	94
5.16 Fréquence de sinistres pour la période de 2050-2075 et le scénario 4.5	94
5.17 Fréquence de sinistres pour la période de 2075-2100 et le scénario 4.5	94
5.18 Fréquence de sinistres pour la période de référence	95
5.19 Fréquence de sinistres pour la période de 2024-2050 et le scénario 8.5	95
5.20 Fréquence de sinistres pour la période de 2050-2075 et le scénario 8.5	95
5.21 Fréquence de sinistres pour la période de 2075-2100 et le scénario 8.5	95
6.1 Évolution des coûts des sinistres pour le département de Haute-Garonne	99
6.2 Fréquence de sinistres pour la période de référence	100
6.3 Fréquence de sinistres pour la période de 2050 et le scénario 8.5	100
6.4 Fréquence de sinistres pour la période de 2075 et le scénario 8.5	100
6.5 Fréquence de sinistres pour la période de 2100 et le scénario 8.5	100
8.1 Application de la méthode quantile-quantile aux données	107
8.2 Schéma du calcul de l'ETP	109

Bibliographie

- [1] ACPR. *Une première évaluation des risques financiers dus au changement climatique - les principaux résultats de l'exercice pilote climatique 2020.* 2021.
- [2] CCR. *Modélisation de l'impact du changement climatique sur les dommages assurés dans le cadre du régime Catastrophes Naturelles.* 2015.
- [3] CNULCD. *La sécheresse en chiffres, 2022 - restaurer afin d'être prêts et résilients.* 2022.
- [4] Cour des COMPTES. *Sols argileux et catastrophes naturelles.* 2022.
- [5] COVEA. *Changement climatique et assurance : quelles conséquences sur la sinistralité à horizon 2050.* 2022.
- [6] Documentation Météo France. 2023. URL : https://donneespubliques.meteofrance.fr/client/document/doc_swi_catnat_268.pdf.
- [7] DRIAS Eau. 2023. URL : <https://www.drias-eau.fr/accompagnement/sections/383>.
- [8] DRIAS Prod. 2023. URL : <https://drias-prod.meteo.fr/okapi/accueil/okapiWebDrias/index.jsp>.
- [9] FFA. *Changement climatique et assurance. Rapport d'étude horizon 2040.* 2015.
- [10] GIEC. *Changements climatiques 2014 : rapport de synthèse. Contribution des Groupes de Travail I, II et III au Cinquième Rapport D'évaluation du Groupe d'Experts Intergouvernemental sur l'Evolution du Climat.* 2014, p. 60-79.
- [11] Géorisques. 2023. URL : <https://www.georisques.gouv.fr>.
- [12] B. IOOSS et al. “Numerical studies of the metamodel fitting and validation processes”. In : *International Journal of Advances in Systems and Measurements* 3 (2010), p. 11-21.
- [13] Météo France. 2023. URL : <https://meteofrance.com>.
- [14] Observatoire des Territoires. 2023. URL : <https://www.observatoire-des-territoires.gouv.fr/nombre-de-logements>.
- [15] Ordonnance n° 2023-78. 2023. URL : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000047106316>.
- [16] “Revue sur l'analyse de sensibilité globale de modèles numériques”. In : *Journal de la Société Française de Statistique* (2010).
- [17] *The Elements of Statistical Learning* par Trevor Hastie, Robert Tibshirani, et Jerome Friedman. 2009.