

Mémoire présenté devant l'Université de Paris-Dauphine  
pour l'obtention du Certificat d'Actuaire de Paris-Dauphine  
et l'admission à l'Institut des Actuaires

le 27/06/2024

Par : Guillaume BERLEMONT

Titre : Réalisation d'un ORSA climatique via des données Open Source

Confidentialité :  Non  Oui (Durée :  1 an  2 ans)

---

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

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

*Membres présents du Jury du Certificat  
d'Actuaire de Paris-Dauphine :*

*Entreprise :*

Nom : KPMG

Signature :   
KPMG ADVISORY  
Siège social  
2 avenue Gambetta - Tour EQHO  
92086 Paris La Défense Cedex  
903 526 165 R.C.S. Nanterre

*Directeur de Mémoire en entreprise :*

Nom : Damien FABRE RUDELLE

Signature : 

---

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

*Secrétariat :*

*Bibliothèque :*

*Signature du responsable entreprise*

  
KPMG ADVISORY  
Siège social  
2 avenue Gambetta - Tour EQHO  
92086 Paris La Défense Cedex  
903 526 165 R.C.S. Nanterre

*Signature du candidat*





## Résumé

---

Le changement climatique est aujourd'hui l'un des enjeux majeurs auxquels les assureurs font face. Cela est dû à l'augmentation rapide du nombre et de l'intensité des catastrophes naturelles, que ce soit des inondations, des sécheresses ou des tempêtes. Ainsi, le coût décennal des catastrophes naturelles dans le monde a été multiplié par 7 entre 1970 et 2010, ce qui est considérable.

Dans ce contexte, le régulateur assurantiel français, l'ACPR, a lancé en 2020 puis en 2023 des exercices pilotes visant à sensibiliser et à permettre la mise en place de processus de gestion des risques climatiques chez les assureurs. De plus, l'ORSA est vu aujourd'hui par les organismes de régulation comme le meilleur élément pour estimer lesdits risques.

Nous nous proposons dans ce mémoire de faire une étude de la sinistralité climatique d'un assureur en suivant les recommandations de l'exercice pilote. Pour cela, nous rassemblerons d'abord une base de données Open Source afin de modéliser la survenance d'arrêts CatNat et la projeter aux différents horizons (2030, 2040 et 2050), selon deux scénarios (RCP 4.5 et 8.5) et selon deux périls (sécheresse et inondation). Nous utiliserons pour cela un modèle *XgBoost* avec *Cross-Validation*.

Ensuite, nous créerons un modèle nous permettant d'évaluer l'évolution moyenne de la sinistralité selon l'horizon, le péril et le scénario. Cette évolution sera alors utilisée pour impacter à la hausse le SCR assurantiel et étudier l'impact bilan du changement climatique dans un cadre ORSA.

---

*Mots-clés* : Climat, ORSA, Sécheresse, Inondation.

## Abstract

---

Climate change is nowadays considered as one of the major challenges that the insurer are facing. This is due to the rapid increase in the number and intensity of natural disasters, like floods, droughts or storms. Between 1970 and 2010, the ten-year cost of natural disasters worldwide was multiplied by a factor of 7.

In this context, France's insurance regulator, the ACPR, has launched two pilot exercises in 2020 and 2023, aimed at raising awareness and enabling insurers to implement climate risk processes. Furthermore, the ORSA is currently seen as the best way of estimating these risks.

In this dissertation, we want to study the climatic claims experience of an insurer, following the recommendations of the pilot exercise. To do this, we will first gather an Open Source database to model the occurrence of CatNat events and project them to different horizons (2030, 2040 and 2050), according to two scenarios (RCP 4.5 and 8.5) and two perils (drought and flood). We will use an XgBoost model with Cross-Validation.

Next, we'll create a model enabling us to evaluate the average evolution of the claims experience according to horizon, peril and scenario. This evolution will then be used to increase the insurance SCR and study the impact of climate change on the insurer balance sheet.

---

*Keywords* : Climate; ORSA; Flood; Drought.

# Note de synthèse

## Introduction

Le changement climatique est aujourd'hui l'un des principaux enjeux auxquels l'Homme fait face. Le climat se réchauffe, et les événements extrêmes sont de plus en plus fréquents. Le risque physique, c'est à dire le risque provoqué par les impacts directs du changement climatique sur les personnes et les biens, voit son importance augmenter pour les assureurs. Parmi ces risques, deux se distinguent de par leurs spécificités et leurs impacts : la sécheresse et l'inondation.

La sécheresse se définit comme un manque d'eau dans une zone donnée. Dans un cadre assurantiel, la sécheresse a notamment comme impact d'endommager les bâtiments par un phénomène de retrait-gonflement des sols, c'est à dire qu'un sol de nature argileuse se rétractera lors d'une sécheresse avant de gonfler si humidifié. Ce phénomène provoquera alors un mouvement de terrain pouvant endommager les bâtiments. L'inondation se définit quant à elle comme la submersion temporaire par l'eau d'une zone habituellement hors d'eau, qu'elle qu'en soit l'origine.

Ces deux risques ont pour spécificité d'être inclus dans un régime unique à la France, le régime CatNat. Dans ce régime, le maire de la commune touchée doit d'abord effectuer une demande de reconnaissance auprès du préfet. Après examen par une commission interministérielle, des arrêtés sont émis et permettent une indemnisation des assurés concernés après expertise. Pour émettre un arrêté, la commission se basera sur des critères météorologiques et géologiques. Dans ce cadre, les assureurs n'ont la main ni sur la décision, ni sur la prime.

En plus de cela, ces périls seront l'objets de fortes évolutions dans les années à venir. Il existe divers scénarios afin de considérer l'évolution du climat, cependant les plus reconnus d'entre eux sont les RCP du GIEC, allant du 1.6 (changement climatique modéré) au 8.5 (changement climatique extrême). Ainsi, selon la CCR, la sinistralité moyenne causée par les inondations de crues augmenterait de 110 % à horizon 2050 selon le scénario 8.5, et de 190 % pour la sécheresse. Ces augmentations considérables sont aujourd'hui un sujet d'inquiétude pour l'ACPR, organisme de contrôle des assurances en France.

Dans ce contexte, ce dernier a lancé courant 2020 un exercice pilote de grande ampleur, regroupant les principaux acteurs du monde de l'assurance et de la banque. Cet exercice, qui a utilisé le scénario 8.5 et des projections à horizon 2050, a confirmé l'impact du changement climatique sur la sinistralité et la nécessité pour les assureurs de se prévenir contre le risque climatique. Dans ce cadre, l'ORSA, élément du pilier 2 du régime Solvabilité 2, a été identifié comme le principal élément d'évaluation des risques. Suite aux résultats de ce premier exercice, l'ACPR a lancé en juillet 2023 (ACPR, [2023b](#)) un nouvel exercice pilote, qui se base cette fois sur le scénario RCP 4.5, plus modéré que le 8.5.

## Problématiques et mode opératoire

Nous cherchons avec cette étude à réaliser un ORSA climatique pour un assureur client de KPMG en prenant les hypothèses suivantes :

- pour les périls, nous nous intéressons à l'inondation ainsi qu'à la sécheresse. Ce choix s'explique par leur importance en France métropolitaine ainsi que par leur appartenance au régime CatNat ;

- pour la période de référence, nous nous penchons sur la période 2018-2022. Cette période de 5 ans nous permet de nous concentrer sur la situation actuelle, notamment en incluant l'année 2022, qui fut très impactante en terme de sécheresse ;
- pour les scénarios de projections, nous prendrons les scénarios RCP 4.5 et 8.5, car ils permettent de comparer les impacts et ont été utilisés lors des exercices pilotes ;
- pour l'horizon de projection des risques, nous prendrons les horizons 2030, 2040 et 2050. Cela nous permettra d'observer l'évolution du climat à intervalles réguliers.

Le plan que nous suivrons afin de réaliser cet objectif est le suivant :

- tout d'abord, nous modéliserons la survenance d'arrêtés CatNat, éléments nécessaires à l'indemnisation dans le cadre du régime. Une fois cela fait, nous voulons pouvoir modéliser cette survenance aux horizons évoqués et selon nos scénarios d'intérêt ;
- ensuite, nous utiliserons cet élément afin de projeter la sinistralité de notre assureur aux différents horizons. Faire une telle projection nous demandera également de nous interroger sur l'inflation du coût des sinistres ;
- enfin, nous utiliserons ces nouveaux éléments afin d'appliquer des chocs à divers éléments comptables, et en particulier les SCR. Faire cela nous permettra alors d'estimer un impact du changement climatique sur le bilan de notre assureur.

Cependant, ce processus soulève pour nous plusieurs problématiques. D'abord, le fait de modéliser une survenance d'arrêtés nécessite de récupérer des données relatives aux critères de décisions. Ces données sont par exemple l'exposition aux argiles d'une commune ou les précipitations totales. Nous devons donc chercher des données en *Open Source* afin de permettre une modélisation des arrêtés pour la période de référence mais également aux horizons d'intérêts. Nous devons ensuite sélectionner notre modèle, ainsi que les différents éléments à utiliser pour projeter la sinistralité.

## Données

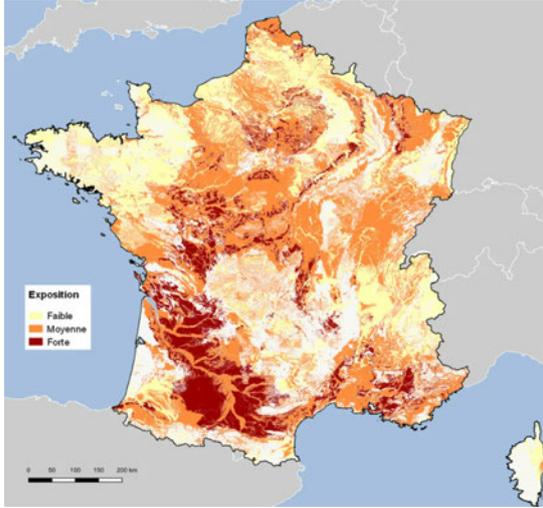
Le début du processus requiert donc de récupérer des données pouvant modéliser la survenance d'arrêtés CatNat sécheresse et inondation. Ces données peuvent être météorologiques, hydrologiques ou géologiques. Concernant la structure de notre base, nous prendrons nos éléments par commune puisqu'il s'agit de la maille de décision, et par année. Nous avons donc un total de 35 798 communes  $\times$  5 années = 178 990 lignes dans notre base.

Concernant les données météorologiques, nous allons nous pencher sur 4 éléments pouvant s'avérer utiles : la température, les précipitations, l'humidité et enfin la vitesse du vent. Pour cela, nous utilisons la base ERA5, qui est une base composée de données météorologiques. Le déclenchement d'arrêtés CatNat étant lié à des événements extrêmes, nous nous concentrons aussi bien sur des valeurs moyennes que des valeurs maximales ou minimales. Nous prenons ces valeurs de façon trimestrielle, car les arrêtés CatNat sont pris par trimestres.

Concernant les données géologiques, nous avons identifié deux éléments d'importance : le SWI (Soil Wetness Index), qui indique le niveau d'humidité des sols jusqu'à 2 mètres de profondeur, et le niveau d'exposition aux argiles du sol. Pour le SWI, nous récupérerons les données sur le site de météo France, et pour l'exposition aux argiles, sur le site du BRGM (2019) (Bureau de Recherches Géologiques et Minières). Ces éléments nous serviront à modéliser le risque sécheresse, même si le SWI peut également s'avérer important pour la modélisation d'inondations.

Concernant cette dernière, il peut être utile d'intégrer à notre base des données hydrologiques comme la présence dans une commune d'un fleuve. Pour cela nous utilisons le site du SANDRE.

Nous présentons ci-dessous la carte du BRGM représentant les niveaux d'exposition aux argiles en France ainsi que la carte du SANDRE représentant les principaux cours d'eau (figure 1).



(a) Carte du BRGM représentant les niveaux d'exposition aux argiles en France



(b) Représentation des cours d'eau de plus de 50 mètres de large en France

FIGURE 1 : Présentation de cartes Open Source

Nous avons donc les variables nécessaires à la modélisation, il nous manque encore la variable d'intérêt, à savoir la présence d'un arrêté sécheresse ou inondation. Ces deux variables prennent la forme de variables binaires : 0 en cas d'absence d'un arrêté pour la commune, 1 sinon.

## Modélisation de la survenance d'un arrêté CatNat

Avec cela, nous pouvons donc réaliser la première étape de notre processus, à savoir la modélisation des arrêtés CatNat sécheresse et inondation pour notre période de référence (2018-2022).

Nous allons nous concentrer sur trois types de modèles : le GLM (Modèle Linéaire Généralisé) avec pénalisation Lasso, la forêt aléatoire et enfin le XgBoost. Afin d'estimer la qualité de la modélisation, nous nous intéresserons à la matrice de confusion obtenue par le modèle. En effet, notre variable d'intérêt étant binaire, la matrice de confusion permet d'estimer avec précision la façon dont notre modèle a pu prédire les arrêtés.

A partir de la matrice de confusion, nous pouvons calculer trois métriques qui nous seront utiles pour évaluer simplement la qualité des modèles :

- Le rappel. Cette métrique correspond au pourcentage de 1 correctement prédits par notre modèle.

$$\text{Rappel} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux négatifs}}$$

- La précision. Cette métrique permet quant à elle de connaître le nombre de prédictions positives correctement effectuées. Sa formule est la suivante

$$\text{Précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux positifs}}$$

- Le F1 score. Cette métrique est la moyenne harmonique du rappel et de la précision, c'est à dire que sa formule est

$$\text{F1 score} = 2 \times \frac{\text{Rappel} \times \text{Précision}}{\text{Rappel} + \text{Précision}}$$

Toutes ces métriques sont comprises entre 0 et 1, et plus elles sont élevées, meilleure sera la prédiction. Nous présentons dans les tables (table 1 et table 2) les résultats de modélisation obtenus.

	Rappel	Précision	F1-score
GLM pénalisé	57,93 %	38,38 %	46,17 %
Forêt aléatoire	56,78%	56,31%	56,55%
XgBoost	61,45 %	53,77 %	57,35 %

TABLE 1 : Comparaison des différents modèles de sécheresse

	Rappel	Précision	F1-score
GLM pénalisé	31.18%	24.78%	27.62%
Forêt aléatoire	49,67%	47,08%	48,34%
XgBoost	52,8 %	45,87 %	49,09 %

TABLE 2 : Comparaison des différents modèles d'inondation

Nous observons que le modèle XgBoost est le meilleur pour chaque péril. La qualité de prédiction des modèles s'explique par la difficulté de prévoir la survenance d'un arrêté CatNat à une maille commune, notamment car la survenance de ces derniers dépend aussi de facteurs politiques. Mais nous avons pu modéliser avec précision le pourcentage de survenance d'arrêté CatNat au niveau métropolitain. La base contient 5,048 % d'arrêtés inondations et 8,485 % d'arrêtés sécheresse. Le modèle prédit quant à lui 5,057 % d'arrêtés inondation et 8,47 % d'arrêtés sécheresse.

## Projection du risque aux horizons d'intérêts

A partir des modèles *Xgboost*, nous procédons alors à la projection des arrêtés CatNat. Le processus est le suivant : nous récupérons une base de données projetées contenant les données *Open Source* et appliquons sur cette dernière les modèles retenus. Ces données nous sont fournies par le DRIAS. En appliquant nos modèles sur la base présentée, nous obtenons alors une évolution du pourcentage de survenance d'arrêtés sécheresse et inondation en France métropolitaine (figure 2).

Sécheresse	RCP 4.5	RCP 8.5	Inondation	RCP 4.5	RCP 8.5
Référence (2018-2022)	8,48 %	8,48 %	Référence (2018-2022)	5,05 %	5,05 %
Horizon 2030	9,37 %	10,4 %	Horizon 2030	5,7 %	7,51 %
Horizon 2040	14,1 %	15,58 %	Horizon 2040	6,35 %	10,47 %
Horizon 2050	18,31 %	19,34 %	Horizon 2050	8,3 %	12,09 %

FIGURE 2 : Evolution de la probabilité d'arrêté inondation/sécheresse par commune en France métropolitaine selon le scénario et l'horizon

Nous constatons une nette évolution à la hausse de la probabilité, qui passe de 8,48 % pour la sécheresse sur la période de référence à 18,31 % pour le scénario 4.5 à horizon 2050. De même pour l'inondation où l'on passe de 5,05 à 8,3 %. Nous notons également l'impact de l'horizon comme du scénario, qui différencie de façon conséquente les valeurs. Avec cela nous avons donc terminé la première partie de notre processus. Interrogeons-nous désormais sur l'impact que cette évolution aura sur la sinistralité de l'assureur.

Nous avons pour objectif de projeter la sinistralité du portefeuille MRH de notre assureur aux

horizons et scénarios décrits précédemment. Pour cela, nous allons nous baser sur la formule

$$Sinistralité\ projetée = \sum_N S \times P \times Nb$$

avec :

- S = la sinistralité moyenne projetée ;
- P = la probabilité de survenance d'un arrêté ;
- N = le nombre de ville touchées par le péril (sécheresse/inondation) ;
- Nb = nombre de sinistres par ville entraînés par un arrêté CatNat.

Trois éléments sont ici projetables :

- La probabilité de survenance d'un arrêté, que nous venons d'effectuer ;
- Le coût moyen d'un sinistre, que nous projetons en nous basant sur un fichier d'inflation fourni par l'ACPR dans le cadre de l'exercice pilote (ACPR, 2023b) ;
- Le nombre de dossiers entraînés par un arrêté CatNat et le nombre de villes touchées par le péril. Cela peut être fait en utilisant un bilan dynamique, mais du fait des évolutions à venir pour notre assureur et par volonté de simplification et de cohérence avec l'exercice pilote, nous restons dans une hypothèse de bilan statique et ne projetons pas ces éléments.

La formule présentée nous donne alors l'évolution de sinistralité suivante, que nous présentons en pourcentage de la sinistralité de référence (figure 3).

Sécheresse	RCP 4.5	RCP 8.5	Inondation	RCP 4.5	RCP 8.5
Horizon 2030	103 %	115 %	Horizon 2030	84 %	110 %
Horizon 2040	117 %	159 %	Horizon 2040	112 %	185 %
Horizon 2050	250 %	269 %	Horizon 2050	180 %	258 %

(a) Evolution de la sinistralité annuelle du péril sécheresse

(b) Evolution de la sinistralité annuelle du péril inondation

FIGURE 3 : Evolution de la sinistralité annuelle des périls sécheresse et inondation

Nous observons pour la sécheresse une hausse de 269 % de la sinistralité d'ici à 2050 et une hausse de 258 % pour l'inondation. Ces chiffres considérables montrent de façon nette l'importance du changement climatique pour les assureurs.

## Réalisation de l'ORSA climatique : calcul de SCR et impact bilan

A partir de l'évolution de sinistralité calculée nous pouvons impacter les modules SCR sécheresse et inondation. Le calcul d'un SCR se fait en prenant le quantile à 99,5 % d'une distribution de sinistralité. Nous avons ici un problème : pour créer cette distribution, il faudrait re-simuler nos données climatiques afin d'obtenir une nouvelle évolution de la sinistralité. Nous établissons donc un proxy : notre modèle de projection nous renvoie des probabilités de survenance pour chaque ville de France. Nous faisons alors varier un seuil à partir duquel la commune est considérée comme ayant un arrêté. En faisant varier ce seuil entre 0 et 1 par pas de 0,001, nous obtenons une distribution du pourcentage de communes en arrêté CatNat. Nous obtenons la distribution suivante (figure 4).

Sécheresse	RCP 4.5	RCP 8.5	Inondation	RCP 4.5	RCP 8.5
Référence (2018-2022)	49,80 %	49,80 %	Référence (2018-2022)	43,53 %	43,53 %
Horizon 2030	61,16 %	60,96 %	Horizon 2030	45,31 %	52,85 %
Horizon 2040	68,17 %	70,10 %	Horizon 2040	46,01 %	54,20 %
Horizon 2050	75,20 %	77,01 %	Horizon 2050	53,96 %	63,68 %

FIGURE 4 : Evolution du quantile à 99,5 % de la probabilité de survenance d'un arrêté CatNat

Cela nous permet alors de calculer une évolution de la sinistralité et donc des sous-modules SCR sécheresse et inondations. Comme nous sommes en bilan statique, nous conservons les autres éléments de bilan identiques à ce qu'ils sont en 2023. Une des priorités de l'ORSA est l'évaluation de la solvabilité assurantielle : pour cela, un ratio est calculé comme suit

$$\text{Ratio de solvabilité} = \frac{\text{Fonds propres disponibles}}{\text{SCR}}$$

Nous regardons donc comment ce ratio est impacté par la hausse du SCR calculée. Ainsi, en combinant les chocs sur la sécheresse et l'inondation nous obtenons le résultat suivant (figure 5).

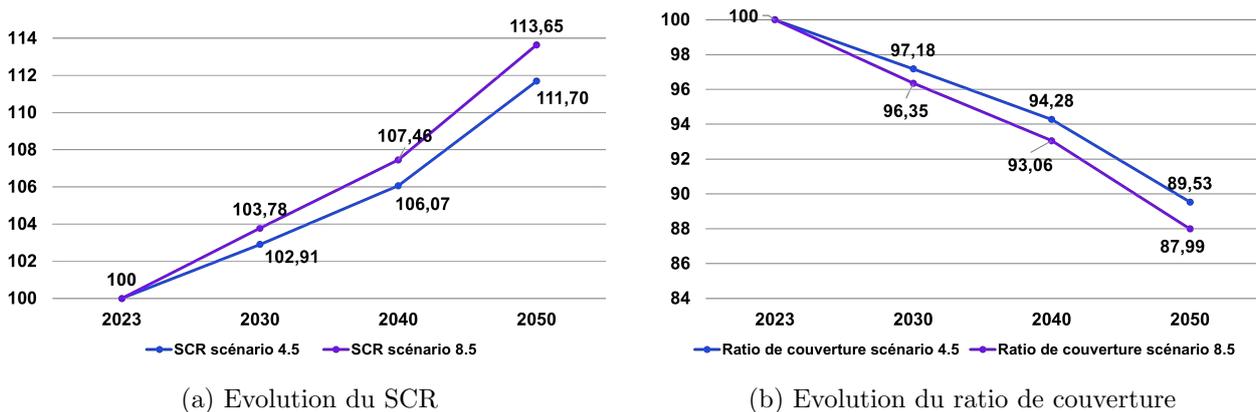


FIGURE 5 : Représentation de l'évolution du SCR ainsi que du ratio de solvabilité dues à l'inondation

Nous observons pour le scénario 8.5 une baisse de 12 % du ratio de couverture Fonds propres/SCR, due à une hausse 14 % du SCR du fait de la forte augmentation des modules sécheresse et inondation. Ce résultat confirme l'impact du changement climatique sur le bilan assurantiel.

## Conclusion

Afin d'effectuer notre étude, nous avons d'abord créé une base de données climatiques grâce à l'*Open Source*. Cette base a permis la modélisation puis la projection d'arrêtés CatNat. Nous avons ensuite établi un modèle de projection de charge faisant intervenir lesdites projections, nous donnant alors l'évolution de la sinistralité sécheresse/inondation de l'assureur aux différents horizons. Une seconde modélisation des arrêtés en prenant les valeurs extrêmes nous a alors permis de calculer l'évolution du SCR assurantiel et donc de mesurer l'impact du changement climatique sur le bilan.

Notre étude présente cependant plusieurs limites : aux niveaux des données assurantielles, nous avons assez peu d'historique, nos F1-scores restent malheureusement assez limités et le proxy utilisé pour le calcul de SCR entraîne nécessairement un biais.

# Synthesis note

## Introduction

Climate change is one of the major challenges facing mankind today. The climate is increasingly warming, and extreme events are becoming more frequent. Metropolitan France is affected by this phenomenon, and the physical risk, i.e. the risk caused by the direct impact of climate change on people and property, such as storms, is seeing a considerable increase in importance for insurers. Among these risks, two stand out for their specificities and impact on the territory: drought and flooding.

Drought is defined as a lack of water in a given area. From an insurer view, drought has the particular impact of damaging buildings through a phenomenon known as RGA, meaning that clay soil will shrink during a drought before swelling when moistened. This phenomenon then causes a ground movement that can damage buildings. Flooding is defined as the temporary submergence by water of an area normally out of water, whatever the cause.

Both of these risks are covered by a unique system in France, the CatNat system. Under this system, the mayor of the commune affected must first submit a request for recognition to the prefect. After examination by a ministerial commission, decrees are issued, allowing compensation to be paid to the policyholders concerned after expert approval. The commission's decisions are based on both meteorological and geological criteria. Within this framework, insurers have no control over either the decision or the premium, which is set at a surcharge of 6% for motor vehicles and 12% for other damage.

Furthermore, these perils are set to change a lot in the years ahead. There are various scenarios for considering climate change, but the most famous and recognized are the IPCC's RCP scenarios, which range from 1.6 (case of moderate climate change) to 8.5 (case of extreme climate change). For example, according to the CCR, the average loss due to flooding would increase by 110% by 2050 under scenario 8.5, and by 190% for drought. These considerable increases are now a huge concern for both insurers and the ACPR, France's insurance regulator.

In this context, the ACPR launched a large-scale pilot exercise during 2021, bringing together the main enterprises in the insurance and banking world. This exercise, which used scenario 8.5 and projections to 2050, confirmed the impact of climate change on claims experience and the need for insurers to prevent and assess their exposure to climate risk. In this context, ORSA, a component of Pillar 2 of Solvency 2, has been identified as the main tool to be used for prevention. Following the encouraging results of this first exercise, the ACPR launched a new pilot exercise in July 2023, this time based on the RCP 4.5 scenario, which is more moderate than the 8.5 scenario.

## Issues and procedure

In this study, we want to carry out a climatic ORSA for one of KPMG's insurance clients, based on the following assumptions:

- for the reference period of our study, we will focus on the period 2018-2022. This 5-year period enables us to match the claims history available to us, but also to focus on the current situation, by including the years 2018, 2021 and 2022, which were particularly impactful in terms of drought or flooding;
- for perils, we will focus on flooding and drought. This choice can be explained by their importance in France but also by their affiliation to the CatNat system, making them very specific;
- for the projection scenarios, we decided to focus on RCP 4.5 and 8.5, as they enable impacts to be compared and were the ones used in the pilot exercises ;
- for the risk projection horizon, we'll take 2030, 2040 and 2050. This will enable us to compare and observe the evolution of climate change.

Our plan to achieve this goal is as follows :

- first, we model the occurrence of CatNat decrees , which are necessary elements to the indemnisation of the policyholders in the CatNat system. Once done, we will project this occurrence to the horizons and scenarios mentioned before ;
- we will then use this element to project our insurer's claims experience over different horizons. Making such a projection will also require us to consider the inflation of a claims cost ;
- finally, we will use those new elements in order to apply shocks to the insurer's balance sheet, especially the SCR. Doing so will enable us to estimate the impact of climate change on our insurer's balance sheet.

But this process raises several issues for us. Firstly, modelling the occurrence of an order requires the retrieval of useful data correlated to the catnat decrees. These data are, for example, the clay exposure of a commune, or the total rainfall. We therefore need to retrieve Open Source data in order to model the orders. On top of this, we plan to project our modeling to different horizons, so we'll need data projected according to the desired horizon and scenario. Next, we'll need to select the model required for our modeling, as well as the various elements to be used to project claims experience, for example in terms of cost inflation.

## Data

The beginning of the process therefore requires going out to retrieve data that can model the occurrence of CatNat drought and flood decrees. This data can be meteorological, hydrological or geological. As for the structure of our database, we'll take our elements by commune, since this is the decision mesh, and by year. We therefore have a total of  $35,798 \text{ communes} \times 5 \text{ years} = 178,990$  lines in our database.

For meteorological data, we're going to focus on 4 elements that may prove useful : temperature, precipitation, humidity and wind speed. To do this, we use the ERA5 database, which is composed of reanalyzed meteorological data, and is sufficiently fine-grained to take into account climatic variations according to location. As CatNat decrees are triggered by extreme events, we focus on average, maximum and minimum values. We take these values on a quarterly basis, as CatNat decrees are issued on a quarterly basis, and elements such as temperature vary considerably depending on the time of year.

With regard to geological data, we have identified two key elements: the SWI (Soil Wetness Index), which indicates the level of soil moisture down to a depth of 2 meters, and the level of exposure to clay

in the soil. For SWI, we'll be retrieving data from the météo France website, and for clay exposure, from the BRGM (2019) (Bureau de Recherches Géologiques et Minières) website. These elements will be used to model the risk of drought, although the SWI may also prove important when modeling flooding.

For the latter, in addition to precipitation, it may be useful to integrate hydrological data into our database, such as the presence of a river or body of water in a given commune. For this, we will use the SANDRE website.

Below is the BRGM map showing clay exposure levels in mainland France, and the SANDRE map showing the main french watercourses (figure 6).

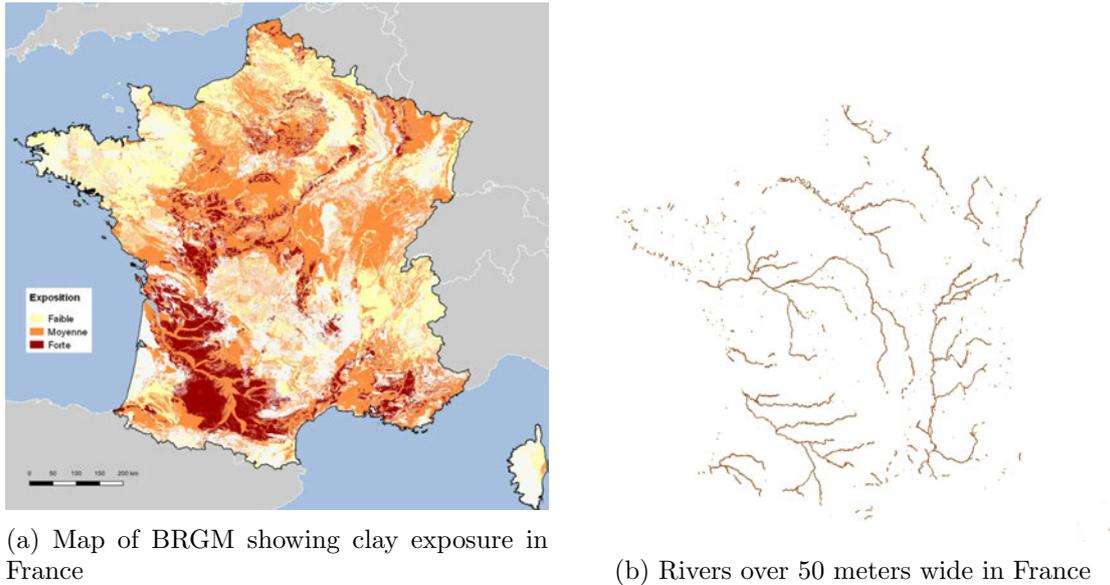


Figure 6: Open Source maps presentation

We therefore have the variables required for modeling, but we still need the variable of interest, i.e. the presence of a drought or flood decree. These two variables take the form of binary variables: 0 in the absence of a decree for the commune, 1 otherwise.

## Modeling the occurrence of a CatNat decree

With this, we can carry out the first stage of our process, which is the modeling of CatNat drought and flood decrees for our reference period (2018-2022).

We will focus on three types of model: the GLM (Generalized Linear Model) with Lasso penalization, the random forest and finally the XgBoost. In order to assess the quality of the modeling, we'll look at the confusion matrix obtained by the model. Indeed, as our variable of interest is binary, the confusion matrix enables us to accurately estimate how well our model was able to predict the stops.

From the confusion matrix, we can calculate three metrics that will be useful for evaluating the quality of the models:

- Recall. This metric corresponds to the percentage of 1s correctly predicted by our model. Its formula is as follows

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- Accuracy. This metric measures the number of positive predictions correctly made. Its formula is as follows

$$Accuracy = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

- F1 score. This metric is the harmonic mean of recall and precision, i.e. its formula is

$$F1\ score = 2 \times \frac{Recall \times Accuracy}{Recall + Accuracy}$$

All these metrics are between 0 and 1, and the higher they are, the better the prediction. In the tables below (table 3 and table 4), we present the modeling results obtained for each model.

	Recall	Accuracy	F1-score
GLM penalized	57,93 %	38,38 %	46,17 %
Random Forest	56,78%	56,31%	56,55%
XgBoost	61,45 %	53,77 %	57,35 %

Table 3: Comparison of different drought models

	Recall	Accuracy	F1-score
GLM penalized	31.18%	24.78%	27.62%
Random Forest	49,67%	47,08%	48,34%
XgBoost	52,8 %	45,87 %	49,09 %

Table 4: Comparison of different inondation models

We observe that in both drought and flood conditions, the XgBoost model proves to be the best. If the predictive quality of the models is rather poor, it's because it's difficult to predict the occurrence of a CatNat decree at a common mesh, notably because the occurrence of these depends on political factors in addition to natural elements. We have, however, been able to accurately model the percentage of CatNat decrees occurring in metropolitan France. The database contains 5,048 % of flood decrees. The model predicts 5.057 % of flood decrees. The database contains 8.485 % drought orders. The model predicts 8.47 % of drought orders.

## Risk projections

From our XgBoost models, we proceed to the projection of our CatNat decrees. The process is as follows: we retrieve a projected database containing SWI and meteorological data, and apply the selected models to it. These data are supplied to us by the DRIAS. By applying our models to the data, we obtain an evolution of the percentage of drought and flood decrees in France (figure 7).

We note a clear increase in probability, from 8.48% for drought in the reference period to 18.31% for scenario 4.5 by 2050. The same applies to flooding, which rises from 5.05% to 8.3%. We also note the impact of both horizon and scenario, which differentiates a lot the values. With this, we have completed the first part of our process. Let's now look at the impact this development will have on the insurer's claims.

Our aim is to project the claims of our insurer's multi-risk home insurance portfolio over the horizons and scenarios described above. To do this, we will use the following formula

$$Projected\ claims = \sum_N L \times P \times Nb$$

where :

Drought	RCP 4.5	RCP 8.5	Flood	RCP 4.5	RCP 8.5
Reference (2018-2022)	8,48 %	8,48 %	Reference (2018-2022)	5,05 %	5,05 %
Horizon 2030	9,37 %	10,4 %	Horizon 2030	5,7 %	7,51 %
Horizon 2040	14,1 %	15,58 %	Horizon 2040	6,35 %	10,47 %
Horizon 2050	18,31 %	19,34 %	Horizon 2050	8,3 %	12,09 %

Figure 7: Change in probability of a flood/drought order by commune in France, by scenario and timeframe

- L = projected average loss experience ;
- P = probability of decree occurrence ;
- N = number of towns affected by the risk (drought/flood) ;
- Nb = number of claims by town caused by a CatNat decree.

Three elements can be projected here:

- First the probability of occurrence of a decree, that we just done here ;
- Then the average cost of a claim, which we project on the basis of an inflation file provided by the ACPR as part of the pilot exercise (ACPR (2023b)) ;
- The number of cases triggered by a CatNat decree and the number of towns affected by peril. This can be done using a dynamic balance sheet. However, because of our insurer ambitions, but also for the goal of simplicity and consistency with the pilot exercise, we'll be sticking to a static balance hypothesis and not projecting these elements.

The formula that we showed give us the evolution of the insurer loss ratio , which we present as a percentage of the reference claims (figure 8).

Drought	RCP 4.5	RCP 8.5	Flood	RCP 4.5	RCP 8.5
Horizon 2030	103 %	115 %	Horizon 2030	84 %	110 %
Horizon 2040	117 %	159 %	Horizon 2040	112 %	185 %
Horizon 2050	250 %	269 %	Horizon 2050	180 %	258 %

(a) Evolution of the annual drought claims

(b) Evolution of the annual flood loss ratio

Figure 8: Annual loss experience for drought and flood risks

We observe that the drought loss experience has an increase of 269 % by 2050 and an increase by 258 % for flood . These numbers clearly demonstrate the importance of climate change for insurers.

## Climatic ORSA: SCR calculation and balance sheet impact

Based on the calculated loss development, we can increase the SCR drought and flood sub-modules.

SCR is calculated by taking the 99.5 % quantile of a claims distribution. Here we come up against a problem: to create this claims distribution, we would have to re-simulate our climatic data in order to obtain a new evolution in the probability of a CatNat decree, and therefore in the claims rate each time. So here we establish a proxy: our projection model returns probabilities of occurrence for each

city. For each model, we can set a threshold to associate 0 or 1 with each city. We then vary the threshold between 0 and 1 in steps of 0.001, giving us 1000 elements. This proxy allows us to obtain a distribution of the probability of CatNat events (figure 9).

Drought	RCP 4.5	RCP 8.5	Flood	RCP 4.5	RCP 8.5
Reference (2018-2022)	49,8 %	49,8 %	Reference (2018-2022)	43,53 %	43,53 %
Horizon 2030	61,16 %	60,96 %	Horizon 2030	45,31 %	52,85 %
Horizon 2040	68,17 %	70,10 %	Horizon 2040	46 %	54,2 %
Horizon 2050	75,2 %	77,01 %	Horizon 2050	53,96 %	63,68 %

Figure 9: Trend in the 99.5% quantile of the probability of a CatNat decree occurring

This allows us to calculate the evolution of the SCR drought and flood sub-modules. As we have a static balance sheet, we keep all other balance sheet items identical to what they will be in 2023. One of ORSA’s priorities is the assessment of insurance solvency : for this, a ratio is calculated as follows

$$Solvency\ ratio = \frac{Available\ equity}{SCR}$$

We therefore look at how this ratio is impacted by the calculated SCR increase. Thus, by combining the drought and flood shocks, we obtain the following result (figure 10).

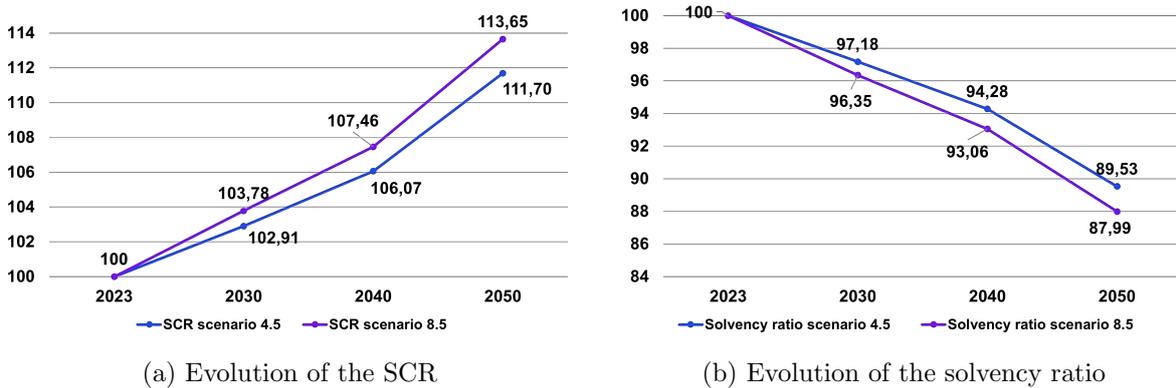


Figure 10: Representation of changes in SCR and solvency ratio due to flooding and drought

For the 8.5 scenario, we observe a decrease by 12 % of the solvency ratio Available equity/SCR, due to an increase by due to a 14 % increase in SCR due to the rise in drought and flood modules. This result confirms the impact of climate change on the insurance balance sheet.

## Conclusion

To carry out our climate study, we first created a climate database using Open Source. This database was used to model and project CatNat events. We then set up a loss projection model using these projections, giving us the insurer’s drought/flood loss experience over different time horizons. A second modelling using extreme values then enabled us to calculate the evolution of the insurance SCR in order to measure the impact of climate change on the balance sheet.

However, our study has a number of limitations: we have little historical data for our insurer loss experience, our F1-scores are unfortunately rather limited, and the proxy used to calculate SCR necessarily entails a bias.

# Remerciements

Je souhaite remercier tous ceux qui ont pu m'aider lors de ce mémoire. Tout d'abord mes tuteurs, Damien Fabre Rudelle et Hamadi Bahri côté entreprise, et Quentin Guibert côté académique. Tout au long de mon mémoire, ils m'ont donné de nombreux conseils et remarques qui m'ont beaucoup aidé, et je n'aurais pas pu faire ce mémoire sans leur aide.

Je voudrais également remercier toute l'équipe de KPMG, qui m'a toujours soutenu et s'est montrée très amicale et confiante envers moi.

Enfin, je voudrais remercier ma famille et mes amis, qui m'ont accompagné tout au long de la rédaction du mémoire.



# Avertissement

Pour des raison de confidentialité, les données utilisées lors de cette étude ont été modifiées. Une représentation en base 100 a également été utilisée dans ce même objectif.



# Table des matières

<b>Résumé</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Note de synthèse</b>	<b>5</b>
<b>Synthesis note</b>	<b>11</b>
<b>Remerciements</b>	<b>17</b>
<b>Avertissement</b>	<b>19</b>
<b>Table des matières</b>	<b>21</b>
<b>Introduction</b>	<b>23</b>
<b>1 Présentation des risques climatiques et du rôle de l'ORSA</b>	<b>25</b>
1.1 Présentation générale des risques climatiques physiques . . . . .	25
1.2 Les évolutions récentes et à venir du climat en France. . . . .	34
1.3 Présentation du régime Solvabilité 2 . . . . .	40
1.4 Modélisation climatique et ORSA climatique . . . . .	44
1.5 Problématiques du mémoire . . . . .	51
<b>2 Analyse des données et création d'une base climatique</b>	<b>53</b>
2.1 Présentation des données à disposition . . . . .	53
2.2 Enrichissement de la base de données à l'aide de données <i>Open Data</i> . . . . .	54
2.3 Arrêtés CatNat sécheresse et inondation . . . . .	63
<b>3 Modélisation des arrêtés CatNat</b>	<b>69</b>

3.1	Sélection des scénarios climatiques à employer . . . . .	69
3.2	Modélisation de la survenance d'arrêtés CatNat . . . . .	71
3.3	Modélisation de la sécheresse . . . . .	83
3.4	Modélisation de l'inondation . . . . .	91
3.5	Projections du risque aux horizons 2030, 2040 et 2050 . . . . .	98
3.6	La sinistralité en présence d'un arrêté CatNat . . . . .	100
<b>4</b>	<b>Etude de la sinistralité assurantielle</b>	<b>105</b>
4.1	Impact du changement climatique sur la charge moyenne . . . . .	105
4.2	Analyse de l'impact du changement climatique sur le bilan . . . . .	108
4.3	Limites et perspectives de l'étude . . . . .	120
	<b>Conclusion</b>	<b>123</b>
	<b>Bibliographie</b>	<b>125</b>
<b>A</b>	<b>Annexes</b>	<b>129</b>
A.1	Les modèles d'arrêtés CatNat . . . . .	129

# Introduction

Parmi les grands risques auxquels font face les assureurs, le risque climatique est aujourd'hui considéré comme l'un des plus importants mais également l'un des plus mal connus. Face aux catastrophes climatiques, la France a mis en place un régime atypique appelé régime CatNat basé sur la mutualisation des sinistres au niveau national et une réassurance garantie par l'Etat. Dans ce régime, l'Etat a un contrôle sur l'indemnisation car il peut émettre en cas de catastrophe naturelle un arrêté CatNat permettant l'indemnisation des assurés après expertise. Cependant ce système est aujourd'hui mis à mal par l'augmentation rapide de la sinistralité, rendant nécessaire pour les assureurs de s'intéresser plus en détail à leurs risques climatiques.

Face à cela, l'ORSA a été identifié par l'EIOPA comme l'outil à privilégier pour identifier et se prémunir contre lesdits risques. Ce processus fait partie de la norme Solvabilité 2.

Depuis 2016, l'Europe a introduit une nouvelle norme pour les assureurs et réassureurs appelée Solvabilité 2. Cette norme vise notamment à permettre à ces derniers une meilleure compréhension et anticipation de leurs risques. Parmi les éléments imposés par cette dernière, on retrouve l'ORSA (*Own Risk Self Assessment*). Ce processus évalue trois éléments : l'évolution du besoin global de solvabilité, le respect des exigences de capital réglementaire, et enfin l'écart entre les hypothèses de calcul de capital réglementaire de l'assureur et son véritable profil de risque. Il s'agit d'un outil de pilotage important pour ce dernier, qui doit lui permettre de se prévenir contre les risques menaçant sa solvabilité en adaptant sa stratégie.

Cependant, l'ORSA climatique s'avère différent des autres ORSA : en effet le risque climatique s'avère très volatile et est amené à évoluer de façon importante. Il est donc nécessaire de changer la temporalité : là où l'ORSA classique projette les risques sur un horizon allant de 3 à 5 ans, l'ORSA climatique devra être fait sur un horizon de court, moyen et long terme. Il devra de plus utiliser plusieurs scénarios climatiques afin de couvrir un large éventail de possibilités. De plus, toujours à cause de ces évolutions attendues, on ne peut pas déduire la sinistralité future à partir de la sinistralité passée.

Dans ce mémoire, nous allons donc procéder à la mise en place d'un processus ORSA pour un portefeuille de contrats MRH. Notre étude débutera par une explication du contexte dans lequel nous nous situons dans le chapitre 1. Nous présenterons ensuite les données de l'assureur et constituerons une base de données climatiques Open Source dans le chapitre 2. A partir de cette base, nous modéliserons la survenance d'arrêtés CatNat puisqu'ils sont nécessaires à l'indemnisation des assurés dans le chapitre 3. Enfin, nous verrons à partir des résultats de modélisation comment cela impacte la sinistralité assurantielle et réaliserons un ORSA climatique dans le chapitre 4.



# Chapitre 1

## Présentation des risques climatiques, de leurs impacts sur les assureurs et du rôle de l'ORSA

L'objectif de cette partie est de poser au mieux le contexte de notre étude, présenter les différentes problématiques et d'où elles découlent, ainsi que définir les différents termes employés. Nous présentons d'abord les différents risques climatiques, en montrant en quoi ces derniers s'avèrent spéciaux pour les assureurs. Nous introduisons également le régime Solvabilité 2 dont le processus ORSA fait partie. Nous nous penchons ensuite sur l'ORSA climatique, que ce soit la méthode de réalisation du processus ou ses spécificités. Enfin, nous voyons plus en détail les risques climatiques et comment ils impactent les assureurs aujourd'hui afin de bien comprendre l'importance de l'ORSA climatique.

### 1.1 Présentation générale des risques climatiques physiques

Présentons les différents risques climatiques. Les risques climatiques sont définis comme l'ensemble des vulnérabilités d'une entreprise aux événements climatiques tels que les tempêtes ou les sécheresses.

#### 1.1.1 Les deux types de risques climatiques

On distingue deux types de risques climatiques :

- Le risque physique, qui est l'impact direct du changement climatique sur les personnes et les biens (exemple : destruction d'une récolte suite à une sécheresse).
- Le risque de transition, qui est la modification du comportement des agents économiques due à la mise en place d'un modèle économique bas carbone (exemple : loi **PACTE**, qui oblige les entreprises à avoir des investissements plus responsables (ACPR, 2022)).

Pour un assureur les risques s'avèrent très différents selon qu'ils soient physique ou de transition. En effet, le risque physique va inclure : la montée des eaux, les inondations, la sécheresse, les incendies, les tempêtes ou encore l'augmentation de la température. Tous ces événements impactent la sinistralité à la hausse et avec un lien de corrélation souvent important : si une inondation détruit des récoltes, la perte de l'agriculteur pourra être liée à ladite inondation.

Les risques physique peuvent être divisés en deux catégories : les risques physiques aigus, conséquences d'événements exceptionnels comme les cyclones ou les incendies, et les risques physiques chroniques, qui sont les conséquences à long terme du changement climatique comme la hausse de la

température.

Le risque de transition va quant à lui regrouper :

- les risques politiques, par exemple dûs au fait de la taxation des produits polluants augmentant le prix de certains biens ;
- Les risques légaux ;
- Les risques technologiques ;
- Le risque de réputation ;
- Le risque de marché.

Il s'avère non seulement difficile à anticiper car cela demande de prévoir les changements sociétaux et de réglementation, mais son impact n'est pas toujours simple à déceler. Par exemple, si une entreprise polluante voit le cours de son action baisser, cela peut être issu de nombreux autres facteurs que son bilan carbone trop élevé.

Notre mémoire va quant à lui se concentrer sur les risques physiques et sur le territoire de la France métropolitaine. Nous nous concentrons donc essentiellement sur ces derniers à partir de maintenant.

### 1.1.2 Les évènements climatiques

#### Les inondations

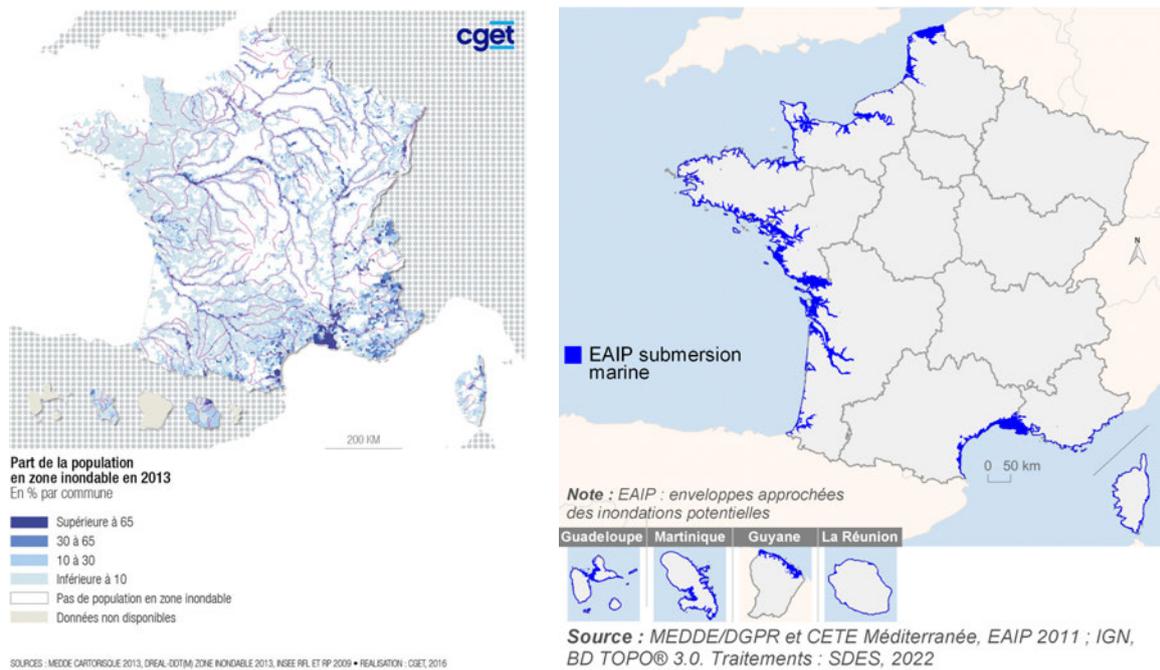
L'inondation désigne la submersion temporaire par l'eau d'une zone habituellement hors d'eau, qu'elle qu'en soit l'origine (définition par EAU FRANCE (2023)).

Les inondations sont un évènement assez courant en France : entre 1982 et 2021, la quasi-totalité des communes furent concernées par au moins une inondation (CCR, 2021). Il existe en tout cinq sortes d'inondations différentes :

1. Crues lentes : débordement de cours d'eau à la suite de précipitations modérées mais répétées.
2. Crues rapides et torrentielles : débordement de cours d'eau à la suite de précipitations intenses mais courtes, pouvant dépasser 100 à 200 millimètres en 24 heures.
3. Ruissellement : l'eau ne peut plus s'infiltrer dans le sol car ce dernier est saturé, ce qui se produit notamment lors de pluies torrentielles qui saturent le sol sans pouvoir s'infiltrer plus en profondeur du fait de leur intensité. A noter qu'on peut également trouver du ruissellement urbain, c'est à dire un ruissellement dû à l'imperméabilité des sols en ville. Si les précipitations sont trop importantes, le réseau d'évacuation peut être saturé, créant un courant qui emprunte les rues jusqu'à un autre réseau d'évacuation ou une rivière.
4. Submersion marine : élévation anormale du niveau de la mer. Ce type d'inondation peut être dû à deux phénomènes météorologiques différents :
  - Une hausse du niveau de la mer qui peut être due à de forts vents.
  - Une houle de forte intensité, la houle étant définie comme un mouvement d'ondulation de la mer. Cette dernière a pour particularité de ne pas être liée à l'action du vent.
5. Remontée de nappe : montée du niveau d'une nappe phréatique jusqu'au sol. Ces dernières étant alimentées par l'infiltration de la pluie, des précipitations particulièrement fortes peuvent provoquer une saturation puis une remontée de la nappe vers la surface.

6. Rupture d'ouvrage : ce type d'inondation est un peu spécial car il ne tire pas nécessairement son origine d'un phénomène météorologique mais de la destruction partielle ou totale d'un ouvrage retenant une importante quantité d'eau, comme un barrage.

Il est important de faire la différence entre ces différents types, car les zones concernées et les impacts sur ces dernières s'avère très différents. L'inondation par submersion marine en particulier s'avère logiquement localisée sur les côtes, mais également sur les fleuves plus en aval comme sur la Loire. Ainsi le gouvernement estime qu'aujourd'hui 1.5 millions de personnes résident dans des zones potentiellement inondables, et à 16.8 millions de personnes au débordement de cours d'eau, ce qui est considérable. Voici la carte représentant lesdites zones 1.1.



(a) Carte de l'exposition aux inondations en France. (Source : ANCT).

(b) Exposition aux submersions marines en France (Source : GOUVERNEMENT FRANÇAIS (2022)).

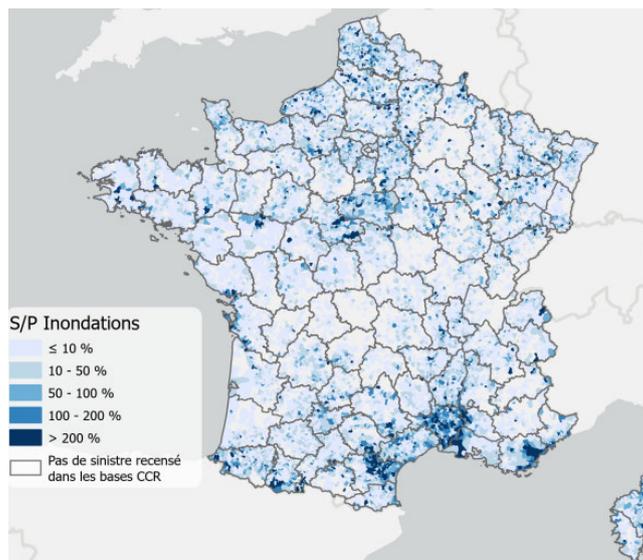
FIGURE 1.1 : Carte des expositions aux inondations et submersion marine en France.

On constate que de nombreuses grandes villes telles que Marseille ou Orléans présentent une forte exposition au risque inondation, ce qui apparaît comme étant logique vu qu'il n'est pas rare qu'une ville soit traversée par un cours d'eau.

Historiquement, l'inondation est un phénomène courant et de grande ampleur. Parmi les grandes inondations françaises, on peut citer la tempête Xynthia en 2010, qui provoqua des submersions marines ainsi que la rupture de plusieurs ouvrages aggravant l'inondation. Concernant le coût pour les assurances, la CCR fournit le S/P des événements inondations par département (figure 1.2).

On constate que la côte Sud, en particulier vers Montpellier, est très touchée par le phénomène, mais n'est pas le seul endroit, avec parfois des S/P aggravés de façon très localisée.

En résumé les inondations sont un phénomène très variable, pouvant impacter une simple commune comme des dizaines, pouvant s'avérer très brève comme durer plusieurs mois. Par conséquent le phénomène inondation recouvre une multitude de réalités.

FIGURE 1.2 : Carte du S/P inondations en France (*Source* : CCR (2021)).

### La sécheresse

La sécheresse est également un évènement climatique d'importance. Une sécheresse se définit comme un manque d'eau dans une zone donnée. Cependant, on peut désigner trois sous-groupes de sécheresse :

1. La sécheresse météorologique, qui se produit en cas de manque de précipitations sur une longue durée.
2. La sécheresse agricole, qui se produit lors d'un déficit hydrique des sols de surface, entre 1 et 2 mètres de profondeur, ce qui correspond à la zone où les racines des plantes prennent l'eau qui leur est nécessaire.
3. La sécheresse hydrologique, qui se produit quand les nappes phréatiques, cours d'eau et lacs présentent un niveau très bas.

Les dégâts causés par la sécheresse peuvent être de natures diverses. Le premier type que nous pouvons identifier est lié aux pertes d'exploitations. Cela est particulièrement vrai pour le secteur agricole, qui souffre de nombreuses pertes en cas de sécheresse.

Cependant, dans un cadre MRH, le type de sinistres sécheresse qui va nous intéresser est le RGA (Retrait-Gonflement des sols Argileux). Les sols argileux désignent les sols composés de particules dont le diamètre est inférieur à deux micromètres. Ces sols ont une structure en feuillets, ce qui fait que l'eau peut s'infiltrer à l'intérieur. Lorsque le sol est humide, le sol argileux va alors gonfler, puis se rétracter en cas d'assèchement. Cette différence peut provoquer un mouvement de terrain qui va alors endommager les biens immobiliers.

Ce phénomène est complexe du fait qu'il se produit sur des périodes de temps très longs : les fissures engendrées par le RGA peuvent apparaître que de nombreuses années après la sécheresse, le phénomène ayant entraîné une fragilisation de la structure qui ne sera pas immédiatement apparente. De plus, on ne peut pas directement associer de fortes températures aux dégâts du RGA, puisque c'est la différence entre un sol argileux sec et un sol argileux gorgé d'eau qui va produire un mouvement de

terrain. Ce mouvement se produira donc si la sécheresse est suivie ou précédée de précipitations bien plus importantes créant un différentiel au niveau du terrain.

Dans ces conditions, il est alors difficile de lier RGA et dommages aux habitations. De plus, il peut être difficile de savoir où prendre des mesures de prévention, c'est pourquoi il a été établi une carte de France (figure 1.3) explicitant l'exposition aux argiles des sols par le BRGM (Bureau de Recherche Géologiques et Minières). Elle comporte plusieurs niveaux d'exposition au RGA : « Faible », « Moyenne », « Forte » ou aucune exposition. Ces niveaux dépendent de deux facteurs, la susceptibilité et la sinistralité. La susceptibilité est issue de l'identification des formations argileuses sujettes au RGA. Ces formations furent ensuite réparties entre trois classes de susceptibilité : faible, moyenne et forte. Ledit classement dépendait de leur lithologie (proportion d'argile, épaisseur et continuité des roches), la minéralogie des argiles et leur comportement géotechnique.

La sinistralité est quant à elle issue d'une base recensant près de 180 000 sinistres sécheresse entre 1989 et 2017. A partir de cette base, il a été calculé une densité de sinistralité par formation argileuse, avec comme catégorie sinistralité faible, moyenne et forte. Au final, 55% des sinistres recensés l'ont été en zone d'exposition forte et 38% en zone d'exposition moyenne, ce qui montre bien l'importance de cette catégorisation.

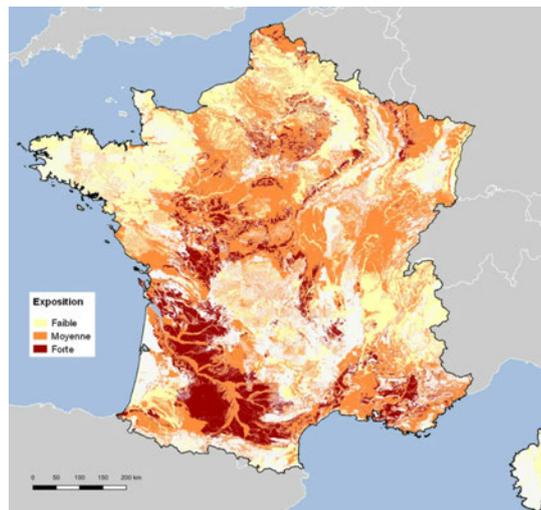


FIGURE 1.3 : Carte de l'exposition aux argiles en France. (Source : BRGM (2019)).

Cette carte permet de savoir quels endroits seront touchés par le RGA, et ainsi adapter les politiques de prévention. On remarque que les parties Sud-Ouest ainsi que le Sud-Est de la France sont très exposées aux argiles, là où le centre et l'Ouest présentent une plus faible exposition.

Aujourd'hui, le RGA reste un phénomène mal connu mais dont la sinistralité ne cesse d'augmenter : ainsi son coût moyen est de plus de 1 milliard d'euros annuels entre 2017 et 2020, contre 445 millions d'euros depuis 1982 (CCR, 2021).

D'une manière générale, le risque sécheresse constitue la principale menace climatique : en effet, sa sinistralité est en train d'augmenter rapidement pour atteindre des niveaux très élevés. Ainsi, la sécheresse de 2022 est la plus coûteuse que la France aie jamais connue, avec des dégâts estimés à 2.5 milliards d'euros (FRANCE ASSUREURS, 2022).

### Les autres évènements climatiques

Parmi les autres évènements, on trouve les séismes, qui sont cependant rares et de faible puissance en France métropolitaine. Ainsi le séisme du 16 juin 2023, de magnitude 5.3, est l'un des plus forts enregistrés en France métropolitaine et a entraîné la reconnaissance de 13 communes en arrêté CatNat, un chiffre peu élevé au regard des 34 945 communes que comporte la France. Historiquement, la principale région sujette aux séismes en France se trouve être les Alpes, région comportant peu de biens assurés (CCR, 2021).

Parmi les autres types d'évènements climatiques, nous pouvons également citer :

- La tempête, qui est définie comme une violente perturbation atmosphérique de grande échelle, accompagnée de vents violents et de fortes précipitations. La France a connu au cours de son histoire de nombreux épisodes de tempêtes, la plus connue étant celle de décembre 1999 appelée Lothar et Martin, qui a entraînée à elle seule des dégâts de 15 milliards d'euros en France métropolitaine, faisant d'elle l'évènement climatique le plus coûteux jamais enregistré en France. Les dégâts causés par les tempêtes sont de nature diverse, puisque tout bien peut potentiellement être détruit. Ainsi la tempête Lothar et Martin a entraîné la destruction de plus d'un millier de pylônes électriques, coupant le courant de 10 millions de Français, et plusieurs grands axes routiers dont des autoroutes furent endommagés.
- La grêle, qui est une précipitation composée de morceaux de glaces appelés grêlons, dont le diamètre est en général compris entre 5 et 50 millimètres. Ces grêlons frappent le sol avec une grande vitesse, comprise entre 30 et 150 km/h, causant d'importants dégâts aux véhicules ou équipements exposés en extérieur.
- La neige. Cet évènement peut causer des dégâts via une accumulation de neige qui va alors endommager des biens par son poids, comme par exemple les toitures des bâtiments. En fondant, la neige peut également entraîner des dégâts en venant mouiller des éléments sensibles à l'eau.

#### 1.1.3 Les scénarios climatiques

Les risques climatiques ont cela de spécial que le risque physique et le risques de transitions'avèrent très différents : dans un premier cas, on étudie les évolutions climatiques, alors que le second cas demande d'étudier les possibles évolutions sociales et de réglementations, ce qui peut être dur à prévoir.

Ces difficultés sont également renforcées par des horizons de projection parfois très lointains, pouvant aller jusqu'à 2100. Cela rend nécessaire pour les assureurs d'utiliser des scénarios de projection climatique afin d'estimer au mieux les risques. Pour cela, de nombreux assureurs vont se baser sur les scénarios *Shared Socioeconomic Pathway* proposés par le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) lors de ses rapports.

Le GIEC est un organisme intergouvernemental regroupant 195 pays et composé de scientifiques ainsi que de représentants des Etats . On y trouve entre autre des scientifiques spécialistes de la météorologie, de la politique ou encore de l'économie, qui ont pour rôle d'étudier le changement climatique et ses impacts. Ces derniers sont divisés en 3 groupes de travail :

1. le premier va étudier les aspects scientifiques du changement climatique ;
2. le second la vulnérabilité des systèmes socio-économiques et naturels au changement climatique ;
3. le troisième les possibilités d'atténuation du changement climatique et les limitations possibles pour l'émission de GES (Gaz à Effet de Serre).

Lors de chaque rapport, l'organisme propose différents scénarios d'évolution de la température mondiale prenant en compte les différents aspects du changement climatique, aussi bien météorologiques que socio-économiques.

Le 6ème rapport, paru entre 2021 et 2023, décrit 6 scénarios :

1. Le scénario SSP1-1.9 suppose une très forte baisse des émissions de gas à effet de serre dès 2025, puis zéro émission nette en 2050, ce qui se traduira par un réchauffement d'1.5°C d'ici à 2100. Ce scénario suppose de fortes émissions négatives (par exemple par les puits de carbone).
2. Le scénario SSP1-2.6 suppose également une forte réduction d'émissions de GES, mais avec un niveau d'émissions négatives moins important, se traduisant par un réchauffement de 2°C d'ici à 2100.
3. Le scénario SSP2-4.5 est un scénario intermédiaire, supposant dans un premier temps une augmentation des émissions de CO2 jusqu'en 2035, puis une stabilisation et enfin une baisse à partir de 2050, mais sans atteindre zéro émission nette d'ici 2100, se traduisant par un réchauffement de 3°C.
4. Le scénario SSP3-7.0 suppose des émissions de GES en hausse, multipliées par deux entre aujourd'hui et 2100, se traduisant par un réchauffement de 3.7°C d'ici 2100.
5. Le scénario SSP5-8.5 suppose des émissions de GES en hausse, multipliées par deux entre aujourd'hui et 2050 et un réchauffement de 4.4°C d'ici 2100. Il s'agit d'un scénario particulièrement pessimiste.

Ces scénarios, basés sur les études scientifiques du GIEC, constituent la base de la projection du risque physique en permettant d'estimer la mesure du changement climatique.

Le graphique 1.4 représente la trajectoire de l'augmentation de la température en °C selon les différents scénarios SSP à horizon 2100, avec comme base la période 1850-1900. Comme chacun de ces scénarios se projette avec une forte incertitude en raison de l'horizon lointain de projection, le GIEC a inclus des intervalles de confiance afin d'englober les différentes possibilités de réchauffement induites par les scénarios.

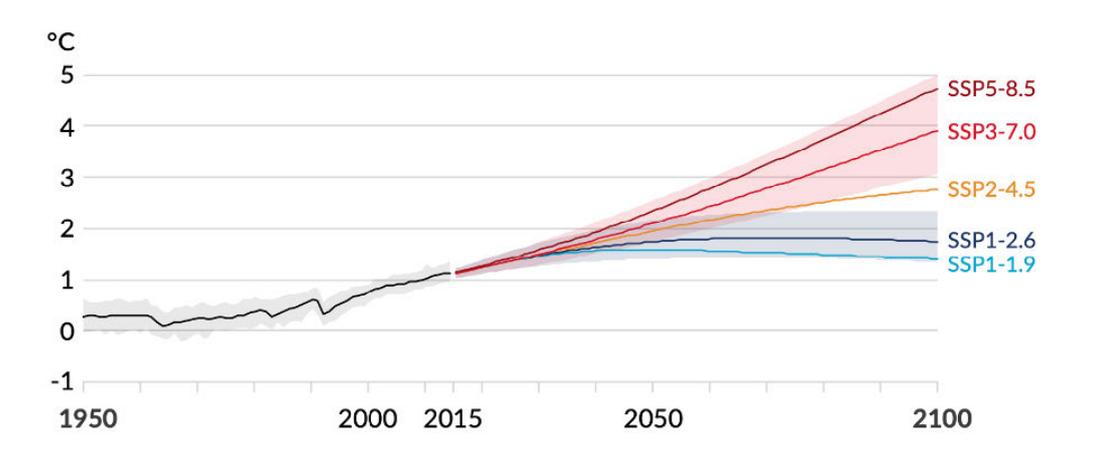


FIGURE 1.4 : Représentation des cinq scénarios climatiques du GIEC (CARBONE4 (2022)).

En plus de cela, il est possible d'utiliser une projection de la population par exemple via les données de l'INSEE, mais aussi des biens concernés par les catastrophes naturelles. Ces projections

supplémentaires servent à prendre en compte une potentielle augmentation de la sinistralité due à une augmentation du nombre et de la valeur des biens assurés.

Pour le risque de transition, on peut également utiliser les scénarios NGFS, qui furent créés par un collectif de banques centrales, et qui envisagent plusieurs types de scénario selon la rapidité et la brutalité de la transition. Dans la mesure où ce mémoire ne traite pas des risques de transition, nous n'approfondirons pas ces scénarios. Le mémoire de PINCEMIN (2021) explique en détail ce que sont ces différents scénarios.

Quel que soit le scénario choisi, on voit sur le graphique que la température de la terre continuera à augmenter au moins jusqu'en 2050. Cela se traduira par une augmentation des sécheresses, des pluies plus intenses causant des inondations plus nombreuses (CCR, 2018).

#### 1.1.4 Les différentes garanties climatiques

Les événements climatiques s'avèrent donc très différents les uns des autres, et avec des conséquences diverses. Aussi retrouvons-nous trois garanties différentes pour ces événements :

1. La garantie MRC (Multi Risques Climatiques) ou assurance récolte. C'est cette dernière qui intervient afin de combler les pertes d'exploitations que peuvent connaître les agriculteurs face au changement climatique, comme lorsqu'une grêle endommage leurs plants.
2. La garantie TGN (Tempête Grêle Neige), qui couvre les événements Tempête, Grêle et Neige et intervient dans les contrats MRH et Auto.
3. Et enfin la garantie CatNat qui va couvrir quant à elle les inondations, la sécheresse et les séismes.

La répartition de la sinistralité des différentes garanties se présente sur le graphique 1.5 comme suit :



FIGURE 1.5 : Répartition des cotisations et de la sinistralité des différentes garanties en 2021 (Source : FRANCE ASSUREURS (2022)).

C'est sur la garantie CatNat que nous allons nous concentrer, en effet elle se trouve être incluse dans un régime très spécial, le régime CatNat.

#### 1.1.5 Le régime CatNat

Le régime CatNat français fut créé en 1982 à la suite de graves inondations en décembre 1981, qui mirent en lumière une couverture insuffisante des catastrophes naturelles de la part des assureurs. Ce

dernier se base sur l’alinéa 12 du préambule de la Consitution de 1947 qui dit ”La Nation proclame la solidarité et l’égalité de tous les Français devant les charges qui résultent des calamités nationales”.

Les piliers du régime sont présentés par la CCR comme étant les suivants (CCR, 2011) :

- couvrir l’intégralité des périls possibles ;
- mettre en place un principe de solidarité où les charges sont supportables par tous ;
- créer un système mixte privé-public ;
- mettre en place une politique de prévention afin d’anticiper au mieux les évènements ;
- garantir un régime solvable et pérenne.

Le régime CatNat a pour objectif l’indemnisation de toutes les victimes des catastrophes naturelles. Il se finance via une surprime de **12%** pour les contrats MRH et de **6%** pour les contrats Auto. Cette surprime s’applique à tout les assurés, qu’ils se trouvent dans une zone à forte sinistralité ou non.

Dans ce régime, l’Etat a un fort pouvoir décisionnel puisque c’est d’abord au niveau ministériel que se prennent les décisions. Le processus d’indemnisation se déroule en plusieurs étapes :

1. A la suite d’un évènement climatique, les assurés déclarent leurs sinistres à leur assureur et préviennent la mairie.
2. Le maire dépose une demande de reconnaissance en catastrophe naturelle auprès du préfet.
3. Le préfet établit un dossier constitué des différentes demandes ainsi que d’un rapport technique sur l’évènement, qu’il transmet au ministère.
4. La direction générale de la sécurité civile et de la gestion des crises présente le dossier à la commission interministérielle qui apprécie le caractère exceptionnel de l’évènement. Pour cela, ils vont se baser sur différents critères. La commission émet alors un avis favorable, défavorable ou d’ajournement publié au journal officiel.
5. Après cela, la préfecture communique les résultats aux mairies qui les communiquent aux assurés. Les assurés peuvent alors déclarer leurs sinistres et peuvent être indemnisés par leurs assureurs après le passage d’un expert.

Le graphe 1.6 explique les différentes étapes du processus de reconnaissance CatNat.



FIGURE 1.6 : Processus d’indemnisation des catastrophes naturelles en France (CCR (2022a)).

Il est important de prendre en compte le fait que les arrêtés CatNat sont émis par communes, ce qui signifie que des communes pourtant voisines et exposées de façon similaires ne seront pas nécessairement reconnues de façon identique, puisqu'une part de politique intervient dans le processus. Les critères de reconnaissance sont les suivants :

- Pour l'inondation, cela se base sur une période de retour de l'évènement de 10 ans. La période retour d'un évènement se définit comme le temps moyen entre deux occurrences dudit évènement. Pour les inondations par ruissellement, le cumul des précipitations à 1h, 3h, 6h ...jusqu'à 72h sont utilisés. Pour les inondations par débordement, les débits de l'eau doivent dépasser une valeur décennale.
- Pour la sécheresse, cela se base sur 2 critères. D'abord la période de retour du SWI(*Soil Wetness Index*) uniforme sur 3 mois de 25 ans. Le SWI est un indice mesurant l'humidité des sols jusqu'à 2 mètres de profondeur, et le SWI utilisé ici est une moyenne glissante de ce dernier sur trois mois, c'est à dire la moyenne sur le mois et les deux mois précédents. Par exemple, la valeur du SWI de mars sera la moyenne des SWI en janvier, février et mars. Ensuite sur l'exposition aux argiles. La surface RGA d'une commune doit être supérieure à 3 % de la surface totale de ladite commune.

En plus de ce processus , tout les assureurs peuvent être réassurés par la Caisse Centrale de Réassurance, réassureur étatique dont les fonds garantis par l'Etat lui permet d'assumer même les coûts les plus élevés des catastrophes naturelles, qui sont très volatiles. Cette réassurance prend la forme d'un traité *Quote-Part*<sup>1</sup> de 50% ainsi qu'un traité *Stop Loss*<sup>2</sup>. Ainsi en 2021 le chiffre d'affaires brut de la CCR au titre des CatNat s'élevait à 1007 millions d'euros (CCR, 2022b).

Ce régime est très spécial pour les assureurs : en effet, ils n'ont la main ni sur la prime, ni sur l'acceptation des sinistres, ni sur les sinistres couverts.

La sinistralité CatNat est donc très particulière pour les assureurs : si les autres évènements peuvent avoir une sinistralité forte, comme la grêle ou les tempêtes, pour ces derniers l'assureur a la possibilité de faire évoluer les primes et les zones couvertes, ce qui lui permet de se protéger face aux augmentations de sinistralité.

## 1.2 Les évolutions récentes et à venir du climat en France.

Depuis la période pré-industrielle (1850-1900), la France a vu sa température moyenne augmenter d'environ 1.5°C . Une des conséquences de ce réchauffement est un renforcement des catastrophes naturelles.

### 1.2.1 Les évènements climatiques récents, témoins d'une évolution rapide du climat actuel

La France a connu une grave sécheresse durant l'été 2022, dont la sinistralité est estimée comme la plus importante de l'histoire du régime. Ainsi, selon FRANCE ASSUREURS (2022), le coût des dégâts de cette sécheresse aux habitations est compris entre 1,9 et 2,8 milliards d'euros, ce qui pourrait donc dépasser les 2,3 milliards d'euros de la sécheresse de 2003, année record en terme de sinistralité.

---

<sup>1</sup>Un traité *Quote-Part* consiste en la prise en charge d'un pourcentage fixe de la sinistralité de l'assureur par le réassureur.

<sup>2</sup>Un traité *Stop Loss* intervient lorsque le montant total des sinistres encourus pendant une période spécifique (généralement un an) dépasse soit un rapport sinistres-primes, soit un excédent (montant défini avec une limite).(Source : CCRE (2021))

Ce fait est aggravé par la sécheresse de février 2023, qui a empêché un bon renouvellement des nappes phréatiques, puisque c'est en hiver que les pluies rechargent au mieux les nappes phréatiques car les plantes sont majoritairement en hibernation et absorbent donc peu d'eau.

Toujours en 2022, la France a aussi connu un épisode de grêle exceptionnel, avec des grêlons pouvant atteindre 9 centimètres de diamètre, ce qui est du jamais vu. On estime à près de 5 milliards d'euros le dégâts causés, ce qui est là encore du jamais vu pour ce phénomène selon le rapport de FRANCE ASSUREURS (2022).

En 2018, la France a connu une grave inondation dans le département de l'Aude, avec des précipitations atteignant 300 millimètres en 6 heures, provoquant une crue de l'Aude et de ses affluents issus de la Montagne Noire. Cette crue extrêmement rapide et violente a entraîné la mort de 14 personnes et des dégâts aux bien assurés de l'ordre de 200 millions d'euros. 257 communes ont été reconnues en arrêté CatNat inondation pour cet évènement.

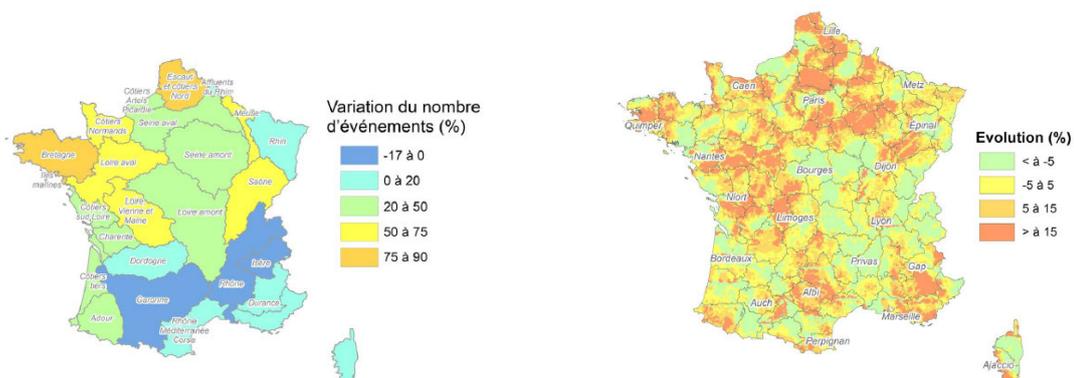
En plus de s'être aggravé récemment, les catastrophes naturelles sont amenées à empirer dans le futur. Pour cela, il faut se demander quel scénario du GIEC considérer, les conséquences s'avérant très différentes selon le scénario choisi. Nous nous concentrerons sur les risques inondations et sécheresse.

### 1.2.2 Evolution des inondations

En tant que principal organisme météorologique, Météo France a élaboré des modèles de projections climatiques à partir de leur base de données, issue des stations météo françaises. L'institution a considéré les scénarios 4.5 et 8.5 du GIEC, le 4.5 étant le plus probable et le 8.5 le plus pessimiste.

A partir de ces données, la CCR a alors produit un rapport sur les évolutions climatiques à venir pour le scénario 8.5. Voyons les résultats obtenus pour le risque inondation. Afin d'estimer au mieux le risque, il est utile de délimiter la France métropolitaine par bassin versant. Un bassin versant correspond à une zone du territoire drainée par un cours d'eau, généralement un fleuve, et ses tributaires, tels que des rivières ou des lacs (OFFICE FRANÇAIS DE LA BIODIVERSITÉ (2022)). Il s'agit donc d'un réseau de cours et de plans d'eau connectés entre eux, ce qui est important lorsque que l'on parle d'inondations puisqu'un trop plein dans l'un des ruisseaux se déversera dans d'autres cours d'eau du bassin versant, étendant l'inondation.

La simulation de la CCR sur les bassins versants montre une claire délimitation entre le sud et le nord. Là où les bassins versants du sud verront leur nombre d'évènements diminuer, ceux du nord et en particulier de l'ouest augmenteront fortement. Ces augmentations sont expliquées par une hausse des précipitations extrêmes dans ces régions. Voici les résultats pour le risque inondation (figure 1.7).



(a) Variation du nombre d'évènements inondations selon le bassin versant.

(b) Représentation de l'extension des surfaces inondées

FIGURE 1.7 : Représentation des variations de l'évènement inondation à horizon 2050. Source : CCR (2018).

La CCR s'est ensuite penchée sur les variations de l'extension des zones inondées d'ici à 2050. On observe que le nord s'avère très touchés, ainsi que les pays de la Loire et le Sud des Alpes. Le Nord des Alpes et le centre sont quant à eux peu touchés, et on voit sur la figure 1.7 une corrélation entre la variation du nombre d'évènements par bassin versants et l'évolution des surfaces inondées.

Nous pouvons également citer le livre blanc "Risques climatiques : quelle prévention ?" publié par Covéa en février 2022 (COVÉA, 2022). En utilisant le scénario 8.5 et une période de référence 2008-2018, Covéa a ainsi évalué que la sinistralité annuelle moyenne augmenterait de 130 % pour les crues torrentielles et par ruissellement, et de 110 % pour les crues par ruissellement à horizon 2050, ce qui est très considérable au vu de la proximité de cet horizon. Voici la carte 1.8 montrant les évolutions par département.

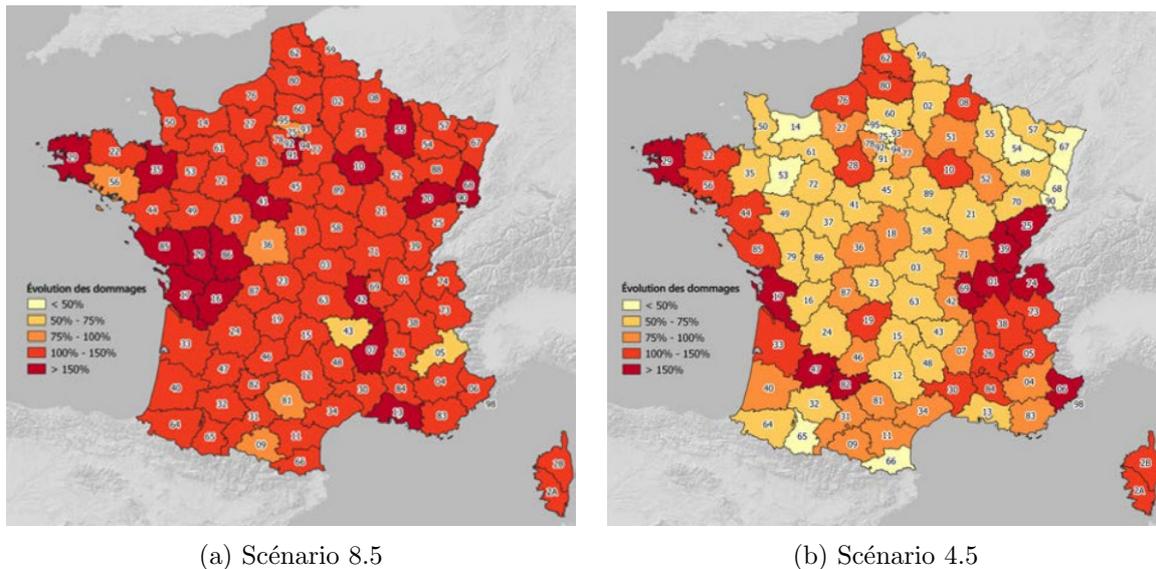


FIGURE 1.8 : Représentation des variations de sinistralité des évènements inondation à horizon 2050. Source : COVÉA (2022).

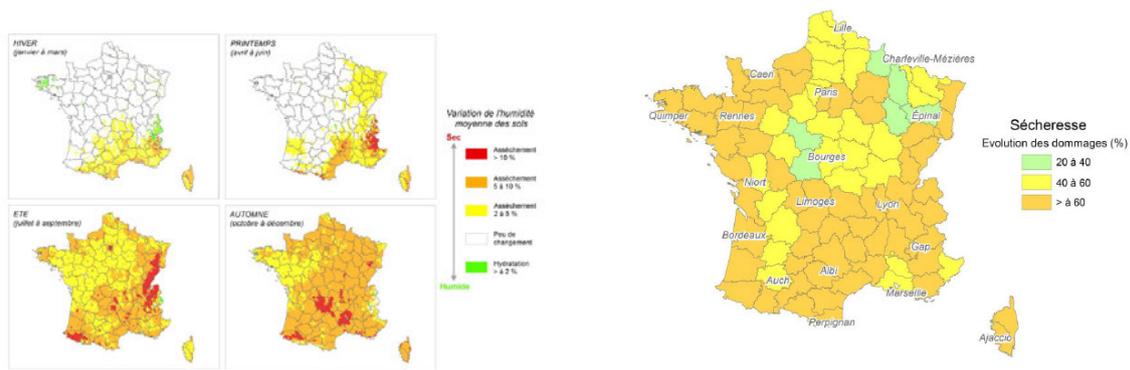
Covéa évalue également que les précipitations décennales, c'est à dire avec une période de retour de 10 ans, se produiront en moyenne tous les 4.5 ans.

Concernant les submersions marines, évènement amené à évoluer fortement du fait de la fonte des glaces et de la dilatation des océans qui entraînent une hausse du niveau de la mer. Selon le GIEC, le rythme de la hausse du niveau marin sera entre 4 et 7 mm par an entre 2010 et 2050 dans le scénario 8.5, conduisant à une élévation comprise entre 18 et 30 centimètres (GIEC, 2022). Dans ces conditions, le ministère de la transition écologique estime que la fréquence des submersions marines dans des villes telles que Nantes ou Brest pourrait être multipliée par 500 d'ici à 2100.

### 1.2.3 Evolution de la sécheresse

Etudions ensuite la sécheresse, et reprenons l'étude de la CCR précédemment citée. Pour ce qui est de la sécheresse, la CCR a utilisé la carte RGA créée par le BRGM ainsi que des données d'humidité des sols produites par Météo France. A partir de là, la CCR a représenté l'évolution de l'humidité des sols pour chaque saison à horizon 2050 selon le scénario 8.5, ainsi que la hausse des dommages qui en découlent, voir le graphique 1.9.

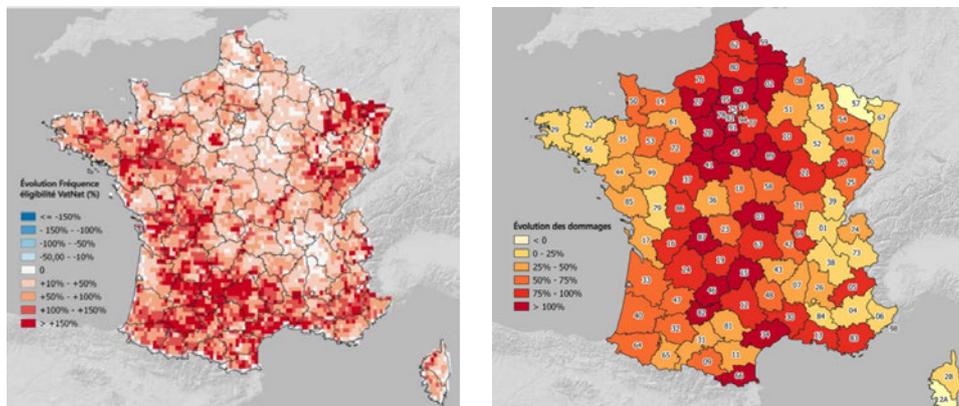
Si l'hiver et le printemps resteront assez proches de ce que l'on connaît aujourd'hui à l'exception du sud, l'été et l'automne seront très impactés peu importe l'endroit. Le sud en particulier présente une très forte augmentation des températures.



(a) Représentation de l'assèchement des sols selon la saison. (b) Evolution des dommages liés à la sécheresse à horizon 2050.

FIGURE 1.9 : Représentation des variations de l'évènement sécheresse à horizon 2050. Source : CCR (2018).

Regardons ensuite les évolutions prévues par Covéa dans son livre blanc. Comme nous l'avons vu dans la description du risque sécheresse, la période de retour du SWI est un élément déterminant concernant la décision de reconnaître en arrêté sécheresse ou non une commune. Covéa a donc projeté cet indice d'humidité afin de voir comment l'éligibilité des communes évoluera à l'horizon 2050. Cette hausse des reconnaissances entraînera logiquement une hausse de la sinistralité, représentée ici par département 1.10.



(a) Evolution du nombre de reconnaissances en arrêté CatNat sécheresse à horizon 2050. (b) Evolution des dommages liés à la sécheresse par département à horizon 2050.

FIGURE 1.10 : Représentation des impacts du changement climatique sur le risque sécheresse en France COVÉA (2022).

Au total, la hausse de la fréquence moyenne d'éligibilité en arrêté sécheresse est de 70 % , provoquant une hausse de la sinistralité de 60 % pour ce péril. La hausse est particulièrement notable pour le nord de la France, pourtant peu exposé aux argiles comme vu précédemment. Enfin, d'après la conférence "Impact des évolutions climatiques sur les périodes de retour des évènements extrêmes" RAYMOND (2023), si nous nous plaçons dans le scénario 8.5 à horizon 2070 la période de retour pour les évènements inondations décennaux passe à 6 ans et les évènements sécheresse décennaux passe à 8 ans. Les évènements extrêmes deviendront par conséquent moins rares.

Voyons maintenant comment les assureurs se situent face à des risques climatiques de plus en plus graves.

### Une sinistralité en forte augmentation

Du fait des évolutions récentes du climat vues précédemment, la sinistralité des événements naturels a logiquement beaucoup augmenté à son tour, atteignant un niveau bien plus important ces 5 dernières années CCR (2021), comme nous allons le voir sur le graphique 1.11, qui représente l'évolution de la sinistralité des événements climatiques en France.

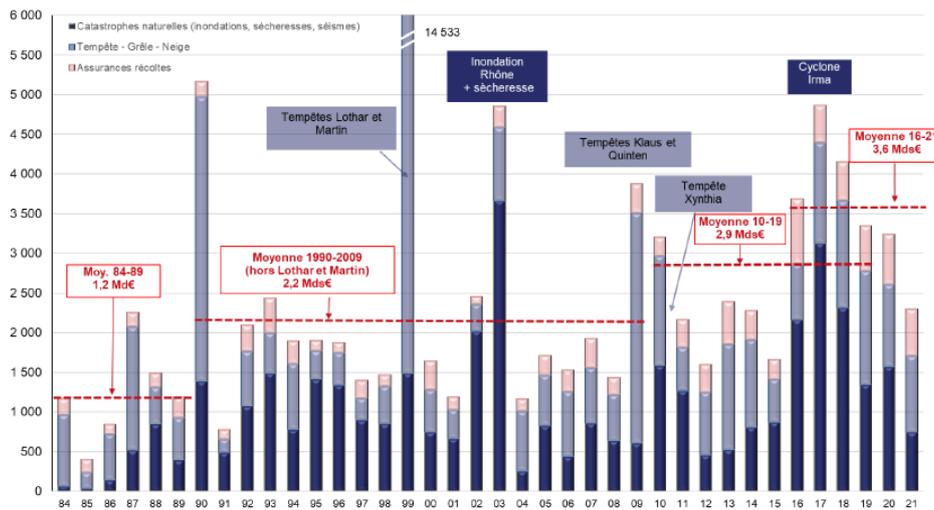


FIGURE 1.11 : Représentation de la sinistralité des événements naturels en France depuis 2022 (FRANCE ASSUREURS, 2022)

On voit sur le graphique que non seulement la sinistralité s'avère extrêmement volatile, pouvant être multipliée par trois d'une année à l'autre, mais en plus que sa moyenne a fortement augmenté, passant de 2.9 Mds € entre 2010 et 2019 à 3.6 Mds € entre 2016 et 2020. Comme évoqué précédemment, l'année 2022 s'est avérée particulièrement forte en terme de sinistralité, puisque les estimations montent à 10 milliards d'euros, en faisant l'année la plus chère depuis 1999, année marquée par des tempêtes de grande ampleur.

Il est à noter que le changement climatique n'est pas le seul responsable de cette inflation. En effet, l'augmentation est en partie due à la hausse de la valeur assurée mais également à une présence parfois accrue dans des zones à risques, particulièrement dans le sud et en bord de mer.

Toujours selon le rapport de la CCR, les pertes annuelles des assureurs seraient aggravées de 35% dans le scénario 8.5 à horizon 2050, avec une augmentation de la sinistralité CatNat d'environ 50%, toutes choses étant égales par ailleurs. L'augmentation due aux submersions marines est particulièrement notable du fait de l'augmentation du niveau de la mer, comme vu dans la section précédente. Voir figure 1.12.

Regardons maintenant les conséquences de cette augmentation de la sinistralité et les adaptations possibles du régime CatNat.

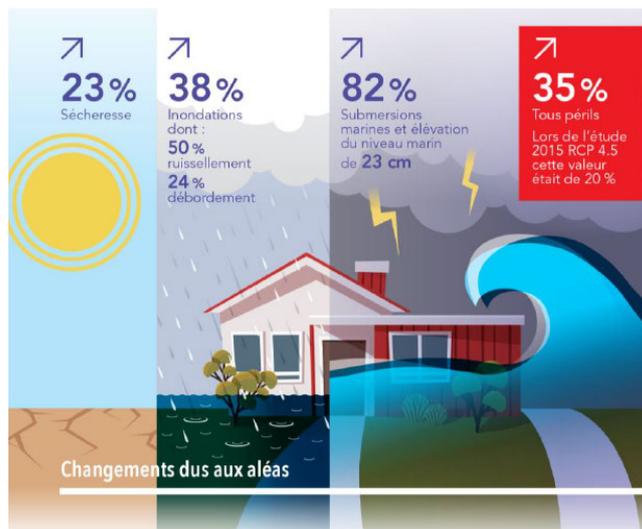


FIGURE 1.12 : Evolution en 2050 des pertes annuelles moyennes par périls liées aux évolutions du climat sur les valeurs assurées actuelles (*Source : CCR (2018)*).

#### 1.2.4 Un régime en difficulté mais en évolution

Cette sinistralité est spéciale du fait du régime CatNat, qui couvre les assureurs tout en les empêchant d'avoir la main sur la gestion de la sinistralité. Cependant, la stabilité de ce régime est fortement menacée par des pertes de plus en plus importantes. Ainsi, le régime CatNat français a enregistré des pertes en 2021 pour la septième année consécutive, le résultat du compte technique atteignant **-105 millions d'euros**, ce qui représente 6% des primes.

Cette pression est accentuée par de nouvelles lois visant à mieux indemniser les assurés et faciliter les reconnaissances CatNat. Ainsi la loi du 28 décembre 2021 a considérablement modifié le régime, par exemple en passant de 18 à 24 mois le délai de dépôt de demande de reconnaissance naturelle ou en montant à 5 ans la prescription relative au RGA, ce qui permettra à plus d'assurés d'obtenir une indemnisation. Ces évolutions peuvent toutefois être relativisées par d'autres régulations instaurées par le régulateur qui pourront permettre une diminution de la sinistralité.

Ainsi la loi Elan de 2019 impose à tout vendeur de terrain non bâti d'informer les potentiels acquéreurs d'un risque argile moyen ou fort, ce qui permet au constructeur de tenir compte de ce paramètre dans la conception du bâtiment.

Face à cela, les assureurs pressent le gouvernement de modifier le régime. La CCR a émis trois propositions afin de résoudre le problème :

1. En renforçant les mesures de prévention auprès des assurés.
2. En augmentant le taux de surprime.
3. En recentrant les critères de reconnaissance CatNat.

La sinistralité actuelle est déjà difficilement supportable pour le régime qui enregistre des pertes de façon récurrente, et nous avons vu que cette dernière n'allait faire qu'augmenter par la suite, faisant de 2022, pourtant considérée comme exceptionnelle, une année annonciatrice.

Maintenant que nous avons expliqué les risques climatiques et leurs impacts, intéressons nous au cadre réglementaire du processus ORSA, qui est ce qui nous intéresse dans ce mémoire.

### 1.3 Présentation du régime Solvabilité 2

Solvabilité 2 est un régime prudentiel qui "s'applique aux entreprises d'assurance directe vie et non-vie qui sont établies sur le territoire d'un État membre ou qui désirent s'y établir (article 2)" depuis le 1er janvier 2016. Établi par la directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009, Solvabilité 2 est un nouveau régime réglementaire européen unifiant la réglementation assurantielle des différents pays de l'Union Européenne (PE, 2009). Ce régime a pour objectif de renforcer la régulation et la prudence des assureurs, en établissant une approche prospective des risques. Ce régime se décompose en trois piliers : exigences qualitatives, exigences quantitatives et obligations de reporting 1.13 .

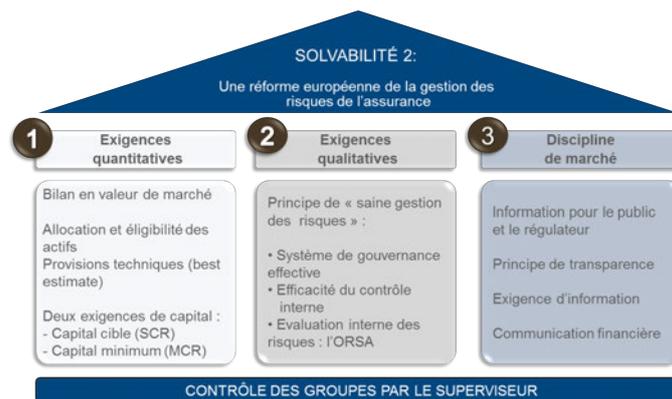


FIGURE 1.13 : Représentation des trois piliers de Solvabilité 2 Source : RISK ATTITUDE (2015).

Nous parlons du premier pilier et des ses différents composants puis du second pilier et en particulier de l'ORSA qui est l'élément qui nous intéresse. Nous terminons par un aperçu du dernier pilier.

Parlons d'abord du premier pilier, qui compile toutes les exigences quantitatives requises pour les assureurs. En effet, bien que le processus ORSA fasse partie du second pilier, nombre d'éléments dont il est question dans ce dernier sont issus du premier pilier, tel que le SCR.

#### 1.3.1 Présentation du pilier 1

Le premier pilier compile les règles de calcul du capital de solvabilité requis, du minimum de capital requis ainsi que les règles de valorisation, que ce soit de l'actif comme du passif.

Sous solvabilité 2, le bilan de l'assureur est un équilibre entre d'un côté, des actifs valorisés à leur valeur de marché, et des passif composé des fonds propres et des provisions techniques.

Dans les fonds propres, nous trouvons le capital de solvabilité requis, ou SCR. Le SCR correspond au montant de fonds propres requis pour l'assureur afin de faire face à une ruine économique à horizon 1 an et avec une probabilité de 99.5%. Sa formule s'écrit

$$SCR = FP(0) - q_{0.5\%}(FP(1)) \times D(1),$$

avec :

- $FP(0)$  le niveau de Fonds Propres initial.
- $q_{0.5\%}(FP(1))$  le quantile à 0.5% des Fonds Propres à un an.
- $D(1)$  le facteur d'actualisation à un an.

Le SCR est lui-même décomposé en plusieurs sous modules, tels que le SCR catastrophe ou le SCR marché . Le SCR peut être calculé soit via la formule standard définie par le régime soit par un modèle interne à l'assureur (ACPR, 2011).

Pour la formule standard, le calcul du SCR est fait en appliquant un choc aux éléments du bilan de l'assureur, puis en agrégeant les différents sous-modules via une formule de corrélation. La représentation générale de ces sous-modules se présente comme sur le graphique 1.14 :

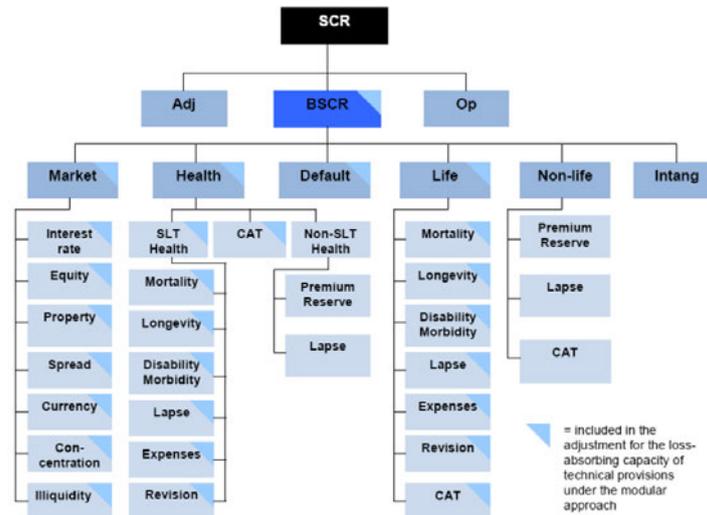


FIGURE 1.14 : Représentation des différents modules et sous-modules du SCR sous Solvabilité 2 (Source : ACPR (2011)).

Les derniers éléments des fonds propres sont :

- Le MCR (minimum de capital requis), qui constitue le minimum de fonds propres que doit avoir l'assurance pour continuer son activité. Son calcul se déroule en trois étapes :
  1. D'abord calcul d'un MCR linéaire, somme d'une partie non vie obtenue par combinaison des primes émises et des *Best Estimate* ainsi que d'une partie vie obtenue par combinaison des provisions techniques et des capitaux sous risques.
  2. Ensuite le MCR combiné, qui ramène le résultat précédent entre 25 et 45 % du SCR via une contrainte réglementaire.
  3. Enfin, comparaison du MCR avec une valeur plancher, le placher MCR absolu ou AMCR. Le MCR est égale au maximum des deux valeurs considérées.
- Le capital excédentaire, qui correspond à la différence entre les éléments de l'actif et du passif précédemment nommés.

Pour plus de détails concernant ces éléments, nous renvoyons vers le mémoire d'actuariat écrit par KARAMAKO (2013).

Au niveau des Provisions techniques, on va trouver la *Risk Margin* ainsi que le *Best Estimate*.

Le *Best Estimate* correspond à la valeur actualisée des cash-flows futurs de passifs. Il peut être considéré avec la *Risk Margin* comme le prix qu'un tiers serait prêt à payer pour racheter le passif de la compagnie. La *Risk Margin* représente le coût que représente l'immobilisation de fonds propres que doit faire l'assureur pour faire face à ses obligations. Son calcul consiste en la projection des SCR futurs. Pour récapituler, voir la figure 1.15.



FIGURE 1.15 : Représentation du bilan assurantiel sous Solvabilité 2.

### 1.3.2 Présentation du pilier 2

#### Présentation générale

Le second pilier du régime solvabilité 2 regroupe les exigences qualitatives de ce dernier et est établi par les articles 41 à 50 de la directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009.

Le premier point du pilier concerne les exigences en matière de gouvernance du régime. L'objectif de Solvabilité 2 en la matière est de garantir une gestion prudente et saine de l'organisme. Au minimum, l'entreprise doit :

- Respecter le principe des quatre yeux, c'est à dire avoir au moins deux dirigeants effectifs.
- Avoir un responsable pour chaque fonction clé, qui sont la fonction actuarielle, la gestion des risques, l'audit interne et la conformité.

Pour chacun de ces postes, une notification devra être envoyée à l'ACPR (Autorité de Contrôle Prudentiel et de Régulation) qui vérifiera l'honorabilité, la compétence et l'expérience des individus considérés.

#### ORSA (Own Risk Solvency Assessment)

Etabli par l'article 45 de la directive Solvabilité 2, l'ORSA désigne l'ensemble des processus mis en place par l'entreprise d'assurance pour évaluer ses risques propres, ainsi que leur évolution et les mesures de gestion desdits risques. De façon générale, l'ORSA est projeté sur un horizon allant de 3 à 5 ans. L'ORSA est un outil de pilotage important pour l'entreprise, qui doit lui permettre d'intégrer l'élément risque à sa stratégie à court terme.

L'ORSA va évaluer au moins trois éléments :

1. le besoin global de solvabilité, qui est une estimation des risques de l'assureur pour lui-même. Le besoin global de solvabilité ne comporte pas que les SCR, mais va également prendre en compte des risques non quantifiables comme le risque de réputation.
2. le respect des exigences de capital : l'assureur doit alors évaluer les potentiels changements futurs dans son profil de risque, la qualité ainsi que la quantité de ses fonds propres et la composition

desdits fonds propres et son évolution potentielle. Ainsi que des exigences relatives au calcul des provisions techniques, avec vérification de leur conformité ainsi que des risques que ces calculs pourraient entraîner.

3. la différence entre les hypothèses de calcul du SCR et le profil de risque réel de l'entreprise.

Ce processus débouche sur un rapport remis à l'ACPR, qui est le superviseur assurantiel, ainsi qu'à l'AMSB (*Administration, management or supervisory body*), qui regroupe le conseil d'administration et la direction générale. Cela permet au superviseur ainsi qu'à la direction d'avoir accès à des informations sur le profil de risque et la solvabilité de l'assureur. Parmi les risques évalués par l'ORSA, on trouve par exemple les risques opérationnels, de liquidité, de réputation ou encore les risques réglementaires.

L'ORSA a donc un aspect quantitatif par le calcul du besoin global de solvabilité, mais également qualitatif en tant qu'outil de pilotage et d'indicateur stratégique. Ce processus doit donc permettre à l'entreprise de comprendre et de mesurer quels éléments de risque sont susceptibles de modifier sa solvabilité et sa situation financière.

### 1.3.3 L'intégration des risques climatiques et de durabilité dans le pilier 2

Le règlement délégué (UE) 2021/1253 du 2 août 2022 a introduit une nouvelle catégorie : le risque de durabilité. Cette catégorie inclut les risques physiques, de transition ou encore les risques de responsabilité, qui sont liés aux potentiels procès pour responsabilité des entreprises dans le changement climatique.

Ce règlement a notamment pour objectif de pousser les assureurs à allonger l'horizon de prise en compte du risque en évaluant l'impact de leurs investissements à long terme, et de refléter les préférences ESG (Environnemental Social Gouvernance) de leurs clients dans leurs choix.

Le risque de durabilité s'est vu intégré dans plusieurs éléments importants pouvant être contrôlés par le régulateur :

- l'adéquation des conseils fournis aux clients ;
- l'identification des potentiels conflits d'intérêts ;
- l'application de la gouvernance des produits.

Cette notion de durabilité doit également être prise en compte dans l'évaluation de la qualité, la durabilité et la quantité de leurs actifs en portefeuille (ACPR, 2019).

### 1.3.4 Présentation du pilier3

Le troisième pilier du régime Solvabilité 2 a trait au reporting, que ce soit au superviseur ou au public. Les rapports qualitatifs doivent parler de la gouvernance, de l'importance des risques, des différents capitaux calculés et expliquer l'histoire derrière les différents chiffres. Ces rapports doivent également expliquer pourquoi ces éléments peuvent différer de la régulation en place.

Il existe deux types de ces rapports :

- Le SFCR (*Solvency Financial Condition Report*), un rapport annuel à destination du public. Ce rapport doit notamment parler des résultats de la compagnie, de son système de gouvernance, son profil de risque, ses méthodes de valorisation ainsi que sa gestion du capital.
- Le RSR (*Regular Supervisory Report*), un document privé à destination du régulateur devant être émis au minimum tout les trois ans.

Pour le quantitatif, chaque organisme est tenu de rapporter au régulateur les éléments suivants : bilan, fonds propres, activité, état détaillé des actifs, des dérivés et état sur la mise en transparence des fonds d'investissement.

Nous allons maintenant nous concentrer sur le rôle de l'ORSA dans les risque climatiques.

## 1.4 Modélisation climatique et ORSA climatique

Notre objectif est d'expliquer comment l'ORSA peut être utilisé afin de modéliser les risques climatiques, et en quoi son rôle s'avère particulier dans la prise en compte desdits risques.

### 1.4.1 Impact des évènements climatiques sur les éléments du bilan assurantiel

#### Impacts sur l'actif

Dans Solvabilité 2, l'actif de l'assureur est comptabilisé en valeur de marché, et se compose essentiellement d'obligations, d'actions ou du valeurs immobilières. Cet actif sera aussi bien impacté par le risque de transition que le risque physique.

Nous avons vu précédemment que le risque de transition pouvait diminuer le cours de l'action d'une entreprise, impactant à la baisse le portefeuille de l'assureur. Certains secteurs comme l'extraction minière pourront être très touchés.

Les obligations seront également impactées, avec une dégradation de la note de la dette souveraine pour les pays les plus impactés par le changement climatique, ces derniers devant faire face à des catastrophes naturelles d'une ampleur très importante (RÉSEAU ACTION CLIMAT, 2022).

Le risque physique va quant à lui mettre en danger l'activité de certaines entreprises comme par exemple les exploitations agricoles, mais également faire baisser la valeur de biens immobiliers exposés à des risques climatiques.

#### Impacts sur le passif

Impact sur les Fonds Propres :

1. Puisque l'assureur fait face à une hausse de la volatilité de la sinistralité, il doit la prendre en compte dans son SCR. Le SCR étant sur un an, cela n'est pas forcément un problème si l'assureur sait correctement ajuster ses primes. Dans la présentation « Prise en compte des risques climatiques dans la mesure de risque des assureurs » (PLANCHET, 2022), il est envisagé deux cas de figures :
  - Soit l'assureur sait correctement adapter ses tarifs, ce qui le forcera à négocier avec son réassureur et ses clients mais son SCR ne sera pas impacté.
  - Soit l'assureur n'arrive pas à adapter ses tarifs, ce qui conduit à la hausse du SCR catastrophes ainsi que du SCR primes et réserves.
2. Nous avons vu que le risque à l'actif serait considérablement augmenté par les risques climatiques, pouvant potentiellement impacter le SCR marché à la hausse.
3. Concernant le MCR, son calcul fait intervenir le montant du SCR , pouvant entraîner une corrélation entre les deux. Une hausse du SCR pourrait alors entraîner une hausse du MCR.

Les Fonds Propres pourraient donc être impactés à la hausse par le changement climatique. Cette idée est renforcée par le guide de l'EIOPA sur l'inclusion des risques climatiques dans le processus ORSA (EIOPA, 2022), qui montre une claire hausse du SCR global en fonction du scénario climatique. Voir table 1.1.

	Baseline	RCP 2.6	RCP 8.5
2020	11.5	na	na
2030	na	12	13
2040	na	14	18

TABLE 1.1 : Tableau d'évolution du SCR d'une compagnie européenne fictive, en millions d'euros (EIOPA, 2022).

Impact sur les provisions techniques :

1. Impact sur la *Risk Margin* : pour rappel, la *Risk Margin* correspond à la valeur actualisée des SCR. Elle sera donc impactée à la hausse de façon proportionnelle à ces derniers.
2. Impact sur le *Best Estimate* : le *Best Estimate* correspond à la valeur actualisée des cash-flows de passifs sous la probabilité risque-neutre. L'augmentation de la sinistralité due au changement climatique impactera donc le *Best Estimate* à la hausse. Cela pourra être aggravé par le fait que certains réassureurs deviennent plus prudents face au risque climatique et refusent d'en assurer certains, ce qui augmentera encore le *Best Estimate*.

On note donc que le changement climatique pourrait impacter d'une manière générale le passif à la hausse et l'actif et la baisse, ce qui aurait pour effet mécanique de faire diminuer le capital excédentaire au bilan.

L'impact du changement climatique peut donc clairement impacter l'assureur dans son bilan, ce qui pousse le régulateur à se pencher en détail sur les risques climatiques.

### 1.4.2 L'exercice pilote ACPR de 2021

Lors des accords de Paris en 2015, les Etats se sont engagés à atténuer le réchauffement climatique et à prendre des mesures d'adaptation afin d'en limiter au mieux les conséquences. Dans ce contexte, l'Etat français a fait passer des lois sur la croissance verte et la transition énergétique, qui visent notamment à mieux orienter les investissements vers le développement durable. Dans cette démarche, l'ACPR (Autorité de Contrôle Prudentiel et de Résolution) a également décidé de renforcer le contrôle sur les risques climatiques auxquels font face les assureurs. De plus, les risques climatiques sont importants mais seront amenés à fortement évoluer dans le futur.

L'ACPR est un organisme étatique ayant pour rôle de contrôler les assureurs français et de s'assurer de leur bonne adéquation avec la régulation.

Du fait de l'importance des risques climatiques et de leurs potentielles évolutions, l'ACPR encourage les assureurs à se pencher sur ces risques, et a pour cela lancé un exercice pilote en 2020. Les deux objectifs principaux de ce dernier sont de sensibiliser les institutions financières françaises aux risques climatiques ainsi que de renforcer leur capacités d'analyse et de gestion d'une part, et obtenir un premier ordre de grandeur de leurs risques et vulnérabilités.

Cet exercice s'est déroulé sur trois trimestres et a mobilisé 9 groupes bancaires et 15 groupes d'assureurs représentant 85% du bilan bancaire et 75% du bilan assurantiel.

## Hypothèses

Pour la réalisation de l'exercice, l'ACPR s'est basée sur des hypothèses ambitieuses. Les risques, qu'ils soient physiques ou de transition, furent projetés à 30 ans, un horizon de temps assez long qui excède largement les durées de projections habituelles.

Dans le contexte du risque de transition, les scénarios NGFS furent utilisés, avec un scénario de référence correspondant à une transition rapide et ordonnée, ainsi que deux scénarios de transition désordonnée. Le premier scénario de transition désordonnée suppose une transition tardive où les objectifs de réduction de carbone en 2030 ne sont pas atteints. Le second correspond à une transition accélérée et suppose quant à lui une très forte hausse du prix du carbone d'ici à 2050.

Pour les risques physiques, le scénario RCP 8.5 du GIEC fut retenu, soit le plus pessimiste de tous. Ce choix s'explique par la volonté d'étudier les impacts les plus forts.

Ensuite, il a été supposé que le bilan des assureurs serait fixe les cinq premières années de projections pour ensuite devenir dynamique, c'est à dire que les participants devaient projeter leurs expositions de crédit selon les différents scénarios. Ces derniers étaient cependant autorisés à réallouer leur portefeuille de crédit. Le graphique 1.16 montre bien la différence des expositions des participants aux différents secteurs, avec une claire baisse de l'exposition dans les industries extractives, vues comme très polluantes.

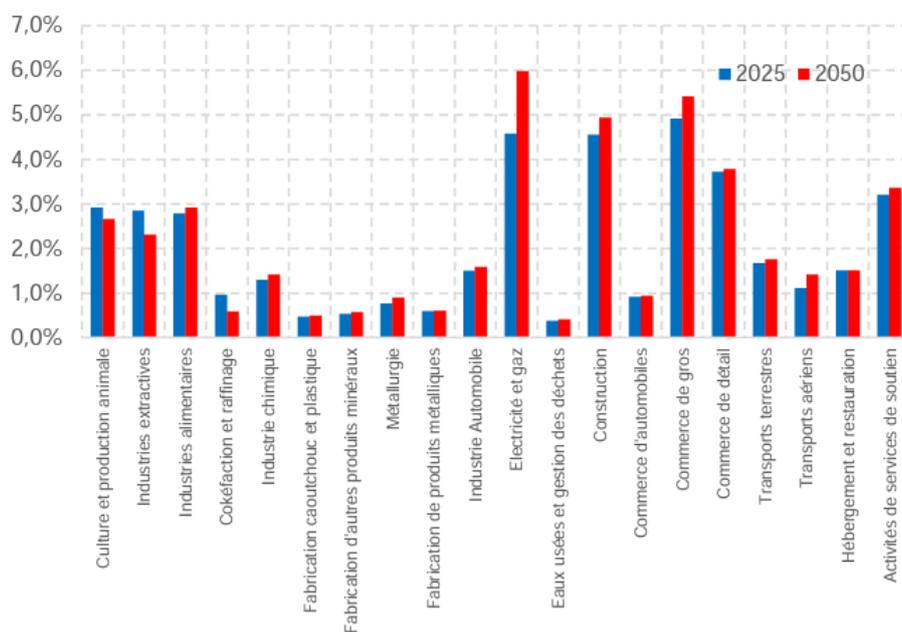


FIGURE 1.16 : Structure sectorielle des expositions de crédits (BANQUE DE FRANCE, 2021).

## Résultats

Concernant le risque physique, la sinistralité des branches entrant dans le régime CatNat augmente de 174 % entre 2019 et 2050. Puisque l'ACPR a laissé la possibilité aux assureurs de modifier leur

politique de souscription à compter de 2025 en fonction de leur S/P, cela s'est traduit par une hausse des primes comprise entre 130 et 200 % des primes sur 30 ans. Cette hausse de la sinistralité est cependant très inégalement répartie, comme le montre la carte suivante, avec certains départements comme la Manche qui voit sa sinistralité multipliée par 5 à horizon 2050 dans le scénarios RCP 8.5, voir figure 1.17.

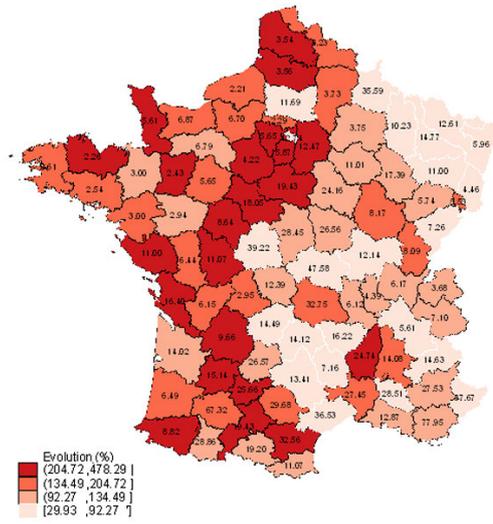


FIGURE 1.17 : Evolution de la sinistralité tous périls entre 2019 et 2050 (*Source* : BANQUE DE FRANCE (2021))

A noter que pour obtenir ce résultat, l'augmentation de la population a également été modélisée selon les prévisions de l'INSEE à horizon 2050.

Concernant le risque de transition, ce dernier s'est avéré avoir un impact plus modéré que prévu, ce qui peut s'expliquer par la possibilité de réallocation du portefeuille des assureurs. Ainsi l'exercice a conclu que les actions d'entreprises émettrices de CO2 chuteront au pire de 26%, ce qui aura un impact faible du fait de la diversité du portefeuille des assurances. En effet, l'exposition des assureurs aux secteurs potentiellement à risque est d'environ 17 % de l'actif de ces derniers. Ces résultats restent cependant à prendre avec précaution en raison des difficultés de projeter les risques de transition.

Une des grands conclusions de l'exercice pilote est l'importance de l'ORSA dans la mesure des risques climatiques, considéré comme l'outil privilégié pour l'étude du risque climatique. Si les résultats se sont avérés intéressants, l'ACPR a néanmoins constaté un grand manque de connaissance des banques et assurances sur leurs risques climatique. Aussi l'institution a prévu de répéter l'exercice de façon régulière, et ce dès cette année. Parlons maintenant du prochain exercice pilote de 2023.

### Les hypothèses du nouvel exercice pilote

Le 6 juillet 2023, l'ACPR a publié le nouveau guide des hypothèses du nouvel exercice pilote ACPR (2023b) dont les résultats seront publiés en mai 2024. Ce nouvel exercice a pour but de continuer à sensibiliser les acteurs à leurs risques climatiques, à affiner et compléter la méthodologie et le processus de prise en compte desdits risques, et enfin à actualiser leur prise en compte en mettant à jour les vulnérabilités des assureurs français.

Cet exercice reprend un bon nombre d'hypothèses du précédent exercice, tout en apportant certaines nouveautés que nous présentons. Cet exercice reposera sur :

- Deux scénarios de long terme, un basé sur un scénario en-dessous de 2°C du NGFS et un autre basé sur un scénario désordonné de transition retardée.
- Un scénario de court terme développé par l'ACPR et la Banque de France et à horizon 2023-2027, scénario combinant des chocs de risques physiques aigus ainsi qu'un choc à l'actif lié à une prise de conscience des marchés aux événements extrêmes.

Pour les scénarios de long terme, des projections seront fournies par l'ACPR en bilan dynamique par pas de 5 ans, de 2025 à 2050. Pour le scénario de court terme, les projections seront faites pour chaque année en bilan statique de 2023 à 2027.

De plus, concernant les scénarios physiques, le scénario 4.5 du GIEC remplace désormais le 8.5. Ce choix d'un scénario moins extrême s'explique par les raisons suivantes :

- il apporte plus de cohérence avec divers scénarios du NGFS qui sont « *below 2°C* » et « *delayed transition* » ;
- à horizon 2050, les différences entre les scénarios 4.5 et 8.5 sont en fait minimales, que ce soit dans la hausse des températures ou la survenance d'événements extrêmes ;
- en utilisant les quantiles, il est possible de simuler des impacts plus extrêmes tout en restant dans le même scénario, comme par exemple en prenant le 95<sup>ème</sup> quantile ;
- enfin, le RCP 8.5 est sujet à controverse du fait de ses hypothèses très pessimistes qui peuvent s'avérer contestables, comme par exemple l'hypothèse du doublement des émissions de GES (ACPR, 2023b).

Nous pouvons donc nous concentrer sur les deux scénarios cités afin de nous placer dans le cadre de l'exercice pilote en respectant ses différentes hypothèses.

Détaillons plus le scénario court terme, ce dernier étant assez intéressant. Tout d'abord, ce dernier reproduit en été 2023 et 2024 des vagues de chaleur similaires à celles rencontrées durant l'été 2022, les assureurs pouvant appliquer à ces années la sinistralité constatée ou estimée en 2022 pour les activités non-vie.

Pour le risque inondation, l'ACPR envisage un scénario de rupture du barrage de Serre-Ponçon au premier semestre 2025, rupture associée à une crue exceptionnelle de la Durance due à des fortes précipitations et d'importantes chaleurs provoquant une fonte des neiges rapide. En effet le barrage n'est conçu que pour contenir l'eau que jusqu'à un niveau déjà dépassé lors des années 1843, 1856 et 1886. Le réchauffement climatique ayant un impact fort sur les précipitations extrêmes et la fonte des neiges (les débits hivernaux pourraient augmenter entre 10 et 50 % d'ici la fin du siècle). En cas de rupture, il est estimé que 17 communes seraient touchées, entraînant des dégâts de 1,3 Md€.

### 1.4.3 L'ORSA climatique

L'ORSA climatique consiste en la prise en compte du changement climatique et du risque climatique dans le processus ORSA. Sa mise en place difficile et sa différence avec les autres risques considérés dans l'ORSA font qu'un nombre insuffisant d'assureurs prennent en compte les risques liés au changement climatique.

Commençons par détailler le processus de mise en place de l'ORSA climatique selon les recommandations émises par l'EIOPA.

Le 2 août 2022, l'EIOPA a publié un Guide d'application pour expliquer le processus de prise en compte du risque climatique dans l'ORSA (EIOPA, 2022), afin d'aider les assureurs à mieux évaluer leurs risques climatiques. Cela est dû au fait que le risque climatique s'avère très différent des autres

de part la prise en compte des scénarios climatiques et de l'horizon très lointain de projection, rendant nécessaire la définition d'un cadre spécifique.

Ces recommandations peuvent être décomposées en deux étapes :

1. Tests de matérialité : Les tests de matérialités ont pour objectif d'établir une hiérarchie entre les risques testés au regard de la stratégie de l'entreprise. Voici le processus de l'assureur pour l'évaluation de sa matérialité :
  - (a) L'assureur évalue son exposition aux risques liés à son activité en définissant le contexte de ses activités.
  - (b) L'assureur liste ensuite les impacts potentiels de ces derniers sur son activité et choisit les indicateurs étudiés (SCR, Fonds Propres...). Pour cette étape, l'assureur peut faire la distinction entre le risque physique et le risque de transition. Nous avons vu au début du chapitre quelles différences il peut y avoir entre les types de risques engendrés par les risques de transition ou physiques.
  - (c) Réaliser les matrices de risques. Pour évaluer la matérialité du risques, trois dimensions doivent être prise en compte : la probabilité, l'impact et l'horizon de temps. Voici un exemple de matrice de risque 1.18.

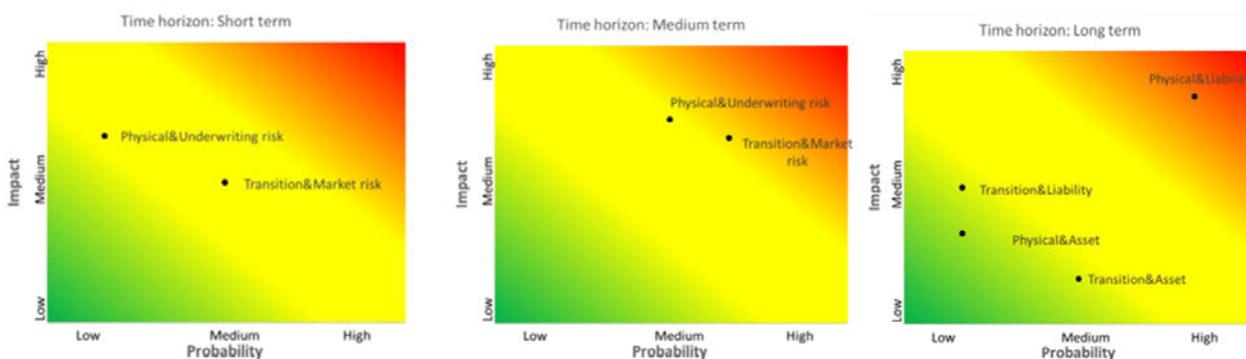


FIGURE 1.18 : Matrice de risque climatique (EIOPA, 2022)

La matrice 1.18 fait référence à trois niveaux de probabilité : faible, moyen et élevé et à trois niveaux d'impacts. Ici, le risque physique est représenté, cependant ces métriques peuvent être faites à un niveau de granularité plus forte, par exemple le risque de sécheresse ou d'inondation dans une région spécifique.

- (d) Justifier la matérialité ou non de chaque risque étudié, c'est à dire sa pertinence dans l'étude. Pour les risque matérialisés, on procède alors à une approche par scénario. Pour les non matérialisés, il faut en expliquer les raisons dans l'ORSA.
2. Une fois la matérialité du risque évalué, on procède à une approche par scénario
    - (a) Sélection de plusieurs scénarios climatiques : au minimum deux, le premier en-dessous de 2°C et de préférence pas plus de 1.5°C, ce qui correspond aux objectifs de l'Union Européenne, et un autre au-dessus de 2°C. Les scénarios pour les risques physiques sélectionnés peuvent être pris auprès du GIEC. Quant aux scénarios des risques de transition, l'assureur peut se fournir auprès de divers organismes tels que le NGFS ou l'IAE. Afin de sélectionner ces scénarios, l'EIOPA recommande de procéder en 4 étapes :
      - Choix des scénarios en fonction des besoins de la société (NGFS, IAE...).

Type d'indicateur	Indicateur	Type de risques climatiques
Bilan	SCR	Physique/Transition
	Modification de l'excès de l'actif sur passif.	Physique/Transition
	Modification de l'actif sur passif.	Physique/Transition
	Valeur stressée ou changement de prix d'actifs sélectionnés.	Transition
	Changement relatif des provisions techniques totales.	Physique
Profit	Ratio de perte.	Physique
	Impact global sur les profits et pertes de la firme.	Physique et transition
Technique	Pertes agrégées (brutes, nettes, cédées).	Physique
	Exposition au risque (somme assurée).	Physique
	Total des actifs soumis au risque de transition .	Transition
	Probabilité d'occurrence annuelle.	Physique
Direct	Emissions carbonées des investissements.	Transition

TABLE 1.2 : Les différentes métriques proposée par l'EIOPA (*Source : EIOPA (2022)*)

- Après avoir choisi le type de scénarios utilisés, on se penche sur les paramètres du dit scénario. Cela peut être les paramètres macroéconomiques tels que l'inflation, des paramètres sur la technologie comme la production de pétrole...
  - L'ambition du scénario, à prendre en compte selon les objectifs de la société.
  - Vitesse de réalisation des scénarios. Les conséquences pour la société pourront en effet être très différents si la transition se fait de façon abrupte ou plus doucement.
- (b) On traduit ces scénarios en métriques d'impacts, aussi bien pour les risques physiques que de transition. Voici les différentes métriques listées 1.2.

On peut y trouver des éléments du bilan comme le SCR, mais également des éléments plus techniques comme les sommes assurées exposées au risque physique ou encore le ratio de perte. Cette diversité de métriques permet à l'assureur de se concentrer sur différents éléments.

- (c) L'assureur utilise alors ces métriques pour évaluer l'impact du changement climatique.

Détaillons plus en détail les difficultés et réflexions relatives à la réalisation de l'ORSA climatique. Concernant les scénarios, nous avons vu lors de la sous-partie précédente que ceux du GIEC ou du NGFS pouvaient être utilisés pour la projection.

De plus, le passif et l'actif s'avèrent différents en terme de sensibilité comme vu précédemment, ce qui implique de mettre en place des approches différentes selon le passif ou l'actif.

L'ORSA climatique doit prendre en compte du risque physique ainsi que du risque de transition, et doit se projeter à différents termes : court, moyen et long. Le court terme allant de 1 à 5 ans, le moyen de 5 à 10 ans, et le long terme entre 15 et 30 ans, 2050 étant généralement la date retenue pour ce terme.

Du fait de la spécificité de chaque risque, il est important pour l'assureur de bien cibler les risques qui lui sont les plus importants. Dans chaque cas, l'historique ainsi que l'évolution de la sinistralité pourront être radicalement différents. Nous avons vu par exemple qu'un assureur fortement exposé dans la partie sud-ouest de la France s'expose à un fort risque sécheresse du fait de la composition argileuse du sol et des fortes températures. Il devra donc accorder une importance particulière à ce

risque dans ses projections.

## 1.5 Problématiques du mémoire

Comme vu précédemment, la connaissance de ses risques climatiques est un enjeu central pour un assureur. Cependant cette connaissance peut s'avérer difficile du fait du caractère différent du risque climatique, qui nécessite d'effectuer des projections dans le futur et d'avoir des éléments météorologiques afin de prévoir l'évolution des risques.

Il faut également pouvoir sélectionner les périls à modéliser selon l'exposition de l'assureur, ces derniers pouvant s'avérer très différents d'une compagnie à l'autre. Enfin il est nécessaire de réfléchir avec soin à la sélection des scénarios de projections, les résultats pouvant s'avérer drastiquement différents selon le scénario choisi.

Il est donc nécessaire de créer des bases de données météorologiques, que ce soit pour modéliser le risque actuel que pour le projeter aux différents horizons d'intérêts. Le choix des horizons et des scénarios sera déterminant car les risques climatiques évoluent très différemment selon ces paramètres.

De plus, puisque nous procédons à une projection à long terme, nous devons mettre en place des proxys afin de simuler l'évolution de l'activité de l'assureur. Ces proxys pourront présenter l'évolution de l'exposition de l'assureur, que ce soit par le nombre de biens assurés ou par leur localisation.



## Chapitre 2

# Analyse des données et création d'une base climatique

L'objectif de ce second chapitre est dans un premier temps d'expliquer quelles données assurantielles sont à notre disposition, de les analyser avec un focus particulier sur leur exposition aux risques climatiques. Une fois cela fait, nous nous penchons sur la création d'une base de données climatique. Cette base a pour objectif de nous permettre de créer un premier modèle prédisant la survenance d'un arrêté CatNat. Notre démarche est de modéliser d'abord la survenance d'un arrêté car sans cela il ne peut y avoir d'indemnisation, puis de modéliser la sinistralité associée aux arrêts.

### 2.1 Présentation des données à disposition

Pour cette étude, nous nous baserons sur deux bases de données fournies par un assureur client de KPMG. Ces bases portent sur les contrats MRH (Multirisques Habitation) dudit assureur.

#### 2.1.1 Répartition de l'exposition

Nous disposons d'une base d'exposition pour les contrats d'assurance. Penchons-nous sur la façon dont les biens sont répartis sur le territoire français. A partir du département de chaque bien, nous avons sommé les valeurs assurées totales afin d'établir un pourcentage. Le graphique 2.1 montre comment le pourcentage évolue selon le département. Pour des raisons de clarté, nous n'avons pas représenté les départements représentant moins de 0.5% de l'exposition.

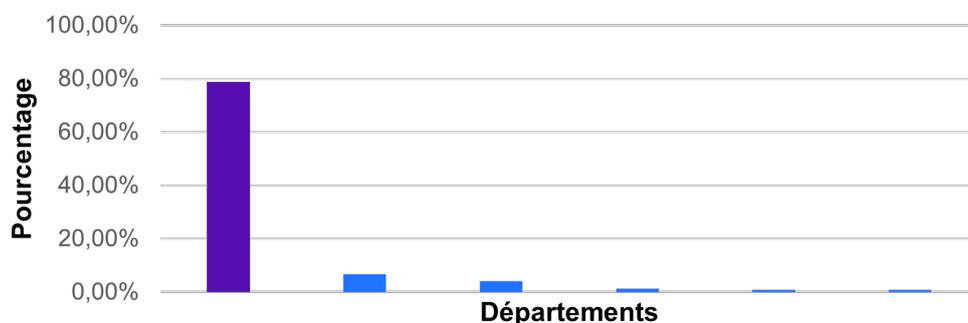


FIGURE 2.1 : Pourcentage de la représentation de la somme assurée par département en France

On constate une forte concentration de l'exposition dans un département spécifique, ainsi qu'une

exposition conséquente dans 2 autres départements. Ces trois départements seront donc le coeur de notre étude.

### 2.1.2 Le grand livre des sinistres

Nous disposons, en plus d'une base de contrats, d'une base de sinistralité recensant les divers sinistres survenus pour l'assureur de façon récente.

Intéressons-nous en particulier aux sinistres climatiques catégorisés dans la base. Nous rassemblons ces derniers sous plusieurs types de périls : inondation, sécheresse, tempête et grêle. Cela nous permet d'avoir un historique de la sinistralité climatique de la base assurantielle.

Une fois cela fait, afin de préparer notre base de sinistre, nous créons une base constituée uniquement des sinistres climatiques. Puis afin d'avoir plus de variables, nous utilisons le numéro de police du contrat afin d'ajouter les variables suivantes :

- Valeur totale du bien ;
- Prime ;
- Code INSEE de la location du bien assuré.

Si l'historique de sinistralité de cette dernière remonte jusqu'en 2000, la grande majorité de la base est comprise entre 2018 et 2022.

Comme expliqué dans la section 1.1.5 , nous nous intéressons aux risques climatiques du régime CatNat. Pour rappel, ceux-ci couvrent les risques inondations, sécheresse, séismes ou encore cyclones.

Cependant les départements qui nous intéressent ne présentent aucune exposition aux risques séismes et cyclones. En revanche le portefeuille assurantiel est exposé aux deux types de risques. Ces deux risques étant représentés de façon importante, dans un cadre ORSA il est intéressant de nous pencher sur les deux. Notre étude portera donc sur les risques sécheresse et inondation.

Cependant, pour rappel notre objectif est de pouvoir modéliser la survenance d'un arrêté CatNat. Or ce n'est pas quelque chose que nous allons pouvoir faire avec nos données actuelles, car nous n'avons pas les différents facteurs qui permettent d'expliquer la survenance d'un arrêté. Il nous faut donc créer une nouvelle base afin de regrouper ces différentes données.

## 2.2 Enrichissement de la base de données à l'aide de données *Open Data*

Puisque nous travaillons sur les risques inondation et sécheresse, nous allons avoir besoin de modéliser des arrêts CatNat puisque il ne peut pas y avoir d'indemnisation sans.

Cependant pour modéliser la survenance d'un arrêté nous avons besoin de données météorologiques, comme par exemple la température, les précipitations ou la vitesse du vent, étant donné que ce sont ces éléments qui peuvent provoquer la survenance d'un arrêté. Ces données n'étant pas fournies avec les bases de l'assureur, il nous est nécessaire d'aller les chercher en *Open Data*.

Cependant, les bases en *Open Source* sont nombreuses et parfois de mauvaise qualité. Nous allons devoir donc faire attention à trois critères.

### **Premier critère :**

Le maillage des données : en effet, les données météorologiques peuvent être soit des observations réelles récupérées par des station météorologiques soit issues de modèles climatiques. Dans le cas des observations réelles, cela veut dire que nous aurions par exemple les données issues des stations météorologiques de France métropolitaine accessibles au téléchargement qui sont au nombre de 44, ce qui s'avère très peu et fait qu'une seule station peut parfois couvrir une région entière.

## 2.2. ENRICHISSEMENT DE LA BASE DE DONNÉES À L'AIDE DE DONNÉES OPEN DATA55

Dans le cadre d'un modèle, ce dernier tente de prévoir la météo en appliquant des modèles mathématiques et physiques a des paramètres météo connus. A partir de cela, des algorithmes sont utilisés afin de prévoir des données selon un maillage plus ou moins fin, généralement répartis de façon uniforme sur le territoire. Par exemple la base du DRIAS, un site étatique qui propose des données publiques, a une maille de 8 kilomètres de côté, ce qui correspond à la présence d'un point pour un carré de 8 kilomètres de large. Voici l'exemple de comment une maille est faite 2.2.

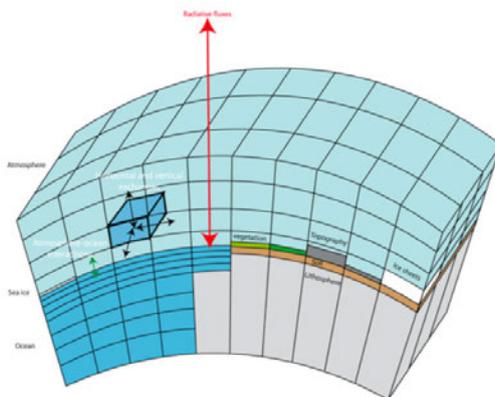


FIGURE 2.2 : Exemple de maillage (*Source : ENCYCLOPÉDIE DE L'ENVIRONNEMENT (2021)*).

Il est important que le maillage soit suffisamment fin pour pouvoir prendre en compte les variations météorologiques sur le territoire. Ainsi nous éliminons les données des stations météorologiques qui sont beaucoup trop espacées entre elles.

### Deuxième critère

La temporalité des données. Comme vu dans dans la section 1.2, le climat évolue rapidement et la France a récemment connu des événements météorologiques d'importance. Il est utile de pouvoir intégrer les données associées à ces événements pour notre modélisation. Nous souhaitons donc avoir des données récentes, idéalement jusqu'en 2022.

### Troisième critère

Le nombre de variables accessibles. Afin d'avoir une modélisation précise, il peut être utile d'avoir de nombreuses variables à disposition telles que l'humidité, les températures, les précipitations ou autres.

## 2.2.1 La base ERA5

Ces critères nous ont conduits à choisir la base ERA5, qui se base sur le modèle météo européen ECMWF (*European Centre for Medium-Range Weather Forecasts*). Cette base est disponible sur le site *Climate Data Store* (EUROPEAN COMMISSION, 2022), qui recense de nombreuses bases de données climatiques en *Open Data*. Ce site européen est issu du projet Copernicus, un programme de l'Union Européenne qui collecte des données sur l'état de la planète. Ces données peuvent servir à gérer la sécurité maritime, assurer le suivi des catastrophes naturelles...

Cette base est issue de données réanalysées, c'est à dire que le modèle utilisé ici combine des observations météo réelles avec des modèles mathématiques en utilisant des lois physiques. Ce principe est inspiré des centres de prédictions météo où régulièrement les prévisions précédentes sont combinées avec les nouvelles données disponibles pour optimiser la modélisation. Dans le cas de l'ECMWF, ce processus, appelé *data assimilation*, est effectué toute les 12 heures. De plus, puisque le modèle ne doit pas régulièrement produire des prédictions, ce dernier a plus de temps pour intégrer des observations supplémentaires.

Cette base correspond bien aux trois critères énoncés précédemment. Tout d'abord, ERA5 présente une maille assez fine de  $0,25 \times 0,25$  degrés. Au niveau de Paris, cela correspond à un carré de 18 kilomètres de large sur 27 kilomètres de long, ce qui s'avère assez fin, et au total ce sont 2310 points qui vont couvrir la France métropolitaine et les territoires situés proche des frontières française. Cette zone est comprise entre degrés de latitude et degrés de longitude.

La figure 2.3 montre au niveau européen à quoi ressemble la maille de notre base.

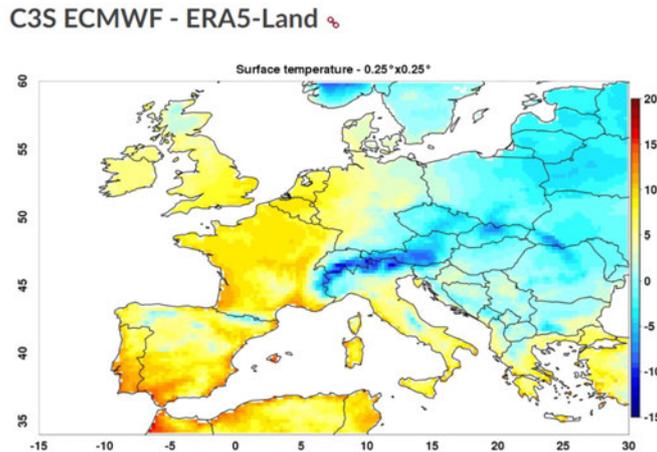


FIGURE 2.3 : Exemple de la maille de la base ERA5 (Source : PYRONEAR (2021)).

Concernant le deuxième critère, la base ERA5 est mise à jour quotidiennement avec une latence de 5 jours, ce qui nous permet d'avoir les données météorologiques les plus récentes possibles. De plus, la base ERA5 propose des observations pour chaque heure depuis l'année 1940, ce qui représente une quantité de données très importante.

Enfin, la base ERA5 comporte de très nombreuses variables. On peut trouver des données de température, de vitesse du vent ou encore d'humidité, mais également des indices sur la présence de neige, sur les radiations UV ou même sur les vagues océaniques.

Nous nous penchons donc sur les données par heure. Nous sélectionnons alors les variables pour chaque jour à 12h, et ce jusqu'en 2018, ce qui correspond à l'horizon de temps jusqu'auquel nous avons les bases de l'assureur. Puisque nous souhaitons prédire les arrêtés CatNat, nous allons prendre chaque variable par trimestre, puisque ces derniers sont émis tous les 3 mois.

Pour constituer notre base, nous avons décidé de nous attarder sur les variables suivantes :

- La température. Dans un cadre climatique, avoir les températures est nécessaire, puisque la projection selon les scénarios du GIEC implique d'appliquer des chocs aux niveaux de températures selon le scénario choisi. Nous allons prendre le minimum, la moyenne et le maximum de chaque trimestre. Dans un cadre sécheresse, la température maximale est importante, puisqu'elle pourra déclencher une forte sécheresse, voir le mémoire CORNEC (2018). La moyenne peut également s'avérer utile car de fortes chaleurs sur de longues périodes sont également facteur de sécheresse, bien que cela soit atténué par le fait que nous prenons des données trimestrielles. Enfin le minimum pourra fournir un complément d'informations par sa potentielle corrélation aux températures moyenne comme maximale.
- Les précipitations. Celles-ci ont un impact sur les arrêtés inondations, puisque de fortes pluies peuvent entraîner des inondations par ruissellement, et que la période de retour des précipitations est un critère de reconnaissance. Puisque ce critère survient logiquement lors de précipitations

extrêmes, nous prenons encore les précipitations maximales au trimestre en plus des précipitations moyennes. Cet élément est confirmé par le mémoire BOYEAU (2022). Nous prenons les précipitations en mètres par jour.

- L'humidité spécifique. Cette variable est corrélée avec la sécheresse et l'inondation, puisque l'humidité est corrélée avec la température de l'air et sa teneur en vapeur d'eau. Nous en prenons la moyenne et la valeur maximale sur les trois mois.
- La vitesse du vent à 10 mètres de hauteur, exprimée en mètres/seconde. Cette dernière impacte notamment les précipitations en les déplaçant mais peut également faire diminuer la température ressentie. Nous en prenons la moyenne sur les trois mois.

Cependant, nous pouvons, à partir de certaines variables, en créer d'autres qui pourront s'avérer intéressantes. Ainsi, nous avons vu dans la section 1.1.5 que les précipitations cumulées sur différentes périodes sont critères de reconnaissance CatNat pour le péril inondation. Nous pouvons donc faire la manipulation suivante : nous prenons les précipitations totales pour chaque heure, et nous les sommes sur 12, 24, 48 et 72 heures. Nous en prenons alors le maximum, ce qui nous permet d'avoir le maximum des précipitations cumulées pour chaque trimestre, soit  $4 \times 4 = 16$  variables.

Nous allons nous concentrer sur la période courant de 2018 à 2022. Cette période a été choisie du fait de l'historique dont nous disposons concernant l'assureur sur lequel est réalisé l'étude. En effet, passé 2018, l'historique de sinistralité se fait plus sporadique et il nous manque de nombreuses données. Puisque nous voulons projeter la sinistralité de ce dernier, il est donc dans notre intérêt de prendre une période météorologique correspondant aux sinistres observés.

De plus, comme vu dans la section 1.2, il y a une forte augmentation de la sinistralité climatique depuis le milieu des années 2010, qui s'accompagne d'années exceptionnelles comme l'an 2018 pour la sécheresse ou 2021 pour l'inondation. Notre objectif est d'avoir une vision se basant sur les années récentes du fait de la forte évolution du climat.

Nous obtenons donc une base constituée de 51 variables, avec à chaque fois la température minimale, moyenne et maximale, l'humidité moyenne et maximale, les précipitations moyennes maximales et moyennes et la vitesse du vent moyenne pour chacun des quatre trimestres, ce qui fait 32 variables auxquelles nous avons ajouté une variable année allant de 2018 à 2022 afin d'identifier à quelle année la variable se situait, et les variables Latitude/Longitude afin d'identifier le point auquel les données sont reliées.

La matrice comporte donc 35 colonnes pour 11 550 lignes (2310 points pour chaque année entre 2018 et 2022, ce qui nous donne  $2310 \times 5 = 11\,550$ ). Cependant cela n'est pas entièrement satisfaisant. En effet, afin de faire notre modélisation, nous voulons avoir une base où chaque commune se voit associer les données météorologiques.

Pour cela, il nous faut d'abord les différentes communes de France, ainsi que leurs coordonnées GPS. Pour cela, nous allons utiliser la base GEOFLA Communes. Cette base, créée par le site géoservices du gouvernement, décrit le découpage administratif de la France (figure 2.4).

En exportant cette base, nous pouvons alors avoir accès à des variables telles que le nom de la commune, son numéro INSEE ou son département, mais également ses coordonnées GPS. Notre objectif est désormais de relier chaque commune aux données météo qui s'en approchent.

Nous allons pour cela utiliser la méthode des barycentres pour 3 points. Voir le site INSTITUT DES MATHÉMATIQUES POUR LA PLANÈTE TERRE (2021) pour plus d'informations sur l'utilisation des barycentres dans un cadre climatique. Cette méthode consiste, pour un point donné, à lui associer la moyenne pondérée des trois points les plus proches de lui. Notre processus sera donc le suivant :

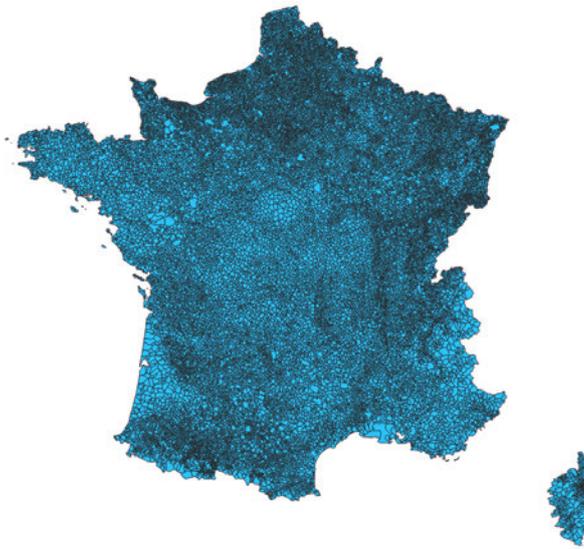


FIGURE 2.4 : Représentation des communes françaises d'après la base GEOFLA.

- Pour chaque commune, calcul des distances entre ses coordonnées et les différents points recueillis dans la base ERA5.
- Classement des points par distance et sélection des 3 plus proches.
- La commune se voit associer les données météorologiques desdits points, selon la formule suivante

$$V = V_a \times \frac{d_a}{d_a + d_b + d_c} + V_b \times \frac{d_b}{d_a + d_b + d_c} + V_c \times \frac{d_c}{d_a + d_b + d_c},$$

avec :

- $V_a$  les valeurs des données météorologiques du point le plus proche de la commune ;
- $V_b$  les valeurs des données météorologiques du deuxième point le plus proche de la commune ;
- $V_c$  les valeurs des données météorologiques du troisième point le plus proche de la commune ;
- $d_a$  la distance entre la commune et le point le plus proche ;
- $d_b$  la distance entre la commune et le deuxième point le plus proche ;
- $d_c$  la distance entre la commune et le troisième point le plus proche.

Faire une moyenne par barycentre permet d'avoir une plus grande diversité de valeurs que si nous avons simplement associé chaque commune au point le plus proche.

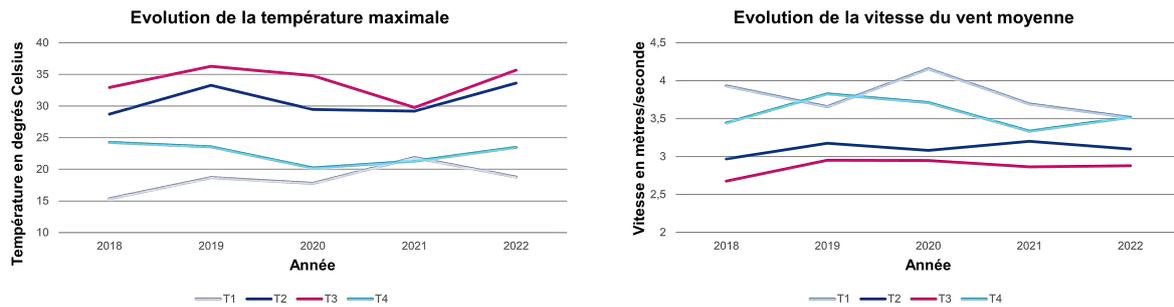
La base GEOFLA recense 35798 communes sur le territoire de la France métropolitaine. Nous obtenons donc une base de 178 990 lignes et variables, 16 météorologiques, 1 pour le nom de la ville, 1 pour son numéro INSEE, et enfin 1 dernière pour l'année.

Dans notre base, la structure est donc la suivante : Un individu : Une année  $\times$  Une commune.

Représentons les différentes variables que nous avons dans notre base afin de visualiser leur évolution selon les années et les trimestres. Voici l'évolution des températures, humidité et précipitations maximales ainsi que de la vitesse du vent (figures 2.5 et 2.6) :

De façon logique, le troisième trimestre, qui concerne notamment août et juillet, comporte les températures maximales les plus élevées. Nous observons également une nette hausse lors de l'année 2022, année exceptionnelle au niveau de ses chaleurs estivales. Pour ce qui est de la vitesse du vent,

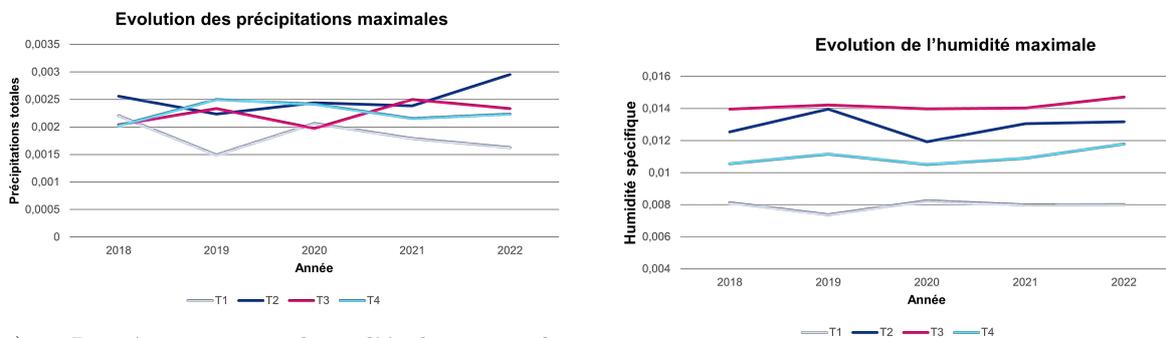
## 2.2. ENRICHISSEMENT DE LA BASE DE DONNÉES À L'AIDE DE DONNÉES OPEN DATA<sup>59</sup>



(a) Représentation de l'évolution des températures maximales en France métropolitaine.

(b) Représentation de l'évolution de la vitesse moyenne du vent en France métropolitaine.

FIGURE 2.5 : Représentation des évolutions de la vitesse moyenne du vent et des températures maximales en France métropolitaine.



(a) Représentation de l'évolution des précipitations maximales en France métropolitaine.

(b) Représentation de l'évolution de l'humidité maximale en France métropolitaine.

FIGURE 2.6 : Représentation des évolutions des précipitations maximales et des humidité maximales en France métropolitaine.

cette dernière varie de façon nette selon les trimestres et années. Ici, le vent est plus rapide lors des premier et deuxième trimestres, et plus lent lors des trimestres les plus chauds.

Cependant ces variables ne sont pas suffisantes, car nous avons par exemple vu pour le péril sécheresse que la survenance de CatNat sécheresse dépend de l'exposition aux argiles, données qui sont fournies par le BRGM.

### 2.2.2 Les données du BRGM

Le BRGM, déjà évoqué dans la section 1.1.2, est une institution d'étude géologique ayant émis une carte (figure 1.3) servant de référence à l'exposition aux argiles en France. Sur le site Géo Risques (MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE, 2022), il nous est ainsi possible de récupérer les données montrées sur ladite carte.

La structure est la suivante : pour chaque commune, nous associons 8 variables, qui correspondent à son exposition aux argiles. Pour rappel, le BRGM distingue quatre niveaux d'expositions différents : fort, moyen, faible et aucune.

Nous avons d'abord le pourcentage de la commune exposée à chaque niveau, ce qui fait quatre variables différentes, puis la superficie de la commune associée à chaque niveau, ce qui nous donne les quatre suivantes.

Ces variables sont très importantes pour la sécheresse, puisqu'une commune majoritairement classée en "aucune exposition" aux argiles ne sera pas reconnue en arrêté sécheresse. Le graphique 2.7 montre les données d'exposition pour la France.

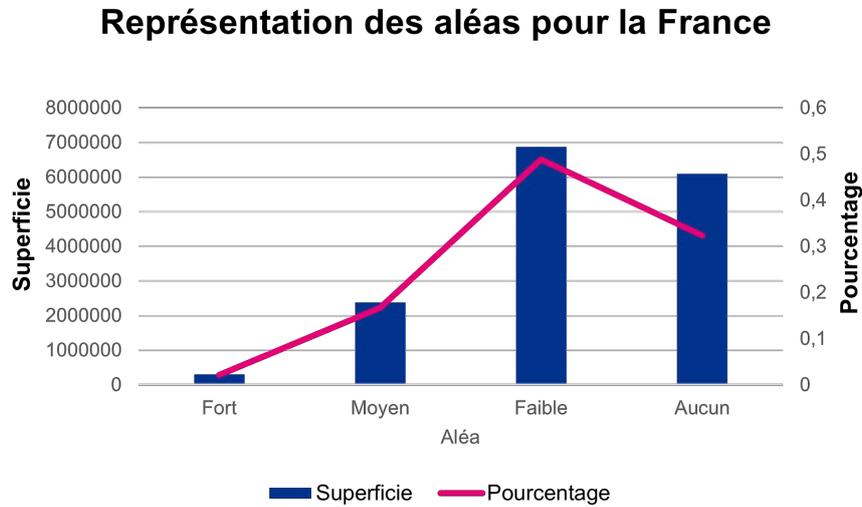


FIGURE 2.7 : Représentation des données BRGM pour la France métropolitaine, avec la surface en mètres carrés et le pourcentage.

Cependant l'exposition aux argiles n'est pas le seul facteur expliquant un arrêté sécheresse.

### 2.2.3 Indice d'humidité des sols(SWI)

Comme vu lors de la section 1.1.5, le SWI (*Soil Wetness Index*) est un critère de reconnaissance en arrêté CatNat sécheresse qui est un indice mesurant l'humidité des sols. Son impact sur la survenance d'arrêté CatNat sécheresse est donc manifeste, mais il pourrait également s'avérer utile lors de la modélisation de l'inondation, puisque qu'un sol très humide pourra être corrélé aux inondations. Son calcul prend la forme suivante

$$SWI = \frac{(\Theta_{fc} - \Theta_{wi})}{(\Theta - \Theta_{wi})},$$

avec :

- $\Theta$  la quantité d'eau contenue dans le sol jusqu'à une profondeur de deux mètres,
- $\Theta_{wi}$  la quantité d'eau contenue dans le sol au point de flétrissement, c'est à dire le point en-dessous duquel l'eau ne peut plus être extraite par les plantes,
- $\Theta_{fc}$  la capacité au champ, c'est à dire le point à partir duquel l'eau ne peut plus être stockée dans le sol et commence à s'infiltrer dans les couches inférieures.

Le SWI est disponible en *Open Data* sur le site de Météo France (MÉTÉO FRANCE, 2022). La base de données fournie est construite à partir du modèle SIM et est employée par la commission interministérielle CatNat. Cette base a donc pour rôle de fournir les données nécessaires à la prise de décision en reconnaissance CatNat et est donc très importante. Le modèle SIM fournit les données du SWI pour une maille de 8 kilomètres de côté, ce qui est très précis et représente un total de 8931 points pour la France métropolitaine.

Puisque le critère de reconnaissance est la valeur du SWI uniforme sur le mois, nous conservons cette structure et récupérons les valeurs du SWI par mois de 2018 à 2022 car c'est cette période qui nous intéresse. Comme pour les données météorologiques, nous avons accès aux coordonnées des points, ce qui nous permet d'associer à chaque commune une valeur de SWI. Le maillage étant très fin, il n'est pas vraiment nécessaire d'utiliser une méthode de barycentre.

Nous avons donc ajouté 12 variables, les valeurs du SWI pour chaque mois par commune, à notre base. Nous avons désormais de nombreuses variables d'importance pouvant avoir un impact sur la décision en arrêté sécheresse telles que les données du BRGM ou le SWI, mais nous n'avons pour les arrêtés inondations que des variables relatives aux précipitations, ce qui peut sembler léger. Une façon intéressante d'avoir plus d'informations sur l'exposition aux inondations peut être de récupérer des données hydrographiques.

### 2.2.4 Les données hydrographiques

Afin de pouvoir prédire au mieux les arrêtés inondations, nous avons rassemblé précédemment les données des précipitations maximales afin de prédire au mieux les inondations par ruissellement pouvant en résulter. Cependant il nous reste également des informations à prendre sur les inondations par débordement, qui sont le résultat de débordement de cours ou de plans d'eau. En effet, ce sont ces éléments liés aux précipitations extrêmes ou à des fontes de neige anormales qui entraînent les inondations par débordement. Pour cela, nous allons nous tourner vers la base de données TOPAGE (SANDRE, 2019). Cette dernière a été créée par la collaboration de l'IGN (institut national de l'information géographique et forestière) et de l'OFB (Office Français de la Biodiversité) et est disponible sur le site du SANDRE (Service d'Administration Nationale des Données et Référentiels sur l'Eau). Cette dernière se décompose en plusieurs jeux de données dont voici ceux que nous allons utiliser :

- Cours d'eau, qui recense les différents cours d'eau français, que ce soit les rivières, les fleuves ou même les canaux ;
- Plans d'eau. Un plan d'eau est défini dans la base comme une surface d'eau de surface, qu'elle soit d'origine humaine ou naturelle. Pour être comptabilisé dans la base, il est nécessaire que le plan d'eau ne soit pas temporaire, c'est à dire qu'il n'ait pas été créé à la suite de fortes précipitations par exemple ;
- Les bassins hydrographiques. La France métropolitaine est divisée en 7 bassins hydrographiques, qui sont :
  1. le bassin Artois-Picardie ;
  2. le bassin Rhin-Meuse ;
  3. le bassin Seine-Normandie ;
  4. le bassin Loire-Bretagne ;
  5. le bassin Adour-Garonne ;
  6. le bassin Rhône-Méditerranée ;
  7. le bassin Corse. Comme expliqué lors de la section 1.2.2, ces bassins peuvent s'avérer avoir une importance significative pour le risque inondation.
- Tronçons hydrographiques. Un tronçon hydrographique se définit comme le "découpage du réseau hydrographique homogène par ses attributs qui peut composer un cours d'eau ou traverser un plan d'eau" (SANDRE, 2019). La base Tronçon hydrographique nous intéresse notamment car elle propose plusieurs classes pour les différents cours d'eau, ces derniers étant divisés selon leur largeur ;

Voici la carte 2.8 montrant les cours et les plans d'eau recensés dans la base TOPAGE :

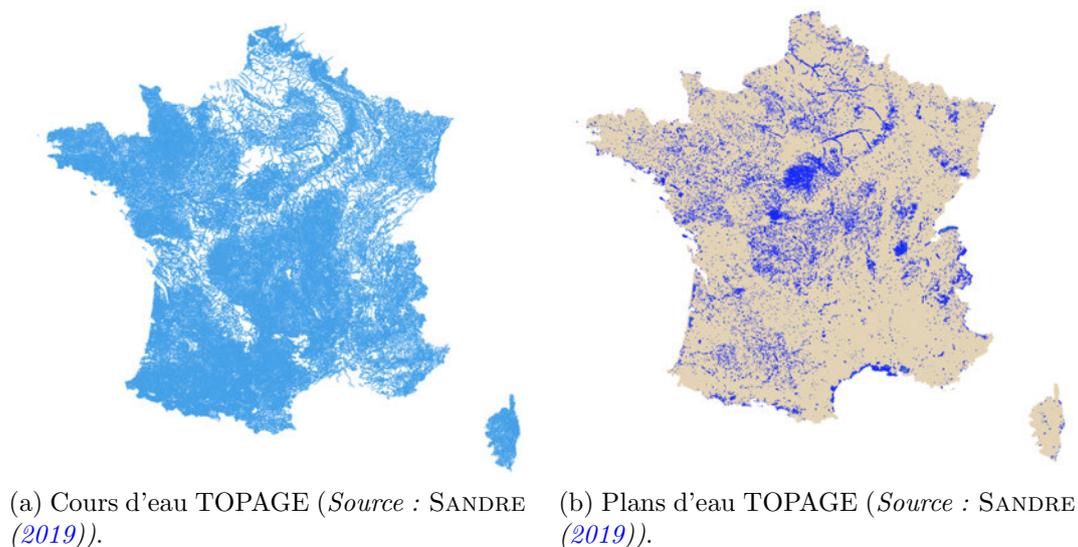


FIGURE 2.8 : Représentation des données TOPAGE

Nous allons créer trois variables :

- le nombre de plans d'eau par commune ;
- la présence ou non d'un fleuve large (largeur supérieure à 50 mètres) sur le territoire de celle-ci ;
- la présence d'un fleuve moyen (entre 15 et 50 mètres) sur le territoire de celle-ci.

Les cours d'eau large et moyens sont quant à eux représentés sur la figure 2.9.

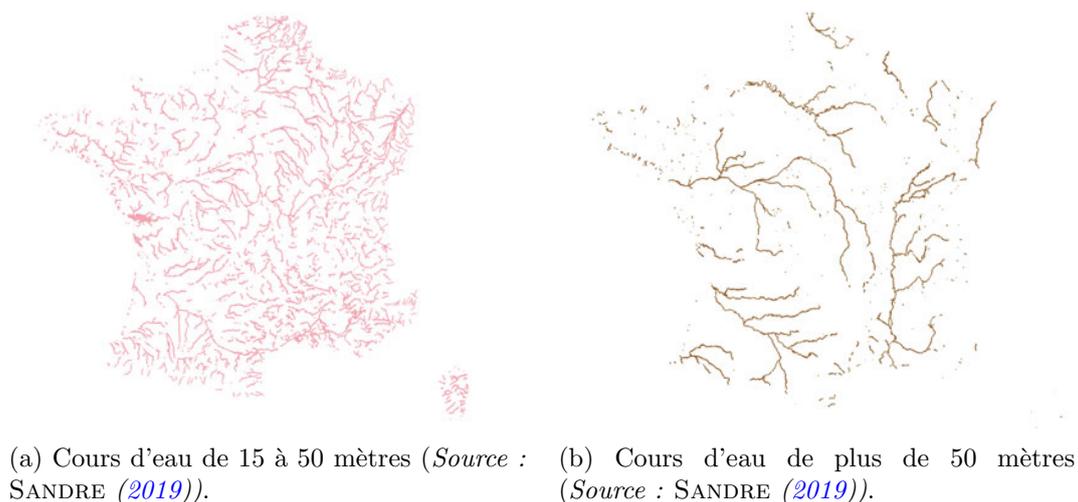


FIGURE 2.9 : Représentation des cours d'eau larges et moyens TOPAGE

Les variables relatives à la présence des cours d'eau moyens et large sont importantes. En effet, ce sont surtout ces cours d'eau qui vont avoir un impact significatif sur la sinistralité de par leur débordement. Par exemple, on peut citer les dernières inondations par débordement les plus graves que sont le débordement de l'Aude en octobre 2018 ou les inondations de mai-juin 2016 où le fleuve le Loing avait été placé en vigilance rouge inondation.

## 2.3 Arrêtés CatNat sécheresse et inondation

Nous avons désormais une base complète de données *Open Source* constituée dans le but de prédire des arrêtés CatNat. Cependant il manque encore à cette base les variables à expliquer, à savoir la présence des arrêtés CatNat sécheresse et inondation.

Ces derniers peuvent être récupérés sur le site de la CCR, qui recense les différents arrêtés parus au Journal Officiel. Nous allons donc télécharger toutes les bases disponibles jusqu'en 2018 pour les périls sécheresse, inondations remontée de nappe et inondations et/ou coulées de boue.

Les fichiers nous indiquent plusieurs éléments : la date de l'évènement, la date de l'arrêté, la date de la parution dudit arrêté au Journal Officiel, le péril et enfin la décision : reconnue ou non reconnue.

Tout d'abord nous allons retirer toutes les décisions ayant conclues à une non-reconnaissance en arrêté CatNat, puisque que dans ce cas il n'y a pas d'indemnisation. Ensuite nous divisons les arrêtés CatNat par péril et par année.

Nous créons alors deux variables binaires :

- Arrêté inondation : vaut 0 en l'absence d'arrêté et 1 en présence d'arrêté.
- Arrêté sécheresse : vaut 0 en l'absence d'arrêté et 1 en présence d'arrêté.

De façon globale (de 2018 à 2022), 8,5 % des communes ont été reconnues en arrêté sécheresse et 5,05 % en arrêté inondation. La figure 2.10 montre le nombre de communes reconnues et le pourcentage qu'elle représente pour toutes les communes françaises par an.

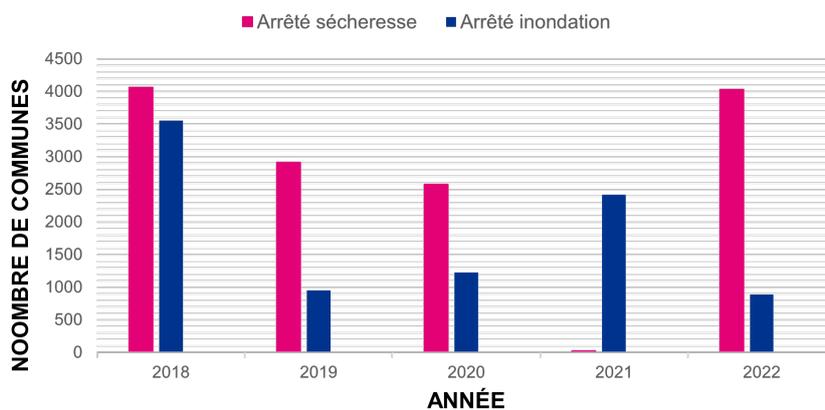


FIGURE 2.10 : Historique du nombre d'arrêtés par an

Les années 2018 et 2022 se détachent fortement, la première ayant compté un nombre important d'arrêtés des deux types et la seconde en raison d'un nombre d'arrêtés sécheresse qui sera probablement record (actuellement, certaines reconnaissances d'arrêtés sécheresse n'ont pas encore été faites). Cela est cohérent avec la sécheresse de grande ampleur ayant eu lieu durant l'été 2022. Regardons comment les différentes variables obtenues peuvent influencer sur la présence d'arrêté inondation ou sécheresse. En sélectionnant les variables du SWI, les températures ainsi que les données du BRGM, nous obtenons ces corrélations (figure 2.11).

Cela n'est pas très significatif, cependant on observe une claire corrélation positive avec les températures maximales et négative avec le SWI, ce qui semble logique : un sol sec déclenchera en moyenne plus d'arrêtés sécheresse.

Concernant l'inondation, nous avons les corrélations suivantes après sélection de variables pouvant être corrélées avec la survenance d'un arrêté 2.12.

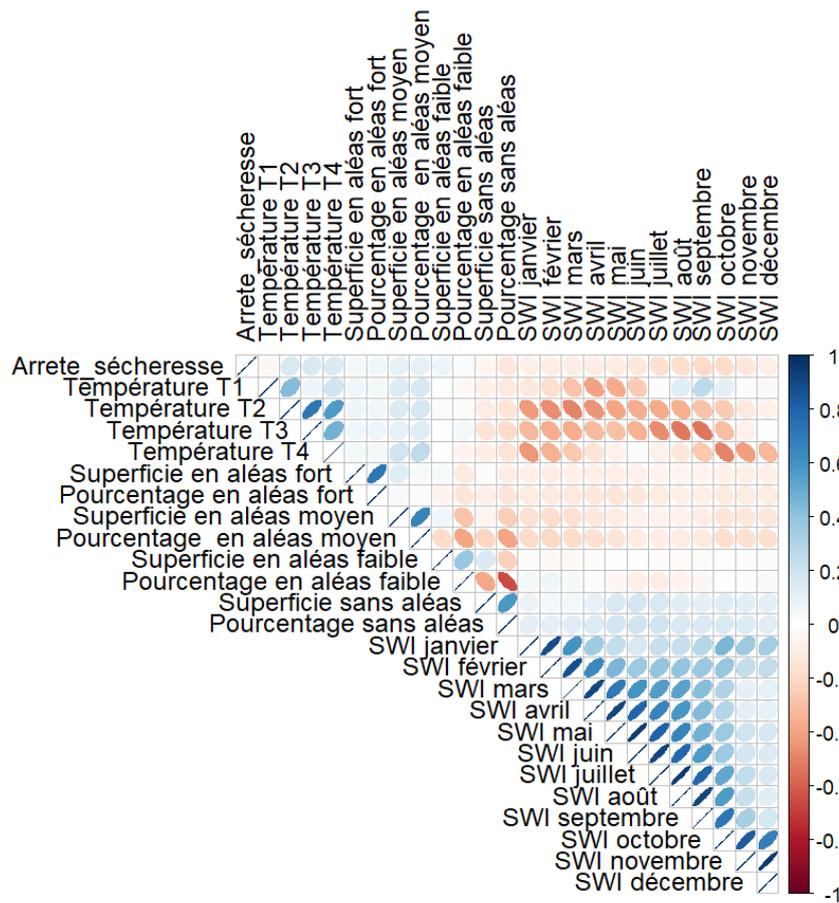


FIGURE 2.11 : Corrélations entre les différentes variables sécheresse et les arrêts sécheresse.

Ici la corrélation entre les différentes variables et les arrêts inondations n'est pas pas très significative, on observe une légère corrélation entre la présence d'un arrêté et la présence d'un fleuve large ou moyen, mais guère plus.

Regardons plus en détail les densités des variables en fonction de la présence d'un arrêté. Notre démarche est la suivante : nous divisons la base en deux selon qu'il y ait un arrêté ou non, puis nous traçons les histogrammes sur un même graphique afin de les comparer. Ainsi, pour la sécheresse, nous pouvons regarder les températures de la figure 2.13.

La différence observée est très significative, avec pour certains trimestres des différences allant jusqu'à 3 degrés, ce qui est considérable.

En ajoutant ces deux variables, nous obtenons donc notre base climatique finale. Cette dernière comporte donc 178 990 lignes correspondant aux communes pour cinq années, et 54 variables dont :

- Le nom, le code INSEE et le département de la commune ;
- La vitesse du vent, la température moyenne et maximale, les précipitations moyennes et maximales et l'humidité moyenne et maximale pour les quatres trimestres de l'année, soit variables ;
- Les variables présence d'un arrêté sécheresse et présence d'un arrêté inondation ;
- Les données du SWI pour chaque mois de l'année, soit 12 variables ;
- Les données du BRGM, soit 8 variables ;

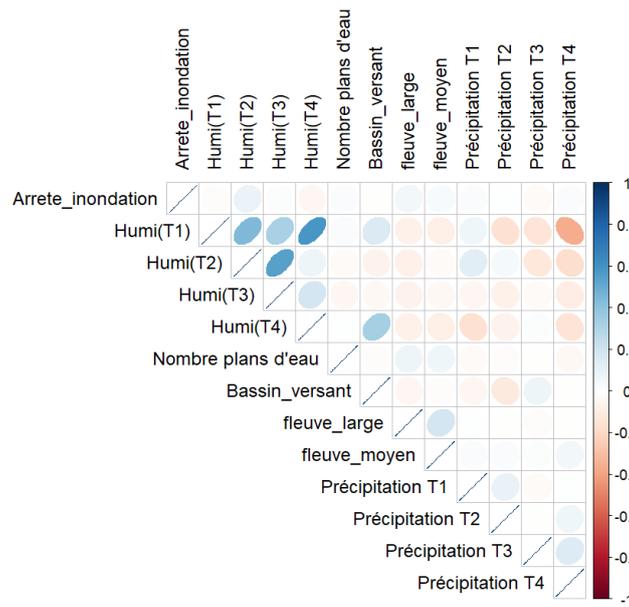


FIGURE 2.12 : Corrélations entre les différentes variables inondation et les arrêts inondations.

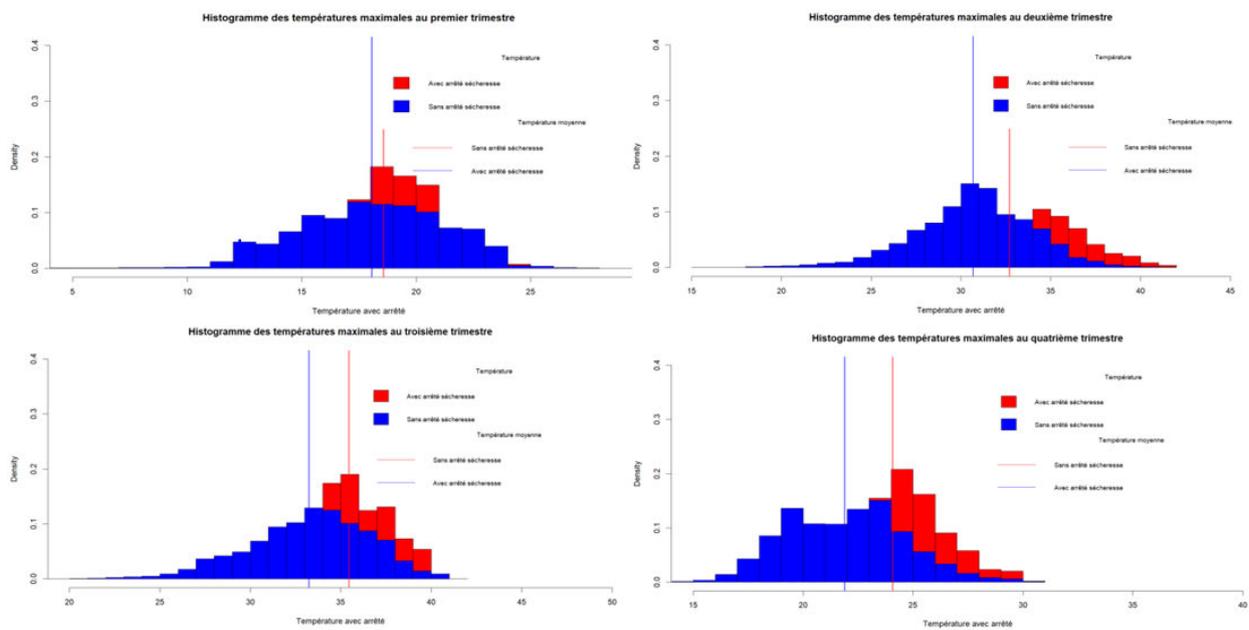


FIGURE 2.13 : Différence entre les températures maximales par trimestre selon la présence d'un arrêté

- Les données hydrographiques qui sont le nombre de plans d'eau, le bassin versant, la présence d'un fleuve large et la présence d'un fleuve moyen ;
- L'année concernée, allant de 2018 à 2022.

Pour les inondations, nous pouvons nous intéresser aux corrélations avec les précipitations observées ainsi que les variables hydrographiques.

Ainsi, nous pouvons regarder le pourcentage de communes présentant un fleuve large ou moyen selon qu'il y ait un arrêté sécheresse ou non :

- 17,1 % des communes avec arrêté inondation comportent un fleuve large 9,8 % pour les communes sans arrêtés.
- 33,8 % des communes avec arrêté inondation comportent un fleuve moyen contre 24,2 % pour les communes sans arrêtés.

La différence entre ces pourcentages est très considérable et souligne bien le fait que les communes comportant un fleuve moyen ou large sont bien plus sujettes aux arrêtés inondations.

A partir de ces résultats, nous pouvons proposer plusieurs idées de traitements sur nos variables :

- Interactions : nous observons une forte corrélation entre certaines de nos variables trimestrielles, comme par exemple l'humidité des deuxièmes et troisièmes trimestre, mais aussi entre des données géologiques comme le SWI. Par exemple, le SWI de janvier est très corrélé avec le SWI février, celui de février avec celui de mars et ainsi de suite. Cela s'explique par le fait que le SWI étudié est une moyenne glissante sur trois mois, le SWI de deux mois proches auront donc des valeurs similaires.

### 2.3.1 Récapitulatif des données *Open Source*

Nous effectuons maintenant un récapitulatif des différentes données *Open Source* que nous avons récupérées pour notre étude. Nous avons donc 58 variables dans notre base :

- 28 variables météorologiques : 4 pour le vent, 8 pour l'humidité, 8 pour la température et 8 pour les précipitations ;
- 8 variables sur la teneur en argile des sols ;
- 12 variables d'indice d'humidité des sols ;
- 4 variables hydrologiques ;
- 4 variables d'indications sur l'année, le nom de la commune, son code INSEE, et son département ;
- 2 variables sur la présence d'arrêté CatNat.

Le résumé de ces dernières est présenté dans la table 2.1.

Nom	Description
Température	La température à 2 mètres au-dessus du sol, exprimée en degrés Celsius. Nous prenons la température maximale et moyenne pour le trimestre.
Précipitations	Les précipitations totales moyennées sur le trimestre. Ces dernières sont exprimées en kilogramme par mètre carré et par seconde. Nous prenons ici également le maximum et la moyenne du trimestre. En effet, quand nous nous intéressons aux inondations par ruissellement, ce sont les précipitations extrêmes qui vont venir saturer le sol et provoquer des inondations.
Humidité	L'humidité relative de l'air. Nous avons pris la moyenne et le maximum par trimestre. Cette dernière se définit comme la masse d'eau dans l'air sur la masse d'air humide. Cette mesure a pour avantage de ne pas dépendre de la température ni de l'altitude, ce qui est le cas pour d'autres mesure d'humidité telle que l'humidité absolue. Son approximation est fournie par la formule suivante $q \approx \frac{0,622e}{p-0,378e}$ avec p la pression de l'air exprimée en Pascal et e la pression partielle de la vapeur d'eau également exprimée en Pascal.
Vent	La vitesse moyenne par trimestre du vent à 10 mètres de hauteur, mesurée en mètres/seconde. Nous avons pris ici la moyenne en raison de la corrélation plus faible entre la vitesse extrême du vent et les inondations/sécheresse.
Arrêté	La présence d'un arrêté sécheresse ou inondation dans la commune pour une année spécifique.
Données BRGM	Les données d'exposition aux argiles du BRGM. Nous avons un total de huit variables : -Superficie en aléas fort, exprimée en mètre carrés ; -Pourcentage de la commune en aléas fort ; -Superficie en aléas moyen ; -Pourcentage de la commune en aléas moyen ; -Superficie en aléas faible ; -Pourcentage de la commune en aléas faible ; -Superficie en sans aléas ; -Pourcentage de la commune en sans aléas ;
SWI	La valeur du SWI telle que modélisée par Météo France. Nous la prenons pour chaque mois, ce qui nous donne 12 variables.

TABLE 2.1 : Les données *Open Source* collectées



## Chapitre 3

# Modélisation de la survenance d'arrêts CatNat et de la sinistralité associée

L'objectif de cette partie va être de modéliser l'augmentation de la sinistralité sécheresse et inondation pour l'assureur étudié. Notre modélisation suit les étapes suivantes :

1. Sélection des hypothèses, c'est à dire : les périls d'intérêts évoqués dans la section 2.1.2 (sécheresse et inondation), les scénarios climatiques considérés, et enfin les différents horizons de projection.
2. Récupération dans la base de données "Grand Livre des sinistres" présentée dans la section 2.1.2 de la sinistralité moyenne pour chaque péril et du nombre de sinistres entraînés par chaque arrêté.
3. A partir de la base climatique constituée dans la section 2.2, nous créons un modèle de prédiction des arrêts sécheresse et un modèle de prédiction des arrêts inondation.
4. Nous créons ensuite une base de données projetées aux horizons d'intérêts.
5. Nous appliquons les modèles créés à l'étape 2 sur lesdites bases.
6. A partir de ces résultats, nous calculons la sinistralité projetée pour chaque péril en faisant la multiplication suivante  $P \times N \times S$ , avec
  - P, la probabilité d'arrêt CatNat ;
  - N, le nombre de sinistres pour l'assureur quand un arrêté survient ;
  - S, la sinistralité moyenne pour l'assureur, i.e. combien coûte un sinistre climatique en moyenne.

Nous étudions une moyenne de la sinistralité, ce qui nous servira alors à la réalisation de l'ORSA climatique. Cette étape sera traitée lors du chapitre 4.

Afin d'illustrer les étapes que nous venons de présenter nous nous référons au graphique 3.1.

Une fois cela fait, nous nous interrogerons sur les conséquences d'une telle hausse sur le bilan assurantiel.

### 3.1 Sélection des scénarios climatiques à employer

Maintenant que nous avons vu et analysé nos données, nous devons nous intéresser à comment sélectionner les scénarios climatiques que nous utiliserons pour la projection dans le futur de nos données climatiques.

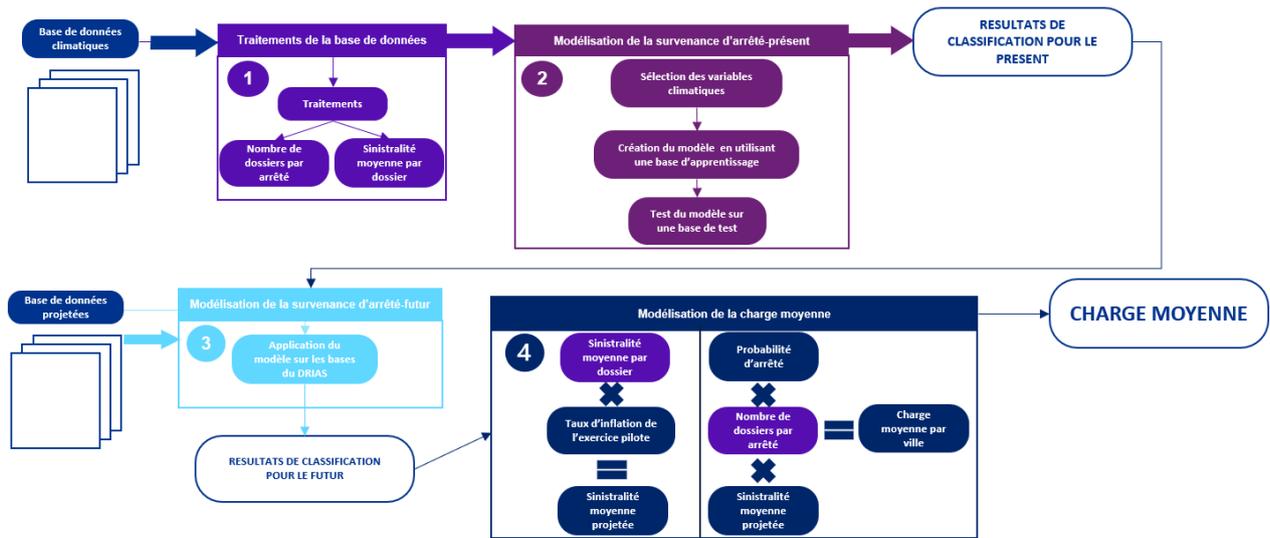


FIGURE 3.1 : Représentation du processus suivi pour le calcul de l'évolution de la charge moyenne

L'exercice pilote de l'ACPR de 2020 a utilisé le scénario RCP 8.5, émis par le GIEC lors de ses rapports. Pour rappel, le scénario RCP 8.5 est le plus pessimiste de tous, avec une aggravation des émissions de gaz à effets de serre et une hausse de  $+3,7^{\circ}\text{C}$  d'ici à 2100. Pour 2050, la hausse de température liée au 8.5 est comprise entre  $1,4$  et  $2,6^{\circ}\text{C}$ .

Le choix du scénario RCP 8.5 pouvait s'expliquer par la volonté d'utiliser le scénario le plus extrême afin d'avoir les impacts les plus significatifs, à l'actif comme au passif.

Cependant l'exercice pilote actuel, lancé en 2023, va utiliser pour le risque physique le scénario RCP 4.5, émis de même par le GIEC.

Le scénario 4.5 est considéré comme un scénario intermédiaire et le plus probable de tous. Pour rappel, ce scénario suppose d'abord une augmentation des émissions jusqu'en 2035, une stabilisation et enfin une baisse à partir de 2050. Sa trajectoire est visible sur le graphique 1.4. Le graphique 3.2 permet d'avoir une comparaison directe entre les deux scénarios.

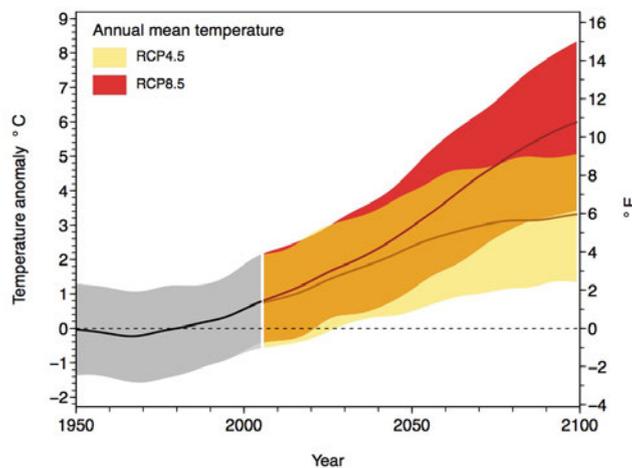


FIGURE 3.2 : Représentation des deux scénarios, 4.5 et 8.5, côte à côte

Comme dit précédemment, la différence s'effectue vraiment à partir de l'horizon 2050, qui sera l'horizon qui va nous intéresser pour cette étude. Cependant, même si la différence n'est pas encore majeure, l'impact n'est pas négligeable, comme le montre par exemple la figure 3.3 qui compare les inondations des deux scénarios. Ce graphique est inclus par l'ACPR dans son rapport sur les hypothèses à utiliser pour l'exercice pilote.

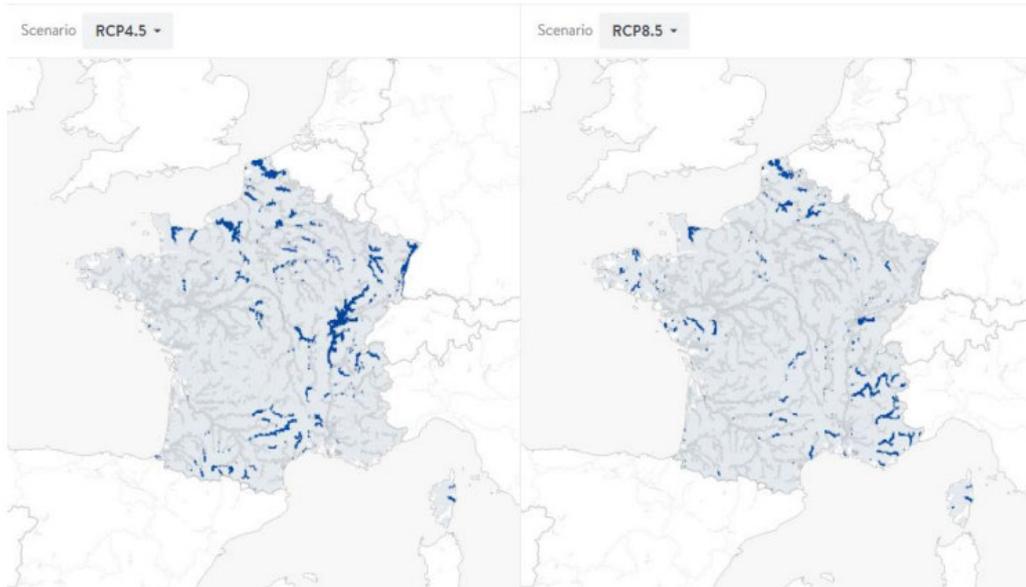


FIGURE 3.3 : Projections de la fraction du territoire exposée annuellement à des inondations fluviales selon les scénarios RCP 4.5 et 8.5 à horizon 2050

Nous allons donc nous concentrer sur les scénarios RCP 4.5 et 8.5, ce qui nous permettra d'être cohérents avec les recommandations de l'EIOPA, qui demande à ce que soient considérés des scénarios supérieurs et inférieurs à  $+2^{\circ}\text{C}$ , ainsi qu'avec les exercices pilotes de l'ACPR.

## 3.2 Modélisation de la survenance d'arrêtés CatNat

Notre objectif est désormais de modéliser la survenance d'un arrêté CatNat inondation ou sécheresse dans une commune à partir de la base climatique précédemment constituée. Nous devons donc prédire une variable catégorielle pouvant prendre la valeur 0 ou la valeur 1.

Pour cela, nous allons tester plusieurs types de modèles : GLM (Modèle Linéaire Généralisé), *Random Forest* et *XgBoost* (*Extreme Gradient Boosting*). Commençons par les décrire.

### 3.2.1 Le Modèle Linéaire Généralisé

Le Modèle Linéaire Généralisé correspond à une généralisation de la régression linéaire. Présentons-en les grands principes que nous pouvons retrouver dans le cours universitaire OBAKRIM et al. (2022).

La régression linéaire consiste en l'établissement d'une relation linéaire entre une variable cible, ici la variable Présence d'un arrêté, et une ou plusieurs variables explicatives, qui peuvent concerner le reste de notre base.

Cela se traduit par une équation de la sorte

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_n X_{i,n} + \epsilon_i,$$

avec :

- $X_{i,1}, X_{i,2}, \dots, X_{i,n}$  les différentes variables explicatives pour la  $i$ ème observation ;
- $Y_i$  la  $i$ ème observation pour la variable cible ;
- $\beta_1, \beta_2, \dots, \beta_k$  les paramètres du modèle linéaire ;
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$  l'erreur entre la somme des variables explicatives et la variable cible.

De façon matricielle, nous pouvons écrire  $Y = X\beta + \epsilon$ , avec

- $Y = (Y_1, Y_2, \dots, Y_n)^\top$  la matrice contenant le vecteur des variables cibles.
- $X = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \dots & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n,1} & X_{n,2} & \dots & X_{n,k} \end{pmatrix}$  la matrice contenant les  $k$  observations des  $n$  variables explicatives.
- $\beta = (\beta_1, \beta_2, \dots, \beta_k)^\top$  le vecteur des paramètres du modèle.
- $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_k)^\top$  le vecteur des erreurs du modèle.

La fonction de vraisemblance du modèle est alors définie comme suit

$$L(\beta, \sigma | y) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left( -\frac{1}{2\sigma^2} (y - X\beta)(y - X\beta)^{-1} \right).$$

L'estimation du  $\hat{\beta}$  optimal se fait alors par maximum de vraisemblance et nous donne  $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ . Maintenant que nous avons défini ce qu'est un modèle linéaire, voyons ce qu'est un modèle linéaire généralisé.

Le GLM (Modèle Linéaire Généralisé) relie quant à lui variables explicatives et variable cible via une fonction dite « de lien », que l'on notera  $g$ . Le GLM est caractérisé par 2 éléments :

- La fonction de lien. Dans une régression linéaire, les valeurs prédites correspondent à la prédiction moyenne des observations. Sauf que dans un GLM elles correspondent à une transformation de la prédiction moyenne selon  $g$ . La fonction de lien se définit alors de la façon suivante :  $g(\mathbb{E}[Y|X]) = X\beta$ .
- Le prédicteur linéaire, qui est la combinaison linéaire des variables explicatives  $X\beta$ .

Le modèle linéaire généralisé permet d'utiliser notamment des données binaires. En effet le modèle linéaire suppose que les variables explicatives sont distribuées selon une loi normale, ce qui n'est pas vrai dans notre cas. Le GLM permet quant à lui d'utiliser la fonction de lien afin de prédire des variables ne suivant pas nécessairement cette loi. Le tableau 3.1 récapitule certaines fonctions de lien.

Nos variables Arrêté Sécheresse et Arrêté inondation étant binaires, nous avons donc besoin d'utiliser les modèles linéaires généralisés, et plus précisément la fonction de lien logit.

Commençons par définir la loi binomiale.

Une variable  $Y$  suit la loi de Bernoulli si elle a comme domaine de définition  $\{0,1\}$  et a pour loi

$$P(Y_i = i) = \pi^i (1 - \pi)^{1-i}$$

On dit que  $Y$  suit une loi de Bernoulli de paramètre  $\pi$ .

Distribution de Y	Fonction de lien	Fonction de la moyenne
Normale	Identité : $\sum_{j=1}^k X_{i,j}\beta_j = \mu$	$\mu = \sum_{j=1}^k X_{i,j}\beta_j$
Poisson	Log : $\sum_{j=1}^k X_{i,j}\beta_j = \log(\mu)$	$\mu = \exp(\sum_{j=1}^k X_{i,j}\beta_j)$
Binomiale	Logit : $\sum_{j=1}^k X_{i,j}\beta_j = \log\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{1}{1 + \exp(-\sum_{j=1}^k X_{i,j}\beta_j)}$
Gamma	Inverse : $\sum_{j=1}^k X_{i,j}\beta_j = -\frac{1}{\mu}$	$\mu = -\frac{1}{\sum_{j=1}^k X_{i,j}\beta_j}$

TABLE 3.1 : Récapitulatif des fonctions de lien

La loi binomiale consiste quant à elle en une somme de lois de Bernoulli indépendantes de paramètre  $p$  et prend ses valeurs dans l'ensemble  $\{0,1,\dots,n\}$ . Sa loi est définie par

$$P(Y_i = i) = \binom{n}{i} \pi^i (1 - \pi)^{n-i}.$$

Expliquons d'abord en quoi le GLM logit va s'avérer utile pour modéliser une variable binaire. Nous avons déjà expliqué que le GLM classique ne pouvait pas fonctionner dans notre cas.

Nous cherchons, dans le cadre d'une modélisation, à expliquer  $Y$  par  $X$ . Cela revient à s'interroger sur la loi de probabilité de  $Y$  sachant  $X$ , que l'on écrit  $\mathbb{P}(Y|X)$ . Nous allons chercher à caractériser cette loi selon une loi binomiale.

La loi binomiale est une somme de lois de Bernoulli indépendantes de paramètre  $p$ . Dans un cadre GLM, nous définissons ce paramètre de la sorte :

$$p(X) = \mathbb{P}(Y = 1|X).$$

Notre objectif est alors de trouver une forme de  $p(x)$  telle que  $X$  puisse être modélisée selon  $Y$ . Pour cela, nous définissons la forme suivante

$$p(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)},$$

ce qui donne

$$\text{logit}(p(x)) = \text{logit}(Y) = \log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n.$$

De façon matricielle, on a :  $\text{logit}(Y) = X\beta$ . Cela nous permet donc de retrouver la fonction de lien précédemment présentée. Le modèle de régression logistique consiste donc à caractériser la loi de  $Y|X = x$  par une loi binomiale de paramètre  $p(x)$ .

La fonction de vraisemblance du modèle est alors

$$L(\beta) = \prod_{i=1}^n \mathbb{P}(Y_i|X_i) = \prod_{i=1}^n p(X_i)^{Y_i} (1 - p(X_i))^{1-Y_i}.$$

Pour estimer les paramètres  $\beta$ , il vaut mieux utiliser la fonction de log-vraisemblance, car la fonction de vraisemblance peut être difficile à manipuler.

$$\begin{aligned} \ell = \log(L) &= \log\left(\prod_{i=1}^n \mathbb{P}(Y_i|X_i)\right) = \log\left(\prod_{i=1}^n p(X_i)^{Y_i} (1 - p(X_i))^{1-Y_i}\right) \\ &= \sum_{i=1}^n Y_i \log(p(X_i)) + (1 - Y_i) \log(1 - p(X_i)) \\ &= \sum_{i=1}^n Y_i X_i \beta + \log\left(1 - \frac{1}{1 + \exp(-X_i \beta)}\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n Y_i X_i \beta + \log\left(\frac{1}{1 + \exp(X_i \beta)}\right) \\
&= \sum_{i=1}^n Y_i X_i \beta - \log(1 + \exp(X_i \beta)).
\end{aligned}$$

Pour analyser la qualité du modèle logit, nous pouvons étudier plusieurs métriques. La première est l'AIC, qui se définit comme suit

$$AIC = 2 \times k - 2 \times \ln(L),$$

avec :

- $k$  le nombre de variables du modèle ;
- $L$  le maximum de la fonction de vraisemblance du modèle.

Cette métrique sert à pénaliser les modèles en fonction du nombre de paramètres. En effet, lorsque l'on évalue un modèle, on se base souvent sur sa vraisemblance. Ajouter un paramètre peut améliorer la vraisemblance mais rend le modèle plus complexe. Se contenter d'utiliser la vraisemblance comme le seul critère est donc insuffisant et c'est ici que l'AIC est intéressant.

De façon similaire à l'AIC, nous introduisons le BIC (*Bayesian Information Criterion*), défini par

$$BIC = -2\log(L) + \log(n)k$$

avec  $n$  le nombre d'observations effectuées. Maintenant que nous avons parlé du GLM logit, intéressons-nous à la méthode Lasso et à son intérêt pour notre étude.

Le GLM peut être pénalisé selon une méthode Lasso (*Least Absolute Shrinkage and Selection Operator*). Dans un modèle linéaire standard, les paramètres  $\beta$  du modèle sont obtenus par minimisation de la somme des carrés des résidus. Dans un modèle Lasso, cette minimisation est obtenue en ajoutant une contrainte supplémentaire. Cette contrainte a pour objectif d'encourager le modèle à enlever les variables peu explicatives

$$\min_{\beta_0, \beta_1, \dots, \beta_k} \frac{1}{2} \sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{i,j})^2 + \lambda \sum_{j=1}^p |\beta_j|,$$

avec  $\lambda \geq 0$

La pénalisation va forcer une réduction des coefficients  $\beta$ , forçant certains à valoir 0. L'avantage de cette méthode est qu'elle supprime certaines variables, réduisant ainsi leur nombre.

Cela est particulièrement utile dans le cas où certaines variables sont corrélées comme dans notre base de donnée Open Source. En effet, la méthode Lasso va avoir tendance à ne retenir qu'une des variables corrélées et éliminer les autres, ce qui peut être visualisé sur la figure 3.4.

### 3.2.2 Les forêts aléatoires

À la base de la forêt aléatoire, on trouve les arbres de décision CART (*Classification And Regression Tree*). Notre description de ces derniers reposera sur l'ouvrage CHESNEAU (2022). Leur objectif est de prédire la variable cible  $Y$  selon les variables explicatives  $X$ . La méthodologie pour un arbre de décision est la suivante : sélection d'une des variables explicatives  $X^j$  et d'une valeur-seuil divisant l'espace en deux parties, avec d'un côté les observations de  $X^j$  supérieures à la valeur-seuil et de l'autre celles inférieures. Cette valeur-seuil peut également être traduite sous forme de question.

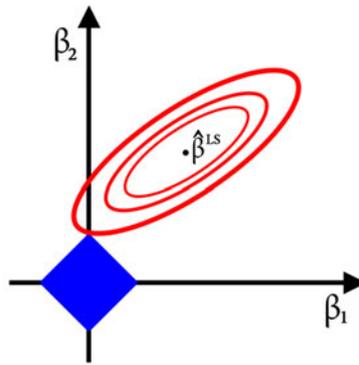


FIGURE 3.4 : Contours de l'erreur(en rouge) et de la fonction de contrainte(en bleu) pour la méthode Lasso en deux dimensions.

Par exemple, si l'on prend une variable taille pour un jeu de données de la population, sélectionner une valeur-seuil de 1,7 pour la variable taille revient à poser la question «L'individu est-il plus grand que 1 mètre 70?». Il est bien sûr possible de diviser plusieurs fois l'ensemble selon divers critères. Le processus de création de l'arbre est donc de :

- Partir du premier élément, appelé racine, qui regroupe l'ensemble des individus du jeu de données.
- Descendre depuis cette racine vers les sous-groupes créés. Les liens qui les relient sont appelés branches et les sous-groupes noeuds. Un arbre se parcourt donc depuis la racine jusqu'aux feuilles.
- Faire partir de chaque noeud 0 ou 2 branches. Les noeuds créés sont appelés noeuds fils et le noeud de départ noeud père.
- Terminer la division de l'espace. Les noeuds finaux sont appelés feuilles.

Voici un exemple d'arbre de décision figure 3.5

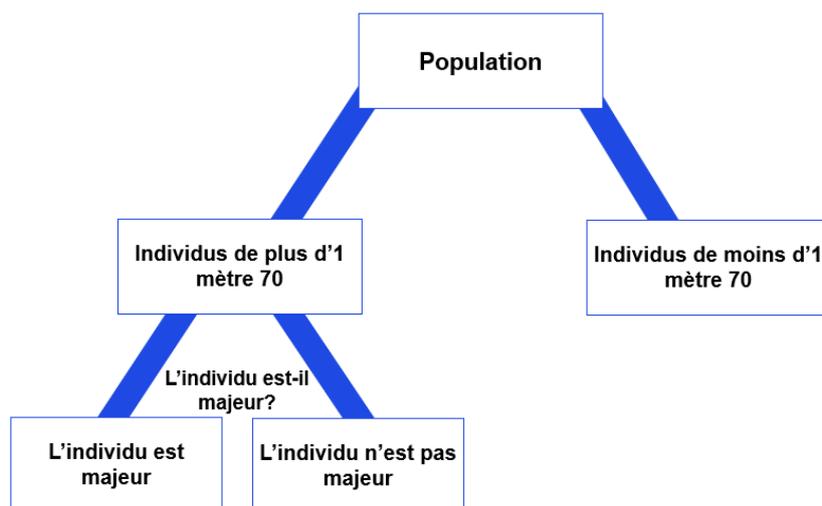


FIGURE 3.5 : Exemple simpliste d'arbre de décision

Lors de la création de l'arbre, il est nécessaire de se poser certaines questions : comment sélectionner la variable de division et comment sélectionner sa valeur seuil ? Quand décidons-nous d'arrêter de créer des divisions ? Notre objectif étant de répartir les individus selon Y afin de permettre une bonne prédiction, nous voulons cette répartition comme étant la plus homogène de ce point de vue. Pour cela, il faut étudier l'erreur créée par une division. Nous introduisons les notations suivantes :

- $n_{j,s,g}$  le nombre d'individus où  $X^j < s$ , avec s la valeur seuil ;
- $n_{j,s,d}$  le nombre d'individus où  $X^j \geq s$ , avec s la valeur seuil ;
- $f_{i,j,s,g}$  la fréquence de la modalité i de Y pour les individus où  $X^j < s$  ;
- $f_{i,j,s,d}$  la fréquence de la modalité i de Y pour les individus où  $X^j \geq s$ .

Pour un nœud, l'erreur ou perte d'information créée par la séparation est donnée par la fonction positive de perte i. Cette fonction prend pour valeur 0 quand la valeur de Y est la même pour tout individu du nœud, on dit alors que la population est homogène, et est maximale lorsque il y a autant d'individus avec pour valeur de Y 1 que d'individus avec pour valeur 0, c'est à dire quand sur 100 individus placés dans le nœud, 50 ont pour valeur de Y 1 et 50 ont pour valeur 0. Sa formule est donnée par

$$i(j, c) = \frac{n_{j,c,g}}{n} i_g(j, c) + \frac{n_{j,c,d}}{n} i_d(j, c)$$

avec :

- $n_k$  les effectifs dans le nœud créé k ;
- n les effectifs dans le nœud précédent ;
- $i_g(j, c)$  la perte d'information créée dans le nœud gauche créé ;
- $i_d(j, c)$  la perte d'information créée dans le nœud droit créé ;

Nous pouvons donner un exemple de fonction mesurant la pureté avec la fonction de Gini :

$$i(j, s) = 1 - \sum_{k=1}^K f_{j,s}^2$$

L'objectif lors de la création de chaque nœud est de minimiser la perte. Nous choisissons donc j et s de façon à ce que cette dernière soit minimale. On note  $j^*$  et  $s^*$  les critères tels que  $i(j^*, s^*) < i(j, s) \forall j, s$ .

La différence de pureté entre les différents nœuds est alors évaluée par le critère d'information  $\Delta$

$$\Delta = \sum_{i=1}^l \frac{n_k}{n} (\text{impureté du } k \text{ ème nœud créé}),$$

avec :

- $n_k$  les effectifs dans le nœud créé k ;
- n les effectifs dans le nœud précédent.

L'opération est ensuite répétée sur les nœuds créés et ainsi de suite. Il existe différents critères d'arrêts :

- Quand le nœud est parfaitement homogène et on ne peut donc plus faire de divisions ;
- Quand le nombre d'individus dans chaque classe passe sous un certain seuil fixé.

Un arbre ainsi créé peut cependant être peu satisfaisant, en effet l'arbre peut s'avérer très complexe du fait d'un grand nombre de noeud. De plus, un arbre trop complexe peut trop s'adapter aux données de l'ensemble d'apprentissage. Il fonctionnera alors moins bien sur un autre ensemble, c'est ce que l'on appelle le sur-apprentissage. On peut alors enlever un certain nombre de noeuds afin de réduire la complexité, ce que l'on appelle l'élagage. L'arbre une fois élagué est désigné comme l'arbre optimal. Le processus d'élagage commence d'abord avec l'arbre créé afin de maximiser l'homogénéité, que l'on appelle arbre maximal. Notre objectif est de construire une suite d'arbres emboîtés à partir de cet arbre et de choisir celui minimisant l'erreur sur un échantillon test. Cette construction repose sur un critère de pénalisation.

Pour chaque sous arbre  $T_1, \dots, T_K$  de l'arbre maximal  $T_{max}$ , composé de  $|T|$  feuilles, la qualité d'ajustement est

$$D(T) = \frac{1}{n} \sum_{i=1}^{|T|} \sum_{j \in k} (y_j - \hat{y}_k)^2$$

La pénalisation prend alors la forme d'une fonction linéaire du nombre de feuilles et nous avons

$$C(T) = D(T) + \theta|T|$$

Le sous-groupe d'arbres est d'abord construit pour  $\theta = 0$ , puis on fait croître la valeur de  $\theta$  jusqu'à ce qu'une des  $K-1$  divisions de l'arbre  $T_K$  ne compense pas l'augmentation de la pénalisation. Les deux feuilles issues de cette division sont regroupées pour former l'arbre  $T_{k-1}$ . On poursuit le processus jusqu'à l'obtention de la séquence d'arbres emboîtés  $T_1 \subset T_2 \dots \subset T_{K-1} \subset T_K$ . Il est également possible de construire la suite  $D(1), \dots, D(K)$  de la qualité d'ajustement en fonction du nombre de feuilles ou de la valeur de  $\theta$ .

Pour identifier l'arbre optimal, on va regarder sur l'échantillon de validation l'arbre qui minimise l'erreur parmi la séquence d'arbres emboîtés créée.

La forêt aléatoire est quant à elle composée de plusieurs arbres binaires CART (nous nous référerons cette fois au mémoire RIHOUEY (2020)). Le problème est que construire  $N$  arbres sur un même échantillon créerait des variables très corrélées, ce qui rend nécessaire l'utilisation de la technique du bootstrap. Le *bootstrap* est un processus segmenté en deux étapes :

- Nous tirons un échantillon aléatoire de taille  $n$  avec remise selon une estimation de la fonction de répartition des variables explicatives que l'on notera  $\hat{F}$  ;
- Nous calculons la moyenne empirique de l'échantillon créé.

Dans le cadre de notre problème, chaque arbre est créé à partir de  $N$  bases créés par *bootstrap* à partir des données initiales. Voici un exemple de forêt aléatoire 3.6.

Concernant les variables à utiliser pour la construction de chaque arbre, l'algorithme *Random Forest* utilise un tirage aléatoire de  $k$  variables parmi le total appelé *feature sampling*. L'algorithme fait donc intervenir deux tirages aléatoires dans sa construction. Le nombre de variables  $k$  à utiliser pour la segmentation est sujet à discussion, mais pour un problème de classification on retient généralement  $k = \sqrt{p}$ , où  $p$  est le nombre de variables explicatives (pour les problèmes de régression, il est recommandé de sélectionner  $\frac{p}{3}$ ). Dans le cadre d'une classification comme c'est le cas ici, la prédiction de la variable cible correspond à la classe la plus représentée parmi tout les arbres.

Pour résumer, la construction d'une forêt aléatoire suit l'algorithme suivant :

1. Création d'un échantillon comportant  $n$  individus par *Bootstrap* ;
2. Sur les  $k$  variables, nous n'en retenons qu'un nombre  $p$  ;
3. Entraînement d'un arbre de décision CART selon le paramètre  $p$ .

Ce processus est alors répété autant de fois que l'on souhaite d'arbres.

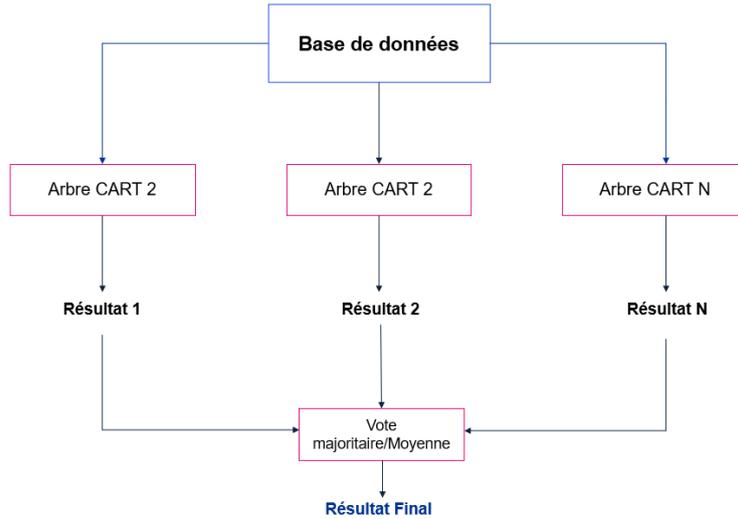


FIGURE 3.6 : Exemple de forêt aléatoire

### 3.2.3 Le Xgboost

Le Xgboost est un modèle de *Machine Learning* pouvant être utilisé aussi bien pour les problèmes de classification que de régression. Les notions présentées ci-dessous sont issues du rapport technique CHEN et GUESTRIN (2016) ainsi que des mémoires AKAFFOU (2020) et SERVEL (2020).

A la base du XgBoost, nous trouvons le boosting dont le principe est le suivant :

- Nous créons un premier modèle que nous entraînons sur les observations. Nous attribuons d'abord un poids égal à chaque observation. Si une observation est mal classée, son poids augmente.
- Nous créons un second modèle afin de corriger les erreurs du premier. Nous l'entraînons avec les données pondérées obtenues lors de la première étape. Nous répétons cette étape jusqu'à ce que toutes les observations soient correctement prédites.

Nous représentons ci-dessous le processus du boosting 3.7 :

Le gradient boosting est une procédure où l'on minimise les erreurs du boosting via l'algorithme de descente de gradient. Expliquons le plus en détail. Comme dit précédemment, la démarche est de fusionner plusieurs modèles de classification en un seul afin d'optimiser au mieux les résultats.

En reprenant les notations  $X$  et  $Y$  précédentes, notre objectif est de construire un régresseur que l'on note  $\hat{f}$  minimisant une fonction de perte  $\mathcal{L}(y_i, \hat{f})$  différentiable par rapport à  $F(x)$ , ce qui se traduit par l'équation

$$\hat{f}(x) = \underset{f(x)}{\operatorname{argmin}} \mathbb{E}_{Y=y, X=x} [\mathcal{L}(y, f(x))]$$

Pour la descente de gradient, le modèle  $\hat{f}$  est amélioré à chaque itération en lui additionnant un élément qui correspond au pas de descente  $\lambda$  multiplié par une approximation de l'opposé du gradient de la fonction de perte obtenue à l'aide d'un arbre de régression. Cela nous donne l'équation suivante

$$\hat{f}_m = \hat{f}_{m-1} - \lambda_m \sum_{i=1}^n \frac{\partial \mathcal{L}(y_i, \hat{f}_{m-1}(x_i))}{f(x_i)}$$

L'algorithme du Gradient Boosting est alors le suivant :

- Initialisation :  $\hat{f}_0 = \underset{\lambda}{\operatorname{argmin}} \sum_{i=1}^n \mathcal{L}(y_i, \lambda)$

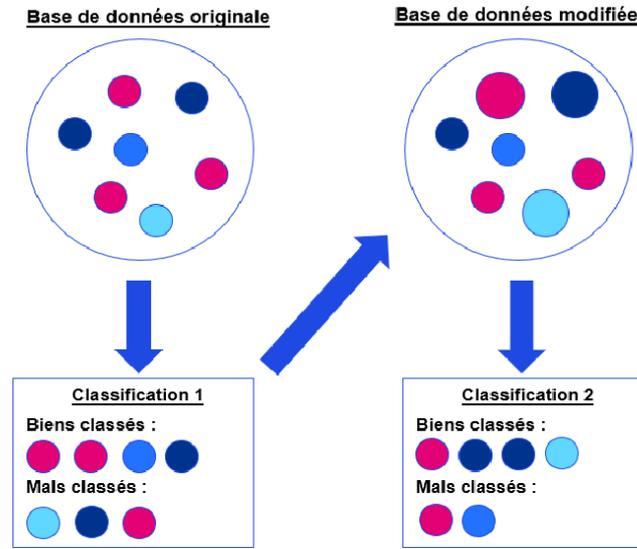


FIGURE 3.7 : Représentation du boosting

- Pour  $m=1, \dots, M$ , on calcule les pseudos-résidus  $r_{mi} = -[\frac{\partial \mathcal{L}(y_i, f(x_i))}{\partial f(x_i)}]_{f=f_{m-1}}$ , pour  $i=1, \dots, m$ . On ajuste ensuite un arbre de régression  $c_m$  en utilisant l'ensemble  $(x_i, r_{mi})_{i=1, \dots, n}$  comme ensemble d'entraînement.  
On détermine ensuite le  $\lambda_m$  optimal en résolvant  $\lambda_m = \underset{\lambda}{\operatorname{argmin}} \sum_{i=1}^n \mathcal{L}(y_i, F_{m-1}(x_i) + \lambda c_m(x_i))$   
Enfin, on met à jour  $\hat{F}$ ,  $\hat{F}_m(x) = \hat{F}_{m-1}(x) + \lambda_m c_m(x_i)$
- Le résultat final est  $\hat{F}_M(x_0)$ .

Le XGboost est une version améliorée et étendue du gradient boosting développée par CHEN et GUESTRIN (2016). Dans cet algorithme, on cherche à minimiser non plus une fonction de perte mais une fonction objectif que l'on notera  $\mathcal{J}$  et dont la formule est

$$\mathcal{J}(\hat{f}(x)) = \sum_{i=1}^n \mathcal{L}(y_i, \hat{f}(x_i)) + \sum_{m=1}^M \Omega(\delta_m)$$

avec :

- $\delta_m$  un arbre de classification CART ;
- $\Omega$  un terme de régularisation dont la formule est  $\Omega(\delta) = \alpha|\delta| + \frac{1}{2}\gamma\|\omega\|^2$ , où  $|\delta|$  désigne le nombre de feuilles de l'arbre, et  $\alpha, \gamma$  des constantes positives.

L'objectif est alors de minimiser  $\mathcal{J}$ . L'idée a été d'utiliser un développement de Taylor au second ordre afin de simplifier l'algorithme. A la  $t^{ème}$  itération, la fonction objectif est alors :

$$\mathcal{J}^{(t)} = \sum_{i=1}^n (g_i \delta_t(x_i) + \frac{1}{2} h_i \omega_j^2) + \alpha |\delta_t| + \frac{1}{2} \lambda \sum_{j=1}^{|\delta_t|} \omega_j^2 + K$$

avec :

- $K$  un terme constant par rapport à  $\delta_t$  ;
- $g_i = \partial_{f^{t-1}} \mathcal{L}(y_i, \hat{f}^{t-1})$  la dérivée partielle de la fonction de perte au premier ordre ;
- $h_i = \partial_{f^{t-1}}^2 \mathcal{L}(y_i, \hat{f}^{t-1})$  la dérivée partielle de la fonction de perte au premier ordre.

A partir de cette fonction de perte, il est alors possible de déterminer une structure optimale pour l'arbre de classification.

Cette technique comporte de nombreux avantages :

- Régularisation : pendant le calcul, un paramètre de régularisation est employé pour contrôler le surajustement ;
- Parallélisation de l'apprentissage : cette méthode divise les données en plusieurs blocs ce qui permet d'utiliser plusieurs coeurs et donc d'accélérer l'apprentissage.
- Blocs de calcul out-of-core : compression des données trop volumineuses afin d'optimiser l'espace disponible sur l'ordinateur.

Maintenant que nous avons décrit les modèles que nous allons employer, réfléchissons à la modélisation et aux problèmes que nous pourrions rencontrer.

### 3.2.4 SMOTE

Un des problèmes auquel nous pouvons nous confronter est le faible nombre d'observations 1 dans notre base. En effet, nous avons 8,5 % de 1 pour la sécheresse et 5,05 % de 1 pour les inondations. Cela est très faible, faisant que notre modèle pourrait ne prédire que des 0 mais tout de même obtenir une excellente précision. Afin de rééquilibrer notre base de test, nous pouvons utiliser la méthode du SMOTE (*Synthetic Minority Oversampling Technique*).

Le rééquilibrage des observations peut se faire par 2 approches bien distinctes :

- Le sur-échantillonnage : nous augmentons le nombre d'individus minoritaires.
- Le sous-échantillonnage : nous retirons une partie des individus majoritaires afin de donner une proportion plus importante aux individus minoritaires.

Dans cette étude, nous nous basons sur le logiciel R CORE TEAM (2023), qui propose deux paramètres concernant le SMOTE : `perc.under` qui contrôle le taux de sous-échantillonnage et `perc.over`, qui contrôle le taux de sur-échantillonnage. Par exemple, supposons que nous posions `perc.under=400` et `perc.over=100`. Le 100 indique que nous doublons le nombre d'individus de la classe minoritaire, c'est à dire que pour un individu minoritaire, un individu synthétique est créé. Le `perc.over=400` indique quant à lui que pour chaque individu minoritaire synthétique, il y a 4 individus de la classe majoritaire.

De façon concrète, de tels paramètres feront que la base contienne 33 % d'individus minoritaires (1 individu minoritaire réel + 1 individu minoritaire synthétique + 4 individus majoritaires associés à l'individu synthétique).

Le SMOTE est donc, comme son nom l'indique, une technique de sur-échantillonnage. Son algorithme est le suivant :

1. Sélection d'un individu minoritaire de façon aléatoire ;
2. Identification de ses  $k$  plus proches voisins,  $k$  étant choisi par l'utilisateur ;
3. Sélection aléatoire d'un de ses voisins ;
4. Génération aléatoire d'un paramètre  $\alpha$  compris entre 0 et 1 ;

- Création d'un nouvel individu minoritaire compris entre l'observation initiale et le voisin choisi selon la valeur du coefficient. Par exemple, si  $\alpha = 0.5$ , alors la nouvelle observation sera à mi-chemin entre l'observation de base et le plus proche voisin sélectionné.

Il est très important d'optimiser le paramètre du nombre de plus proches voisins. En effet, ce nombre a une influence considérable sur la création des individus minoritaires synthétiques, comme montré sur la figure 3.8

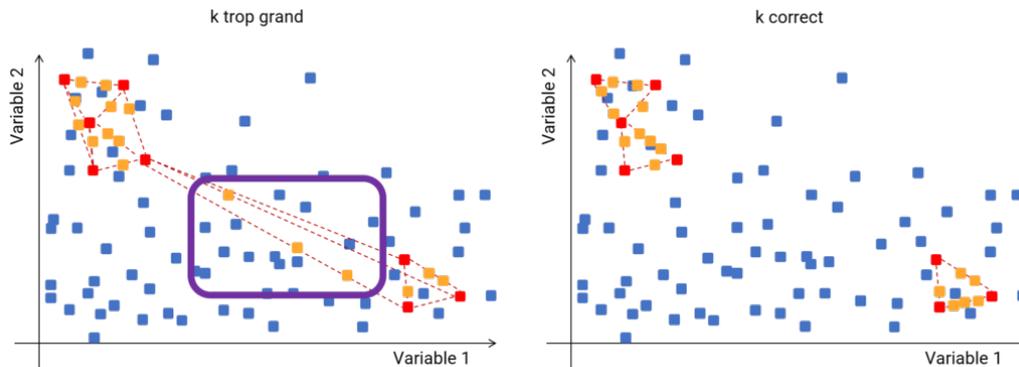


FIGURE 3.8 : Exemple de SMOTE avec différents k (Source : KOBIA (2022))

Les individus minoritaires réels sont en rouge et les individus minoritaires synthétiques en jaune. Ici, le k de la figure de gauche, égal à 8, est trop grand. En effet, des individus minoritaires sont créés dans la zone violette qui n'est pas représentative des zones où se trouvent réellement les individus minoritaires. Cela est dû à un nombre de k voisins trop important.

Le principal défaut du SMOTE est qu'il peut entraîner des problèmes de sur-apprentissage, c'est-à-dire une bonne prédiction sur l'ensemble d'apprentissage mais une mauvaise sur l'ensemble de test. Cela est dû au fait que nous ne modifions pas l'ensemble de test mais seulement celui d'apprentissage. Le modèle aura alors une base équilibrée sur laquelle apprendre mais risque de ne pas nécessairement bien prédire sur une base très déséquilibrée comme la nôtre.

### 3.2.5 Evaluation des modèles

Parlons maintenant de comment nous allons évaluer nos modèles. Puisque nous nous intéressons à un problème de classification, une bonne méthode d'évaluation consiste à nous intéresser à la matrice de confusion des prédictions du modèle. Cette dernière va recenser quatre éléments :

- Le nombre de vrais négatifs, c'est à dire les 0 correctement prédits ;
- Le nombre de vrais positifs, c'est à dire les 1 correctement prédits ;
- Le nombre de faux positifs, c'est à dire le nombre de 0 qui ont été prédits comme des 1 ;
- Le nombre de faux négatifs, c'est à dire le nombre de 1 qui ont été prédits comme des 0.

Le tableau 3.2 montre à quoi ressemble une matrice de confusion.

Afin de vérifier nos résultats, nous allons utiliser plusieurs métriques basées sur la matrice de confusion afin de vérifier la bonne conduite du modèle. Présentons quelques unes de ces métriques :

- Le rappel. Cette métrique correspond au pourcentage de 1 correctement prédits par notre modèle.

$$Rappel = \frac{Vrais\ Positifs}{Vrais\ Positifs + Faux\ négatifs}$$

Observations Prédictions	0	1
0	VN	FN
1	FP	VP

TABLE 3.2 : La matrice de confusion

Le rappel est compris entre 0 et 1. Plus il est élevé, meilleure est la prédiction. Cette métrique est importante dans notre cas, puisque nous souhaitons pouvoir correctement prédire les 1, c'est à dire la présence d'arrêté CatNat. Cependant le rappel a une limite, c'est qu'il ne prend pas en compte le nombre de Faux Positifs. Or un nombre important de Faux Positifs nous pousserait à surestimer grandement la sinistralité CatNat.

- La précision. Cette métrique permet quant à elle de connaître le nombre de prédictions positives correctement effectuées. Sa formule est la suivante

$$\text{Précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux positifs}}$$

Une précision haute signifie que la majorité des observations positives du modèle sont correctement prédites, ce qui implique une bonne prédiction des 0. Cette métrique est intéressante car comme dit précédemment, nous ne souhaitons pas prédire trop de 1.

Nous avons donc désormais deux métriques. Cependant, elles ne sont pas suffisantes. En effet, lorsqu'on les considère de façon séparée elles peuvent s'avérer trompeuses. Un modèle prédisant tout le temps 1 aura un rappel élevé, tandis que s'il ne le prédit jamais, alors ce sera sa précision qui sera élevée. Nous devons donc combiner ces deux métriques afin d'avoir une évaluation claire du modèle.

- Le F1 score. Cette métrique est la moyenne harmonique du rappel et de la précision, c'est à dire que sa formule est

$$F1 \text{ score} = 2 \times \frac{\text{Rappel} \times \text{Précision}}{\text{Rappel} + \text{Précision}}$$

Le F1 score permet d'évaluer la performance de notre modèle en prenant en compte aussi bien le rappel que la précision. C'est donc sur cette métrique que nous allons nous concentrer.

En plus du F1 score, nous pouvons également nous pencher sur le Kappa de Cohen. En effet, ce dernier a pour objectif de mesurer l'accord entre deux variables qualitatives ayant les mêmes modalités, ici 0 et 1. Ce dernier est défini par la formule suivante

$$Kappa = \frac{C_{observée} - C_{aléatoire}}{1 - C_{aléatoire}},$$

avec :

- $C_{observée}$  la concordance observée, c'est à dire le nombre de vrais positifs/ négatifs sur le total des observations  $\frac{VP+VN}{VP+VN+FP+FN}$  ;
- $C_{aléatoire}$  la concordance aléatoire, c'est-à-dire que l'observation et la prédiction soient en accord de façon aléatoire. Cela se traduit par la formule suivante  $\frac{VP+FN}{VP+VN+FP+FN} \times \frac{VP+FP}{VP+VN+FP+FN}$ .

Le Kappa est inclut entre -1 et 1 et peut se traduire de la façon suivante : plus ce dernier est proche de 1, plus l'accord entre les observations et les prédictions est important. Un Kappa négatif quant à lui exprime un désaccord.

Afin de créer nos modèles et de s'assurer de leur bonne qualité, nous allons employer un processus de validation croisée.

Ce processus a pour but d'évaluer au mieux la qualité du modèle sur les données d'entraînement. Pour cela, la base de données est séparée en K parties de façon aléatoire, le K étant choisi par l'utilisateur. Le modèle est entraîné sur K-1 parties et est testé sur le dernier. L'opération est répétée K fois jusqu'à ce que tout les parties soient testés. La moyenne des scores enregistrés sert de mesure de performance. La figure 3.9 montre le processus correspondant

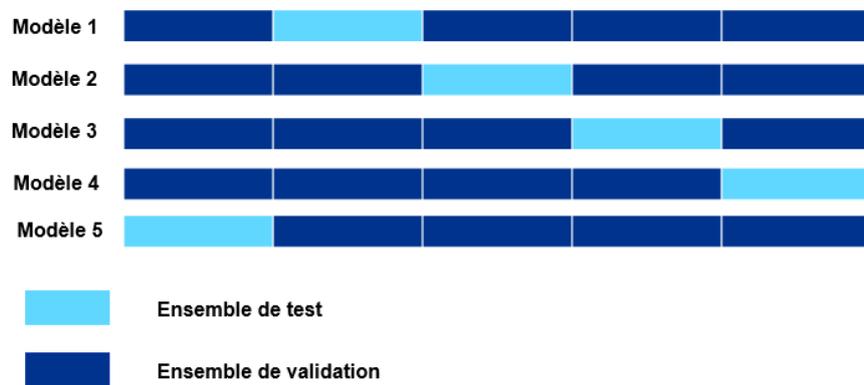


FIGURE 3.9 : Exemple de validation croisée avec 5 parties

L'avantage principal de cette méthode est d'éviter le côté aléatoire de la répartition entre base de test et base d'entraînement, qui peut entraîner un problème de sur-apprentissage : une répartition peut ainsi permettre d'avoir d'excellent résultats sur l'ensemble d'entraînement mais médiocre sur l'ensemble de test. Ce processus permet d'éviter cela puisque toutes les données sont utilisées pour l'entraînement et pour le test à tour de rôle.

Procédons désormais à la modélisation proprement dite.

Pour cela nous procédons à une division apprentissage/test. La base d'apprentissage sert à créer notre modèle de classification. La base de test quand à elle a pour objectif de tester le modèle sur une autre partie de l'ensemble des données afin de s'assurer de la bonne qualité de prédiction du modèle. Dans notre cas, nous diviserons la base en mettant 70% des données dans la base d'apprentissage et 30% dans la base de test.

### 3.3 Modélisation de la sécheresse

#### 3.3.1 Modélisation avec GLM pénalisé

Nous avons dans notre base 54 variables différentes, dont 4 ne sont pas utilisables (nom de la commune, année, code insee et département). Dans le cadre d'une modélisation, ce nombre important de variables peut entraîner une trop grande complexité du modèle. Nous allons donc chercher à identifier l'importance de chaque variable.

Nous faisons donc d'abord un modèle complet avec toutes les variables sur la base d'apprentissage,

puis nous enlevons les moins significatives afin d'obtenir notre modèle complet. Puisque la fonction calcule l'AIC sans la variable, plus cette dernière aura un score élevé, plus elle sera significative.

Voici le graphique 3.10 contenant les différences d'AIC pour chacune de nos variables. Afin de rendre la représentation plus claire, nous avons soustrait à toutes les valeurs d'AIC la valeur associée à la variable la moins explicative. Cela veut dire que si le score d'AIC associé à ladite variable est de 40 000, nous lui associerons la valeur 0. Pour ce graphique comme pour tout les suivants, nous représenterons pas l'intégralité des variables afin de rendre le graphique lisible.

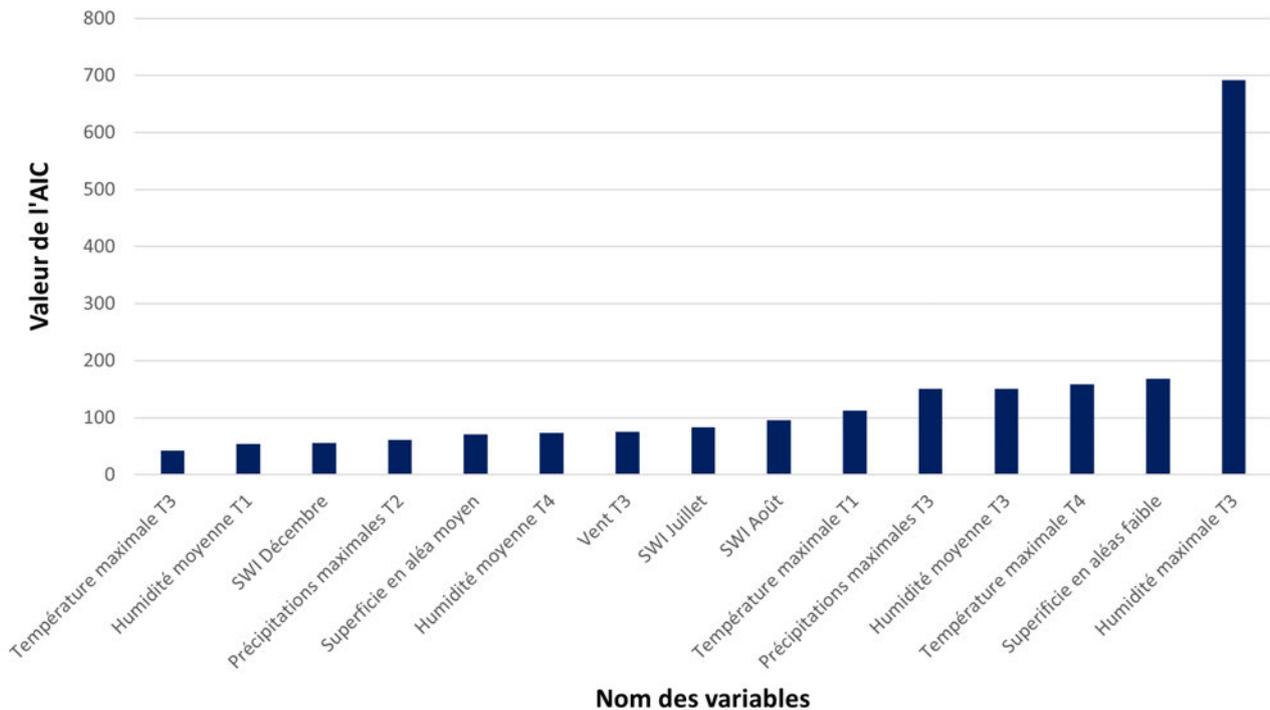


FIGURE 3.10 : Représentation de l'importance des variables de notre base dans le modèle pour la sécheresse

On remarque que pour la modélisation de la sécheresse, le modèle accorde une très grande importance pour la variable humidité au troisième trimestre. Ce résultat, bien qu'inattendu, est explicable.

En effet l'humidité est fortement corrélée avec la température mais aussi les précipitations. Une température, même élevée, ne provoquera pas de sécheresse si des pluies surviennent régulièrement. L'humidité permet de contourner ce problème, de plus le troisième trimestre est souvent le plus sujet aux inondations, puisque c'est durant ce dernier qu'on trouve les mois les plus chauds (juillet, août). Nous notons également une forte importance des variables SWI des mois de septembre, octobre et novembre, ce qui est quelque chose que nous avons pu voir lors de l'étude de la corrélation avec les arrêtés sécheresse. En effet, les mois suivants l'été sont très importants : puisque les mois les plus chauds sont juin, juillet et août, il y aura une tendance à avoir un début de sécheresse durant ces mois là. Néanmoins un mois de septembre ou d'octobre pluvieux pourra fortement limiter l'impact de cette sécheresse. Au contraire, un automne sec viendra aggraver la sécheresse.

Parmi les autres variables importantes pour le modèle, nous remarquons la superficie en aléas argile faible. Cela confirme l'importance de la composition des sols, ce qui est cohérent avec le fait qu'il s'agisse d'un critère de déclenchement d'arrêtés sécheresse.

Enfin, nous notons plusieurs variables pouvant servir à prédire l'inondation, comme les précipitations

du deuxième trimestre. Ces variables sont négativement corrélées à la sécheresse, puisque de fortes pluies viendront réduire la sécheresse, ce qui peut expliquer leur présence.

Notre modèle est donc le suivant (table 3.3) :

Variables
Humidité maximale T3+Superficie en aléas faible+Température maximale T4
+Humidité moyenne T3+Précipitations maximales T3
+Températures maximales T1+SWI août+SWI Juillet+ Vitesse du vent T3+Humdiité moyenne T4
+Superficie en aléas moyen+Précipitations maximales T2
+SWI Décembre+Humidité moyenne T1+Température maximale T3

TABLE 3.3 : Modèle GLM lasso sécheresse

Désormais, nous allons nous concentrer sur l’élaboration du meilleur modèle de pénalisation Lasso. Pour cela, nous allons nous intéresser au cas de figure sans SMOTE et avec afin de voir l’impact de ce dernier sur notre modélisation.

Sans SMOTE, nous obtenons la matrice de confusion suivante(table 3.4) ce qui nous donne un

Classes prédites	Classes réelles	
	0	1
0	32548	2665
1	245	340

TABLE 3.4 : Matrice de confusion du GLM sécheresse

Kappa de 26,69 % et un F1-score de 30,98 %.

Les résultats sont assez peu satisfaisants, aussi nous demandons-nous si l’utilisation du SMOTE pourrait améliorer le modèle. Nous appliquons donc un SMOTE avec pour paramètres perc.under=550, perc.over=100, ce qui nous donne une base comportant 26,67 % de 1 au lieu de 8,5 %. Nous entraînons alors notre modèle sur cette nouvelle base d’apprentissage.

Nous appliquons alors ce modèle à notre base de test, les coefficients dudit modèle se trouvant en annexe A.1. Nous regardons la matrice de confusion afin d’évaluer ce dernier (table 3.5).

Observations Prédications	0	1
0	29 935	1 278
1	2 825	1 760

TABLE 3.5 : Matrice de confusion du modèle GLM sécheresse

Résultats :

- Rappel = 57,93% ;
- Précision = 38,38% ;
- F1-score = 46,17% ;
- Kappa = 40,05 %.

Ces résultats peuvent être interprétés de la façon suivante : le modèle prédit correctement 58 % des arrêts sécheresse, mais sur 100 arrêts prédits par le modèle, il n’y en a que 38 qui sont réellement

survenus. Que la précision soit plus faible que le rappel est quelque chose d'attendu du fait de l'important déséquilibre de notre base : même si le modèle ne prédit mal qu'un faible nombre de 0, cela aura un impact fort sur la précision.

Nous arrivons alors à un F1-score de 46,17 %, ce qui est peu satisfaisant. Nous remarquons cependant une très forte amélioration de la prédiction en utilisant le SMOTE, qui semble donc efficace dans un cadre GLM pénalisé. Nous regardons ensuite la modélisation selon la forêt aléatoire, modèle pouvant donner de meilleurs résultats qu'un GLM.

### Modélisation avec forêt aléatoire

Pour procéder à la modélisation avec forêt aléatoire, il nous faut d'abord sélectionner les différentes variables à utiliser pour la modélisation. Pour cela, nous allons d'abord créer une forêt aléatoire avec l'intégralité de nos variables. Une fois cela fait, nous regardons le critère d'importance de Gini (table 3.11) de chaque variable. Ce critère correspond à la baisse du critère d'impureté de Gini entraîné par la variable, éléments que nous avons évoqués dans la partie 3.2.2.

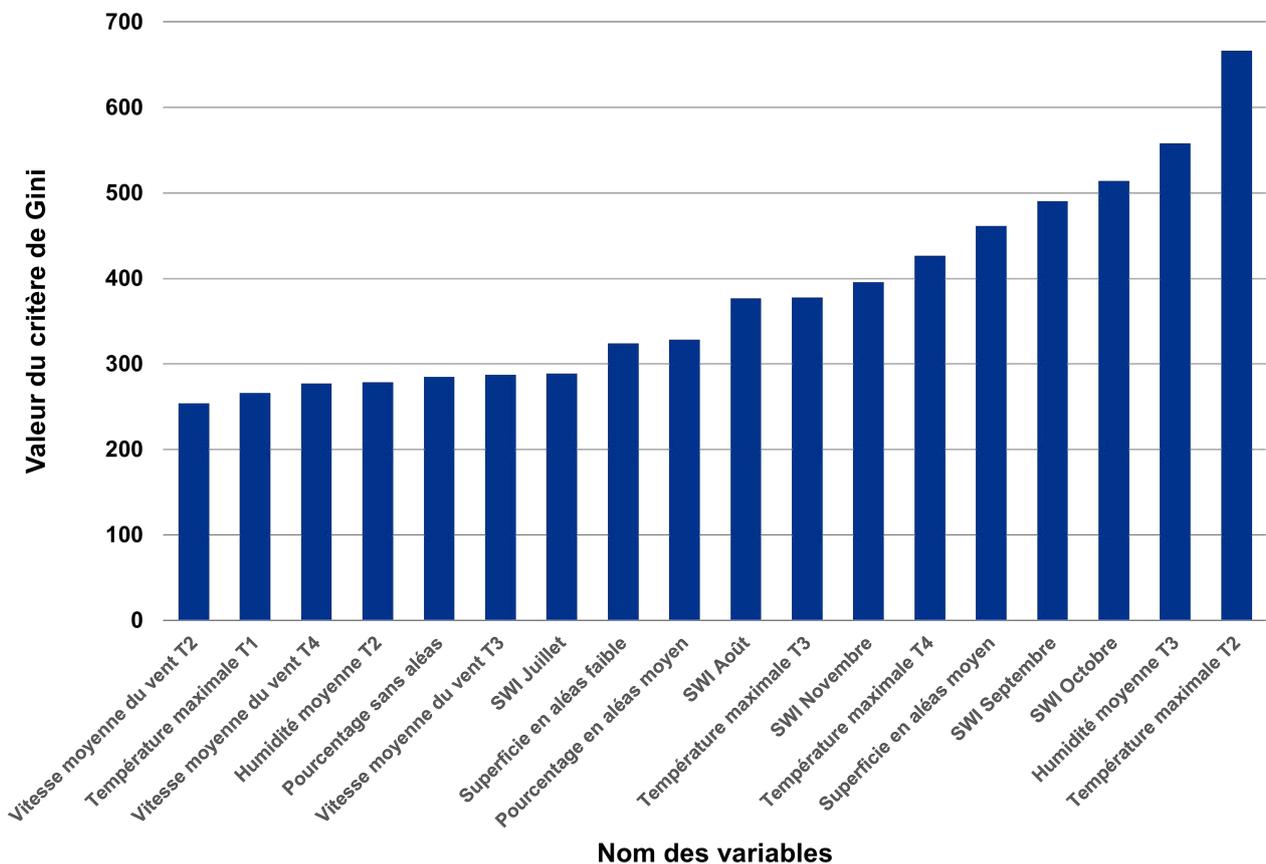


FIGURE 3.11 : Représentation de l'importance des variables pour la sécheresse selon le critère de Gini

Nous constatons que la variable température maximale est la plus importante de toutes, ce qui est cohérent pour la sécheresse. Nous aurions en revanche pu nous attendre à ce que ce soit la température du troisième trimestre qui s'avère importante, puisque ce trimestre comporte les mois les plus chauds. L'humidité est également importante, ce qui paraît logique du fait de sa corrélation avec les températures et les précipitations. Enfin, nous remarquons la présence de nombreuses variables

SWI, en particulier celles d'été et automne (pour rappel le SWI étudié ici est la moyenne des trois dernier mois, ce qui signifie que le SWI d'octobre regroupe les mois d'août, septembre et octobre).

Nous prenons donc le modèle ci-dessous (table 3.6).

Variables
Humidité maximale T3+SWI Octobre+Température maximale T2
+Superficie de la commune en aléas moyen+Température maximale T3+SWI Août
+SWI Novembre+SWI Septembre+SWI juin+ Pourcentage de la commune en aléas moyen
+Vitesse du vent T4 +SWI Juillet+Humidité maximale T2+SWI Mai
+Vitesse du vent T1+SWI mars+SWI Décembre
+Température maximale T1+Vitesse du vent T3+Température maximale T4

TABLE 3.6 : Modèle Forêt aléatoire sécheresse

Nous allons devoir faire un hyperparamétrage afin de créer le meilleur modèle possible. Cet hyperparamétrage portera sur les éléments suivants :

- le nombre d'arbres qui vont composer la forêt aléatoire ;
- le nombre de variables  $p$  utilisées pour la décision ;
- la profondeur maximale de chaque arbre.

Afin d'optimiser lesdits paramètres, nous procédons à une *cross-validation* ayant pour objectif de maximiser au mieux la valeur de notre Kappa. Nous commençons d'abord par chercher le nombre de variables optimal, qui est capital dans la création d'un modèle forêt aléatoire performant. Pour cela, nous utilisons une validation croisée à 10 feuilles, en cherchant le nombre maximisant le Kappa du modèle. Nous avons alors le graphique suivant montrant l'évolution du Kappa selon le nombres de variables 3.12 :

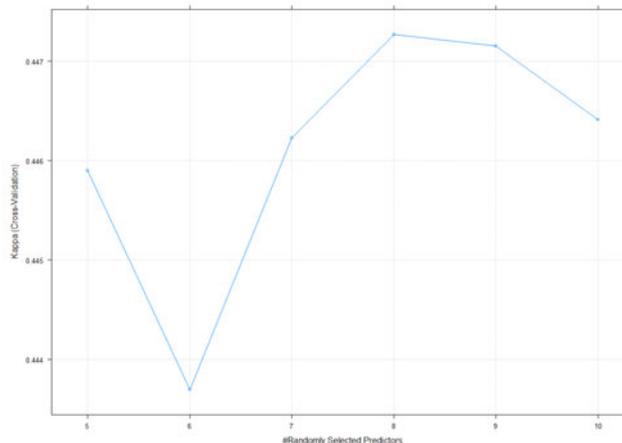


FIGURE 3.12 : Représentation de l'évolution du Kappa selon le nombre de variables

Une fois cela fait, nous construisons une seconde validation croisée à 10 feuilles. Cette validation croisée, réalisée sur R CORE TEAM (2023), effectue le processus autant de fois qu'il y a de combinaisons d'hyperparamètres, ce qui permet de repérer la combinaison maximisant le kappa. Nous représentons l'évolution du Kappa ci-dessous 3.13

Cette dernière nous donne les valeurs optimales suivantes (table 3.7).

L'utilisation des paramètres et du modèle présentés plus haut appliqués à notre base de test nous donne la matrice de confusion suivante (table 3.8).

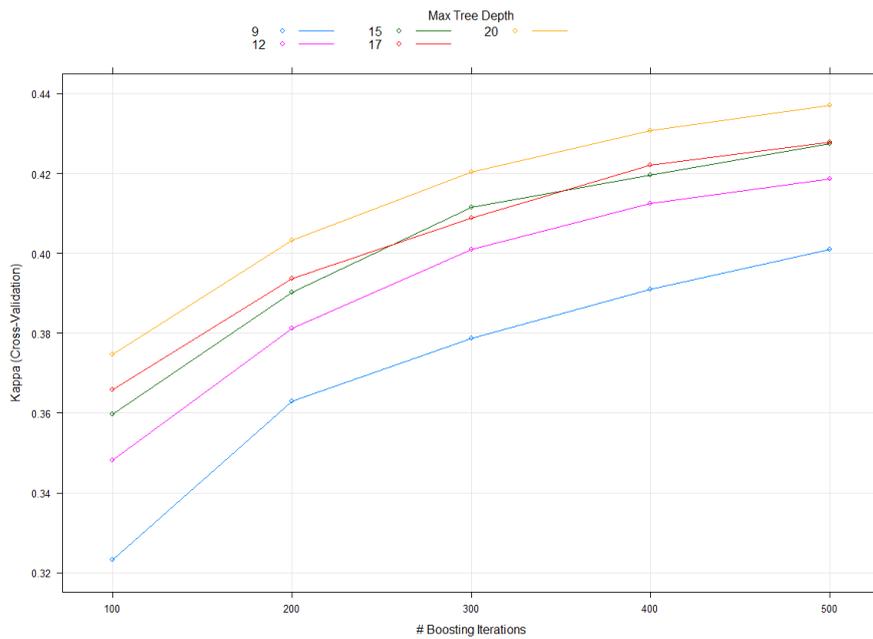


FIGURE 3.13 : Représentation de l'évolution du Kappa selon le nombre d'arbres ainsi que de la profondeur maximale de chaque arbre

Paramètre	Valeur optimale
Nombre d'arbres	500
Nombre de variables utilisées	10
Profondeur maximale de chaque arbre	9

TABLE 3.7 : Valeurs optimales des paramètres

Classes prédites	Classes réelles	
	0	1
0	31 723	1 416
1	1 037	1 622

TABLE 3.8 : La matrice de confusion de la forêt aléatoire sécheresse

Résultats de la forêt aléatoire sécheresse :

- Rappel = 53,39 %;
- Précision = 61,00 %;
- F1-score = 56,94 %;
- Kappa = 53,24 %.

Nous constatons une forte augmentation des différentes métriques, la forêt aléatoire est donc un meilleur type de modèle à considérer.

### Modélisation avec XgBoost

Nous étudions maintenant les modèles XgBoost. La première étape est, comme pour le GLM ainsi que la forêt aléatoire, de sélectionner les variables que nous utiliserons par la suite. Pour cela nous

procédons d'une façon assez similaire avec ce que nous avons fait précédemment :

- nous construisons un modèle XgBoost intégrant l'ensemble des variables de notre base ;
- nous utilisons la fonction *xgb.importance* du logiciel R CORE TEAM (2023). Cette fonction attribue un score d'importance, compris entre 0 et 1, à chaque variable. Ce score représente la contribution relative de chaque variable aux gains de précision obtenus sur les divisions de ladite variable. Ce gain de précision peut par exemple désigner le critère de gini, utilisé pour mesurer la qualité du modèle.

Nous représentons maintenant le score d'importance des 20 variables les plus significatives pour le modèle (figure 3.14).

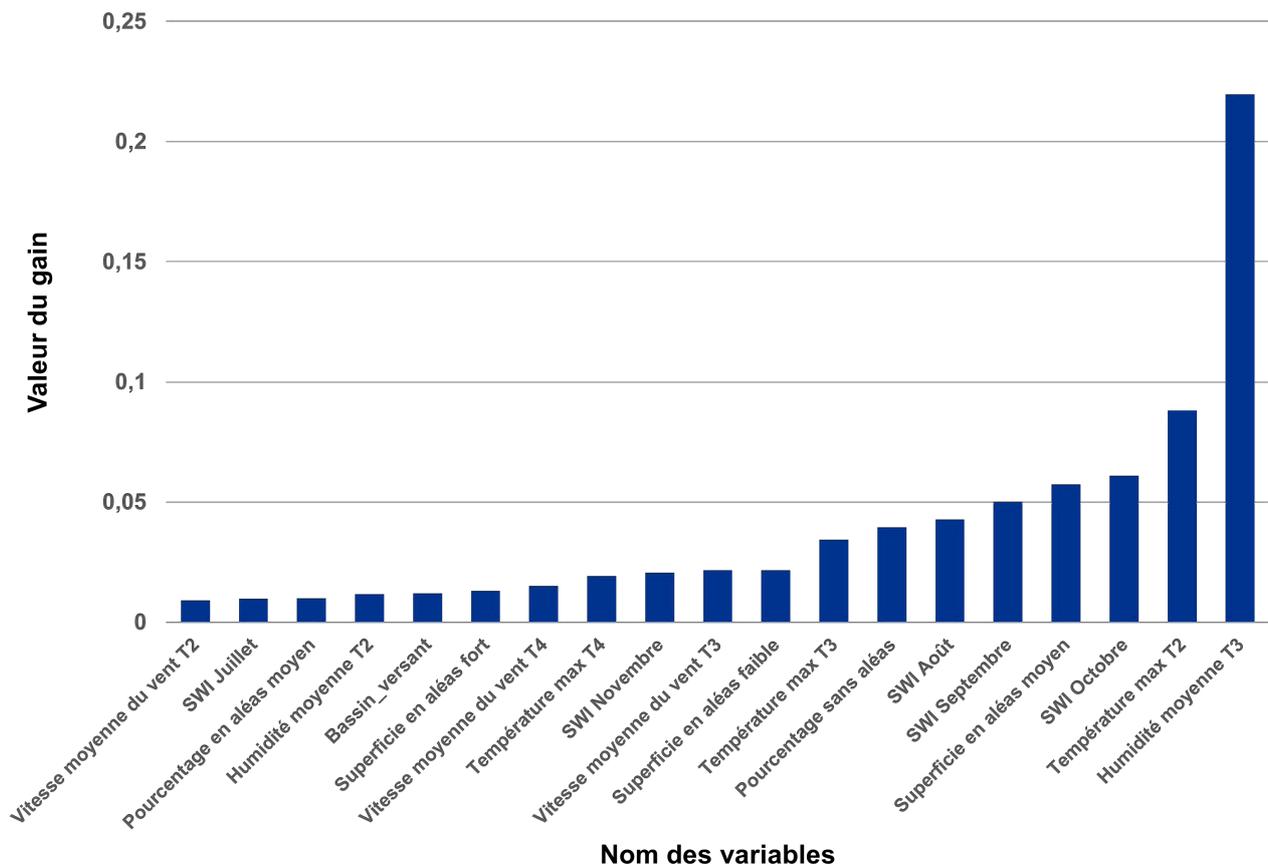


FIGURE 3.14 : Représentation de l'importance des variables pour le modèle XgBoost

Cela nous donne alors le modèle suivant (table 3.9).

Une nouvelle fois, l'humidité du troisième trimestre s'avère importante dans la régression, ce qui est cohérent avec ce que nous avons pu observer pour les forêts aléatoires. Nous observons que le XgBoost tend à accorder beaucoup d'importance aux mêmes variables que la forêt aléatoire vue précédemment, comme pour les données du BRGM ou le SWI, même si nous observons moins de variables SWI.

Une fois, le modèle créé, nous procédons comme avant à un hyperparamétrage. Pour le modèle XgBoost, il va nous falloir optimiser trois hyperparamètres en particulier :

- le nombre d'arbre de décision utilisés dans le modèle final ;
- la profondeur maximale de chacun de ces arbres ;

Variables

Humidité moyenne T3+Température max T2+SWI Octobre+Superficie de la commune en aléas moyen  
 +SWI Septembre+SWI Août+Pourcentage de la commune sans aléas+Température max T3  
 +Superficie de la commune en aléas faible+Vitesse moyenne du vent T3+SWI Novembre  
 +Température max T4+Vitesse moyenne du vent T4+Superficie de la commune en aléas fort  
 +Bassin versant+Humidité moyenne T2+Pourcentage de la commune en aléas moyen  
 +SWI Juillet+Vitesse moyenne du vent T2

TABLE 3.9 : Modèle XgBoost sécheresse

- le taux d'apprentissage, paramètre contrôlant la contribution de chacun des arbres au modèle.

Nous utilisons une nouvelle fois la validation croisée pour optimiser le paramètre Kappa, voici la façon dont ce dernier évolue selon l'hyperparamétrage (figure 3.15) :

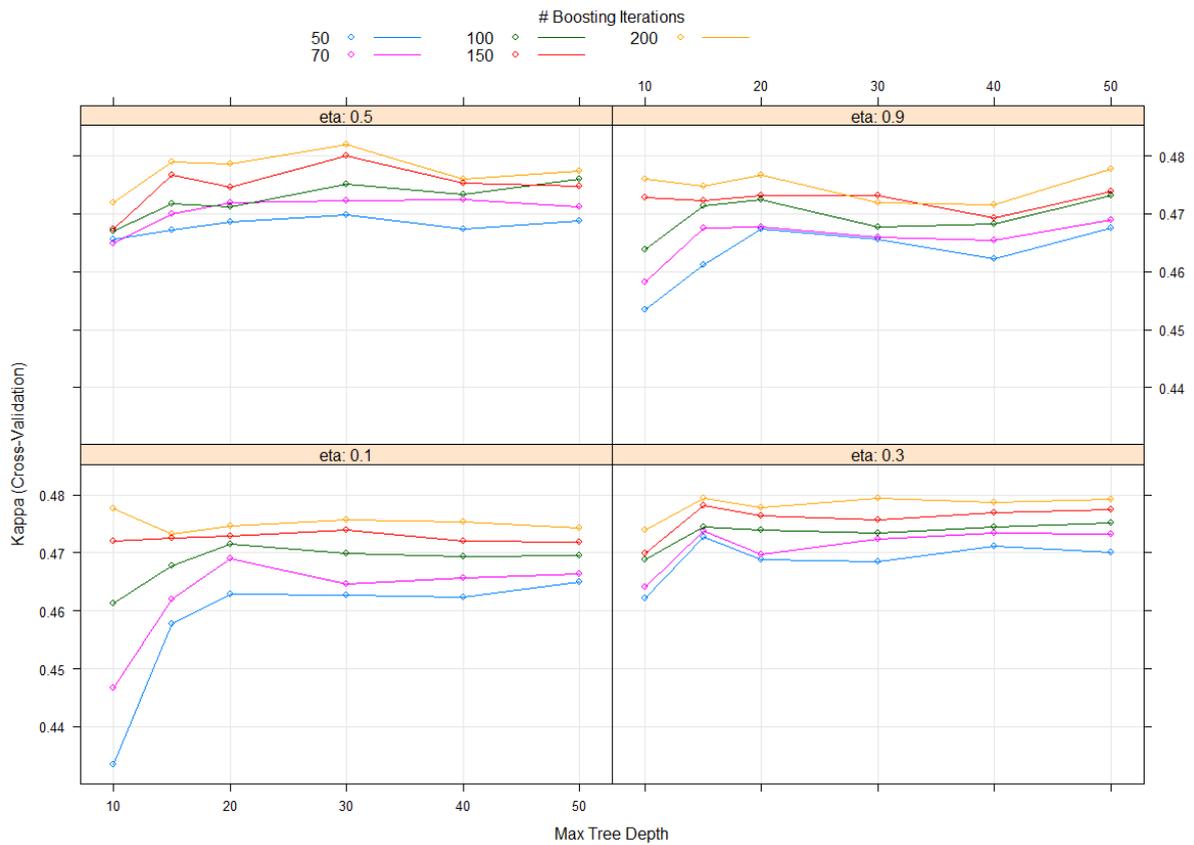


FIGURE 3.15 : Evolution du Kappa selon les hyperparamètres pris par le XgBoost sécheresse

Cela nous permet d'obtenir les hyperparamètres optimaux présentés dans la table 3.10

Paramètre	Valeur optimale
Nombre d'arbres	100
Profondeur maximale	20
Taux d'apprentissage	0.5

TABLE 3.10 : Valeurs des paramètres optimaux

Nous avons ajouté la valeur du Kappa par fold dans l'annexe A.1.

Nous avons désormais un modèle additif ainsi que ses hyperparamètres optimaux. Nous entraînons ce modèle et le testons sur notre base de test, nous donnant une matrice de confusion (table 3.11).

Observations Prédictions	0	1
0	31 155	1 171
1	1 605	1 867

TABLE 3.11 : Matrice de confusion du modèle XgBoost sécheresse

Ce qui nous donne les résultats suivants :

- Rappel = 61,45 %
- Précision = 53,77 %
- F1-score = 57,35 %
- Kappa = 53,11 %

Nous observons une nouvelle fois une augmentation des métriques. Le XgBoost est ici le meilleur des trois types de modèles testés.

## 3.4 Modélisation de l'inondation

### 3.4.1 Modélisation avec GLM pénalisé

De même que pour la sécheresse, nous allons nous concentrer sur le critère d'amélioration de l'AIC afin de déterminer l'importance des différentes variables (figure 3.16).

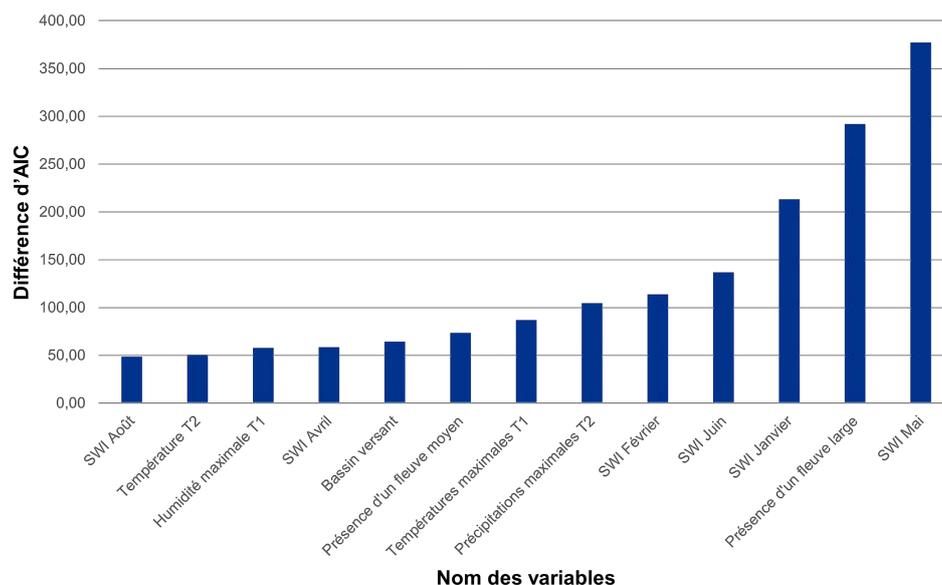


FIGURE 3.16 : Représentation de l'importance des variables pour l'inondation selon le critère de Gini

Ici, il n'y a pas de variable bien plus importante que les autres. Nous remarquons cependant l'importance de la présence d'un fleuve large, ce qui est cohérent pour les inondations par débordement, mais également celle du SWI. En effet, même si le SWI est utilisé comme indicateur pour le déclenchement d'un arrêté sécheresse, l'humidité des sols est corrélée avec les inondations : une très forte humidité peut être associée à des fortes pluies. Cependant les mois concernés ne sont pas les mêmes : là où pour la sécheresse il s'agissait du SWI pour les mois d'octobre ou novembre, ici nous notons la présence des mois de janvier ou d'avril.

Le modèle prend donc la forme suivante (table 3.12).

Variables	
SWI mai+Présence d'un fleuve large+SWI janvier+SWI juin+SWI février+Précipitations maximales T1	
+ Présence d'un fleuve moyen+ Bassin versant+ SWI Avril	
+Humidité maximale T1+Température maximale T2+SWI Août	

TABLE 3.12 : Modèle GLM lasso inondation

Une fois que nous avons sélectionné nos variables, nous pouvons procéder à une *Cross-Validation* à 10 feuilles pour créer notre modèle prédictif. Nous commençons d'abord avec le SMOTE. Nos paramètres sont perc.over=100 et perc.under=600, ce qui représente 25% de 1 dans notre base. Nous montrons la matrice de confusion obtenue (table 3.13).

Observations Prédictions	0	1
0	32 270	1 247
1	1 715	566

TABLE 3.13 : Matrice de confusion du modèle GLM inondation

Résultats :

- Rappel= 31,18 %
- Précision = 24,78%
- F1-score = 27,62%
- Kappa=23,29 %

Nous constatons que le GLM fonctionne nettement moins bien qu'avec la sécheresse, avec un F1-score aux alentours de 27,6 %. Cependant, nous avons vu que le GLM pénalisé est peu satisfaisant pour expliquer la sécheresse, et que la forêt aléatoire ainsi que le XgBoost sont de bien meilleures alternatives. Nous allons donc nous pencher dessus afin d'essayer d'améliorer notre F1-score.

### Modélisation avec la forêt aléatoire

Comme pour la sécheresse, nous procédons à la sélection de variables par coefficient de Gini, puis nous procédons à une cross-validation afin de sélectionner nos meilleurs hyperparamètres.

L'importance des variables ne présentant pas d'éléments notables, nous ne présenterons pas le graphique de Gini ici, qui peut cependant être trouvé en annexe A.1. Nous passons donc directement au modèle choisi ci-présent, dont les variables sont citées par ordre d'importance pour le modèle (table 3.14).

Variables

Températures maximales T1+Humidité maximale T1+Humidité maximale T2+Humidité maximale T4  
 +Humidité moyenne T1+Bassin versant+Températures maximales T2  
 +SWI Juin+Vitesse du vent T4+SWI Octobre  
 +SWI Avril+Températures maximales T4+SWI Mai+Températures maximales T3  
 +Humidité moyenne T3+SWI Septembre+Humidité maximale T3+Humidité moyenne T2  
 +SWI Mars+Précipitations maximales T4

TABLE 3.14 : Modèle Forêt aléatoire inondation

Nous remarquons que les températures ainsi que l'humidité viennent jouer un rôle d'importance dans notre modèle, mais que les précipitations ne sont pas très présentes, ce qui est surprenant. Il est possible que la corrélation avec l'humidité et la température puisse rendre ces variables superflues.

Comme pour la forêt aléatoire sécheresse, nous procédons à l'évaluation du nombre de variables optimal (figure 3.17).

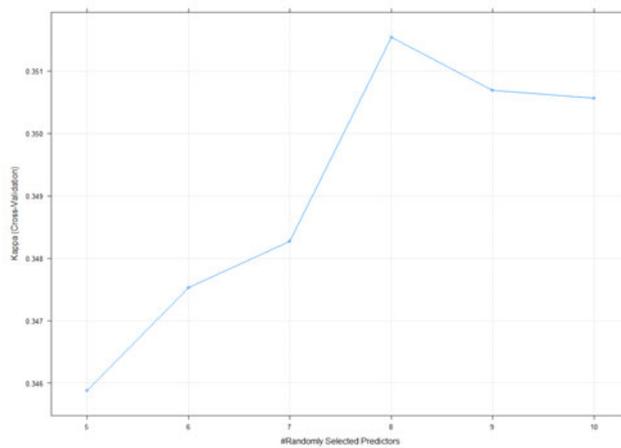


FIGURE 3.17 : Représentation de l'évolution du Kappa selon le nombre d'arbres

Notre nombre de variables optimal est de 8, nous représentons l'évolution du Kappa (figure 3.18).

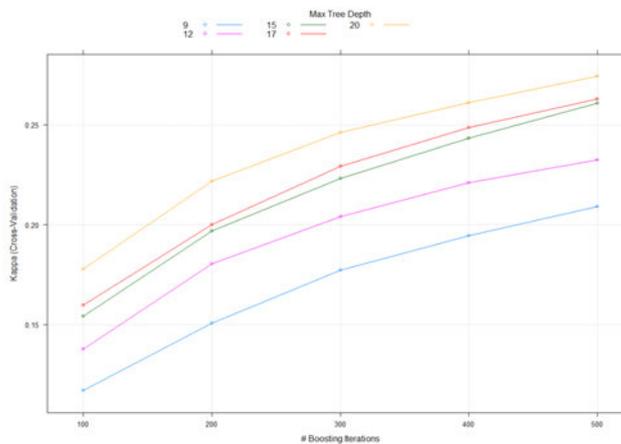


FIGURE 3.18 : Représentation de l'évolution du Kappa selon le nombre d'arbres ainsi que de la profondeur maximale de chaque arbre

Cela nous donne alors le tableau 3.15.

Paramètre	Valeur optimale
Nombre d'arbres	500
Nombre de variables utilisées	8
Profondeur maximale de chaque arbre	9

TABLE 3.15 : Valeurs optimales des paramètres

Comme précédemment, nous appliquons ce modèle sur notre base de test afin d'obtenir notre matrice de confusion (table 3.16).

Classes prédites	Classes réelles	
	0	1
0	33 032	926
1	946	882

TABLE 3.16 : Matrice de confusion de la forêt aléatoire inondation

Résultats :

- Rappel= 48,78 %
- Précision = 48,25 %
- F1-score= 48,51 %
- Kappa= 45,5 %

Si l'amélioration est très nette par rapport au GLM, elle n'est pas pleinement satisfaisante, restant en deça des 50 %.

### 3.4.2 Modélisation avec XgBoost

De nouveau, l'utilisation du gain de précision nous permet de créer notre modèle optimal, que nous représentons ci-dessous (table 3.17). Nous avons cette fois mis l'importance des variables dans l'annexe A.1.

Variables
Température maximale T3 + Température maximale T4 + Température maximale T1 + Bassin versant + Température maximale T2 + Superficie de la commune en aléas faible + SWI Mars + SWI Octobre + Vitesse du vent moyenne T3 + SWI Janvier + Précipitations moyennes T1 + SWI Juin + SWI Juillet + Superficie de la commune en aléas moyen + Humidité maximale T2 + Humidité maximale T3 + Fleuve large + SWI Novembre + SWI Août + SWI Septembre

TABLE 3.17 : Modèle XgBoost inondation

Nous notons que les variables les plus importantes s'avèrent être les variables de températures maximales. Si l'on pourrait penser que ces variables sont surtout essentielles pour la sécheresse, elles peuvent avoir également un effet important sur l'inondation, puisque de fortes températures peuvent être corrélées à de faibles précipitations et donc à une absence d'inondations.

Nous procédons ensuite à l'hyperparamétrage (figure 3.19).

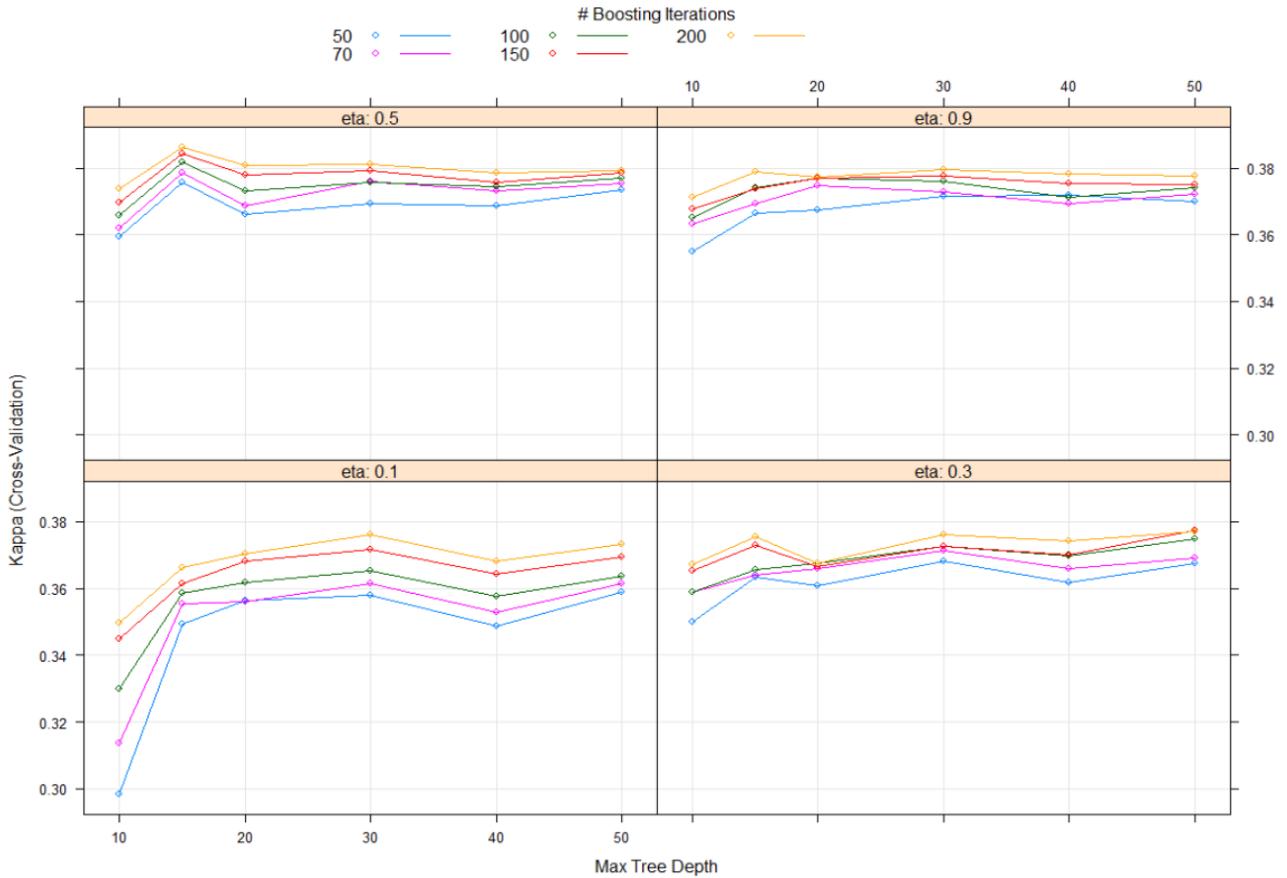


FIGURE 3.19 : Evolution du Kappa selon les hyperparamètres pris par le XgBoost inondation

Paramètre	Valeur optimale
Nombre d'arbres	200
Profondeur maximale	15
Taux d'apprentissage	0.5

TABLE 3.18 : Valeurs optimales des paramètres pour l'inondation

Cet hyperparamétrage nous donne alors les hyperparamètres optimaux ci-dessus (table 3.18).

Nous utilisons alors ces hyperparamètres afin d'établir notre modèle final. Nous présentons la matrice de confusion obtenue (figure 3.20) :

Observations Prédictions	0	1
	0	32 649
1	1331	1027

FIGURE 3.20 : Evolution du Kappa selon les hyperparamètres pris par le XgBoost inondation

Cela nous donne alors les résultats suivants :

- Rappel = 56,49 % ;
- Précision = 43,55 % ;
- F1-score = 49,19 % ;
- Kappa = 46,17 %.

Nous concluons donc que le meilleur modèle pour l'inondation est le XgBoost, comme pour la sécheresse.

De plus, il est intéressant de noter que le SMOTE ne permet pas d'amélioration du F1-score pour ce modèle. En effet, les tests de SMOTE effectués ont causé une diminution du F1-Score de 1 à 2% selon les paramètres sélectionnés. D'une manière générale, le SMOTE s'est avéré décevant, permettant une amélioration seulement pour le GLM lasso sécheresse. Si cela est dommageable, cela se comprend car cette technique ne va pas nécessairement améliorer un modèle. En effet, en ne modifiant que la base d'apprentissage, le SMOTE peut engendrer un problème de sur-apprentissage.

On constate également que la prédiction s'avère bien meilleure pour la sécheresse que l'inondation. Cela peut s'expliquer notamment par le fait que nous ayons pour la sécheresse les données utilisées pour le déclenchement d'un arrêté que sont le SWI et l'exposition aux argiles, là où l'inondation possède des critères de reconnaissance plus complexes comme expliqué précédemment.

### 3.4.3 Backtesting de nos modèles

Maintenant que nous avons testé nos différents modèles, nous pouvons nous intéresser à leur pertinence pour la prédiction du futur. En effet, nous avons depuis le début entraîné nos modèles sur l'ensemble de notre base. Cependant, du fait de la variabilité du climat, cela ne veut pas forcément dire que ces derniers sont adaptés pour prédire les arrêtés CatNat à venir. Pour rappel, notre historique va de 2018 à 2022.

Aussi faisons-nous la division de notre base en deux : d'un côté, l'année 2022 en base de test et toutes les années précédentes en base d'apprentissage. Nous reprenons nos deux meilleurs modèles afin de voir comment ces derniers prédisent la survenance d'arrêtés en 2022. Cela est particulièrement intéressant car l'année 2022 est désormais une année record en terme d'arrêtés sécheresse, il est donc utile d'évaluer la capacité de notre modèle à s'adapter aux années exceptionnelles.

#### Sécheresse

Nous utilisons le modèle XgBoost créé précédemment, puisque nous avons pu voir qu'il nous donnait les meilleurs résultats. Nous testons alors son efficacité sur la base de test constituée, et obtenons les résultats suivants :

- Rappel = 58,06 % ;
- Précision = 28,04 % ;
- F1-score = 37,82 %.

Si le rappel reste à un niveau proche de ce que l'on avait précédemment, nous observons une très grande diminution de la précision du modèle, c'est-à-dire que le modèle prédit beaucoup de faux positifs. Cela peut s'expliquer par le fait que le modèle ait appris sur des années comportant moins de sinistres sécheresse que l'année 2022, qui fut très grave concernant ce péril.

## Inondation

Nous utilisons le modèle XgBoost créé précédemment. Nous testons alors son efficacité sur la base de test constituée, et obtenons les résultats suivants :

- Rappel = 15,11 % ;
- Précision = 7,11 % ;
- F1-score = 12,2 %.

Ici le modèle n'est pas bon, échouant à prédire l'année 2022 de façon correcte, avec un f1-score de seulement 12 %, ce qui est très faible. Cela peut s'expliquer encore une fois par une faible profondeur historique, mais également via le fait que les zones géographiques concernées par les phénomènes inondations sont assez différentes selon l'année.

Cependant cette explication n'est pas pleinement satisfaisante puisque la sécheresse présente aussi ce genre d'obstacles. Là où nous constatons une différence plus importante, c'est lorsque nous nous focalisons au niveau des communes touchées. Là où les communes concernées par la sécheresse ont pour une bonne partie déjà connu des arrêtés sécheresse, notamment en 2018, ce n'est pas le cas pour l'inondation, ce qui fait que le modèle est moins bien calibré sur les variables associées aux communes concernées. Un autre élément d'importance pourrait concerner la multitude des critères de reconnaissance des inondations, lesdits critères étant bien plus nombreux que pour la sécheresse.

### 3.4.4 Comparaison des différents modèles, état de l'art et pistes d'améliorations

Nous comparons ici chacun de nos modèles afin de sélectionner le meilleur. Nous commençons avec les modèles sécheresse (table 3.19).

	Rappel	Précision	F1-score	Kappa
GLM pénalisé	57,93 %	38,38 %	46,17 %	40,05 %
Forêt aléatoire	53,39 %	61,00%	56,94%	53,24 %
XgBoost	61,45 %	53,77 %	57,35 %	53,11 %

TABLE 3.19 : Comparaison des différents modèles de sécheresse

Puis avec les modèles d'inondation (table 3.20).

	Rappel	Précision	F1-score	Kappa
GLM pénalisé	31.18%	24.78%	27.62%	23.29%
Forêt aléatoire	48,78 %	48,25%	48,51%	45,5%
XgBoost	56,49 %	43,55 %	49,19 %	46,17%

TABLE 3.20 : Comparaison des différents modèles d'inondation

Finalement, nous retenons donc les modèles XgBoost pour l'inondation et la sécheresse car ces derniers ont le meilleur F1-Score.

Nous avons donc deux modèles aux F1-score respectifs de 49% et 57%. Le fait de prédire correctement les arrêtés par commune est important, car du fait de la faible exposition de notre assureur, l'évolution de la probabilité d'arrêt CatNat sera pour lui plus localisé. Le score de nos modèles est donc peu satisfaisant. Cependant, nous pouvons nuancer cela par le fait que nos modèles seront projetés selon le pourcentage de survenance d'un arrêté. Nous pouvons donc nous interroger sur la capacité du modèle à prédire correctement le pourcentage de communes en arrêté au niveau national.

Voici la comparaison des nos prédictions et des observations d'arrêtés :

- La base contient 5,048 % d'arrêtés inondations. Le modèle prédit quant à lui 5,057 % d'arrêtés.
- La base contient 8,485 % d'arrêtés sécheresse. Le modèle prédit quant à lui 8,47 % d'arrêtés.

La différence s'avère donc très faible entre la probabilité moyenne prédite et la probabilité moyenne observée, ce qui est un bon point pour notre projection aux différents horizons.

Il peut également être intéressant de comparer nos résultats à ceux d'études ayant effectué un travail similaire au nôtre. Ainsi, nous pouvons citer la thèse HERANVAL (2022), qui a procédé à une modélisation des arrêtés CatNat sécheresse sur la période 2003-2018. Voici les F1-score présentés dans l'étude pour différents modèles :

- GLMnet : 50,3 % ;
- Forêt aléatoire : 57 % ;
- XgBoost : 57,3 % ;
- Aggregate : 57,6 % .

Nous remarquons que le F1-score est très proche de ce que nous avons obtenu, et que là aussi nous observons une augmentation du F1-score en passant du GLM au XgBoost.

Pour l'inondation, nous n'avons pas trouvé d'études faisant intervenir de métriques similaires aux nôtres, ce qui rend la comparaison difficile.

Nos scores sont, au moins pour la sécheresse, dans la lignée de ce qui a pu être fait précédemment. Cependant ces derniers restent améliorables. Nous allons donc évoquer des pistes d'améliorations pour ces derniers :

- L'ajout de nouvelles variables. Nous avons déjà beaucoup de variables dans nos bases, et en ajouter serait sans doute source de confusion. Cependant certaines variables, comme l'imperméabilité des sols ou des variables liées à la population des villes pourraient être intéressantes. Certaines variables liées au SWI pourraient également être utiles, comme par exemple prendre le maximum sur certains mois au lieu de la moyenne.
- Nous pourrions également nous concentrer sur certaines régions comme par exemple l'Ouest où des arrêtés inondations sont très souvent reconnus, ce qui nous permettrait d'avoir une plus grande base à disposition. Cependant cela ne correspondrait pas au territoire couvert par notre assureur, et introduirait un biais pour la prédiction quant au reste du territoire.

### 3.5 Projections du risque aux horizons 2030, 2040 et 2050

Nous avons désormais deux modèles afin de prédire les arrêtés, qu'ils soient sécheresse ou inondation. Cependant cela est insuffisant, car il nous faut désormais nous projeter aux différents horizons prévus. Notre objectif est donc de projeter la probabilité d'arrêtés CatNat aux différents horizons considérés. Nous supposons tout au long de cette partie que les normes de reconnaissance en arrêté CatNat n'évolueront pas. En effet, la reconnaissance en arrêté CatNat se base sur différents critères météorologiques et géologiques comme expliqué dans la partie 1.1.5. Ces critères sont donc susceptibles d'évoluer avec le temps, cependant nous resterons dans l'hypothèse qu'ils resteront fixes. Cela est dû à la difficulté de les projeter, qui dépendront notamment d'une volonté politique pouvant les faire évoluer fortement.

Pour cela, nous devons récupérer des données climatiques projetées à nos horizons d'intérêts. Nous nous intéressons aux données du DRIAS MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE (2023) (Donner accès aux scénarios climatiques Régionalisés français pour l'Impact et l'Adaptation de nos Sociétés et environnement) : ces dernières proposent des projections pour le vent, la température, les précipitations

et enfin l'humidité. Cela nous intéresse car il s'agit de données que nous avons incluses dans notre base de modélisation. Présentons rapidement le DRIAS et son travail.

Le DRIAS est un organisme du ministère de la transition écologique visant à rendre disponible des simulations climatiques au grand public. Pour cela, le DRIAS utilise différents types de modèles climatiques, un régional et l'autre global. Tout d'abord, le modèle climatique global vise à représenter les différentes interactions entre les différentes parties de la Terre. Pour cela, ce dernier se base sur une maille découpant la planète en carré de côtés d'environ 100 kilomètres. A partir de ces modèles, les scientifiques peuvent élaborer des modèles régionaux, plus précis et qui tiennent ainsi compte d'éléments ignorés par le modèle global, comme le relief. Voici un graphique résumant le fonctionnement dudit modèle (figure 3.21).

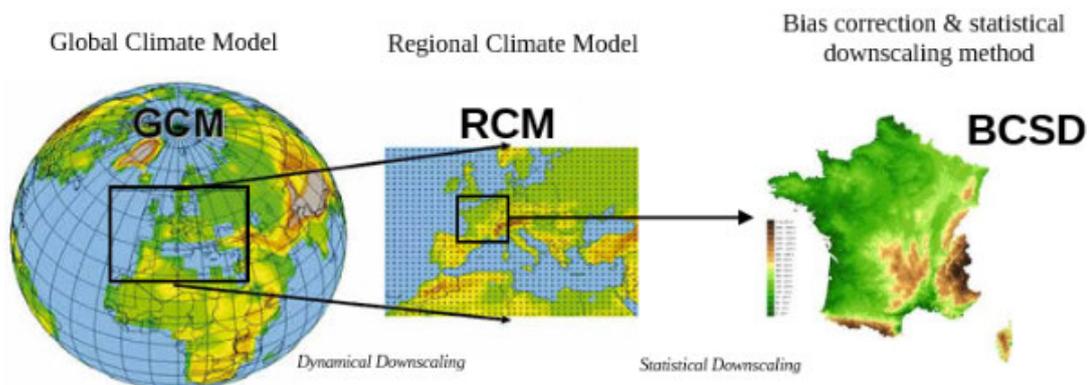


FIGURE 3.21 : Présentation du passage du modèle global au modèle régional DRIAS (2014)

Dans la section 2.2, nous avons identifié 3 éléments d'importance quant il s'agit de sélectionner des données Open Source, à savoir la maille, la temporalité et le nombre de données accessibles. Intéressons-nous à ces 3 critères pour le DRIAS :

- La maille. Le DRIAS propose une maille de 8 kilomètres sur 8 kilomètres, ce qui représente 8981 points équi répartis sur le territoire métropolitain. Il s'agit donc d'une maille très précise, plus que les données du DRIAS que nous avons utilisé. C'est donc satisfaisant.
- La temporalité. Nous nous penchons sur la projection dans le futur jusqu'à 2050, et le DRIAS propose des simulations jusqu'en 2100, avec possibilité de sélectionner les années. Cela est donc satisfaisant pour nous.
- Le nombre de variables. Comme dit précédemment, nous voulons procéder à la projection des variables températures, humidité, précipitations et vitesse du vent. Le DRIAS proposant chacune de ces variables, c'est encore une fois un bon point pour nous.

Nous allons donc prendre 30 ans de projection, de 2025 à 2055, et nous concentrer sur les horizons nous intéressant, à savoir 2030, 2040 et 2050. Afin d'avoir un historique suffisant, nous prendrons une moyenne sur 5 ans pour chaque horizon, c'est-à-dire entre 2028 et 2032 pour 2030, 2038 et 2042 pour 2040, et enfin 2048 et 2052 pour 2050. Nous ferons alors une moyenne globale de tous ces éléments. Cela nous permettra de minimiser les aléas liés au modèle et d'avoir la simulation la plus stable possible.

Représentons la différence de nos moyennes en fonction du scénario ainsi que de l'horizon pour comprendre comment le changement climatique vient modifier ces paramètres. Voici la façon dont les températures évoluent en fonction des horizons H1(2030), H2(2040) et H3(2050) 3.22.

Nous constatons une nette évolution positive de la température maximale. Pour le scénario 4.5, nous observons jusqu'à 2°C de hausse par rapport à notre période de référence (2018-2022), ce qui

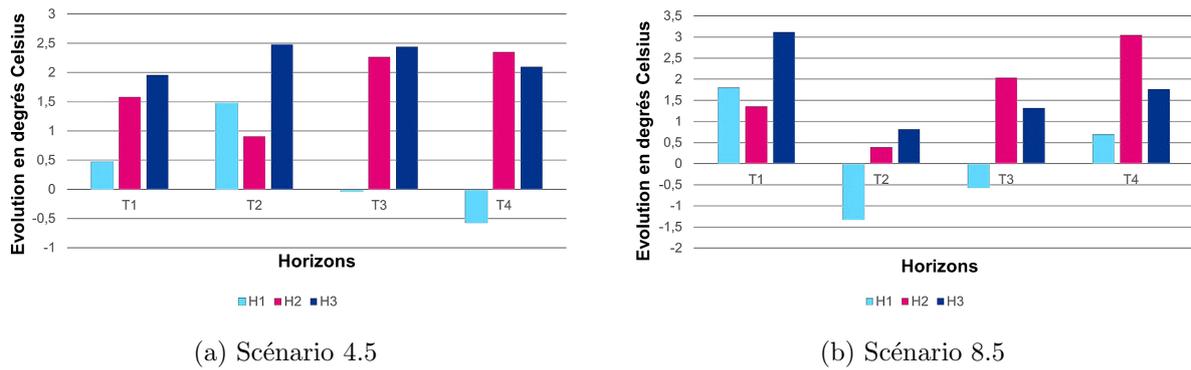


FIGURE 3.22 : Représentation de la différence de température entre les horizons et la période de référence (2018-2022)

est cohérent avec le GIEC. La principale surprise vient du scénario 8.5 pour l'horizon 2050, où nous observons une claire baisse des températures pour les troisième et quatrième trimestres. Cependant, même en moyennant sur une période de 5 ans, il est possible que l'évolution du climat ne soit pas parfaitement linéaire.

Les autres données sont celles d'hydrographie, d'exposition aux argiles et enfin le SWI.

L'hydrographie peut ne pas être projetée. En effet, même s'il est possible que certains plans d'eau disparaissent du fait du changement climatique, cet impact ne devrait pas être suffisant pour causer la disparition de mares, lacs ou fleuve de taille moyenne ou large. De plus, modéliser une disparition de ces éléments serait source de complexité et risquerait d'être très contestable.

Les données du BRGM resteront fixes, puisque la composition du sol ne change pas avec le changement climatique.

Il nous reste donc à réfléchir sur la projection du SWI, élément de grande importance pour la sécheresse comme pour l'inondation. Pour cela, nous pouvons utiliser une nouvelle fois le DRIAS, qui propose des projections de ce dernier jusqu'en 2100 et selon les scénarios du GIEC. Dans ce cas là, nous utiliserons la même méthode que pour les variables atmosphériques, à ceci près que nous nous baserons sur les années hydrologiques, qui vont d'août à juillet. Par exemple, nous ferons une moyenne d'août 2027 à juillet 2032 pour l'horizon 2030.

Pour récapituler, nous avons dans chaque base climatique les variables d'humidité, de précipitations, de vitesse du vent et de température par trimestre, les variables d'hydrologie, de composition des sols et enfin le SWI pour chaque mois. Cela est une bonne nouvelle car la majorité de nos variables d'importance ont pu être projetées aux différents horizons.

### 3.6 La sinistralité en présence d'un arrêté CatNat : données actuelles et projections.

Afin d'estimer la sinistralité climatique projetée de l'assureur, nous avons besoin de trois éléments :

- La probabilité de survenance d'un arrêté CatNat à l'horizon voulu ;
- Le nombre de sinistres engendrés par la survenance d'un arrêté. Pour cela, nous prenons le Grand Livre des sinistres évoqué dans la section 2.1.2 de l'assureur afin de calculer un taux de destruction.
- Le coût moyen d'un sinistre climatique. Les chiffres actuels sont fournis par la CCR ou France Assureurs, nous appliquons un facteur d'inflation à ces chiffres afin de matérialiser l'augmentation de la sinistralité.

Cela nous donne donc la formule suivante pour estimer l'évolution de la charge sinistre

$$Charge = \sum_{i=1}^n \mathbb{P}(\text{Arrêté}) \times N \times S,$$

avec :

- $N$  = nombre de dossiers de sinistralité engendrés par un arrêté CatNat ;
- $n$  = le nombre de villes présentant un historique de sinistralité pour notre assureur ;
- $S$  = Sinistralité moyenne par dossier.

### 3.6.1 Projection du coût d'un sinistre via l'inflation

Pour le facteur d'inflation, nous utilisons un fichier fourni par l' ACPR (2023b) pour le scénario de long terme de l'exercice climatique de 2023. Ce dernier prend la forme d'un tableau fournissant un taux d'inflation pour chaque année jusqu'en 2050. Nous avons d'abord une inflation de 6,2 % pour l'année 2022 basée sur l'inflation réelle s'étant produite cette année, puis le scénario suppose un retour progressif à une inflation modérée aux alentours de 2 %, avec une moyenne 2023-2050 s'établissant à 2,03 %. Nous représentons cette évolution sur la figure 3.23.

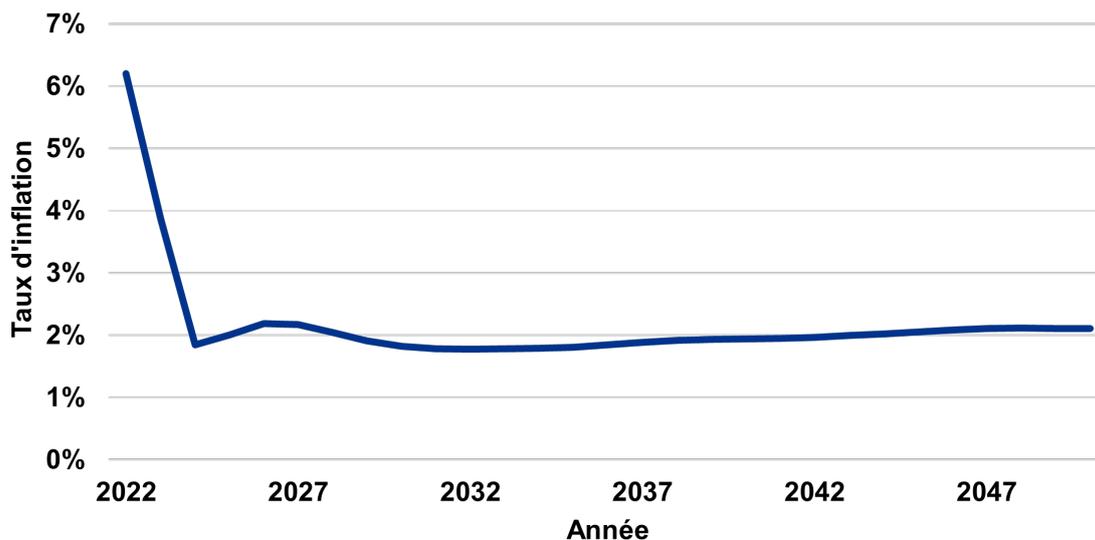


FIGURE 3.23 : Représentation de l'inflation en France selon le scénario long terme ACPR (2023b)

Pour rappel, le coût moyen d'un sinistre RGA est estimé par la cour des comptes à 16 300 € sur la période 1989-2019, ce qui, en appliquant le taux d'inflation, nous donne un coût moyen de 30 382 € en 2050, ce qui est une augmentation très significative. Cette courbe d'inflation est cependant contestable, notamment car elle suppose un retour à une faible inflation dès 2023, ce qui ne semble pas se confirmer pour le moment, et que cette dernière restera peu élevée jusqu'en 2050.

### 3.6.2 Projection de la probabilité d'arrêt CatNat aux différents horizons

Nous disposons désormais de bases pour chacun des horizons et scénarios d'intérêts contenant nos variables climatiques. Nous allons maintenant prendre les meilleurs modèles trouvés et les appliquer auxdites bases pour effectuer une prédiction des arrêts. Nous allons pour cela utiliser le type

«prob» de la fonction *predict* de R CORE TEAM (2023), qui renvoie la probabilité d'appartenance à une classe des individus, ici 0 ou 1.

Concernant les variables, nous nous baserons sur les données du GIEC évoquées précédemment, ainsi que sur les données ne changeant pas avec le temps : les données hydrographique et la composition du sol.

Nous avons dans notre base 8,5 % de communes en arrêté sécheresse et 5,05 % en arrêté inondation. Nous avons vu précédemment que la probabilité moyenne prédite était très proche de la réalité, notre base de départ est donc satisfaisante.

Nous allons présenter ici l'évolution de la probabilité selon l'horizon, le scénario et le péril. Afin de prendre en compte les possibles erreurs, nous avons créé 1000 bases par bootstrap afin de créer un intervalle dont nous présenterons ici deux tableaux contenant les prédictions moyennes maximales et minimales ainsi que la médiane, qui sera l'élément que nous retiendrons pour notre projection. Nous mettrons également l'erreur standard de notre échantillon dans un tableau pouvant être retrouvé en annexe A.1 . Voici les tableaux contenant les valeurs de nos projections, avec en parenthèse l'intervalle d'erreur calculé.

Sécheresse	RCP 4.5	RCP 8.5	Inondation	RCP 4.5	RCP 8.5
Référence (2018-2022)	8,4 %	8,4 %	Référence (2018-2022)	5,05 %	5,05 %
Horizon 2030	9,37 (9,03-9,62) %	10,4 (10,07-10,65) %	Horizon 2030	5,7 (5,27-6,13) %	7,51 (6,98-7,86) %
Horizon 2040	14,1 (13,67-14,75) %	15,58 (15,15-16,07) %	Horizon 2040	6,35 (5,97-6,91) %	10,47 (9,96-10,97) %
Horizon 2050	18,31 (18,01-18,61) %	19,34 (18,84-19,87) %	Horizon 2050	8,3 (7,81-8,84) %	12,09 (11,43-12,7) %

TABLE 3.21 : Evolution de la probabilité d'arrêt inondation/sécheresse par commune en France Métropolitaine selon le scénario et l'horizon

Représentons maintenant l'évolution de la probabilité d'arrêtés sécheresse/inondation selon les horizons (figure 3.24).

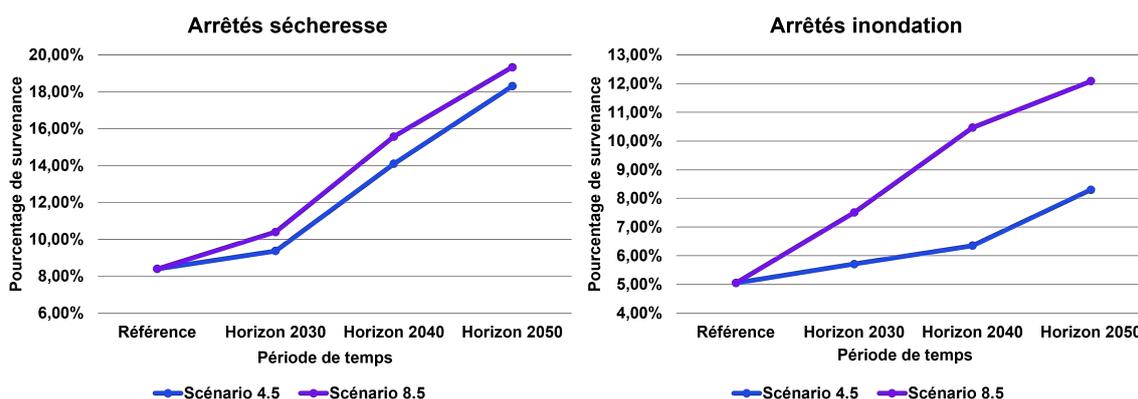


FIGURE 3.24 : Evolution de la probabilité d'arrêtés sécheresse/inondation par commune en France Métropolitaine selon le scénario et l'horizon

Pour la sécheresse, nous observons ici une importante augmentation de la probabilité de survenance. Nous notons que cette évolution dépend non seulement de l'horizon, mais également du scénario. Cependant, la différence entre les deux scénarios est assez faible. Cela est dû au fait qu'à l'horizon 2050, les deux scénarios sont encore assez peu différenciés, la différence commençant à se faire ressentir

au-delà de cet horizon.

Pour l'inondation, la différence s'avère en revanche plus marquée, ce qui peut être dû à une plus forte hausse du niveau de la mer, un élément qui s'avère assez différent selon chaque scénario.

Notre modèle s'avère donc cohérent concernant l'évolution des périls à un niveau national. Cependant, il est intéressant de représenter à un niveau plus fin les différentes cartes représentant l'évolution de la probabilité de survenance d'un arrêté CatNat entre notre période de référence et l'horizon 2050. Si nos prédictions ont été faites à un niveau communal, nous représentons ici lesdites cartes par département pour des raisons de lisibilité (figures 3.25 et 3.26).

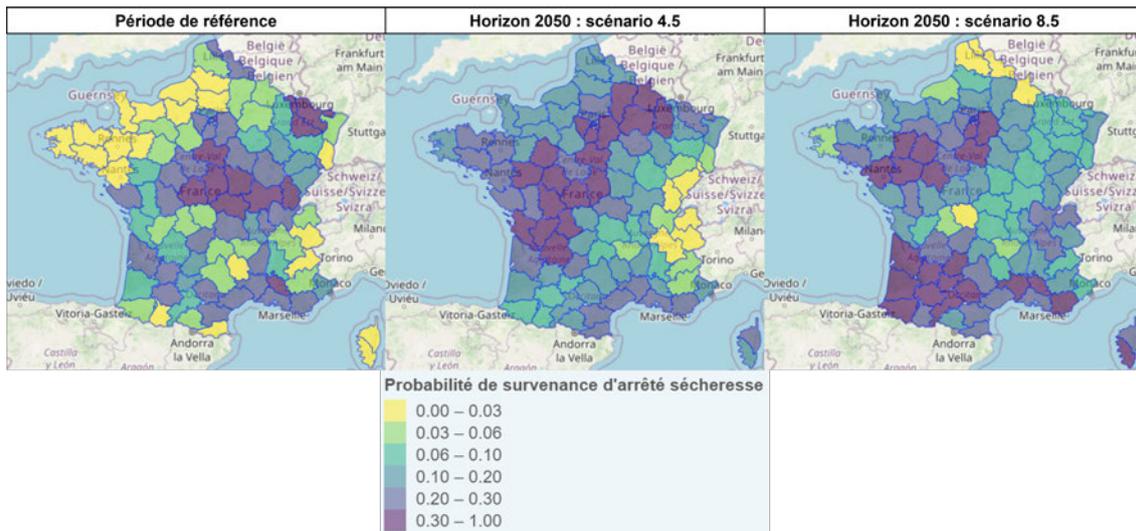


FIGURE 3.25 : Evolution de la probabilité de survenance d'un arrêté sécheresse en France : à gauche la période de référence, au milieu l'horizon 2050 pour le scénario 4.5 et à droite pour le scénario 8.5

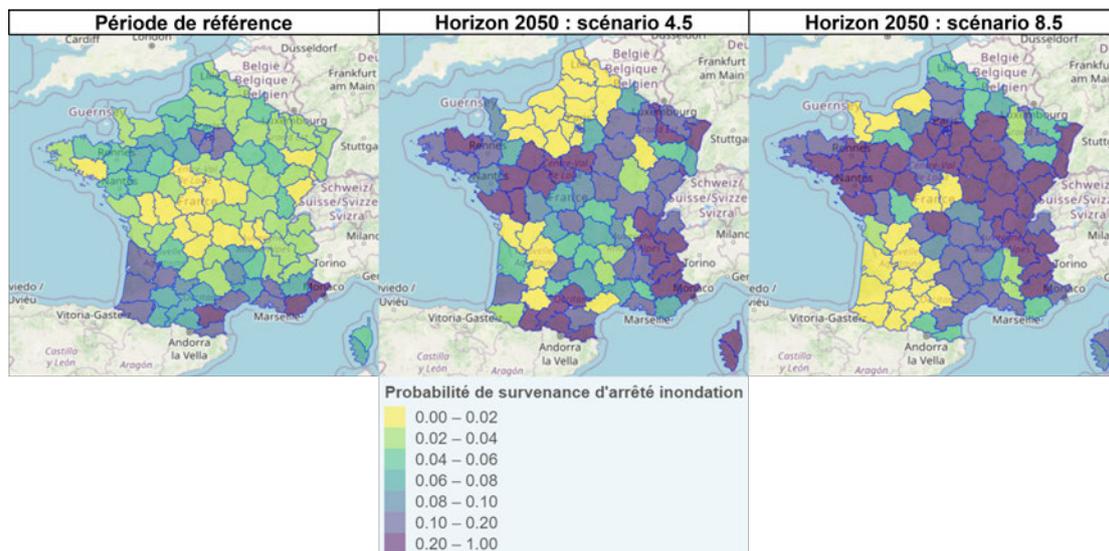


FIGURE 3.26 : Evolution de la probabilité de survenance d'un arrêté inondation en France : à gauche la période de référence, au milieu l'horizon 2050 pour le scénario 4.5 et à droite pour le scénario 8.5

Pour la sécheresse, nous observons une importante similarité entre les cartes, avec évidemment une forte augmentation des probabilités de survenance à horizon 2050. Dans le scénario 4.5 en particulier, le centre s'avérera très touché comme pour la période de référence, ce qui est explicable par la présence d'argiles dans la région. Pour le scénario 8.5, il est intéressant de voir que le Sud-Ouest sera très touché, ce qui est cohérent avec les épisodes de sécheresse récents où ce département fut très touché par le risque sécheresse.

Pour la carte inondation, nous constatons que le scénario 8.5 entraîne bien une sur-représentation du nombre de départements à forte probabilité d'incidence. C'est notamment le cas pour l'Île-de-France, qui s'avère bien plus touchée dans le scénario 8.5 que 4.5. En revanche, le Sud-Ouest est très peu touché dans ce scénario. Le fait que des départements limitrophes aient la même classe de probabilité peut s'expliquer par le fait que nous utilisons de nombreuses variables météorologiques, qui en moyenne peuvent ne pas être très différentes dans une région donnée, en particulier pour la température ou les précipitations.

Il peut également être intéressant de mettre en parallèle nos résultats avec ceux d'autres études. Concernant la sécheresse, nous nous baserons sur l'étude effectuée par COVÉA (2022), à laquelle nous avons fait déjà référence dans la section 1.2.2. Cette étude se penche sur la fréquence d'éligibilité moyenne des communes aux arrêtés CatNat via l'étude d'un indice d'humidité du sol dépassant une durée de retour de 25 ans, ce qui est un des critères pour la survenance des arrêtés sécheresse. Dans un scénario 8.5 à horizon 2050, ces derniers ont conclu à une hausse moyenne de la fréquence d'éligibilité de 70 %. Rapporté à notre étude, cela donnerait une évolution allant d'une probabilité de 8,4 % à une probabilité de 14,28 %. Pour rappel notre étude avait conclu à une évolution allant jusqu'à 19,34 %. Cette différence peut s'expliquer par la période historique, qui va de 2008 à 2018 pour Covéa et n'intègre donc pas les épisodes de sécheresse récents.

Pour l'inondation, nous pouvons également nous pencher sur l'étude de Covéa, bien que cette dernière étudie de façon plus indirecte la survenance d'arrêté CatNat. Nous avons vu que les critères de survenance d'arrêté CatNat se basaient sur des précipitations décennales. Or dans ce cadre Covéa a évalué que les précipitations décennales se produiraient tous les 4,5 ans en moyenne. Une modification aussi importante impacterait considérablement à la hausse le nombre de communes pouvant être reconnues en arrêté CatNat inondation, potentiellement en faisant plus que doubler ce nombre. Pour rappel, la probabilité d'arrêté CatNat que nous avons calculée pour le scénario 8.5 à horizon 2050 représentait une hausse de 139 % par rapport à la période de référence. Nous nous positionnons donc dans une perspective proche à celle calculée dans le livre blanc de Covéa.

Maintenant que nous avons étudié les évolutions climatiques et leur impact sur la survenance d'arrêtés, nous procédons à l'étude de la sinistralité de notre assureur ainsi qu'à la réalisation d'un ORSA à partir des évolutions remarquées.

## Chapitre 4

# Etude de la sinistralité assurantielle et réalisation de l'ORSA climatique

Notre objectif dans ce chapitre est d'utiliser l'augmentation de la probabilité de survenance d'arrêts climatiques afin de calculer une augmentation de sinistralité pour les périls sécheresse et inondations. Une fois cela fait, nous voulons utiliser cette hausse afin d'impacter le bilan assurantiel dans le cadre de l'ORSA.

### 4.1 Impact du changement climatique sur la charge moyenne

Nous allons désormais procéder au calcul de l'évolution de la charge moyenne de la sécheresse et de l'inondation pour l'assureur. Pour rappel, nous nous basons sur trois éléments :

- La probabilité d'arrêt climatique, projetée dans la partie 3.6.2;
- Le nombre de dossiers générés en moyenne par un arrêt;
- Et enfin le coût moyen d'un dossier CatNat pour l'assureur.

Comme dit dans le chapitre 2, nous concentrerons notre étude sur la sinistralité climatique de 3 départements.

#### 4.1.1 Impact du changement climatique sur la sinistralité

Nous procédons d'abord à un traitement de la base des sinistres. Pour cela, nous associons ensemble les différents paiements associés à chaque sinistre et effectuons une somme du règlement de l'année, des règlements des années précédentes et la provision du sinistre afin d'avoir une estimation la plus complète possible de la charge associée.

Une fois cela fait, nous regardons la date de survenance de sinistre et l'associons avec l'arrêt correspondant.

Enfin, nous rassemblons les différents sinistres par commune. Nous notons alors le nombre total d'arrêts survenus dans lesdites villes.

Deux éléments restent à considérer : l'évolution du bilan, facteur d'évolution du nombre de dossiers par sinistre ainsi que l'inflation des coûts des sinistres, facteur de l'évolution de leur coût moyen.

Nous nous basons sur une hypothèse de bilan statique. En effet, l'assureur que nous étudions n'a pas forcément d'importantes évolutions de portefeuille à venir, et projeter de façon trop imprécise rendrait notre estimation hasardeuse. Il est donc préférable que nous restions dans une hypothèse de bilan statique.

Il nous faut cependant appliquer aux sinistres le choc d'inflation afin de prendre en compte l'augmentation de leur coût. Voici le tableau récapitulant l'inflation du coût des sinistres aux horizons d'intérêt (tableau 4.1). Pour des raisons de confidentialité, nous travaillons en base 100.

Horizon	Coût
Période de référence (2018-2022)	100
Horizon 2030	127
Horizon 2040	152
Horizon 2050	186

TABLE 4.1 : Sinistralité CatNat en y appliquant un facteur d'inflation

Nous mesurons ensuite la sinistralité annuelle moyenne de chaque péril afin d'avoir une comparaison par rapport aux résultats du modèle. De nouveau, nous restons sur une base 100 de sinistralité.

Par exemple, pour la sécheresse à horizon 2030 du scénario 4.5 le calcul prend la forme suivante

$$\begin{aligned} & \text{Sinistralité moyenne projetée en 2030} \times 9.37 \% (\text{Probabilité d'arrêt sécheresse à horizon 2030}) \\ & \quad \times \text{Nombre de villes} \times \text{Nombre de sinistres par ville} \end{aligned}$$

et pour l'inondation cette forme là :

$$\begin{aligned} & \text{Sinistralité moyenne projetée en 2030} \times 5.7 \% (\text{Probabilité d'arrêt inondation à horizon 2030}) \\ & \quad \times \text{Nombre de villes} \times \text{Nombre de sinistres par ville} \end{aligned}$$

Nous avons alors les tableaux 4.2 qui décrivent l'évolution de la sinistralité en référence à la période historique 2018-2022. Cela veut dire que si le résultat est de 120 %, la sinistralité moyenne sera 20 % supérieure à celle observée durant la période de référence.

Sécheresse	RCP 4.5	RCP 8.5
Horizon 2030	103 %	115 %
Horizon 2040	117 %	159 %
Horizon 2050	250 %	269 %

(a) Evolution de la sinistralité annuelle du péril sécheresse

Inondation	RCP 4.5	RCP 8.5
Horizon 2030	84 %	110 %
Horizon 2040	112 %	185 %
Horizon 2050	180 %	258 %

(b) Evolution de la sinistralité annuelle du péril inondation

TABLE 4.2 : Evolution de la sinistralité annuelle des périls sécheresse et inondation

Les tableaux 4.2 confirment l'impact important de l'horizon ainsi que du scénario choisi, puisque dans le cadre de l'inondation, nous observons une différence manifeste. Si pour la sécheresse cet impact est moins important, il n'en demeure pas moins clair. Nous observons pour la sécheresse une hausse de 269 % de la sinistralité climatique d'ici à 2050 et une hausse de 258 % pour l'inondation. Ces chiffres, très considérables, montrent de façon nette l'importance du changement climatique pour les assureurs.

#### 4.1.2 Modélisation de la sinistralité selon le type de bâtiment

Nous avons maintenant une médiane pour l'évolution de la sinistralité du portefeuille assurantiel. Cependant, nous souhaiterions également pouvoir modéliser les différences d'impact selon le type de bâtiment. Il nous faut donc pour cela séparer notre base de sinistralité selon le type de bâtiment mais

également selon l'exposition. En faisant cela, nous aurons un coût de sinistralité moyen et un nombre de dossiers par arrêté CatNat différents selon le type de bâtiment.

Nous pouvons distinguer 3 types de bâtiments : maison, appartement ou immeuble. Nous procédons alors de la sorte : tout d'abord, nous utilisons la variable "Nature du bâtiment" de notre base de données afin de la diviser en 3 : les sinistres associés à des immeubles, ceux associés à des appartements et enfin ceux des maisons. Comme expliqué précédemment, nous utilisons la date pour associer le sinistre à un arrêté CatNat, puis calculons le nombre de sinistres par arrêté puis nous calculons la sinistralité moyenne. Pour résumer, nous avons créé trois bases pour chaque type de bâtiment et recalculé pour chacune d'entre elle le nombre de sinistre par arrêté et la sinistralité moyenne.

Cependant, la sinistralité appartement étant peu importante du fait de la structure du portefeuille, nous choisirons de nous concentrer sur les maisons et les immeubles.

Nous présentons les résultats à horizon 2050 dans les tableaux suivants 4.3 :

Sécheresse			Inondation		
Type de bâtiment	RCP 4.5	RCP 8.5	Type de bâtiment	RCP 4.5	RCP 8.5
Maison	247 %	261 %	Maison	131 %	191 %
Immeuble	223 %	236 %	Immeuble	201 %	304 %

TABLE 4.3 : Sinistralité CatNat à horizon 2050 selon le type de bâtiment

Nous observons sur les tableaux 4.3 que dans le cadre de la sécheresse une assez faible différence de l'évolution de la sinistralité est constatée, là où pour les inondations la différence entre maison et immeuble est manifeste. Cela peut s'expliquer notamment par le fait que les immeubles assurés se concentrent plus dans des villes d'importance traversées par un fleuve ou une grande rivière, ce qui les rend plus exposés aux inondations. Concernant le fait que la sinistralité totale sécheresse soit supérieure à la sinistralité combinée des maisons et immeubles, cela est dû à la non-représentation des appartements qui font tendre la sinistralité à la hausse. De fait, il peut être intéressant de nous interroger sur l'impact de l'exposition du portefeuille sur le nombre de dossiers engendrés par un arrêté CatNat et la sinistralité qui y est associée. Pour cela, nous calculons l'exposition de l'assureur pour chaque ville, puis nous séparons notre base en deux via la médiane des expositions par ville. Pour la sécheresse l'impact est le suivant : le nombre de dossiers est multiplié par un facteur de 1,53, et la sinistralité par un facteur de 3,15 entre la classe la moins exposée et celle concentrant la majeure partie de l'exposition. Pour l'inondation, le nombre de dossiers est multiplié par un facteur de 1,1, et la sinistralité par un facteur de 1,65. Cette variable a donc un impact conséquent sur la façon dont nos chiffres évoluent.

Interrogeons-nous sur la plausibilité des résultats obtenus. Pour cela, nous comparons notre médiane avec des résultats d'autres études afin d'en estimer la cohérence. Selon une étude du BRGM à destination du SÉNAT FRANÇAIS (2023), la sinistralité sécheresse cumulée entre 2020 et 2050 représentera un coût de 43 milliards d'euros, soit le triple des trois décennies précédentes. Nous sommes donc assez proches de ce résultat, puisque notre étude donne une multiplication par un facteur 2,7 de la sinistralité sécheresse.

Concernant l'inondation, le livre blanc de covéa COVÉA (2022) prédit quant à lui une augmentation de 110 % de sinistralité pour les crues de plaine et 130 % pour les crues éclair à l'horizon 2050 et selon le scénario 8.5. Notre augmentation est quant à elle de 158 %, ce qui s'avère plus élevé mais reste dans un ordre d'idée similaire.

Nos résultats s'avèrent donc plausibles, interrogeons-nous maintenant sur les conséquences possibles de ces derniers sur l'assureur.

## 4.2 Analyse de l'impact du changement climatique sur le bilan

Nous nous demandons désormais comment l'évolution de la sinistralité que nous venons de calculer va pouvoir impacter le bilan de l'assureur. Pour cela nous voulons observer la façon dont le bilan sera impacté par le changement climatique.

### 4.2.1 Présentation de la situation de base de notre assureur

Nous commençons par expliciter la situation de départ de notre assureur, c'est à dire présenter son bilan et montrer quelles parties de ce dernier pourraient être sujettes à l'impact du réchauffement climatique. Nous avons déjà pu présenter la norme Solvabilité 2 dans la partie 1.3, aussi nous concentrons-nous ici sur les éléments les plus importants pour notre étude.

Nous nous sommes placés au cours de cette étude dans une étude du passif via la sinistralité. Lors de la section 1.4.1, nous avons vu que le changement climatique peut impacter les fonds propres via le SCR ou les provisions techniques via le *Best Estimate*. Cependant, le *Best Estimate* correspondant à la valeur actualisée des flux de cashflow futur, il semble complexe de le projeter jusqu'à l'horizon 2050 avec nos calculs de charges de sinistralité. En revanche, le SCR correspond tout à fait à ce que nous souhaitons étudier, son calcul faisant intervenir la distribution de la sinistralité et disposant de sous-modules relatifs à la sécheresse et à l'inondation. Notre étude va donc se concentrer sur l'impact sur les fonds propres du changement climatique.

### Le SCR (Solvency Capital Requirement)

Nous avons pu présenter brièvement le SCR lors du chapitre 1.3, en expliquant notamment que ce dernier était divisé en différents modules selon le risque. Ici, nous le présenterons de façon plus spécifique de manière à nous concentrer sur ce qui va nous intéresser dans le cadre de notre étude. Le SCR est calculé selon la formule suivante

$$SCR = BSCR + Ad + SCR_{op}$$

avec :

- BSCR : l'agrégation des SCR de modules ainsi que du SCR de module des actifs intangibles. Nous avons décrit les modules ainsi que la formule de calcul dans la partie 1.3.1 ;
- Ad : ajustement calculé à partir du BSCR et du BSCR net, tenant compte de la capacité d'absorption des pertes des provisions techniques ;
- $SCR_{op}$  : le SCR opérationnel, c'est à dire le SCR associé aux risques de pertes directes ou indirectes dues à des erreurs humaines ou à des défaillances des procédures de l'entreprise.

Parmi ces trois éléments, celui qui nous intéressera s'avère être le BSCR. En effet, si le SCR opérationnel pourrait être impacté par le changement climatique, notre étude ne porte pas sur ce dernier et quantifier l'impact serait complexe. Quant à l'ajustement, ce dernier dépend du BSCR.

Comme expliqué précédemment, le BSCR est divisé en modules liés à différents risques :

- le risque de marché. Cet élément regroupe tous les risques auxquels l'assureur est exposé de par ses actifs, comme l'immobilier, les obligations ou encore les actions ;
- le risque de contrepartie. Ce risque traduit la possibilité qu'une des contreparties de l'assureur fasse défaillance et ne puisse donc pas respecter ses engagements ;
- le risque de souscription. Ce risque correspond aux risques que prend l'assureur en distribuant des contrats d'assurance.

Les risques sécheresse et inondation vont faire partie du risque de souscription, dans notre cas en non-vie. Le SCR de souscription non-vie est alors divisé en sous-modules, qui sont :

- le risque de primes et de provisionnement. Cet élément correspond au risque que les primes ou les provisions s'avèrent insuffisantes au regard de la sinistralité ;
- le risque catastrophe, qui regroupe les risques provenant d'évènements extrêmes et irréguliers ;
- le risque de rachat, qui couvre la possibilité que les assurés rachètent leur contrat dans un contexte défavorable. Si l'on pense d'abord à l'assurance vie pour ce risque, il existe cependant quelques types de contrats non-vie pluri-annuels.

Nous allons donc nous intéresser au SCR Catastrophe. Nous présentons ici les différents sous-modules (figure 4.1) de ce dernier.

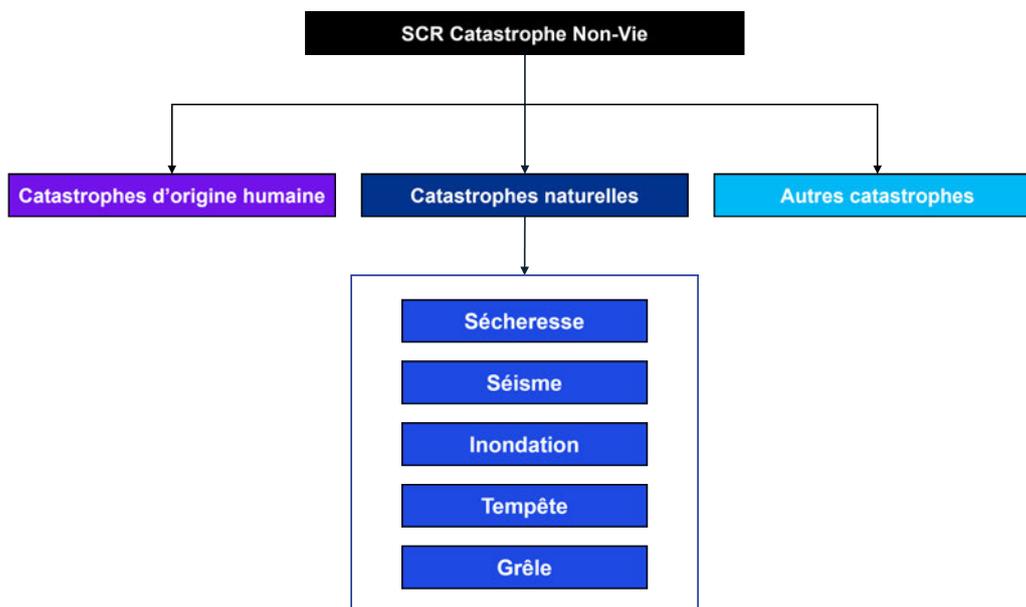


FIGURE 4.1 : Présentation des modules du SCR catastrophe non-vie

Notre étude portera donc sur l'évolution des sous-modules sécheresse et inondation causée par l'augmentation de sinistralité calculée précédemment.

Maintenant que nous avons pu expliciter ce qui allait nous intéresser dans le bilan, expliquons la situation de base de notre assureur en nous concentrant sur le SCR. Pour des raisons de confidentialité, les chiffres ont été modifiés.

### SCR assurantiel et méthodes de calcul de ce dernier

Pour l'année 2023, le SCR total de notre assureur s'élève à 132,5 M € et son BSCR à 150,4 M €. Représentons la façon dont ce dernier est réparti, ainsi que la façon dont le SCR souscription est réparti (figure 4.2) (pour des raisons de simplicité, nous ne représenterons pas l'effet de diversification) :

Nous voyons donc que parmi les trois modules du SCR total, le risque de marché et de souscription non-vie sont de loin les plus importants. Pour ce qui est du risque de souscription, le module catastrophe représente ici la majorité du risque. Comme montré dans la figure 4.1, ce module est lui-même divisé en plusieurs sous-modules : catastrophes d'origine humaine, catastrophes naturelle et autres catastrophes, qui est ici d'un montant nul. A noter que nous ne parlons pas ici de l'effet de diversification, qui va venir diminuer le SCR.

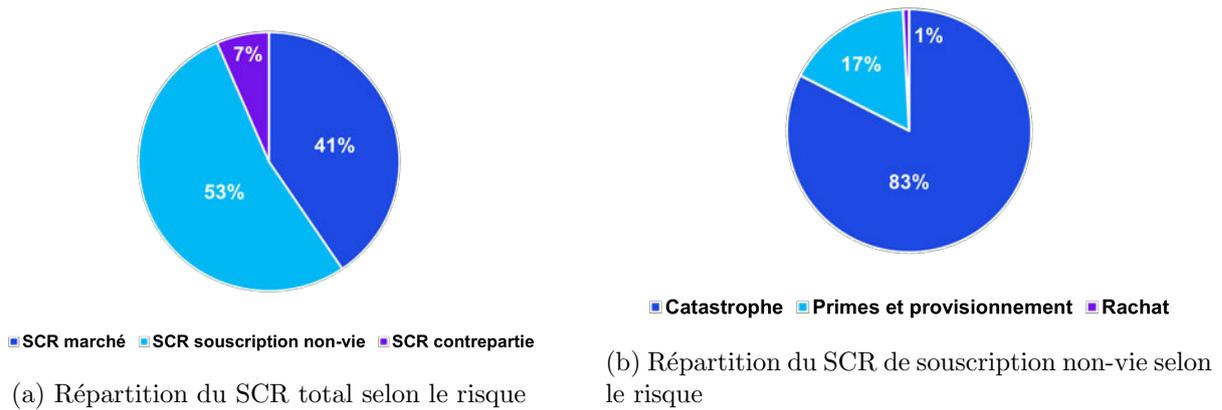


FIGURE 4.2 : Répartitions du SCR total puis du SCR souscription non-vie

Le SCR catastrophes naturelles non-vie est quant à lui réparti de la sorte (figure 4.3) :

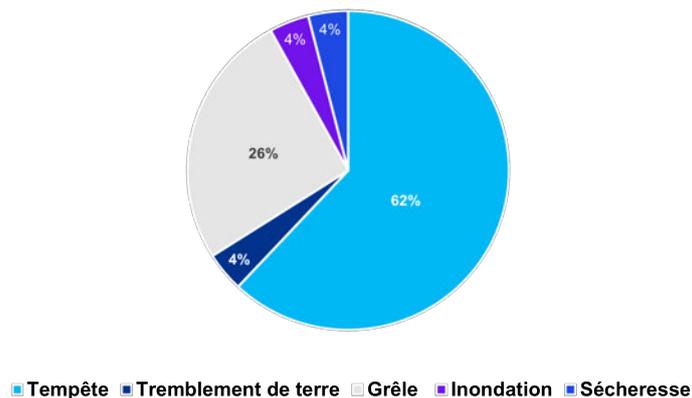


FIGURE 4.3 : Répartition du SCR catastrophe non-vie naturel

Nous observons que la tempête et la grêle représentent la majorité du SCR catastrophes naturelles, tandis que les risques inondation et sécheresse représentant un total de 8 % sur le SCR concerné. Si ces risques ne sont pas les plus importants, ils n'en demeurent pas moins conséquents dans le SCR cat et donc dans le SCR global. Sans tenir compte de l'effet de diversification, ces deux risques représentent 6 % du BSCR, ce qui fait qu'un important choc à la hausse de ces deux sous-modules aurait un impact non négligeable sur le SCR général.

Concernant le calcul du SCR cat, notre assureur se base sur les résultats de la cinquième étude quantitative d'impact (QIS 5), les études quantitatives d'impact étant menées par la Commission Européenne afin de tester les exigences quantitatives de solvabilité 2.

Dans cette étude, les assureurs calculent le risque CAT en utilisant le CRESTA (*Catastrophe Risk Evaluation and Standardising Target Accumulations*), qui est un zonage géographique se basant sur les départements français. Chaque département se voit attribuer un coefficient selon sa vulnérabilité à l'aléa climatique. La carte ci-dessous représente les différents coefficients CRESTA pour la France métropolitaine (figure 4.4) :

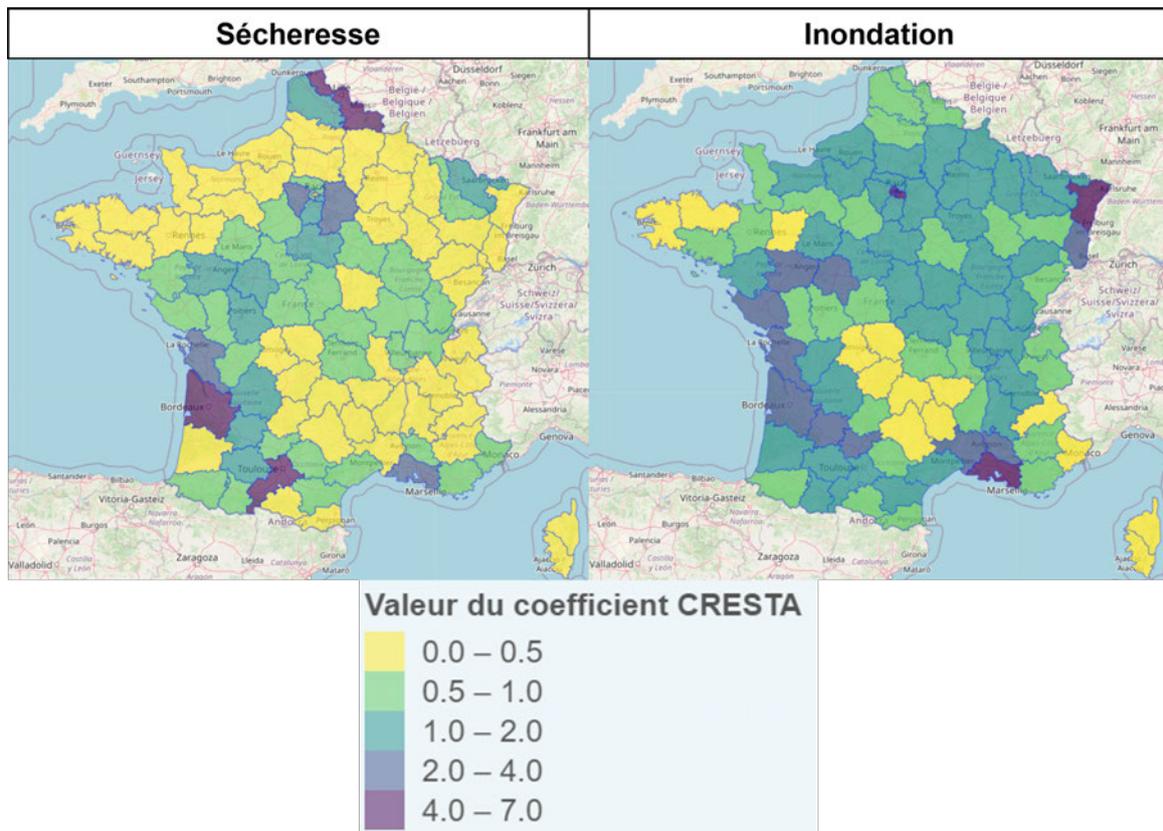


FIGURE 4.4 : Présentation de la carte des coefficients CRESTA pour la France métropolitaine

A partir de cet élément et des indications de QIS 5, l'assureur suit le processus de calcul suivant pour le SCR inondation :

- dans une première étape, l'assureur calcule son exposition dans chaque département français. Cela lui permet de calculer une charge de risque brute Cat à l'aide d'une matrice d'agrégation des départements ainsi que de l'indicateur de relativité de zone du CRESTA ;
- dans une seconde étape, deux scénarios comportant chacun deux évènements sont considérés :
  - pour le premier scénario, deux coefficients correspondant respectivement aux évènements 1 et 2 sont appliqués sur la charge de risque ;
  - de même dans le second scénario, avec cette fois-ci des coefficients différents. A partir de ces derniers, l'on obtient une charge de risque pour chaque évènement de chaque scénario, auquel on soustrait la réassurance Cat de notre assureur.
  - Enfin, le montant maximum des deux scénarios est retenu en tant que SCR inondation avant réassurance.

Le processus décrit précédemment est similaire pour le SCR sécheresse, cependant pour ce dernier un seul scénario à un seul évènement est considéré.

Dans les deux cas, la réassurance s'avère particulièrement importante, puisqu'elle va venir diminuer fortement le montant de sinistralité associé à la catastrophe naturelle. Notre assureur, comme la grande majorité des assureurs français, a un programme de réassurance auprès de la CCR. Comme expliqué dans la partie 1.1.5, ce programme prévoit une quote part de 50 %, ce qui veut dire que 50 % de

la sinistralité sera assumée par le réassureur. En plus de cela, le programme inclut un programme *stop loss*, ce qui empêche le SCR d'être trop élevé et fait que les SCR sécheresse et inondations s'avèrent proches car limités par le traité.

Maintenant que nous avons parlé du SCR de notre assureur, expliquons son rôle dans la solvabilité de ce dernier. Comme dit précédemment, le SCR fait partie des fonds propres, qui se divisent en deux parties : le SCR et les fonds propres disponibles. A partir du SCR, l'assureur calcule deux ratios de solvabilité : un vis à vis du MCR et l'autre vis à vis du SCR. Pour rappel, le MCR est l'exigence minimale de fonds propres, en-dessous de laquelle le régulateur intervient automatiquement. Son montant absolu est fixé à 4 M € dans la norme Solvabilité 2. Les deux ratios se calculent en divisant le montant des fonds propres disponibles par le montant du SCR puis par le montant du MCR. Les fonds propres disponibles sont classés en plusieurs Tier selon leur disponibilité :

- le Tier 1, aussi appelé fonds propres de base, est la partie la plus solide, constituée par exemple d'actions ou de participations minoritaires ;
- le Tier 2 désigne les fonds propres complémentaires, comme les plus values latentes ou les provisions ;
- enfin, le Tier 3 consiste en les actifs d'impôts différés et les autres instruments de capital comme les comptes subordonnés de mutuelle.

La norme stipule qu'il est nécessaire que les fonds propres Tier 1 représentent plus de 50 % du SCR, et que les fonds propres Tier 2 ne représentent pas plus de 15 % de ce dernier. Pour notre assureur, les fonds propres sont constitués en majorité de Tier 1. Dans le cadre assurantiel, le ratio de solvabilité entre les fonds propres disponibles et le SCR s'avère particulièrement important, puisqu'il s'agit d'un indicateur clé quant à la résistance de l'assureur aux chocs.

Pour notre assureur, ce ratio se situe entre 250 et 300 %, ce qui est proche de la moyenne française, qui se situe au premier semestre 2023 à 255 % selon le rapport de l'ACPR (2023a). Un des enjeux principaux lié au changement climatique est la baisse de ce ratio induite par la hausse de sinistralité.

En plus de cela, chaque assureur se doit de définir son appétence aux risques vis à vis de ses différents modules. Dans le cas de notre assureur, ce dernier est averse aux risques et souhaite limiter au mieux son SCR afin de préserver son ratio de solvabilité. Cela signifie que même une baisse peu importante de ce ratio pourra être préoccupante.

Le ratio de solvabilité est donc un élément d'importance pour l'assureur. Dans les cas de chocs importants, ce dernier peut se voir impacter fortement du fait de la hausse du SCR qui s'accompagne par la baisse des fonds propres disponibles. L'ORSA est justement le processus qui a pour rôle de tester la solidité de l'assureur, notamment en calculant l'évolution des ratios de solvabilité en cas de forts chocs.

#### 4.2.2 ORSA assurantiel

Maintenant que nous avons expliqué la situation de départ de notre assureur, notamment en matière de bilan, présentons l'ORSA de ce dernier. Expliquer le processus habituel d'ORSA nous permettra d'en comprendre les enjeux.

Comme expliqué dans la partie 1.3.2, l'ORSA évalue au moins trois éléments :

- le besoin global de solvabilité ;
- le respect des exigences de capital ;
- la différence entre les hypothèses de calcul du SCR et le profil de risque réel de l'assureur.

Nous nous concentrons ici sur les deux premiers éléments, les différences d'hypothèses étant difficiles à jauger pour nous. Afin de jauger le respect des exigences de capital, plusieurs scénarios sont envisagés : un scénario central, correspondant aux hypothèses du business plan de l'assureur, et plusieurs scénarios de chocs visant à vérifier la solidité de l'assureur par rapport aux risques. Ces scénarios vont s'appliquer aux différents éléments de l'actif ainsi qu'au passif, touchant l'immobilier, les actions... Ces scénarios sont projetés entre 3 et 5 ans, ce qui veut dire que pour l'ORSA 2022 les projections des différents éléments se font au moins jusqu'en 2025. Notre assureur étant averse au risque, son désir est que son ratio reste au-delà des 150 %, peu importe le scénario choisi.

Selon le document d'orientation ORSA fourni par l'INSTITUT DES ACTUAIRES (2014), nous pouvons distinguer au moins quatre méthodes de calibrage des scénarios de chocs :

- le calibrage à dire d'expert, c'est à dire la construction de scénarios plus ou moins probables mais cohérents. Cette méthode présente une part de subjectivité, mais a l'avantage de présenter un aspect réaliste et facile à comprendre ;
- le calibrage sur des données historiques. Cette méthode permet d'éviter le côté subjectif du calibrage à dire d'expert, mais a pour défaut d'être potentiellement trop prudente. Dans un cadre de changement climatique, elle apparaît comme étant peu appropriée pour cette raison ;
- le choix de niveaux de stress relatif ou absolu :
  - Un stress relatif est un choc qui ne dépend pas de la situation actuelle mais a été calibré sur des variations passées. Une fois de plus, le désavantage principal de la méthode est de trop se baser sur le passé et d'avoir une vision pro-cyclique du risque ;
  - Le stress absolu est un choc se basant sur des niveaux historiques absolus. Une fois de plus, cette méthode n'est pas vraiment intéressante pour nous du fait du changement climatique à venir.
- les stress de tendances supposent une variation durable des paramètres. Ce type de scénario traduit une prévision d'un changement constant à long terme. Dans un cadre climatique cela peut être intéressant mais présente le risque de ne pas être suffisamment adaptatif.

De toutes les méthodes que nous venons de présenter, le calibrage à dire d'expert semble être la plus appropriée dans un cadre de changement climatique rapide tel que nous le vivons aujourd'hui.

Décrivons plus en détail le scénario central retenu mais également les scénarios de stress climatiques introduits dans l'ORSA. Dans le cadre du scénario central, de nombreuses hypothèses sont retenues pour la projection. Ces hypothèses sont issues soit de la stratégie de l'assureur, par exemple pour l'évolution des primes, soit de la réglementation. Ces hypothèses couvrent de nombreux sujets, que ce soit la comptabilisation de la réassurance, de plus ou moins-values latentes ou encore de la courbe des taux.

En utilisant ces hypothèses, il est alors possible de projeter l'ensemble du compte de résultat technique de l'assureur : les primes, les provisions techniques, les cessions en réassurance... Mais également le compte de résultat non-technique comme le résultat exceptionnel ou le produits des placements non transférés. Ces résultats sont utiles car ils permettent à l'assureur de visualiser de façon crédible la manière dont son bilan va être amené à évoluer. Cependant, ces résultats se basent sur des hypothèses prudentes et raisonnables, qui ne viennent pas tester la solvabilité. C'est donc pour cela que l'on considère également pendant l'ORSA des scénarios de stress, qui se concentrent sur l'évolution des fonds propres et des ratios de solvabilité associés. Parlons maintenant d'un scénario de stress pouvant s'avérer intéressant pour nous.

Parmi les scénarios de stress considérés durant l'ORSA, nous retrouvons le scénario de stress sécheresse. Ce scénario se base sur la pire année de sécheresse relevée par la CCR, à savoir 2003 avec 2,024 milliards d'euros (l'année 2022 a depuis battu ce record avec un montant estimé 2,9 milliards d'euros, mais les sinistres sécheresse mettent souvent beaucoup de temps à se manifester, rendant ce montant sujet à évolution et empêchant l'intégration directe de l'année dans les scénarios de stress). Cette sinistralité est alors rapportée à la part de marché de l'assureur, ce qui se traduit par un choc très important sur plusieurs années consécutives.

Ce scénario se base sur les hypothèses de l'exercice pilote de l'ACPR lancé en 2023, dont nous avons parlé dans la partie 1.4.2. En effet, le scénario de court terme sécheresse suppose que les vagues de chaleur de l'été 2022 se répliquent en 2023 et 2024, créant donc trois années consécutives de forte sinistralité sécheresse.

Dans le cadre de notre assureur, ce scénario catastrophe a un très fort impact, puisque le ratio de solvabilité de l'assureur passe nettement en-dessous de la barre des 200 %, ce qui est une chute importante. Cela est d'autant plus notable que la sécheresse est un risque peu important pour notre assureur, ne représentant qu'une faible part de sa sinistralité en temps normal. Nous présentons ici l'évolution en base 100 pour notre assureur des fonds propres disponibles, du SCR et du ratio de solvabilité associé selon le scénario central et le scénario de stress sécheresse (figure 4.5).

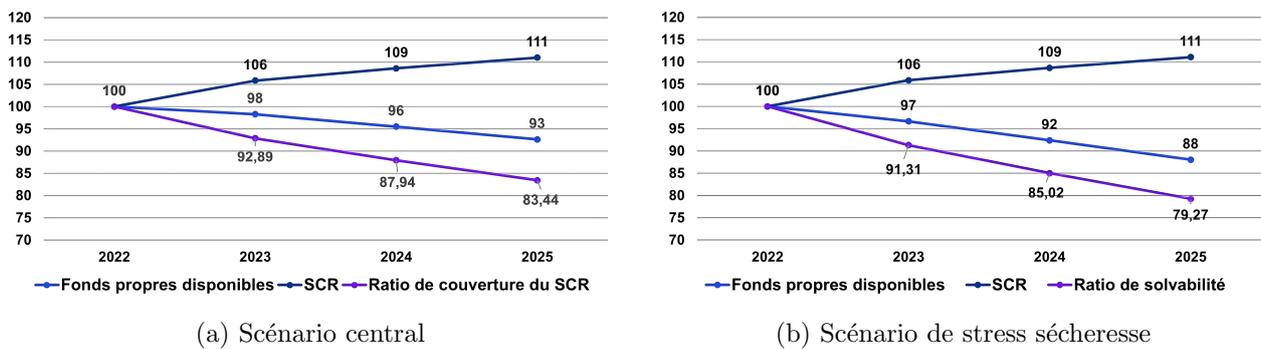


FIGURE 4.5 : Trajectoires des fonds propres disponibles, du SCR et du ratio de solvabilité

La figure 4.5 nous montre que concernant le scénario central, le SCR augmente de 11 % et les fonds propres disponibles baissent de 7 %, ce qui est expliqué par les pertes anticipées par l'assureur.

Dans le cadre du scénario sécheresse, le SCR global s'est vu impacter de près de 11 % de hausse tandis que les fonds propres disponibles pour couvrir le SCR diminuaient de 13 %, ce qui fait chuter de façon importante le ratio de couverture du SCR. Dans le cas de notre assureur, la chute des fonds propres disponibles s'explique par des résultats techniques négatifs, aggravés par la forte sinistralité sécheresse du scénario.

Maintenant que nous avons pu présenter le contexte réglementaire, la situation de base de notre assureur et son ORSA classique, nous pouvons procéder à nos calculs.

### 4.2.3 Réalisation de l'ORSA climatique

Nous avons présenté la situation de notre assureur et la façon dont ce dernier réalise l'ORSA classique, notre objectif est d'effectuer un ORSA climatique. Pour cela, nous nous concentrons sur l'évolution des SCR sécheresse et inondations, en évaluant pour chaque horizon de temps et scénario la façon dont ces derniers évoluent. Cela nous permet alors de créer plusieurs scénarios d'évolution du SCR et du ratio de solvabilité, avec différents niveaux de chocs.

### Calcul des SCR sécheresse et inondation selon le scénario

Pour rappel, nous avons calculé l'évolution de la sinistralité selon trois facteurs :

- la probabilité d'arrêt CatNat ;
- le nombre de dossiers de sinistralité engendré par un arrêté ;
- le coût moyen d'un dossier.

Comme nous sommes dans une hypothèse de bilan statique, la probabilité d'arrêt CatNat est l'élément sur lequel nous nous concentrons afin d'impacter la sinistralité, les deux autres éléments restants fixes dans ce cadre. Pour rappel, notre base *Open Source* contient un pourcentage de 5,05 % de communes en arrêtés inondations et 8,48 % de communes en arrêtés sécheresse. Notre objectif est de créer une distribution de ladite probabilité dans une vision modèle interne.

Nous sommes ici confrontés à un problème : pour obtenir une distribution de la probabilité, il nous faudrait être en mesure de re-simuler nos variables *Open Source*. Or cela impliquerait de faire intervenir un modèle Cat, comme dans le mémoire HASSANI (2017). Nous n'avons cependant pas accès à l'un de ces modèles, aussi va-t-il nous falloir établir un proxy.

Comme nous l'avons expliqué dans la section 3.6.2, nos projections de probabilité se basent sur la fonction *prob* du logiciel R CORE TEAM (2023). Pour chaque commune, notre modèle renvoie une probabilité, et nous maximisons la qualité du modèle en fonction d'un seuil, compris entre 0 et 1, qui détermine si la commune se voit affecter la valeur 0 ou 1. Le proxy pouvant nous intéresser va justement impliquer ce seuil. Nous faisons prendre des valeurs au seuil entre 0 et 1 par pas de 0.001, nous donnant un millier de seuils différents. Faire varier ce seuil nous donne une distribution de la probabilité. Comme nous nous intéressons au SCR, nous prenons le quantile à 99,5 %, c'est à dire la probabilité associée au seuil 0,005, chaque commune ayant une probabilité au-dessus dudit seuil se voyant associer un arrêté.

Notre proxy induit forcément un biais puisqu'il ne se base pas sur la réglementation concernant le calcul du SCR, mais se place plutôt dans une vision de modèle interne. La qualité du proxy est à juger sur la qualité des résultats qu'il peut nous fournir, nous devons donc nous interroger sur leur plausibilité une fois obtenus.

Le proxy nous donne les résultats suivants (figure 4.6).

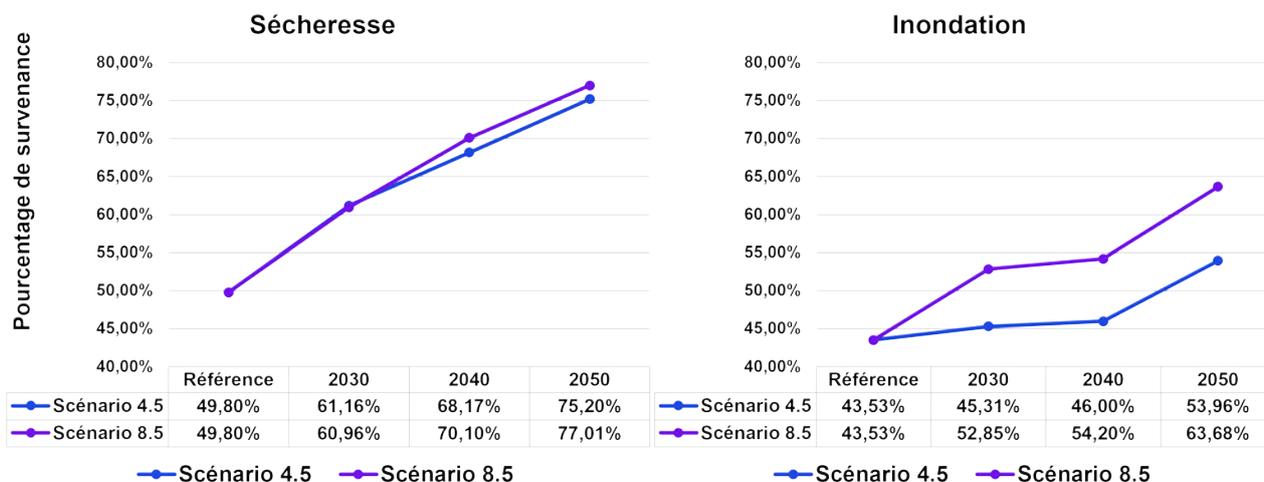


FIGURE 4.6 : Evolution du quantile à 99,5 % de la probabilité de survie d'un arrêté CatNat

Concernant les résultats, nous observons sur la figure 4.6 une forte augmentation de la probabilité selon l'horizon et le scénario, cohérente avec ce à quoi nous nous attendions. Nous remarquons que les scénarios sécheresse sont très proches, élément que nous avons déjà observé lors de la section 3.6.2. Nous avons des probabilités maximales à :

- 77 % pour la sécheresse. Ce chiffre très élevé se justifie par des prédictions pessimistes concernant la température de la France. Ainsi, la sécheresse de 2022 a entraîné la reconnaissance en arrêté sécheresse de 15,6 % des communes françaises, et 35 % des sols furent en situation de sécheresse. Cependant, cette sécheresse n'est pas vue comme une anomalie mais une année annociatrice, et des vagues de chaleur extrêmes pourront être attendues en France en 2050. Par exemple, le record de chaleur de France métropolitaine est actuellement de 46 °C, mais un rapport de l'Agence européenne pour l'environnement chiffre un possible record à 55°C en 2050. Il est donc possible d'envisager un scénario extrême en 2050 provoquant une sécheresse sur tout le territoire. Cependant, la probabilité de reconnaissance reste limitée par l'exposition aux argiles, puisque c'est de par leur existence que la sinistralité survient. Ainsi, 36 % des communes françaises ne présentent aucune exposition forte ou moyenne aux argiles, or comme expliqué dans la section 1.1.2, 93 % des sinistres proviennent de zones d'expositions forte ou moyenne. Le chiffre trouvé semble donc très haut, mais se justifie par la méthode de calcul, qui utilise un quantile à 99,5 % et par l'impact du changement climatique.
- 64 % pour l'inondation. Ce chiffre est très élevé, pour des raisons similaires à celles évoquées pour la sécheresse : à horizon 2050, le niveau de la mer sera monté, menaçant les villes côtières, et nous avons pu voir dans la section 1.2.2 que la période de retour d'événements de précipitations extrêmes se réduira fortement en 2050. De plus, l'inondation est un péril touchant la majorité des communes en France : d'après le rapport CCR (2021), 27 907 communes furent reconnues en arrêté inondation pour l'année 1999, soit 76 % des communes de France métropolitaine. Cependant le régime CatNat était assez différent d'aujourd'hui, notamment au niveau des critères de reconnaissance, de plus nous nous basons sur la période 2018-2022, où nous nous situons autour des 5 % de reconnaissance. Si nous nous basons sur l'historique entier du régime, nous arrivons vers une moyenne de 8,5 %. De façon générale, il est logique que les arrêtés inondation arrivent à un niveau moins élevé que les arrêtés sécheresse car ils sont moins présents dans notre base.

Nos résultats s'avèrent donc cohérents, il nous faut maintenant les utiliser pour calculer une évolution du SCR.

En nous basant sur ces évolutions et en reprenant la méthode de calcul employée dans la section 4.1.1, nous pouvons alors proposer une évolution des SCR sécheresse et inondation bruts de réassurance en pourcentage par rapport à la base (figure 4.7).

Sécheresse	RCP 4.5	RCP 8.5	Inondation	RCP 4.5	RCP 8.5
Horizon 2030	146	146	Horizon 2030	124	145
Horizon 2040	196	202	Horizon 2040	151	178
Horizon 2050	265	271	Horizon 2050	218	257

FIGURE 4.7 : Niveau du SCR par rapport à l'origine, selon le scénario, le péril et l'horizon de temps

Nous constatons une forte augmentation du SCR, qui va se situer dans les mêmes ordres de grandeur que l'évolution de la sinistralité présentée dans la figure 4.2. La différence principale est que

l'augmentation est plus forte pour l'horizon 2030, ce qui peut s'expliquer par le fait que nous nous concentrons ici sur les événements extrêmes.

Cependant il s'agit ici du SCR brut de réassurance, ce qui est problématique puisque cette dernière a un fort impact sur le montant du SCR. Or nous sommes confrontés ici à un nouveau problème : le traité de réassurance inclut un traité *stop loss*, qui va venir empêcher la sinistralité de dépasser un certain seuil. Ce seuil étant déjà atteint en 2023, le garder tel quel ferait que les SCR resteraient parfaitement identiques pour chaque horizon et scénario, ce qui aurait peu d'intérêt. Non seulement cela nous empêcherait d'évaluer l'impact du changement climatique, mais en plus cela ne serait pas cohérent : le traité dépendant des primes CatNat acquises, celles-ci ne resteraient pas fixes en cas de hausse de sinistralité.

Nous allons donc considérer deux éléments :

- d'abord, une évolution des primes CatNat que nous indexons sur l'inflation montrée figure 4.1 ;
- ensuite, un choc sur les limites des contrats de réassurance. Actuellement le traité comporte un *Stop-Loss* à 200 %. Nous supposons que ce dernier passe à 300 %. Pour ce chiffre, nous nous sommes basés sur un rapport du SÉNAT FRANÇAIS (2009) suite à la sécheresse de 2003, qui proposait un relèvement du *Stop-Loss* du fait de l'augmentation de la sinistralité.

### Impact sur le bilan assurantiel global

Nous voulons observer l'impact de la hausse du SCR sur notre assureur. Comme expliqué avant, un des enjeux principaux de l'ORSA est de calculer l'évolution du ratio de solvabilité selon un scénario central et plusieurs scénarios de chocs. Dans notre cas, nos scénarios correspondent aux augmentations des SCR calculées précédemment. Comme nous l'avons montré, les sous-modules sécheresse et inondations sont assez peu importants dans le SCR, mais nos calculs montrent une très forte augmentation de ces derniers. Etant en bilan statique, nous gardons les montants de 2023 concernant les différents éléments de bilan excepté pour les sous-modules sécheresse et inondation. Nous allons considérer trois scénarios : le premier ne faisant intervenir que l'augmentation du SCR sécheresse, le deuxième que l'augmentation du SCR inondation, et le dernier combinant les deux.

Nous représentons pour chaque figure l'évolution du SCR et du ratio de solvabilité de l'assureur en base 100.

- en considérant seulement l'augmentation du SCR sécheresse pour les deux scénarios (figure 4.8) ;

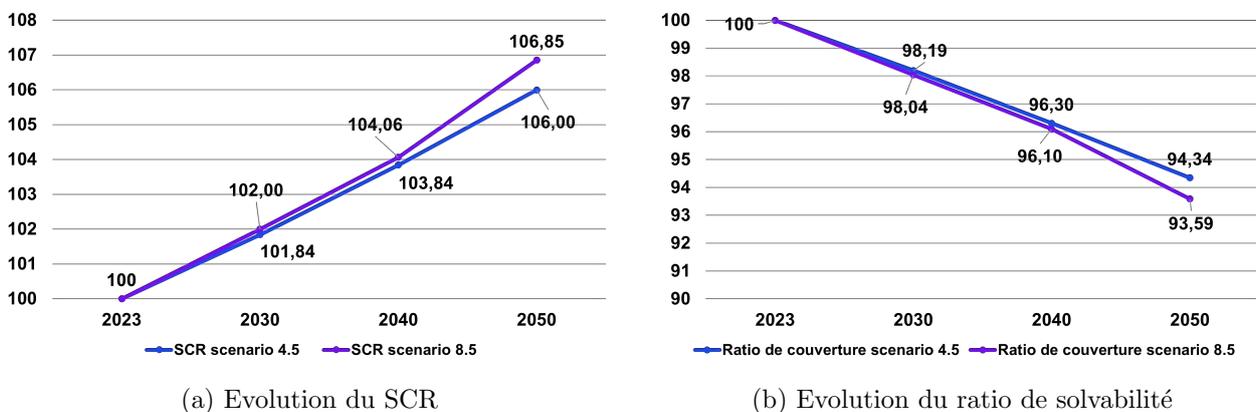


FIGURE 4.8 : Représentation de l'évolution du SCR, ainsi que du ratio de solvabilité, due à la sécheresse

Nous observons sur la figure 4.8 une baisse de plus de 6 % du ratio de solvabilité. Cette baisse confirme le fait que malgré l'importance peu élevée du SCR sécheresse dans le bilan, son évolution peut tout de même s'avérer problématique pour l'assureur. Nous notons également la proximité des deux scénarios, ce qui est cohérent avec ce que nous avons trouvé précédemment ;

- en considérant seulement l'augmentation du sous-module SCR inondation pour les deux scénarios (figure 4.9) ;

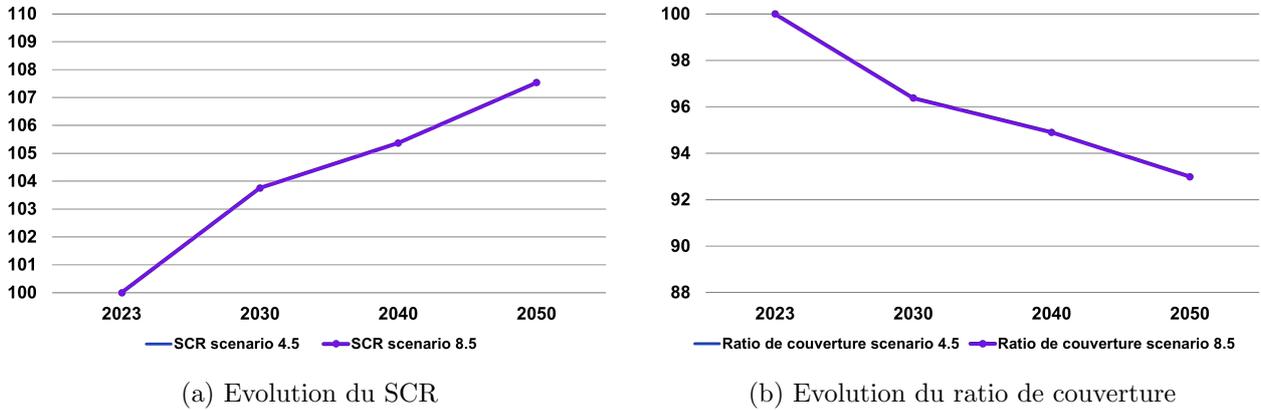


FIGURE 4.9 : Représentation de l'évolution du SCR, ainsi que du ratio de solvabilité, due à l'inondation

Nous observons sur la figure 4.9 qu'il n'y a ici aucune différence entre les deux scénarios (les deux courbes sont confondues), alors que ces derniers sont normalement différenciés. Cela est dû au fait que l'intégralité de la sur-sinistralité liée au 8.5 est ici absorbée par la réassurance.

- en considérant l'augmentation combinée du SCR inondation et du SCR sécheresse pour les deux scénarios. Ce scénario peut être considéré dans un cadre ORSA comme un scénario de choc, en particulier dans le cadre du scénario 8.5, puisqu'il va venir impacter le plus fortement le SCR (figure 4.10).

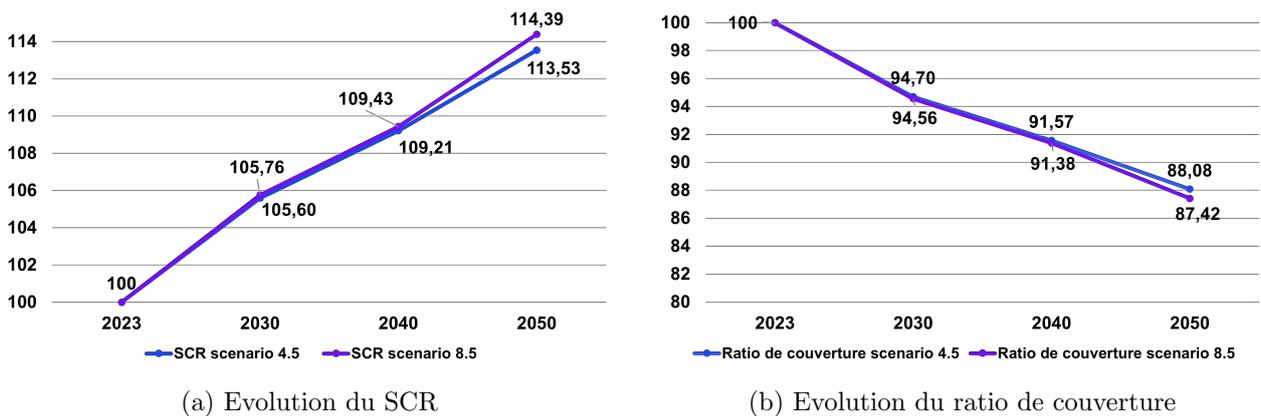


FIGURE 4.10 : Représentation de l'évolution du SCR ainsi que du ratio de solvabilité dues aux chocs combinés

En combinant les chocs, nous observons alors sur la figure 4.10 une baisse de 12,5 % du ratio de solvabilité. Il s'agit donc d'une baisse conséquente compte tenu de la taille des sous-modules sécheresse et inondation dans le SCR global.

Nous voyons que le changement climatique, plus précisément les périls sécheresse et inondation, ont un clair impact sur le bilan de notre assureur, et que leur augmentation fera augmenter de façon très conséquente le SCR, diminuant par la même occasion le ratio de solvabilité, élément de grande importance pour les assureurs. Nous pouvons cependant nous interroger sur la cohérence des résultats concernant cet impact bilan, que ce soit au niveau du SCR ou du ratio de solvabilité. Premièrement, l'impact global de la sécheresse s'avère bien moins important dans notre étude que dans l'ORSA classique. Cela est dû au fait que notre assureur est peu exposé à la sécheresse et a donc peu de sinistralité du fait de ce péril. De plus l'ORSA considérait en plus de la hausse du SCR une forte diminution des fonds propres. Un autre élément important est la similarité des deux scénarios, ce qui s'explique par le fait qu'and dans le cadre de l'inondation, l'augmentation du SCR due au scénario est absorbée entièrement par la réassurance.

Concernant la crédibilité de la hausse du SCR, nous pouvons nous intéresser à l'étude effectuée par la CCR (2023), qui présente des périodes de retour pour des événements extrêmes, élément important quand on parle de SCR. Ainsi, la période de retour de la sécheresse de 2022 passerait de 25 ans à 12 ans selon le scénario 4.5 et 6 ans selon le scénario 8.5. Il y a aurait donc une multiplication par 2 dans le scénario 4.5 et par 4 dans le scénario 8.5, là où nous avons trouvé des facteurs multiplicatifs de 2,6 et 2,7 pour les scénarios. Si nous ne pouvons pas comparer ces résultats de façon exacte, nous constatons tout de même que les ordres de grandeur s'avèrent proches. Pour l'inondation le livre blanc de COVÉA (2022) prédit que les événements cinquantennals surviendront tout les 11 ans à horizon 2050 pour le scénario 8.5, soit une hausse par un facteur de 4,5. Cette fois notre augmentation de SCR inondation est assez inférieure puisqu'il s'agit d'un facteur de 2,57 dans notre cas. Cependant les différences de notre étude avec la leur semblent suffisamment importantes pour expliquer cela.

### **Analyse qualitative de l'ORSA : pilotage et adaptations**

Nous avons donc pu proposer une quantification de l'impact du changement climatique sur le bilan assurantiel. Cet impact devra logiquement être suivi d'actions de la part de l'assureur afin de s'adapter et de limiter ses risques. Nous proposons donc plusieurs types d'actions possibles. Nous les distribuons en trois types : les actions au niveau de la souscription, au niveau de la prévention auprès des assurés, au niveau du transfert de risque et enfin au niveau de l'exclusion du risque.

- Concernant la souscription :
  - L'assureur pourra répercuter la hausse de la sinistralité sur l'assuré. Cela peut passer par l'augmentation des primes, ce qui permet de faire supporter la hausse de sinistralité par l'assuré. Une autre façon de faire serait la mise en place de franchises et de limites plus contraignantes. Cependant cette politique présente le risque d'augmenter les résiliations, et savoir comment modéliser un tel élément peut s'avérer complexe.
  - L'assureur pourrait également se tourner vers l'assurance paramétrique. Pour rappel, l'assurance paramétrique permet le remboursement forfaitaire automatique de l'assuré dès la survenance d'un certain événement. Ce type d'assurance permet de limiter considérablement le risque lié au changement climatique, puisque le montant de remboursement est fixé à l'avance.
- Concernant la prévention :
  - L'assureur pourra prodiguer des conseils aux assurés, par exemple sur les potentiels travaux à entreprendre ou les zones à risques selon le péril.
  - Il pourra également mettre en place des mesures d'accompagnement pour les assurés : des compensations financières auprès de l'assuré pour des travaux réduisant le risque, des cellules d'alerte afin de prévenir les assurés d'événements dangereux ou encore le financement de nouvelles technologies réduisant les coûts.

- Concernant le transfert de risque : L'assureur pourra optimiser sa couverture afin de réduire au mieux le risque. Cependant cette option pose le problème évident de l'acceptation dudit risque par les réassureurs, qui pourront refuser de couvrir plus de risques alors que ces derniers sont déjà très importants.
- Concernant l'exclusion du risque, l'assureur pourra revoir son offre en excluant certains risques ou garanties trop coûteuses, ou en définissant certaines zones comme étant inassurables. Il y a néanmoins un risque concernant l'image et la politique de l'assureur, qui peuvent rendre cette stratégie difficile à appliquer.

Pour conclure, nous voyons que l'assureur dispose de nombreux moyens d'adaptations face aux changements climatiques et à la hausse de la sinistralité. Mais chaque mesure présente des défauts, que ce soit concernant l'image de l'entreprise ou les difficultés de modélisation.

## 4.3 Limites et perspectives de l'étude

### 4.3.1 Limites

Nous diviserons notre étude des limites de ce mémoire en deux parties : les limites liées aux données et celles liées aux modèles développés.

#### Limites des données

Concernant nos données, nous avons deux types de sources : l'assureur étudié et les données *Open Source*, qu'elles soient géologiques, atmosphériques ou hydrologiques.

Concernant les données de l'assureur, ces dernières se sont avérées comme ayant peu de profondeur. En effet, nous n'avons qu'un faible nombre de sinistres sécheresse ou inondation, et nous n'avons que peu de sinistres avant 2016 dans nos bases. Cela rend les résultats de notre étude sensibles à une évolution importante de la sinistralité pour notre assureur.

Concernant les données *Open Source*, bien que ces dernières viennent de bases créées par des organismes reconnus, elles peuvent néanmoins produire des résultats étranges ou peu satisfaisants. De plus nous n'avons pas extrait certaines données pouvant être utiles, telles que l'imperméabilité des sols ou leur élévation, éléments utiles lorsque l'on parle d'inondations car l'imperméabilité d'un sol fait que l'eau ruisselle sur ce dernier. Pour la sécheresse, nous aurions pu intégrer des données liées aux nappes phréatiques, qui jouent un rôle d'importance lors des épisodes de sécheresse.

Enfin nous pouvons nous interroger sur la période historique considérée lors de cette étude. En effet, nos raisons de considérer 5 ans de profondeur historique se justifiaient par une volonté conservatrice et par le peu d'historique de sinistralité à notre disposition, nous poussant à faire correspondre les historiques entre eux. Cependant de nombreuses études climatiques prennent un historique bien plus important, comme par exemple 30 ans, afin de lisser au maximum les variations du climat. En effet ce dernier n'est pas linéaire. Un historique comme le nôtre ne lisse pas suffisamment lesdites variations, et représente plus un moment précis du climat.

#### Limites des modèles

Concernant les modèles, nos résultats ne sont pas pleinement satisfaisants. En effet, nous n'avons pas pu excéder les 57 % de F1-score pour le risque sécheresse et 49 % pour l'inondation. Ces scores peuvent notamment s'expliquer par les limites des données que nous venons d'exposer. Nous aurions pu aussi considérer, en plus des GLM lasso, forêt aléatoire et XgBoost, des réseaux de neurones afin d'avoir le meilleur résultat possible. Ce type de méthode est en revanche très long à simuler pour un

ordinateur. Nous aurions également pu nous pencher sur une validation croisée répétée, méthode très coûteuse en temps mais pouvant améliorer le résultat.

Cette limite est néanmoins à relativiser par le fait qu'il y a un facteur politique dans la décision d'arrêter, élément impossible à prendre en compte dans nos modèles, mais également par le fait que nous ayons pu obtenir des résultats similaires à d'autres études réalisées sur le sujet.

Une autre limite identifiable peut se trouver dans le modèle de charge. Premièrement, nous avons utilisé la courbe d'inflation de l'EIOPA, ce qui peut être mal adapté à la situation de notre assureur. Concernant le bilan statique, si ce dernier se justifie par volonté de cohérence avec l'exercice pilote, il est cependant contestable car le nombre de sinistres comme de villes touchées pourra être impacté par le changement climatique.

Concernant nos calculs de SCR, la limite la plus évidente reste la nécessité que nous avons eu de développer un proxy. Ce dernier ne correspond pas à ce qui aurait dû être fait dans le meilleur des cas, à savoir re-simuler de nombreuses fois nos données afin d'obtenir une distribution. Cette limite reste une nouvelle fois à relativiser car les résultats obtenus furent cohérents avec ce à quoi nous pouvions nous attendre.

### 4.3.2 Perspectives

Concernant les perspectives, notre étude a été réalisée en même temps que le nouvel exercice pilote lancé par l'ACPR. Cet exercice, dont nous avons parlé au cours de notre mémoire, est de grande ampleur et s'avère très important pour l'adaptation des assureurs aux risques climatiques. Par rapport à notre étude, plusieurs éléments de cet exercice seront intéressants à considérer :

- Premièrement, nous pourrions comparer les différents résultats obtenus lors dudit exercice aux nôtres, afin de voir les différences et les similarités qui se dégagent. Il sera intéressant d'étudier en quoi les différences d'hypothèses et de méthodes entre notre étude et l'exercice pilote peuvent causer des différences de résultats. Un intérêt particulier pourra être apporté aux résultats concernant la considération des risques sécheresse et inondations dans l'exercice pilote. Pour rappel, l'exercice pilote envisage les scénarios suivants :
  1. Pour la sécheresse : la sécheresse de l'été 2022 se reproduit en 2023 et en 2024.
  2. Pour l'inondation : une rupture du barrage de Serre-Ponçon se produit en 2025, inondant la vallée en contre-bas.

Si notre assureur n'a aucune exposition aux alentours du barrage de Serre-Ponçon, il présente en revanche une claire exposition à des fleuves importants, qui possèdent également le long de leur courant des barrages et des retenues d'eau susceptibles de déborder. Les résultats obtenus lors de l'exercice pilote pourraient être alors étudiés et mis en parallèle avec une potentielle rupture desdits barrages.

Concernant la sécheresse, les résultats pourraient également être intéressants à analyser, notamment en mettant en parallèle les résultats avec la sinistralité de notre assureur pour les années concernées.

De plus, nous nous intéresserons aux conclusions de l'exercice concernant l'impact sur le bilan assurantiel du changement climatique, particulièrement sur nos deux périls d'intérêts. Nous observons comment le SCR pourra être impacté et notamment si d'autres postes du bilan pourront être touchés de façon importante. En effet, nous n'avons pas pu pour des raisons de faisabilité faire évoluer tous les postes du bilan, cependant certains d'entre eux comme les fonds propres pourront être sujets à des fortes évolutions du fait du changement climatique.

- Deuxièmement, l'exercice pilote aura pour conséquence prévisible de permettre l'instauration de processus d'évaluation climatique plus poussés chez les assureurs. Il s'agit d'ailleurs d'un

des objectifs principaux de l'ACPR pour l'exercice pilote. Dans notre cas, notre étude et sa méthodologie pourront servir de base à une démarche d'étude des risques climatiques pour notre assureur.

De façon plus générale, notre étude se place dans un contexte de forte augmentation des risques climatiques et d'une importante prise de conscience de ces derniers par les acteurs financiers.

# Conclusion

La question des risques climatiques est aujourd'hui un enjeu majeur pour les assureurs. De part leur évolution rapide, ceux-ci pourront s'avérer très dangereux pour les assurances dans un futur proche. Face à cela, les différents acteurs du monde de l'assureur se mobilisent afin d'anticiper au mieux les impacts des risques climatiques à venir. Ainsi, l'exercice pilote de l'ACPR a été lancé courant été 2023.

Dans ce contexte, notre mémoire a eu pour objectif d'effectuer une analyse climatique sur les données d'un assureur afin de saisir l'évolution de sa sinistralité climatique dans un contexte ORSA. Pour cette étude, nous nous sommes concentrés sur les risques inondation et sécheresse, considérés comme étant les plus importants au vu de l'exposition de l'assureur étudié.

Nous avons procédé en plusieurs étapes. Tout d'abord nous avons constitué une base de données climatiques afin de pouvoir modéliser la survenance d'arrêts, qu'ils soient sécheresse ou inondation. Cette base, constituée à partir de données *Open Source*, rassemble des données atmosphériques telles que l'humidité, géologiques telles que la composition du sol, et enfin hydrographiques telles que la présence de cours d'eau. Ces données furent collectées sur l'intégralité de la France métropolitaine.

Nous avons alors testé différents modèles afin de comparer leur qualité de prédiction, ce qui nous a permis de sélectionner le meilleur modèle possible. Ces différents modèles furent les modèles GLM lasso, forêt aléatoire et enfin XgBoost, avec utilisation de la méthode de la validation croisée pour optimiser nos hyperparamètres.

Une fois cela fait, nous avons alors constitué une nouvelle base climatique en utilisant les données du DRIAS, cette fois selon les scénarios 4.5 et 8.5 du GIEC. Pour cette base, nous avons défini les horizons d'intérêts comme étant 2030, 2040 et 2050. Les variables projetées concernaient l'humidité, la température, la vitesse du vent et enfin les précipitations.

En appliquant les modèles précédemment créés, nous avons pu ainsi modéliser l'évolution de la probabilité de survenance d'arrêt CatNat, cette dernière passant de 8,5% pour la période de référence à 19,3 % pour le scénario RCP 8.5 sécheresse, ce qui est extrêmement considérable.

De plus, nous avons pu projeter le coût moyen d'un sinistre Cat en utilisant des données d'inflation fournies par l'ACPR. A partir de cela, nous avons créé le modèle suivant :

*Probabilité d'arrêt CatNat*  $\times$  *Nombre de dossier par arrêt*  $\times$  *Sinistralité moyenne par dossier*

Ce modèle nous a permis d'évaluer l'évolution de la sinistralité de l'assureur.

Les résultats obtenus nous montrent une claire augmentation de la sinistralité, cohérente avec les horizons et les scénarios, qui pourront servir à l'assureur à anticiper et à se couvrir face à des périls climatiques de plus en plus importants. Ainsi, selon le scénario 8.5, la sinistralité sécheresse sera multipliée par un facteur 2,7 et la sinistralité inondation par un facteur 2,5 . A partir de cette hausse, nous avons simulé une distribution de sinistralité, ce qui nous a permis d'impacter à la hausse le SCR. Cette augmentation s'est traduite dans le bilan par une claire baisse du ratio de solvabilité de l'assureur, allant jusqu'à 12 %, ce qui est considérable. Sur ces bases, il pourra être envisagé une étude incluant des variations sur l'intégralité du bilan, comme par exemple au niveau des fonds propres.

Ces résultats nous permettent donc de confirmer l'impact du changement climatique sur la sinistralité comme sur le bilan assurantiel. Dans un contexte de changement rapide du climat mais également des exigences réglementaires vis à vis du sujet, bien illustrées par les exercices pilotes, les assureurs doivent se préparer et anticiper le changement climatique. Dans ce contexte, notre étude pourra être comparée aux résultats futurs de l'exercice pilote et alimenter la réflexion autour de ces sujets d'importance.

# Bibliographie

- ACPR (2011). Analyse et synthèses - Solvabilité 2 : principaux enseignements de la cinquième étude quantitative d'impact. Rapp. tech. URL : [https://acpr.banque-france.fr/sites/default/files/201103-acpr-solvabilite-2-enseignements-de-qis5\\_0.pdf](https://acpr.banque-france.fr/sites/default/files/201103-acpr-solvabilite-2-enseignements-de-qis5_0.pdf).
- ACPR (2019). Vers une meilleure prise en compte des risques environnementaux dans Solvabilité 2. Rapp. tech. URL : [https://acpr.banque-france.fr/sites/default/files/medias/documents/20190910\\_revue\\_acpr\\_septembre\\_finance\\_durable\\_revgtalbc.pdf#:~:text=L%20autorit%C3%A9%20europ%C3%A9enne%20des%20assurances,\(DDA\)%20et%20Solvabilit%C3%A9%20](https://acpr.banque-france.fr/sites/default/files/medias/documents/20190910_revue_acpr_septembre_finance_durable_revgtalbc.pdf#:~:text=L%20autorit%C3%A9%20europ%C3%A9enne%20des%20assurances,(DDA)%20et%20Solvabilit%C3%A9%20).
- ACPR (2022). La gouvernance des risques liés au changement climatique dans le secteur de l'assurance. Rapp. tech. URL : <https://acpr.banque-france.fr/la-gouvernance-des-risques-lies-au-changement-climatique-dans-le-secteur-de-lassurance>.
- ACPR (2023a). La situation des assureurs soumis à Solvabilité II en France au premier semestre 2023. Rapp. tech. URL : [https://acpr.banque-france.fr/sites/default/files/medias/documents/20231121\\_as\\_assureurs\\_s2\\_2023\\_q2.pdf](https://acpr.banque-france.fr/sites/default/files/medias/documents/20231121_as_assureurs_s2_2023_q2.pdf).
- ACPR (2023b). Scénarios et hypothèses principales de l'exercice de stress test climatique 2023. Rapp. tech. URL : <https://acpr.banque-france.fr/scenarios-et-hypotheses-principales-de-lexercice-de-stress-test-climatique-2023>.
- AKAFFOU, D. H. (2020). Méthode alternative de tarification santé: GLM/XGBoost. Mémoire d'actuariat. EURIA.
- BANQUE DE FRANCE (2021). Résultats de l'exercice pilote. Rapp. tech. Banque de France.
- BOYEAU, C. (2022). Évaluation de l'impact du changement climatique sur le risque inondation en France métropolitaine. Mémoire d'actuariat. ISUP.
- BRGM (2019). Carte du BRGM. Info terre BRGM [carte en ligne]. URL : <https://infoterre.brgm.fr/actualites/exposition-au-retrait-gonflement-argiles>.
- CARBONE4 (juill. 2022). Les scénarios SSP. Rapp. tech. URL : <https://www.carbone4.com/publication-scenarios-ssp-adaptation>.
- CCR (2011). Le régime d'indemnisation des catastrophes naturelles. Rapp. tech. CCR. URL : <https://www.ccr.fr/-/indemnisation-des-catastrophes-naturelles-en-france>.
- CCR (2018). Conséquences du changement climatique sur le coût des catastrophes naturelles en France à l'horizon 2050. Rapp. tech. CCR. URL : <https://www.ccr.fr/documents/35794/35836/Etude+Climatique+2018+version+complete.pdf/6a7b6120-7050-ff2e-4aa9-89e80c1e30f2?t=1536662736000>.
- CCR (2021). Rapport CatNat 1982/2021. Rapp. tech.
- CCR (2022a). L'indemnisation des Catastrophes Naturelles en France. Rapp. tech.
- CCR (2022b). Rapport d'activité. Rapp. tech. CCR. URL : <https://ccr.foleon.com/rapport-d-activite-2022/rapport-d-activite-2022/>.
- CCR (2023). Conséquences du changement climatique sur le coût des catastrophes naturelles en France à l'horizon 2050. Rapp. tech. CCR. URL : <https://www.ccr.fr/documents/35794/1255983/CCR+Etude+climat+BAG+23102023+page+22mo.pdf/68b95f6e-8238-4dcc-6c56-025fa410257b?t=1698161402128>.

- CCRE (2021). Qu'est ce que la réassurance en excédent de perte annuelle ? [en ligne]. URL : <https://blog.ccr-re.com/fr/qu-est-ce-que-la-reassurance-en-excedent-de-perte-annuelle>.
- CHEN, T. et GUESTRIN, C. (2016). XGBoost: A Scalable Tree Boosting System. Rapp. tech. URL : [arXiv:1603.02754](https://arxiv.org/abs/1603.02754).
- CHESNEAU, C. (2022). Statistique : méthodes et applications avec R. Cassini.
- CORNEC, D. L. (2018). Etude d'un modèle de coûts du risque sécheresse en France. Mémoire d'actuariat. Dauphine.
- COVÉA (2022). Risques climatiques : quelle prévention? Rapp. tech. URL : [https://www.covea.com/sites/default/files/2023-05/livre\\_blanc\\_covea\\_risque\\_climatique\\_quelles\\_preventions\\_202305.pdf](https://www.covea.com/sites/default/files/2023-05/livre_blanc_covea_risque_climatique_quelles_preventions_202305.pdf).
- DRIAS (2014). Simulations Jouzel-2014. Rapp. tech. Ministère de la transition écologique.
- EAU FRANCE (2023). Les différents types d'inondations. Rapp. tech. URL : <https://www.eaufrance.fr/les-inondations-et-les-submersions-marines>.
- EIOPA (2022). Application guidance on running climate change materiality assessment and using climate change scenarios in the ORSA. Rapp. tech. URL : [https://www.eiopa.europa.eu/publications/application-guidance-climate-change-materiality-assessments-and-climate-change-scenarios-orsa\\_en](https://www.eiopa.europa.eu/publications/application-guidance-climate-change-materiality-assessments-and-climate-change-scenarios-orsa_en).
- ENCYCLOPÉDIE DE L'ENVIRONNEMENT (2021). Modèles de climat. [en ligne]. URL : <https://www.encyclopedie-environnement.org/climat/modeles-climat/>.
- EUROPEAN COMMISSION (2022). Climate Data Store. [en ligne]. URL : <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- FRANCE ASSUREURS (2022). L'assurance des événements naturels en 2021. Rapp. tech.
- GIEC (2022). Synthèse du sixième rapport du GIEC. Rapp. tech. GIEC.
- GOVERNEMENT FRANÇAIS (2022). Enjeux en zones de submersion marine. Rapp. tech.
- HASSANI, H. E. (2017). Modélisation stochastique des inondations en France et Applications en réassurance. Mémoire d'actuariat. ENSAE.
- HERANVAL, A. (2022). Contributions des données de l'assurance à l'étude des risques naturels : application de méthodes d'apprentissage statistique pour l'évaluation de la nature et du coût des dommages assurés liés aux événements naturels en France. Thèse. Sorbonne Université.
- INSTITUT DES ACTUAIRES (2014). Document d'orientation ORSA. Rapp. tech.
- INSTITUT DES MATHÉMATIQUES POUR LA PLANÈTE TERRE (2021). Transport optimal pour la prévision météorologique de la production renouvelable. [en ligne]. URL : <https://impt.math.cnrs.fr/projets-impt/2021-flamary-tantet/>.
- KARAMAKO, V. (2013). Etude Comparative des référentiels IFRS 4 phase 2, Solvabilité 2 et MCEV : Exemple d'un contrat d'épargne Mono support euro. Mémoire d'actuariat. CNAM.
- KOBIA (2022). Comprendre le SMOTE et éviter ses pièges. [en ligne]. URL : <https://kobia.fr/imbalanced-data-smote/>.
- MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE (2022). Géo Risques. [en ligne]. URL : <https://www.georisques.gouv.fr/retrait-gonflement-des-argiles>.
- MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE (2023). DRIAS, les futurs du climat. [en ligne]. URL : <https://www.drias-climat.fr/>.
- MÉTÉO FRANCE (2022). Soil Wetness Index. [en ligne]. URL : [https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=301&id\\_rubrique=40#:~:text=Le%20SWI%20\(de%20l'anglais,l'alimentation%20des%20plantes\)..](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=301&id_rubrique=40#:~:text=Le%20SWI%20(de%20l'anglais,l'alimentation%20des%20plantes)..)
- OBAKRIM, S., AILLIOT, P., MONBET, V. et RAILLARD, N. (2022). EM algorithm for generalized Ridge regression with spatial covariates. Rapp. tech. URL : <https://arxiv.org/abs/2208.04754>.
- OFFICE FRANÇAIS DE LA BIODIVERSITÉ (2022). Le bassin versant. Rapp. tech.
- PARLEMENT EUROPÉEN ET CONSEIL DE L'UNION EUROPÉENNE (2009). Directive 2009/138/CE du Parlement Européen et du Conseil du 25 novembre 2009. Régime réglementaire assurantiel.
- PINCEMIN, G. (2021). Risques climatiques et mortalité, impact du risque canicule à l'horizon 2070. Mémoire d'actuariat. EURIA.

- PLANCHET, F. (2022). Prise en compte des risques climatiques dans la mesure de risque des assureurs. Rapp. tech. ISFA.
- PYRONEAR (2021). C3S ECMWF - ERA5-Land. [en ligne]. URL : [https://pyronear.org/pyro-risks/overview/datasets/C3S-ECMWF\\_ERA5LAND.html](https://pyronear.org/pyro-risks/overview/datasets/C3S-ECMWF_ERA5LAND.html).
- R CORE TEAM (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL : <https://www.R-project.org>.
- RAYMOND, M. (16 jan. 2023). Impact des évolutions climatiques sur les périodes de retour des évènements extrêmes. Rapp. tech. URL : <https://www.youtube.com/watch?v=ir8MfbHBCF8>.
- RIHOUEY, N. (2020). Application des méthodes d'apprentissage à la modélisation de la prime pure en santé collective. Mémoire d'actuariat. ISFA.
- RISK ATTITUDE (2015). Représentation du régime solvabilité 2. URL : <https://riskattitude.net/solvabilite-2-risk-broker-vous-aide-a-votre-conformite/>.
- RÉSEAU ACTION CLIMAT (2022). Réformer le système financier international pour lutter contre le changement climatique. Rapp. tech. Réseau action climat.
- SANDRE (2019). Base du Sandre. Sandre Eau France [en ligne]. URL : <https://www.sandre.eaufrance.fr>.
- SERVEL, M. (2020). Une approche individuelle du provisionnement des sinistres corporels automobiles. Mémoire d'actuariat. ISFA.
- SÉNAT FRANÇAIS (2009). Sécheresse de 2003 : un passé qui ne passe pas. URL : <https://www.senat.fr/rap/r09-039/r09-03911.html>.
- SÉNAT FRANÇAIS (2023). La sécheresse ébranle les fondations du régime CatNat. URL : [https://www.senat.fr/rap/r22-354/r22-354\\_mono.html](https://www.senat.fr/rap/r22-354/r22-354_mono.html).



# Annexe A

## Annexes

### A.1 Les modèles d'arrêt CatNat

#### A.1.1 Les modèles GLM sécheresse/inondations

##### Courbe ROC

Nous présentons ici les courbes ROC des modèles GLM Lasso.  
Pour la sécheresse A.1 :

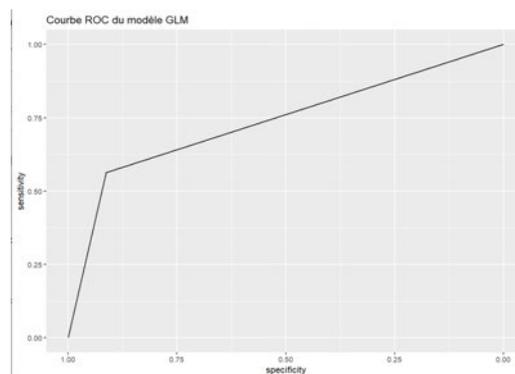


FIGURE A.1 : Représentation de la courbe ROC pour le GLM lasso sécheresse

Pour l'inondation A.2 :

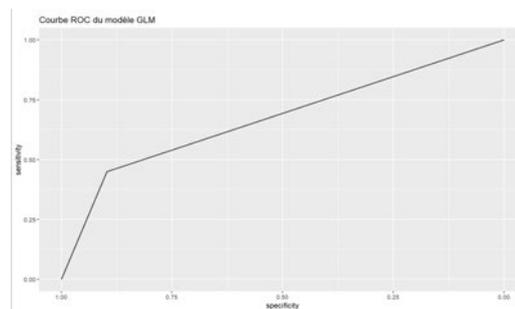


FIGURE A.2 : Représentation de la courbe ROC pour le GLM lasso inondation



### A.1.3 Le modèle XgBoost inondations

Nous représentons l'importance des variables dans le modèle XgBoost inondation A.5.

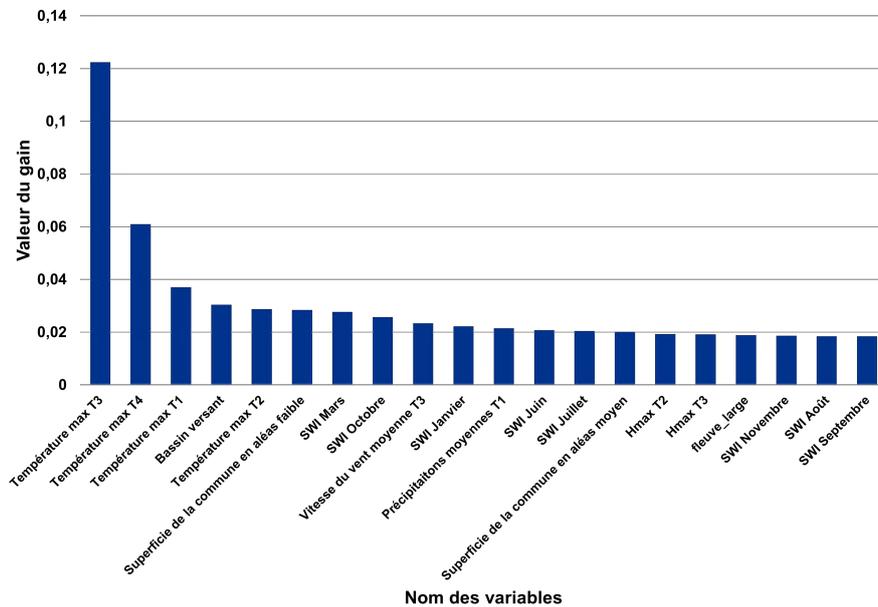


FIGURE A.5 : Représentation de l'importance des variables dans le modèle XgBoost inondation

Nous montrons ensuite la valeur du Kappa pour chaque fold lors de la validation croisée à 10 folds pour la sécheresse (figure A.6).

Fold	Kappa
1	0.4729263
2	0.4781025
3	0.4606249
4	0.4842077
5	0.4902407
6	0.4889781
7	0.4779430
8	0.4684812
9	0.4755665
10	0.4831650

FIGURE A.6 : Représentation du Kappa selon la feuille de CV

De même pour la précision (figure A.7) :

Fold	Précision
1	0.9313500
2	0.9297437
3	0.9305866
4	0.9297437
5	0.9319087
6	0.9310754
7	0.9316293
8	0.9316245
9	0.9325372
10	0.9307961

FIGURE A.7 : Représentation de la précision selon la feuille de CV

#### A.1.4 Les erreurs standards des modèles

Nous représentons ci-dessous l'erreur standard associée au bootstrap réalisé pour la sécheresse lors du chapitre 3 A.8.

Scénario	Horizon	Erreur standard
Scénario 4.5	2030	3.061691e-05
Scénario 4.5	2040	5.410749e-05
Scénario 4.5	2050	3.362197e-05
Scénario 8.5	2030	2.44007e-05
Scénario 8.5	2040	4.400014e-05
Scénario 8.5	2050	4.948921e-05

FIGURE A.8 : Représentation de l'erreur standard pour la sécheresse

Nous représentons ci-dessous l'erreur standard associée au bootstrap réalisé pour l'inondation lors du chapitre 3 A.9.

Scénario	Horizon	Erreur standard
Scénario 4.5	2030	4.219954e-05
Scénario 4.5	2040	4.071827e-05
Scénario 4.5	2050	4.729467e-05
Scénario 8.5	2030	4.448421e-05
Scénario 8.5	2040	5.175573e-05
Scénario 8.5	2050	6.405822e-05

FIGURE A.9 : Représentation de l'erreur standard pour l'inondation