

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

Par : Vincent MAURER

Titre : Assurance forêt : modélisation mathématique du risque tempête et création d'un zonier communal pour la tarification

Confidentialité :  NON  (Durée :  1 an  2 ans)

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

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

.....

.....

.....

*Membres présents du jury de  
l'ISFA*

.....

.....

.....

*Entreprise :*

*Nom :* Actuelia

**SAS ACTUELIA**  
51 Rue Rennequin  
75017 PARIS  
SAS au capital de 30 000 €  
R.C.S. Nanterre 795 239 151  
TVA Intracom. : FR 40 795239151

*Signature :*

*Directeur de mémoire en  
entreprise :*

*Nom :* Louis-Anselme de Lamaze

*Signature :* 

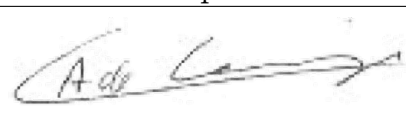
*Invité :*

*Nom :*

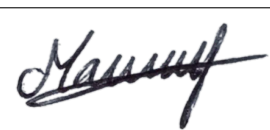
*Signature :*

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

Signature du responsable entreprise



Signature du candidat





# Résumé

Dans un contexte de dérèglement climatique qui accroît l'exposition aux aléas venteux et à leurs conséquences, ce mémoire traite de la gestion du risque tempête en assurance forêt. Il propose une refonte objective de la tarification de cette garantie. L'analyse s'inscrit dans un environnement assurantiel en évolution (exigences prudentielles, attentes de transparence tarifaire, etc.) et dans un marché historiquement peu assuré, où la compréhension fine des expositions locales est déterminante.

L'étude s'appuie sur des bases de données historiques de sinistres et de primes ainsi que sur des données géophysiques agrégées commune par commune sur l'ensemble de la France métropolitaine. Ces données explicatives sont harmonisées dans un même référentiel SIG afin de permettre une modélisation et une interprétation cohérentes du risque.

Le but de ce mémoire est de comprendre comment la création d'un zonier à la maille communale du risque de déracinement des arbres peut fiabiliser les tarifs proposés par les assureurs de ce domaine.

La démarche repose sur la modélisation mathématique du risque tempête. Celle-ci s'effectue dans un premier temps à l'aide de modèles linéaires généralisés puis à l'aide de modèles de *machine learning*. Les prédictions issues de ces travaux sont la base de la construction du zonier, ce dernier étant réparti par niveaux de risque. Une segmentation géographique et stratégique fine du risque est ainsi obtenue.

L'ensemble des travaux de ce mémoire vise à fournir au marché français de l'assurance forêt un cadre méthodologique reproductible et un outil concret de fiabilisation des tarifs, aligné avec les enjeux climatiques contemporains.

**Mots clés** : Catastrophe Naturelle, Assurance forêt, Tempête, Tarification, Modèle Linéaire Généralisé, *Machine Learning*, Zonier, Période de retour, Valeurs extrêmes, Copule



# Abstract

In a context of climate change that increases exposure to wind hazards and their consequences, this thesis addresses storm risk management in forest insurance. It proposes a data-driven redesign of pricing of this guarantee. The analysis is set against the backdrop of a changing insurance environment (prudential requirements, expectations of pricing transparency, etc.) and a historically underinsured market, where a detailed understanding of local exposures is necessary.

The study draws on historical databases of claims and premiums, as well as aggregated geophysical data for each municipality across mainland France. This explanatory data is harmonized within a single GIS reference system to enable consistent risk modeling and interpretation.

The aim of this thesis is therefore to understand how the conception of a commune-level zoning map for the risk of tree uprooting can enhance the credibility and robustness of pricing offered by insurers in this field.

The approach is based on mathematical modeling of storm risk. This is initially carried out using generalized linear models and then using machine learning models. The predictions generated by this model form the basis for the construction of the zoning system, which is divided into risk levels. This zoning system allows for detailed geographical and strategic segmentation.

The overall aim of this thesis is to provide the French forest insurance market with a reproducible methodological framework and a concrete tool for improving the reliability of rates, in line with contemporary climate issues.

**Keywords :** Natural Disaster, Forest Insurance, Storm, Pricing, Generalized Linear Model, Machine Learning, Zoning, Return Period, Extreme Value Theory, Copula



# Note de synthèse

## Contexte et problématique

Ce mémoire s'ancre dans un contexte de dérèglement climatique. En effet, les dernières années sont caractérisées par une augmentation de la fréquence et de la sévérité des sinistres liés aux catastrophes naturelles. Cette évolution menace l'équilibre financier et la projection stratégique des compagnies d'assurance garantissant ces aléas.

Le secteur forestier n'échappe pas à ces enjeux. Malgré son développement encore limité sur le territoire français métropolitain, l'assurance forestière se doit d'anticiper les évolutions et de mettre à jour ses méthodes tarifaires afin d'éclairer ses clients sur la nature des phénomènes couverts et la justification des primes facturées. La garantie tempête, couverture majeure proposée, requiert à ce titre une analyse approfondie.

Les travaux effectués visent à déterminer de quelle manière l'analyse du risque tempête à l'échelle communale en assurance forêt peut contribuer à optimiser la fiabilité des tarifs élaborés par les assureurs spécialisés.

## Cadre et données utilisées

L'assurance forêt désigne une assurance de dommages aux biens. Elle garantit la valeur des peuplements et infrastructures forestiers contre les risques naturels assurables tels que l'incendie et les tempêtes.

Dans ce mémoire, seul le risque tempête est modélisé. Une tempête est qualifiée comme telle lorsque le vent moyen atteint ou dépasse 89 km/h sur une période d'au moins 10 minutes. En France, les tempêtes ne sont pas aussi destructrices que dans certaines autres régions du monde (Philippines, Amérique du Nord...). Toutefois, les tempêtes Lothar et Martin en 1999 ainsi que Klaus en 2009 ont laissé localement des impacts et des coûts d'indemnisation importants pour les citoyens ainsi que pour les assureurs.

Bien que des forêts soient victimes de déracinement lors d'événements importants comme ceux cités précédemment, ou simplement au quotidien, le taux d'assurance des forêts privées en France reste très faible (4 %) comparé à d'autres pays (95 % en Suède). Une meilleure compréhension du risque de déracinement des arbres par l'impact des tempêtes apparaît primordiale.

En se basant sur *Windthrow Hazard Classification*, une étude de 1985 cherchant à modéliser le risque de déracinement au Royaume-Uni, il a été possible d'identifier plusieurs variables pouvant expliquer le phénomène.

Nom	Explication	Unité
Sol	Nature du sol prédominant dans la commune	—
Élévation	Altitude moyenne de la commune	m
Essence	Catégorie des arbres présents : conifères, feuillus ou mélange des deux dans la commune	—
Vent	Vitesse maximale du vent relevée dans le département l'année du sinistre	km/h
Abri Topographique	Mesure de la protection offerte par le relief environnant, calculée dans un rayon de 5 km autour du point d'intérêt	°

TABLE 1 – Variables retenues : Explication technique

L'étude britannique a été complétée et enrichie par diverses discussions avec des spécialistes du domaine forestier. En premier lieu, des échanges ont été effectués avec Barry Gardiner, scientifique senior à l'Institut Européen de la Forêt Cultivée (IEFC), qui a pu aider dans la recherche de données à la maille communale souhaitée. De plus, Groupama Forêts Assurances, acteur majeur du sujet en France, a fourni sa base de données de sinistres et de primes, qui a été modifiée pour ces travaux. Néanmoins, la zone d'implantation majeure des données demeure le Bassin aquitain et les sinistres les plus fréquents se retrouvent bien en 1999 et 2009. Au total, 2701 sinistres sont référencés.

Cette base de données historique permet d'avoir un passif fiable de déracinement des arbres dans certaines communes, que nous pouvons expliquer par les données géophysiques citées en table 1. Le lien trouvé peut ensuite être développé pour toutes les communes de France métropolitaine.

## Modélisation du risque

Ce lien entre la variable à expliquer, à savoir le ratio surface sinistrée / surface assurée, et les différentes variables explicatives est capturé par des modèles adaptés pour cela. En premier lieu, une représentation du risque similaire à celle de l'étude britannique précitée a été testée. Cette modélisation attribue un score de risque à chaque variable explicative pour une localisation donnée. Le total des scores définit le risque global porté à cet emplacement.

Afin de prendre en compte les interactions entre les variables et de mettre en perspective les résultats du modèle précédent, des modèles linéaires généralisés (MLG) ont ensuite été implémentés. Parmi ceux testés, celui ayant fourni les meilleurs résultats est un MLG quasi-binomial avec une fonction lien logit.

Les coefficients obtenus viennent confronter certaines intuitions, qui avaient été confortées par le modèle précédent. Par exemple, l'altitude n'augmente pas le risque de déracinement mais le réduit, ce qui n'était pas capturé par le modèle du score. Ces modélisations fines viennent donc améliorer la compréhension du risque tempête en assurance forêt.

Le MLG quasi-binomial fournit les métriques de la table 2.

Mesure	Valeur
Pseudo- $R^2$	0,116
RMSE	0,325
Précision	0,492
Rappel	0,756
$F_1$ -score	0,596
AUC PRC	0,584

TABLE 2 – MLG Quasi-binomial : Mesures de performance

Ces résultats sont une référence acceptable mais la capacité prédictive ainsi que l'explication des données initiales peuvent être améliorées. Étant susceptibles d'offrir une flexibilité accrue face aux non-linéarités et aux interactions, les modèles de *machine learning* ont ensuite été considérés, à savoir *Random Forest* ainsi que *XGBoost*. Les deux modèles ayant donné des résultats similaires, *Random Forest*, meilleur prédicteur, est présenté par la suite.

*Random Forest* requiert l'optimisation d'hyperparamètres pour que le modèle soit le plus efficace possible. La configuration retenue est la suivante :

- $B = 1000$  arbres
- $mtry = 2$  variables par division
- $nodesize = 5$  observations minimales par nœud terminal

Les métriques du modèle sont présentées dans la table 3 et sont issues du modèle ainsi optimisé, qui améliore les résultats par rapport au modèle de base non optimisé.

Mesure	Valeur	Comparaison MLG
$R^2$	0,350	+ 0,234
RMSE	0,268	- 0,057
Précision	0,784	+ 0,292
Rappel	0,730	- 0,026
$F_1$ -score	0,756	+ 0,160
AUC PRC	0,823	+ 0,239

TABLE 3 – *Random Forest* : Mesures de performance

Une nette amélioration des métriques par rapport au MLG quasi-binomial est observée. Le coefficient de détermination  $R^2$  traduit une meilleure capacité explicative du modèle, tandis que le  $F_1$ -score met en évidence une amélioration de l'équilibre entre précision et rappel. L'ensemble de ces résultats est confirmé par la courbe précision-rappel de la figure 1, dont l'aire sous la courbe atteint 0,823.

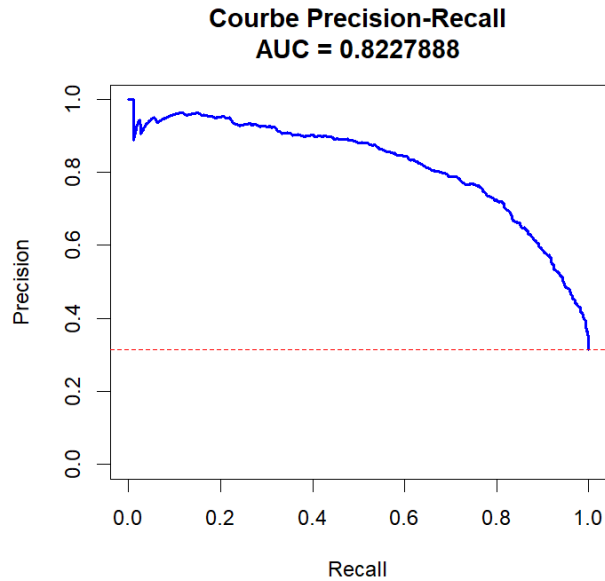


FIGURE 1 – *Random Forest* : Courbe précision-rappel

La pertinence du modèle *Random Forest* est renforcée par la proximité entre la moyenne des ratios de destruction prédits et la moyenne observée dans la base de sinistres. Cette moyenne estimée était nettement plus faible pour le MLG quasi-binomial notamment.

Moyenne Ratio Prédit <i>Random Forest</i>	Moyenne sinistralité historique
33,08 %	32,33 %

TABLE 4 – *Random Forest* : Moyenne prédite vs Moyenne observée du ratio

Une fois le modèle optimal estimé et la prédiction retranscrite à toutes les communes françaises, il est possible de visualiser le risque via un zonier.

## Zonage

Un zonier est un référentiel de zones géographiques homogènes. Chacune de ces zones est rattachée à une catégorie de risque. C'est un outil actuariel efficace pour délimiter un périmètre (ici la France métropolitaine) par niveaux de risque équivalents. Les modélisations ont été effectuées à l'aide du logiciel de cartographie QGIS. Le zonier agrégé en 10 zones de la figure 2 présente les ratios prédits par le modèle *Random Forest*.

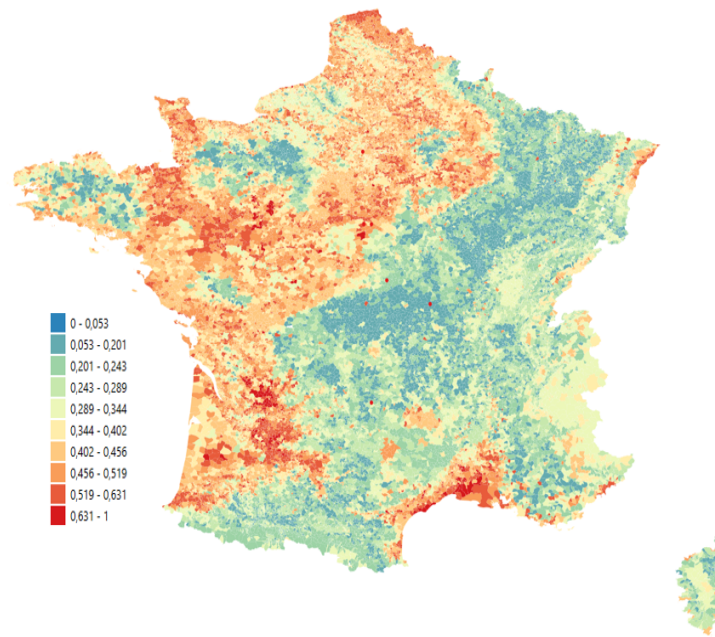


FIGURE 2 – Zonier agrégé en 10 zones : *Random Forest*

Pour une lecture commerciale, le risque est regroupé en 4 zones géographiques. Les effets d'essence sont déjà pris en compte lors de la modélisation, et le zonier applique un ajustement différent selon l'essence : trois grilles de 4 zones, selon l'essèce majoritaire de la commune (conifères, feuillus, mixte) sont ainsi obtenues, sans double comptage de l'impact de l'essèce majoritaire.

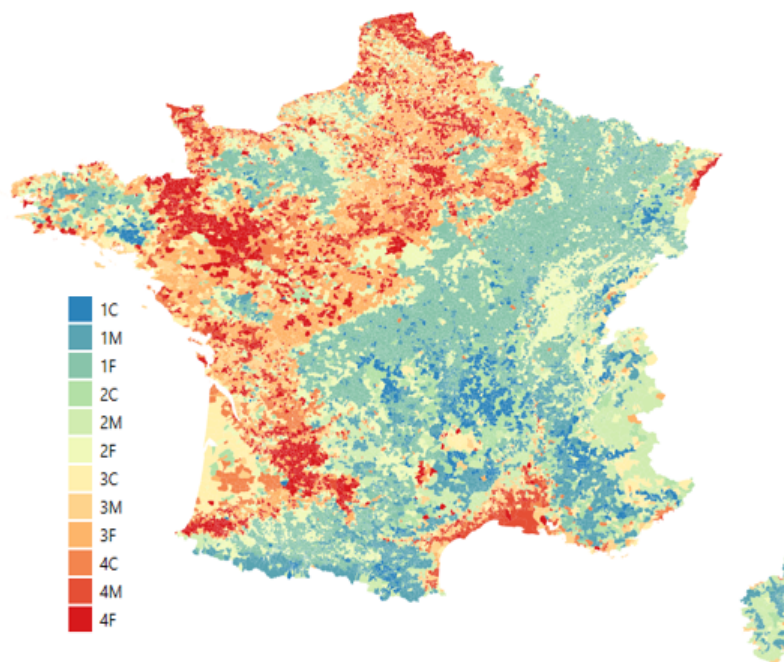


FIGURE 3 – Zonier agrégé en 12 zones : *Random Forest*

Les deux zoniers montrent clairement que le Bassin aquitain est exposé significativement : les vents venant de l'Atlantique y frappent souvent sans beaucoup d'abri, ce qui explique la forte présence dans la région de l'un des principaux assureurs du domaine forestier français. Les zones

côtières sont aussi plus risquées, car elles sont directement exposées et peu protégées. À l'inverse, les massifs et, plus largement, l'est de la France sont globalement moins à risque grâce au relief et à l'éloignement des vents océaniques.

## Tarification

Ce zonier est ensuite adapté de manière opérationnelle pour servir à la tarification.

Tout d'abord, il est primordial de noter que, dans la base de données étudiée, il existe un déséquilibre très important entre les S/P dits attritionnels (années hors catastrophes) et le S/P catastrophe en 2009, année de passage de la tempête Klaus dans le Bassin aquitain, lieu d'implantation majeur de l'assureur étudié. Ceci est mis en avant dans la table 5, qui présente les ratios Sinistres / Primes de ces années.

S/P attritionnel	S/P catastrophe
6 %	1 155 %

TABLE 5 – Ratios S/P : Distinction entre attritionnel et catastrophe

L'assureur a une cible annuelle de S/P supposée égale à 80 % dans le Bassin aquitain. Cette cible de 80 % a été fixée pour garantir la rentabilité du portefeuille tout en constituant des réserves suffisantes pour absorber les sinistres extrêmes, couvrir les frais de gestion et de réassurance, et maintenir une marge de sécurité face à la volatilité du risque tempête.

Pour connaître le S/P annuel de l'assureur et la différence par rapport à la cible de 80 %, le S/P catastrophe doit être lissé sur plusieurs années, et être additionné au S/P attritionnel. Pour savoir comment lisser cette partie catastrophe, une période de retour doit être calculée, comme l'explique le calcul ci-dessous :

$$S/P_{\text{Annuel}} = S/P_{\text{Attritionnel}} + \frac{S/P_{\text{Catastrophe}}}{\text{Période de retour}}$$

Une période de retour est la durée moyenne entre deux occurrences d'un événement d'intensité donnée. Elle correspond à l'inverse de la probabilité annuelle d'occurrence. Dans ce cas d'étude, la période recherchée est la période de retour d'une tempête similaire à la tempête Klaus sur le Bassin aquitain, celle-ci étant la dernière à avoir touché sévèrement la zone étudiée du portefeuille. Pour cela, la base de données de Météo-France SYNOP est utilisée, et plus précisément la variable indiquant la vitesse maximale de rafale mesurée pendant une période, qui est ici la journée.

Les données SYNOP sont disponibles à partir de novembre 2003 pour les rafales. Trois stations météorologiques sont localisées dans le Bassin aquitain. Les stations de Pointe de Chassiron, Bordeaux Mérignac et Mont-de-Marsan sont ainsi à l'étude.

Après avoir étudié chaque station individuellement, une première approche envisagée pour considérer l'impact sur les trois stations simultanément a consisté à créer une station prenant la rafale maximale journalière entre les trois stations du Bassin aquitain. Pour un niveau de retour  $x = 134,88$  km/h, qui est la rafale maximale moyenne observée dans les stations du Bassin

aquitain le jour majeur du passage de Klaus, la période de retour estimée du maximum est de **11,28979 ans**.

Néanmoins, cette approche ignore la dépendance entre les variables et sous-estime souvent le risque combiné. Pour mieux capturer le phénomène observé, les copules sont introduites : elles relient entre elles les distributions marginales après les avoir transformées en variables uniformes sur  $[0,1]$ , ce qui permet de modéliser précisément les interactions. Les corrélations entre stations sont montrées par la figure 4.

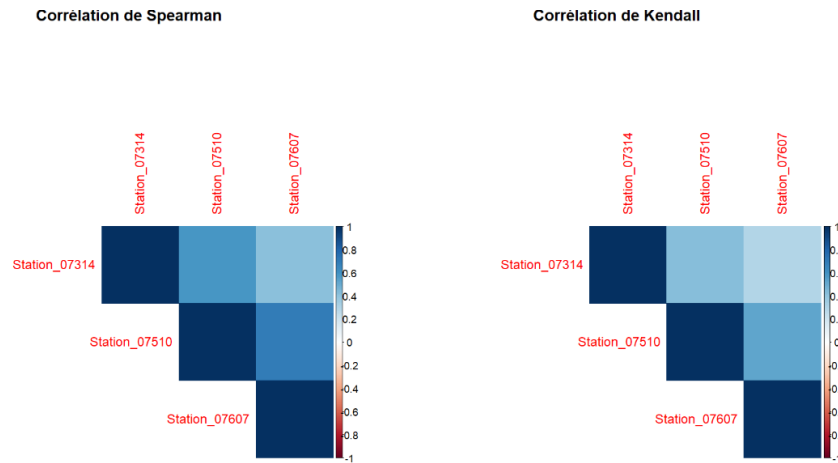


FIGURE 4 – Analyse des stations : Corrélations

Par critère de l’AIC et du BIC, la copule correspondant le mieux aux données est une copule de Gumbel avec pour paramètre  $\alpha = 1,607$ .

La période de retour calculée est alors de **14,5 ans**.

Pour lisser la charge exceptionnelle de 2009 ( $S/P = 1155\%$ ), le coût catastrophe est annualisé :

$$S/P_{\text{Annuel}} = S/P_{\text{Attritionnel}} + \frac{S/P_{\text{Catastrophe}}}{\text{période de retour}} = 5,90\% + \frac{1155\%}{14,5} = 5,90\% + 79,66\% = \boxed{85,56\%}$$

L’écart à la cible de 80 % dans le Bassin aquitain est donc de 5,56 points de pourcentage. Cela implique une hausse tarifaire moyenne de 6,95 % dans cette zone géographique pour atteindre l’objectif.

Le zonier obtenu précédemment intervient comme passerelle opérationnelle entre la modélisation et les tarifs. Pour chaque classe de risque, on compare le ratio de destruction moyen prédit à la moyenne globale par essence du Bassin aquitain. La figure 5 récapitule le processus permettant d’obtenir le tarif souhaité.

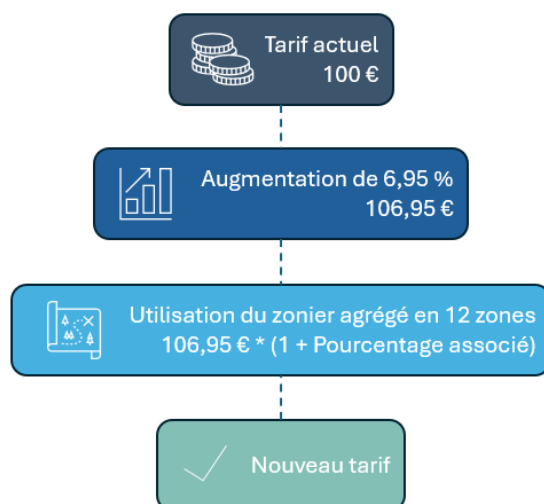


FIGURE 5 – Détermination du nouveau tarif : Schéma

Cette mécanique garantit deux éléments. Tout d’abord, l’évolution tarifaire moyenne au sein du Bassin aquitain atteint +6,95 %. Ceci permet d’atteindre la cible de S/P de 80 %. Deuxièmement, à l’intérieur du Bassin aquitain, les zones plus risquées que la moyenne voient leurs tarifs augmenter et les zones moins risquées voient leurs tarifs diminuer.

Outre le Bassin aquitain, dont l’évolution tarifaire cible a motivé les évolutions tarifaires, les mêmes augmentations de tarif sont retenues pour toutes les communes de France métropolitaine. Cela permet à l’assureur de visualiser les zones à faible risque (zones 1 et 2) afin de s’y implanter avec un tarif qui soit compétitif.

Les évolutions tarifaires se retrouvent dans la table 6.

Classe	Variation
1C	-20,32 %
2C	-8,76 %
3C	+1,40 %
4C	+15,30 %
1F	-11,20 %
2F	-2,00 %
3F	+10,82 %
4F	+22,83 %
1M	-12,86 %
2M	-4,19 %
3M	+8,79 %
4M	+18,15 %

TABLE 6 – Bassin aquitain : Évolutions tarifaires par classe

L’évolution souhaitée de 6,95 % dans le Bassin aquitain est obtenue. Celle-ci valide l’objectif de refonte de la tarification du risque tempête en se basant sur un zonier à la maille communale. Cette refonte est opérationnelle et lisible pour les assureurs comme pour les assurés, ce qui était le but recherché.

## Conclusion

La figure 6 récapitule et lie les différents travaux de ce mémoire. Dans un premier temps, la constitution des bases de données a permis de modéliser le risque tempête. En se basant sur cette modélisation, des prédictions du ratio de déracinement des arbres ont été obtenues. Le risque plus ou moins élevé indiqué par le ratio a conduit à la constitution d'un zonier. Ce dernier a servi à délimiter le risque et à repenser la tarification mise en place jusqu'alors.

L'outil proposé permet un développement stratégique à une échelle supérieure à l'implantation actuelle.

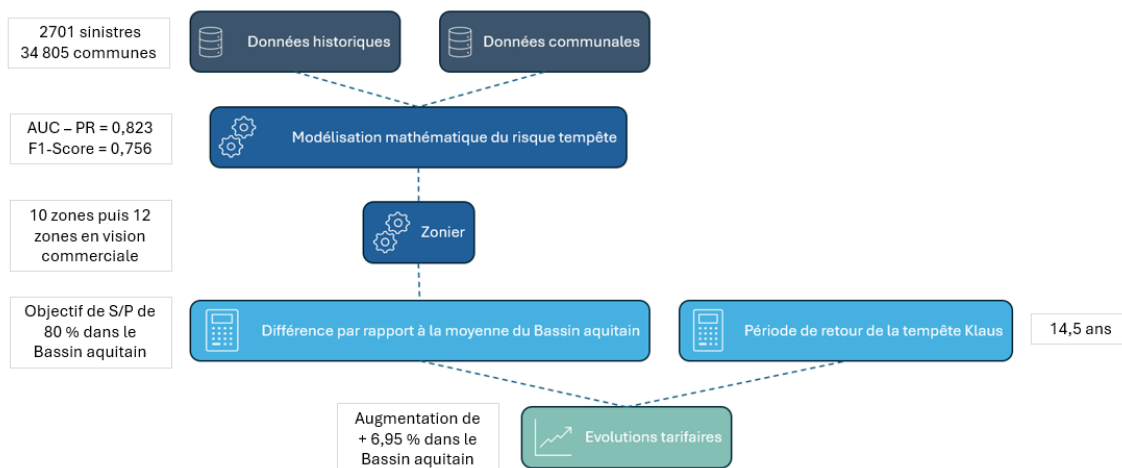


FIGURE 6 – Récapitulatif du mémoire : Schéma

Sur le terrain, le zonier et la proposition d'ajustement tarifaire constituent un outil opérationnel dont la mise en œuvre nécessite une validation par les équipes commerciales pour s'assurer de l'acceptabilité des évolutions proposées. La fiabilité du modèle dans le temps repose sur une actualisation régulière des données géophysiques et de sinistralité, ainsi que sur l'enrichissement progressif de l'historique hors Bassin aquitain, actuellement sous-représenté dans le portefeuille.

La méthodologie employée est avant tout à retenir dans ce mémoire. Celle-ci est innovante pour l'assurance forêt et se base sur des données crédibles. La modélisation du risque basée sur ces données précises permet une compréhension du risque à une échelle fine. Les impacts tarifaires proposés sont donc maîtrisés et explicables à tous.

Ce mémoire répond aux problématiques actuelles de l'assurance forêt et fournit une base de travail potentielle à de nouvelles études dans ce domaine. La méthodologie employée est également transposable à d'autres objets d'étude.



# Executive summary

## Context and issues

This thesis is written in a context marked by the climate emergency. Indeed, recent years have been characterized by an increase in the frequency and in the severity of natural disaster-related claims. This trend threatens the financial stability and the strategic planning of insurance companies that cover such risks.

The forestry sector is no exception to these challenges. Despite its still limited development in France, forestry insurance must anticipate changes and update its pricing methods in order to inform its customers about the nature of the phenomena covered and the justification for the premiums charged. Storm coverage, a major type of coverage offered, requires in-depth analysis.

This research aims to determine how storm risk analysis at the municipal level in forest insurance can help in order to optimize the reliability of prices set by forest insurers.

The approach adopted consists first of explaining the main concepts, then presenting all the data used. Once this information has been brought together, it is used to mathematically model storm risk. This approach results in the establishment of municipal zoning which, after operational aggregation, enables pricing practices to be redesigned.

## Framework and data used

Forest insurance refers to property damage insurance. It covers the value of forest stands and infrastructure against insurable natural risks such as fire and storms.

In this thesis, storm risk is the only one considered. A storm is defined as occurring when the average wind speed reaches or exceeds 89 km/h over a period of at least 10 minutes. In France, storms are not as destructive as in some other parts of the world (the Philippines, North America, etc.). However, storms Lothar and Martin in 1999 and Klaus in 2009 had a significant local impact and resulted in substantial compensation costs for citizens as well as for insurers.

Although forests are subject to uprooting during major events such as those mentioned above, or simply in everyday life, the insurance take-up rate for private forests in France remains very low (4%) compared to other countries (95% in Sweden for example). A better understanding of the risk of tree uprooting due to the impact of storms therefore appears essential.

Based on *Windthrow Hazard Classification*, a 1985 study seeking to model the risk of uprooting in the United Kingdom, it has been possible to identify several variables that could explain the phenomenon.

Name	Explanation	Unit
Soil	Predominant soil type in the municipality	—
Elevation	Average altitude of the municipality	m
Tree species	Category of trees present : conifers, broadleaves, or a mix of both in the municipality	—
Wind	Maximum wind speed recorded in the department in the year of the claim	km/h
Topographic shelter	Measure of the protection offered by the surrounding terrain, calculated within a 5 km radius around the municipality	°

TABLE 7 – Variables considered in this thesis : Technical details

The British study was completed and enriched by various discussions with specialists in the forestry field. First, exchanges were held with Barry Gardiner, senior scientist at the *Institut Européen de la Forêt Cultivée (IEFC)*, who was able to help us in our search for data at the desired municipal level. In addition, Groupama Forêts Assurances, a major player in French forestry insurance, provided us with their database of claims and premiums, which was modified for this study. Nevertheless, the main area covered by the data remains the Aquitaine Basin, and the most frequent claims occurred in 1999 and 2009. A total of 2,701 claims are listed.

This historical database provides a record of tree uprooting in certain municipalities, which we can explain using the geophysical data cited in table 7. The link found can then be developed for all municipalities in France.

## Risk modeling

This link between the variable to be explained, namely the ratio of damaged area to insured area, and the various explanatory variables is captured by models adapted for this purpose. First, a representation of risk similar to that used in the aforementioned British study was tested. This model assigns a risk score to each explanatory variable for a given location. The total scores define the overall risk at that location.

In order to take into account interactions between variables and put the results of the previous model into perspective, generalized linear models (GLM) were then implemented. Among those tested, the one that provided the best results was a quasi-binomial GLM with a logit link function.

The coefficients obtained challenge certain assumptions that had been reinforced by the previous model. For example, altitude does not increase the risk of uprooting but reduces it, which was not captured by the scoring model. These detailed models therefore improve our understanding of storm risk in forest insurance.

The quasi-binomial GLM provides the metrics shown in Table 8.

<b>Metric</b>	<b>Value</b>
Pseudo- $R^2$	0.116
RMSE	0.325
Precision	0.492
Recall	0.756
$F_1$ -score	0.596
AUC PRC	0.584

TABLE 8 – Quasi-binomial GLM : Performance metrics

These results are an acceptable basis, but the predictive capacity and explanation of the initial data can be improved. Machine learning models, which are likely to offer greater flexibility in dealing with non-linearities and interactions, were then considered, namely Random Forest and XGBoost. As both models produced similar results, Random Forest, which is the better predictor, is presented below.

Random Forest requires hyperparameter optimization to ensure that the model is as effective as possible. The configuration chosen is shown below :

- $B = 1000$  trees
- $mtry = 2$  variables per split
- $nodesize = 5$  minimum observations per terminal node

The model metrics are presented in table 9 and are based on the optimized model, which improves results compared to the unoptimized baseline model.

<b>Metric</b>	<b>Value</b>
$R^2$	0.350
RMSE	0.268
Precision	0.784
Recall	0.730
$F_1$ -score	0.756
AUC PRC	0.823

TABLE 9 – Random Forest : Performance metrics

There is a clear improvement in metrics compared to the quasi-binomial GLM. The coefficient of determination  $R^2$  reflects the model's improved explanatory power, while the  $F_1$ -score highlights an improvement in the balance between precision and recall. All of these results are confirmed by the precision-recall curve in figure 7, whose area under the curve reaches 0.823.

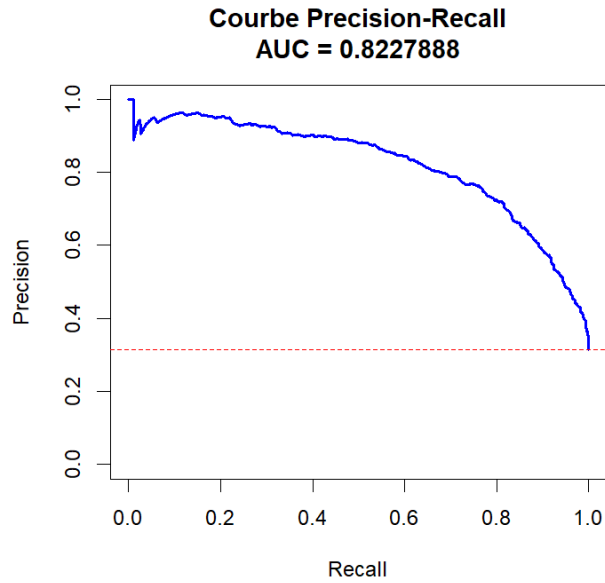


FIGURE 7 – Random Forest : Precision-recall curve

The relevance of the Random Forest model is reinforced by the proximity between the average predicted destruction ratios and the average observed in the claims database. This estimation was significantly lower for the quasi-binomial GLM for example.

<b>Average Predicted Ratio</b> <i>Random Forest</i>	<b>Average Historical Ratio</b>
33.08%	32.33%

TABLE 10 – *Random Forest* : Average predicted vs. average observed ratio

Once the optimal model has been estimated and the prediction applied to all French municipalities, it is possible to visualize the risk using a zoning map.

## Zoning

A zoning map is a reference system of homogeneous geographical areas. Each of these areas is linked to a risk category. It is an effective actuarial tool for defining a perimeter (in this case, metropolitan France) by equivalent risk levels. The modeling was performed using QGIS mapping software. The zoning system, aggregated into 10 zones in figure 8, presents the ratios predicted by the Random Forest model.

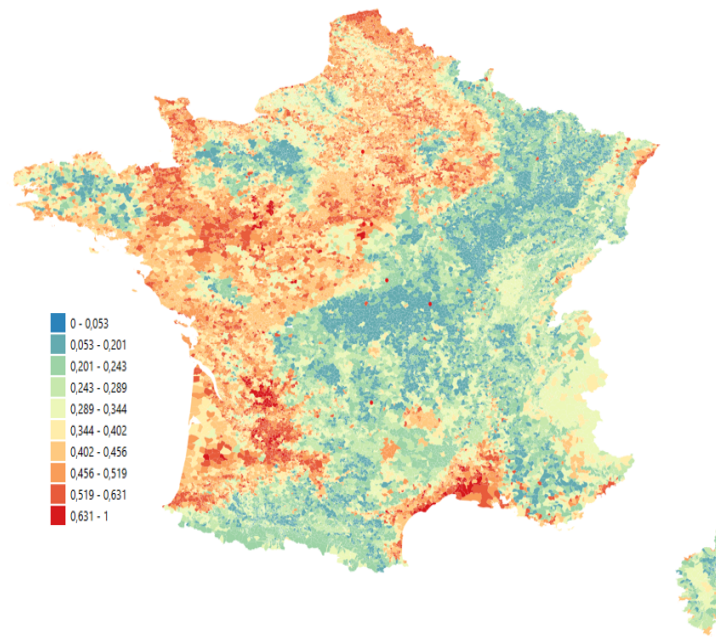


FIGURE 8 – Zoning map aggregated into 10 areas : *Random Forest*

For commercial purposes, risk is grouped into four geographical zones. The effects of tree species are already taken into account during modeling, and the zoning system applies a different adjustment according to species : three grids of four zones, depending on the predominant species in the municipality (conifers, broadleaves, mixed), are thus obtained, without double counting the impact of the predominant species.

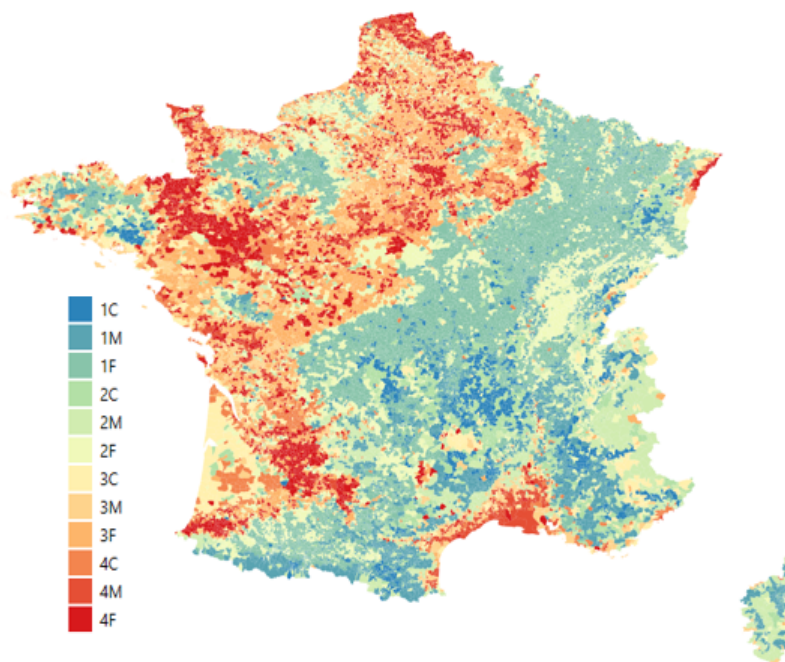


FIGURE 9 – Zoning map aggregated into 12 areas : *Random Forest*

The two maps clearly show that the Aquitaine Basin is significantly exposed : winds from the Atlantic often hit the region without much shelter, which explains the strong presence in the region of one of the leading insurers in the French forestry sector. Coastal areas are also more at

risk, as they are directly exposed and offer little protection. By contrast, mountain areas and, more broadly, eastern France are generally less at risk thanks to the terrain and distance from ocean winds.

## Pricing

This zoning map is then adapted for operational use in pricing.

Firstly, it is important to note that, in the database studied, there is a significant imbalance between the so-called attrition loss ratio (years without disasters) and the disaster loss ratio in 2009, the year in which Storm Klaus hit the Aquitaine Basin, a major location for the insurer studied. This is highlighted in table 11, which shows the loss ratios for these years.

Attritional loss ratio	Catastrophe loss ratio
6%	1155%

TABLE 11 – Loss ratios : Distinction between attritional and catastrophe

The insurer has an annual loss ratio target of 80% in the Aquitaine Basin. This 80% target was set to ensure the portfolio’s profitability while building sufficient reserves to absorb extreme claims, cover management and reinsurance costs, and maintain a safety margin in the face of storm risk volatility.

To determine the insurer’s annual loss ratio and the difference from the 80% target, the catastrophe loss ratio must be smoothed over several years and added to the attrition loss ratio. To determine how to smooth this catastrophe portion, a return period must be calculated, as explained in the calculation below :

$$\text{Loss Ratio}_{\text{Total}} = \text{Loss Ratio}_{\text{Attritional}} + \frac{\text{Loss Ratio}_{\text{Catastrophe}}}{\text{Return Period}}$$

A return period is the average time between two occurrences of an event of a given intensity. It corresponds to the inverse of the annual probability of occurrence. In this case study, the period sought is the return period of a storm similar to Storm Klaus in the Aquitaine Basin, which was the last storm to severely affect the area studied in the portfolio. To do this, the Météo-France SYNOP database is used, specifically the variable indicating the maximum gust speed measured during a period, which in this case is one day.

SYNOP gust data has been available since November 2003, providing a large historical record. Three weather stations are located in the Aquitaine Basin : Pointe de Chassiron, Bordeaux Mérignac, and Mont-de-Marsan.

After studying each station individually, an initial approach to consider the impact on the three stations simultaneously consisted of creating a station that took the maximum daily gust between the three stations in the Aquitaine Basin. For a return level of  $x = 134.88$  km/h, which is the average maximum gust observed at the stations in the Aquitaine Basin on the day Klaus made landfall, the estimated return period of the maximum is **11.28979 years**.

However, this approach ignores the dependence between variables and often underestimates the combined risk. To better capture the observed phenomenon, copulas are used : they link

marginal distributions after transforming them into uniform variables on  $[0,1]$ , which allows interactions to be modeled accurately, the correlations of which are shown in figure 10.

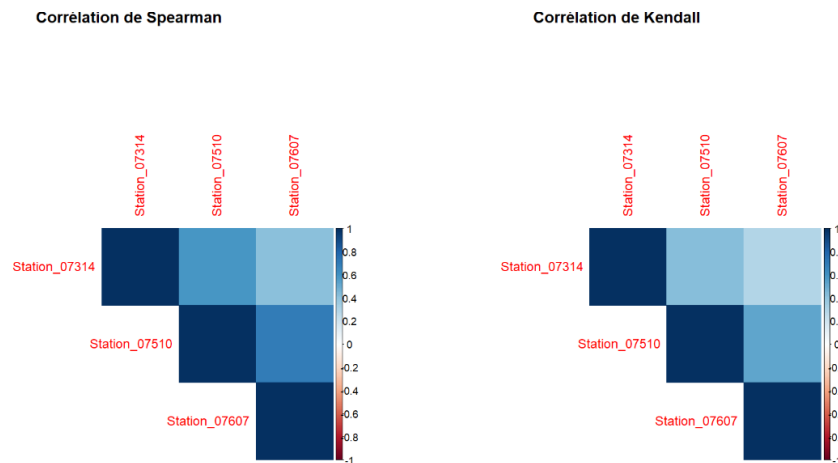


FIGURE 10 – Station analysis : Correlation

According to the AIC and BIC criteria, the copula that best fits the data is a Gumbel copula with parameter  $\alpha = 1.607$ .

After calculations with joint probabilities under the chosen copula, the calculated return period is **14.5 years**.

To smooth the exceptional burden from 2009 (Loss Ratio = 1,155%), the catastrophe cost is annualized :

$$\text{Loss Ratio}_{\text{annual}} = \text{Loss Ratio}_{\text{attritional}} + \frac{\text{Loss Ratio}_{\text{catastrophe}}}{\text{return period}} = 5.90\% + \frac{1155\%}{14.5} = \boxed{85.56\%}$$

The gap to the 80% target in the Aquitaine Basin is therefore 5.56 percentage points. This implies an average rate increase of 6.95% in this geographic area to reach the objective.

The zoning map obtained previously acts as an operational bridge between modeling and pricing. For each risk class, the predicted average destruction ratio is compared to the overall average by species for the Aquitaine Basin. Figure 11 summarizes the process used to obtain the desired price.

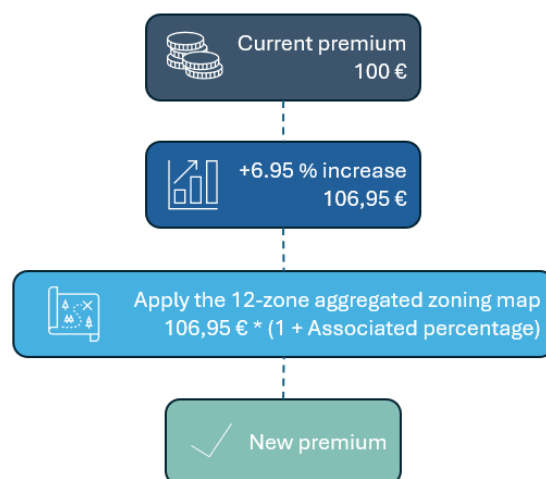


FIGURE 11 – New Rate Determination : Diagram

This approach guarantees two things. First, the average rate increase within the Aquitaine Basin reaches +6.95%. This allows the loss ratio target of 80% to be achieved. Second, within the Aquitaine Basin, areas with above-average risk see their rates increase, while areas with below-average risk see their rates decrease.

Rate rises are obtained for all French cities outside of the Aquitaine Basin, whose goal rate evolution prompted the price adjustments. This enables the insurer to locate low-risk regions (zones 1 and 2) and provide competitive rates there.

The pricing changes are indicated in table 12.

Class	Change
1C	−20,32 %
2C	−8,76 %
3C	+1,40 %
4C	+15,30 %
1F	−11,20 %
2F	−2,00 %
3F	+10,82 %
4F	+22,83 %
1M	−12,86 %
2M	−4,19 %
3M	+8,79 %
4M	+18,15 %

TABLE 12 – Aquitaine Basin : Pricing changes by class

The targeted increase of 6.95% in the Aquitaine Basin has been achieved. It validates the objective of overhauling storm risk pricing based on a municipal zoning system. This overhaul is operational and clear to both teams and policyholders, which was the intended goal.

## Conclusion

The diagram in figure 12 shows the structure of this thesis. The creation of a municipal zoning system, based on statistical and machine learning models, makes rates more reliable by directly linking the local risk level to the rate applied. This zoning system improves the geographical relevance of pricing, reduces imbalances linked to extreme events, and provides insurers with an operational and reproducible tool to adjust their premiums in a transparent and robust manner, while developing strategically.

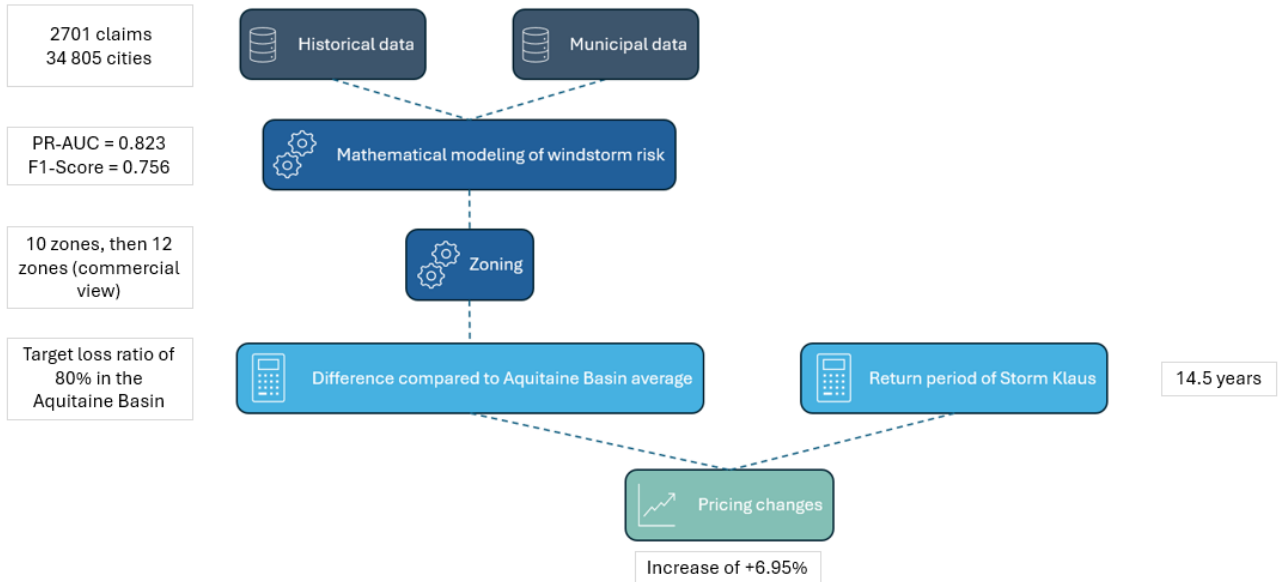


FIGURE 12 – Summary of the thesis : Diagram

Operationally, the zoning map and the proposed rate adjustment constitute an operational tool whose implementation requires validation by the sales teams to ensure the acceptability of the proposed changes. The reliability of the model over time depends on regular updates of geophysical and loss data, as well as the gradual enrichment of the history outside the Aquitaine Basin, which is currently underrepresented in the portfolio.

Beyond addressing the issue, this thesis proposes a replicable methodology : starting with the collection and aggregation of several sources at the municipal level to arrive at a mathematical risk model and a zoning system. The latter provides geographical input for pricing based on a portfolio return period that challenges the standard values used. The various key specificities were generally overcome : scarcity and imbalance of claims, correlation of extreme values, and commercial explainability constraints.

This thesis addresses current issues in forest insurance and provides a potential basis for further research in this field. The methodology used can also be applied to other areas of study.



# Remerciements

Tout d'abord, je tiens à remercier David Fitouchi, Frank Boukobza, Louis-Anselme de Lamaze et Camille Blanc-Vannet, associés du cabinet Actuelia, pour m'avoir donné l'opportunité de rejoindre leur équipe.

Je remercie tout particulièrement Louis-Anselme de Lamaze et Alban Petey, qui ont encadré ce mémoire au sein du cabinet. Leurs conseils pertinents et originaux ainsi que leur suivi attentif m'ont permis d'avancer sereinement tout en laissant place à mon initiative et à ma créativité.

Merci à l'ensemble de mes collègues d'Actuelia pour cette alternance et ces premières semaines de CDI. Leurs conseils et la bonne humeur quotidienne de chacun sont très appréciés pour cette première expérience professionnelle.

Je remercie Maud Thomas, ma tutrice pédagogique ISFA, pour ses retours rapides et ses conseils clairs et utiles. Je remercie par ailleurs l'ensemble des professeurs de l'ISFA pour les trois ans de formation passés dans cette école.

Merci à Groupama Forêts Assurances pour leur confiance. Merci pour les conseils et la bienveillance de tous lors des réunions faites en commun.

Merci à tous les membres de ma promotion pour ces trois années d'apprentissage et pour l'ensemble des bons moments partagés.

De manière générale, merci à tous mes proches : famille, amis et toutes les personnes qui ont pu échanger avec moi à propos de ce mémoire. Discuter avec eux dans diverses situations a toujours été enrichissant et a conforté la pertinence de mes recherches. Dans ce sens, j'adresse un merci tout particulier à Moïses.

Merci à mon frère et à tous ceux qui ont pris du temps pour relire ce mémoire!

Enfin, un grand merci à Barry Gardiner pour son expertise et sa réactivité tout au long de ce mémoire.



# Sommaire

Résumé .....	I
Abstract .....	III
Note de synthèse.....	V
Executive summary.....	XV
Remerciements .....	XXV
Sommaire.....	XXVII
Liste des abréviations .....	2
Introduction.....	3
I Des risques climatiques globaux à l'assurance forêt .....	5
I.1 Le changement climatique et ses enjeux .....	5
I.1.1 Définitions et indicateurs .....	5
I.1.2 Tendances historiques et projections.....	6
I.1.3 Enjeux macroéconomiques et impacts sociétaux .....	7
I.2 Répercussions sur le secteur de l'assurance .....	8
I.2.1 Les catastrophes naturelles et leurs impacts .....	8
I.2.2 Régime Catastrophes Naturelles : fonctionnement et panorama....	10
I.2.3 Impacts sur la tarification, les provisions et le capital économique .	11
I.2.4 Stratégies d'adaptation des assureurs.....	11
I.3 Focus sur le risque tempête : du phénomène physique à la modélisation actuarielle .....	12
I.3.1 Nature du risque tempête.....	12
I.3.2 Historique des tempêtes majeures et de leurs impacts.....	13
I.3.3 Modélisation et cartographie du risque tempête .....	15
I.3.4 Spécificités des dommages tempêtes sur les forêts .....	17
I.4 L'assurance forêt : intérêt, produits et challenges .....	17
I.4.1 Intérêt de l'assurance .....	18
I.4.2 Historique et cadre réglementaire de l'assurance forêt.....	19
I.4.3 Étude de marché et présentation de Groupama Forêts Assurances .	19
I.4.4 Enjeux actuariels.....	20
I.5 Synthèse de la première partie et réflexions conséquentes pour ce mémoire	21
II Élaboration d'une base de données adaptée à l'étude.....	22
II.1 Étude de référence : <i>Windthrow Hazard Classification</i> .....	22
II.1.1 Contexte de l'étude .....	22
II.1.2 Objectifs d'utilisation de cette étude .....	22
II.1.3 Travaux réalisés par K.F. Miller.....	23
II.1.4 Enseignements clés et variables retenues pour la modélisation.....	26
II.2 Objectifs et périmètre adopté.....	27
II.2.1 Besoins analytiques .....	27
II.2.2 Démarche adoptée.....	27
II.3 Base de données historique.....	29
II.3.1 Présentation de la base .....	29
II.3.2 Analyse des données .....	31

II.3.3	Travaux réalisés sur la base de données.....	32
II.3.4	Base de primes historique.....	34
II.3.5	Agrégation avec la base de données communale .....	35
II.4	Base de données par commune .....	35
II.4.1	Données relatives à l'altitude.....	35
II.4.2	Données relatives au sol .....	37
II.4.3	Données relatives à l'abri topographique .....	38
II.4.4	Données relatives à l'essence de l'arbre .....	40
II.4.5	Données relatives au vent.....	42
II.4.6	Prise de recul sur les données alternatives susceptibles d'enrichir l'étude.....	43
II.5	Synthèse de la deuxième partie et agrégation de cette base de données ....	44
III	Modélisation du risque et construction d'un zonier à la maille communale .....	45
III.1	Fonction et intérêt des zoniers .....	45
III.1.1	Définition et avantages.....	45
III.1.2	Zonier administratif.....	46
III.1.3	Élaboration d'un premier zonier par le calcul d'un score.....	46
III.2	Modélisation du risque par Modèles Linéaires Généralisés (MLG).....	48
III.2.1	Introduction aux MLG.....	48
III.2.2	Gestion en cas de données déséquilibrées.....	50
III.2.3	Premiers essais et MLG quasi-binomial.....	52
III.2.4	<i>Backtesting</i> et validation des modèles .....	58
III.3	Modélisation du risque par Modèles <i>machine learning</i> .....	59
III.3.1	Introduction aux modèles <i>machine learning</i> .....	59
III.3.2	<i>Random Forest</i> .....	60
III.3.3	<i>XGBoost</i> .....	62
III.3.4	<i>Backtesting</i> .....	65
III.3.5	Comparaison et choix du meilleur modèle.....	66
III.4	Production du zonier .....	67
III.4.1	Cartographie.....	67
III.4.2	Mise en place opérationnelle.....	69
III.5	Synthèse de la troisième partie et utilisation future du zonier .....	69
IV	Recalibration de la tarification du risque tempête .....	71
IV.1	Mise en place stratégique et technique .....	71
IV.1.1	Retour sur la tarification en assurance forêt .....	71
IV.1.2	Objectifs tarifaires et commerciaux .....	71
IV.1.3	Ancienne tarification .....	72
IV.1.4	Utilisation du zonier communal .....	73
IV.2	Travaux sur la période de retour des sinistres catastrophe .....	73
IV.2.1	Impact de la tempête Klaus et hypothèses .....	73
IV.2.2	Introduction à la théorie des valeurs extrêmes .....	74
IV.2.3	Analyse univariée pour une station .....	75
IV.2.4	Introduction à la théorie des copules .....	78
IV.2.5	Analyse multivariée via copules pour plusieurs stations .....	79
IV.3	Impact de la nouvelle tarification.....	82
IV.3.1	Effets statistiques et sur les ratios .....	82
IV.3.2	Impact commercial et stratégique.....	85

IV.4	Synthèse de la quatrième partie et ouverture.....	86
V	Conclusion .....	87
V.1	Synthèse des résultats .....	87
V.2	Mise en place opérationnelle, limites et recommandations.....	88
	Bibliographie.....	90
A	Différents types de sols au Royaume-Uni et en France.....	94
B	Paramètres inclus dans les fichiers de données SYNOP.....	95
C	Classification conifères/feuillus en fonction de la nature de l'arbre de la base de données de sinistre.....	96
D	Vents maximaux moyens par département.....	97
E	Statistiques descriptives générales des données.....	98
F	Illustrations des types de sol et des types d'arbres.....	100
G	Choix des hyperparamètres et mise en œuvre du modèle <i>Random Forest</i> .....	103
H	Implémentation de l'algorithme d'optimisation de Nelder-Mead .....	106
	Table des figures.....	108
	Liste des tables .....	110

# Liste des abréviations

<b>ABI</b>	Association of British Insurers	<b>INRAE</b>	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
<b>AIC</b>	Critère d'information d'Akaike		
<b>ANOVA</b>	Analyse de la Variance	<b>INSEE</b>	Institut National de la Statistique et des Études Économiques
<b>AUC</b>	Aire sous la Courbe	<b>IORP</b>	Institution de Retraite Professionnelle
<b>BFGS</b>	Broyden–Fletcher–Goldfarb–Shanno	<b>LOCDDR</b>	Louisiana Office of Community Development – Disaster Recovery
<b>BIC</b>	Critère d'information Bayésien		
<b>CART</b>	Arbres de Classification et de Régression	<b>MAE</b>	Erreur Absolue Moyenne
<b>Cat-Nat</b>	Catastrophe Naturelle	<b>MISSE</b>	Mutuelle Initiée par les Sylviculteurs du Sud-Ouest
<b>CCR</b>	Caisse Centrale de Réassurance	<b>MLG</b>	Modèle Linéaire Généralisé
<b>CIFA</b>	Compte d'Investissement Forestier et d'Assurance	<b>MSE</b>	Erreur Quadratique Moyenne
<b>CO<sub>2</sub></b>	Dioxyde de Carbone	<b>OOB</b>	Out Of Bag
<b>EIOPA</b>	European Insurance and Occupational Pensions Authority	<b>POT</b>	Peaks Over Threshold
<b>GEV</b>	Generalized Extreme Value Distribution	<b>PRC</b>	Courbe Précision–Rappel
<b>GIEC</b>	Groupe d'Experts Intergouvernemental sur l'évolution du Climat	<b>Q–Q Plot</b>	Quantile–Quantile Plot
<b>GPD</b>	Generalized Pareto Distribution	<b>RMSE</b>	Écart Quadratique Moyen
<b>GVIF</b>	Indices de Variance d'Inflation Généralisés	<b>ROC</b>	Receiver Operating Characteristics
<b>IEFC</b>	Institut Européen de la Forêt Cultivée	<b>S/P</b>	Sinistres / Primes
<b>IFRS</b>	International Financial Reporting Standards	<b>SANN</b>	Simulated Annealing
<b>IGN</b>	Institut National de l'Information Géographique et Forestière	<b>SIG</b>	Système d'Information Géographique
		<b>SSP</b>	Shared Socio-economic Pathways
		<b>TOPEX</b>	Abri Topographique
		<b>XGBoost</b>	Extreme Gradient Boosting

# Introduction

Les catastrophes naturelles, dont la fréquence et l'intensité s'accroissent avec le changement climatique, constituent aujourd'hui l'un des défis les plus pressants de notre époque. Leur caractère imprévisible et leur portée mondiale en font une menace omniprésente, tandis que leurs conséquences s'étendent désormais à des secteurs qui n'étaient pas vulnérables il y a encore quelques décennies.

Au cœur de toutes ces problématiques et afin de répondre au besoin de protection de ces différents secteurs, le monde de l'assurance a pour mission de s'adapter constamment à l'environnement dans lequel il se trouve. Néanmoins, les tempêtes, tsunamis et autres séismes produisent des dégâts matériels et financiers qui se répètent, en témoignent les plus de 80 000 sinistres Cat-Nat indemnisés en 2023 en France (CCR [2024]). L'actuaire peut donc utiliser cette sinistralité historique pour modéliser et estimer les évolutions futures. Il doit toutefois rester vigilant face aux changements rapides dus au dérèglement climatique ainsi qu'aux spécificités du domaine étudié, afin d'éviter des travaux inadaptés.

Afin de contrôler cela, l'actualité assurantielle récente montre que différents éléments réglementaires sont mis en place pour intégrer les risques climatiques et de durabilité (AMOUCHE ET THILL [2024]). L'EIOPA, en coordination avec d'autres entités, a publié en 2024 un scénario climatique stressé inédit dans le cadre de l'initiative *Fit for 55* (EIOPA [2024]). Cette initiative européenne vise à réduire de 55 % les émissions de gaz à effet de serre d'ici 2030 et à atteindre la neutralité carbone d'ici 2050. D'autres réformes se mettent en place, comme la révision de la formule standard de la directive Solvabilité II prévue en 2026, le règlement SFDR à l'échelle européenne ou l'article 29 de la Loi Énergie-Climat à l'échelle française. Ce mémoire s'inscrit donc dans une période où la question climatique est au cœur des réflexions du domaine actuariel.

Le secteur forestier est bien entendu lui aussi soumis à ces risques Cat-Nat. Toutefois, l'assurance des biens forestiers, ou assurance forêt, n'est utilisée que par 4 % des propriétaires privés de forêts (BRUNETTE ET HANEWINKEL [2023]). Il n'en demeure pas moins primordial pour les acteurs de ce marché d'être à la pointe des innovations afin de proposer notamment à leurs clients un tarif qui soit juste, et donc basé sur des justifications mathématiques et actuarielles solides et recevables. Une tarification plus fine permettrait non seulement de réduire les primes pour les zones peu exposées, mais aussi d'inciter davantage de propriétaires à souscrire une assurance.

Groupama Forêts Assurances, acteur majeur du marché de l'assurance forêt en France, a mandaté le cabinet Actuelia en fin d'année 2024 pour examiner et ajuster ses pratiques tarifaires relatives à la garantie tempête. À cette fin, ils ont fourni leurs bases de données historiques. Pour des raisons de confidentialité, ces données sont modifiées dans ce mémoire. Les valeurs retenues restent toutefois représentatives, fondées sur des ordres de grandeur fiables et directement issus du marché étudié.

Ce mémoire s'inscrit au cœur d'une problématique concrète d'entreprise : développer l'assurance forêt en France, comme cela a pu être fait dans d'autres pays (95 % de couverture forestière en Suède par exemple (BRUNETTE ET HANEWINKEL [2023])). Il poursuit plusieurs objectifs : identifier les spécificités et défis propres à ce marché, construire une base de données claire et

cohérente, et apporter une véritable valeur ajoutée à Groupama Forêts Assurances comme au secteur, en modélisant actuariellement le risque tempête. Il vise ainsi à repérer des variables explicatives pertinentes, au-delà de la seule vitesse du vent, qui s'était révélée insuffisante en elle-même lors d'études préalables internes au cabinet. Le modèle est appliqué à l'échelle communale en France métropolitaine afin de définir des classes de risque et d'étayer une nouvelle tarification. L'ambition est donc d'aboutir à des résultats solides grâce à une démarche originale.

Cette démarche pourrait se résumer dans la problématique suivante : Comment la création d'un zonier à la maille communale, issu de la modélisation du risque tempête en assurance forêt, permet de fiabiliser les tarifs proposés par les assureurs de ce domaine ?

La première partie a pour but de poser le contexte du sujet. Elle débute par une approche globale en rappelant ce qu'est le réchauffement climatique et quels sont ses impacts chiffrés. Puis, progressivement, le sujet se rapporte au domaine de l'assurance, au risque tempête et à l'impact de ce dernier sur les forêts. Chaque section de cette partie est nécessaire pour comprendre la suivante et pour pouvoir, ainsi, aborder la suite de ce mémoire avec tous les éléments nécessaires à sa bonne compréhension.

La deuxième partie met en avant l'étude britannique *Windthrow Hazard Classification*, qui a inspiré les premiers modèles de ce mémoire. Cette étude examine le risque de déracinement des arbres en se basant sur diverses variables explicatives. La suite de la partie traite de la collecte des données nécessaires pour les modélisations qui suivent.

La troisième partie vise à modéliser le risque de déracinement des arbres sous l'effet des tempêtes à partir des variables explicatives précédemment obtenues. Plusieurs approches (score de risque, modèles linéaires généralisés, *machine learning*) sont estimées et comparées afin de retenir le modèle offrant le meilleur compromis entre capacité explicative et prédictive. Le modèle sélectionné sert ensuite de base à la classification des communes en niveaux de risque et à la construction du zonier.

La quatrième partie applique les résultats des sections précédentes à la refonte de la tarification du risque tempête, à partir des données modifiées de Groupama Forêts Assurances. La refonte s'appuie sur le zonier communal créé, mais aussi sur l'utilisation d'outils mathématiques avancés, tels que les copules ou la théorie des valeurs extrêmes, afin de définir les risques auxquels est confronté l'assureur.

Ces différents travaux répondent à un enjeu fort : parvenir à une tarification fine d'un risque assez peu assuré en France métropolitaine.

# I - Des risques climatiques globaux à l'assurance forêt

## I.1 Le changement climatique et ses enjeux

### I.1.1 Définitions et indicateurs

Le climat est une notion prépondérante dans les enjeux sociétaux, géographiques et assurantiels actuels. À la différence de la météorologie, qui est décrite à un instant donné, le climat s'applique sur un temps long. On y étudie la moyenne des conditions météorologiques observées pour une région donnée sur une période d'au moins 30 ans. Des conclusions générales peuvent être tirées de l'étude du climat de certaines zones géographiques. Les actualités de ces dernières décennies, telles que la fonte des glaces ou la hausse des températures, remettent en question certains acquis et amènent de nouvelles problématiques par rapport aux évolutions climatiques.

À ce titre, le GIEC, organisme intergouvernemental de référence sur l'ensemble de ces enjeux, définit le changement climatique comme « tout changement de climat dans le temps, qu'il soit dû à la variabilité naturelle ou aux activités humaines » (GIEC [2023]). De manière générale, parler de dérèglement ou de changement climatique apporte nuance et précision par rapport au terme de réchauffement climatique, car certaines régions peuvent connaître des saisons froides plus marquées et d'autres régions voient leurs cycles de précipitations grandement modifiés.

De nombreux indicateurs, dont certains sont présentés dans la table I.1, permettent de quantifier et d'analyser les évolutions climatiques récentes. Ces indicateurs constituent la base de toute analyse de risque climatique.

Indicateur	Rôle
Température	Indicateur de référence de l'évolution climatique.
Cumul des émissions de CO <sub>2</sub>	Indicateur de calibrage des scénarios climatiques et du risque de transition.
Étendue de la banquise	Indicateur du réchauffement global.
Précipitations moyennes	Indicateur pour les garanties inondation et submersion.
Déracinement des arbres	Indicateur central du mémoire.

TABLE I.1 – Indicateurs climatiques : Usages actuariels

La table met en lumière l'utilité actuarielle que peut avoir tout indicateur biophysique. À ce titre, ces différents éléments peuvent être étudiés séparément puis projetés dans des scénarios de *stress* (à savoir des scénarios, parfois extrêmes), avant de tenter de trouver des analyses communes entre ces indicateurs à court, moyen ou long terme. Le GIEC, par exemple, décline ses projections climatiques en différents SSP, qui sont des scénarios de *stress* qui décrivent l'évolution de la Terre face aux choix possibles de développement des êtres humains. Parmi eux se trouve le SSP1-2.6, qui donne une vision optimiste tandis que le scénario intermédiaire SSP2-4.5 prolonge les politiques actuelles et que le scénario extrême SSP5-8.5 place la Terre dans le cas d'une dépendance extrême aux énergies fossiles. Ces différents *stress* permettent de

fournir une vision globale du risque dans le but d'éveiller une prise de conscience collective et de fournir des éléments de solutions.

## I.1.2 Tendances historiques et projections

D'après le plus récent rapport d'expertise du GIEC, intitulé AR6 et publié en 2023 (GIEC [2023]), il est établi que le réchauffement du climat à l'échelle mondiale dû aux activités humaines est un fait. La dernière décennie complète est à ce titre la plus chaude jamais enregistrée depuis plus de 125 000 ans. De plus, en 2019, la concentration de CO<sub>2</sub> dans l'atmosphère a atteint un taux qui n'avait pas été atteint depuis 2 millions d'années.

Les scénarios de *stress* évoqués précédemment montrent que la hausse du niveau de réchauffement global de 1,5 °C par rapport à l'époque pré-industrielle sera atteinte au début des années 2030, et ce quels que soient les efforts de réduction immédiate des émissions mondiales de CO<sub>2</sub>. Des efforts sont alors à faire au plus vite étant donné que 3,3 milliards de personnes vivent dans des zones qui sont déjà vulnérables au changement climatique.

On observe que le scénario intermédiaire, soit celui correspondant à l'impact des politiques actuelles, conduit à une stabilité des émissions de CO<sub>2</sub> (Table I.2) et à une hausse conséquente de la température (Figure I.1). Les actions actuelles ne sont donc pas suffisantes face à l'essor du dérèglement climatique.

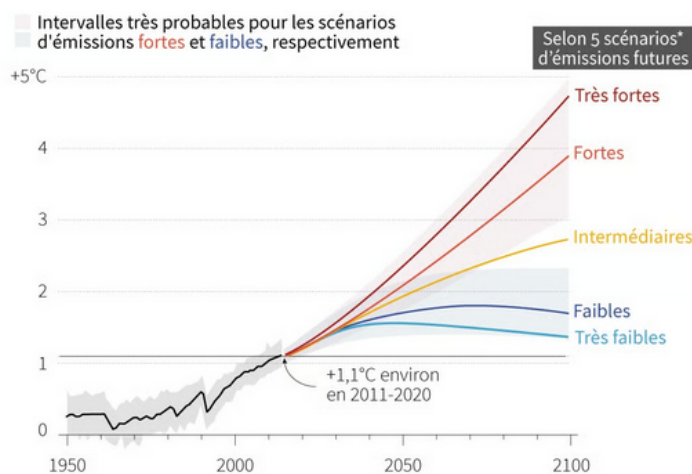


FIGURE I.1 – Température : Évolution selon les scénarios SSP (NOVETHIC [2023])

Scénario	Trajectoire d'émissions nettes de CO <sub>2</sub>	$\Delta T$ 2100 <sup>a</sup>	Événements extrêmes <sup>b</sup>
SSP1-1.9 (très bas)	Pic vers 2020 et zéro net vers 2050	1,4 °C [1,0–1,8]	×1,5
SSP1-2.6 (bas)	–50 % en 2050 et quasi-zéro en 2100	1,8 °C [1,3–2,4]	×1,7
SSP2-4.5 (médiann)	Stable avant 2050 puis –25 % en 2100	2,7 °C [2,1–3,5]	×2,2
SSP3-7.0 (élevé)	+70 % d'ici 2100	3,6 °C [2,8–4,6]	×3,0
SSP5-8.5 (très élevé)	+100 % d'ici 2050 ; +200 % en 2100	4,4 °C [3,3–5,7]	×3,8

<sup>a</sup> Réchauffement médian et intervalle de confiance entre 2081-2100 et 1850-1900.

<sup>b</sup> Multiplicateur de fréquence d'événements de précipitations extrêmes (ou chaleur extrême).

TABLE I.2 – Scénarios SSP à horizon 2100 : Synthèse (GIEC [2023])

Des engagements ont été tenus par les États mais ces derniers ne sont pas suffisants. L'écart entre les objectifs et la réalité ne ferait que s'élargir dans les décennies futures, comme l'illustre le graphique I.2, également issu du rapport du GIEC :

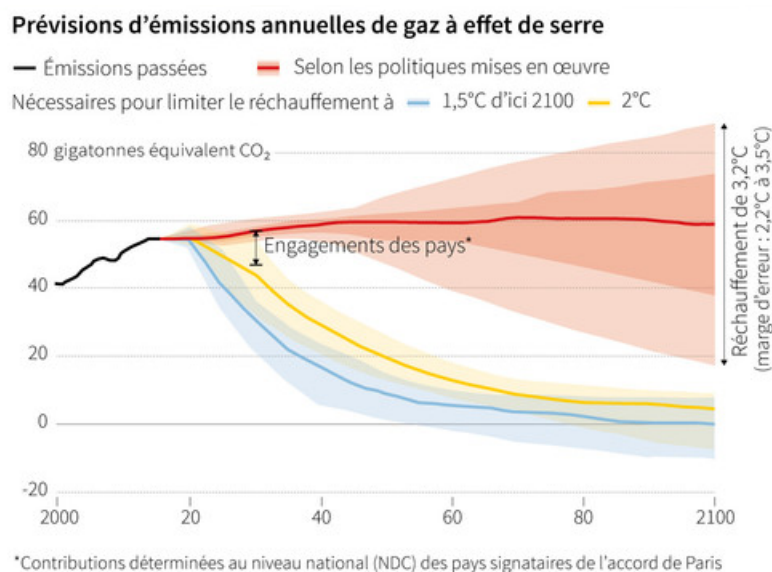


FIGURE I.2 – Objectifs d'émission : Insuffisance par rapport aux objectifs (NOVETHIC [2023])

### I.1.3 Enjeux macroéconomiques et impacts sociétaux

Le dérèglement climatique, ainsi que l'ensemble des risques qui en résultent, sont ainsi un défi majeur du monde contemporain. Leurs effets directs touchent de plus en plus de pays et de personnes à travers le monde, avec une sévérité accrue. Des impacts sont visibles sur la santé publique avec par exemple une mortalité accrue due aux canicules plus sévères. En effet, Santé Publique France a recensé deux fois plus de recours aux soins d'urgence lors des épisodes de canicule en 2023, soit 20 000 hospitalisations d'urgence. (SANTÉ PUBLIQUE FRANCE [2024])

À l'échelle de la France métropolitaine, une aggravation des écarts entre régions vulnérables et centres économiques protégés pourrait être constatée, ce qui causerait un risque d'appauvrissement rural. Une surexposition à la chaleur pourrait également réduire la force de travail via une diminution des heures travaillées ou un travail réalisé dans des conditions extrêmes. D'un point de vue économique, cette surexposition entraînerait une baisse du PIB potentiel liée aux pertes agricoles, aux dommages aux infrastructures et à la moindre productivité. Ces situations de crise augmentent le risque d'inflation et de volatilité accrue des marchés.

Plus généralement, des tensions pourraient aussi voir le jour d'un point de vue géopolitique autour de l'eau, des terres cultivables et de bien d'autres points. À titre d'exemple, l'ouragan Katrina aux États-Unis a, en 2005, causé la mort de 1 836 personnes et 141 500 sinistrés ont dû être hébergés dans des foyers d'accueil. De nombreuses terres ont été détruites dans plus de 18 états américains. Au-delà de ces éléments apparents, de nombreuses conséquences plus discrètes ont émergé : marché pétrolier américain à l'arrêt, transport commercial fluvial perturbé, tourisme brutalement interrompu, inflation massive sur les produits de première nécessité et l'électricité ou encore pillages dans les zones endommagées (LOCDDR [2005]).

Les chiffres exposés ci-dessus démontrent que le réchauffement climatique n'est pas une menace abstraite : il se traduit déjà par des événements extrêmes. Ces derniers peuvent causer des pertes humaines, une pression accrue sur les infrastructures et une hausse notable des coûts. Il est donc nécessaire de s'interroger sur la façon dont les chocs peuvent être absorbés. La section suivante se penche plus précisément sur la dimension assurantielle relative à ces problématiques afin de passer du constat scientifique et social aux enjeux actuariels, en particulier ceux liés aux forêts.

## I.2 Répercussions sur le secteur de l'assurance

### I.2.1 Les catastrophes naturelles et leurs impacts

Une catastrophe naturelle se définit comme « un événement résultant de l'activité anormale d'un agent naturel, et ce lorsque les mesures habituelles prévenant ces dommages n'ont pas réussi à les empêcher ou à les prédire correctement » (INSEE [2016]). Ces événements naturels peuvent prendre des formes diverses, ayant une probabilité d'occurrence plus ou moins élevée à travers les différentes régions du monde. Les principaux sont, entre autres :

- Les inondations
- La sécheresse
- Les tempêtes
- Les séismes
- Les tsunamis
- Les avalanches et glissements de terrain

Ces catastrophes modifient profondément le paysage assurantiel et soulèvent des difficultés auxquelles l'assureur doit faire face. L'augmentation de la fréquence et de l'intensité des sinistres, ainsi que de leurs coûts d'indemnisation s'est indéniablement accentuée au cours de la dernière décennie (Figure I.3).

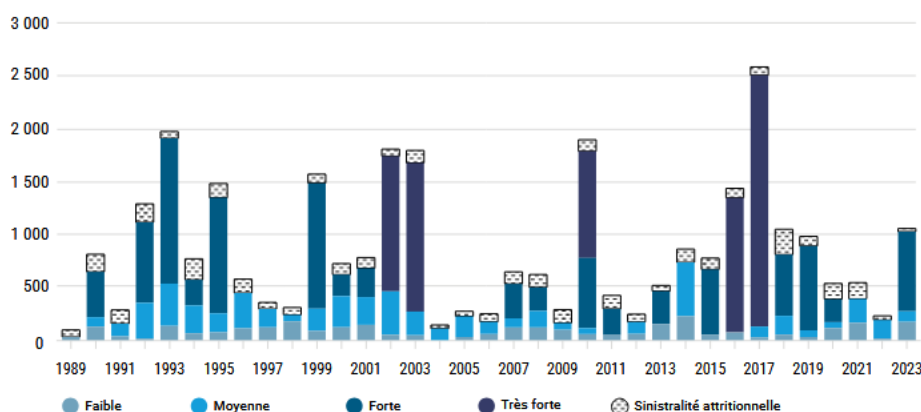


FIGURE I.3 – Sinistralité hors sécheresse par gravité : Évolution (en M €) (CCR [2024])

Ce graphique met en évidence le fait que ce sont les sinistres graves voire extrêmes qui ressortent de l'indemnisation des assureurs. De plus, le risque de dérèglement climatique est référencé, dans la cartographie prospective de l'assurance par France Assureurs (FRANCE ASSUREURS [2024]), comme celui avec le score fréquence-sévérité le plus élevé en France et les événements naturels exceptionnels se classent également parmi les cinq risques principaux. La tendance à la hausse de l'impact des risques climatiques sur le secteur de l'assurance est visible dans la figure I.4.

RANG	RISQUES	SCORE MOYEN	SCORE (Fréq.; Sév.)
1	Cyberattaques	4,1	(4,2; 4,1)
1	Dérèglement climatique	4,1	(4,0; 4,3)
3	Environnement économique	3,8	(3,7; 4,0)
4	Environnement politique*	3,6	(3,6; 3,6)
5	Événement naturel exceptionnel	3,3	(2,8; 3,9)

FIGURE I.4 – Score fréquence-sévérité : Cinq risques les plus élevés pour le secteur de l’assurance en France (FRANCE ASSUREURS [2024])

D’après la même cartographie (FRANCE ASSUREURS [2024]), le dérèglement climatique est la principale menace pour la société française du point de vue du secteur de l’assurance et de la réassurance, devant les tensions sociales et politiques ou les cyberattaques, comme le montre le graphique de la figure I.5.

Source: Cartographie 2025 de la profession de l’assurance et de la réassurance, France Assureurs

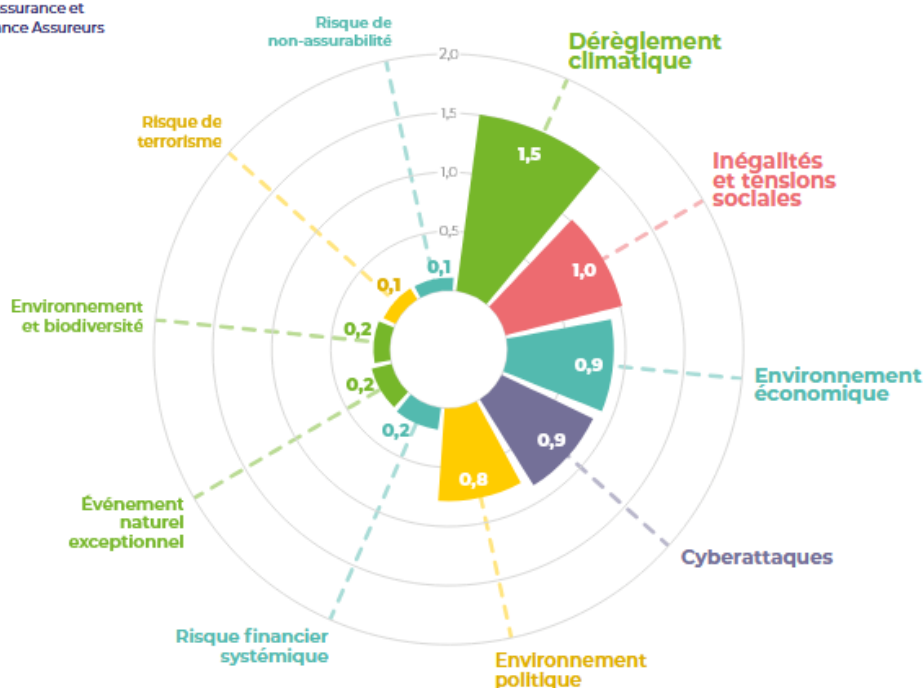


FIGURE I.5 – Menaces majeures pour la société française : Vision assurance/réassurance (FRANCE ASSUREURS [2024])

En considérant cette fois l’échelle mondiale, les inondations et les tempêtes représentent à elles deux plus de 70 % des catastrophes naturelles survenues en 2022. La tendance française se retrouve. En effet, les catastrophes naturelles ont engendré 280 milliards de dollars de dégâts en 2023 dans le monde, dont 108 milliards de dollars ont été couverts par les compagnies d’assurance, selon le réassureur Swiss Re, qui alerte sur une probable multiplication par deux d’ici 10 ans.

## I.2.2 Régime Catastrophes Naturelles : fonctionnement et panorama

La France a mis en place un dispositif spécial, appelé « régime Cat-Nat » et instauré par la Loi du 13 juillet 1982. Ce partenariat privé-public inclut la responsabilité de l'État et garantit à tous les citoyens une indemnisation adéquate en cas de sinistre causé par un phénomène naturel. Le but principal était de prendre en considération ces risques, largement sous-assurés jusqu'alors. Les objectifs plus précis sont détaillés dans la table I.3.

Objectif	Détails
Couverture étendue	Protéger tous les types de risques naturels, sauf Tempête, Grêle et Neige
Principe de solidarité	Surprimes uniformisées : 12 % sur les biens (hors véhicules), 6 % sur les garanties vol/incendie
Coopération État-Assureurs	Partenariat public-privé efficace
Pérennité du système	Système solvable et durable sur le long terme

TABLE I.3 – Régime Cat-Nat : Objectifs et mécanismes associés

L'indemnisation se déroule selon le schéma suivant :

### Protection des assurés en cas de phénomènes naturels reconnus comme catastrophes naturelles

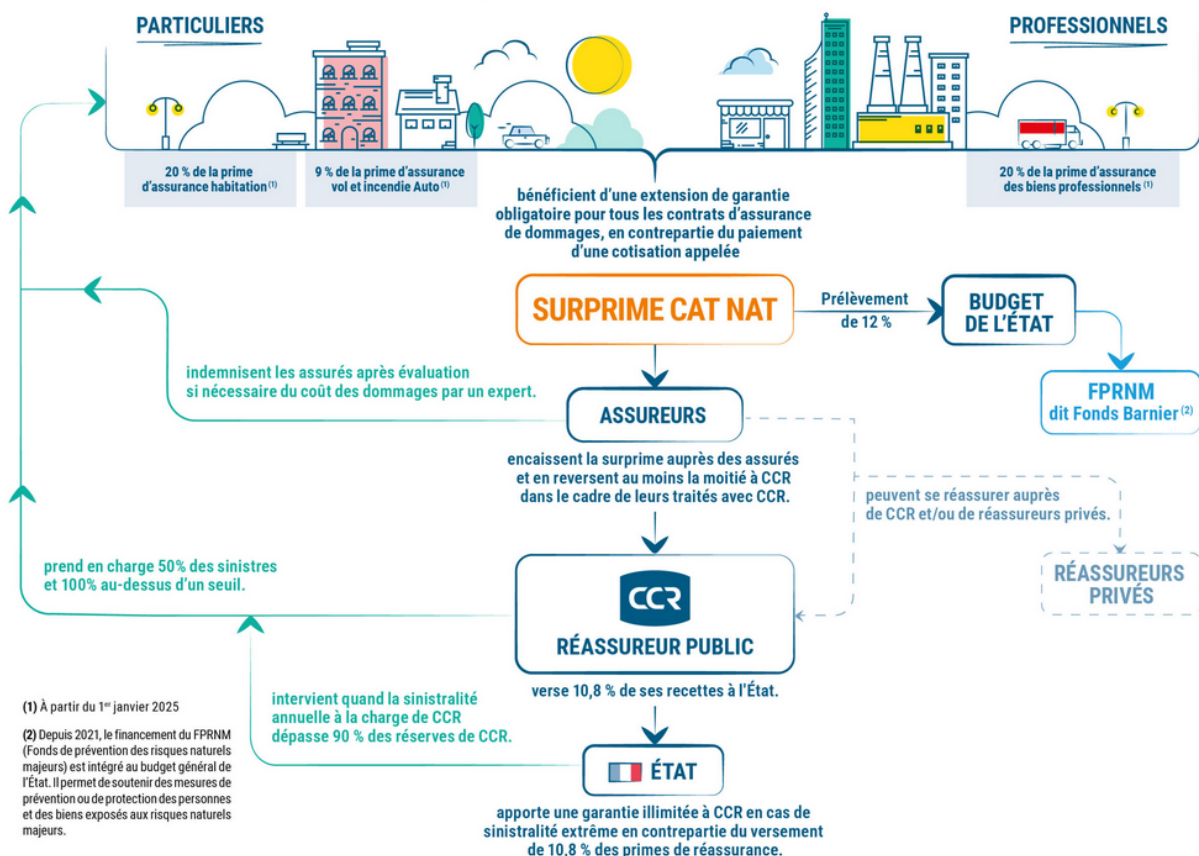


FIGURE I.6 – Catastrophes naturelles en France : Schéma d'indemnisation (CCR [2024])

Ce schéma permet de bien mettre en avant le partenariat entre assureurs privés et réassureur public. De manière plus détaillée, la garantie Cat-Nat s'applique après reconnaissance par

arrêté interministériel, pris à la suite d'une demande de la commune et publié au *Journal officiel*. Le dispositif de réassurance de la CCR repose sur une double couverture : une première proportionnelle en quote-part, où les assureurs cèdent 50 % des primes à la CCR en échange d'une prise en charge équivalente des sinistres, et une seconde non proportionnelle sur rétention, qui intervient lorsque la sinistralité dépasse un seuil prédéfini. En moyenne, entre 1982 et 2023, la CCR a couvert 51 % des sinistres Cat-Nat. L'État garantit également la CCR en cas de sinistralité exceptionnelle, avec un seuil d'intervention fixé à 90 % de ses réserves. Cette garantie n'a été activée qu'une seule fois. Ce mécanisme couvre aujourd'hui environ 95 % du marché et fait l'objet de révisions régulières pour rester adapté aux enjeux actuels (CCR [2024]).

### **I.2.3 Impacts sur la tarification, les provisions et le capital économique**

Du fait de la hausse massive des catastrophes naturelles lors des dernières décennies, l'historique de données utilisé pour établir la tarification ou les scénarios devient de moins en moins représentatif de la réalité et pourrait ainsi fournir des résultats biaisés. Un recalibrage des coefficients de tarification va entraîner la hausse des coûts et des tarifs, ce qui est attendu dans l'activité d'assurance. Cependant, il est difficile de le quantifier précisément.

Pour se prémunir d'une baisse du ratio de solvabilité, les organismes d'assurance ont intérêt à revoir à la hausse les primes afin que celles-ci transcrivent au mieux le risque lié aux catastrophes naturelles. Cela revient à une hausse des exigences réglementaires en matière de solvabilité. Ce besoin d'actualisation continue pousse également les assureurs à réévaluer régulièrement les paramètres de fréquence et de sévérité dans leurs modèles.

Les provisions constituées par les assureurs sont également impactées par la hausse des catastrophes naturelles. Le caractère incertain et potentiellement extrême des sinistres mène à des choix prudents. Les provisions sont ainsi calculées en prenant en considération une forme de prudence, afin de rester conforme aux réglementations.

Par ailleurs, plusieurs travaux de *stress tests* climatiques ont été menés au niveau européen, notamment par l'EIOPA. Ceux-ci intègrent des trajectoires SSP pour projeter la fréquence et la gravité des événements extrêmes sur des portefeuilles d'assureurs. Ces *stress tests* servent de base pour tester la solidité des assureurs et recalibrer les exigences en capital économique (SCR) et les politiques tarifaires en fonction des décisions climatiques prises.

### **I.2.4 Stratégies d'adaptation des assureurs**

Avec la fréquence et l'intensité des sinistres Cat-Nat qui augmentent, les assureurs s'interrogent sur la couverture de certains risques. Les retraits envisagés peuvent être issus de questions géographiques, comme lorsque Farmers Insurance s'est retiré du marché de la Floride en 2023 du fait de la surexposition aux ouragans et aux tempêtes. Ces retraits peuvent aussi être la conséquence de garanties non suffisantes de la part de l'assuré. Un exemple est la situation où Generali refuse d'indemniser un foyer en cas de reconnaissance de catastrophe naturelle si les mesures fixées par le PPRI (Plan de Prévention des Risques d'Inondations) ne sont pas respectées ou si l'habitation se situe sur une zone dite « inconstructible ». L'assureur estime alors que le risque est quasiment certain. À l'avenir, de nouveaux acteurs, parmi lesquels les États, pourraient intervenir et s'engager dans l'indemnisation des sinistres de catastrophes naturelles.

Il est important de ne pas considérer uniquement les faits extrêmes concrets, telles que les tempêtes ou les inondations. Les phénomènes beaucoup plus lents pouvant atteindre leur seuil de rupture et entraîner des dégâts conséquents, comme l'érosion des littoraux, doivent être pris en compte. Il est ainsi nécessaire de pratiquer des mesures de prévention et de sensibilisation plus accrues.

De plus, pour faire face à l'éclosion des risques climatiques, de nouveaux produits assurantiels ont vu le jour : c'est le cas de l'assurance paramétrique. Ce produit novateur en assurance repose sur le dépassement d'un ou plusieurs indices définis en amont tels que la vitesse du vent ou le cumul de précipitations. Contrairement à l'assurance « classique », elle ne nécessite pas d'expertise de terrain et se base uniquement sur des grilles de seuils, ce qui permet un versement des prestations plus rapide. Ce type de produit est particulièrement adapté aux événements climatiques extrêmes, souvent difficilement assurables. Il est de plus en plus utilisé dans les secteurs agricoles et forestiers. Cela en fait une solution intéressante dans un contexte de dérèglement climatique croissant. Le marché mondial de l'assurance paramétrique a été évalué à 16,2 milliards de dollars en 2024 et devrait augmenter de plus de 10 % dans les dix prochaines années (GLOBAL MARKET INSIGHTS [2025]).

## **I.3 Focus sur le risque tempête : du phénomène physique à la modélisation actuarielle**

### **I.3.1 Nature du risque tempête**

Les tempêtes sont qualifiées comme telles dès lors que le vent moyen atteint ou dépasse 89 km/h sur une période d'au moins 10 minutes (GÉORISQUES [2025]). Ce phénomène météorologique se produit lorsque des courants d'air chaud provenant de la mer rencontrent des courants d'air froid issus de la terre. Le choc entre ces masses d'air, aux teneurs en eau différentes, engendre des vents violents. Parmi les dégâts causés figurent les habitations endommagées, les arbres arrachés, ou encore les coupures de courant et de réseaux. Les tempêtes étant généralement accompagnées de fortes précipitations, des inondations (crue ou ruissellement) sont également une conséquence possible de ces phénomènes. Lors du passage d'une tempête, les dégâts peuvent s'observer sur une zone très vaste (la zone inondée en 2010 après la tempête Xynthia s'étend sur près de 500 km<sup>2</sup> dans les départements de Charente-Maritime et de Vendée).

Pour caractériser les tempêtes, l'outil généralement utilisé est l'échelle de Beaufort, dont les seuils et la description sont indiqués dans la table I.4. Le seuil de 89 km/h retenu pour qualifier une tempête y est retrouvé.

Force	Description	Vitesse (km/h)
0	Calme	< 1
1	Très légère brise	1–5
2	Légère brise	6–11
3	Petite brise	12–19
4	Jolie brise	20–27
5	Bonne brise	28–38
6	Vent frais	39–49
7	Grand frais	50–61
8	Coup de vent	62–74
9	Fort coup de vent	75–88
10	Tempête	89–102
11	Violente tempête	103–120
12	Ouragan	> 120

TABLE I.4 – Échelle de Beaufort : Correspondance force du vent / vitesse

Le vent moyen, calculé en moyennant la vitesse du vent sur plusieurs minutes (référence de l'échelle de Beaufort) doit être distingué des rafales, qui ne sont que des pointes de vitesse durant quelques secondes.

### I.3.2 Historique des tempêtes majeures et de leurs impacts

Au niveau international, les zones les plus exposées aux tempêtes sont les côtes Atlantique et Pacifique Nord (États-Unis et Mexique) ainsi que le Pacifique Sud (Philippines et Indonésie).

La tempête Katrina a été évoquée dans la partie précédente. La table I.5 présente d'autres tempêtes parmi les plus dévastatrices au cours du dernier siècle, ainsi que leurs dates, leurs localisations et certaines de leurs caractéristiques.

Nom	Localisation	Date	Caractéristiques
Bhola	Bangladesh	nov. 1970	Le plus meurtrier : environ 500 000 morts dus à des infrastructures peu adaptées.
Andrew	États-Unis (Sud-est), Bahamas	août 1992	Vents supérieurs à 240 km/h. Coûts estimés à 26,5 milliards de dollars.
Nargis	Myanmar	mai 2008	Submersions agricoles. 138 000 morts, 2,4 millions de sinistrés.
Sandy	États-Unis (Côte Est), Caraïbes	oct. 2012	Destructions étendues. Large zone touchée, surnommée <i>Superstorm</i> .
Haiyan	Philippines	nov. 2013	Vents continus à 315 km/h, pointes à 380 km/h. Plus de 4 millions de déplacés.
Ian	Floride, Cuba, Carolines	sept. 2022	Vents jusqu'à 240 km/h. Inondations majeures, plus de 100 morts et 100 milliards de dollars de dégâts.

TABLE I.5 – Tempêtes majeures mondiales : Caractéristiques

À l'échelle nationale, la France a également été touchée par des tempêtes. L'échelle est certes moindre, mais les dégâts restent bien réels. Les sinistres les plus marquants sont résumés dans le tableau suivant :

Nom	Localisation	Date	Caractéristiques
Lothar	Nord de la France	25–26 déc. 1999	Rafales de 150 à 170 km/h sur une large zone nord.
Martin	Centre-sud et sud-est	27–28 déc. 1999	Zone moins étendue que Lothar mais rafales jusqu'à 200 km/h.
Klaus	Sud-ouest	24 janv. 2009	Vents proches de 200 km/h sur une zone étendue (dont fait partie le Bassin aquitain).
Xynthia	Ouest	27–28 févr. 2010	Meurtrière sur les littoraux.
Ciaran	Nord-ouest	1–2 nov. 2023	Submersions en zones côtières basses et exposées.

TABLE I.6 – Tempêtes majeures en France métropolitaine : Caractéristiques

Comme l'indiquent les tables I.5 et I.6, la France ne figure pas parmi les pays historiquement les plus exposés aux tempêtes. Pour autant, les dommages y sont bien réels. Il paraît essentiel que la prévention et la couverture assurantielle des risques y revêtent une importance notable, afin de transformer autant que possible des chocs rares et massifs en charges prévisibles et finançables.

Afin de motiver une répartition du risque à une maille fine, la figure I.7 synthétise quarante tempêtes majeures depuis 1980 et la surface nationale exposée à différents seuils de rafales.

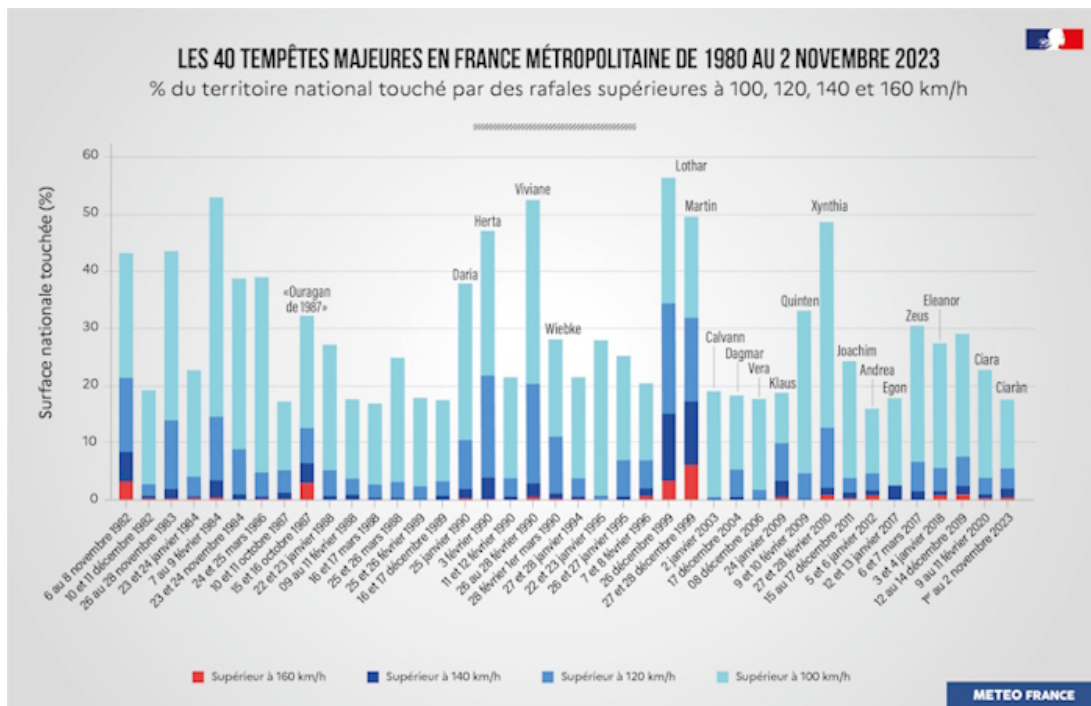


FIGURE I.7 – Tempêtes en France métropolitaine : Quarante événements majeurs (MÉTÉO-FRANCE [2024])

Cette variabilité d'étendue et d'intensité justifie une approche spatialisée de la tarification. Les tempêtes les plus puissantes peuvent entraîner des vents forts sur plus de la moitié du territoire français. L'étude de ces événements extrêmes permet de mieux prévoir et d'interpréter les évolutions possibles.

### I.3.3 Modélisation et cartographie du risque tempête

Pour prédire et interpréter le risque tempête, il est nécessaire de s'appuyer sur des données robustes. En France métropolitaine, les données fiables les plus régulières (pointages toutes les trois heures) sont celles des stations météorologiques de Météo-France. C'est la base de données SYNOP. Entre autres, il est possible d'obtenir la date d'observation, la température en °C, l'humidité en % ou la vitesse du vent et des rafales en km/h. La base SYNOP englobe 42 stations en France métropolitaine, réparties uniformément sur l'ensemble du territoire.

Il existe bien plus de 42 stations Météo-France en France métropolitaine, mais sont considérées ici uniquement des stations dont toutes les informations de l'**Annexe B** (rafales, température, points de rosée, etc.) sont disponibles quotidiennement.

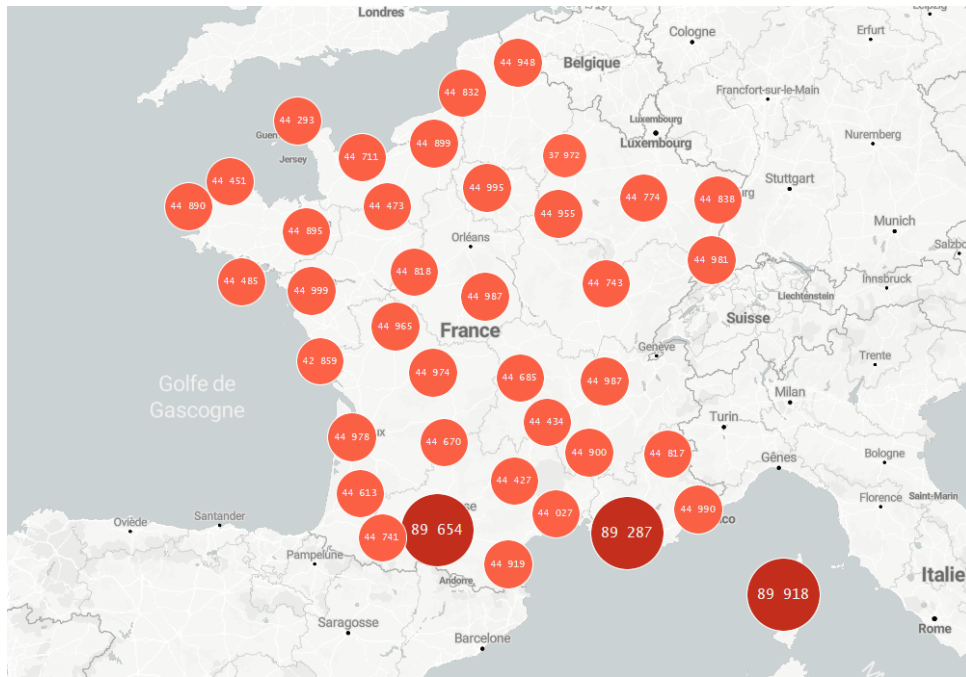


FIGURE I.8 – Stations météorologiques en France métropolitaine : Implantation géographique (OPENDATA SOFT [2024])

Sur la figure I.8, les points oranges indiquent qu’une seule base est présente dans la zone géographique concernée. Cet indicateur se porte à deux bases lorsqu’un point est rouge foncé. Le nombre inscrit dans le cercle précise les données disponibles. Si ces données sont très fiables, elles présentent néanmoins une limite majeure dans le cas de la création d’un zonier communal. En effet, les données sont collectées par station météorologique, et la France métropolitaine ne compte que 42 stations à relevés quotidiens précis pour 97 départements. Pour pallier ce manque de données, des regroupements sont effectués, ce qui induit inévitablement un manque de précision dans l’analyse qui en découle. De plus, certaines données sont manquantes dans la base étudiée. La table I.7 recense le pourcentage de valeurs manquantes dans la base de données SYNOP en 2023.

Paramètre	% de valeurs manquantes
Température	1,0
Humidité	1,4
Pression	1,6
Précipitation	4,4
Vitesse du vent	0,7
Vitesse des rafales	18,6

TABLE I.7 – Données SYNOP : Pourcentages de valeurs manquantes (FOMEN [2024])

C’est principalement pour ces deux raisons que la base SYNOP ne sera pas utilisée dans la modélisation du risque tempête, ce qui est fait habituellement. La modélisation fera ainsi intervenir davantage de variables explicatives, comme nous pourrons le développer dans les prochaines parties.

La valeur élevée pour la vitesse des rafales dans la table I.7 est due au fait que cette donnée n'est enregistrée que depuis novembre 2003, alors que la base débute en janvier 1996. Cette variable est utilisée plus tard pour le calcul de la période de retour, qui est un calcul qui nécessite un historique régulier dans le temps. Malgré l'absence de 18,6 % des données liées aux sept années non documentées, l'historique est régulier depuis novembre 2003 donc cela est suffisant pour cette étude.

Depuis plusieurs décennies, la modélisation actuarielle s'appuie sur une assise théorique plus solide. La validation des modèles repose sur des critères comme l'AUC, le  $F_1$ -score ou encore un *backtesting* important. Les assureurs parviennent à mieux évaluer le risque tempête grâce à des données et des méthodes d'évaluation plus précises. Il est alors possible de dresser des cartographies de vulnérabilité, d'établir des primes ajustées au niveau de risque, ou encore des produits mieux adaptés aux diverses zones géographiques. L'objectif recherché est l'amélioration de la couverture des assurés face aux tempêtes, garantissant la pérennité des systèmes de couverture, malgré les enjeux posés par le réchauffement climatique évoqués lors de la sous-partie précédente

### **I.3.4 Spécificités des dommages tempêtes sur les forêts**

Jusqu'ici, l'analyse a porté sur une approche générale du risque tempête. Ainsi a été évoqué l'impact du vent et de sa puissance, couplé avec certains phénomènes pouvant entraîner des inondations ou des dégâts agricoles. Pour l'assurance forêt, le problème doit être vu sous un angle différent. Outre la vitesse du vent, il est légitime de penser que d'autres variables expliquent le déracinement.

En regardant plus précisément les zones les plus susceptibles d'être touchées par le déracinement des arbres en France, on note que les régions les plus soumises aux risques, qui sont la basse vallée du Rhône (Cévennes), le Bassin aquitain ou encore le nord-ouest de la France, présentent des caractéristiques très différentes, que ce soit en termes d'exposition ou d'essence d'arbre majoritaire. De plus, l'hypothèse d'une intensité accrue des tempêtes en haute altitude doit être nuancée au regard de l'adaptation des essences concernées par ces régions, généralement des mélèzes et des pins cembro. En effet, ces espèces sont adaptées aux conditions rudes voire extrêmes. Leurs racines sont profondes et favorisent le non-déracinement.

La suite du mémoire a pour but d'évaluer la pertinence de plusieurs facteurs explicatifs du risque tempête en assurance forêt.

## **I.4 L'assurance forêt : intérêt, produits et challenges**

Dans la suite du mémoire, l'« assurance forêt » désigne une assurance de dommages aux biens : elle garantit la valeur des peuplements et infrastructures forestières contre les risques naturels assurables tels que l'incendie et les tempêtes, conformément aux articles du Titre II du Livre I du Code des assurances.

### I.4.1 Intérêt de l'assurance

Aujourd'hui, en France, 75 % des forêts sont privées. Le quart restant est public et se répartit principalement entre forêts domaniales et forêts communales. La gestion de toutes les forêts françaises est encadrée par le Code forestier. La part de forêts privées est nettement plus élevée dans l'ouest de la France. Elle dépasse 90 % pour les régions Pays de la Loire, Nouvelle Aquitaine et Bretagne. La région Grand Est est la seule où la forêt privée est minoritaire (45 %). La figure I.9 confirme ces données en présentant le pourcentage de la surface forestière publique par département.

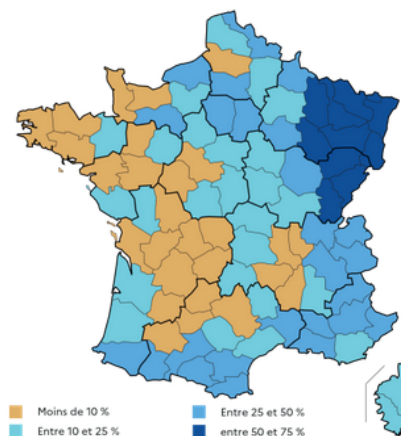


FIGURE I.9 – Pourcentage de la surface forestière publique par département : Carte (IGN [2023])

La forêt privée française est très segmentée. Avec 3,3 millions de propriétaires en 2024, ceux-ci possèdent généralement de petites parcelles. Les propriétaires sont majoritairement retraités, achetant tardivement et ayant du temps pour s'en occuper. Ce sont également des biens qui se morcellent beaucoup durant l'héritage.

Les motivations pour acheter une forêt peuvent être multiples :

- Avoir un endroit calme pour soi
- Entretenir un patrimoine familial
- Intérêt financier en louant sa forêt pour de la chasse par exemple
- Placement sur le long terme
- Contribuer à la préservation de l'environnement

Il paraît alors intéressant d'assurer sa parcelle forestière afin d'anticiper le coût des dégâts au vu de l'augmentation des risques climatiques, comme expliqué précédemment. Une propriété forestière est exposée à des risques climatiques tels que la chute d'un arbre en cas de tempête ou encore un incendie d'origine naturelle (foudre) ou humaine (mégot de cigarette jeté par terre). Un propriétaire d'une forêt, en tant que responsable des dommages physiques qui peuvent être causés à un tiers par un de ces arbres, est dans l'obligation de posséder une assurance Responsabilité Civile (FAVRE [2023]).

Frédéric Duboscq, président de Groupama Forêts Assurances, donne l'enjeu d'une assurance forêt fiable, ce qui est le but de ce mémoire. Pour lui, **assurer sa forêt**, « **c'est se donner les moyens de reconstituer le peuplement en cas de sinistre et de sécuriser la performance du placement patrimonial que constitue ce terrain** ».

## I.4.2 Historique et cadre réglementaire de l'assurance forêt

L'assurance forêt ne constitue pas une nouveauté. Les pays scandinaves ont la plus ancienne tradition d'assurance forestière avec le premier contrat d'assurance incendie proposé en Norvège en 1898. Un état de l'art complet de l'assurance forêt et de son historique a été réalisé en 2023 par Marielle Brunette et Stéphane Couture, tous deux membres de l'INRAE de Lorraine (BRUNETTE ET HANEWINKEL [2023]). Cet article recense et met en avant les principaux points ressortant de 38 articles publiés dans divers pays entre 1928 et 2021. La plupart des documents traitant de l'assurance forêt ont été publiés dans la dernière décennie (27 sur les 38 articles). Les principaux sujets de recherches sont le risque incendie aux États-Unis et le risque tempête en Europe.

En effet, les garanties majeures en assurance forêt couvrent ces deux risques que sont l'incendie et la tempête. Néanmoins, des garanties spéciales peuvent être appliquées en fonction des spécificités des pays. Ainsi, une garantie contre les dégâts causés par les insectes est proposée au Danemark et une garantie contre la perte de carbone dans les forêts est proposée en Nouvelle-Zélande. À l'échelle européenne, au cours de la période 1950-2000, une moyenne annuelle de 35 millions de mètres cubes de bois a été endommagée par des perturbations. Les tempêtes sont responsables de 53 % des dommages totaux, les incendies de 16 %, la neige de 3 % et les facteurs biotiques de 16 % (BRUNETTE ET COUTURE [2023]).

Le taux d'assurance des forêts privées en France est très faible (4 %) comparé à d'autres pays (95 % en Suède, 55 % en Nouvelle-Zélande ou encore 50 % en Chine) (BRUNETTE ET HANEWINKEL [2023]), en partie à cause d'un manque d'indicateurs fiables (probabilités de survie des essences, fréquence et gravité des perturbations sous changement climatique) pour évaluer précisément le risque. De plus, un problème de sélection adverse demeure : tous les propriétaires paient la même prime même si certains ont pris des mesures pour réduire les risques encourus sur leur patrimoine.

Du côté de la demande, la petitesse de la plupart des parcelles de la gestion forestière dissuade les propriétaires d'investir une part de leurs revenus dans une assurance. Mutualiser une surface forestière importante serait une solution pour des primes plus abordables et une meilleure répartition des risques. Pour compenser le problème de sélection adverse du aux primes uniformes évoqué précédemment, prendre en compte dans le calcul des primes les efforts de réduction des risques pourrait stimuler l'assurance en valorisant les pratiques jugées vertueuses.

L'État a tenté de prendre en charge ce problème de faible assurance. Ainsi, depuis le 1er janvier 2017, l'État a indiqué qu'il n'interviendrait plus pour aider les sylviculteurs à reconstituer leurs forêts après le passage d'une tempête, dans la mesure où ce risque est assurable. Des dispositifs fiscaux comme le CIFA ont d'ailleurs été mis en place pour assurer la promotion de l'assurance en forêt.

## I.4.3 Étude de marché et présentation de Groupama Forêts Assurances

Sur le marché français, plusieurs assureurs distribuent des produits relatifs à l'assurance de forêts. Parmi eux, on retrouve notamment :

- Sylvassur : assurance née du partenariat entre Fransylva Services et Verspieren
- XLB Assurances : courtier
- Groupama Forêts Assurances (anciennement MISSO)

La table I.8 détaille les informations principales relatives à l'un de ces acteurs, Groupama Forêts Assurances. Comme indiqué dans l'introduction, cet acteur est à l'origine de la mission ayant initié la réflexion de ce mémoire. Les diverses données de sinistres et de primes utilisées par la suite dans les travaux du mémoire sont des données issues de cet assureur. Certaines sont anonymisées et cela est précisé aux instants concernés.

Information	Détail
Origine	Créée en 1941 sous la forme de mutuelle régionale (MISSO), rebaptisée « Groupama Forêts Assurances ».
Clientèle cible	Propriétaires privés, collectivités et coopératives forestières, partout en France, mais principalement dans le Bassin Aquitain.
Garanties de base	Incendie, tempête-ouragan-cyclone, RC.
Options fréquentes	Grêle, givre, frais de reboisement, etc.
Distribution	Réseau d'agences Groupama et plateforme en ligne.
Points forts	Solidité financière du groupe, expertise historique, présence régionale.
Limites usuelles	Exclusion des très petites parcelles, motivation faible des sylviculteurs à s'assurer.

TABLE I.8 – Groupama Forêts Assurances : Fiche d'identité

#### I.4.4 Enjeux actuariels

Le premier point à soulever demeure la recherche des variables permettant d'expliquer le risque tempête en assurance forêt. Il faut donc concevoir une base de données solide et spatialisée intégrant des données physiques et environnementales tout en créant un modèle de risque qui reste réaliste par rapport à la réalité empirique.

De plus, l'impact des tempêtes n'est pas homogène sur le territoire et certaines zones sont naturellement plus vulnérables. Pourtant, les classifications du risque par les assureurs se font à une maille trop grande (région ou département) pour rendre vraiment compte de la variabilité locale du risque. Un zonier à l'échelle communale permettrait de mieux cerner la réalité et de mieux répartir le risque. Ce découpage géographique est donc une des clés à optimiser dans ce mémoire. Avoir une réponse mathématique serait par ailleurs une réponse fiable aux critiques sur le caractère arbitraire des prix perçus par l'assuré dans ce domaine de l'assurance.

Le principe de prudence entraîne une hausse des tarifs, faute de modélisation suffisamment fine. Ce mémoire cherche à fournir des tarifs fiables, fondés sur des justifications objectives. Cela permettrait non seulement de rassurer les assurés mais aussi de conquérir une clientèle plus large. Une meilleure compréhension des prix et des garanties pourrait faire émerger une demande d'assurance des forêts aujourd'hui encore faible en France.

Enfin, si l'on s'en tient à la dimension technique, le meilleur modèle de tarification est aussi une opportunité commerciale. Si l'assureur sait précisément distinguer les zones à risque modéré, il pourra intervenir efficacement sur de nouveaux segments de clientèle, ciblant des propriétaires non couverts. Une offre claire, lisible, adaptée et géolocalisée pourrait faire un véritable produit d'appel pour un propriétaire forestier encore hésitant, à condition d'affiner la mutualisation,

d'atteindre la rentabilité du portefeuille et de se positionner sur un marché encore peu structuré. L'innovation actuarielle devient ainsi un véritable levier stratégique de croissance pour le secteur.

## **I.5 Synthèse de la première partie et réflexions conséquentes pour ce mémoire**

Cette première partie a établi le cadre de ce mémoire en démontrant la nécessité d'une refonte des pratiques relatives au risque tempête en assurance forêt. Trois constats majeurs émergent de cette analyse contextuelle.

Premièrement, les effets du dérèglement climatique ces dernières décennies amplifient l'exposition aux tempêtes. Les sinistres Cat-Nat augmentent à ce titre de façon continue depuis 1982. Cette tendance invalide progressivement les bases historiques de tarification car les repères historiques deviennent progressivement non conformes à la réalité.

Deuxièmement, le marché français de l'assurance forêt reste peu développé : malgré une exposition croissante aux risques climatiques, seuls 4 % des propriétaires privés souscrivent une couverture, contre 95 % en Suède. Cette sous-assurance massive révèle l'inadéquation des outils tarifaires actuels, souvent perçus comme arbitraires ou disproportionnés par rapport au risque réel.

Troisièmement, l'hétérogénéité spatiale du risque tempête en France métropolitaine nécessite une approche géodifférenciée. Les zoniers administratifs classiques (départementaux ou régionaux) s'avèrent insuffisants pour capter les variations locales de vulnérabilité forestière.

En réponse à ces défis, ce mémoire développe une méthodologie actuarielle innovante articulée comme suit : modélisation mathématique du risque tempête puis création d'un zonier à la maille communale afin d'effectuer une refonte géographique de la tarification en France métropolitaine. Pour pouvoir effectuer cela, la partie suivante s'intéresse à la construction d'une base de données constituée de variables géophysiques à la maille communale.

En synthèse, cette première partie fournit les bases ainsi que les enjeux d'une assurance forestière pérenne en France, adaptée aux défis du climat.

## II - Élaboration d'une base de données adaptée à l'étude

### II.1 Étude de référence : *Windthrow Hazard Classification*

#### II.1.1 Contexte de l'étude

Partant du constat posé en partie I sur l'exigence d'une tarification fiable, la revue de la littérature a identifié une référence centrale : l'étude *Windthrow Hazard Classification* de K. F. Miller (1985), scientifique britannique. Cette étude fournit le cadre nécessaire à la constitution d'une base de données robuste. Sauf mention contraire, toutes les données présentées dans la sous-partie II.1 en sont issues (MILLER [1985]).

Pour recontextualiser, le domaine assurantiel au Royaume-Uni se trouve en 1985 dans une période charnière de son histoire (ARNAUD [1997]). C'est d'ailleurs l'année de la création de l'*Association of British Insurers* (ABI), qui unifie les différentes branches britanniques afin d'avoir une voix unique lors des pourparlers européens. Ce sont aussi des années qui nécessitent une réflexion autour des nouveautés technologiques et sociales, et de leur intégration dans la prise en compte du risque.

À cette période, au Royaume-Uni, le mot d'ordre en finance et en assurance est l'autoréglementation. Malgré l'avènement de plus en plus régulier de normes et de réglementations européennes, l'assurance britannique montre un désir d'indépendance notable. À ce titre, la réforme de la régulation des *Lloyd's* en 1982 ou encore le *Big Bang* de 1986, qui attire les assureurs étrangers du fait de la déréglementation qui s'y opère, démontre la volonté d'instaurer une nouvelle ère de l'assurance britannique. Cette volonté repose sur des arguments solides. En effet, à cette période, le Royaume-Uni réalise 26 % des primes globales d'assurance européennes, ce qui en fait le deuxième pays le plus important à ce sujet en Europe, après l'Allemagne de l'Ouest.

Cette dynamique autour du risque dans les secteurs financiers a son pendant dans la gestion du risque forestier, domaine dans lequel s'inscrit l'étude de K.F. Miller, chercheur rattaché à la *Forestry Commission*. Il rédige ainsi son étude, présentant un zonage du risque de déracinement des arbres au Royaume-Uni. Si le pays n'est pas le plus réputé pour ses tempêtes, il n'en demeure pas moins nécessaire de classer ce risque à l'échelle nationale, pour prévenir d'événements graves.

#### II.1.2 Objectifs d'utilisation de cette étude

Bien qu'écrite il y a 40 ans, cette étude demeure pertinente car elle mobilise des facteurs biophysiques intemporels et sert encore de référence fiable pour les modèles actuels.

Dans un premier temps, l'objectif, afin de créer un zonier commune par commune en France, est de réutiliser cette étude, qui a été concluante au Royaume-Uni, tout en l'adaptant aux

caractéristiques de la France. En effet, il est assez cohérent de penser que, malgré une influence certaine, le comportement des variables explicatives n'est pas exactement le même en France et au Royaume-Uni. Un premier travail consiste à observer les travaux réalisés par K.F. Miller et à examiner si les variables explicatives sont utilisables à l'échelle française (données fiables, variables intéressantes dans le nouveau référentiel, possibles nouvelles variables à intégrer...) en les exploitant de la manière la plus juste possible.

L'altitude, à titre d'exemple, peut jouer un rôle très différent dans ces deux pays. En comparant les statistiques topographiques, l'altitude moyenne en France est de 375 mètres tandis que celle au Royaume-Uni est de 162 mètres. De manière encore plus significative, l'altitude maximale n'atteint que 1345 m au Royaume-Uni au niveau du Ben Nevis en Écosse. En France, l'altitude maximale est de 4806 m, sachant que de nombreux massifs, parmi lesquels les Alpes et les Pyrénées, présentent des altitudes très élevées.

Cet exemple donne l'intuition que l'altitude va influencer différemment le risque de déracinement des arbres au Royaume-Uni et en France. Si l'altitude a été prise en exemple en raison de sa représentation claire et directe, le même raisonnement logique peut s'appliquer à différentes variables.

Néanmoins, avant d'entamer une analyse propre au cas français, il convient d'analyser en détail les travaux menés par K.F. Miller concernant le Royaume-Uni.

### II.1.3 Travaux réalisés par K.F. Miller

Le but de Miller est d'explicitier, à l'aide de quatre variables explicatives, le risque de déracinement dans les plantations de conifères au Royaume-Uni. Ainsi, une carte du niveau de risque de déracinement sur l'ensemble du Royaume-Uni est obtenue.

Les quatre variables explicatives évoquées précédemment sont :

- **La force du vent de la zone géographique considérée** : Cette caractéristique est déterminée à partir du zonier présenté ci-dessous.

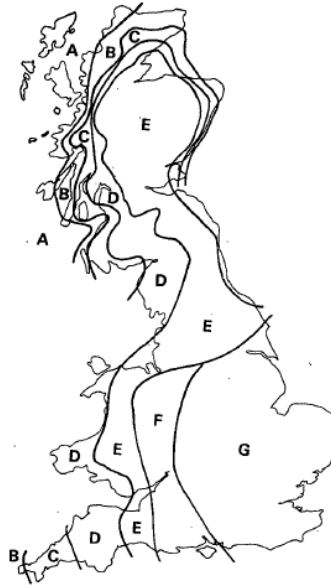


FIGURE II.1 – Étude britannique : Zonier du vent

En fonction de la zone dans laquelle le point est situé (de A à G), un score est attribué. Ce score est compris entre 0 et 13, comme indiqué Table II.1, et correspond donc à l'impact du vent sur le déracinement.

Zone de Vent	A	B	C	D	E	F	G
Score Associé	13	11	9,5	7,5	2,5	0,5	0

TABLE II.1 – Étude britannique : Score associé au vent

- **L'altitude** : La même idée est reprise pour l'altitude. Un score est attribué en fonction de l'élévation au-dessus du niveau de la mer. Cette donnée est fournie ici par l'*Ordnance Survey*, qui est l'agence exécutive du gouvernement du Royaume-Uni en charge de la cartographie du pays.

La séparation des altitudes et leurs scores associés sont les suivants :

Altitude (m)	Score
541+	10
466-540	9
406-465	8
361-405	7
316-360	6
286-315	5
256-285	4
226-255	3
191-225	2
141-190	1
61-140	0.5
0-60	0

TABLE II.2 – Étude britannique : Score associé à l'altitude

La faible différence d'altitude fait que quelques dizaines de mètres peuvent beaucoup influencer sur le score, ce qui ne sera pas le cas en France étant donné l'écart d'altitudes plus conséquent.

- **L'abri topographique** : Également appelé TOPEX, il s'agit du degré de protection offert par l'environnement autour de la zone étudiée. Un abri topographique élevé peut protéger du vent par le biais de montagnes, de collines ou encore via d'autres abris naturels tels que des infrastructures créées par l'homme.

TOPEX (total °)	Score
0-9	10
10-15	9
16-17	8
18-19	7
20-22	6
23-24	5
25-27	4
28-40	3
41-70	2
71-100	1
101+	0

TABLE II.3 – Étude britannique : Score associé à l'abri topographique

- **Le type de sol** : L'étude utilise aussi les informations relatives au type de sol, et en particulier la capacité d'enracinement sur la zone considérée. La correspondance entre les types de sols et leur niveau de risque se trouve en **Annexe A**. La table II.4 récapitule grossièrement ces scores.

Capacité d'enracinement (cm)	Score
+ de 45	0
Entre 25 et 45	5
- de 25	10

TABLE II.4 – Étude britannique : Score associé à la capacité d'enracinement

Au final, dans l'étude britannique, l'attribution du niveau de risque (qui va du risque mineur (1) au risque majeur (6)) pour la zone considérée se fait par le biais d'un score total qui cumule les 4 scores par addition.

Le résultat final de l'étude, qui suit donc un système empirique global, établit les scores comme suit :

Score total	Classe de hasard
0-7,5	1
8-13,5	2
14-19,0	3
19,5-24,5	4
25-30,5	5
$\geq 31$	6

TABLE II.5 – Étude britannique : Classes de hasard

Avec ces classes, il est alors possible de décomposer un territoire, tel qu'une forêt, par zones de risque. Un exemple d'une parcelle de forêt est donné avec la figure II.2 :

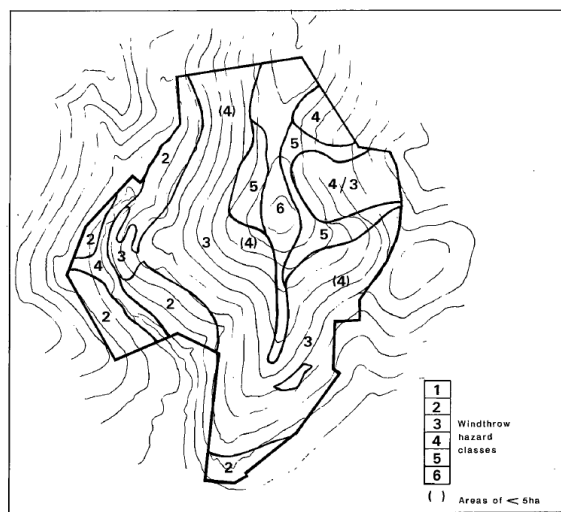


FIGURE II.2 – Étude britannique : Zonier du niveau de risque dans une forêt

La maille de la parcelle est fine et le score évolue entre risque faible et élevé. Cela justifie une modélisation communale pour notre étude.

## II.1.4 Enseignements clés et variables retenues pour la modélisation

L'idée d'attribuer un score aux variables offre une première approche de modélisation intéressante à mettre en place. Cela fournit une première fondation de travail et souligne surtout que, au-delà de son caractère actuel, la modélisation présentée dans ce mémoire aborde une problématique qui persiste depuis plusieurs années. De plus, l'état de l'art actuel met à disposition des outils plus efficaces qu'en 1985 pour répondre à ces problématiques.

Bien que pionnière dans son domaine, la grille de Miller s'appuie sur des données antérieures à l'intensification climatique observée depuis les années 1990. En fixant le risque dans quatre variables physiques et des classes empiriques, elle ignore la dynamique des vents, la migration des essences et la modification des bilans hydriques. Or, les contrastes topographiques et climatiques entre le Royaume-Uni et la France font qu'une transposition identique risque de biaiser la tarification. Il faut donc être attentif à bien calibrer chacune des variables choisies pour cette étude.

Afin d'effectuer un zonier via calcul du score dans un premier temps, cinq variables sont conservées, fortement liées à l'étude de Miller. La table suivante présente les variables gardées ainsi que leurs unités :

Nom	Explication	Unité
Sol	Nature du sol prédominant dans la commune	—
Élévation	Altitude moyenne de la commune	m
Essence	Catégorie des arbres présents : conifères, feuillus ou mélange des deux dans la commune	—
Vent	Vitesse maximale du vent relevée dans le département l'année du sinistre	km/h
Abri Topographique	Mesure de la protection offerte par le relief environnant, calculée dans un rayon de 5 km autour du point d'intérêt	°

TABLE II.6 – Variables retenues : Zonier par score

Étant issues d'une étude fiable et qui englobe précisément les problématiques de ce mémoire, les quatre variables de Miller sont conservées, avec, en supplément, une distinction entre l'essence d'arbre majoritaire (conifères, feuillus et mélange des deux) dans la commune considérée, ce qui est une distinction majeure en France, au regard des différents échanges menés avec des experts du domaine forestier. Ces variables sont à la base de la constitution de la base de données.

## II.2 Objectifs et périmètre adopté

### II.2.1 Besoins analytiques

Après avoir sélectionné les variables utiles à l'analyse dans la partie précédente, il est donc nécessaire de collecter ces données à la maille communale. Cette base de données se doit d'être complète, c'est-à-dire qu'elle s'adapte à la base de données de sinistres : à chaque commune sinistrée peut être associée des données géographiques. Une clé qui puisse relier la base de sinistres et la base de données communale doit donc être trouvée.

Après la collecte de ces données, une attention toute particulière est apportée à la qualité de ces données trouvées avec un contrôle de cohérence au moins visuel. Ensuite, celles-ci doivent être retravaillées et implémentées dans des modèles dans le langage de programmation R avant d'être cartographiées.

### II.2.2 Démarche adoptée

Pour mener à bien les travaux de recherche de données, en plus des recherches personnelles préliminaires, des conseils et des sources ont été fournis par Monsieur Barry Gardiner, scientifique senior à l'Institut Européen de la Forêt Cultivée (IEFC) et titulaire d'un doctorat de l'Université de Manchester (1983). Cet expert des problématiques forestières a fourni son expertise dans la collecte de données et a pu, par la suite, confirmer ou infirmer la véracité des résultats obtenus

au fur et à mesure du mémoire. Ce mémoire se base donc sur des données vérifiées, qui seront présentées dans la sous partie II.4.

Concernant la cartographie, le logiciel utilisé s'intitule QGIS. C'est un SIG en libre accès. L'interface du logiciel est montrée ci-dessous, avec l'importation d'une carte de France segmentée en communes (DATA NETWORK TEAM [2025]), sur laquelle les travaux futurs se basent afin d'effectuer les zoniers.

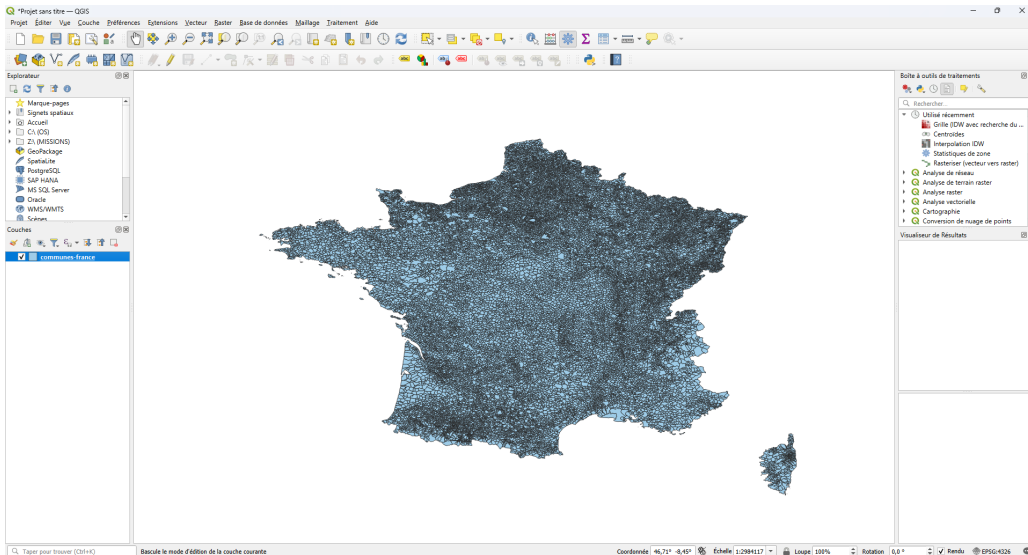


FIGURE II.3 – QGIS : Aperçu de l'interface

Dès cette première carte simplement importée dans QGIS, de nombreuses et précieuses informations sont obtenues sur les 34 805 communes de France métropolitaine, avec, entre autres, les données suivantes obtenues pour trois communes : Gosselming, Montsapey et Cudos :

Ville	Gosselming	Montsapey	Cudos
Latitude	48,789873327525	45,53128640648	44,378001471980
Département	Moselle (57)	Savoie (73)	Gironde (33)
Arrondissement	Sarrebouurg-Château-Salins	Saint-Jean-de-Maurienne	Langon
Communauté de communes	Sarrebouurg Moselle Sud	Porte de Maurienne	Bazadais
Code INSEE	57255	73270	33144
Aire montagneuse	NON	OUI	NON

TABLE II.7 – QGIS : Comparaison de valeurs pour trois communes françaises

Ces trois communes serviront de références en tant qu'exemples tout au long de ce mémoire. Elles ont été choisies pour représenter une commune dans la zone du portefeuille en Gironde, une ville à risque supposé modéré en Moselle ainsi qu'une commune dans une zone d'altitude en Savoie.

De manière générale, la base de données se présente de la façon suivante dans le logiciel :

	geo_point_2d	reg_code	reg_name	dep_code	dep_name	arrdep_code	arrdep_name	bv2022_code	bv2022_name	epci_name	com_code	com_name
2338	lat: 47.7460032...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	41269	Vendôme	Communauté d...	41293	Villiersfaux
2339	lat: 47.7981949...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	41269	Vendôme	Communauté d...	41003	Areines
2340	lat: 47.9623241...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	41143	Mondoubleau	Communauté d...	41060	Cormenon
2341	lat: 47.7788098...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	72269	Saint-Calais	Communauté d...	41250	Sougé
2342	lat: 47.7864292...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	41269	Vendôme	Communauté d...	41259	Thoré-la-Roche...
2343	lat: 47.9591765...	24	Centre-Val de L...	41	Loir-et-Cher	412	Vendôme	41269	Vendôme	Communauté d...	41048	Chauvigny-du-...

FIGURE II.4 – QGIS : Extrait de la base de données communale

De nombreuses possibilités sont envisageables avec QGIS. Grâce au croisement ou à l’empilage de couches vectorielles comme une carte communale et de couches raster telles qu’une grille des vents moyens, il est possible de confronter de nombreuses variables. Le logiciel permet ensuite d’enchaîner avec des requêtes spatiales ciblées, afin d’isoler, par exemple, les communes où les rafales excèdent 120 km/h tout en reposant sur des sols à faible enracinement. Ces sélections peuvent être enrichies par le calcul automatique d’indicateurs statistiques pour extraire la médiane, la somme ou encore un score. Une fois ces traitements réalisés, QGIS offre des outils de cartographie avancée de haute résolution, directement exploitables et modulables selon l’utilité qui doit en être faite.

Maintenant que les besoins et la démarche pour y répondre sont exposés, les deux bases de données vont être présentées, en commençant par la base de données historique des sinistres.

## II.3 Base de données historique

### II.3.1 Présentation de la base

Conformément à ce qui a été présenté, les équipes de Groupama Forêts Assurances, acteur majeur du domaine de la protection assurantielle forestière, ont donné accès à leurs données dans le cadre d’une mission client. Aucune information confidentielle n’est publiée dans ce mémoire, mais les sources proviennent bien d’une origine professionnelle et ancrée dans le cadre de l’assurance forêt.

Groupama Forêts Assurances étant implanté majoritairement dans le Bassin aquitain, la base de données de sinistres de ce mémoire est composée majoritairement de sinistres issus de cette zone. Le Bassin aquitain s’étend sur 8 départements, dans une zone segmentée au sud-ouest de la France et présentée ci-dessous :

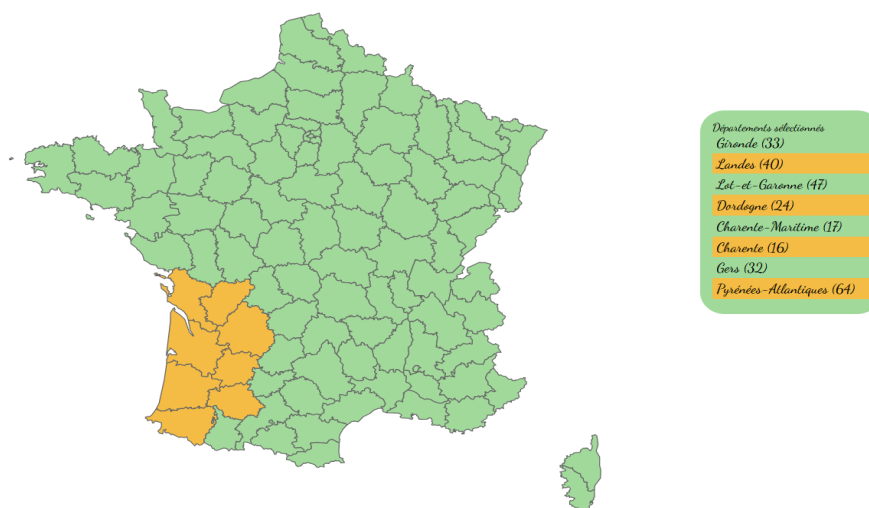


FIGURE II.5 – Bassin aquitain : Départements concernés

D'autres sinistres sont présents, issus de diverses régions françaises. Au total, 2 701 sinistres sont recensés depuis 1997 pour la garantie tempête. La base de données présente des informations diverses, dont cinq en particulier sur lesquelles cette étude peut s'appuyer, que la table suivante explicite à l'aide d'un exemple. Celui-ci s'est produit dans l'une des trois communes retenues comme cas d'étude tout au long du mémoire.

Donnée	Exemple
Département	Gironde (33)
Année de survenance	2009
Commune	Cudos
Surface assurée totale (ha)	12,920
Surface sinistrée totale (ha)	12,540

TABLE II.8 – Base sinistres : Exemple de données d'entrée

En s'intéressant aux coûts cumulés des sinistres par année de l'historique récent (entre 2008 et 2023), une année sort du lot de manière nette. Il s'agit de 2009, durant laquelle la tempête Klaus a ravagé la partie sud-ouest de la France, soit le lieu d'implantation principal de Groupama Forêts Assurances.

Année	Sinistres	Année	Sinistres
2023	175 400	2015	77 600
2022	857 700	2014	16 000
2021	56 700	2013	26 000
2020	28 700	2012	79 500
2019	9 400	2011	10 800
2018	91 400	2010	27 000
2017	51 800	2009	16 172 800
2016	0	2008	17 900

TABLE II.9 – Coût des sinistres : Évolution annuelle

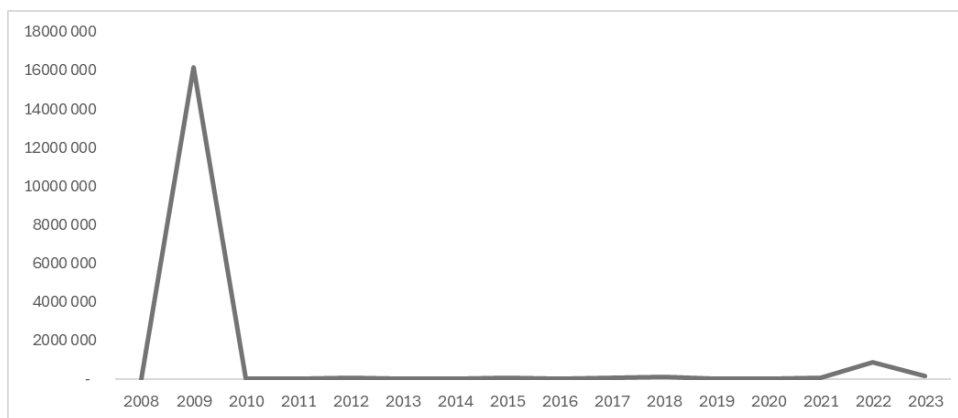


FIGURE II.6 – Coût des sinistres : Évolution annuelle

La base de données actuelle ne fait pas la distinction entre les années selon qu’elles comptent un, aucun ou plusieurs sinistres majeurs. Une première intuition est qu’il faudra veiller à bien distinguer les exercices dominés par des sinistres attritionnels (c’est-à-dire des sinistres courants), de ceux marqués par des catastrophes, comme la tempête Klaus en 2009.

### II.3.2 Analyse des données

Les sinistres sont recensés majoritairement en 1999 (tempêtes Lothar et Martin) et 2009 (tempête Klaus). Cela est assez explicite dans la table II.10. Les tempêtes de 1999 étaient les plus fortes et ont touché de manière plus étendue la France tandis que Klaus a été très localisée dans le sud-ouest.

Année	Sinistres	Année	Sinistres
2023	16	2009	583
2022	23	2008	4
2021	3	2007	2
2020	10	2006	29
2019	3	2005	2
2018	15	2004	8
2017	9	2003	32
2016	0	2002	2
2015	2	2001	13
2014	6	2000	12
2013	3	1999	1 891
2012	9	1998	12
2011	1	1997	1

TABLE II.10 – Nombre de sinistres : Évolution annuelle

Parmi les 2701 sinistres, 1877 étaient situés géographiquement dans le Bassin aquitain. Les communes les plus sinistrées sont toutes situées dans ce périmètre géographique, comme le montre la table II.11 ainsi que la carte de répartition des sinistres en figure II.7.

Département	Commune	Nombre de sinistres
Gironde (33)	ST JEAN D'ILLAC	37
Landes (40)	PISSOS	34
Gironde (33)	CESTAS	29
Gironde (33)	SALLES	27
Landes (40)	LOSSE	26

TABLE II.11 – Nombre de sinistres : Cinq communes les plus sinistrées

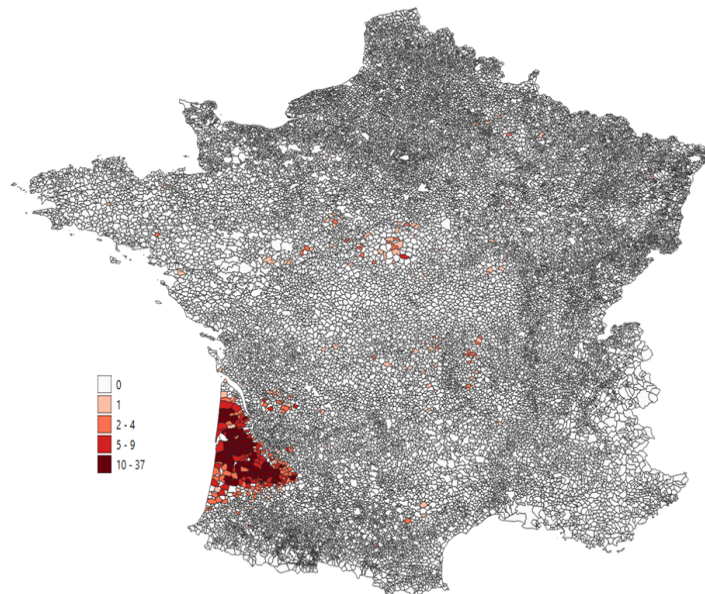


FIGURE II.7 – Nombre de sinistres par commune : Zonier

La table II.12 met de son côté en avant la portion moyenne de terrain qui est assurée et la portion moyenne qui est sinistrée.

Surface assurée moyenne (ha)	Surface sinistrée moyenne (ha)
136,412	35,063

TABLE II.12 – Surfaces assurées et sinistrées : Moyennes

### II.3.3 Travaux réalisés sur la base de données

À l'aide de ces données, divers travaux sont menés afin d'obtenir une base harmonieuse et ajustée.

- **Calcul du ratio de destruction** : Étant la variable à expliquer dans les modèles de ce mémoire, cette variable est calculée prioritairement. Elle s'établit selon la formule :

$$\text{Ratio de destruction} = \frac{\text{Surface sinistrée}}{\text{Surface assurée}}$$

Dans l'exemple du sinistre à Cudos présenté dans la table II.8, le ratio de destruction est de 97,06 %. Le ratio de destruction moyen sur la base de sinistres est, après gestion des cas extrêmes (ratios non compris entre 0 et 1, etc.), de 32,33 %.

- **Association de chaque commune à son code INSEE :** Un deuxième travail important est de faire correspondre chaque commune au code INSEE qui lui est propre.

Sur ce travail, quatre points d'attention sont à retenir en particulier :

- Tout d'abord, la gestion des tirets, des espaces et des accents n'est pas la même entre la base de sinistres et la base de l'INSEE. Ainsi, tout a dû être uniformisé avant de réaliser n'importe quel travail.
- En France, certaines villes portent le même nom, comme Angerville, qui peut se localiser à la fois dans le Calvados avec le code INSEE 14012 ou bien dans l'Essonne avec le code 91016. Un test sur la conformité des deux premiers chiffres du code INSEE avec le département du sinistre est réalisé, et des modifications entreprises en cas de test erroné.

À titre d'exemple, dans la base de données de sinistres se trouvent ces 2 lignes suivantes, dont l'une aurait été assimilée au mauvais code INSEE sans ce travail préliminaire.

Donnée	Exemple 1	Exemple 2
Département	18	89
Année de survenance	1999	2009
Commune	<b>Gron</b>	<b>Gron</b>
Surface assurée totale	42,14	1,33
Surface sinistrée totale	20,37	0,20

TABLE II.13 – Travaux sur la base : Gestion de deux communes avec le même nom

- Groupama Forêts Assurances possède une connaissance fine du sud-ouest de la France, où elle est implantée. Des noms de lieux-dits ou de forêts précises sont parfois mentionnés comme emplacement du sinistre. Des recherches supplémentaires sont menées pour ces cas précis où l'emplacement était donné de manière trop fine, afin de trouver la commune rattachée.
- Enfin, le cas où certaines communes n'existent plus car celles-ci ont été absorbées ou fusionnées s'est présenté très ponctuellement. Là aussi, des recherches complémentaires sont mises en œuvre.

Ce travail préparatoire est primordial pour faire des liens avec les autres bases, qui seront présentées dans les parties et sous-parties suivantes.

## II.3.4 Base de primes historique

La base de primes a également été mise à disposition par Groupama Forêts Assurances à partir de 2008. Cela ne fournit donc pas les données de cotisations pour l'année 1999, celle de Lothar et Martin. Pour des raisons de confidentialité, la base est quelque peu anonymisée. Une chronique de primes fictive, mais se basant sur des données de primes cohérentes avec la réalité, est donc considérée. L'évolution des primes est à ce titre expliquée dans le paragraphe ci-après.

Les primes totales annuelles en 2008 et en 2009 sont de 1 400 000 €. Puis, l'impact de la tempête Klaus a mené à la prise de conscience de l'intérêt de la garantie tempête en assurance forêt et, en 2010 et 2011, les primes se sont élevées à 2 000 000 €. Ces primes baissent progressivement jusqu'en 2016. Enfin, l'augmentation générale des catastrophes naturelles, couplée à la prise de conscience des risques engendrés par le dérèglement climatique fait augmenter les primes progressivement jusqu'à aujourd'hui, atteignant en 2023 1 800 000 €. Une représentation visuelle de cette évolution est donnée par la figure II.8.

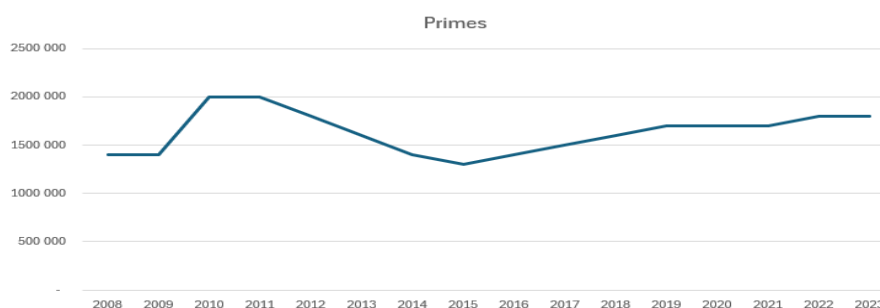


FIGURE II.8 – Primes reçues : Évolution annuelle

Les ratios S/P peuvent être calculés à partir des valeurs annuelles de primes et de sinistres.

Année	Sinistres	Primes	S/P
2023	175 400	1 800 000	10 %
2022	857 700	1 800 000	48 %
2021	56 700	1 700 000	3 %
2020	28 700	1 700 000	2 %
2019	9 400	1 700 000	1 %
2018	91 400	1 600 000	6 %
2017	51 800	1 500 000	3 %
2016	0	1 400 000	0 %
2015	77 600	1 300 000	6 %
2014	16 000	1 400 000	1 %
2013	26 000	1 600 000	2 %
2012	79 500	1 800 000	4 %
2011	10 800	2 000 000	1 %
2010	27 000	2 000 000	1 %
2009	16 172 800	1 400 000	1 155 %
2008	17 900	1 400 000	1 %

TABLE II.14 – Sinistres, primes et ratios S/P : Évolution annuelle

À nouveau, ces chiffres mettent en avant une séparation immédiate entre 2009 et les autres années. Cela engendre une séparation entre S/P attritionnel, qui équivaut à la moyenne des S/P pour toutes les années hormis 2009, et S/P catastrophe qui est donc égal au S/P de l'année 2009.

S/P attritionnel	S/P catastrophe
6 %	1 155 %

TABLE II.15 – Ratios S/P : Distinction entre attritionnel et catastrophe

Ces S/P sont par la suite utilisés lors de la mise en place d'une tarification. Il est toutefois intéressant de noter dès maintenant la dérive de sinistralité observée au cours des deux années d'historique les plus récentes. Bien qu'il n'y ait pas eu de sinistres majeurs sur la zone d'implantation de Groupama Forêts Assurances, cela peut être expliqué par différents facteurs : un effet collatéral d'autres tempêtes plus puissantes en dehors du portefeuille (Ciaran en 2023, etc.), une augmentation des sinistres attritionnels ou encore une population qui déclare davantage ses sinistres. Ces deux exercices, marqués par une augmentation du S/P, sont à l'origine de la mission confiée qui a donné lieu à ce mémoire. La modulation tarifaire qui sera proposée vise à ramener le S/P à un niveau plus conforme aux attentes.

### II.3.5 Agrégation avec la base de données communale

La base de sinistres doit maintenant être agrégée avec les données géographiques afin de pouvoir réaliser la modélisation mathématique du risque tempête et le zonier par commune. Pour cela, le lien se fait par le code INSEE des communes, d'où l'importance du travail présenté plus haut.

Cette base de données par commune est présentée dans la sous-partie suivante. Celle-ci a pour but d'explicitier l'origine des données qui sont utilisées par la suite, et qui ont été introduites précédemment, à savoir le type de sol, l'altitude, le vent maximal, l'abri topographique et l'essence des arbres.

## II.4 Base de données par commune

Les statistiques descriptives (moyenne, minimum, maximum, espèce majoritaire, etc.) propres à chaque variable présentée dans cette section sont détaillées dans l'**Annexe E**. On constate que le Bassin aquitain possède une altitude inférieure à la moyenne nationale, procurant un abri topographique plus limité.

### II.4.1 Données relatives à l'altitude

Conformément à ce qui a été dit précédemment, les données de ce mémoire sont le fruit d'une réflexion et d'échanges communs entre Actuelia, Groupama Forêts Assurances et Barry Gardiner.

Après avoir réfléchi ensemble aux données intéressantes et aux formats qu’elles devaient prendre, des cartes d’une grande précision ont pu être obtenues. À ce titre, l’extraction des données d’altitude s’est appuyée sur une carte SIG transmise par Barry Gardiner. Celle-ci donne l’altitude et la bathymétrie en France métropolitaine. Seule l’altitude est réellement utile dans le cadre de l’étude. Initialement, les données sont sous la forme de la figure II.9. Pour extraire les données par commune, une deuxième carte donnant cette séparation par commune a été importée (DATA NETWORK TEAM [2025]), comme l’illustre la figure II.10.



FIGURE II.9 – Altitude et bathymétrie en France métropolitaine : Carte

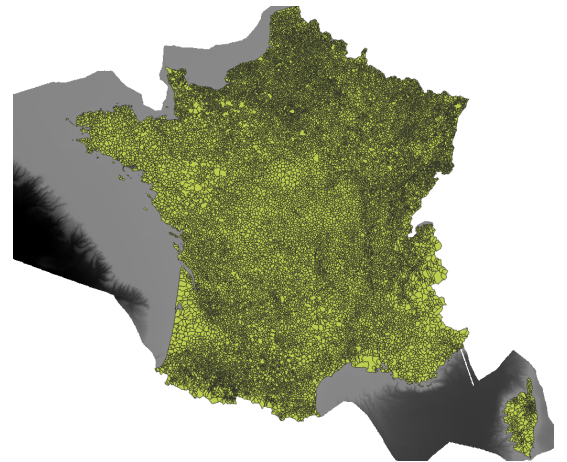


FIGURE II.10 – Altitude et communes : Cartes superposées

Depuis QGIS, il est possible d’extraire divers éléments par commune via l’outil « Statistiques de zone », comme la moyenne des données d’altitude rattachée à la commune. Il est également possible de déterminer, pour chaque commune, l’altitude minimale, maximale et médiane, ainsi que l’écart-type ou encore l’altitude la plus fréquemment observée.

Étant donné la forte disparité possible entre les altitudes maximales et minimales observées dans certaines communes, l’altitude moyenne est conservée pour la suite de ce mémoire. On présente les altitudes dans les tables et figures suivantes, où l’on retrouve clairement des zones de hautes altitudes dans les Alpes et les Pyrénées, ce qui conforte la cohérence des cartes.

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Altitude moyenne (m)	255,22	1569,33	110,42

TABLE II.16 – Exemple de trois communes françaises : Altitude moyenne

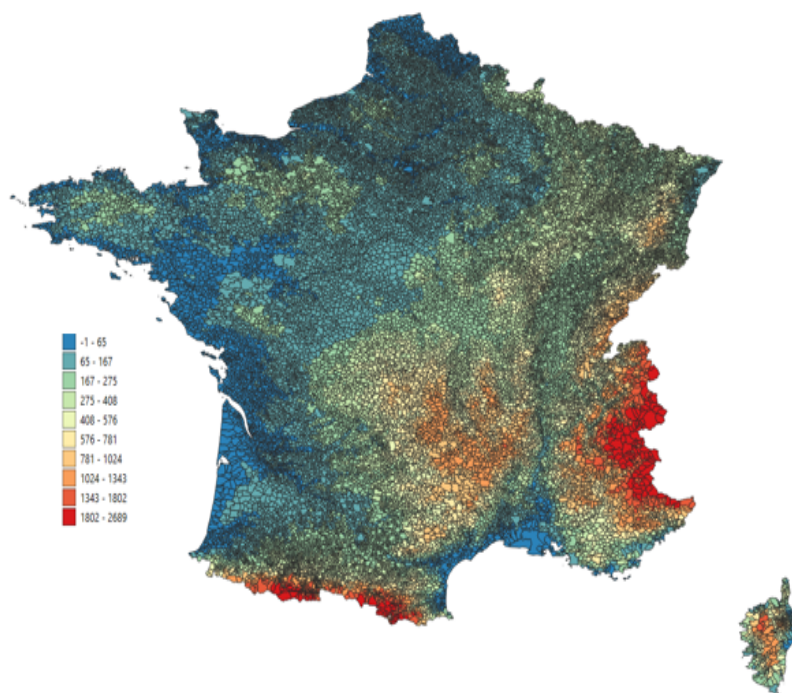


FIGURE II.11 – Altitude par commune en France métropolitaine : Zonier

## II.4.2 Données relatives au sol

Pour le type de sol, une carte précise à l'échelle européenne a été transmise par le chercheur Barry Gardiner. Celle-ci a ensuite été retravaillée pour extraire les données à l'échelle de la France métropolitaine.



FIGURE II.12 – Sol majoritaire en France métropolitaine : Carte

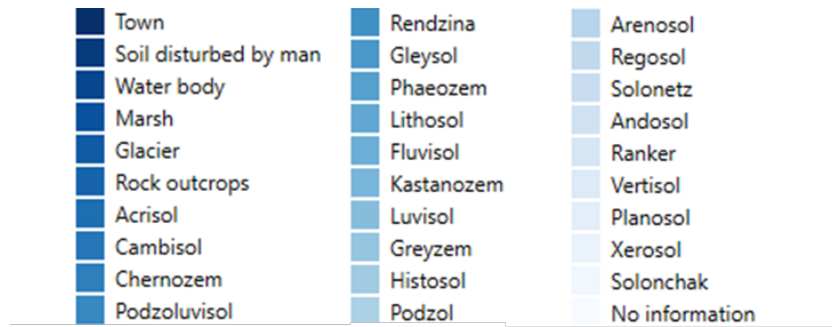


FIGURE II.13 – Sol majoritaire en France métropolitaine : Légende

Cette fois-ci, ce n'est plus la moyenne qui est prise en considération mais le sol majoritaire au sein de la commune, c'est-à-dire le plus fréquent.

Il y a 30 types de sols recensés comme étant majoritaires en France métropolitaine, comme le montre la figure II.13.

Enfin, dans les trois villes prises en exemple depuis le début des travaux, l'analyse QGIS montre que trois sols différents sont majoritaires (Table II.17). C'est donc bien une variable qui évolue et qui a de l'impact en France.

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Type de sol majoritaire	Cambisol	Podzol	Luvisol

TABLE II.17 – Exemple de trois communes françaises : Type de sol majoritaire

Les résultats sont dans l'ensemble cohérents. En effet, Gosselming est une terre agricole en Moselle et le cambisol, grâce son contenu élevé en minéraux, est idéal pour l'agriculture. Cudos est majoritairement composé de luvisol, qui est favorable pour les terres agricoles, mais aussi viticoles, bien connues en Gironde. Enfin, le podzol est un sol typique des régions à longs hivers froids, ce qui correspond bien à une ville savoyarde comme Montsapey.

### II.4.3 Données relatives à l'abri topographique

Dans cette section est considéré l'abri topographique à 5 kilomètres. Il s'agit du niveau de protection qui entoure un point dans un rayon de cinq kilomètres. Il s'exprime en degrés. Plus l'abri topographique est faible, moins la commune est abritée. La carte présentée figure II.14 est également fournie par Barry Gardiner. Trois données étaient disponibles : l'abri topographique à 2, 5 et 10 kilomètres. Les résultats ont été par la suite obtenus avec les trois données et, ces résultats étant globalement similaires, le choix s'est porté sur la valeur médiane, à savoir l'abri à 5 kilomètres.

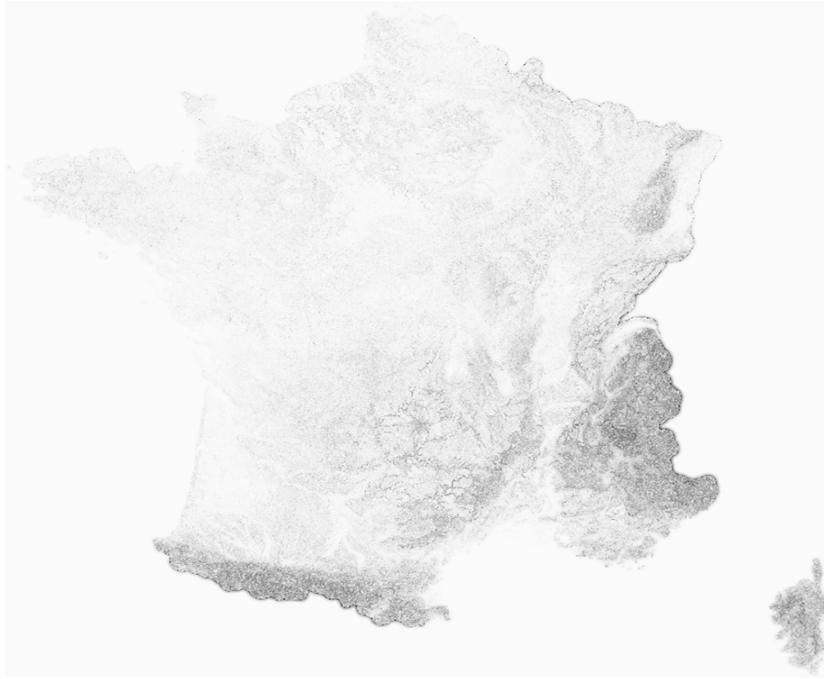


FIGURE II.14 – Abri topographique à 5 kilomètres en France métropolitaine : Carte

Dans cette figure, plus le point est foncé, plus l’abri est élevé. Le zonier met en avant que les zones montagneuses (Alpes, Pyrénées, Corse...) sont les mieux abritées. Pour agréger par commune, une moyenne des points par commune est de nouveau utilisée.

Pour les trois villes prises en exemples tout au long de cette partie, il est possible de souligner directement un abri similaire entre Gosselming et Cudos et un abri nettement supérieur, dû aux montagnes notamment, pour Montsapey, qui a pour rôle de protéger les espèces forestières qui s’y trouvent. Un zonier donné figure II.15 précise l’abri topographique par commune.

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Abri topographique à 5 km (en °)	15,90	125,99	10,87

TABLE II.18 – Exemple de trois communes françaises : Abri topographique

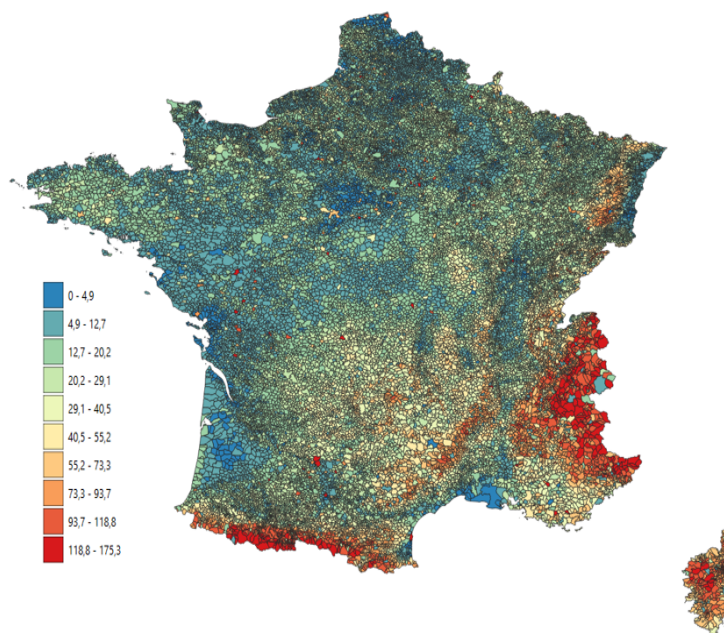


FIGURE II.15 – Abri topographique par commune en France métropolitaine : Zonier

La lecture croisée des cartes II.11 et II.15 montre un motif comparable entre altitude et TOPEX, mais la concordance n'est que partielle. Dans les plaines ouvertes (Beauce, Bassin aquitain), l'altitude modeste s'accompagne le plus souvent d'un TOPEX inférieur à 15°, traduisant une exposition marquée au vent. Pourtant, certaines buttes littorales ou reliefs isolés bénéficient d'un abri élevé malgré une cote proche du niveau 0, preuve que le TOPEX saisit l'horizon angulaire plutôt que la hauteur absolue. En somme, altitude et abri restent liés mais les considérer conjointement évite de sur-pondérer l'un ou l'autre lors d'une modélisation mathématique.

#### II.4.4 Données relatives à l'essence de l'arbre

Dans cette section, l'intérêt principal réside dans le fait de séparer les communes selon que les forêts qui sont sur leur territoire détiennent principalement des arbres feuillus ou des conifères.

Pour cela, l'expérience historique prime et, pour la plupart des sinistres, une note donnant un élément sur la nature de l'arbre figure dans la base de données. Pour toutes les communes au sein desquelles un sinistre s'est déroulé et pour lesquelles une note de l'assureur est rédigée, il est possible de trouver l'essence, comme explicité en détail dans l'**Annexe C**.

Néanmoins, pour les communes sans note spécifiée, la donnée est issue là aussi d'une carte des types d'arbre par forêt en France. La table II.19 montre qu'il est nécessaire d'introduire une notion de mélange, lorsque le recouvrement majoritaire des forêts est un mélange entre conifères et feuillus.

#	Arbre / Type de peuplement	Classification
1	Futaie de feuillus	Feuillus
2	Futaie de conifères	Conifères
3	Futaie mixte	Mélange
4	Mélange de futaie de feuillus et taillis	Feuillus
5	Mélange de futaie de conifères et taillis	Conifères
6	Taillis	Mélange
30	Forêt ouverte	Mélange
40	Lande	Mélange
50	Peupleraie	Feuillus

TABLE II.19 – Peuplements forestiers : Typologie et classification

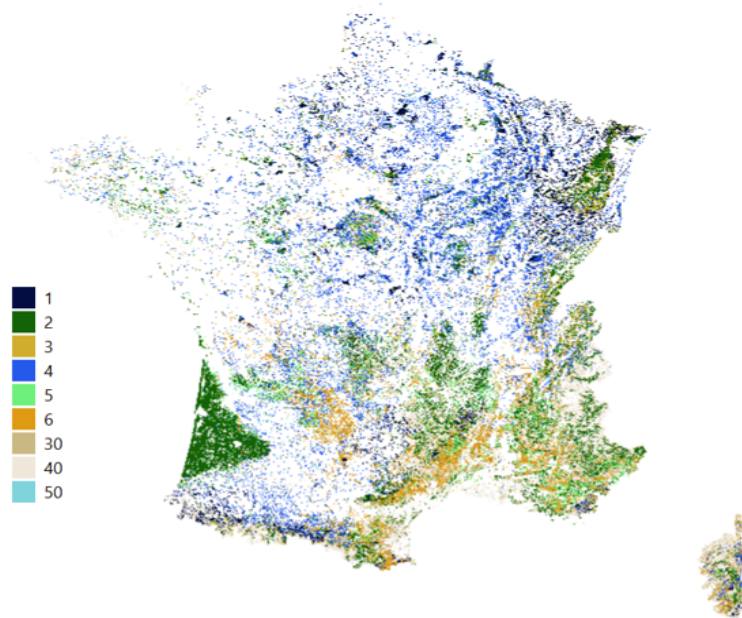


FIGURE II.16 – Types d'arbre par forêt en France métropolitaine : Carte

Sur la figure II.16, le vert représente les communes avec une majorité de conifères tandis que le bleu représente les feuillus et le jaune les mélanges. Cependant, comme il n'y a pas de forêts dans toutes les communes de France, il faut cette fois-ci extraire l'essence majoritaire non pas par commune mais par département (DATAGOUV [2022]). Le Bassin aquitain est fortement recouvert de forêts, tout comme une large partie est de la France métropolitaine.

Pour les trois communes prises en exemple au fil de ce mémoire, les données sont les suivantes :

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Numéro	1	40	2
Type de peuplement	Futaie de feuillus	Lande	Futaie de conifères
Classification	Feuillus	Mélange	Conifères

TABLE II.20 – Exemple de trois communes françaises : Essence d’arbre majoritaire

À Cudos en Gironde, les conifères sont majoritaires, comme dans la plupart de la zone du Bassin aquitain, qui est, pour rappel, la zone d’étude majeure de ce mémoire.

## II.4.5 Données relatives au vent

La question du vent est le cas le plus délicat. En effet, beaucoup de données sont disponibles mais peu s’adaptent au cas d’étude. Trouver l’agrégation la plus pertinente se révèle ainsi être un véritable défi. Les données SYNOP ne suffisent pas à couvrir l’ensemble des villes françaises, étant situées pour la plupart dans des emplacements excentrés et non représentatifs d’un territoire global.

Une option initialement envisagée consistait à utiliser le *Global Wind Atlas* (DTU WIND ENERGY [2025]), carte référençant le vent moyen et accessible en *open data*. Néanmoins, prendre la moyenne ne reflète pas assez la notion de tempête car les valeurs journalières du vent influent de manière trop forte sur le calcul. Barry Gardiner a alors transmis une carte raster représentant la vitesse du vent que l’on s’attendrait à voir dépasser une fois tous les 50 ans en moyenne en France. Cette carte correspondait globalement à ce qui était recherché mais elle présentait des écueils qui font qu’elle n’était pas exploitable. En effet, la donnée recherchée est un vent maximal mais ce n’est pas possible de trouver cette information à une maille aussi précise que la maille communale sur un historique assez long pour correspondre à la base de sinistres souhaitée. Cette information est trouvable pour quelques stations uniquement (SACRE et al. [2007]).

Les données sélectionnées ont donc été celles de Météo-France concernant le vent maximal par département pour chaque année entre 1999 et 2023 (L’INTERNAUTE [2024]). Les chiffres présentés sont estimés à partir des données des bulletins climatiques libres d’accès de Météo-France portant sur plus d’une centaine de stations météorologiques réparties dans toute la France métropolitaine. Cela est certes une perte de précision par rapport aux données communales recherchées. Cependant, par expérience, lorsqu’un vent maximal touche une commune, il touche également une vaste zone alentours dans des proportions similaires, spécialement dans le cadre des tempêtes. Le département est donc une maille tout à fait convenable pour cette variable. Ainsi, chaque sinistre peut être associé au vent maximal enregistré dans le département l’année de survenance, permettant ainsi de constituer un historique de 25 ans, ce qui est satisfaisant.

Pour ce qui est de chaque commune en France, la moyenne de ces maximums journaliers est calculée. La table les représentant est consultable en **Annexe D**. Le vent maximal moyen dans les trois communes analysées est relativement similaire. Les départements qui ressortent comme excédentaires sont l’Aveyron, le Gard, la Lozère et l’Hérault, représentés en rouge sur la figure II.17.

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Vent maximal moyen	106,6	112,56	121,4

TABLE II.21 – Exemple de trois communes françaises : Vent maximal moyen sur 25 ans

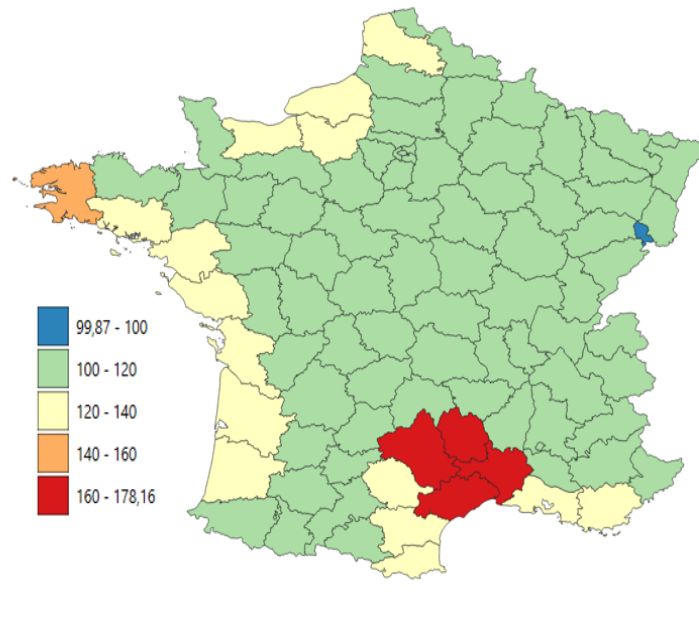


FIGURE II.17 – Vent maximal moyen sur 25 ans en France métropolitaine : Zonier

#### II.4.6 Prise de recul sur les données alternatives susceptibles d'enrichir l'étude

Les données mobilisées pour cette étude proviennent de sources scientifiques issues directement du domaine forestier, gage de fiabilité. Il aurait certes été possible d'intégrer d'autres variables telles que l'humidité du sol, la hauteur dominante des arbres ou encore la température pour affiner encore l'analyse. Toutefois, le jeu de variables explicatives déjà retenu constitue une base suffisante pour bâtir un modèle statistiquement solide et interprétable.

En outre, les recherches de K. F. Miller et l'ensemble des publications existantes à ce jour ne démontrent pas que l'inclusion d'autres variables améliore notablement les performances de modélisation du risque tempête. Leur absence n'affecte donc pas la validité de l'approche. Elle favorise à l'inverse la reproductibilité et la compréhension du modèle. Cela minimise aussi la diffusion d'incertitudes quant à la qualité des données.

## II.5 Synthèse de la deuxième partie et agrégation de cette base de données

Après avoir posé le contexte et les bases théoriques du risque tempête en assurance forêt dans la partie I, cette deuxième partie entendait opérer une bascule entre un concept initial et des données chiffrées relatives à l'étude qui soient à la fois opérationnelles et vérifiées.

Consolider une base unique pour l'ensemble des communes de France métropolitaine est un élément primordial. Cinq variables explicatives y figurent : vent maximal moyen, altitude moyenne, abri topographique, type de sol majoritaire et nature de la parcelle couverte (conifères/feuillus). Chaque variable provient de sources publiques ou de données d'un expert forestier puis est géocodée dans le logiciel QGIS. Ce passage permet de repérer d'éventuelles valeurs manquantes ainsi que de contrôler la cohérence spatiale.

Le croisement entre cette base nouvellement créée et la base de sinistres en possession s'effectue par le code INSEE. Ces deux bases de données (34 805 communes cartographiées et 2 701 sinistres) sont désormais le fondement du travail qui va être mené dans la partie suivante. Une modélisation mathématique par calcul du score comme dans l'étude de Miller va être testée. Celle-ci sera comparée à des modèles linéaires généralisées ainsi qu'à des modèles de *machine learning* afin de trouver le modèle correspondant le mieux aux données recensées. Le but recherché avec ces modèles est d'estimer un ratio de destruction pour chacune des communes françaises. Cela permettra de créer un zonier du niveau de risque qui soit le plus adapté à la réalité.

En synthèse, la base de données obtenue est composée de variables explicatives pertinentes pour notre cas d'étude. Cette base est directement opérationnelle pour les modélisations qui vont suivre.

# III - Modélisation du risque et construction d'un zonier à la maille communale

## III.1 Fonction et intérêt des zoniers

### III.1.1 Définition et avantages

Un zonier est une carte qui divise un territoire en zones homogènes, établies d'après leur niveau d'exposition au risque. Il se construit en général par garantie d'assurance (zonier inondation, zonier sécheresse ou zonier de déracinement des arbres dans le cadre de ce mémoire pour la garantie tempête).

Les zoniers ont été introduits en assurance dans un premier temps pour des raisons commerciales. Ils permettent d'éviter des écarts de tarification entre deux zones similaires pour une même variable, lorsque tous les autres critères sont égaux. Leur intérêt consiste également en la vérification de la suffisance de la diversification géographique, afin de se développer dans des endroits à faible risque et assurer sa rentabilité, même si la prime doit y être plus faible. Ainsi, les arguments présentés aux équipes commerciales reposent sur une justification mathématique solide, ce qui facilite le travail des agents généraux.

Ces agents généraux doivent par ailleurs être consultés une fois le zonier réalisé car ils ont une connaissance client issue du terrain qui permet de confirmer ou infirmer les résultats mathématiques obtenus par le zonier. Par exemple, si le zonier mène à une hausse massive des tarifs sur le portefeuille historique, le risque de perdre des clients importants se développe.

Par ailleurs, l'élaboration d'un zonier conduit à créer des classes de risque géographiques. Cet élément réduit donc le nombre de modalités par rapport à une modélisation où l'emplacement géographique serait intégré en tant que variable explicative.

En effet, dans toute démarche de zonage actuariel, la première étape consiste à isoler l'effet géographique avant de l'intégrer au tarif. Soit  $N_i$  le nombre de sinistres déclarés par l'individu  $i$  et  $e_i$  son exposition. Avec  $\lambda_i$  le paramètre d'une loi de Poisson, il est supposé que :

$$N_i \sim \mathcal{P}(\lambda_i),$$

et l'on calibre un modèle de fréquence de type MLG log-Poisson :

$$\ln(\mathbb{E}[N_i | x_i]) = \ln(e_i) + \beta_0 + x_i' \beta,$$

où  $x_i = (x_i^1, \dots, x_i^p)$  regroupe les variables tarifaires non géographiques. Le jeu de données est scindé en un échantillon d'apprentissage et un échantillon de validation plus petit, idéalement répartis de manière uniforme sur le territoire. Une fois le modèle ajusté sans variable géographique, les résidus peuvent être analysés afin de détecter l'effet spatial résiduel. C'est à partir de cette structure de dépendance qu'est ensuite construit le zonier.

Le choix de la technique de zonage dépend de l'objectif poursuivi, mais l'objectif demeure dans tous les cas d'estimer, pour chaque zone, une surfréquence ou une surprime géographique en cohérence avec le modèle initial.

### III.1.2 Zonier administratif

En réalité, l'ensemble des zoniers montrés dans la partie précédente correspond à un type particulier : les zoniers administratifs. Ces derniers sont issus d'un zonage par agrégation territoriale. Il est possible de retrouver des zoniers administratifs à la maille des pays, des régions, des départements (Figure II.17), des cantons ou, comme c'est le but dans ce mémoire, à la maille communale (Figure II.15).

Le zonier administratif est un dispositif de tarification qui regroupe les entités géographiques citées précédemment en un nombre fini de classes de risque homogènes. Trois grandes étapes se distinguent dans sa conception :

Tout d'abord, il est nécessaire d'isoler l'effet spatial. Cela s'effectue en calibrant tout d'abord un MLG (souvent log-Poisson) sans variable géographique ; les paramètres  $\hat{\beta}_{2,\dots,p}$  obtenus servent à extraire, pour chaque individu  $i$ , un risque spatial résiduel  $R_i$  et, pour chaque commune  $k$ , le risque spatial résiduel associé  $\hat{r}_k$ .

$$R_i = \frac{N_i}{e_i \exp(\hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip})}, \quad \hat{r}_k = \frac{\sum_{i \in k} e_i R_i}{\sum_{i \in k} e_i},$$

Cependant, si le maillage choisi est jugé trop faible, son résidu est remplacé par celui du niveau supérieur (canton, département ...). Un vecteur robuste  $\hat{r}_k$  est ainsi obtenu pour toutes les communes.

Au final, les communes sont ordonnées selon leur niveau de risque puis regroupées selon une méthode prédéfinie (quantiles, ruptures naturelles, intervalles égaux, etc.). La carte obtenue constitue le zonier.

### III.1.3 Élaboration d'un premier zonier par le calcul d'un score

Un premier zonier a été réalisé afin d'étudier la pertinence d'un modèle similaire à celui de *Windthrow Hazard Classification*. Pour vérifier cela, un score a été attribué aux cinq variables à la manière de l'étude britannique.

Le tableau français de comparaison entre le type de sol et le niveau de risque associé est présenté en **Annexe A**. Il est le fruit de recherches variées et d'une validation par Barry Gardiner. Par exemple, les histosols ont un risque à 9 car ils ne font souvent que quelques centimètres d'épaisseur et imposent donc un enracinement très superficiel. Au contraire, les cambisols ont un score de 3 car ce sont des sols qui peuvent atteindre une profondeur de plus d'un mètre. Ils sont bien structurés, présent dans les matériaux calcaires et donnent une grande cohésion. Une représentation visuelle de ces sols est fournie en **Annexe F**, afin d'étayer ces propos de manière concrète.

Les conifères ayant plus tendance à se déraciner que les feuillus, les scores attribués selon l'essence de l'arbre sont ceux de la table III.1. Des exemples d'arbres de ces deux essences sont également trouvables en **Annexe F** pour se rendre compte visuellement de ces affirmations.

Type majoritaire	Score
Conifères	10
Mélange	5
Feuillus	0

TABLE III.1 – Score associé à l’essence majoritaire : Étude menée

Pour le vent, l’altitude ainsi que l’abri topographique, les quantiles ont été calculés sur les données géographiques communales afin d’attribuer un score entre 1 et 10 selon l’intervalle dans lequel chaque commune se situe vis-à-vis de la variable, comme le montre la table III.2.

Score	Borne Max Vent	Borne Max TOPEX	Borne Max Altitude
1	107,696	7,53515726	56,1669523
2	115,5253333	10,1133667	90,57025761
3	123,3546667	12,55267	118,7746294
4	131,184	15,1447769	147,0525375
5	139,0133333	18,0405957	189,1077124
6	146,8426667	21,6437978	236,6933284
7	154,672	26,5946105	297,0225632
8	162,5013333	34,8411459	391,2437888
9	170,3306667	55,0963156	624,9016289
10	Au-delà	Au-delà	Au-delà

TABLE III.2 – Scores associés au vent, au TOPEX et à l’altitude : Étude menée

Pour les villes de Gosselming, Montsapey et Cudos, les scores suivants ont été obtenus :

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Score Sol	3	6	3
Score Essence	0	5	10
Score Vent	1	2	3
Score TOPEX	6	1	8
Score Altitude	7	10	3
<b>SCORE Total</b>	<b>15</b>	<b>25</b>	<b>27</b>

TABLE III.3 – Exemple de trois communes françaises : Synthèse des scores

Un score élevé est retrouvé à Cudos en Gironde, mais aussi en haute altitude par l'intermédiaire de la commune de Montsapey. En agrégeant à l'échelle nationale, les résultats globaux du score représentant le risque de déracinement des arbres sont obtenus sur le zonier de la figure III.1.

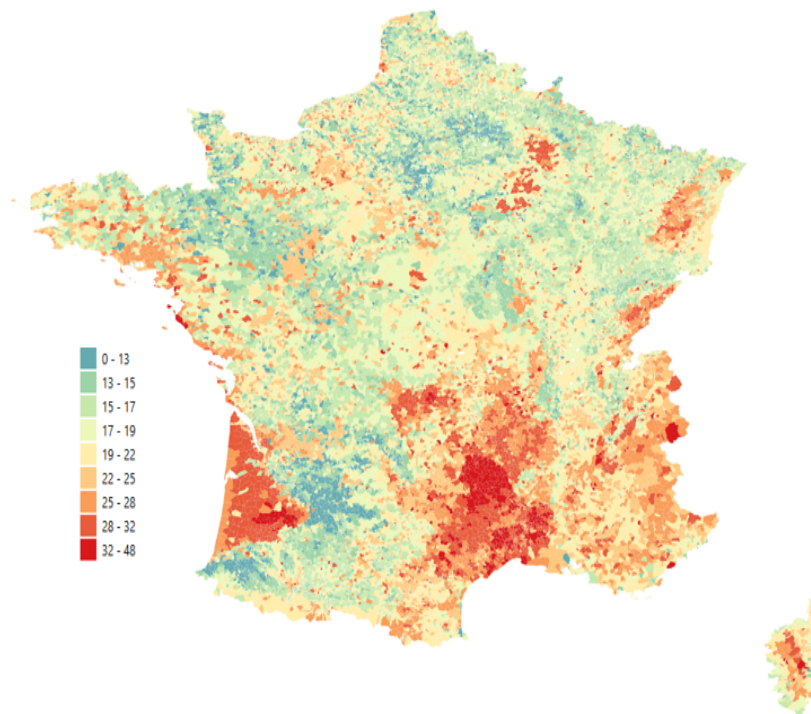


FIGURE III.1 – Zonier agrégé en 10 zones : Score

Ce zonier permet de tirer de premiers enseignements à propos du risque de déracinement des arbres en France métropolitaine. Les régions côtières figurent parmi les plus risquées, ce qui est un attendu vu les vents forts qui y soufflent et le peu d'abri qui s'y trouve. Les régions d'altitude sont également particulièrement à risque d'après ce zonier. Enfin, il est intéressant de noter que la zone majoritaire du portefeuille présente un risque majeur de déracinement dans ce premier modèle comparé au reste de la France métropolitaine.

L'intérêt est désormais de comparer ce modèle issu de l'étude britannique avec des modèles avancés tenant compte des interactions entre variables. Un MLG va, en plus de tenir compte de l'interaction, se baser sur un historique de sinistralité connu en lien avec l'assurance forêt.

## III.2 Modélisation du risque par Modèles Linéaires Généralisés (MLG)

### III.2.1 Introduction aux MLG

Un modèle linéaire généralisé (MCCULLAGH ET NELDER [1989]) est un outil permettant de modéliser la relation entre une variable que l'on cherche à expliquer, généralement notée  $Y$ , et

un ensemble de variables explicatives de dimension  $d$ . Pour chaque observation  $i = 1, \dots, n$ , on note le vecteur de covariables

$$x_i = (x_{i1}, \dots, x_{id})^\top \in \mathbb{R}^d, \quad \text{et} \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top \in \mathbb{R}^d.$$

Dans le cadre de ce mémoire, la variable réponse est le taux de déracinement des arbres par commune et les variables explicatives ont été présentées dans la partie II.

Le MLG est une généralisation du modèle linéaire « classique ». Il définit l'espérance de la variable à expliquer  $Y_i$  comme une combinaison linéaire des variables explicatives. Ceci se fait à l'aide d'une fonction de lien :

$$g(\mathbb{E}[Y_i]) = \beta_0 + \sum_{j=1}^d \beta_j x_{ij},$$

où  $g(\cdot)$  est une fonction réelle, strictement monotone et déterministe.

Un MLG repose principalement sur trois éléments :

1. **Une composante aléatoire** : Pour chaque observation  $i$ , la loi de  $Y_i$  appartient à la famille exponentielle, c'est-à-dire

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}.$$

Les moments vérifient

$$\mathbb{E}[Y_i] = b'(\theta_i) \quad \text{et} \quad \text{Var}(Y_i) = a(\phi) b''(\theta_i).$$

2. **Un prédicteur linéaire** :

$$\eta_i = x_i^\top \boldsymbol{\beta} + \beta_0 = \beta_0 + \sum_{j=1}^d \beta_j x_{ij}.$$

3. **Une fonction de lien** : Elle relie l'espérance de  $Y_i$  au prédicteur

$$g(\mathbb{E}[Y_i]) = \eta_i \quad \iff \quad \mathbb{E}[Y_i] = g^{-1}(\eta_i).$$

Les fonctions de lien usuelles pour certaines des grandes familles de MLG sont détaillées dans la table III.4.

Distribution	Lien	Fonction lien
Gamma	Inverse	$g(y) = \frac{1}{y}$
Normale	Identité	$g(y) = y$
Bernoulli	Logit	$g(y) = \ln\left(\frac{y}{1-y}\right)$
Poisson	Logarithme	$g(y) = \ln(y)$
Bêta	Logit	$g(y) = \ln\left(\frac{y}{1-y}\right)$

TABLE III.4 – MLG : Fonctions de lien classiques

Les paramètres  $\beta$  et  $\phi$  sont généralement estimés par maximum de vraisemblance, qui s'écrit :

$$L(\mathbf{y}; \theta, \phi) = \exp\left[\sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right],$$

La log-vraisemblance en découle :

$$\ell(\mathbf{y}; \theta, \phi) = \ln L(\mathbf{y}; \theta, \phi) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi).$$

Enfin, le maximum est obtenu en annulant pour tout  $j$  le gradient  $\partial \ell / \partial \beta_j$ .

Les critères d'informations AIC et BIC sont utilisés pour obtenir le modèle le plus pertinent possible. En comparant ces modèles, il faut minimiser ces critères :

$$\text{AIC} = -2\ell + 2k, \quad \text{BIC} = -2\ell + k \ln(n),$$

où  $k$  est le nombre de paramètres et  $n$  le nombre d'observations.

De plus, plusieurs indices sont analysés pour évaluer la qualité de l'ajustement du modèle. La déviance mesure la qualité d'ajustement du modèle en comparant la vraisemblance du modèle considéré à celle d'un modèle saturé (modèle reproduisant exactement les données). Une déviance faible indique un bon ajustement. La dispersion permet elle d'évaluer si la variance observée est conforme à celle attendue par le modèle. Une sur-dispersion ou une sous-dispersion peut indiquer une mauvaise spécification.

Le  $R^2$  est généralement utilisé pour évaluer la capacité prédictive d'un modèle. Toutefois, ce coefficient n'est pas toujours défini. Pour l'étude de certains MLG, une approximation courante est le pseudo- $R^2$  de McFadden, défini comme :

$$R_{McFadden}^2 = 1 - \frac{\ell_{\text{modèle}}}{\ell_{\text{nul}}},$$

où  $\ell_{\text{modèle}}$  représente la log-vraisemblance du modèle ajusté, et  $\ell_{\text{nul}}$  celle du modèle nul (modèle sans aucune variable explicative). Le modèle est meilleur si cette valeur est proche de 1.

Un  $R^2$  élevé signifie un bon ajustement sur l'échantillon d'apprentissage, mais n'implique ni causalité ni bonne performance prédictive. Par exemple, un  $R^2$  de 0,4 indique que sur l'échantillon d'ajustement, le modèle explique environ 40 % de la variabilité du ratio de destruction, la variable à expliquer. Ce niveau peut être acceptable ou non selon le domaine et on l'interprète surtout en comparaison de modèles concurrents.

Dans le cadre de l'étude en cours, qui consiste à estimer un ratio compris entre 0 et 1, deux distributions semblent particulièrement adaptées : un MLG Bêta ainsi qu'un MLG quasi-binomial.

### III.2.2 Gestion en cas de données déséquilibrées

Avant d'étudier ces modèles, il convient de réaliser un focus sur les outils d'évaluation de la performance en présence de données déséquilibrées. En effet, dans les données utilisées, les ratios sont plus regroupés aux alentours de 0 % que de 100 %, avec des valeurs en dessous de 50 % majoritaires, comme le montre la table III.5. Le *boxplot* présenté en figure III.2 montre la

répartition des ratios plus globalement, d'où il est possible de tirer une tendance basse. Afin de ne pas perdre de pouvoir prédictif pour les ratios prédits élevés, il est nécessaire d'utiliser des métriques d'utilisation qui prennent en considération cet aspect de déséquilibre binaire dans l'évaluation.

Nombre de ratios inférieurs à 0,5	Nombre de ratios supérieurs ou égaux à 0,5
1489	715

TABLE III.5 – Ratio de sinistres : Répartition

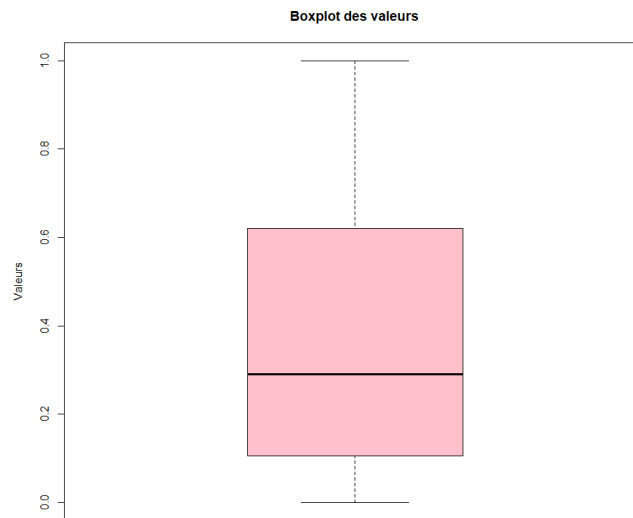


FIGURE III.2 – Ratios de destruction de la base de sinistres : *Boxplot*

Les méthodes usuelles d'évaluation de la prédiction passent en général par le calcul du taux de vrais positifs et de vrais négatifs, ainsi que par la représentation de courbes ROC qui donne le taux de vrais positifs par rapport au taux de faux positifs. Ces méthodes peuvent présenter des résultats biaisés lorsque les classes sont déséquilibrées. La solution envisagée dans le cas binaire présent est alors d'utiliser le  $F_1$ -score ainsi que les courbes Précision-Rappel (PRC), dont les calculs sont donnés ci-après. (HERANVAL et al. [2022])

$$\text{Précision}(p_c) = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}}$$

$$\text{Rappel}(p_c) = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Négatifs}}$$

où  $p_c$  est une probabilité critique variant entre 0 et 1.

La précision quantifie le nombre de prédictions positives correctes parmi toutes les prédictions positives effectuées, tandis que le rappel quantifie le nombre de prédictions positives correctes parmi toutes les prédictions positives qui auraient pu être effectuées. Grâce à ces deux indicateurs, le  $F_1$ -score est obtenu. Il s'écrit :

$$F_1 = 2 \cdot \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

Plus le  $F_1$ -score est proche de 1, plus le modèle de prédiction est performant.

La courbe PRC affiche les valeurs de Précision et de Rappel lorsque la probabilité critique  $p_c$  varie entre 0 et 1. La courbe PRC d'un modèle satisfaisant tend vers le coin supérieur droit. À l'inverse, la courbe d'un modèle non adapté correspond à une droite horizontale dont l'ordonnée à l'origine est proportionnelle au nombre de positifs dans l'ensemble de données. Pour un jeu de données équilibré, cette valeur tourne généralement autour de 0,5. (THOMAS [2024])

### III.2.3 Premiers essais et MLG quasi-binomial

Le premier modèle à être testé est un MLG Bêta. En effet, la loi Bêta est particulièrement adaptée pour modéliser des variables comprises entre 0 et 1, ce qui en fait une première intuition de modélisation naturelle pour les ratios de destruction forestière. Après avoir exclu les valeurs aberrantes ou manquantes (ratio supérieur à 1, valeurs *NA* des différentes variables...) et mis en forme les variables, le but est d'expliquer la variable de déracinement, notée *Ratio\_SinAss*, à l'aide des différentes variables explicatives. Via R, il est également possible de modéliser des interactions entre les variables explicatives.

Néanmoins, le MLG Bêta, tout comme d'autres modèles qui ont été testés, s'avère inadapté à notre étude. Dans ce cas, le modèle est trop complexe par rapport à la taille de l'échantillon disponible. Différentes transformations ont été envisagées afin d'obtenir des résultats cohérents, telles que des transformations logarithmiques et racine carrée des variables, l'ajout d'interactions supplémentaires ou de termes polynomiaux, des fonctions liens alternatives (probit) ou encore un regroupement de sols en cinq groupes afin de réduire les modalités. Des valeurs plus cohérentes ont été trouvées mais les résultats n'ont pas permis d'améliorer significativement l'ajustement.

Cette dernière intuition se confirme en regardant les valeurs prédites de ratio pour les communes françaises. La majorité des ratios estimés valent 0 % ou 100 %, ce qui n'est pas cohérent avec la réalité. Une solution aurait pu être d'étoffer la base avec du *data pooling* ou de la création de nouvelles données fictives notamment mais tester de nouveaux modèles semble prioritaire. Par conséquent, un modèle quasi-binomial est considéré dans la sous-partie suivante. Cela semble être un bon compromis dans ce contexte de données limitées.

Un MLG quasi-binomial, contrairement à une loi binomiale, conserve la même moyenne mais estime un paramètre  $\phi$  supplémentaire pour corriger la sur-dispersion ( $\phi > 1$ ) ou la sous-dispersion ( $\phi < 1$ ) sans changer l'interprétation des coefficients. On l'emploie principalement dans des problématiques de proportion ou une modélisation binomiale serait trop stricte, comme dans ce cas d'étude. Dans un MLG quasi-binomial, la fonction lien la plus fréquemment utilisée est la fonction logit, la relation du schéma III.3 est donc à l'étude dans cette partie :

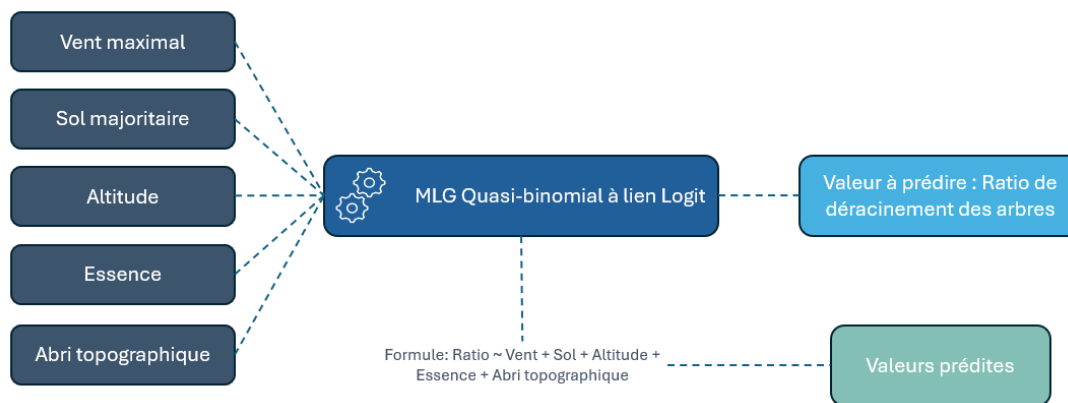


FIGURE III.3 – MLG Quasi-binomial : Schéma explicatif

Une nouvelle fois, le ratio de déracinement est expliqué dans ce modèle par les variables explicatives comme l'altitude ou encore le type de sol majoritaire. Une analyse ANOVA (Table III.6) permet d'analyser si les différentes variables ont un véritable apport ou non.

Variable	Df	Déviante	$p$ -valeur	Signif.
Wind_Max25y	1	29,7352	$1,64 \times 10^{-15}$	***
Soil_Majority	13	31,5766	$2,42 \times 10^{-9}$	***
Elevation_Mean	1	1,9979	$3,89 \times 10^{-2}$	*
Conifer_Bool	1	8,1142	$3,17 \times 10^{-5}$	***
Topex5k	1	0,7391	$2,09 \times 10^{-1}$	

TABLE III.6 – MLG Quasi-binomial : ANOVA

En analysant cette table, chacune des variables est bien significative, hormis l'abri topographique qui est très corrélé avec l'altitude, et apporte de la compréhension au modèle dans son ensemble. Ce modèle est donc positif de ce point de vue.

De plus, les indices GVIF (Table III.7) pondérés par les degrés de liberté sont tous inférieurs à 2. L'indice le plus élevé, à savoir 1,62 pour Elevation\_Mean, reste très inférieur aux valeurs critiques de 5 ou 10 qui signaleraient une multicolinéarité élevée. Il n'y a donc pas d'information redondante susceptible de dégrader la précision des coefficients et il n'est donc pas nécessaire d'éliminer ou de transformer l'une des variables à ce stade.

Variable	GVIF	Df	$GVIF^{1/(2Df)}$
Wind_Max25y	1,2147	1	1,1021
Soil_Majority	2,9580	13	1,0426
Elevation_Mean	2,6199	1	1,6186
Conifer_Bool	1,2939	1	1,1375
Topex5k	2,2743	1	1,5081

TABLE III.7 – MLG Quasi-binomial : Indices de variance d'inflation généralisés

La table III.8 montre une légère sous-dispersion : la variance observée est légèrement plus faible

que sous une loi binomiale équivalente. Les tests sont donc un peu plus conservateurs, mais le quasi-binomial reste adapté : il corrige la dispersion sans modifier les coefficients significatifs.

Déviance	Degrés de liberté	Dispersion
1232,836	2186	0,5639

TABLE III.8 – MLG Quasi-binomial : Déviance et dispersion

Le pseudo- $R^2$  de McFadden est de 0,11529 et ajouter des interactions n’améliore pas le modèle. Malgré la hausse par rapport au modèle Bêta, cela reste une valeur modérée, ce qui s’explique par la taille d’échantillon limitée et l’hétérogénéité du phénomène. Un objectif par la suite est néanmoins de chercher à augmenter le pseudo- $R^2$  avec des modèles de *machine learning* notamment.

Les modalités d’une approche quasi-binomiale ainsi que les coefficients obtenus sont répertoriés dans la table III.9.

Variable	Estimate	Std. Error	$t$	Pr(>   $t$  )
(Intercept)	-3,3313	0,4247	-7,848	< $10^{-6***}$
Wind_Max25y	0,0146	0,00265	5,495	$4,37 \times 10^{-8***}$
Soil_Majority10	0,3068	0,1717	1,787	0,073
Soil_Majority11	0,6440	0,1969	3,277	0,0011**
Soil_Majority17	0,7310	0,1458	5,031	$2,9 \times 10^{-7***}$
Soil_Majority21	0,6383	0,2235	2,854	0,0043**
Soil_Majority22	1,2192	0,2027	6,011	< $10^{-6***}$
Soil_Majority24	0,3547	0,1534	2,314	0,021 *
<i>(...autres modalités de Soil_Majority omises pour la lisibilité)</i>				
Elevation_Mean	-0,00019	0,00027	-0,63	0,53
Conifer_Bool	0,00457	0,00114	4,01	$6,6 \times 10^{-5***}$
Topex5k	-0,00502	0,00409	-1,24	0,216

TABLE III.9 – MLG Quasi-binomial : Coefficients

Contrairement au modèle basé sur les scores, l’altitude fait ici diminuer le niveau de risque. En effet, l’estimateur est de -0,00019, ce qui veut dire qu’un mètre d’altitude fait perdre cette valeur au logit du ratio prédit. Le MLG apporte alors de la nuance et du recul par rapport aux premières intuitions. En haute altitude, les arbres se déracinent moins. Cela peut être expliqué par plusieurs facteurs : une meilleure acclimatation dans le temps, un sol plus robuste ou encore un abri topographique plus élevé du fait du relief.

Les signes des autres estimateurs vont dans le sens de la première modélisation et confirment l’intuition. Le vent accroît le risque de déracinement, tandis qu’un abri topographique prononcé le réduit ; ces tendances vont bien dans le sens attendu. Par ailleurs, les conifères se déracinent plus fréquemment que les feuillus, un constat bien établi en sylviculture. Enfin, par rapport à la modalité de référence, qui est le cambisol (n° 8), seuls deux types de sol diminuent encore davantage le risque de déracinement : le regosol (n° 22), caractérisé par une roche meuble, et le solonchak (n° 29), à forte salinité. Les sols présentant le risque le plus élevé sont l’histosol

(n° 19), riche en débris végétaux, et le ranker (n° 25), un sol siliceux. Les résultats observés confirment les intuitions initiales et les images de l'**Annexe F** viennent également appuyer visuellement ces résultats.

À l'aide des coefficients de la table III.9, il est possible d'obtenir le logit du ratio de destruction via le calcul suivant :

$$\begin{aligned} \text{logit}(\text{Ratio\_SinAss}) = & -3,3331944 + 0,0145379 \text{Wind\_Max25y} - 0,0050717 \text{Topex5k} \\ & + 0,3607925 \mathbf{1}_{\{\text{Soil\_Maj}=10\}} + 0,6449119 \mathbf{1}_{\{\text{Soil\_Maj}=11\}} + 0,4880385 \mathbf{1}_{\{\text{Soil\_Maj}=12\}} \\ & + 0,2147141 \mathbf{1}_{\{\text{Soil\_Maj}=15\}} + 0,7333020 \mathbf{1}_{\{\text{Soil\_Maj}=17\}} + 1,2711157 \mathbf{1}_{\{\text{Soil\_Maj}=19\}} \\ & + 0,2294820 \mathbf{1}_{\{\text{Soil\_Maj}=20\}} + 0,6383275 \mathbf{1}_{\{\text{Soil\_Maj}=21\}} - 0,1347833 \mathbf{1}_{\{\text{Soil\_Maj}=22\}} \\ & + 0,4105515 \mathbf{1}_{\{\text{Soil\_Maj}=24\}} + 2,3524699 \mathbf{1}_{\{\text{Soil\_Maj}=25\}} + 0,3999421 \mathbf{1}_{\{\text{Soil\_Maj}=27\}} \\ & - 0,5321116 \mathbf{1}_{\{\text{Soil\_Maj}=29\}} - 0,0004192 \text{Elevation} + 0,4496387 \mathbf{1}_{\{\text{Essence}=\text{Conifères}\}} \end{aligned}$$

Dans la base de données historique, il n'y a pas de cas dont l'essence majoritaire est un mélange. Cette composante n'intervient donc pas dans le MLG. Comme les feuillus, modalité de référence, entraînent une augmentation de 0 du  $\text{logit}(\text{Ratio\_SinAss})$  et que les conifères entraînent une augmentation de 0,4496387, on émet l'hypothèse que le mélange entraîne une hausse à hauteur de la moyenne de ces deux facteurs, à savoir 0,2248193. L'expression finale est donc :

$$\text{logit}(\text{Ratio\_SinAss}) = -3,3331944 + (\dots) + 0,4496387 \mathbf{1}_{\{\text{Essence}=\text{Conifères}\}} + 0,2248193 \mathbf{1}_{\{\text{Essence}=\text{Mélange}\}}$$

Le ratio seul est lui obtenu via le calcul simple suivant :

$$\text{Ratio\_SinAss} = \frac{1}{1 + e^{-\text{logit}(\text{Ratio\_SinAss})}}$$

Il est désormais possible d'obtenir le ratio pour toutes les communes de France. Les statistiques sur les ratios obtenus sont détaillées dans la table III.10 et les ratios des trois communes prises en exemple sont détaillées dans la table III.11.

Moyenne	Médiane	1 <sup>er</sup> quartile	3 <sup>e</sup> quartile	Min	Max
19,27 %	17,18 %	13,15 %	24,20 %	3,11 %	82,62 %

TABLE III.10 – MLG Quasi-binomial : Statistiques descriptives de *Ratio\_SinAss*

Gosselming	Montsapey	Cudos
Moselle 12,22 %	Savoie 7,31 %	Gironde 38,07 %

TABLE III.11 – MLG Quasi-binomial : Ratio de destruction - Exemple de trois communes françaises

Pour comprendre plus naturellement cette formule, voici l'explication du ratio calculé pour Gosselming, Montsapey et Cudos :

La table III.12 rappelle tout d'abord les valeurs des données utiles pour chacune de ces trois communes. Les calculs pour les trois communes sont ensuite présentés, avec des coefficients arrondis à trois décimales pour une meilleure lisibilité.

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Altitude	255,22	1569,33	110,42
Sol majoritaire	Cambisol (8)	Podzol (20)	Luvisol (17)
Abri topographique	15,90	125,99	10,87
Vent maximal moyen	106,60	112,56	121,40
Essence	Feuillus	Mélange	Conifères

TABLE III.12 – Exemple de trois communes françaises : Données

### Exemple de Gosselming :

$$\begin{aligned}
 \text{logit}(\text{Ratio\_SinAss}) &= -3,333 + 0,015 \times 106,60 - 0,005 \times 15,90 \\
 &\quad - 0,0004 \times 255,22 + 0,450 \times 0 \\
 &= -\mathbf{1,971}
 \end{aligned}$$

Et ainsi, par *logit* inverse, le ratio de destruction vaut **12,22%** pour Gosselming.

### Exemple de Montsapey :

$$\begin{aligned}
 \text{logit}(\text{Ratio\_SinAss}) &= -3,333 + 0,015 \times 112,56 - 0,005 \times 125,99 \\
 &\quad + 0,229 \times 1 - 0,0004 \times 1569,33 + 0,225 \times 1 \\
 &= -\mathbf{2,539}
 \end{aligned}$$

Et ainsi, par *logit* inverse, le ratio de destruction vaut **7,31%** pour Montsapey.

### Exemple de Cudos :

$$\begin{aligned}
 \text{logit}(\text{Ratio\_SinAss}) &= -3,333 + 0,015 \times 121,40 - 0,005 \times 10,87 \\
 &\quad + 0,733 \times 1 - 0,0004 \times 110,42 + 0,450 \times 1 \\
 &= -\mathbf{0,487}
 \end{aligned}$$

Et ainsi, par *logit* inverse, le ratio de destruction vaut **38,07%** pour Cudos.

Les ratios prédits sont majoritairement concentrés entre 10 et 25 %. Néanmoins, pour certaines zones plus risquées d'après les variables, ce ratio de destruction explose, comme dans le cas de Cudos en Gironde.

La capacité prédictive s'évalue via le calcul de la précision et du rappel. Le ratio prédit via la formule de Ratio\_SinAss est comparée avec le ratio présent dans la base de données initiale. En fixant le seuil à 50 %, afin de séparer en classification binaire les risques faibles et les risques élevés, les résultats de la table III.13 sont obtenus.

Mesure	Effectif	% du total
Vrais Positifs (VP)	328	14,9 %
Vrais Négatifs (VN)	1432	65,0 %
Faux Positifs (FP)	338	15,3 %
Faux Négatifs (FN)	106	4,8 %
<b>Total</b>	<b>2204</b>	<b>100,0 %</b>

TABLE III.13 – MLG Quasi-binomial : Effectifs et pourcentages de la matrice de confusion

Cela permet de distinguer efficacement les valeurs prédites correctement, qui représentent 80 % des valeurs dans ce cas. L'ensemble de ces valeurs trouvent leur analyse dans le calcul de la précision, du rappel et du  $F_1$ -score, obtenus dans la table III.14 :

Mesure	Valeur
Précision	0,492
Rappel	0,756
$F_1$ -score	0,596

TABLE III.14 – MLG Quasi-binomial : Mesures globales de performance

La précision de 0,492 indique que sur 100 prédictions « risque élevé », 49 sont correctes et le rappel de 0,756 indique que 76 % des vraies communes à risque élevé sont détectées. C'est l'objectif poursuivi en fixant le seuil à 50 % : le modèle capture très bien les seuils élevés mais a tendance à sur-alerter. Le  $F_1$ -score de 0,596 indique un aspect prédictif équilibré correct, qui peut néanmoins être amélioré.

L'aire sous la courbe (AUC) résume, sur tous les seuils de décision, le compromis entre rappel et précision, une valeur plus élevée indiquant que le modèle détecte mieux les classes positives tout en limitant les faux positifs. La courbe précision-rappel (Figure III.4) montre une AUC de 0,584, ce qui indique une capacité de discrimination modérée de la part du modèle, qui peut être améliorée. La ligne pointillée rouge est égale à la prévalence de la classe positive. L'AUC n'est que modestement supérieur par rapport à l'aléatoire et à cette *baseline*.

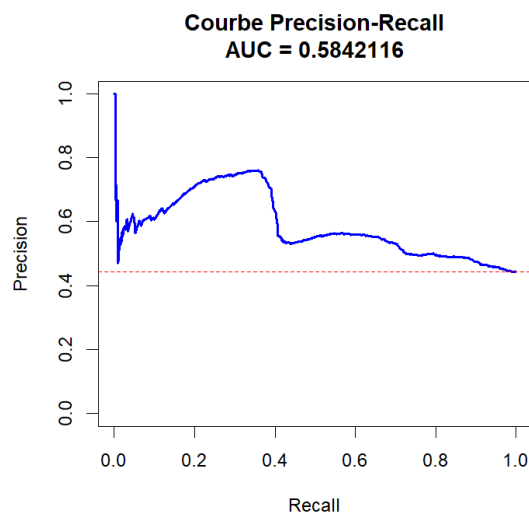


FIGURE III.4 – MLG Quasi-binomial : Courbe précision-rappel

### III.2.4 *Backtesting* et validation des modèles

Avant de présenter les résultats, rappelons ce qu'est le *backtesting* et la validation croisée :

Le *backtesting* sert à évaluer un modèle sur des données non utilisées à l'estimation pour vérifier la stabilité et la calibration. La validation croisée à  $K$  plis sert, quant à elle, à partitionner les données en  $K$  sous-ensembles pour, à chaque itération, apprendre sur  $K-1$  plis et tester sur le pli restant.

Le *backtesting* a ici pour but de vérifier la robustesse et la validité du MLG quasi-binomial. Il est nécessaire de s'assurer que le modèle possède à la fois une capacité de généralisation satisfaisante, une stabilité dans le temps ainsi qu'une calibration adaptée. Cela est essentiel d'un point de vue actuariel afin d'être sûr que les coefficients estimés sont fiables et que le modèle puisse être déployé opérationnellement pour cartographier un zonier tempête en assurance forêt.

Quelques éléments ont été mis en place afin d'effectuer cette validation :

- Tout d'abord, une **validation croisée** à 10 plis a été implémentée pour évaluer la capacité de généralisation du modèle. L'échantillon a été séparé en dix sous-ensembles tout en s'assurant une répartition équitable des ratios faibles et élevés.
- Les données ont été **divisées en deux groupes** selon une répartition 75 %/25 %. 75 % des données sont entraînées et 25 % servent de test afin d'évaluer la capacité de généralisation du modèle.
- 500 itérations de **bootstrap** ont été effectuées pour évaluer la stabilité des performances du modèle et construire des intervalles de confiance.
- Des tests de diagnostic statistique ont été menés comme des tests de normalité des résidus (Shapiro-Wilk) et d'homoscédasticité afin de vérifier les hypothèses du MLG.

Les résultats démontrent une performance cohérente du modèle quasi-binomial selon différentes méthodes d'évaluation :

Métrique	Validation Croisée	Test Final	Bootstrap
Pseudo $R^2$	$0,1155 \pm 0,0357$	0,1237	$0,1146 \pm 0,0117$
MSE	$0,105954 \pm 0,003957$	0,104492	$0,105406 \pm 0,001310$
MAE	$0,273781 \pm 0,005874$	0,276980	$0,277792 \pm 0,001924$
RMSE	$0,325455 \pm 0,006079$	0,323253	–

TABLE III.15 – MLG Quasi-binomial : Validation par trois stratégies d'évaluation

La cohérence entre les performances en validation croisée ( $R^2 = 0,1155$ ) et sur l'ensemble de test final ( $R^2 = 0,1237$ ) avec un écart de 0,0082, démontre une bonne capacité de généralisation du modèle. De plus, la performance bootstrap ( $R^2 = 0,1146$ ) est cohérente avec la validation croisée.

Le test de Shapiro–Wilk ( $W = 0,9373$ ,  $p = 1,5 \times 10^{-9}$ ) rejette la normalité des résidus, comme attendu : la variable étant bornée dans  $[0, 1]$ , les résidus d'un MLG quasi-binomial ne suivent pas une loi normale. Des intervalles de confiance robustes s'imposent donc. La corrélation |résidus|/prédictions vaut 0,0754 ( $p = 0,173$ ), ne signalant pas d'hétéroscédasticité marquée. C'est un élément positif pour le modèle

La corrélation prédictions/observations demeure compatible avec les données actuarielles forestières françaises. La MSE (0,104 492) et la MAE (0,276 980) indiquent une précision satisfaisante au regard de la complexité du risque tempête. Le quasi-binomial constitue ainsi une base soutenable pour le zonier, sous réserve des limites inhérentes aux phénomènes naturels complexes.

L'élément qui va faire qu'il est primordial d'étudier au-delà des modèles linéaires généralisés est que, en plus du  $F_1$ -score et de l'AUC qui sont toujours à améliorer, le MLG a tendance à sous-estimer les ratios prédits, mis en avant par la table III.16. Cela peut être expliqué du fait du nombre important de sinistres avec une faible parcelle sinistrée.

Moyenne Ratio Prédit MLG	Moyenne sinistralité historique
19,27 %	32,33 %

TABLE III.16 – MLG Quasi-binomial : Moyenne prédite vs Moyenne observée du ratio

Bien que cela n'ait pas d'impact direct sur la tarification car la séparation du zonier se fera par classes de risques à définir, un modèle mieux ajusté aux données de sinistres est requis pour garantir à la fois calibration et pouvoir discriminant. Des variantes de MLG (Bêta-inflaté, Tweedie) ont été testées sans amélioration notable des critères. En prolongement logique, la section suivante s'intéresse aux modèles de *machine learning*, susceptibles d'offrir une flexibilité accrue face aux non-linéarités et aux interactions.

### III.3 Modélisation du risque par Modèles *machine learning*

#### III.3.1 Introduction aux modèles *machine learning*

Le *machine learning*, traduit en français par apprentissage automatique, est un domaine scientifique qui consiste à laisser les algorithmes trouver des motifs récurrents dans les données qui lui sont fournies. Avec ces motifs identifiés, il est alors possible de prédire des comportements futurs et de mieux prendre en compte les différents éléments environnants. Défini pour la première fois par l'informaticien Arthur Samuel en 1959, le *machine learning* a connu une expansion fulgurante, en particulier depuis 2010.

Le *machine learning* aide à proposer des solutions pratiques là où les méthodes classiques s'arrêtent, à condition de disposer de données pertinentes. Il peut aussi donner des intuitions pour des modèles que l'on sait résoudre mais qui sont difficiles à formaliser en termes algorithmiques. Enfin, il peut aider dans le cas de problèmes que l'on sait résoudre, mais qui ne sont pas optimisés. Le *machine learning* est donc utilisé quand les données sont disponibles mais les connaissances sur le sujet peu accessibles ou peu développées (AZENCOTT [2022]). C'est précisément le cas dans ce mémoire.

Au sein du *machine learning*, l'apprentissage supervisé a pour but d'apprendre à faire des prédictions, à partir d'une liste d'exemples étiquetés, c'est-à-dire accompagnés de la valeur à

prédire. Une représentation simplifiée de ce qu'est l'apprentissage supervisé est donnée dans la figure III.5.

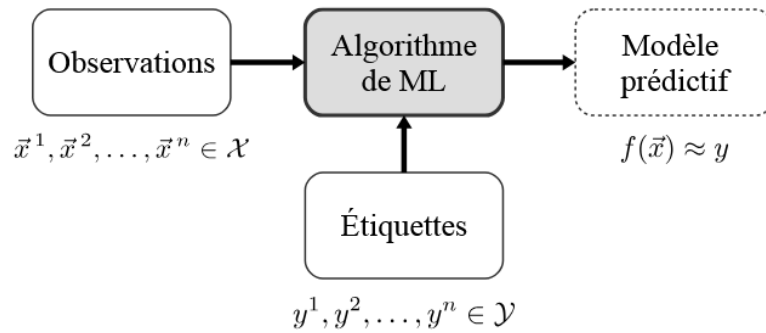


FIGURE III.5 – Apprentissage supervisé : Fonctionnement (AZENCOTT [2022])

Parmi les types de modèles utilisant l'apprentissage supervisé se distinguent principalement les algorithmes *Random Forest* et *XGBoost*. Ces algorithmes sont particulièrement adaptés aux problématiques actuarielles car ils gèrent efficacement les données déséquilibrées, fréquentes en assurance, comme dans le périmètre d'étude. Contrairement aux MLG qui reposent sur des hypothèses distributionnelles strictes, les méthodes de *machine learning* offrent une plus grande flexibilité pour capturer des relations non-linéaires complexes entre variables explicatives. Cette capacité peut s'avérer primordiale dans le contexte forestier où les interactions entre facteurs environnementaux (vent, sol, topographie) peuvent présenter des effets de seuil difficiles à modéliser par des approches plus classiques. De plus, ces algorithmes intègrent naturellement des mécanismes de validation croisée et de limitation du surajustement.

### III.3.2 *Random Forest*

*Random Forest* a été introduit par Breiman (BREIMAN [2001]), afin de pallier les inconvénients des arbres CART, tout en préservant leurs avantages. Comme indiqué par son nom, une forêt aléatoire est un ensemble d'arbres. Les forêts aléatoires sont basées sur 3 principes fondamentaux, qui visent à atteindre les principes de diversité et d'indépendance :

- Les prédicteurs combinés sont des arbres CART.
- Chaque arbre est entraîné sur un jeu de données différent, obtenu par tirage aléatoire avec remise de l'ensemble d'entraînement.
- Lors de la construction de chaque arbre, à chaque itération, le critère CART est optimisé seulement sur les coupures faisant intervenir des variables explicatives tirées au hasard.

Dans la plupart des cas, *Random Forest* gère mieux le déséquilibre des données que d'autres modèles comparables car chaque arbre est entraîné sur un échantillon différent, ce qui peut naturellement rééquilibrer les classes. D'un point de vue pratique, le nombre d'arbres  $B$  ainsi que le nombre de variables explicatives choisies aléatoirement sont des hyperparamètres, qui sont à optimiser par validation croisée. Leurs valeurs par défaut sont robustes, par exemple  $B = 500$  pour le nombre d'arbres.

Dans notre cas, la base de données initiale a été soumise au même filtrage que pour les MLG afin de garantir la qualité des données. Dans le cadre des modèles *machine learning*, le mélange

ne peut être une solution pour l'essence étant donné qu'il n'y en a pas dans la base de test et les variables concernées ont été considérées comme feuillus, ce qui influe peu sur le résultat final, comme le montre la table G.1 et facilite le travail. De plus, l'optimisation des hyperparamètres s'est faite par des méthodes variées, détaillées en **Annexe G**.

La configuration optimale retenue pour les hyperparamètres est la suivante :

- $B = 1000$  arbres
- $mtry = 2$  variables par division
- $nodesize = 5$  observations minimales par nœud terminal

La table III.17 démontre que le modèle ainsi optimisé améliore les performances du MLG quasi-binomial.

Modèle	RMSE	$R^2$
MLG Quasi-binomial	0,3254	0,1155
<i>Random Forest</i> optimisé	0,2683	0,3502

TABLE III.17 – Comparaison des performances : MLG vs *Random Forest*

*Random Forest* apporte un gain net. Les niveaux ( $R^2 = 0,35$ ,  $RMSE = 0,27$ ) sont raisonnables pour un phénomène très bruité même s'il reste une marge d'amélioration importante.

Quant à l'évaluation de la capacité prédictive du modèle *Random Forest*, la même méthode que précédemment est employée et les résultats de la table III.18 sont obtenus pour l'évaluation des classes en dessous et au-dessus du seuil 50 %. *Random Forest* prédit bien mieux les valeurs au-dessus du seuil, donc les valeurs hautes.

Mesure	Effectif	% du total
Vrais Positifs (VP)	522	23,7 %
Vrais Négatifs (VN)	1345	61,0 %
Faux Positifs (FP)	144	6,5 %
Faux Négatifs (FN)	193	8,8 %
<b>Total</b>	<b>2204</b>	<b>100,0 %</b>

TABLE III.18 – *Random Forest* : Effectifs et pourcentages de la matrice de confusion

Là encore, ces valeurs trouvent leur analyse dans le calcul de la précision, du rappel et du  $F_1$ -score, obtenus dans la table III.19 :

Mesure	Valeur
Précision	0,784
Rappel	0,730
$F_1$ -score	0,756

TABLE III.19 – *Random Forest* : Mesures globales de performance

La capacité prédictive est nettement améliorée par rapport au modèle précédent, confirmant l'intérêt de l'étude de ce modèle. La courbe précision-rappel (Figure III.6) montre une AUC de 0,823, ce qui indique une capacité de discrimination bien meilleure comparée à l'aléatoire et au MLG. La courbe précision-rappel montre que le modèle *Random Forest* surpasse significativement un modèle simple 0/1 (ligne horizontale en pointillés) ainsi que l'aléatoire. Le modèle démontre donc une capacité prédictive robuste sur l'ensemble de la plage, validant ainsi son utilité pratique pour prédire le ratio sinistres/assurés.

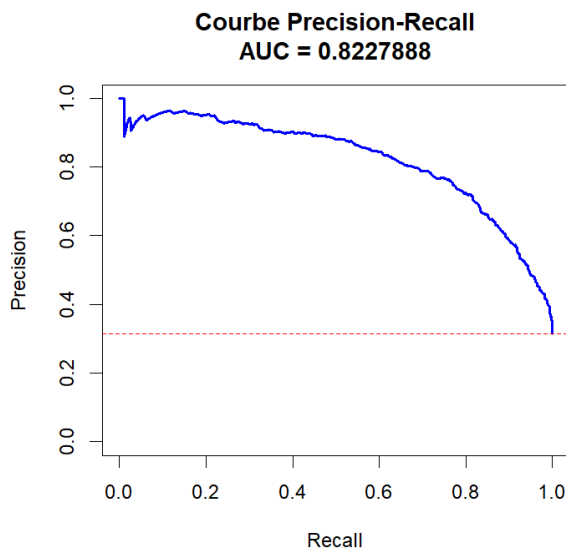


FIGURE III.6 – *Random Forest* : Courbe précision-rappel

Reste à comparer maintenant *Random Forest* avec l'autre modèle majeur du *machine learning* : *XGBoost*, afin de savoir quel modèle garder pour les prédictions futures des ratios par commune.

### III.3.3 *XGBoost*

*XGBoost*, présenté en détail par Chen et Guestrin en 2016 (CHEN ET GUESTRIN [2016]), est une alternative au modèle *Random Forest*, qui appartient à la famille des modèles de *boosting*. Dans cette famille, un ensemble d'estimateurs est construit puis agrégé. Dans le cas de cette étude, la variable d'intérêt correspond toujours au taux de déracinement des arbres. C'est une méthode séquentielle, contrairement à *Random Forest*.

On note  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$  le jeu de données, où  $x_i \in \mathbb{R}^p$  est le vecteur des variables explicatives (vent, sol majoritaire, abri topographique, etc.). On note  $y_i \in [0, 1]$  le taux de déracinement observé. Les  $m_t$  sont des arbres de décision et  $\varepsilon_t$  sont les résidus au pas  $t$  (avec  $\varepsilon_t = y - m^{(t)}(x)$ ). Les grandes étapes du *boosting* sont les suivantes :

1. On estime le modèle  $m_1$  (arbre faible) pour  $y$  à partir de  $x$ . On en déduit le vecteur d'erreurs  $\varepsilon_1$ .
2. On estime le modèle  $m_2$  pour  $\varepsilon_1$  à partir de  $x$  (apprentissage sur les pseudo-résidus). On en déduit  $\varepsilon_2$ .
3. On effectue cela  $k$  fois, avec  $k$  fixé au préalable, afin d'obtenir l'estimateur :

$$m^{(k)}(x) = m_1(x) + m_2(x) + \dots + m_k(x) = m^{(k-1)}(x) + m_k(x).$$

*XGBoost* utilise la descente de gradient, un algorithme d'optimisation visant à minimiser une fonction objectif différentiable notée  $\mathcal{L}(\theta)$ .

Il existe différents hyperparamètres permettant de contrôler simultanément la complexité et la performance du modèle. Le bon choix de ceux-ci, via optimisation bayésienne, est primordial afin d'optimiser les résultats et la prédiction du modèle. L'optimisation bayésienne est une méthode efficace qui utilise les résultats des tests effectués précédemment pour prédire quelles combinaisons d'hyperparamètres ont le plus de chances de donner de bons résultats. Au lieu d'explorer les paramètres de manière aléatoire ou systématique, elle construit un modèle probabiliste qui apprend à chaque itération où chercher les meilleurs paramètres, ce qui la rend beaucoup plus efficace.

En implémentant cela sur R, les paramètres obtenus sont montrés dans la table III.20 et les métriques qui découlent du modèle optimisé sont donnés dans la table III.21 :

Hyperparamètre	Rôle	Valeur optimale
<code>eta</code>	Taux d'apprentissage	0,1963
<code>max_depth</code>	Profondeur maximale des arbres individuels	7
<code>nrounds</code>	Nombre total d'itérations	146
<code>min_child_weight</code>	Poids minimal cumulé par feuille	3,2191
<code>subsample</code>	Fraction d'observations échantillonnées par arbre	0,9744
<code>colsample_bytree</code>	Fraction de variables par arbre	0,8099
<code>colsample_bylevel</code>	Fraction de variables par niveau	0,7550
<code>gamma</code>	Gain minimal pour scinder un nœud	0,1621
<code>alpha</code>	Régularisation L1 des feuilles	0,5968
<code>lambda</code>	Régularisation L2 des feuilles	0,8083

TABLE III.20 – *XGBoost* : Hyperparamètres optimaux par optimisation bayésienne

Modèle	RMSE	$R^2$
MLG Quasi-binomial	0,3254	0,1155
<i>Random Forest</i> optimisé	0,2683	0,3502
<i>XGBoost</i> optimisé	0,2790	0,3088

TABLE III.21 – Comparaison des performances : MLG vs *Random Forest* vs *XGBoost*

*XGBoost* est légèrement moins performant que *Random Forest* selon les critères du RMSE et du  $R^2$  mais reste cependant tout à fait performant. Il convient maintenant d'analyser la capacité prédictive de ce modèle.

Mesure	Effectif	% du total
Vrais Positifs (VP)	439	19,9 %
Vrais Négatifs (VN)	1356	61,5 %
Faux Positifs (FP)	133	6,0 %
Faux Négatifs (FN)	276	12,5 %
<b>Total</b>	<b>2204</b>	<b>100,0 %</b>

TABLE III.22 – *XGBoost* : Effectifs et pourcentages de la matrice de confusion

La précision reste similaire en comparaison au *Random Forest*. Cependant, le rappel diminue, ce qui indique que les vraies communes à risque élevé sont moins systématiquement détectées que dans le cas du *Random Forest*. Cela fait donc aussi diminuer le  $F_1$ -score, comme indiqué dans la table III.23 :

Mesure	Valeur
Précision	0,768
Rappel	0,614
$F_1$ -score	0,682

TABLE III.23 – *XGBoost* : Mesures globales de performance

La capacité prédictive est donc en baisse vis-à-vis du modèle précédent. C'est néanmoins en hausse par rapport au MLG ce qui confirme l'intérêt de l'utilisation des modèles *machine learning*. La courbe précision-rappel (Figure III.7) montre une AUC de 0,760, ce qui indique toujours une capacité de discrimination meilleure comparée à l'aléatoire.

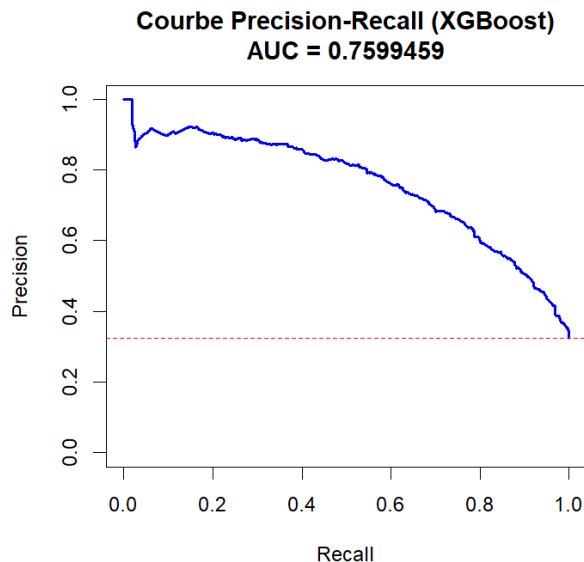


FIGURE III.7 – *XGBoost* : Courbe précision-rappel

Le modèle *Random Forest* semble donc plus adapté que *XGBoost*, modèle étudié dans cette section. De ce fait, seul le *backtesting* du modèle *Random Forest* est effectué dans la section qui suit. Le choix final du modèle sera détaillé et explicité après avoir détaillé ce *backtesting*.

### III.3.4 Backtesting

Afin de confirmer la pertinence du modèle *Random Forest*, il convient de réaliser une analyse de *backtesting* approfondie, comme dans le cas des MLG. L'avantage avec *Random Forest* est que cette analyse est en partie inhérente au code étant donné la séparation entre base d'entraînement et base de test.

Tout d'abord, l'optimisation des hyperparamètres a son importance, dans le sens où les performances s'améliorent entre le modèle de base et le modèle paramétré (Table III.24)

Modèle	RMSE	$R^2$
<i>Random Forest</i> non optimisé	0,2719	0,3371
<i>Random Forest</i> optimisé	0,2683	0,3502

TABLE III.24 – Comparaison des performances : *Random Forest* basique et optimisé

Une validation sur 5 plis a été réalisée. Concrètement, les données sont ordonnées chronologiquement, et, à chaque pli, noté  $i$ , les observations passées constituent l'échantillon d'apprentissage tandis qu'un bloc fixe de 440 communes sert de base de test. Deux indicateurs sont calculés pour chaque pli : RMSE et  $R^2$ .

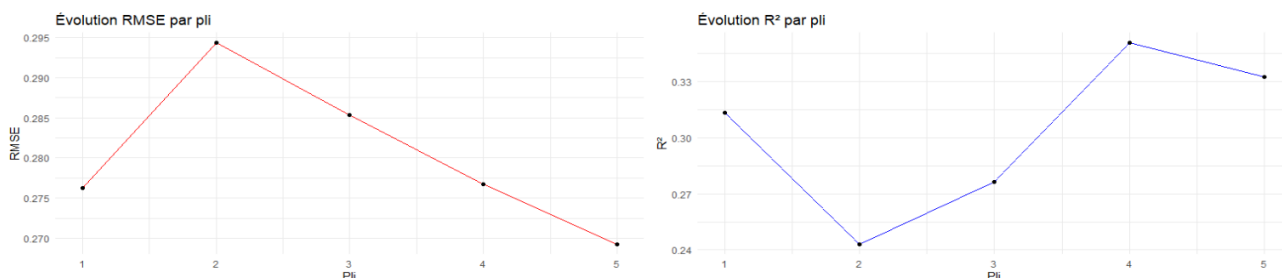


FIGURE III.8 – *Random Forest* : Évolution du RMSE et du  $R^2$  par pli

Les cinq plis affichent une RMSE moyenne de 0,2804 (écart-type de 0,0097) et un  $R^2$  moyen de 0,3031 (écart-type de 0,0433 ici). Ainsi, le modèle explique environ un tiers de la variance du ratio sinistres/surface assurée, avec une variabilité faible. Cette stabilité confirme la pertinence du modèle.

Les ratios prédits sont revus à la hausse par rapport au MLG et s'approchent avec quasi-exactitude des ratios moyens de la sinistralité. En effet, une différence de moins de 1 % est observée entre la moyenne de *Random Forest* et celle de la sinistralité. Pour comparaison, la moyenne obtenue est de 32,16 % pour *XGBoost*.

Moyenne Ratio Prédit <i>Random Forest</i>	Moyenne sinistralité historique
33,08 %	32,33 %

TABLE III.25 – *Random Forest* : Moyenne prédite vs Moyenne observée du ratio

De son côté, la figure III.9 indique que les valeurs faibles sont bien prédites (correspondance avec la ligne rouge) et qu’une dispersion plus importante pour les communes à haut risque se retrouve. Ce résultat est cohérent avec l’aire sous la courbe précision–rappel (AUC-PR).

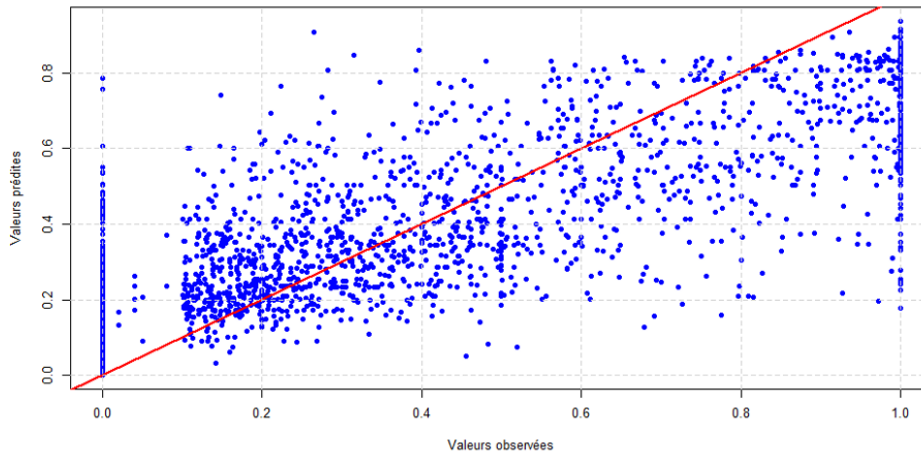


FIGURE III.9 – *Random Forest* : Valeurs observées et valeurs prédites

### III.3.5 Comparaison et choix du meilleur modèle

Les résultats du *backtesting* du modèle *Random Forest* sont convaincants. La table III.26 récapitule les valeurs des AUC et des  $F_1$ -score des trois modèles principaux étudiés et confirme la pertinence du modèle *Random Forest* pour la suite de ce mémoire.

Modèle	AUC ROC	AUC PRC	$R^2$	RMSE	$F_1$ -score
MLG Quasi-binomial	0,647	0,584	0,116 <sup>a</sup>	0,325	0,596
<i>XGBoost</i> optimisé	0,791	0,760	0,309	0,279	0,682
<b><i>Random Forest</i> optimisé</b>	<b>0,904</b>	<b>0,823</b>	<b>0,350</b>	<b>0,268</b>	<b>0,756</b>

<sup>a</sup> Pour le MLG, le  $R^2$  correspond à l’indice de McFadden.

TABLE III.26 – Comparaison des performances : Métriques finales pour les trois modèles

Ce modèle est donc choisi pour les prédictions ainsi que la construction du zonier. Les statistiques générales des prédictions sont présentées en table III.27 et les prédictions pour les trois exemples de communes se retrouvent en table III.28.

Moyenne	Médiane	1 <sup>er</sup> quartile	3 <sup>e</sup> quartile	Min	Max
33,08 %	30,65 %	22,73 %	42,99 %	9,59 %	83,27 %

TABLE III.27 – *Random Forest* : Statistiques descriptives de *Ratio\_SinAss*

Gosselming	Montsapey	Cudos
Moselle	Savoie	Gironde
18,49 %	31,37 %	40,55 %

TABLE III.28 – *Random Forest* : Ratio de destruction - Exemple de trois communes françaises

Les valeurs sont conformes avec les attentes de la sinistralité historique. Il est alors possible de passer à la réalisation du zonier.

## III.4 Production du zonier

### III.4.1 Cartographie

Via QGIS, il est possible de faire différentes modélisations et différents choix pour la séparation en classes. Celle retenue dans ce mémoire est la méthode d'optimisation de Jenks par rupture naturelle (JENKS [1967]). Il s'agit d'une méthode de regroupement de données conçue pour déterminer la meilleure disposition des valeurs dans différentes classes.

Dans un premier temps, et afin de bien comprendre les réponses géographiques à la problématique de ce mémoire, 10 zones ont été gardées. Cela permet de bien prendre en considération les zones à risque françaises. La figure III.10 présente les résultats obtenus avec le modèle *Random Forest*.

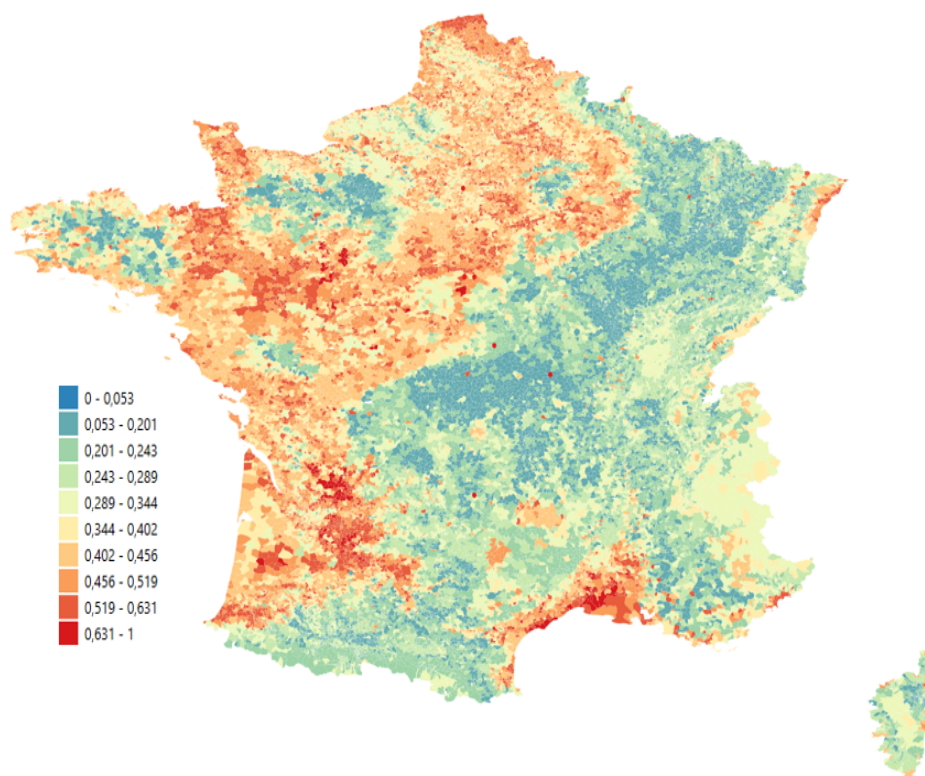


FIGURE III.10 – Zonier agrégé en 10 zones : *Random Forest*

Il est possible de comparer avec les zoniers obtenus avec *XGBoost* (Figure III.11) et MLG (Figure III.12). Pour la lisibilité, une échelle unique est utilisée pour tous les zoniers : celle du modèle retenu, *Random Forest*.

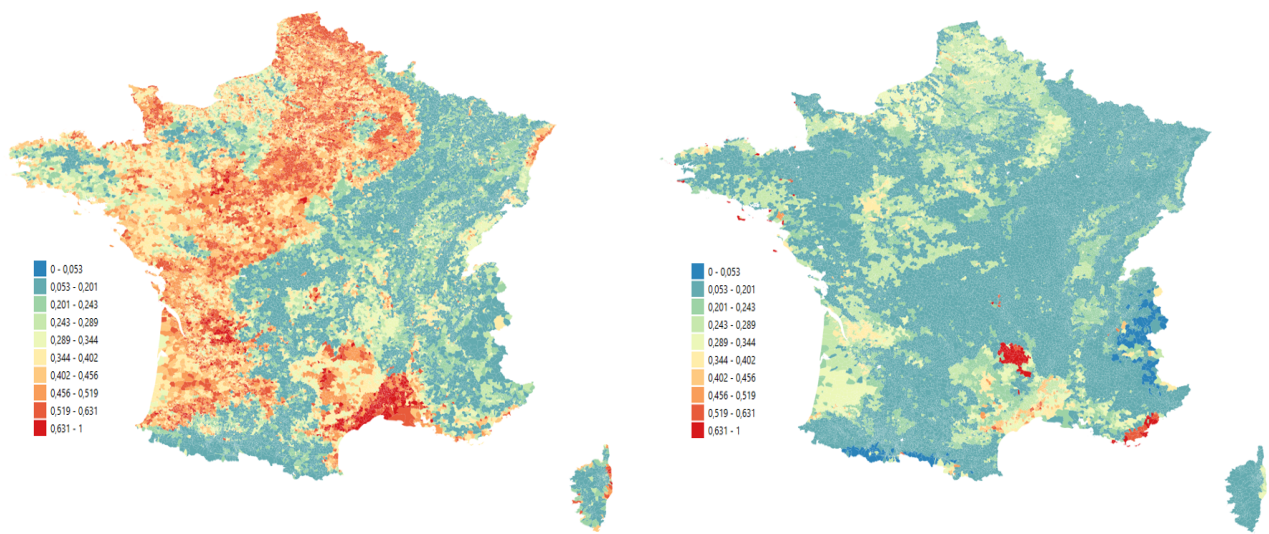


FIGURE III.11 – Zonier agrégé en 10 zones : *XGBoost*      FIGURE III.12 – Zonier agrégé en 10 zones : MLG Quasi-binomial

Les résultats sont ressemblants entre les deux modèles de *machine learning*. Ils diffèrent pour le MLG mais les ratios prédits sont plus faibles donc l'échelle ne rend pas compte pleinement des similitudes à première vue. En regardant plus en détail, un constat immédiat est que les mêmes zones à risque sont considérées dans les trois modèles.

La carte de la figure III.10 ainsi retenue met en évidence une première zone à risque sur la façade méditerranéenne vers le Languedoc-Roussillon. Cette vulnérabilité est prolongée jusqu'aux côtes méditerranéennes, régulièrement éprouvées par les vents hivernaux ou estivaux. De plus, la zone du portefeuille, située le long du littoral aquitain, est également risquée : la forêt des Landes, sur sols sableux, subit bien les tempêtes atlantiques. Plus au nord, Vendée, Charente-Maritime et Poitou forment un autre lieu à haut déracinement, qui remonte vers l'Île-de-France et le Nord. Les côtes bretonnes et celles de la Manche sont aussi sujettes à ce risque.

À l'inverse, l'Est, des Ardennes au Jura en passant par la Bourgogne et l'Alsace, reste majoritairement peu risqué (hormis le nord-est alsacien). En Corse, seule la côte orientale montre quelques touches orangées, tandis que l'intérieur montagneux demeure globalement préservé du déracinement. Enfin, les zones montagneuses, comme attendu, présentent un risque faible de déracinement. Ceci est dû notamment à la robustesse des arbres qui s'y situent, au type de sol dans lequel les arbres sont plantés ou encore à l'abri topographique important procuré par le relief environnant.

Il est intéressant de regarder les zoniers à la lumière de la carte de la figure II.16. Le sol français métropolitain n'étant pas recouvert exclusivement de forêts, certaines zones ont été considérées à risque selon leurs critères géophysiques alors qu'aucun peuplement forestier ne se trouve dans cette commune. Lors d'une mise en application commerciale de cette étude, il est judicieux de regarder les zones où le risque est faible tout en ayant une densité forestière élevée. C'est d'ailleurs cette agrégation commerciale qui justifie la nouvelle décomposition du zonier présentée dans la section suivante.

### III.4.2 Mise en place opérationnelle

Si un zonier à 10 zones est plus parlant dans une considération actuarielle afin d'analyser les résultats, il est préférable d'agréger selon moins de zones. Cela répond à une démarche commerciale : la différenciation est alors plus simple et ce nouveau zonier, montré figure III.13 répond à un souci de clarté pour le client. Quatre zones sont retenues dorénavant, ce choix faisant suite à des discussions avec les équipes de Groupama Forêts Assurances.

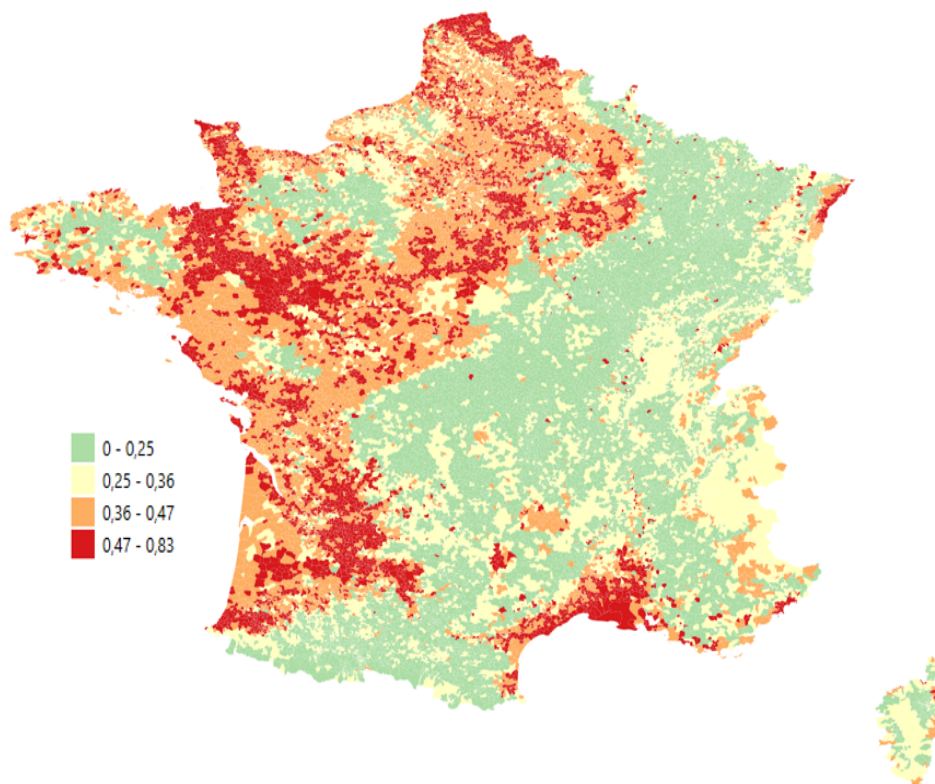


FIGURE III.13 – Zonier agrégé en 4 zones : *Random Forest*

### III.5 Synthèse de la troisième partie et utilisation future du zonier

Dans cette partie, un premier modèle inspiré de l'étude britannique *Windthrow Hazard Classification* a servi à repérer les zones à risque élevé en France métropolitaine. Toutefois, une modélisation actuarielle plus précise, prenant en compte la sinistralité passée ainsi que les interactions entre variables explicatives, a révélé, pour l'altitude, un effet de signe opposé à celui suggéré par le premier modèle.

Deux MLG ont été estimés. Le premier, à distribution bêta, a été écarté en raison de performances insuffisantes et de prédictions trop proches des bornes. Le second, quasi-binomial à lien logit, s'est révélé plus adapté à notre cas d'étude. Néanmoins, la comparaison des métriques et des prédictions met en évidence la supériorité des modèles de *machine learning*.

Le modèle *Random Forest* a en particulier fourni des résultats largement satisfaisants. La capacité prédictive du modèle ainsi que le *backtesting* ont confirmé la pertinence de ce choix, en témoignent l'AUC-PR de 0,823 ou le  $F_1$ -score de 0,756. Grâce aux prédictions de ce modèle, un zonier en 10 zones pour analyse actuarielle et géographique, puis en 4 zones pour compréhension commerciale a été créé. Les résultats obtenus, globalement fidèles aux attentes, confirment la pertinence de l'approche retenue.

Le zonier fournit la différenciation géographique du risque. Mais pour l'ancrer dans un cadre tarifaire réaliste, il reste à résoudre une question critique : comment lisser sur plusieurs années l'impact sur le portefeuille de la tempête Klaus en 2009 (S/P = 115%) ? Cela nécessite d'estimer sa période de retour, objet de la section IV.2., avant donc d'utiliser le zonier dans la tarification.

En synthèse, cette partie présente les différents aspects de la modélisation mathématique du risque tempête et de la création d'un zonier à la maille communale, qui sera utilisé dans la suite pour la tarification.

# IV - Recalibration de la tarification du risque tempête

## IV.1 Mise en place stratégique et technique

### IV.1.1 Retour sur la tarification en assurance forêt

L'assurance forêt, étant très corrélée géographiquement avec les catastrophes naturelles, donne des résultats assez variables selon les années et le portefeuille étudié. Dans le cadre du portefeuille de ce mémoire, localisé dans le Bassin aquitain, un diagnostic des S/P, déjà présenté au sein de la figure II.14, met en évidence un constat immédiat.

En effet, un déséquilibre entre les S/P dits attritionnels (années hors catastrophes) et le S/P dit catastrophe en 2009, année de passage de la tempête Klaus dans le Bassin aquitain, est mis en exergue. Nous ne possédons cet historique de S/P que depuis 2008. Dans le cas contraire, 1999 (tempêtes Lothar et Martin) aurait également été une année intéressante à étudier et à prendre en compte.

S/P attritionnel	S/P catastrophe
6 %	1 155 %

TABLE IV.1 – Ratios S/P : Distinction entre attritionnel et catastrophe

Il est donc nécessaire d'étudier séparément les années sans catastrophe, où l'absence de sinistres donne des S/P très favorables à l'assureur, et les années avec catastrophes, pour lesquelles le S/P doit être amorti sur plusieurs années, via le calcul d'une période de retour propre au portefeuille étudié. Celle-ci est estimée à l'aide de la théorie des valeurs extrêmes ainsi que des copules, dans l'objectif de fiabiliser l'amortissement. Le calcul du S/P annuel est alors le suivant :

$$S/P_{\text{Annuel}} = S/P_{\text{Attritionnel}} + \frac{S/P_{\text{Catastrophe}}}{\text{Période de retour}}$$

Les nouveaux jeux de données géographiques et les travaux réalisés précédemment qui en ont découlé, donnant plus d'informations à l'échelle communale, justifient la relance de nouveaux tarifs qui soient plus adaptés et fiables.

### IV.1.2 Objectifs tarifaires et commerciaux

Fournir une justification actuarielle de la tarification, via modélisation mathématique du risque et production d'un zonier, permet d'obtenir des valeurs cohérentes, et d'adapter, si nécessaire, des tarifs d'expérience plus adaptés au cadre actuel. Dans le paysage assurantiel, il est primordial d'être toujours en mesure d'expliquer d'où proviennent les chiffres.

Cela permet d'assurer l'acceptabilité des clients, mais aussi de se positionner de manière compétitive sur le marché en se repositionnant de manière juste face aux offres des autres assureurs spécialisés dans l'assurance forêt. Il y a ainsi une convergence des intérêts dans cette recherche de justesse commerciale.

Afin d'assurer les résultats de l'entreprise, le ratio S/P annuel ciblé par l'assureur est considéré égal à 80 % dans le Bassin aquitain. Cette cible est déterminée afin de se protéger contre les catastrophes extrêmes (au-delà de l'historique du portefeuille) tout en gardant une marge incluant les frais de réassurance. Ce choix est fondé sur la forte volatilité du portefeuille forêt d'une année à l'autre. Il vise à garantir une tarification compétitive en année normale, tout en protégeant l'entreprise contre les années extrêmes par le lissage des événements catastrophiques.

La cible de S/P est fixée par rapport au Bassin aquitain et les évolutions tarifaires trouvées seront appliquées aux autres communes françaises. C'est une hypothèse prise du fait de la forte représentation de cette zone géographique dans le portefeuille. Le but est d'ancrer la calibration des tarifs sur un échantillon riche en données et donc statistiquement plus stable, plutôt que d'essayer de faire des calculs généraux avec des données trop fines, ce qui pourrait biaiser les résultats.

La hausse ou la baisse des tarifs proposés s'opère donc à partir de la différence entre le S/P cible et le S/P annuel lissé obtenu. Pour cela, présentons dans un premier temps les éléments principaux de la tarification en place pour le moment. Celle-ci ne sera pas entièrement dévoilée pour des raisons de confidentialité mais les idées majeures sont retransmises.

### IV.1.3 Ancienne tarification

Le tarif s'obtient selon 3 critères : l'essence ou les essences de l'espace à assurer, l'âge des arbres qui composent cet espace ainsi que la surface à assurer en hectares. Pour chaque décomposition essence/catégorie d'âge, un facteur est multiplié par la surface à assurer afin d'obtenir le tarif souhaité.

Considérons un exemple. Un sylviculteur possède une forêt avec les caractéristiques de la table IV.2 :

Essence	Âge	Surface concernée (en ha)
Pins maritimes	6 ans	7
Pins maritimes	4 ans	3
Eucalyptus	12 ans	4

TABLE IV.2 – Répartition d'une forêt : Exemple pour tarification

En parallèle, la grille de tarification donne les informations suivantes :

Essence	Âge	Prix par hectare
Pins maritimes	6–10 ans	4,93
Pins maritimes	0–5 ans	2,46
Eucalyptus	10 ans et plus	9,86

TABLE IV.3 – Prix par hectare : Exemple pour tarification

Ainsi, cet exemple fictif donne un tarif pour l'assuré de 81,33 €. Il est possible de bien séparer les essences afin de reconstituer une forêt entière. Le but est maintenant de voir comment le zonier précédemment mis en place va influencer dans la tarification.

#### IV.1.4 Utilisation du zonier communal

Une fois le S/P annuel obtenu, avec le S/P catastrophe lissé, il est possible d'obtenir une différence par rapport au S/P cible, qui est pour rappel de 80 %. Ce S/P cible est fixé au niveau du Bassin aquitain. Le coût hors-cat et l'impact  $R$  d'un événement majeur (type Klaus) annualisé via sa période de retour sont séparés :

$$E[S]_{\text{bassin}} = S_{\text{bassin}}^{\text{non-cat}} + \frac{S_{\text{bassin}}^{\text{Klaus}}}{R}, \quad S/P_{\text{bassin}}^{\text{hist}} = \frac{E[S]_{\text{bassin}}}{P_{\text{bassin}}}.$$

Chaque commune  $i$  est comparée à la moyenne historique dans le Bassin aquitain. L'écart par commune est :

$$\Delta_i = SP_i^{\text{hist}} - SP_{\text{bassin}}^{\text{hist}}.$$

Les communes sont classées en 4 zones tarifaires  $z$  selon  $\Delta_i$ , ce qui a été fait dans la partie zonier. À chaque zone  $z$  est donc associée un facteur multiplicatif. En appliquant ces facteurs à la tarification, un nouveau tarif justifié est obtenu, ce qui constitue l'objectif de cette étude.

L'essence majoritaire étant déjà intégrée comme une variable explicative dans la modélisation du risque et afin de ne pas double-compter l'impact de celle-ci, une séparation est effectuée selon l'essence dans le zonier. Ainsi, le zonier est séparé en 12 zones, et non en 4 (zones 1, 2, 3 et 4 pour les conifères et idem pour le mélange et les feuillus).

Avant d'utiliser ce zonier agrégé en 12 zones, le S/P annuel doit être calculé. Pour cela, il est nécessaire d'estimer la période de retour de la tempête Klaus sur le Bassin aquitain. La section suivante explicite ces points.

## IV.2 Travaux sur la période de retour des sinistres catastrophe

### IV.2.1 Impact de la tempête Klaus et hypothèses

La tempête Klaus est une tempête exceptionnelle par son envergure, qui a principalement touché le sud-ouest de la France, soit la zone considérée du portefeuille, entre le 23 et le 25 janvier 2009. Les spécialistes la classent comme la plus dévastatrice en France depuis les tempêtes de 1999.

Les rafales maximales ayant touché les stations météorologiques du sud-ouest de la France lors de la journée du samedi sont détaillées dans la carte de la figure IV.1.

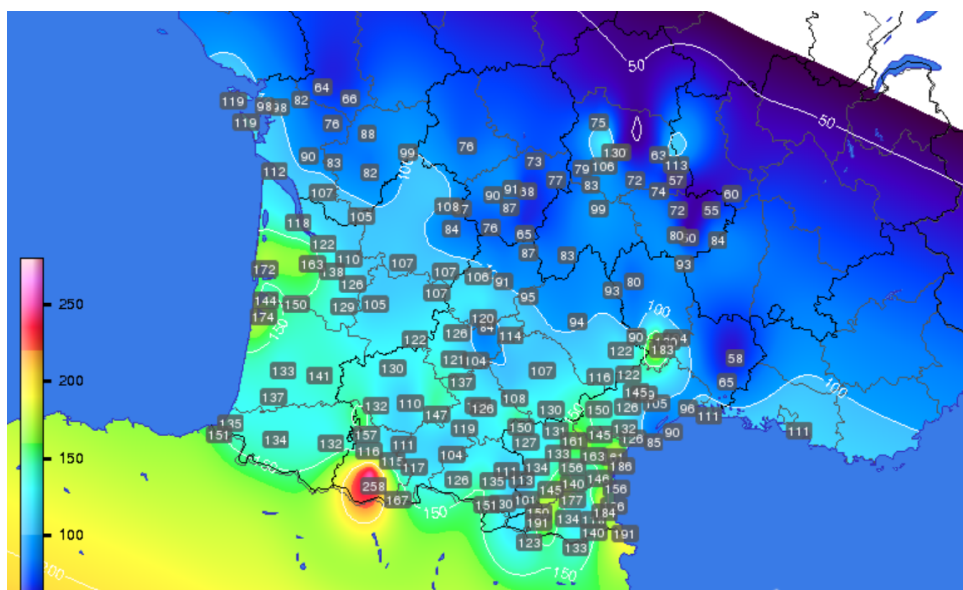


FIGURE IV.1 – Tempête Klaus (24 janvier 2009) : Rafales journalières (INFOCLIMAT [2009])

En considérant les huit départements du Bassin aquitain, la rafale maximale moyenne observée dans les stations retenues est de 134,88 km/h. C'est ce qui est appelé le **niveau de retour**. Il est question de voir à quelle fréquence ce niveau est atteint à nouveau. Pour cela, les données SYNOP, présentées dans la première partie, sont utilisées. La variable considérée est *rafper*, qui est la vitesse maximale de rafale mesurée pendant une période, qui est définie dans ce cas comme la journée.

Les données SYNOP de rafales sont prises en considération depuis le 1<sup>er</sup> novembre 2003, ce qui fournit un historique conséquent. Trois stations météorologiques sont localisées dans le secteur du Bassin aquitain. Celles-ci sont présentées dans la table IV.4, ainsi que leur département.

Numéro de la station	Emplacement	Département
07314	Pointe de Chassiron	Charente-Maritime
07510	Bordeaux Mérignac	Gironde
07607	Mont-de-Marsan	Landes

TABLE IV.4 – Base SYNOP : Stations du Bassin aquitain

Après avoir converti les m/s en km/h afin de comparer des grandeurs équivalentes, les données sont prêtes à être utilisées. Au préalable, rappelons les notions essentielles de la théorie des valeurs extrêmes.

## IV.2.2 Introduction à la théorie des valeurs extrêmes

La théorie des valeurs extrêmes fournit un cadre probabiliste pour modéliser la fréquence et l'intensité des phénomènes rares. Dans l'approche des maxima par blocs, si  $(X_1, \dots, X_n)$  sont i.i.d. de loi  $F$  et  $M_n = \max(X_1, \dots, X_n)$ , le théorème de Fisher–Tippett–Gnedenko montre que,

pour des suites  $(a_n > 0, b_n)$ , la loi de  $M_n$  converge vers une des trois familles limites (Fréchet, Weibull, Gumbel), unifiées par la loi  $GEV(\mu, \sigma, \xi)$  :

$$G(x) = \begin{cases} \exp\left\{-\left[1 + \xi \frac{x-\mu}{\sigma}\right]^{-1/\xi}\right\}, & 1 + \xi \frac{x-\mu}{\sigma} > 0, \quad \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases}$$

où  $\mu$  est un paramètre de position,  $\sigma > 0$  d'échelle et  $\xi$  de forme. L'épaisseur de queue est lue via  $\xi$  :  $\xi > 0$  (Queue lourde ou Fréchet),  $\xi = 0$  (Gumbel),  $\xi < 0$  (Queue bornée ou Weibull).

L'approche POT est privilégiée car elle exploite tous les excès au-dessus d'un certain seuil  $u$ . Pour un seuil  $u$  suffisamment élevé, la loi  $Y = X - u \mid X > u$  est alors approchée par une distribution GPD (Pareto généralisée) de paramètres  $(\sigma_u, \xi)$  :

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-1/\xi}, \quad y \geq 0, \quad 1 + \xi \frac{y}{\sigma_u} > 0.$$

Sous stationnarité, les franchissements du seuil suivent en première approximation un processus de Poisson d'intensité annuelle  $\lambda_u$  (nombre moyen d'excès par an). Les niveaux de retour  $z_T$  sont ainsi déduits pour une période de retour  $T$  (en années) :

$$z_T = u + \begin{cases} \frac{\sigma_u}{\xi} \{(\lambda_u T)^\xi - 1\}, & \xi \neq 0, \\ \sigma_u \log(\lambda_u T), & \xi = 0, \end{cases}$$

et, réciproquement, il est possible d'obtenir les périodes de retour d'un niveau  $x \geq u$  :

$$T(x) = \frac{1}{\lambda_u} \begin{cases} \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{1/\xi}, & \xi \neq 0, \\ \exp\left(\frac{x-u}{\sigma_u}\right), & \xi = 0. \end{cases}$$

Dans ce mémoire, les paramètres  $(\xi, \sigma_u)$  sont estimés par maximum de vraisemblance via l'optimisation Nelder–Mead, détaillée dans l'**Annexe H** (les méthodes SANN et BFGS ont aussi été testées et donnent des résultats très similaires). De plus, le seuil  $u$  est choisi à partir d'outils de diagnostic POT. Le premier est le *Mean Excess Plot*, avec lequel on retient une zone où la courbe est à peu près linéaire, signe que la loi des excès au-dessus de  $u$  est bien approchée par une GPD. La valeur est affinée ou confirmée avec le *Threshrange Plot* en ré-estimant  $\xi$  et  $\sigma$  sur une plage de seuils voisins. On cherche une stabilité (paramètres quasi constants dans un intervalle de  $u$ ). Le choix final est validé par un diagnostic d'ajustement : un *QQ-plot* proche de la diagonale est souhaité.

### IV.2.3 Analyse univariée pour une station

Dans un premier temps, la démarche a été effectuée pour chacune des stations individuellement. La démarche est illustrée maintenant sur la station 07314, située à la Pointe de Chassiron, sur l'île d'Oléron.

Les tracés du *Mean Excess Plot* (figure IV.2) ainsi que du *Threshrange Plot* (figure IV.3) conduisent à considérer un seuil de  $u = 85$  km/h pour la station. Cela se confirme par le fait

que les paramètres  $\xi$  et  $\sigma$  sont stables autour du seuil choisi. Comparativement, pour la station 07510, on considère un seuil de  $u = 73$  km/h et ce dernier est de  $u = 65$  km/h pour la station 07607.

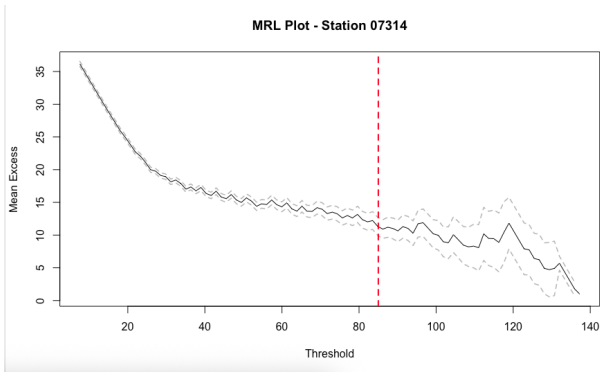


FIGURE IV.2 – Station 07314 : *Mean Excess Plot*

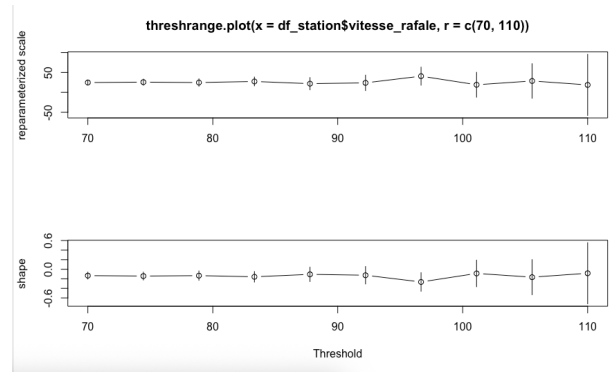


FIGURE IV.3 – Station 07314 : *Threshrange Plot*

Avec le seuil de  $u = 92$  km/h pour la station 07314, le *QQ-plot* des observations est bien aligné, avec quelques rares exceptions pour les extrêmes, ce qui confirme la pertinence du seuil utilisé.

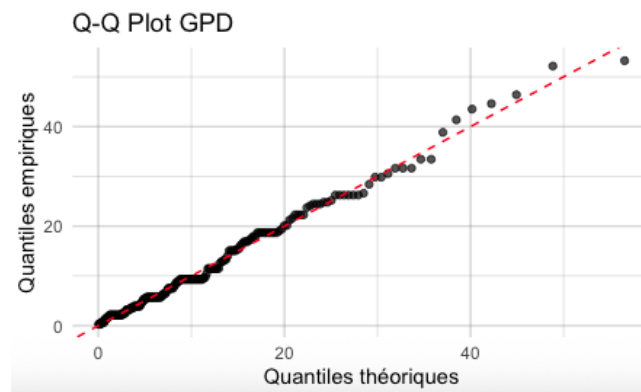


FIGURE IV.4 – Station 07314 : *Quantile-Quantile Plot*

Finalement, les informations suivantes sont procurées pour la station de Pointe de Chassiron :

Sur les excès  $X - u$ , on estime  $(\xi, \sigma_u)$  par maximum de vraisemblance et on calcule  $\lambda_u$  comme le nombre d'excès divisé par la durée d'observation (en années). Ici :

$$\hat{\xi} = -0,1362, \quad \hat{\sigma}_u = 12,5872 \text{ km/h}, \quad \hat{\lambda}_u \simeq 5,9 \text{ excès/an.}$$

À partir des formules POT, les niveaux de retour pour 07314 aux années 2, 10, 50 et 100 ans obtenus sont :

$$z_2 \approx 118,39 \text{ km/h}, \quad z_{10} \approx 131,38 \text{ km/h}, \quad z_{50} \approx 141,82 \text{ km/h}, \quad z_{100} \approx 145,66 \text{ km/h.}$$

Ces niveaux de retour se retrouvent sur la courbe de la figure IV.5.

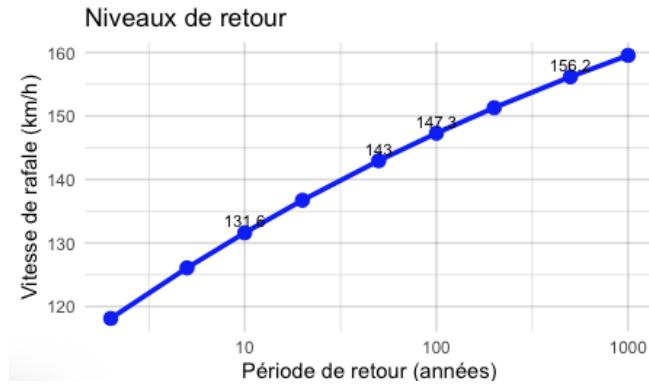


FIGURE IV.5 – Station 07314 : Niveaux de retour

Réciproquement, pour un niveau  $x = 134,88$  km/h, qui est pour rappel l'ordre de grandeur associé à Klaus dans le Bassin aquitain, la période de retour estimée est de **16,50298 ans** pour la station 07314.

Une première approche pour considérer l'impact sur les trois stations simultanément a consisté à créer une station, nommée 99999, prenant la rafale maximale journalière entre les trois stations du Bassin aquitain. Le *Mean Excess Plot* est le suivant :

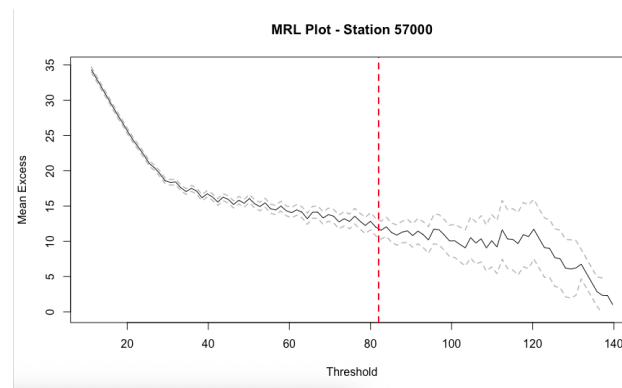


FIGURE IV.6 – Station 99999 : *Mean Excess Plot*

Avec un seuil de  $u = 82$  km/h, les résultats obtenus sont :

$$\hat{\xi} = -0,0964, \quad \hat{\sigma}_u = 13,0791 \text{ km/h}, \quad \hat{\lambda}_u \simeq 14,43 \text{ excès/an.}$$

Les résultats prennent la forme d'une série temporelle, comme l'indique la figure IV.7.

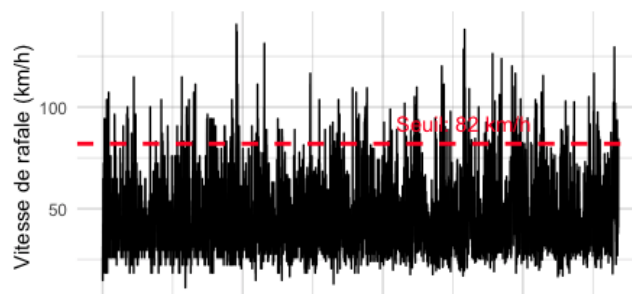


FIGURE IV.7 – Station 99999 : Série temporelle

Les niveaux de retour pour 99999 aux années 2, 5, 10, 50 et 100 ans obtenus sont :

$$z_2 \approx 119,56 \text{ km/h}, \quad z_{10} \approx 133,66 \text{ km/h}, \quad z_{50} \approx 145,73 \text{ km/h}, \quad z_{100} \approx 150,38 \text{ km/h}.$$

Ces niveaux de retour se retrouvent sur la courbe de la figure IV.8.

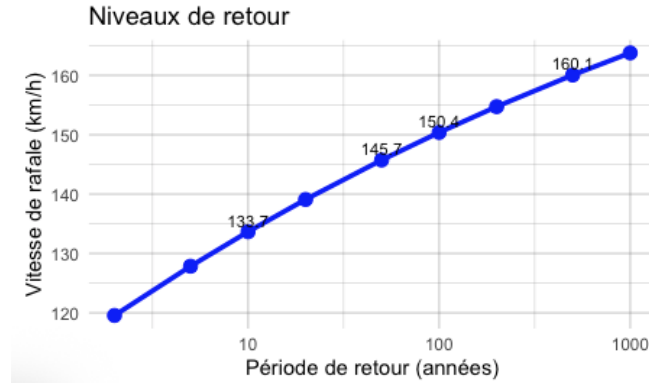


FIGURE IV.8 – Station 99999 : Niveaux de retour

Pour un niveau  $x = 134,88 \text{ km/h}$ , la période de retour estimée du maximum est de **11,28979 ans**.

Des résultats cohérents sont obtenus avec cette première approche, mais cette dernière comporte deux limites principales :

Premièrement, elle ne tient pas compte des interactions entre variables. Deuxièmement, la période de retour est vraisemblablement sous-estimée : une tempête pouvant durer plusieurs jours dans le Bassin aquitain, des indicateurs relevés sur deux jours distincts peuvent en réalité être un seul et même événement. Il est alors nécessaire d'introduire les copules pour modéliser la dépendance entre les stations.

#### IV.2.4 Introduction à la théorie des copules

Une copule de dimension  $d$  est une fonction de répartition sur  $[0, 1]^d$  avec des lois marginales uniformes sur  $[0, 1]$ .

On note  $C(\mathbf{u}) = C(u_1, \dots, u_d)$  pour les fonctions de répartition qui sont des copules

$$\rightarrow C \text{ est une application de la forme } C : [0, 1]^d \rightarrow [0, 1]$$

Afin de pouvoir appliquer la théorie des copules, il faut donc transformer les  $n$  données à disposition en données uniformes. Pour chaque station, il y a un seuil  $u_i$  et on note  $\hat{p}_{i,u} = \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{X_{it} \leq u_i\}$  la masse empirique sous le seuil. La fonction de répartition hybride

$$\hat{F}_i(x) = \begin{cases} \hat{p}_{i,u} \hat{F}_{i,\leq u}(x), & x \leq u_i, \\ \hat{p}_{i,u} + (1 - \hat{p}_{i,u}) \hat{H}_{i,u}(x - u_i), & x > u_i, \end{cases}$$

combine la fonction de répartition empirique tronquée  $\hat{F}_{i,\leq u}$  et la GPD ajustée  $\hat{H}_{i,u}(y) = 1 - \left(1 + \hat{\xi}_i y / \hat{\sigma}_{i,u}\right)^{-1/\hat{\xi}_i}$  (ou  $1 - e^{-y/\hat{\sigma}_{i,u}}$  si  $\hat{\xi}_i = 0$ ). Les pseudo-observations  $U_{it} = \hat{F}_i(X_{it})$  sont ainsi uniformes.

Sur la matrice  $\mathbf{U} = (U_{1t}, U_{2t}, U_{3t})_{t=1, \dots, n}$ , on calcule les corrélations de rang (Spearman  $\rho_S$ , Kendall  $\tau$ ) et on inspecte les nuages de points bivariés.

Les coefficients de dépendance de queue à gauche et à droite, notés respectivement  $\lambda_L$  et  $\lambda_U$ , sont définis par :

$$\lambda_L = \lim_{q \rightarrow 0} \frac{C(q, q)}{q},$$

$$\lambda_U = \lim_{q \rightarrow 1} \frac{1 - 2q + C(q, q)}{1 - q}.$$

Ces coefficients suivent les inégalités suivantes :  $0 \leq \lambda_L, \lambda_U \leq 1$ .  $\lambda = 0$  indique une absence de dépendance dans la queue considérée, à l'inverse de  $\lambda > 0$ .

Les copules testées sont les copules usuelles, listées ci-dessous :

- Gaussienne
- $t$ -Student
- Clayton
- Gumbel
- Frank

La copule la plus adaptée selon le critère de l'AIC et, en cas de similitudes, du BIC, est choisie. Pour rappel, ces critères doivent être minimisés.

## IV.2.5 Analyse multivariée via copules pour plusieurs stations

Concrètement, après avoir reformaté les résultats obtenus pour chaque station dans le cadre univarié, afin qu'ils soient adaptés au cadre multivarié, les étapes ci-dessous sont mises en place.

Avant d'estimer la copule optimale, la dépendance entre les variables uniformisées est examinée. La figure IV.9 affiche un nuage de points 3D des pseudo-observations dans le cube unité. Chaque point correspond à une observation journalière transformée. En cas d'indépendance, les points seraient répartis de façon homogène dans  $[0, 1] \times [0, 1] \times [0, 1]$ .

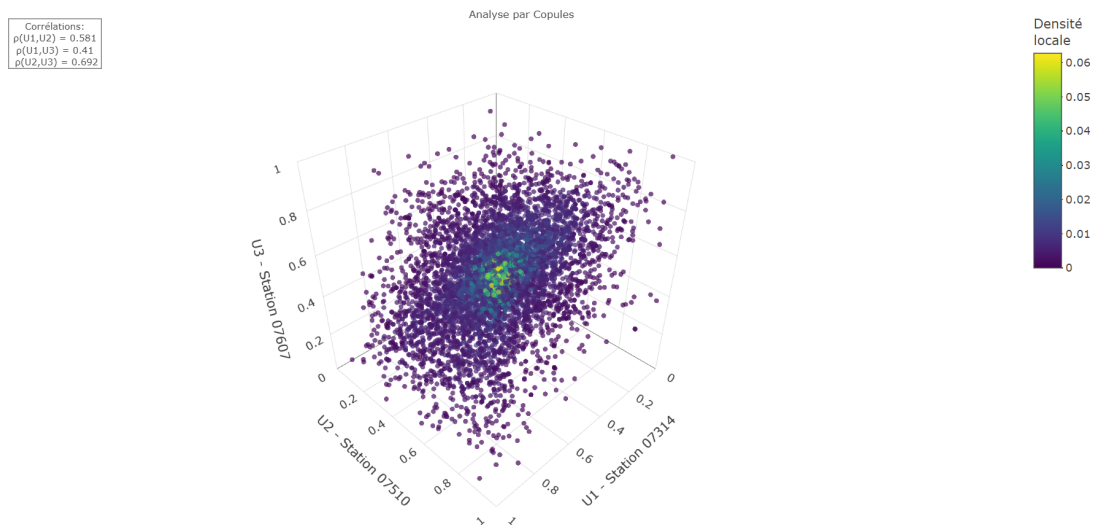


FIGURE IV.9 – Analyse des stations : Représentation des distributions uniformisées

La figure révèle les zones de concentration des observations, permettant d'identifier visuellement la structure de dépendance entre les trois stations. Les corrélations de Spearman calculées ( $\rho(U1,U2) = 0,581$ ,  $\rho(U1,U3) = 0,41$ ,  $\rho(U2,U3) = 0,692$ ) se traduisent par une répartition non uniforme dans le cube, avec des zones de plus forte densité indiquant la dépendance entre stations. Cela justifie l'utilisation d'une copule.

La transformation vers une distribution uniforme sur  $[0,1]$  est ensuite menée pour chacune des marginales et les résultats satisfaisants de la table IV.5 sont obtenus, dans le sens où la moyenne et la médiane sont proches de 0,5 et la variance proche de  $1/12$ , ce qui vaut environ 0,0833 :

Station	Moyenne	Médiane	Variance
07314	0,509	0,503	0,082
07510	0,512	0,504	0,082
07607	0,514	0,500	0,081

TABLE IV.5 – Analyse des stations : Vérification des marginales uniformes

La figure IV.10 montre la corrélation de Spearman et de Kendall entre les stations. La corrélation importante conforte le choix d'utiliser les copules dans cette section.

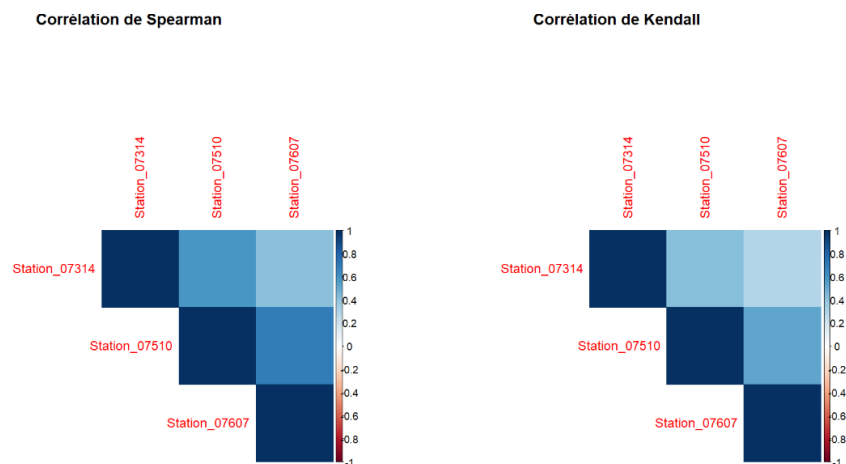


FIGURE IV.10 – Analyse des stations : Corrélation

La copule la plus adaptée, par le critère de minimisation de l'AIC, est celle de Gumbel (Table IV.6). C'est le résultat attendu en théorie car c'est la copule qui est la plus compatible pour modéliser une dépendance de queue supérieure, c'est-à-dire dans ce cas la co-occurrence d'événements extrêmes sur plusieurs stations. Le paramètre de la meilleure copule de Gumbel est  $\alpha = 1,607$ .

Copule	AIC	BIC
Gaussienne	-7163, 61	-7156, 72
t-Student	-7436, 71	-7422, 94
Clayton	-4219, 97	-4213, 09
Gumbel	-7474, 34	-7467, 45
Frank	-6397, 04	-6390, 15

TABLE IV.6 – Analyse des stations : Choix de la copule adaptée

Désormais, si  $C$  est la copule trivariée ajustée de Gumbel et  $C_{ij}$  ses marginales bivariées, alors on a les résultats suivants dans le cadre du calcul des probabilités conjointes sous la copule :

$$\begin{aligned} \Pr(\text{aucune station} > s) &= C(u_1, u_2, u_3), \\ \Pr(\text{au moins une station} > s) &= 1 - C(u_1, u_2, u_3), \\ \Pr(X_i > s_i, X_j > s_j) &= 1 - u_i - u_j + C_{ij}(u_i, u_j), \\ \Pr(\text{toutes} > s) &= 1 - \sum_{i=1}^3 u_i + \sum_{1 \leq i < j \leq 3} C_{ij}(u_i, u_j) - C(u_1, u_2, u_3). \end{aligned}$$

Ces égalités sont cohérentes avec les bornes de Bonferroni.

Finalement, les périodes de retour pour l'intersection (toutes les stations  $> 134,88$  km/h) et l'union (au moins une station  $> 134,88$  km/h) sont :

Mesure	Valeur
Probabilité	$2,836262 \times 10^{-6}$
Fréquence annuelle	$1,017519 \times 10^{-3}$ événements/an
Période de retour	982,8 ans

TABLE IV.7 – Période de retour : Intersection

Mesure	Valeur
Probabilité	$1,93 \times 10^{-4}$
Fréquence annuelle	0,069 événements/an
Période de retour	14,5 ans

TABLE IV.8 – Période de retour : Union

La dépendance raccourcit la période de retour de l'intersection : les événements extrêmes simultanés sont plus fréquents que sous indépendance. Et la dépendance allonge la période de retour de l'union : au moins un événement extrême est moins fréquent que sous indépendance.

Dans le cas de cette étude, c'est l'union qui est à conserver et la période de retour obtenue d'une tempête similaire à la tempête Klaus sur le Bassin aquitain est de **14,5 ans**.

Pour rappel, le S/P en 2009 du portefeuille étudié était de 1 155 %. Un lissage du S/P est donc à réaliser. Chaque année, le S/P catastrophe lissé est de  $1\,155\% / 14,5 = 79,66\%$ .

Le S/P annuel historique est de  $79,66\% + 5,90\%$  (attritionnel) = **85,56 %**

Néanmoins, dans le cas présent, on se couvre contre une tempête de la même intensité que celle de 2009. Si une tempête d'envergure plus grande se produisait, un traité de réassurance de type *Excess of Loss* doit être mis en place afin de se protéger face aux pertes possibles. Il est supposé dans ce mémoire que les frais de réassurance déjà engagés par Groupama Forêts Assurances avant cette refonte tarifaire demeurent inchangés.

## IV.3 Impact de la nouvelle tarification

### IV.3.1 Effets statistiques et sur les ratios

Le S/P annuel est de 85,56 %. La différence par rapport à la cible de 80 % dans le Bassin aquitain est ainsi de 5,56 points de pourcentage. Pour atteindre la cible de S/P, une hausse des tarifs de **6,95 %** dans le Bassin aquitain doit être appliquée. Cette hausse découle du calcul suivant :

$$85,56 = 80 \times (1 + \text{hausse}) \Rightarrow 1 + \text{hausse} = \frac{85,56}{80} = 1,0695 \Rightarrow \text{hausse} = 0,0695 = \mathbf{6,95\%}$$

Pour atteindre l'objectif dans le Bassin aquitain, l'attention est portée sur les huit départements de cette zone géographique. En faisant appel aux ratios prédits par *Random Forest*, les moyennes dans cette zone géographique selon l'essence sont les suivantes :

Essence	Moyenne du ratio
Conifères	47,57 %
Feuillus	38,19 %
Mélange	40,53 %

TABLE IV.9 – Bassin aquitain : Moyennes de ratios par type d'essence

Pour chaque commune du Bassin aquitain, on considère son essence majoritaire et on calcule l'écart au ratio de cette essence dans la table IV.9. On regroupe ensuite les communes par catégorie de zonier et par essence majoritaire, puis on prend, pour chaque couple (zonier, essence), la moyenne de ces écarts par rapport à la moyenne du Bassin aquitain. Ces écarts sont indiqués dans la table IV.10.

Classe	Variation
1C	-25,50 %
2C	-14,69 %
3C	-5,19 %
4C	+7,81 %
1F	-16,97 %
2F	-8,37 %
3F	+3,61 %
4F	+14,85 %
1M	-18,52 %
2M	-10,42 %
3M	+1,72 %
4M	+10,47 %

TABLE IV.10 – Bassin aquitain : Différences à la moyenne par classe

La zone 1 représente la classe de risque du zonier III.13 avec les ratios les plus bas et 4 la zone la plus risquée. C correspond aux conifères, M au mélange et F aux feuillus.

Le schéma de la figure IV.11 présente les étapes de l'évolution tarifaire. Sur ce schéma, le tarif actuel est supposé égal à 100 €.

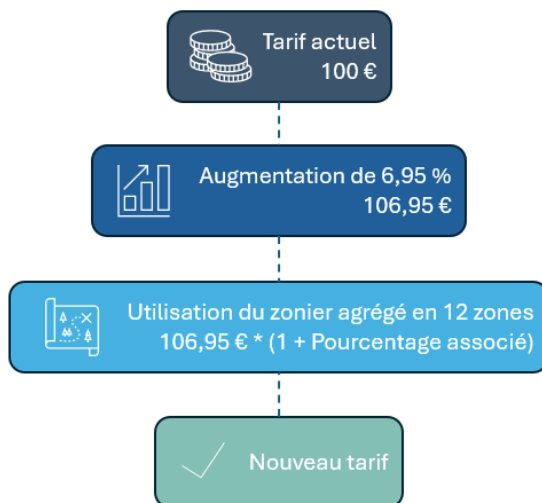


FIGURE IV.11 – Détermination du nouveau tarif : Schéma

Ainsi, les évolutions tarifaires pour chaque classe de risque sont résumées dans la table suivante :

Classe	Variation
1C	-20,32 %
2C	-8,76 %
3C	+1,40 %
4C	+15,30 %
1F	-11,20 %
2F	-2,00 %
3F	+10,82 %
4F	+22,83 %
1M	-12,86 %
2M	-4,19 %
3M	+8,79 %
4M	+18,15 %

TABLE IV.11 – Bassin aquitain : Évolutions tarifaires par classe

Elles proviennent de la formule suivante :

$$\Delta \text{tarif}_{\text{classe}} = (1 + 0,0695) \times (1 + \Delta \text{ratio}_{\text{classe}}) - 1$$

Exemple pour 1C :  $(1 + 0,0695) \times (1 - 0,2550) - 1 = -21,36 \%$ .

En moyenne, il y a nécessairement une augmentation des tarifs de 6,95 % dans le Bassin aquitain, entraînant une baisse du ratio S/P de 5,56 %, ce qui était l'objectif poursuivi. En France

métropolitaine, l'évolution moyenne suit les mêmes pourcentages. Au global, l'évolution est légèrement à la hausse : +0,23%. Toutefois, ce chiffre n'est pas réellement significatif étant donné que la grande majorité du portefeuille se trouve dans le Bassin aquitain. Il démontre la capacité de se développer dans d'autres régions françaises. Les figures IV.12 et IV.13 récapitulent ces évolutions tarifaires sous la forme d'un zonier.

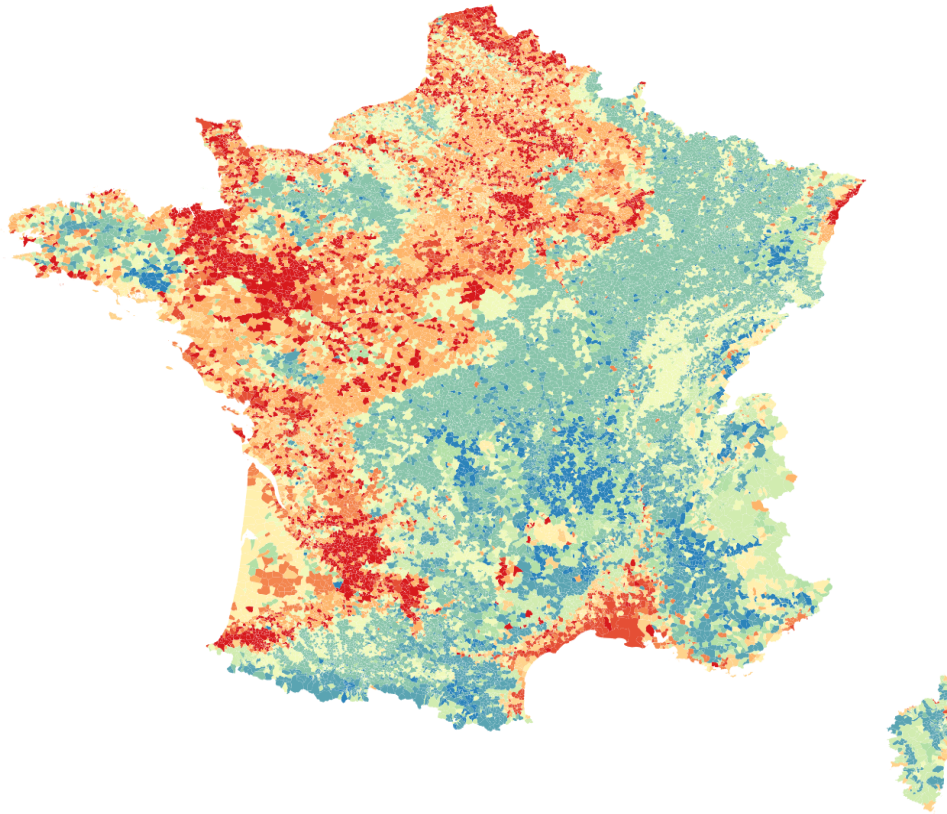


FIGURE IV.12 – Synthèse des évolutions tarifaires : Zonier

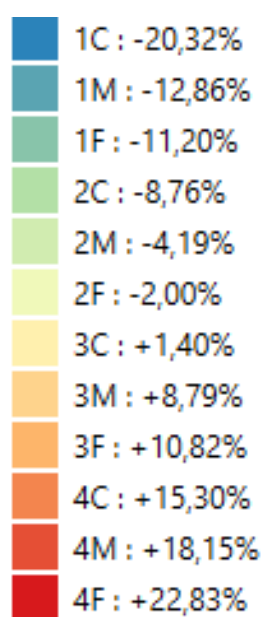


FIGURE IV.13 – Synthèse des évolutions tarifaires : Légende

Les peuplements de feuillus présentent des évolutions tarifaires plus marquées que les conifères. L'assureur disposait déjà d'une bonne maîtrise du tarif sur les conifères (très présents dans le Bassin aquitain), ce qui a conduit à des ajustements plus modérés pour cette essence. À l'inverse, les hausses plus importantes observées sur les feuillus reflètent avant tout une compensation de primes initialement sous-tarifées.

À titre d'exemple, pour les trois villes dont l'exemple est suivi dans ce mémoire, les augmentations tarifaires sont les suivantes :

Ville	Gosselming	Montsapey	Cudos
Département	Moselle	Savoie	Gironde
Type	1F	2M	3C
Différence	-16,97 %	-10,42 %	-5,19 %
<b>Évolution tarifaire</b>	<b>-11,20 %</b>	<b>-4,19 %</b>	<b>+1,40 %</b>

TABLE IV.12 – Exemple de trois communes françaises : Synthèse

La hausse tarifaire constatée en Gironde permet de proposer une baisse des tarifs à d'autres endroits en France tout en maintenant un objectif de S/P cohérent. Ce choix a des conséquences commerciales et stratégiques non négligeables.

### IV.3.2 Impact commercial et stratégique

La nouvelle tarification s'appuie sur un zonier en douze zones. Il s'appuie également sur des écarts de S/P par rapport à la moyenne du Bassin aquitain. Ces écarts sont par essence et par zone, ce qui permet d'afficher des tarifs compréhensibles pour les assurés.

L'impact stratégique de ce mémoire est important pour le secteur de l'assurance forêt :

- **Offres ciblées** proposées dans les zones 1 et 2 avec des prix compétitifs dans des régions historiquement peu assurées.
- **Protection du portefeuille** dans les zones 3 et 4 par une hausse des prix mesurée et justifiée.
- **Argumentaire simple pour les commerciaux** : Baisse des tarifs là où le risque est plus faible et hausse là où il est objectivement plus élevé, ce qui est encore une fois fondé sur le zonier et sur les métriques de S/P lissées par la période de retour.

Ainsi, la cible est atteinte avec une base mathématique solide, ouvrant la voie à un développement dans de nouvelles régions (prix compétitifs en zones moins exposées) tout en renforçant la rentabilité là où le risque est structurellement plus élevé. La table IV.11 fournit la grille d'ajustements par famille à déployer en tarification et en discours commercial.

## IV.4 Synthèse de la quatrième partie et ouverture

La recherche d'un tarif plus juste, fondée sur des bases mathématiques et actuarielles éprouvées, a donc conduit à calculer une période de retour : celle d'une tempête similaire à la tempête Klaus sur la zone géographique du Bassin aquitain. La valeur trouvée est de 14,5 ans.

Dans le cadre d'une recherche encore plus approfondie sur le calcul d'une période de retour, il aurait été pertinent d'utiliser des extrêmes multivariés, tout en précisant encore davantage le cadre spatio-temporel. Néanmoins, au sein de ce mémoire, l'utilisation de copules pour représenter l'interaction entre les stations permet d'ores et déjà de fournir des données réalistes et qui permettent de mener à son terme la démarche. Cette nouvelle perspective de recherche constitue toutefois un développement possible du travail présent.

Grâce à la période de retour et au zonier construit dans la partie précédente, des classes de risques par essence ont pu être créées. Ainsi, de nouveaux tarifs, en phase avec les objectifs de S/P de l'assureur, ont pu être proposés. L'évolution moyenne est bien celle souhaitée dans le Bassin aquitain, avec une augmentation moyenne des tarifs de 6,95 %. Cela permet de réduire les tarifs dans certaines zones encore peu touchées et considérées à risque plus relatif, afin de pouvoir s'y implanter durablement et ainsi créer un portefeuille plus développé.

Néanmoins, cette augmentation théorique doit être confirmée par une expérience du terrain, conformément aux échanges avec Groupama Forêts Assurances. En effet, si les hausses de tarif sont trop élevées dans certaines régions historiques, une partie importante du portefeuille pourrait se retirer. À l'inverse, si les baisses de tarif sont trop importantes sur les nouvelles régions où l'assureur souhaite se positionner, cela peut être un frein à la rentabilité au lieu d'être un atout. C'est donc un équilibre entre théorie et pratique qui est à rechercher continuellement.

Le sujet, qu'il relève de la tarification ou bien du calcul de la période de retour, est vaste et de nombreuses pistes restent ouvertes afin de trouver des résultats toujours plus précis. En synthèse, la cible de S/P est atteignable sur une base mathématique solide et avec un dispositif opérationnel prêt à être déployé.

# V - Conclusion

## V.1 Synthèse des résultats

Ce travail a pour objectif de fournir un cadre actuariel robuste et opérationnel pour tarifier le risque tempête en assurance forêt. La démarche a consisté à modéliser le déracinement des arbres à une maille géographique fine, à savoir la commune, afin de refléter l'hétérogénéité spatiale de l'exposition en France métropolitaine et d'aboutir à une tarification plus juste et plus lisible pour les assurés comme pour l'assureur.

La première étape a consisté à créer une base de données adaptée aux enjeux de ce mémoire. En plus de l'historique de sinistres et de primes fourni par Groupama Forêts Assurances, acteur majeur du marché de l'assurance forêt, une base communale a été créée en combinant des variables pertinentes pour expliquer le déracinement des arbres. Ces variables retenues sont l'altitude, la nature des sols, l'abri topographique, l'essence dominante et le vent maximal. Cette base, obtenue et traitée à l'échelle communale, a servi de socle à toutes les étapes suivantes.

Sur cette base, une première approche inspirée de la littérature a permis de segmenter le risque à partir d'un score. Afin de confirmer le premier zonier obtenu, la modélisation a été approfondie. Côté modèles linéaires généralisés, un MLG Bêta s'est révélé peu opérant en pratique (prédictions trop proches des bornes), tandis qu'un MLG quasi-binomial avec lien logit a offert des résultats plus stables et interprétables. Côté *machine learning*, les modèles (*Random Forest* et *XGBoost*) ont été calibrés et *backtestés*. Leurs performances prédictives de bonne qualité, mesurées par des métriques de classification usuelles tel que le  $F_1$ -score, ont conduit à privilégier l'approche *Random Forest* ( $F_1$ -score = 0,756 et AUC = 0,823) pour la suite des travaux.

La production cartographique, via le logiciel QGIS, a fourni un zonier en dix classes, pertinent pour l'analyse actuarielle. Ce zonier a ensuite été séparé en quatre classes pour l'usage commercial et pour permettre une meilleure lisibilité. Les résultats obtenus confirment un risque plus élevé que la moyenne nationale dans le Bassin aquitain et sur l'ensemble des côtes atlantique et méditerranéenne, ainsi que dans certaines zones du nord-ouest du pays. Au contraire, des régions de l'Est et les massifs apparaissent globalement moins exposées.

La dernière étape a consisté à rattacher le zonier aux objectifs économiques et à l'expérience historique de sinistralité extrême de l'assureur. La tempête Klaus, événement majeur pour le portefeuille étudié, a servi de référence pour estimer une période de retour de 14,5 ans. Ces travaux permettent d'atteindre une cible de S/P de 80 % dans le Bassin aquitain. Intégrés au zonier, ces éléments se traduisent en évolutions tarifaires déployables à l'échelle nationale par le biais de classes de risque.

Les principales conclusions sont les suivantes. D'une part, la modélisation à la maille communale, combinée au modèle *Random Forest*, améliore significativement la capacité de discrimination géographique du risque et fournit un zonier fiable utilisable en pratique pour l'assureur. D'autre part, la période de retour estimée pour Klaus, couplée à la cible S/P, a permis de recalibrer les tarifs de manière ciblée : en moyenne, une hausse de 6,95 % est obtenue dans le Bassin aquitain, ce qui était l'objectif afin d'aboutir au S/P cible.

## V.2 Mise en place opérationnelle, limites et recommandations

En pratique, le zonier devient un outil stratégique. Il permet d'identifier des zones faiblement sinistrées où un développement commercial est pertinent, bien que la prime y soit plus faible. À l'inverse, il permet de maîtriser le risque dans les zones historiquement plus touchées via une hausse des primes ou une politique plus prudente. Cet usage doit s'accompagner d'un échange systématique avec les réseaux commerciaux : une hausse trop marquée sur des zones historiques du portefeuille peut entraîner une perte importante de clients.

Au-delà de la refonte tarifaire, un assureur forestier doit être en mesure de se prémunir contre des pertes susceptibles d'excéder l'expérience historique observée (comme la tempête Klaus en 2009 dans le cas de ce mémoire).

Ces travaux ouvrent la voie à de nouvelles réflexions. En effet, il est important de souligner que la structure des forêts ainsi que les données environnementales évoluent chaque année et peuvent ainsi modifier localement le risque. Cette variabilité demande donc une réactualisation régulière des données et du zonier. Une chaîne annuelle de mise à jour pourrait capter les tendances, y compris celles liées au changement climatique, et proposer des évolutions tarifaires adaptées à une fréquence régulière. Ce cycle de refonte tarifaire doit s'appuyer sur une gouvernance de données et des objectifs clairs et stables de la part de l'assureur.

Il serait également prudent d'élargir le périmètre de marché de l'assurance forêt. Cette démarche permettrait de capitaliser sur les résultats de ce mémoire tout en limitant les biais potentiels dus au nombre important de données issues du Bassin aquitain. L'accès à des données plus générales, à l'échelle nationale ou davantage, renforcerait les conclusions et permettrait la conception de nouveaux travaux sur la base de cette étude.

Enfin, bien que la maille communale constitue un compromis original et efficace dans ce mémoire, il pourrait être intéressant d'explorer une maille plus précise de type « forêts » issue d'un référentiel forestier dédié. Obtenir des données de vent maximal à la maille communale serait également un atout certain pour la précision des études menées. Il convient également de demeurer attentif à l'intégration progressive de nouvelles variables explicatives à ce modèle afin d'en maintenir la pertinence.

Au-delà de la réponse à la problématique, ce mémoire propose une méthodologie répliquable. En partant de la collecte et de l'agrégation de plusieurs cartes à la maille communale, les travaux de modélisation du risque tempête ont mené à la construction d'un zonier. Ce dernier est à la base d'une tarification appuyée sur une période de retour des sinistres extrêmes qui challenge les valeurs usuelles. Ainsi, cette méthodologie appliquée à l'assurance forêt constitue un socle solide et transposable à d'autres objets d'étude.



# Bibliographie

- AMOUCHE L., THILL J. [2024]. « Intégration des risques de durabilité dans la gestion des risques, ce que disent les normes ». Présentation, « 100% Actuaire — Data Science — Durabilité », Institut des actuaires.
- ARNAUD D. [1997]. *L'Assurance en Grande-Bretagne : Histoire, Économie et Société*. Sous la dir. de ÉCONOMICA. Préface de François Ewald.
- AZENCOTT C.-A. [2022]. *Introduction au Machine Learning*. 2<sup>e</sup> éd. InfoSup. Version PDF accessible en ligne. Paris : Dunod. ISBN : 978-2100834761. URL : [https://cazencott.info/dotclear/public/lectures/IntroML\\_Azencott.pdf](https://cazencott.info/dotclear/public/lectures/IntroML_Azencott.pdf).
- BOUCHOUCHI I. [2024]. « Défi climatique et durabilité, vers les limites de l'assurabilité ? » Mémoire de fin d'études. Actuelia.
- BREIMAN L. [2001] « Random forests ». *Machine Learning* 45, 5-32. URL : <https://doi.org/10.1023/A:1010933404324>.
- BRUNETTE M., COUTURE S. [2023] « L'assurance contre les risques naturels en forêt : une synthèse de la littérature en économie ». *INRAE*. URL : <https://hal.inrae.fr/hal-04145961>.
- BRUNETTE M., HANEWINKEL M. [2023] « Assurance financière et assurance naturelle : une application à la forêt ». *Revue forestière française*. URL : <https://revueforestierefrancaise.agroparistech.fr/article/view/7603>.
- CCR. [2024]. « Les catastrophes naturelles en France : Bilan 1982–2023 ». URL : <https://www.ccr.fr/les-catastrophes-naturelles-en-france-bilan-1982-2023/>.
- CHATGPT BY OPENAI. « ChatGPT (o3 et 4-o) ». Utilisé pour la correction linguistique et orthographique ainsi que pour la configuration LaTeX.
- CHEN T., GUESTRIN C. [2016]. « XGBoost : A Scalable Tree Boosting System ». In : *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 785-794. URL : <https://doi.org/10.1145/2939672.2939785>.
- CLAUDE BY ANTHROPIC. « Claude Sonnet 4 ». Utilisé pour l'aide au code.
- DATA NETWORK TEAM. [2025]. « Communes — France, Jeu de données du « référentiel géographique » Opendatasoft ». Version 2025. Opendatasoft. URL : <https://public.opendatasoft.com/explore/dataset/georef-france-commune/>.
- DATA.GOUV. [2022]. « Carte des départements ». Fichier GeoJSON des départements français. En collaboration avec l'IGN. URL : <https://www.data.gouv.fr/fr/datasets/carte-des-departements-2-1/>.

- DTU WIND ENERGY. [2025]. « Global Wind Atlas ». Données développées par DTU Wind Energy (Technical University of Denmark) en partenariat avec le World Bank Group / ESMAP. URL : <https://globalwindatlas.info/en/about/TermsOfUse>.
- EIOPA. [2024]. « Fit-for-55 climate scenario analysis ». Publié le 19 Novembre 2024. URL : [https://www.eiopa.europa.eu/document/download/6a90e0a7-135e-41f0-af69-22065a0bf8cd\\_en?filename=Report%20on%20fit%20for%2055%20climate%20scenario%20analysis.pdf](https://www.eiopa.europa.eu/document/download/6a90e0a7-135e-41f0-af69-22065a0bf8cd_en?filename=Report%20on%20fit%20for%2055%20climate%20scenario%20analysis.pdf).
- FAVRE J. [2023]. « Assurance forêt : pourquoi assurer sa forêt ? » Goodassur by Selectra. URL : <https://www.goodassur.com/assurance-foret>.
- FOMEN P. [2024]. « Analyse du taux de surprime du régime des Catastrophes Naturelles et impact des tempêtes sur la tarification en assurance MRH ». Mémoire de fin d'études. Actuelia.
- FRANCE ASSUREURS. [2024]. « Cartographie prospective 2025 de l'assurance ». URL : [https://www.franceassureurs.fr/wp-content/uploads/fa\\_cartographie-prospective-2025.pdf](https://www.franceassureurs.fr/wp-content/uploads/fa_cartographie-prospective-2025.pdf).
- GÉORISQUES. [2025]. « Tempête - Un risque amplifié par le changement climatique, 2025 ». Géorisques. URL : <https://www.georisques.gouv.fr/minformer-sur-un-risque/tempele>.
- GIEC. [2023] « Résumé à l'intention des décideurs ». *Changement climatique 2023 : Rapport de synthèse*. URL : [https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC\\_AR6\\_SYR\\_SPM.pdf](https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf).
- GLOBAL MARKET INSIGHTS. [2025]. « Parametric Insurance Market Size & Share, Growth Analysis 2034 ». URL : <https://www.gminsights.com/fr/industry-analysis/parametric-insurance-market>.
- HERANVAL A., LOPEZ O., THOMAS M. [2022] « Application of machine learning methods to predict drought cost in France ». *European Actuarial Journal* 13.2, 731-753. URL : <https://doi.org/10.1007/s13385-022-00327-z>.
- HIA T. [2020]. « Exploitation de données à l'adresse pour un zonier technique climatique ». Mémoire de fin d'études. Pacifica.
- IGN. [2023]. « Forêt privée vs forêt publique : à qui appartient la forêt française ? » URL : <https://www.ign.fr/mag/foret-privee-vs-foret-publique-qui-appartient-la-foret-francaise>.
- INFOCLIMAT. [2009]. « HistorIC – Tempête Klaus ». Données de Météo-France : vitesse de rafales, détails d'évolution. URL : <https://www.infoclimat.fr/historic-details-evenement-36-tempete-klaus.html>.
- INSEE. [2016]. « Définition d'une catastrophe naturelle ». URL : <https://www.insee.fr/fr/metadonnees/definition/c1727>.

- JENKS G. F. [1967] « The Data Model Concept in Statistical Mapping ». *International Yearbook of Cartography* 7, 186-190.
- L'INTERNAUTE. [2024]. « Les départements les plus venteux de France — Classement de 1999 à 2023 ». Données issues des bases Météo-France. URL : <https://www.linternaute.com/voyage/climat/classement/departements/vent>.
- LOCDDR. [2005]. « Hurricanes Katrina and Rita Disaster Recovery Programs ». URL : <https://locddr.la.gov/programs/hurricanes-katrina-and-rita>.
- MCCULLAGH P., NELDER J. A. [1989]. *Generalized Linear Models*. 2<sup>e</sup> éd. London : Chapman et Hall. ISBN : 978-0412317606.
- MÉTÉO-FRANCE. [2024]. « Les tempêtes remarquables en France ». URL : <https://meteofrance.com/actualites-et-dossiers/comprendre-la-meteo/les-tempetes-remarquables-en-france>.
- MILLER K. F. [1985]. *Windthrow Hazard Classification*. © Crown copyright. Forestry Commission.
- NELDER J. A., MEAD R. [1965] « A simplex method for function minimization ». *The Computer Journal* 7.4, 308-313.
- NOVETHIC. [2023]. « Les cinq graphiques à retenir du dernier rapport du GIEC ». URL : <https://www.novethic.fr/actualite/environnement/climat/isr-rse/les-cinq-graphiques-a-retenir-du-dernier-rapport-du-giec-151417.html>.
- OPENDATASOFT. [2024]. « Données SYNOP essentielles - OMM ». Données Météo-France ouvertes accessibles sur la plateforme OpenDataSoft. URL : <https://public.opendatasoft.com/explore/dataset/donnees-synop-essentielles-omm/>.
- PLATEL J. [2022]. « Risque tempête : refléter la densité d'évènements dommageables à l'aide d'un indice construit sur de l'open data ». Mémoire de fin d'études. Prim'Act.
- SACRE C., MOISSELIN J.-M., SABRE M., FLORI J.-P., DUBUISSON B. [2007] « A new statistical approach to extreme wind speeds in France ». *Journal of Wind Engineering and Industrial Aerodynamics*.
- SANTÉ PUBLIQUE FRANCE. [2024] « Bilan canicule et santé : un été 2023 marqué par 4 épisodes de canicule avec un impact sanitaire important ».
- TECHNOSCIENCE. [2025]. « Bassin aquitain — Définition et Explications ». URL : <https://www.techno-science.net/glossaire-definition/Bassin-aquitain.html>.
- THOMAS M. [2024]. « Contributions à la théorie des valeurs extrêmes et à la gestion du risque ». Habilitation à Diriger des Recherches. Paris, France : Sorbonne Université.

# ANNEXES

# A - Différents types de sols au Royaume-Uni et en France

Soil type	Soil code	Score
Brown earths	1, 1d, 1u, 1z	0
Podzols	3	0
Intergrade and podzolic ironpan soils	4b, 4z	0
Ironpan soils	4	2.5
Ground water gley	5	7.5
Peaty gleys	6	10
Peaty podzolic gley	6z	7.5
Brown gleys, podzolic gleys	7b, 7z	5
Surface water gleys	7	10
<i>Juncus</i> bogs	8a, 8b, 8c, 8d	10*
<i>Molinia</i> bogs	9a, 9b, 9c, 9d, 9e	10*
<i>Sphagnum</i> bogs	10a, 10b	10*
<i>Calluna/Eriophorum/Trichophorum</i> bogs	11a, 11b, 11c, 11d	10*
Eroded bogs	14, 14h, 14w	10*
Mining spoil, stony or coarse textured	2s	0
Mining spoil, shaly or coarse textured	2m	5
Calcareous (rendzina)	12a	5
Calcareous (argillic brown earths)	12b, 12t	0
Brown and podzolic rankers	13b, 13z	7.5-10**
Scree	13s	0-10***
Peaty and gley rankers	13r, 13g, 13p	10
Well drained littoral soils	15s, 15d, 15e	0
Poorly drained littoral soils	15i	2.5
	15g	5
	15w	10

FIGURE A.1 – Typologie des sols et niveau de risque associé : Royaume-Uni (MILLER [1985])

N°	Type de sol	Risq.	N°	Type de sol	Risq.
1	Town	1	16	Kastanozem	3
2	Soil disturbed by man	7	17	Luvisol	3
3	Water body	10	18	Greyzem	4
4	Marsh	9	19	Histosol	9
5	Glacier	10	20	Podzol	6
6	Rock outcrops	8	21	Arenosol	7
7	Acrisol	6	22	Regosol	5
8	Cambisol	3	23	Solonetz	5
9	Chernozem	2	24	Andosol	4
10	Podzoluvisol	5	25	Ranker	8
11	Rendzina	5	26	Vertisol	5
12	Gleysol	8	27	Planosol	6
13	Phaeozem	2	28	Xerosol	6
14	Lithosol	9	29	Solonchak	2
15	Fluvisol	4	30	No information	–

TABLE A.1 – Typologie des sols et niveau de risque associé : France

## B - Paramètres inclus dans les fichiers de données SYNOP

Descriptif	Mnémonique	Type *	Unité
indicatif OMM station	numer_sta	car	
date (UTC)	date	car	AAAAMMDDHHMISS
pression au niveau mer	pmer	int	Pa
variation de pression en 3 heures	tend	int	Pa
type de tendance barométrique	cod_tend	int	code (0200)
direction du vent moyen 10mn	dd	int	degré
vitesse du vent moyen 10mn	ff	réel	m/s
température	t	réel	K
point de rosée	td	réel	K
humidité	u	int	%
visibilité horizontale	vv	réel	mètre
temps présent	ww	int	code (4677)
temps passé 1	w1	int	code (4561)
temps passé 2	w2	int	code (4561)
nébulosité totale	n	réel	%
nébulosité des nuages de l'étage inférieur	nbas	int	octa
hauteur de la base des nuages de l'étage inférieur	hbas	int	mètre
type des nuages de l'étage inférieur	cl	int	code (0513)
type des nuages de l'étage moyen	cm	int	code (0515)
type des nuages de l'étage supérieur	ch	int	code (0509)
pression station	pres	int	Pa
niveau barométrique	niv_bar	int	Pa
géopotentiel	geop	int	m <sup>2</sup> /s <sup>2</sup>
variation de pression en 24 heures	tend24	int	Pa
température minimale sur N heures	tnN	réel	K
température maximale sur N heures	txN	réel	K
température minimale du sol sur 12 heures	tminsol	réel	K
méthode mesure tw	sw	int	code (3855)
température du thermomètre mouillé	tw	réel	K
rafales sur les 10 dernières minutes	raf10	réel	m/s
rafales sur une période	rafper	réel	m/s
période de mesure de la rafale	per	réel	minute
état du sol	etat_sol	int	code (0901)
hauteur totale de la couche de neige, glace, autre, au sol	ht_neige	réel	mètre
hauteur de la neige fraîche	ssfrai	réel	mètre
Période de mesure de la neige fraîche	perssfrai	réel	1/10 heure
Précipitations dans les N dernières heures	rrN	réel	mm
Phénomène spécial	phenspeN	réel	code (3778)
Nébulosité couche nuageuse N	nnuageN	int	octa
Type de nuage N	ctypeN	int	code (0500)
Hauteur de base de nuage N	hnuageN	int	mètre

\* **car** : caractère ASCII, **int** : nombre entier, **réel** : nombre réel (avec décimale).

Les nombres entre parenthèses après le mot « code » sont les numéros de table de code de l'OMM (Organisation Mondiale de la Météorologie)

FIGURE B.1 – Données SYNOP : Variables disponibles

# C - Classification conifères/feuillus en fonction de la nature de l'arbre de la base de données de sinistre

Arbre	Classification	Arbre	Classification
Chênes d'Amérique	Feuillus	Pins du Sud-Ouest	Conifères
Douglas	Conifères	Pins sylvestres	Conifères
Epicéas	Conifères	DOUGLAS ET TSF	Conifères
Futaie feuillue hors peupliers	Feuillus	Peupliers	Feuillus
Futaies résineuses	Conifères	Epicéas	Conifères
Pins sans précision	Conifères	Mélange résineux divers	Conifères
Peupliers	Feuillus	Pins laricios de Corse	Conifères
Sapins F.R.	Conifères	Pins laricios	Conifères
Sapins Douglas	Conifères	Peupliers Aquitaine	Feuillus
Epicéas F.R.	Conifères	Futaie régulière de chênes	Feuillus
Pins laricios – Pins noirs d'Au.	Conifères	Douglas + Sapins	Conifères
Futaie Feuillue	Feuillus	Peupliers hors Aquitaine	Feuillus
Pin maritime (Maj. II semis/plan)	Conifères	Plantation Douglas	Conifères
Châtaigniers	Feuillus	Résineux	Conifères
PINS DE 25 À 39 ANS	Conifères	Epicéas de Sitka	Conifères
Douglas épiceas sapins	Conifères	Douglas – Epicéas	Conifères
Pins sylvestres, épiceas, sapins	Conifères	Pins maritimes + CR FAITE	Conifères
Feuillue hors peupliers	Feuillus	Epicéas sitka F.R.	Conifères
Pins noirs d'Autriche	Conifères	Sapins pectinés	Conifères
Epicéas	Conifères	Epicéas Douglas	Conifères
Pins du Sud-Ouest (cdegi)	Conifères	Résineux divers	Conifères
Epicéas-sapins	Conifères	Pins sylvestres et maritimes	Conifères
Pins laricios et douglas	Conifères	Pin maritime F.R.	Conifères
Futaies résineuses, sapins	Conifères	Pins maritimes divers	Conifères
Mel. épiceas Douglas	Conifères	Mélange feuillus divers	Feuillus
Mélange résineux	Conifères	Mélange résineux feuillus	Conifères
Pins et sapins	Conifères	Epicéas–Futaies résineuses	Conifères
Douglas + sapins	Conifères	Douglas et autres résineux	Conifères
Divers feuillus	Feuillus	Mél. épiceas sapins	Conifères
Futaies feuillues	Feuillus	Sapins + épiceas	Conifères
Peupliers reconnus agronomiques	Feuillus	Epicéas – sapins	Conifères
Mélange sapins épiceas	Conifères	Douglas / Epicéas	Conifères
Pins maritimes, Douglas	Conifères	Douglas sapins laricios	Conifères
Pins + épiceas	Conifères	Douglas et sapins	Conifères
Mélange résineux feuillu	Conifères	Douglas et épiceas	Conifères
Douglas + épiceas	Conifères	Mélange sapins et épiceas	Conifères
Douglas – Epicéas-sapins	Conifères	Mélange résineux–Douglas	Conifères
Douglas – sapins	Conifères	Epicéas et sapins	Conifères
Sapins – Epicéas	Conifères	Douglas + épiceas	Conifères
Douglas et Sapins	Conifères	Pins et douglas	Conifères
Douglas Sapins	Conifères	Sapins, épiceas	Conifères
Douglas-sapins	Conifères	Douglas sapins	Conifères
Douglas et sapins	Conifères	Mélange résineux divers	Conifères
Mélange résineux divers	Conifères	Mélange résineux résineux	Conifères
Epicéas, sapins, Douglas	Conifères	Mélange résineux et feuillus	Conifères
Résineux divers	Conifères	Mélange résineux Douglas	Conifères
Résineux	Conifères	Résineux feuillus	Conifères
Pins Sylvestres, Douglas	Conifères	Douglas et sapins divers	Conifères
Pins sylvestres douglas sapins	Conifères	Epicéas sapins douglas	Conifères

(...)

TABLE C.1 – Essences d'arbres : Liste partielle et classification

# D - Vents maximaux moyens par département

Code	Département	Wind25y_Max	Code	Département	Wind25y_Max
01	Ain	114,68	02	Aisne	114,68
03	Allier	117,72	04	Alpes-de-Haute-Provence	110,48
05	Hautes-Alpes	103,79	06	Alpes-Maritimes	102,44
07	Ardèche	116,72	08	Ardennes	108,43
09	Ariège	109,44	10	Aube	112,32
11	Aude	123,36	12	Aveyron	178,16
13	Bouches-du-Rhône	124,36	14	Calvados	134,64
15	Cantal	113,88	16	Charente	117,88
17	Charente-Maritime	129,28	18	Cher	118,84
19	Corrèze	118,84	2A	Corse-du-Sud	119,84
2B	Haute-Corse	129,16	21	Côte-d'Or	110,16
22	Côtes-d'Armor	137,44	23	Creuse	113,48
24	Dordogne	111,92	25	Doubs	105,28
26	Drôme	119,16	27	Eure	125,48
28	Eure-et-Loir	128,44	29	Finistère	154,24
30	Gard	121,68	31	Haute-Garonne	114,60
32	Gers	112,72	33	Gironde	124,36
34	Hérault	122,52	35	Ille-et-Vilaine	132,12
36	Indre	120,04	37	Indre-et-Loire	122,08
38	Isère	119,32	39	Jura	105,28
40	Landes	129,84	41	Loir-et-Cher	122,76
42	Loire	114,28	43	Haute-Loire	108,36
44	Loire-Atlantique	132,04	45	Loiret	118,32
46	Lot	111,04	47	Lot-et-Garonne	113,56
48	Lozère	121,12	49	Maine-et-Loire	123,20
50	Manche	158,20	51	Marne	126,40
52	Haute-Marne	109,12	53	Mayenne	123,60
54	Meurthe-et-Moselle	108,40	55	Meuse	103,92
56	Morbihan	146,36	57	Moselle	105,56
58	Nièvre	112,48	59	Nord	129,24
60	Oise	123,80	61	Orne	130,20
62	Pas-de-Calais	141,12	63	Puy-de-Dôme	119,88
64	Pyrénées-Atlantiques	115,88	65	Hautes-Pyrénées	109,08
66	Pyrénées-Orientales	136,36	67	Bas-Rhin	106,04
68	Haut-Rhin	104,28	69	Rhône	103,32
70	Haute-Saône	107,64	71	Saône-et-Loire	109,88
72	Sarthe	123,68	73	Savoie	106,88
74	Haute-Savoie	102,84	75	Paris	119,24
76	Seine-Maritime	134,75	77	Seine-et-Marne	113,12
78	Yvelines	109,08	79	Deux-Sèvres	115,32
80	Somme	114,88	81	Tarn	123,88
82	Tarn-et-Garonne	110,04	83	Var	122,40
84	Vaucluse	119,96	85	Vendée	129,16
86	Vienne	115,36	87	Haute-Vienne	101,24
88	Vosges	115,16	89	Yonne	109,04
90	Territoire de Belfort	99,87	91	Essonne	111,64
92	Hauts-de-Seine	107,12	93	Seine-Saint-Denis	104,76
94	Val-de-Marne	108,48	95	Val-d'Oise	111,68

TABLE D.1 – Rafale maximale moyenne sur 25 ans : Échelle départementale

## E - Statistiques descriptives générales des données

<b>Statistique</b>	<b>Valeur (en m)</b>
Moyenne	279,380
Médiane	189,108
1 <sup>er</sup> quartile	104,920
3 <sup>ème</sup> quartile	337,282
Minimum	-1,105
Maximum	2689,088
Moyenne Bassin aquitain	128,973

TABLE E.1 – Statistiques descriptives : Altitude moyenne par commune

<b>Classement</b>	<b>Sol majoritaire</b>
1	Cambisol (8)
2	Luvisol (17)
3	Rendzina (11)
4	Fluvisol (15)
5	Podzoluvisol (10)

TABLE E.2 – Statistiques descriptives : Sol majoritaire dans les communes de France métropolitaine

<b>Statistique</b>	<b>Valeur (en °)</b>
Moyenne	25,847
Médiane	18,038
1 <sup>er</sup> quartile	11,331
3 <sup>ème</sup> quartile	30,039
Minimum	0,813
Maximum	175,332
Moyenne Bassin aquitain	20,216

TABLE E.3 – Statistiques descriptives : Abri topographique par commune

Statistique	Valeur (en km/h)
Moyenne	117,227
Médiane	113,600
1 <sup>er</sup> quartile	109,070
3 <sup>ème</sup> quartile	118,060
Minimum	99,867
Maximum	178,160
Moyenne Bassin aquitain	116,630

TABLE E.4 – Statistiques descriptives : Vent maximal par département moyen sur 25 ans

Département	Commune	Ratio sinistre	Ratio prédit	Différence
Gironde (33)	ST JEAN D'ILLAC	46,88%	46,67%	-0,20%
Puy-de-Dôme (63)	PASLIERES	21,74%	21,63%	-0,11%
Creuse (23)	DONTREIX	15,72%	15,66%	-0,06%
Landes (40)	MANO	50,00%	49,98%	-0,02%
Landes (40)	CALLEN	55,13%	55,23%	+0,10%
Indre-et-Loire (37)	SENNEVIERES	45,99%	46,23%	+0,23%
Gironde (33)	ST MAGNE	40,97%	41,20%	+0,23%
Puy-de-Dôme (63)	ISSERTEAUX	19,59%	19,86%	+0,27%
Gironde (33)	ST SYMPHORIEN	40,00%	40,29%	+0,29%
Cher (18)	INEUIL	40,64%	40,93%	+0,29%

TABLE E.5 – Statistiques descriptives : Communes avec l'écart ratio sinistre vs ratio prédit via *Random Forest* le plus faible

## F - Illustrations des types de sol et des types d'arbres

Cette annexe présente une sélection de sols et d'arbres, cités et utilisés dans ce mémoire, afin de faciliter la compréhension des termes employés ainsi que des résultats obtenus.

Sols présentant un risque élevé de déracinement :



FIGURE F.1 – Illustration des types de sol : Histosol



FIGURE F.2 – Illustration des types de sol : Ranker

Sols présentant un risque faible de déracinement :



FIGURE F.3 – Illustration des types de sol : Solonchak



FIGURE F.4 – Illustration des types de sol : Cambisol

Arbres présentant un risque faible de déracinement :



FIGURE F.5 – Illustration des types d'arbres : Chêne (feuillu)

Arbres présentant un risque élevé de déracinement :



FIGURE F.6 – Illustration des types d’arbres : Sapins de Douglas (conifères)

Les images de cette annexe proviennent des ressources suivantes :

- **Histosol** : *Triple Performance*.
- **Ranker** : *Wikipedia - Photographie personnelle de Hausmaus, 2009*.
- **Solonchak** : *Photosoil*.
- **Cambisol** : *Imaggeo - European Geosciences Union*.
- **Douglas** : *Postposmo*.
- **Chêne** : *Salamandre — La revue des curieux de nature*.

# G - Choix des hyperparamètres et mise en œuvre du modèle *Random Forest*

L'optimisation de *Random Forest* repose sur l'ajustement de deux paramètres principaux :

- $B$  : nombre d'arbres dans la forêt
- $mtry$  : nombre de variables testées à chaque division

Ces paramètres influencent directement les performances du modèle et nécessitent un calibrage adapté aux données. L'optimisation permet d'améliorer significativement la capacité prédictive tout en limitant le surapprentissage.

Pour le nombre d'arbres, la valeur par défaut est de 500. L'évolution de l'erreur *Out Of Bag* (OOB) a donc été compilée selon plusieurs valeurs du nombre d'arbres (entre 50 et 1500). L'erreur OOB présente l'avantage de fournir une estimation non biaisée de l'erreur de généralisation sans nécessiter de jeu de validation séparé. L'erreur est minimale en  $B = 1000$ , comme le montre la figure G.1. Au-delà de cette valeur, les améliorations deviennent marginales tandis que le temps de calcul continue d'augmenter linéairement.

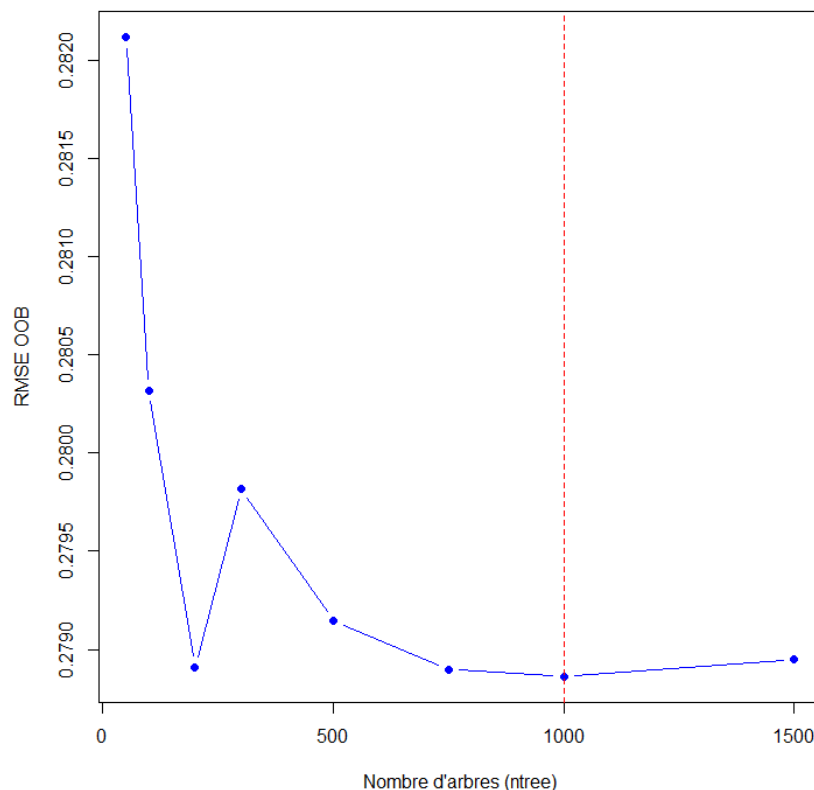


FIGURE G.1 – *Random Forest* : Évolution de l'erreur OOB selon le nombre d'arbres

De la même manière,  $mtry$  représente le nombre de variables testées à chaque division. Ce paramètre contrôle le degré de randomisation : une valeur trop faible réduit la diversité entre arbres, tandis qu'une valeur trop élevée peut conduire à une corrélation excessive entre les

prédicteurs. Le paramètre *mtry* par défaut suit la règle  $\sqrt{p}$  pour la régression, où  $p$  est le nombre de variables explicatives. Dans notre cas avec 5 variables, cela donne  $\sqrt{5} \approx 2$ . Le nombre de variables candidates à chaque nœud optimal est finalement de 2 après observation des performances pour *mtry* compris entre 1 et 5. Cette valeur confirme la pertinence de la règle théorique standard.

Un troisième hyperparamètre est considéré. Il correspond au nombre minimal d'observations trouvées par nœud terminal. Le meilleur compromis entre erreur et variance est recherché. Ce paramètre, nommé *nodesize* est optimal quand il est égal à 5.

À titre indicatif, la figure G.2 représente un arbre CART individuel extrait de notre modèle. Les prédictions finales sont obtenues par la moyenne des prédictions des  $B = 1000$  arbres qui composent le modèle *Random Forest*.

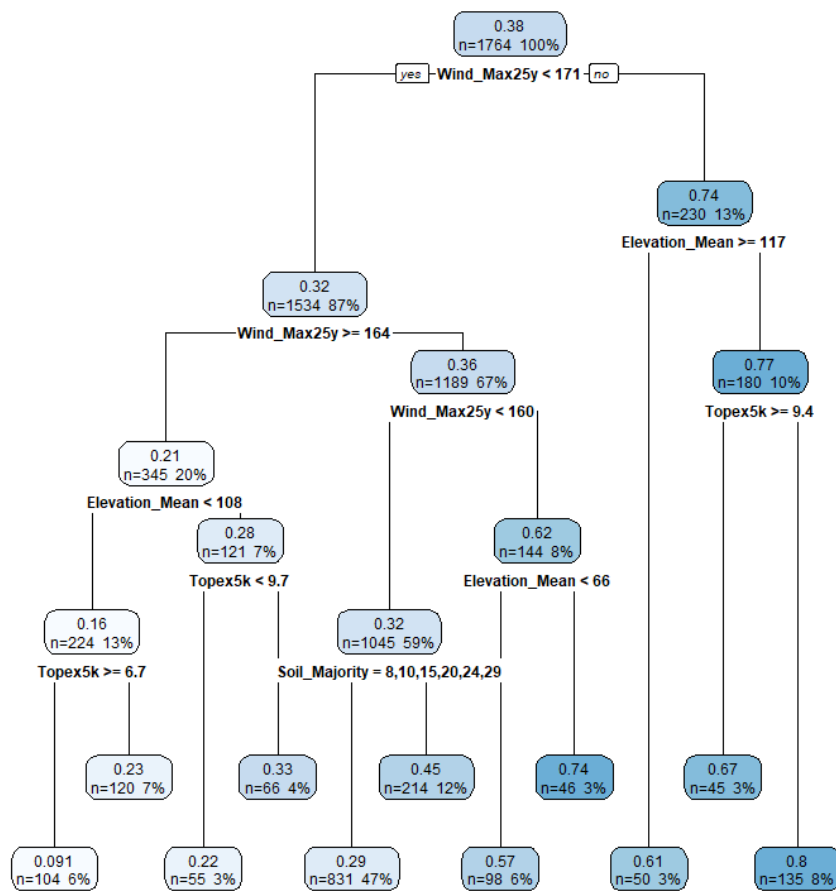


FIGURE G.2 – *Random Forest* : Arbre CART individuel

Le modèle *Random Forest* optimisé permet également d'évaluer l'importance relative des variables explicatives. Les résultats (Figure G.1) montrent que :

1. *Wind\_Max25y* (vitesse maximale du vent) : importance la plus élevée
2. *Elevation\_Mean* (altitude moyenne) : deuxième facteur explicatif
3. *Topex5k* (abri topographique) : troisième importance
4. *Soil\_Majority* (type de sol majoritaire) : influence modérée
5. *Conifer\_Bool* (essence des arbres) : importance plus faible mais significative

<b>Variable</b>	<b>%IncMSE</b>	<b>IncNodePurity</b>
Wind_Max25y	156,96	62,00
Elevation_Mean	45,58	41,18
Topex5k	41,94	36,32
Soil_Majority	40,29	13,81
Conifer_Bool	13,51	3,24

TABLE G.1 – *Random Forest* : Importance des variables

%IncMSE représente l'importance extrinsèque et IncNodePurity l'importance intrinsèque du modèle. Les résultats suivent l'intuition géographique. La vitesse maximale du vent est la plus influente sur le déracinement des arbres. Néanmoins chaque variable a son intérêt à sa manière. Il est normal qu'un booléen comme l'essence de l'arbre influe de façon moindre. Cette hiérarchisation des variables confirme l'intuition géophysique selon laquelle les facteurs météorologiques et topographiques dominent le risque de déracinement des arbres.

Le choix d'hyperparamètres adaptés et optimisés est primordial dans le cadre de la tarification du risque tempête en assurance forêt.

# H - Implémentation de l'algorithme d'optimisation de Nelder-Mead

L'algorithme de Nelder-Mead (NELDER ET MEAD [1965]) est une méthode d'optimisation sans dérivées, particulièrement adaptée à l'estimation par maximum de vraisemblance des paramètres de la distribution de Pareto généralisée dans le cadre de la théorie des valeurs extrêmes. Cette méthode présente l'avantage de ne pas nécessiter le calcul explicite du gradient de la fonction de vraisemblance, ce qui la rend robuste dans des situations où la fonction objectif peut présenter des irrégularités locales.

## Principe général :

L'algorithme cherche à minimiser une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Cela se fait en faisant évoluer un simplexe de  $n + 1$  points dans l'espace des paramètres. À chaque itération  $t$ , le simplexe  $S_t = \{x^{(1)}, \dots, x^{(n+1)}\}$  est rangé selon les valeurs de la fonction objectif  $f$  :

$$f(x^{(1)}) \leq f(x^{(2)}) \leq \dots \leq f(x^{(n+1)}),$$

où  $x^{(1)}$  représente le meilleur point et  $x^{(n+1)}$  le plus mauvais.

## Fonction objectif (GPD) et contraintes :

Soient un seuil  $u$  et des excès  $z_i = y_i - u > 0$  pour  $i = 1, \dots, N$ . Nous estimons  $(\xi, \sigma_{\text{GPD}})$  par maximum de vraisemblance en *minimisant* la négative log-vraisemblance *normalisée* :

$$\bar{\mathcal{L}}(\xi, \sigma_{\text{GPD}}) = \frac{1}{N} \sum_{i=1}^N \left[ \log \sigma_{\text{GPD}} + \left(1 + \frac{1}{\xi}\right) \log \left(1 + \frac{\xi z_i}{\sigma_{\text{GPD}}}\right) \right],$$

définie sur le domaine  $\sigma_{\text{GPD}} > 0$  et  $1 + \frac{\xi z_i}{\sigma_{\text{GPD}}} > 0 \forall i$  (la limite  $\xi \rightarrow 0$  redonne le cas exponentiel).

Pour garantir  $\sigma_{\text{GPD}} > 0$ , nous ré-paramétrisons  $\sigma_{\text{GPD}} = \exp(\theta)$  et optimisons sur  $x = (\xi, \theta)$ . Nous utilisons une barrière de domaine :

$$f(x) = \begin{cases} \bar{\mathcal{L}}(\xi, e^\theta) & \text{si } 1 + \xi z_i e^{-\theta} > 0 \forall i, \\ +\infty & \text{sinon.} \end{cases}$$

*Optionnel* : remplacer la barrière par une pénalisation quadratique  $\lambda \sum_{i=1}^N [\min\{0, 1 + \xi z_i e^{-\theta}\}]^2$  avec  $\lambda$  grand.

## Opérations géométriques :

Quatre opérations géométriques sont utilisées, avec les paramètres  $\alpha > 0$ ,  $\gamma > 1$ ,  $0 < \rho < 1$  et  $0 < \sigma < 1$  :

- **Réflexion** :  $x_r = \bar{x} + \alpha(\bar{x} - x^{(n+1)})$
- **Expansion** :  $x_e = \bar{x} + \gamma(x_r - \bar{x})$
- **Contraction extérieure** :  $x_{oc} = \bar{x} + \rho(x_r - \bar{x})$
- **Contraction intérieure** :  $x_{ic} = \bar{x} - \rho(x_r - \bar{x})$

- **Réduction** :  $x^{(i)} \leftarrow x^{(1)} + \sigma(x^{(i)} - x^{(1)})$

où  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x^{(i)}$  est la moyenne des  $n$  meilleurs points.

### Règles de décision :

L'algorithme suit une séquence de tests. Le but est de trouver quelle séquence adopter.

1. Si  $f(x_r) < f(x^{(1)})$  : conserver  $\arg \min\{f(x_e), f(x_r)\}$  ,
2. Si  $f(x^{(1)}) \leq f(x_r) < f(x^{(n)})$  : accepter,
3. Si  $f(x^{(n)}) \leq f(x_r) < f(x^{(n+1)})$  : contraction extérieure si une amélioration est notée, sinon réduction,
4. Sinon : contraction intérieure si une amélioration est notée, sinon réduction.

### Critères d'arrêt (valeurs retenues) :

Nous travaillons sur l'objectif *normalisé*  $\bar{\mathcal{L}}$  (indépendant de  $N$ ) et fixons :

- **Taille du simplexe** :  $\max_j \left( \max_i x_j^{(i)} - \min_i x_j^{(i)} \right) \leq \text{xtol}$ , avec  $\text{xtol} = 10^{-6}$  sur  $(\xi, \theta)$  ;
- **Dispersion des valeurs** :  $\text{sd}(f(x^{(1)}), \dots, f(x^{(n+1)})) \leq \text{ftol}$ , avec  $\text{ftol} = 10^{-8}$  sur  $\bar{\mathcal{L}}$  ;
- **Nombre maximal d'itérations** : arrêt si  $\text{maxit} = 5000$  est atteint.

Cette implémentation de l'algorithme de Nelder-Mead permet d'obtenir des estimations fiables des paramètres de la distribution GPD, essentielles pour le calcul des périodes de retour présentées dans la partie IV de ce mémoire.

# Table des figures

1	<i>Random Forest</i> : Courbe précision-rappel . . . . .	VIII
2	Zonier agrégé en 10 zones : <i>Random Forest</i> . . . . .	IX
3	Zonier agrégé en 12 zones : <i>Random Forest</i> . . . . .	IX
4	Analyse des stations : Corrélation . . . . .	XI
5	Détermination du nouveau tarif : Schéma . . . . .	XII
6	Récapitulatif du mémoire : Schéma . . . . .	XIII
7	Random Forest : Precision-recall curve . . . . .	XVIII
8	Zoning map aggregated into 10 areas : <i>Random Forest</i> . . . . .	XIX
9	Zoning map aggregated into 12 areas : <i>Random Forest</i> . . . . .	XIX
10	Station analysis : Correlation . . . . .	XXI
11	New Rate Determination : Diagram . . . . .	XXII
12	Summary of the thesis : Diagram . . . . .	XXIII
I.1	Température : Évolution selon les scénarios SSP (NOVETHIC [2023]) .	6
I.2	Objectifs d'émission : Insuffisance par rapport aux objectifs (NOVETHIC [2023]) . . . . .	7
I.3	Sinistralité hors sécheresse par gravité : Évolution (en M €) (CCR [2024])	8
I.4	Score fréquence-sévérité : Cinq risques les plus élevés pour le secteur de l'assurance en France (FRANCE ASSUREURS [2024]) . . . . .	9
I.5	Menaces majeures pour la société française : Vision assurance/réassurance (FRANCE ASSUREURS [2024]) . . . . .	9
I.6	Catastrophes naturelles en France : Schéma d'indemnisation (CCR [2024])	10
I.7	Tempêtes en France métropolitaine : Quarante événements majeurs (MÉTÉO-FRANCE [2024]) . . . . .	15
I.8	Stations météorologiques en France métropolitaine : Implantation géographique (OPENDATASOFT [2024]) . . . . .	16
I.9	Pourcentage de la surface forestière publique par département : Carte (IGN [2023]) . . . . .	18
II.1	Étude britannique : Zonier du vent . . . . .	24
II.2	Étude britannique : Zonier du niveau de risque dans une forêt . . . . .	26
II.3	QGIS : Aperçu de l'interface . . . . .	28
II.4	QGIS : Extrait de la base de données communale . . . . .	29
II.5	Bassin aquitain : Départements concernés . . . . .	30
II.6	Coût des sinistres : Évolution annuelle . . . . .	31
II.7	Nombre de sinistres par commune : Zonier . . . . .	32
II.8	Primes reçues : Évolution annuelle . . . . .	34
II.9	Altitude et bathymétrie en France métropolitaine : Carte . . . . .	36
II.10	Altitude et communes : Cartes superposées . . . . .	36
II.11	Altitude par commune en France métropolitaine : Zonier . . . . .	37
II.12	Sol majoritaire en France métropolitaine : Carte . . . . .	37
II.13	Sol majoritaire en France métropolitaine : Légende . . . . .	38
II.14	Abri topographique à 5 kilomètres en France métropolitaine : Carte .	39
II.15	Abri topographique par commune en France métropolitaine : Zonier .	40
II.16	Types d'arbre par forêt en France métropolitaine : Carte . . . . .	41
II.17	Vent maximal moyen sur 25 ans en France métropolitaine : Zonier . .	43
III.1	Zonier agrégé en 10 zones : Score . . . . .	48

III.2	Ratios de destruction de la base de sinistres : <i>Boxplot</i> . . . . .	51
III.3	MLG Quasi-binomial : Schéma explicatif . . . . .	53
III.4	MLG Quasi-binomial : Courbe précision-rappel . . . . .	57
III.5	Apprentissage supervisé : Fonctionnement (AZENCOTT [2022]) . . . . .	60
III.6	<i>Random Forest</i> : Courbe précision-rappel . . . . .	62
III.7	<i>XGBoost</i> : Courbe précision-rappel . . . . .	64
III.8	<i>Random Forest</i> : Évolution du RMSE et du $R^2$ par pli . . . . .	65
III.9	<i>Random Forest</i> : Valeurs observées et valeurs prédites . . . . .	66
III.10	Zonier agrégé en 10 zones : <i>Random Forest</i> . . . . .	67
III.11	Zonier agrégé en 10 zones : <i>XGBoost</i> . . . . .	68
III.12	Zonier agrégé en 10 zones : MLG Quasi-binomial . . . . .	68
III.13	Zonier agrégé en 4 zones : <i>Random Forest</i> . . . . .	69
IV.1	Tempête Klaus (24 janvier 2009) : Rafales journalières (INFOCLIMAT [2009]) . . . . .	74
IV.2	Station 07314 : <i>Mean Excess Plot</i> . . . . .	76
IV.3	Station 07314 : <i>Threshrange Plot</i> . . . . .	76
IV.4	Station 07314 : Quantile-Quantile Plot . . . . .	76
IV.5	Station 07314 : Niveaux de retour . . . . .	77
IV.6	Station 99999 : <i>Mean Excess Plot</i> . . . . .	77
IV.7	Station 99999 : Série temporelle . . . . .	77
IV.8	Station 99999 : Niveaux de retour . . . . .	78
IV.9	Analyse des stations : Représentation des distributions uniformisées . . . . .	79
IV.10	Analyse des stations : Corrélation . . . . .	80
IV.11	Détermination du nouveau tarif : Schéma . . . . .	83
IV.12	Synthèse des évolutions tarifaires : Zonier . . . . .	84
IV.13	Synthèse des évolutions tarifaires : Légende . . . . .	84
A.1	Typologie des sols et niveau de risque associé : Royaume-Uni (MILLER [1985]) . . . . .	94
B.1	Données SYNOP : Variables disponibles . . . . .	95
F.1	Illustration des types de sol : Histosol . . . . .	100
F.2	Illustration des types de sol : Ranker . . . . .	100
F.3	Illustration des types de sol : Solonchak . . . . .	101
F.4	Illustration des types de sol : Cambisol . . . . .	101
F.5	Illustration des types d'arbres : Chêne (feuillus) . . . . .	101
F.6	Illustration des types d'arbres : Sapins de Douglas (conifères) . . . . .	102
G.1	<i>Random Forest</i> : Évolution de l'erreur OOB selon le nombre d'arbres . . . . .	103
G.2	<i>Random Forest</i> : Arbre CART individuel . . . . .	104

# Liste des tables

1	Variables retenues : Explication technique . . . . .	VI
2	MLG Quasi-binomial : Mesures de performance . . . . .	VII
3	<i>Random Forest</i> : Mesures de performance . . . . .	VII
4	<i>Random Forest</i> : Moyenne prédite vs Moyenne observée du ratio . . . . .	VIII
5	Ratios S/P : Distinction entre attritionnel et catastrophe . . . . .	X
6	Bassin aquitain : Évolutions tarifaires par classe . . . . .	XII
7	Variables considered in this thesis : Technical details . . . . .	XVI
8	Quasi-binomial GLM : Performance metrics . . . . .	XVII
9	Random Forest : Performance metrics . . . . .	XVII
10	<i>Random Forest</i> : Average predicted vs. average observed ratio . . . . .	XVIII
11	Loss ratios : Distinction between attritional and catastrophe . . . . .	XX
12	Aquitaine Basin : Pricing changes by class . . . . .	XXII
I.1	Indicateurs climatiques : Usages actuariels . . . . .	5
I.2	Scénarios SSP à horizon 2100 : Synthèse (GIEC [2023]) . . . . .	6
I.3	Régime Cat-Nat : Objectifs et mécanismes associés . . . . .	10
I.4	Échelle de Beaufort : Correspondance force du vent / vitesse . . . . .	13
I.5	Tempêtes majeures mondiales : Caractéristiques . . . . .	14
I.6	Tempêtes majeures en France métropolitaine : Caractéristiques . . . . .	14
I.7	Données SYNOP : Pourcentages de valeurs manquantes (FOMEN [2024]) . . . . .	16
I.8	Groupama Forêts Assurances : Fiche d'identité . . . . .	20
II.1	Étude britannique : Score associé au vent . . . . .	24
II.2	Étude britannique : Score associé à l'altitude . . . . .	24
II.3	Étude britannique : Score associé à l'abri topographique . . . . .	25
II.4	Étude britannique : Score associé à la capacité d'enracinement . . . . .	25
II.5	Étude britannique : Classes de hasard . . . . .	26
II.6	Variables retenues : Zonier par score . . . . .	27
II.7	QGIS : Comparaison de valeurs pour trois communes françaises . . . . .	28
II.8	Base sinistres : Exemple de données d'entrée . . . . .	30
II.9	Coût des sinistres : Évolution annuelle . . . . .	30
II.10	Nombre de sinistres : Évolution annuelle . . . . .	31
II.11	Nombre de sinistres : Cinq communes les plus sinistrées . . . . .	32
II.12	Surfaces assurées et sinistrées : Moyennes . . . . .	32
II.13	Travaux sur la base : Gestion de deux communes avec le même nom . . . . .	33
II.14	Sinistres, primes et ratios S/P : Évolution annuelle . . . . .	34
II.15	Ratios S/P : Distinction entre attritionnel et catastrophe . . . . .	35
II.16	Exemple de trois communes françaises : Altitude moyenne . . . . .	36
II.17	Exemple de trois communes françaises : Type de sol majoritaire . . . . .	38
II.18	Exemple de trois communes françaises : Abri topographique . . . . .	39
II.19	Peuplements forestiers : Typologie et classification . . . . .	41
II.20	Exemple de trois communes françaises : Essence d'arbre majoritaire . . . . .	42
II.21	Exemple de trois communes françaises : Vent maximal moyen sur 25 ans . . . . .	43
III.1	Score associé à l'essence majoritaire : Étude menée . . . . .	47
III.2	Scores associés au vent, au TOPEX et à l'altitude : Étude menée . . . . .	47
III.3	Exemple de trois communes françaises : Synthèse des scores . . . . .	47
III.4	MLG : Fonctions de lien classiques . . . . .	49

III.5	Ratio de sinistres : Répartition . . . . .	51
III.6	MLG Quasi-binomial : ANOVA . . . . .	53
III.7	MLG Quasi-binomial : Indices de variance d'inflation généralisés . . . . .	53
III.8	MLG Quasi-binomial : Déviance et dispersion . . . . .	54
III.9	MLG Quasi-binomial : Coefficients . . . . .	54
III.10	MLG Quasi-binomial : Statistiques descriptives de <i>Ratio_SinAss</i> . . . . .	55
III.11	MLG Quasi-binomial : Ratio de destruction - Exemple de trois communes françaises . . . . .	55
III.12	Exemple de trois communes françaises : Données . . . . .	56
III.13	MLG Quasi-binomial : Effectifs et pourcentages de la matrice de confusion . . . . .	57
III.14	MLG Quasi-binomial : Mesures globales de performance . . . . .	57
III.15	MLG Quasi-binomial : Validation par trois stratégies d'évaluation . . . . .	58
III.16	MLG Quasi-binomial : Moyenne prédite vs Moyenne observée du ratio . . . . .	59
III.17	Comparaison des performances : MLG vs <i>Random Forest</i> . . . . .	61
III.18	<i>Random Forest</i> : Effectifs et pourcentages de la matrice de confusion . . . . .	61
III.19	<i>Random Forest</i> : Mesures globales de performance . . . . .	61
III.20	<i>XGBoost</i> : Hyperparamètres optimaux par optimisation bayésienne . . . . .	63
III.21	Comparaison des performances : MLG vs <i>Random Forest</i> vs <i>XGBoost</i> . . . . .	63
III.22	<i>XGBoost</i> : Effectifs et pourcentages de la matrice de confusion . . . . .	64
III.23	<i>XGBoost</i> : Mesures globales de performance . . . . .	64
III.24	Comparaison des performances : <i>Random Forest</i> basique et optimisé . . . . .	65
III.25	<i>Random Forest</i> : Moyenne prédite vs Moyenne observée du ratio . . . . .	65
III.26	Comparaison des performances : Métriques finales pour les trois modèles . . . . .	66
III.27	<i>Random Forest</i> : Statistiques descriptives de <i>Ratio_SinAss</i> . . . . .	66
III.28	<i>Random Forest</i> : Ratio de destruction - Exemple de trois communes françaises . . . . .	67
IV.1	Ratios S/P : Distinction entre attritionnel et catastrophe . . . . .	71
IV.2	Répartition d'une forêt : Exemple pour tarification . . . . .	72
IV.3	Prix par hectare : Exemple pour tarification . . . . .	73
IV.4	Base SYNOP : Stations du Bassin aquitain . . . . .	74
IV.5	Analyse des stations : Vérification des marginales uniformes . . . . .	80
IV.6	Analyse des stations : Choix de la copule adaptée . . . . .	80
IV.7	Période de retour : Intersection . . . . .	81
IV.8	Période de retour : Union . . . . .	81
IV.9	Bassin aquitain : Moyennes de ratios par type d'essence . . . . .	82
IV.10	Bassin aquitain : Différences à la moyenne par classe . . . . .	82
IV.11	Bassin aquitain : Évolutions tarifaires par classe . . . . .	83
IV.12	Exemple de trois communes françaises : Synthèse . . . . .	85
A.1	Typologie des sols et niveau de risque associé : France . . . . .	94
C.1	Essences d'arbres : Liste partielle et classification . . . . .	96
D.1	Rafale maximale moyenne sur 25 ans : Échelle départementale . . . . .	97
E.1	Statistiques descriptives : Altitude moyenne par commune . . . . .	98
E.2	Statistiques descriptives : Sol majoritaire dans les communes de France métropolitaine . . . . .	98
E.3	Statistiques descriptives : Abri topographique par commune . . . . .	98
E.4	Statistiques descriptives : Vent maximal par département moyen sur 25 ans . . . . .	99

E.5	Statistiques descriptives : Communes avec l'écart ratio sinistre vs ratio prédit via <i>Random Forest</i> le plus faible . . . . .	99
G.1	<i>Random Forest</i> : Importance des variables . . . . .	105