

Mémoire présenté devant l'ENSAE Paris  
pour l'obtention du diplôme de la filière Actuariat  
et l'admission à l'Institut des Actuaires  
le 09/03/2023

Par : **Alexandra DOUTHAUD**

Titre : **Analyse de la survenance d'un arrêté Cat-Nat de type  
submersion marine face au changement climatique à partir  
de l'Open Data et du Machine Learning**

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

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

*Membres présents du jury de la filière :*

*Nom : Christian-Yann Robert*

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

*Antoine Mantel  
Julie Surget Sun  
Nicolas Zec*

*Entreprise : Deloitte **D.***

*Signature : CYRIL CHALIN*

*Directeur du mémoire en entreprise :*

*Nom : Joseph DELAWARI*

*Signature :*

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

Secrétariat :

Bibliothèque :

Signature du responsable entreprise

CYRIL CHALIN

Signature du candidat

*Douthaud*

---

## Résumé

Les autorités de régulation et de supervision souhaitent accélérer la transformation de la prise en compte du risque climatique par les sociétés d'assurance. Pour ce faire, certains superviseurs ainsi que certains régulateurs développent progressivement des exercices de tests de résistance liés au climat. Il s'agit de s'assurer de la résilience du secteur et d'accélérer la mise en œuvre de futurs exercices obligatoires. L'objectif pour les assureurs est alors de pouvoir démontrer comment ils pourront aligner leurs stratégies aux défis climatiques.

Le risque d'inondation est l'un des risques des Catastrophes Naturelles (Cat-Nat) auquel la population et les compagnies d'assurance françaises sont le plus exposées. Parmi les types d'inondations, la submersion marine est peu étudiée dans la littérature actuarielle. Il n'existe pas de base de données complète et accessible rassemblant les variables explicatives dues aux inondations par submersion marine en France. Cependant, son étude est de plus en plus pertinente et devient un sujet crucial dans le cadre du changement climatique. Ce type d'inondation est rare et très dangereux, principalement en raison de l'interdépendance des phénomènes tels que le vent, la pression atmosphérique, et le niveau de la mer. Cette interdépendance en fait un événement très difficile à prédire.

Cette étude présente une première approche permettant d'appréhender ce risque. Le modèle issu de cette étude repose sur l'évolution des inondations par submersion marine à partir de la fréquence des arrêtés Cat-Nat sur le littoral atlantique français. En d'autres termes, le but de ce mémoire est d'évaluer la fréquence des submersions marines à horizon 2050 au travers d'une prédiction des arrêtés Cat-Nat avec une prise en compte du changement climatique. Les conditions climatiques futures sont déterminées dans une large mesure par les émissions actuelles et futures de gaz à effet de serre. Une trajectoire de concentration représentative (RCP) pour chacun des scénarios d'émissions, RCP 4.5 (*Business as Usual*) et 8.5 (plus pessimiste), du Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC) ont été utilisées.

Avant de pouvoir réaliser un modèle reposant sur des méthodes de *machine learning*, le premier objectif est de construire une base à partir de données intégralement libres d'accès. Le second objectif est ensuite d'utiliser ces modèles afin de prédire la survenance des futurs arrêtés Cat-Nat. La variable modélisée est la déclaration/occurrence d'un arrêté Cat-Nat pour un jour J et un canton C. Le calibrage du modèle a nécessité la constitution d'une base de données d'apprentissage comprenant l'historique de 1990 à 2005 de l'ensemble des variables. Les projections de 2006 à 2020 des variables explicatives sont ensuite utilisées en entrée du modèle afin de déterminer la fréquence des submersions sur cette période.

Afin de valider le modèle, une base de validation historique de la variable à expliquer de 2006 à 2020 a été exploitée. A partir d'une validation statistique, le modèle de *Gradient Boosting* a été retenu parmi trois méthodes d'algorithmes supervisés. Il a alors pu être utilisé pour projeter la fréquence des arrêtés Cat-Nat à horizon 2050 sur le littoral Atlantique français. Il apparaît que les départements les plus touchés par la tempête Xynthia de 2010 (Charente-Maritime et Vendée) restent très exposés au risque de submersion marine. L'ensemble de ces éléments ouvrent la voie à l'étude de ce risque atypique et à l'émergence de bonnes pratiques pour le secteur. En effet, la comparaison de RCP est liée à l'étude de scénarios qui devrait être généralisée dans les pratiques de gestion des risques.

*Mots clés : Submersion marine, changement climatique, aléa, machine learning, données libres d'accès, Copernicus*

## Abstract

Regulators and supervisors want to accelerate the transformation of insurance companies' consideration of climate risk. To this end, several of them are gradually developing climate-related stress testing exercises. The aim is to ensure the resilience of the sector and accelerate the implementation of future mandatory exercises. The objective for insurers is then to be able to demonstrate how they will be able to align their strategies to climate challenges.

Flood risk is one of the Natural Catastrophes (Nat-Cat) to which the French population and insurance companies are most exposed. On the other hand, marine submersion remains a type of flooding that is hardly studied in the actuarial domain. Moreover, there is no complete or accessible database that gathers the explanatory variables of marine flooding in France. However, its study is increasingly relevant and is a crucial topic in the sphere of climate change. This type of flooding is rare and very dangerous, mainly due to the interdependence of different phenomena such as the wind, the atmospheric pressure, and the sea level, and it is this interdependence that makes it all the more difficult to predict.

This study presents a first approach to understand this risk. The model proposed in this study uses Nat-Cat decrees to construct the frequency of coastal flooding on the French Atlantic coast. In other words, the aim of this thesis is to evaluate the frequency of marine flooding throughout the years, until 2050. The modeled variable is the occurrence of a Nat-Cat decree for a day  $D$  and a canton  $C$ . This is accomplished by using two different climate change scenarios to predict the previously established Nat-Cat decrees. One trajectory for each of the Representative Concentration Pathways (RCP), the RCP 4.5 (Business as Usual) and 8.5 (more pessimistic), emissions scenarios of the Intergovernmental Panel on Climate Change (IPCC) were used.

Before being able to implement a model based on machine learning methods, the first objective is to build a database from open source data. The second objective is then to use these models to predict the occurrence of future Nat-Cat events. The calibration of the model required the establishment of a training database with data between 1990 to 2005 for all the variables. Projections of the explanatory variables from 2006 to 2020 are then used as an input to the model in order to determine the frequency of flooding over this period.

In order to validate the model, a test was carried on the historical database from 2006 to 2020 of the variable to be explained. Using a statistical test for validation, the Gradient Boosting model was selected among three supervised algorithms methods. It could then be used to project the frequency of Nat-Cat decrees by 2050 in the French Atlantic coast.

It appears that the departments most affected by the Xynthia storm of 2010 (Charente-Maritime and Vendée) remain highly exposed to the risk of coastal flooding. All of these elements open the way to the study of this atypical risk and the emergence of good practices for the sector. Indeed, the RCPs comparison is linked to the scenario study that should be generalized in risk management practices.

*Key words : Coastal flood, Copernicus, climate change, hazard, machine learning, open data*

## Synthèse

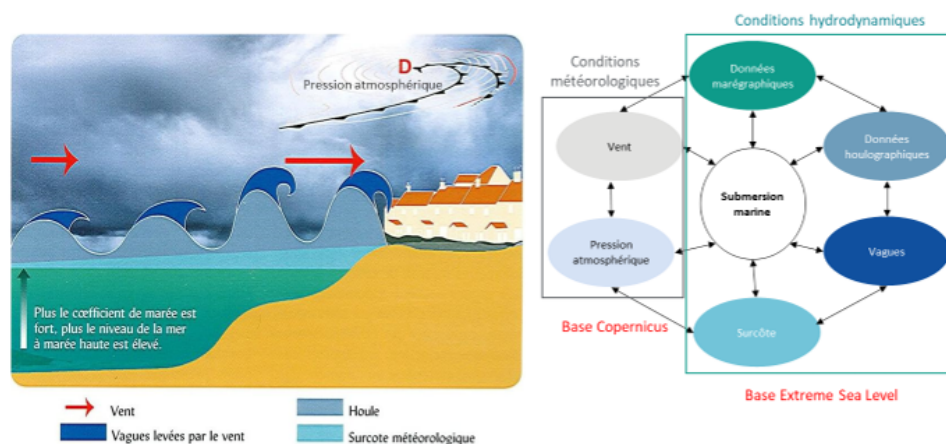
Sur ces dernières décennies, les activités productives humaines ont connu une accélération exponentielle. Cette véritable explosion de la croissance s'est accompagnée d'une exploitation massive des énergies carbonées. En effet, les émissions de gaz à effet de serre (**GES**) sont intrinsèquement liées au développement des sociétés modernes.

En 1988, le Programme des Nations Unies pour l'environnement et l'Organisation météorologique mondiale a créé un groupe international d'experts du climat, le **GIEC**. Son objectif est de rassembler et de synthétiser des milliers de travaux scientifiques sur les changements climatiques, leurs causes et leurs répercussions potentielles. L'augmentation des concentrations des **GES** est partagée par la quasi-totalité de la communauté des climatologues. En outre, il existe des impacts visibles indéniables liés au dérèglement climatique tels que le recul des glaciers ou encore la perte de masse des calottes glaciaires. La hausse des températures est donc vérifiée et contribue à l'élévation du niveau des mers.

Dans le cadre du changement climatique, les événements extrêmes ou catastrophes naturelles (Cat-Nat) risquent fort de se révéler de plus en plus fréquents. Même si ces événements restent rares, il est essentiel de les projeter sur un horizon de temps assez long afin de suivre l'augmentation de celles-ci. Dans la perspective de l'élaboration du 5e rapport d'évaluation du **GIEC**, quatre scénarios de référence (ou **RCP**, pour *Representative Concentration Pathways*) de **GES** ont été effectués pour le XXIe siècle et au-delà. Ces scénarios correspondent à l'ampleur des efforts à réaliser pour réduire les émissions de **GES** au niveau mondial.

Selon le **GIEC**<sup>1</sup>, le niveau marin à horizon 2050 serait en hausse de 20 cm pour un scénario sans changement (*business as usual* ou **RCP 4.5**) et 23 cm pour le scénario le plus pessimiste (**RCP 8.5**). Ces hausses éventuelles sont directement liées à la fréquence des catastrophes naturelles de type submersion marine. Il s'agit d'un type d'inondation présent dans les zones côtières et causé par des conditions marégraphiques et météorologiques particulièrement défavorables.

Pour pouvoir correctement les modéliser, il est indispensable de comprendre leurs causes. La représentation suivante permet de représenter le phénomène de manière schématique :



**Figure 1** – Les paramètres liés à la submersion marine d'après les données REFMAR

1. IPCC, 2014 : Climate Change 2014 : Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

La qualité des résultats de la modélisation dépend des données de construction du modèle. Pour cette étude elles sont de deux types :

- Les conditions météorologiques proviennent des résultats de l'initiative **EURO-CORDEX** (Base Copernicus). Ce programme regroupe le travail de diverses institutions européennes afin de construire une série de modèles climatiques particulièrement adaptés à cette zone géographique. Les modèles utilisés se basent sur plusieurs équations qui cherchent à représenter les différentes relations existantes entre l'atmosphère, les océans et la terre ferme. La vitesse du vent et la pression atmosphérique moyenne au niveau de la mer ont alors été sélectionnées dans cette base de données.
- Les conditions hydrodynamiques quant à elles sont extraites de la base *Extreme Sea Level*. Les projections sont basées sur différents travaux de la *National Oceanic and Atmospheric Administration*. Différentes variables sont extraites comme la hauteur des vagues, ou l'importance de la surcôte.

Les projections des données sont réalisées sur plusieurs scénarios :

- Historique : les données telles qu'observées dans le passé grâce à plusieurs stations dotées de capteurs installées tout au long de la côte.
- **RCP 4.5** : projection sur le futur en supposant une augmentation relativement contrôlée des températures.
- **RCP 8.5** : projection sur un futur conservant le même modèle de croissance et donc un impact important sur les températures futures.

Ces données climatiques sont libres d'accès et ont été récupérées via des API<sup>2</sup>. Une fois les données traitées, les modèles ont été construits.

L'objectif recherché est la construction d'un modèle de projection de la survenance des submersions marines à l'horizon 2050 sur le littoral atlantique français. La variable modélisée est la déclaration/occurrence d'un arrêté Cat-Nat pour un jour J et un canton C. Le schéma suivant permet de résumer la modélisation effectuée :

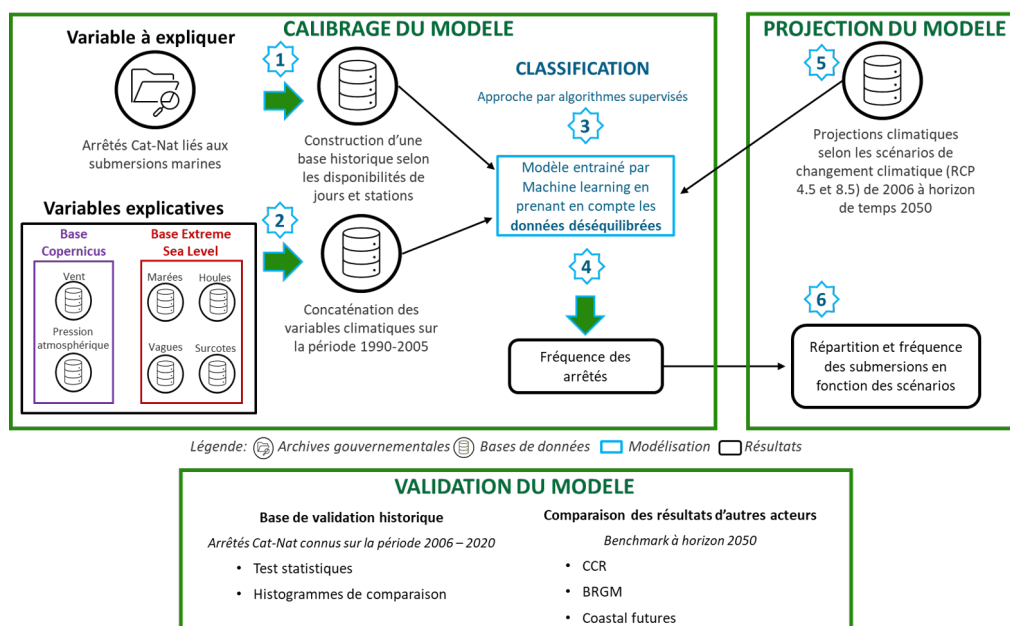


Figure 2 – Synthèse de méthodologie

2. Application Programming Interface ou interface de programmation. Cette technique permet à deux applications distinctes telles qu'une page internet et une interface python d'échanger des données de manière automatique.

Les étapes de modélisation peuvent être résumées de la façon suivante :

1. **La construction de la base de données contenant la variable à expliquer** selon les maillages temporels (journalier) et géographique choisis (canton). À noter que la maille choisie pour le modèle dépend de la précision des données à disposition. En utilisant des modèles **RCM**, il n'est pas possible d'utiliser une maille géographique plus fine que le canton. La maille temporelle journalière est quant à elle justifiée par l'étude de la durée d'un événement submersion. Cette variable est constituée des demandes d'arrêtés Cat-Nat fournies par la **CCR**. La publication d'un tel arrêté au Journal Officiel est une condition nécessaire à l'intervention de la **CCR** et si besoin de l'état. Cette décision dépend d'une commission interministérielle qui se réunit à la demande de ou des mairies impactées localement. En effet, dans le système français, un partenariat est formé entre l'état, la **CCR** et les compagnies d'assurance en cas d'événements identifiés par l'état comme des catastrophes naturelles. Les données recueillies dans cette base de données vont de 1990 à 2020.
  
2. **La construction de la base de données contenant les variables explicatives** sélectionnées à partir des paramètres constitutifs d'une submersion marine. Le même maillage temporel (journalier) et géographique (canton) est à nouveau appliqué. Ces variables sont constituées des conditions météorologiques et conditions hydrodynamiques décrites précédemment. Les données recueillies dans cette base de données vont de 1990 à 2005.
  
3. **La construction de quatre bases de données finales.** La première est une base historique allant de 1990 à 2005 contenant l'ensemble des variables pour le modèle d'apprentissage. La seconde est une base historique de validation allant de 2006 à 2020 sur la variable à expliquer. Elle permet de vérifier le meilleur modèle de *machine learning* à partir de données réelles. La troisième est une base de variables explicatives pour le scénario **RCP 4.5** à horizon 2050. La dernière est une base de variables explicatives pour le scénario **RCP 8.5** à horizon 2050. Ces deux dernières bases sont donc utilisées pour la projection du modèle de *machine learning*.  
La construction de ces bases nécessite de réaliser divers traitements avant de pouvoir les utiliser. Ainsi, un travail de localisation des stations ou encore des traitements liés au caractère déséquilibré et saisonnier des données a été effectué.
  
4. **L'utilisation de trois modèles de *machine learning* pour classifier les demandes d'arrêtés Cat-Nat :**
  - Le premier modèle est le *Gradient Boosting (XGB)*. Cet algorithme pondère les observations au fur et à mesure de l'apprentissage. L'amélioration itérative des prédictions permet d'obtenir un optimum en sommant les différents arbres de décision construits.
  - Le second est l'algorithme *Random Forest*. La classification effectuée se base sur un regroupement, ou « forêt », d'arbres de décisions indépendants. Les données d'apprentissage sont sous divisées en plusieurs ensembles sélectionnés aléatoirement pour construire les arbres.
  - Le troisième est le modèle *Support Vector Machines (SVM)*. Cet algorithme cherche à classifier les données en les distinguant à l'aide d'une frontière. Elles sont représentées dans un espace à  $n$  dimensions et la frontière est construite grâce à un hyperplan.

Les modèles sont entraînés sur une base d'apprentissage couvrant la période 1990 à 2005. Les projections de 2006 à 2020 des variables explicatives sont ensuite utilisées en entrée du modèle afin de déterminer la fréquence des submersions sur cette période.

Afin de pouvoir valider le modèle, une base de données sur la variable à expliquer est disponible sur la période 2006 à 2020. Elle va permettre de choisir le meilleur modèle en comparant cette réalité avec les prédictions faites par les trois modèles évoqués.

La validation du modèle se déroule ensuite en deux étapes. La première est la sélection de la meilleure stratégie d'ajustement de chaque modèle. Le ré-échantillonnage de la classe majoritaire (*resampling majority*) est la stratégie la plus efficace. Elle permet aussi d'exclure le modèle **SVM** qui présente des résultats bien inférieurs aux deux autres modèles. Une explication possible est son inefficacité sur les bases de grande dimension comme celle construite ici. La deuxième étape est la comparaison des résultats obtenus selon deux méthodes de validation sur les meilleurs modèles de *machine learning* retenus (*Gradient Boosting* et le *Random Forest*). La comparaison s'effectue sur l'ensemble de la base de validation du littoral atlantique français sans distinction par cantons.

La première méthode est graphique, des histogrammes par année pour l'ensemble de la base du littoral Atlantique sont réalisés en fonction du type d'arrêt Cat-Nat (CLASS=1 : submersion non reconnue par la commission interministérielle et CLASS=2 : submersion reconnue par la commission interministérielle). De manière générale on se rend compte que le modèle prédit d'avantage de CLASS=0 correspondant au fait qu'il n'y ait pas eu de submersion marine. Cela est représentatif de la sélectivité des critères appliqués lors de cette discrimination. Le modèle **XGB** semble mieux prédire les résultats que le *Random Forest*.

Afin de prendre une décision, une deuxième méthode est effectuée en fonction de deux métriques statistiques adaptées aux données déséquilibrées comme celles-ci. Ces métriques sont le score de précision ou *accuracy* et le F1 score qui se fonde sur la moyenne harmonique entre le taux de faux positifs et celui de faux négatifs.

Les tableaux suivants permettent de résumer l'ensemble des scores obtenus en fonction du scénario **RCP** regardé :

Bases	XGB		Random Forest	
	Apprentissage	Test	Apprentissage	Test
Accuracy	89.4 %	74.9%	87.41%	69.8%
F1 score	89%	85.6%	87%	82.1%

**Table 1** – Comparaison des résultats sur le scénario **RCP** 4.5

Bases	XGB		Random Forest	
	Apprentissage	Test	Apprentissage	Test
Accuracy	89.4 %	75.8%	87.4%	70%
F1 score	89%	86.2%	87%	82.3%

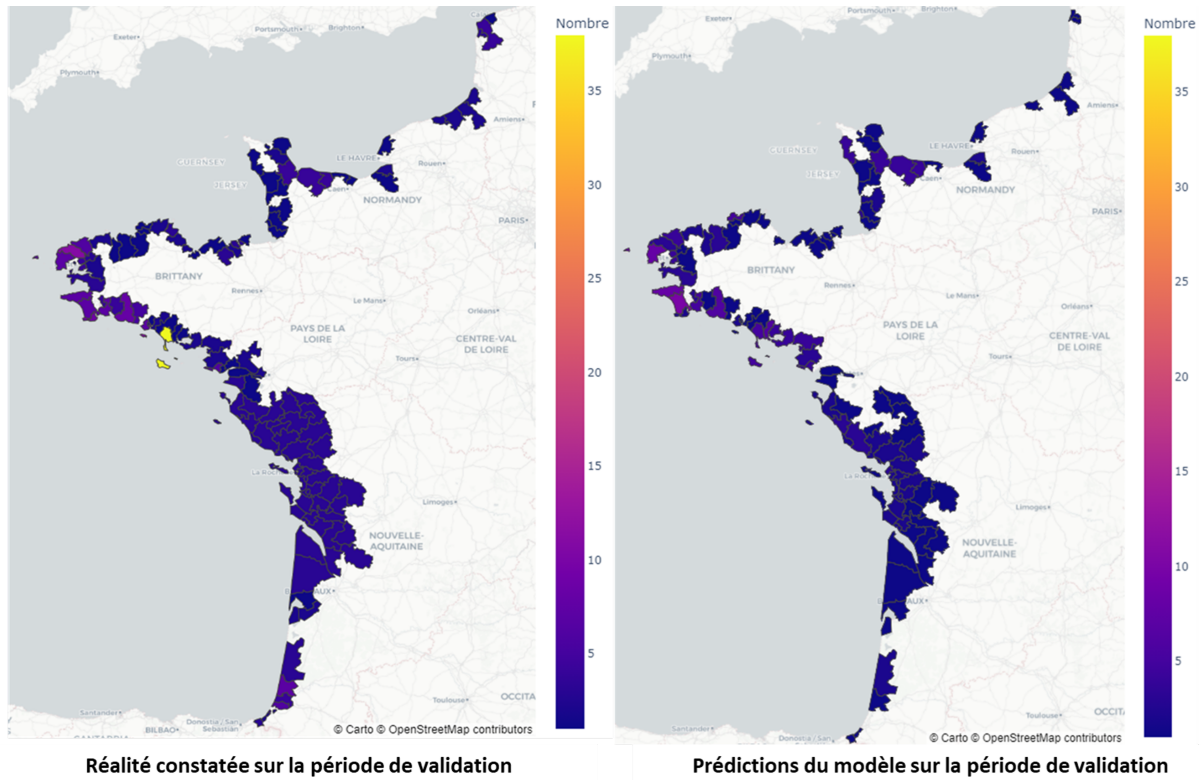
**Table 2** – Comparaison des résultats sur le scénario **RCP** 8.5

Le modèle **XGB** est finalement retenu car il présente des métriques de validations supérieures quel que soit le scénario.

- 5. La validation sur le modèle final choisi.** Dans notre cas, sur le modèle **XGB**, une nouvelle validation a lieu mais cette fois-ci en prenant une maille cantonale. Il est alors possible de faire une comparaison entre les projections du modèle sur la période de validation (2006-2020) et les résultats réels sur cette période.

Cette validation a permis de confirmer que la fréquence des arrêtés type 2 représentant les submersions marines reconnues est mieux prédite par le modèle que le type 1 représentant les submersions marines non reconnues. Dans le cadre de ce mémoire, cela représente la modalité qui est conservée pour notre objet d'étude.

Les cartes suivantes permettent de comparer les prédictions du modèle RCP 8.5 à la réalité :



**Figure 3** – Nombre de jours de submersions marines reconnues

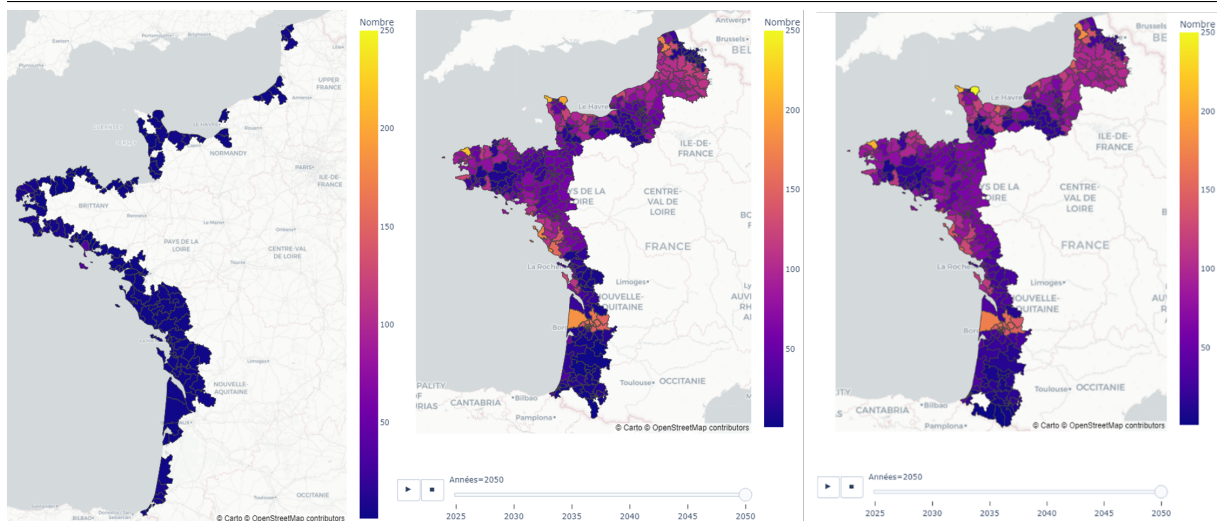
Il est possible de remarquer que le modèle localise bien les zones soumises aux submersions marines mais sous-estime sur la période leur fréquence.

La région Bretagne est ainsi correctement identifiée comme la principale victime des submersions marines. Toutefois, les événements prédits sont bien plus répartis sur tout son littoral et notamment sur sa pointe. En réalité, on peut observer qu'ils se sont concentrés sur sa façade nord.

De plus, il est observable que la région Nouvelle Aquitaine reste celle la moins bien prédite par le modèle, notamment à l'intérieur des terres.

6. **L'analyse de l'évolution des submersions marines.** Elle est faite entre la réalité connue à 2020 et les prédictions faites par le modèle à partir des figures suivantes :





**Figure 4** – Réalité connue à 2020 (1ère carte) par rapport aux prédictions à 2050 selon les **RCP 4.5** (2ème figure) et **RCP 8.5** (3ème figure)

Le premier constat est qu'à horizon 2050 le modèle prévoit des submersions qui touchent beaucoup plus l'intérieur des terres. Cela s'observe sur toutes les zones modélisées. Ainsi, l'impact des submersions marines pourrait atteindre la frontière de la région Ile de France qui n'a pourtant pas de façade littorale dans les deux scénarios. Ce constat est révélateur d'un choix de modélisation, qui était de ne pas restreindre le modèle aux uniques zones déjà sinistrées et de pouvoir affecter l'ensemble du littoral.

Les hypothèses entre les scénarios **RCP 4.5** et **8.5** ne s'écartent réellement qu'à partir de 2100. Par conséquent, l'augmentation de la fréquence entre ces deux scénarios doit être relativisée. A horizon 2050, l'augmentation est de 20% entre ces deux scénarios sur l'ensemble de la zone du littoral étudiée mais devrait s'intensifier significativement à horizon 2100.

Le modèle présenté permet d'avoir une vision à la maille cantonale des événements journaliers de type submersion marine à horizon 2050 selon les scénarios **RCP 4.5** et **8.5**.

Les résultats de cette étude ont révélé que le modèle fournit des résultats comparables à d'autres acteurs (**CCR**, **BRGM**, *Coastal Futures*). On retrouve des résultats comparables en termes de localisation pour les deux scénarios climatiques étudiés. Les départements de Charente-Maritime et Vendée affectés par Xynthia demeurent parmi les départements les plus risqués. Un point de vigilance est également noté pour les départements du Pas-de-Calais, de la Manche et de la Gironde.

La transition climatique associée aux événements extrêmes a fait apparaître la nécessité de réaliser des stress tests climatiques. Ces tests peuvent avoir un impact conséquent à la fois sur le comportement des clients et sur les marchés financiers. Cette étude s'inscrit alors dans le cadre de ces tests comme un outil de modélisation pour mieux anticiper l'impact des risques climatiques. Cela permet dès lors d'aider à accroître la résilience du secteur financier.

## Executive Summary

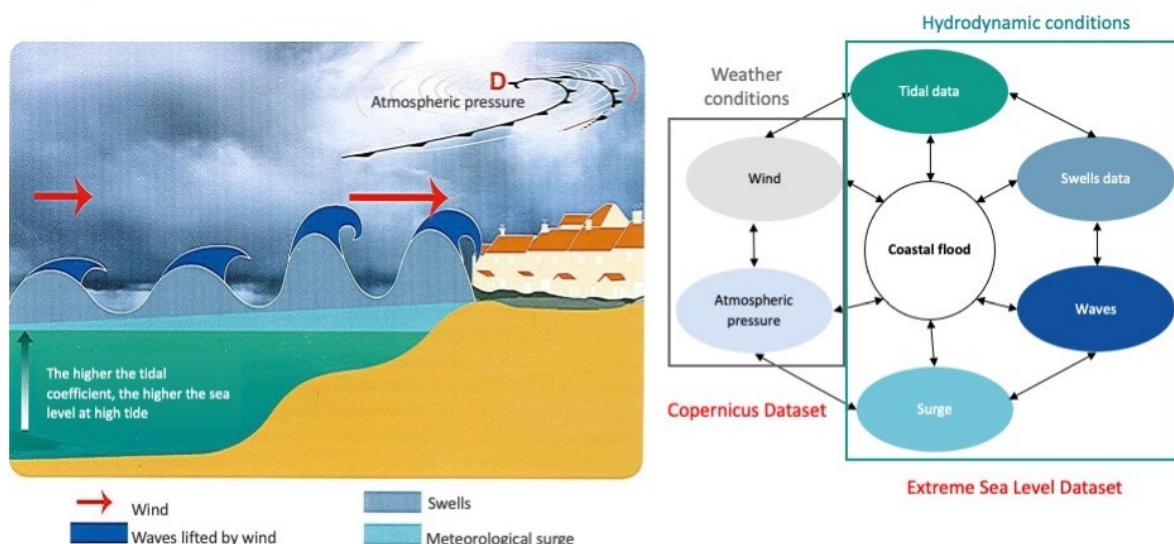
Over the last decades, human production activities have accelerated exponentially. This explosive growth has been made possible by the exploitation of carbon-based energies. This only goes to reinforce that greenhouse gas (GHG) emissions are intrinsically linked to the development of modern societies.

In 1988, the United Nations Environment Program and the World Meteorological Organization created an international group of climate experts, the IPCC. Its objective is to gather and summarize thousands of scientific papers on climate change, its causes, and its potential repercussions. The increase in GHG concentrations is shared by the majority of the community of climatologists. Nevertheless, there are undeniable visible impacts due to climatic disruptions such as the retreat of glaciers or the loss of mass of ice caps. The rise in temperature is therefore verified, and contributes to the rise in sea level.

In the context of climate change, extreme events or natural catastrophes (Nat-Cat) are likely to become more and more frequent. Since these events are still rare, it is essential to project them over a long enough period in order to establish if this increase in frequency is observable. For the preparation of the IPCC 5th Assessment Report, four reference scenarios (or **RCP**, for Representative Concentration Pathways) of GHGs have been carried out for the 21st century and beyond. These scenarios correspond to the level of investments made to reduce GHG emissions at the global level.

According to the IPCC<sup>3</sup>, the sea level would rise by 20 cm by 2050 for the scenario with no external intervention (business as usual or **RCP 4.5**), and 23 cm for the most pessimistic scenario (**RCP 8.5**). These possible rises are directly linked to the frequency of coastal flooding catastrophes. This is a type of flooding that occurs in coastal areas and is caused by particularly adverse tidal and meteorological conditions.

In order to model them correctly, it is essential to understand their causes. The following representation allows to represent the phenomenon in a schematic way :



**Figure 5** – Representation of the coastal flood based on REFMAR data

3. IPCC, 2014 : Climate Change 2014 : Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

The quality of the modeling results depends on the model construction data. For this study they are of two types :

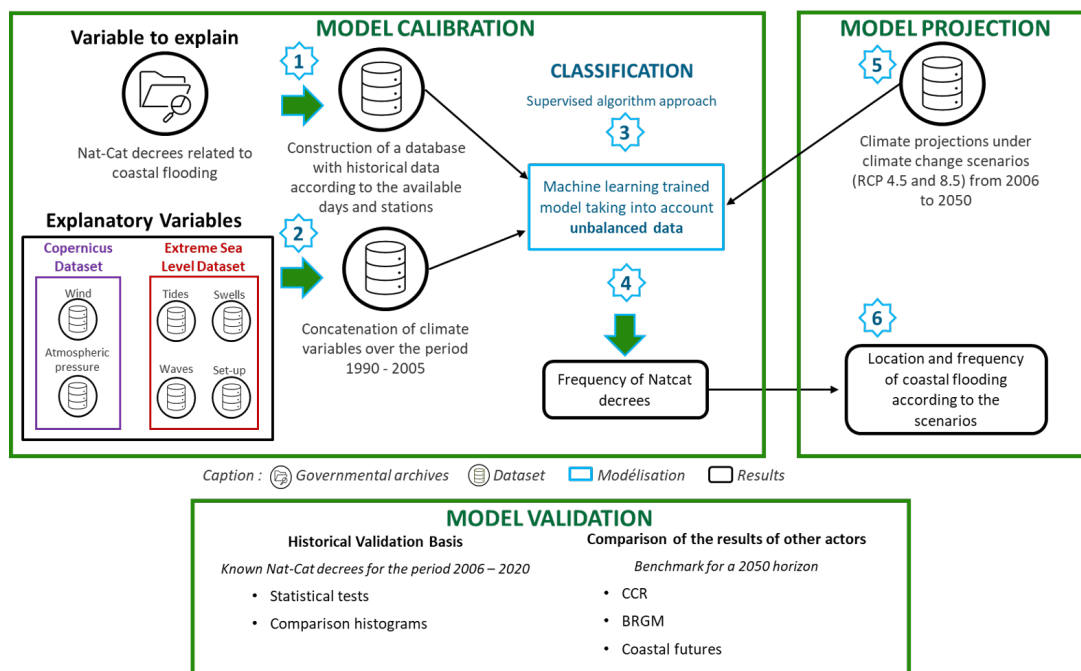
- The meteorological conditions come from the results of the **EURO-CORDEX** initiative (Copernicus database). This program brings together the work of various European institutions to build a series of climate models particularly suited to this geographical area. The models used are based on several equations that represent the different relationships between the atmosphere, the oceans, and the land. The wind speed and the mean atmospheric pressure at sea level were then selected from this database.
- The hydrodynamic conditions are extracted from the Extreme Sea Level database. The projections are based on different works of the National Oceanic and Atmospheric Administration. Different variables are extracted such as wave heights, or the importance of the surge.

The data projections are performed on several scenarios :

- Historical : the data as observed in the past thanks to several stations with sensors installed along the coast.
- **RCP 4.5** : projection to the future assuming a relatively controlled increase in temperature.
- **RCP 8.5** : projection to the future with the same growth pattern and therefore significant impact on future temperatures.

This climate data is open source and has been retrieved via APIs<sup>4</sup>. Once the data was processed, the models were built.

The objective is to build a model to project the occurrence of marine flooding by 2050 on the French Atlantic coast. The modeled variable is the occurrence of a Nat-Cat decree for a day D and a canton C. The following diagram summarizes the modeling undertaken :



**Figure 6** – Summary of methodology

4. Application Programming Interface. This technique allows two distinct applications such as a web page and a python interface to exchange data automatically

The modeling steps can be summarized as follows :

1. **Construction of the database containing the variable to explain** according to the temporal (daily) and geographical grids chosen (canton). It should be noted that the grids of the model depend on the accuracy of the available data. When using **RCM** models, it is not possible to use a geographical grid finer than the canton. The daily time grid is justified by the study of the duration of a flood event. This variable consists of the requests for Nat-Cat decrees provided by the **CCR**. The publication of such an order in le Journal Officiel is a necessary condition for the **CCR**, and if needed state, intervention. This decision depends on an interministerial commission which meets at the request of the town(s) affected locally. In the case of the French system, a partnership is formed between the state, the **CCR** and the insurance companies in the case of a natural disaster. The data collected in this database are from 1990 to 2020.
2. **Construction of the database containing the explanatory variables** selected from the constituent parameters of a marine submersion. The same temporal (daily) and geographical (canton) grid is again applied. The data collected in this database are from 1990 to 2005.
3. **Construction of four final databases**. The first one is a database with data from 1990 to 2005 containing all the variables for the learning model. The second is a historical validation database from 2006 to 2020 on the explanatory variable. It allows the verification of the best machine learning model from real data. The third one is a database of explanatory variables for the **RCP 4.5** scenario for 2050. The last one is a database of explanatory variables for the **RCP 8.5** scenario by 2050. These last two bases are used for the projection of the machine learning model.

The construction of these databases requires various treatments before they can be used. Thus, work was carried out to locate the stations and to process the unbalanced and seasonal nature of the data.

4. **The calibration and selection of Machine Learning models** using a supervised algorithm approach. Three machine learning models are used to classify Nat-Cat :
  - The first model is Gradient Boosting (**XGB**). This algorithm weights the observations as it learns. Iterative improvement of the predictions allows to obtain an optimum by summing the different decision trees built.
  - The second is the Random Forest algorithm. The classification performed is based on a grouping, or "forest", of independent decision trees. The training data are subdivided into several randomly selected sets to build the trees.
  - The third is the Support Vector Machines (**SVM**) model. This algorithm tries to classify the data by distinguishing them with a boundary. The data is represented in an n-dimensional space and the boundary is constructed using a hyperplane.

The models are trained on a training database covering the period from 1990 to 2005. Projections of the explanatory variables from 2006 to 2020 are then used as an input to the model in order to determine the frequency of flooding over this period.

In order to validate the model, a database on the variable to be explained is available for the period 2006 to 2020. It will allow us to choose the best model by comparing the predictions made by the three

models mentioned with the real-world data.

The validation of the model then takes place in two steps. The first is the selection of the best fitting strategy for each model. It can be seen that the resampling of the majority class is the most efficient strategy. It also allows us to exclude the **SVM** model which presents with much inferior results to the two other models. A possible explanation to this is its inefficiency on high dimensional databases like the one constructed here.

The second step is the comparison of the results obtained according to two validation methods on Gradient Boosting and Random Forest. The comparison is made on the whole validation base of the French Atlantic coast without distinction by cantons. The first method is graphical, where histograms by year for the whole Atlantic coastline database are drawn out according to the type of Nat-Cat decrees (CLASS=1 : submersion not recognized by the interministerial commission, and CLASS=2 : submersion recognized by the interministerial commission). In general, we see that the model predicts more CLASS=0 corresponding to the fact that there was no coastal flooding. This reflect the harsh selection criterias used to descriminate. The **XGB** model seems to predict the results better than the Random Forest. In order to choose between the two, a second method is performed based on two statistical metrics suitable for unbalanced data such as these. These metrics are the accuracy score, and the F1 score which is based on the harmonic mean between the false positive and false negative rates.

The following tables give a summary of these results on the whole validation database :

Database	XGB		Random Forest	
	Learning	Test	Learning	Test
Accuracy	89.4 %	74.9%	87.41%	69.8%
F1 score	89%	85.6%	87%	82.1%

**Table 3** – Comparison of results on the scenario **RCP 4.5**

Database	XGB		Random Forest	
	Learning	Test	Learning	Test
Accuracy	89.4 %	75.8%	87.4%	70%
F1 score	89%	86.2%	87%	82.3%

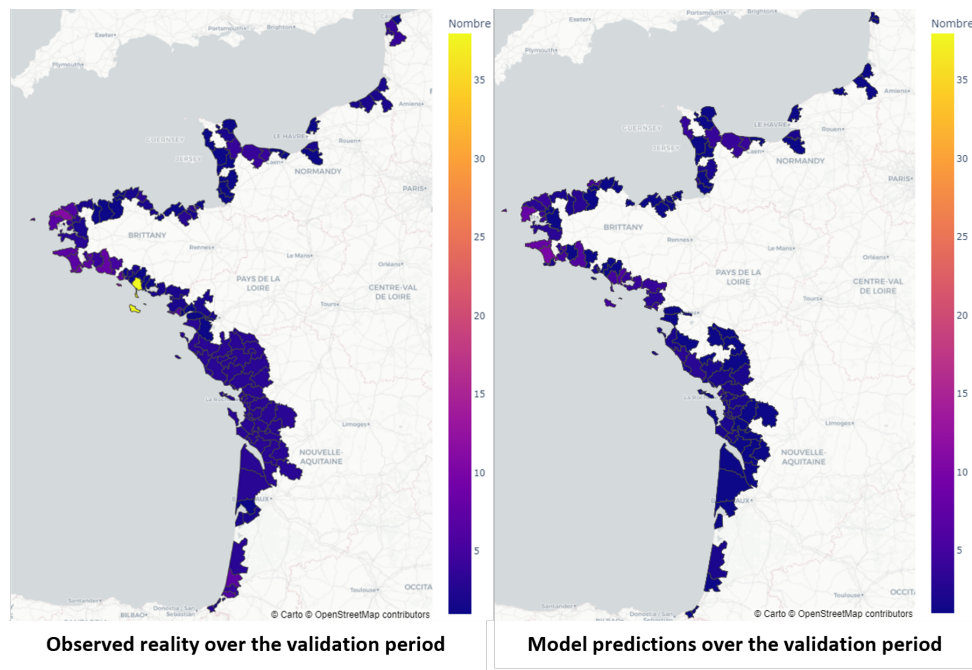
**Table 4** – Comparaison des résultats sur le scénario **RCP 8.5**

The **XGB** model is finally retained because it presents superior validation metrics for both scenarios.

- 5. Validation of the final model chosen.** In our case, a new validation is performed on the **XGB** model, but this time by taking a cantonal grid. It is then possible to make a comparison between the model projections over the validation period (2006-2020), and the actual results over this period.

This validation confirmed that the second type of marine floods (representing recognized marine floods) is better predicted by the model than the first type (representing not recognized marine floods). For the purpose of this dissertation, this represents the modality that is retained for our study.

The following maps allow us to compare the predictions of RCP 8.5 with reality :



**Figure 7** – Coastal submersions over the historical validation period (2006-2020) for the Nat-Cat type 2 decrees (recognized submersion)

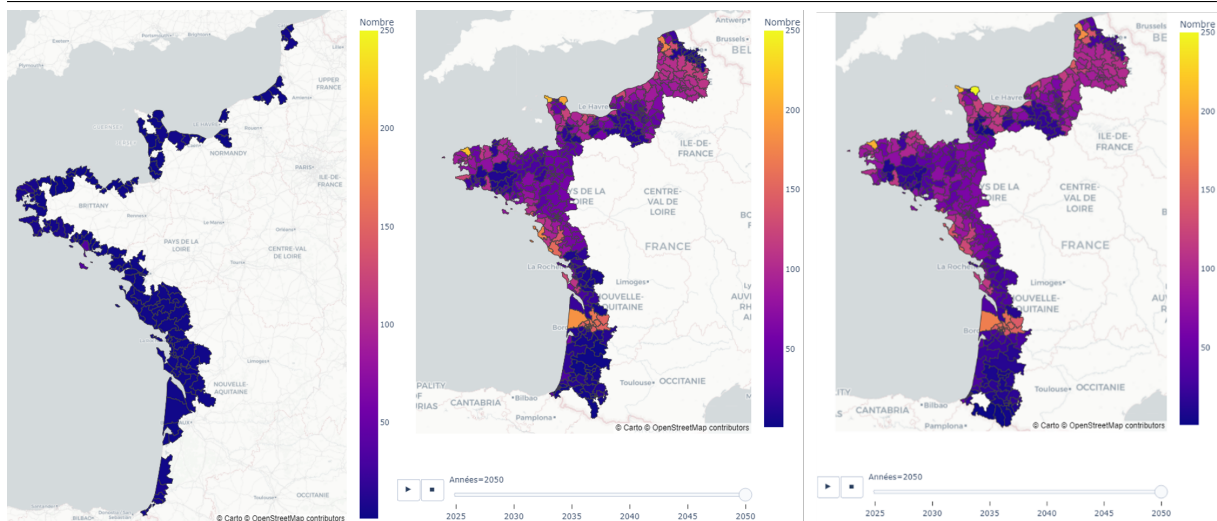
These maps show that the model is well suited to locate the areas subjected to coastal flooding but underestimates their frequency over the period.

The Brittany region is thus correctly identified as the main victim of coastal flooding. However, the predicted events are much more spread out over the entire coastline, especially on its tip. In reality, we can observe that they are concentrated on its northern facade.

In addition, it is observed that the New Aquitaine region remains the least well predicted by the model, especially inland.

## 6. Analysis of the results according to the climate scenarios.

It is possible to see the evolution of coastal flooding between the known reality in 2020 and the predictions made by the model from the following figures :



**Figure 8** – Known reality in 2020 (1st map) compared to predictions in 2050 according to **RCP 4.5** (2nd figure) and **RCP 8.5** (3rd figure)

The first observation is that by 2050 the model predicts flooding that affects the interior much more. This is observed in all the areas modeled. Thus, the impact of coastal flooding could reach the border of the Ile de France region, which does not have a coastline, in both scenarios. This observation is indicative of a modeling choice, which was not to restrict the model to the only areas already affected and to be able to affect the entire coastline.

The assumptions between the **RCP 4.5** and **8.5** scenarios do not really differ until 2100. Therefore, the increase in frequency between these two scenarios must be put into perspective. By 2050, the increase is 20% over the entire coastal area studied between these two scenarios but should intensify significantly by 2100.

The model presented allows the visualization at the cantonal level of the daily events of coastal flooding type at horizon 2050 according to the **RCP 4.5** and **8.5** scenarios.

The results of this study revealed that the model provides comparable results to other actors (**CCR**, **BRGM**, Coastal Futures). We find comparable results in terms of location for the two climate scenarios studied. The departments of Charente-Maritime and Vendée affected by Xynthia remain among the riskiest departments. A point of vigilance is also brought to the departments of Pas-de-Calais, Manche, and Gironde.

The climate transition associated with extreme events has brought to light the need to perform climate stress tests. These tests can have a significant impact on both customer behavior and financial markets. This study is part of these tests as a modeling tool to better anticipate climate risks. It can therefore help to increase the resilience of the finance sector.

---

## Remerciements

Avant tout développement, je tenais à remercier toutes les personnes qui m'ont beaucoup appris et à celles qui ont eu la gentillesse de faire de mon stage de fin d'études un moment très profitable.

Je voudrais commencer par remercier l'initiateur de l'offre Climat, Cyril Chalin, sans qui cette étude n'aurait pas vu le jour. Je tiens également à remercier Claude Chassain et Baptiste Brechot de m'avoir permis d'effectuer mon stage de fin d'études au sein de Deloitte. Sans eux, je n'aurais pas eu l'opportunité de pouvoir allier les missions de conseil et de recherche sur un risque climatique.

De plus, je remercie tout particulièrement, Joseph Delawari, mon tuteur d'entreprise, pour avoir lancé l'idée des sujets liés aux risques climatiques physiques et pour m'avoir accompagnée tout au long de la réalisation de ce mémoire. Je tiens aussi à remercier Manuela Gómez Torres pour m'avoir accompagnée avec beaucoup de patience et d'attention. Je suis également très honorée d'avoir pu travailler sur le projet climat avec mes collègues Ines Boussim, Manuella Mouafo et Mathieu Clavier. Merci à Mihir John, Axel Rosier et Axel Elbaz pour leur aide précieuse dans la relecture de mon mémoire. Merci à eux pour les agréables moments de discussion ensemble.

Je tiens aussi à remercier mon tuteur académique, Christian-Yann Robert, toujours disponible pour répondre à mes questionnements et orienter ma réflexion.

J'exprime ma profonde gratitude à mes très chers parents, qui me réservent toujours soutien et amour. Un grand merci à Mathias Baudry de m'avoir aidée à chaque fois que j'en avais besoin et pour sa présence inestimable. Une pensée particulière à mes amies Siran Huang, Katia Metna et Jinlin Lin pour leur confiance et d'être à mes côtés. Je tenais à remercier particulièrement ma petite sœur, Estelle, pour son soutien indéfectible et à toutes nos soirées sans fin sur nos ordinateurs.

J'en profite aussi pour remercier les membres de l'équipe actuariat de Deloitte pour leur accueil et leur aide, chacun à leur manière.



# Table des matières

<b>Résumé</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Synthèse</b>	<b>3</b>
<b>Executive Summary</b>	<b>9</b>
<b>Remerciements</b>	<b>15</b>
<b>Introduction</b>	<b>20</b>
<b>1 Le cadre de l'étude</b>	<b>23</b>
1.1 L'importance d'appréhender le risque climatique . . . . .	23
1.1.1 La classification du risque climatique . . . . .	23
1.1.2 Un impact sur l'ensemble de la chaîne de valeur et des risques pour l'assureur . . .	23
1.1.3 Les enseignements tirés de l'étude du risque climatique . . . . .	24
1.1.3.1 Chronologie des principales publications liées au changement climatique .	25
1.1.3.2 Les résultats de l'exercice pilote de l'ACPR en France . . . . .	25
1.1.3.3 Les risques liés au changement climatique dans l'ORSA (avis EIOPA – ORSA avril 2021) . . . . .	26
1.1.3.4 Le 6ème rapport d'évaluation du GIEC (juin 2021) . . . . .	26
1.2 Le système français du régime des Catastrophes Naturelles (Cat-Nat) . . . . .	27
1.2.1 Le fonctionnement du système . . . . .	28
1.2.1.1 Un système français égalitaire et unique . . . . .	28
1.2.1.2 Liste courante des périls couverts par le régime . . . . .	29
1.2.2 Comprendre la couverture du risque Cat-Nat et le rôle de la CCR . . . . .	29
1.3 Rappels généraux sur les inondations . . . . .	31
1.3.1 Typologies d'inondations . . . . .	31
1.3.2 Les submersions marines . . . . .	32
1.3.2.1 Typologie et paramètres des submersions marines . . . . .	32
1.3.2.2 Les dangers liés aux submersions marines et les facteurs aggravants . . .	34
1.3.2.3 Exemples de submersions marines marquantes en France ces dernières décennies . . . . .	36
1.4 Rappels sur la modélisation du risque inondation . . . . .	36
1.4.1 La modélisation des Cat-Nat . . . . .	37
1.4.2 Méthodes d'analyse du risque d'inondation . . . . .	37
1.4.2.1 Méthodes historiques et paléohydrologiques . . . . .	38
1.4.2.2 Méthodes hydrologiques et hydrauliques . . . . .	38
1.4.2.3 Méthodes géologiques et géomorphologiques . . . . .	38
1.4.3 Modélisations possibles des Submersions marines en France . . . . .	38
1.4.3.1 Les équations de Saint-Venant comme préambule au modèle déterministe	38
1.4.3.2 Modèle déterministe . . . . .	40
1.4.3.3 Modèles probabiliste et stochastique . . . . .	41
1.4.3.4 Un choix relevant des besoins de l'étude réalisée . . . . .	42
1.4.3.4.1 Les besoins de projection . . . . .	42

1.4.3.4.2	Le modèle choisi . . . . .	43
<b>2</b>	<b>Présentation de la modélisation relative aux submersions marines</b>	<b>45</b>
2.1	Introduction . . . . .	45
2.2	Calibrage du modèle de submersion marine . . . . .	45
2.2.1	Les déclarations de catastrophes naturelles . . . . .	46
2.2.1.1	Le choix de base de données . . . . .	46
2.2.1.2	Présentation de la BDD de la CCR . . . . .	47
2.2.1.3	Les limites à l'utilisation des arrêtés de catastrophes naturelles . . . . .	48
2.2.2	Les variables climatiques issues de COPERNICUS . . . . .	49
2.2.2.1	Présentation du modèle GCM . . . . .	49
2.2.2.2	Présentation du modèle RCM . . . . .	51
2.2.2.3	Présentation du Programme Mondial de Recherche sur le Climat (PMRC) . . . . .	53
2.2.2.4	Présentation du modèle utilisé au travers de l'initiative CORDEX . . . . .	54
2.2.3	Présentation de la base de données relatives au niveau de mer extrême (ESL) . . . . .	56
2.2.3.1	L'intérêt de la base (ESL) . . . . .	56
2.2.3.2	Présentation de la méthodologie adoptée . . . . .	57
2.2.3.2.1	Présentation générale . . . . .	57
2.2.3.2.2	Focus sur l'analyse des valeurs extrêmes . . . . .	58
2.2.3.2.3	Ajustement et période de retour . . . . .	59
2.2.3.2.4	Comparaison des différentes méthodes d'estimations de quantiles élevés . . . . .	60
2.2.3.3	Traitement du maillage temporel et géographique . . . . .	60
2.2.3.3.1	Maillage temporel . . . . .	60
2.2.3.3.2	Maillage géographique . . . . .	63
2.2.4	Limites . . . . .	65
2.3	L'agrégation des données . . . . .	66
2.3.1	La maille géographique cantonale . . . . .	66
2.3.2	La maille temporelle journalière . . . . .	67
2.3.3	Création des bases d'apprentissage et de validation . . . . .	67
2.4	Présentation et sélection des modèles . . . . .	67
2.4.1	Rappels généraux sur le <i>machine learning</i> . . . . .	68
2.4.1.1	Un modèle par classification . . . . .	68
2.4.1.2	Le choix de l'approche par algorithme supervisé . . . . .	69
2.4.2	Traitements préalables face à des données déséquilibrées et mise en place du modèle de classification . . . . .	69
2.4.2.1	Initialisation de la BDD . . . . .	69
2.4.2.1.1	Le caractère déséquilibré des données . . . . .	69
2.4.2.1.2	Le caractère saisonnier des données . . . . .	70
2.4.2.2	Séparation des données . . . . .	71
2.4.3	Calibration du modèle . . . . .	71
2.4.3.1	<i>Gradient Boosting</i> (XGB) . . . . .	71
2.4.3.1.1	Présentation générale . . . . .	71
2.4.3.1.2	Présentation de l'algorithme et visualisation . . . . .	72
2.4.3.1.3	Application et considération du modèle . . . . .	73
2.4.3.2	<i>Random Forest</i> . . . . .	74
2.4.3.2.1	Présentation générale . . . . .	74

2.4.3.2.2	Présentation de l'algorithme et visualisation	74
2.4.3.2.3	Application et considération du modèle	76
2.4.3.3	SVM	76
2.4.3.3.1	Présentation du modèle	76
2.4.3.3.2	Visualisation du modèle	77
2.4.3.3.3	Avantages et inconvénients du modèle	78
2.4.3.4	Présentation des résultats	78
2.4.3.4.1	Critères de validité choisi	78
2.4.3.4.2	Résultats du <i>Gradient Boosting</i>	80
2.4.3.4.3	Résultats du Random Forest	81
2.4.3.5	Résultats du SVM	83
2.4.4	Choix du meilleur modèle	85
2.4.4.1	Le modèle <i>Gradient Boosting</i>	86
2.4.4.1.1	La base de prédiction RCP 4.5	86
2.4.4.1.2	La base de prédiction RCP 8.5	87
2.4.4.2	Le modèle Random Forest	88
2.4.4.2.1	La base de prédiction RCP 4.5	88
2.4.4.2.2	La base de prédiction RCP 8.5	90
2.4.5	Interprétation du meilleur modèle	91
2.4.6	Validation du meilleur modèle	92
2.5	Comparaison des résultats	95
2.5.1	Présentation des résultats à l'horizon 2050	96
2.5.2	Evolution des résultats au cours des années jusqu'à 2050	100
2.5.2.1	Scénario 4.5	100
2.5.2.2	Scénario 8.5	101
2.5.3	Études de la CCR	102
2.5.3.1	Étude de 2015	102
2.5.3.2	Etude de 2018	104
2.5.4	Outil de visualisation en ligne du BRGM	105
2.5.5	Outil de visualisation en ligne <i>Coastal Futures</i>	106
<b>Conclusion</b>		<b>108</b>
<b>3 Références</b>		<b>110</b>
<b>Annexes</b>		<b>112</b>
	Annexe : Rappels généraux sur la réassurance	112
	Annexe A : Descriptif des variables base EURO-CORDEX	114
	Annexe B : Descriptif des variables base ESL	116
	Annexe C : Lois usuelles en hydrologie	117
	Gumble MAX	117
	Log Pearson 3 (LPIII)	117
	Annexe D : Calcul pour la loi GEV	118
	Annexe E : Exemples d'exploitation des données du modèle CMIP5	119
	Annexe F : Ventilation des contributions projetées de l'ESL et de leur incertitude	121
	Annexe G : Projection des changements des contributions d'ESL selon différentes régions	122
	Annexe H : Projections des modèles individuels	123
	Annexe I : k plus proche voisin	123

Annexe J : Algorithme de classification . . . . .	124
Annexe K : Visualisation exhaustive de la base de validation historique des arrêtés Cat-Nat (1990-2020) . . . . .	124
Annexe L : Carte des départements . . . . .	127
Annexe M : Sorties de l'outil coastal futures sur le deuxième jeu de données . . . . .	128
Annexe N : Base historique d'apprentissage . . . . .	129
Annexe O : Base historique de validation . . . . .	130
Annexe P : Superposition de la carte de prédiction des cantons affectés par la submersion marine à la carte de la France . . . . .	131
<b>Table des figures</b>	<b>132</b>
<b>Liste des tableaux</b>	<b>134</b>

## Introduction

Sécheresses et vagues de chaleur, cyclones et tempêtes, pluies torrentielles et inondations, tous ces événements extrêmes vont encore s'intensifier. La fonte des glaces et la montée des eaux sont déjà irréversibles : voici les conclusions du troisième volet du dernier rapport d'évaluation du **Groupe d'experts Intergouvernemental sur l'Evolution du Climat (GIEC)**. Cet organisme est ouvert à tous les membres de l'ONU. Il a été créé en 1988 en vue de fournir des évaluations détaillées de l'état des connaissances scientifiques, techniques et socio-économiques sur les changements climatiques, leurs causes et leurs répercussions.

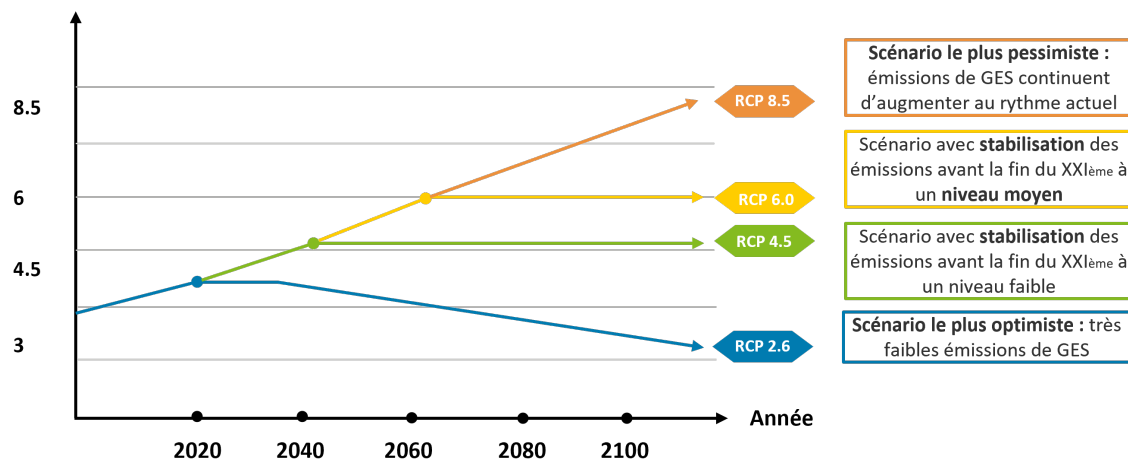
Le troisième volet du rapport, publié au mois d'avril 2022, expose cette dure réalité de l'augmentation des émissions de **Gaz à Effet de Serre (GES)**. Le rapport *Climate Change 2022 : Mitigation of Climate Change* montre également que les émissions de **GES** entre 2010 et 2019 ont été plus élevées que pendant toute autre décennie dans l'Histoire de l'Humanité.

Ainsi, le rapport du **GIEC** offre des aperçus de différents futurs possibles. En effet, dans la perspective du 5e rapport d'évaluation publié en 2013-2014, le **GIEC** a établi quatre profils *Representative Concentration Pathway (RCP)* représentatifs des évolutions de concentration de **GES** :

- **RCP 2.6** : seul scénario à être aligné avec les objectifs des accords de Paris <sup>5</sup>.
- **RCP 4.5 et 6.0** : scénarios intermédiaires qui ne sont plus alignés avec les objectifs des accords de Paris avec des pics d'émissions respectifs en 2030 et 2060.
- **RCP 8.5** : scénario aligné avec un futur sans changement de politique par rapport aux émissions de **GES**. Il est généralement appelé le scénario « Business As Usual ».

Le graphique suivant permet de représenter ces scénarios à horizon 2100 :

Forçage radiatif dû aux activités humaines ( $W.m^{-2}$ )



**Figure 9** – Les différents scénarios RCP d'après le GIEC

Il est primordial de mentionner l'effet de l'inertie climatique visible sur le graphique ci-dessus. Entre 2020 et 2040 les différents scénarios ne sont pas encore largement diversifiés. On peut tout de même souligner les options d'atténuation qui peuvent permettre de réduire les émissions de **GES**, d'atténuer les effets du changement climatique et de parvenir à un développement plus durable. Des solutions sont disponibles dans les secteurs de l'énergie, de l'agriculture, de l'industrie, du bâtiment ou encore des

5. Maintient de l'augmentation de la température à une cible de 1,5°C par rapport à l'année 2016

transports pour réduire de plus de moitié les émissions d'ici 2030.

L'un des effets majeurs du dérèglement climatique est l'augmentation des températures. Depuis 150 ans environ (début de l'ère industrielle), la température à la surface de la Terre a augmenté de 1°C, et, pourrait atteindre les 2°C d'ici 2040. On comprend alors aisément que le changement climatique va transformer les cycles traditionnels et participer à l'élévation du niveau de la mer. La hausse estimée du niveau marin pourrait ainsi atteindre 7 mètres si les calottes glacières fondent entièrement<sup>6</sup>. Dans le cadre de la prise en compte de ce changement climatique, il convient également de rappeler que les risques climatiques peuvent être décomposés en trois grandes catégories : physique, de transition ou encore de responsabilité. Ce mémoire, par sa nature, traite du premier risque qui résulte des dommages directement causés par les phénomènes météorologiques et climatiques à travers des événements extrêmes ou des évolutions progressives.

Les inondations vont alors devenir plus fréquentes, plus violentes et moins prévisibles. Même si ces dernières ont été source de fertilité ou encore de richesse en accompagnant l'essor de certaines civilisations, allant de l'Égypte à la Chine, elles sont le plus souvent rattachées à la destruction et au déclin. Des sociétés ont ainsi été amenées à disparaître, comme en témoignent les ruines de la ville de Liangzhu (3 300 – 2 300 de n. è.). En effet, cette civilisation néolithique chinoise particulièrement développée – surnommée la « Venise de l'Orient » – s'est effondrée à la suite d'inondations dévastatrices.

De nos jours, les inondations sont principalement cause de dommages économiques. Ces dommages sont fondamentaux dans l'évaluation des risques par les compagnies d'assurance. En effet, elles doivent les évaluer afin de proposer une offre de polices cohérentes aux habitants et aux entreprises des zones à risque.

La question de savoir si nous pouvons prédire la survenance du risque d'inondation face au changement climatique se pose ainsi légitimement. On peut dénombrer trois grands types d'inondations : la crue de rivière (*fluvial flood*), l'inondation éclair (*flash flood*) et la submersion marine (*coastal flood*).

Ce mémoire traite de cette dernière catégorie qui représente une inondation de courte durée (quelques heures à quelques jours) localisée en zone côtière par des conditions océaniques et météorologiques défavorables telles qu'une forte marée ou un vent fort. En effet, il s'agit ici d'un risque peu étudié dans le milieu actuariel<sup>7</sup> avec un potentiel de dommage économique élevé.

Pour ce faire, un découpage géographique a été choisi pour cette étude qui se concentre sur la métropole française, et plus particulièrement son littoral océanique. L'amplitude de la marée en Méditerranée est faible, mais elle pourrait faire l'objet d'une étude ultérieure.

La méthodologie adoptée pour cette étude répond au besoin de modéliser une projection de long terme (jusqu'à 2050) de la sinistralité liée aux submersions marines par une approche dynamique. Plus que la fiabilité du montant de l'impact, l'objectif visé est de différencier les scénarios RCP. En effet, nous sommes à un moment expérimental où l'objectif est de faire émerger de bonnes pratiques. Les résultats ont donc pour vocation d'être un outil d'aide à la décision stratégique.

Il convient d'abord de dresser le cadre de l'étude à travers divers rappels allant du régime des catastrophes naturelles (Cat-Nat) aux modélisations existantes dans le cadre des inondations.

6. D'après *Climate Change 2021 – the physical science basis*

7. Mémoires de Mr Gahbiche (Estimation de la prime pure Cat-Nat au travers de données géographiques) et Mme Prévot (Projet de gestion géographique des risques) portant sur le sujet de la gestion des risques, mais dans un contexte géographique.

Si le spectre de méthodes est relativement large et propose des solutions tout à fait élaborées et propres à la géophysique marine, la décision de l'actuaire se porte souvent sur des techniques relativement moins complexes. Ainsi, il convient de présenter une méthodologie applicable pour l'actuaire et relevant de la modélisation stochastique. Pour ce faire, les données utilisées proviennent intégralement de données libres d'accès. En outre, elles sont issues d'un programme de recherche européen et d'un papier de recherche. Des modèles de *machine learning* ont ensuite été développés afin de pouvoir prédire les submersions marines à horizon 2050. L'originalité de cette étude réside dans la création intégrale d'une base de données applicable aux submersions marines face au changement climatique à une maille géographique cantonale.

# 1 Le cadre de l'étude

## 1.1 L'importance d'appréhender le risque climatique

Afin de pouvoir appréhender le risque climatique, il apparaît important de d'abord définir cette notion (1.1.1). Puis, de voir l'interaction qu'il existe entre ce risque et la chaîne de valeur d'un assureur (1.1.2). Pour finir, il est intéressant d'avoir un aperçu des principaux enseignements tirés de l'étude du risque climatique (1.1.3).

### 1.1.1 La classification du risque climatique

Le risque climatique recouvre trois familles de risque :

- **Le risque physique** mesure l'impact direct du changement climatique ou les pertes associées aux dommages causés par les aléas climatiques sur les personnes et les biens. *Exemples : inondations, incendies, ouragans, canicules, froids extrêmes, sécheresse ...*
- **Le risque de transition** résulte d'une modification du comportement des agents économiques et financiers en réponse à la mise en place d'évolution réglementaires, de politiques énergétiques ou de changements technologiques. En d'autres termes, ce risque représente les conséquences économiques liées au passage à une économie bas carbone. *Exemples : réglementation, technologie, réputation, marchés ...*
- **Le risque de responsabilité** mesure les impacts financiers incertains résultant d'éventuelles poursuites en justice pour avoir contribué au changement climatique ou ne pas avoir suffisamment pris en compte les risques climatiques. *Exemples : communication biaisée ou incomplète pouvant aller jusqu'à des scandales liés au greenwashing ...*

Ces deux derniers risques ne seront pas traités dans le cadre de ce mémoire. En effet, l'étude de la submersion marine implique de s'intéresser au risque physique. Par ailleurs, ce dernier est le principal risque pour une compagnie d'assurance non-vie.

### 1.1.2 Un impact sur l'ensemble de la chaîne de valeur et des risques pour l'assureur

Le changement climatique pousse depuis plusieurs années l'ensemble des acteurs économiques à intégrer la dimension **ESG**, dans le cadre de leur communication et de la gestion quotidienne de leur entreprise.

La crise sanitaire et la multiplication des catastrophes naturelles actuelles confirment l'impact significatif que peuvent avoir les risques émergents sur la société et les assureurs. Ainsi, répondre à l'enjeu du risque climatique nécessite une prise de conscience collective. D'autant plus que le changement climatique impact l'ensemble de la chaîne de valeur assurance. Les graphiques ci-dessous permettent de souligner l'importance d'étudier le risque climatique puisqu'il impact l'ensemble de la chaîne de valeur d'un assureur à travers un nombre important de périls.



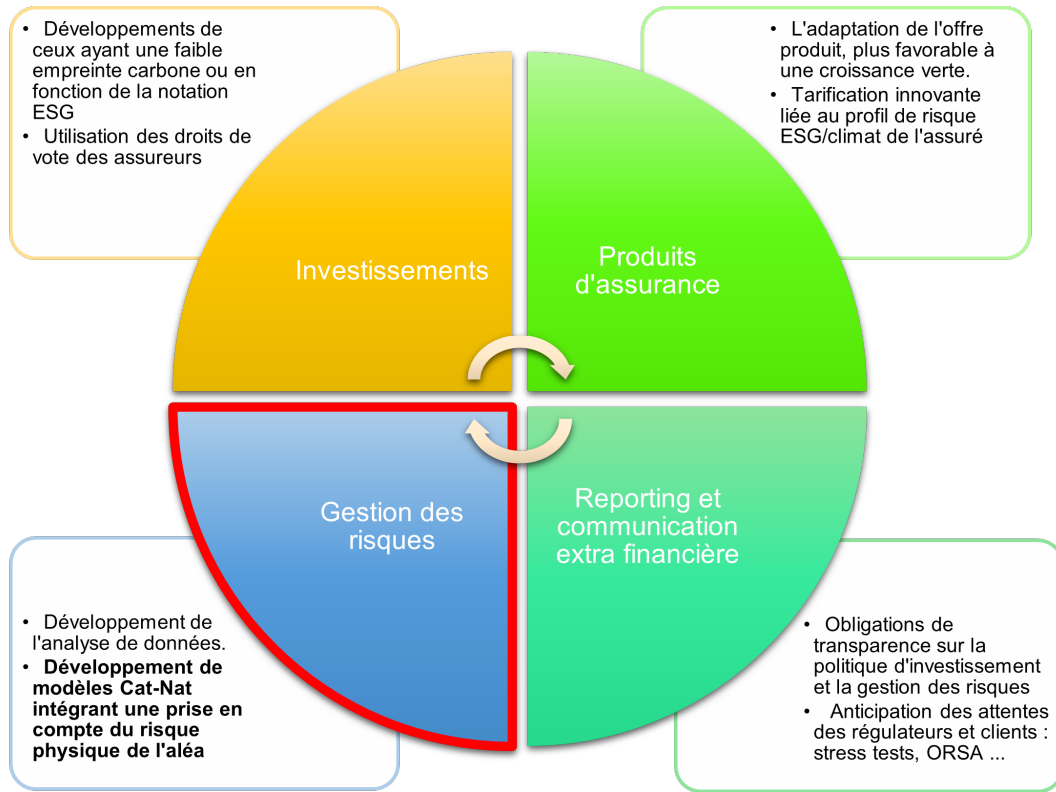


Figure 10 – La chaîne de valeur assurance impactée par le changement climatique

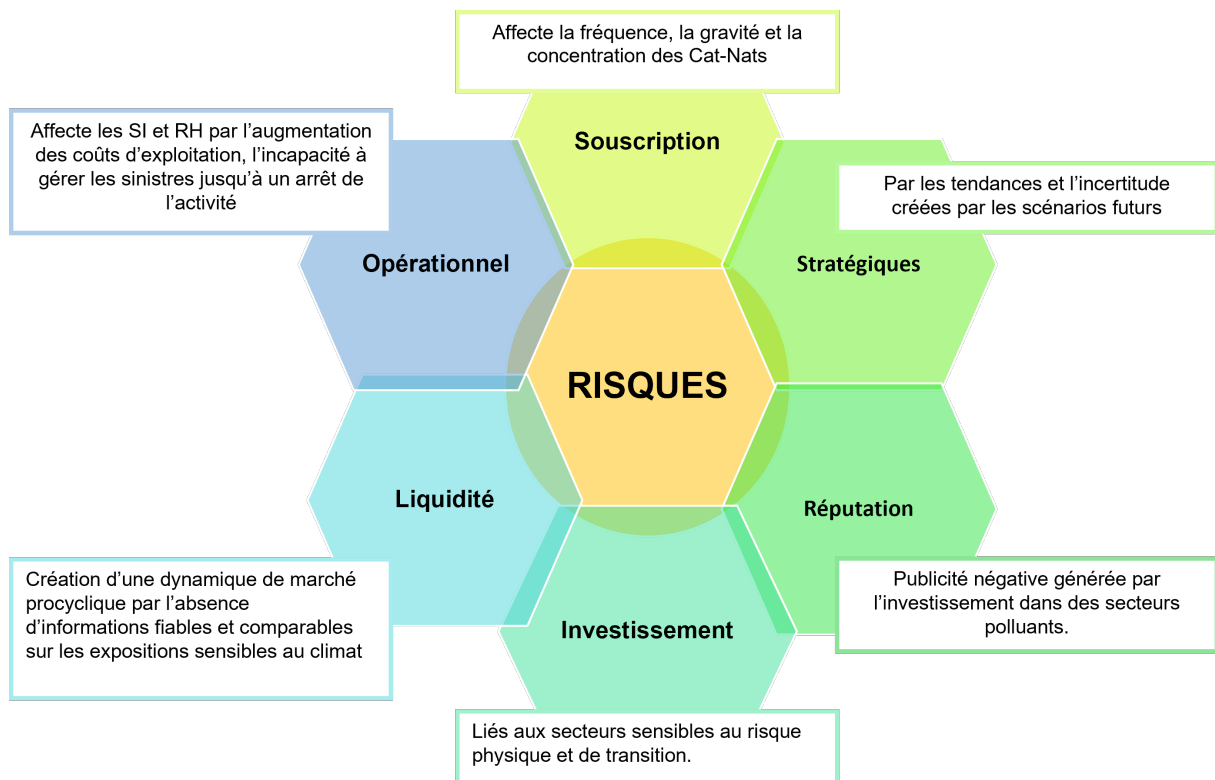


Figure 11 – Les risques impactés par le changement climatique

### 1.1.3 Les enseignements tirés de l'étude du risque climatique

### 1.1.3.1 Chronologie des principales publications liées au changement climatique

Le tableau suivant permet de visualiser les principales publications faites ces dernières années et de mettre en avant le renouvellement de l'exercice climatique.

Années	Publications
2015	Accords de Paris (Article 173-VI de la loi de 2015 sur la transition écologique et la croissance verte).
2019	<b>ACPR</b> – Etats des lieux : les assureurs français face au risque de changement climatique
2020	Lancement du <i>Climate transparency Hub</i> recensant tous les rapports <b>ESG</b> des institutions financières avec une partie sur les stress-tests climatiques.
2020	<b>ACPR</b> – Exercice pilote climatique
2021	SCOR – <i>Mind the sustainability Gap</i>
2021	<b>GIEC</b> – Publication du « 6ème rapport d'évaluation » qui est une mise à jour de la 5ème évaluation qui datait de 2014.
2021	<b>EIOPA</b> – Avis sur l'encadrement des scénarios de risques liés au changement climatique dans <b>ORSA</b>
2021	<b>EIOPA</b> – Consultation sur l'évaluation de la matérialité du changement climatique
2021	COP26 : engagement de la Banque de France/ <b>ACPR</b>
2021	France Assureurs - changement climatique : quel impact sur l'assurance à l'horizon 2050 ?
2022	France Assureurs – Réussir la transition écologique et renforcer la résilience face aux risques climatiques
2022	<b>ACPR</b> – Etats des lieux : Gouvernance des risques liés au changement climatique
2022	<b>EIOPA</b> – Principes méthodologiques pour les stress-tests
2023	<b>ACPR</b> – Renouvellement de l'exercice climatique

**Table 5** – Principales publications sur le risque climatique au cours des 10 dernières années

### 1.1.3.2 Les résultats de l'exercice pilote de l'ACPR en France

L'Autorité de Contrôle Prudentielle et de Résolution (**ACPR**) a publié le 04 mai 2021 les résultats de l'exercice pilote climatique réalisé en cordée par les groupes bancaires et d'assurance. Ce rapport souligne notamment la nécessité pour les établissements financiers et les superviseurs d'amplifier et d'accélérer leurs actions face à l'urgence climatique actuelle.

Cet exercice s'est révélé être inédit, de par son périmètre de couverture (assurance, banque), et par son horizon long (des projections sur une trentaine d'années). Il s'agit d'un des premiers exercices européens, faisant suite aux exercices menés par le PRA<sup>8</sup> au Royaume-Uni en 2019 et aux Pays-Bas en 2018.

Au total, cet exercice a engagé 9 groupes bancaires et 15 groupes d'assurance, sur plusieurs mois, fin 2020. Ces derniers représentant respectivement 85% des banques et 75% des assureurs Européens en termes de volume d'actifs.

Ainsi, les principaux enjeux de cet exercice pilote furent de mesurer les risques physiques et de transition, de permettre un dialogue avec les organisations sur les méthodologies, et d'ouvrir un dialogue banque-assurance.

L'exercice pilote a également permis de réaliser une première mesure des effets des actions correctives qui sont envisagées par les assureurs, parmi elles : hausses tarifaires, stratégies de fin de certaines activités

8. PRA : *Prudential Regulation Authority*

émettrices de GES, volonté de financer l'économie ou encore de préserver une relation client.

Cet exercice a été l'objet d'une mobilisation, de manière transversale entre les équipes et les moyens, mais a également pu confirmer le fait qu'il existe un manque de données, une insuffisance des modèles et des méthodologies actuelles.

D'après cette étude, deux chiffres clés sont à retenir :

- 35 % des pertes totales causées par des événements extrêmes sont actuellement assurées en Europe.
- 42 % des assureurs disposent d'une évaluation des risques liés au changement climatique dans leur rapport ORSA.

### 1.1.3.3 Les risques liés au changement climatique dans l'ORSA (avis EIOPA – ORSA avril 2021)

L'EIOPA a initié un plan d'actions sur les thématiques ESG dans le contexte de sa mission de définition du cadre réglementaire prudentiel et de revue des pratiques des superviseurs européens. Dans ce contexte, l'EIOPA a publié en avril 2021 un avis quant à la supervision de la prise en compte des scénarios climatiques par les assureurs dans l'ORSA (processus interne d'évaluation des risques et de la solvabilité par l'organisme). Cet avis préconise en effet l'intégration du risque du changement climatique dans l'ORSA, à court et long terme, selon divers scénarios.

Ainsi, l'EIOPA propose d'une part la prise en compte d'au moins deux scénarios climatiques :

- Une augmentation de la température globale en dessous de 2°C.
- Une augmentation de la température globale supérieur à 2°C.

D'autre part, un horizon à long terme au-delà des projections actuelles de l'ORSA :

- Un changement climatique court terme : 5/10 ans.
- Un changement climatique moyen terme : 30 ans.
- Un changement climatique long terme : 80 ans.

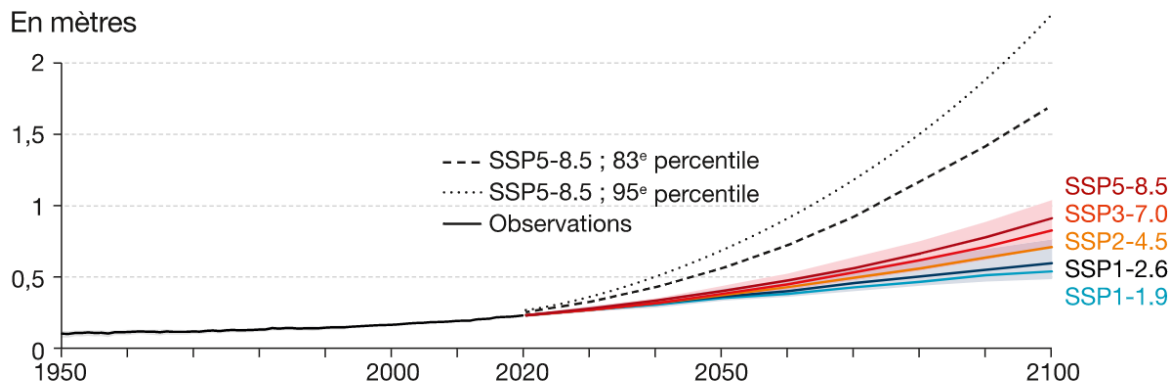
### 1.1.3.4 Le 6ème rapport d'évaluation du GIEC (juin 2021)

Ce dernier rapport résulte des travaux de 3 groupes de travail rassemblant des milliers d'experts et de scientifiques à travers le monde. Sur la modélisation climatique il est en effet primordial d'assurer le lien entre les profils académiques et les praticiens en intégrant pleinement les actuaires dans le processus. Son principal objectif est la mise en évidence des mécanismes physiques à l'œuvre dans l'évolution du climat. Il prend également en compte les causes potentielles d'évolution avec en particulier le rôle des activités humaines. Ainsi, sont exposés les températures, les concentrations en GES, l'élévation du niveau de la mer ou encore l'évolution des événements extrêmes.

Des projections sont faites sous la forme de scénarios codifiés par probabilité d'occurrence. Dans ce dernier rapport ce sont les projections SSP qui sont évoqués à la place des trajectoires RCP. En effet, 6 scénarios de projection SSP viennent remplacer les 4 scénarios de projections RCP de l'Assessment Report ou modèle d'évaluation (AR)<sup>5</sup>. Les scénarios SSP intègrent plusieurs combinaisons sur des évolutions économiques, technologiques, démographiques et institutionnelles futures. Ils comprennent un éventail plus large de projections des émissions de GES et des nouveaux scénarios de réduction des émissions de CO<sub>2</sub>.

Globalement, ces trajectoires offrent un cadre commun pour toute simulation climatique. Chacun des scénarios est un profil d'évolution des concentrations en GES sur le très long terme. C'est également par l'intermédiaire de la notion de forçage radiatif que sont mesurées de manière quantitative ces évolutions.

A partir de ces données il est par exemple possible d'avoir la projection de la hausse moyenne du niveau des mers par rapport à 1900 :



Note : les lignes pleines montrent les projections médianes. Les régions ombrées montrent les plages probables pour SSP1-2.6 et SSP3-7.0.

**Figure 12** – Projection de la hausse moyenne du niveau des mers par rapport à 1900 d'après le GIEC

À l'horizon 2100, le niveau moyen des mers et des océans devrait augmenter quel que soit le SSP 5-8,5 considéré par rapport à la moyenne de 1995-2014. Dans le cadre de ce mémoire, compte tenu de la disponibilité des trajectoires sur les variables étudiées, ce sont les trajectoires RCP qui seront utilisées. En effet, pour le moment, il n'y a pas de données disponibles des projections des SSP à une maille assez fine.

## 1.2 Le système français du régime des Catastrophes Naturelles (Cat-Nat)

Si les pays européens sont principalement exposés aux inondations, ils le sont également à d'autres événements extrêmes. Ainsi, il est primordial de définir correctement le terme de catastrophe naturelle. En France, dans la loi 82-600 du 13 juillet 1982, le premier article permet de définir une catastrophe naturelle (Cat-Nat). « Sont considérés comme les effets des catastrophes naturelles les dommages matériels directs ayant eu pour cause déterminante l'intensité anormale d'un agent naturel, lorsque les mesures habituelles à prendre pour prévenir ces dommages n'ont pu empêcher leur survenance ou n'ont pu être prises. ».

Cette loi du 13 juillet 1982 a été votée à la suite de violentes inondations produites à la fin de l'année 1981 et au début de 1982 dans les vallées de la Garonne, de la Saône et du Rhône. En effet, ces catastrophes ont permis de mettre en lumière le fait qu'il n'existe presque aucune garantie d'assurance contre les risques naturels et donc les inondations.

Il est important de comprendre comment les gouvernements ont mis en place des mécanismes de protection contre ce risque extrême et inhabituel. En effet, il est fortement corrélé et volatil. C'est pourquoi, le fonctionnement du système Cat-Nat sera d'abord évoqué (1.2.1) pour permettre de comprendre la couverture du risque Cat-Nat et le rôle de la CCR (1.2.2).

## 1.2.1 Le fonctionnement du système

### 1.2.1.1 Un système français égalitaire et unique

La France, en raison de son histoire centralisé et étatique, a décidé que la solidarité et l'égalité face aux catastrophes naturelles devraient être un droit constitutionnel (paragraphe 12 de la Constitution du 27 octobre 1946). Elle a ainsi créé un partenariat unique entre l'État et les compagnies d'assurances.

L'État français exerce beaucoup de responsabilités et de pouvoirs dans ce partenariat. L'idée est de créer une protection commune et inclusive tout en garantissant sa durabilité, sa solvabilité et son efficacité. Il est financé par la fixation d'une prime supplémentaire à tous les contrats d'assurance. En ce qui concerne les contrats, à l'exclusion des véhicules motorisés, il s'agit d'une prime de 12 %. En ce qui concerne la protection des véhicules motorisés contre le vol ou l'incendie, elle est de 6%. Cela représente un coût moyen très acceptable : 21 euros pour les véhicules non-professionnels, et 2,6 euros pour les automobiles. Les franchises sont également fixées par la loi et sont réparties entre les biens à usage professionnel et les autres. Lissés sur tout le territoire français, quels que soient les risques, ils assurent l'inclusion de tous ses habitants. Ce travail en collaboration avec la planification des risques au niveau municipal, principalement liés aux inondations, permet de prévenir la construction dans les zones les plus risquées. Les municipalités qui ne mettent pas en œuvre de tels plans font face à des franchises plus élevées.

Ce partenariat dépend de deux conditions qui déterminent la possibilité de paiements d'assurance :

- **Condition étatique** : un arrêté de constatation doit être publiée au Journal officiel. Cela signifie que l'État français doit reconnaître légalement l'événement comme une catastrophe naturelle. Ainsi, dans la pratique, dès l'apparition d'un événement naturel dommageable, les maires formulent une demande de reconnaissance par l'État de l'événement produit telle qu'une catastrophe naturelle. Cette demande est elle-même transmise au préfet du département. Une commission interministérielle<sup>9</sup> est alors chargée d'énoncer un avis quant à l'état, ou l'absence de catastrophe naturelle. Cela s'effectue à partir de dossier contenant des rapports transmis par les organismes scientifiques compétents, et ce, en fonction des périls. Ce processus est donc initié par le maire de la localité concernée et passe ensuite au niveau ministériel pour la décision finale, sur la base de critères complexes et techniques.
- **Condition d'assurance** : les biens en question doivent être couverts par une assurance. Les experts évalueront les coûts, et la compagnie d'assurance fournira les paiements initiaux. Cela garantit une réponse réactive, professionnelle et locale à tout événement.

Après avoir été fortement critiqués de par les lenteurs administratives ou encore le manque de communication sur les décisions prises, le régime Cat-Nat a été modifié. En effet, une nouvelle loi a été promulguée le 28 décembre 2021 et a été publiée dans le « Journal Officiel » le lendemain. La loi facilite la reconnaissance de l'état de catastrophe naturelle, améliore et accélère l'indemnisation des victimes, tout en renforçant la transparence des procédures. Il comprend également des mesures sur le risque de sécheresse-réhydratation des sols.

---

9. Elle est représentée par les ministères ci-après et la CCR en assure le secrétariat.  
 - Le ministère de l'Intérieur, de l'Outre-Mer, et des collectivités territoriales et de l'Immigration  
 - Le ministère de l'Économie, des Finances et de l'Industrie  
 - Le ministère du Budget, des Comptes publics, de la Fonction publique et de la Réforme de l'État  
 - Le ministère de l'Écologie, du Développement Durable, des Transports et du Logement

Le système français à l'État en son cœur, ce qui n'est pas le cas dans chaque pays. À titre d'exemple, aux Etats-Unis, toute personne peut choisir d'acheter une protection ou non. Au niveau fédéral, seulement 12 % des ménages disposent d'une protection spécifique contre les inondations. Cette protection est alors tarifée en fonction des risques existants.

### 1.2.1.2 Liste courante des périls couverts par le régime

Ce système couvre tous les événements naturels extrêmes, qui sont :

- Les **inondations**.
- La **subsidence** – également appelé la **sécheresse** –, il s'agit en réalité de mouvements de terrains argileux qui font suite à la baisse de la teneur en eau des sols.
- Les **mouvements de terrains** (glissements de terrains, chute de blocs ou encore les effondrements de cavités souterraines).
- Les **ruissellements d'eau et de boue** .
- Tout phénomène relié à **l'action de la mer**, c'est-à-dire submersion marine, recul du trait de côte par érosion marine, etc.
- Les **avalanches**.
- Les **séismes**.

En raison de ses nombreux territoires partout dans le monde, la France est également exposée à des événements plus tropicaux qui sont parmi les plus coûteux, comme :

- Les **vents cycloniques en Outre-Mer**
- Les **effets du volcanisme actif**

L'événement individuel le plus coûteux est l'ouragan Irma qui a dévasté les Antilles françaises en 2017 et a coûté environ 2 milliards d'euros. Les deux inondations les plus coûteuses sont en 4ème et 5ème position avec environ 1,085 milliards de dommages et sont survenus respectivement en 2002 et 2003. L'année 2003 a été particulièrement sévère avec des inondations combinées avec une sécheresse et une vague de chaleur extrêmement graves qui ont causé les deuxièmes plus gros dégâts à 1,9 milliard d'euros. Tout cela permet d'indiquer qu'il existe une potentielle relation entre les inondations et la sécheresse. Cette potentielle corrélation sera étudiée au travers de la partie 3.

Il est également important de souligner le fait que cette liste est susceptible d'évoluer, dans le cadre où certains événements peuvent encore s'abattre sur le pays (missiles ou terrorisme par exemple).

## 1.2.2 Comprendre la couverture du risque Cat-Nat et le rôle de la CCR

En France, plusieurs protections existent lorsqu'une catastrophe naturelle se produit : l'assurance individuelle, la protection des cultures et du bétail assurée par le ministère de l'Agriculture, le Fonds pour la prévention des grands risques naturels qui fournit de l'argent lorsque des expropriations sont nécessaires, et le régime d'assurance contre les catastrophes naturelles qui couvre les dommages restants. Ce dernier régime relève de la responsabilité de la **CCR**.

Les inondations sont les catastrophes naturelles qui ont historiquement le plus d'impact socio-économique après les sécheresses. Dans cette optique, les mesures prises par les administrations publiques

doivent être orientées vers une gestion adéquate des risques liés aux inondations afin de les minimiser et d'en atténuer les conséquences.

Fondée par la loi du 25 avril 1946, après les troubles de la Seconde Guerre mondiale, la **CCR** est une société non cotée détenue à 100% par l'État français. La **CCR** est classée parmi les 20 premiers réassureurs mondiaux, et bénéficie d'une notation AA<sup>10</sup> par l'agence *Standard and Poor's*<sup>11</sup> et A+<sup>12</sup> par l'agence *AM Best*<sup>13</sup> (avec perspective stable). Le choix de ces deux agences vient du fait qu'en Europe, *Standard Poor's* a une influence prépondérante alors que sur le marché américain *AM Best* occupe cette position de leader du marché. Les notations attribuées par ces agences de notations sont essentielles dans le monde de la réassurance. Ainsi, la plupart des cédantes refuseront de placer leurs programmes RC à des réassureurs ne disposant pas d'un *rating* jugé suffisant. Ce succès est l'héritage d'une longue histoire, au cours de laquelle la **CCR** a développé ses activités de couvertures de réassurances des catastrophes naturelles. La **CCR** a une garantie de l'État, et se distingue de ses concurrents en proposant des couvertures illimitées pour des branches spécifiques au marché du territoire français. En effet, c'est la loi du 13 juillet 1983 qui habilite la **CCR** à couvrir en réassurance – avec garantie de l'État – tous les risques résultant des conséquences des catastrophes naturelles.

En cas d'événement vraiment extrême, la **CCR** agira et assurera les compagnies d'assurance avec des paiements très importants. En ce qui concerne la garantie liée aux catastrophes naturelles, la **CCR** propose un schéma de réassurance, négocié avec la FFSA<sup>14</sup> et le GEMA<sup>15</sup>. Elle comporte une quote-part à 50 % et un *stop-loss*<sup>16</sup> sur rétention de portée limitée dont la priorité est le plus souvent de 200%. Cette couverture non-proportionnelle est tarifée cédante par cédante. Si cette garantie n'est pas suffisante, l'État français pourra agir comme un garant en dernier ressort.

Néanmoins, cette garantie de l'État ne confère pas à la **CCR** l'exclusivité de la réassurance des catastrophes naturelles. Effectivement, tout assureur reste libre de se garantir auprès du réassureur de son choix. Il a même la possibilité de porter le risque de ne pas se réassurer du tout, tout en respectant les besoins en capital

Si l'événement n'est pas reconnu, la compagnie d'assurance exécutera ses activités habituelles et pourra acheter une protection auprès de la division commerciale du marché libre de **CCR**.

La **CCR** gère également certains Fonds Publics pour le compte du gouvernement français qui sont liés à des risques extrêmes :

- Le **Fonds National de Gestion des Risques en Agriculture**. (FNGRA).
- Le **Fonds de Prévention des Risques Naturels Majeurs**. (FPRNM).
- Le **Fonds de Compensation des Risques de l'Assurance de la Construction** (FCAC).
- Le **Fonds de Garantie des Risques liés à l'Épandage** agricole des boues d'épuration urbaines et industrielles. (FGRE)
- Le FSCI : le **Fonds de Sécurisation du Crédit Inter-entreprises**. (FSCI)

10. Source : [\[en ligne\]](#)

11. *Standard & Poor's* (S&P) est une agence américaine (filiale de l'entreprise McGraw-Hill) publiant des analyses financières quant aux actions et obligations.

12. Source : [\[en ligne\]](#)

13. *A. M Best* est également une agence de notation internationale, spécialisée dans le secteur de l'assurance.

14. Fédération Française des Sociétés d'Assureurs

15. Groupement des Entreprises Mutuelles d'Assurance

16. Excédent de perte annuelle représentant une réassurance non-proportionnelle qui protège l'assureur contre de mauvais résultats. Les concepts de base de la réassurance sont présentés en annexe A

— Le **Fonds de Garantie des dommages consécutifs a des Actes de Prévention, de Diagnostique ou de Soins** dispensés par des professionnels de santé exerçant à titre libéral. (FAPDS)

La **CCR** reste toutefois la seule des entreprises de son secteur d'activité a proposer un large éventail de formules de réassurance, et ce, a couverture illimitée. Ceci représente un avantage pour les assureurs, puisqu'ils bénéficient d'une sécurité maximale dans l'hypothèse d'un sinistre majeur aux dommages multiples (inondation, sécheresse ...).

Son principal devoir est donc d'offrir des polices de réassurance contre les risques extrêmes aux compagnies d'assurance françaises. Ces risques incluent les catastrophes naturelles mais aussi les attaques terroristes. Sa garantie d'État permet de fournir un soutien digne de confiance contre des événements qui peuvent s'avérer très coûteux. Une autre de ses missions consiste à recueillir des données et à élaborer des modèles de risque. Ces modèles sont le fruit d'une collaboration avec la communauté scientifique.

## 1.3 Rappels généraux sur les inondations

Cette partie a pour vocation de permettre de présenter les différents types d'inondations existants (1.3.1) en apportant une attention particulière à l'inondation étudiée : la submersion marine (1.3.2).

### 1.3.1 Typologies d'inondations

Les inondations appartiennent à la catégorie des catastrophes naturelles hydrologiques, c'est-à-dire relevant du cycle de l'eau. Une inondation est définie comme une montée et une propagation de l'eau – qu'elle soit rapide ou lente – sur des terres qui ne sont habituellement pas submergées.

Deux types d'inondations sont catégorisées :

- Les **inondations marines**, c'est-à-dire impliquant une inondation par l'eau de la mer.
- Les **inondations terrestres**, c'est-à-dire impliquant une inondation par des eaux douces.

Parmi les inondations marines, il existe trois grands types qui se différencient par leur fait générateur.

Premièrement, il existe ce que l'on appelle la **crue de rivière** – *fluvial flood* ou *riverine flood* –. La crue de rivière est causée par de fortes pluies ou d'importantes fontes de neige, faisant déborder un cours d'eau de son lit. Elle est notamment caractérisée par sa rareté, sa longue durée (jusqu'à plusieurs semaines), et le fait qu'elle soit particulièrement bien localisée. Il s'agit de l'inondation marine la plus courante en France, et concerne souvent les grands bassins versants (comme la Saône, la Loire, la Meuse ou encore la Seine). Il n'est pas sans rappeler la célèbre crue de la Seine en janvier 1910. Cet événement avait coûté plus de 400 millions de franc-or de l'époque en dommages directs, ce qui équivaldrait à 1,6 milliard d'euros<sup>17</sup>.

Deuxièmement, il existe l'**inondation éclair** – *pluvial flood* ou *flash flood* –. L'inondation éclair est causée par de très fortes précipitations en un intervalle de temps très court, entraînant crues et ruissellements. Elle se caractérise par un fort débit qui est extrêmement dévastateur. L'inondation se concentre dans les petits bassins versants (dans les cours d'eau des montagnes ou encore vers les rivières méditerranéennes). Les inondations éclair se localisent principalement sur la côte sud-est, comme

17. D'après Auguste Pawlowski, Albert Radoux, Les crues de Paris (vie-xxe siècle). Causes, mécanisme, histoire, dangers, la lutte contre le fléau, Berger-Levrault et Cie, Paris, 1910 : [en ligne](#)



en témoignent les inondations du Gard (30) en septembre 2002, celles du Rhône (69) en décembre 2003 ou bien celles ayant touchés le Var (83) en juin 2010. Le premier exemple cité figure parmi les événements les plus coûteux de l'histoire de l'assurance française en coûtant plus de 800 millions d'euros<sup>18</sup>.

Enfin, le troisième et dernier grand type d'inondation marine est la **submersion marine – coastal flood**. Une submersion marine est définie comme une inondation temporaire de la zone côtière par la mer, dans des conditions marégraphiques et météorologiques qui sont défavorables. En effet, cette dernière est causée par l'inondation de terres le long des côtes, elle-même due à une marée anormalement haute, pouvant être amplifiée à la fois par la pluie et le vent. Elle affecte les terres peu élevées, et peut être occasionnée par la force d'une tempête du fait des paramètres atmosphériques en vigueur (la température, l'humidité, la direction et la force du vent, la pression atmosphérique et la hauteur des précipitations ...).

Le tableau suivant permet de synthétiser les points ci-dessus :

	Côtier	Intérieur des terres	
	Submersion marine	Inondation éclair	Crue de rivière
Sources	Eau de mer	Surfaces non drainables et rivières mineures	Rivière de taille moyenne à importante, lacs
Causes	Vents d'orage sur la terre ferme poussant l'eau contre une côte pendant des heures.	Précipitations extrêmes en peu de temps, excès d'infiltration du sot et systèmes de drainage submergés	Fortes pluies pendant des jours ou des semaines mais aussi fonte des neiges, embâcles de glaces
Caractéristiques	Hauteur des ondes influencée par les caractéristiques des tempêtes et des côtes, interaction avec les marées	Soudaine et de courte durée (crue soudaine), débits élevés dans les terrains escarpés et les débris	Les vagues de crue peuvent s'accumuler rapidement (crue éclair) ou graduellement, durent souvent de longues périodes
Zones à risque	Zones côtières	Partout y compris dans les zones urbaines	Zones proches des rivières, plaines inondables

**Table 6** – Les principaux types d'inondations

## 1.3.2 Les submersions marines

### 1.3.2.1 Typologie et paramètres des submersions marines

Cette étude a pour vocation d'analyser les inondations marines, et plus particulièrement celles par vagues submersives.

Il existe trois différents types de submersions marines (Cariolet, 2011) :

- La **submersion par débordement** : cela signifie que le niveau d'eau est supérieur à l'altitude du rivage (cote de crête ou terrain naturel).
- La **submersion par franchissement** : elle a lieu si des paquets de mer liés au déferlement des vagues viennent inonder les aménagements ou les ouvrages<sup>19</sup>. Par ailleurs, ces franchissements par paquets de mer peuvent s'accompagner de projections très importantes de galets lorsque le phénomène prend place dans une plage de galets, amplifiant ainsi son fort impact destructif.
- La **submersion par rupture du système de protection** : cela se produit lorsque les terres en arrière d'un ouvrage se trouvent inondées si ce dernier se rompt. Cette défaillance peut naître

18. D'après le site de l'argus de l'assurance : [\[en ligne\]](#)

19. Les ouvrages hydrauliques sont des systèmes de protection contre les inondations/ submersions, regroupant les barrages, canaux, digues...

de plusieurs causes : un déséquilibre sédimentaire<sup>20</sup> du cordon littoral, une érosion<sup>21</sup> chronique importante ou encore un mauvais entretien des ouvrages de protection. Ce type de submersion ne sera pas étudié dans ce mémoire.

Les deux premiers types de submersions ne seront pas distingués par manque d'information topographique. Leur formation est donc une combinaison de plusieurs phénomènes aléatoires.

Tout d'abord, les vagues-submersives peuvent affecter des portions vulnérables des littoraux dans un contexte hydrologique et météorologique très particulier. En effet, ces dernières sont liées à une élévation extrême du niveau de la mer, elle-même due à la combinaison de plusieurs phénomènes.

Premièrement, la submersion est directement reliée à l'**intensité de la marée** – l'oscillation quotidienne de la mer dont le niveau monte et descend alternativement –. Effectivement, plus le coefficient est important, plus le niveau de la mer à marée haute est élevé.

Dans un second temps, la surélévation du niveau marin peut être produite après le passage d'une tempête : on l'appelle la **surcote**. Cette surcote se crée selon trois processus principaux :

- D'abord **les vagues** qui sont liées à la forte houle, contribuent à augmenter la hauteur de l'eau.
- Puis, le **vent** exerçant des frottements à la surface de l'eau génère une modification des courants et ainsi du niveau de la mer.
- Enfin, il y a une diminution de la **pression atmosphérique** : cela signifie que, puisque le poids de l'air décroît à la surface de la mer, mécaniquement, le niveau de cette dernière monte. Par ailleurs, une diminution de la pression atmosphérique d'un hectopascal (hPa)<sup>22</sup>, est approximativement égale à l'élévation d'environ un centimètre de la hauteur d'eau. Prenons l'exemple suivant ; une dépression de 950 hPa génère une surélévation d'approximativement 65 cm (soit une différence de 65 hPa par rapport à la dépression atmosphérique moyenne qui est de 1015 hPa).

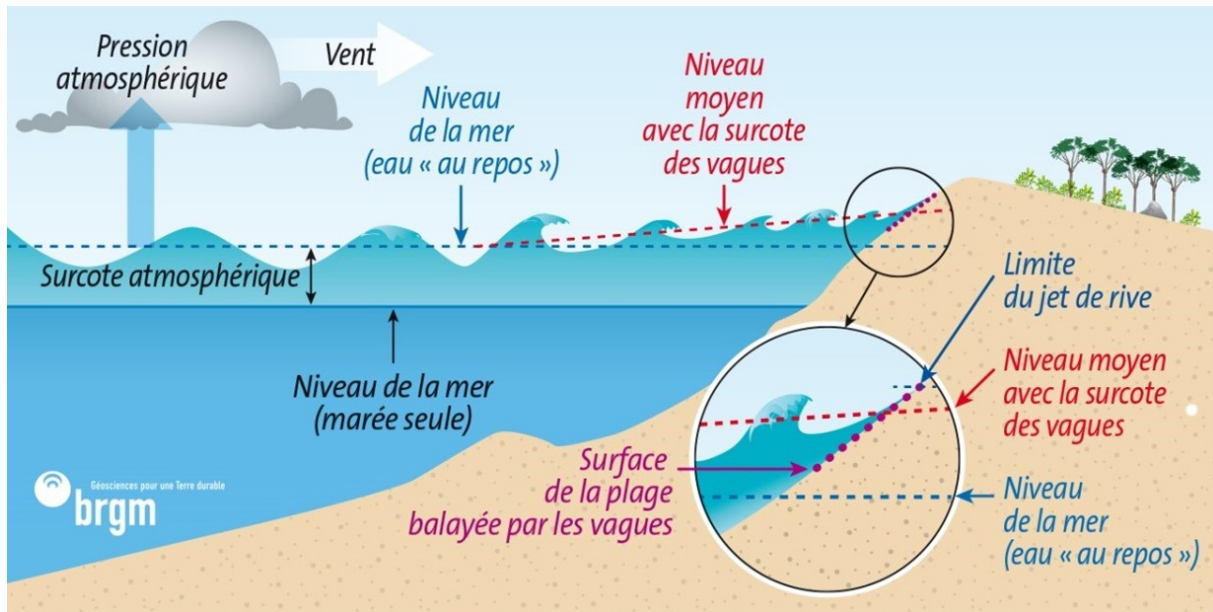
Pour finir, peut se rajouter aux différents phénomènes cités précédemment celui du **déferlement des vagues** à la côte. Ce dernier se traduit par une projection violente de masses d'eau.

20. Les roches sédimentaires sont par exemple le charbon ou le calcaire.

21. L'érosion est une usure/ transformation que les eaux et actions atmosphériques font subir à l'écorce terrestre (sont enlevés par le vent, la pluie, les glaciers...).

22. Unité de mesure de la pression atmosphérique, équivalant à cent pascals.

Le schéma<sup>23</sup> ci-dessous permet d'illustrer les principaux paramètres entrant en compte dans l'évaluation des niveaux marins extrêmes.



**Figure 13** – Schéma illustratif des principaux paramètres d'une submersion marine, source : BRGM

### 1.3.2.2 Les dangers liés aux submersions marines et les facteurs aggravants

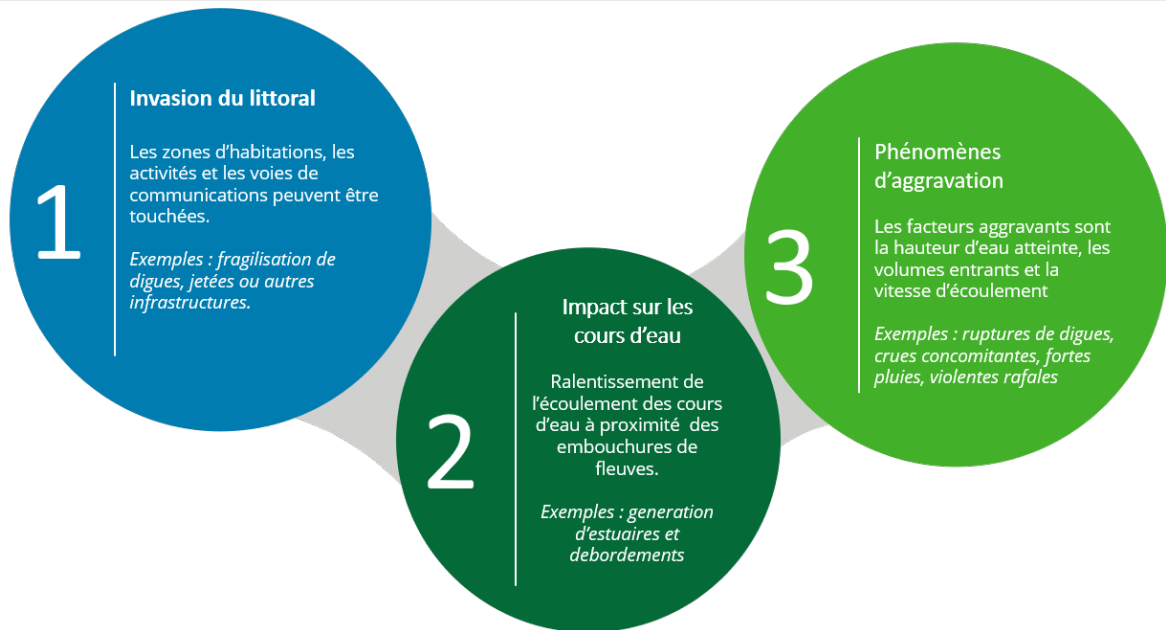
Il est tout d'abord à noter que les vagues submersives touchent principalement les zones basses proches du littoral. En revanche, les inondations – qui sont elles-mêmes dues aux submersions marines – sont dans la capacité d'atteindre une hauteur d'eau de plusieurs mètres. Ainsi, elles peuvent envahir le littoral sur une distance de plusieurs kilomètres, à l'intérieur même des terres. Dans cette perspective, les zones d'habitations, d'activités ou encore les voies de communication sont susceptibles d'être endommagées en très peu de temps. Les digues, jetées ou autres infrastructures côtières peuvent aussi être franchies et fragilisées. De plus, à proximité des embouchures de fleuves : des estuaires et des débordements peuvent être générés, conséquence du ralentissement de l'écoulement des cours d'eau. Enfin, les dégâts peuvent également être aggravés en cas de ruptures de digues, de crues concomitantes, de violentes rafales de vent ou encore de fortes pluies.

L'existence-même de la simultanéité des phénomènes cités précédemment aggrave la submersion. Cela entraîne l'accroissement des débordements et permet ainsi à la mer d'atteindre des zones qui sont habituellement abritées. Cette gravité des débordements varie en fonction de trois facteurs directement liés aux phénomènes décrits précédemment :

Le schéma ci-après permet d'avoir une vision synthétique des éléments évoqués :

- la hauteur d'eau atteinte
- les volumes entrants
- la vitesse d'écoulement des eaux

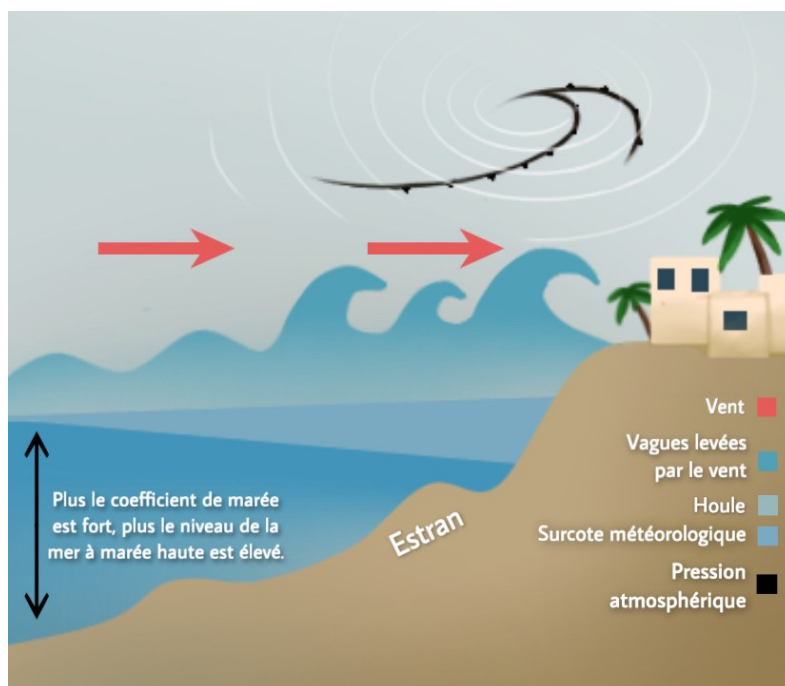
23. "Établissement public de référence dans les applications des sciences de la Terre pour gérer les ressources et les risques du sol et du sous-sol dans une perspective de développement durable" [\[en ligne\]](#)



**Figure 14** – Schéma synthétique des dangers liés aux submersions marines et facteurs aggravants

L'intensité de la submersion marine dépend également de la configuration des fonds marins, de la zone couverte et découverte par la marée, communément appelée l'estran, et d'autres caractéristiques géographiques des côtes. Trois grandes caractéristiques géographiques sont à relever. La diminution de la profondeur de la mer (à l'arrivée sur la côte, l'énergie des vagues se transforme en surélévation du niveau d'eau) la nature des fonds (qui va accélérer ou freiner la propagation de la vague vers la côte grâce à du sable ou des galets par exemple). Enfin, la troisième et dernière caractéristique géographique est l'orientation de la côte. Cette dernière s'établit en fonction de la direction de propagation de la houle et des vagues.

Le schéma ci-après permet d'avoir une vision synthétique des éléments évoqués :



**Figure 15** – Schéma illustratif des principaux phénomènes influant sur une submersion marine

Il est également à souligner que, selon l'INSEE, 4,5 millions d'habitants supplémentaires sont attendus sur le littoral français d'ici 2040. Cela pourrait alors augmenter les conséquences économiques de tels événements.

### 1.3.2.3 Exemples de submersions marines marquantes en France ces dernières décennies

Plusieurs événements mortels de type submersion marine se sont produits en France au cours des dernières décennies. La tempête *Xynthia* reste à ce jour l'événement le plus récent et destructeur sur le littoral français. La conjonction de vents violents et de fortes marées a donné lieu à une onde de tempête et une submersion marine, occasionnant d'importantes inondations. Cette tempête s'abattit sur plusieurs pays européens entre le 26 février et le 1er mars 2010. Elle toucha la France dans la nuit du 27 au 28 février 2010, balayant le littoral atlantique, propageant l'eau très rapidement dans les terres, et ce, malgré la présence de structures de défense. 11 départements furent touchés, et plus particulièrement la Vendée (85), la Charente-Maritime (17), les Deux-Sèvres (79) et la Vienne (86). La tempête tua plus de 50 Français et engendra près d'un milliard d'euros de dommages. Lors du passage de la tempête, l'eau de la mer a atteint près de 2 mètres dans les habitations de certains endroits. Les conditions atmosphériques réunies dans la nuit de la tempête *Xynthia* provoquèrent une surcote météorologique à marée haute, avec un coefficient de 102 et une forte houle. La Rochelle (17) par exemple, enregistra une surcote exceptionnelle d'1m53, ce qui provoqua la rupture de nombreuses digues. À ce moment-là, la mer avait d'ores et déjà dépassé de plus d'un mètre le niveau des plus grandes marées observées. De plus, lors de son passage, la tempête *Xynthia* entraîna des rafales de vent atteignant près de 160 km/h.

Le facteur aggravant de la tempête *Xynthia* est son coefficient de marée, ce qui la différencie par exemple des tempêtes *Lothar* et *Martin* (décembre 1999) et *Klaus* (24 janvier 2009). En effet, ces tempêtes se caractérisèrent par l'intensité de la vitesse du vent, ainsi que la présence d'une forte houle et une dépression atmosphérique importante. En effet, en décembre 1999, la France fut ravagée par deux tempêtes successives : **Lothar et Martin**. *Lothar*, la première, frappa le 26 décembre 1999 les régions de la moitié nord du pays, et le lendemain, le 27 décembre 1999, *Martin* balaya la moitié sud du territoire. Considérées encore aujourd'hui comme la « tempête du siècle », cette catastrophe naturelle reste la plus coûteuse et la plus meurtrière en France (92 morts au total). Le vent atteignait jusqu'à 169 km/h dans Paris, 200 km/h à l'Île de Ré (17) ou encore 210 km/h au pont de Tancarville (76). Ainsi, de nombreuses habitations et infrastructures furent inondées en Charente-Maritime (17) et en Gironde (33), provoquant même un incident à la centrale nucléaire du Blayais (33).

Le 24 janvier 2009, le Golfe de Gascogne subit une forte dépression atlantique : il s'agit de la tempête *Klaus*. Cette dernière traversa la France, de la Charente-Maritime (17) à la région lyonnaise et à la Côte d'Azur. La catastrophe fit 12 victimes, et engendra 1,2 milliards d'euros de dégâts matériels. La tempête généra des vents particulièrement violents, notamment sur le quart Sud-Ouest. De nombreux records locaux ont été battus au Languedoc-Roussillon (91), conséquences d'une violente tramontane. Cette dernière fut accompagnée d'importantes crues sur les bassins de l'Adour-Garonne (26) et de la Charente (16).

## 1.4 Rappels sur la modélisation du risque inondation

Cette partie présente tout d'abord la modélisation générale des catastrophes naturelles (1.4.1). Puis, un aperçu des principales techniques utilisées dans le cas du risque inondation sont présentées (1.4.2).

Pour finir, les principales approches envisageables dans le cas spécifique de la submersion marine seront discutées (1.4.3).

### 1.4.1 La modélisation des Cat-Nat

Afin d'évaluer un risque physique comme l'inondation, il faut pouvoir déterminer sa probabilité d'occurrence et la gravité de ce risque. En outre, une analyse doit être effectuée, ce qui implique une étude détaillée des facteurs de risque : l'exposition, l'aléa, et la vulnérabilité. Ces facteurs de risques sont les trois modules principaux d'un modèle Cat-Nat.

Le **module exposition** recense l'ensemble du phénomène avec ses caractéristiques et les valeurs assurées. La cartographie produite servira ensuite de référence pour le module aléa.

Le **module aléa** a pour objectif de caractériser l'intensité du phénomène physique étudié. En d'autres termes, il permet d'évaluer les variables physiques les plus à même d'expliquer les dégâts qui sont engendrés par un événement. Il permet donc de représenter les zones les plus touchées par l'événement. Dans ce contexte, la cartographie des risques d'inondation est un élément de base des études d'analyse. Elle permet d'évaluer efficacement la distribution spatiale des divers éléments de gravité (comme le niveau de la surface de l'eau, la vitesse du débit, le transport des sédiments ou les temps caractéristiques) et la fréquence (périodes de retour ou probabilité de dépassement) du phénomène d'inondation.

Le **module vulnérabilité** est un croisement des deux modules précédents permettant de déterminer les pertes associées au phénomène. Il requiert généralement une base de données des sinistres provenant des assureurs.

Le module de vulnérabilité ne sera pas traité ici mais pourrait faire l'objet d'une analyse spécifique pour améliorer le modèle. Du fait de l'analyse des risques de submersions marines liés au changement climatique, la principale composante ici étudiée est l'évolution des aléas climatiques.

Par ailleurs, différentes approches sont envisageables en ce qui concerne la modélisation du changement climatique. Ces dernières sont présentées succinctement par degré de complexité ci-après :

- L'**approche simplifiée** est strictement basée sur des facteurs proportionnels simples :

$$\text{Sinistralité en 2050} = \text{Sinistralité en 2020} \times (1 + \text{facteur aléa}) \times (1 + \text{facteur exposition}) \times (1 + \text{facteur vulnérabilité})$$

France Assureurs avait mis en place cette approche il y a quelques années au moment de la COP21. Les résultats produits se sont trouvés être aberrants et justifiaient les deux approches suivantes.

- Une **approche intermédiaire** serait d'utiliser des sources de données publiques, en approche stochastique afin d'affiner l'évaluation de la fréquence des événements Cat-Nat.
- Une **approche intégrale** serait d'avoir un catalogue stochastique d'événements pour le climat présent et futur. L'intégration des données Cat-Nat permettrait une prise en compte explicite de la vulnérabilité présente et future. Cette approche est la plus complexe du fait des modèles nécessaires et de la quantité et qualité de données nécessaires.

### 1.4.2 Méthodes d'analyse du risque d'inondation

Différentes méthodes existent pour tenir compte du risque inondation (Díez-Herrero et al., 2009). Les modèles peuvent être plus ou moins simplifiés (uni, bi, ou tri dimensionnels) pour la prise en compte

des différents paramètres comme la profondeur et la vélocité. Toutefois, la modélisation peut s'avérer complexe et les coûts de calculs peuvent vite devenir exorbitants.

#### 1.4.2.1 Méthodes historiques et paléohydrologiques

- **Historique** : étude de documentations, d'archives et des marques physiques laissées pour référencer les différentes inondations sur une période donnée.
- **Paléohydrologique** : informations compilées grâce à l'étude géologiques de marques et de dépôts laissées par des inondations passées. Elles peuvent être datés grâce à diverses techniques.

A travers une analyse simple, ces deux méthodes permettent de cartographier des zones d'inondations historiques. Une analyse plus complexe vise à convertir les niveaux historiques en débits circulant afin de pouvoir les utiliser dans des modèles hydrauliques classiques.

#### 1.4.2.2 Méthodes hydrologiques et hydrauliques

- **Hydrologique** : estime les rejets générés dans un bassin versant. Des analyses statistiques sont réalisées sur les pics de débits ou de précipitations à travers des modèles hydrométéorologiques de conversion pluie-débit.
- **Hydraulique** : cherche à calculer et modéliser la profondeur et la vélocité des rejets en se basant sur différentes hypothèses et simplifications du flux de l'eau.

Ces deux méthodes permettent d'obtenir un hydrographe. Cela représente une étape fondamentale car l'hydrographe permet d'assigner des probabilités d'occurrence à de nombreux paramètres physiques.

#### 1.4.2.3 Méthodes géologiques et géomorphologiques

- **Géologique** : étudie la nature et les caractéristiques des milieux pour comprendre la propagation des inondations.
- **Géomorphologique** : délimite les zones géomorphologiquement actives le long des rives et donc plus propices aux inondations en se fondant sur l'étude des sédiments

Cette partie représente une vision simplifiée des principaux groupes de méthodes d'analyses existantes pour traiter de l'inondation. Ces méthodes ne sont pas exclusives et doivent au contraire être utilisées de façon complémentaire.

### 1.4.3 Modélisations possibles des Submersions marines en France

Afin de prévoir l'importance des événements potentiels futurs, J-P Naulin et al<sup>24</sup> de la CCR, ont montré que deux approches peuvent être envisagées pour la modélisation des submersions marines.

#### 1.4.3.1 Les équations de Saint-Venant comme préambule au modèle déterministe

Lors de la présentation d'un compte-rendu à l'académie des sciences en 1871, Saint-Venant propose un premier modèle pour l'écoulement des canaux à surface libre. Dans sa contribution hydrodynamique, il évoque la théorie des équations générales du mouvement non permanent des eaux courantes. En outre, cela justifie que ces équations soient utilisées dans le cadre de l'océanographie côtière ou des écoulements à faible profondeur telles que les inondations, les ondes de marée ou encore de crue.

---

24. Source : [\[en ligne\]](#)

Les équations de Saint-Venant sont également connues sous l'appellation anglo-saxonne "*Shallow water equations*". Cette dénomination permet de comprendre que la profondeur est petite par rapport à la longueur dans les canaux.

Compte tenu de la dépendance de la hauteur et de la vitesse de l'eau au temps et à l'espace, ce modèle est régi par un système d'équations aux dérivées partielles.

Dans le modèle développé, tous les points d'une même tranche d'eau se déplacent à la même vitesse. L'écoulement est donc considéré comme monodimensionnel.

En effet, on considère :

- Un temps quelconque  $t$  dans un canal ou une rivière
- $U$  la vitesse moyenne des eaux à travers une section transversale dont  $s$  est l'abscisse, sensiblement horizontale comptée à partir d'un endroit quelconque de son cours

Le niveau de la surface fluide peut hausser ou baisser en chaque point sans qu'il y ait d'apport ou d'absorption d'eau par les rives dans l'étendue finie considérée de ce cours

Mathématiquement, deux lois de conservation constituent le modèle de Saint-Venant : celle de la masse et celle de la quantité de mouvement.

### Première équation concernant la masse

Le volume  $\omega ds$  (1) d'une couche comprise entre deux sections avec pour abscisses  $s$  et  $s + ds$  ne doit pas être changé lorsque, après un petit temps  $\Delta t$ , ses abscisses sont devenues  $s + U\Delta t$  et  $s + ds + (U + \frac{dU}{ds} ds)\Delta t$ .

L'épaisseur devient alors :  $ds + \frac{dU}{ds} ds \Delta t$  (2)

La section a ainsi comme superficie :

$$\omega + \left( \frac{d\omega}{dt} + \frac{d\omega}{ds} \times \frac{ds}{dt} \right) \times \Delta t = \omega + \frac{d\omega}{dt} \times \Delta t + \frac{d\omega}{ds} U \Delta t \quad (3)$$

Après égalisation du produit de cette superficie (3) par l'épaisseur (2) au volume (1) et en divisant par  $\Delta t ds$ , on obtient la première équation :

$$d\omega/dt + \frac{d(\omega U)}{ds} = 0$$

Quand la largeur  $a$  est constante et en considérant la profondeur  $y$  on a bien  $\omega = ay$ .

### Deuxième équation concernant la quantité de mouvement

On cherche maintenant l'équilibre dynamique pour le temps  $\Delta t$ .

Soient :

- $\xi$  l'ordonnée verticale de la superficie au dessus d'un plan horizontal ou hauteur de la topographie
- $\chi$  le périmètre mouillé
- $\rho$  la densité ou  $\rho g$  le poids de l'unité de volume



—  $\rho g F$  le frottement

Saint-Venant propose trois quantités de mouvement en fonction de :

— La pesanteur et donc de la topographie (l'angle du fond) :

$$\rho g \omega ds \frac{d(\xi+y)}{ds} t$$

— L'écart entre les pressions des 2 cotés

$$\rho g a [y \frac{y}{2} - 1/2 (y + \frac{dy}{ds} ds)^2] dt = -\rho g \omega ds \frac{dy}{ds} \Delta t$$

— Le frottement du fond :  $-\rho g F \chi ds$

La quantité de mouvement effective, en appelant  $d_c$  une différentielle complète est égale à l'accélération totale :

$$\rho \omega ds \frac{d_c U}{dt} \Delta t = \rho \omega ds (\frac{dU}{dt} + \frac{dU}{ds} U) \Delta t$$

En égalisant la somme des trois premières quantités à la troisième et en divisant tout par  $\rho g \omega ds \Delta t$ , on a la deuxième équation :

$$\frac{d\xi}{ds} = \frac{1}{g} \frac{dU}{dt} + \frac{U}{g} \frac{dU}{ds} + \frac{\chi}{\omega} \frac{F}{\rho g}$$

Dans le cas d'une direction horizontale, la composante du poids de la tranche est nulle.

La principale propriété est que tout comme les équations de Riemann, certaines quantités fonctions de la hauteur  $h$  et de la vitesse de l'eau  $v$ , se propagent comme des ondes à des vitesses et gravités  $g$  différentes. Ces quantités nommées « invariants de Riemann » permettent de mieux comprendre l'écoulement de l'eau le long des canaux. Dans le cas d'une remontée de courant, l'onde circule à la vitesse  $v - \sqrt{gh}$  où  $\sqrt{gh}$  est la célérité de propagation d'après une formule connue de Lagrange.

### 1.4.3.2 Modèle déterministe

Le modèle déterministe est construit en respectant le même schéma que la modélisation modulaire Cat-Nat présentée précédemment : aléa, vulnérabilité et exposition.

Le **modèle d'aléa** cherche à estimer les hauteurs maximales atteintes par l'eau dans les terres pendant l'évènement. Le modèle comporte trois étapes :

- **L'estimation des niveaux d'eau en mer** : Elle dépend de l'effet combiné de la marée et de la surcote météorologique induite par la tempête. La modélisation est réalisée en se basant sur le système d'équations de Saint-Venant pour les écoulements quasi horizontaux sous hypothèses de pression hydrostatique et d'ondes longues. Plusieurs solutions de résolution existent, dont la simulation en interne grâce au modèle Telemac-2D<sup>25</sup> ou l'utilisation du site **MARC**<sup>26</sup> (Modélisation

25. Développé par le Laboratoire National d'Hydraulique et Environnement (LNHE) et la direction des recherches et développements d'EDF pour permettre la résolution de ce cas à deux dimensions d'espace horizontales.

26. Auparavant connu sous le nom de projet PREVIMER, il a comme but de fournir des modélisations et prévisions en temps réel. Pour les zones côtières françaises, de nombreuses variables de prévisions sont disponibles telles que la direction et l'intensité des courants, la hauteur, la fréquence et direction des vagues. Pour plus d'informations : [\[en ligne\]](#)

et Analyse pour la Recherche Côtière), plus moderne et robuste mais limité à l'Atlantique Nord Est. Les principaux résultats du modèle Telemac-2D résident dans le fait d'avoir en chaque point du maillage de résolution, la hauteur d'eau et la vitesse moyenne. Dans ce modèle, la position des vitesses maximales peut différer tout comme présenter une certaine sous-estimation. Ainsi, l'utilisation de Telemac-3D peut être justifiée afin de mieux représenter les écoulements et vecteurs de vitesses. Ce dernier modèle permet de résoudre les équations de Navier Stokes tridimensionnelles de l'hydraulique à surface libre. En outre, il calcule les vitesses dans les trois directions de chaque point du maillage tridimensionnel. Le modèle Telemac est moins privilégié car basé sur des données de moins bonnes qualités. Les erreurs sont plus faibles avec le modèle **MARC** et présentent également des dispersions moins importantes que celles obtenues avec le modèle Télémac. Des écarts importants avec les observations peuvent exister et sont ajustés grâce aux données marégraphiques (ex : 0.4 à 0.5 mètre dans le Perthuis Charentais pendant Xynthia)

- **L'estimation de la surcote de vague** : Les vagues génèrent également une surélévation lors de leur déferlement sur le littoral. Celle-ci dépend entre autres de la configuration du littoral, de la hauteur et de la périodicité des vagues.
- **L'estimation de la propagation de l'eau dans les terres** : Un modèle d'inondation reposant sur le principe du *modèle Lisflood*<sup>27</sup> a été développé. Ce modèle reposant sur la formule de *Manning Strikler* et dont le mode de calcul est basé sur les différences finies, propage l'eau sur un Modèle Numérique de Terrain (MNT)<sup>28</sup> présentant une résolution spatiale de 25 m.

Le **modèle d'exposition** se base sur les portefeuilles des différentes entreprises d'assurance. Il permet donc d'offrir une vision complète de l'exposition au risque. La base de données répertorie les différents biens, leur géolocalisation, l'estimation de leur valeur ainsi que la distance à la cote.

Le **modèle de vulnérabilité** estime le coût de l'évènement ( $C_{(est,i)}$ ) grâce au produit pour chaque bien  $i$  de la probabilité du sinistre ( $P_{sin,i}$ ), par le taux de destruction ( $T_{dest,i}$ ) et par la valeur du bien assuré ( $V_{ass,i}$ ). En d'autres termes :  $C_{(est,i)} = P_{sin,i} \times T_{dest,i} \times V_{ass,i}$ . La probabilité est primordiale car les analyses ont montré un grand nombre de biens assurés mais non sinistrés dans les zones inondées. Pour tenir compte de ce constat et pallier les défauts du modèle d'aléa, une loi de sinistralité, permettant d'estimer la probabilité de sinistre, peut être calibrée à partir de deux variables explicatives : la hauteur d'eau estimée à la position du bien considéré et la surélévation au large. Le taux de destruction sert quant à lui à déterminer la part de valeur assurée endommagée au regard de l'aléa simulé et de l'exposition.

### 1.4.3.3 Modèles probabiliste et stochastique

La différence entre une méthode probabiliste et stochastique est que lorsque l'on utilise la première, on fait un catalogue sur l'historique sans chercher des valeurs extrêmes hors de ce dernier. Il s'agit donc d'un catalogue probable car les événements ont déjà eu lieu dans le passé.

Un catalogue stochastique représente un aléa où chaque paramètre à son intervalle défini avec un échantillonnage intelligent. En combinant tous les espaces on fait un plan d'expérience sachant qu'il y a des paramètres mutuellement excluant.

Ainsi, chaque scénario stochastique est une combinaison d'aléa où les paramètres varient. Il y a donc des

27. Modèle utilisé pour les études sur l'eau et le climat mais aussi pour la prévision des inondations et des sécheresses. Il simule le cycle complet de l'eau par grandes zones en prenant en compte les effets combinés des changements climatiques et météorologiques

28. Hauteur des biens par rapport à l'eau

incertitudes sur tous les paramètres que l'on ne retrouve pas dans une méthode probabiliste. Par exemple, un changement de surcote en passant de 32 cm à 34 cm pour voir son impact.

Le **modèle d'aléa** des modèles probabiliste et stochastique se différencie du modèle déterministe en ne se basant plus sur un événement réel mais en simulant des événements fictifs.

Les simulations se doivent d'être cohérentes avec les observations. Une première version de ce type de modèle a été proposée par JP Naulin et Al. en se basant sur les données marégraphiques historiques des stations du littoral. Les résultats du modèle probabiliste peuvent offrir une première vision de l'exposition du territoire français au risque de submersion marine.

Les **modèles d'exposition et de vulnérabilité** du modèle probabiliste sont repris du modèle déterministe.

### 1.4.3.4 Un choix relevant des besoins de l'étude réalisée

#### 1.4.3.4.1 Les besoins de projection

Il est question de modéliser le risque physique en prenant compte du changement climatique dans le cadre de l'assurance non-vie. Pour cela, il faut quantifier les pertes assurantielles, et cela à partir des scénarios d'émissions de GES en lien avec les dernières informations du GIEC (scénarios RCP et SSP<sup>29</sup>).

Afin de capturer la sensibilité des risques à l'utilisation des scénarios, il convient à minima d'utiliser deux scénarios. Lors de cette étude, les scénarios suivants sont étudiés :

- 8.5 : plus défavorable, niveau de risque serait le plus important
- 4.5 : passage à l'action

Pourquoi le besoin de modéliser le risque physique ? car il existe un impact croissant du changement climatique sur les assureurs, mais également une accélération de la part de l'ACPR quant aux bonnes pratiques à exercer pour les assureurs (*stress-tests*<sup>30</sup>, ORSA...).

Pour se faire, il existe deux alternatives pour les assureurs : développer une modélisation interne ou faire appel à des solutions existantes très coûteuses.

Ainsi, les assureurs se doivent d'accompagner les équipes *Risk Management* client dans l'analyse de scénarios type RCP et SSP formulés par le GIEC, et proposer une alternative à la solution proposée par la CCR et aux modèles Cat-Nat.

Cette dernière ayant pour objectif d'avoir plus de transparence, ainsi qu'une meilleure compréhension des sorties des modèles CCR et Cat-Nat.

29. SSP : *Shared Socioeconomic Pathways*, ou scénarios d'évolution socio-économique. Sont élaborés en parallèle des RCP.

30. Une opération de *stress-test* consiste à définir plusieurs scénarios à un horizon d'un ou deux ans qui seront appliqués aux portefeuilles des banques (crédits, placements, dette) afin de mesurer leur évolution

### 1.4.3.4.2 Le modèle choisi

À la suite des travaux de réassureurs tels que la **CCR** et pour avoir un meilleur contrôle des modèles, certains acteurs du monde de l'assurance sont amenés à rechercher ou développer des outils permettant de comparer les différents modèles climatiques.

Cette méthodologie de la modélisation prédictive du risque physique climatique a pour principaux objectifs :

- D'accompagner les équipes de Risk Management client dans l'implémentation des stress-tests et l'analyse de scénarios type **RCP/SSP**.
- De proposer une alternative à des solutions de type **CCR** et aux modèles Cat-Nat.
- D'enrichir la prise de décision stratégique à partir d'un outil de visualisation des risques.

Ce modèle prédictif repose essentiellement sur la recherche des sources de données climatiques. En effet, la modélisation prédictive du climat est fortement dépendante des données utilisées en entrée des modèles pour définir l'aléa climatique. En d'autres termes, la modélisation développée est fondée sur des variables explicatives qui prennent en compte le changement climatique.

En effet, en assurance, les données sont exposées à un cadre réglementaire de plus en plus strict et exigeant. Par soucis de solvabilité et de transparence des informations, elles doivent être exactes, complètes et à jour.

Les variables explicatives doivent donc respecter les principes suivants :

- **L'utilité** : Tous les résultats trouvés sont fondés sur l'exploitation des données. S'il n'y a pas de données, il est alors impossible de mener une étude en y apportant des résultats et des analyses.
- **La disponibilité** : Une multitude de variables peut être utilisée lors d'une étude. Ici, les données ont l'avantage d'être accessibles librement. Il est commun de se trouver face à un manque ou un défaut d'informations du fait d'une saisie manuelle telles que :
  - La variable existe mais elle est mal ou non renseignée.
  - La variable n'est renseignée que pour certaines périodes.
  - La variable est inexistante.
- **La fiabilité** : Dans le cas contraire, les résultats pourraient être erronés et faire perdre de la crédibilité à l'étude. Dans le cadre de variables climatiques, un important travail de recherche et de traitement est nécessaire. Il est donc crucial de rechercher de manière continue ces sources qui sont régulièrement mise à jour par des sources principalement académiques. Ainsi, certaines bases de données ou variables peuvent être écartées du périmètre de l'étude.
- **La complexité** : Elle est propre à chaque variable climatique. Les sources de données sont publiques mais nécessitent un travail important de nettoyage et d'analyse préliminaires avant même de les insérer dans les modèles. En effet, la nature des données climatiques d'aléa (historiques et prospectives), implique un travail d'harmonisation à travers du *data reengineering*.

Ainsi, le choix d'une approche stochastique peut être justifiée de deux manières. D'une part, compte tenu des données à disposition, relevant uniquement de l'*Open Source*, et du fait que les modèles utilisés dans l'approche déterministe et le modèle probabiliste reposent sur des données passées, pour un historique restreint, le choix se porte vers le modèle stochastique. D'autre part, du fait

de l'impact du changement climatique à prendre en compte et de la variabilité des paramètres, le choix se tourne à nouveau vers une approche stochastique.

Pour finir, afin de pouvoir prendre en compte le changement climatique, la dernière consultation de l'**EIOPA** a servi de référence.

Dans cette dernière, l'intégration du changement climatique dans l'**ORSA** peut être faite à travers 4 méthodes d'analyse de scénarios climatiques face aux risques physiques :

- L'utilisation de l'**outil NGFS Climate Impact Explorer** (open source) avec une précision qui n'est pas encore celle du marché
- L'utilisation des sorties de **Modèles Cat-Nat** : approche la plus précise car elle repose sur les bases de données des agences de modélisation. Néanmoins, cette solution reste coûteuse car elle nécessite de payer les licences annuelles de ces sociétés.
- L'utilisation des **données publiques de stress-tests passés** : approche peu précise avec une granularité géographique qui n'est pas adaptée aux besoins de modélisation.
- L'utilisation des **données de modèles climatiques de type GCM/RCM** : approche qui combine précision et bas niveau d'investissements dans des licences mais qui nécessite le plus de développements internes (approche privilégiée dans la méthodologie). A l'issue de ces développements, la société devient propriétaire de sa méthodologie ce qui est un avantage certain.

## 2 Présentation de la modélisation relative aux submersions marines

### 2.1 Introduction

La méthodologie adoptée repose sur 4 phases principales :

- 1) **L'identification de la variable à expliquer** : Pour mener à bien ce projet, il est important de définir la variable que l'on souhaite prédire. L'objectif final est de projeter des survenances d'évènements climatiques de type submersion marine. Les états Cat-Nat représentent cette variable à prédire. Les données utilisées proviennent de la **CCR** et sont accessibles publiquement.
- 2) **La Sélection des variables explicatives** : Cette étape consiste à identifier les variables explicatives du modèle de prédiction de la submersion marine. Les variables explicatives sont tirées de l'étude de papiers académiques. Elles sont constituées des informations relatives aux niveaux d'eau et à des conditions météorologiques extrêmes.

Ces deux premières étapes font partie du calibrage du modèle à travers la construction des BDD. Elles feront l'objet de la deuxième section de ce chapitre (2.2). L'ensemble de ces variables seront ensuite agrégé en troisième partie (2.3)

- 3) **La détermination de la survenance des évènements par des algorithmes de classification** : Cette étape consiste à mettre en place et utiliser des simulations stochastiques pour prédire les arrêts Cat-Nat de type submersion marine. Cette étape permet de projeter le modèle et fera l'objet de la quatrième section de ce chapitre (2.4).
- 4) **La validation des sorties de modèle par une approche de *back-testing*** : Cette étape nous permet de nous assurer de la cohérence des modèles. Pour cela, des analyses comparatives seront présentées en dernière partie de ce chapitre (2.5).

### 2.2 Calibrage du modèle de submersion marine

Afin de pouvoir prédire les arrêts Cat-Nat de type submersion marine, au regard du changement climatique, il est nécessaire de former deux types de bases.

D'une part, **une base de données historiques** doit être constituée pour construire le modèle de prédiction de submersion marine. Cette base permet de prédire la survenance des submersions de 2021 à 2050. Les données sont divisées en 3 types, issues de sources différentes :

- 1) Les déclarations de catastrophes naturelles
- 2) Les variables climatiques issus de COPERNICUS
- 3) Les variables de niveaux de mer extrêmes provenant d'un papier de recherche<sup>31</sup>

D'autre part, **une base de données prospectives** doit être constituée à l'aide des trois derniers types de sources évoqués précédemment. Cette base de données représente la projection dans 50 ans des variables climatiques explicatives.

---

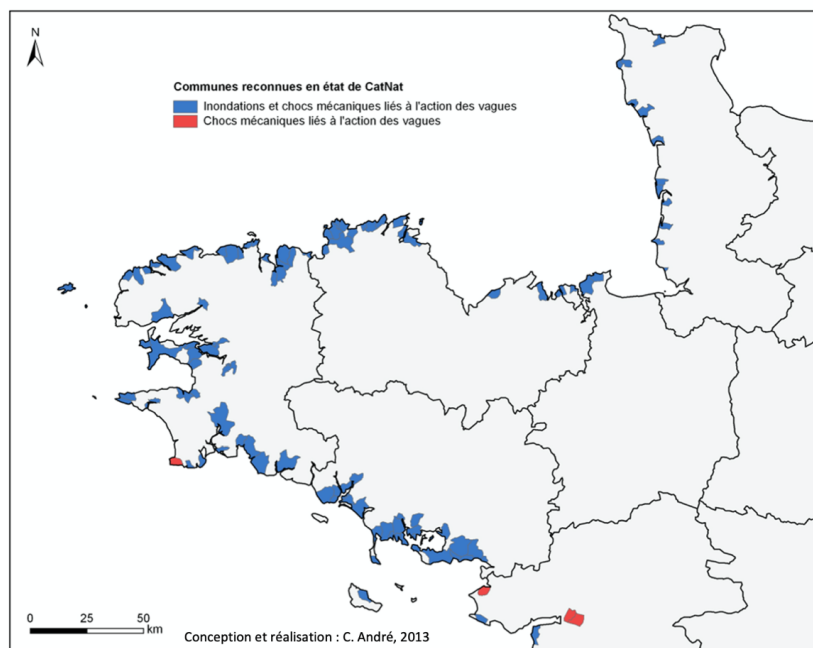
31. Vousdoukas MI, Mentaschi L, Voukouvalas E, Verlaan M, Jevrejeva S, Jackson LP, Feyen L. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard.[\[en ligne\]](#)

## 2.2.1 Les déclarations de catastrophes naturelles

### 2.2.1.1 Le choix de base de données

Lorsque survient un évènement naturel, les victimes ont la possibilité de faire une requête auprès de la **CCR** afin que cet évènement soit classé en tant que catastrophe naturelle. Si l'état de catastrophe naturelle est accepté (Évènements reconnus), les dommages survenus seront alors pris en compte par l'État. La **CCR** met à disposition toutes les reconnaissances d'état de catastrophe naturelle pour toutes les communes de France depuis 1990.

Il serait envisageable de se demander s'il est possible d'utiliser la base de données **GASPAR**<sup>32</sup> (Gestion ASsistée des Procédures Administratives relatives aux Risques naturels), gérée par le ministère de l'Environnement. En effet, elle contient la liste des arrêtés de reconnaissance de l'état de catastrophe naturelle des communes depuis la création régime Cat-Nat. En revanche, l'un des problèmes de la base de données Gaspar, mentionné par J. Douvinet (2006 et 2010), réside dans la codification de l'aléa (dit « nature du phénomène »). On retrouve les phénomènes dit « Inondations et coulées de boue, Chocs mécaniques liés à l'action des vagues et vents cycloniques ». Cela prend donc en compte des événements qui ne sont pas uniquement de type submersion marine. La carte suivante permet d'avoir une représentation de cette base dans le cadre de la tempête Johanna :



**Figure 16** – Carte des communes reconnues en état de catastrophe naturelle après la tempête Johanna (source base de données Gaspar)

Les données fournies par le site de la **CCR** permettent d'établir une base de survenance de submersion marine en dissociant la nature du phénomène.

32. Base GASPAR : [\[en ligne\]](#)

### 2.2.1.2 Présentation de la BDD de la CCR

Cette base de données a été mise en place à l'initiative de la CCR. Cette dernière recense l'intégralité des demandes et les reconnaissances de l'état de catastrophes naturelles auprès du ministère de l'intérieur pour tous types de périls. Dans le cadre des submersions marines, la catégorie retenue s'intitule « Chocs liés aux vagues ».

Dans un premier temps, cette base a pour vocation de calculer les franchises à appliquer dans chaque commune, et ce après la survenance d'un événement qualifié de Cat-Nat.

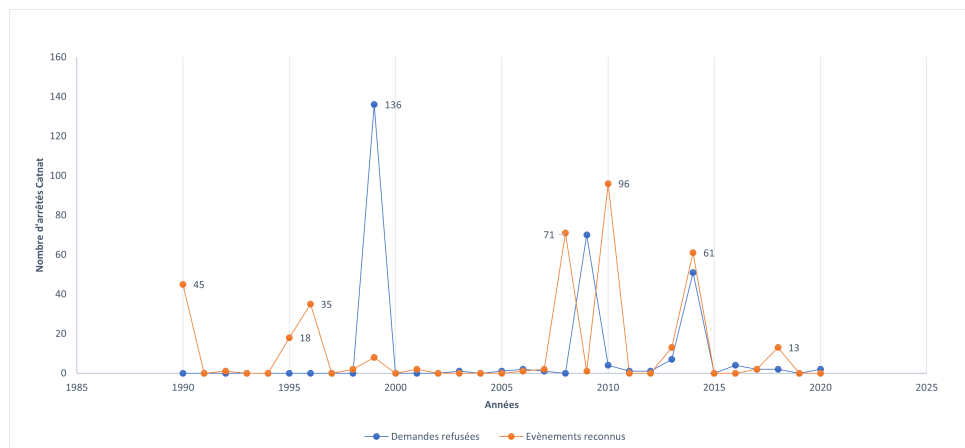
Dans un second temps, cette base a pour mission d'identifier les communes impactées par les Cat-Nat, de caractériser les événements, notamment pour son étendue géographique et sa période de survenance.

Dans notre cas, elle constitue un socle de données historiques non biaisées par l'exposition des assureurs permettant une vision générale du risque submersion marine en France.

Un filtre a été appliqué en considérant pour le besoin de l'étude seulement les départements du littoral. En effet, des observations rares sont créatrices d'incertitude.

Une solution possible est d'utiliser l'information provenant de sites statistiquement similaires<sup>33</sup> afin d'avoir une meilleure précision<sup>34</sup>.

Or, la Méditerranée et l'Atlantique sont des sources de données homogènes. Cela explique le choix de traiter uniquement de la façade atlantique. Le graphique suivant permet de représenter le nombre de ces arrêtés de 1990 à 2020 sur la façade atlantique française<sup>35</sup> :



**Figure 17** – Nombre d'arrêtés Cat-Nat de type submersion marine au cours du temps

Pour le besoin de l'étude, nous avons besoin d'avoir une base complète sur le littoral atlantique français. En effet, le but de la classification est justement de faire la distinction entre les cantons touchés et les cantons non touchés.

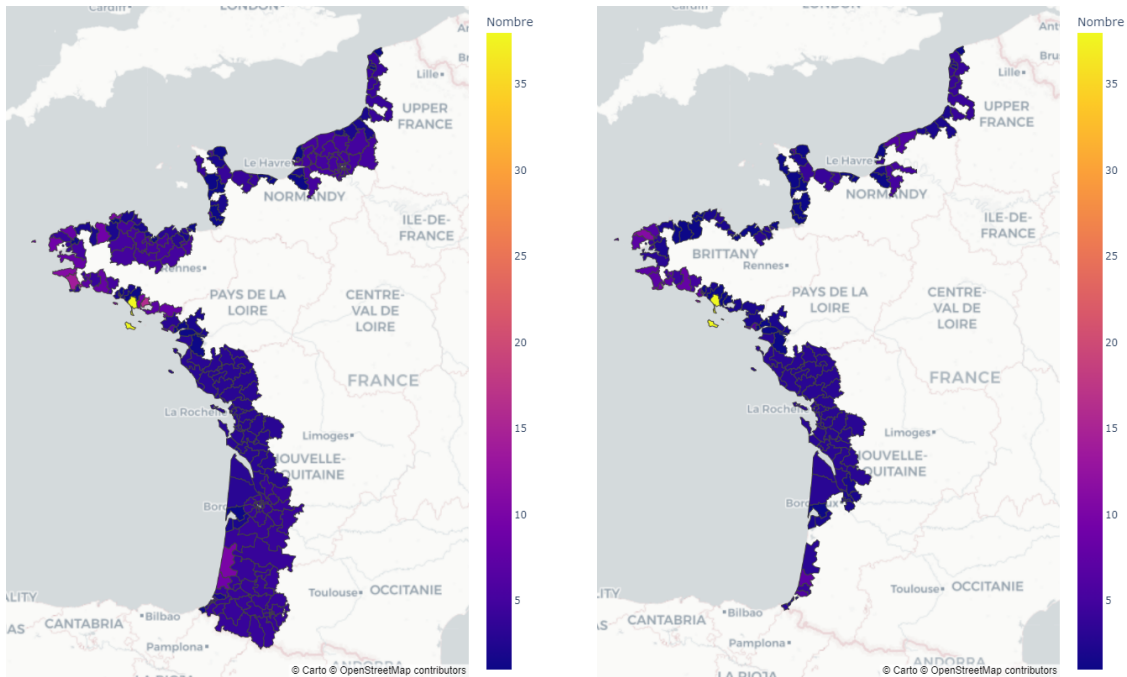
33. Bernardara, P., Andreewsky M. and M. Benoit (2011), Application of the Regional Frequency Analysis to the estimation of extreme storm surges, *Journal of Geophysical Research*, 116, C02008, 1-11

34. Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag

35. Source : données CCR



Les cartes suivantes représentent la base des d'arrêtés Cat-Nat en fonction des cantons français<sup>36</sup> et des différentes classes. La base compile les demandes retenues ou non de 1990 à 2020. On peut identifier une certaine concentration géographique des évènements :



**Figure 18** – Nombre de jours de submersions marines déclarés (reconnues ou non) et reconnues (droite) par canton de 1990 à 2020

On remarque que 5 régions sont affectées dans la base historique :

- 53 : Bretagne
- 32 : Hauts de France
- 28 : Normandie
- 75 : Nouvelle Aquitaine
- 52 : Pays de la Loire

On se rend compte que la commune de Bignan (code canton 56017) située dans le Morbihan en Bretagne est plus impactée que les autres cantons du littoral. Elle ne représente pas une valeur aberrante mais reflète la présence d'un événement de longue durée (du 8/01/14 au 04/02/14).

### 2.2.1.3 Les limites à l'utilisation des arrêtés de catastrophes naturelles

La principale limite réside dans le fait que la base des arrêtés ne donne aucune information concernant l'intensité des évènements (aléa et dommages). En effet, la base permet seulement d'indiquer que les communes – reconnues par l'état – s'accordent aux critères de la commission interministérielle.

De plus, la profondeur historique de l'information se trouve grandement limitée. Effectivement, cette dernière ne remonte qu'à l'année 1990.

36. Source : données CCR et INSEE

Néanmoins, la profondeur historique est suffisante d'autant plus les observations de l'océan en dessous de 50m (qui contiennent l'essentiel de la mémoire de l'océan) sont rares et leur couverture spatiale et temporelle peu homogène<sup>37</sup> (déplacement des postes, modification des sites de mesure, de l'instrumentation, méthodes de calcul des paramètres) avant la fin des années 1970 (date de lancement des premiers satellites), et avant l'émergence des flotteurs ARGO au début des années 2000. Plusieurs articles avec différents degrés de vulgarisation ont été écrits à ce sujet<sup>38</sup>.

## 2.2.2 Les variables climatiques issues de COPERNICUS

Des requêtes d'extraction provenant de bases historiques et de projections climatiques ont été effectuées à partir des modèles académiques directement liés aux travaux du GIEC appelés « *Global Circulation Models* ». Il est alors important de présenter la provenance des résultats utilisés et l'évaluation de ces modèles.

### 2.2.2.1 Présentation du modèle GCM

Un modèle de circulation global, ou *Global Circulation Model* (**GCM**), est un modèle cherchant à simuler la circulation générale de l'atmosphère planétaire et ou des océans. Le terme de circulation générale correspond aux mouvements atmosphériques et ou océaniques de grande ampleur, avec leurs composantes permanentes et transitoires, à plusieurs échelles.

Pour modéliser ces dynamiques, les **GCM** se basent sur une combinaison d'équations. Les équations de Navier-Stokes sont utilisés pour les dynamiques des fluides newtoniens (gaz et la plupart des liquides) appliquées à une sphère en rotation (représentant la planète). Elles sont associées à des équations d'équilibre de la thermodynamique qui permettent de prendre en compte les sources d'énergie, notamment le rayonnement solaire et les changements d'états liés. Ces processus sont complexes et multidisciplinaires.

Le graphique ci dessous représente les différentes relations modélisées par les équations évoquées précédemment :

37. Benchmarking homogenization algorithms for monthly data : [\[en ligne\]](#)

38. Cassou C. et Mignot J. (2013), Enjeux, méthodes et fondamentaux de prévisibilité et prévision décennale. La Météorologie, 81, 23-31, DOI : 10.4267/2042/51099

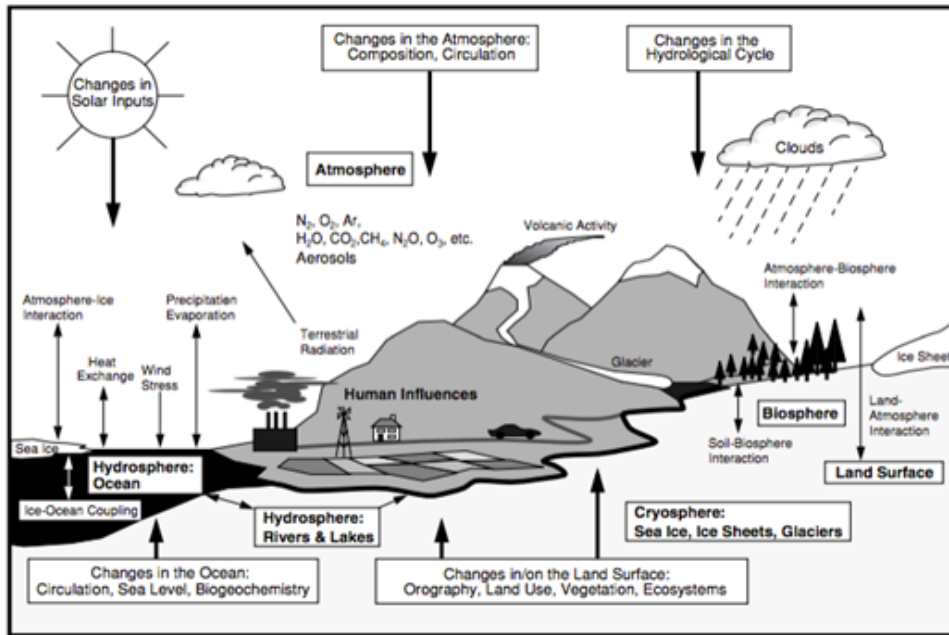
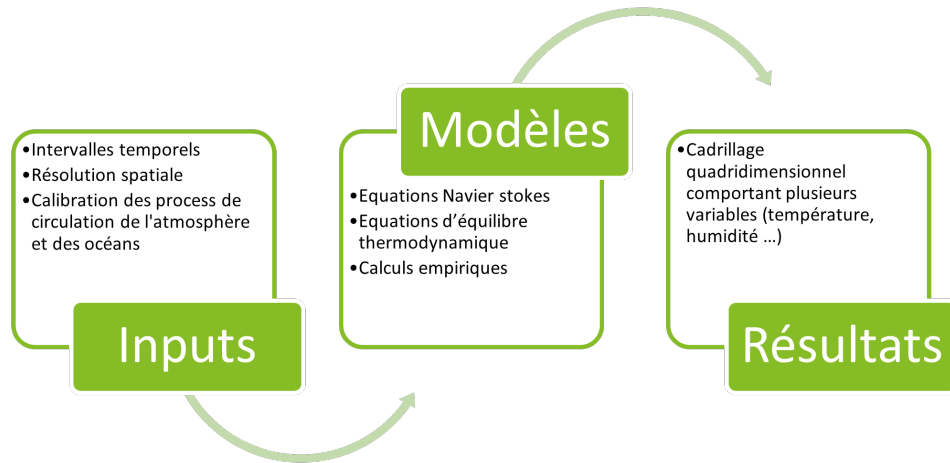


Figure 19 – Représentation des relations modélisées, source : [\[en ligne\]](#)

Afin d'obtenir un modèle climatique global (GCM), il est nécessaire d'associer des modèles de circulation globaux individuels. En effet, certains sont spécifiés sur l'atmosphère, d'autres sur les océans ou la surface terrestre. Les résultats obtenus sont associés à un quadrillage planétaire en trois dimensions. Des grilles sont superposées pour représenter verticalement différents niveaux de l'atmosphère et des océans. La résolution horizontale est relativement faible, les côtés de chaque carré des grilles variant de 100 à 500km. La temporalité des résultats est une quatrième dimension indispensable pour réaliser des simulations climatiques. En fonction des modèles choisis plusieurs types de variables peuvent être obtenues : température, humidité, radiation de surface, précipitations ... Ces modèles ont de nombreuses applications mais sont notamment utilisés pour construire des simulations d'évolution du climat planétaire.

Le fonctionnement d'un GCM peut être résumé de la façon suivante<sup>39</sup> :

39. Source Model Primer : [\[en ligne\]](#)



**Figure 20** – Représentation schématique du fonctionnement d'un GCM

La résolution des quadrillages des GCM est une faiblesse particulièrement notable. En effet, une zone de 500 sur 500 kilomètres peut comporter de nombreux microclimats différents du fait d'une topographie variée. Toutefois, une faible résolution est nécessaire pour pouvoir répliquer un climat planétaire du fait des coûts de calculs importants. Pour pallier ce défaut, des modèles régionaux à la résolution plus importante ont été créés.

En utilisant un processus itératif et différentes hypothèses, il est possible d'utiliser ces modèles pour générer une évolution séquentielle potentielle du climat futur. En utilisant les différents scénarios RCP il est ainsi possible d'estimer des tendances climatiques montrant l'impact de l'évolution des émissions de carbone par les activités humaines.

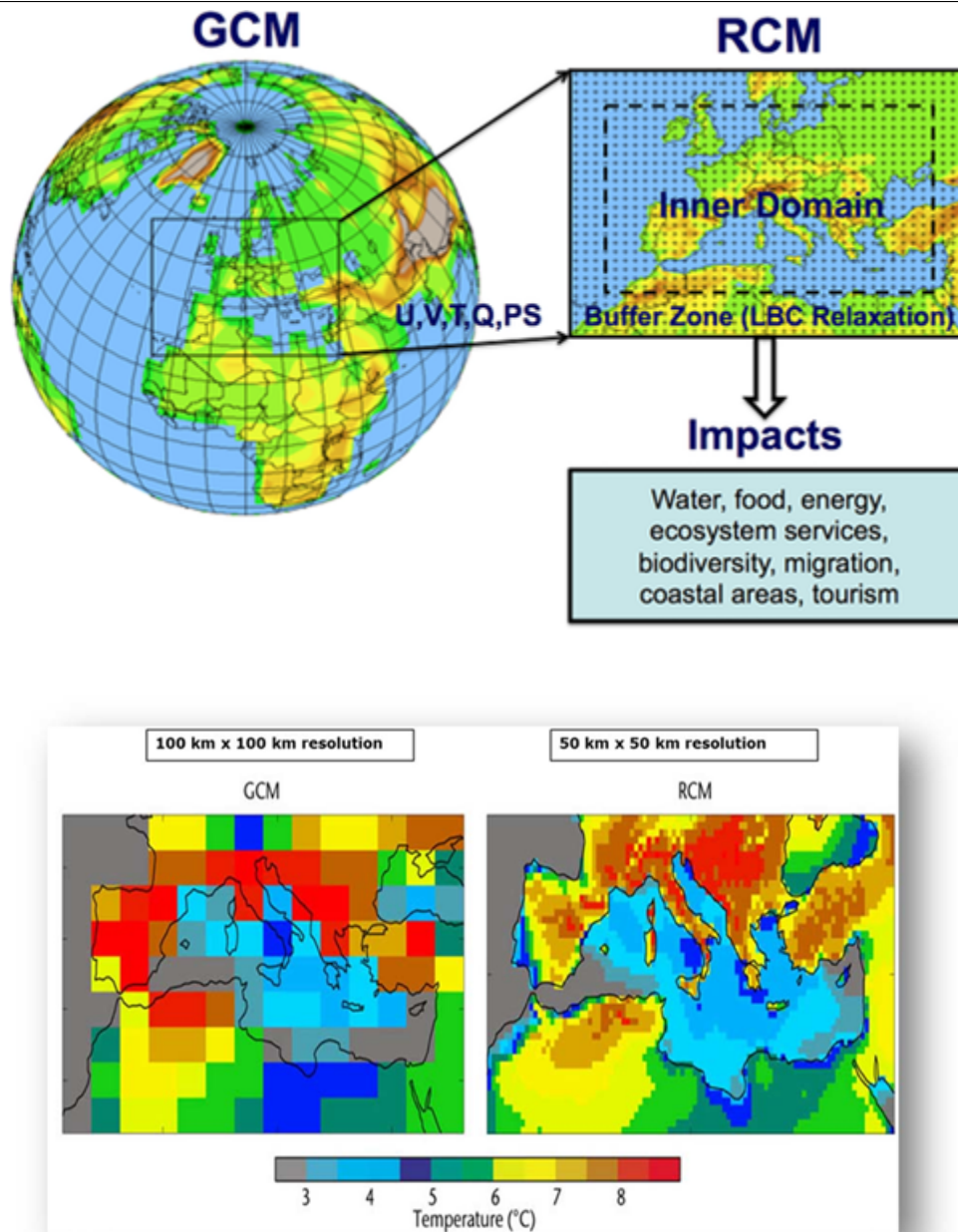
Un exemple de modèle GCM est le GCM ARPEGE produit par le groupe de modélisation CNRM, avec une résolution de 150km. C'est un modèle construit à partir des années 90, fruit de la collaboration en Météo France et le Centre Européen de Prévision. Il fait partie du cadre développé par l'initiative CORDEX.

### 2.2.2.2 Présentation du modèle RCM

Les modèles climatiques régionaux ont été développés pour pouvoir réduire la taille des quadrillages et pouvoir offrir une meilleure granularité de résultats. Cette précision est nécessaire pour l'étude des spécificités climatiques locales et pour estimer avec plus de finesse les vulnérabilités aux changements climatiques. L'échelle des grilles passe ainsi de la centaine de kilomètres à la dizaine de kilomètres.

Les graphiques ci-dessous illustrent le changement d'échelle réalisé<sup>40</sup> :

40. Sources : Graphique 1 [\[en ligne\]](#) Graphique 2 [\[en ligne\]](#)



**Figure 21** – Illustration du changement d'échelle entre **GCM** et **RCM**

Pour qualifier l'utilisation de ces modèles régionaux, on parle de *dynamical downscaling* ou encore descente dynamique. Toutefois, il ne faut pas les voir par opposition aux modèles **GCM** décrits précédemment. Ils sont au contraire fortement complémentaires. Le modèle **GCM** vient décrire les phénomènes de grande ampleur, comme l'effet des gaz à effet de serre ou de variations du rayonnement solaire. Le modèle **RCM** vient quant à lui affiner les résultats du **GCM** en représentant des informations locales plus complexes (topographie, cotes, dynamiques climatiques locales ...). L'association des deux types de modèles permet d'optimiser la granularité voulue tout en gardant sous contrôle le coût en temps de calcul.

Les résultats des modèles **RCM** sont basés sur une réduction d'échelle des modèles **GCM** à partir d'une réduction d'échelle physique. Les résultats des modèles **RCM** ne comportent donc pas une composante aléatoire par rapport au **GCM** de référence. Bien qu'offrant une résolution plus élevée, la résolution des modèles **RCM** reste inférieure à la résolution des modèles CAT actuellement utilisées par les assureurs.

La contrepartie **RCM** du modèle **GCM** AREPEGE évoqué précédemment est le **RCM** ALADIN. Développé à partir des années 2000, il s'agit d'un modèle bi-spectral à aire limitée. Il partage une grande partie du cadre de modélisation du **GCM** mais s'applique à une résolution plus précise. Son intérêt principal est de limiter le coût de calcul des simulations par rapport à des modèles de climat globaux de même résolution. Il fait également parti du programme **CORDEX**.

### 2.2.2.3 Présentation du Programme Mondial de Recherche sur le Climat (PMRC)

Dans le cadre de ce programme, le groupe de travail sur la modélisation couplée (WGCM) a établi le projet de comparaison de modèles couplés (**CMIP**) en 1995. Ces modèles permettent d'étudier les résultats des modèles de circulation générale couplés atmosphère-océan (AOGCM). Ils permettent au climat simulé de s'adapter aux changements de forçage climatique, tels que l'augmentation du dioxyde de carbone atmosphérique (**RCPs**).

Ils fournissent également le diagnostic, la validation, la comparabilité, la documentation et l'accès aux données des modèles climatiques. Ce cadre permet donc à une communauté diversifiée de scientifiques d'analyser les **GCM**. L'effort **CMIP** est financé par le programme de modélisation du climat régional et mondial (RGCM) de la Division des sciences du climat et de l'environnement du ministère américain de l'énergie, programme de recherche biologique et environnementale (BER).

La phase cinq du **CMIP** (**CMIP5**) a fourni une grande partie de données accessibles<sup>41</sup> au cinquième rapport d'évaluation (**AR5**) du **GIEC**. Le **CMIP6** se différencie du **CMIP5** sur divers aspects. Tout d'abord, **CMIP6** compte une centaine de modèles climatiques différents provenant de presque 50 groupes de modélisation différents. Il s'agit ici d'une augmentation significative du nombre de modélisateurs et modélisations par rapport à **CMIP5** qui comptait uniquement une quarantaine de modèles différents provenant d'une vingtaine de groupes de modélisation différents.

En termes de méthodologie, **CMIP6** utilise des résolutions spatiales plus élevées et présente des améliorations en termes de processus physiques.

Ces évolutions résultent en une plus grande sensibilité des résultats. Toutefois, cette augmentation de la sensibilité n'impacte pas de manière significative les résultats et les conclusions établies à partir de **CMIP5**.

Une plus grande amplitude ne signifie pas forcément plus de volatilité. Au contraire une plus grande sensibilité des scénarios permet de renforcer le niveau d'information dans les queues de distribution.

A titre illustratif, les variables suivantes peuvent être utilisées pour représenter certains phénomènes impactant le niveau de l'eau :

- Précipitations extrêmes (R95p) dans le cadre du scénario 8.5
- Vitesse du vent (SfcWind) dans le cadre du scénario 8.5

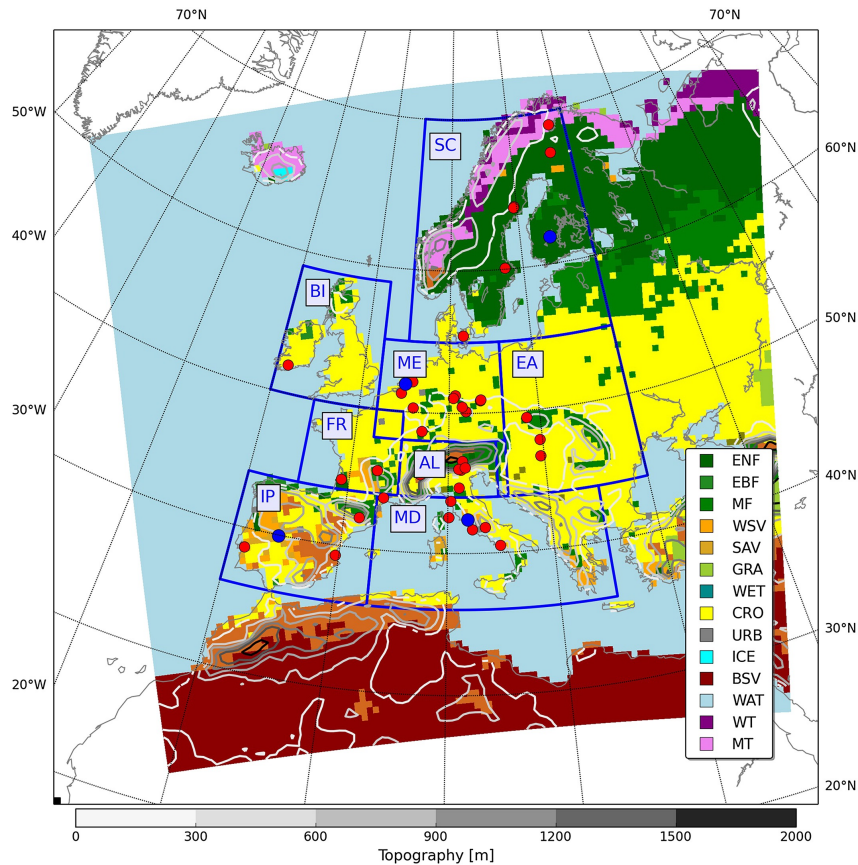
Des représentations graphiques après exploitation de ces données sont présentés en Annexe E.

41. Extraction possible via le lien suivant : [\[en ligne\]](#). Les variables sont également visualisables à travers ce lien : [\[en ligne\]](#)

### 2.2.2.4 Présentation du modèle utilisé au travers de l'initiative CORDEX

L'initiative *Coordinated Regional Climate Downscaling Experiment (CORDEX)* est le premier programme international visant à développer un cadre commun pour simuler le climat historique et futur au niveau régional<sup>42</sup>. Les simulations futures sont basées sur différents scénarios évoqués précédemment dans ce mémoire : **RCP 4.5** et **RCP 8.5**<sup>43</sup>. Des données climatiques sont ainsi générées grâce à un ensemble de modèles **RCM** qui sont eux-mêmes forcés par plusieurs modèles **GCM** provenant de l'initiative **CMIP5**.

Une sous-division de ce programme, **EURO-CORDEX**, est concentrée sur la zone Européenne et le pourtour méditerranéen<sup>44</sup> :



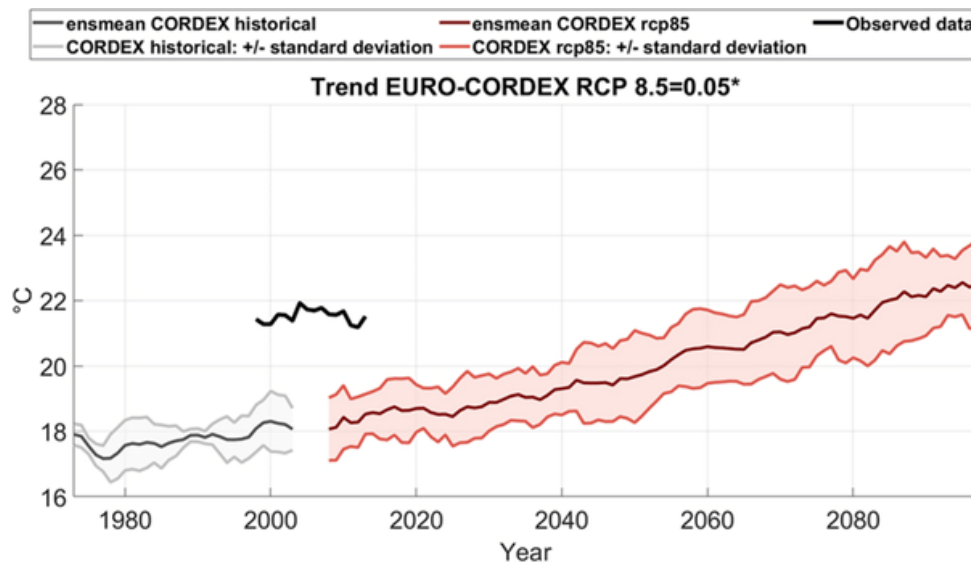
**Figure 22** – Zone modélisée par EURO-CORDEX

42. L'ensemble des données utilisées sont disponibles via le lien suivant : [\[site\]](#)

43. L'ensemble des scénarios **RCP** disponibles sont les suivants : 2.6, 4.5, 6, et 8.5

44. Source : [\[en ligne\]](#)

Des projections sont disponibles sur toute la zone jusqu'en 2100 en fonction de différents RCP et avec une résolution fine de 12.5 kilomètres. Ci-dessous se trouve une prédiction de la tendance des températures maximum en Europe en se basant sur un scénario RCP 8.5<sup>45</sup> :



**Figure 23** – Résultats pour le RCP 8.5

Ce mémoire se fonde en partie sur des variables provenant de l'ensemble de modèles EURO-CORDEX. Si plus d'une vingtaine<sup>46</sup> d'entre elles sont disponibles. A noter qu'étant donné la taille des données, une procédure de téléchargement peut être réalisé uniquement pour 12 variables sur une période de 4 ans quand la maille journalière est choisie. Dans le cadre de la prédiction des submersions marines, deux variables explicatives ont été sélectionnées :

- **Vitesse du vent à 10m (m/s)** : magnitude de la vitesse horizontale deux dimensionnelles du vent. Il s'agit d'une moyenne sur la période agrégée en mètres par seconde.
- **Pression moyenne au niveau de la mer** : pression atmosphérique au niveau de la mer. Dans les régions où la surface de la Terre est au-dessus du niveau de la mer, la pression de surface est utilisée pour calculer la pression atmosphérique qui existerait au niveau de la mer étant donnée une température de l'air constante de la surface à ce niveau. Les données représentent la moyenne sur la période d'agrégation.

La qualité des résultats du programme a été le sujet de plusieurs travaux académiques : Schmidli et al. (2007), Kotlarski et al. (2014), Smiatek et al. (2016) ... Ces derniers confirment une capacité des RCM à représenter les caractéristiques basiques du climat européen. Les biais de températures saisonniers et régionaux restent limités à  $\pm 1.5$  degrés Celsius. Toutefois les biais concernant les précipitations sont plus importants et se situent entre  $\pm 40\%$ . De façon plus générale, la complexité des modèles induit de nombreuses hypothèses qui sont ressenties dans les biais des résultats. Une étude des différents modèles du programme<sup>47</sup> conclut que les températures sont en général mieux simulées que les précipitations. Les résultats restent de qualité suffisante pour réaliser des simulations tendancielle pour l'évolution du climat. Attendre des prévisions précises avec plusieurs décennies d'avance reste illusoire.

45. Source : [\[en ligne\]](#)

46. Une liste des variables est disponible en Annexe A

47. Evaluation of EURO-CORDEX historical simulations by high-quality observational datasets in southern Italy : insights on drought assessment [\[en ligne\]](#)



## 2.2.3 Présentation de la base de données relatives au niveau de mer extrême (ESL)

### 2.2.3.1 L'intérêt de la base (ESL)

Le réchauffement climatique entraîne l'augmentation du niveau de la mer qui est, elle-même liée à celle du risque d'inondation par submersion marine<sup>48</sup>. En effet, d'après la quatrième évaluation nationale du climat<sup>49</sup> : depuis 1900, le niveau moyen mondial de la mer a augmenté d'environ 16 à 21 cm dont près de la moitié de cette hausse depuis 1993 avec le réchauffement des océans.

Ces constats ont amené la *National Oceanic and Atmospheric Administration* (NOAA) à élargir ses résultats potentiels en 2100 par rapport aux études précédentes pour qu'ils se situent entre 0,3 et 2,5 m d'élévation du niveau mondial de la mer. Dans le papier de recherche étudié, des projections probabilistes des niveaux extrêmes de la mer sont faites en tenant compte des changements du niveau moyen de la mer, des marées, des vagues de vent et des ondes de tempête.

Il y est notamment évoqué que la hausse des niveaux extrêmes de la mer est due à la dilatation thermique<sup>50</sup> entraînant une perte de la masse des glaciers et des nappes au Groenland et en Antarctique. A moins que des mesures d'adaptation ne soient prises rapidement, cela implique des niveaux de risque d'inondation par submersion marine sans précédent.

L'intérêt de cet article porte sur la disponibilité des données à travers la mise à disposition du référentiel 'LISCoAsT'<sup>51</sup> (*Large scale Integrated Sea-level and Coastal Assessment Tool*). Ce référentiel d'évaluation est intégré à grande échelle pour le niveau de la mer et du littoral. Il a été développé par le centre commun de recherche de la commission européenne.

L'un de ces objectifs est d'aider à l'analyse des impacts climatiques sur les zones côtières en Europe et dans le monde.

La majorité des études portant sur l'intensification des niveaux de mer extrêmes (ESL) et des submersions marines suppose un climat stationnaire en négligeant en outre les effets des vagues. Dans l'article étudié, des simulations dynamiques ont été combinés pour toutes les composantes du niveau extrême de la mer. Ainsi, ce niveau est fonction de l'élévation du niveau moyen de la mer (MSL), des niveaux d'eau induits par les marées ( $\eta_{tide}$ ), et des conditions météorologiques extrêmes générées par les vagues et les ondes de tempête constituant des extrêmes climatiques ( $\eta_{CE}$ ).

La surcote marine, onde de tempête ou *storm surge* est une élévation temporaire du niveau de la mer par rapport au niveau théorique attendu du a la marée astronomique. Le niveau d'eau total atteint lors d'une tempête est d'autant plus élevé que la surcote et le coefficient de marée sont importants, et que ces deux phénomènes sont concomitants dans le temps. La surcote se décompose en deux parties appelées la surcote atmosphérique<sup>52</sup> et la surcote liée aux vagues<sup>53</sup> (Krien et al., 2010 ; Cariot, 2011).

48. The challenge of unprecedented floods and droughts in risk management [en ligne]

49. US Global Change Research Program [en ligne]

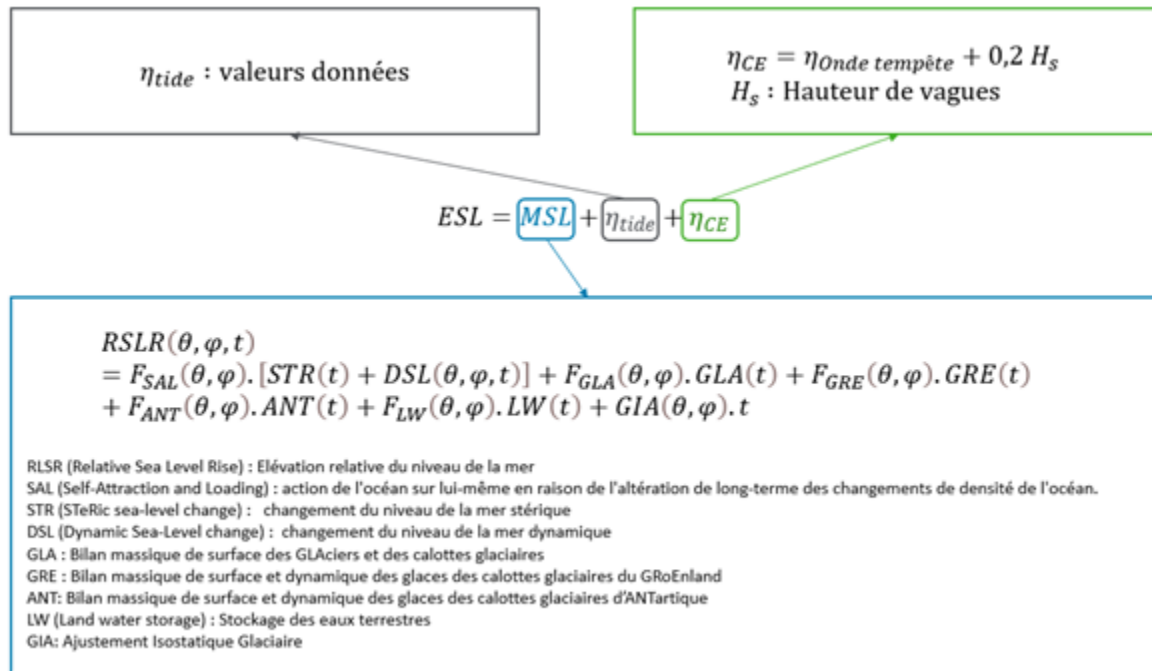
50. Les changements de température font changer la matière de volume. Un réchauffement entraîne un déplacement de plus en plus actif des particules. Elles ont alors besoin d'un espace supplémentaire et cela agrandit la taille du matériau.

51. Global Extreme Sea Level projections [en ligne]

52. Elle est liée à 2 processus. L'un lié au système dépressionnaire entraînant une élévation d'eau par effet de baromètre inverse. L'autre lié à l'effet du vent en exerçant une contrainte à la surface de l'eau.

53. Phénomène due au déferlement de la houle face à la diminution de la profondeur des fonds marins à l'approche des côtes.

Soit, mathématiquement :



## 2.2.3.2 Présentation de la méthodologie adoptée

### 2.2.3.2.1 Présentation générale

La procédure pour générer des projections **ESL** est résumé dans le tableau suivant :

Composantes sous forme de distribution de probabilité	Modèles	Procédures
MSL	Monte Carlo (MC) : échantillonnage aléatoire répété pour chaque variable 5000 fois.	1) Calcul de la valeur relative de l'augmentation du niveau de la mer (RSLR) 2) Projections probabilistes générées par simulations de MC pour déterminer des valeurs en fonction du temps et de la localisation

53. AVISO : [\[en ligne\]](#)

53. Egbert, G. D., and S. Y. Erofeeva (2002), Efficient inverse modeling of barotropic ocean tides, J. Atmos. Oceanic Tech., 19(2), 183–204

Composantes sous forme de distribution de probabilité	Modèles	Procédures
$\eta_{tide}$	FES 2014 <sup>54</sup> et des simulations de marées sous RSLR TOPEX/POSEIDON Global Inverse Solution <sup>55</sup>	<ol style="list-style-type: none"> <li>1) Comme on s'intéresse aux événements extrêmes, on prend le maximum</li> <li>2) Étude de la base de référence pour la période 1980-2014</li> <li>3) Étude du changement relatif 2010-2100.</li> <li>4) Intégration de l'erreur modèle océanique</li> </ol> <p>Les valeurs de l'eau à marée haute sont estimées pour différentes périodes de retour en fonction du temps et de la localisation.</p>
$\eta_{CE}$	ERA-INTERIM pour le vent et la pression atmosphérique historique et CMIP5 GCM pour les projections Vagues : modèle d'onde spectrale de 3ème génération WW3 <sup>56</sup> Onde de tempête : DFLOW FM <sup>57</sup> et Delft3D-FLOW	<ol style="list-style-type: none"> <li>1) Réanalyse de la base historique afin d'obtenir la contribution des vagues / Modélisation océan</li> <li>2) Intégration/correction des effets des cyclones tropicaux (TC) en fonction des données altimétriques satellitaires.</li> <li>3) Analyse des valeurs extrêmes (EVA) sur les séries temporelles (Baseline : 1980 – 2014)</li> <li>4) Projections des variables</li> <li>5) EVA pour obtenir les valeurs futures et comparaison historique vs le futur pour avoir les changements relatifs (ST : 1980 – 2100)</li> </ol> <p>Les valeurs <math>\eta_{CE}</math> sont estimées pour différentes périodes de retour en fonction du temps et de la localisation.</p>

**Table 7** – Résumé des différents modèles et procédures employés dans le cadre des projections ESL

### 2.2.3.2.2 Focus sur l'analyse des valeurs extrêmes

Lorsqu'un événement n'a jamais ou très rarement été observé, comme c'est le cas pour la submersion marine, les méthodes statistiques usuelles ne peuvent pas être utilisées. C'est ce cadre théorique qui est donné par la Théorie des Valeurs Extrêmes (TVE) et qui est confirmé par la pratique. La TVE est une branche de la statistique qui se consacre à la modélisation de l'occurrence des événements rares. En effet, le besoin de construire des événements intenses et extrêmes est justifié par le fait de travailler sur des valeurs

57. Hydrodynamic simulations : Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Feyen, L. Extreme sea levels on the rise along Europe's coasts. *Earth's Future* 5, 304–323 (2017).

57. Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H. Ward, P. J. A global reanalysis of storm surges and extreme sea levels. *Nat. Commun.* 7, 11969 (2016) ou agers, B., Rego J. L.; Verlaan, M.; Lalic, A.; Genseberger, M.; Friocourt, Y.; van der Pijl, S. (2014), A global tide and storm surge model with a parallel unstructured-grid shallow water solver, in American Geophysical Union, Fall Meeting 2014, San Francisco, Calif.

maximales plutôt que moyennes. En d'autres termes, une approche pour quantifier le risque climatique est de s'intéresser à la loi de probabilité du maximum annuel des variables climatiques. La loi de ces extrêmes peut être aisément obtenu si on dispose d'un historique assez large.

Les domaines d'application de la **TVE** sont nombreux : la finance, l'actuariat (Solvabilité 2, Exigences de Fonds propres avec le SCR) ou encore l'hydrologie. En effet, les études océanographiques et d'ingénierie côtière s'appuient souvent sur la **TVE** pour décrire la fréquence des vagues extrêmes, les événements de niveau d'eau, les impacts des inondations, et pour comprendre les effets de l'augmentation du niveau de la mer. Plus ce niveau augmente, plus la probabilité qu'une submersion marine arrive augmente. De manière équivalente, la période de retour ou l'intervalle de récurrence de ces inondations diminue<sup>58</sup>. Les données utilisées sont issues d'une analyse faite à partir de la **TVE** mais il pourrait être intéressant de voir l'impact d'une utilisation d'autres lois usuelles en hydrologie<sup>59</sup>.

### 2.2.3.2.3 Ajustement et période de retour

Pour chacune des stations étudiées et pour chaque mois de l'année, des lois peuvent être ajustées par différentes méthodes. Dans l'idéal il faudrait ajuster les lois de probabilité sur l'ensemble des stations. En revanche, pour des raisons de significativité, il est possible de faire le choix de ne traiter que de stations disposant d'un historique suffisant.

Parmi les méthodes d'estimation des quantiles fondées sur les séries maximales annuelles, deux sont communément utilisées (Ferrer, 1992) :

- La méthode du dépassement de seuil (POT) [GPD] : Cette méthode alternative consiste en garder toutes les observations qui dépassent un niveau élevé puis à ajuster une loi appropriée à ces dépassements représentés par les événements extrêmes.
- La méthode des maxima par bloc (**AM**, Annual Maxima) [GEV] : Cette méthode traite du comportement stochastique des valeurs extrêmes dans un processus. Des blocs de taille identique dont seulement le maximum de chacun d'eux est utilisé pour ajuster une loi **GEV**. Pour les données météorologiques, des blocs de taille un an sont utilisés afin de gommer les effets saisonniers.

L'article de recherche étudié utilise la seconde méthode en utilisant les distributions de valeurs extrêmes généralisées (**GEV**) dont la fonction de distribution est la suivante :

$$F(x; \mu, \sigma, k) = \begin{cases} \exp[-(1 + k \times (\frac{x-\mu}{\sigma}))^{-1/k}] & \text{pour } k \neq 0 \\ \exp[-\exp-(\frac{x-\mu}{\sigma})] & \text{pour } k = 0; \end{cases}$$

Où  $F$  est la probabilité que le niveau d'eau  $x$  ne soit pas dépassé quelque soit la période de retour d'un an.  $\mu, \sigma, k$  sont respectivement les paramètres de localisation, d'échelle et de forme.

Elle comprend également des cas spécifiques avec des distributions de Gumbel, Frechet ou encore Weibul<sup>60</sup>.

58. Vitousek, S., Barnard, P., Fletcher, C. et al. Doubling of coastal flooding frequency within decades due to sea-level rise. *Sci Rep* 7, 1399 (2017). [\[en ligne\]](#)

59. Pour plus d'informations, se référer à l'annexe C sur les lois usuelles en hydrologie.

60. Papier de Samantaray et al, pouvant être utilisé comme un guide pour l'estimation de modèles **GEV**.

En utilisant la **TVE**, la borne supérieure n'est pas identifiée et doit donc être projetée. La fréquence espérée est définie par l'espérance :

$$E(X) = 1 - F$$

En ingénierie côtière, le niveau de retour des aléas maritimes est incontournable. Il est possible de définir cette période de retour de la manière suivante :

$$Tr = 1/E(X)$$

Par exemple, si l'on considère une hauteur de vague et que le résultat vaut 100, cela signifie que cette hauteur est dépassée en moyenne une fois tous les 100 ans.

#### **2.2.3.2.4 Comparaison des différentes méthodes d'estimations de quantiles élevés**

Il existe 3 méthodes principales :

- Méthode des moments (MOM) dans l'espace réel ou logarithmique (LMOM) : Elle n'utilise pas toutes les données de l'échantillon de manière exhaustive.
- Maximum de vraisemblance (ML) : C'est souvent l'approche la plus efficace (moins de variance dans les paramètres estimés), mais elle est très sensible au choix du modèle de distribution. Pour la loi **GEV** il suffit de rappeler la fonction caractéristique, la densité et la fonction de vraisemblance<sup>61</sup>.
- Moments de probabilité pondérée (PWM) : Donne plus de poids aux valeurs les plus élevées de la série, ce qui a pour effet de produire des valeurs plus prudentes.

Dans l'étude choisie, c'est la méthode du maximum de vraisemblance qui est choisie. Pour tenir compte de l'incertitude des estimations des paramètres **GEV**, une simulation de Monte-Carlo avec 100 000 réalisations a été appliquée pour chaque point de grille. Chaque réalisation génère des valeurs aléatoires de  $\mu$ ,  $\sigma$  et  $k$  basées sur les intervalles de confiance à 95% découlant des estimations du maximum de vraisemblance.

Il est possible dans d'autres cas de se référer aux travaux de Tiogum et Delcaillau. Ils permettent d'estimer le niveau de retour à 100 ans de données environnementales. En outre, cela permet de savoir quelles méthodes et estimateurs utiliser en fonction de la taille du jeu de données et de ses caractéristiques.

Finalement, les projections probabilistes mondiales de l'**ESL** sont obtenues à partir de simulations de Monte-Carlo sur les trois composantes d'**ESL** ( $MSL$ ,  $\eta_{tide}$ ,  $\eta_{CE}$ ).

### **2.2.3.3 Traitement du maillage temporel et géographique**

#### **2.2.3.3.1 Maillage temporel**

**Présentation des informations présentes dans l'article de recherche :**

En ce qui concerne la temporalité des bases de données il en existe une par scénario :

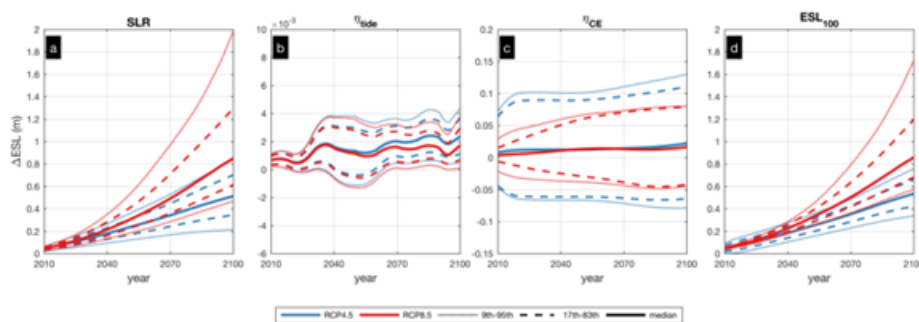
---

61. Éléments présents en annexe D

Périodes	Scenarii
01/12/1969 – 30/11/2004 (35 ans)	Historique de 35 ans
01/12/2009 - 30/11/2099 (100 ans)	RCP 4.5
01/12/2009 - 30/11/2099 (100 ans)	RCP 8.5

**Table 8** – Récapitulatif des différentes périodes étudiées par scénarios pour la variable **ESL**

En effet, dans le cadre de la temporalité étudiée, les données utilisées vont jusqu'à 2050. D'après l'article étudié, entre 2000 et 2100, la moyenne mondiale d'(ESL sur 100 ans augmente de 34 à 76 cm et de 58 à 172cm selon respectivement les scénarios RCP 4.5 et 8.5. D'ici 2050, l'article évoque également une augmentation de la moyenne mondiale d'ESL100 de 14-34 cm et 24-41 cm sous RCP 4.5 et RCP 8.5, respectivement :



**Figure 24** – Changements prévus dans l'ESL mondial

On remarque que le scénario RCP le plus pessimiste intègre des incertitudes plus importantes concernant les nappes glaciaires de l'Antarctique et du Groenland.

### Retraitement spécifique au scénario historique :

Après étude de la base de données historique<sup>62</sup>, il a été remarqué que la variable de temporalité n'était pas présente. Afin de remédier à ce problème une autre base de données<sup>63</sup> a été utilisée afin de la reconstituer.

Cette base de données s'intéresse au niveau côtier de la mer en Europe. Elle permet une meilleure compréhension hydrodynamique des potentiels futurs pour ces zones clés dans le cadre du changement climatique.

Ces indicateurs sont produits grâce à une combinaison de modèles. Le modèle *Deltares Global Tide and Surge Model* (GTSM) sert de base. Il repose sur des équations d'eau peu profonde intégrées verticalement ou barotropes. C'est un modèle hydrodynamique moyenné par la profondeur, avec une couverture mondiale, qui simule dynamiquement les marées et les ondes de tempête. La calibration commence par les eaux profondes océaniques puis simule la dissipation de l'énergie vers les eaux moins profondes et plus proches des côtes. Grâce à une gestion dynamique de la résolution, décroissante de l'éloignement des côtes, il permet d'obtenir une précision importante pour un nombre de calculs relativement réduit. En effet, maintenir une résolution aussi fine pour les cotes que pour les océans n'est ni nécessaire ni faisable en termes de calculs. La résolution passe ainsi de 25km à 1.25km pour les côtes européennes.

Il est utilisé avec un forçage climatique régional généré par le modèle HIRHAM5. Ce dernier est une version locale du modèle climatique plus global *EC-EARTH*, qui est conservé pour avoir des résultats

62. Extreme Sea level - Historical [\[en ligne\]](#)

63. Site des bases : [\[en ligne\]](#) ou [\[en ligne\]](#)

cohérents aux frontières des modèles. Ce maillage plus précis permet d'obtenir des résultats plus consistants et exacts localement. Ces modèles correspondent aux RCM et GCM décrit précédemment dans ce mémoire. Trois scénarios différents sont donnés aux modèles :

- Un scénario historique représentant le climat historique de 1977 à 2005
- Le scénario RCP 8.5, beaucoup plus pessimiste, allant de 2040 à 2070
- Et le scénario RCP 4.5, supposant une diminution des émissions, à partir de 2070

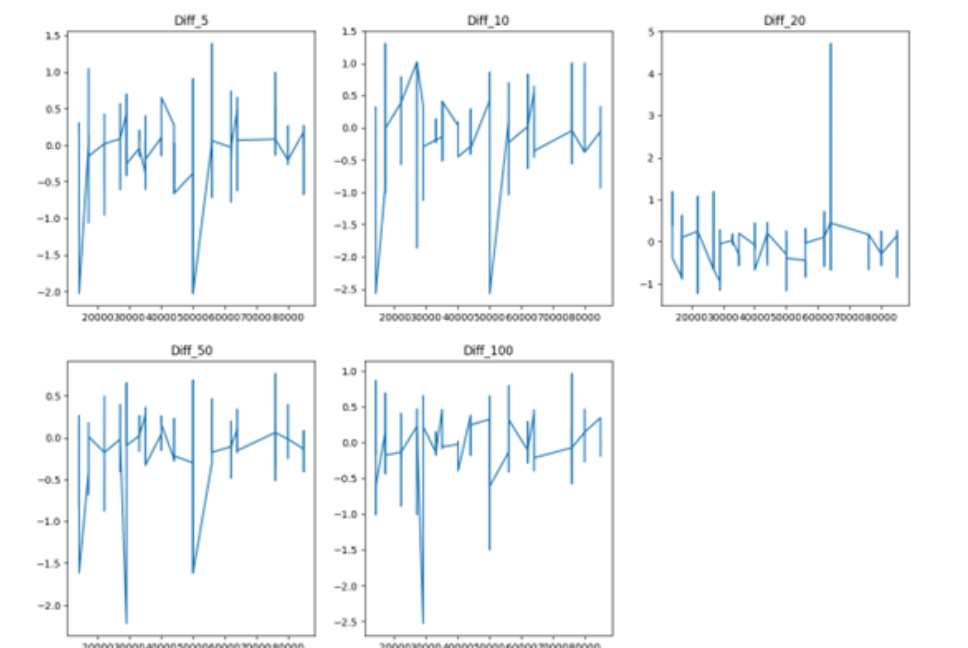
A eux trois, ils couvrent les évolutions jusqu'en 2100, forçant des valeurs entre 2.6 et 8.5 W/m<sup>2</sup>. Les différents scénarios ne sont pas associés à une probabilité d'occurrence. Le scénario RCP 4.5 est appliqué post 2070 et reflète l'hypothèse que l'augmentation des événements climatiques extrêmes poussera la société à s'adapter et à réduire son impact. Les simulations de possibles climats futurs commencent à partir de 2040 car l'impact sur le niveau de la mer, en particulier les niveaux extrêmes, est détectable après plusieurs décennies.

La base de données contient de nombreuses variables<sup>64</sup> décrivant sur quatre dimensions les caractéristiques des marées et des surcotes.

Les variables sélectionnées pour pouvoir compléter la base historique sont celles constituant le niveau d'eau extrême (EEWL) à savoir le niveau d'eau moyen, la surcote et la marée. Il est ensuite possible à partir d'EEWL de retrouver ESL car la variable est bien présente dans la base du papier de recherche.

Un test est ensuite effectué sur la base de données historique alors constituée afin de vérifier que la valeur présente dans la base de données initiale est bien présente dans cette nouvelle base de données historique.

Les graphiques ci-dessous montrent la variation en pourcentage pour chaque période de retour des cantons :



**Figure 25** – Variation en pourcentage des valeurs historiques (ordonnée) par canton (abscisse) en fonction des périodes de retour

64. Une présentation des variables est présente en Annexe B

Pour les bases de données RCP 4.5<sup>65</sup> et RCP 8.5<sup>66</sup> une variable de temporalité était bien présente (« decades »). Il a donc été possible de faire une interpolation linéaire du fait de la faible variabilité des données.

On obtient alors une valeur par canton et par période de retour pour chaque scénario.

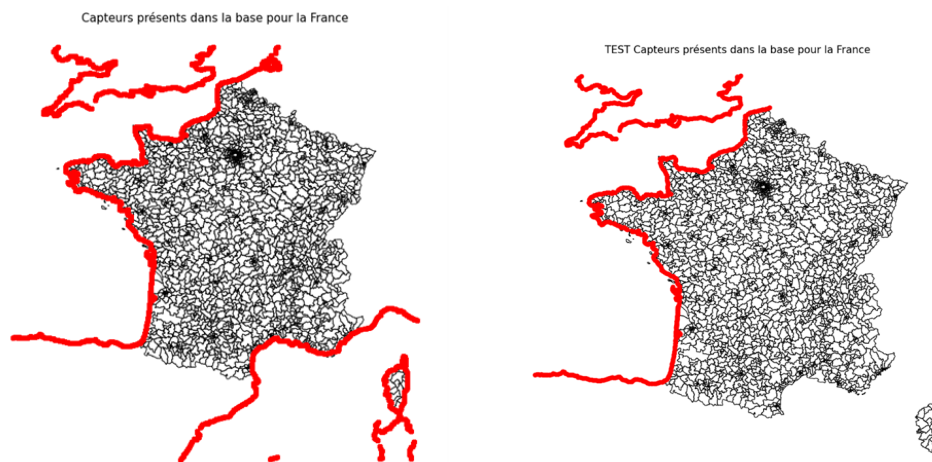
### 2.2.3.3.2 Maillage géographique

**À partir des bases issues d'ESL à l'exception du scénario historique :**

Dans le cadre du mémoire présenté, nous souhaitons étudier le littoral atlantique français. Un découpage géographique a donc dû avoir lieu pour récupérer ce maillage. Une répartition géographique des données ESL est disponible en annexe G. L'intérêt est porté sur la zone J (Europe du nord). Une autre étude traitant directement de la zone géographique européenne (annexe H) est également disponible. Les données ont pu être téléchargé sous format NETCDF<sup>67</sup>. La première étape consiste à les retransformer pour avoir un format csv.

Puis, un découpage doit avoir lieu afin de pouvoir traiter de la France. C'est comme si le monde était divisé à travers une maille. Dans chaque carré des données sont présentes. Pour obtenir la France, il faut découper un carré au sein de l'Europe avec une maille géographique 0.11.

La figure ci-après permet de montrer l'ensemble des capteurs disponibles après le découpage territorial<sup>68</sup> :



**Figure 26** – Localisation des capteurs côtiers

Nous retirons les capteurs présents sur la Méditerranée et la Corse à partir des coordonnées géographique (longitude, latitude). En revanche, cette méthode ne peut pas s'appliquer à certains cantons du territoire britannique. Une piste d'amélioration serait de pouvoir effectuer un retraitement plus fin.

Les capteurs présents dans la base ne sont pas localisés sur le territoire français mais en mer compte tenu de la nature hydrologique des variables étudiées. Il est nécessaire d'affecter à chacun de ces capteurs le canton du littoral le plus proche.

65. Extreme Sea level - RCP 4.5 [\[en ligne\]](#)

66. Extreme Sea level - RCP8.5 [\[en ligne\]](#)

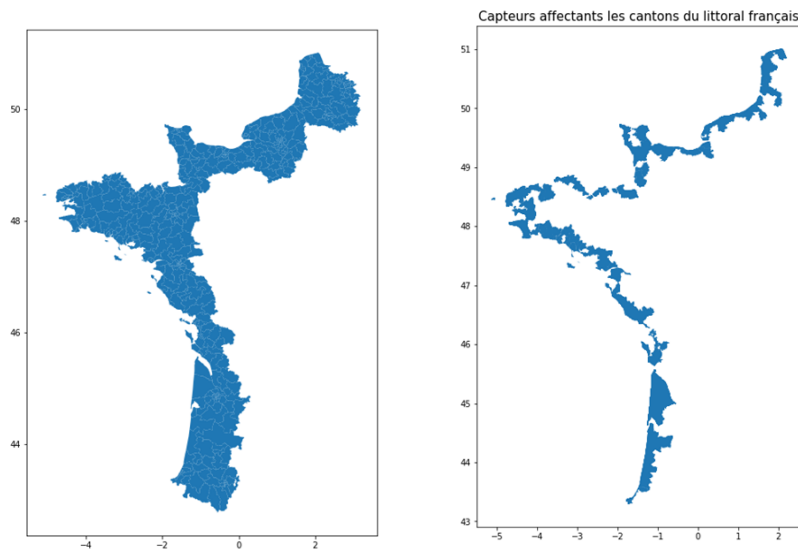
67. Type de fichier étant une matrice qui représente le monde découpé en grille.

68. Source : données des capteurs issus d'ESL



Pour ce faire, l'utilisation de la base cantons du littoral<sup>69</sup> est faite à partir des données INSEE, chaque capteur se voit alors affecter au canton le plus proche. Pour ceux ne disposant pas de mesure, une utilisation des k plus proches voisins a été faite.

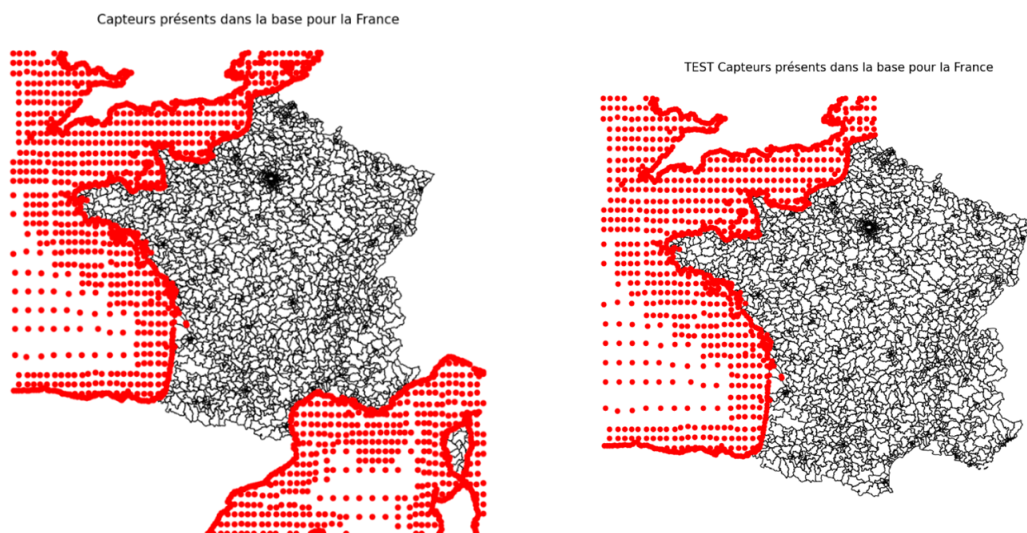
Les graphiques ci-dessous présentent le retraitement qui a été fait afin de passer de l'ensemble des cantons présents pour les départements du littoral (partie gauche) aux cantons du littoral de première ligne (partie droite).



**Figure 27** – Représentation des cantons du littoral

### Traitement de la base historique COPERNICUS :

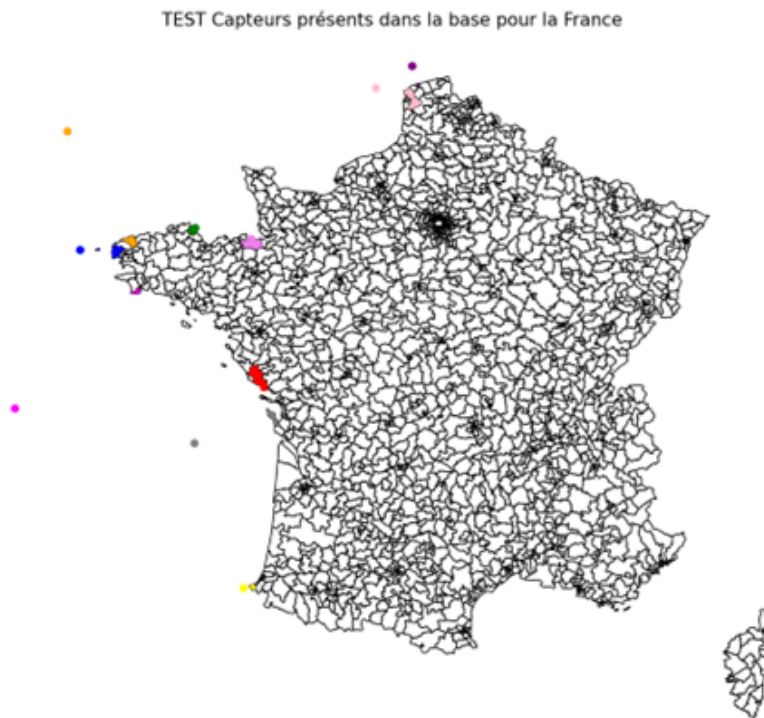
Les figures suivantes permettent de représenter les capteurs issus de la base CORPERNICUS.



**Figure 28** – Représentation de la localisation des capteurs

69. Les cantons ont été déterminés à partir des départements du littoral : 14, 17, 22, 27, 29, 33, 35, 40, 44, 50, 56, 62, 64, 76, 80 et 85.

Afin de vérifier l'affectation des capteurs aux cantons, nous choisissons 10 capteurs aléatoirement en les représentant avec leurs cantons associés :



**Figure 29** – Localisation des 10 capteurs sélectionnés aléatoirement

On remarque que l'affectation s'est bien effectuée.

### 2.2.4 Limites

Les incertitudes actuelles de l'**ESL** sont liées à la compétence prédictive des modèles océaniques, ainsi qu'aux erreurs d'ajustement lors de l'analyse des valeurs extrêmes de la série chronologique  $\mu_{CE}$ .

Toutes les variables **ESL** sont traités comme indépendantes. Or, une modélisation où les variables sont corrélées seraient intéressante mais cela va au-delà des capacités actuelles de modélisation et de calcul. En revanche, des études antérieures ont démontré la validité de l'hypothèse pour les projections relatives aux changements climatiques<sup>70</sup>, car l'erreur résultante est compensée par d'autres sources d'incertitude.

Dans l'approche proposée, nous ignorons les anomalies du niveau moyen de la mer (MSLA) dues aux effets saisonniers et aux cycles climatiques.

L'échelle temporelle et spatiale choisie s'accompagne de certains compromis inévitables en termes de méthodologie, de résolution spatiale et, par conséquent, de précision, notamment par rapport aux études locales et aux prévisions opérationnelles.

Pour finir, il est important de rappeler que certaines variables comme la topographie ou encore l'absorption des sols n'ont pas été modélisées dans l'approche choisie. En revanche, il serait intéressant

<sup>70</sup>. Perez, J., Menendez, M., Camus, P., Mendez, F. J. Losada, I. J. Statistical multi-model climate projections of surface ocean waves in Europe. *Ocean Model.* 96, 161–170 (2015).

d'essayer de pouvoir les prendre en compte dans un exercice futur. Ces éléments permettent d'accentuer l'importance de prendre en compte la distance à la mer au travers de la topographie et de l'absorption des sols.

## 2.3 L'agrégation des données

À la suite de la collecte de données d'aléa fiables et le nettoyage des bases, il est nécessaire de définir et de tester des méthodes d'agrégation des données.

Le tableau suivant permet de récapituler l'ensemble des bases de données disponibles pour la modélisation :

BDD	Temporalité	Variables
<b>CCR</b>	Historique (1990-2005)	Code canton, Date, Numéros d'arrêtés, Classification de l'arrêté (2 si l'arrêté est reconnu, 1 s'il ne l'est pas)
COPERNICUS	Historique (1990-2005), <b>RCP 4.5</b> , <b>RCP 8.5</b>	Pression atmosphérique (psl), Vent (sfc wind)
<b>ESL</b>	Historique, <b>RCP 4.5</b> , <b>RCP 8.5</b>	<b>ESL</b> par période de retour : 5,10,20,50 et 100, EEWL par période de retour : 5,10,20,50 et 100

**Table 9** – Différentes bases de données disponibles pour la modélisation

Les données climatiques utilisées proviennent de différentes sources et nécessitent d'avoir la même granularité pour l'intégration dans les modèles. En effet, ces données doivent respecter les deux mailles choisis pour la modélisation à savoir :

- La maille géographique cantonale
- La maille temporelle journalière

### 2.3.1 La maille géographique cantonale

Afin d'établir la base des cantons du littoral, il est nécessaire de sélectionner les bons cantons. Une première possibilité aurait été de considérer uniquement les cantons sinistrés de la base des arrêtés CCR. En revanche, cette méthode résulterait en une restriction de la base. L'alternative envisagée a donc été de prendre la maille supérieure afin de déterminer les cantons à choisir. En d'autres termes, l'ensemble des cantons des départements contenant des cantons sinistrés ont été choisis.

Le modèle étudié repose donc sur la première ligne des départements du littoral Atlantique pour constituer l'ensemble des cantons du littoral. Ces départements du littoral appartiennent à cinq régions du littoral<sup>71</sup> :

71. voir l'annexe P pour la superposition de la carte de prédiction des cantons affectés par la submersion marine à la carte de la France.

Régions	Code INSEE	Départements
La Bretagne	53	Côtes-d'Armor, Côtes-du-Nord, Finistère, Ille-et-Villaine et Morbihan
Les Hauts-de-France	32	Nord, Pas-de-Calais et Somme
La Normandie	28	Calvados, Eure, Manche et Seine-Maritime
La Nouvelle Aquitaine	75	Charente-Maritime, Gironde, Landes et Pyrénées-Atlantiques
Le Pays de la Loire	52	Loire-Atlantique et Vendée

### 2.3.2 La maille temporelle journalière

Les événements de type submersion marine sont traités à la maille journalière du fait de leur durée très courte. Cette affirmation a pu être vérifiée à l'aide d'une étude préalable sur la durée d'une submersion marine provenant de la base des arrêtés Cat-Nat. Le graphique suivant permet de montrer la durée moyenne en jours de ces événements par année :

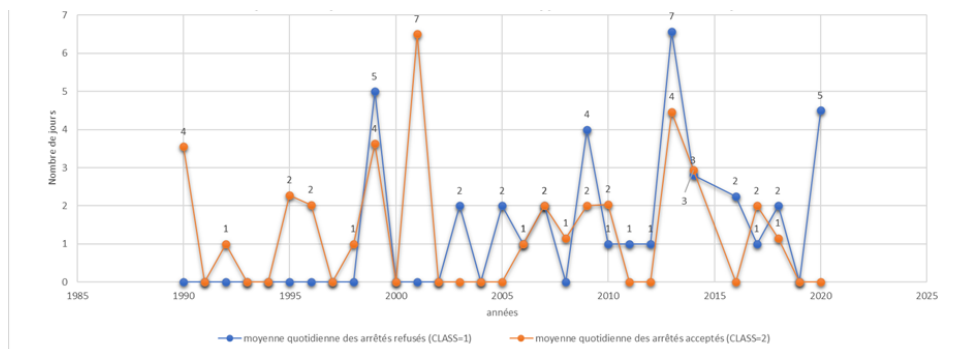


Figure 30 – Durée moyenne en jours des événements de type submersion marine par année

D'après les éléments de la base historique, une submersion marine dure au maximum 7 jours. Cet événement est donc de très courte durée avec une moyenne de 2 jours. En revanche, ce constat est applicable uniquement sur la base historique. Une aggravation des conditions climatiques laisse présager une augmentation de la durée des submersions marines. Il convient alors de prendre à minima, la durée maximale enregistrée sur la base historique afin de pouvoir comptabiliser ces événements extrêmes.

### 2.3.3 Création des bases d'apprentissage et de validation

Elles doivent également pouvoir être agrégées en fonction des temporalités. La base CCR a du être séparée en deux bases pour respecter l'agrégation des données avec COPERNICUS qui a un historique plus restreint. Ainsi, la base de données COPERNICUS peut être décomposée de la façon suivante :

- Historique (1990-2005)
- Historique pour validation (2006-2020)

## 2.4 Présentation et sélection des modèles

Une fois la récupération des variables effectuée, grâce à des programmes intégralement développés en langage Python, il est alors nécessaire de les traiter afin de pouvoir construire un modèle approprié.

L'objectif recherché est la construction d'un modèle de projection de la survenance de submersions marines, à horizon 2050, dans la façade atlantique française face au changement climatique. En d'autres termes, la variable à expliquer est constituée des demandes d'arrêtés Cat-Nat en fonction de variables explicatives climatiques. Pour répondre à cette problématique, une fonction de lien devait être définie. La modélisation par *machine learning* s'est alors avérée appropriée. En effet, le *machine learning* est une technique moderne et performante permettant d'apprendre et de prédire des relations entre des données d'entrées et des données de sorties. Le lien entre ces données n'est pas explicitement défini et est alors implicite. À noter qu'afin de pouvoir utiliser l'apprentissage automatique, les classes de la variable explicative ont été transformées en données catégorielles numériques.

### 2.4.1 Rappels généraux sur le *machine learning*

L'objectif général de l'apprentissage statistique (*machine learning*) est d'élaborer des procédures automatiques qui permettent de mettre en évidence des règles générales à partir d'exemples. Il s'agit donc d'imiter le fonctionnement inductif du cerveau humain dans le but de développer des systèmes d'intelligence artificielle.

Le *machine learning* ou apprentissage statistique est associé à des algorithmes regroupant un panel de modèles, généralement sans distribution, qui peuvent être utilisés pour extraire des informations de données ou aider à la prévision.

L'absence d'hypothèse de distribution dans l'apprentissage automatique implique d'avoir suffisamment de données pour capturer la loi du phénomène. C'est souvent la raison pour laquelle l'apprentissage automatique est associé au "*Big Data*".

#### 2.4.1.1 Un modèle par classification

En intelligence artificielle et en apprentissage automatique, la classification fait référence à la capacité de la machine à attribuer les instances à leurs bons groupes. Pour que la machine puisse décider comment attribuer une instance à son groupe, elle doit apprendre les modèles de cette affectation à partir des fonctionnalités d'apprentissage disponibles dans un ensemble de données d'apprentissage étiqueté.

Dans le cas de la classification, le phénomène à prédire est discret, contrairement à la régression où le phénomène est continu lorsque l'apprentissage est supervisé. En d'autres termes, on parlera de régression lorsque la variable prédictive est réelle et peut prendre un nombre infini de valeurs et de classification lorsque cette dernière est catégorielle et peut prendre un nombre fini de valeurs. Nous avons deux types de classification : la classification binaire et la classification multi-classe. La première, permet à la machine de classer une instance selon deux classes possibles ; oui/non, 1/0 ou vrai/faux.

Dans le second, la machine doit classer une instance comme l'une des trois classes ou plus.

Dans le cas étudié, il s'agit de ce type de classification puisque la variable à expliquer (Arrêtés Cat-Nat) peut prendre 3 modalités :

- 0 quand il n'y a pas eu d'événements de type submersion marine.
- 1 quand il y a eu un tel événement mais il n'a pas été reconnu comme un état de catastrophe naturelle.

— - 2 quand un tel événement a eu lieu et a été reconnu.

### 2.4.1.2 Le choix de l'approche par algorithme supervisé

Il existe deux familles principales d'algorithmes, supervisé ou non supervisé :

- **Supervisé** : Algorithme dont l'ajustement requiert une base de données labellisée. Pour chaque individu  $X_i$  observé, un label  $Y_i$  doit être disponible. L'objectif est d'expliquer. Il s'agit par exemple des méthodes *CART*, *SVM*, *Neural Networks*, *GLM*, *Knn*.
- **Non-supervisé** : Algorithme qui apprend le plus souvent en observant des individus isolés  $X_i$ . Il permet entre autres de définir des classes d'individus similaires. L'objectif est donc de segmenter ou classifier. Par exemple, *CAH*, *K-means*.

Dans le cas étudié, l'objectif est de prédire et relève donc naturellement des algorithmes supervisés.

L'ensemble des algorithmes proviennent d'un programme d'optimisation plus ou moins difficile à résoudre selon la nature des données.

Soit une BDD composée de  $n$  couples  $Z_i = (X_i, Y_i)$ , que nous supposons être des réalisations indépendantes d'une même loi  $P$  inconnue. Alors  $Z_i = (X_i, Y_i) \sim^{iid} P$ .

Les  $X_i$  appartiennent à un espace  $\mathcal{X}$  et s'appellent les entrées ou les *features*. En général,  $\mathcal{X} = \mathbb{R}^d$  pour un grand entier  $d$ .

Les  $Y_i$  appartiennent à un espace  $\mathcal{Y}$ , et s'appellent les sorties ou les étiquettes. En général,  $\mathcal{Y}$  est fini ou est un sous-ensemble de  $\mathbb{R}$ .

Le but de l'apprentissage supervisé est de prévoir l'étiquette  $Y$  associée à toute nouvelle entrée  $X$ . Cette réalisation étant indépendante des réalisations précédemment observées.

Une fonction de prédiction est alors une fonction (mesurable) de  $\mathcal{X}$  dans  $\mathcal{Y}$ .

L'ensemble de toutes les fonctions de prédiction est noté  $\mathcal{F}(\mathcal{X}, \mathcal{Y})$ .

L'ensemble d'apprentissage est composée de la BDD  $Z_1, \dots, Z_n$

Un algorithme d'apprentissage est une fonction qui a tout ensemble d'apprentissage renvoie une fonction de prédiction.

## 2.4.2 Traitements préalables face à des données déséquilibrées et mise en place du modèle de classification

### 2.4.2.1 Initialisation de la BDD

#### 2.4.2.1.1 Le caractère déséquilibré des données

Avant de pouvoir mettre en place un modèle de classification, il est souvent nécessaire d'effectuer des traitements préalables. Du fait du phénomène extrême et très rare étudié, la base de données construite

comporte des données fortement déséquilibrées. En effet, la classe 0 (Pas d'événements de type submersion marine) représente 90% de notre BDD compte tenu du maillage choisi (maille journalière et cantonale).

Toute comme la standardisation des données est courante en *machine learning*, ici une stratégie d'échantillonnage ou *sampling* a du avoir lieu pour réduire le nombre de classes 0 dans la base de données.

En outre, la standardisation des données n'a pas été effectuée ici du fait du caractère extrême des données. En revanche, il est possible dans les problèmes où l'on souhaite accorder plus d'importance à certaines classes d'utiliser la fonction *RandomUnderSampling* et choisir une stratégie d'échantillonnage.

Ainsi, deux stratégies ont été testées dans la méthode d'ajustement. La première « *majority* » rééchantillonne uniquement la classe majoritaire et sera présentée sous l'appellation « Cas 2 ». La seconde « *all* », rééchantillonne toutes les classes et sera présentée sous le « Cas 3 ». Le premier cas, « cas 1 » représente les résultats sans stratégie d'échantillonnage.

### 2.4.2.1.2 Le caractère saisonnier des données

Une autre solution envisagée face aux données déséquilibrées de la BDD était de prendre en compte le caractère saisonnier des submersions marines. En effet, l'un des paramètres de la submersion marine est la marée et l'attraction de la lune et du soleil sur les océans permet de prédire avec précision leur occurrence. La difficulté ne réside alors pas dans la temporalité mais dans l'amplitude. Ainsi, il n'est pas nécessaire de connaître la forme de tous les fonds marins pour pouvoir estimer la hauteur des marées car il existe des modèles mathématiques par l'estimation statistique<sup>72</sup>. Pratiquement, il s'agit d'alléger la BDD en supprimant les données sur des périodes où il n'y a jamais eu de submersions marines. Le graphique suivant permet de représenter l'ensemble des submersions marines en fonction du type d'arrêtés Cat-Nat (Demandes refusées ou acceptées) et de leur mois dans l'année.

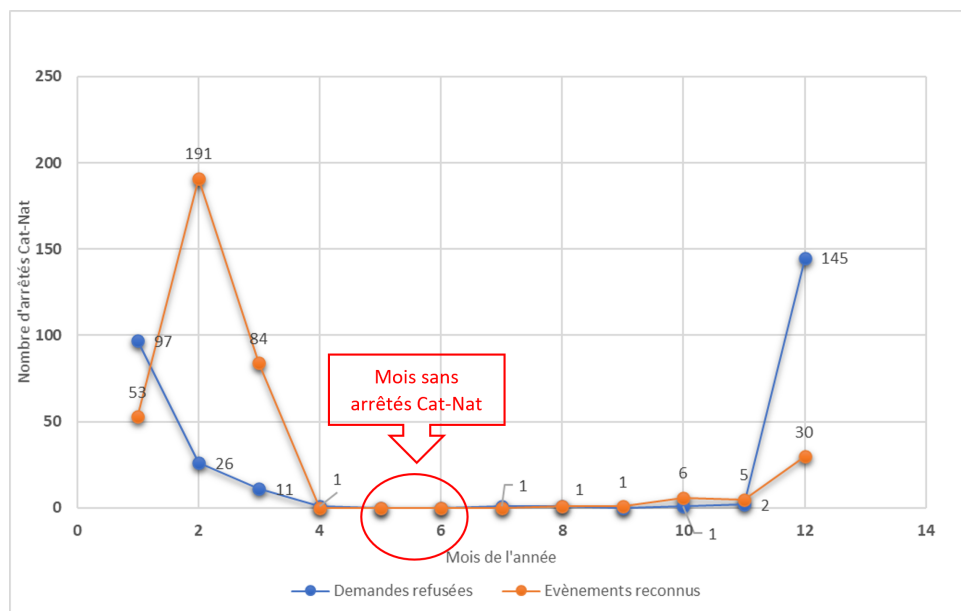


Figure 31 – Concentration mensuelle des événements de type submersion marine

Le graphique permet de se rendre compte que les mois de mai et juin n'ont jamais eu aucune

72. Back Model reduced variational data assimilation for shallow water flow models : [\[en ligne\]](#)

demande d'arrêtés Cat-Nat.

Une base de données où ces mois ont été retraités a été effectuée. En revanche, les résultats ne sont pas gardés car après application de cette base saisonnière, les résultats ne sont pas améliorés. (Cas 4 présent en Annexe I).

Par ailleurs, cette démarche constitue une limite importante. En outre, en appliquant ce retraitement on supprime de la base d'apprentissage et donc de prédiction des mois qui ne sont certes pas touchés à travers l'historique des submersions marines, mais qui pourrait l'être à l'avenir, compte tenu du changement climatique.

### 2.4.2.2 Séparation des données

Une fois la BDD rééquilibrée, la séparation des données est une étape préalable à tout modèle de *machine learning*. En outre, un découpage entre base d'apprentissage et de test doit être effectué. Le choix suivant a été fait pour le découpage de la base totale historique (1990-2005) :

- 75% de la base historique a été utilisée pour constituer la base d'entraînement (*Train*) du modèle (voir Annexe N).
- Le reste constituera la base de test sur laquelle la validation sera enregistrée (voir Annexe O).

En d'autres termes, la prédiction des arrêtés **CCR** de type submersion marine se fait à partir des données historiques (BASE TRAIN). La validation du modèle se fait sur la base de test et dans notre cas sur la base de validation historique. En effet, la profondeur historique disponible sur les arrêtés Cat-Nat est plus récente que sur les autres variables. En outre, les arrêtés Cat-Nat de 2006 à 2020 sont disponibles pour valider le modèle. Une fois cette base de validation historique des arrêtés Cat-Nat utilisée, les bases **RCP** 4.5 et 8.5 sont utilisés pour la projection du modèle (BASE TEST).

### 2.4.3 Calibration du modèle

Afin de pouvoir déterminer la fréquence des arrêtés Cat-Nat, il faut tout d'abord sélectionner des modèles de classification cohérents.

Trois modèles ont été testés afin de de pouvoir calibrer le modèle le plus adapté possible : le *Gradient Boosting*, le *SVM* et le *Random Forest*.

#### 2.4.3.1 Gradient Boosting (XGB)

##### 2.4.3.1.1 Présentation générale

Le *Gradient Boosting* est une technique de *machine learning* qui associe le *boosting* et la descente du *gradient*.

Le *boosting* est un algorithme d'apprentissage supervisé qui consiste à « agréger des classifieurs séquentiellement sur un échantillon d'apprentissage dont les poids [de performance] des individus sont corrigés au fur et à mesure<sup>73</sup> ». Ainsi, à partir d'un ensemble de données, on va augmenter (respective-

73. D'après les notes de cours de Pr. Ricco RAKOTOMALALA de l'Université Lumière Lyon 2, lien : [\[en ligne\]](#)



ment diminuer) le poids de chaque observation compliquée à classer (respectivement sans difficulté de classification). On va développer un nouvel arbre de décision à partir de la version précédente, afin de mieux prédire les erreurs de classification. Le dernier modèle d'ensemble aura des prédictions pondérées obtenues par tous les anciens modèles d'arbres.

La descente du *gradient* est une « technique itérative qui permet d'approcher la solution d'un problème d'optimisation <sup>74</sup> ». En Intelligence Artificielle, on entend parler de cet algorithme lorsqu'on doit trouver le minimum (ou maximum) « local » d'une fonction objective. On prend une valeur d'un paramètre «  $x$  » au hasard et au lieu de tester toutes ses valeurs possibles, on fait varier ce «  $x$  » par rapport à la pente de la fonction objective jusqu'à se rapprocher petit à petit du minimum (ou maximum). La descente de gradient permet d'obtenir les valeurs des paramètres d'une fonction permettant d'atteindre le minimum (ou maximum) local.

### 2.4.3.1.2 Présentation de l'algorithme et visualisation

Cette méthode itérative est donc fondée sur la notion d'apprentissage faible.

La séparation de la base sert à obtenir un nombre fini de nœuds finaux. L'algorithme est qualifié de « gourmand » : à chaque étape, il prend le meilleur découpage de nœud possible. A la fin, plusieurs sous espaces sont disponible pour chaque feuille de l'arbre.

Pour chaque région, une valeur constante est associée :  $x \in R_j \rightarrow T(\Theta, x) = v_j$  où  $\Theta = \{R_j, v_j\}_{j=1}^N$

La formule générale pour un arbre est :  $T(\Theta, x) = \sum_{j=1}^N v_j I(x \in R_j)$

Les paramètres sont obtenus en minimisant le risque empirique :

$$\tilde{\Theta} = \arg_{\Theta} \min \sum_{j=1}^J \sum_{x_i \in R_j} L(y_i, v_j)$$

$v_j$  est en général la classe modale des points de la feuille d'un arbre de classification. Les régions sont difficiles à calculer directement et l'algorithme « gourmand » est utilisé.

Un modèle d'arbre boosté est la somme d'arbres :  $f_M(x) = \sum_{m=1}^M T(x, \Theta_m)$

A chaque étape on résout la formule suivante :  $\tilde{\Theta}_m = \arg_{\Theta_m} \min \sum_{m=1}^M L(y_i, f_{m-1}(x) + T(x_i, \Theta_m))$

Les régions et les constantes sont ensuite calculés comme pour un arbre unique.

Pour une fonction de perte cela est représentée par la fonction suivante :

$$grad_{im} = \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right] f(x_i) = f_{m-1}(x_i)$$

On veut utiliser un nouvel arbre pour approximer le gradient. On résout alors :

$$\tilde{\Theta}_m = \arg_{\Theta} \min \sum_{i=1}^N (-grad_{im} - T(x_i, \Theta))^2$$

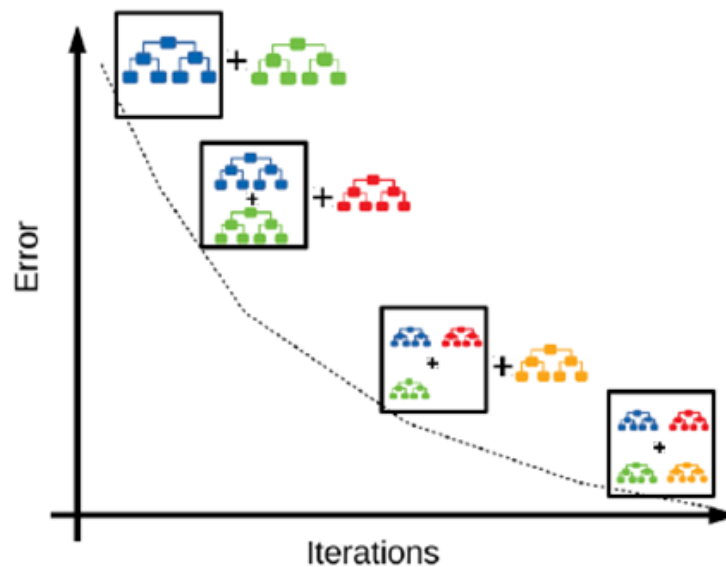
Pour  $k$  class, le gradient de classification est :

$$-grad_{ikm} = I(y_i = k) - p_k(x_i)$$

74. D'après les notes de cours de Pr. Ricco RAKOTOMALALA de l'Université Lumière Lyon 2, lien : [\[en ligne\]](#)

Une présentation d'un récapitulatif de l'algorithme de classification est disponible en Annexe J.

Ce principe de sommer différents arbres afin de décroître l'erreur d'apprentissage peut être représenté par le schéma suivant :



**Figure 32** – Réduction de l'erreur d'apprentissage avec le Gradient Boosting

Le schéma d'apprentissage est donc le suivant :

- Choix de  $N$  et du taux d'apprentissage  $\alpha$
- Ajuster un arbre  $f_0$
- Définir une nouvelle cible telle que  $Y_1 = Y - f_0(x)$
- Ajuster  $f_1$  comme une nouvelle cible définie par  $\alpha \times Y_1$
- Itérer jusqu'à atteindre  $N$

### 2.4.3.1.3 Application et considération du modèle

Il est possible d'appliquer cette méthode à partir de la bibliothèque Python « XGBoost » qui fournit un cadre d'amplification ou *boosting* de gradient. Le renforcement des gradients est une technique d'apprentissage automatique courante pour les problèmes de classification.

Il s'agit d'une procédure flexible, portable, précise et hautement efficace. En effet, les performances sont améliorées en combinant plusieurs modèles en un seul modèle pour corriger les erreurs commises par les modèles existants. Cette méthode optimise le modèle en faisant des prédictions.

Les 2 paramètres les plus importants de l'estimateur *GradientBoostingClassifier* sont *n\_estimators* et *learning\_rate*. Cet estimateur supporte la classification binaire et multi-classes.

Le paramètre *n\_estimators* permet de contrôler le nombre d'apprenants faibles. Ils travaillent sur le principe qu'un apprenant faible fait de mauvaises prédictions lorsqu'il est seul, mais produit la meilleure prédiction lorsqu'il est en équipe. De façon pratique, cela permet l'induction d'arbres de régression à

chaque itération quand la classification est à plus de 2 classes. *Gradient Boosting* crée alors un apprenant fort basé sur des apprenants faibles. Il ajoute des modèles séquentiellement. Par conséquent, les erreurs des modèles faibles sont corrigées par les prochains modèles de la chaîne pour obtenir une solution optimisée, ce qu'on appelle l'ensemblage <sup>75</sup>.

Le taux d'apprentissage est un hyperparamètre dans la plage  $(0,0, 1,0]$  qui contrôle le débordement via le retrait. Ce taux a été fixé à 30%.

Les principales raisons d'utiliser ce modèle sont sa grande vitesse d'exécution, l'amélioration du rendement du modèle et la réduction des erreurs de modèle.

Les avantages sont donc la robustesse de la conception, sa bonne capacité de généralisation, et son bon pouvoir prédictif. En revanche, cela est considéré comme une boîte noire et nécessite d'autres outils pour pouvoir correctement contrôler le modèle ce qui lui fait perdre en interprétabilité.

### 2.4.3.2 *Random Forest*

#### 2.4.3.2.1 *Présentation générale*

L'algorithme des forêts aléatoires (*Random Forest*) est un algorithme de classification basé sur l'assemblage d'arbres de décisions indépendants. Cet algorithme a été proposé par Leo Breiman et Adèle Cutler en 2001.

Le défaut de l'arbre de décision résidait dans la forte dépendance de l'échantillon initial. Afin d'améliorer la performance des résultats, on peut effectuer un apprentissage sur de multiples (une dizaine voire centaine) arbres de décisions (une forêt d'arbres).

Le *Random Forest* repose sur 2 types de tirages aléatoires : le *bagging* et le *feature sampling*.

Le *bagging* est une méthode reposant sur le découpage de la base de données en plusieurs sous-ensembles. Ici, il s'agit de construire plusieurs arbres de décisions en tirant aléatoirement plusieurs échantillons d'observations. On entraîne ensuite nos différents sous-ensembles (soit les arbres de décisions). Enfin on rassemble les résultats issus des modèles entraînés.

Le *feature sampling* est un tirage aléatoire sur les variables explicatives à chaque nœud qui aboutit à des arbres non corrélés. Le but est effectivement d'obtenir des arbres complètement décorrélés pour que la stratégie d'agrégation soit performante. En effet, le *Random Forest* réside dans le principe suivant : l'union fait la force. Les estimateurs pris individuellement sont de moins bonne qualité mais réunis ils permettent d'obtenir une prédiction extrêmement performante.

#### 2.4.3.2.2 *Présentation de l'algorithme et visualisation*

L'algorithme suivant permet de définir un arbre de décision. A chaque nœud est associée une variable selon un critère défini par l'utilisateur (variance, Gini) afin de déterminer des classes aussi homogènes que

---

<sup>75</sup>. Plus d'informations sont disponibles via le lien suivant : [\[en ligne\]](#)

possible.

Le schéma d'apprentissage est le suivant :

- Choix de la variable et du seuil de la valeur associée qui génère la meilleure séparation.
- Tester l'amélioration sur l'erreur de prédiction dans la base test.
- Répéter jusqu'à ce que le critère d'erreur soit atteint.

Un ensemble diversifié de classificateurs est donc créé en introduisant le caractère aléatoire dans la construction du classificateur. La prédiction de l'ensemble est donnée comme la prédiction moyenne des classificateurs individuels.

Comme les autres classificateurs, les classificateurs forestiers doivent être équipés de deux tableaux : un tableau X de forme éparsée ou dense ( $n\_samples$ ,  $n\_features$ ) contenant les échantillons d'apprentissage, et un tableau Y de type ( $n\_samples$ ,) contenant les valeurs cibles (étiquettes de classe) pour les échantillons de formation.

Comme les arbres de décision, les forêts d'arbres s'étendent également à des problèmes de production multiple (si Y est un tableau de type ( $n\_samples$ ,  $n\_outputs$ )).

Dans les forêts aléatoires (*RandomForestClassifier*), chaque arbre de l'ensemble est construit à partir d'un échantillon tiré avec remplacement (c.-à-d., un échantillon *bootstrap*) de l'ensemble de formation.

En outre, lors de la division de chaque nœud lors de la construction d'un arbre, le meilleur *split* est trouvé soit à partir de toutes les caractéristiques d'entrée ou d'un sous-ensemble aléatoire de taille  $max\_features$ .

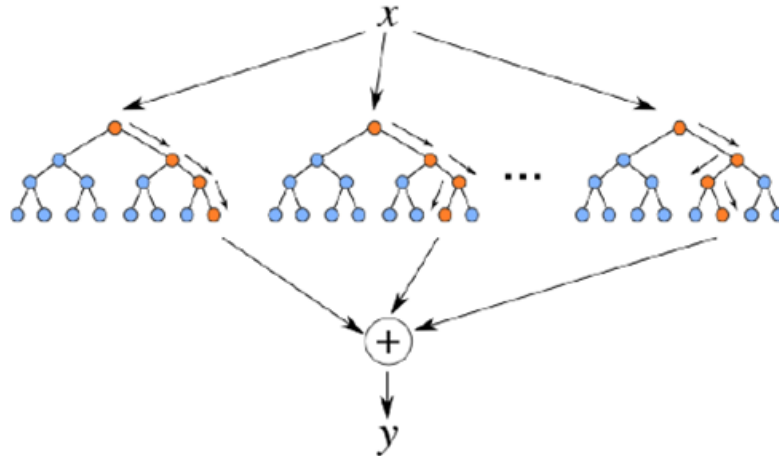
Le but de ces deux sources de hasard est de diminuer la variance de l'estimateur 'forestier'. En effet, les arbres de décision individuels présentent généralement des écarts élevés et ont tendance à déborder. Le caractère aléatoire injecté dans les forêts donne des arbres de décision avec des erreurs de prédiction quelque peu découplées. En prenant une moyenne de ces prédictions, certaines erreurs peuvent s'annuler. Les forêts aléatoires réduisent la variance en combinant divers arbres, parfois au prix d'une légère augmentation du biais. En pratique, la réduction de la variance est souvent significative, ce qui donne un meilleur modèle global.

Les principaux paramètres à ajuster lors de l'utilisation de ces méthodes sont  $n\_estimators$  et  $max\_features$ . Le premier est le nombre d'arbres dans la forêt. Plus il est grand, mieux c'est, mais aussi plus il faudra de temps pour calculer. De plus, notez que les résultats cesseront de s'améliorer considérablement au-delà d'un nombre critique d'arbres. Cette dernière est la taille des sous-ensembles aléatoires de fonctionnalités à prendre en compte lors du fractionnement d'un nœud. Plus la réduction de la variance est faible, plus l'augmentation du biais est importante. De bons résultats sont souvent obtenus lors du réglage de  $max\_depth=None$  en combinaison avec  $min\_samples\_split=2$  (c'est-à-dire lors du développement complet des arbres. Les meilleures valeurs de paramètres doivent toujours être validées par recoupement. De plus, dans les forêts aléatoires, les échantillons *bootstrap* sont utilisés par défaut (`bootstrap=True`)).

### 2.4.3.2.3 Application et considération du modèle

Le principe d'un algorithme de forêt aléatoire est que pour un nombre de fois  $N$  défini par l'utilisateur, la BDD est rééchantillonnée. Dans chaque nouvelle base, un *Gradient Boosting* est calibré. La prédiction moyenne de ces  $N$  modèles fournit la prédiction finale.

Cet algorithme peut être représenté par la figure suivante :



**Figure 33** – Algorithme Gradient Boosting

Ses avantages résident dans la robustesse de sa conception, sa facilité de parallélisation et sa gestion de nombreux cas possible. Par ailleurs, sa simplicité d'interprétation est liée à une lecture facile des résultats.

Ces inconvénients sont également son côté boîte noire, et la nécessité d'avoir d'autres outils pour contrôler le modèle ce qui lui fait perdre en interprétabilité.

## 2.4.3.3 SVM

### 2.4.3.3.1 Présentation du modèle

Les machines à vecteurs supports, plus connues sous le sigle anglais **SVM** (*Support Vector Machines*) sont un cas particulier du minimiseur du risque empirique convexe.

Développé dans les années 1990, le **SVM** est un algorithme supervisé de *machine learning*. Il peut être utilisé à la fois pour de la classification, de la régression ou encore la détection de valeurs aberrantes. A noter qu'il reste plus utilisé dans le premier cas et c'est également celui qui est employé dans ce mémoire.

Dans les faits, chaque donnée est tracée comme un point dans un espace à  $n$ -dimension (où  $n$  est le nombre de caractéristiques que nous avons) avec la valeur de chaque caractéristique comme valeur d'une coordonnée particulière. La classification peut ensuite avoir lieu en cherchant l'hyperplan qui différencie le mieux les trois classes.

Le principe est simple, il s'agit de séparer les données en classes à l'aide d'une frontière. Les vecteurs de support sont simplement des coordonnées de l'observation individuelle. Le classificateur **SVM** est une frontière qui sépare au mieux les 3 classes. La distance est également appelée « marge » et doit être

maximale. Les SVM sont les données les plus proches de la frontière.

Le type de SVM paramétré correspond à une *cross validation*. Il s'agit d'une répétition du schéma apprentissage/validation sur différentes fractions constituées à partir des données initiales. Ce schéma est alors moins soumis à la variance de la base d'exemples.

Les calculs de séparation des points de données dépendent d'une fonction du noyau. Il existe différentes fonctions du noyau : linéaire, polynomiale, gaussienne, fonction de base radiale (RBF) et sigmoïde. En termes simples, ces fonctions déterminent la fluidité et l'efficacité de la séparation des classes, et jouer avec leurs hyperparamètres peut conduire à un surajustement ou à un sous-ajustement.

Dans son format le plus simple, le SVM ne prend pas en charge la classification multi-classe. Il prend en charge la classification binaire et la séparation des points de données en deux classes. Pour la classification multi-classe, le même principe est utilisé après avoir décomposé le problème multi-classe en plusieurs problèmes de classification binaire.

### 2.4.3.3.2 Visualisation du modèle

L'idée est de mapper les points de données à l'espace de grande dimension pour obtenir une séparation linéaire mutuelle entre toutes les deux classes. C'est ce qu'on appelle une approche *One to One*, qui décompose le problème multi classe en plusieurs problèmes de classification binaire. Un classificateur binaire pour chaque paire de classes.

Dans l'approche *One-to-One*, nous avons besoin d'un hyperplan pour séparer toutes les deux classes, en négligeant les points de la troisième classe. Cela signifie que la séparation ne prend en compte que les points des deux classes dans la division actuelle. Par exemple, la ligne rouge-bleu tente de maximiser la séparation uniquement entre les points bleus et rouges. Cela n'a rien à voir avec les points verts :

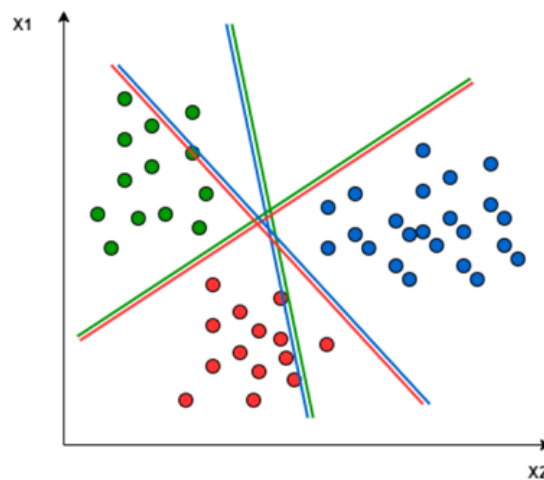
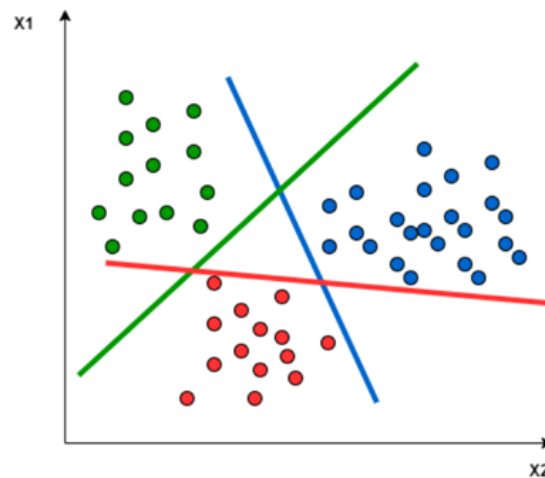


Figure 34 – SVM approche One-to-One

Une autre approche que l'on peut utiliser est *One-to-Rest*. Dans cette approche, la répartition est définie sur un classificateur binaire pour chaque classe.

Dans l'approche *One-to-Rest*, nous avons besoin d'un hyperplan pour séparer une classe de toutes les autres à la fois. Cela signifie que la séparation prend en compte tous les points, les divisant en deux

groupes ; un groupe pour les points de classe et un groupe pour tous les autres points. Par exemple, la ligne verte tente de maximiser la séparation entre les points verts et tous les autres points à la fois.



**Figure 35** – SVM approche One-to-Rest

Le schéma d'apprentissage est donc le suivant :

- Choix du kernel ou noyau (Gaussien, sigmoïde, polynomial)
- Choix de la fonction de coût
- Résolution du programme d'optimisation (Primal ou dual)

#### **2.4.3.3 Avantages et inconvénients du modèle**

Les principaux avantages de ce modèle sont d'une part son efficacité grâce à son fort pouvoir prédictif dans les espaces de dimensions raisonnables. Puis, son efficacité dans les cas où le nombre de dimensions est supérieur au nombre d'échantillons. De plus, il utilise un sous-ensemble de points d'entraînement dans la fonction de décision (appelés vecteurs de support), donc il est également efficace dans la gestion de mémoire. Pour finir, il est polyvalent : différentes fonctions du noyau peuvent être spécifiées pour la fonction de décision. Des noyaux communs sont fournis, mais il est également possible de spécifier des noyaux personnalisés.

En revanche, il existe deux principaux inconvénients au SVM. Si le nombre de caractéristiques est beaucoup plus grand que le nombre d'échantillons, il faut éviter de trop ajuster le choix des fonctions Kernel et le terme de régularisation est crucial. Choisir le bon noyau n'est pas toujours aisé. De plus, les SVM ne fournissent pas directement d'estimations de probabilité, elles sont calculées à l'aide d'une validation croisée coûteuse à cinq (voir Scores et probabilités, ci-dessous). Pour finir, tout comme le premier modèle, il est considéré comme une boîte noire qui nécessite d'autres outils pour contrôler correctement le modèle.

#### **2.4.3.4 Présentation des résultats**

##### **2.4.3.4.1 Critères de validité choisis**

Les avantages et inconvénients propres à chaque méthode peuvent être résumés par le tableau suivant ayant cette légende :

Moyen	Correct	Bien	Très bien
-------	---------	------	-----------

Algorithme	Pré-traitement	Modularité	Convergence	Prédiction	Interpretation
XGB	Très bien	Correct	Correct	Moyen	Bien
RF	Très bien	Correct	Correct	Bien	Moyen
SVM	Moyen	Moyen	Correct	Bien	Moyen

**Table 10** – Synthèse des avantages et inconvénients des modèles étudié

Afin de juger de la performance du modèle, il est important de définir un score.

En général, des tests d'adéquation ou de conformité tels que Khi-deux ou encore Kolmogorov-Smirnov sont choisis. En revanche, dans le cadre du *machine learning* où la fonction de lien est implicite pour un événement extrême, il n'est pas adéquat de les employer. En effet, le test de Khi-deux montre une relation avec le Théorème Central Limite et donc la loi normale, ce qui n'est pas le cas avec les lois extrêmes. Par ailleurs, le test de conformité n'est pas sensible aux comportements de queue du modèle, ce qui caractérise les lois extrêmes.

Ainsi, parmi les métriques connues pour valider les modèles d'apprentissage automatiques supervisés, trois critères ont été choisis :

- Le score de précision ou *accuracy*
- Le F1 Score : il calcule la moyenne harmonique entre la précision et le rappel, et les deux dépendent du faux positif et du faux négatif. Il est donc utile de calculer le score f1 lorsque l'ensemble de données n'est pas équilibré.
- La matrice de confusion : la mesure d'erreurs en classification par matrice de confusion est une méthode qui permet de voir le nombre et le types d'erreurs commises par le modèle. Il ne s'agit pas d'un indicateur ni d'un score mais plusieurs indicateurs de score sont créés grâce à la matrice de confusion. Ci-dessous un exemple de matrice de confusion :

		Classes Réelles		
		0	1	2
Classes prédites	0	Vrais arrêts Cat-Nat de type 0	Mauvaise prédiction de classe : type 0 au lieu de 1	Mauvaise prédiction de classe : type 0 au lieu de 2
	1	Mauvaise prédiction de classe type 1 au lieu de 0	Vrais arrêts Cat-Nat de type 1	Mauvaise prédiction de classe : type 1 au lieu de type 2
	2	Mauvaise prédiction de classe : type 2 au lieu de type 0	Mauvaise prédiction de classe : type 2 au lieu de 1	Vrais arrêts Cat-Nat de type 2

**Table 11** – Matrice de confusion à 3 dimensions appliquées à l'étude

La méthode de matrices de confusion est utilisée sur l'échantillon d'apprentissage et de test. Cette méthode nous permet de voir la sensibilité entre les classes.



Les tests de score de qualité de modèle sont réalisés sur les données non apprises. La qualité du modèle dépend du score obtenu sur la base de test.

Le choix du modèle a été effectué en deux temps. D'une part, la base de validation historique composée des arrêts Cat-Nat de 2006 à 2020 a été utilisée pour vérifier la qualité de prédiction du modèle. En effet, cela permet de comparer la réalité aux prédictions. Une fois la première sélection effectuée, l'étape suivante est le choix du modèle en comparant les meilleurs résultats de chaque modèle.

### 2.4.3.4.2 Résultats du Gradient Boosting

En fonction des cas étudiés, les résultats du modèle sur la base d'apprentissage peuvent être résumés par la figure suivante :

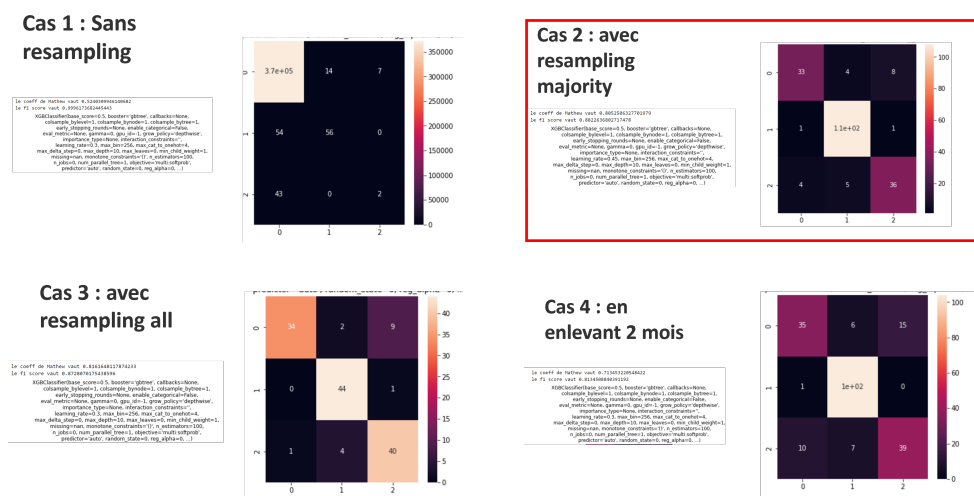


Figure 36 – Résultats sur la base d'apprentissage

Le premier modèle semble effectué un sur-apprentissage en ne prédisant que des classes 0. Le modèle ayant les meilleurs résultats sur la base d'apprentissage est le modèle effectuant un rééchantillonnage de type « majorité ». Cela semble cohérent avec le fait que notre base de données soit déséquilibré par une classe.

Il est maintenant intéressant de regarder les résultats de ce modèle en fonction des RCP étudiés. Comme le montre les résultats suivants représentant respectivement le RCP 4.5 et 8.5, les mêmes conclusions peuvent être apportées.

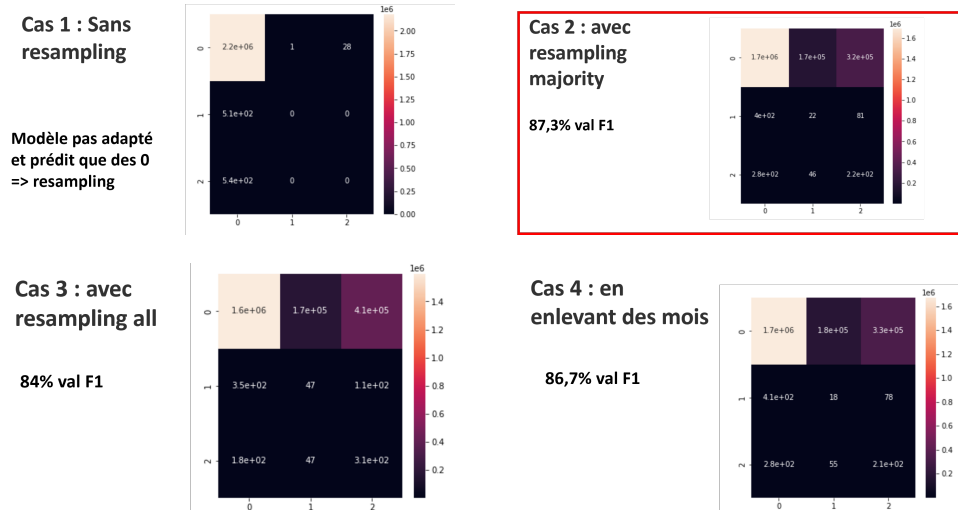


Figure 37 – Résultats sur la base de validation historique RCP 8.5

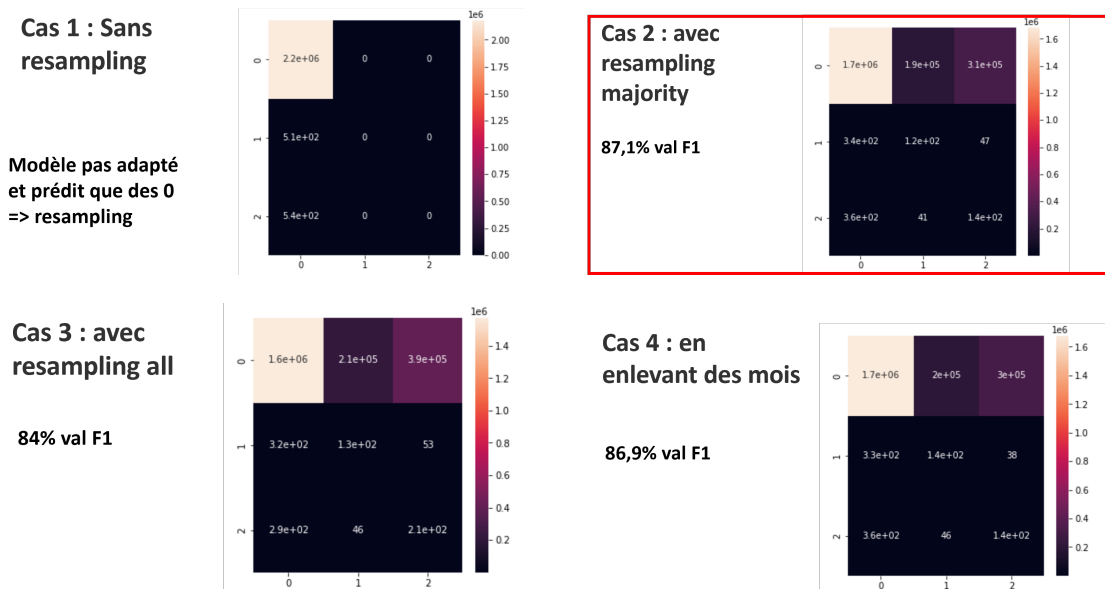


Figure 38 – Résultats sur la base de validation historique RCP 4.5

Le modèle *Gradient Boosting* choisi est celui étant rééchantillonné selon la classe majoritaire.

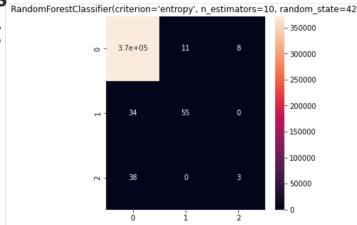
### 2.4.3.4.3 Résultats du Random Forest

Les résultats sur la base d'apprentissage peuvent être résumés par la figure suivante :

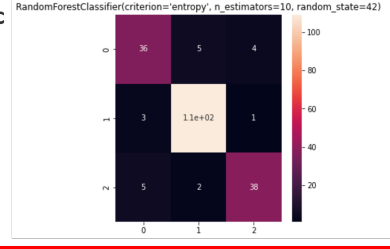
le coeff de Mathew vaut 0.5796227485675695  
le f1 score vaut 0.9997104902470135

le coeff de Mathew vaut 0.8324068974551972  
le f1 score vaut 0.9006905573763822

**Cas 1 : Sans resampling**

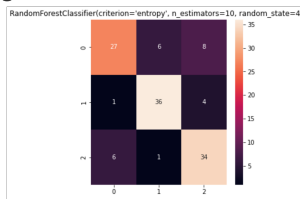


**Cas 2 : avec resampling majority**



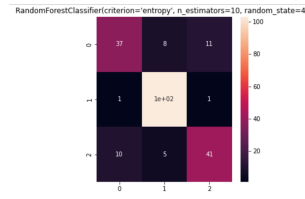
**Cas 3 : avec resampling all**

le coeff de Mathew vaut 0.6855829496241486  
le f1 score vaut 0.7862506841817186



**Cas 4 : en enlevant 2 mois**

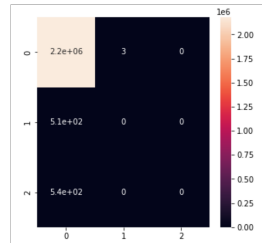
le coeff de Mathew vaut 0.7355639673970379  
le f1 score vaut 0.8287921538274061



**Figure 39** – Résultats sur la base d'apprentissage

Il est possible de faire les mêmes conclusions que sur le *Gradient Boosting*. Les résultats sur la base de validation historique sont les suivants respectivement pour les scénarios RCP 4.5 et 8.5 :

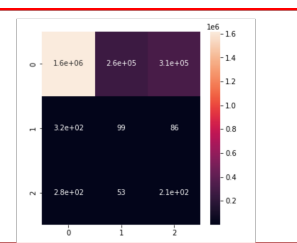
**Cas 1 : Sans resampling**



Modèle pas adapté et prédit que des 0 => resampling

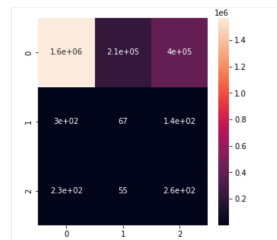
**Cas 2 : avec resampling majority**

85% val F1



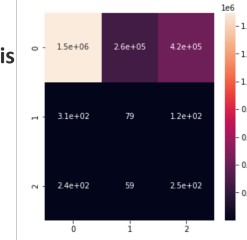
**Cas 3 : avec resampling all**

83,6% val F1

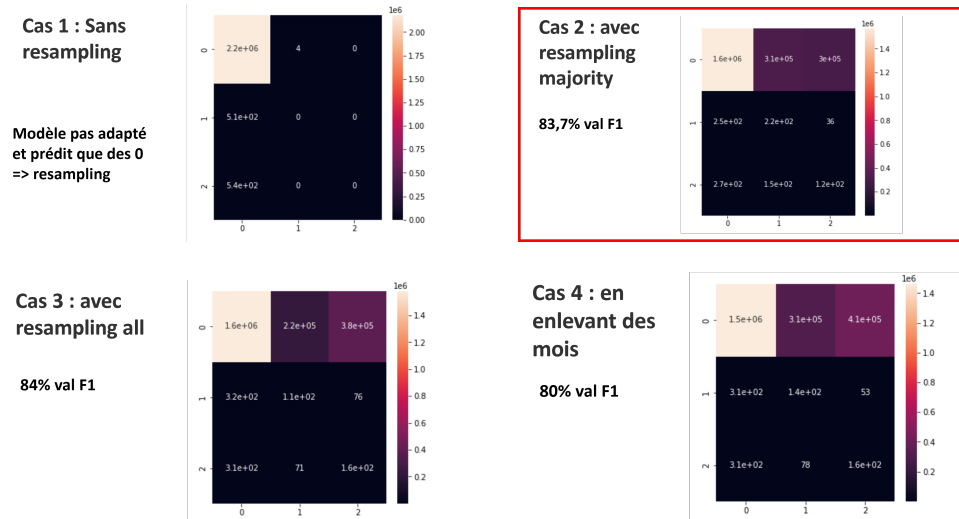


**Cas 4 : en enlevant des mois**

81% val F1



**Figure 40** – Résultats sur la base de validation historique RCP 8.5



**Figure 41** – Résultats sur la base de validation historique RCP 4.5

A nouveau, le modèle retenu pour la phase de sélection du meilleur modèle est le deuxième cas avec stratégie de rééchantillonnage sur la classe majoritaire.

### 2.4.3.5 Résultats du SVM

Les cas sans rééchantillonnage et avec la base retraitée des mois sans submersions marines ne sont pas pertinents et sont donc exclus pour les raisons vues précédemment. En outre, du fait d'une classe particulièrement déséquilibré, il est nécessaire d'opérer un rééchantillonnage des données. De plus, enlever deux mois de l'année représente une limite indéniable.

Ainsi, seuls les cas de rééchantillonnage « *majority* » et « *all* » seront étudiés. Par ailleurs, deux types de choix de noyaux seront étudiés : RBF (*Radial Basis Function*) et polynomial. Ces noyaux sont courants dans le cadre d'une classification quand la variable à expliquer est multimodale.

Les résultats sur la base d'apprentissage sont alors les suivants :

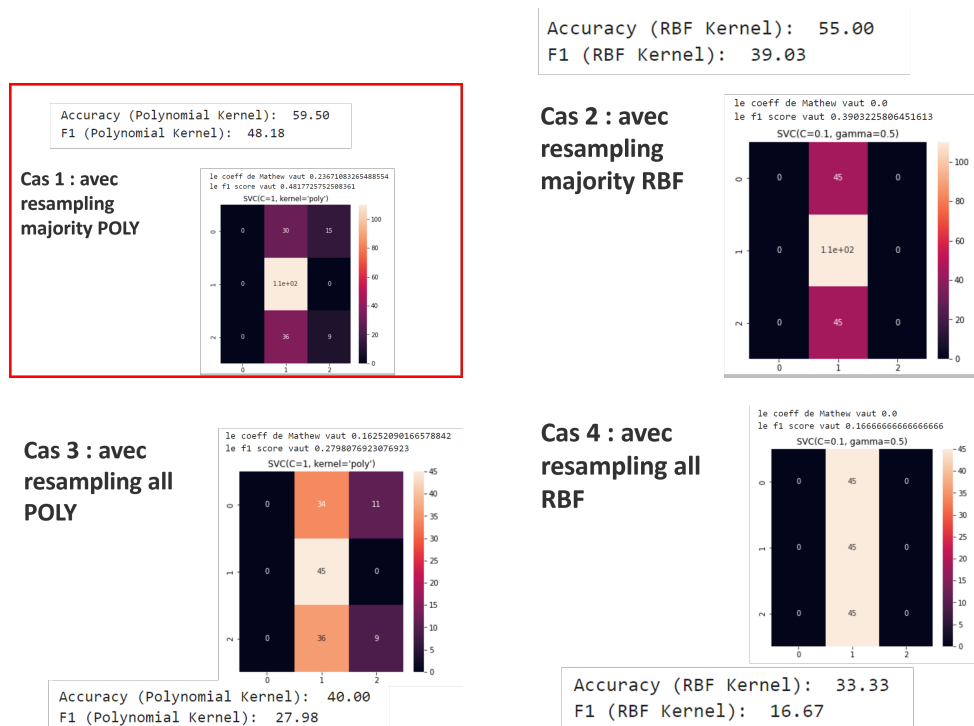


Figure 42 – Résultats sur la base d'apprentissage

Dans le cas du SVM il est également possible d'appliquer une option « *balanced* » permettant de prendre en compte le caractère déséquilibré des données. L'utilisation de ce paramètre permet d'améliorer les résultats comme le montre la figure suivante :

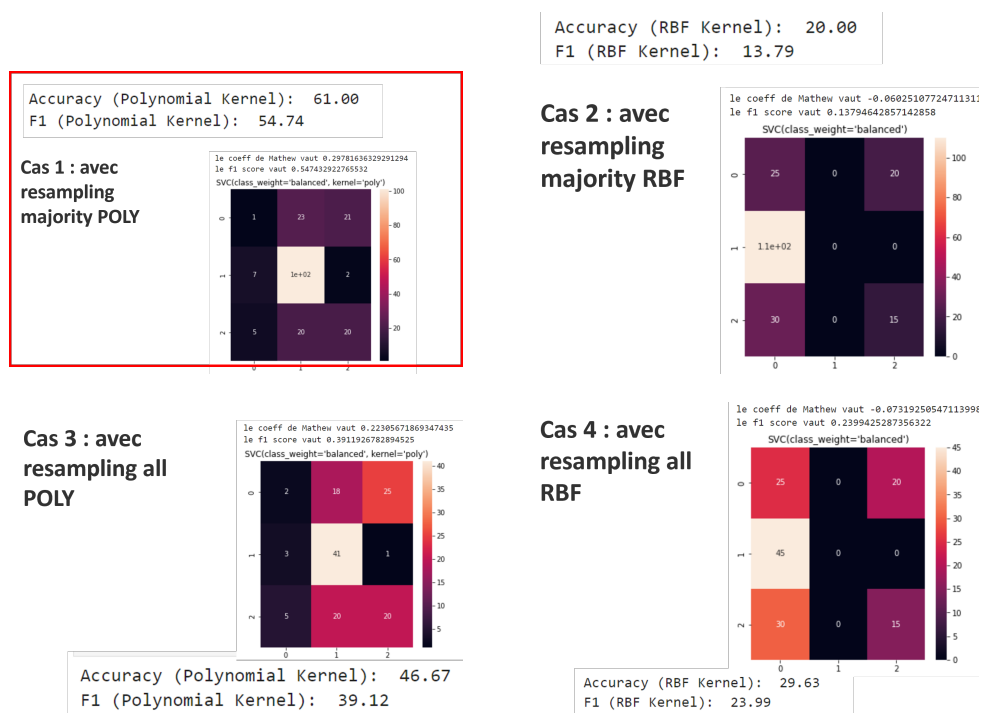


Figure 43 – Résultats sur la base d'apprentissage avec option « *balanced* »

Le modèle retenu pour la partie suivante traitant de sélection du meilleur modèle est le cas polynomial avec à nouveau le rééchantillonnage sur la classe majoritaire « *majority* ».

La validation du modèle sur la base historique de validation peut être représentée par la figure suivante en fonction des RCP 4.5 et 8.5 pour le meilleur modèle évoqué :

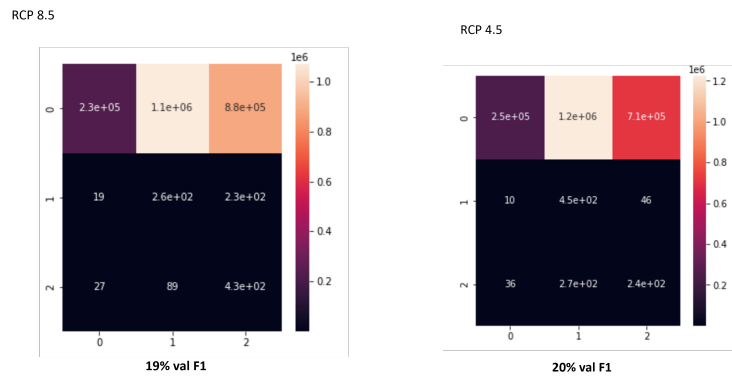


Figure 44 – Validation du modèle sur la base historique

Il est possible d’observer que les résultats sur la base de test sont trop faibles. Le SVM ne sera donc pas retenu pour la sélection du meilleur modèle.

Une des raisons pouvant expliquer que dans notre cas ce modèle ne soit pas optimal est la taille de la base de données. En effet, cette méthode est très efficace avec des bases de données modérées (c’est-à-dire quand il ne s’agit pas d’un trop grand nombre d’observations). Ce qui n’est pas le cas dans le traitement de nos bases de données.

#### 2.4.4 Choix du meilleur modèle

Cette étape permet de mesurer la capacité prédictive du modèle sur la base de validation. En d’autres termes, les deux derniers modèles restants (*Gradient Boosting* et *Random Forest*) sont utilisés pour faire des prédictions sur la base de validation et mesurer l’erreur commise sur cette base.

La base d’apprentissage de chaque modèle a été utilisée sur la période 1990-2005. En effet, les variables climatiques dans le cadre de COPERNICUS ne vont que jusqu’à 2005. En revanche, la base contenant la variable explicative des arrêtés Cat-Nat dispose d’une temporalité plus longue. En effet, elle va de 1990 à 2020 comme le montre le graphique suivant présentant l’ensemble des jours d’arrêtés Cat-Nat déclarés :

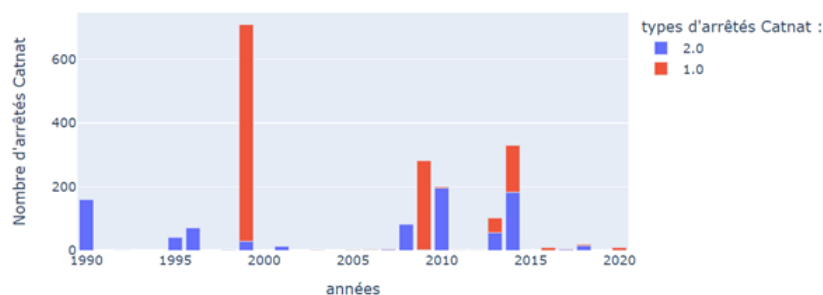
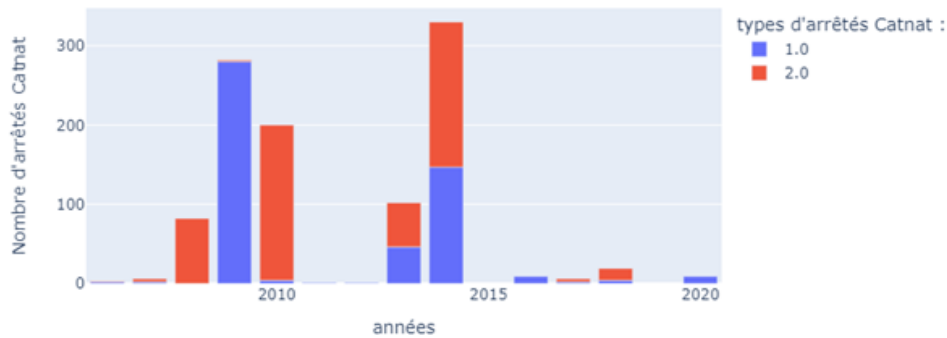


Figure 45 – Ensemble des jours d’arrêtés Cat-Nat déclarés

Les détails par années sont présents en annexe Q

Afin de vérifier la qualité de prédiction de chaque modèle la période de validation représentant la réalité entre 2006 et 2020 est utilisée.

Le graphique suivant permet de représenter cette base de validation à partir de la variable à prédire des arrêtés Cat-Nat selon le maillage journalier :



**Figure 46** – Base de validation historique sur la période 2006 à 2020

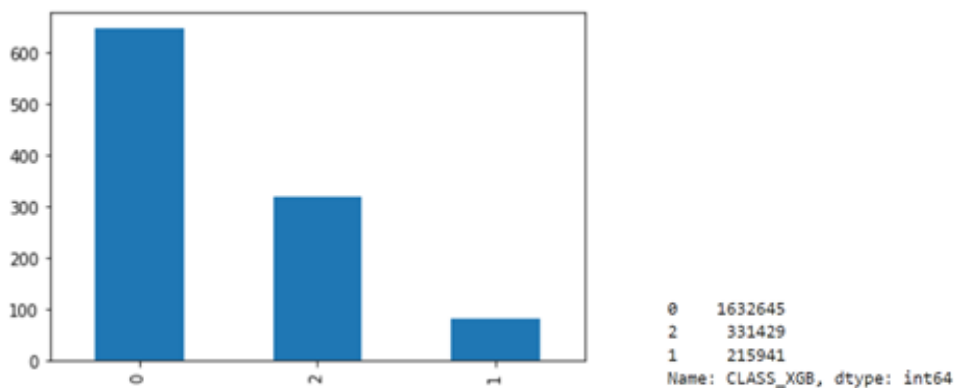
La première étape est de considérer la classe déséquilibrée (CLASS=0) signifiant qu'il n'y a pas de submersions marines.

Parmi les 2181065 observations composant la base historique journalière des cantons du littoral atlantique français, 2180015 sont de type 0, soit presque la totalité de la base. Ainsi, il est nécessaire de traiter la modalité « CLASS=0 » de manière différenciée.

### 2.4.4.1 Le modèle *Gradient Boosting*

#### 2.4.4.1.1 La base de prédiction RCP 4.5

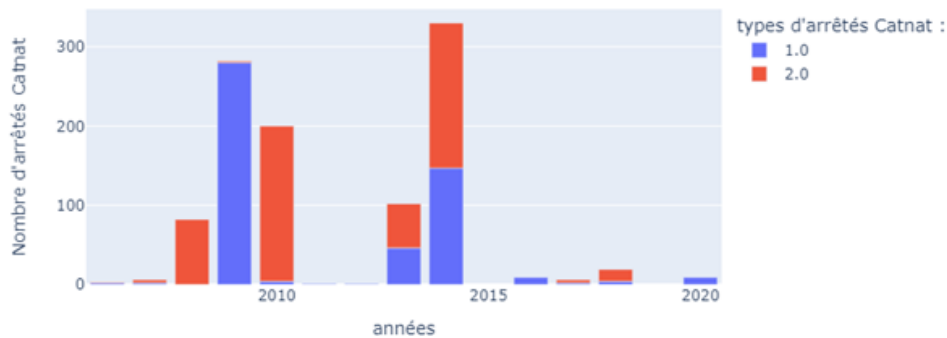
Les résultats du modèle *Gradient Boosting* dans la base de validation quand il n'y a pas de submersion marine sont les suivants :



**Figure 47** – Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS\_XGB ») du modèle *Gradient Boosting* RCP 4.5 sur la base recensant l'absence de submersion marine

L'erreur de prédiction sur le modèle XGB est d'environ 25% en résolvant l'opération suivante :  $\frac{331429+215941}{2180015}$  pour la modalité « CLASS=0 ».

Il est maintenant possible de regarder l'occurrence des événements de type submersion marine pour le modèle *Gradient Boosting* :



**Figure 48** – Résultat du modèle XGB pour le scénario RCP 4.5

On se rend que le modèle a du mal à distinguer les types 0 (pas de submersion marine) et 1 (submersion marine déclarée mais pas acceptée). Sachant que les experts chargés de valider un arrêté Cat-Nat se fondent également sur des métriques liées à des variables naturelles, il est possible de se demander si cela n'est pas cohérent de se retrouver avec une majorité de type 0 au lieu de 1.

En revanche le type 2 est plutôt bien prédit. Afin de prendre en compte le déséquilibre de la classe majoritaire il est possible de regarder les résultats du F1 score et du taux de précision (*accuracy*) pour le modèle :

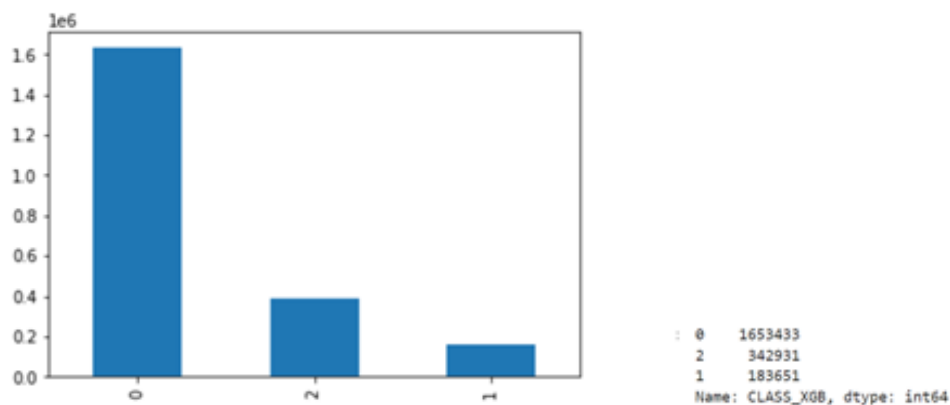
	Gradient Boosting Base d'apprentissage	Gradient Boosting Base de validation
Accuracy	89.4%	74.9%
F1-score	89%	85.6%

**Table 12** – Résultats du modèle XGB pour le scénario RCP 4.5

On remarque que le modèle effectue de meilleures prédictions sur la base d'apprentissage mais conserve de bons résultats compte tenu du déséquilibre de la base.

#### 2.4.4.1.2 La base de prédiction RCP 8.5

Les résultats du modèle XGB dans la base de validation quand il n'y a pas de submersion marine sont les suivants :



**Figure 49** – Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS\_XGB ») du modèle Gradient Boosting RCP 8.5 sur la base recensant l'absence de submersion marine

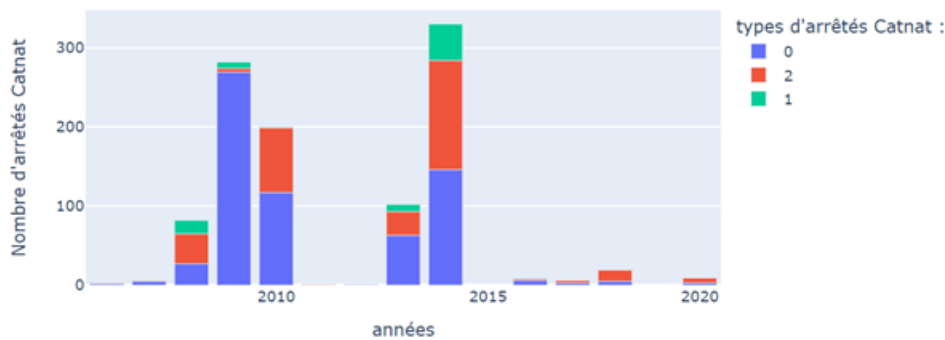


	XGB Base d'apprentissage	XGB Base de validation
Accuracy	89.4%	75.8%
F1-score	89%	86.2%

**Table 13** – Résultats du modèle XGB pour le scénario RCP 8.5

Ainsi, l'erreur de prédiction sur le modèle *Gradient Boosting* est d'environ 24% en résolvant l'opération suivante :  $\frac{342931+183651}{2180015}$  pour la modalité « CLASS=0 ».

Il est maintenant possible de regarder l'occurrence des événements de types 1 ou 2 représentant une déclaration de submersion marine pour le modèle *Gradient Boosting* :



**Figure 50** – Résultat du modèle XGB pour le scénario RCP 8.5

On se rend que le modèle a du mal à distinguer les types 0 (pas de submersion marine) et 1 (submersion marine déclarée mais pas acceptée). Sachant que les experts chargés de valider un arrêté Cat-Nat se fondent également sur des métriques liées à des variables naturelles, il est possible de se demander si cela n'est pas cohérent de se retrouver avec une majorité de type 0 au lieu de 1.

En revanche le type 2 est plutôt bien prédit. Afin de prendre en compte le déséquilibre de la classe majoritaire il est possible de regarder les résultats du F1 score et du taux de précision (*accuracy*) pour le modèle :

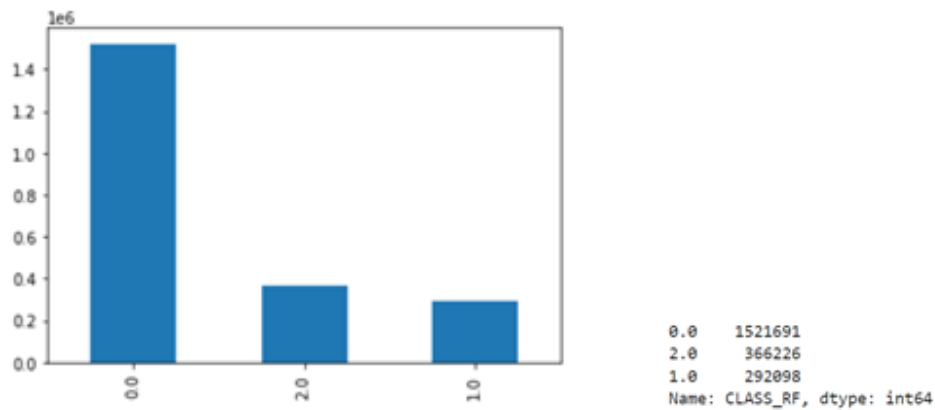
On remarque que les résultats du scénario RCP 8.5 sont comparables au scénario RCP 4.5. Cela est cohérent compte tenu du fait que dans des horizons court terme comme celui-ci les deux trajectoires sont comparables.

#### 2.4.4.2 Le modèle Random Forest

Le même processus est effectué sur ce modèle afin de pouvoir choisir le meilleur modèle.

##### 2.4.4.2.1 La base de prédiction RCP 4.5

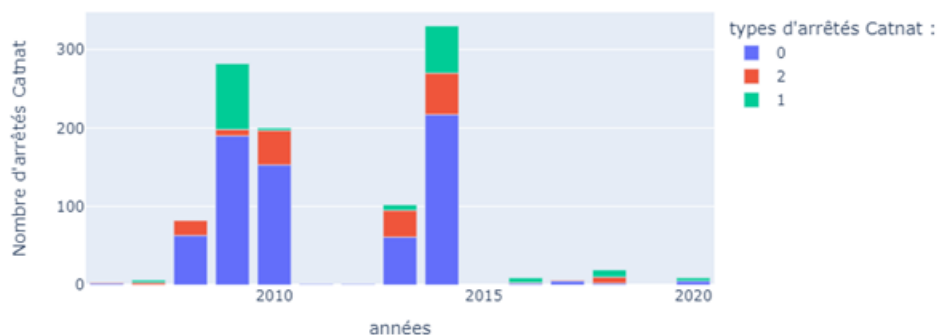
Les résultats du modèle *Random Forest* dans la base de validation quand il n'y a pas de submersion marine sont les suivants :



**Figure 51** – Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS\_RF ») du modèle *Random Forest* RCP 4.5 sur la base recensant l'absence de submersion marine

Ainsi, l'erreur de prédiction sur le modèle *Random Forest* est d'environ 30,2% en résolvant l'opération suivante :  $\frac{366226+292098}{2180015}$  pour la modalité « CLASS=0 ».

Il est maintenant possible de regarder l'occurrence des événements de types 1 ou 2 représentant une déclaration de submersion marine pour le modèle *Random Forest* :



**Figure 52** – Résultat du modèle *Random Forest* pour le scénario RCP 4.5

La modalité de type 1 semble mieux prédite que dans le modèle précédent au détriment du type 2. Afin de pouvoir déterminer quel est le meilleur modèle il semble donc opportun d'utiliser à nouveau des métriques de validation :

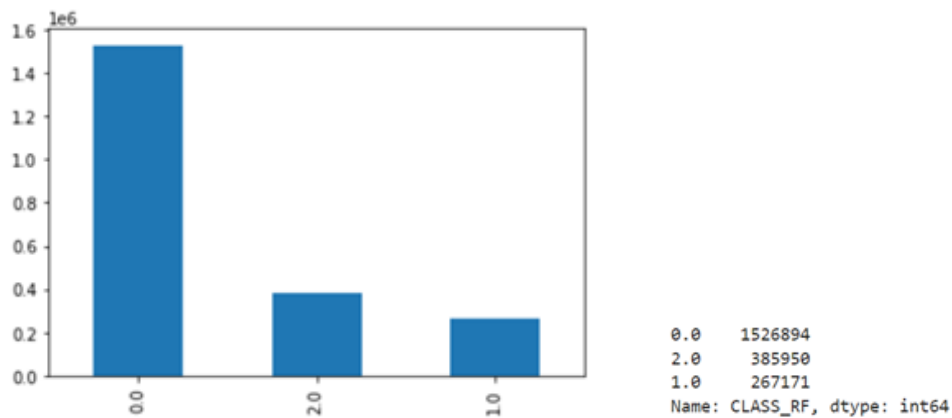
	<i>Random Forest</i> Base d'apprentissage	<i>Random Forest</i> Base de validation
Accuracy	87.4%	69.8%
F1-score	87%	82.1%

**Table 14** – Résultats du modèle *Random Forest* pour le scénario RCP 4.5

Il est possible de remarquer que quel que soit la métrique, le modèle *Gradient Boosting* est préférable pour le scénario RCP 4.5. Le modèle de *Random Forest* est moins précis et effectue donc de moins bonnes prédictions.

**2.4.4.2.2 La base de prédiction RCP 8.5**

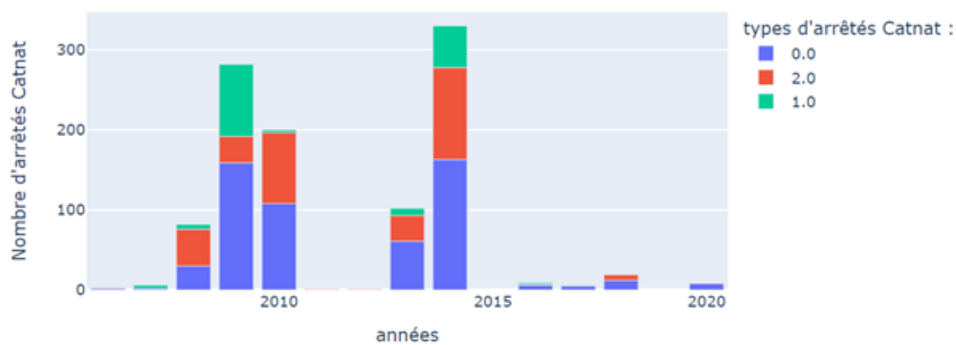
Les résultats du modèle *Random Forest* dans la base de validation quand il n’y a pas de submersion marine sont les suivants :



**Figure 53** – Répartition des modalités de la variable type d’arrêtés Cat-Nat (« CLASS\_XGB ») du modèle *Random Forest* RCP 8.5 sur la base recensant l’absence de submersion marine

Ainsi, l’erreur de prédiction sur le modèle *Random Forest* est d’environ 30% en résolvant l’opération suivante :  $\frac{385950+267171}{2180015}$  pour la modalité « CLASS=0 ».

Il est possible de remarquer que quel que soit le scénario, le taux d’erreur est plus important sur le *Random Forest*.



**Figure 54** – Résultat du modèle *Random Forest* pour le scénario RCP 8.5

Les mêmes conclusions peuvent être faites sur le scénario RCP 8.5. Afin de pouvoir écarter de manière définitive ce modèle, il est nécessaire de regarder les métriques de validation :

	<i>Random Forest</i> Base d’apprentissage	<i>Random Forest</i> Base de validation
Accuracy	87.4%	70%
F1-score	87%	82.3%

**Table 15** – Résultats du modèle *Random Forest* pour le scénario RCP 8.5

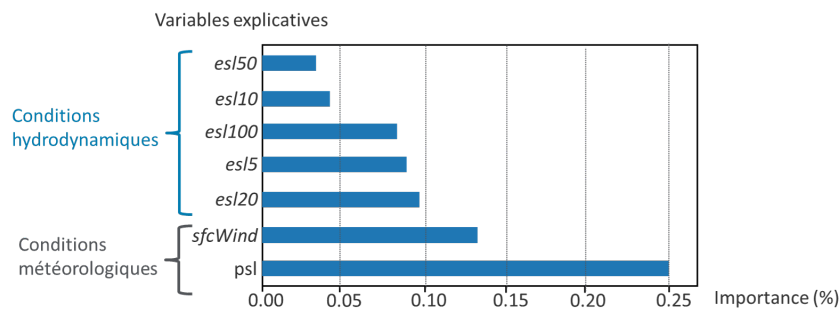
A nouveau, quelque soit l’aspect regardé, le modèle de *Random Forest* est moins performant que le *Gradient Boosting*. Finalement, le modèle *Gradient Boosting* est celui qui sera retenu dans le cadre de la comparaison des projections à horizon 2050 avec les résultats d’autres organismes dans la section suivante.

## 2.4.5 Interprétation du meilleur modèle

Afin de comprendre la combinaison des variables explicatives génératrices des arrêts submersions marines, il est possible de regarder leur importance dans le modèle. En d'autres termes, il s'agit de faire référence aux techniques qui attribuent un score aux caractéristiques d'entrée (variables explicatives) en fonction de leur utilité pour prédire une variable cible (variable à expliquer).

Les scores d'importance des caractéristiques jouent un rôle important dans les études de modélisation prédictive. En outre, cela fournit un aperçu des données, du modèle et la base pour la réduction de la dimensionnalité et la sélection des caractéristiques qui peuvent améliorer l'efficacité et l'efficacité d'un modèle prédictif sur le problème. La plupart des scores d'importance sont calculés par un modèle prédictif qui a été ajusté sur l'ensemble de données. L'inspection du score d'importance donne un aperçu de ce modèle spécifique et des caractéristiques qui sont les plus importantes et les moins importantes pour le modèle lors de la prédiction. Il s'agit d'un type d'interprétation de modèle qui peut être effectué pour les modèles qui le supportent.

Dans le cadre de cette étude, le modèle de *Gradient Boosting* est le meilleur modèle sélectionné. Après avoir été ajusté, le modèle fournit une propriété *feature\_importances* à laquelle on peut accéder pour récupérer les scores d'importance relative pour chaque caractéristique d'entrée. Les résultats sont présentés avec la figure ci-après :



Où *esl* = *extreme sea level* ; *sfcWind* = vent ; *psl* = pression atmosphérique

**Figure 55** – Score d'importance des variables explicatives dans le modèle de *Gradient Boosting*

Les variables explicatives sont de deux types : hydrodynamique (cinq premières) et météorologique (deux dernières).

Dans le cadre de la première catégorie, la variable utilisée est le niveau de mer extrême ou *extreme sea level (ESL)*. Cette variable a été déclinée d'après l'article de recherche utilisée en période de retour. Une période de retour caractérise le temps statistique entre deux occurrences de l'événement submersion en fonction de l'intensité.

Les périodes de retours présent en compte sont : 5, 10, 20, 50 et 100 ans. On peut remarquer que la période de retour de 20 ans est la plus significative (environ 10%). Elle peut être interprétée comme une probabilité statistique. Cette période de retour de 20 ans signifie qu'une submersion s'est produite statistiquement à la fréquence d'une fois tous les 20 ans. Il ne faut pas y voir qu'elle se produira régulièrement tous les 20 ans mais que statistiquement, elle a 20% de chance de se produire durant une année particulière.

Dans le cadre de la seconde catégorie, deux variables ont été utilisées : le vent (psl) et la pression atmosphérique (psl). On remarque que les variables météorologiques sont les plus explicatives du modèle avec respectivement 14 et 25% pour le vent et la pression atmosphérique. Finalement, les scores d'importance présentés confirment que les variables sélectionnées pour la modélisation sont pertinentes à l'étude de la submersion marine.

### 2.4.6 Validation du meilleur modèle

Une fois le meilleur modèle sélectionné, ici *le Gradient Boosting*, il est possible d'effectuer une dernière étape de validation. En outre, il est possible d'utiliser la base de données de validation historique pour comparer les résultats prédits à la réalité. Cette base contient la variable à expliquer ou arrêts Cat-Nat de type submersion marine sur la période 2006 à 2020.

Lorsqu'une submersion marine est déclarée, deux types de décisions d'arrêtés Cat-Nat sont possibles :

- « CLASS 1 » : la demande de submersion marine a été refusée
- « CLASS 2 » : la demande de submersion marine a été acceptée

Les figures ci-après permettent de comparer les résultats obtenus en fonction des types de décisions prises sur la période de validation (2006-2020) en fonction du scénario RCP.

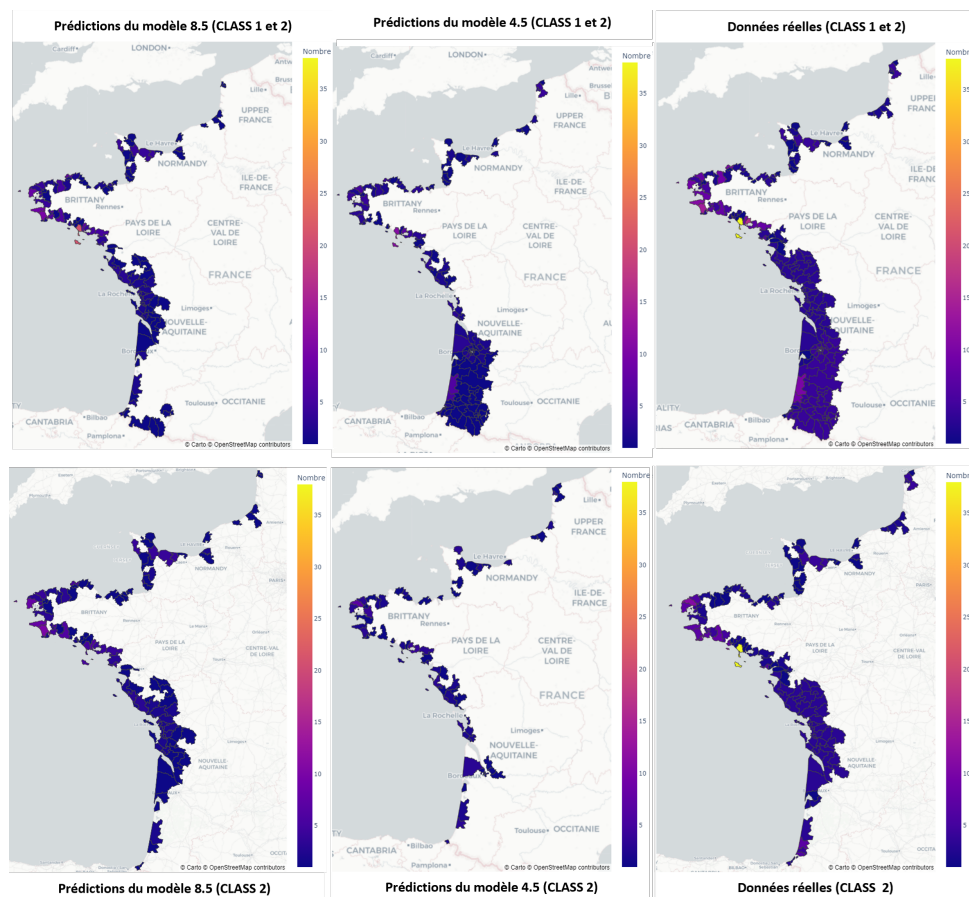


Figure 56 – Comparaison des prédictions à la réalité pour les demandes d'arrêtés Cat-Nat

Dans le cas des demandes refusées, il est possible de réaliser plusieurs observations. Premièrement, pour la partie nord du littoral, allant jusqu'au Golfe de la Gironde, le modèle 8.5 est plus proche de la

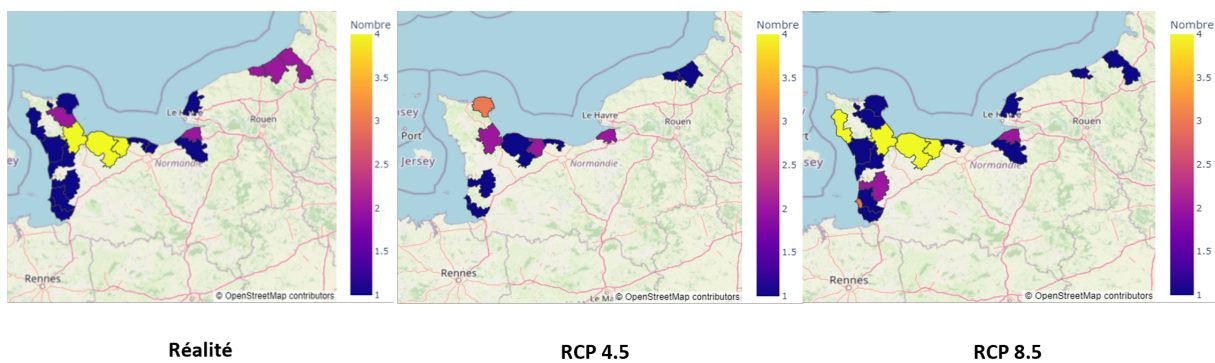
réalité. Pour le reste de la façade atlantique c'est le modèle 4.5 qui est le plus réaliste. En effet, la Nouvelle Aquitaine présente des arrêtés qui sont beaucoup plus présents dans les terres et que le modèle n'a pas su prédire. La meilleure représentation par le modèle 8.5 s'explique par son entraînement sur des événements plus extrêmes ce qui implique des critères plus sévères et donc une plus grande tendance à refuser la qualification que le modèle 4.5. La justification établissant qu'un arrêté soit reconnu relève en effet d'un jugement d'expert.

Pour les demandes acceptées, le modèle 8.5 est clairement plus performant que le 4.5. Cela s'applique à tout le littoral, puisqu'il arrive mieux à représenter même les événements à l'intérieur des terres. Toutefois, si localisation des événements est plutôt bien respectée, le modèle sous-estime la fréquence des arrêtés Cat-Nat. Sur la période de validation 536 jours de submersions ont été recensés. Le modèle 4.5 n'en identifie que 132 donc seulement 25% quand le modèle 8.5 en prédit 231. Le modèle 8.5 est donc relativement plus performant mais dans les deux cas la fréquence est largement sous-estimée au niveau global de la côte atlantique française. Il reste à noter une exception qui est que la pointe de la Bretagne présente un plus grand nombre d'évènements qu'en réalité pour les deux modèles.

À partir de ces premiers constats, il est possible d'affiner la sélection du modèle. Seulement le type 2 concernant les submersions marines effectivement reconnues sera conservé. Ce choix permet d'éviter de se retrouver avec des cantons qui peuvent être affectés par d'autres types d'inondation résultant d'une submersion marine tel que le débordement d'un cours d'eau.

À partir de la base sélectionnée, il est possible d'effectuer une analyse régionale. Cinq régions sont ainsi observables :

— La Normandie (28) :

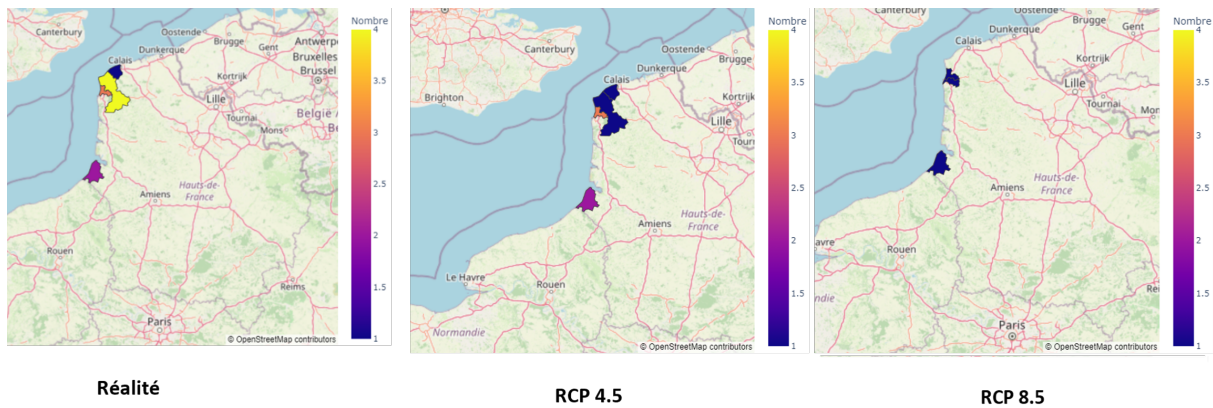


**Figure 57** – Comparaison des prédictions à la réalité par cantons en Normandie

Il est clair que les variables du scénario 8.5 sont plus adaptées à cette région. Notamment sur les cantons Carentan, Trévières et Bayeux qui sont parfaitement prédits avec 4 jours de submersions marines chacun.

En réalité on observe 34 jours de submersions marines dans cette région. Le modèle 4.5 n'en prédit que 17 soit 50% de moins et le modèle 8.5 33. Ce dernier est donc plus adapté géographiquement et au niveau du nombre d'évènements.

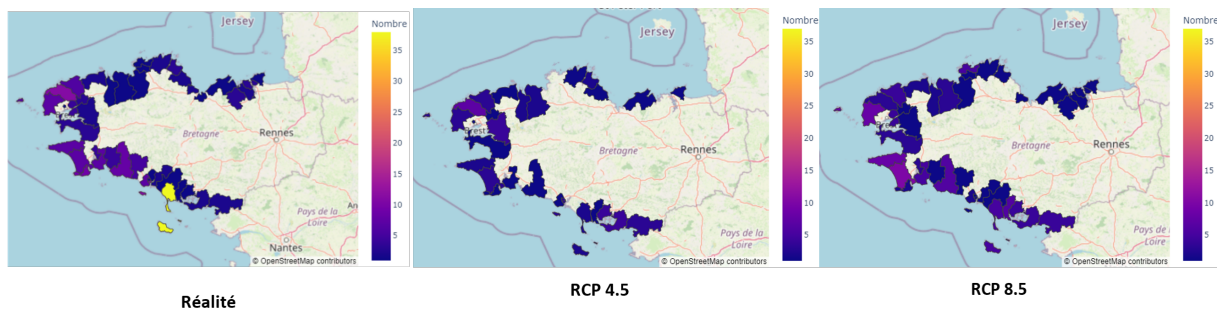
— Les Hauts de France (32) :



**Figure 58** – Comparaison des prédictions à la réalité par cantons dans Les Hauts de France

Dans le cas de cette région les deux modèles sont relativement peu performants. Le modèle 4.5 identifie légèrement mieux les zones impactées. 10 jours de submersions marines sont notés dans la région. Le modèle 4.5 en prédit 7 quand le modèle 8.5 n’en identifie que 3. Le modèle 4.5 est donc relativement le meilleur modèle pour cette région. Il s’agit de la seule région où ce modèle arrive plus proche de la réalité. Toutefois, cette région est celle qui compte le plus faible nombre de jours de submersion au global.

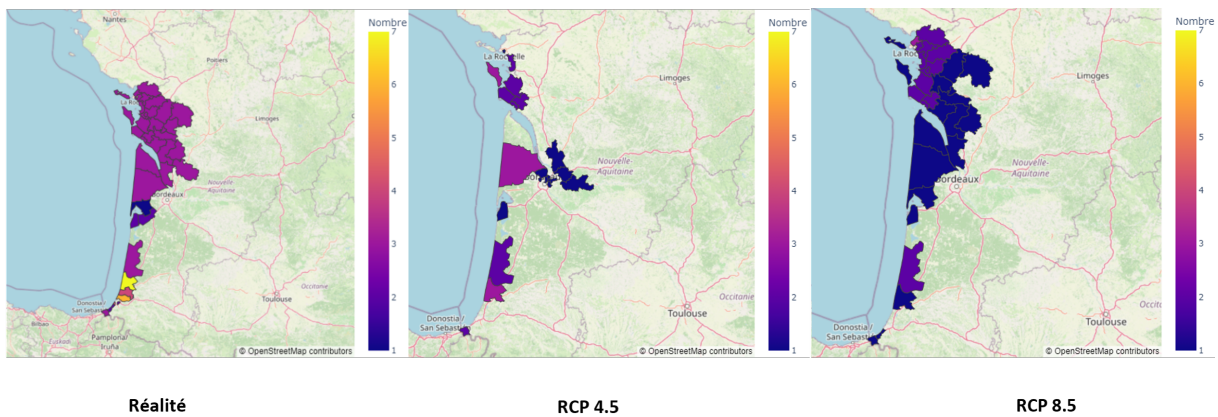
— La Bretagne (53) :



**Figure 59** – Comparaison des prédictions à la réalité par cantons en Bretagne

Une première observation est que la fréquence des submersions est très largement sous-estimée et tout particulièrement pour le modèle 4.5. Ce dernier ne voit que 63 jours soit 22% des 290 réalisés. Le modèle 8.5 est relativement meilleur mais sous-estime toujours en prédisant 132 jours. Les deux modèles ont également tendance à sur-affecter la pointe de la Bretagne, alors qu’en réalité la façade Sud est celle la plus affectée notamment avec le canton de Quiberon (38 jours).

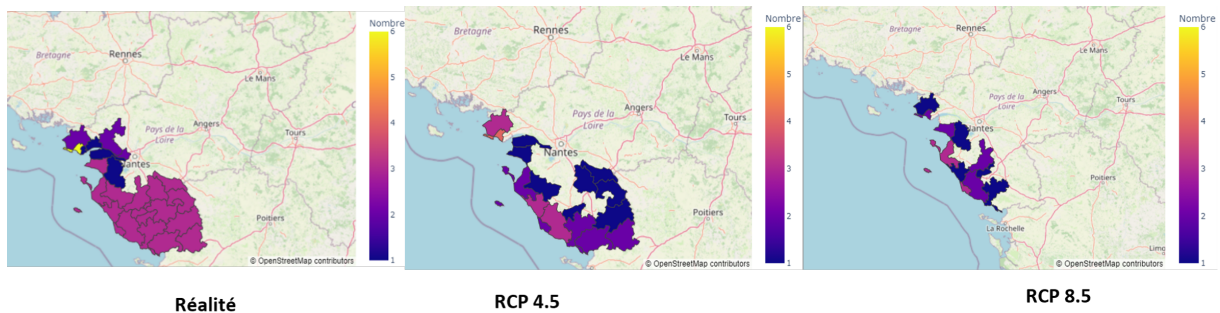
— La Nouvelle Aquitaine (75) :



**Figure 60** – Comparaison des prédictions à la réalité par cantons en Nouvelle-Aquitaine

Pour la région Nouvelle Aquitaine, le modèle 8.5 semble être le plus performant. Il l’est notamment sur les submersions impactant l’intérieur des terres, au Nord du Golfe de la Garonne. La fréquence est également globalement sous-évaluée par les deux modèles. Des 110 jours réalisés les modèles 4.5 et 8.5 n’en prédisent respectivement que 29 et 42, soit 23% et 38%. Il est à noter que le modèle 4.5 identifie correctement le canton Marensin-Sud comme un de ceux les plus impactés.

— Les Pays de la Loire (52) :



**Figure 61** – Comparaison des prédictions à la réalité par cantons en Pays de la Loire

Globalement les deux modèles sous-estiment la fréquence globale et la présence à l’intérieur des terres des submersions marines. 92 jours de submersion ont été recensé dans la réalité pendant cette période. Le modèle 4.5 en donne seulement 22 et le modèle 8.5 31, soit respectivement 24% et 34%. Le modèle 8.5 reste le plus efficace et identifie correctement le canton de Bole-Escoublac comme le plus impacté de la zone.

## 2.5 Comparaison des résultats

L’analyse des scénarios climatiques est un outil efficace d’aide à la décision stratégique face à l’aléa climatique. Pour se faire, il est essentiel de déterminer le niveau d’ambition de l’assureur et de choisir des scénarios climatiques pertinents. Afin d’être pertinents, ces scénarios devront être comparables aux scénarios existants à maille et horizon temporels équivalents. Cette étude porte alors sur la comparaison de deux scénarios issus des études du **GIEC** : les **RCP 4.5** et **8.5** à horizon 2050.

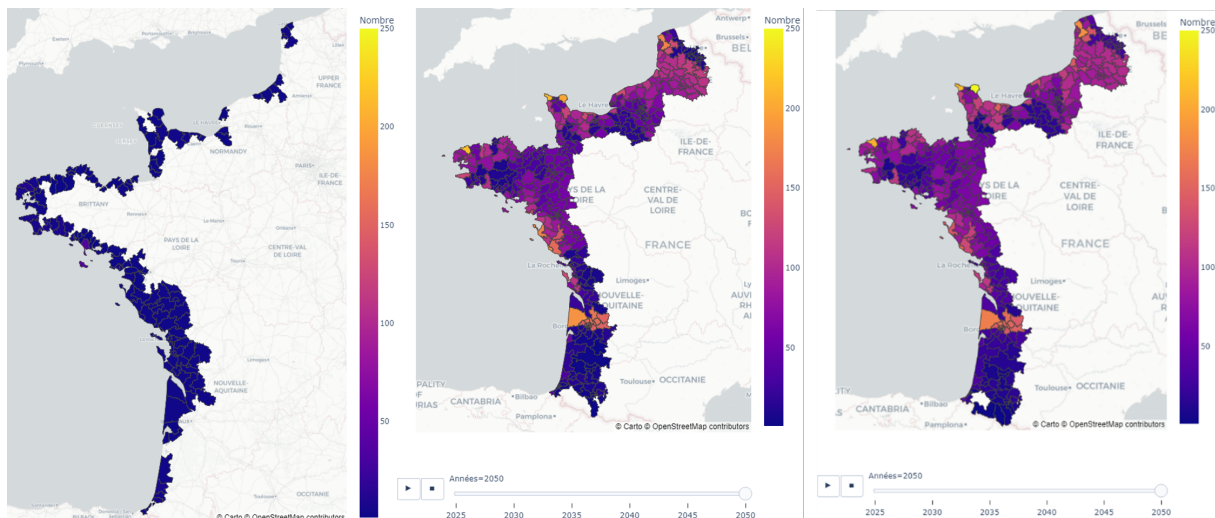


Cette dernière section porte sur la production de comparaisons des résultats de la méthodologie présentée avec ceux du marché. Cela permet alors de vérifier la démarche et la cohérence des modèles avec des résultats déjà publiés.

### 2.5.1 Présentation des résultats à l'horizon 2050

Comme présenté en section précédente, le modèle retenu contient uniquement les arrêtés Cat-Nat ayant été reconnus (CLASS 2). Les résultats portent donc sur la déclaration d'une submersion marine au sens strict : uniquement si un arrêté à été publié au Journal Officiel.

Les graphiques suivants permettent de représenter l'évolution du nombre de jours d'arrêtés submersions marines déclaré à la maille cantonale selon les RCP 4.5 et 8.5 à horizon 2050 :



**Figure 62** – Réalité connue à 2020 (1ère carte) par rapport aux prédictions à 2050 selon les RCP 4.5 (2ème figure) et RCP 8.5 (3ème figure)

Le premier constat est que l'on retrouve bien la cohérence entre les deux scénarios. En effet, le scénario 8.5 est un scénario pessimiste et qui se doit d'être une aggravation du scénario 4.5. C'est bien ce que l'on retrouve ici avec une augmentation du nombre de submersion marine journalière. De plus, on observe également que l'emprise inondée est plus étendue dans ce scénario pessimiste.

De manière générale, certains cantons peuvent paraître loin du littoral pour être exposés. En revanche, ces résultats sont révélateurs du choix effectué en début de modélisation de prendre les cantons de la première ligne des départements du littoral. Ce choix est également justifié par le souhait d'obtenir la démarche la plus prudente possible afin de ne pas sous-estimer le risque de submersion.

La modélisation du risque de submersion marine est un processus complexe qui implique de nombreux facteurs interdépendants. Dans cette étude, les conditions météorologiques et hydrodynamiques ont été prises en compte. En revanche, il est possible de rendre le modèle plus sophistiqué et précis en ajoutant des facteurs géologiques ou encore topographiques.

Dans cette étude, l'absorption des sols n'a pas été prise en compte. Or, les sols peuvent jouer un rôle important dans la propagation des eaux de submersions. Les sols sableux peuvent permettre une absorption rapide et une infiltration des eaux. Tandis que les sols argileux peuvent limiter l'infiltration et causer une stagnation des eaux. Ainsi, ne pas prendre en compte cette variable peut conduire à des

prévisions imprécises.

De même, la topographie est un facteur clé dans la modélisation du risque de submersion marine. Ne pas prendre en compte ce facteur peut sous-estimer les zones les plus vulnérables. En outre, les hauteurs et les pentes de terrains peuvent jouer un rôle important dans la propagation des eaux. Cela peut ainsi créer des zones de concentration des eaux et d'autres zones de moindre risque.

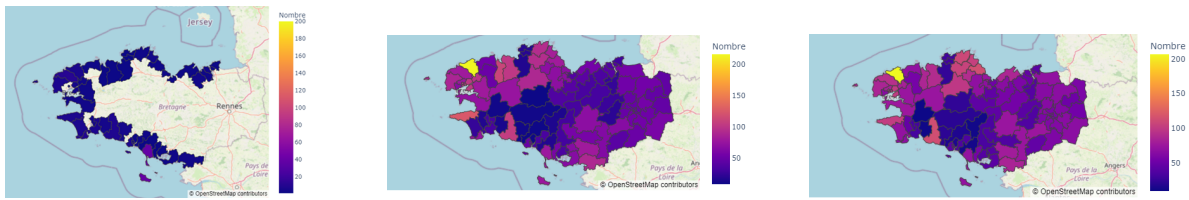
Les cantons les plus éloignés restent tout de même peu exposés au risque de submersion. Par ailleurs, une submersion marine peut avoir un impact plus loin dans les terres compte tenu du nombre d'affluents présents en France. En effet, lorsqu'une submersion marine se produit, les eaux peuvent se propager dans les bassins versants des cours d'eau et se diffuser dans les terres en suivant leur lit. Pour que cet effet de diffusion soit pris en compte, il serait intéressant de considérer les débits des cours d'eau et la pluviométrie potentielle dans les bassins versants. Les débits des cours d'eau dépendent des précipitations, de l'évapotranspiration et de l'écoulement de l'eau dans le bassin versant. La pluviométrie potentielle relève quant à elle de la quantité de précipitations qui peuvent tomber dans une zone donnée en fonction de son relief et de son climat.

De façon plus précise, on peut identifier plusieurs nouvelles dynamiques. Premièrement, les estuaires des grands fleuves français devraient être bien plus impactés qu'historiquement et ce dans les deux scénarios. Cela s'observe tout particulièrement sur celui de la Gironde, mais également pour ceux de la Seine et de la Loire. Deuxièmement, on peut remarquer que les zones affectées ne changent que peu entre les deux scénarios, seule la fréquence des événements diffère. On peut ainsi déduire de nouvelles zones qui pourraient être victimes de submersions marines. La région Haut-de-France se distingue ainsi particulièrement. Historiquement relativement peu affectée, elle pourrait devenir une de celles les plus touchées et ce profondément dans son arrière-pays. Une explication potentielle est la faiblesse du relief et un niveau moyen proche de celui de la mer exposant donc fortement à l'élévation future de son niveau. Troisièmement, les submersions marines pourraient avoir un impact important sur des zones pourtant assez éloignées du littoral. Ainsi, des cantons à la frontière de la région Île-de-France ou dans les marches des Pyrénées pourraient connaître des épisodes de submersions marines. La région Bretagne dans son entièreté serait quant à elle exposée quel que soit les scénarios réalisés. Plusieurs facteurs aggravants expliqueraient cette situation : une façade littorale au Nord, à l'Ouest ainsi qu'au Sud surexposant aux submersions marines et une faiblesse généralisée du relief empirant la situation.

Ces différentes observations amènent à penser que le coût des événements de type submersion marine devrait augmenter très fortement à l'avenir. L'Etat risque alors de devoir augmenter les cotisations permettant de financer ce système et/ou d'augmenter l'exigence des critères utilisés pour discriminer les événements. Ces réformes seraient indispensables pour pouvoir financer les reconstructions et de nouveaux ouvrages permettant de réduire l'impact des submersions.

Les cartes suivantes permettent de mieux appréhender les prédictions réalisées par les deux scénarios à horizon 2050. La comparaison à la situation historique permet de pouvoir représenter l'évolution au cours du temps de la probabilité de survenance d'une submersion. Afin de pouvoir mieux comprendre les graphiques du littoral atlantique (figure 60), une étude régionale a été entreprise :

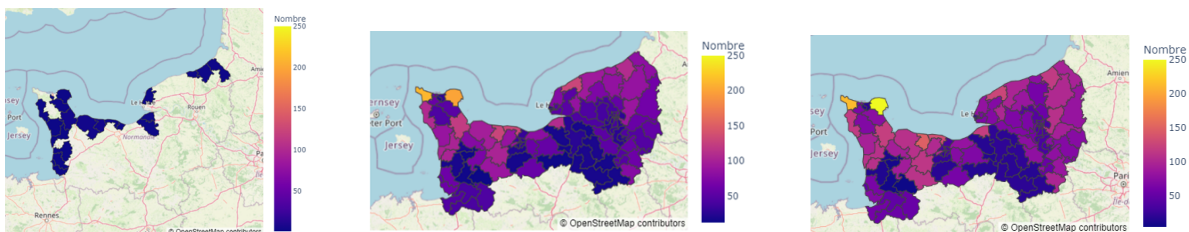
— La Bretagne (53) :



**Figure 63** – Comparaison de l'évolution de la fréquence des submersions par canton en Bretagne entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite)

En étant l'une des régions les plus exposée au risque de submersion, leur évolution est significative en Bretagne quel que soit le scénario RCP envisagé. Il est possible de remarquer une empreinte géographique plus étendue dans les projections. Ce constat est révélateur de la méthodologie choisie. En effet, même les cantons de l'intérieur des terres sont ainsi impactés. Par exemple, le canton de Vitré recense 47 jours de submersions pour le scénario 4.5 contre 54 avec le scénario 8.5, soit 15% de plus, alors qu'il n'était pas impacté en 2020. La répartition géographique est très similaire mais le scénario 8.5 comprend plus de jours cumulés de submersions (6335 contre 7895). Ces événements semblent plus concentrés sur la façade nord de la région par rapport à l'historique. Ainsi le canton de Lesneven devient le plus concerné avec plus de 200 sinistres pour les deux scénarios contre 10 en 2020. Le canton le plus touché de 2020 (Quiberon), passe ainsi de 35 à respectivement 78 et 85 jours cumulés en 2050 soit des augmentations respectives de 123% et 143% pour les scénarios 4.5 et 8.5.

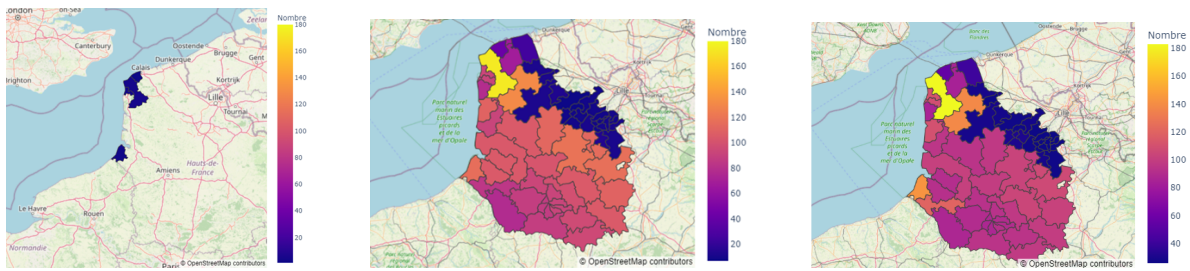
— La Normandie (28) :



**Figure 64** – Comparaison de l'évolution de la fréquence des submersions par canton en Normandie entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite)

Les cantons de l'intérieur des terres de la région Normandie sont également très impactés par les submersions marines. Le cours de la Seine est identifiable comme vecteur de cette propagation, mais plus généralement toute la région est impactée. Par exemple, le canton de Pont-Audemer recenserait respectivement 17 et 55 jours d'occurrence alors qu'il n'était pas concerné jusqu'à présent. La fréquence reste bien plus faible que sur les cantons du littoral. Pour les deux scénarios, les submersions devraient se concentrer sur les pointes du Cotentin. Le canton de Val-de-Saire compte ainsi respectivement 200 et 250 jours alors qu'il n'en avait qu'un en 2020. Ces dernières sont géographiquement très propices aux submersions car bordées sur plusieurs côtés par la Manche. Le scénario 8.5 envisage une fréquence plus élevée que le 4.5 culminant à 250 jours cumulés en une année contre 200 pour le scénario 4.5.

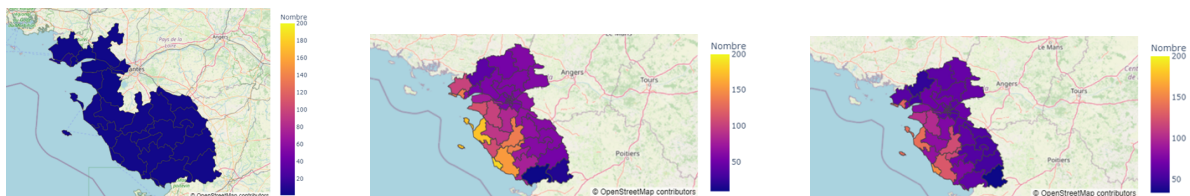
— Les Hauts-de-France (32) :



**Figure 65** – Comparaison de l'évolution de la fréquence des submersions par canton dans les Hauts-de-France entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite)

D'après l'occurrence historique des arrêtés de catastrophes naturelles de type submersion marine, la région des Hauts-de-France semble épargnée. En effet, elle n'a été que très peu impactée dans la base historique. Ce constat est très différent dans les cartes de projections. Si le principal canton touché (Desvres) ne change pas, tous les autres cantons pourraient être affectés à l'horizon 2050 quel que soit le scénario envisagé. Le canton de Desvres passe de 4 jours de sinistres à respectivement 174 et 183 jours selon les scénarios 4.5 et 8.5. Tandis que le canton de Friville-Escarbotin, précédemment vierge de tout dégât, compterait respectivement 116 et 143 jours de submersions. La fréquence des submersions pourrait ainsi s'accroître fortement, passant de moins d'une dizaine à plus d'une centaine selon les scénarios respectifs 4.5 et 8.5.

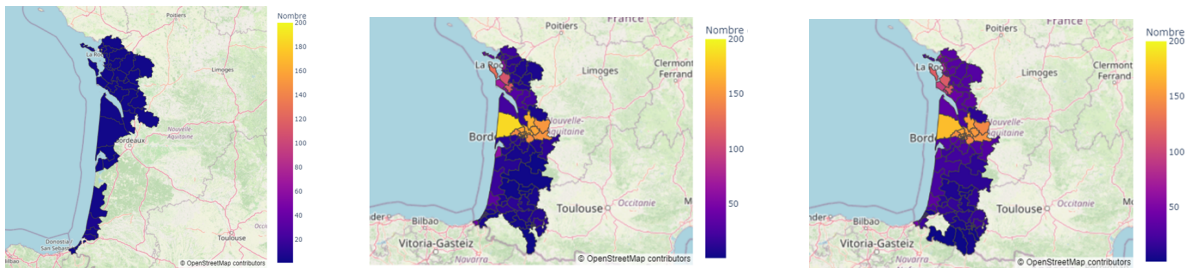
— Les Pays de la Loire (52) :



**Figure 66** – Comparaison de l'évolution de la fréquence des submersions par canton dans les Pays de la Loire entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite)

Similairement aux régions évoquées précédemment, les cantons non-côtiers de la région des Pays de la Loire sont très impactés dans les deux scénarios. Un élément diffère cependant des autres régions. En effet, le scénario 4.5 prédit ici un canton touché par plus de jours de submersions que le scénario 8.5. Ce dernier en cumule plus au globale mais les répartit plus sur l'entièreté du territoire. Le canton de la Baule Escoubac, le plus touché historiquement, passe de 6 jours à respectivement 138 et 134 jours. Les cantons de l'agglomération nantaise qui étaient in affectés jusqu'à présent pourraient être submergés respectivement 64 et 72 jours.

— La Nouvelle-Aquitaine (75) :



**Figure 67** – Comparaison de l'évolution de la fréquence des submersions par canton en Nouvelle-Aquitaine entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite)

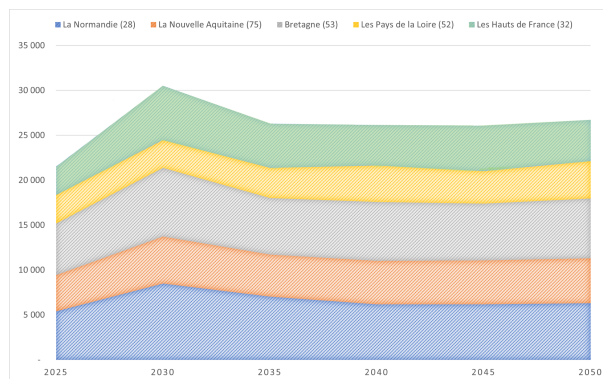
Dans le cas de la région Nouvelle-Aquitaine, la fréquence et le nombre cumulé des évènements sont en fort essor quel que soit le scénario. L'impact sur l'agglomération bordelaise est toutefois remarquable. Concentrant près d'un million d'habitants, elle pourrait être touchée significativement par les conséquences du changement climatique. Ainsi, les 5 cantons de la ville de Bordeaux présentent chacun respectivement près de 110 et 160 jours de submersions contre 0 en 2020. La présence de la Gironde et la faiblesse de l'élévation de ces cantons étant des facteurs facilitant la propagation de ces catastrophes. Toutefois, même les cantons pyrénéens pourraient être impactés par ces évènements dans le scénario 8.5. Par exemple, le canton de la Montagne Basque est impacté 1 fois contre 0 pour le scénario 4.5 et l'historique. Cela reste toute de même exceptionnel et peu probable.

### 2.5.2 Evolution des résultats au cours des années jusqu'à 2050

Afin de mieux comprendre la tendance des résultats évoqués précédemment, il est important de suivre l'évolution de l'occurrence des submersions au fil des années. Pour se faire, les scénarios 4.5 et 8.5 de 2025 à 2050 avec un pas de temps de 5 ans seront analysées aussi bien sur l'aspect géographique que fréquentiel.

#### 2.5.2.1 Scénario 4.5

Le scénario 4.5 suppose une stabilité des émissions de gaz à effet de serre. Les projections envisagées sont représentées dans le graphique ci-dessous :



**Figure 68** – Évolution du nombre de jours d'arrêtés de submersion au cours du temps pour le scénario 4.5

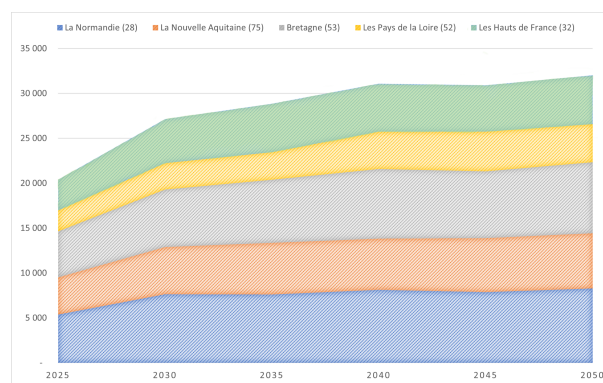
Après un pic en 2030, suivi d'une réduction du nombre global d'évènements, la tendance reste très stable entre 2030 et 2050. Le nombre total de jours de submersions augmente ainsi de 24% sur toute la période contre 44% entre 2025 et 2030. Ce pic peut s'expliquer par le caractère aléatoire des paramètres climatiques qui présentent ainsi un impact disproportionné pour cette année particulière pour ce scénario.

En effectuant une analyse régionale, il est possible de constater l'existence de deux groupes distincts. La Normandie et la Bretagne cumulent de loin le nombre d'occurrence le plus important, respectivement 6043 et 5641 en 2050. Toutefois, ce dernier connaît une croissance plus faible par rapport à l'ensemble du littoral. Les trois autres régions restent globalement bien moins impactées, cumulant environ 4000 jours chacune, mais connaissent une croissance du nombre de jours de submersion bien plus importante.

Pour des raisons opérationnelles, un pas de temps de 5 ans a été appliqué. Cette contrainte, peut donner l'impression d'un lissage des courbes et gagnerai à avoir une granularité plus précise.

### 2.5.2.2 Scénario 8.5

Le graphique ci-dessous permet d'observer l'évolution du nombre cumulé de jours de submersions marines pour chacune des régions du littoral français. Il s'agit des résultats du scénario le plus pessimiste :



**Figure 69** – Évolution du nombre de jours d'arrêtés de submersion au cours du temps pour le scénario 8.5

Une première remarque est la tendance commune à la hausse du nombre de ces évènements sur la période 2025 - 2050. Au niveau global, le passage de 2025 à 2030 est le plus significatif avec un bond de 33% du nombre de jours impactés. La croissance reste importante, bien que moindre, entre 2030 et 2040 atteignant 15%. Sur la période 2040 - 2050 une stabilisation du nombre de jours impactés apparaît. Toutefois, de 2025 à 2050 une hausse de 57% est observée sur l'ensemble du littoral.

En effectuant une analyse régionale, les Pays de la Loire (52) apparaissent comme la zone connaissant la plus importante croissance du nombre d'évènements, avec une hausse de 82%. Elle reste toutefois celle la moins impactée au total. En effet, le classement des régions se maintient entre 2025 et 2050. La région Normandie reste celle la plus impactée, malgré une hausse de 54% inférieure à celle du littoral étudié pris dans sa globalité. La Bretagne conserve également sa deuxième place et l'écart de plus de 1500 jours avec la région suivante est remarquable. Il est évident que ces régions sont et resteront disproportionnellement affectées par ces évènements.

## 2.5.3 Études de la CCR

Le **GIEC** a estimé comme probable l'augmentation de la fréquence de submersion marine à la suite de la hausse du niveau de la mer. Ce constat a mené les études à se multiplier. En outre, afin de pouvoir estimer l'impact des scénarios du **GIEC**, la **CCR** a mené en partenariat avec Météo-France deux études.

### 2.5.3.1 Étude de 2015

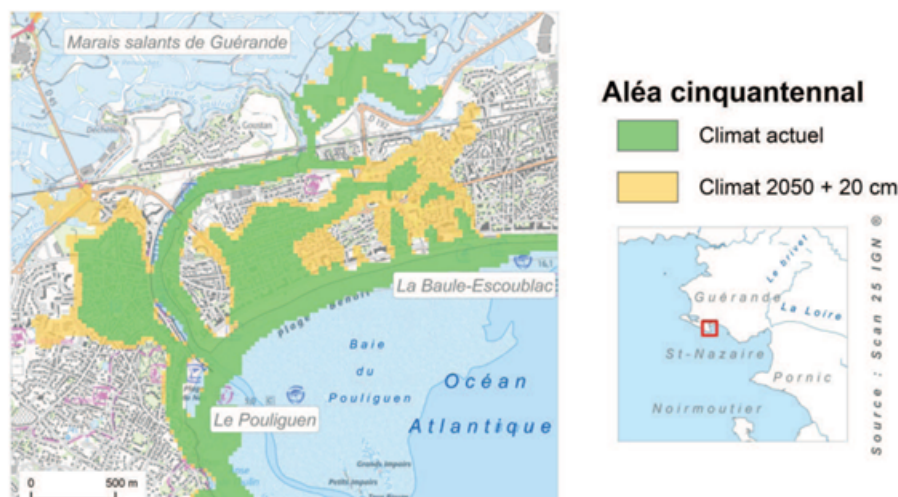
Cette première étude, réalisée en 2015<sup>76</sup>, portait sur le scénario dit « BaU : Business as Usual » ou **RCP** 4.5. Il s'agit donc d'une poursuite de l'augmentation des émissions de **GES** au rythme actuel.

A noter que le périmètre de l'étude de 2015 ne portait que sur le littoral atlantique français pour lequel les modélisations ARPEGE ont été réalisées en priorité. Ce périmètre est donc bien comparable à celui de ce mémoire.

L'étude de 2015 a porté sur trois aspects de l'exposition à horizon 2050 du littoral atlantique français.

D'une part, l'effet de l'évolution du climat n'a pas montré de tendances à un renforcement de la fréquence des submersions marines. Puis, l'impact de l'élévation du niveau de la mer (20 cm) a été beaucoup plus visible avec une augmentation significative des dommages. D'après l'étude, « Ce phénomène est particulièrement vrai dans les secteurs où le relief est le moins marqué ».

L'étude ne propose pas de carte de la France spécifique aux submersions marines mais la figure suivante est disponible :

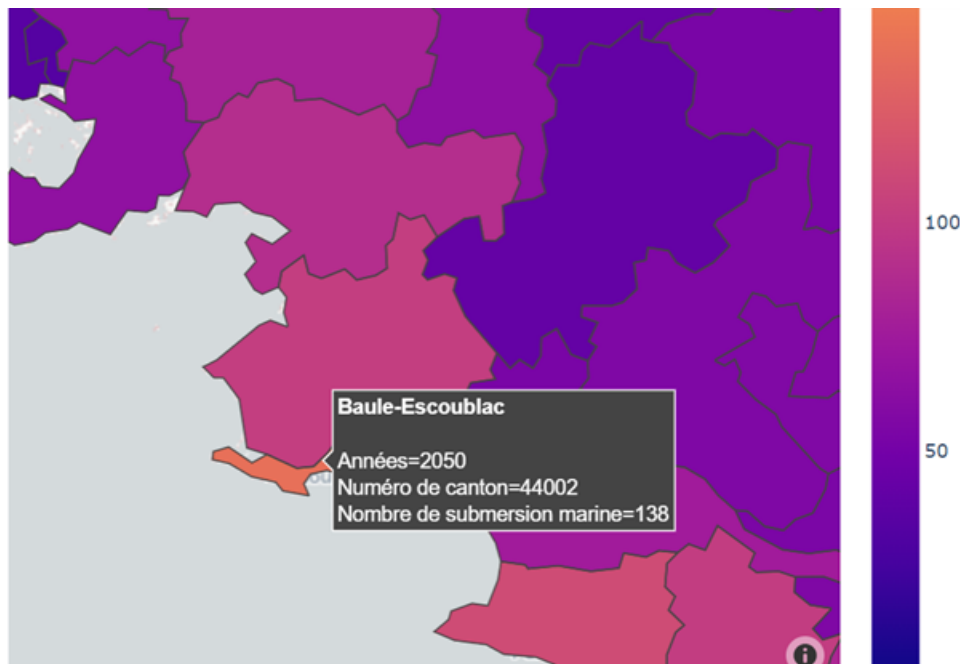


**Figure 70** – Illustration de l'aléa cinquantennal dans le secteur de la baie du Pouliguen (44), Source : figure 12 de l'étude **CCR** 2015

A partir du scénario établi, il est possible de regarder la commune de Pouliguen dans le département de la Loire-Atlantique. Cette commune est représentée par le canton 44002 mais il n'est pas possible de faire une comparaison au même niveau de précision que l'étude de la **CCR**. En revanche, il s'agit bien d'un canton impacté par les submersions marines de façon significative à horizon 2050 selon le scénario

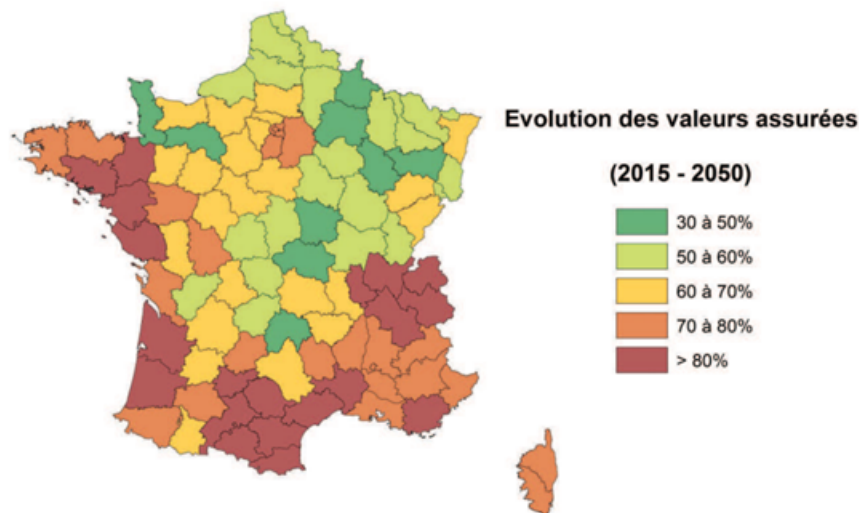
76. Lien : [\[en ligne\]](#)

## RCP 4.5.



**Figure 71** – Représentation du canton contenant la commune du Pouliguen d’après le scénario RCP 4.5

Même s’il n’est pas possible d’avoir une carte spécifique aux submersions marines dans cette étude, il est possible de vérifier la cohérence générale des résultats sur le littoral atlantique français. En outre, l’étude de 2015, fournie la figure suivante :



**Figure 72** – Cartographie de l’évolution des valeurs assurées par département entre 2015 et 2050, Source : figure 7 de l’étude CCR 2015

Les départements<sup>77</sup> les plus impactés sont bien représentés par le scénario. En revanche, le niveau de détail ne permet pas de pouvoir définir la relation directe avec la submersion marine. L’étude de 2018 est donc utilisée pour permettre une meilleure comparaison.

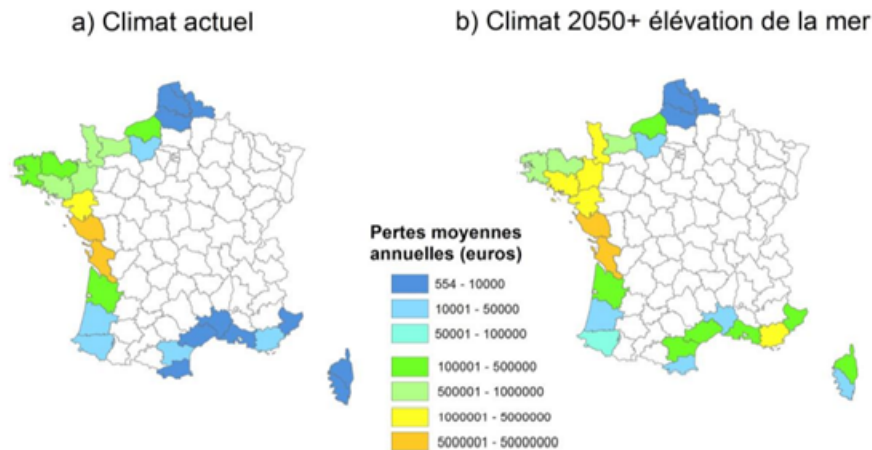
77. Une carte des départements français peut être trouver en Annexe L.



### 2.5.3.2 Etude de 2018

La seconde étude, réalisée en 2018<sup>78</sup>, portait sur le scénario le plus pessimiste ou RCP 8.5. L'une des principales conclusions de cette dernière est la prévision d'une augmentation de 82% des submersions marines.

La figure ci-dessous provient de l'étude de 2018 et montre de fortes disparités territoriales :



**Figure 73** – Impact du climat et de l'élévation de la mer sur les pertes annuelles moyennes, Source : figure 23 de l'étude RCP 2018

D'après la figure ci-dessus, les départements du littoral atlantique les plus affectés par l'élévation du niveau de la mer sont la Vendée et la Charente-Maritime<sup>79</sup>. Il est également possible de remarquer que les départements suivants : la Manche, l'Ille-et-Vilaine, le Morbihan, les Cotes-d'Armor, le Finistère et les Pyrénées-Atlantiques subissent une augmentation significative de leurs pertes.

Afin de pouvoir comparer les résultats de cette étude avec la modélisation présentée, il est nécessaire de regarder les cantons par rapport à leur localisation départementale. A partir de la cartographie du scénario RCP 8.5, il est possible de remarquer que certains cantons sont particulièrement exposés aux submersions marines. La liste suivante permet de souligner les cantons les plus affectés par département :

- **Le Pas-de-Calais** : Desvres (153), Lumbres (100)
- **La Manche** : Hague (204), Val de Saire (246)
- **Le Finistère** : Lesneven (207)
- **La Vendée** : Ile d'Yeu (123), Sables- d'Olonne (157)
- **Les Côtes-d'Armor** : Paimpol (120)
- **La Charente-Maritime** : Ile d'Oleron (113)
- **La Gironde** : Sud-Médoc (157)

Remarque : les chiffres entre parenthèses correspondent au nombre de submersion marine journalière à horizon 2050 pour le scénario 8.5.

Les résultats trouvés sont dans l'ensemble cohérents avec les résultats de l'étude de la CCR et Météo France. A noter que l'on ne retrouve pas l'Ille-et-Vilaine et le Morbihan parmi les départements les

78. Lien : [en ligne](#) p21 partie 4.3 « La submersion »

79. Zones touchées par la Tempête Xynthia

plus exposés, au profit du Pas-de-Calais et de la Gironde. Le Pas-de-Calais constitue la véritable différence de résultats avec l'étude de la CCR qui apparaît comme peu exposé.

### 2.5.4 Outil de visualisation en ligne du BRGM

Le BRGM a mis en place une méthode permettant de choisir la meilleure approche possible pour évaluer la submersion marine à une échelle régionale. Cette méthode est fonction des spécificités de différents types de côtes exposés à l'aléa (Garcin et al., 2012). La méthodologie se base sur la plus haute mer astronomique. Ce niveau correspond au niveau le plus élevé que la mer peut atteindre lors d'évènements astronomiques favorables. Ce niveau est déterminé par interpolation linéaire entre les divers capteurs dispersés dans les différents ports de la côte française.

A ce niveau préalable, il convient d'ajouter l'effet du changement climatique sur la hausse du niveau de la mer ainsi que celui de la surcote. Cette dernière peut atteindre les 4 mètres ce qui représente un impact important en fonction des caractéristiques topographiques du littoral. Ces dernières se basent sur des mesures à haute résolution réalisées par l'IGN. En fonction des paramètres entrés par l'utilisateur sur le niveau de la marée, le site calcule l'impact par rapport aux données topographiques fournies par IGN.

La carte suivante<sup>80</sup> permet de représenter ces zones les plus exposées à la submersion marine :



**Figure 74** – Zones exposées aux submersions marines, Source : BRGM

Avant toute chose, il apparaît opportun de commencer par une remarque préliminaire. Il est possible de remarquer que la visualisation de la carte du BRGM prend en compte le niveau de la mer. En revanche, il n'est malheureusement pas possible de choisir une valeur entre 0 et 0.5m. D'après le GIEC, l'élévation du niveau de la mer devrait être à horizon 2050 d'une vingtaine de centimètres. Un paramètre à 0.5m par rapport à 2016 a donc été choisi ici par défaut. Il n'est donc pas possible de faire de comparaison entre les scénarios RCP 4.5 et 8.5 du fait de ce curseur.

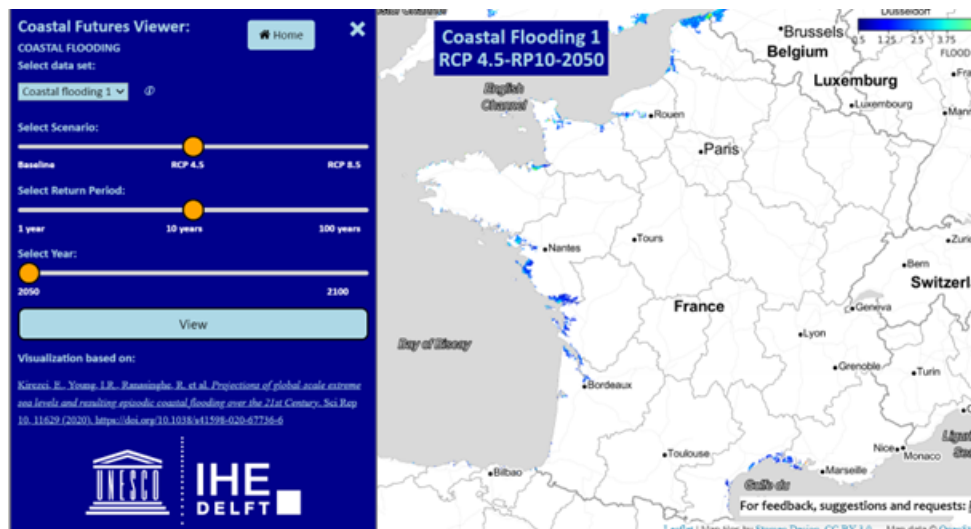
80. Source : [\[en ligne\]](#)

Toutefois, cette carte permet de remarquer que contrairement à l'étude de la CCR, les régions du Nord-Pas-de Calais et de la Gironde sont également bien présentes comme zones à risques.

Finalement, le modèle présenté apparaît comme pertinent avec les résultats du BRGM. En d'autres termes, les cantons les plus touchés par les submersions marines sont bien présents dans la carte du BRGM.

### 2.5.5 Outil de visualisation en ligne *Coastal Futures*

Il est possible de pallier au manque de distinction sur les scénarios RCP 4.5 et 8.5 de l'outil en ligne précédemment utilisé. En effet, un autre outil<sup>81</sup> fondé sur un travail de recherche<sup>82</sup> permet de visualiser les submersions marines dans le monde. *Coastal futures* est un visualisateur interactif en ligne qui combine plusieurs ensembles de données publiées sur les projections du 21e siècle des facteurs d'impact climatique côtier évalués par un groupe de travail du sixième rapport d'évaluation (AR6) du GIEC. En d'autres termes, il permet aux utilisateurs de voir les résultats de l'impact des submersions marines. Cet outil a été créé et est entretenu par l'IHE Delft, aux Pays-Bas par l'intermédiaire de l'équipe CoFu dans le cadre d'une collaboration entre le département des risques et de la résilience des zones côtières et urbaines, et l'Institut de recherche sur le développement et l'urbanisme, et le département d'hydro-informatique et d'innovation socio-technique.



**Figure 75** – Zones exposées aux submersions marines à horizon 2050 selon le scénario RCP 4.5, Source : Coastal futures

81. Source : [\[en ligne\]](#)

82. Source : [\[en ligne\]](#)

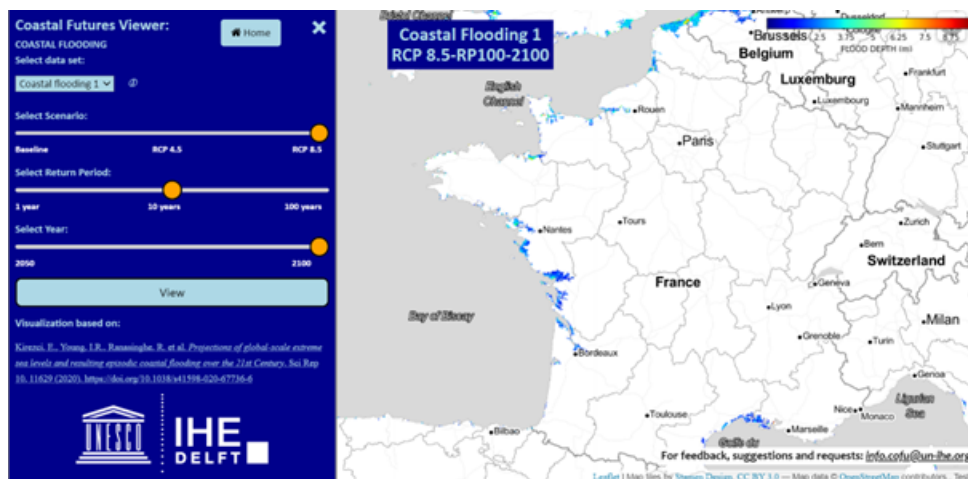


**Figure 76** – Zones exposées aux submersions marines à horizon 2050 selon le scénario RCP 8.5, Source : Coastal futures

Il est possible de remarquer que deux jeux de données sont disponibles sur l’outil en ligne. En revanche, la composition de ces deux jeux n’est pas précisée et aucune différence n’est notée. Les résultats de ce deuxième jeu de données sont présents en annexe M.

Un premier constat apparaît sur ces cartes faisant la distinction entre scénario RCP 4.5 et 8.5. A horizon 2050, les zones touchées par les submersions sont les mêmes. C’est également le constat global qu’il est possible de faire à partir des projections réalisées sur la modélisation présentée. Ce constat vient du fait que l’horizon de projection reste de court terme et les trajectoires ne sont pas sensiblement différentes à cet horizon. Il pourrait être intéressant d’effectuer cet exercice à plus long terme tel que l’horizon 2100.

En revanche, d’après l’outil en ligne, les résultats restent similaires :



**Figure 77** – Zones exposées aux submersions marines à horizon 2100 selon le scénario RCP 8.5, Source : Coastal futures

Finalement, tout comme le montrait l’outil en ligne précédent, le Nord-Pas-de-Calais et la Gironde restent des départements affectés par les submersions marines. Les résultats de ces deux derniers outils en ligne sont similaires. Ainsi, la modélisation présentée couvre bien les mêmes zones à risque de submersion marine. En revanche, aucune indication sur la fréquence des arrêtés n’est disponible dans ces deux outils en ligne.

## Conclusion

L'ACPR note une prise de conscience générale et souligne le degré de maturité très hétérogène selon les institutions sur la prise en compte du changement climatique et les améliorations méthodologiques nécessaires. À ce stade, les exercices entrepris par les acteurs du marché assurantiel ont des objectifs exploratoires (*learning exercise*). En effet, il n'y a pas encore de consensus du marché sur la prise en compte du réchauffement climatique. Ces exercices ont la vertu de participer à la mise en place de bonnes pratiques émergentes.

La submersion marine est l'une des conséquences majeures du changement climatique et de l'élévation du niveau de la mer. À ce titre, son étude mérite une attention particulière puisque les côtes seront de plus en plus impactées. L'objectif principal de ce mémoire était de pouvoir contribuer aux connaissances actuarielles du phénomène de submersion marine. Dans le but de maîtriser les enjeux inhérents, des outils et méthodes de construction ont dû être appliqués. Ainsi, à partir de modèles climatiques académiques, la fréquence du risque de submersion marine a pu être étudiée à horizon 2050 pour les scénarios du GIEC.

Une autre amélioration envisageable serait d'utiliser non plus les scénarios RCP mais les nouveaux scénarios SSP. Cette mise à jour des scénarios et des observations utilisées permettrait alors la prise en compte explicite des facteurs économiques, technologiques, démographiques et institutionnels.

Ce mémoire a pour vocation de proposer une méthodologie complète et flexible à l'aide de modèles de *machine learning*. Cette méthodologie permet de réaliser des prédictions de submersions marines en fonction des cantons du littoral atlantique français à horizon 2050. L'une des principales difficultés rencontrées est le manque de données climatiques « clés en main ». Ce constat explique que les prédictions ont dû être réalisées à partir de trois sources de données en libre accès sur internet.

Le modèle présenté a été construit à partir des cantons appartenant à la première ligne des départements du littoral. Certains cantons sont éloignés de la côte mais ont des fréquences de survenance d'arrêtés Cat-Nat de type submersion marine importantes. L'absence de données telles que la topographie ou encore l'absorption des sols sont autant d'effets à prendre en compte pour résoudre certains biais causés par le modèle.

Par ailleurs, il serait également intéressant de prendre en compte la diffusion possible des eaux dans les terres via les cours d'eau. Les cours d'eau peuvent agir comme canaux pour les eaux de submersion, les conduisant loin de la côte et les faisant remonter les vallées. Cela peut alors être créateur de zones d'inondations importantes.

Cependant, il est important de noter que les modèles sont limités par la précision des données disponibles et des hypothèses utilisées pour estimer la propagation des eaux. Il est donc difficile de prendre en compte les changements climatiques inhérents tels que les modifications du relief et des cours d'eau.

Chaque limite constitue autant de pistes d'amélioration envisageables. De nombreux points d'améliorations résident dans l'utilisation de données plus précises. Il est par exemple possible d'imaginer l'utilisation de sources de données complémentaires de modèles CAT avec un maillage plus précis, ou d'envisager un modèle à une échelle non plus régionale mais locale. D'autant plus qu'une résolution plus fine permettrait d'identifier des ouvrages de protections présents sur le littoral. Il est possible par exemple de citer les structures naturelles (dunes) et artificielles (digues).

Par ailleurs, les variables pourraient avoir une disponibilité temporelle plus importante. En outre, les variables climatiques issues de la base COPERNICUS ne vont que jusqu'à 2005. De plus, la maille cantonale choisie est plus large que la précision demandée par les assureurs (adresse). En revanche, la modélisation utilisée repose sur des modèles RCM, seuls modèles disponibles en open source. Cette source ne présente pas à ce jour de maille exploitable plus fine du fait de la puissance de calcul nécessaire à sa réalisation. Ce constat relève certes une limite mais aussi une opportunité au développement de modèles plus fins.

Malgré les bons résultats obtenus sur le meilleur modèle de *machine learning* (XGB), le caractère déséquilibré de la base reste un frein au modèle. En effet, compte tenu de cet événement extrême très rare, il est difficile de ne pas avoir une prépondérance de l'événement « aucune submersion marine n'a été déclarée ». L'ensemble de ces éléments fait qu'il est difficile de prédire la submersion marine avec un modèle général de *machine learning*. Il pourrait être intéressant de construire un modèle personnalisé qui est créé spécialement pour ce phénomène. Il pourrait être envisageable d'utiliser un réseau de neurones par exemple. Un tel modèle nécessite des données précises comme évoquées dans le paragraphe précédent.

Par ailleurs, ce mémoire traite exclusivement du module aléa. Il pourrait être intéressant à l'aide d'un portefeuille de données d'un assureur de pouvoir traiter des autres modules CAT (exposition et vulnérabilité). Cet élément constitue une limite car il s'agit de la propriété des assureurs.

La cartographie réalisée permet d'être un outil d'aide à la décision et d'avoir les premières orientations de cet événement. A l'échelle de la France, il apparaît que les départements les plus touchés par la tempête Xynthia de 2010 (Charente-Maritime et Vendée) restent très exposés au risque de submersion marine. En effet, la topographie relativement plane de certaines zones côtières conduit à délimiter des emprises étendues exposées au risque de submersion. De plus, le modèle met en évidence que les départements du Pas-de-Calais, de la Manche et du Finistère sont particulièrement exposés au risque de submersion marine. Ce constat est cohérent avec les cartes diffusées par d'autres acteurs. D'autant plus lorsque l'on sait que le Nord Pas-de-Calais et la Normandie sont proies à des glissements ou effondrement des côtes à falaises. Et que la Vendée et les landes sont également victimes de phénomènes d'érosion importants.

Ces conclusions sont d'autant plus importantes au regard de la dernière conférence annuelle de l'ACPR le 05 décembre 2022. L'un des sujets traités était « l'exposition des acteurs financiers au risque climatique : un état des lieux ». Lors de cet événement l'ACPR a rappelé que les difficultés à surmonter pour prendre en compte le risque climatique tels que le manque de données ou les difficultés de modélisation ne doivent pas être un frein. Il est donc important que chaque acteur du secteur assurantiel puisse entreprendre des travaux dans le cadre d'un processus itératif d'amélioration continue.

### 3 Références

- [1] **AKBANI, R., KWEK, S. JAPKOWICZ, N.**, "Applying support vector machines to imbalanced datasets", *Machine Learning : ECML*, vol. 3201, Springer, Berlin, 2004, p.39-50. [\[en ligne\]](#)
- [2] **BAMBER J. L., ASPINALL W. P.**, "An expert judgement assessment of future sea level rise from the ice sheets", *Nature Climate Change*, vol. 3, Nature, Bristol, 2013, p. 424-427. [\[en ligne\]](#)
- [3] **BELLOTTI, T., CROOK, J.**, "Support vector machines for credit scoring and discovery of significant features", *Expert Systems with Applications*, vol.36, Elsevier, England, 2009, p.3302-3308. [\[en ligne\]](#)
- [4] **BISHOP C.M.**, *Pattern Recognition and Machine Learning*, chap.7.2, Springer, Cambridge, 2006, p.345-357. [\[en ligne\]](#)
- [5] **ANDRÉ C.**, *Analyse des dommages lies aux submersions marines et evaluation des couts induits aux habitations a partir de donnees d'assurance : perspectives apportees par les tempetes Johanna (2008) et Xynthia (2010)*, Géographie, Université de Bretagne occidentale, Brest, 2013. [\[en ligne\]](#)
- [6] **CAISSE CENTRALE DE RÉASSURANCE, MÉTÉO FRANCE**, *Conséquences du changement climatique sur le cout des catastrophes naturelles en France à horizon 2050*, CCR, Paris, 2018. [\[en ligne\]](#)
- [7] **CEREMA** *Changement climatique : adapter les territoires littoraux*, L'Essentiel, Cerema, Paris, 2021. [\[en ligne\]](#)
- [8] **CEREMA**, *Submersions marines : mieux comprendre pour s'en protéger*, L'essentiel, Cerema, Paris, 2017. [\[en ligne\]](#)
- [9] **DELOITTE**, *Risque climatique : l'appréhender en pratique dans l'assurance*, Paris, 2021. [\[en ligne\]](#)
- [10] **FISCHER E.M., SIPPEL S., KNUTTI R.**, "Increasing probability of record-shattering climate extremes", *Nature Climate Change*, vol.11, Nature, Bristol, 2021, p.689-695. [\[en ligne\]](#)
- [11] **JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R.**, *An Introduction to Statistical Learning : with Applications in R*, Springer, 2013. [\[en ligne\]](#)
- [12] **GARNER, A. J. et al.**, "Impact of climate change on New York City's coastal flood hazard : Increasing flood heights from the preindustrial to 2300 CE", *Proceedings of the National Academy of Sciences*, vol. 114, University of Victoria, Victoria, 2017. [\[en ligne\]](#)
- [13] **GEORISQUES.GOUV**, *Inondations ; s'informer pour mieux se protéger*, [\[en ligne\]](#)
- [14] **EL HASSANI H.**, *Modélisation stochastique des inondations en France et applications en réassurance*, ENSAE Paristech - 22 février 2017. , [\[en ligne\]](#)
- [15] **HAN, J., KAMBER, M.**, *Data Mining : Concepts and Techniques*, 3rd Edition, Morgan Kaufmann Books, Amsterdam, 2011, [\[en ligne\]](#)
- [16] **JACKSON L. P. JEVREJEVA S.**, "A probabilistic approach to 21st century regional sea-level projections using RCP and High-end scenarios", *Global and Planet Change*, vol. 146, Elsevier, Amsterdam, 2016, p.179-189. [\[en ligne\]](#)
- [17] **LEGIFRANCE.GOUV.FR**, *Loi n° 82-600 du 13 juillet 1982 relative à l'indemnisation des victimes de catastrophes naturelles* [\[en ligne\]](#)
- [18] **MÉTÉO FRANCE.GOUV**, *les vagues submersions*, 2 mars 2020. [\[en ligne\]](#)
- [19] **MÉTÉO FRANCE.GOUV**, *Tempêtes en France métropolitaine*, 14 mars 2019. [\[en ligne\]](#)
- [20] **RAKOTOMALALA, R.**, *Cours « Gradient Boosting »*, Université Lumière Lyon 2. [\[en ligne\]](#)
- [21] **ROKACH, L., MAIMON, O.**, *Data Mining and Knowledge Discovery Handbook*, Springer-Verlag, New York, 2005, p.165-192. [\[en ligne\]](#)

- [22] **SMOLA A. J., SCHÖLKOPF B.**, “A tutorial on support vector regression”, *Statistics and Computing*, vol.14, Kluwer Academic Publishers, Amsterdam, 2004, p.199-222. [\[en ligne\]](#)
- [23] **TIOGUIM I., DELCAILLAU D.**, *Modélisation d’extrêmes de séries temporelles : une étude empirique*, EURIA-Euro-Institut d’Actuaria, Bretagne, 2018. [\[en ligne\]](#)
- [24] **PRÉVOT I.**, *Projet de gestion géographique des risques*, Université Paris Dauphine, 2015. [\[en ligne\]](#)
- [25] **QUANTIN A.**, *Modélisation du péril inondation : passage d’une approche déterministe à une approche probabiliste en vue d’une tarification à l’exposition des couvertures de réassurance dans le cadre du régime des Catastrophes Naturelles*, Centre d’Études Actuarielles, 2011. [\[en ligne\]](#)
- [26] **VIE-PUBLIQUE.GOUV**, *Rapport 2022 du GIEC ; une nouvelle alerte face au réchauffement climatique*, 1 mars 2022. [\[en ligne\]](#)
- [27] **VITOUSEK S., BARNARD P. L., FLETCHER C. H., FRAZER N., ERIKSON L. STORLAZZI C. D.**, “Doubling of coastal flooding frequency within decades due to sea-level rise” *Scientific Reports*, vol. 1399, Nature, England, 2017. [\[en ligne\]](#)
- [28] **VOUSDOUKAS M., MENTASCHI L., MONGELLI I., CISCAR J-C., HINKEL J., WARD P., GOSLING S., FEYEN L.**, *Adapting to rising coastal flood risk in the EU under climate change*, Publications Office of the European Union, Luxembourg, 2020. [\[en ligne\]](#)
- [29] **VOUSDOUKAS M., MENTASCHI L., VOUKOUVALAS E., VERLAAN M. FEYEN L.**, “Extreme sea levels on the rise along Europe’s coasts”, *Earth’s Future*, Vol. 5, United States, 2017, p.304-323. [\[en ligne\]](#)
- [30] **VOUSDOUKAS M., MENTASCHI L., VOUKOUVALAS E., VERLAAN M., JACKSON L. P. FEYEN L.**, “Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard”, *Nature Communications*, vol.2360, 2018. [\[en ligne\]](#)
- [31] **WWF**, *GIEC : un troisième rapport qui doit sonner comme une inspiration pour le ou la futur.e président.e* – communiqué de presse, 4 avril 2022. [\[en ligne\]](#)

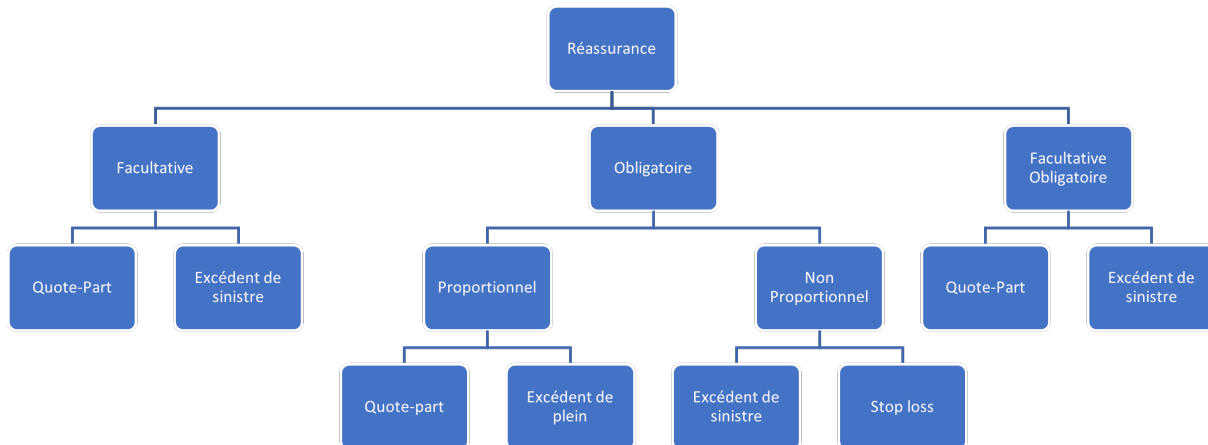


## Annexes :

### Annexe : Rappels généraux sur la réassurance

D'après Picard et Besson, « Une opération de réassurance est un contrat sur lequel un réassureur (dit cessionnaire) vis-à-vis d'un assureur professionnel (dit cédant) qui répond seul et intégralement vis-à-vis des assurés des risques par lui assurés, prend en charge moyennant rémunération, tout ou partie des sommes dues ou versées aux assurés à titre de sinistres ».

Le graphique suivant permet d'avoir une vision d'ensemble du fonctionnement de la réassurance :



**Figure 78** – Les différents types de réassurance et leurs formes

Ainsi, il existe trois types de réassurance :

- **Facultatif** : l'assureur présente s'il le souhaite un risque au réassureur qui a toute liberté après analyse d'accepter ou de refuser.

Au-delà de cette vision parfaite du risque, un autre avantage à ce type de réassurance est la liberté de souscription. En revanche, une antisélection est possible si l'assureur ne montre que ce qu'il veut et garde uniquement les "bons risques". De plus, la gestion est lourde (beaucoup de dossiers à analyser), les rendements sont incertains et le portefeuille peut être déséquilibré lorsqu'il nécessite beaucoup de facultatif pour mutualiser un seul risque.

Les risques concernés sont les grands risques industriels ou atypiques de par leur exposition ou rareté (3% des risques cédés).

- **Obligatoire** : il existe un traité qui prédéfinit les obligations réciproques dans un cadre fixé (nature des risques, zone géographique, montant et exclusions). L'assureur a l'obligation de céder et le réassureur a l'obligation d'accepter.

Les principaux avantages de ce type de réassurance sont la sécurité de transférer les risques et la gestion allégée qu'elle procure. En revanche, ses principaux inconvénients sont l'existence d'exclusions et l'obligation systématique de céder.

- **Facultatif/Obligatoire** : l'assureur est en facultatif, le réassureur est en obligatoire. Ce traité est rare et concerne moins de 1% des traités. Cela fonctionne quand le réassureur a de longs contrats

avec l'assureur et lui fait confiance.

Le tableau suivant permet d'expliciter les éléments présentés en figure A.1 :

Risque proportionnelle	Risque non proportionnelle
<p>Principe d'une même proportion entre les primes et les sinistres. Il y a donc proportionnalité entre d'un côté la prime et la somme assurée d'un risque et de l'autre le montant du sinistre pris en charge par le réassureur.</p> <ul style="list-style-type: none"> <li><b>Quote-part (QP)</b> : il existe un taux de cession unique sur un portefeuille prédéfini</li> </ul> <p>Ex : 40% de QP est de la cession <math>\Leftrightarrow</math> 40% cédés au réassureur : on cède 40% des primes et 40% des sinistres.</p> <p>Il est possible (pas systématique) que le réassureur introduise une <b>limite par événement</b> (notamment par catnat) ou une <b>limite par an</b> (ex : « Max de 150% de S/P »)</p> <p>La quote-part peut-être sur le « brut » (intervenir en 1<sup>er</sup>) ou sur le « net » intervenir après du non proportionnel.</p> <p>Il faut faire attention à l'ordre d'intervention des protections.</p> <p><b>Avantage : facile à comprendre et à appliquer</b>  <b>Inconvénient : Force à céder un morceau des bons petits risques.</b></p> <ul style="list-style-type: none"> <li><b>Excédent de plein</b> : moins fréquent, l'assureur détermine d'un montant (plein de rétention de somme assuré en deça duquel il conserve 100% du risque. L'assureur va céder, pour les risques qui excèdent ce plein, l'excédent de sommation.</li> </ul> <p>Le taux de cession est déterminé risque par risque.                      Ex : rétention 1M, risque A = 5M <math>\rightarrow</math> taux de cession 80% (4/5). NB : il y a toujours un plafond.</p> <p><b>Avantage : plus optimal si le sujet de notre portefeuille est la volatilité (petit/moyen risque) <math>\rightarrow</math> cette méthode permet de garder que les petits risques et céder un morceau des grands et atypiques.</b></p>	<p>« Déphasage » entre les primes cédées et les sinistres cédées. Le montant de sinistre pris en charge par le réassureur est déterminé a priori avec un seuil d'intervention et une limite maximum, sans lien direct avec la prime originale et la somme assurée.</p> <ul style="list-style-type: none"> <li><b>Excédent de sinistre (Excess of loss)</b>: l'assureur écrète et l'excédent de sinistralité est transféré en 2 schémas possibles d'excédents de sinistre (noté <math>x_s</math>)                             <ul style="list-style-type: none"> <li><b>Par sinistre</b> : autant de franchises que de bâtiment</li> <li><b>Par événements</b> : on agrège les sinistres générés par le même fait générateur.</li> </ul> </li> </ul> <p>L'assureur fait toujours les deux ensemble.</p> <p>Exemple : « <math>15x_s, 5</math> » : 15 c'est le montant de portée, 5 c'est le montant de priorité et <math>x_s</math> c'est l'excédent de sinistre en réassurance non proportionnelle.</p> <ul style="list-style-type: none"> <li><b>Excédent de perte annuelle (stop loss)</b>                      C'est le même principe que l'excédent de sinistre, mais ici on couvre la sinistralité annuelle agrégée. Cela revient à couvrir le rapport S/P de l'assureur (en l'écrétant).</li> </ul> <p>Exemple : Si <math>S/P &gt; \text{Priorité}</math> : cela correspond à un excédent de perte. Le réassureur prendra l'excédent au-dessus de la priorité. Ainsi, dans le cas « 50% <math>\times</math> 110% » on couvre bien la sinistralité annuelle et pas un sinistre précis.</p>

## Annexe A :

Variables	Descriptions
Vitesse du vent à 10m	Magnitude de la vitesse horizontale deux dimensionnelles du vent. Il s'agit d'une moyenne sur la période agrégée en mètres par seconde.
U-composante du vent à 10m	Magnitude de la composante est du vent à 10m au-dessus de la surface.
V-composante du vent à 10m	Magnitude de la composante nord du vent à 10m au-dessus de la surface.
200hPa U-composante du vent à 10m	Magnitude de la composante est du vent à 10m au-dessus de la surface à 200hPa.
200hPa V-composante du vent à 10m	Magnitude de la composante nord du vent à 10m au-dessus de la surface à 200hPa.
Humidité relative à 2m	Ratio en pourcentage de la masse de vapeur d'eau à la masse de vapeur d'eau au point de saturation compte tenu de la température locale. Les données représentent la moyenne sur la période d'agrégation à 2m au-dessus de la surface.
Humidité spécifique à 2m	Quantité d'humidité dans l'air à 2 m au-dessus de la surface divisée par la quantité d'air ajoutée à l'humidité locale.
Température à 2m	Température de l'air ambiant. Les données représentent la moyenne sur la période d'agrégation à 2m au-dessus de la surface.
500 hPa hauteur géopotentielle	Energie gravitationnelle potentielle par unité de masse normalisée par la gravité standard à 500 hPa à la même latitude.
850hPa U-composante du vent	Magnitude de la composante est de la vitesse horizontale bidimensionnelle de l'air à 850hPa.
850hPa V-composante du vent	Magnitude de la composante nord de la vitesse horizontale bidimensionnelle de l'air à 850hPa.
Fraction de surface terrestre	Pourcentage de cellule de grille occupée par la surface terrestre. Les données sont indépendantes du temps.
Température maximale à 2 m au cours des dernières 24 heures	Température maximale de l'air proche de la surface. Les données représentent le maximum quotidien à 2m au-dessus de la surface.

Variables	Descriptions
Température minimale à 2 m au cours des dernières 24 heures	Température minimale de l'air près de la surface. Les données représentent le minimum quotidien à 2m au-dessus de la surface.
Flux d'évaporation moyen	Masse d'eau liquide de surface et souterraine par unité de surface et de temps, qui s'évapore du sol. Les données incluent la sublimation et représentent la moyenne sur la période d'agrégation.
Flux moyen de précipitations	Dépôt d'eau à la surface de la Terre sous forme de pluie, de neige, de glace ou de grêle. Le flux de précipitations est la masse d'eau par unité de surface et de temps. Les données représentent la moyenne sur la période d'agrégation
Pression moyenne au niveau de la mer	Pression atmosphérique au niveau de la mer. Dans les régions où la surface de la Terre est au-dessus du niveau de la mer, la pression de surface est utilisée pour calculer la pression atmosphérique qui existerait au niveau de la mer étant donné une température de l'air constante de la surface à ce niveau. Les données représentent la moyenne sur la période d'agrégation.
Orographie	Hauteur au-dessus du géoïde (soit 0,0 au-dessus de l'océan). Les données sont indépendantes du temps.
Pression de surface	Pression de l'air à la limite inférieure de l'atmosphère.
Rayonnement solaire de surface descendante	Flux radiatif d'énergie en ondes courtes orientées vers le bas par unité de terrain. Les données représentent la moyenne sur la période d'agrégation à la surface.
Rayonnement thermique de surface descendante	Flux radiatif d'énergie à ondes longues incandescent sur la surface par unité de terrain.
Rayonnement d'onde courte ascendant de surface	Flux radiatif d'énergie d'onde courte à la surface par unité de terrain.
Couverture nuageuse totale	Fait référence à l'ensemble de la colonne d'atmosphère, telle que vue de la surface, le sommet de l'atmosphère. La couverture nuageuse fait référence à la fraction de la surface horizontale occupée par les nuages.
Flux de ruissellement total	Masse d'eau liquide de surface et souterraine par unité de terrain et de temps, qui s'écoule du sol.

**Table 16** – Descriptif des variables base EURO-CORDEX

Variables	Descriptions
Niveau de surcote pour différents centiles	Niveau défini comme la différence entre les simulations de marée pure et de hauteur d'eau totale, pour les centiles suivants : 10e, 25e, 50e, 75e et 90e.
Niveau de surcote pour différentes périodes de retour	Niveau défini comme la différence entre les simulations de marée pure et de hauteur d'eau totale, pour les périodes de retour suivantes : 2, 5, 10, 25, 50 et 100 ans. La période de retour est une manière standard de décrire la probabilité et la gravité d'un événement. Il décrit l'intervalle de temps estimé entre des événements d'une taille ou d'une intensité similaire.
Marnage	Marnage moyen observé sur la période de 30 ans simulée.
Niveau d'eau total pour différents centiles	Niveau d'eau, y compris le niveau de marée pure et d'onde de tempête, pour les centiles suivants : 10e, 25e, 50e, 75e et 90e.
Niveau d'eau total pour différentes périodes de retour	Niveau d'eau total, y compris la marée, le niveau de surcote et la prise en compte de l'élévation future du niveau de la mer, pour les périodes de retour suivantes : 2, 5, 10, 25, 50 et 100 ans.

**Table 17** – Descriptif des variables base **ESL**

## **Annexe B : Descriptif des variables base ESL**

Variables	Descriptions
Marée haute la plus élevée annuellement	Marée haute, y compris le niveau moyen de la mer et l'élévation du niveau de la mer. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Marée basse la plus basse annuellement	Marée basse, y compris le niveau moyen de la mer et l'élévation du niveau de la mer. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Moyenne annuelle des plus importantes marées hautes	Moyenne annuelle de la marée haute la plus élevée de chaque jour de marée (fenêtre de 25 heures), y compris le niveau moyen de la mer et l'élévation du niveau de la mer. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Moyenne annuelle de la marée basse la plus basse	Moyenne annuelle de la marée basse la plus basse de chaque jour de marée, y compris le niveau moyen de la mer et l'élévation du niveau de la mer. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Moyenne sur la période des plus hautes marées hautes	Moyenne sur la période de 30 ans simulée. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Moyenne sur la période des plus basses marées basses	Moyenne sur la période de 30 ans simulée. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.
Marée astronomique la plus haute	Marée astronomique la plus haute sur la période de 30 ans simulée.
Marée astronomique la plus basse	Marée astronomique la plus basse sur la période de 30 ans simulée.
Niveau moyen de la mer	Niveau moyen de la mer incluant son élévation sur la période de 30 ans simulée. Les ondes de tempête causées par le forçage atmosphérique ne sont pas prises en compte.

## Annexe C : Lois usuelles en hydrologie

### Gumble MAX

Il s'agit d'un type de distribution statistique qui ne présente aucune restriction quel que soit le côté. Cette distribution des valeurs extrêmes (basse ou haute) est utilisée dans différentes distributions pour modéliser des pics. Notamment dans la prédiction de catastrophe naturelle (séisme, inondation ...).

Pour la période de retour ( $t$ ) voulue, la variable abrégée ( $Y_t$ ) est la suivante :

$$Y_t = \ln(\ln(\frac{t}{t-1}))$$

La moyenne et l'écart-type abrégés ont été obtenus à partir de la table de distribution de Gumbel pour l'échantillon de taille  $N$  donné. Le facteur de fréquence est utilisé en utilisant la formule suivante :

$$K_t = \frac{Y_t - Y_n}{S_n}$$

$K_t$  est le facteur de fréquence,  $Y_t$  est la variable abrégée,  $Y_n$  la moyenne abrégée,  $S_n$  l'écart type abrégé.

Ainsi le déversement prédit ( $Q_p$ ) est calculé en utilisant la formule de distribution normale standard pour diverses périodes de retours selon différentes saisons :

$$Q_p = \mu + K_t \times \sigma$$

### Log Pearson 3 (LPIII)

Il existe également des distributions basées sur la fonction Gamma de 2 ou 3 paramètres dans l'espace réel ou logarithmique tels que Pearson de type III (PIII), Log-Pearson de type III (LPIII) et leurs cas spécifiques de distribution log normale et généralisée de Pareto (GPD) à deux paramètres.

Dans le cas de la LPIII, la méthode statistique utilisée permet de prédire l'inondation sur peu de site ou sur une rivière spécifique. La fréquence de distribution est construite après avoir calculé les données statistiques à un endroit particulier.

Cette méthode particulière aide pour extrapoler des périodes de retour plus extrême. Après avoir trouvé l'écoulement actuel ( $Z$ ), il faut calculer son logarithme naturel pour trouver la moyenne du logarithme standard ( $\mu$ ) and son écart type pour chaque saison :

$$Z = \log_{10} Q$$

Le coefficient de d'asymétrie ( $C_s$ ) est calculé en utilisant le logarithme de l'écoulement  $Z$  et pour la période de retour requise  $T$ , la probabilité ( $P$ ) est ensuite calculée en pourcentage avec la formule suivante :

$$P = \frac{1}{T}(\%)$$

A partir de la table de distribution normal standard, par interpolation, la déviation normale standard  $z$  est ensuite calculé. Le facteur de fréquence dépend du coefficient de d'asymétrie et de la période de retour.

Quand  $C_s = 0$ , le facteur de fréquence est égal à  $z$  et est calculé comme dans le cas d'une déviation. Sinon, le facteur de fréquence ( $K_t$ ) est modifié en utilisant les formules développées par Kite (1977) :

$$K_t = z + (z^2 - 1)k + \frac{1}{3} \times (z^3 - 6z) \times k^2 - (z^2 - 1) \times k^3 + zk^4 + \frac{1}{3} \times k^5$$

Où  $k = C_s/6$

Pour prédire le logarithme de l'écoulement la formule suivante est utilisée :

$$q_p = \mu + K_t \times \sigma$$

Où  $q_p$  représente l'écoulement logarithme prédit,  $\mu$  la moyenne du logarithme standard et  $\sigma$  sa déviation.

Ainsi l'écoulement prédit ( $Q_p$ ) est calculé en prenant l'antilog de  $q_p$  :

$$Q_p = \text{antilog}(q_p)m^3/s$$

## Annexe D : Calcul pour la loi GEV

Fonction caractéristique d'une loi GEV :

$$\forall x \in \mathbb{R}, F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \times \left( \frac{x - \mu}{\sigma} \right)_+^{-1/\xi} \right] \right\}$$

En la dérivant, la densité associée obtenue est :

$$\forall x \geq 0, \text{ tel que : } 1 + \kappa \times \frac{x}{\sigma} \geq 0 :$$

$$f(x; \mu, \sigma, \kappa) = \begin{cases} \frac{1}{\sigma} \left[ 1 + \kappa \times \left( \frac{x - \mu}{\sigma} \right) \right]^{(-1/\kappa) - 1} \exp \left\{ - \left[ 1 + \kappa \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\kappa} \right\} & \text{si } \kappa \neq 0 \\ \frac{1}{\sigma} \times \exp \left( - \frac{x - \mu}{\sigma} \right) \times \exp \left[ - \exp \left( - \frac{x - \mu}{\sigma} \right) \right] & \text{si } \kappa = 0 \end{cases}$$

Fonction de vraisemblance de la loi GEV :

$$\ln(L(x|\theta)) = l(x|\theta) = -n \ln(\sigma) + \sum_{i=1}^n \left[ \left( \frac{1}{\kappa} - 1 \right) \ln \left[ (y_i) - (y_i)^{1/\kappa} \right] \right]$$

avec :  $\theta = (\mu, \sigma, \kappa)$  et  $y_i = \left[ 1 - \frac{\kappa}{\sigma} (x_i - \mu) \right]$ .

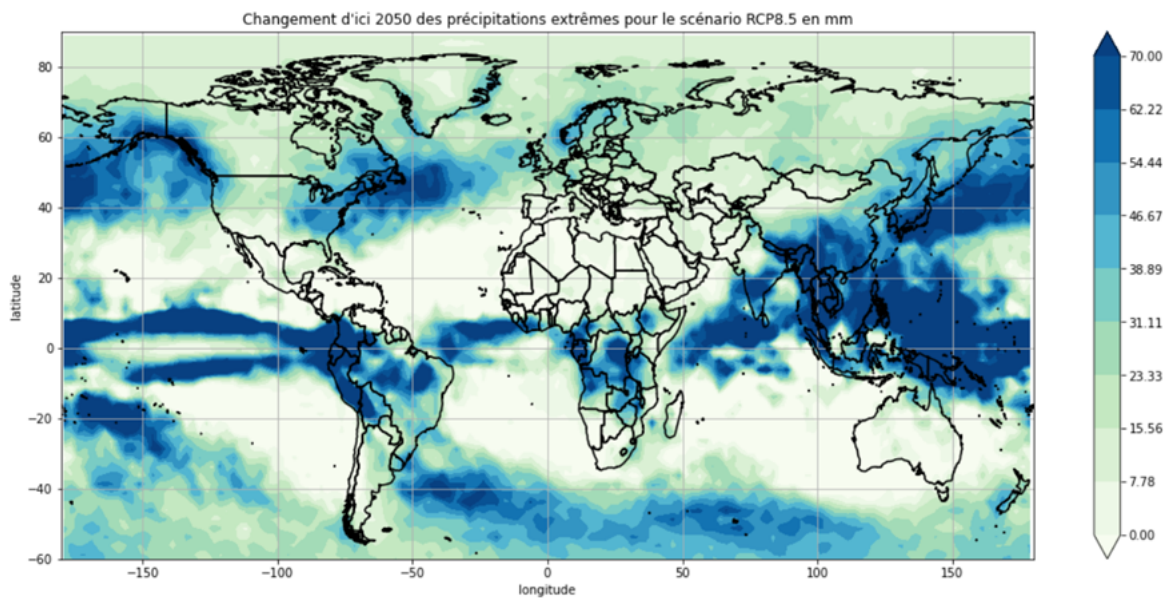
Dériver successivement par  $\mu, \sigma$  et  $\kappa$ , avec le maximum de vraisemblance en  $\theta = \hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\kappa})$ , permet d'obtenir le système d'équations suivant :

$$\left\{ \begin{array}{l} \frac{1}{\sigma} \sum_{i=1}^n \frac{1 - \kappa - y_i^{1/\kappa}}{y_i} = 0 \\ -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n \frac{1 - \kappa - y_i^{1/\kappa}}{y_i} \times \frac{x_i - \mu}{\sigma} = 0 \\ -\frac{1}{\kappa^2} \sum_{i=1}^n \ln(y_i) (1 - \kappa - (y_i)^{1/\kappa}) + \frac{1 - \kappa - (y_i)^{1/\kappa}}{y_i} \times \kappa \times \frac{x_i - \mu}{\sigma} = 0 \end{array} \right.$$

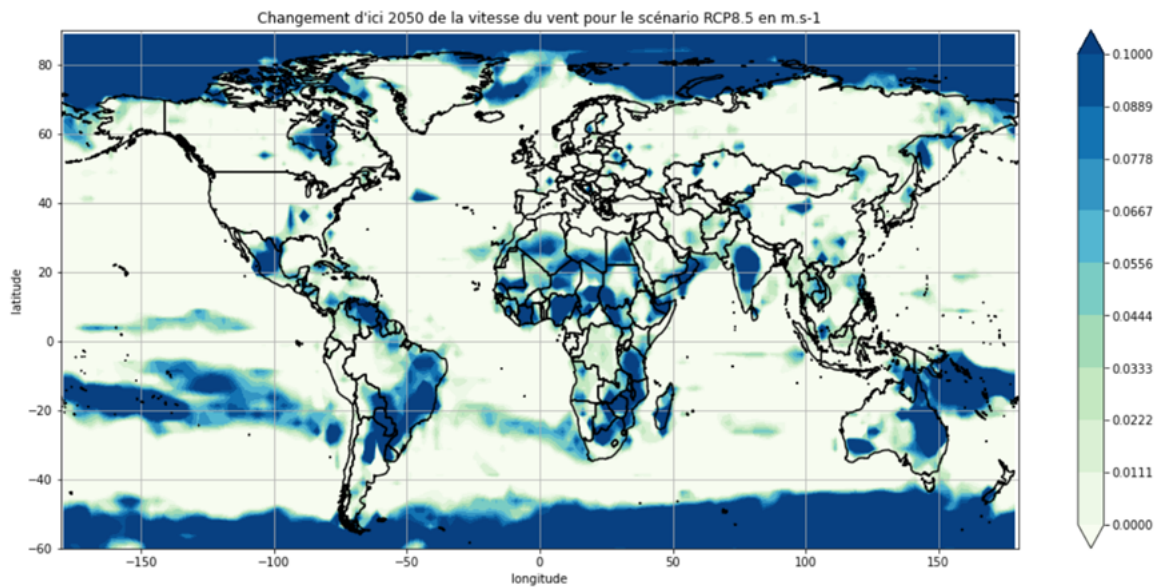
Il faut ensuite utiliser la méthode de Newton pour résoudre cette équation. Pour se faire, il faut vérifier qu'il s'agit bien d'un maximum global.

A noter qu'il n'existe pas toujours de solution à ce système. Ainsi, pour étudier les performances des estimateurs il est possible de rajouter un taux d'erreur

## Annexe E : Exemples d'exploitation des données du modèle CMIP5



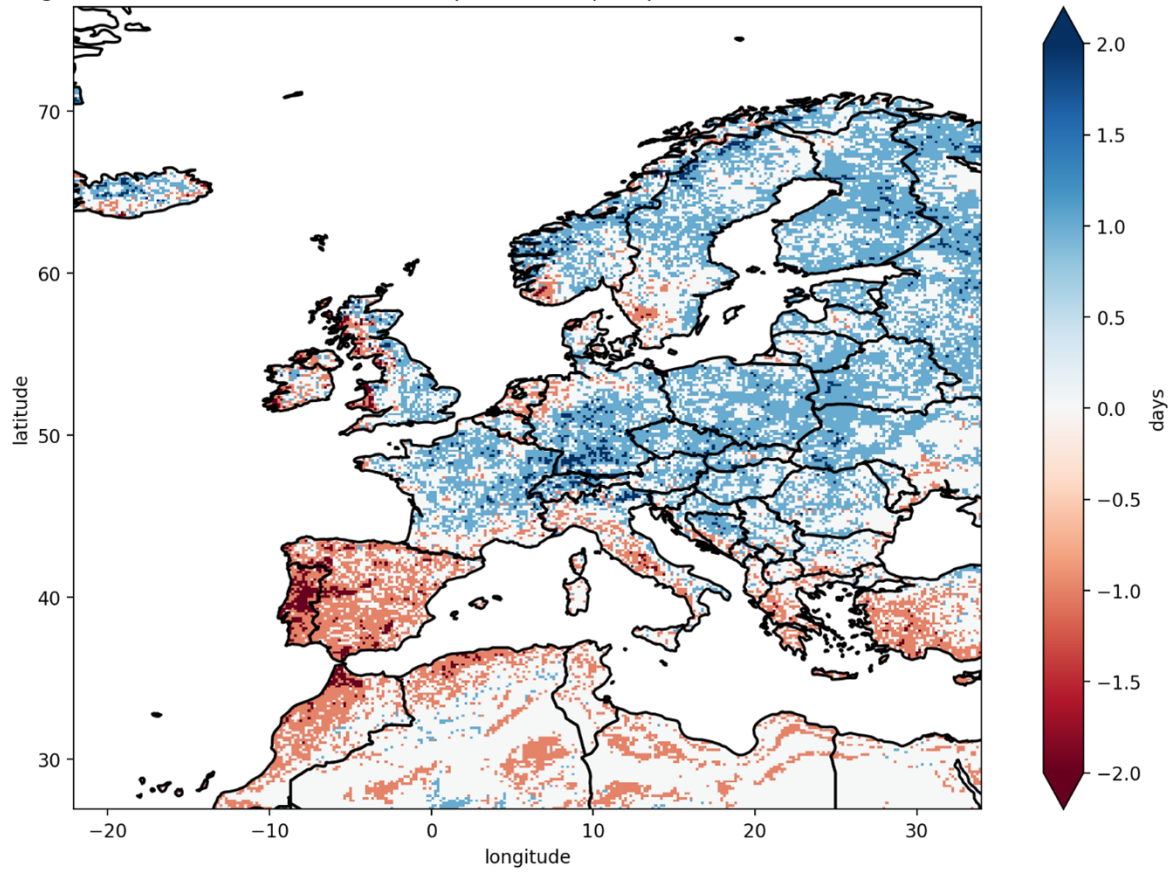
**Figure 79** – Sorties du modèle sur les précipitations



**Figure 80** – Sorties du modèle sur la vitesse du vent

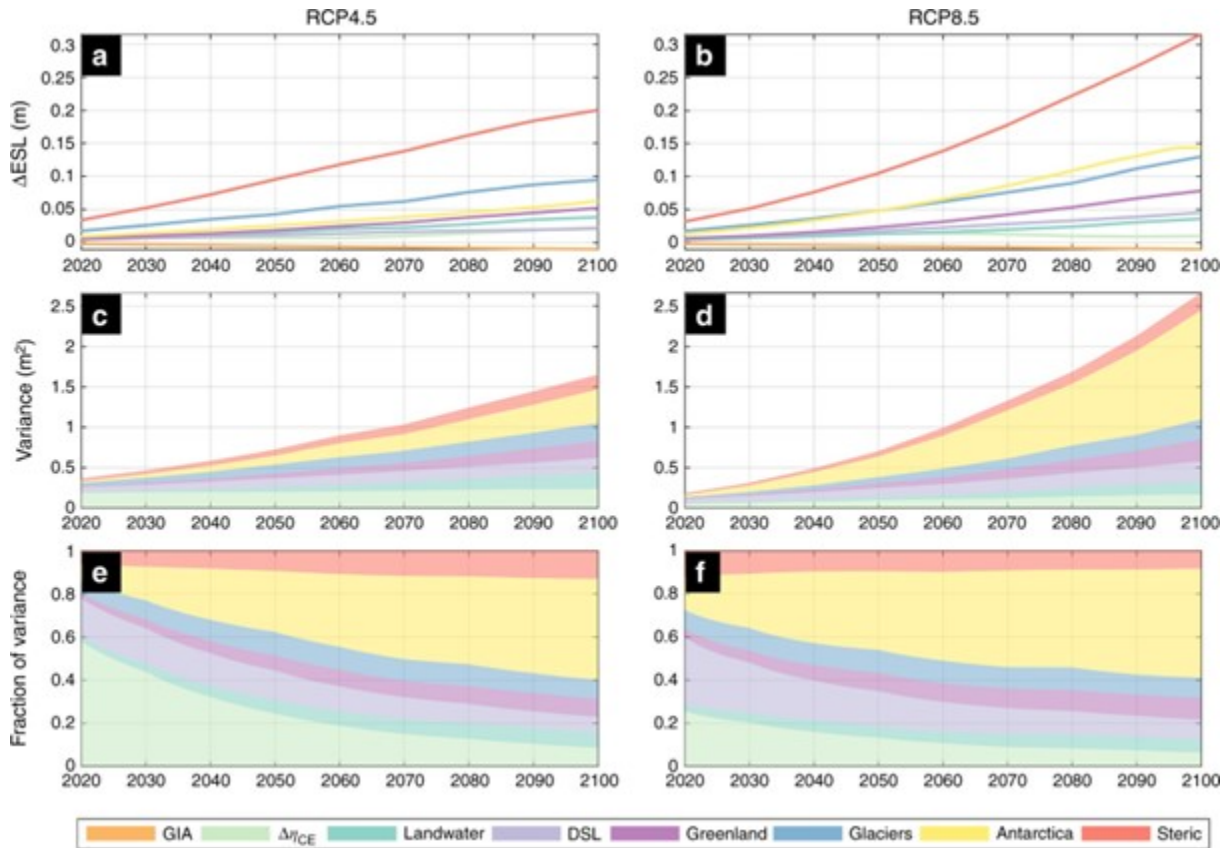


Changement dans la durée maximale des périodes de précipitations entre 1990 et 2040 - RCP 8.5



**Figure 81** – Sorties du modèle sur les précipitations

## Annexe F : Ventilation des contributions projetées de l'ESL et de leur incertitude



**Figure 82** – Ventilation des contributions projetées de l'ESL et de leur incertitude

La composante stérique est la variable la plus significative de l'augmentation de l'ESL 100 au 21e siècle. En 2100, elle exprime 37% de la moyenne mondiale ESL 100 dans les deux RCP. Les deuxième et troisième contributeurs les plus importants varient selon les RCP, car la fonte des glaciers est plus importante que celle de l'Antarctique (18 % contre 12 % de ESL100, respectivement) selon le RCP 4.5, tandis que sous la RCP 8.5, l'Antarctique prévaut après 2050 (en 2100, l'Antarctique : 17 %, les glaciers : 15 %).

## Annexe G : Projection des changements des contributions d'ESL selon différentes régions

	RCP4.5		RCP8.5	
	2050	2100	2050	2100
<b>NEWFOUNDLAND &amp; LABRADOR</b>	3 [-7-14]	5 [-4-15]	4 [-3-12]	8 [5-11]
<b>WEST N ATLANTIC</b>	1 [-10-13]	4 [-11-18]	6 [-4-18]	7 [3-12]
<b>CENTRAL AMERICA</b>	-6 [-12--0]	-7 [-11--1]	-4 [-4--3]	-5 [-7--3]
<b>SOUTH S. AMERICA</b>	0 [-8-8]	5 [-7-14]	-3 [-6-0]	-9 [-10--7]
<b>SOUTHERN OCEAN</b>	10 [6-14]	9 [4-14]	12 [9-16]	19 [16-22]
<b>NORTH EUROPE</b>	2 [0-4]	4 [-2-10]	11 [4-19]	17 [8-25]
<b>S EUROPE &amp; NW AFRICA</b>	3 [1-5]	4 [0-8]	-2 [-2--1]	-4 [-5--2]
<b>E AFRICA</b>	-11 [-17--4]	-18 [-25--10]	-12 [-16--6]	-21 [-32--10]
<b>S &amp; SE ASIA, AUSTRALIA</b>	-5 [-29-25]	-3 [-30-29]	-5 [-19-11]	-22 [-43-2]
<b>E CHINA &amp; JAPAN SEA</b>	17 [-26-65]	36 [-6-82]	10 [-0-21]	27 [11-42]
<b>OKHOTSK &amp; W BERING</b>	3 [-17-28]	-1 [-13-13]	-11 [-20--3]	-15 [-25--7]
<b>ALASKA-E BERING</b>	4 [-0-9]	5 [1-9]	4 [1-7]	11 [6-16]
<b>S PACIFIC</b>	-6 [-19-8]	-5 [-14-5]	-0 [-4-5]	10 [-1-20]

**Figure 83** – Projection des changements des contributions d'ESL selon différentes régions

Les valeurs sont exprimées en moyenne selon 14 zones géographiques régionales définies par leur proximité géographique et leurs similarités. Les valeurs médianes sont exprimées entre crochets selon un intervalle de confiance à 67%.

## Annexe H : Projections des modèles individuels

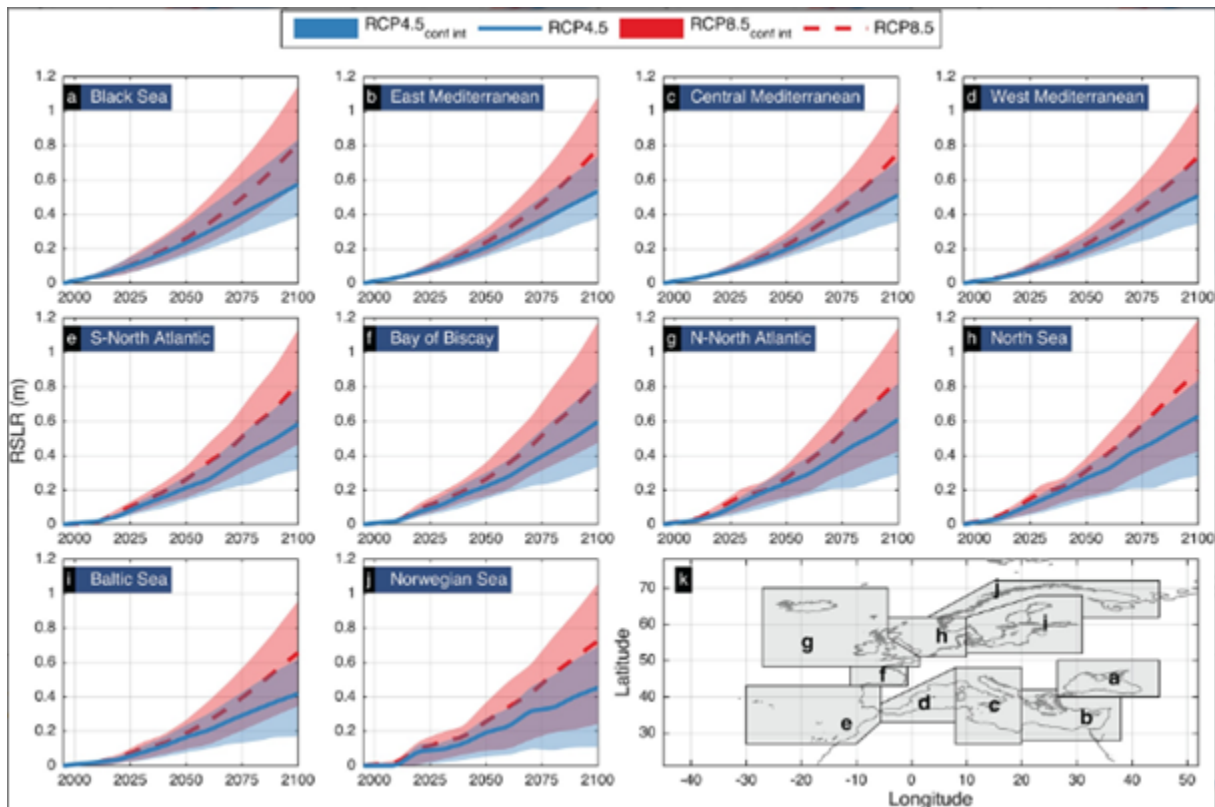


Figure 84 – Projections des modèles individuels

Source : Michalis Vousdoukas, Lorenzo Mentaschi, Evangelos Voukouvalas, Martin Verlaan, Luc Feyen (in press 2017). Extreme sea levels on the rise along Europe’s coasts. *Earth’s Future*. [\[en ligne\]](#)

## Annexe I : k plus proche voisin

Il s’agit d’un algorithme supervisé utilisé à des fins de prédiction. A partir de l’observation d’un nouvel individu, l’algorithme permet de l’attribuer au même groupe (ou classe) en fonction de sa distance.

Etapes :

- 1 Choix d’une mesure commune (souvent euclidienne)
- 2 Choix du nombre k de voisin à considérer
- 3 Affectation a un groupe

Le principal avantage est qu’il s’agit d’un algorithme d’une grande simplicité mais il ne convient pas au problème ayant plus de 10 variables et est très sensible aux données bruitées.

Dans le cadre de ce mémoire, les inconvénients exposés ne s’appliquaient pas compte tenu du faible nombre de variables et des séries temporelles complètes.

## Annexe J : Algorithme de classification

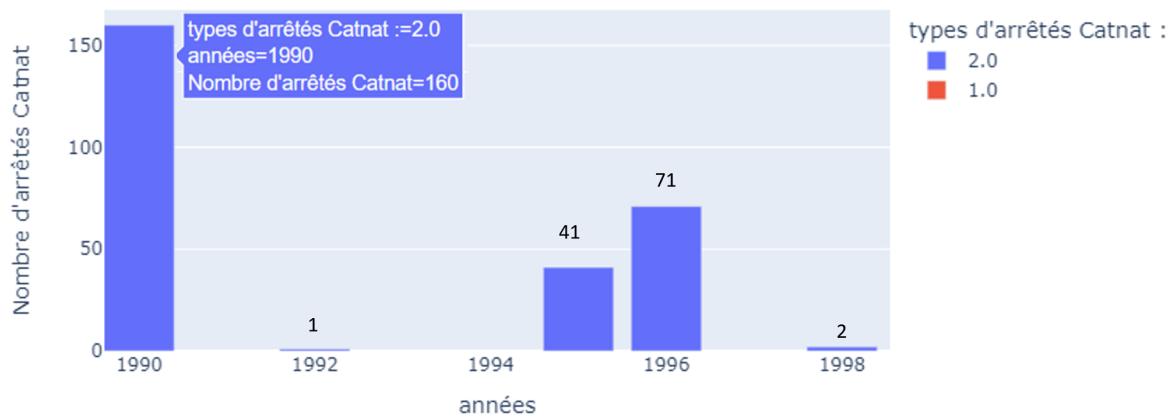
Initialisation :  $f_{k0} = 0, k = 1, 2, \dots, K$

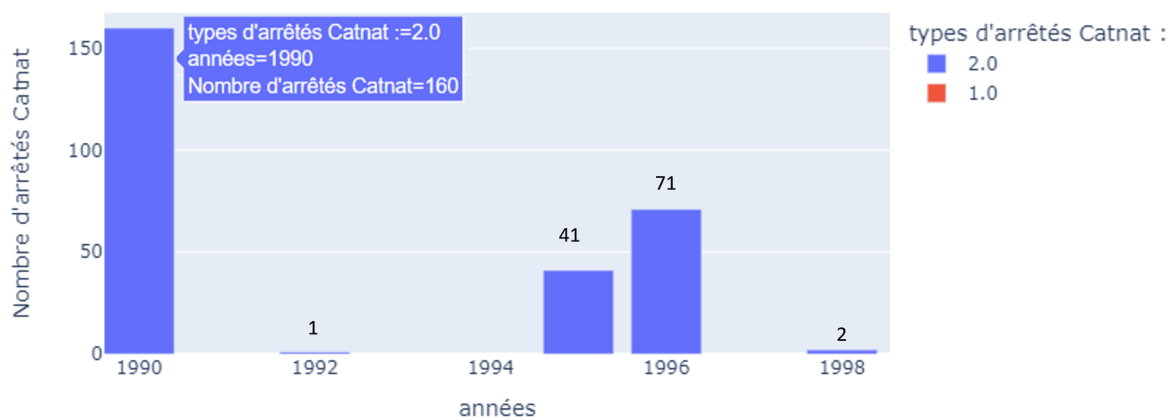
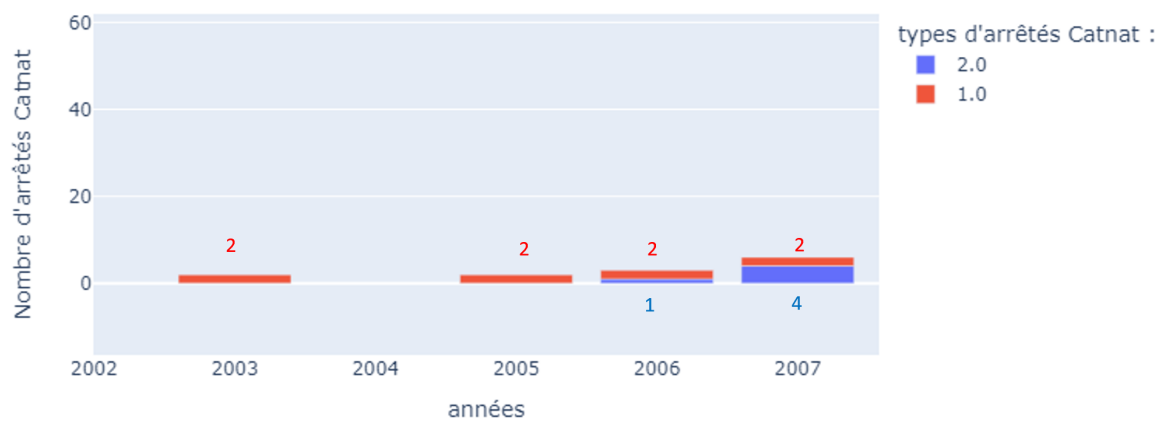
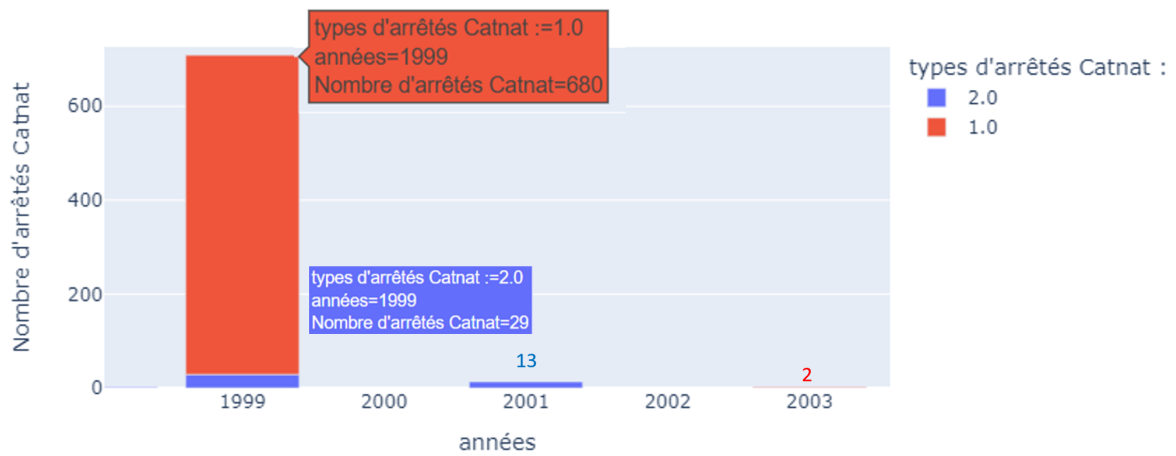
De  $m = 1$  à  $M$  :

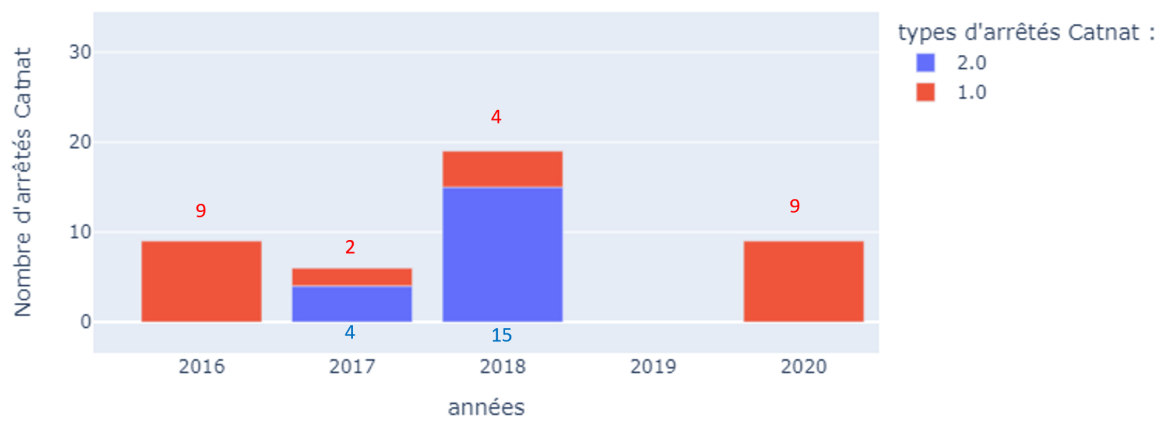
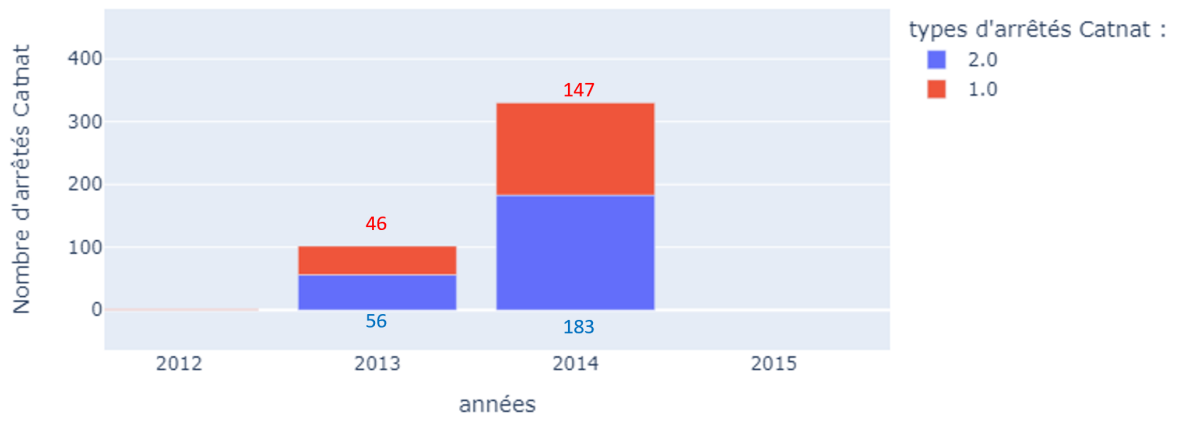
- Update :  $p_k(x) = \frac{\exp(f_k(x))}{\sum_{j=1}^K \exp(f_j(x))}, k = 1, 2, \dots, K$
- pour  $k = 1, 2, \dots, K$
- définir  $r_{ikm} = y_{ik} - p_k(x_i), i = 1, 2, \dots, N$
- calibrer un arbre pour les cibles
- calculer :  $\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} r_{ikm}}{\sum_{x_i \in R_{jkm}} |r_{ikm}|(1-|r_{ikm}|)}, j = 1, 2, \dots, J_m$
- mettre à jour :  $f_{k,m}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jkm} I(x \in R_{jkm})$

Résultat :  $\hat{f}_k(x) = f_{kM}(x), k = 1, 2, \dots, K$

## Annexe K : Visualisation exhaustive de la base de validation historique des arrêtés Cat-Nat (1990-2020)





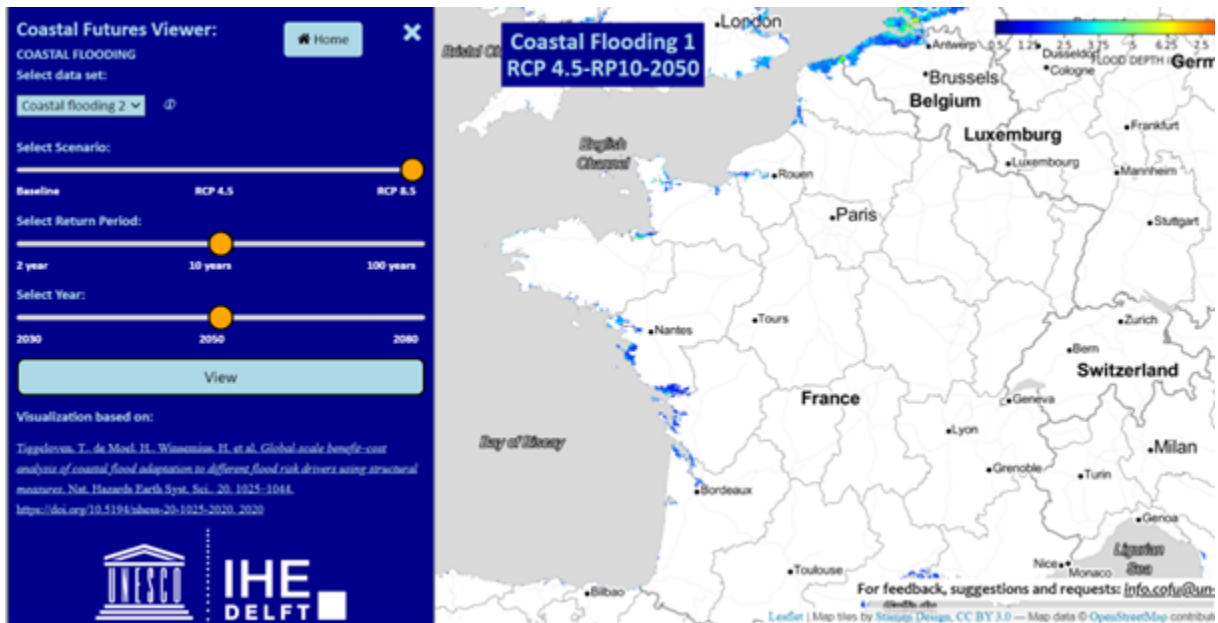
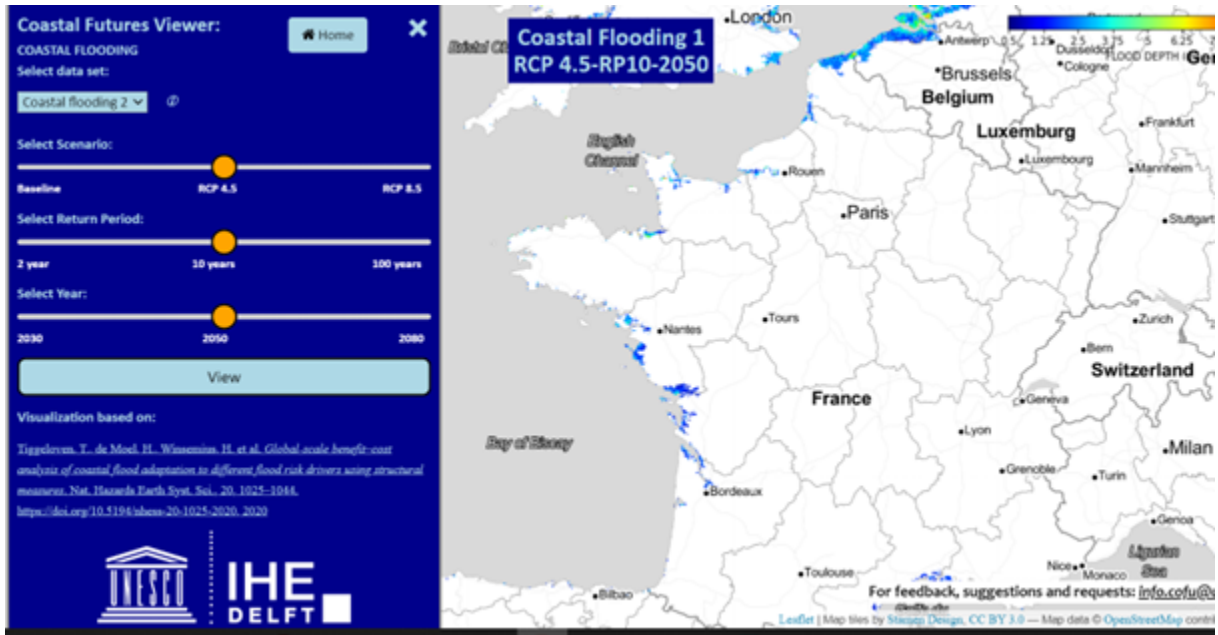


## Annexe L : Carte des départements



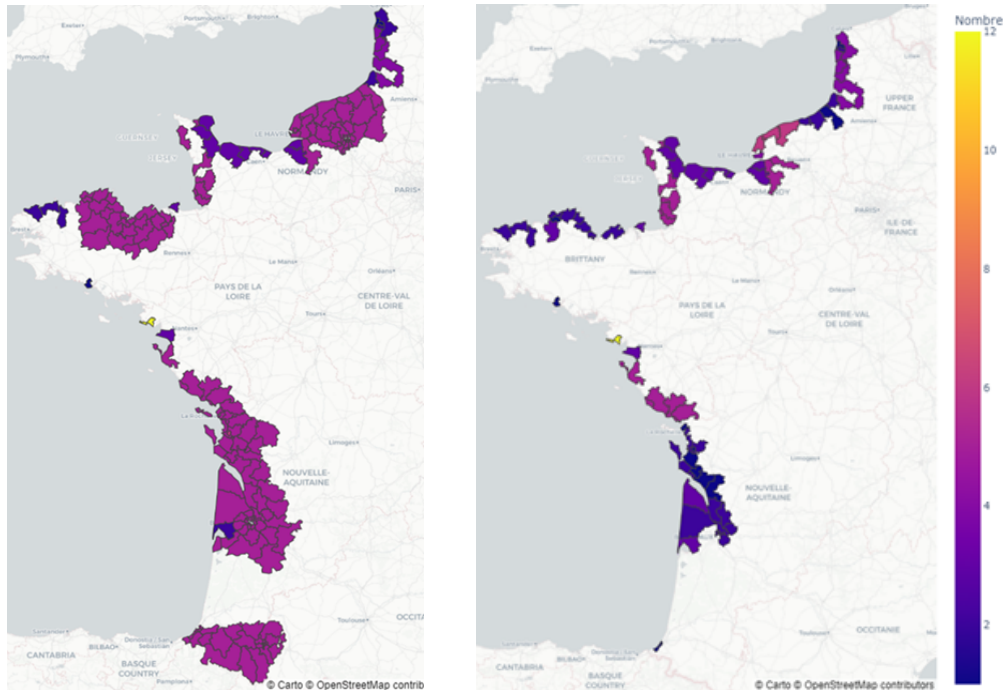


## Annexe M : Sorties de l'outil coastal futures sur le deuxième jeu de données



## Annexe N : Base historique d'apprentissage

Les cartes suivantes représentent le nombre d'arrêtés Cat-Nat en fonction des cantons français et des différentes classes pour la base d'apprentissage donnée aux différents modèles. La classe 1 couvre la période 1990 à 2005 et la classe 2 se restreint aux années 1990 à 2001. Elles permettent d'identifier une certaine concentration géographique des événements :

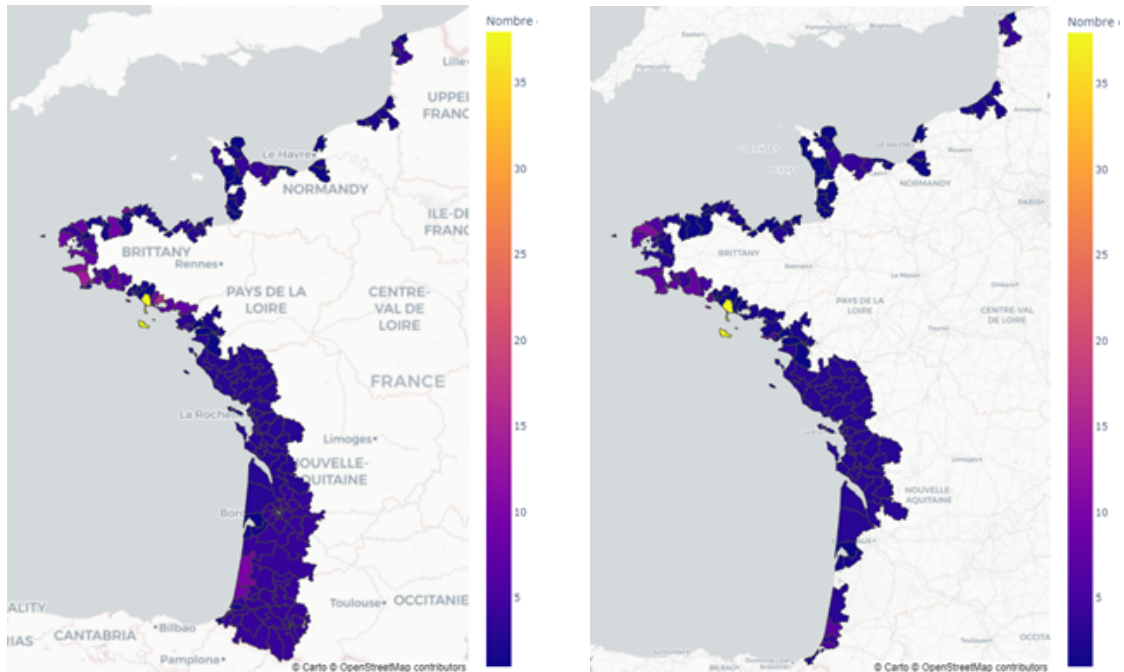


**Figure 85** – Nombre de jours de submersions marines déclarés (reconnues ou non) (gauche) et reconnues (droite) par canton sur la période d'apprentissage 1990-2005

On peut remarquer que le nombre de jour concernant les demandes refusées (classe 1) sont bien plus présentes à l'intérieur des terres. Ceci est particulièrement le cas pour les départements des Pyrénées-Atlantiques, des Landes, des Côtes-d'Armor et de Seine-Maritime. Cela reflète bien le caractère déséquilibré de la base et le niveau d'exigence des commissions interministérielles chargées de la décision.

## Annexe O : Base historique de validation

Les cartes suivantes représentent le nombre d'arrêtés Cat-Nat en fonction des cantons français et des différentes classes pour la base de validation utilisée pour comparer les différents modèles. La classe 1 couvre la période 2006 à 2020 et la classe 2 se restreint aux années 2006 à 2018. Elles permettent d'identifier une certaine concentration géographique des événements :



**Figure 86** – Nombre de jours de submersions marines déclarés (reconnues ou non) (gauche) et reconnues (droite) par canton sur la période de validation historique 2006-2020

On peut observer que comme sur la base d'apprentissage, les demandes se concentrent sur le littoral sud de la région Bretagne.

## Annexe P : Superposition de la carte de prédiction des cantons affectés par la submersion marine à la carte de la France



**Figure 87** – Superposition de la carte de prédiction des cantons affectés par la submersion marine à la carte de la France

On remarque que les cantons choisis sont uniquement ceux de la première ligne de département du littoral.

## Table des figures

1	Les paramètres liés à la submersion marine d'après les données REFMAR . . . . .	3
2	Synthèse de méthodologie . . . . .	4
3	Nombre de jours de submersions marines reconnues . . . . .	7
4	Réalité connue à 2020 (1ère carte) par rapport aux prédictions à 2050 selon les RCP 4.5 (2ème figure) et RCP 8.5 (3ème figure) . . . . .	8
5	Representation of the coastal flood based on REFMAR data . . . . .	9
6	Summary of methodology . . . . .	10
7	Coastal submersions over the historical validation period (2006-2020) for the Nat-Cat type 2 decrees (recognized submersion) . . . . .	13
8	Known reality in 2020 (1st map) compared to predictions in 2050 according to RCP 4.5 (2nd figure) and RCP 8.5 (3rd figure) . . . . .	14
9	Les différents scénarios RCP d'après le GIEC . . . . .	20
10	La chaîne de valeur assurance impactée par le changement climatique . . . . .	24
11	Les risques impactés par le changement climatique . . . . .	24
12	Projection de la hausse moyenne du niveau des mers par rapport à 1900 d'après le GIEC . . . . .	27
13	Schéma illustratif des principaux paramètres d'une submersion marine, source : BRGM . . . . .	34
14	Schéma synthétique des dangers liés aux submersions marines et facteurs aggravants . . . . .	35
15	Schéma illustratif des principaux phénomènes influant sur une submersion marine . . . . .	35
16	Carte des communes reconnues en état de catastrophe naturelle après la tempête Johanna (source base de données Gaspar) . . . . .	46
17	Nombre d'arrêtés Cat-Nat de type submersion marine au cours du temps . . . . .	47
18	Nombre de jours de submersions marines déclarés (reconnues ou non) et reconnues (droite) par canton de 1990 à 2020 . . . . .	48
19	Représentation des relations modélisées, source : <a href="#">[en ligne]</a> . . . . .	50
20	Représentation schématique du fonctionnement d'un GCM . . . . .	51
21	Illustration du changement d'échelle entre GCM et RCM . . . . .	52
22	Zone modélisée par EURO-CORDEX . . . . .	54
23	Résultats pour le RCP 8.5 . . . . .	55
24	Changements prévus dans l'ESL mondial . . . . .	61
25	Variation en pourcentage des valeurs historiques (ordonnée) par canton (abscisse) en fonction des périodes de retour . . . . .	62
26	Localisation des capteurs côtiers . . . . .	63
27	Représentation des cantons du littoral . . . . .	64
28	Représentation de la localisation des capteurs . . . . .	64
29	Localisation des 10 capteurs sélectionnés aléatoirement . . . . .	65
30	Durée moyenne en jours des événements de type submersion marine par année . . . . .	67
31	Concentration mensuelle des évènements de type submersion marine . . . . .	70
32	Réduction de l'erreur d'apprentissage avec le Gradient Boosting . . . . .	73
33	Algorithme Gradient Boosting . . . . .	76
34	SVM approche One-to-One . . . . .	77
35	SVM approche One-to-Rest . . . . .	78
36	Résultats sur la base d'apprentissage . . . . .	80
37	Résultats sur la base de validation historique RCP 8.5 . . . . .	81
38	Résultats sur la base de validation historique RCP 4.5 . . . . .	81
39	Résultats sur la base d'apprentissage . . . . .	82

40	Résultats sur la base de validation historique RCP 8.5 . . . . .	82
41	Résultats sur la base de validation historique RCP 4.5 . . . . .	83
42	Résultats sur la base d'apprentissage . . . . .	84
43	Résultats sur la base d'apprentissage avec option « <i>balanced</i> » . . . . .	84
44	Validation du modèle sur la base historique . . . . .	85
45	Ensemble des jours d'arrêtés Cat-Nat déclarés . . . . .	85
46	Base de validation historique sur la période 2006 à 2020 . . . . .	86
47	Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS_XGB ») du modèle <i>Gradient Boosting</i> RCP 4.5 sur la base recensant l'absence de submersion marine . . . . .	86
48	Résultat du modèle XGB pour le scénario RCP 4.5 . . . . .	87
49	Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS_XGB ») du modèle <i>Gradient Boosting</i> RCP 8.5 sur la base recensant l'absence de submersion marine . . . . .	87
50	Résultat du modèle XGB pour le scénario RCP 8.5 . . . . .	88
51	Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS_RF ») du modèle <i>Random Forest</i> RCP 4.5 sur la base recensant l'absence de submersion marine . . . . .	89
52	Résultat du modèle <i>Random Forest</i> pour le scénario RCP 4.5 . . . . .	89
53	Répartition des modalités de la variable type d'arrêtés Cat-Nat (« CLASS_XGB ») du modèle <i>Random Forest</i> RCP 8.5 sur la base recensant l'absence de submersion marine . . . . .	90
54	Résultat du modèle <i>Random Forest</i> pour le scénario RCP 8.5 . . . . .	90
55	Score d'importance des variables explicatives dans le modèle de <i>Gradient Boosting</i> . . . . .	91
56	Comparaison des prédictions à la réalité pour les demandes d'arrêtés Cat-Nat . . . . .	92
57	Comparaison des prédictions à la réalité par cantons en Normandie . . . . .	93
58	Comparaison des prédictions à la réalité par cantons dans Les Hauts de France . . . . .	94
59	Comparaison des prédictions à la réalité par cantons en Bretagne . . . . .	94
60	Comparaison des prédictions à la réalité par cantons en Nouvelle-Aquitaine . . . . .	95
61	Comparaison des prédictions à la réalité par cantons en Pays de la Loire . . . . .	95
62	Réalité connue à 2020 (1ère carte) par rapport aux prédictions à 2050 selon les RCP 4.5 (2ème figure) et RCP 8.5 (3ème figure) . . . . .	96
63	Comparaison de l'évolution de la fréquence des submersions par canton en Bretagne entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite) . . . . .	98
64	Comparaison de l'évolution de la fréquence des submersions par canton en Normandie entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite) . . . . .	98
65	Comparaison de l'évolution de la fréquence des submersions par canton dans les Hauts-de-France entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite) . . . . .	99
66	Comparaison de l'évolution de la fréquence des submersions par canton dans les Pays de la Loire entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite) . . . . .	99
67	Comparaison de l'évolution de la fréquence des submersions par canton en Nouvelle-Aquitaine entre l'historique (gauche) et les scénarios 4.5 (centre) et 8.5 (droite) . . . . .	100
68	Évolution du nombre de jours d'arrêtés de submersion au cours du temps pour le scénario 4.5	100
69	Évolution du nombre de jours d'arrêtés de submersion au cours du temps pour le scénario 8.5	101
70	Illustration de l'aléa cinquantennal dans le secteur de la baie du Pouliguen (44), Source : figure 12 de l'étude CCR 2015 . . . . .	102
71	Représentation du canton contenant la commune du Pouliguen d'après le scénario RCP 4.5	103
72	Cartographie de l'évolution des valeurs assurées par département entre 2015 et 2050, Source : figure 7 de l'étude CCR 2015 . . . . .	103
73	Impact du climat et de l'élévation de la mer sur les pertes annuelles moyennes, Source : figure 23 de l'étude RCP 2018 . . . . .	104

74	Zones exposées aux submersions marines, Source : BRGM . . . . .	105
75	Zones exposées aux submersions marines à horizon 2050 selon le scénario RCP 4.5, Source : Coastal futures . . . . .	106
76	Zones exposées aux submersions marines à horizon 2050 selon le scénario RCP 8.5, Source : Coastal futures . . . . .	107
77	Zones exposées aux submersions marines à horizon 2100 selon le scénario RCP 8.5, Source : Coastal futures . . . . .	107
78	Les différents types de réassurance et leurs formes . . . . .	112
79	Sorties du modèle sur les précipitations . . . . .	119
80	Sorties du modèle sur la vitesse du vent . . . . .	119
81	Sorties du modèle sur les précipitations . . . . .	120
82	Ventilation des contributions projetées de l'ESL et de leur incertitude . . . . .	121
83	Projection des changements des contributions d'ESL selon différentes régions . . . . .	122
84	Projections des modèles individuels . . . . .	123
85	Nombre de jours de submersions marines déclarés (reconnues ou non) (gauche) et reconnues (droite) par canton sur la période d'apprentissage 1990-2005 . . . . .	129
86	Nombre de jours de submersions marines déclarés (reconnues ou non) (gauche) et reconnues (droite) par canton sur la période de validation historique 2006-2020 . . . . .	130
87	Superposition de la carte de prédiction des cantons affectés par la submersion marine à la carte de la France . . . . .	131

## Liste des tableaux

1	Comparaison des résultats sur le scénario RCP 4.5 . . . . .	6
2	Comparaison des résultats sur le scénario RCP 8.5 . . . . .	6
3	Comparison of results on the scenario RCP 4.5 . . . . .	12
4	Comparaison des résultats sur le scénario RCP 8.5 . . . . .	12
5	Principales publications sur le risque climatique au cours des 10 dernières années . . . . .	25
6	Les principaux types d'inondations . . . . .	32
7	Résumé des différents modèles et procédures employés dans le cadre des projections ESL . . . . .	58
8	Récapitulatif des différentes périodes étudiées par scénarios pour la variable ESL . . . . .	61
9	Différentes bases de données disponibles pour la modélisation . . . . .	66
10	Synthèse des avantages et inconvénients des modèles étudié . . . . .	79
11	Matrice de confusion à 3 dimensions appliquées à l'étude . . . . .	79
12	Résultats du modèle XGB pour le scénario RCP 4.5 . . . . .	87
13	Résultats du modèle XGB pour le scénario RCP 8.5 . . . . .	88
14	Résultats du modèle <i>Random Forest</i> pour le scénario RCP 4.5 . . . . .	89
15	Résultats du modèle <i>Random Forest</i> pour le scénario RCP 8.5 . . . . .	90
16	Descriptif des variables base EURO-CORDEX . . . . .	115
17	Descriptif des variables base ESL . . . . .	116

## Liste d'acronymes

- ACPR** Autorité de Contrôle Prudentiel et de Résolution. 16, 25, 42, 108, 109
- AM** Annuam Maxima. 59
- AR** Assessment Report ou modèle d'évaluation. 26, 53, 106
- BRGM** Bureau de Recherches Géologiques et Minières. 14, 34, 132
- CCR** Caisse Centrale de Réassurance. 5, 8, 11, 14, 16–18, 27–31, 38, 42, 43, 45–48, 66, 67, 71, 102–106, 133
- CMIP** Coupled Model Intercomparison Project. 18, 53, 54, 58, 119
- CORDEX** Coordinated Regional Climate Downscaling Experiment. 4, 10, 17, 18, 51, 53–55, 115, 132, 134
- EIOPA** European Insurance and Occupational Pensions Authority (Autorité Européenne des Assurances et des Pensions Professionnelles (AEAPP)). 16, 25, 26, 44
- ESG** Environnementale, Societale et de Gouvernance. 23, 25, 26
- ESL** Extreme Sea Level ou Niveau de Mer extrême. 17, 18, 56–58, 60–63, 65, 66, 91, 116, 121, 122, 132, 134
- GCM** Global Circulation Model. 17, 44, 49–54, 58, 62, 132
- GES** Gaz à Effet de Serre. 3, 20, 26, 27, 42, 102
- GEV** Generalized Extreme Value (Distribution). 59, 60
- GIEC** Groupe d'experts Intergouvernemental sur l'Evolution du Climat. 1, 3, 16, 20, 25–27, 42, 49, 53, 95, 102, 106, 108, 132
- MARC** Modélisation et Analyse pour la Recherche Côtière. 40, 41
- ORSA** Own Risk and Solvency Assessment (processus interne d'évaluation des risques et de la solvabilité par l'organisme). 16, 25, 26, 42, 44
- RCM** Regional Circulation Model. 5, 11, 17, 44, 51–55, 62, 109, 132
- RCP** Representative Concentration Pathway. 1–12, 14, 18, 20, 21, 26, 27, 42, 43, 51, 53–55, 61–63, 66, 71, 80–83, 85–90, 92, 95, 96, 98, 102–108, 121, 132–134
- SSP** *Shared Socioeconomic Pathways*. 26, 27, 42, 43, 108
- SVM** *Support Vector Machines*. 5, 6, 11, 12, 18, 69, 71, 76–79, 83–85, 132
- TVE** Théorie des Valeurs Extrêmes. 58–60
- XGB** *Gradient Boosting*. 5, 6, 11, 12, 17, 71, 79, 86–88, 109, 133, 134