

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

le

Par : Nicxan HENSMAN STALIN

Titre : Modélisation d'une carte d'exposition au risque ouragan aux Etats-Unis

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

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

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

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

Entreprise :

Nom : Exiom Partners

Signature :



EXIOM PARTNERS SAS
24-26, rue Notre Dame des Victoires
75002 PARIS
SIRET : 851 841 944 00034
APE 7022Z RCS Paris

Directeur de Mémoire en entreprise :

Nom : Chadi HANNA

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 :

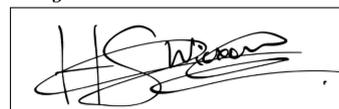
Bibliothèque :

Signature du responsable entreprise



EXIOM PARTNERS SAS
24-26, rue Notre Dame des Victoires
75002 PARIS
SIRET : 851 841 944 00034
APE 7022Z RCS Paris

Signature du candidat



Résumé

Le risque cyclonique aux États-Unis, en perpétuelle mutation face aux défis climatiques croissants, demeure un sujet de préoccupation majeur. Cette étude consiste à mieux appréhender le comportement complexe des ouragans dans le bassin Atlantique Nord. En effet, ces phénomènes météorologiques ont un impact dévastateur sur les populations, les biens et l'économie, justifiant ainsi la nécessité de mieux comprendre leur dynamique pour une gestion des risques plus efficace. L'objectif de cette étude est de fournir des outils précieux à l'industrie de l'assurance et aux décideurs, afin d'améliorer l'évaluation et la gestion des risques cycloniques. Elle met également en lumière les défis croissants auxquels le secteur de l'assurance lié aux ouragans aux États-Unis est confronté, tels que l'augmentation des primes d'assurance habitation, le retrait de certains assureurs et les répercussions financières des ouragans récents.

La modélisation des saisons cycloniques se décline en trois phases distinctes : la fréquence, la trajectoire et l'intensité des tempêtes. Chaque phase repose sur une méthodologie rigoureuse, élaborée à partir d'une revue exhaustive des modèles statistiques existants. Cette approche vise à créer un cadre solide pour la compréhension de la formation, du déplacement et de l'intensification des ouragans.

Pour estimer les pertes financières liées aux ouragans, une base de données des coûts a été constituée, répertoriant les ouragans ayant occasionné plus d'un milliard de dollars de dommages par comté aux États-Unis entre 1982 et 2022. Des modèles statistiques avancés, notamment XG-Boost, ont été employés pour simuler les pertes annuelles des saisons cycloniques. Cette approche a permis d'estimer les pertes annuelles moyennes par comté et de mener une analyse approfondie du risque cyclonique dans les régions les plus exposées aux États-Unis, en particulier en explorant des scénarios extrêmes.

Mots-clés : Risque cyclonique, Ouragan, États-Unis, Assurabilité, Indice de risque national, Défis climatiques

Abstract

Cyclonic Risk in the United States, constantly evolving in the face of growing climate challenges, remains a major cause for concern. This study aims to gain a better understanding of the complex behavior of hurricanes in the North Atlantic basin. Indeed, these meteorological phenomena have a devastating impact on populations, property, and the economy, justifying the need for a deeper understanding of their dynamics for more effective risk management. The objective of this study is to provide valuable tools to the insurance industry and policymakers to enhance the assessment and management of cyclonic risks. It also sheds light on the growing challenges facing the hurricane-related insurance sector in the United States, such as rising home insurance premiums, the withdrawal of some insurers, and the financial repercussions of recent hurricanes.

The modeling of cyclonic seasons unfolds in three distinct phases: frequency, trajectory, and intensity of storms. Each phase relies on a rigorous methodology developed through a comprehensive review of existing statistical models. This approach aims to create a solid framework for understanding hurricane formation, movement, and intensification.

To estimate the financial losses associated with hurricanes, a cost database was compiled, listing hurricanes that caused more than \$1 billion in damage per county in the USA between 1982 and 2022. Advanced statistical models, including XGBoost, were used to simulate annual losses during hurricane seasons. This approach made it possible to estimate average annual losses per county, and to conduct an in-depth analysis of hurricane risk in the most exposed regions of the United States, in particular by exploring extreme scenarios.

Keywords : Cyclone Risk, Hurricane, United States, Insurability, National Risk Index, Climate Challenges

Note de Synthèse

Contexte et objectifs

L'étude s'inscrit dans un contexte où les ouragans, en tant que phénomènes météorologiques redoutables, suscitent un intérêt croissant tant sur le plan scientifique que dans la gestion des risques. Les États-Unis, en particulier les États côtiers du golfe du Mexique et de la côte Est, demeurent particulièrement vulnérables aux ravages causés par ces tempêtes tropicales, une vulnérabilité accentuée par la tendance à l'augmentation de l'intensité des ouragans liée au réchauffement climatique observé au cours des dernières décennies (INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, 2023 ; KOSSIN et al., 2020).

Le risque cyclonique est la catastrophe naturelle la plus coûteuse pour les États-Unis, avec des coûts totaux pouvant atteindre jusqu'à des centaines de milliards de dollars pour un seul ouragan (INSURANCE INFORMATION INSTITUTE, 2023b), soulignant ainsi l'importance cruciale de cette problématique, notamment pour l'industrie de l'assurance. De plus ces dernières décennies sont marquées par une augmentation très marquante des coûts liés aux ouragans. En moyenne, un ouragan coûte près de 50 milliards de dollars depuis la dernière décennie, une hausse significative par rapport aux années 80, où le coût moyen était inférieur à 10 milliards de dollars (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2022).

Face à des défis croissants tels que l'augmentation des coûts liés aux ouragans, les retraits d'assureurs des zones les plus exposées et l'augmentation des primes, cette étude poursuit plusieurs objectifs. En premier lieu, elle vise à évaluer de manière précise et rigoureuse l'exposition au risque d'ouragan aux États-Unis et à identifier les facteurs contribuant à cette exposition, fournissant ainsi des outils essentiels pour quantifier et mesurer l'exposition du portefeuille d'un assureur. En second lieu, l'étude s'emploie à orienter la politique de souscription des assureurs envers différentes zones en identifiant les régions les plus vulnérables. Enfin, elle ambitionne d'apporter des premiers éléments de réponse quant à l'assurabilité de certaines zones, mettant en lumière les régions à risque élevé qui pourraient nécessiter une réévaluation des politiques d'assurance.

Simulation des saisons cycloniques

La méthodologie de cette étude repose sur des étapes cruciales, débutant par l'utilisation de la base de données mondiale IBTrACS (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b), fondamentale pour modéliser 10 000 saisons cycloniques afin de capturer la diversité des phénomènes cycloniques et fournir une représentation complète des risques potentiels aux États-Unis. La base utilisée couvre le bassin Atlantique Nord sur les 40 dernières années, de 1982 à 2022, en tenant compte des avancées technologiques et de l'impact du réchauffement climatique depuis les années 1980. Chaque étape de la modélisation fait l'objet d'une revue approfondie des modèles utilisés dans la littérature scientifique, permettant ainsi de situer l'approche adoptée dans le contexte de cette étude.

Fréquence annuelle

La première étape consiste en une analyse de la fréquence annuelle des tempêtes cycloniques. Pour ce faire, une loi binomiale négative a été mis en oeuvre de paramètres $\hat{m}_{EMV} = 15.8$ et $\hat{p}_{EMV} = 0.58$ générant 15.7 tempêtes par an en moyenne sur 10 000 saisons cycloniques, en accord avec l'observation historique de 15.8 tempêtes par an.

Points de genèse

La simulation des points de genèse des tempêtes est réalisée via un lissage spatial par noyau gaussien, intégrant les données historiques pour créer une distribution spatiale potentielle. L'approche considère les facteurs géographiques et environnementaux propices aux tempêtes, avec une hyperparamétrisation basée sur la validation croisée pour déterminer la fenêtre de lissage optimale.

Trajectoire

La modélisation des trajectoires des tempêtes repose sur des processus auto-régressifs afin de saisir les variations longitudinales et latitudinales historiques, avec une flexibilité pour anticiper les évolutions futures. Le modèle intègre des termes linéaires et non linéaires pour représenter les déplacements des tempêtes, notamment l'éloignement de l'Équateur. Les coefficients sont estimés par tranche de 5 degrés de latitude, engendrant 22 modèles de régression, 11 chacun pour la latitude (ϕ_t) et la longitude (ξ_t) représentés à travers les équations suivantes

$$\Delta\xi_t = a_0 + a_1\Delta\xi_{t-1} + \varepsilon_t^\xi,$$

$$\Delta\phi_t = b_0 + b_1\Delta\phi_{t-1} + \frac{b_2}{\phi_{t-1}} + \varepsilon_t^\phi.$$

La durée de vie des tempêtes est simulée en fonction de leur lieu de genèse, avec deux zones identifiées : près de la côte ouest africaine et en dehors de cette zone. La fin de la trajectoire d'une tempête est liée à son intensité : si elle n'atteint pas sa durée maximale et perd en intensité en touchant la terre jusqu'à un seuil spécifique (fixé à 68 km/h), elle est considérée comme éteinte. La figure 1 compare 50 trajectoires simulées avec 50 trajectoires empiriques tirées de la base de données IBTrACS. Les trajectoires simulées reproduisent fidèlement l'aspect parabolique caractéristique des trajectoires de tempêtes historiques.

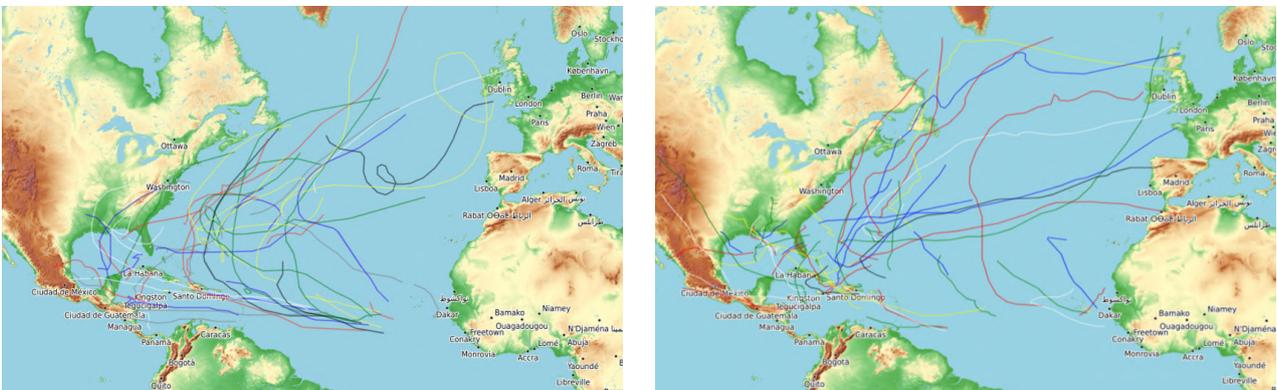


FIGURE 1 : Echantillon de 50 trajectoires empiriques (à gauche) et simulées (à droite)

Intensité

Pression Centrale

La puissance d'une tempête tropicale est principalement déterminée par son déficit de pression, exprimé comme la différence entre la pression atmosphérique ambiante et la pression au centre de la tempête, mesurée au niveau de la mer. Deux modèles ont été utilisés pour modéliser la pression centrale en milieu océanique : le modèle de VICKERY et al. (2000), basé sur l'intensité relative, prédit le logarithme de cette mesure en fonction de données antérieures et de la température de surface de l'eau. Un deuxième modèle de CUI et CARACOGLIA (2019) utilise un processus autorégressif pour modéliser les variations de l'intensité relative. Les coefficients sont estimés à une résolution de $5^\circ \times 5^\circ$ dans le bassin Atlantique Nord, puis interpolés bilinéairement en $1^\circ \times 1^\circ$. En matière de provision pour le risque ouragan en assurance, le modèle de CUI et CARACOGLIA (2019) est privilégié en raison de sa prudence systématique face à la progression des catégories d'ouragans (inférieur à 980 hPa). En environnement terrestre, la décroissance exponentielle de la pression centrale post-atterrissage est modélisée selon KAPLAN et DEMARIA (1995). Ce modèle considère l'évolution du déficit de pression depuis l'atterrissage, le temps écoulé depuis cet instant, et une constante de désintégration. La modélisation de cette constante est réalisée pour quatre zones distinctes : Atlantique Nord, Golfe du Mexique, Floride et Mer des Caraïbes.

Vitesse maximale des vents

Les études scientifiques montrent que la vitesse maximale des vents augmente à mesure que la pression centrale diminue (HOLLAND, 2008). Ainsi le modèle utilisé dans le cadre de cette étude est un modèle de régression polynomiale d'ordre 2 basé sur le déficit de pression centrale. Les résultats montrent que le modèle est en mesure de saisir de manière précise les caractéristiques spécifiques à chaque catégorie de pression. Il est cependant remarquable que plusieurs points sont surestimés, en particulier pour les catégories d'ouragans majeurs, ce qui est encourageant car cela traduit une certaine prudence maintenue par le modèle.

Rayon de la vitesse maximale des vents

Un modèle novateur, proposé par DANIEL et JOHN (2022), intègre des concepts physiques fondamentaux pour modéliser le rayon de la vitesse maximale des vents. En utilisant une régression linéaire avec trois variables clés (la vitesse maximale des vents, la latitude et la moyenne des rayons maximaux atteints par les vents de 17,5 m/s dans chaque quadrant), le modèle démontre une bonne concordance avec les rayons empiriques, notamment pour les valeurs inférieures à 100 km, caractéristiques des ouragans de forte intensité. Bien que l'adéquation soit moins marquée pour les rayons plus grands, le modèle conserve une approche prudente en les sous-estimant.

Gradient de vitesse

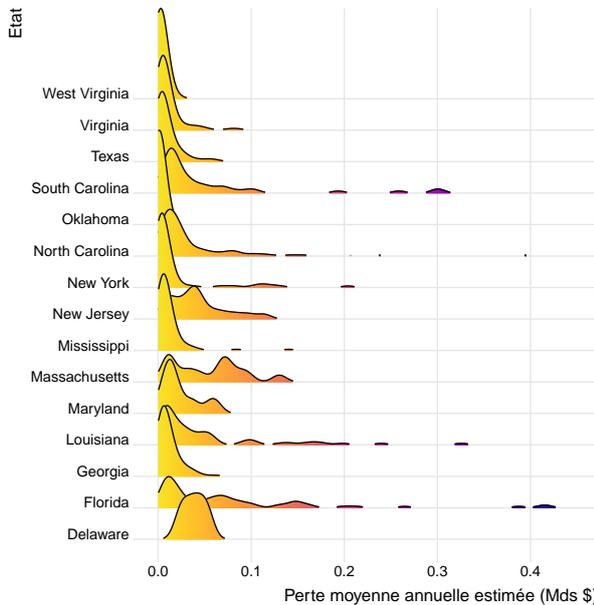
Cette phase vise à élaborer un modèle déterminant la vitesse maximale des vents sur toute la superficie d'un ouragan à un moment précis, exprimant le gradient de la vitesse maximale des vents pour une position donnée. Le modèle de gradient de vitesse choisi, développé par WILLOUGHBY et al. (2006), se divise en trois phases distinctes : une phase initiale modélisant l'augmentation de la vitesse des vents en fonction d'une puissance du rayon à l'intérieur de l'œil, une deuxième phase présentant une transition polynomiale douce à travers le mur de l'œil, suivie d'une troisième phase de décroissance exponentielle à l'extérieur de l'œil.

Estimation et analyse des pertes financières associées

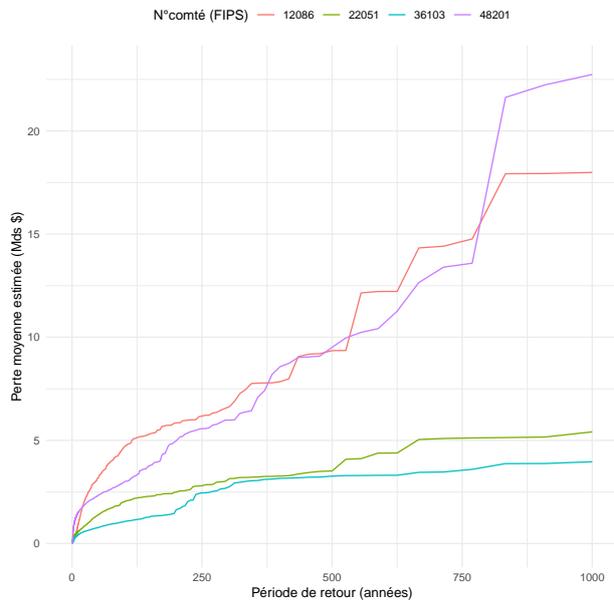
Cette dernière étape de la méthodologie se concentre sur la simulation des pertes financières liées aux ouragans. Une base de données des coûts a été établie, recensant les ouragans ayant engendré plus d'un milliard de dollars de dommages par comté aux États-Unis entre 1982 et 2022. Les coûts, initialement calculés par État, sont adaptés par comté à l'aide d'un indice de risque ouragan, noté EAL , représentant la perte économique annuelle moyenne d'un comté due aux aléas naturels, y compris le risque d'ouragan (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023c). Ainsi, avec i représentant un comté et j l'État dans lequel se situe le comté, la formule suivante est utilisée

$$\text{Coût}_i^j = \text{Coût}_j \times \frac{EAL_i^j}{\sum_{k=1}^N EAL_k^j},$$

où Coût_i^j représente le coût total réel dans le comté i de l'État j , la variable d'intérêt de l'étude, Coût_j représente le coût total réel dans l'État j , EAL_i^j la perte économique moyenne annuelle associée, et N le nombre de comtés dans l'État j . Ensuite, la base de données est enrichie de variables explicatives liées aux caractéristiques spécifiques des ouragans et des comtés. Pour modéliser une fonction de perte due aux ouragans, le modèle XGBoost a été identifié comme le plus performant et retenu pour simuler les pertes annuelles des saisons cycloniques. Cependant malgré une qualité de prédiction globalement correcte, le modèle XGBoost montre des performances non robustes, se manifestant par une tendance à sous-estimer les valeurs prédites, en particulier pour les événements à coûts élevés. Cette sous-estimation, préoccupante en gestion des risques en assurance, pourrait être attribuée au nombre limité de variables dans la base d'entraînement. Bien que l'intensité des ouragans et l'exposition des comtés aux risques d'ouragan soient incluses, l'ajout de variables significatives pourrait être nécessaire. En revanche l'analyse de l'importance des variables souligne le poids des caractéristiques liées à l'intensité des ouragans, telles que la vitesse maximale des vents, sur la variable cible, en cohérence avec l'idée que l'intensité d'un ouragan est directement liée aux dommages et, par conséquent, aux coûts finaux.



(a) Distribution des pertes moyennes annuelles estimées par Etat



(b) Courbes AEP de 4 comtés aux Etats-Unis

Ensuite, la perte moyenne annuelle est calculée en agrégeant les coûts bruts totaux de tous les ouragans pour chaque saison cyclonique, puis en calculant la moyenne de ces sommes sur l'ensemble

des saisons. La figure 2a compare les distributions des pertes moyennes annuelles entre les États, mettant en évidence ceux les plus exposés au risque d'ouragan, tels que la Floride et la Louisiane, avec des pertes moyennes particulièrement élevées. Cette observation s'aligne de manière cohérente avec les constatations actuelles sur l'impact cyclonique aux États-Unis.

La figure 2b représente la courbe AEP du comté affichant la perte moyenne annuelle la plus élevée dans différents États. Les comtés, identifiés par leur code FIPS, composé des deux premiers chiffres représentant l'État, reflètent des profils de risque cyclonique distincts au sein des États de la Floride (12), de la Louisiane (22), du Texas (48) et de New York (36). La localisation géographique des assurés devient cruciale pour les assureurs, car l'exposition au risque varie considérablement d'un comté à l'autre. Par exemple, assurer des biens dans le comté de la Floride entraînerait un SCR proche de 6 milliards de dollars, tandis qu'une couverture similaire en Louisiane se traduirait par un SCR d'environ 2.5 milliards de dollars. Les comtés de la Floride et du Texas affichent également des quantiles extrêmes nettement plus élevés que ceux de la Louisiane et de New York.

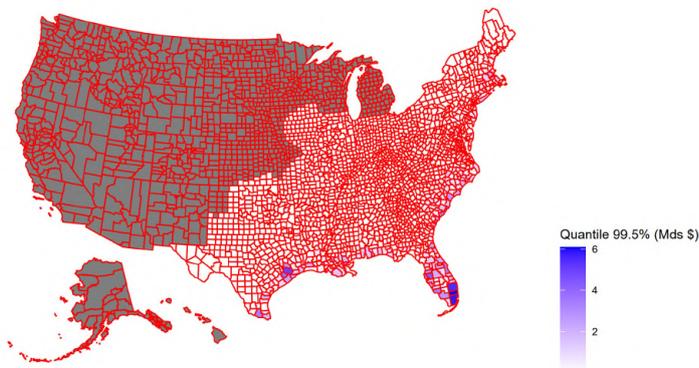


FIGURE 3 : Quantile de niveau 99.5% de la perte moyenne total estimée aux États-Unis

L'analyse cartographique en figure 3 révèle de manière significative que les comtés affichant le quantile de niveau de 99.5% le plus élevé sont localisés le long des côtes du golfe du Mexique et du bassin Atlantique Nord, particulièrement au sein des régions de la Floride, de la Louisiane, du Texas, ainsi que des Caroline du Nord et du Sud. Il est intéressant de souligner que les comtés en Floride se distinguent par les quantiles les plus élevés, oscillant entre 2 et 6 milliards de dollars, suivi de près par ceux de la Louisiane et du Texas, où certains comtés côtiers enregistrent des quantiles se rapprochant de 4 milliards de dollars. Enfin l'estimation du coût moyen annuel par foyer et

par comté est réalisée selon la formule suivante

$$\text{Coût moyen annuel par foyer} = \frac{\text{Perte moyenne annuelle au sein du comté}}{\text{Nombre de foyer au sein du comté}}$$

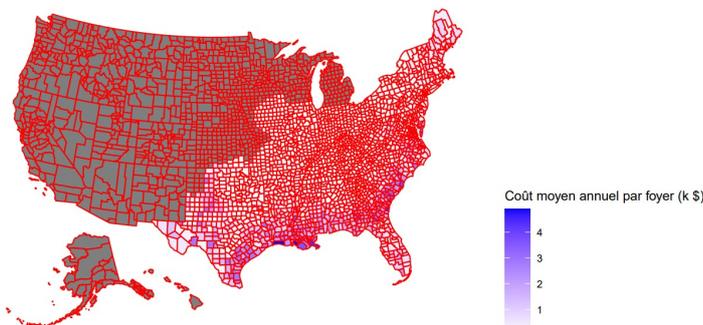
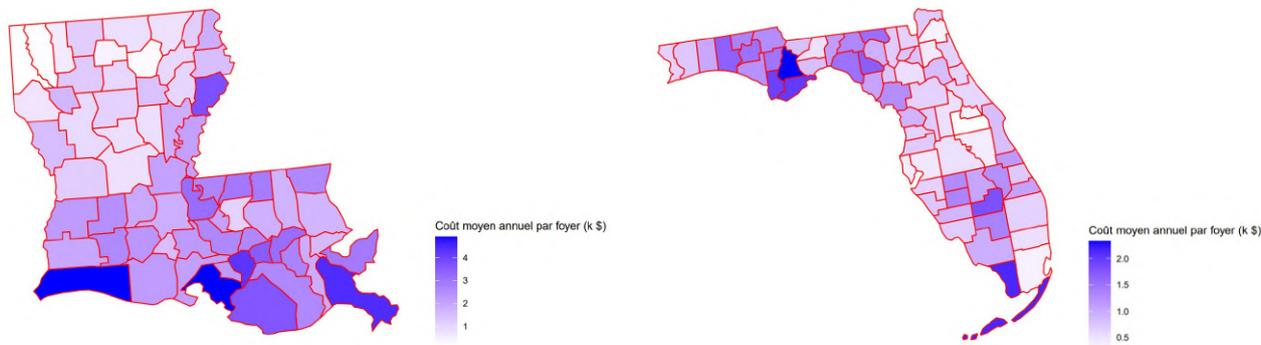


FIGURE 4 : Coût moyen annuel estimé par comté

présentant des coûts moyens annuels particulièrement élevées. Selon UHLIG (2023), le coût moyen d'une couverture d'assurance complète contre les ouragans aux États-Unis s'élève à environ 2 555 dollars par an. L'approche reproduit cette moyenne observée avec des coûts moyens annuels par foyer variant

Il est important de souligner que **cette approche ne correspond pas strictement à la réalité de la tarification en assurance MRH avec des primes individualisées**, notamment en raison de l'hypothèse d'une souscription obligatoire de tous les foyers au sein d'un comté. Ainsi, le coût moyen annuel calculé s'apparente davantage à un coût mutualisé à l'échelle du comté, plutôt qu'à une prime d'assurance au sens usuel. La figure 4 met en évidence des disparités significatives entre les comtés, notamment ceux côtiers du golfe du Mexique et de la côte du bassin Atlantique,

jusqu'à environ 5 000 dollars. Cependant, une limite de cette estimation réside dans son lien avec l'estimation préalable de la perte moyenne annuelle par comté, dérivée du modèle XGBoost. En effet la variable représentant la population exposée au risque d'ouragan par comté a un impact limité sur la prédiction du coût. Ainsi, les prédictions de coût sont moins précises pour les comtés à population moins élevée, en particulier dans l'État du Texas.



(a) Coût moyen annuel estimé en Louisiane

(b) Coût moyen annuel estimé en Floride

Pour la Louisiane (figure 5a), les coûts moyens annuels par foyer s'élèvent à 4 000 à 5 000 dollars le long de la côte du golfe du Mexique, reflétant une exposition élevée aux ouragans. En 2022, les primes d'assurance ont augmenté pour atteindre en moyenne 4 700 dollars par an. En Floride (figure 5b), malgré des pertes moyennes annuelles élevées, les coûts estimés par foyer (jusqu'à 3 000 dollars) restent relativement abordables en raison de la densité de population élevée, malgré le risque élevé d'ouragan. Une préoccupation majeure des assureurs dans ces zones à risques n'est pas tant la perte moyenne occasionnée par ces phénomènes dévastateurs mais plutôt la considérable variance de ces pertes moyennes, entraînant des faillites d'assureurs en Floride, avec par exemple plus de 15 compagnies quittant l'État après l'ouragan Laura en 2020. Les comtés du Golfe du Mexique, particulièrement en Floride, présentent des pertes maximales estimées pouvant atteindre 40 milliards de dollars selon les scénarios simulés. Cette variabilité conduit à une hausse des primes, l'*Insurance Information Institute* estimant actuellement une prime d'assurance habitation moyenne en Floride à environ 6 000 dollars par an, bien au-dessus de la moyenne nationale de 1 700 dollars aux États-Unis (GABE et al., 2023). Face à cette augmentation des coûts et à la menace persistante des phénomènes cycloniques, des mesures potentielles sont envisagées, telles que l'augmentation de la franchise, la réduction de la concentration du risque, et la promotion de mesures de protection, visant à équilibrer les coûts croissants des catastrophes naturelles tout en maintenant l'accessibilité à l'assurance pour les résidents des zones à risque.

Limites et pistes d'amélioration

En conclusion, cette étude souligne les défis inhérents à la modélisation du risque cyclonique, appelant à explorer des approches plus avancées comme l'intégration de modèles de Machine Learning ou de Deep Learning. L'absence de considération de la dépendance entre les variations des trajectoires et l'intensité des ouragans constitue une limite, tout comme l'omission des projections climatiques futures du GIEC (IPCC, 2022). Dans la modélisation des coûts financiers, des simplifications au niveau des comtés ont été nécessaires pour rendre le modèle praticable dans le contexte assurantiel. L'exploration de modèles de Machine Learning plus sophistiqués, tels que les réseaux de neurones, pourrait offrir des solutions aux défis liés à l'estimation des dommages. De plus, l'influence des facteurs externes sur les coûts, tels que les politiques gouvernementales, mérite une exploration approfondie, incitant à intégrer un ensemble plus complet de variables explicatives.

Synthesis note

Context and objectives

The study takes place against a backdrop of growing scientific and risk management interest in hurricanes as severe weather phenomena. The United States, in particular the coastal states of the Gulf of Mexico and the East Coast, remain particularly vulnerable to the devastation caused by these tropical storms, a vulnerability accentuated by the trend towards increased hurricane intensity linked to global warming observed over the last few decades Intergovernmental Panel on Climate Change (2023) and Kossin et al. (2020).

Hurricane risk is the most costly natural disaster for the United States, with total costs reaching hundreds of billions of dollars for a single hurricane, underlining the crucial importance of this issue, particularly for the insurance industry. In recent decades, hurricane-related costs have risen sharply. On average, a hurricane has cost almost \$50 billion over the last decade, a significant increase on the 1980s, when the average cost was less than \$10 billion (National Oceanic and Atmospheric Administration, 2022).

Faced with increasing challenges such as rising hurricane-related costs, insurer withdrawals from the most exposed areas and rising premiums, this study has several objectives. Firstly, it aims to accurately and rigorously assess hurricane risk exposure in the United States and identify the factors contributing to this exposure, thus providing essential tools for quantifying and measuring the exposure of an insurer's portfolio. Secondly, the study aims to guide insurers' underwriting policies towards different zones by identifying the most vulnerable regions. Finally, it aims to provide some initial answers as to the insurability of certain zones, highlighting high-risk regions that may require a reassessment of insurance policies.

Simulating cyclonic seasons

The methodology of this study is based on crucial steps, starting with the use of the IBTrACS global database (National Oceanic and Atmospheric Administration, 2023b), fundamental to model 10,000 cyclonic seasons in order to capture the diversity of cyclonic phenomena and provide a comprehensive representation of potential risks in the United States. The database used covers the North Atlantic basin over the last 40 years, from 1982 to 2022, taking into account technological advances and the impact of global warming since the 1980s. Each step of the modelling process is reviewed in detail, to situate the approach adopted in the context of this study.

Annual frequency

The first step is to analyze the annual frequency of cyclonic storms. To this end, a negative binomial distribution was implemented with parameters $\hat{m}_{EMV} = 15.8$ and $\hat{p}_{EMV} = 0.58$ generating 15.7 storms per year on average over 10 000 cyclonic seasons, in line with historical observation of 15.8 storms per year.

Genesis points

Simulation of storm genesis points is performed via spatial Gaussian kernel smoothing, integrating historical data to create a potential spatial distribution. The approach considers geographic and environmental factors conducive to storms, with hyperparameterization based on cross-validation to determine the optimal smoothing window.

Trajectory

The modeling of storm tracks is based on autoregressive processes to capture historical longitudinal and latitudinal variations, with flexibility to anticipate future evolutions. The model incorporates linear and non-linear terms to represent storm displacements, including distance from the Equator. Coefficients are estimated in 5-degree latitude steps, generating 22 regression models, 11 each for latitude (ϕ_t) and longitude (ξ_t) represented by the following equations

$$\Delta\xi_t = a_0 + a_1\Delta\xi_{t-1} + \varepsilon_t^\xi,$$

$$\Delta\phi_t = b_0 + b_1\Delta\phi_{t-1} + \frac{b_2}{\phi_{t-1}} + \varepsilon_t^\phi.$$

The lifetime of storms is simulated as a function of where they originate, with two zones identified: near the West African coast and outside this zone. The end of a storm's trajectory is linked to its intensity: if it does not reach its maximum duration, and loses intensity as it makes landfall up to a specific threshold (set at 68 km/h), it is considered extinct. The figure 6 compares 50 simulated trajectories with 50 empirical trajectories taken from the IBTrACS database. The simulated trajectories faithfully reproduce the parabolic aspect characteristic of historical storm tracks.



Figure 6: Sample of 50 empirical (left) and simulated (right) trajectories

Intensity

Central pressure

The power of a tropical storm is mainly determined by its pressure deficit, expressed as the difference between the ambient atmospheric pressure and the pressure at the center of the storm, measured at sea level. Two models have been used to model central pressure in the ocean environment: the Vickery et al. (2000) model, based on relative intensity, predicts the logarithm of this measure as a function of past data and water surface temperature. A second Cui and Caracoglia (2019) model uses an autoregressive process to model variations in relative intensity. Coefficients are estimated at a resolution of $5^\circ \times 5^\circ$ in the North Atlantic basin, then bilinearly interpolated in $1^\circ \times 1^\circ$. For insurance hurricane risk provisioning, the Cui and Caracoglia (2019) model is favored for its systematic conservatism in the face of increasing hurricane categories (below 980 hPa). In a terrestrial environment, the exponential decay of post-landing central pressure is modelled according to Kaplan and DeMaria (1995). This model considers the evolution of the pressure deficit since landing, the time elapsed since that moment, and a decay constant. This constant is modeled for four distinct zones: North Atlantic, Gulf of Mexico, Florida and Caribbean Sea.

Maximum wind speed

Scientific studies show that maximum wind speed increases as central pressure decreases. Thus, the model used in this study is a 2nd-order polynomial regression model based on the central pressure deficit. The results show that the model is able to accurately capture the specific characteristics of each pressure category. However, it is notable that several points are overestimated, particularly for the major hurricane categories, which is encouraging as it reflects a degree of caution maintained by the model.

Range of Maximum Wind Speed

An innovative model, proposed by Daniel and John (2022), integrates fundamental physical concepts to model the radius of maximum wind speed. Using linear regression with three key variables (maximum wind speed, latitude and average maximum radii reached by 17.5 m/s winds in each quadrant), the model demonstrates good agreement with empirical radii, particularly for values below 100 km, characteristic of high-intensity hurricanes. Although the fit is less marked for larger radii, the model maintains a conservative approach by underestimating them.

Speed gradient

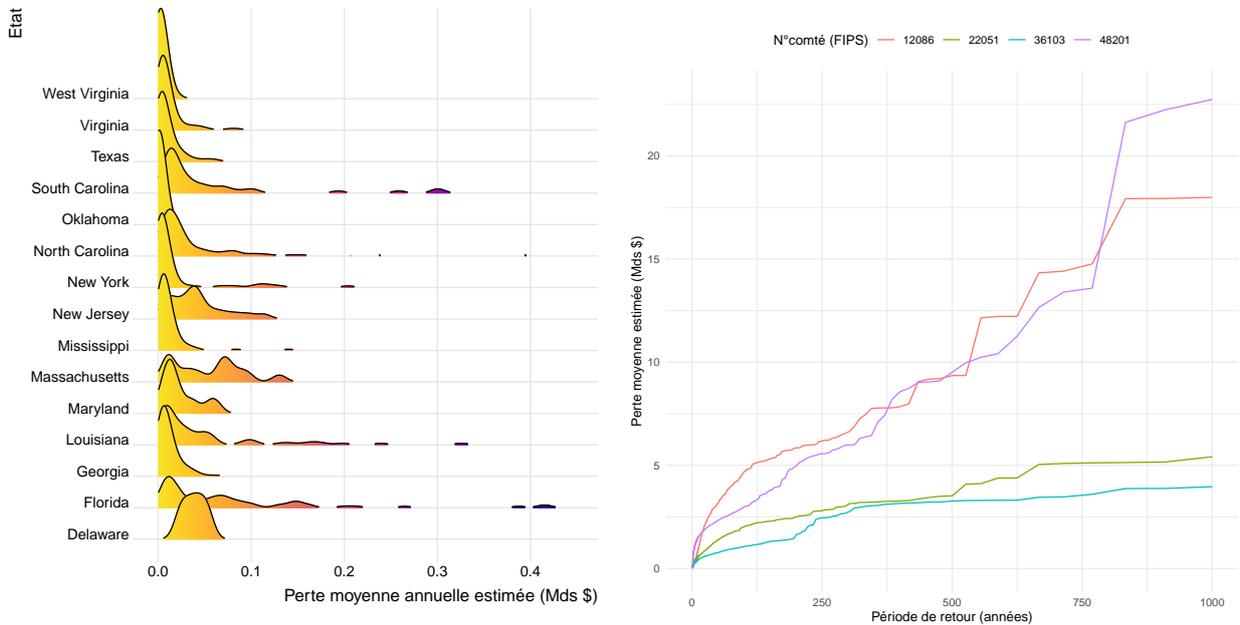
The aim of this phase is to develop a model that determines the maximum wind speed over the entire area of a hurricane at a given time, expressing the maximum wind speed gradient for a given position. The chosen velocity gradient model, developed by Willoughby et al. (2006), is divided into three distinct phases: an initial phase modeling the increase in wind speed as a function of radius power inside the eye, a second phase presenting a smooth polynomial transition across the eye wall, followed by a third phase of exponential decay outside the eye.

Estimation and analysis of associated financial losses

This final stage of the methodology focuses on simulating the financial losses associated with hurricanes. A database of costs has been established, listing hurricanes that generated more than \$1 billion in damage per county in the United States between 1982 and 2022. The costs, initially calculated by state, are adapted by county using a hurricane risk index, denoted EAL , representing a county's average annual economic loss due to natural hazards, including hurricane risks (Federal Emergency Management Agency, 2023c). Thus, with i representing a county and j the state in which the county is located, the following formula is used

$$\text{Cost}_i^j = \text{Cost}_j \times \frac{EAL_i^j}{\sum_{k=1}^N EAL_k^j},$$

where Cost_i^j represents the actual total cost in county i of state j , the variable of interest in the study, Cost_j represents the actual total cost in state j , EAL_i^j the associated average annual economic loss, and N the number of counties in state j . The database is then enriched with explanatory variables linked to the specific characteristics of hurricanes and counties. To model a hurricane loss function, the XGBoost model was identified as the best performer and selected to simulate annual hurricane season losses. However, despite its generally good predictive quality, the XGBoost model's performance is not robust, as evidenced by its tendency to underestimate predicted values, particularly for high-cost events. This underestimation, which is a cause for concern in insurance risk management, could be attributed to the limited number of variables in the training base. Although hurricane intensity and county exposure to hurricane risk are included, the addition of significant variables may be necessary. On the other hand, the analysis of the importance of variables highlights the weight of characteristics linked to hurricane intensity, such as maximum wind speed, on the target variable, consistent with the idea that hurricane intensity is directly linked to damage and, consequently, to final costs.



(a) Distribution of estimated average annual losses by state

(b) AEP curves for 4 counties in the United States

Next, the average annual loss is calculated by aggregating the total gross costs of all hurricanes for each hurricane season, then averaging these sums over all seasons. Figure 7a compares the distributions of annual average losses between states, highlighting those most exposed to hurricane risk, such as

Florida and Louisiana, with particularly high average losses. This observation aligns consistently with current findings on hurricane impact in the USA.

Figure 7b shows the AEP curve for the county with the highest average annual loss in different states. The counties, identified by their FIPS code, made up of the first two digits representing the state, reflect distinct hurricane risk profiles within the states of Florida (12), Louisiana (22), Texas (48) and New York (36). The geographical location of policyholders becomes crucial for insurers, as risk exposure varies considerably from one county to another. For example, insuring property in Florida County would result in an SCR close to \$6 billion, while similar coverage in Louisiana would result in an SCR of around \$2.5 billion. Florida and Texas counties also show significantly higher extreme quantiles than Louisiana and New York.

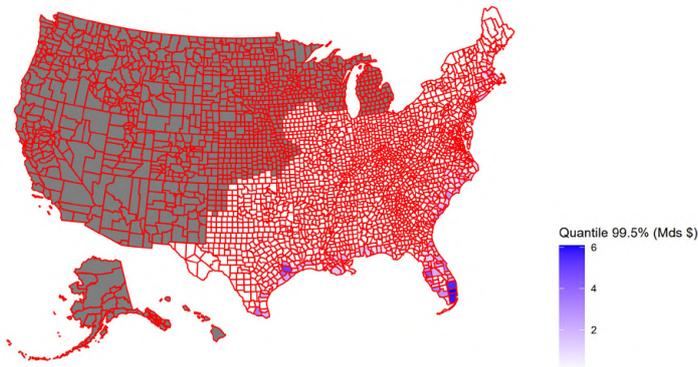


Figure 8: Quantile level 99.5% of estimated total average loss in the U.S.

The map analysis in figure 8 shows that the counties with the highest 99.5% quantile level are located along the coasts of the Gulf of Mexico and the North Atlantic basin, particularly in the regions of Florida, Louisiana, Texas, North Carolina and South Carolina. It is interesting to note that counties in Florida stand out with the highest quantiles, oscillating between 2 and 6 billion dollars, closely followed by those in Louisiana and Texas, where certain coastal counties record quantiles approaching 4 billion dollars. Finally, the average annual cost per household and per county is estimated using the following formula

$$\text{Annual average cost per household} = \frac{\text{Annual average loss within the county}}{\text{Number of households within the county}}$$

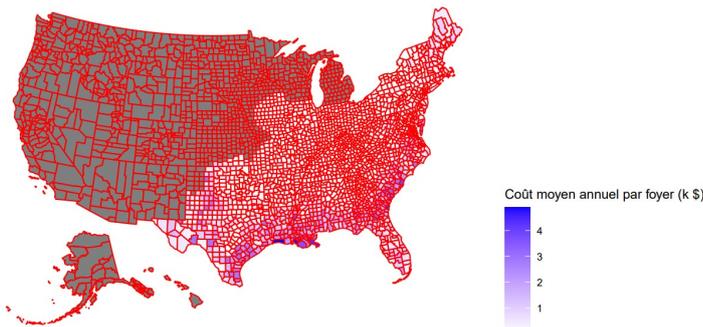
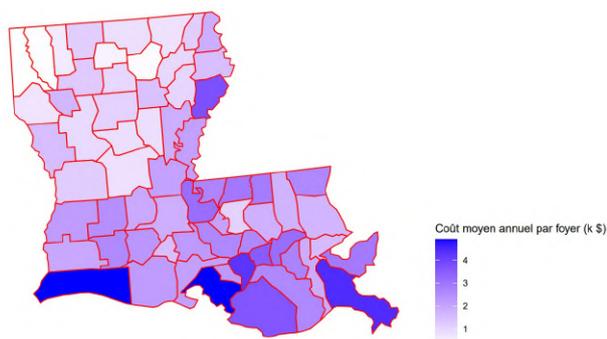


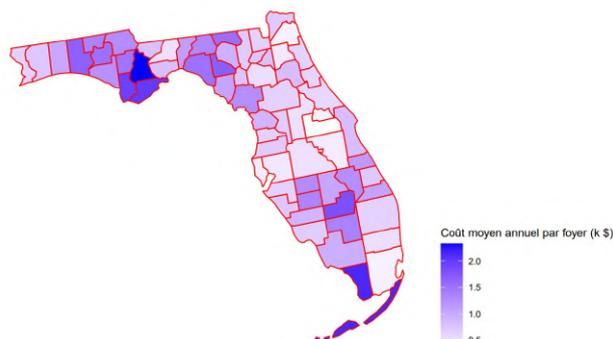
Figure 9: Estimated average annual cost per county

According to Uhlig (2023), the average cost of comprehensive hurricane insurance coverage in the United States is around \$2,555 per year. The approach reproduces this observed average, with average annual costs per household ranging up to around \$5,000. However, a limitation of this estimate lies in its link with the prior estimate of the average annual loss per county, derived from the XGBoost model. Indeed, the variable representing the population exposed to hurricane risk by county has a limited impact on cost prediction. As a result, cost predictions are less accurate for counties with lower populations, particularly in the state of Texas.

It's important to stress that **this approach doesn't strictly correspond to the reality of household insurance pricing with individualized premiums**, notably due to the assumption of compulsory subscription by all households within a county. As a result, the average annual cost calculated is more akin to a county-wide mutualized cost, rather than an insurance premium in the usual sense. Figure 9 highlights significant disparities between counties, particularly those on the Gulf Coast and Atlantic Basin coast, with particularly high average annual costs.



(a) Estimated average annual cost in Louisiana



(b) Estimated average annual cost in Florida

For Louisiana (figure 10a), average annual costs per household are \$4,000 to \$5,000 along the Gulf Coast, reflecting high exposure to hurricanes. By 2022, insurance premiums have risen to an average of \$4,700 per year. In Florida (figure 10b), despite high average annual losses, estimated costs per household (up to \$3,000) remain relatively affordable due to high population density, despite high hurricane risk. A major concern for insurers in these high-risk areas is not so much the average loss caused by these devastating phenomena, but rather the considerable variance in these average losses, leading to insurer bankruptcies in Florida, with over 15 companies leaving the state after Hurricane Laura in 2020, for example. Gulf of Mexico counties, particularly in Florida, have estimated maximum losses of up to \$40 billion, depending on the scenarios simulated. This variability is leading to higher premiums, with the Insurance Information Institute currently estimating an average home insurance premium in Florida at around \$6,000 per year, well above the U.S. national average of \$1,700 (Gabe et al., 2023). Faced with these rising costs and the persistent threat of cyclonic phenomena, potential measures are being considered, such as increasing deductibles, reducing risk concentration, and promoting protective measures, aimed at balancing the rising costs of natural disasters while maintaining accessibility to insurance for residents of at-risk areas.

Limitations and sources of improvement

In conclusion, this study highlights the challenges inherent in hurricane risk modeling, calling for the exploration of more advanced approaches such as the integration of Machine Learning or Deep Learning models. The absence of any consideration of the dependence between track variations and hurricane intensity is a limitation, as is the omission of future IPCC climate projections (IPCC, 2022). In modeling financial costs, county-level simplifications were necessary to make the model workable in the insurance context. The exploration of more sophisticated Machine Learning models, such as neural networks, could offer solutions to the challenges associated with damage estimation. In addition, the influence of external factors on costs, such as government policies, merits further exploration, prompting the integration of a more comprehensive set of explanatory variables.

Remerciements

Je tiens à exprimer ma sincère gratitude à toutes les personnes qui ont contribué à la réalisation de cette étude. Leur soutien et leur expertise ont été essentiels pour mener à bien ce travail.

Je tiens tout particulièrement à adresser mes remerciements à Chadi HANNA, mon tuteur de stage, dont la disponibilité, les conseils éclairés, la pédagogie, et la bienveillance ont été d'une valeur inestimable. Son expertise et son engagement ont été des atouts majeurs dans la réalisation de cette étude.

Je souhaite également exprimer ma sincère reconnaissance envers Adam SMIAROWSI et Coralie CHARBONNEL pour leur disponibilité et la confiance qu'ils m'ont accordée tout au long de ma période de stage.

Enfin je souhaite également remercier Quentin GUIBERT, mon tuteur académique, pour ses conseils précieux, son expertise, et sa disponibilité tout au long de la réalisation de ce mémoire.

Table des matières

Résumé	3
Abstract	4
Note de Synthèse	5
Synthesis note	11
Remerciements	17
Glossaire	21
Introduction	23
1 Notions sur les tempêtes et ouragans	25
1.1 Contexte et définitions	25
1.2 Cycle de vie d'un ouragan	26
1.3 Caractéristiques et signes précurseurs d'un ouragan	34
1.4 Les ouragans dans le bassin Atlantique Nord	37
1.5 Impacts assurantiels des ouragans sur le continent américain	40
1.6 Présentation des bases de données	47
2 Simulation de la survenance des ouragans	53
2.1 Cadre et périmètres	54
2.2 Modélisation de la fréquence annuelle	59
2.3 Modélisation des points de genèse	62
3 Simulation de la dynamique des ouragans	71

3.1	Modélisation des trajectoires	71
3.2	Modélisation de l'intensité	95
4	Modélisation du coût financier	127
4.1	Construction d'une fonction de perte au risque ouragan aux Etats-Unis	128
4.2	Application aux saisons cycloniques simulées	145
4.3	Limites de l'étude et perspectives d'amélioration	154
	Conclusion	157
	Bibliographie	159
A	Annexes	167
A.1	Echelle de Beaufort	167
A.2	Rappels théoriques sur les chaînes de Markov	168
A.3	Rappels théoriques sur les séries temporelles	170
A.4	Qualité d'ajustement de la modélisation des trajectoires	172
A.5	Calcul de l'intensité relative	178
A.6	Analyse comparative des modèles de V_{max}	179
A.7	Rappels sur les modèles linéaires généralisés, XGBoost et Random Forest	182
A.8	Gradient de vitesse maximale des vents	185
A.9	Résultats des modèles GLM et ajustement de la variable coût	188

Glossaire

AIC Akaike Information Criterion. 61, 138, 140, 183

CCRIF Caribbean Catastrophe Risk Insurance Facility. 46, 47

CPHC Central Pacific Hurricane Center. 48

EAL Expected Annual Value. 131

FEMA Federal Emergency Management Agency. 42, 45, 131–133

GIEC Groupe d'experts Intergouvernemental sur l'Evolution du Climat. 40, 60

IBTrACS International Best Track Archive for Climate Stewardship. 37, 38, 48–51, 55–57, 60, 71, 73–81, 88, 92, 98, 103, 108, 109, 113, 133

JTWC Joint Typhoon Warning Center. 48

MPI Maximum Possible Intensity. 96, 99

MSW Maximum Sustained Wind (Vitesse maximale des vents soutenus noté V_{max}). 35

NFIP National Flood Insurance Program. 42, 45, 46

NHC National Hurricane Center. 38, 47, 53

NOAA National Oceanic and Atmospheric Administration. 50, 53, 128, 130

PC Pression Centrale noté également P . 35, 113, 114

RMSE Root Mean Square Error. 59, 140, 142, 144, 184

RMW Radial Maximum Wind (Rayon de la vitesse maximale des vents soutenus noté R_{max}). 35

RSMC Regional Specialised Meteorological Center. 49, 50

SS Echelle de Saffir-Simpson. 38

SST Sea Surface Temperature (Température à la surface de la mer noté T). 98

WMO World Meteorological Organization (Organisation Météorologique Mondiale noté OMM). 48, 49

Introduction

Les ouragans, ces phénomènes météorologiques redoutables, continuent de susciter un intérêt croissant tant sur le plan scientifique que dans le domaine de la gestion des risques. Leur capacité à engendrer des perturbations majeures et dévastatrices dans certaines régions du globe est très impressionnante. Dans ce contexte, les États-Unis, en particulier les États côtiers du golfe du Mexique et de la côte Est, sont particulièrement vulnérables aux ravages causés par ces tempêtes tropicales. En effet, le risque cyclonique est la catastrophe naturelle la plus coûteuse pour les États-Unis, avec des coûts totaux pouvant atteindre jusqu'à des centaines de milliards de dollars pour un seul ouragan (INSURANCE INFORMATION INSTITUTE, 2023b). Par exemple, l'ouragan Katrina survenu en 2005 demeure la catastrophe naturelle la plus coûteuse jamais survenue aux États-Unis, ayant engendré au total près de 193 milliards de dollars de dégâts et plus de 1800 décès (KNABB et al., 2023). De plus, les experts en climatologie observent une tendance à l'augmentation de l'intensité de ces phénomènes dévastateurs au cours des dernières décennies, une tendance qui pourrait s'accroître à l'avenir en raison principalement du réchauffement climatique qui affecte la Terre depuis quelques décennies (INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, 2023; KOSSIN et al., 2020).

L'impact financier et social des ouragans est significatif. Les pertes matérielles, les dommages aux infrastructures, et les perturbations économiques qu'ils engendrent font des ouragans un sujet de préoccupation majeur, non seulement pour les autorités gouvernementales mais également pour le secteur privé. Parmi les acteurs concernés, l'industrie de l'assurance occupe une place centrale, car elle est chargée d'évaluer et de gérer le risque financier associé aux dommages causés par les ouragans (INSURANCE INFORMATION INSTITUTE, 2023a). Ces dernières décennies sont marquées par une augmentation très marquante des coûts liés aux ouragans. En moyenne, un ouragan coûte près de 50 milliards de dollars depuis la dernière décennie, une hausse significative par rapport aux années 80, où le coût moyen était inférieur à 10 milliards de dollars (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2022). Dans ce contexte, le marché de l'assurance contre le risque cyclonique aux États-Unis est devenu de plus en plus incertain. De nombreuses compagnies d'assurance se voient contraintes de se retirer des zones les plus exposées et d'augmenter leurs primes en raison du risque de plus en plus élevé associé à ces phénomènes destructeurs, contribuant ainsi à une situation de crise dans le secteur de l'assurance (CHRISTOPHER et al., 2023). Cette problématique complexe est exacerbée par d'autres facteurs tels que le réchauffement climatique, la montée de l'inflation, l'attrait continu de la population pour les régions côtières, et les coûts de construction des infrastructures de plus en plus élevés. Tous ces éléments soulèvent de plus en plus de questions sur la viabilité de l'assurance contre le risque cyclonique dans ces régions extrêmement exposées (QUILLEN, 2022; PHILBRICK et WU, 2022; UNITED STATES BUREAU OF LABOR STATISTICS, 2023).

Au cœur de cette problématique se posent plusieurs questions cruciales : Comment évaluer de manière précise et rigoureuse l'exposition au risque d'ouragan aux États-Unis ? Quels sont les facteurs qui contribuent à rendre cette exposition si importante ? Comment calculer avec fiabilité les pertes potentielles causées par un ouragan dans une région donnée ? Ces interrogations constituent le fil conducteur de cette étude.

L'objectif central de cette étude est de développer une modélisation avancée visant à cartographier

de manière précise et détaillée l'exposition au risque d'ouragan aux États-Unis, en mettant particulièrement l'accent sur les régions les plus vulnérables face à ce risque cyclonique. Cette carte d'exposition projetée permettra d'offrir une vision exhaustive des zones présentant le plus haut niveau de risque, des éléments qui contribuent à ce risque, et des conséquences potentielles en cas de survenue d'une tempête majeure. Ces informations revêtent une importance capitale pour l'industrie de l'assurance, car elles lui fournissent une vue d'ensemble de l'exposition de son portefeuille aux risques ouragan ainsi que des implications financières qui en découlent. L'approche adoptée repose sur une méthodologie rigoureuse et l'utilisation de données actualisées, dans le but ultime de réaliser une évaluation fine et précise des risques associés aux ouragans. Le but final de cette démarche est de mettre à disposition des compagnies d'assurance, des responsables de la gestion des catastrophes, et des décideurs, des outils et des informations essentiels pour quantifier et mesurer l'exposition de leur portefeuille au risque cyclonique afin d'anticiper et mieux gérer les conséquences financières et humaines liées aux ouragans.

L'étude est structurée en quatre parties distinctes. La première partie pose les fondements nécessaires à la modélisation en introduisant les concepts essentiels relatifs aux phénomènes météorologiques en question (chapitre 1). Elle fournit également une description détaillée de l'impact de ces phénomènes sur les régions étudiées, tout en présentant de manière synthétique les bases de données disponibles en libre accès et utiles à l'étude. Les deuxième et troisième parties de l'étude se consacrent à la modélisation du risque cyclonique (chapitres 2 et 3). Le deuxième chapitre se penche sur la modélisation de la fréquence annuelle des tempêtes cycloniques ainsi que sur les zones de genèse de ces phénomènes. Le troisième chapitre approfondit l'analyse en se concentrant sur le développement des trajectoires et de l'intensité des ouragans, en mettant en lumière leurs caractéristiques physiques fondamentales telles que la pression centrale, la vitesse maximale des vents, ou encore le gradient de vitesse. Enfin, la dernière partie de cette étude est consacrée à l'évaluation de l'impact financier des saisons cycloniques simulées (chapitre 4). Pour ce faire, la modélisation d'une fonction de perte liée aux ouragans aux États-Unis est présentée. Cette modélisation permettra de relier de manière explicite les facteurs représentant l'intensité d'un ouragan avec les caractéristiques des zones touchées, notamment les comtés au sein d'un État, et les dommages causés. L'étude se conclut par une analyse assurantielle des coûts bruts annuels dans les zones les plus exposées, ainsi que par l'estimation du coût moyen annuel par foyer pour l'ensemble des régions analysées, offrant ainsi une vue détaillée des conséquences financières de ces phénomènes dévastateurs.

Chapitre 1

Notions sur les tempêtes et ouragans

Les tempêtes et ouragans sont des phénomènes naturels qui ont marqué l'histoire de l'humanité depuis des siècles (EMANUEL, 2005). Ces phénomènes météorologiques ont été observés depuis des millénaires, mais leur étude scientifique n'a commencé que récemment.

L'étymologie du terme ouragan, *hurricane* en anglais, est dérivé du terme espagnol *huracan*, venant lui-même de *hunnaken* le « Dieu des tempêtes » au sein de la civilisation Maya (AMERICAN METEOROLOGICAL SOCIETY, 2023). Historiquement, la notion d'ouragan a été mentionnée pour la première fois à travers un écrit sur le second voyage de Christophe Colomb vers l'Amérique en 1494 (LUDLUM, 1963). Depuis, les observations sur ces phénomènes particuliers se sont intensifiées au fil des siècles et en 1873, le National Weather Service aux Etats-Unis émet son premier avis officiel d'ouragan.

Ce n'est qu'à partir du 19ème siècle que les scientifiques ont commencé à étudier les tempêtes et les ouragans de manière plus systématique, en utilisant des instruments de mesure et en collectant des données sur les vents, la pression atmosphérique et les précipitations (EMANUEL, 2018). Ces études ont permis de mieux comprendre les mécanismes qui régissent ces phénomènes et de développer des modèles de prévision plus précis pour mieux protéger les populations et les infrastructures.

L'objectif de ce chapitre est de fournir une description complète des phénomènes naturels que sont les tempêtes et les ouragans, ainsi que des outils indispensables pour mener à bien l'étude portant sur l'exposition et l'impact financier d'un ouragan sur le territoire des Etats-Unis. Les notions fondamentales liées à ces événements météorologiques, incluant leurs définitions, le processus de formation et de développement, leur classification, ainsi que quelques chiffres clés associés seront abordées dans ce chapitre. Il est important de souligner que cette étude adopte une approche simplifiée dans l'exposition de ces notions, en évitant de se plonger dans des détails physiques complexes qui nécessiteraient des connaissances plus spécialisées. Pour les lecteurs désireux d'approfondir ces aspects, des sources plus pointues sont fournies (EMANUEL, 2018 ; ELSNER et KARA, 1999), leur permettant ainsi d'accéder à une connaissance plus approfondie. Enfin, la présentation de deux bases de données majeures, HURDAT2 et IBTrACS, regroupant chacune un ensemble de données sur les tempêtes et ouragans depuis le 19ème siècle, sera effectuée. L'historique, le contenu, ainsi que les limites et inconvénients respectifs de ces bases de données seront décrits, permettant ainsi de procéder à une analyse comparative et d'orienter le choix vers la plus adaptée pour la modélisation finale portant sur l'estimation des pertes financières liée au risque d'ouragan sur le sol américain.

1.1 Contexte et définitions

En météorologie, une tempête peut être définie comme une zone étendue de vents violents générés par un système de basses pressions (METEO-FRANCE, 2020). Deux domaines de la météorologie apportent des informations plus précises :

- Météorologie marine : une tempête correspond à la force 10 de l'échelle Beaufort. La force 10, qualifiée de tempête, correspond à des vents moyens de 89 à 102 km/h et des rafales supérieures à 110 km/h.
- Météorologie tropicale : le terme « tempête tropicale » désigne une dépression observée au niveau des latitudes tropicales ou subtropicales dont les vents moyens sont compris entre **62** et **117 km/h**. Au-delà de ces valeurs, la dépression devient un « cyclone tropical ».

L'étude se focalise sur le domaine de la **météorologie tropicale**. Dans ce cadre, les tempêtes tropicales se définissent comme des **perturbations atmosphériques**¹ qui s'organisent en **dépression tropicale**, avec des **vents de surface** tournant cycloniquement (sens anti-horaire dans l'hémisphère Nord, sens horaire dans l'hémisphère Sud) autour du centre dépressionnaire et un **maximum de vitesse moyenne des vents** soutenus pendant une minute dépassant au moins 17 m/s soit **62 km/h**. Ces tempêtes tropicales peuvent continuer à s'intensifier et une fois que le maximum de vitesse moyenne des vents soutenus dépasse le seuil de **118 km/h**, les tempêtes tropicales passent à leur forme la plus évoluée connu sous le terme de **cyclone tropical** (BEUCHER, 2010).

Il est à noter que les termes cyclones, ouragans ou typhons désignent la même réalité et ne diffèrent uniquement par leur lieu de genèse. La figure 1.1 présente l'appellation en fonction du lieu de genèse du phénomène.

Selon Météo-France, « *Les termes typhon, ouragan et cyclone tropical recouvrent tous les trois la même réalité. Ils désignent un phénomène tourbillonnaire des régions tropicales (entre 30°N et 30°S) accompagné de vents dont la vitesse est supérieure ou égale à 64 nœuds, c'est-à-dire 118 km/h (soit une force 12 sur l'échelle de Beaufort). La désignation adoptée dépend simplement de l'endroit du globe où se produit le phénomène* » (METEO-FRANCE, 2023).

L'étude se portant spécifiquement sur le bassin Atlantique Nord, le terme ouragan sera donc adopté dans toute la suite des travaux. Toutefois, les termes cyclones et typhons seront utilisés si la zone considérée est différente.



FIGURE 1.1 : Carte de localisation des noms attribués selon la zone de genèse (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004)

1.2 Cycle de vie d'un ouragan

Dans cette section consacrée au cycle de vie d'un ouragan, les différentes étapes composant ce processus complexe seront examinées en détail. Les sous-parties suivantes présenteront successivement la genèse d'une perturbation, la formation d'un ouragan, son développement et sa maturité, la structure qu'il atteint à son apogée, ainsi que les mécanismes de dissipation et de déclin qui marquent la fin

¹Les perturbations atmosphériques des régions tropicales sont appelées perturbations tropicales car elles sont des précurseurs des phénomènes extrêmes comme les ouragans

de ce phénomène tourbillonnaire. Chaque étape est essentielle pour comprendre la naissance, l'évolution et la fin du cycle de vie d'un ouragan, offrant ainsi une approche approfondie de ce phénomène météorologique majeur (ELSNER et KARA, 1999).

1.2.1 Genèse d'une perturbation

Les régions tropicales sont fortement marquées par la rencontre entre des masses d'air froides en provenance des pôles et des masses d'air plus chaudes en provenance des tropiques (ROUX, 2018). Dans l'hémisphère nord, l'air froid se déplace vers le sud tandis que l'air chaud se dirige vers le nord. Lorsque l'air chaud entre en contact avec l'air froid, des phénomènes météorologiques complexes se mettent en place. L'air chaud, étant plus léger que l'air froid, se soulève au-dessus de cette masse d'air plus dense, créant ainsi une zone de basse pression atmosphérique, communément appelée **dépression atmosphérique** ou zone de basse pression atmosphérique. La rencontre entre l'air chaud et l'air froid entraîne également le refroidissement de l'air chaud, ce qui favorise la formation de nuages. Il en résulte la naissance d'une **perturbation tropicale**.

1.2.2 Formation d'un ouragan

La perturbation tropicale résultante peut se dissiper rapidement ou se renforcer pour devenir une **dépression tropicale**. Une **dépression tropicale**² est une perturbation ayant un mouvement cyclonique³ autour d'un centre de basse pression, avec une présence de vents maximaux soutenus sur une minute inférieurs à 17 m/s (soit environ 63 km/h). La dépression tropicale peut ensuite évoluer pour devenir une **tempête tropicale**, puis en un **ouragan** ou **cyclone tropical** en fonction de l'augmentation de la vitesse maximale des vents comme évoquée précédemment. La vitesse maximale des vents est donc un critère important de classification des différents états du phénomène cyclonique.

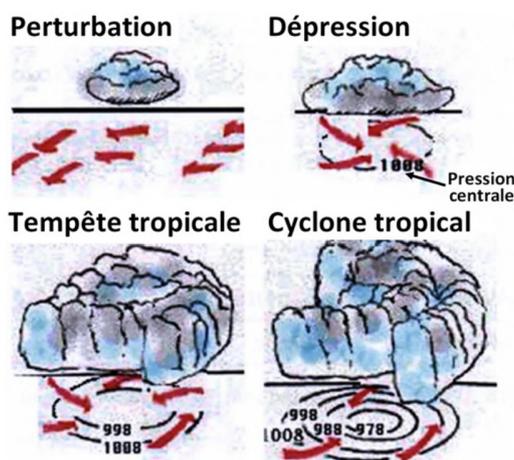


FIGURE 1.2 : Les différents stades de développement d'un cyclone (DEPARTMENT OF ATMOSPHERIC SCIENCES, 2010)

L'étude de la genèse des ouragans suscite un profond intérêt dans le domaine de la recherche scientifique, et bien que leur formation demeure un sujet complexe, les scientifiques s'accordent sur plusieurs conditions nécessaires à leur développement (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004) :

²À distinguer de la dépression atmosphérique

³Mouvement de rotation dans le sens horaire dans l'hémisphère sud, anti-horaire dans l'hémisphère nord

1.2.2.1 Facteurs internes

Les facteurs mentionnés ci-dessous sont généralement présents dans les bassins tropicaux lors de la saison cyclonique, ce qui favorise la formation d'ouragans. Cependant, bien qu'ils définissent globalement les conditions nécessaires à la naissance et au développement des ouragans, ils ne garantissent pas nécessairement la formation certaine d'un ouragan. De plus, ces facteurs ne sont pas nécessairement cumulatifs, et il est possible que dans de rares cas, des ouragans se forment en l'absence de l'une des conditions mentionnées ci-dessous, comme cela sera expliqué ultérieurement (ELSNER et KARA, 1999).

Conditions thermiques : il est nécessaire que la température de la mer soit supérieure à 26°C sur une profondeur d'au moins 50 mètres. Cette chaleur est fournie par l'évaporation de grandes quantités d'eau en surface, qui alimente le système de machine à vapeur qu'est un ouragan. Si l'eau est trop froide, l'ouragan ne peut pas se former ou s'il était déjà formé auparavant, il s'affaiblit et perd ses caractéristiques cycloniques tropicales.

Conditions géographiques : être suffisamment éloigné de l'Équateur (5° de latitude, soit une distance voisine de 550 km) de façon à ce que la force de Coriolis⁴ (ÉLIE, 2009) ne soit pas nulle. Cette force est causée par la rotation de la Terre et entraîne une déviation des vents vers la droite dans l'hémisphère nord et vers la gauche dans l'hémisphère sud. Elle est absente à l'équateur et est responsable de l'initiation du mouvement tourbillonnaire initial. Dans une bande de plus ou moins 5° de latitude, la force de Coriolis est trop faible pour engendrer un tel mouvement, ce qui explique l'absence d'ouragans entre 5° nord et 5° sud.

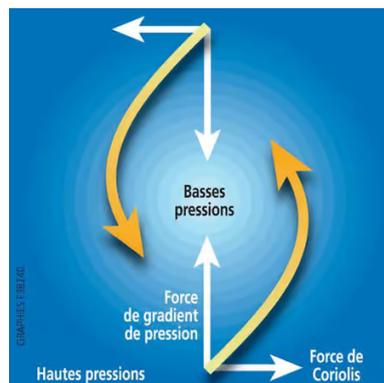


FIGURE 1.3 : Mouvement des forces engendrant la rotation d'un ouragan

Lorsque les vents se déplacent entre les zones de haute et de basse pression, un équilibre est créé entre plusieurs forces. La force de Coriolis agit comme une force centrifuge, éloignant ainsi les vents de la zone de basse pression. Parallèlement, la force de gradient de pression⁵ attire les vents vers le « creux » de la dépression, où la pression est la plus basse. Ces forces concurrentes génèrent un mouvement d'enroulement des vents autour de la zone de basse pression, suivant approximativement les lignes de pression égale. Les mouvements résultants de ces forces sont représentés sur la figure 1.3 (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004).

Conditions atmosphériques : une forte humidité qui doit être supérieure à 40% mais inférieure à 70% pour permettre la formation de cumulonimbus, qui sont des nuages convectifs responsables des

⁴Pour une latitude θ (en radian) la force de Coriolis notée f_C (en Newton) est donnée par : $f_C(\theta) = 1,454 \times 10^{-4} \sin(\theta)$

⁵Lorsqu'il y a une différence de pression entre deux points, l'air se déplace naturellement de l'endroit où la pression est la plus élevée vers l'endroit où elle est la moins élevée. Cette transition d'une haute pression vers une basse pression définit la force de gradient de pression.

orages. Il faut également la pré-existence d'une zone dépressionnaire⁶, d'un amas nuageux, d'une ligne de grains⁷ ou d'une onde tropicale⁸ associée à de la convection et à un faible mouvement d'air convergent de basses couches, créant des mouvements ascendants pour l'air humide. De plus, la température de l'air doit diminuer rapidement avec l'altitude pour créer des conditions de stabilité dans l'atmosphère et favoriser cette convection. Et enfin, la présence de vents en altitude est essentielle, jusqu'à environ 15 km d'altitude. Cependant, ces vents ne doivent pas présenter de cisaillement, c'est-à-dire une variation de force et de direction selon l'altitude, car cela pourrait détruire la structure verticale de l'ouragan. Les effets du cisaillement vertical du vent sur la formation et le développement des ouragans sont complexes et ne sont pas entièrement compris au sein de la communauté scientifique. En général, pendant la genèse, un cisaillement fort peut entraver la rotation et la structure verticale de l'ouragan naissant, rendant difficile le développement et l'intensification de l'ouragan. Si l'ouragan parvient tout de même à se former, le cisaillement vertical du vent continue de jouer un rôle déterminant dans la vie de l'ouragan. Des valeurs élevées de cisaillement peuvent perturber la structure symétrique et verticale de l'ouragan, entraînant une diminution rapide de son intensité.

1.2.2.2 Facteurs externes

Exceptionnellement, des ouragans peuvent se former en dehors de ces conditions habituelles. En 2001, le typhon Vamei a émergé à seulement 1,5° de latitude nord de l'équateur, en profitant d'une perturbation atmosphérique préexistante et de conditions relativement fraîches liées à la mousson⁹ dans le Pacifique nord-ouest. Les facteurs qui ont contribué à la formation de ce typhon sont estimés se produire seulement tous les 100 à 400 ans (CHANG et al., 2003). Il est également arrivé que des ouragans se soient formés avec des températures de surface de la mer aussi basses que 25°C, comme l'ouragan Vince en 2005 dans le bassin Atlantique (FRANKLIN, 2006).

D'autres facteurs externes peuvent influencer l'activité cyclonique notamment le phénomène climatique connu sous le nom de **El Niño** (SARACHIK et CANE, 2010). El Niño est un phénomène océanique qui se traduit par une hausse anormale de la température à la surface de l'eau sur la zone du Pacifique Central jusqu'aux côtes du Pérou et de l'Equateur. Il survient généralement tous les 2 à 7 ans et dure entre 12 et 18 mois (exceptionnellement quelques années comme dans les années 90).

Bien que son origine reste encore mystérieuse pour les scientifiques, les prévisions concernant son arrivée sont pour la plupart certaines. Début mai 2023, l'Organisation Météorologique Mondiale (OMM) déclare que le **phénomène El Nino sera de retour durant la saison 2023** avec une probabilité de 60% entre mai et juillet, de 70% entre juin et août, et **80%** entre juillet et septembre. L'OMM indique qu'un risque important de températures mondiales records pour cette saison est à prévoir (DEMEURE, 2023).

Si ces conditions climatiques décuple le risque d'ouragans sur la zone du Pacifique Est, l'effet est inverse dans le bassin Atlantique Nord (SCHNEYER, 2023). L'atmosphère est plus stable, la température du bassin est un peu plus faible que la normale et un cisaillement du vent plus marqué avec un vent d'altitude dominant présent durant toute l'année. Ainsi **lors des années El Niño la saison cyclonique de l'Atlantique est moins intense**. Cependant les prévisions d'ouragans durant cette période particulière sont compliquées en raison de l'atténuation d'El Niño sur la saison cyclonique, rendant les prédictions d'ouragans plus délicat et incertain.

⁶Région où la pression atmosphérique est basse par rapport à celle du voisinage au même niveau. Au niveau de la mer, un niveau de pression inférieure à 1010 hPa sera dite dépressionnaire (MÉTÉO-FRANCE, 2003)

⁷Bande nuageuse constituée de nuages orageux

⁸Perturbation tropicale associée à un axe dépressionnaire des couches basses et moyennes de la troposphère (plus basse couche de l'atmosphère de la Terre), circulant d'est en ouest

⁹Régime météorologique particulier avec alternance d'une saison humide et d'une saison sèche et une inversion du sens des vents dominants, soufflant dans les basses couches de l'océan vers le continent en été, du continent vers l'océan en hiver. (ENCYCLOPÉDIE-ENVIRONNEMENT, 2022)

1.2.3 Développement et maturité

Dans la phase de développement et de maturité d'un ouragan, plusieurs aspects essentiels contribuent à la compréhension de ce phénomène cyclonique. Cette étape cruciale du cycle de vie d'un ouragan comprend le cycle énergétique et l'intensité du système, sa structure à maturité, ainsi que l'analyse de sa trajectoire. En analysant ces différentes facettes, cette étude approfondit la compréhension du comportement complexe et dynamique des ouragans, permettant ainsi d'améliorer les capacités de prévision et d'atténuation des risques liés à ces puissants événements météorologiques.

1.2.3.1 Cycle énergétique et intensité

Un ouragan est considéré comme une **machine thermique** colossal qui tire son énergie du **déséquilibre thermodynamique entre l'atmosphère et l'océan**, qui agissent respectivement comme source froide et source chaude. De manière simplifiée, le processus thermodynamique (ROUX, 2018) s'explique de la manière suivante : la grande partie de l'énergie d'un ouragan provient de l'**évaporation à la surface de l'océan**, forcée par des vents d'intensité croissante. En effet, lorsqu'un ouragan se forme, l'air chaud et humide au-dessus de l'océan s'élève et se refroidit, formant des nuages. Les **courants ascendants** qui en résultent transportent ces nuages jusqu'à une quinzaine de kilomètres d'altitude, où l'air s'éloigne du centre de l'ouragan dans un **mouvement divergent et anticyclonique**¹⁰. À cet endroit, l'air chaud et humide de l'ouragan perd de l'énergie sous forme de **rayonnement thermique**¹¹ vers l'espace environnant. Ce processus de rayonnement permet à l'air de se refroidir et contribue à l'équilibre énergétique de l'ouragan. C'est un mécanisme essentiel pour maintenir le système en mouvement et éviter qu'il ne surchauffe. Le rayonnement thermique est un des mécanismes de dispersion de l'énergie accumulée dans l'ouragan, garantissant ainsi son équilibre thermodynamique tout au long de son cycle de vie. Finalement, l'air redescend vers la surface de l'océan à travers le **flux descendant** qui prédomine à grande distance. Dans l'œil de l'ouragan, le faible mouvement descendant de l'air joue un rôle crucial : il réchauffe l'air qui se trouve déjà au centre de l'œil de l'ouragan. Ce réchauffement de l'air au cœur de l'ouragan renforce la **dépression centrale**, qui est la zone de pression atmosphérique très basse située au centre de l'ouragan. En renforçant cette dépression centrale, l'air descendant attire l'air extérieur des basses couches vers l'intérieur de l'ouragan. Cette circulation de l'air contribue ainsi au maintien du cycle de l'ouragan, en assurant un approvisionnement continu en air chaud et humide. Cet apport constant d'énergie thermique nourrit le système cyclonique et maintient sa force et son intensité. Une illustration du processus thermodynamique d'un ouragan à maturité est représenté sur la figure 1.4.

1.2.3.2 Structure à maturité

Un ouragan est une vaste dépression à centre chaud caractérisée par une masse nuageuse presque circulaire d'un diamètre d'environ 500 km, pouvant s'étendre jusqu'à 1 000 km. Les vents dans l'ouragan peuvent atteindre des vitesses considérables, augmentant progressivement vers le centre. L'**œil** de l'ouragan est un centre de rotation entouré de **bandes spirales** et d'une zone dense de nuages très compacts appelée **mur de l'œil**. Les vents d'ouragan peuvent s'étendre jusqu'à 150 km de diamètre depuis l'œil de l'ouragan. Cette vitesse élevée des vents est due à une diminution de la densité de l'air à l'intérieur de la dépression, créant un appel d'air où l'air environnant (l'air extérieur des basses couches évoquée précédemment) s'engouffre dans cet espace de basse pression et subit une forte accélération.

Ainsi la structure d'un ouragan peut se résumer à travers les éléments suivants (ELSNER et KARA, 1999).

¹⁰Dans le sens contraire au mouvement cyclonique de l'hémisphère Nord ou Sud

¹¹Transfert d'énergie sous forme d'ondes électromagnétiques émises par un corps chaud vers l'espace environnant, où il se propage sans besoin de support matériel

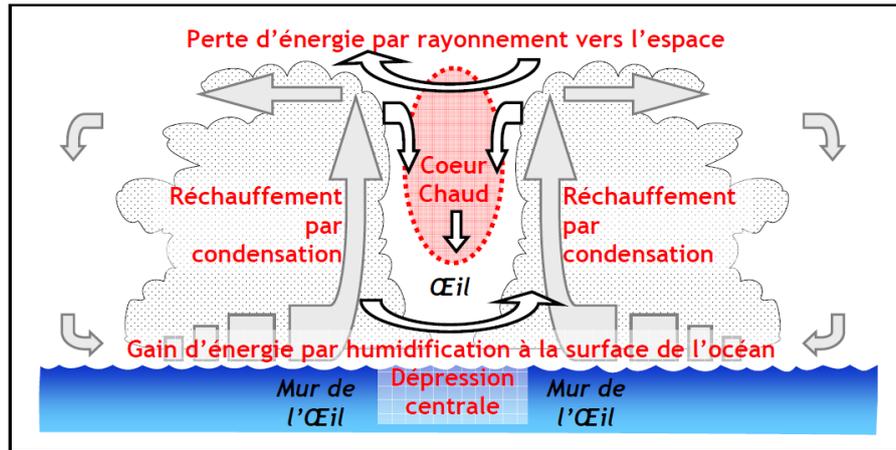


FIGURE 1.4 : Cycle thermodynamique d'un ouragan à maturité (ROUX, 2018)

Les bandes spirales : une structure organisée de lignes de grains qui enroulent le centre de l'ouragan en spirale. Ces bandes comportent généralement une bande principale et des bandes secondaires, pouvant s'étendre sur plusieurs centaines de kilomètres autour de l'œil. Elles sont constituées principalement de nuages convectifs tels que des cumulus et des cumulonimbus, immergés dans une zone nuageuse étendue qui produit des précipitations intenses et des orages violents.

L'œil de l'ouragan : région centrale de l'ouragan représentant le point de convergence où la pression atmosphérique est la plus basse. L'œil de l'ouragan est un cercle d'un diamètre moyen de 30 à 60 kilomètres, mais il peut atteindre plus de 200 km dans certains cas exceptionnels. Elle se caractérise par des vents descendants ou quasi-inexistants ainsi que par quelques nuages de faible hauteur. En raison de la forte subsidence de l'air¹² dans cette zone, le ciel y est généralement peu nuageux, les précipitations sont rares voire inexistantes, et la température est supérieure à celle de l'extérieur. Cette région est généralement marquée par un calme apparent, sans pluie ni vent.

Le mur de l'œil : région nuageuse très active en forme de cercle qui entoure l'œil d'un ouragan. Il s'agit d'une zone de cumulonimbus très développés, souvent organisés en bandes spirales resserrées ou en anneaux, qui peuvent atteindre une hauteur de 15 km ou plus et qui culminent jusqu'à 5 km d'altitude. Cette région est constituée des aspects les plus dévastateurs d'un ouragan : les vents et les précipitations. Les vents les plus violents atteignent fréquemment 200 km/h, dépassant parfois 300 km/h. Les mouvements verticaux de l'air y sont particulièrement rapides, entraînant une turbulence extrêmement forte.

La figure 1.5 schématise l'ensemble de la structure d'un ouragan à maturité et représente les principaux éléments évoqués tels que les bandes spirales, l'œil, le mur de l'œil et les vitesses des vents. Un exemple visuel est donnée avec la figure 1.6 où est représenté la structure à maturité de l'ouragan Irma survenu en 2017 sur le bassin Atlantique Nord par une image satellitaire et sa trajectoire jusqu'aux étendues terrestres.

¹²Le déplacement d'air vers le sol dans l'atmosphère

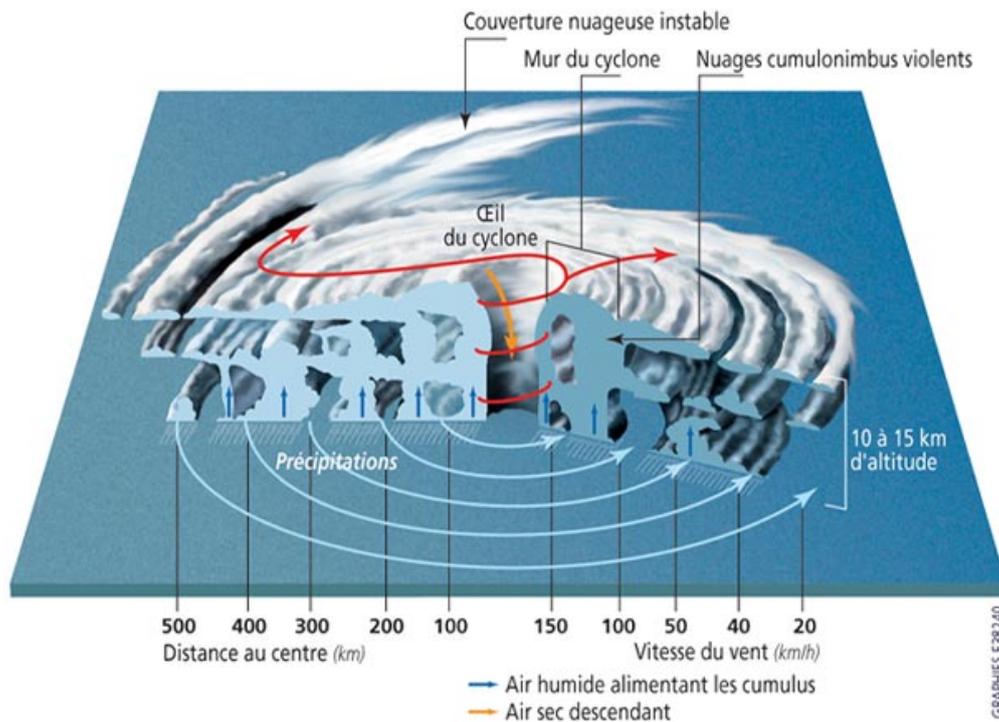


FIGURE 1.5 : Structure d'un ouragan à maturité (CYCLONE EXTRÊME, 2023)

1.2.3.3 Trajectoire

Les trajectoires des ouragans sont influencées par deux régimes de circulations atmosphériques distincts¹³ (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004) :

- une circulation équatoriale d'est en ouest aux latitudes tropicales (sauf dans l'océan Pacifique sud-ouest où elle est d'ouest en est)
- l'entrée des cyclones dans la circulation des régions tempérées après un changement de direction plus ou moins brutal. L'incurvation de la trajectoire vers le pôle est due à la force de Coriolis, qui dévie les cyclones vers le sud dans l'hémisphère nord et vers le nord dans l'hémisphère sud.

En théorie, ces deux composantes confèrent aux ouragans une **trajectoire parabolique** comme celle observé sur l'ouragan Irma 1.6, mais en réalité, les trajectoires observées peuvent être plus ou moins complexes, présentant des boucles, des retours en arrière, etc... C'est le cas de l'ouragan Mitch dont la trajectoire est représenté sur la figure 1.7b, qui a frappé l'Amérique centrale en 1998. Il a été caractérisé par des changements de direction très intenses et inhabituels. Contrairement à la trajectoire parabolique théorique des cyclones, les mouvements de Mitch ont été très erratiques, avec des changements de direction vers le nord, le sud-ouest, puis de nouveau vers le nord, et ainsi de suite. Il est considéré comme le deuxième ouragan le plus meurtrier avec plus de 20 000 morts (plus de 9000 morts officiels et plus de 20 000 en rajoutant les disparus) et l'un des plus puissants ouragans enregistrés dans le bassin Atlantique, avec des vents atteignant 290 km/h (GUINEY et LAWRENCE, 1999).

Au commencement de leur formation, les ouragans se déplacent à une vitesse plutôt modérée, habituellement inférieure ou égale à 20 km/h. Toutefois, au fur et à mesure de leur développement, ils connaissent une accélération significative lorsqu'ils changent de trajectoire, notamment lorsque la

¹³des mouvements de l'air influençant la trajectoire des ouragans



(a) Image satellitaire d'Irma sur le nord des Petites Antilles le 6 septembre 2017, à son maximum, (NATIONAL AERONAUTICS AND SPACE ADMINISTRATION - SATELLITE AQUA, 2017)



(b) Trajectoire de l'ouragan Irma, (NATIONAL HURRICANE CENTER, 2017)

FIGURE 1.6 : Ouragan Irma

composante dominante de leur mouvement s'oriente d'ouest en est (un événement qui se produit généralement autour de 23° en latitude). À ce stade, la vitesse de déplacement peut dépasser les 45 km/h (ELSNER et KARA, 1999).

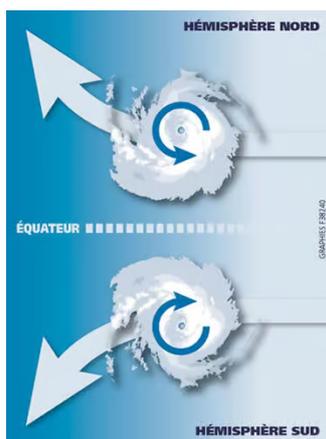
1.2.4 Dissipation et déclin

Les ouragans sont tributaires de sources d'énergie qui leur permettent de maintenir leur intensité. Ainsi, dès lors qu'une de ces sources s'atténue ou disparaît, l'ouragan commence à s'affaiblir (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004). Les causes les plus récurrentes qui contribuent à cet affaiblissement sont les suivantes.

Baisse de la température des eaux de surface océanique : Lorsque l'ouragan se déplace vers des océans où les eaux en surface ne sont pas suffisamment chaudes, l'alimentation en vapeur d'eau ne sera plus suffisante pour satisfaire les besoins conséquents d'un tel système. En conséquence, l'ouragan s'affaiblit progressivement, mais sa dissipation complète peut prendre plusieurs jours.

Impact du cisaillement vertical du vent : Lorsque l'ouragan rencontre des conditions où les vents varient en direction et en intensité avec l'altitude, son organisation structurelle se détériore. Cela se traduit par une déformation de sa structure, affectant des caractéristiques essentielles telles que l'œil, la rotation complète et la vitesse des vents.

Trajectoire vers l'Équateur : À mesure que l'ouragan se rapproche de l'Équateur, la force de Coriolis, qui joue un rôle crucial dans le maintien de la rotation cyclonique, diminue. En conséquence,



(a) Trajectoire habituelle des ouragans dans chaque hémisphère



(b) Trajectoire atypique de l'ouragan Mitch en 1988

FIGURE 1.7 : Trajectoire d'un ouragan (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004)

l'ouragan s'affaiblit naturellement.

Impact de l'atterrissage : Lorsqu'un ouragan traverse la terre, il est couramment désigné par le terme « atterrissage ». À ce stade, l'ouragan perd son principal carburant, à savoir les eaux chaudes de l'océan, ce qui entraîne souvent une diminution significative de son intensité. Ce phénomène est particulièrement remarquable dans la région des Caraïbes. Cependant, les ouragans les plus puissants peuvent conserver suffisamment d'énergie pour traverser des étendues terrestres et se régénérer lorsqu'ils entrent à nouveau en contact avec un océan, à condition que les conditions favorables à leur renforcement soient réunies. Un exemple notable est l'ouragan Irma en 2017 illustré sur la figure 1.6 (JOHN et al., 2021), qui a balayé les îles des Caraïbes, provoquant des dégâts majeurs sur son passage, avant de poursuivre sa trajectoire en direction des États-Unis. Malgré une traversée des zones terrestres, Irma a conservé une intensité remarquable et est devenu l'un des ouragans les plus dévastateurs jamais enregistrés dans l'océan Atlantique.

Cette liste de causes n'est pas exhaustive et de manière générale, les facteurs cités précédemment peuvent simultanément conduire à la dissipation d'un ouragan. De plus, il convient de souligner qu'au-delà de **30° de latitude** en moyenne, les ouragans perdent de leur intensité et ceux de petite taille se dissipent rapidement, mais les plus puissants et les plus étendus peuvent conserver leur organisation et une certaine intensité pendant plusieurs jours. Certains ouragans atteignant les latitudes moyennes interagissent avec la circulation d'ouest et évoluent vers une structure caractéristique des tempêtes des zones tempérées. La probabilité de telles transitions varie en fonction de la configuration du bassin océanique, par exemple environ **un ouragan sur trois dans le bassin Atlantique Nord subit cette transformation** d'après les estimations (ROUX, 2018). En début d'automne, il arrive ainsi que d'anciens ouragans des Caraïbes se transforment, en quelques jours, en tempêtes balayant les côtes ouest de l'Europe.

1.3 Caractéristiques et signes précurseurs d'un ouragan

Afin de poursuivre cette étude, il est important de définir certaines caractéristiques propres à un ouragan qui seront essentielles pour la modélisation (AMERICAN METEOROLOGICAL SOCIETY, 2023).

Parmi ces caractéristiques, figurent notamment :

Pression centrale : noté PC ou P , pression régnant au centre de l'ouragan et au niveau de la mer, mesurée en hectopascals (hPa). C'est à ce niveau-là que la pression est la plus basse pour un ouragan. En d'autres termes, plus la pression centrale est basse et plus l'ouragan est destructeur.

Vitesse maximale des vents soutenus : notée MSW (Maximum Sustained Wind) ou V_{max} , mesurée en noeuds¹⁴ (kt), correspond à la vitesse moyenne du vent la plus élevée mesurée en un point et à un moment non obstrué à une altitude de 10 mètres pendant une durée de 1 minute (selon la définition du United States National Weather Service et du Joint Typhoon Warning Center) ou de 10 minutes (selon la définition de l'Organisation météorologique mondiale). La durée retenue dans le cadre de cette étude est de 1 minute, conformément à son utilisation dans les bases de données d'ouragans qui sont présentées dans la section 1.6.

Rayon de vitesse maximale des vents : noté RMW (Radial Maximum Wind), représente la distance entre le centre de l'œil de l'ouragan et le mur d'un ouragan, étant donné que c'est à cette distance que la vitesse des vents est la plus élevée.

La figure 1.8 présente une analyse détaillée de l'évolution des différentes caractéristiques d'un ouragan, notamment la pression et la vitesse des vents en fonction de leur position. Ce graphique met en évidence des corrélations entre ces caractéristiques. En effet, l'intensité d'un ouragan est directement liée à sa valeur de pression centrale P_{min} , qui tend à être plus basse pour les ouragans les plus intenses. Cela se traduit par des vitesses et des rayons maximums des vents, notés respectivement V_{max} et R_{VM} , particulièrement importants.

Il est essentiel de souligner que les vents qui atteignent les côtes en premier sont généralement les moins intenses de l'ouragan. Cette observation met en évidence le fait que les effets les plus destructeurs de l'ouragan se concentrent principalement à la frontière entre l'œil de l'ouragan et le mur de l'œil. Cette zone critique est caractérisée par des conditions météorologiques extrêmes, où la combinaison des vents violents et de la pression atmosphérique basse peut causer des dégâts significatifs.

Les ouragans menacent davantage les îles et les régions côtières en raison des risques maritimes engendrés et de ces caractéristiques destructrices (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004).

Le premier indicateur précurseur annonçant la formation d'un ouragan pour les zones côtières se manifeste sous la forme d'une **houle cyclonique**¹⁵ qui se déplace à une vitesse supérieure à celle du phénomène. La hauteur, la fréquence et la direction de cette houle fournissent des indications précieuses sur la distance et l'orientation de la trajectoire de l'ouragan. La houle cyclonique peut parfois être observée jusqu'à 1 000 km en avant de l'ouragan et ses effets peuvent continuer à se faire sentir après son passage.

Le second indicateur est **l'intensification des vents** au niveau des côtes. Les ravages causés par les vents violents et imprévisibles sont considérables. Leur intensité dépasse aisément 150 km/h et peut exceptionnellement atteindre près de 370 km/h à proximité de l'œil d'un ouragan.

Le cumul pluviométrique peut atteindre des proportions considérables (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004), même pour des cyclones de faible intensité en termes de vitesse du vent, tandis que certains cyclones de grande envergure peuvent être peu générateurs de précipitations. Un autre signe précurseur est donc la présence de **pluies abondantes** qui peuvent provoquer des inondations, des glissements de terrain et des coulées boueuses d'une ampleur variable.

¹⁴ $1kt \approx 1.852m/s$

¹⁵ Désigne, dans la périphérie d'un ouragan, la formation de vagues qui peuvent atteindre jusqu'à 20 mètres de hauteur, (DICTIONNAIRE-ENVIRONNEMENT, 2010)

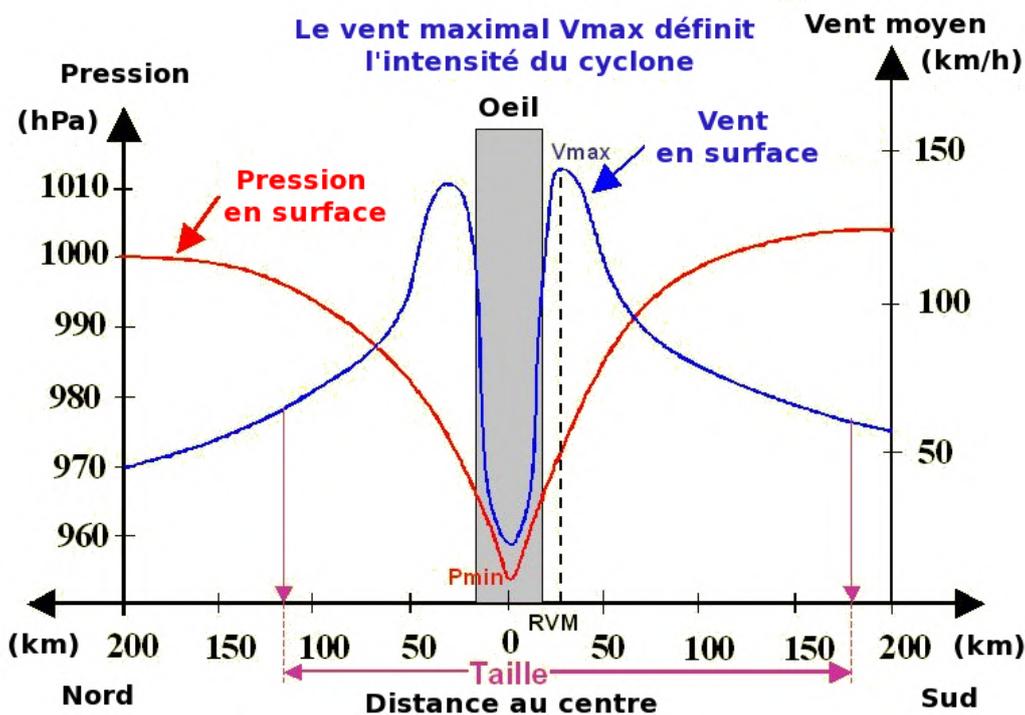


FIGURE 1.8 : Variation des caractéristiques dans un ouragan (CENTRE MÉTÉOROLOGIQUE RÉGIONAL SPÉCIALISÉ CYCLONES DE LA RÉUNION, 2023)

Enfin un dernier signe précurseur notable à prendre en compte est **l'onde de tempête**, également connue sous le nom de **marée de tempête**. Elle représente le phénomène de montée rapide du niveau de la mer près des côtes lorsque se rapproche une tempête. En effet, les vents violents du large poussent l'eau vers les côtes tandis que la pression très basse près de l'œil du ouragan aspire l'eau vers le haut (phénomène d'intumescence). Les zones basses sont les plus vulnérables aux effets de l'onde de tempête, tandis que les régions où le relief s'élève rapidement ne sont généralement pas touchées. Une illustration comparant une marée haute normale et marée de tempête est représenté sur la figure 1.9.

Les préjudices causés par les ouragans peuvent être multiples, en particulier leurs effets sur les êtres humains, l'économie et l'environnement. Malgré les avancées en matière de prévention et de protection, ces phénomènes météorologiques continuent de provoquer un nombre important de décès, de blessures et de sans-abri chaque année, principalement en raison des noyades causées par les montées des eaux et des rivières, des glissements de terrain et des coulées de boue. Les dommages matériels, quant à eux, dépendent de l'intensité des vents et du potentiel de pluie de l'ouragan, et peuvent engendrer des perturbations économiques considérables, tels que la destruction des infrastructures, la détérioration des outils de production industrielle, les impacts sur les réseaux d'eau, de téléphone et d'électricité, ainsi que l'interruption de la vie économique. Enfin, les conséquences environnementales peuvent être graves, incluant la dévastation des forêts, la pollution du littoral et des terres suite à des naufrages ou des dégâts aux infrastructures de transport.

L'ouragan IRMA (1.6) considéré comme l'une des pires catastrophes naturelles dans le bassin Atlantique totalise un bilan de 137 décès, plus de 1000 blessés et un montant total des dommages estimé à plus de 65,8 milliards de dollars US (THIERRY et al., 2018).

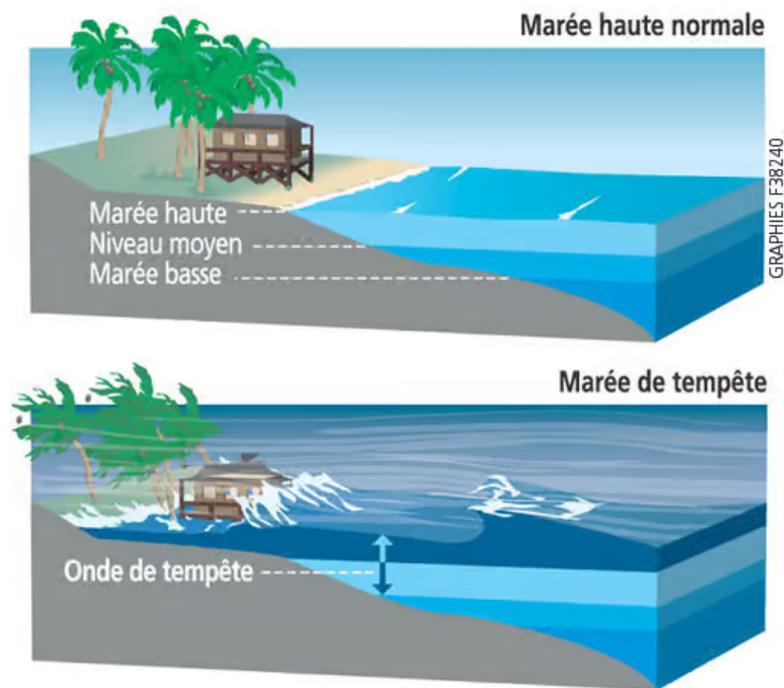


FIGURE 1.9 : Marée haute normale et marée de tempête (DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES, 2004)

1.4 Les ouragans dans le bassin Atlantique Nord

De nombreux bassins cycloniques sont répartis à travers le monde, tels que le Pacifique Nord-Ouest, le Pacifique Nord-Est, l’océan Indien Nord, l’océan Indien Sud-Ouest, l’océan Atlantique Sud, et bien d’autres encore.

Parmi ces bassins cycloniques mondiaux, le bassin Atlantique Nord est l’un des plus actifs et les plus étudiés après ceux du Pacifique Nord-Ouest et Nord-Est (CYCLONE EXTRÊME, 2023). Situé au large des côtes de l’Amérique du Nord, de l’Amérique centrale et des Caraïbes, ce bassin est le théâtre de la formation et de l’intensification de nombreux ouragans chaque année. Les eaux chaudes de l’Atlantique, combinées à des conditions atmosphériques favorables, offrent un environnement propice à la genèse et au développement de ces phénomènes dévastateurs.

1.4.1 Saison cyclonique

La saison cyclonique du bassin Atlantique-Nord s’étend généralement du début du mois de juin à la fin du mois de novembre, période durant laquelle des phénomènes tropicaux peuvent se former à tout moment. Toutefois, la plus grande fréquence de ces phénomènes est principalement observée entre les mois d’août et d’octobre. L’histogramme représenté sur la figure 1.10¹⁶ présente le nombre de tempêtes tropicales formées par mois entre les années 1982 et 2022 dans le bassin Atlantique Nord.

¹⁶Établi à partir des données de la base IBTrACS (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b)

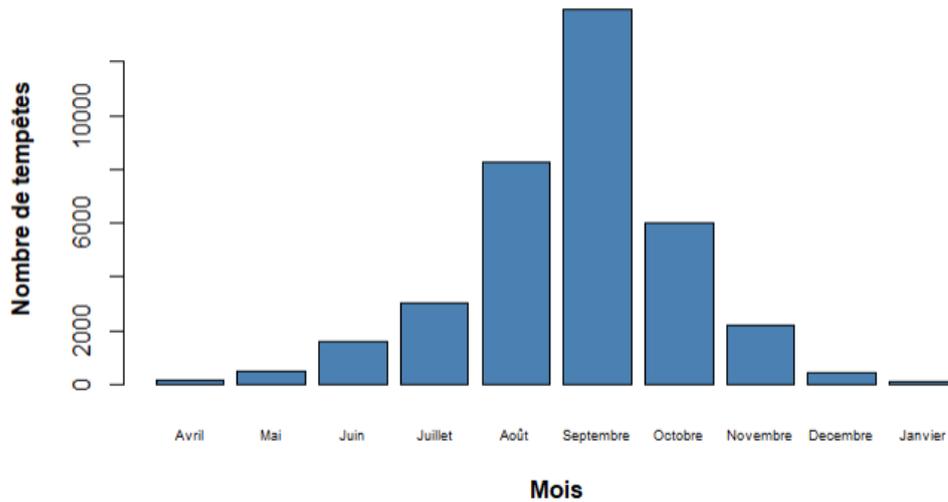


FIGURE 1.10 : Evolution du nombre de tempêtes par mois dans le bassin Atlantique Nord (1982-2022)

1.4.2 Echelle de Saffir-Simpson

L'échelle de Saffir-Simpson (noté SS par la suite, représenté sur la figure 1.1), mise en place en 1969 à la demande de l'Organisation Météorologique Mondiale par deux météorologues, l'ancien directeur du National Hurricane Center (NHC) Robert Simpson et l'ingénieur civil Herbert Saffir, est un système de classification de la puissance des ouragans basé sur la vitesse de leurs vents (ELSNER et KARA, 1999).

L'objectif initial de ce classement était d'évaluer les éventuels dommages pouvant affecter les populations et les infrastructures. Cette échelle est appliquée dans des bassins spécifiques, notamment le bassin Atlantique, couvrant une vaste zone s'étendant de l'ouest de la longitude 0° jusqu'à la longitude 180° . Pour classer un système sur cette échelle, la vitesse des vents est enregistrée pendant une minute à une hauteur de 10 mètres.

Des polémiques concernant la classification des ouragans selon l'échelle de Saffir-Simpson ont engendré de vives controverses au sein de la communauté scientifique américaine ces dernières années, notamment en ce qui concerne la possible introduction d'une Catégorie 6 (METEO-TROPICALE, 2023).

En effet, la saison 2005 a marqué un tournant dans l'histoire des cyclones dans le bassin Atlantique. Cette année-là, à la fois le nombre et l'intensité des systèmes cycloniques ont remis en question la pertinence du système de classification en vigueur, en suggérant l'intégration de critères tels que les précipitations ou la taille des cyclones. Plus récemment, les ouragans Irma et Maria frappant les Antilles lors de la saison 2017 ainsi que l'ouragan Dorian en 2019 qui a atteint une vitesse maximale des vents soutenus proche de 300 km/h et des rafales de vents allant jusqu'à plus de 350 km/h (MASTERS, 2019) ont refait surgir la polémique. Néanmoins les créateurs de l'échelle de Saffir-Simpson ont estimé qu'il n'était pas nécessaire de la modifier et qu'elle répondait parfaitement aux besoins en matière de prévention. Ils ont argumenté que tout système météorologique dépassant les 252 km/h (Catégorie 5) était déjà extrêmement dévastateur et meurtrier, et qu'ajouter une Catégorie 6 pourrait donner l'impression que les systèmes de Catégorie 5 sont moins dangereux. A ce jour, aucune modification n'est prévue et l'échelle de Saffir-Simpson continue d'être utilisée par les organisations météorologiques mondiales.

Afin d'avoir une idée de la représentation de l'échelle de Saffir-Simpson, le tableau 1.2¹⁷ détaille

¹⁷Établi à partir des données de la base IBTrACS (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b)

Catégorie	Vitesse maximale des vents	Pression centrale	Dégâts Potentiels
Dépression Tropicale	< 62 km/h	> 995 hPa	Pas de dommages Beaucoup de pluie
Tempête tropicale	63 - 118 km/h	> 995 hPa	Dommages négligeables Beaucoup de pluie
Ouragan Cat 1	119 - 153 km/h	980 - 994 hPa	Dommages limités aux maisons mobiles et à la végétation Aucun dégât structurel significatif aux bâtiments
Ouragan Cat 2	154 - 177 km/h	965 - 979 hPa	Dommages possibles aux toitures, portes et fenêtres des habitations Dommages importants aux maisons mobiles
Ouragan Cat 3	178- 208 km/h	945 - 964 hPa	Dommages moyens à la structure des bâtiments non renforcés Les maisons mobiles sont détruites Dommages importants à la végétation Inondations importantes, habitations côtières endommagées
Ouragan Cat 4	209 - 251 km/h	921 - 944 hPa	Dommages importants à la structure des bâtiments non renforcés Les maisons mobiles sont détruites Dommages importants à la végétation Inondations importantes, évacuation d'urgence dans un rayon de 3 km des côtes
Ouragan Cat 5	> 252 km/h	< 920 hPa	Destruction des toitures, des portes et des fenêtres Des habitations complètement détruites Evacuation d'urgence dans un rayon de 8 à 16 km des côtes

TABLE 1.1 : Echelle de Saffir-Simpson

le nombre de phénomènes observés selon leur catégorie au cours de la période 1982-2022 sur le bassin Atlantique Nord. Le taux de conversion d'une tempête à un ouragan est également présenté pour chaque catégorie. Ce taux est simplement calculé en faisant le ratio entre le nombre d'ouragans d'une catégorie et le nombre de tempêtes tropicales. Ainsi, un ouragan de Catégorie 5 est passé par toutes les catégories de 1 à 4 avant d'atteindre la Catégorie 5 et sera donc comptabilisé dans chaque catégorie.

Le nombre total de tempêtes tropicales s'élève à 649 sur la période observée, soit une moyenne d'environ 16 tempêtes tropicales par année. Le taux de conversion d'une tempête à un ouragan est d'environ 42%, soit une moyenne de 7 ouragans par année entre 1982 et 2022. Les ouragans de catégorie 1 sont les plus représentés avec un total de 279 contrairement à ceux de catégories 5 au nombre de 19 avec un taux de conversion de seulement 3%, chiffre plutôt bas et rassurant pour les zones et populations affectées par ces phénomènes. Cependant, malgré un chiffre relativement bas, le bassin Atlantique a connu 13 ouragans de catégorie 5 durant ces 20 dernières années, soit un tiers de ce qui s'est produit pendant le 20ème siècle. Cette observation ne tend pas à diminuer et peut être expliquée par l'intensification du réchauffement climatique au cours des dernières décennies (LE MONDE, 2019).

	Tempête Tropicale	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
Nombre	649	279	162	116	72	19
Tempête à ouragan (%)	100%	42%	24%	17%	11%	3%

TABLE 1.2 : Nombre et pourcentage des tempêtes et ouragans sur la période 1982-2022 dans le bassin Atlantique Nord

En effet, d’après le dernier rapport du GIEC (IPCC, 2022) établi en 2022, le réchauffement climatique ne devrait pas augmenter le nombre de tempêtes et d’ouragans sur le bassin Atlantique Nord, la tendance pourrait même être à la baisse d’après les experts. cependant la température élevée des océans pourrait contribuer à intensifier la puissance de ces phénomènes météorologiques et leurs effets destructeurs. Ainsi le GIEC prévoit que la proportion mondiale de cyclones tropicaux qui atteignent des niveaux très intenses (catégorie 4-5), ainsi que leurs pics de vents et de précipitations, devraient augmenter avec le réchauffement climatique. Le bassin Atlantique Nord pourrait donc être confronté à une **augmentation de l’intensité des ouragans dans les prochaines décennies**, soit une présence significative d’ouragans de catégorie 4 et 5, sans qu’il soit précisé de données chiffrées. En étudiant la littérature scientifique KOSSIN et al. (2020), qui contribue au sixième rapport du GIEC, analyse l’évolution de l’intensité des ouragans majeurs¹⁸ pour la période allant de 1979 à 2017, couvrant tous les bassins cycloniques mondiaux. Les auteurs définissent une *probability of major hurricane exceedance*, c’est-à-dire la probabilité qu’un ouragan passe d’une catégorie 2 à 3, avec une vitesse maximale des vents dépassant 178 km/h selon l’échelle de Saffir-Simpson. L’analyse montre que la moyenne globale de cette probabilité pour tous les bassins cycloniques mondiaux est de 0.08 par décennie. Toutefois, il est important de souligner que le bassin Atlantique Nord présente la probabilité la plus élevée, s’élevant à 0.49 par décennie. Cela signifie qu’**au cours des dix dernières années, la probabilité de passer d’un ouragan de catégorie 2 à 3 dans le bassin Atlantique Nord a augmenté de 49%, avec un niveau de confiance supérieur à 99%**. Ce résultat est cohérent avec l’observation d’une tendance à la hausse de la proportion d’ouragans majeurs dans le bassin Atlantique Nord, représentant une augmentation moyenne de 42% par décennie sur la période 1979 à 2017, comme illustré sur la figure 1.11. Cette augmentation est la plus élevée parmi tous les bassins cycloniques mondiaux. Chaque point du graphique de la figure 1.11 est calculé par tranche de 3 années, couvrant la période de 1979 à 2017.

1.5 Impacts assurantiels des ouragans sur le continent américain

Cette section se focalise sur les impacts assurantiels des ouragans dans différentes zones spécifiques du continent américain, à savoir les États-Unis et la région des Caraïbes. L’analyse approfondie se concentre principalement sur les États-Unis, en examinant les garanties touchées par les ouragans, les coûts estimés des dégâts, les répercussions sur les sociétés et les assureurs opérant dans cette région, ainsi que les différentes stratégies adoptées par les assureurs en matière de diversification du risque. En ce qui concerne la région des Caraïbes, une analyse plus synthétique est réalisée sans entrer dans les détails spécifiques.

1.5.1 Etats-Unis

1.5.1.1 Stratégies et dispositifs de protection

Aux Etats-Unis, les Etats fédérés peuvent solliciter du président américain une déclaration de catastrophe majeure, *major disaster*, suite à tout évènement naturel, notamment un ouragan (FEDERAL

¹⁸supérieurs ou égaux à la catégorie 3 sur l’échelle de Saffir-Simpson

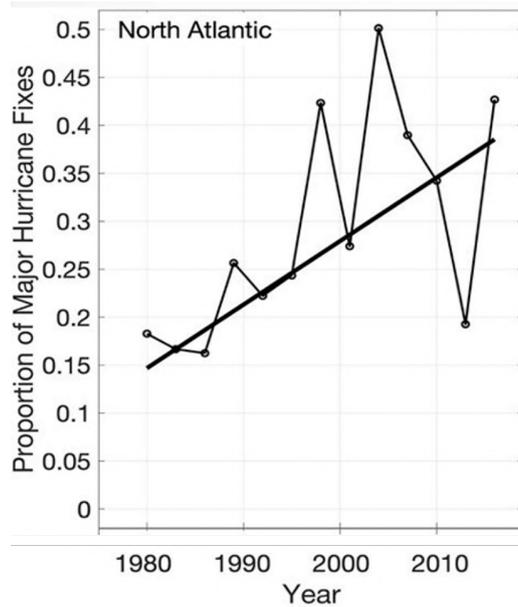


FIGURE 1.11 : Evolution de la proportion des ouragans majeures par décennie (KOSSIN et al., 2020)

EMERGENCY MANAGEMENT AGENCY, 2023b). Cette déclaration est effectuée à la demande du gouverneur de l'État concerné. Elle permet de mobiliser des aides fédérales en cas de dégâts causés par des événements naturels dépassant la capacité d'intervention de l'État et des administrations locales. Cependant, malgré cette assistance, l'aide moyenne accordée en cas de catastrophe est de l'ordre de 5000\$ par foyer, ce qui peut rapidement se révéler insuffisant face aux dommages causés par un ouragan (SCHLICHTER, 2023b).

En conséquence, les assurances privées jouent un rôle essentiel pour pallier ces insuffisances. Bien qu'il n'existe pas de police d'assurance spécifique pour les ouragans, il est possible de souscrire des assurances couvrant les deux principales sources de dégâts causés par ces phénomènes climatiques, à savoir les inondations et les vents violents (SCHLICHTER, 2023a).

Toutefois, la couverture d'assurance peut varier selon les États. En général, les polices d'assurance habitation standard couvrent les dommages causés par le vent, y compris ceux provoqués par un ouragan. Néanmoins, pour les résidents des régions côtières de l'Atlantique ou du Golfe du Mexique, où le risque d'ouragan est plus élevé, il est fortement recommandé de souscrire une assurance spécifique pour les vents et les inondations. Cette assurance peut être obtenue auprès d'une compagnie d'assurance privée ou auprès d'un pool d'assurance géré par l'État, qui est donc public. Il est également important de noter que les inondations sont généralement exclues de la plupart des polices d'assurance habitation (SCHLICHTER, 2023a). Ainsi, pour les résidents des régions exposées au risque d'ouragan, il est essentiel de souscrire à des polices d'assurance distinctes afin de bénéficier d'une couverture adéquate face à ces phénomènes climatiques dévastateurs. Cette précaution permet de mieux se prémunir contre les conséquences financières souvent importantes liées aux dommages causés par les ouragans. Il existe deux catégories de polices d'assurances pour faire face aux dommages causés par les ouragans aux États-Unis (SCHLICHTER, 2023a).

Assurance contre les tempêtes de vents : Cette assurance est proposée dans 19 États (Alabama, Connecticut, Delaware, Floride, Georgie, Hawaï, Louisiane, Maine, Maryland, Massachusetts, Mississippi, New Jersey, New York, Caroline du Nord, Pennsylvanie, Rhode Island, Caroline du Sud, Texas, Virginie) ainsi qu'à Washington D.C représentée sur la figure 1.12. Elle peut prendre la forme d'une police d'assurance-vent autonome, à souscrire distinctement d'une assurance habitation, ou être incluse

dans une police d'assurance-habitation, offrant ainsi une garantie intégrée. Dans les deux cas, cette assurance supplémentaire est soumise à une franchise spécifique pour les ouragans, les tempêtes nommées ou les tempêtes de vent, en plus de la franchise tous risques de l'assurance habitation. Contrairement aux franchises d'assurance habitation, les franchises liées aux tempêtes de vents sont généralement calculées en pourcentage de la valeur assurée, oscillant entre 1 et 5% (HARBOUR, 2022). Ces franchises se déclinent en trois catégories distinctes.

Franchise tempête : également appelée franchise vent/grêle, elle couvre les dommages causés par les ouragans, ainsi que les dégâts résultant de tornades et autres vents violents.

Franchise pour tempête nommée : cette franchise entre en vigueur lorsque la maison de l'assuré est endommagée par une tempête nommée par le National Weather Service ou le National Hurricane Center. Toutefois, elle n'est pas activée en cas de tornade ou de tempête de vent violent.

Franchise en cas d'ouragan : cette franchise est généralement déclenchée lorsque la tempête présente des vents suffisamment violents pour être qualifiée d'ouragan (supérieure à 118 km/h). Si un assuré souscrit une assurance contre les tempêtes de vents avec une franchise en cas d'ouragan, mais que la tempête qui a endommagé sa maison présente des vitesses de vent correspondant à une tempête tropicale, l'indemnisation excédera la franchise tous risques (le montant convenu de payer pour tout type de dommages couverts, y compris les tempêtes de vents, par son assurance habitation) et ne sera pas soumise à la franchise en cas d'ouragan.

Assurance contre les inondations : cette assurance couvre les dommages spécifiques causés par l'eau aux bâtiments et aux biens. Le National Flood Insurance Program (NFIP), le plus grand fournisseur d'assurance inondation du pays, est administré par l'agence fédérale de gestion des urgences, la *Federal Emergency Management Agency* (FEMA), et permet aux propriétaires, entrepreneurs et locataires de se procurer une assurance inondation soutenue par le gouvernement fédéral, donc publique (DIVISION DE LA LÉGISLATION COMPARÉE DE LA DIRECTION DE L'INITIATIVE PARLEMENTAIRE ET DES DÉLÉGATIONS, 2017). Cependant, l'assurance proposée par le gouvernement offre une couverture limitée, jusqu'à 250 000\$ pour les bâtiments et 100 000\$ pour les biens (SCHLICHTER, 2023b). Une couverture plus étendue et des limites plus élevées peuvent être obtenues auprès de compagnies d'assurance privées qui ne travaillent pas avec le NFIP.

1.5.1.2 Coûts estimés et perspectives pour l'assurabilité face aux ouragans

Le tableau 1.3 répertorie les dix ouragans ayant occasionné les coûts totaux et les pertes assurées estimés les plus élevés au cours de leur année respective (INSURANCE INFORMATION INSTITUTE, 2023b), ainsi que leur valeur ajustée selon l'indice des prix à la consommation de 2023 et de 2022 (NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION, 2023).

Les notations suivantes, exprimées en milliards de dollars (Md \$) sont adoptées :

- CT : Coût total estimé de l'ouragan durant l'année d'occurrence ;
- CT_{2023} : Coût total estimé de l'ouragan ajusté à l'indice des prix à la consommation de 2023 ;
- PA : Perte assurée totale estimée causée par l'ouragan durant l'année d'occurrence ;
- PA_{2022} : Perte assurée totale estimée ajustée à l'indice des prix à la consommation de 2022.

Les ouragans sont classés par ordre décroissant de la valeur du coût total ajusté selon l'indice des prix à la consommation de 2023, CT_{2023} . Ces chiffres permettent de mettre en lumière l'impact



FIGURE 1.12 : États avec franchise séparée pour les ouragans et les tempêtes de vent (INSURANCE, 2018)

économique considérable que ces ouragans ont eu et illustrent les défis auxquels les assurances et les résidents des zones touchées doivent faire face pour gérer ces événements climatiques dévastateurs.

Rang	Année	Ouragan	CT (Md \$)	CT ₂₀₂₃ (Md \$)	PA (Md \$)	PA ₂₀₂₂ (Md \$)
1	2005	Katrina	125	193.8	65	98.727
2	2017	Harvey	125	153.8	30	36.45
3	2022	Ian	113.1	115.2	52.5	52.5
4	2017	Maria	91.6	110.7	29.5	35.654
5	2012	Sandy	68.7	85.2	30	38.688
6	2021	Ida	75.2	81.7	36	39.256
7	2017	Irma	77.2	61.5	29.55	35.714
8	1992	Andrew	27.3	58.3	16	33.875
9	2008	Ike	38	41.7	18.2	24.815
10	2004	Ivan	26.1	32.8	7.11	11.1

TABLE 1.3 : Classement des 10 ouragans les plus coûteux aux États-Unis (INSURANCE INFORMATION INSTITUTE, 2023b)

Au cours des 15 dernières années, les pertes assurées liées aux ouragans ont considérablement augmenté. En ajustant ces coûts à l’inflation, il est observable que neuf des dix ouragans les plus coûteux de l’histoire des États-Unis ont touché le pays depuis 2004. De plus, il est important de souligner que ces neuf ouragans figurent également parmi les dix catastrophes naturelles les plus coûteuses aux États-Unis (INSURANCE INFORMATION INSTITUTE, 2023d). Cette tendance s’explique principalement par le fait que de plus en plus de personnes se sont installées dans des zones à risque depuis les années 1940, en particulier le long des côtes (INSURANCE INFORMATION INSTITUTE, 2023c).

Le Bureau du recensement des États-Unis rapporte que les maisons construites aujourd’hui sont plus grandes et plus chères qu’auparavant, ce qui entraîne des coûts de remplacement plus élevés en cas de sinistre. L’accroissement de la taille des maisons et de la valeur des biens qu’elles contiennent contribue ainsi à la hausse des pertes assurées. Les données suggèrent également que les changements démographiques jouent un rôle de plus en plus important dans l’augmentation des coûts des catastrophes naturelles, en plus des facteurs climatiques (INSURANCE INFORMATION INSTITUTE, 2023c).

Une étude réalisée par Aon¹⁹ (AON, 2023) en collaboration avec l’université de Columbia a estimé que, selon certains scénarios, les pertes assurées liées aux ouragans aux États-Unis pourraient augmenter d’au moins 10% sur une période de 20 ans en raison des changements climatiques. Cependant, il est essentiel de noter que cette estimation ne prend pas en compte l’augmentation des risques due à des facteurs non climatiques, tels que les nouveaux développements et l’inflation. Les conclusions de cette étude reposent sur l’évolution de la recherche climatique et pourraient être révisées à mesure que de nouvelles connaissances seront acquises. Cette situation soulève des questions importantes quant à l’avenir de l’assurabilité face aux ouragans et à la nécessité d’adapter les stratégies d’assurance pour faire face à ces défis croissants.

La Figure 1.13 présente l’évolution du coût total des ouragans par décennie, ayant dépassé le seuil du milliard de dollars corrigés de l’inflation. Le nombre d’ouragans de l’Atlantique d’une valeur d’un milliard de dollars a presque doublé entre les années 1980 et les années 2010, et les coûts corrigés de l’inflation ont été multipliés par plus de onze.

Ces augmentations importantes sont attribuées à divers facteurs. Tout d’abord, le changement climatique est considéré comme l’un des principaux contributeurs, car il a augmenté l’intensité des ouragans, entraînant ainsi des dégâts plus importants (INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, 2023). Ensuite, l’augmentation du nombre de personnes s’installant dans les zones côtières à haut risque a également joué un rôle significatif dans l’augmentation des coûts des ouragans. Les régions côtières attirent de plus en plus de population en raison de divers facteurs, tels que l’attrait des zones côtières pour la résidence et le tourisme, ce qui augmente la vulnérabilité des communautés face aux ouragans (PHILBRICK et WU, 2022).

Enfin, la hausse spectaculaire des coûts de construction (QUILLEN, 2022) depuis 1980 a également contribué à accroître les coûts totaux des ouragans. La valeur croissante des biens immobiliers (POTTER, 2022), associée à une augmentation de la taille des maisons et à l’inflation (UNITED STATES BUREAU OF LABOR STATISTICS, 2023), a entraîné des coûts de remplacement plus élevés en cas de sinistre (INSURANCE INFORMATION INSTITUTE, 2023b).

Face à la grande volatilité du risque ouragan et aux coûts importants engendrés par ces catastrophes, des dispositifs de réassurance (INSURANCE INFORMATION INSTITUTE, 2023a) ont été mis en place pour permettre aux assureurs de faire face à de telles situations. Cependant, l’ouragan Andrew, qui a frappé le sud de la Floride en 1992 et entraîné 16 milliards de dollars de pertes assurées à l’époque (représenté sur le tableau 1.3), a révélé que les assureurs américains avaient sérieusement sous-estimé l’étendue de leur responsabilité face à de telles méga-catastrophes. Avant cet événement, le secteur de l’assurance estimait qu’une perte de 8 milliards de dollars constituait le maximum possible en cas de catastrophe (INSURANCE INFORMATION INSTITUTE, 2023a). Cette expérience a conduit les réassureurs à reconsidérer leur position, ce qui a incité les compagnies d’assurance primaires à réévaluer leurs besoins en matière de réassurance catastrophe.

Depuis lors, les rétentions (c’est-à-dire les montants que les assureurs sont prêts à supporter eux-mêmes avant de faire appel à la réassurance) et la coassurance (partage du risque à différents niveaux entre assureurs) ont augmenté de manière significative au cours des dernières décennies. Ainsi lorsque les prix de la réassurance sont élevés et les capacités rares en raison du risque élevé de catastrophes cycloniques, certaines compagnies d’assurances se tournent vers les marchés des capitaux (INSURANCE

¹⁹Acteur majeur dans les domaines de la gestion des risques, du courtage d’assurance et de réassurance, du conseil en ressources humaines et du management et de la souscription d’assurances spécialisées

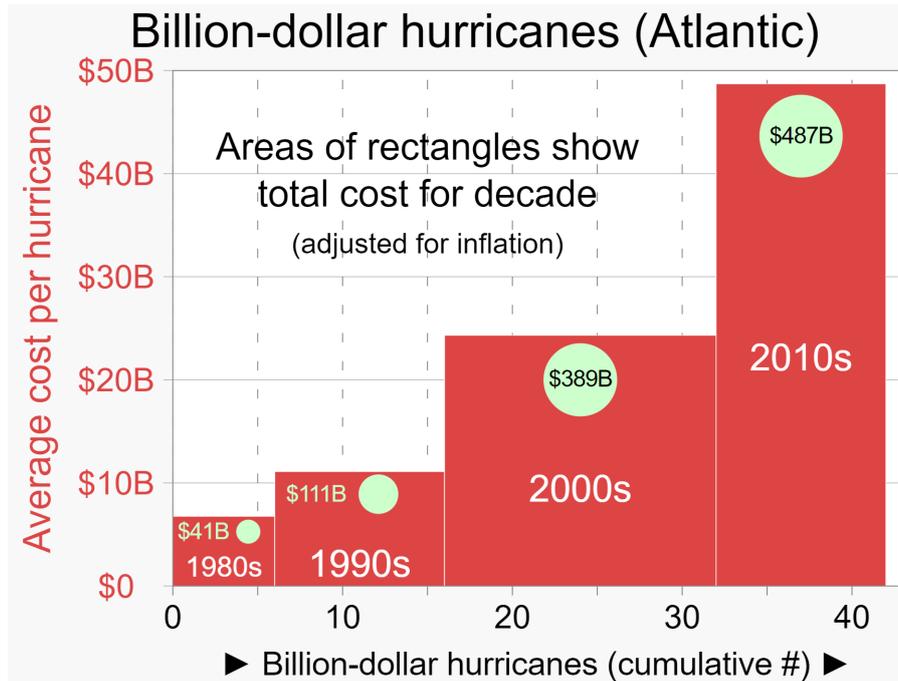


FIGURE 1.13 : Evolution du coût total des ouragans par décennie (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2022)

INFORMATION INSTITUTE, 2023a) pour trouver des arrangements financiers innovants telles que la titrisation des risques d'assurance, des options de vente d'actions ou des obligations catastrophes²⁰. Ces stratégies financières innovantes permettent aux compagnies d'assurance de mieux gérer leur exposition aux risques liés aux ouragans tout en trouvant des moyens de financer les coûts élevés associés à ces catastrophes naturelles dévastatrices. Cependant, il est essentiel de noter que ces approches comportent également des défis et des risques, notamment en ce qui concerne la capacité des investisseurs à absorber ces risques et à faire face à de multiples catastrophes simultanées ou répétées. Ainsi, l'industrie de l'assurance continue de chercher des solutions innovantes pour mieux gérer les risques liés aux ouragans et garantir une protection financière adéquate aux assurés face à ces événements climatiques extrêmes²¹.

L'assurabilité face aux ouragans aux États-Unis est confrontée à de nombreuses sources d'incertitudes et de défis pour l'avenir (CHO, 2022). Les coûts croissants des ouragans au cours des dernières années ont mis à rude épreuve de nombreuses compagnies d'assurances, tant privées que publiques. De nombreuses assureurs ont du mal à payer leurs sinistres, entraînant des faillites dans le secteur privé et des augmentations significatives des primes dans le secteur public et privé (CHRISTOPHER et al., 2023).

Un exemple concret est le programme NFIP géré par la FEMA, qui en 2021, a commencé à fixer des taux équivalents au risque réel d'inondation pour mieux communiquer le véritable danger aux propriétaires et limiter les pertes pour le gouvernement. Cela s'est traduit par des hausses de prix considérables pour de nombreux assurés, rendant l'accès à l'assurance de plus en plus difficile pour certains propriétaires. Certaines primes pour les propriétés côtières ont grimpé de 700 ou 800\$ à 4000 ou 5000\$ (DORE, 2022).

Selon Swiss Re (CHO, 2022), les primes des polices d'assurance habitation devraient continuer à augmenter de manière significative dans les années à venir, environ 5,3% par an d'ici 2040. Par exemple,

²⁰ titre spécialisé qui augmente la capacité des assureurs à fournir une protection d'assurance en transférant le risque aux investisseurs de l'obligation

²¹ Plus de renseignements disponibles sur le site officiel *Insurance Information Institute* : <https://www.iii.org>

les primes ont augmenté de 12,1 % aux États-Unis entre 2021 et 2022. Les taux d'augmentation des primes aux États-Unis sont déjà élevés, en particulier dans les régions où les catastrophes naturelles sont plus fréquentes. Cela rendra l'assurance de plus en plus coûteuse pour les propriétaires, ce qui peut entraîner une diminution de la demande d'assurance ou une couverture insuffisante pour certains propriétaires.

En décembre 2022, le commissaire aux assurances de la Louisiane, James Jim Donelon déclare que le marché de l'assurance de l'État « est en crise » suite aux dégâts causés par les ouragans ces dernières années. En Louisiane, de petites compagnies d'assurance soutenues par Citizens, une compagnie mandatée par l'État pour fournir une couverture contre les tempêtes aux propriétaires qui ne pouvaient pas trouver d'assurance privée, ont dû augmenter leurs primes de 63% pour atteindre une moyenne de 4700\$ par an (CHRISTOPHER et al., 2023).

Aujourd'hui, de nombreuses compagnies d'assurances quittent certains États où elles exerçaient car elles sont devenues insolubles, et d'autres refusent de souscrire de nouveaux propriétaires en raison des risques élevés associés aux ouragans. Les ouragans récents ont entraîné une vague de faillites dans le secteur de l'assurance, mettant en évidence les difficultés croissantes auxquelles sont confrontées les compagnies d'assurances pour gérer les risques liés aux ouragans. Récemment, suite au passage de l'ouragan Laura en Louisiane en 2020, plus de 50 compagnies d'assurance ont cessé d'accepter de nouveaux clients, et 11 se sont déclarées en faillite. De même, en Floride, plus de 15 compagnies d'assurance ont fait faillite suite à cet événement (SACKS, 2023).

Enfin, le programme NFIP a accumulé une dette importante dépassant les 20 milliards de dollars, affaiblissant davantage le secteur public en matière d'assurance (CHO, 2022).

Dans l'ensemble, l'avenir de l'assurabilité face au risque d'ouragan aux États-Unis est incertain. Les coûts croissants, les difficultés financières des assureurs, les faillites dans le secteur de l'assurance et l'accumulation de dettes du NFIP sont autant de facteurs qui soulèvent des questions sur la viabilité à long terme de l'assurance contre les ouragans. Pour faire face à ces défis, des solutions novatrices seront nécessaires pour mieux gérer les risques liés aux ouragans et garantir une protection adéquate aux résidents face à ces événements climatiques extrêmes (CHO, 2022).

1.5.2 Région des Caraïbes

En 2004, les ouragans ont provoqué des dommages économiques importants dans plusieurs petites îles des Caraïbes, entraînant des pertes estimées à plus de 4 milliards de dollars. Face à cette situation, les gouvernements de la région ont sollicité l'aide de la Banque mondiale pour faciliter l'accès à des mécanismes d'assurance contre les catastrophes. C'est ainsi qu'est née la CCRIF, *Caribbean Catastrophe Risk Insurance Facility*, ce projet lancé en juin 2007 après deux années de planification, a été créé dans le but de limiter l'impact financier des désastres naturels (INSURANCE INFORMATION INSTITUTE, 2023a).

Financée par le Japon, le Canada, l'Union Européenne et la Banque de Développement de la Caraïbe (CDB), la CCRIF vise à assurer une réponse rapide suite à une catastrophe et à maintenir les fonctions de l'État. Elle a été conçue comme un fonds régional de catastrophe destiné aux gouvernements des Caraïbes, offrant une source de liquidités à court terme lorsqu'une situation catastrophique se déclenche. Son objectif principal est de fournir un soutien financier immédiat pour faire face aux conséquences des ouragans et des tremblements de terre dans la région (POPOVIC, 2019).

Les 21 pays assurés par la CCRIF comprennent Anguilla, Antigua et Barbuda, les Bahamas, Barbade, Bermudes, les Îles Vierges Britanniques, les Îles Caïmans, la Dominique, la Grenade, Haïti, la Jamaïque, Montserrat, Saint Kitts et Nevis, Sainte Lucie, Saint Maarten, Saint Vincent et les Grenadines, Trinidad et Tobago, Turks et Caïcos, le Panama et le Nicaragua (POPOVIC, 2019).

Fonctionnant comme une société d'assurance mutuelle, la CCRIF permet aux pays membres de mutualiser leurs risques dans un portefeuille diversifié et d'acquérir des produits de réassurance ou d'autres mécanismes de transfert de risques sur les marchés financiers internationaux, avec des économies pou-

vant atteindre jusqu'à 50% par rapport à un achat individuel de protection contre les catastrophes pour chaque pays. De plus, étant donné qu'un ouragan ou un tremblement de terre n'affecte en moyenne qu'un à trois pays des Caraïbes au cours d'une année donnée, chaque pays contribue moins à la réserve commune que s'il détenait ses propres réserves (INSURANCE INFORMATION INSTITUTE, 2023a).

Les membres paient des primes en fonction de leur niveau probable d'utilisation des fonds de la réserve. À mesure que les pays améliorent leurs normes de construction pour se protéger davantage contre les catastrophes, les primes diminuent en conséquence.

La CCRIF utilise une approche d'assurance paramétrique pour déterminer les paiements d'indemnisation, ce qui permet des versements rapides en cas de sinistre. Dans ce système, les paiements sont déclenchés par la survenue d'un événement spécifique, qui peut être objectivement vérifié, tel qu'un ouragan atteignant une certaine vitesse de vent ou un tremblement de terre atteignant un certain seuil de secousses, plutôt que par une évaluation des pertes réelles effectuée par un expert, un processus qui peut prendre plusieurs mois. Les montants des indemnisations sont calculés à partir de modèles qui estiment l'impact financier de la catastrophe (INSURANCE INFORMATION INSTITUTE, 2023a).

Par exemple, suite aux cyclones tropicaux Maria et Irma, la CCRIF a versé un total de 55 millions de dollars US à neuf gouvernements membres (CCRIF SPC, 2019).

1.6 Présentation des bases de données

Dans cette section, deux bases de données majeures largement utilisées dans la littérature scientifique sont présentées pour l'étude des phénomènes cycloniques.

1.6.1 Base HURDAT2

La base de données HURDAT2 est un ensemble de données disponible en open-source sur le site officiel du National Hurricane Center (NATIONAL HURRICANE CENTER, 2023a), qui recense de manière systématique les caractéristiques et les trajectoires des phénomènes cycloniques depuis le milieu du 19ème siècle sur le bassin Atlantique Nord et depuis le milieu du 20ème siècle sur le bassin Pacifique Est.

HURDAT2 offre une vision historique détaillée des événements cycloniques mineurs et majeurs de leur naissance à leur extinction, permettant une analyse approfondie de leur fréquence, intensité, trajectoire et leurs impacts. Les données contenues dans cette base de données comprennent des informations telles que la vitesse des vents maximum, la pression atmosphérique, la localisation géographique ou encore la date des différentes observations météorologiques associées aux cyclones sur un intervalle de 6h (UTC).

Par ailleurs, cette base est considéré comme le meilleur relevé de trajectoire historique des phénomènes cycloniques, en anglais *Best Track Data*. Elle est unanimement reconnue comme la référence incontournable pour l'étude des ouragans dans le bassin Atlantique-Nord et est largement utilisée par les chercheurs, les scientifiques et les acteurs de la communauté météorologique et climatologique à travers le monde en raison de sa qualité, de sa précision et de sa valeur scientifique indéniable.

Une version étendue et améliorée de HURDAT2, connue sous le nom de *Tropical Cyclone Extended Best Track Dataset* (NATIONAL HURRICANE CENTER, 2023b), développée par le NHC a été mis en place afin d'avoir des informations plus poussées sur les tempêtes et ouragans. Ces informations regroupent des caractéristiques supplémentaires sur la pression et les rayons des vents ainsi que l'ajout du rayon maximum des vents et du diamètre de l'oeil. Cependant, ces données ne sont disponibles que depuis 1988 en raison de l'absence d'outils technologiques permettant de mesurer ces grandeurs avant cette date.

1.6.2 Base IBTrACS

La base de données IBTrACS (International Best Track Archive for Climate Stewardship) est une ressource essentielle pour l'étude des ouragans à l'échelle mondiale. Elle constitue une compilation exhaustive des trajectoires, des caractéristiques et des paramètres associés aux cyclones tropicaux. IBTrACS a été développé en collaboration avec tous les centres météorologiques régionaux spécialisés de l'Organisation météorologique mondiale (OMM), et d'autres organisations et scientifiques du monde du climat. La base de données est disponible sur le site de l'agence américaine d'observation océanique et atmosphérique (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b).

IBTrACS offre une vision globale et unifiée des événements cycloniques. Elle est considérée comme la collection mondiale de cyclones tropicaux la plus complète qui soit. La base fusionne des données récentes et historiques sur les cyclones tropicaux provenant de plusieurs agences afin de créer un ensemble de données unifié, accessible au public, facilitant ainsi les comparaisons et les analyses entre les différents bassins océaniques. En résumé, la base de données IBTrACS offre une compilation exhaustive des données sur les tempêtes tropicales, en combinant les données de HURDAT2 et d'autres bases complémentaires provenant d'autres centres météorologiques. Cela permet d'avoir une perspective globale et détaillée des phénomènes cycloniques, permettant aux chercheurs et scientifiques d'avoir une base de référence solide pour leurs études.

Sur le tableau 1.4, sont exposées quelques caractéristiques de l'ouragan IRMA (1.6), extraites de la base IBTrACS.

Année	Nom	Date/Heure	Latitude	Longitude	MSW (kt)	PC (hPa)	SS
2017	IRMA	31/08/2017 06 :00	16.4	-32.5	65	994	1
2017	IRMA	31/08/2017 09 :00	16.5	-32.9	72	988	1
2017	IRMA	31/08/2017 12 :00	16.7	-33.4	80	983	1
2017	IRMA	31/08/2017 15 :00	16.9	-33.8	87	976	2
...

TABLE 1.4 : Quelques caractéristiques de l'ouragan IRMA dans la base de données IBTrACS

1.6.3 Choix de la base d'étude

Dans le cadre de cette étude, la base de données IBTrACS est choisie pour plusieurs raisons.

Tout d'abord, la base de données IBTrACS offre une portée plus large en incluant les ouragans du monde entier, ce qui permet d'intégrer des événements provenant de différentes régions géographiques. Cette caractéristique pourrait s'avérer cruciale pour construire et calibrer une fonction de coût robuste (chapitre 4) en prenant en compte des taux de destruction spécifiques à chaque pays. La disponibilité de ces informations contribue à la construction d'une courbe de vulnérabilité plus précise, ce qui est essentiel pour évaluer les risques et déterminer des stratégies de tarifications de police d'assurances adaptées.

En outre, cette approche permet d'obtenir une vision globale et complète des phénomènes cycloniques, enrichissant ainsi l'analyse des risques. Un autre avantage important de la base de données IBTrACS est la résolution temporelle plus fine, avec un intervalle de relevé de 3 heures, par rapport à l'intervalle de 6 heures de la base de données HURDAT2. Cette précision temporelle accrue fournit des données plus détaillées sur les trajectoires et les intensités des tempêtes, améliorant ainsi la précision des analyses et des modèles.

Enfin, la base de données IBTrACS consolide non seulement les informations de HURDAT2, mais également celles provenant de diverses agences météorologiques renommées, telles que le JTWC (*Joint Typhoon Warning Center*), le CPHC (*Central Pacific Hurricane Center*), le WMO (*World Meteorological*

Organization), et d'autres encore. Cette collaboration entre différentes sources de données réputées garantit la fiabilité et la qualité des informations utilisées dans l'analyse des tempêtes.

Toutefois, il convient de souligner que le choix de la base de données IBTrACS, par opposition à HURDAT2, ne sous-entend en aucun cas que cette dernière est dépourvue d'utilité pour l'étude en question, ou qu'elle aurait été inadéquate. Au contraire, la base de données HURDAT2 aurait également pu être largement suffisante pour mener à bien cette étude. Cependant, après une analyse comparative, la préférence s'est portée sur la base de données IBTrACS, en raison des avantages spécifiques qu'elle offre, notamment l'avantage significatif d'inclure des données cycloniques sur les bassins mondiaux, ce qui pourrait s'avérer être d'une grande utilité pour les étapes ultérieures de cette recherche.

1.6.4 Analyse approfondie de la base IBTrACS

Le tableau 1.5 présente une synthèse des chiffres clés issus de l'analyse de la base de données IBTrACS, en accord avec les paramètres spatiaux et temporels définis dans le chapitre suivant 2.1. Ces données constituent des informations essentielles qui serviront de base pour les analyses ultérieures.

Bases de données	IBTrACS
Période observée	1982-2022
Nombre total de tempêtes	649
Moyenne annuelle	15.8
Ecart-type	5.3
Nombre total d'ouragans	279
% de conversion (tempête \Rightarrow ouragan)	43%

TABLE 1.5 : Chiffres clés de la base de données IBTrACS

La base de données IBTrACS comprend un total de 163 variables. Cependant, de nombreuses colonnes sont vides pour une grande majorité de ces variables. En effet cette base de données est une compilation de différentes bases établies par l'Organisation Météorologique Mondiale (OMM), *World Meteorological Organization (WMO)* dans chaque bassin océanique, avec la contribution d'une ou plusieurs agences météorologiques. Dans ce cadre, l'OMM désigne une agence responsable de la collecte des données pour les intégrer dans IBTrACS, identifiée par la variable *WMO_AGENCY* (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2019).

En tout, la base de données contient 14 ensembles de données élaborés par des scientifiques ou des agences météorologiques internationales impliquées dans IBTrACS. Ces ensembles de données sont répartis comme suit.

USA_Agency_Information : Fournit une liste de variables provenant d'agences américaines telles que le *National Hurricane Center*, *Joint Typhoon Warning Center*, *Central Pacific Hurricane Center* ou le Centre météorologique régional spécialisé de l'OMM à Miami et Honolulu.

RSMC Tokyo : *Regional Specialized Meteorological Center Tokyo*, agence météorologique japonaise chargée des prévisions officielles concernant les typhons dans le Pacifique Nord.

CMA Shanghai Typhoon Institute : *Chinese Meteorological Administration Shanghai Typhoon Institute*, fournit des informations sur les typhons dans l'ouest du Pacifique Nord.

Hong Kong Observatory (HKO) : Fournit des informations sur les typhons dans l'ouest du Pacifique Nord, en mettant l'accent sur les typhons qui s'approchent de Hong Kong.

RSMC New Delhi (IMD) : *Regional Specialised Meteorological Center* à New Delhi, est chargé des prévisions relatives aux cyclones tropicaux dans le nord de l’océan Indien.

RSMC La Reunion (MeteoFrance) : *Regional Specialised Meteorological Center* à La Réunion et géré par MétéoFrance, est responsable des rapports sur les cyclones tropicaux dans le sud de l’océan Indien.

Bureau de météorologie (BoM) : Centre d’alerte aux cyclones tropicaux de l’OMM autour de grandes villes côtières australiennes : Perth, Darwin et Brisbane.

TCWC Wellington : Service météorologique de Nouvelle-Zélande, responsable des rapports sur les cyclones tropicaux dans la partie sud du Pacifique Sud.

RSMC Nadi (Fidji) : *Regional Specialised Meteorological Center* à Nadi, responsable des cyclones tropicaux dans la partie nord du Pacifique Sud.

DataSet 824 (DS824) : Compilation de données sur les tempêtes élaborée par le *National Center for Atmospheric Center / University Corporation for Atmospheric Research* (NCAR/UCAR) Research Data Archive (RDA) dans les années 1960 et 1970.

TapeDeck (TD) 9636 : Base de données mondiale élaborée par la NOAA et le *National Climatic Center* dans les années 1960 et 1970. Cette base de données n’a pas été mise à jour depuis les années 1980 et est utilisée dans IBTrACS en dehors de l’Atlantique Nord et avant 1980.

TapeDeck (TD) 9635 : Base de données élaborée dans les années 1970 par la Marine des Etats-Unis (*United States Navy* et la NOAA, fournissant des informations sur les tempêtes de 1945 à 1976.

Neumann Southern Hemisphere Dataset : Ensemble de données collectées par le météorologue Charlie Neumann sur les meilleures trajectoires de tempêtes de l’hémisphère sud de 1960 à 2007.

Chenoweth Dataset : Michael L. Chenoweth a réalisé une réanalyse des ouragans historiques dans l’océan Atlantique Nord, indépendante des données de la base HURDAT2 et s’étend de 1851 à 1898. Ainsi sur les 163 variables disponibles dans la base IBTrACS, il est important de noter que de nombreuses variables ne contiennent aucune donnée puisque chaque ensemble de données est spécifique à une zone océanique précise parmi les 14 ensemble de données constituant la base IBTrACS.

L’étude se concentrant sur le bassin Atlantique Nord, le tableau 1.6 présente une description des variables les plus pertinentes dans ce cadre extraites de la base. Bien que cette liste ne soit pas exhaustive, elle englobe toutes les variables susceptibles d’être utilisées dans cette étude.

Après avoir exposé les concepts liés aux tempêtes et aux ouragans, ainsi que les bases de données qui enregistrent leurs caractéristiques, et avoir présenté une synthèse de l’impact financier résultant de ces phénomènes dévastateurs sur le continent américain, le chapitre suivant se concentre sur la modélisation statistique des ouragans. Cette étape vise à développer une modélisation optimale pour 10 000 saisons cycloniques dans le bassin Atlantique Nord. Cette démarche s’inscrit dans le cadre de l’objectif final, à savoir évaluer le risque cyclonique sur le territoire des États-Unis. La problématique centrale de cette étude est de modéliser une carte d’exposition au risque d’ouragan aux États-Unis. Cette carte, basée sur des données fiables et des modèles statistiques avancés, permettra de mieux comprendre comment les ouragans peuvent affecter différentes régions du pays et quels sont les facteurs

Variable	Description
SID	Identifiant du système (<i>Storm Identifier</i>)
SEASON	Année d'occurrence
BASIN	Bassin océanique (exemple : NA pour <i>North Atlantic</i>)
NAME	Nom du système fourni par l'agence
ISO_TIME	Date et Heure sous la norme ISO en coordonnées UTC
WMO_AGENCY	Agence responsable des données telles qu'elles sont répertoriées
USA_AGENCY	Le fichier de l'agence fournissant les informations (exemple : HURDAT_ATL pour les données de la base HURDAT2 du bassin Atlantique Nord)
USA_ATCF_ID	Identifiant du système dans la base source de l'agence WMO_AGENCY
USA_LAT	Latitude en degré nord
USA_LON	Longitude en degré est
DIST2LAND	Distance du centre de la tempête à la côte terrestre la plus proche (en km)
LANDFALL	Distance à la côte terrestre la plus proche entre la position actuelle et la prochaine position signalée dans trois heures (en km). Pour une position donnée : <ul style="list-style-type: none"> – Si LANDFALL = 0, alors le système traverse une zone terrestre ; – Si DIST2LAND = LANDFALL, alors le système s'éloigne de la zone terrestre (car la position actuelle est la plus proche de la zone terrestre pour les trois prochaines heures) ; – Si LANDFALL < DIST2LAND, alors le système se déplace en direction d'une zone terrestre (car la prochaine position signalée sera plus proche de la zone terrestre que la position actuelle).
USA_RECORD	Indique la nature du système au point enregistré (Perturbation, Dépression Tropicale, Tempête Tropicale, Ouragan, etc. . .)
USA_WIND	Vitesse maximale des vents soutenus 1.3 (en noeuds)
USA_PRES	Pression minimale au centre de l'ouragan et au niveau de la mer 1.3 (en hPa)
USA_SSHS	Classification sur l'échelle de Saffir-Simpson
USA_Rx_y	$x \in \{34, 50, 64\}$ et $y \in \{NE, SE, NW, SW\}$, Rayon maximum atteint par les vents de vitesse x (en noeuds) selon la direction du quadran y (Nord-Est, Sud-Est, Nord-Ouest, Sud-Ouest)
USA_RMW	Distance entre le centre du système et sa bande de vents les plus forts
USA_EYE	Diamètre de l'oeil du système, s'il a atteint le stade d'ouragan
IFLAG	Chaîne de 14 caractères indiquant la nature de la donnée issue de chacun des 14 ensemble de données (exemple : P_____O__I_V__). La signification de chaque caractère correspond à : <ul style="list-style-type: none"> – _ = les données manquantes. Aucune information fournie ; – O = les données sont telles que fournies par l'agence ; – I = les données sur la position sont telles que fournies par l'agence, mais les données d'intensité (et probablement d'autres) ont été interpolées ou complétées ; – V = les données sur la position et l'intensité sont telles que fournies par l'agence, mais quelques données ont été interpolées ou complétées ;

TABLE 1.6 : Liste non exhaustive des variables de la base IBTrACS

qui influencent le risque. De plus, cette modélisation aura un impact significatif sur le secteur assurantiel, en permettant aux compagnies d'assurance d'estimer avec précision leurs engagements financiers en cas de catastrophe cyclonique. Ainsi, le prochain chapitre abordera la modélisation statistique des ouragans en trois phases distinctes : la fréquence des ouragans, leurs trajectoires et leur intensité. Ces modèles avancés permettront d'approfondir la compréhension des phénomènes cycloniques et d'affiner la capacité à prédire leur comportement futur. A l'issue de cette modélisation il sera possible de simuler un échantillon de saisons cycloniques du bassin Atlantique Nord qui servira d'appui pour analyser les impacts financiers liée au risque ouragan, en particulier pour le secteur assurantiel (chapitre 4).

Chapitre 2

Simulation de la survenance des ouragans

Dans la littérature scientifique, diverses approches de modélisation ont été utilisées pour explorer et analyser les caractéristiques des tempêtes, en se focalisant sur leur fréquence, leur trajectoire et leur intensité. Ces modèles, élaborés par des chercheurs de différentes disciplines, ont apporté des avancées significatives dans la compréhension des phénomènes cycloniques. Selon les objectifs recherchés, ces modèles peuvent se regrouper en deux grandes catégories : dynamique et statistique (NATIONAL HURRICANE CENTER, 2019).

Les modèles dynamiques : ces modèles sont extrêmement complexes et utilisent des superordinateurs pour résoudre les équations mathématiques qui régissent la physique et le mouvement de l'atmosphère. Ils se divisent en deux sous-catégories :

Les modèles de circulation générale (MCG) : ces modèles sont utilisés pour étudier les variations climatiques à long terme dans une région, ainsi que les changements potentiels dans les schémas de tempêtes et les caractéristiques des ouragans. Ils intègrent des équations mathématiques complexes telles que les équations de Navier-Stokes, de la mécanique des fluides, et les lois de la thermodynamique pour représenter la circulation atmosphérique, les échanges de chaleur et d'humidité, et d'autres processus physiques. Ils sont utilisés principalement pour des études climatiques à long terme et permettent de comprendre les mécanismes complexes régissant le système climatique.

Les modèles de prévision numérique du temps (MPNT) : ces modèles sont utilisés pour effectuer des prévisions à court terme de la trajectoire et de l'intensité des tempêtes. Ils se basent sur des équations de la dynamique des fluides et des équations thermodynamiques pour simuler l'évolution de l'atmosphère. Parmi les MPNT couramment utilisés, on trouve le modèle américain Global Forecast System (GFS) développé par la NOAA et le modèle européen ECMWF (European Centre for Medium-Range Weather Forecasts). Ils sont utilisés par les agences météorologiques nationales et internationales telles que le NHC pour fournir des prévisions météorologiques au grand public.

Les modèles statistiques : ces modèles exploitent les relations statistiques entre les paramètres des tempêtes telles que leur position, la vitesse des vents ou la pression centrale. En s'appuyant sur des méthodes statistiques et l'analyse des données historiques, ces modèles permettent de générer un échantillon de tempêtes. Plusieurs méthodes statistiques sont utilisées dans la littérature scientifique, par exemple des modèles de régression (VICKERY et al., 2000), des chaînes de Markov (RISI, 2005) ou encore des séries temporelles (BLOEMENDAAL et al., 2020). En raison de contraintes de faisabilité et de limitations des outils disponibles, la modélisation des tempêtes dans cette étude est réalisée via des modèles statistiques plutôt que des modèles dynamiques. L'objectif de ce chapitre consiste à exposer les différentes méthodes statistiques employées, ainsi que les hypothèses sous-jacentes, en vue de la

modélisation de la fréquence et de la genèse d'un ouragan. Les résultats issus de ces approches sont également présentés et analysés. L'étude concernant la dynamique des trajectoires et de l'intensité des ouragans est traitée dans le chapitre 3.

2.1 Cadre et périmètres

2.1.1 Choix de modélisation

Avant de se lancer dans la modélisation, il est essentiel de déterminer avec précision ce que l'on souhaite exactement modéliser. L'un des objectifs majeurs de cette étude consiste à modéliser de manière exhaustive les tempêtes, en englobant leur genèse, leur trajectoire ainsi que leur intensité. En adoptant cette approche globale, le but est de saisir tous les aspects importants et les étapes clés du processus de développement des tempêtes. Ce choix découle de plusieurs observations fondamentales, illustré à travers l'exemple de l'ouragan Florence ci-dessous.

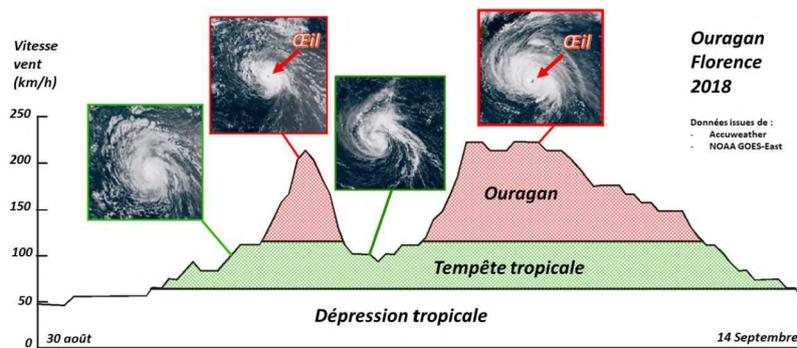


FIGURE 2.1 : Comportement de l'ouragan Florence en été 2018 (PLUVIOSE, 2019)

La figure 2.1 représentant l'évolution de l'ouragan Florence en fonction de sa vitesse de vent, est apparu durant l'été 2018 dans le bassin Atlantique Nord pour ensuite atteindre les côtes du Sud-Est des Etats-Unis, plus précisément l'Etat de la Caroline.

L'observation de ce phénomène naturel se révèle intéressante car son évolution se décompose en 3 phases. Dans un premier temps (**phase 1**) le système suit une évolution classique d'une dépression tropicale en tempête tropicale pour ensuite atteindre le stade d'ouragan. Puis dans un second temps (**phase 2**) il arrive à se rétrograder de lui-même en tempête tropicale selon les conditions météorologiques qu'il rencontre pendant son trajet pour finalement repasser une seconde fois au stade d'ouragan (**phase 3**), le tout sur une période très courte de l'ordre de quelques jours.

Ainsi une question cruciale se pose lors de l'étude de ces phénomènes : doit-on se concentrer sur les trajectoires des ouragans déjà formés, ou bien sur les trajectoires des tempêtes qui pourraient potentiellement devenir des ouragans ?

La modélisation basé uniquement sur les ouragans déjà formés semble être un choix plus évident si l'objectif final est d'analyser les impacts liées à ces phénomènes. Cependant, cette approche peut limiter la compréhension de la formation et de l'évolution des ouragans, car elle ne prend pas en compte les étapes précoces de leur formation et de leur développement. Plusieurs interrogations risquent d'intervenir :

- Quel sera le point de genèse considéré pour un ouragan ? Par exemple dans le cas de l'ouragan Florence montrant deux phases de genèses, considère t-on celui de la phase 1 ou 3 ?
- Quel filtre appliqué pour la sélection des ouragans dans la base de données ? Vitesse des vents supérieures à 118 km/h avec ou sans possibilité de rétrogradation ?

- Sans les étapes précoces de formation et de développement, les ouragans simulés seront-ils cohérents et réalistes ?

En revanche, la modélisation des trajectoires des tempêtes qui peuvent devenir des ouragans présente des avantages pour mieux comprendre les conditions environnementales qui conduisent à la formation et à l'intensification des ouragans, ainsi que les facteurs qui les font s'intensifier ou s'affaiblir. Cette approche peut également aider les prévisionnistes à mieux anticiper les risques pour les communautés côtières et les populations affectées.

De plus, comme l'a montré l'évolution du comportement de l'ouragan Florence sur la figure 2.1, un ouragan peut repasser au stade de tempête tropicale avant de redevenir un ouragan. Ainsi, il peut être difficile de savoir si la tempête qui a finalement donné naissance à l'ouragan a commencé avant ou après le point initial choisi pour l'analyse. Cette incertitude dans les données peut affecter les résultats de nos modélisations et rendre plus difficile la prédiction de la trajectoire et de l'intensité des ouragans.

Afin d'éviter ces incertitudes et de capturer l'ensemble du processus cyclonique, il est choisi de **modéliser les tempêtes dans leur globalité, depuis leur naissance jusqu'à leur disparition**. Cette approche permet de mieux appréhender les mécanismes sous-jacents à la formation et à l'évolution des tempêtes, tout en prenant en compte les interactions complexes entre les différents paramètres atmosphériques. Une telle démarche offre une vision plus complète et détaillée des tempêtes du bassin Atlantique Nord, contribuant ainsi à une meilleure compréhension de ces phénomènes naturels.

2.1.2 Périmètre temporel

La base de données IBTrACS fournit des données sur les tempêtes depuis 1851. Cependant la période temporelle retenue dans le cadre de l'étude sera de **1982 à 2022**, ce qui correspond à une période de 40 ans d'observations historiques. Ce choix spécifique de période est motivé par deux facteurs importants.

Tout d'abord, il convient de souligner les **avancées technologiques significatives** survenues à partir des années 1970. Ces progrès technologiques ont considérablement amélioré la capacité à observer et à enregistrer les caractéristiques des tempêtes (EMANUEL, 2018). En effet, l'analyse du graphique 2.2¹ du nombre de tempêtes en fonction de l'année révèle une tendance marquée depuis les années 1970. Une augmentation significative du nombre de tempêtes au cours de cette période est observée, ce qui peut être attribué en partie à une meilleure détection et à une meilleure documentation des tempêtes grâce aux avancées technologiques. Les méthodes d'observation, telles que les satellites météorologiques et les instruments de mesure plus sophistiqués, ont permis d'obtenir des données plus précises et plus complètes sur les tempêtes. En se concentrant sur la période postérieure à ces avancées technologiques, l'étude intègre les données les plus fiables et les plus détaillées.

Le second facteur essentiel à prendre en compte dans l'analyse des tempêtes est l'impact du **réchauffement climatique**. Les effets du changement climatique ont été de plus en plus prononcés à partir des années 1980, et il est essentiel d'intégrer cette influence dans l'étude. Le graphique 2.3 représente l'évolution de la température moyenne annuelle mondiale entre 1850 et 2020 sous forme d'écart à la moyenne sur la période de référence 1961-1990. Jusqu'en 1940, l'écart est fortement négatif, suivi d'une tendance généralement négative jusqu'aux années 1980. Cependant, à partir du début des années 1980, une nette augmentation de la température se manifeste, avec un écart presque systématiquement positif. Cette évolution met en évidence un changement significatif du climat mondial, marqué par une transition vers des températures plus élevées à partir des années 1980. En limitant la période étudiée à partir de cette époque, une meilleure compréhension de l'évolution des tempêtes dans un contexte de changement climatique est rendue possible.

En se focalisant spécifiquement sur le bassin d'étude, l'évolution de la température moyenne annuelle du bassin Atlantique Nord entre 1850 et 2023 est représenté sur la figure 2.4 sous forme d'écart à

¹Source : Etabli à partir de la base IBTrACS

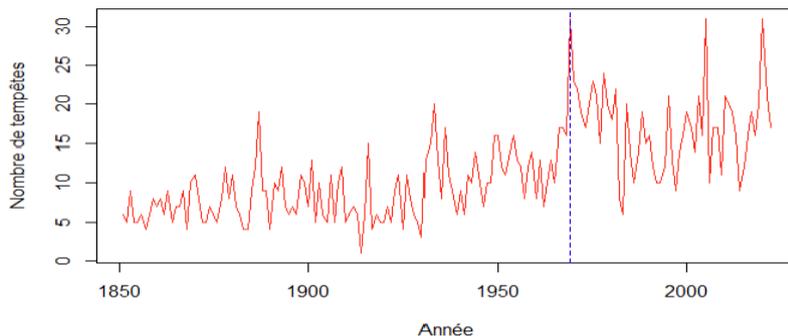


FIGURE 2.2 : Evolution du nombre de tempêtes entre 1851 et 2021

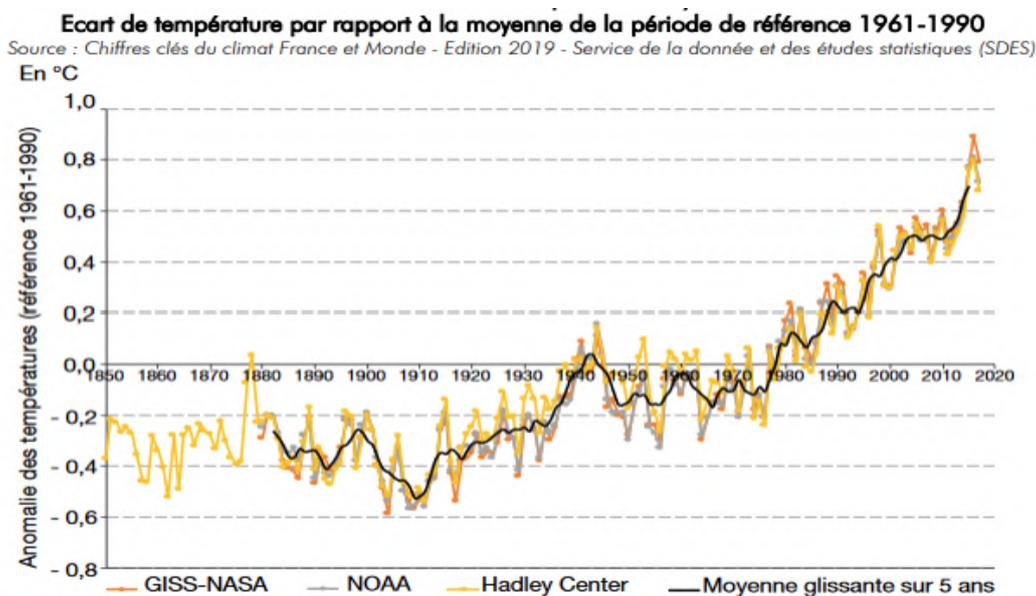


FIGURE 2.3 : Evolution des températures mondiales entre 1850 et 2020 (COMMISARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE, 2019)

la moyenne sur la période de référence 1971-2000. De même que sur le graphique précédent, il est observable une nette augmentation de la température, avec un écart presque systématiquement positif à partir des années 1990.

2.1.3 Périmètre spatial

Les tempêtes de la base IBTrACS concernant le bassin Atlantique Nord entre 1982 et 2022 sont toutes contenues dans une zone géographique délimitée de 6.2 à 70.7 degrés de latitude nord et de -107.7 à 13.5 degrés de longitude ouest à est. Toutefois, dans le cadre de cette étude, il est décidé de restreindre légèrement la zone de longitude, en la limitant de -107.7 à 0 degrés. Cette décision est prise car l'intérêt se porte principalement sur les régions terrestres du continent américain, tandis que les points de tempêtes s'étendant de 0 à 13.5 degrés de longitude est correspondent à l'arrivée des tempêtes tropicales en Europe. cette étude ne s'intéressant pas à l'impact des tempêtes sur le continent européen, il est choisi de restreindre légèrement le périmètre spatial afin d'avoir une zone

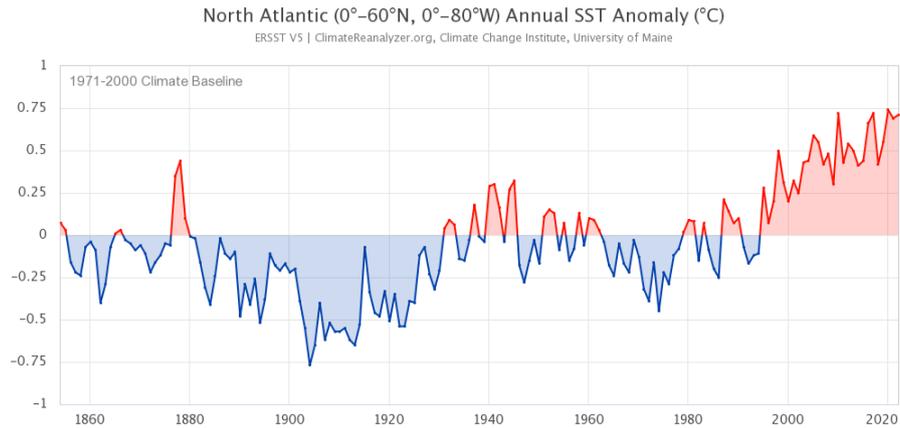
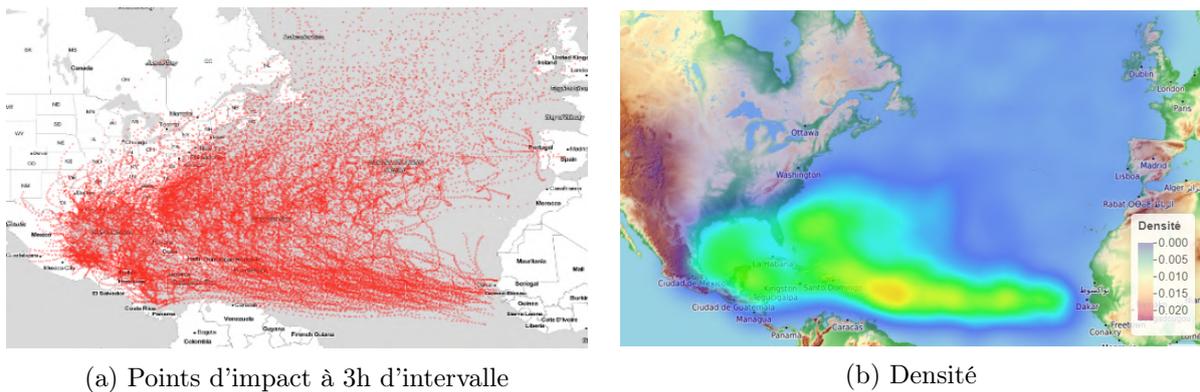


FIGURE 2.4 : Evolution des températures du bassin Atlantique Nord entre 1850 et 2023 (CLIMATE CHANGE INSTITUTE et UNIVERSITY OF MAINE, 2023)

d'étude plus ciblée. Le périmètre spatial retenu est donc $[-110^\circ, 0^\circ] \times [6^\circ, 71^\circ]$. Toutefois dans le cadre de certaines modélisations exposées par la suite, le périmètre spatial peut légèrement changer en raisons d'un manque d'observations dans certaines zones du bassin Atlantique Nord, notamment au nord du bassin. Ces modifications sont précisées à chaque fois qu'elles sont réalisées.



(a) Points d'impact à 3h d'intervalle

(b) Densité

FIGURE 2.5 : Visualisation des tempêtes du bassin Atlantique Nord entre 1982 et 2022 (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b)

La figure 2.5 présente les données extraites de la base IBTrACS, qui correspondent aux périmètres spatial et temporel spécifiés dans le cadre de l'étude.

2.1.4 Méthodologie

Le choix de modélisation, ainsi que la délimitation précise du périmètre spatial et temporel, étant établis, les différentes étapes de la modélisation peuvent être abordées. L'objectif final de cette démarche est de simuler un grand nombre de saisons cycloniques en se basant sur les données historiques existantes.

La simulation de multiples saisons cycloniques permettra d'obtenir une vision plus complète et réaliste des phénomènes cycloniques qui n'ont probablement encore jamais été observées mais qui restent tout à fait possible. Cela permettra de mieux comprendre les caractéristiques des ouragans, tels que leur fréquence, leur intensité et leur trajectoire, et d'estimer les risques associés. L'objectif est donc de simuler 10 000 saisons cycloniques en procédant selon les étapes suivantes, également synthétisées

sur la figure 2.6.

Etape 1 : Modéliser la fréquence annuelle de survenance d'une tempête dans le bassin Atlantique Nord. Cela permettra ainsi de simuler un nombre annuel de tempêtes pour chacune des 10 000 saisons cycloniques.

Etape 2 : Simuler une trajectoire pour chaque tempête d'une saison cyclonique obtenu à l'étape 1. Cette modélisation se fait en 2 phases.

Phase 1 : Simulation d'un point de genèse d'une tempête, c'est-à-dire sa position en longitude et en latitude ainsi que son mois de survenance.

Phase 2 : Simulation de la trajectoire de la tempête, c'est-à-dire sa position en longitude et en latitude jusqu'à sa dissipation.

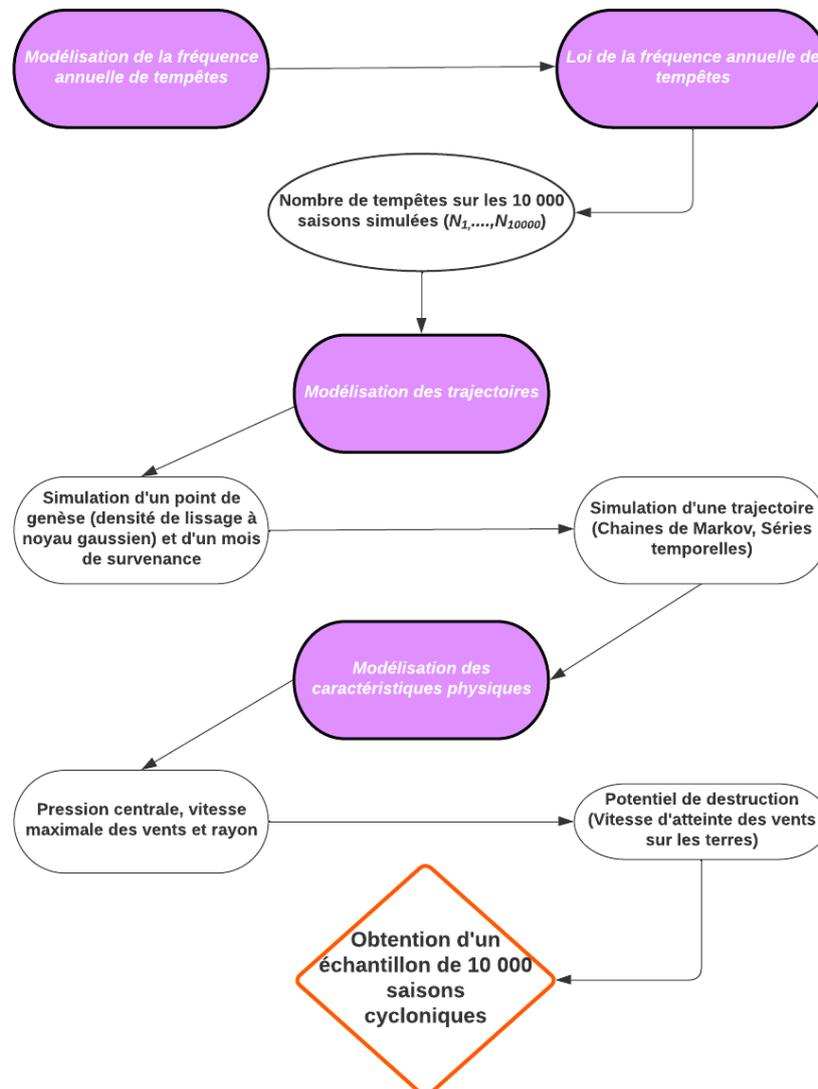


FIGURE 2.6 : Méthodologie de la modélisation de 10 000 saisons cycloniques

Étape 3 : Modélisation des caractéristiques physiques (pression centrale, vitesse maximale des vents, etc...) de chaque tempête afin de déterminer son intensité lors de son passage sur des étendues terrestres et d'évaluer son potentiel de destruction. Le potentiel de destruction d'une tempête qui se transforme en ouragan est en grande partie associé à sa vitesse des vents maximum comme observé sur l'échelle de Saffir-Simpson dans le tableau 1.1. Ainsi l'objectif final de cette étape sera de déterminer une vitesse d'atteinte des territoires pour chaque tempête simulée.

2.2 Modélisation de la fréquence annuelle

La modélisation de la fréquence annuelle de survenance des tempêtes dans le bassin Atlantique Nord est une étape cruciale pour comprendre leur comportement et évaluer leur impact sur le territoire américain. La fréquence annuelle des ouragans est influencée par de nombreux facteurs, tels que les conditions océaniques, atmosphériques et climatiques. Pour mieux appréhender cette relation complexe, des études scientifiques ont été menées pour explorer les modèles statistiques appropriés.

2.2.1 État de l'art des modèles statistiques

Au sein de la littérature scientifique, différentes approches ont été explorées pour modéliser la fréquence annuelle des tempêtes cycloniques. Une première approche consiste à expliquer cette fréquence en fonction de variables exogènes. Par exemple, SANKAR et al. (2016) ont utilisé trois types de réseaux de neurones artificiels pour prédire l'activité d'une saison cyclonique dans le bassin Nord Indien :

- Perceptron multicouche (*multilayer perceptron (MLP)*);
- Réseaux de neurones à base radiale (*radial basis function (RBF)*);
- Réseaux de neurones à régression généralisée (*Generalized regression neural network (GRNN)*).

Ces modèles ont été entraînés sur les données cycloniques de 1971 à 2002 et testés sur la période de 2003 à 2013 du bassin Nord Indien. Les résultats ont montré que cinq variables, telles que la hauteur géopotentielle² à 500 hPa, l'humidité relative à 500 hPa, la pression au niveau de la mer, le vent zonal³ à 700 hPa et 200 hPa, étaient retenues par les trois modèles. Les performances des modèles ont été évaluées à l'aide de métriques telles que le RMSE (*Root Mean Square Error*), l'écart type et le coefficient de corrélation, et le réseau de neurones à perceptron multicouche s'est avéré plus robuste pour prédire l'activité d'une saison cyclonique grâce à une validation croisée de type *Leave-One-Out Cross-Validation (LOOCV)*

Une deuxième approche couramment utilisée consiste à ajuster une loi de probabilité à la fréquence annuelle des tempêtes observées historiquement dans un bassin. Par exemple, KEES et al. (2021) ou BLOEMENDAAL et al. (2020) ont ajusté une loi de Poisson sur la fréquence annuelle de tempêtes dans le bassin Atlantique Nord. VICKERY et al. (2000) ont quant à eux utilisé une loi binomiale négative.

Des travaux similaires ont été réalisés dans des mémoires d'actuariat, DENISE-BAILLON (2013) et YOMBI (2016), où la fréquence annuelle des tempêtes et des ouragans dans le bassin Atlantique Nord a été modélisée en utilisant respectivement une loi binomiale négative et une loi de Poisson.

2.2.2 Choix de l'approche et résultats

Dans le cadre de cette étude, la **modélisation de la fréquence annuelle de tempête par ajustement à une loi de probabilité** a été retenue. Ce choix repose sur plusieurs justifications. Tout d'abord, en ajustant une loi de probabilité à ces données, il est possible d'obtenir une représentation

²désigne l'altitude à laquelle on atteint un potentiel égal de gravité

³vent dominant à une latitude correspondant au climat zonal

statistique robuste de la fréquence annuelle des tempêtes, ce qui permet de mieux comprendre et anticiper leur occurrence. De plus, en utilisant cette méthodologie, l'attention est directement portée sur la fréquence des tempêtes cycloniques en elle-même, évitant ainsi la nécessité de spécifier et de collecter un ensemble complet de variables exogènes. En effet la première approche proposée par SANKAR et al. (2016) nécessite de rassembler un nombre important de variables physiques et climatiques afin d'avoir une vision globale de leurs effets sur la fréquence cyclonique. Ainsi le choix de la deuxième approche simplifie la modélisation et évite les problèmes liés à la disponibilité et à la qualité des données exogènes.

Le tableau 1.5 révèle que la base de données IBTrACS recense un total de 649 tempêtes dans le bassin Atlantique Nord pour la période de 1982 à 2022. Cela correspond à une moyenne annuelle de 15.8 tempêtes, avec un écart-type de 5.3. La variation du nombre de tempêtes par année et la fréquence en fonction du nombre de tempêtes sont représentées sur la figure 2.7. Une observation remarquable est la présence de deux années exceptionnelles, 2004 et 2020, au cours desquelles un nombre record de 31 tempêtes a été enregistré. Bien que certaines indications suggèrent une tendance à l'augmentation du nombre de tempêtes à l'avenir, il convient de noter que les experts de l'OMM, comme mentionné dans le chapitre précédent tiré du rapport du dernier rapport du GIEC, ne partagent pas cet avis. Selon leurs prévisions, le nombre de tempêtes devrait rester relativement constant, mais il est prévu que la proportion de tempêtes atteignant un stade d'ouragan majeur augmente durant les prochaines décennies.

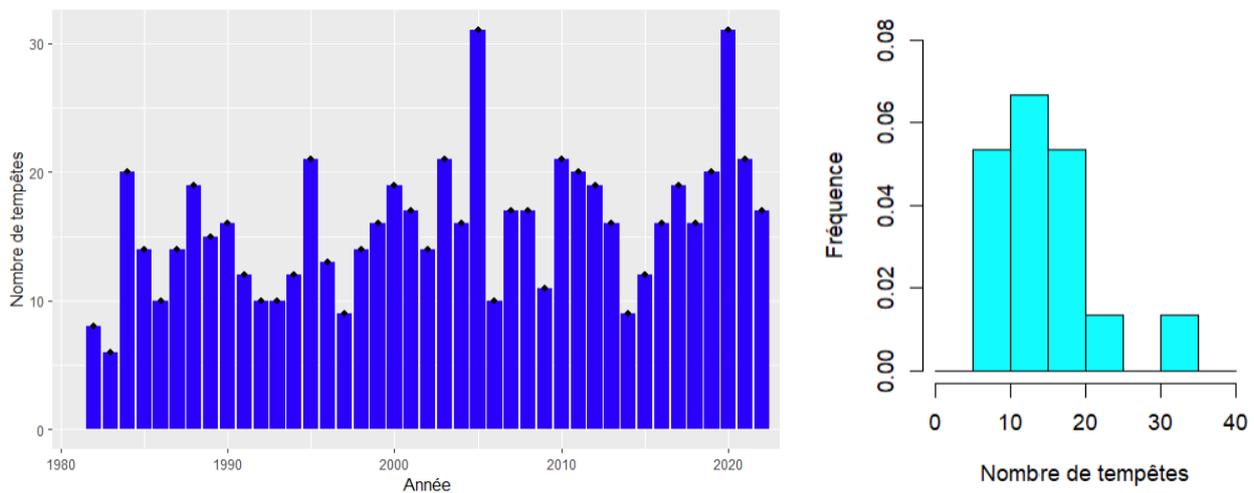


FIGURE 2.7 : Nombre de tempêtes par année et fréquence associée

Les variations du nombre de tempêtes en fonction de l'année et l'histogramme associé laisse suggérer une modélisation de la fréquence par une loi de Poisson ou par une loi Binomiale Négative en raison de sa caractéristique de sur-dispersion. La sur-dispersion signifie que la variance des données est plus grande que ce qui serait attendu sous une distribution de Poisson, ce qui indique que les données présentent une plus grande variabilité que ce que la loi de Poisson pourrait expliquer.

Soit N la variable aléatoire représentant la fréquence annuelle de tempêtes. En posant $n = 40$ le nombre d'années historiques (1982-2022) et (n_1, \dots, n_{40}) l'échantillon empirique à disposition avec $\sum_{i=1}^{40} n_i = 649$:

– Si $N \sim \mathcal{P}(\lambda)$ alors

$$\forall k \in \mathbb{N}, \mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \lambda > 0$$

La loi de Poisson est caractérisée par une équi-dispersion ($\mathbb{E}[N] = \mathbb{V}[N] = \lambda$), ce qui ne semble pas adaptée à l'échantillon empirique.

L'estimateur du maximum de vraisemblance de la loi de Poisson est sa moyenne empirique,

$$\hat{\lambda}_{EMV} = \bar{N} = \frac{1}{n} \sum_{i=1}^n n_i$$

– Si $N \sim \mathcal{BN}(m, p)$ alors

$$\forall k \in \mathbb{N}, \mathbb{P}(N = k) = \frac{\Gamma(m+k)}{k! \Gamma(m)} p^m (1-p)^k, m > 0 \text{ et } 0 < p < 1$$

La loi Binomiale Négative est caractérisée par une sur-dispersion ($\mathbb{E}[N] = \frac{m(1-p)}{p} < \mathbb{V}[N] = \frac{m(1-p)}{p^2}$), ce qui semble plus adaptée à l'échantillon empirique.

Il n'existe pas de formules fermées pour les estimateurs des paramètres de la loi Binomiale Négative ($\hat{m}_{EMV}, \hat{p}_{EMV}$) par la méthode du maximum de vraisemblance. Par la suite, ces estimateurs sont déterminés numériquement à l'aide du package *fitdistrplus*, (DELIGNETTE-MULLER et DUTANG, 2015) disponible sur le logiciel R, (R CORE TEAM, 2023).

La méthode d'estimation retenue afin de choisir entre ces deux lois est le maximum de vraisemblance et cela pour plusieurs raisons. Tout d'abord pour son efficacité, cette méthode permet d'obtenir des estimateurs asymptotiquement sans biais et sous certaines conditions, qui atteignent la limite de la variance minimale (borne de Cramér-Rao). Ensuite pour sa consistance, ce qui signifie que lorsque la taille de l'échantillon augmente, les estimateurs convergent vers les vraies valeurs des paramètres du modèle. Les résultats sont présentés dans le tableau 2.1 ci-dessous :

Distribution	Log-vraisemblance	Paramètres estimés	AIC
$\mathcal{P}(\lambda)$	-128.6	$\hat{\lambda}_{EMV} = 15.8$	259.1
$\mathcal{BN}(m, p)$	-124.8	$\hat{m}_{EMV} = 15.8 ; \hat{p}_{EMV} = 0.58$	253.6

TABLE 2.1 : Résultats obtenus par maximum de vraisemblance

Conformément aux attentes, la loi binomiale négative est celle qui maximise la log-vraisemblance. De plus, le critère d'information Akaike (AIC) est également présenté. Ce critère permet de comparer les deux modèles et d'identifier celui qui offre le meilleur compromis entre ajustement aux données et parcimonie. Il tient compte de la qualité de l'ajustement du modèle aux données, représentée par la log-vraisemblance, tout en pénalisant les modèles plus complexes avec leur nombre de paramètres. Ainsi en choisissant le modèle avec le plus petit AIC, la simplicité et l'interprétabilité sont favorisées, tout en préservant la précision de l'ajustement. En notant k le nombre de paramètres du modèle et L sa log-vraisemblance, le critère AIC est donnée par la formule définie comme suit

$$AIC = 2k - 2 \log(L).$$

La minimisation de l'AIC confirme que la loi binomiale négative est plus appropriée pour décrire la fréquence annuelle de tempêtes dans le cadre de l'étude. Maintenant que le choix de la loi est fixée, la qualité d'adéquation de la loi simulée à nos données empiriques reste à vérifier. L'hypothèse nulle suivante est donc à tester

$$H_0 : N \sim \mathcal{BN}(\hat{m}_{EMV}, \hat{p}_{EMV}).$$

Un test classique et largement utilisé pour évaluer si une distribution discrète correspond bien à un ensemble de données observées est le test d'adéquation du khi-deux (CORDER et FOREMAN, 2014).

Test d'adéquation du khi-deux : Le principe du test du khi-deux repose sur la comparaison entre les fréquences observées et les fréquences attendues selon la distribution théorique. La statistique de test, notée D , est calculée en utilisant la formule suivante

$$D = \sum_{k=1}^J \frac{(O_k - E_k)^2}{E_k},$$

où O_k représente la fréquence observée dans la classe k et E_k représente la fréquence attendue selon la distribution théorique dans la même classe, $k \in \{1, \dots, J\}$ avec J égal au nombre de classes. La somme est effectuée sur toutes les classes de la variable étudiée. Sous l'hypothèse nulle, la statistique de test D suit approximativement une distribution du khi-deux avec $J - 1$ degrés de liberté. Un test asymptotique de niveau α est réalisable en comparant la valeur de D calculée à partir d'un échantillon et le quantile d'ordre $1 - \alpha$ d'une loi du khi-deux à $J - 1$ degrés de liberté noté $\chi_{1-\alpha}^2$. L'hypothèse d'adéquation est alors rejetée avec un seuil de risque α si $D > \chi_{1-\alpha}^2$.

Dans le cadre de l'étude, le nombre de classes différentes observé dans l'échantillon empirique est égal à 15, ce qui est beaucoup trop élevé sachant que de nombreuses classes comportent très peu d'observations (inférieures à 5). Des regroupements de classes par tranche de 5 allant de 5 à 31 (la dernière classe allant de 26 à 31) représentant le nombre de tempêtes observés sur une année, sont effectuées afin d'obtenir des classes avec un nombre important d'observations. Cela donne finalement un nombre de classes $J = 5$ pour un ensemble d'observations $n = 40$.

Les résultats du test du khi-deux sont présentés dans le tableau 2.2 ci-dessous. Suite aux résultats du test, la p-valeur obtenue est d'environ 80%, ce qui est largement supérieur au seuil de signification de 5%. Cela signifie que l'hypothèse selon laquelle la **fréquence annuelle de tempêtes suit une loi binomiale négative avec $\hat{m}_{EMV} = 15.8$ et $\hat{p}_{EMV} = 0,58$ est retenu.**

Test	Distribution	Statistique de Test	P-value
χ^2	$\mathcal{BN}(\hat{m}_{EMV}, \hat{p}_{EMV})$	1.52	0.82

TABLE 2.2 : Résultats du test d'adéquation du χ^2

Enfin, pour vérifier graphiquement l'adéquation de la distribution choisie avec les données empiriques, deux histogrammes sont présentés sur la figure 2.8. Le premier est l'histogramme théorique d'une loi binomiale négative avec les paramètres obtenues par la méthode du maximum de vraisemblance ($\mathcal{BN}(\hat{m}_{EMV}, \hat{p}_{EMV})$), tandis que le deuxième est l'histogramme empirique du nombre de tempêtes observées.

Finalement, le nombre de tempêtes annuelles sur 10 000 saisons cycloniques est simulable en générant 10 000 réalisations d'une variable aléatoire suivant la loi binomiale négative $\mathcal{BN}(\hat{m}_{EMV}, \hat{p}_{EMV})$. Le total des tempêtes simulées sur ces 10 000 saisons cycloniques s'élève à 157 689, ce qui correspond à une moyenne de 15,7 tempêtes par année. Cette moyenne est cohérente avec la moyenne annuelle de 15,8 tempêtes observée au cours de la période étudiée.

2.3 Modélisation des points de genèse

Dans cette section, la modélisation des points de genèse est effectuée pour chaque tempête simulée dans la section précédente. À ce point de genèse simulé, un mois de survenance est également attribué. Le mois de survenance aura un impact significatif sur l'intensité du phénomène cyclonique (3.2) car selon le mois, la pression atmosphérique et la température du bassin océanique seront différentes, deux caractéristiques importantes influençant l'intensité d'un ouragan.

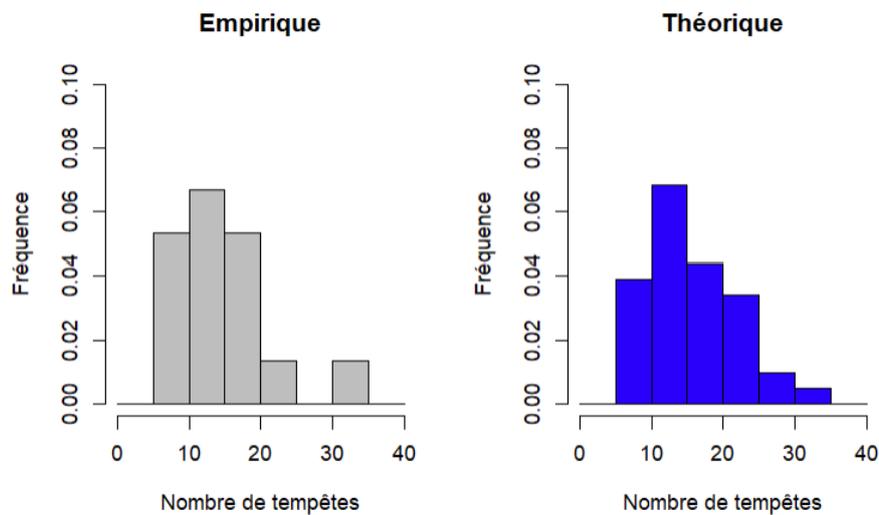


FIGURE 2.8 : Histogramme empirique et simulée de la fréquence annuelle de tempêtes

2.3.1 Zones de cyclogénèse

Le processus par lequel une circulation cyclonique prend naissance ou s'intensifie dans une région de l'atmosphère se nomme la **cyclogénèse** (ELSNER et KARA, 1999). Sur la période étudiée, les points de formation historique des tempêtes sont représentés sur la figure 2.9. L'observation de cette figure révèle la présence de deux principales zones de cyclogénèse parmi les 649 points de genèse historique des tempêtes. La première zone se situe près de la côte ouest africaine, tandis que la deuxième zone se trouve près du Golfe du Mexique.

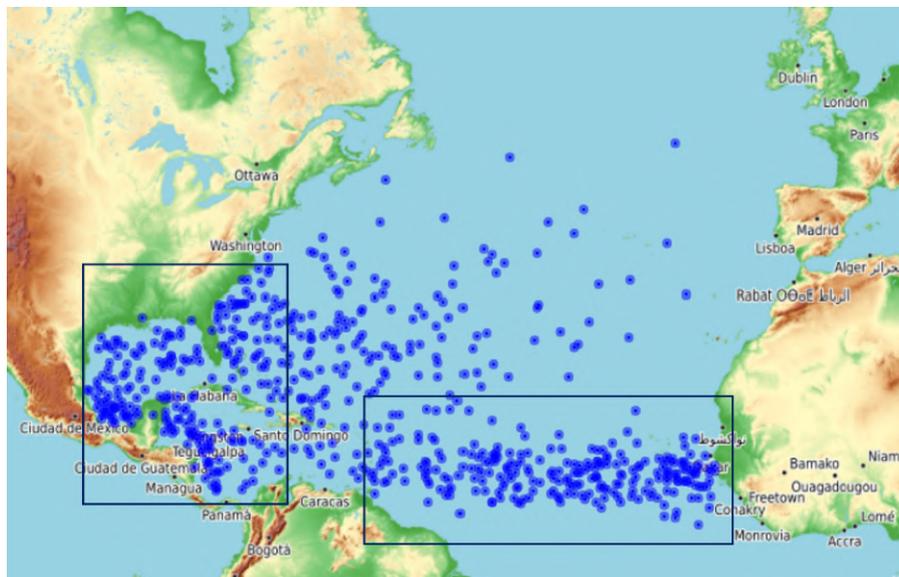


FIGURE 2.9 : Points de formation des tempêtes

2.3.2 État de l'art des modèles statistiques

Dans la littérature scientifique, tout comme pour la modélisation de la fréquence annuelle, plusieurs approches ont été utilisées afin de simuler des points de formations de tempêtes. Par la suite une liste

de quelques approches couramment utilisées dans les revues scientifiques est présentée.

Une première approche plutôt simpliste consiste tout simplement à effectuer un tirage aléatoire avec remise parmi l'ensemble des points initiaux historiques. Cette méthode a par exemple été utilisée chez VICKERY et al. (2000) ou encore DENISE-BAILLON (2013) dans son mémoire d'actuariat.

Une deuxième approche consiste à effectuer un lissage spatial des points de formation historiques. Cette méthode est une amélioration de la première approche car sous certaines conditions, il est possible de simuler une tempête en un point du globe qui ne fait pas partie des points initiaux historiques. Cette approche offre ainsi une plus grande flexibilité et une meilleure représentation de la variabilité spatiale du phénomène. Cette méthode a été utilisée par exemple chez HALL et JEWSON (2007) ou encore chez YOMBI (2016) dans son mémoire d'actuariat.

Une troisième approche consiste à effectuer des interpolations à partir des points de formation historiques et la maille de résolution. Par exemple, dans l'article BLOEMENDAAL et al. (2020), les points de formations historiques sont comptés selon un quadrillage en résolution $5^\circ \times 5^\circ$ des bassins océaniques, puis une interpolation cubique a été effectuée afin d'obtenir des valeurs sur un maillage plus fin ($1^\circ \times 1^\circ$). La valeur de ces mailles est ensuite utilisée comme pondération lors de l'échantillonnage des lieux de genèse. Enfin, l'emplacement d'un point de genèse simulé est échantillonné en sélectionnant un emplacement aléatoire (à $0,1^\circ$ près) à l'intérieur d'une des cases de la grille $1^\circ \times 1^\circ$.

2.3.3 Choix de l'approche et résultats

Dans le cadre de ce mémoire, la **modélisation des points de formation de tempêtes par un lissage spatial des points de genèse historiques** a été retenue. Ce choix est motivé par sa capacité à capturer la variabilité spatiale du phénomène par rapport aux 2 autres méthodes.

2.3.3.1 Lissage des points de genèse historiques

Le lissage spatial a été réalisé en trois étapes :

- Le choix du noyau (K) ;
- L'hyper-paramétrage de la fenêtre de lissage (L) ;
- L'estimation non-paramétrique de la densité spatiale des points initiaux.

Noyau : pour cette étude, des noyaux non uniformes sont plus adaptés car il est préférable d'obtenir une pondération qui décroît avec la distance au point d'estimation. De plus, d'après plusieurs études telles que RENAUD et al. (2018) ou FAUCHER et al. (2002), le choix du noyau n'a pas une influence significative sur la qualité du lissage. Pour leur modélisation de tempêtes, (HALL et JEWSON, 2007) utilise un noyau gaussien⁴ tandis que (RUMPF et al., 2007) utilise un noyau d'Epanechnikov⁵. Le noyau gaussien utilisé par (HALL et JEWSON, 2007) sera retenu par la suite.

Fenêtre de lissage : contrairement au noyau, le choix de la fenêtre de lissage a un impact significatif sur la qualité du lissage. Toujours d'après les mêmes études, la taille de la fenêtre de lissage conditionne fortement le phénomène de dispersion des points simulés en fonction des points historiques. L'hyper-paramétrage de la fenêtre de lissage est donc une étape cruciale afin d'avoir une densité de lissage spatiale cohérente et réaliste.

⁴Noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$.

⁵Noyau d'Epanechnikov : $K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{u \geq 1}$.

En notant $\mathbf{r} = (r_x, r_y)$ un point du globe repéré par sa longitude r_x et sa latitude r_y , $\mathbf{r}_i = (r_{i,x}, r_{i,y})$ un point de genèse historique, la fonction de densité à noyau gaussien en un point r (HALL et JEWSON, 2007) s'écrit

$$f(\mathbf{r}) = \frac{1}{2\pi NL^2} \sum_{i=1}^N \exp \frac{-d(\mathbf{r}_i, \mathbf{r})}{2L^2},$$

où N représente le nombre de points de genèse historiques ($N = 649$), L la fenêtre de lissage en degrés et d connue sous le nom de *great circle distance formula*, représente la distance orthodromique qui est la plus courte distance entre deux points sur une sphère, (ici la Terre). Soient deux points de la Terre repérés par leur longitude et leur latitude en radians, $\mathbf{r} = (r_x, r_y)$, $\mathbf{r}' = (r'_x, r'_y)$ et R_T le rayon de la Terre, la distance orthodromique d est donnée par la relation suivante (RAJPOOT, 2016)

$$d(\mathbf{r}, \mathbf{r}') = R_T \arccos(\cos r_x \cos r'_x \cos(r_y - r'_y) + \sin r_x \sin r'_x). \quad (2.1)$$

L'hyper-paramétrage de la fenêtre de lissage L a été réalisé en utilisant le critère de validation croisée par optimisation de la log-vraisemblance, noté *LCV* (*Likelihood Cross-Validation*). Cette méthode consiste à estimer la fonction de densité maximisant la log-vraisemblance en laissant un seul point de données à l'extérieur de l'échantillon et en utilisant les autres points de données pour le modèle. Ce processus est répété en laissant chaque point de données à l'extérieur une fois, ce qui donne une série d'estimations de la fonction de densité.

Différentes expressions du critère *LCV* sont disponibles dans la littérature pour choisir de manière optimale la fenêtre de lissage L (HABBEMA et al., 1974). Pour une fenêtre de lissage L , on note $\hat{f}_{-i,L}(r_i)$ la fonction de densité estimée en r_i sans l'observation i , \hat{f}_L la fonction de densité estimée globale, n le nombre d'observations et χ le domaine de lissage, fixé dans ce cas de 0.1 à 50. Une approche simpliste, proposée par HABBEMA et al. (1974), consiste à sélectionner la valeur de L qui maximise le produit de toutes $\hat{f}_{-i,L}(r_i)$, soit

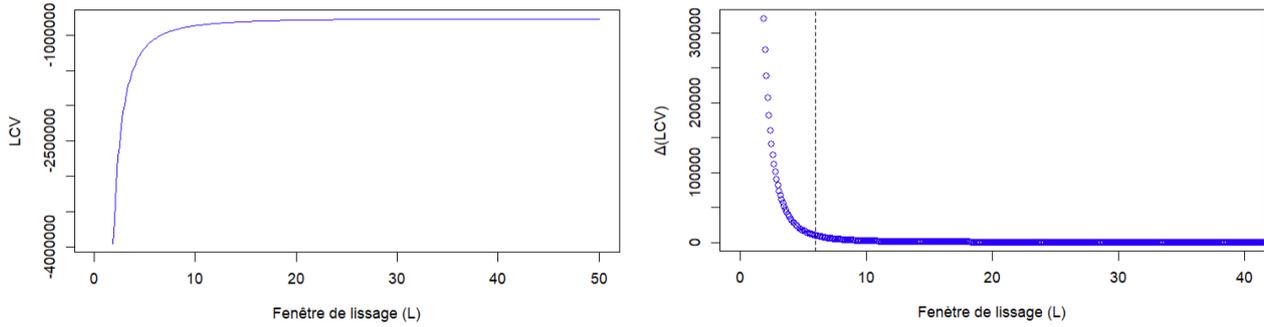
$$LCV_1 = \prod_{i=1}^n \hat{f}_{-i,L}(r_i).$$

Un LCV_1 élevé indique une proximité accrue entre la densité estimée et les observations réelles. Cependant, une version améliorée de ce critère a été proposée par les auteurs, en introduisant un terme de pénalisation à la somme des logarithmes de toutes les $\hat{f}_{-i,L}(r_i)$, soit

$$LCV_2 = \sum_{i=1}^n \log \hat{f}_{-i,L}(r_i) - n \left(\int_{\chi} \hat{f}_L(\mathbf{u}) d\mathbf{u} - 1 \right).$$

Ce terme de pénalisation tient compte du nombre d'observations et correspond à la pénalisation du lissage excessif. L'idée est que si le lissage est trop important, la densité estimée sera trop uniforme et donnera une mauvaise approximation des données. Afin d'obtenir un lissage spatial robuste et cohérent avec nos données historiques, le critère LCV_2 est retenu. Ce critère prend en compte à la fois la proximité avec les observations et la pénalisation du lissage excessif, fournissant ainsi une estimation plus fiable et adaptée de la densité spatiale. Les résultats de l'hyper-paramétrisation de L sont présentés sur les graphiques 2.10.

Le premier graphique présente l'évolution du critère *LCV* en fonction de la fenêtre de lissage. Une croissance très forte est observée pour les valeurs de L comprises entre 0 et 10, suivie d'une augmentation plus modérée qui finit par se stabiliser pour les valeurs élevées de L . Il s'avère donc complexe de sélectionner la valeur optimale de \hat{L} uniquement à partir de cette représentation graphique. Afin de faciliter cette tâche, la construction d'un second graphique présentant l'évolution de la différence entre les valeurs de *LCV* en fonction de L est effectué. Ce dernier graphique permet d'identifier plus

FIGURE 2.10 : Choix de la fenêtre optimale \hat{L}

précisément à partir de quelle valeur de L le critère LCV cesse d'augmenter de manière significative. L'analyse révèle que pour $L = 6$, le critère présente une augmentation marginale. Par conséquent, après avoir comparé plusieurs résultats graphiques pour des valeurs de L comprises entre 5 et 10, une valeur optimale de \hat{L} égale à 6 est retenue. Cette sélection a été effectuée en prenant en compte la stagnation observée dans la croissance du critère LCV pour des valeurs de L supérieures à 6, ce qui indique une diminution de la sensibilité de la méthode de lissage spatial pour des fenêtres plus larges.

Une représentation graphique illustrant le lissage spatial des points de genèse historiques, réalisé à l'aide de la valeur optimale \hat{L} , est présentée dans la figure 2.11. Le premier graphique, sans superposition des points de genèse historiques, met en évidence l'effet du lissage qui a pris en compte de manière appropriée les deux principales zones de formation historique, à savoir le Golfe du Mexique et la côte ouest africaine, comme observé précédemment dans la figure 2.9. Cette représentation graphique témoigne de la prise en compte de la variabilité spatiale du phénomène de formation de tempêtes. De manière complémentaire, la figure suivante présente la superposition des points de genèse historiques, renforçant ainsi l'observation précédente. Cette représentation confirme la capacité du lissage spatial à capturer l'hétérogénéité spatiale inhérente au processus de formation des tempêtes. Ainsi, cette analyse graphique valide l'approche du lissage spatial et sa capacité à rendre compte des variations spatiales complexes dans le processus de formation des tempêtes.

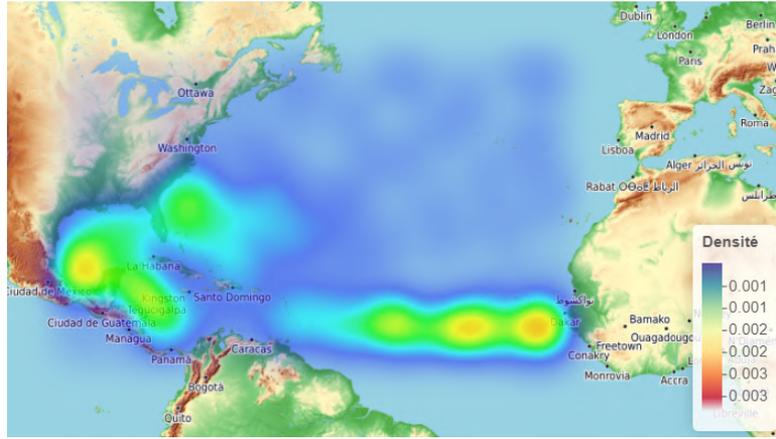
2.3.3.2 Simulation des points de genèse

La densité spatiale des points de genèse estimée, notée $\hat{f}_{\hat{L}}$, a permis de déterminer les points de genèse pour la simulation des 10 000 années cycloniques. La méthode employée pour cette simulation est celle de l'inversion de la fonction de répartition. Ce choix de méthode s'est avéré adéquat compte tenu de la nature discrète de la distribution $\hat{f}_{\hat{L}}$, caractérisée par plusieurs localisations r_i très proches les unes des autres.

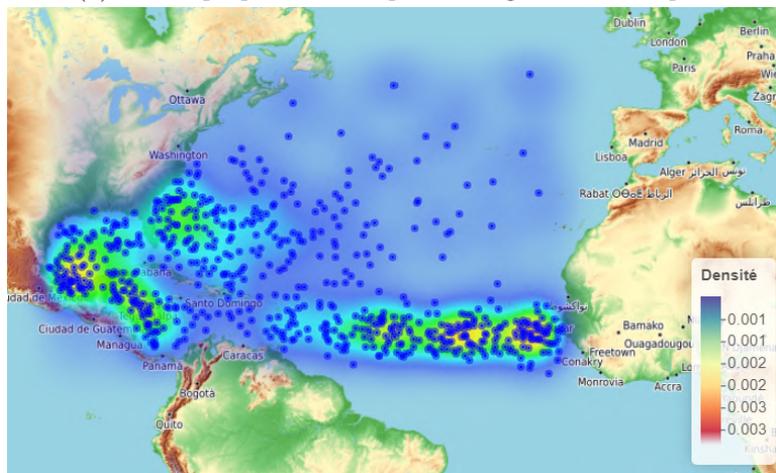
La méthode d'inversion de la fonction de répartition offre l'avantage de permettre la génération de valeurs aléatoires qui suivent la distribution $\hat{f}_{\hat{L}}$. Cette approche a été largement utilisée dans la littérature scientifique pour simuler des phénomènes discontinus ou à localisations multiples. C'est le cas de HALL et JEWSON (2007) et également YOMBI (2016) dans son mémoire d'actuariat. En utilisant cette méthode, des points initiaux de manière cohérente sont générés, en respectant la distribution spatiale représentée par $\hat{f}_{\hat{L}}$.

Normalisation de la fonction de densité

Le lissage spatial précédemment effectué a permis d'obtenir une représentation de l'intensité des points de genèse des tempêtes, c'est-à-dire le nombre de points par unité de surface sur une grille couvrant l'ensemble des points de genèses historiques. Cette grille est construite en utilisant une résolution de 1° et va donc de -97 à -16 degrés en longitude et de 7 à 47 degrés en latitude. Afin d'obtenir une mesure de densité plus appropriée, une normalisation de cette intensité est réalisée, en la fixant à 1.



(a) sans superposition des points de genèse historiques



(b) avec superposition des points de genèse historiques

FIGURE 2.11 : Lissage spatial des points de genèse historiques ($\hat{L} = 6$)

Soit R la variable aléatoire représentant une position r du globe et n le nombre de points obtenus après le lissage spatial. On a $R \sim \hat{f}_L$, la fonction de densité de lissage normalisée est définie de la manière suivante

$$\mathbb{P}[R = r_i] = \frac{\hat{f}_L^*(r_i)}{\sum_n \hat{f}_L^*(r_i)}, \quad \forall i \in 1, \dots, n.$$

Algorithme de simulation

En notant $p_i = \mathbb{P}[R = r_{(i)}]$ la probabilité de sélectionner la position avec la i -ème plus petite intensité $r_{(i)}$, $i \in \{1, \dots, n\}$. Soit $U \sim \mathcal{U}[0, 1]$, on peut établir la relation suivante

$$R = r_{(1)}\mathbb{1}_{U < p_1} + r_{(2)}\mathbb{1}_{p_1 \leq U < p_1 + p_2} + \dots + r_{(n)}\mathbb{1}_{p_1 + \dots + p_{n-1} \leq U < 1}. \quad (2.2)$$

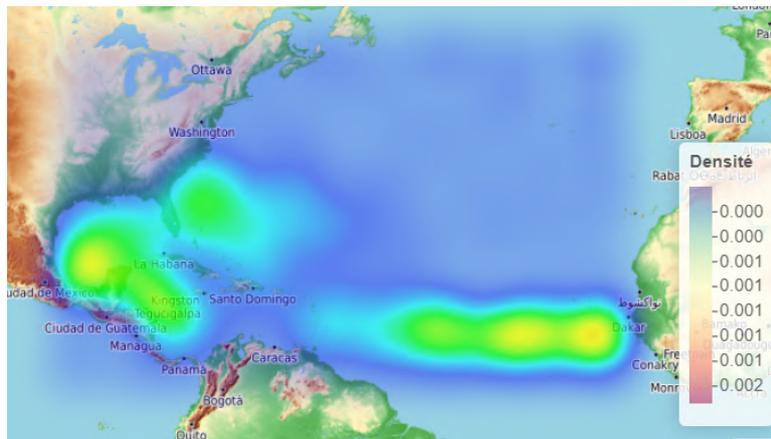
La fonction de répartition bivariee est alors définie comme suit

$$F(x, u) = \mathbb{P}[R \leq x, U \leq u],$$

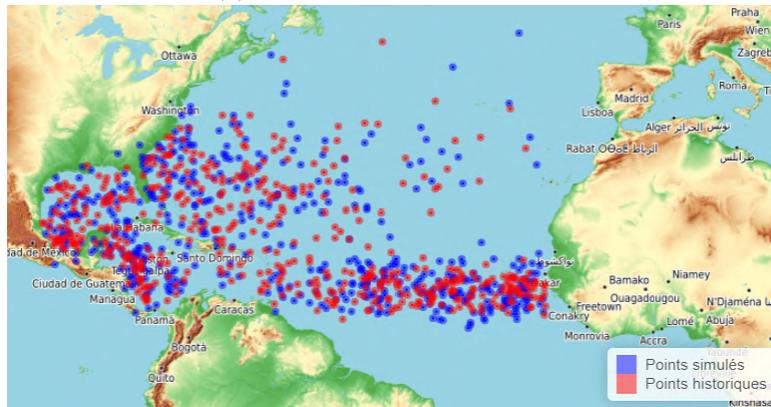
et peut être calculée de la manière suivante

$$F(x, u) = \begin{cases} 0 & \text{si } u < p_1 \\ r_{(1)} & \text{si } p_1 \leq u < p_1 + p_2 \\ r_{(2)} & \text{si } p_1 + p_2 \leq u < p_1 + p_2 + p_3 \\ \dots & \\ r_{(n)} & \text{si } p_1 + p_2 + \dots + p_{n-1} \leq u \leq 1. \end{cases}$$

La simulation de 157 689 points de genèses de tempêtes, correspondant au nombre total de tempêtes sur les 10 000 saisons cycloniques simulées, implique donc la génération de 157 689 réalisations d'une variable aléatoire uniforme sur l'intervalle $[0,1]$, et à partir de là, déduire les valeurs de R en utilisant la relation (2.2).



(a) Densité des points simulés



(b) 500 points de genèse simulés et historiques

FIGURE 2.12 : Résultats de la simulation des points de genèse de tempêtes

Les résultats de la simulation sont présentés sur la figure 2.12. L'analyse du premier graphique, qui représente la densité des points simulés, révèle que la simulation a préservé la répartition spatiale des points initiaux observés empiriquement, comme illustré sur la figure 2.9. En effet, une intensité plus élevée persiste près des côtes ouest-africaines, et l'existence d'une concentration spatiale significative dans le Golfe du Mexique et à l'est des côtes américaines est toujours observable. Il convient de noter que le nombre de points simulés dépasse largement celui des points initiaux empiriques, ce qui se traduit par une échelle de mesure de l'intensité beaucoup plus étendue que celle présentée sur la figure 2.11.

Le deuxième graphique en figure 2.12, qui représente un échantillon de 500 points de genèse simulés parmi les 157 689 points et 500 points de genèse historique, a été créé dans le but de comparer visuellement la répartition spatiale des deux ensembles de données. Cette comparaison permet de vérifier si la simulation reproduit de manière cohérente les schémas observés dans les données historiques. En d'autres termes, en superposant les points simulés et les points historiques, une évaluation de la capacité à produire une répartition spatiale des points de genèse de tempêtes cycloniques observées dans le passé de manière cohérente est réalisable. L'observation importante est que les points de simulation et les points historiques se concentrent principalement dans les mêmes zones. Les deux ensembles de points présentent une concordance spatiale significative renforçant ainsi la validité de l'approche de simulation.

En conclusion, pour chaque point de genèse simulé, un mois de survenance est assigné en utilisant une distribution historique basée sur les occurrences mensuelles passées, comme illustré sur la figure 1.10. Cette approche, déjà adoptée par BLOEMENDAAL et al. (2020) et DENISE-BAILLON (2013) dans leurs travaux de modélisation du phénomène cyclonique, permet de considérer l'influence significative du mois de survenance sur l'intensité du phénomène cyclonique.

Chapitre 3

Simulation de la dynamique des ouragans

Le chapitre précédent a établi les bases de cette étude en se concentrant sur la modélisation de la survenance des ouragans. Ce troisième chapitre plonge au cœur de la complexité de ces phénomènes naturels en explorant leur dynamique, notamment leur trajectoire et leur intensité. Cette exploration repose sur une analyse approfondie de leurs caractéristiques fondamentales, telles que la pression centrale, la vitesse maximale des vents, le rayon de la vitesse maximale des vents et le gradient de vitesse. L'objectif ultime de cette démarche est la simulation précise des saisons cycloniques du bassin Atlantique Nord, tout en préservant les caractéristiques intrinsèques de ces phénomènes. Cette approche détaillée revêt une importance capitale pour une meilleure compréhension du comportement des ouragans, ainsi que pour l'analyse précise des prévisions et des estimations financières.

Le chapitre se divise en deux sections principales. La première section se penche sur la trajectoire des ouragans, tandis que la seconde se concentre sur leur intensité. Chacune de ces sections explore diverses approches pour modéliser la dynamique des ouragans. En complément des résultats et des hypothèses sous-jacents à cette modélisation, un état de l'art des méthodes statistiques présentes dans la littérature scientifique est présenté, afin de situer l'approche adoptée dans le cadre de cette étude.

3.1 Modélisation des trajectoires

Les points de genèse initiaux de chaque tempête simulée étant établis, l'étape suivante consiste à simuler leur trajectoire. La figure 3.1 présente une visualisation graphique des 649 trajectoires de tempêtes enregistrées dans la base de données IBTrACS pour la période 1982-2022 dans le bassin Atlantique Nord.

Cette représentation permet de visualiser l'ensemble des schémas généraux de déplacement des tempêtes cycloniques dans la région. Une observation attentive de ces trajectoires révèle une tendance dominante de déplacement d'est en ouest, en direction des côtes du continent américain, avant de se diriger vers l'océan Atlantique en direction du pôle Nord ou de l'Europe.

Il est intéressant de noter que certaines trajectoires de tempêtes se dissipent progressivement au sein du continent américain, perdant leur intensité en raison des interactions avec les terres. Cependant, d'autres tempêtes, souvent plus violentes, parviennent à maintenir leur circulation et à regagner l'océan Atlantique après avoir touché les côtes du continent américain. Ces tempêtes persistent parfois et peuvent même se développer davantage en se déplaçant vers le nord en direction du pôle Nord ou vers l'Europe.

Ces observations mettent en évidence l'importance des conditions atmosphériques et océaniques prédominantes dans la région du bassin Atlantique Nord, qui favorisent ce schéma de déplacement caractéristique. En analysant ces schémas, il est possible d'identifier les voies de déplacement privilégiées des tempêtes ainsi que les zones de dissipation courantes, fournissant ainsi des informations précieuses pour la compréhension et la prévision des tempêtes cycloniques dans cette région.

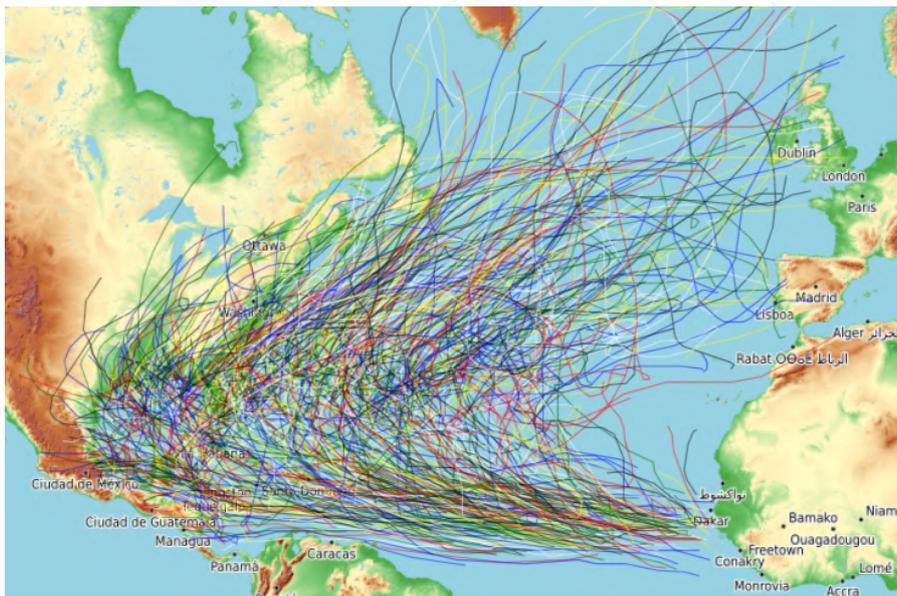


FIGURE 3.1 : Trajectoires des tempêtes historiques (1982-2022)

3.1.1 État de l'art des modèles statistiques

De nombreuses approches statistiques ont été développées dans la littérature pour la modélisation des trajectoires de tempêtes dans le bassin Atlantique Nord. Ces modèles ont été largement étudiés et repris par d'autres chercheurs dans des travaux plus récents. Quelques-uns de ces modèles éminents sont évoqués ci-dessous.

Le modèle CLIPER (Climatologie et Persistence), initialement instauré par NEUMANN (1972) dans les années 70, figure parmi les premiers modèles largement connus et basiques. Il repose sur une approche de régression multiple prenant en compte des paramètres tels que le mouvement actuel et passé (sur les 12 et 24 heures précédentes), le mouvement précédent, la date, la latitude, la longitude et l'intensité initiale. L'hypothèse sous-jacente au modèle CLIPER est que les tempêtes cycloniques ont tendance à suivre des trajectoires climatologiques moyennes ou à persister dans leur trajectoire actuelle sur de courtes périodes (NEUMANN, 1972). Le modèle CLIPER a été initialement conçu pour la prévision en temps réel de trajectoires particulières. Toutefois, dans le cadre de la modélisation visant à synthétiser un grand nombre de trajectoires diverses et aléatoires, cette approche n'est pas retenue.

Des travaux menés par VICKERY et al. (2000) ont exploité des modèles de régression linéaire multiple, avec une résolution spatiale de $5^\circ \times 5^\circ$, pour prédire la vitesse de translation et la direction angulaire des tempêtes. Les variables explicatives comprennent la latitude, la longitude, la vitesse de translation et la position angulaire au temps précédent. D'autres chercheurs, tels que RUMPF et al. (2007) et KEES et al. (2021) ont également adopté cette méthode pour leurs simulations de trajectoires.

Une approche différente, proposée par JAMES et MASON (2005) repose sur l'utilisation de processus auto-régressifs $AR(1)$ appliqués aux séries temporelles de latitude et de longitude. Leurs recherches ont révélé que les variations longitudinales et latitudinales suivent une tendance linéaire, justifiant ainsi cette approche. Le modèle de trajectoire développé par BLOEMENDAAL et al. (2020) est basé sur cette méthode.

(RISI, 2005) utilise des chaînes de Markov pour estimer les probabilités de transition des variations de la vitesse de translation et de la direction angulaire, en fonction de la position actuelle, de la vitesse de translation et de la direction angulaire au temps précédent. Par ailleurs, DENISE-BAILLON (2013) et YOMBI (2016) ont également adopté une approche basée sur les Chaînes de Markov, mais en calculant les probabilités de transition d'une position à une autre, en se basant sur les changements de positions

historiques.

Ces différents modèles statistiques énumérés précédemment représentent seulement un échantillon restreint des approches existantes pour la modélisation des trajectoires de tempêtes dans le bassin Atlantique Nord. Néanmoins, il convient de noter que le domaine de la prévision des trajectoires cycloniques est en constante évolution, et de nombreuses autres méthodes statistiques ont été développées et explorées par les chercheurs. Parmi ces approches supplémentaires, on retrouve les méthodes basées sur les processus stochastiques, les modèles bayésiens, les modèles de régression non linéaire, les modèles à base de données géospatiales¹, et bien d'autres encore. Chaque approche présente ses propres avantages et limitations. La recherche dans ce domaine est donc continuellement animée par la quête d'amélioration des performances de ces modèles afin de mieux anticiper les déplacements des tempêtes (CHEN et al., 2020).

3.1.2 Approche par chaînes de Markov

Une synthèse des principes fondamentaux des Chaînes de Markov utile pour l'étude est présentée dans l'annexe A.2 (RUCH et CHABANOL, 2012). Dans cette section l'application de ces concepts dans le contexte spécifique de cette étude est abordée, ainsi que les résultats obtenus et les éventuelles limites identifiées.

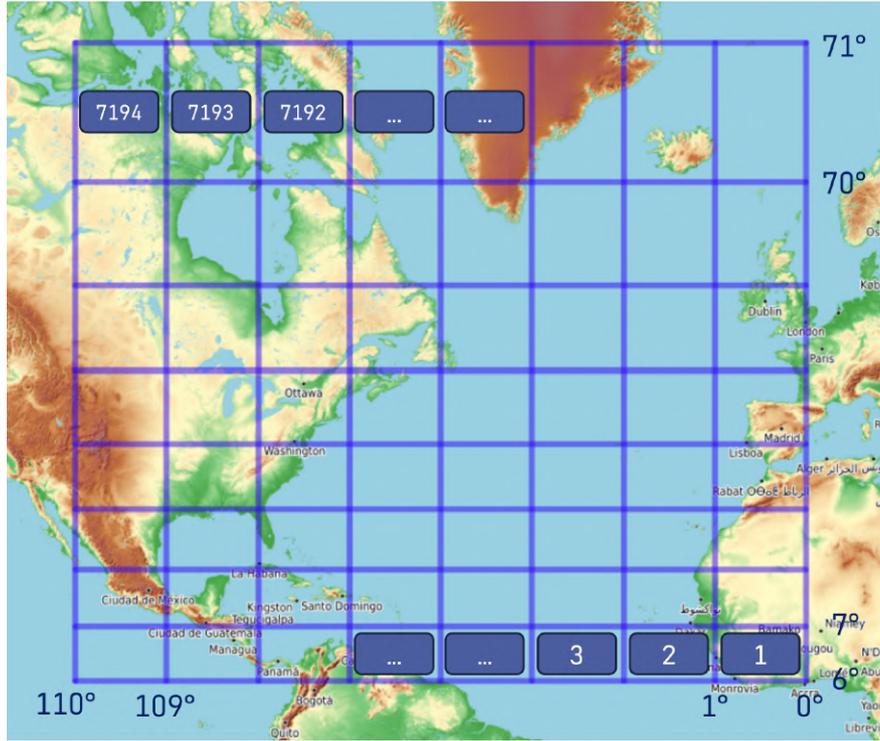
3.1.2.1 Application

La modélisation repose sur le fait que **la trajectoire d'une tempête est une chaîne de Markov homogène et discrète dans le temps**, notée $(T_n)_{n \geq 0}$. Cette hypothèse est basée sur l'idée que **la trajectoire future d'une tempête tropicale n'est pas affectée par son historique, mais dépend uniquement de sa position actuelle**. Cette approche est similaire à celle utilisée par DENISE-BAILLON (2013) et YOMBI (2016) dans leur mémoire d'actuariat.

Le périmètre spatial défini en 2.1.3 est discrétisé en carrés d'aire $l_x \times l_y$, où l_x représente la résolution en latitude (nord) et l_y en longitude (ouest). Dans cette étude, une résolution de 1° en latitude et longitude est choisie, ce qui permet d'obtenir un quadrillage assez fin avec 7194 cases (109×66), tout en restant proche de la résolution utilisée dans la base de données IBTrACS qui est de 0.1° . Ce choix de résolution repose sur la nécessité de trouver un équilibre entre la capacité de la méthode à produire des résultats réalistes (robustesse) et l'efficacité de la paramétrisation des données pour obtenir ces résultats. En diminuant trop la résolution, il y aurait un risque de manque de données sur chaque case, ce qui pourrait compromettre la qualité des résultats. Ainsi, le choix d'une résolution de 1° permet de garantir à la fois une quantité suffisante de données et une paramétrisation efficace pour obtenir des résultats fiables et significatifs. Il est important de noter que, en considérant la Terre comme parfaitement sphérique, un degré de latitude ou de longitude correspond à environ 111km. Chaque case du quadrillage est ensuite numérotée de 1 à 7194, en commençant par l'extrémité sud-est et en balayant la zone de gauche à droite, comme illustré dans la figure 3.2. De plus, la transition d'un état à un autre s'effectuera en utilisant un intervalle de temps de 3 heures, en accord avec la base de données IBTrACS. Ainsi, l'instant $t = 0$ correspondra à l'heure 0, et l'instant $t = 1$ correspondra à l'heure 3, et ainsi de suite.

Ainsi, une tempête qui se retrouve en un point $\mathbf{r} = (r_x, r_y)$, où $r_x \in [6; 71]$ représente la coordonnée en longitude et $r_y \in [-110; 0]$ la coordonnée en latitude, appartient à un état de la chaîne de Markov identifié par un numéro de localisation compris entre 1 et 7194. Ce numéro de localisation pour chaque point d'une tempête historique est attribué par la formule suivante :

¹Données qui sont associées à des emplacements géographiques spécifiques sur la Terre. Ces données combinent des informations spatiales (comme des coordonnées géographiques ou des limites de zones) avec des données locales (telles que des informations sur la population, le climat, l'utilisation des terres, etc...)

FIGURE 3.2 : Quadrillage du bassin Atlantique Nord en résolution $1^\circ \times 1^\circ$

$$\text{LOC} = 110 \times \lfloor r_y - 6 \rfloor + \lfloor |r_x| \rfloor + 1 \quad (3.1)$$

où $\lfloor x \rfloor$ et $|x|$ désignent respectivement la partie entière et la valeur absolue de x et LOC représente le numéro de la case issue du quadrillage définie sur la figure 3.2 sur laquelle se situe le point \mathbf{r} .

Enfin, pour chaque localisation correspondant à un point de tempête historique, les coordonnées du centre de la case en longitude et en latitude sont attribuées. Par exemple, en appliquant cette méthode au point de genèse de l'ouragan IRMA 1.6, le tableau 3.1 est obtenu :

Coordonnées initiales	LOC $1^\circ \times 1^\circ$	Nouvelles coordonnées
$(16.1^\circ ; -26.9^\circ)$	1107	$(16.5^\circ ; -26.5^\circ)$

TABLE 3.1 : Application de la méthode sur le point de genèse de l'ouragan IRMA

Une fois toutes les localisations historiques obtenues, les probabilités de transition de la chaîne de Markov $(T_n)_{n \geq 0}$ peuvent maintenant être calculées afin d'obtenir la matrice de transition P de dimension 7194×7194 , soit 51 753 636 éléments. Cependant, dans la pratique, certaines lignes et colonnes de la matrice sont composées de valeurs nulles en raison de l'absence d'historique pour certaines localisations. De plus, le stockage d'une matrice aussi volumineuse dans un logiciel informatique pose des difficultés. Pour ces raisons, les valeurs de localisations sont calculées uniquement à partir de l'historique des tempêtes disponibles dans la base de données IBTrACS.

Soit E l'ensemble des états de la chaîne de Markov et $i, j \in E$, où j peut être égal ou différent de i . La probabilité de transition de l'état i à j est définie par :

$$p_{i,j} = \frac{\text{Nombre de localisations suivantes correspondant à } j \text{ en partant de } i}{\text{Nombre total de localisations correspondant à } i} \quad (3.2)$$

Ainsi, à partir de l'historique des tempêtes, un vecteur de probabilités est obtenu pour chaque

localisation i , indiquant les probabilités de transition vers les localisations suivantes potentielles, avec une somme totale égale à 1. Les localisations suivantes non répertoriées sont associées à une probabilité de transition nulle, et la probabilité de passage de l'état mort à n'importe quel autre état est fixée à 0, ce qui marque la fin de la trajectoire. L'état mort correspond historiquement à une localisation où un point d'impact d'une tempête a été enregistré pour la dernière fois dans la base de données IBTrACS.

La simulation de la chaîne $(T_n)_{n \geq 0}$ est réalisée selon la procédure suivante. Dans un premier temps, le point de genèse T_0 d'une tempête est obtenu en utilisant la formule (2.2), basée sur le lissage par noyau gaussien des points de genèse historiques. Ensuite, une valeur de localisation est calculée en utilisant la formule (3.1). Par la suite, en se référant à la formule (3.2), le vecteur de probabilités $V_0 = (p_{0,j_1^0}, \dots, p_{0,j_{t_0-1}^0}, p_{0,mort})$ est associé au point T_0 , où t_0 représente la taille du vecteur V_0 et $(j_1^0, \dots, j_{t_0-1}^0, mort)$ représente l'ensemble des localisations suivantes potentielles répertoriées en partant de T_0 .

Une fois ces étapes préliminaires effectuées, l'intervalle $]0, 1]$ est subdivisé en t_0 sous-intervalles selon la méthode suivante :

- $I_{j_1^0} =]0; p_{0,j_1^0}]$
- $I_{j_2^0} =]p_{0,j_1^0}; p_{0,j_1^0} + p_{j_1^0,j_2^0}]$
- ...
- $I_{j_{mort}^0} =]p_{0,j_1^0} + \dots + p_{0,j_{t_0-1}^0}; \underbrace{p_{0,j_1^0} + \dots + p_{0,j_{t_0-1}^0} + p_{0,j_{mort}^0}}_{=1}]$

Ensuite, une variable aléatoire U est générée selon une distribution uniforme sur l'intervalle $]0, 1]$. Les situations suivantes peuvent alors se présenter :

- Si $U \in I_{j_1^0}$ alors $T_1 = j_1^0$
- Sinon si $U \in I_{j_2^0}$ alors $T_1 = j_2^0$
- ...
- Sinon si $U \in I_{j_{mort}^0}$ alors $T_1 = mort$

Ainsi, en appliquant cette approche, le point T_1 est obtenu, et les étapes décrites précédemment sont répétées à partir de ce nouveau point. La simulation de la chaîne se termine lorsque l'un des points de la chaîne, T_i avec $i \geq 0$, correspond à l'état *mort*.

3.1.2.2 Résultats

En utilisant cette méthode, il est possible de simuler les 157 689 trajectoires de tempêtes pour les 10 000 saisons cycloniques. La figure 3.3 présente un échantillon de 50 trajectoires simulées et empiriques tirés au hasard respectivement parmi les simulations et la base IBTrACS.

Plusieurs observations peuvent être formulées à la suite de ces résultats. Tout d'abord, les trajectoires simulées conservent bien l'aspect parabolique caractéristique des trajectoires de tempêtes historiques. Cependant, toutes les trajectoires simulées présentent des changements de direction plus brusques et une courbure moins prononcée. Cette irrégularité peut s'expliquer par la résolution choisie pour cette étude ($1^\circ \times 1^\circ$), qui est moins précise que celle de la base de données IBTrACS ($0.1^\circ \times 0.1^\circ$).

De plus, il convient de noter que la méthode d'inférence utilisée pour estimer les probabilités de transition entre les positions des tempêtes est non paramétrique. Cette approche, bien que flexible, peut introduire des variations importantes dans les trajectoires simulées en raison de son adaptabilité



FIGURE 3.3 : Echantillon de 50 trajectoires empiriques (à gauche) et simulées (à droite)

excessive aux données. Une alternative prometteuse pour améliorer les résultats consisterait à envisager une méthode de lissage des probabilités de transition autour de chaque point de la trajectoire. Ce lissage pourrait atténuer les changements de direction brusques et permettre de mieux capturer la courbure caractéristique des trajectoires de tempêtes.

La méthode utilisée présente ainsi des limites significatives pour l'étude. En particulier, le problème majeur réside dans le fait que **tous les points de passage d'une trajectoire simulée correspondent à des points de passage de trajectoires empiriques**. Cette limitation pose un problème car le modèle ne prend pas en compte la variabilité inhérente au phénomène cyclonique. En d'autres termes, il n'y a pas de possibilité d'introduire une variabilité ou une diversité dans les trajectoires simulées, ce qui peut être une lacune importante pour la représentation réaliste des cyclones.

Une modélisation similaire basé sur une approche non-paramétrique des chaînes de Markov est proposé par Arthur Charpentier (CHARPENTIER, 2014). L'auteur propose de calculer une probabilité de « mourir » sur chaque case de la grille. Pour chaque case i de la grille, cette probabilité est définie comme suit

$$p_i^{mort} = \frac{\text{Nombre de tempêtes mortes dans la case } i}{\text{Nombre total de tempêtes}}$$

Ainsi, pour chaque case de la grille, une probabilité est associée à la possibilité d'une tempête de « mourir » dans cette case. De plus, un vecteur de localisation contenant toutes les localisations potentielles suivantes, répertoriées dans la base de données historique lui est également attribué.

La simulation de la chaîne $(T_n)_{n \geq 0}$ est réalisée selon la procédure suivante. Supposons une position initiale $T_0 = i$ avec i une localisation parmi les 7194 disponibles sur la grille. Une variable aléatoire U est alors générée selon une distribution uniforme sur l'intervalle $]0, 1]$, et deux cas de figure se présentent :

- Si $U \leq p_i^{mort}$ alors $T_1 = mort$ et la simulation de la chaîne est terminée.
- Si $U > p_i^{mort}$ alors T_1 est tirée aléatoirement parmi toutes les localisations potentielles suivantes contenues dans le vecteur de localisation associée à l'état i , qui peuvent très bien être égales à i elle-même ou correspondre à l'état $mort$.

De manière similaire à l'approche précédente, il est également possible de simuler les 157 689 trajectoires de tempêtes pour les 10 000 saisons cycloniques par cette méthode. La figure 3.4 présente un échantillon de 50 trajectoires simulées et empiriques sélectionnées aléatoirement parmi les simulations et la base de données IBTrACS respectivement. Tout comme dans la première approche, l'aspect parabolique des trajectoires de tempêtes est préservé. De plus, l'utilisation de la probabilité de « mourir

» sur une case en comparaison avec une réalisation d'une loi uniforme permet de mieux capturer les fins de trajectoires observées empiriquement. En effet, dans l'approche précédente, la probabilité de « mourir » (lorsqu'elle était présente) était directement incluse dans le vecteur de probabilités de transition, ce qui permettait à une tempête de ne pas s'arrêter. En revanche, dans cette nouvelle approche, la comparaison préalable entre la réalisation de la loi uniforme et la probabilité de « mourir » permet de mieux saisir le phénomène de dissipation de la tempête.



FIGURE 3.4 : Echantillon de 50 trajectoires empiriques (à gauche) et simulées (à droite)

Cependant, cette nouvelle méthode présente des limites similaires à l'approche précédente, notamment **l'absence de diversité dans les points de passages des trajectoires simulées** qui sont toutes entièrement analogues aux points de passages des trajectoires empiriques sans apporter de variations significatives.

Afin de mieux saisir la diversité des trajectoires cycloniques, il est nécessaire d'adopter d'autres méthodes de modélisation plus adaptées intégrant les caractéristiques temporelles des déplacements des tempêtes. Par la suite, une modélisation des trajectoires de tempêtes en utilisant des séries temporelles est proposé, offrant ainsi un cadre plus approprié pour capturer la variabilité et les schémas de déplacement observés.

3.1.3 Approche par Séries Temporelles

Une synthèse des principes fondamentaux des séries temporelles (SYLVAIN, 2023) utile pour l'étude sont présentés dans l'annexe A.3. Dans cette section, l'application de ces concepts dans le contexte spécifique de cette étude est abordée, ainsi que les résultats obtenus et les éventuelles limites identifiées.

3.1.3.1 Application

La modélisation adoptée repose sur l'hypothèse selon laquelle **les déplacements des tempêtes, exprimés par leurs coordonnées en longitude et latitude, peuvent être représentés par des processus auto-régressifs**. Cette méthodologie découle des travaux antérieurs de JAMES et MASON (2005), qui ont employé cette méthode pour modéliser les trajectoires de tempêtes dans la mer de Corail, située dans le Sud-ouest de l'océan Pacifique. Dans le contexte de cette étude, la démarche s'inspire des recherches réalisées par BLOEMENDAAL et al. (2020) dans ce domaine spécifique. Ces chercheurs ont appliqué la méthodologie développée par JAMES et MASON (2005) pour générer une base de données de trajectoires de tempêtes couvrant l'ensemble des bassins cycloniques mondiaux, incluant le bassin Atlantique Nord.

L'approche adoptée repose sur l'analyse de plusieurs observations empiriques. Pour chaque tempête de la base de données IBTrACS, les variations longitudinales et latitudinales observés sont regroupés en séries temporelles distinctes, à savoir $(\xi_t)_{t \geq 0}$ pour la longitude et $(\phi_t)_{t \geq 0}$ pour la latitude. Les différences

longitudinales et latitudinales, représentées respectivement par $\Delta\xi_t = \xi_t - \xi_{t-1}$ et $\Delta\phi_t = \phi_t - \phi_{t-1}$, sont calculées entre deux points successifs le long de chaque trajectoire de tempête. L'évolution de ces différences est illustrée dans la figure 3.5, où chaque trajectoire de la base IBTrACS est représentée.

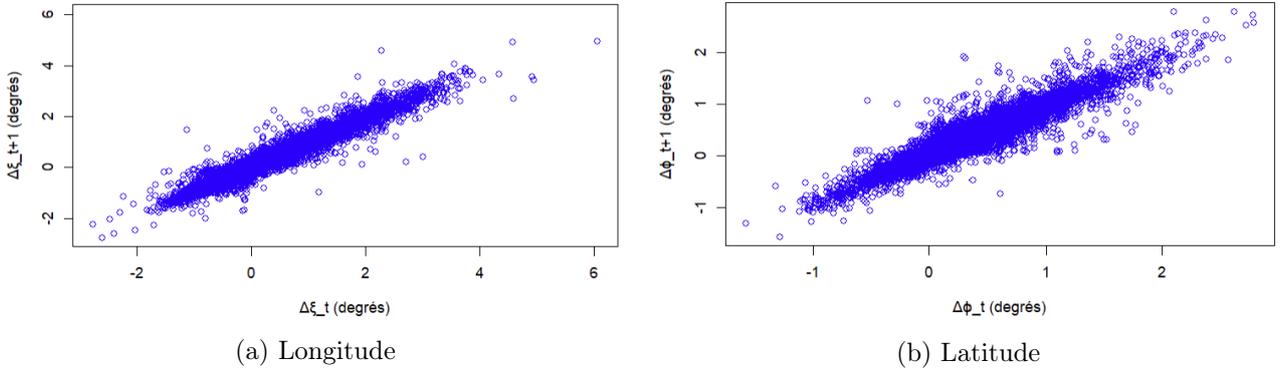


FIGURE 3.5 : Évolution des différences longitudinales et latitudinales entre points successifs d'une trajectoire de tempête

Une linéarité évidente est observable lors de l'analyse des deux graphiques représentant les différences successives des coordonnées de longitude et de latitude. Cette forte linéarité suggère que les différences longitudinales et latitudinales entre deux points successifs d'une trajectoire de tempête peuvent être expliquées par des processus auto-régressifs d'ordre 1

$$\Delta\xi_t = a_0 + a_1\Delta\xi_{t-1} + \varepsilon_t^\xi, \quad (3.3)$$

$$\Delta\phi_t = b_0 + b_1\Delta\phi_{t-1} + \varepsilon_t^\phi, \quad (3.4)$$

avec a_0 , a_1 , b_0 , b_1 des coefficients à estimer et ε_t^ξ , ε_t^ϕ des bruits blancs centrés de variance constante et indépendant respectivement de $\Delta\xi_{t-1}$ et $\Delta\phi_{t-1}$.

Le choix de travailler sur des séries temporelles différenciées des variations longitudinales et latitudinales s'inscrit dans la recherche de la stationnarité des données. La stationnarité est une hypothèse fondamentale dans l'analyse des séries temporelles, notamment lorsque l'on souhaite utiliser des modèles statistiques pour capturer les comportements temporels. L'hypothèse de stationnarité repose sur l'idée que les propriétés statistiques d'une série temporelle demeurent constantes à travers le temps. En d'autres termes, la moyenne, la variance et la structure de dépendance de la série restent inchangées sur différentes périodes. Cette hypothèse est essentielle pour appliquer des modèles prédictifs, car elle permet de supposer que les relations statistiques observées dans le passé continueront à être valables à l'avenir. Cependant, dans le cas des trajectoires de tempêtes, les données longitudinales et latitudinales brutes peuvent souvent présenter des tendances ou des comportements non stationnaires. Par exemple, les tempêtes peuvent présenter des cycles saisonniers, des variations à long terme liées au changement climatique, ou d'autres irrégularités temporelles. Travailler sur des séries temporelles différenciées permet de supprimer ces tendances temporelles non stationnaires, transformant ainsi les données en des séries plus stationnaires. En se concentrant sur des séries temporelles différenciées, il est possible de mieux isoler les composantes stochastiques des mouvements des tempêtes, ce qui facilite la modélisation à l'aide de méthodes telles que les processus autorégressifs. En fin de compte, cette approche permet de capturer de manière plus robuste les comportements aléatoires sous-jacents des trajectoires de tempêtes, ce qui est essentiel pour la simulation réaliste des déplacements futurs des tempêtes cycloniques.

D'après les expériences réalisées par JAMES et MASON (2005), il a été constaté que des ordres plus élevés pour les processus décrits par les équations (3.3) et (3.4) n'apportent que peu ou pas

d'avantages supplémentaires, et ces modèles sont suffisants dans le cadre de l'étude. Cependant, les auteurs soulignent que ces modèles de base ne parviennent pas à reproduire certains aspects importants du comportement des tempêtes observés dans les données historiques.

Tout d'abord, l'une des principales caractéristiques des tempêtes est leur rareté à des latitudes très basses. Dans tous les bassins cycloniques mondiaux, il y a très peu de trajectoires d'ouragans qui se forment à environ 10 degrés de l'Équateur, et encore moins à environ 5 degrés de l'Équateur. Dans le bassin Atlantique Nord, aucune tempête ne s'est formée à moins de 5 degrés de l'Équateur. De plus, toutes les trajectoires qui se sont formées à des latitudes plus élevées se sont toutes dirigées vers le nord et/ou l'ouest, comme observé sur la carte 3.1 des trajectoires historiques des tempêtes. Ainsi, l'analyse historique des trajectoires de tempêtes est convaincante en ce qui concerne la tendance des ouragans à s'éloigner de l'Équateur, une tendance qui est particulièrement dominante à des latitudes très basses, et qui diminue progressivement à mesure que la tempête s'en éloigne.

Les auteurs proposent donc d'intégrer cette tendance au modèle de base pour la latitude (3.4) en ajoutant un terme non linéaire qui renforce considérablement la tendance de ϕ_t à se diriger vers le nord lorsque le cyclone se rapproche de l'Équateur. Ce terme non linéaire peut prendre différentes formes, et dans le cadre de cette étude, un terme variant de manière inversement proportionnelle à la latitude est choisi. Ainsi, ce terme sera d'autant plus grand lorsque la tempête sera proche de l'Équateur, ce qui entraînera une plus grande variation latitudinale et forcera la tempête à s'éloigner davantage de l'Équateur. Dans le cadre de l'étude, les modèles considérés pour expliquer les différences longitudinales et latitudinales entre deux points successifs d'une trajectoire de tempête sont donc les suivants

$$\Delta\xi_t = a_0 + a_1\Delta\xi_{t-1} + \varepsilon_t^\xi, \quad (3.5)$$

$$\Delta\phi_t = b_0 + b_1\Delta\phi_{t-1} + \frac{b_2}{\phi_{t-1}} + \varepsilon_t^\phi, \quad (3.6)$$

avec un nouveau coefficient b_2 à estimer.

3.1.3.2 Construction du modèle

La construction du modèle s'inspire des travaux réalisés par BLOEMENDAAL et al. (2020). Les variations longitudinales et latitudinales sont regroupées par section de 5 degrés de latitude du bassin Atlantique Nord. Cette approche est justifiée par le fait qu'une tempête suit généralement une trajectoire parabolique au cours de sa vie et qu'un mouvement commun de trajectoire est souvent observable à chaque section de 5 degrés de latitude.

Le but de cette approche est de capturer cette tendance générale de déplacement des tempêtes du bassin Atlantique Nord tout en intégrant une variation stochastique obtenue à partir de l'étude des résultats des modèles. L'objectif est d'obtenir des trajectoires réalistes et représentatives de la variabilité du phénomène de déplacement des tempêtes. Ainsi, cette approche permettra d'obtenir un échantillon de trajectoires de tempêtes plus diversifié, tout en maintenant les caractéristiques essentielles des trajectoires observées empiriquement. Cela contribuera à mieux saisir la diversité des trajectoires cycloniques et à améliorer la modélisation des phénomènes de déplacement.

Le périmètre spatial en latitude défini en section 2.1.3 est modifié et fixé à $[5^\circ, 60^\circ]$ en raison d'un nombre d'observations faible pour des latitudes supérieures à 60° . Le découpage en section de 5 degré de latitude est représenté sur la figure 3.6. Sur chacune de ces sections, les modèles définis en (3.5) et (3.6) sont entraînés en utilisant les variations longitudinales $\Delta\xi_t$ et latitudinales $\Delta\phi_t$ historiques issues de la base de données IBTrACS.

Ainsi l'analyse repose sur un total de 22 modèles de régression, comprenant 11 modèles pour la latitude et 11 modèles pour la longitude. Les coefficients a_0 , a_1 , b_0 , b_1 et b_2 sont estimés en utilisant les méthodes de régression linéaires.

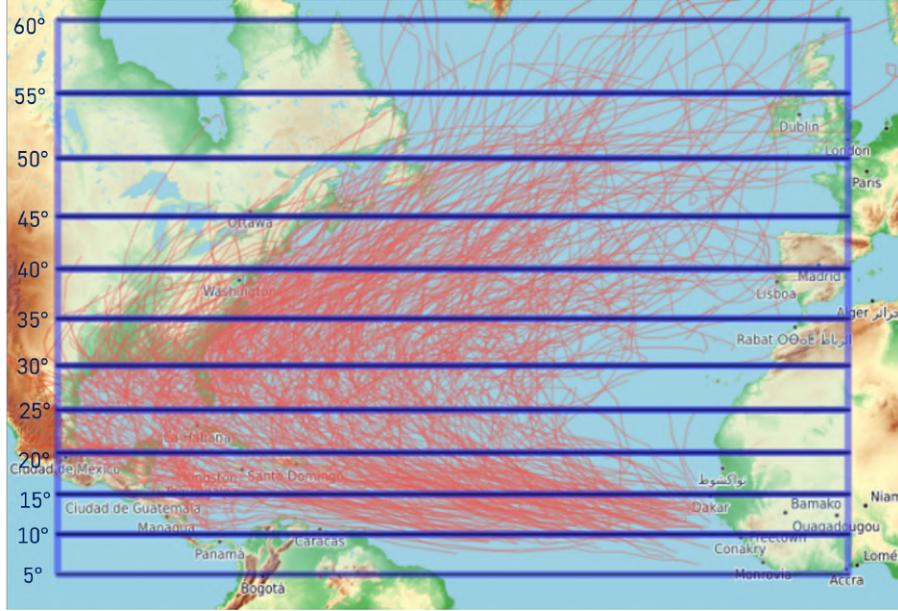


FIGURE 3.6 : Découpage du bassin Atlantique Nord par section de 5° en latitude

Le tableau 3.2 présente les coefficients obtenus pour chaque modèle associé à chaque section de latitude de la figure 3.6 ainsi que les écart-types des résidus de chacun de ces modèles. La significativité des résidus est analysée dans la section présentant les résultats finaux de la méthode en section 3.1.3.5.

En examinant le tableau 3.3, il est remarquable que les écart-types des coefficients estimés demeurent généralement minimales, notamment dans les sections de latitude où la base de données IB-TrACS répertorie un grand nombre d'impacts cycloniques (c'est-à-dire en dessous de 35 degrés de latitude). Cela souligne la robustesse des modèles dans ces régions densément couvertes par les données cycloniques. Cependant, il est intéressant de noter que les écart-types des coefficients estimés augmentent significativement dans les zones où les points d'impact cycloniques sont moins fréquents, spécifiquement au-dessus de 35 degrés de latitude. Cette variation dans l'incertitude des coefficients pourrait être attribuée à une disponibilité limitée de données dans ces régions, ce qui rend l'estimation plus délicate. Pour une évaluation plus approfondie de la fiabilité et de la pertinence des modèles de régression utilisés, des analyses complémentaires telles que l'étude des résidus et la réalisation de tests de prédiction s'avèrent nécessaires.

Section de 5°	a_0	a_1	σ_ξ	b_0	b_1	b_2	σ_ϕ
[55° ; 60°]	-0.0187	1.0561	0.0868	-0.5478	0.9532	29.2829	0.0508
[50° ; 55°[-0.0190	1.0021	0.0860	-0.2245	0.9536	11.7013	0.0578
[45° ; 50°[0.1039	0.9346	0.0923	-0.2875	0.9701	13.4501	0.0654
[40° ; 45°[0.0213	1.0081	0.0945	-0.0244	1.0156	10.1940	0.0831
[35° ; 40°[0.0278	1.0126	0.1079	-0.0243	0.9960	1.2904	0.0945
[30° ; 35°[0.0240	1.0015	0.1222	0.0425	0.9807	-0.9850	0.1054
[25° ; 30°[0.0135	0.9820	0.1402	-0.0443	0.9649	1.6697	0.1154
[20° ; 25°[0.0021	0.9792	0.1992	0.0437	0.9544	-0.5991	0.1535
[15° ; 20°[-0.01158	0.9577	0.3196	0.0076	0.9569	0.0392	0.2005
[10° ; 15°[-0.0138	0.9714	0.3754	0.0145	0.9366	0.0077	0.2161
[5° ; 10°[-0.4531	0.4440	0.6869	0.0190	0.8328	0.4405	0.2315

TABLE 3.2 : Résultats des estimations des 22 modèles de régression

Section de 5°	σ_{a_0}	σ_{a_1}	σ_{b_0}	σ_{b_1}	σ_{b_2}
[55° ; 60°]	0.0557	0.0395	0.6079	0.0265	34.90
[50° ; 55°[0.0294	0.0161	0.3619	0.0149	18.85
[45° ; 50°[0.0197	0.0125	0.2240	0.0111	10.58
[40° ; 45°[0.0075	0.0065	0.1077	0.0067	4.524
[35° ; 40°[0.0029	0.0038	0.0484	0.0045	1.802
[30° ; 35°[0.0016	0.0035	0.0309	0.0042	0.9962
[25° ; 30°[0.0014	0.0036	0.0236	0.0043	0.6503
[20° ; 25°[0.0017	0.0041	0.0188	0.0052	0.4136
[15° ; 20°[0.0019	0.0036	0.0106	0.0048	0.1827
[10° ; 15°[0.0027	0.0040	0.0077	0.0058	0.0982
[5° ; 10°[0.0464	0.0515	0.0524	0.0033	0.4771

TABLE 3.3 : Ecart-type des coefficients estimés

3.1.3.3 Méthode de simulation

De même que l'approche markovienne, la transition d'un état à un autre s'effectue en utilisant un intervalle de temps de 3 heures, en accord avec la base de données IBTrACS. En conséquence, l'instant $t = 0$ correspondra à l'heure 0, tandis que l'instant $t = 1$ correspondra à l'heure 3, et ainsi de suite. Il est donc nécessaire de fixer un nombre maximal de points simulés, noté T , pour une trajectoire de tempête. Pour fixer ce nombre de points maximal, il est d'abord essentiel d'avoir un aperçu de la durée maximale des tempêtes en fonction de leur point de genèse. La figure 3.7 représente la durée des tempêtes observées historiquement en fonction de leur lieu de genèse. Il est observable que les tempêtes qui naissent près de la côte ouest africaine ont une durée en moyenne plus longue que celles qui naissent près du continent américain, soit au Golfe du Mexique ou sur la côte Est des Etats-Unis.

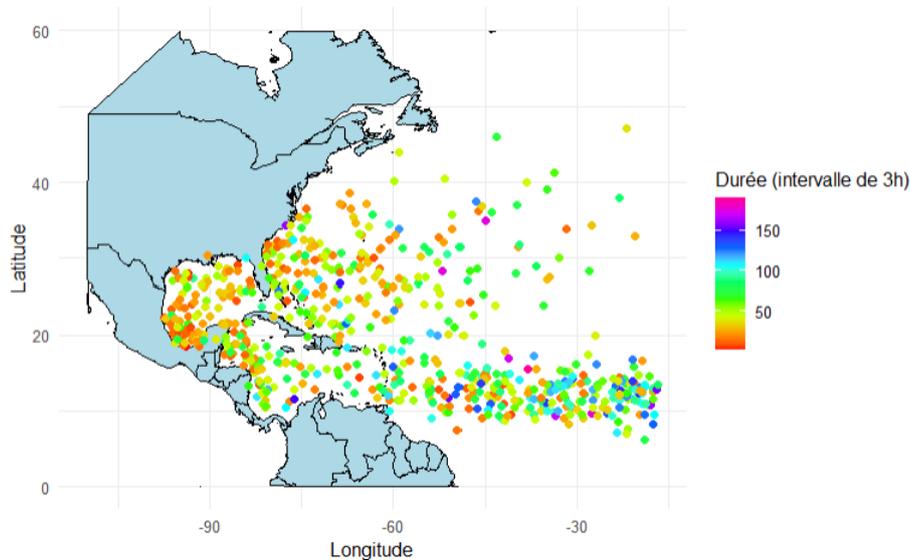


FIGURE 3.7 : Carte de la durée des tempêtes historiques en fonction du lieu de genèse (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b)

Ainsi pour fixer T de manière efficace, il est choisi de diviser l'échantillon de durée des tempêtes historiques en deux zones : les durées historiques des tempêtes qui sont nés près de la côte ouest africaine, noté **zone 1**, et le reste des tempêtes, noté **zone 2**. Ainsi pour le groupe 1, toutes les

durées de tempêtes historiques dont les coordonnées des points de genèse en longitude et latitude sont supérieures et inférieures respectivement à - 60 et 25 degrés. La figure 3.8 représente les histogrammes obtenues des durées de tempêtes historiques en fonction des deux zones établies.

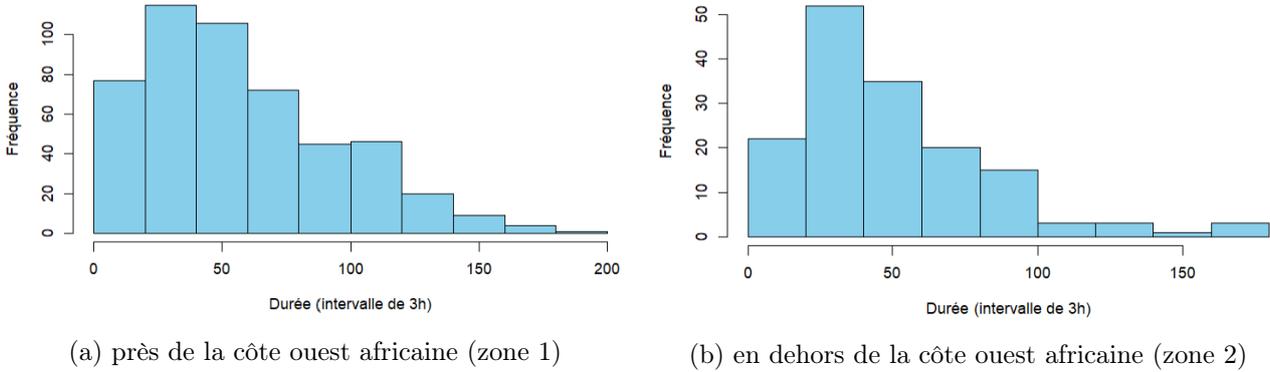


FIGURE 3.8 : Histogramme des durées de tempêtes historiques en fonction de la zone de genèse

Ensuite, une analyse d'ajustement à deux distributions de probabilité, à savoir la distribution gamma et la distribution log-normale, est effectuée en utilisant la méthode du maximum de vraisemblance.

Pour rappel si $X \sim \mathcal{G}(\alpha, \beta)$ alors la fonction de densité de X est donnée par :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

où $x > 0$, $\alpha > 0$ est le paramètre de forme, $\beta > 0$ est le paramètre d'échelle, et Γ est la fonction gamma.

Le principe de la méthode du maximum de vraisemblance pour estimer les paramètres de la loi gamma à partir d'un échantillon $\mathcal{D} = (x_1, x_2, \dots, x_n)$ est de trouver les valeurs de $\hat{\alpha}_{EMV}$ et $\hat{\beta}_{EMV}$ qui maximisent la fonction de vraisemblance $L(\alpha, \beta | \mathcal{D})$ définie comme suit

$$L(\alpha, \beta | \mathcal{D}) = \prod_{i=1}^n f(x_i; \alpha, \beta).$$

En pratique, il est plus courant de travailler avec le logarithme de la fonction de vraisemblance pour faciliter les calculs, ce qui donne la log-vraisemblance

$$\ln L(\alpha, \beta | \mathcal{D}) = \sum_{i=1}^n \left[\ln \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right) + (\alpha - 1) \ln(x_i) - \beta x_i \right].$$

Les estimations des paramètres $\hat{\alpha}_{EMV}$ et $\hat{\beta}_{EMV}$ sont alors obtenues en résolvant les équations suivantes

$$\frac{\partial}{\partial \alpha} \ln L(\alpha, \beta | \mathcal{D}) = 0, \quad \frac{\partial}{\partial \beta} \ln L(\alpha, \beta | \mathcal{D}) = 0.$$

Ces équations permettent de déterminer les valeurs de $\hat{\alpha}_{EMV}$ et $\hat{\beta}_{EMV}$ qui maximisent la fonction de vraisemblance et fournissent ainsi les estimations des paramètres de la loi gamma à partir de l'échantillon observé.

Les formules pour les estimations des paramètres sont les suivantes

$$\hat{\alpha}_{EMV} = \frac{\bar{X}^2}{s^2}, \quad \hat{\beta}_{EMV} = \frac{\bar{X}}{s^2},$$

où \bar{X} est la moyenne de l'échantillon \mathcal{D} et s^2 est la variance de l'échantillon.

La loi log-normale est souvent utilisée pour modéliser des variables aléatoires positives qui ont une distribution logarithmique normale. Si X suit une loi log-normale, sa fonction de densité est définie comme suit

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}},$$

où $x > 0$, μ est la moyenne logarithmique et σ est l'écart-type logarithmique.

Pour estimer les paramètres $\hat{\mu}_{EMV}$ et $\hat{\sigma}_{EMV}$ de la loi log-normale à partir d'un échantillon $\mathcal{D} = (x_1, x_2, \dots, x_n)$, la méthode du maximum de vraisemblance consiste à maximiser la fonction de vraisemblance $L(\mu, \sigma|\mathcal{D})$ définie comme suit

$$L(\mu, \sigma|\mathcal{D}) = \prod_{i=1}^n f(x_i; \mu, \sigma).$$

En pratique, il est plus courant de travailler avec le logarithme de la fonction de vraisemblance, ce qui donne la log-vraisemblance

$$\ln L(\mu, \sigma|\mathcal{D}) = -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \sum_{i=1}^n \left[\frac{(\ln(x_i) - \mu)^2}{2\sigma^2} \right].$$

Les estimations des paramètres $\hat{\mu}_{EMV}$ et $\hat{\sigma}_{EMV}$ sont obtenues en résolvant les équations suivantes, dérivées de la log-vraisemblance

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma|\mathcal{D}) = 0, \quad \frac{\partial}{\partial \sigma} \ln L(\mu, \sigma|\mathcal{D}) = 0.$$

Les formules pour les estimations des paramètres sont les suivantes

$$\hat{\mu}_{EMV} = \frac{1}{n} \sum_{i=1}^n \ln(x_i), \quad \hat{\sigma}_{EMV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\ln(x_i) - \hat{\mu}_{EMV})^2}.$$

Ces estimations des paramètres permettent de caractériser la loi log-normale à partir de l'échantillon observé et de l'utiliser dans la modélisation statistique des données.

Les résultats de ces ajustements mettent en évidence une meilleure concordance avec une distribution log-normale pour les durées historiques des tempêtes dont le point de genèse est situé près de la côte africaine. En revanche, une distribution gamma est privilégiée pour l'autre catégorie de tempêtes. Les graphiques illustrant ces ajustements sont présentés dans la figure 3.9, avec une distribution gamma pour la zone 1 et une distribution log-normale pour la zone 2.

Il est notable que les graphiques comparant les densités, les QQ-plots, les PP-plots, ainsi que les fonctions de répartition entre les données prédites et les données empiriques sont satisfaisants et montrent une concordance remarquable. Par conséquent, les distributions précédemment identifiées par chaque zone sont retenues pour les analyses ultérieures. Les résultats numériques de la calibration sont présentés dans le tableau 3.4, tandis que le tableau 3.5 expose les résultats des tests de Kolmogorov-Smirnov.

Le principe mathématique du test de Kolmogorov-Smirnov repose sur la comparaison de la fonction de répartition empirique $F_n(x)$ de l'échantillon avec la fonction de répartition théorique $F(x)$ de la loi gamma ou log-normale (selon l'échantillon retenue) (CORDER et FOREMAN, 2014). La statistique de test D_n est calculée en prenant la plus grande différence verticale entre ces deux fonctions :

$$D_n = \sup_x |F_n(x) - F(x)|$$

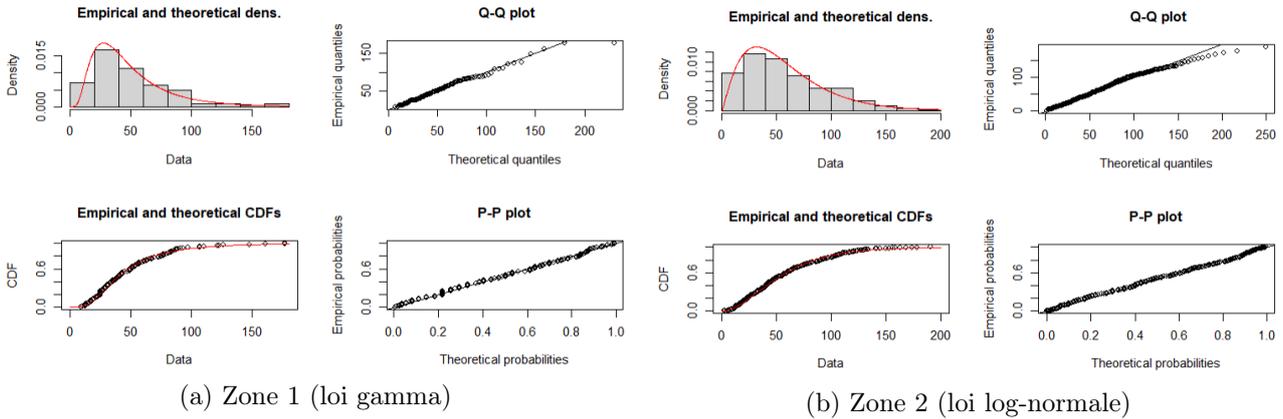


FIGURE 3.9 : Résultats des calibrations réalisées

Si la statistique de test D_n est inférieure à une certaine valeur critique, généralement déterminée par un seuil de significativité (fixée à 5%), alors l'hypothèse nulle selon laquelle l'échantillon suit la loi gamma ou log-normale n'est pas rejetée. En revanche, si D_n dépasse la valeur critique, cela suggère que la loi retenue ne convient pas bien aux données.

Zone	Distribution	Log-vraisemblance	Paramètres estimés
1	$\mathcal{G}(\alpha, \beta)$	-2425.66	$\hat{\alpha}_{EMV} = 2.238$; $\hat{\beta}_{EMV} = 0.038$
2	$\mathcal{LN}(\mu, \sigma)$	-719.83	$\hat{\mu}_{EMV} = 3.708$; $\hat{\sigma}_{EMV} = 0.635$

TABLE 3.4 : Résultats obtenus par maximum de vraisemblance

Zone	Distribution	Statistique de Test	P-value
1	$\mathcal{G}(\hat{\alpha}_{EMV}, \hat{\beta}_{EMV})$	0.044	0.29
2	$\mathcal{LN}(\hat{\mu}_{EMV}, \hat{\sigma}_{EMV})$	0.048	0.85

TABLE 3.5 : Résultats des tests d'adéquation de Kolmogorov-Smirnov

En examinant les résultats des tests d'adéquation de Kolmogorov-Smirnov pour les deux groupes de durées de tempêtes, il est observable que les distributions ajustées (la loi Gamma pour la zone 1 et la loi log-normale pour la zone 2) présentent des statistiques de test relativement faibles, accompagnées de p-values significatives. Ces résultats indiquent que les distributions proposées sont en accord avec les données historiques. En d'autres termes, les modèles de loi de probabilité sélectionnés semblent bien représenter les durées de tempêtes observées dans les deux zones considérées.

Avec la détermination précise de la valeur de T , il est désormais possible de procéder à la simulation complète d'une trajectoire de tempête. Cette méthode de simulation s'articule autour des étapes suivantes.

Etape 1 : Détermination du point de genèse et de la durée d'une trajectoire

Au temps initial $t = 0$, le point de genèse est simulé et noté (ξ_0, ϕ_0) ainsi que son mois de survenance. Cette simulation repose sur la formule 2.12, qui implique un lissage par noyau gaussien des points de genèse historiques et l'utilisation de la distribution historique des mois de survenance décrite dans la figure 1.10. La simulation du mois de survenance est nécessaire car elle représente une variable importante dans la section 3.2 dédiée à la modélisation de l'intensité des tempêtes cycloniques. De plus (ξ_0, ϕ_0) est associé à une section de latitude notée l_0 , définie sur la figure 3.6, avec l_0 appartenant à

l'ensemble $\mathcal{L} = \left\{ [5^\circ; 10^\circ[, [10^\circ; 15^\circ[, \dots, [55^\circ; 60^\circ] \right\}$. Ensuite, en fonction de la zone géographique dans laquelle se situe le point de genèse simulé, la durée de vie maximale T de la tempête est tiré selon une loi gamma ou une loi log-normale, avec les paramètres obtenus lors des calibrations qui sont présentés en détail dans le tableau 3.4.

Etape 2 : Détermination du point suivant la genèse

Au temps $t = 1$, le point suivant de la trajectoire (ξ_1, ϕ_1) est calculé comme suit.

Calcul de ξ_1 : Il faut tout d'abord déterminer la variation longitudinale $\Delta\xi_1$ à l'aide de la formule de régression (3.5), ce qui donne

$$\Delta\xi_1^{l_0} = a_0^{l_0} + a_1^{l_0} \Delta\xi_0^{l_0}. \quad (3.7)$$

Les coefficients $a_0^{l_0}$ et $a_1^{l_0}$ proviennent des modèles de régression de longitude associé à la section de latitude l_0 , regroupés dans le tableau 3.2.

Le terme $\Delta\xi_0^{l_0}$, est obtenu en effectuant un tirage aléatoire d'une loi normale d'espérance et variance égales à celles de $(\Delta\xi_{t-1}^{l_0})_{t \geq 1}$, représentant les variations longitudinales au temps $t - 1$ précédant chaque variation longitudinale à l'instant t associé à la section l_0 . Cette approche est justifiée par le fait que chaque groupe $(\Delta\xi_{t-1}^l)_{t \geq 1}$, pour tout $l \in \mathcal{L}$, peut être modélisé à l'aide d'une distribution normale dont l'espérance et la variance sont égales à celles du groupe.

Une fois $\Delta\xi_0^{l_0}$ déterminé, $\Delta\xi_1^{l_0}$ est obtenu en utilisant (3.7). La position longitudinale ξ_1 au temps $t = 1$ se déduit ainsi comme suit

$$\xi_1 = \xi_0 + \Delta\xi_1^{l_0} + \varepsilon_0^{\xi, l_0},$$

avec ξ_0 la longitude à $t = 0$, $\Delta\xi_1^{l_0}$ la variation longitudinale entre les dates 0 et 1, et ε_0^{ξ, l_0} le terme d'erreur aléatoire. Ce terme d'erreur aléatoire est issue de l'étude des résidus du modèle de régression de longitude définie en (3.5) qui sont supposés gaussiens et indépendants. Ainsi le terme ε_0^{ξ, l_0} est simulé selon une réalisation d'une loi normale d'espérance et de variance égales à celles des résidus associés à la section de latitude l_0 .

Calcul de ϕ_1 : De manière similaire à ξ_1 , il faut tout d'abord déterminer la variation latitudinale $\Delta\phi_1$ à l'aide de la formule de régression (3.6), ce qui donne

$$\Delta\phi_1^{l_0} = b_0^{l_0} + b_1^{l_0} \Delta\phi_0^{l_0} + \frac{b_2^{l_0}}{\phi_0}. \quad (3.8)$$

Les coefficients $b_0^{l_0}$, $b_1^{l_0}$ et $b_2^{l_0}$ proviennent des modèles de régression de latitude de la section l_0 , regroupés dans le tableau 3.2.

De même que pour $\Delta\xi_0^{l_0}$, le terme $\Delta\phi_0^{l_0}$ est obtenu en effectuant un tirage aléatoire à partir d'une distribution normale d'espérance et variance égales à $(\Delta\phi_{t-1}^{l_0})_{t \geq 1}$, représentant les variations latitudinales au temps $t - 1$ précédant chaque variation latitudinale à l'instant t associé à la section l_0 .

Une fois $\Delta\phi_0^{l_0}$ déterminé, $\Delta\phi_1^{l_0}$ est calculable en utilisant la relation définie en (3.8) et ainsi il est possible de déduire ϕ_1 à l'aide de la formule de récurrence suivante

$$\phi_1 = \phi_0 + \Delta\phi_1^{l_0} + \varepsilon_0^{\phi, l_0},$$

avec ϕ_0 la longitude à $t = 0$, $\Delta\phi_1^{l_0}$ la variation longitudinale entre les dates 0 et 1, et $\varepsilon_0^{\phi, l_0}$ le terme d'erreur aléatoire issue de l'étude des résidus du modèle de régression de latitude définie en (3.6), simulé de manière similaire à ε_0^{ξ, l_0} .

Cette étape revêt une importance cruciale, car selon les auteurs (BLOEMENDAAL et al., 2020) l'approche aléatoire utilisée pour déterminer $\Delta\xi_1^{l_0}$ et $\Delta\phi_1^{l_0}$ permet d'obtenir des trajectoires de tempêtes plus diversifiées et aléatoires, en accord avec les caractéristiques réelles des phénomènes observés.

Etape 3 : Détermination de tous les points suivants

Pour tout $1 < t \leq T$, les positions suivantes de la trajectoire $(\xi_t, \phi_t)_{1 < t \leq T}$ sont obtenues en réitérant les formules de régressions utilisées à l'étapes précédentes en tenant compte des changements de section de latitude, entraînant des variations dans les résidus pris en considération ainsi que dans les différents coefficients des formules de régression. Les formules de récurrence sont les suivantes

$$\xi_t = \xi_{t-1} + \Delta \xi_t^{l_{t-1}} + \varepsilon^{\xi, l_{t-1}}, \quad \varepsilon^{\xi, l_{t-1}} \sim \mathcal{N}(\mu^{\xi, l_{t-1}}, \sigma^{\xi, l_{t-1}}), \quad (3.9)$$

$$\phi_t = \phi_{t-1} + \Delta \phi_t^{l_{t-1}} + \varepsilon^{\phi, l_{t-1}}, \quad \varepsilon^{\phi, l_{t-1}} \sim \mathcal{N}(\mu^{\phi, l_{t-1}}, \sigma^{\phi, l_{t-1}}), \quad (3.10)$$

avec $l_{t-1} \in \mathcal{L}$ la section de latitude correspondante au temps $t - 1$.

Des contraintes supplémentaires sont introduites dans le modèle afin d'éliminer certaines trajectoires potentiellement irréalistes. L'analyse des trajectoires historiques révèle qu'un nombre très limité de trajectoires ont été observées aux basses latitudes. En particulier, la figure 3.1 démontre que toutes les trajectoires situées en dessous de 10 degrés de latitude s'orientent vers le nord. De plus, au-delà de 45 degrés de latitude, la grande majorité des trajectoires présentent une variation longitudinale importante en direction de l'est. Quelques rares trajectoires présentent une variation longitudinale vers l'ouest, mais elles sont peu nombreuses et se situent presque en dehors du domaine d'étude (au-dessus de 60 degrés en latitude). Ainsi, plusieurs scénarios sont pris en compte dans le modèle :

- Si $\phi_{t-1} > 45^\circ$ et $\Delta \xi_{t-1}^{l_{t-1}} < 0$ pour tout $1 \leq t \leq T$ alors

$$\Delta \xi_{t-1}^{l_{t-1}} < 0,$$

puis la formule de récurrence en 3.9 est appliquée ;

- Si $\phi_{t-1} < 10^\circ$ pour tout $1 \leq t \leq T$ alors la formule de récurrence est modifiée et devient

$$\phi_t = \xi_{t-1} + \Delta \phi_t^{l_t} + |\varepsilon^{\phi, l_{t-1}}|, \quad \varepsilon^{\phi, l_{t-1}} \sim \mathcal{N}(\mu^{\phi, l_{t-1}}, \sigma^{\phi, l_{t-1}}),$$

où $|x|$ désigne la valeur absolue de x .

Sinon la formule de récurrence en 3.10 est appliquée.

Etape 4 : Détermination du moment d'arrêt de la trajectoire

À présent que la trajectoire peut être modélisée à tout instant, il est essentiel de déterminer le moment où elle prend fin, c'est-à-dire quand la tempête atteint l'état de « mort ». Plusieurs options sont envisageables.

Dans leur modélisation, BLOEMENDAAL et al. (2020) proposent trois approches distinctes pour mettre fin à une trajectoire de tempête : la première approche consiste à considérer qu'une tempête est « morte » dès lors qu'elle sort du domaine de définition délimité par $[-110^\circ, 0^\circ] \times [5^\circ, 60^\circ]$. La deuxième approche consiste à considérer qu'une tempête est « morte » lorsqu'elle entre en contact avec une zone terrestre et y demeure pendant au moins trois intervalles de temps successifs, soit un total de 9 heures sur une étendue terrestre. Cependant, cette approche présente un problème majeur, car elle entraîne la suppression de trajectoires qui ne reflètent pas fidèlement la réalité des phénomènes observés. En effet, l'analyse des trajectoires historiques révèle qu'un nombre important de tempêtes restent sur des étendues terrestres pendant plus de 9 heures et peuvent continuer à causer des dégâts considérables. Par exemple, l'ouragan Irma, représenté sur la figure 1.6, a touché la Floride le 10 septembre 2017, mais ne s'est complètement dissipé que le 13 septembre 2017 à 18h, soit une durée bien supérieure à

9h jours (JOHN et al., 2021). Ainsi, cette approche est problématique, car l'étude s'intéresse également à l'évolution de la trajectoire des tempêtes ainsi qu'à leur intensité associée sur les terres. La troisième approche utilisée est prise naturellement en compte par le modèle car pour chaque tempête simulée est associée une durée de vie T en heures tirée de la distribution des durées de tempêtes historiques 3.7.

Une approche alternative plus cohérente consisterait à faire « mourir » une tempête en fonction de l'évolution de son intensité. En effet, l'une des caractéristiques essentielles d'un ouragan est la vitesse maximale de ses vents, et comme expliqué en section 1.2.4, une tempête évoluant en ouragan se dissipe instantanément lorsqu'elle entre en contact avec une terre. Ainsi, une solution plus adéquate serait de mettre fin à la trajectoire de la tempête dès lors que son intensité est fortement diminuée, c'est-à-dire lorsque la vitesse maximale de ses vents passe en dessous d'un certain seuil où les dégâts liés à la tempête sont minimales ou inexistantes. Cette approche a été adoptée par RUMPF et al. (2007) dans sa modélisation des trajectoires de tempêtes dans le bassin Pacifique nord-ouest.

Dans le cadre de cette étude, **trois approches distinctes ont été sélectionnées pour déterminer le moment d'arrêt d'une trajectoire de tempête**. Pour rappel, chaque tempête simulée se voit attribuer aléatoirement une durée de vie maximale T , tirée d'une distribution illustrée dans la figure 3.7. Les trois méthodes de fin d'une trajectoire de tempête sont les suivantes :

1. **Atteint de la durée de vie maximale T** : Lorsqu'une tempête atteint sa durée de vie maximale T , sa trajectoire se termine immédiatement quel que soit son emplacement, qu'il soit sur terre ou en mer ;
2. **Sortie du domaine de définition** : Cette deuxième approche s'applique si la durée de vie maximale T n'est pas atteinte, mais que la tempête quitte le domaine de définition. Ce domaine est délimité par des frontières géographiques spécifiques définies par $[-110^\circ, 0^\circ] \times [5^\circ, 60^\circ]$;
3. **Diminution de l'intensité** : Cette troisième approche est utilisée si la durée de vie maximale T n'est pas atteinte et que la tempête ne quitte pas le domaine de définition, mais qu'elle perd en intensité lorsqu'elle entre en contact avec la terre, jusqu'à atteindre un seuil spécifique où les dommages potentiels deviennent négligeables. Ce seuil est fixé à 68 km/h, correspondant à la catégorie de dépression tropicale sur l'échelle de Saffir-Simpson. Ce traitement est effectué lors de la modélisation des vitesses maximales des vents détaillé dans la section 3.2.2.1.

3.1.3.4 Nouvelle approche de modélisation de la fin d'une trajectoire

Bien que la trajectoire et l'intensité d'une tempête puissent sembler interdépendants, une tentative de modélisation du moment d'arrêt de la trajectoire indépendamment de l'intensité de la tempête est proposée par la suite. Autrement dit, une analyse plus fine des trajectoires qui entrent en contact avec la terre est réalisée afin de déterminer leur moment d'arrêt.

Pour déterminer le moment d'arrêt d'une trajectoire de tempête après son atterrissage sur le continent américain, l'observation historique sur la carte de la figure 3.1 révèle que ces trajectoires passent par des régions telles que la mer des Caraïbes, le Golfe du Mexique ou l'Océan Atlantique avant d'atteindre des côtes terrestres, notamment en Amérique centrale, au Mexique ou sur la côte sud-est des États-Unis, comme en Floride. Durant cette traversée terrestre, la tempête parcourt une certaine distance avant de soit regagner l'océan, soit se dissiper complètement sur les terres. Afin de modéliser ce phénomène, une approche consiste à délimiter une zone terrestre maximale dans laquelle la tempête pourra se déplacer une fois qu'elle atterrit sur une côte. Cette zone permettra à la tempête de poursuivre sa trajectoire sur les terres du continent américain si elle réussit à sortir de cette zone, ou de regagner l'océan si elle ne parvient pas à atteindre l'extrémité de la zone. La construction de cette zone terrestre maximale s'effectue en suivant les étapes suivantes.

Dans un premier temps, un masque Terre-Océan est créé pour identifier les tempêtes qui atterissent sur une côte terrestre, également appelé *landfall* en anglais. De plus, il est observé historiquement

qu'une tempête traversant les Caraïbes a tendance à continuer son parcours vers le continent américain. Par conséquent, l'intérêt est porté uniquement sur les tempêtes qui atterrissent effectivement sur le continent américain. Pour cette étape, un masque Terre-Océan en résolution 0.1° ($\approx 11.1\text{km}$) est construit et représenté sur la figure 3.10.

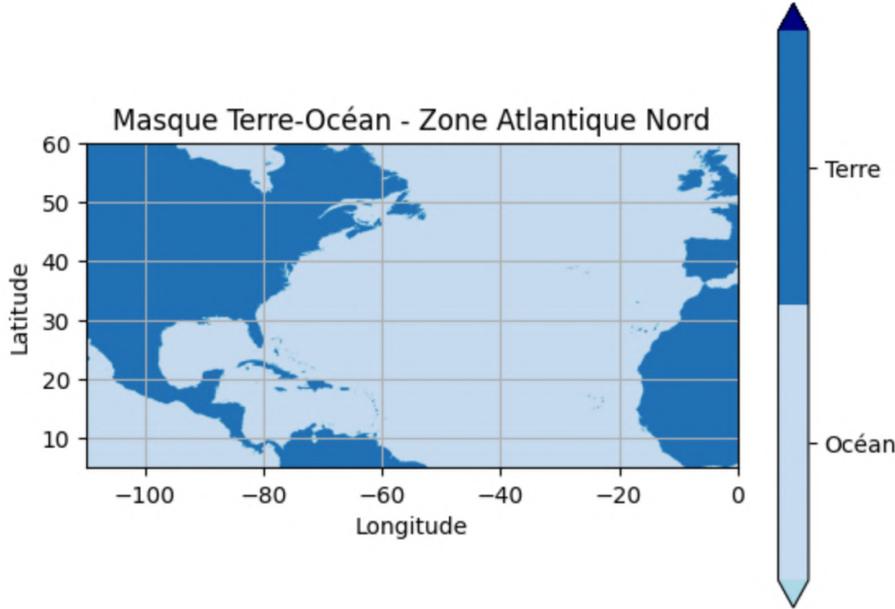


FIGURE 3.10 : Masque Terre-Océan en résolution 0.1°

L'étape suivante consiste à déterminer la distance maximale notée d_{max} entre la trajectoire de chaque tempête historique et la côte terrestre la plus proche. Il est important de préciser que d_{max} ne représente pas la distance maximale parcourue par une tempête depuis la côte terrestre la plus proche, mais plutôt la distance maximale atteinte par la tempête à partir de cette côte. Pour cela, tous les points d'impact des tempêtes historiques sur le continent américain sont collectés. Cela permettra d'obtenir un échantillon des distances maximales atteintes par chaque tempête historique lors de sa traversée sur les terres américaines.

La figure 3.11 illustre la répartition des points d'impact des tempêtes historiques sur le continent américain au cours de la période allant de 1982 à 2022. Sur cette période, la base de données IBTrACS enregistre un total de 649 tempêtes historiques. Parmi celles-ci, 280 tempêtes ont effectivement touché les côtes du continent américain, générant un total de 3597 points d'impact enregistrés.

Afin de démontrer le processus de calcul de la distance maximale pour une tempête historique donnée, les coordonnées des côtes du continent américain, allant du Panama à la province canadienne de la Nouvelle-Écosse, sont récupérées à partir du site officiel de Natural Earth, une source de bases de données cartographiques (NATURAL EARTH, 2023). Pour chaque point d'impact de la tempête, la distance minimale jusqu'à la côte la plus proche est calculée. Cela permet d'obtenir un échantillon de distances minimales pour chaque point d'impact de la tempête par rapport à la côte la plus proche. Enfin, la distance maximale est déterminée en prenant la valeur maximale parmi toutes ces distances minimales, représentant ainsi la distance maximale atteinte par une tempête depuis la côte terrestre la plus proche. Un exemple concret de calcul de d_{max} pour les points d'impact d'une trajectoire historique issue de la base de données IBTrACS est illustré à travers la figure 3.12.

À l'issue de cette étape, un échantillon de 280 valeurs, notée $\mathcal{D} = (d_{max,1}, \dots, d_{max,280})$, est obtenu. Cet échantillon contient les distances maximales atteintes par toutes les tempêtes historiques jusqu'à la côte terrestre la plus proche parmi tous leurs points d'impacts sur le continent américain.

La troisième étape de cette étude consiste à ajuster une loi de probabilité sur l'échantillon \mathcal{D} obtenu



FIGURE 3.11 : Points d'impact sur le continent américain des tempêtes historiques (1982-2022)



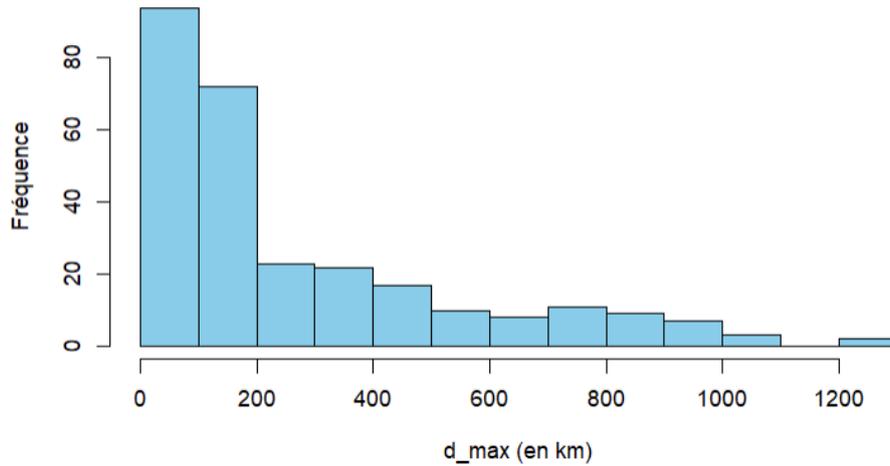
FIGURE 3.12 : Exemple de calcul de d_{max}

à l'étape précédente. Pour cela, l'histogramme issu de l'échantillon des distances maximales calculées précédemment est tracé, comme illustré sur la figure 3.13.

La distribution des distances maximales atteintes par les tempêtes historiques diminue progressivement à mesure que la distance augmente. Cette observation est cohérente étant donné que rares sont les tempêtes ayant atteint une grande distance depuis la côte terrestre la plus proche. Afin de modéliser cette fréquence, une loi gamma a été choisie pour plusieurs raisons. Tout d'abord, la forme de la distribution observée sur l'histogramme 3.13 montre une décroissance à mesure que la distance augmente, et semble se modéliser par une loi gamma. La méthode de calibration utilisée est celle du maximum de vraisemblance. Les résultats de la calibration par une loi gamma sont présentés sur la figure 3.14 et le tableau 3.6.

Distribution	Log-vraisemblance	Paramètres estimés
$\mathcal{G}(\alpha, \beta)$	-1830.6	$\hat{\alpha}_{EMV} = 0.893$; $\hat{\beta}_{EMV} = 0.003$

TABLE 3.6 : Résultats obtenus par maximum de vraisemblance

FIGURE 3.13 : Histogramme de l'échantillon \mathcal{D}

Les résultats de la calibration de la loi gamma aux données de distances maximales atteintes par les tempêtes historiques montrent un ajustement satisfaisant tant au niveau de la densité que de la fonction de répartition. L'histogramme de l'échantillon des distances maximales (figure 3.13) suggère une décroissance de la fréquence des distances à mesure que celles-ci augmentent, ce qui est cohérent avec les caractéristiques des tempêtes historiques. La loi gamma, avec sa queue longue, semble bien capturer cette décroissance et constitue donc une approximation raisonnable pour modéliser la distribution complexe des distances maximales.

Le P-P plot présente une bonne concordance entre les probabilités observés et les probabilités théoriques de la loi gamma, indiquant une adéquation raisonnable entre la distribution empirique des distances maximales et la distribution théorique de la loi gamma.

Cependant, le QQ-plot révèle un ajustement moins satisfaisant pour les quantiles extrêmes. Cela peut s'expliquer par le fait que la loi gamma pourrait ne pas être la meilleure approximation pour modéliser les valeurs extrêmes des distances maximales. D'autres distributions ou modèles plus appropriés pourraient être envisagés pour mieux représenter ces quantiles extrêmes.

En conclusion, la calibration de la loi gamma aux données de distances maximales des tempêtes historiques montre globalement un bon ajustement, mais des améliorations pourraient être apportées pour modéliser les quantiles extrêmes de manière plus précise. Il est essentiel de souligner qu'une modélisation statistique par une loi gamma de la fréquence des distances maximales atteintes par les tempêtes historiques reste une simplification de la réalité. D'autres distributions ou modèles pourraient être envisagés en fonction des spécificités des données et des connaissances sur le processus physique sous-jacent. Il est donc important d'approfondir l'analyse et d'examiner d'autres alternatives avant de conclure définitivement sur le choix de la loi gamma pour cette modélisation. Dans le cadre de cette étude, la loi gamma semble être une approximation raisonnable pour décrire cette distribution complexe et sera donc retenue par la suite.

Afin de quantifier formellement l'adéquation de la loi gamma à l'échantillon, un test de Kolmogorov-Smirnov a été réalisé. Dans ce contexte, le test vérifie si l'échantillon de distances maximales suit effectivement une loi gamma.

Test	Distribution	Statistique de Test	P-value
KS	$\mathcal{G}(\hat{\alpha}_{EMV}, \hat{\beta}_{EMV})$	0.07	0.67

TABLE 3.7 : Résultats du test d'adéquation de Kolmogorov-Smirnov

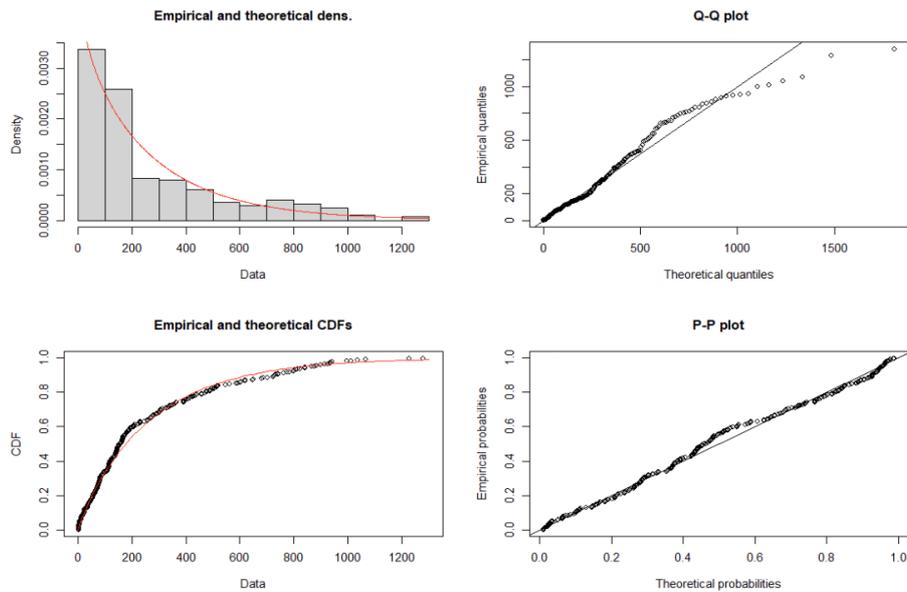


FIGURE 3.14 : Résultats de la calibration par loi gamma

La p-value étant supérieure au niveau de signification fixée à 5%, il n'y a pas suffisamment de preuves pour rejeter l'hypothèse nulle. Par conséquent, selon ce test, il n'y a pas de preuves significatives pour affirmer que les distances maximales observées et les données simulées proviennent de distributions différentes.

En conclusion, l'ajustement satisfaisant de la loi gamma aux données de distances maximales atteintes par les tempêtes historiques suggère que cette distribution constitue une approximation raisonnable pour modéliser la complexité de ces distances. Toutefois, il est important de garder à l'esprit que cette modélisation reste une simplification de la réalité, et qu'il convient d'explorer d'autres alternatives et d'approfondir l'analyse pour obtenir une représentation encore plus précise du phénomène étudié. La loi gamma demeure cependant un choix pertinent et sera retenue pour la suite de l'étude.

Après avoir effectué les étapes mentionnées précédemment, la détermination de la zone terrestre maximale pour une tempête qui atteint une côte du continent américain suit la procédure suivante. Dans un premier temps, un tirage aléatoire est réalisé pour obtenir une réalisation de la loi gamma avec des paramètres $\hat{\alpha}_{EMV} = 0.893$ et $\hat{\beta}_{EMV} = 0.003$. Cette étape permet d'obtenir la distance maximale en kilomètres que la tempête peut atteindre à partir de la côte la plus proche. Ensuite, la zone terrestre est construite en calculant la position maximale que la tempête pourrait atteindre à partir de cette distance maximale. Pour ce faire, la distance maximale est convertie en degrés². Ainsi, un quadrillage est créé pour représenter la zone terrestre maximale, où les extrémités de chaque diagonale correspondent à la position initiale d'atterrissage de la tempête et à la position maximale atteignable par la tempête. La tempête est considérée comme étant « morte » dès qu'elle sort de cette zone terrestre maximale déterminée.

3.1.3.5 Résultats et qualité d'ajustement du modèle

En utilisant la méthode de simulation d'une trajectoire préalablement exposée et la méthode de fin de trajectoire d'une tempête entrant sur terre décrite dans la section 3.1.3.4, il devient envisageable de générer des trajectoires de tempêtes au nombre de 157 689 pour les 10 000 saisons cycloniques. Pour illustrer cette démarche, la figure 3.15 présente un échantillon de 50 trajectoires obtenues par

²1° \approx 111km

simulation ainsi que 50 trajectoires tirées au hasard dans la base de données empiriques IBTrACS. Ces exemples permettent de comparer et d'évaluer la pertinence des trajectoires simulées en les confrontant aux données réelles.



FIGURE 3.15 : Echantillon de 50 trajectoires empiriques (à gauche) et simulées (à droite)

Plusieurs observations importantes peuvent être formulées suite à ces résultats. Tout d'abord, les trajectoires simulées présentent une bonne concordance avec l'aspect parabolique typique des trajectoires de tempêtes historiques.

Cependant, comparativement à l'approche markovienne, l'approche basée sur les séries temporelles offre certains avantages. L'un des principaux atouts de cette approche réside dans sa capacité à générer des mouvements de trajectoires de manière aléatoire tout en conservant les caractéristiques observées empiriquement. Contrairement à l'approche markovienne, les trajectoires simulées ne présentent pas de changements brusques de direction, car il n'y a pas de quadrillage du bassin océanique dans cette méthode. De plus, cette approche permet une simulation avec une résolution plus précise, atteignant jusqu'à 0.1° .

Néanmoins, il est important de souligner quelques limitations de cette approche. Tout d'abord, la détermination du moment d'arrêt de la trajectoire pose un défi. La simulation de la distance maximale atteignable par une tempête dès son atterrissage sur une côte terrestre constitue une approximation grossière pour déterminer la fin de la trajectoire. De plus, la modélisation de ce phénomène à travers une loi gamma suscite certaines incertitudes, étant donné sa simplicité. Bien que la distance parcourue par une tempête sur terre joue un rôle crucial dans sa dissipation et son affaiblissement, l'approche actuelle ne permet pas une modélisation précise de sa disparition. Une approche plus adéquate serait de considérer l'intensité de la tempête au moment de son atterrissage sur une côte terrestre, comme expliqué dans l'étape 4 de la méthode de simulation en section 3.1.3.3. Cette piste est actuellement envisagée pour améliorer la modélisation des trajectoires de tempêtes et traitée dans la section 3.2 dédiée à la modélisation de l'intensité des tempêtes.

Les résultats obtenus pour l'ajustement de nos modèles de trajectoires de tempêtes aux données empiriques sont représentés à travers le tableau 3.8 et sont très encourageants. Le coefficient de détermination R^2 est un indicateur clé de la qualité de l'ajustement d'un modèle aux données observées. Dans le contexte de l'étude, tous les R^2 obtenus sont supérieurs à 0.8, ce qui suggère que plus de 80% de la variance des données est expliquée par nos modèles. Cette forte corrélation entre les trajectoires simulées et les données réelles démontre la capacité de nos modèles à reproduire fidèlement le comportement des tempêtes historiques.

Pour évaluer la validité statistique des résultats obtenus, des tests de significativité ont été effectués. La p-valeur globale du test de Fisher, qui permet d'évaluer si l'ajustement global du modèle est significatif, est inférieure à 0.01 pour tous les cas. Cela signifie que les modèles sont statistiquement significatifs et qu'ils s'ajustent de manière satisfaisante aux données empiriques.

Section de 5°	R_ξ^2	R_ϕ^2	p_ξ^{global}	p_ϕ^{global}
[55° ; 60°]	0.87	0.84	< 0.01	< 0.01
[50° ; 55°[0.88	0.89	< 0.01	< 0.01
[45° ; 50°[0.85	0.89	< 0.01	< 0.01
[35° ; 40°[0.92	0.92	< 0.01	< 0.01
[35° ; 40°[0.94	0.92	< 0.01	< 0.01
[30° ; 35°[0.93	0.89	< 0.01	< 0.01
[25° ; 30°[0.92	0.89	< 0.01	< 0.01
[20° ; 25°[0.92	0.87	< 0.01	< 0.01
[15° ; 20°[0.92	0.86	< 0.01	< 0.01
[10° ; 15°[0.92	0.84	< 0.01	< 0.01
[5° ; 10°[0.89	0.83	< 0.01	< 0.01

TABLE 3.8 : Résultats de la qualité d'ajustement

Cependant, les résultats des tests d'homoscédasticité et d'autocorrélation des résidus ne permettent pas de parvenir à des conclusions définitives. Ces observations suggèrent que les modèles ne satisfont pas entièrement aux hypothèses d'homoscédasticité (constance de la variance des résidus) et d'absence d'autocorrélation des résidus. Ces résultats indiquent que les modèles pourraient présenter certaines lacunes dans la représentation de certaines caractéristiques des données empiriques.

Malgré ces résultats mitigés, des analyses complémentaires ont été menées pour mieux comprendre la nature des résidus, présentées en annexe A.4. L'analyse des histogrammes des résidus montre qu'ils sont bien centrés autour de zéro. Cette observation suggère que les prédictions du modèle sont en moyenne proches des valeurs réelles observées, ce qui implique une absence de biais systématique dans les prédictions. De plus, l'analyse des fonctions d'autocorrélation et d'autocorrélation partielle des résidus révèle une diminution progressive à mesure que le décalage augmente. Cette baisse significative des fonctions d'autocorrélation et d'autocorrélation partielle vers zéro indique une faible ou nulle corrélation entre les résidus à différents points dans le temps. Ce résultat positif suggère que le modèle a bien capturé la plupart des dépendances temporelles présentes dans les données.

Ces observations amènent à interpréter les résultats des tests d'homoscédasticité et d'autocorrélation avec prudence. Bien que ces tests n'aient pas abouti à des résultats significatifs, les propriétés des résidus identifiées à travers les histogrammes et les fonctions d'autocovariance indiquent que nos modèles parviennent à capturer certaines caractéristiques importantes des données empiriques.

Sur la base de ces constatations, les conclusions suivantes concernant la qualité des résidus sont établies et utilisées dans le cadre de l'étude :

- **Absence d'autocorrélation** : Les faibles valeurs des fonctions d'autocorrélation et d'autocorrélation partielle indiquent que les résidus ne présentent pas de corrélation significative avec leurs propres valeurs décalées ;
- **Centré** : Le fait que les résidus aient une moyenne proche de zéro indique que, en moyenne, les prédictions du modèle ne sont pas biaisées et ne surestiment ou sous-estiment pas systématiquement les données ;
- **Homoscédasticité** : La dispersion des résidus est raisonnablement constante dans l'ensemble des valeurs prédites ;
- **Gaussianité** : Les résidus sont supposés être approximativement gaussiens.

3.1.4 Comparaison des méthodes et pistes d'améliorations

Dans cette section, une analyse comparative des différentes approches de simulation de trajectoires est réalisée, c'est-à-dire celles envisagées par les chaînes de Markov paramétrique et non-paramétrique et celle envisagée par les séries temporelles.

Chacune de ces méthodes présente des caractéristiques distinctes et des avantages spécifiques pour la modélisation des trajectoires de tempêtes. Les chaînes de Markov sont reconnues pour leur capacité à capturer les transitions discrètes entre les états et à modéliser les déplacements stochastiques. D'autre part, les séries temporelles offrent une perspective temporelle continue qui peut être utile pour représenter les variations à long terme et les tendances observées dans les données. Sur la figure 3.16, les trajectoires simulées des trois méthodes de simulation et les trajectoires empiriques sont représentés afin de pouvoir confronter les différentes méthodes.

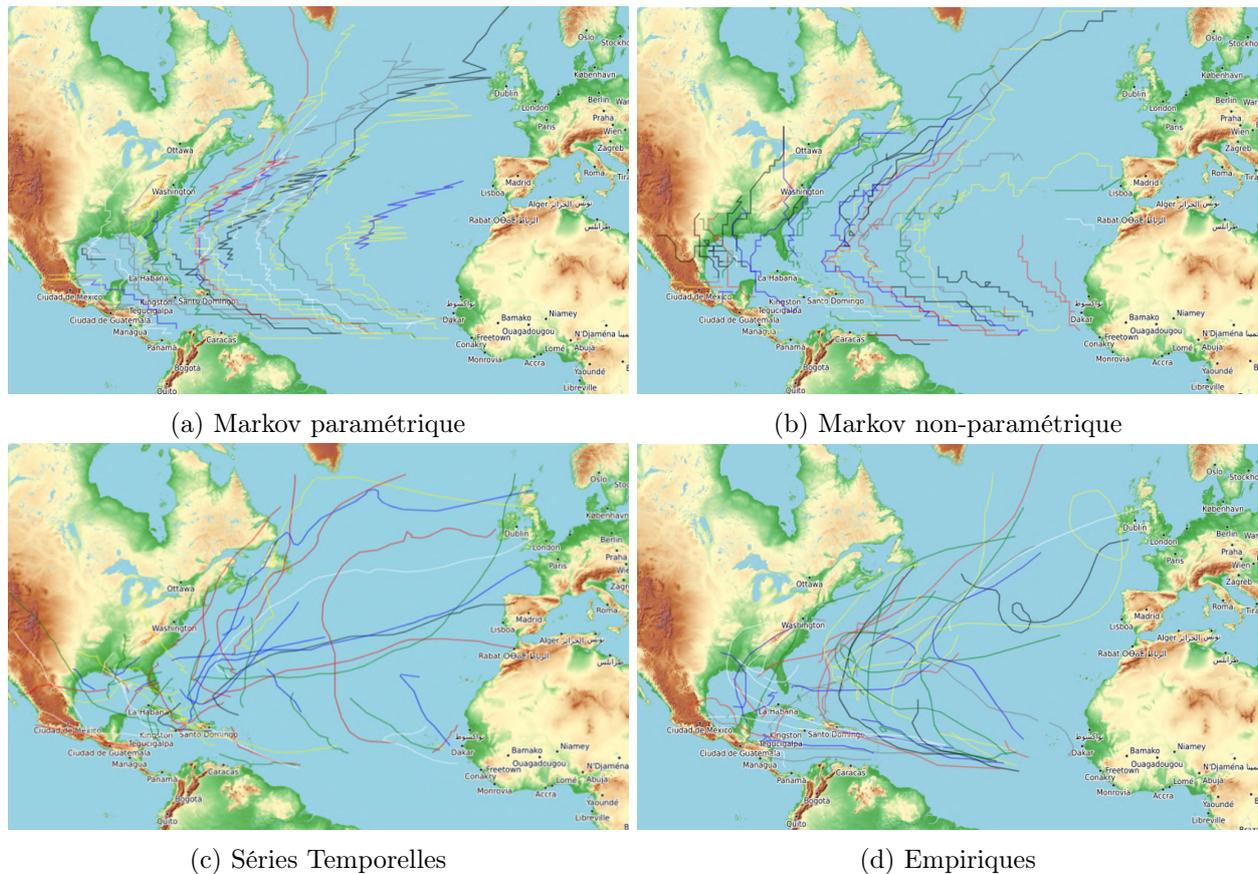


FIGURE 3.16 : Comparaison des trois méthodes de simulation de trajectoires

L'observation graphique révèle que la modélisation par les séries temporelles produit des trajectoires qui présentent une proximité spatiale remarquable avec celles observées historiquement par rapport aux deux autres méthodes. Ces trajectoires parviennent à reproduire de manière réaliste les mouvements des tempêtes tout en introduisant de la diversité dans leurs déplacements. Ainsi, l'approche retenue pour la modélisation des trajectoires des 10 000 saisons cycloniques est celle reposant sur les séries temporelles, suivant toutes les étapes décrites dans la méthode de simulation en section 3.1.3.3.

Une validation supplémentaire aurait pu consister à quantifier les écarts entre les trajectoires prédites et observées afin de comparer les performances des différentes approches. Des mesures de distance telles que la distance euclidienne ou la déformation temporelle dynamique, connue sous le nom de *Dynamic Time Warping* (DTW), auraient pu être envisagées à cet effet (YAGUANG et al., 2021).

Malheureusement, ces validations n'ont pas été réalisées dans le cadre de cette étude en raison de contraintes de temps. Cependant, elles peuvent fournir des informations pertinentes pour évaluer la robustesse des méthodes envisagées.

De plus, dans le cadre de cette étude, une approche qui modélise les séries longitudinales et latitudinales de manière autonome est adoptée. Cette approche a montré sa capacité à capturer les caractéristiques des trajectoires de tempêtes de manière réaliste. Cependant, il est important de reconnaître qu'elle repose sur l'hypothèse d'indépendance entre les séries longitudinales et latitudinales ainsi que sur l'indépendance entre les bruits blancs associés à ces séries représentés à travers les équations (3.5) et (3.6). Cette hypothèse d'indépendance peut être considérée comme une simplification importante du processus sous-jacent.

Il est tout à fait plausible qu'il existe des co-mouvements entre les variations longitudinales et latitudinales des tempêtes, et il n'est pas nécessairement vrai que les bruits blancs associés à ces séries soient indépendants entre eux. Par conséquent, une perspective intéressante pour une amélioration future de cette approche serait d'explorer des modèles plus complexes, tels que les modèles à vecteurs autorégressifs (VAR) (HAMILTON, 1994). Les modèles VAR permettraient de capturer les relations de dépendance entre les deux séries tout en prenant en compte les interactions potentielles entre les bruits blancs.

L'adoption d'un modèle VAR pourrait ainsi offrir une représentation plus fidèle du processus de formation des trajectoires de tempêtes en prenant en compte les relations entre les dimensions longitudinales et latitudinales. Cela pourrait permettre de mieux saisir les phénomènes complexes qui sous-tendent les déplacements des tempêtes cycloniques et d'améliorer la précision des prédictions futures. Cette perspective de modélisation plus sophistiquée représente un domaine de recherche prometteur pour des études futures dans le domaine de la modélisation des tempêtes cycloniques.

Après avoir abordé en détail la modélisation des trajectoires des tempêtes, la section suivante se penche sur l'évolution de leur intensité tout au long de leur parcours.

3.2 Modélisation de l'intensité

L'étude de l'intensité des tempêtes tropicales et de leur évolution temporelle constitue un axe de recherche essentiel dans le domaine météorologique (CHEN et al., 2020). Ces phénomènes, dont l'ampleur peut engendrer des conséquences dévastatrices, requièrent une compréhension approfondie de leurs caractéristiques fondamentales. Dans cette étude, l'attention est portée sur quatre caractéristiques : pression centrale, vitesse maximale des vents, rayon de la vitesse maximale des vents et gradient de vitesse des vents. Ces quatre variables sont parmi les caractéristiques les plus cruciales pour décrire la puissance et l'intensité d'un ouragan. Elles sont fréquemment utilisées dans le domaine climatologique pour comprendre et prédire leur comportement (CHEN et al., 2020). Des études (ELSNER et KARA, 1999) ont démontré que ces variables sont étroitement liées à l'évolution des ouragans. Par exemple, la pression centrale et la vitesse maximale des vents sont des indicateurs directs de leur puissance, comme le montre leur utilisation à travers l'échelle de Saffir-Simpson sur le tableau 1.1. De plus, ces variables sont généralement bien documentées et largement disponibles grâce aux observations météorologiques, aux satellites et aux données de modèles numériques (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2019), comme expliqué en section 1.6.4. Cela facilite leur utilisation dans les modèles d'intensité. Et enfin en se concentrant sur ces quatre variables clés, il est possible de simplifier le modèle tout en capturant les principaux facteurs qui influencent l'intensité de l'ouragan. Cela rend le modèle plus interprétable et plus facile à utiliser (VICKERY et al., 2000).

L'objectif de cette section est d'introduire une méthodologie visant à modéliser ces quatre caractéristiques majeures. Dans un premier temps, un panorama des modèles statistiques préexistants dans la littérature scientifique est exposé pour chaque caractéristique. Cette exploration permettra de situer l'approche utilisée au sein de la littérature scientifique.

Par la suite, la mise en œuvre des modèles choisis pour chaque caractéristique sera détaillée. Chaque phase de modélisation sera présentée, incluant les hypothèses sous-jacentes, les paramètres sélectionnés, ainsi que les résultats à travers une analyse approfondie, permettant de mesurer l'adéquation de l'approche aux données empiriques. À l'issue de cette section, chaque trajectoire de tempête simulée à la section précédente 3.1 est associée aux caractéristiques modélisées spécifiques à l'évolution de son intensité.

3.2.1 Modélisation de la pression centrale

La caractéristique fondamentale qui détermine la puissance d'une tempête tropicale réside dans son déficit de pression, défini comme l'écart entre la pression atmosphérique environnante et la pression au centre de la tempête, mesurée au niveau de la mer. L'ampleur de ce déficit de pression est directement proportionnelle à l'intensité dévastatrice de la tempête. La littérature scientifique recense diverses approches de modélisation de la pression centrale. Ces modèles se distinguent notamment en fonction de la zone géographique considérée, qu'elle soit océanique ou terrestre.

En effet, une distinction pertinente se dessine lors de l'analyse de l'évolution de la pression au sein d'une tempête, à l'exemple de l'ouragan Irma 1.6 (section 1.2.3.2) illustré sur la figure 3.17. Il est ainsi observable que la pression chute rapidement lorsque l'ouragan évolue au-dessus de l'océan, tandis qu'elle augmente dès qu'il atteint une zone terrestre. Une légère variation apparaît également entre les interactions avec différentes étendues terrestres, notamment lorsque l'ouragan frôle ou touche des territoires de petites superficies tels que les îles et les états insulaires de la région des Caraïbes. Dans ces cas, bien que la pression augmente lors du passage sur les terres, cette augmentation n'est généralement pas suffisante pour compenser la chute observée pendant son parcours océanique. L'augmentation de pression sont particulièrement notables lorsqu'une tempête entre en contact avec une zone terrestre plus étendue, comme illustré lors de l'atterrissage de l'ouragan Irma en Floride le 10 septembre.

3.2.1.1 Etat de l'art des modèles statistiques

Dans l'environnement océanique, divers modèles statistiques ont été développés par des spécialistes. Parmi les plus courants, le modèle proposé par DEMARIA et KAPLAN (1994) se distingue. Ces chercheurs ont établi une relation empirique entre la température de surface de la mer et l'intensité maximale d'une tempête. Cette relation permet de déterminer le *Maximum Possible Intensity* (MPI) qui correspond à la valeur minimale de la pression centrale qu'une tempête pourrait atteindre à un moment donné. En d'autres termes, le MPI représente le degré d'intensité maximale qu'une tempête pourrait potentiellement atteindre à un instant donné. Dans leur modélisation, la vitesse maximale des vents est utilisée comme indicateur d'intensité maximale de la tempête. Cependant, il est également envisageable d'employer le déficit de pression comme mesure d'intensité maximale, étant donné que ces deux caractéristiques sont étroitement liées (ce lien est expliqué en détail dans la section dédiée à la modélisation de la vitesse maximale des vents 3.2.2.1). Dans leurs travaux, BLOEMENDAAL et al. (2020) ont opté pour l'utilisation du déficit de pression.

Un autre modèle largement répandu dans la littérature pour la modélisation de la pression centrale dans le contexte océanique est celui élaboré par VICKERY et al. (2000), se fondant sur le concept d'intensité relative. Cette approche s'impose comme la méthode la plus fréquemment adoptée pour simuler les risques liés aux ouragans au sein de la communauté scientifique (CUI et CARACOGIA, 2019) et a également été appliqué dans des mémoires d'actuariat (DENISE-BAILLON, 2013; YOMBI, 2016). L'idée d'intensité relative a d'abord été développée par DARLING (1991), dans son modèle visant à prédire les vitesses de vent des tempêtes cycloniques du bassin Atlantique Nord. Il définit l'intensité relative comme le rapport entre la pression centrale effective et la pression centrale minimale théoriquement atteignable, compte tenu des conditions thermodynamiques en vigueur à un moment donné durant une tempête. Cette grandeur théorique varie entre 0 et 1, et plus elle se rapproche de 1,

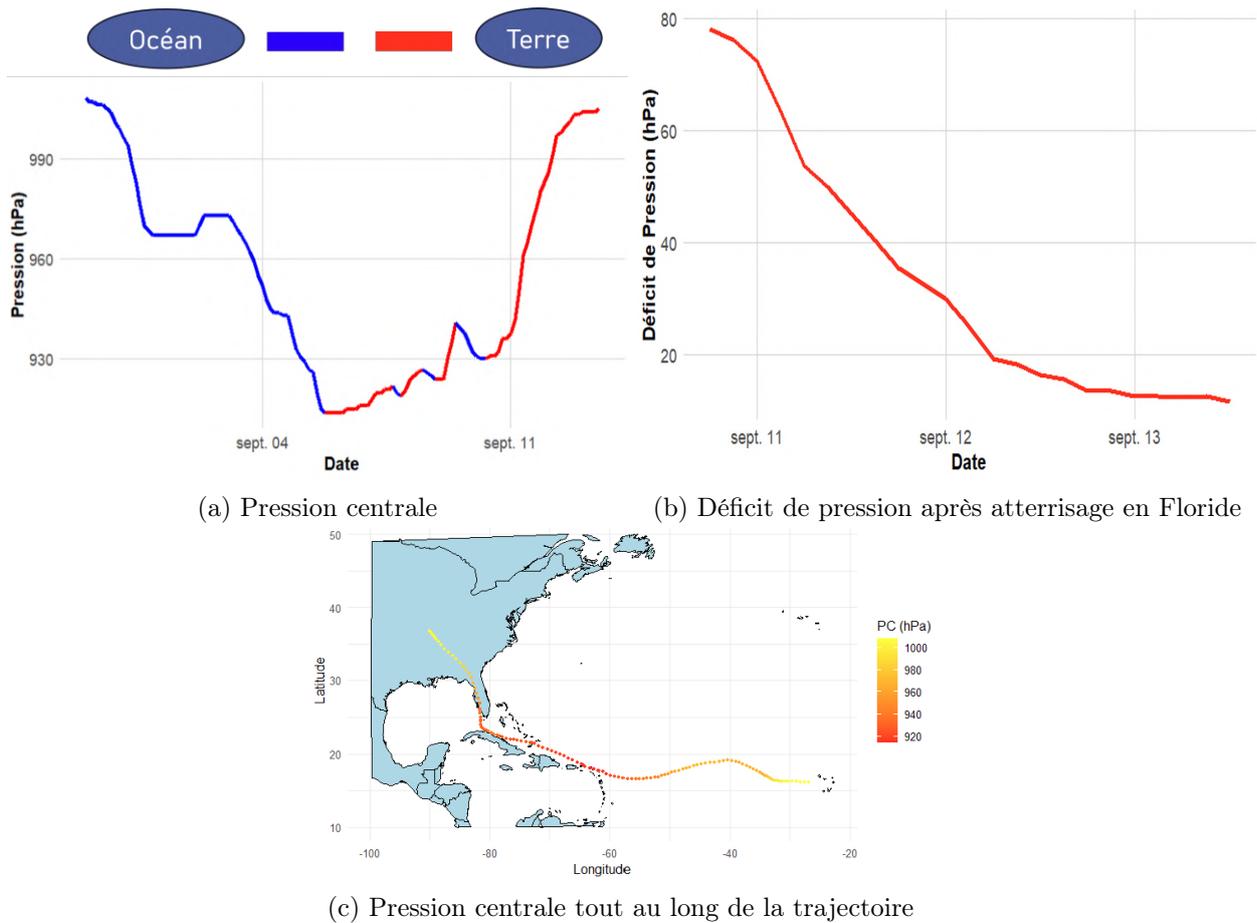


FIGURE 3.17 : Evolution de la pression centrale de l'ouragan Irma

plus l'ouragan est potentiellement destructeur. Plusieurs chercheurs continuent de recourir au modèle proposé par VICKERY et al. (2000) en raison de sa complexité et de sa capacité à prendre en compte diverses conditions environnementales et thermodynamiques. À titre d'exemple, ce modèle a été adapté à d'autres bassins océaniques pour prédire les vitesses de vent des typhons le long de la côte nord-ouest du Pacifique (HONG et al., 2016).

Dans l'environnement terrestre, l'un des premiers modèles proposés pour appréhender l'évolution de la pression centrale des tempêtes a été élaboré par BATTIS, M.E. et al. (1980). Ces chercheurs se sont attelés à modéliser le déficit de pression sur les terres en fonction du temps passé sur le sol et de l'angle formé entre l'ouragan et la ligne côtière au point d'atterrissage. Cependant, ce modèle présente des difficultés à saisir la chute très rapide du déficit de pression lors de l'atterrissage. C'est dans cette optique que KAPLAN et DEMARIA (1995) ont introduit une correction à ce modèle, en incorporant un terme exponentiel capable de prendre en considération cette décroissance très prompte du déficit de pression après l'atterrissage. Ce modèle continue d'être utilisé très largement au sein de la communauté scientifique en raison de sa simplicité et de son efficacité (CUI et CARACOGIA, 2019; LIU et PANG, 2011). Un exemple illustrant de manière explicite ce choix de modèle, accompagné de détails supplémentaires dans la section dédiée à la modélisation de la pression centrale terrestre (section 3.2.1.3), est présenté dans la figure 3.17. Cette figure expose l'évolution du déficit de pression de l'ouragan Irma après son impact sur les côtes de la Floride le 10 septembre 2017. Ce schéma illustre une tendance typique de ces phénomènes cycloniques, caractérisée par une chute extrêmement rapide du déficit de pression en quelques heures, se rapprochant rapidement de zéro.

3.2.1.2 Modèle de pression centrale dans l’océan

Trois modèles ont été soumis à des tests pour la modélisation de la pression centrale en milieu océanique. La sélection s’est portée sur le modèle qui parvient à établir un équilibre satisfaisant entre la qualité d’ajustement du modèle et les performances de prédiction. Par la suite, les différentes démarches adoptées pour chaque modèle sont exposées.

En premier lieu, la modélisation proposée dans les travaux de BLOEMENDAAL et al. (2020) est mise en œuvre. Pour cela, la température moyenne mensuelle ainsi que la pression atmosphérique moyenne mensuelle sont nécessaires. Ces deux variables sont extraites de la base de données ERA-5³ (EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS, 2023). Ces variables sont disponibles avec une résolution de 0.25° . Il est choisi de collecter ces informations pour la région géographique du bassin Atlantique Nord, telle qu’il a été définie lors de l’application du modèle de trajectoire par séries temporelles (section 3.1.3), soit $[-110^\circ, 0^\circ] \times [5^\circ, 60^\circ]$. Idéalement, l’utilisation des données de température et de pression atmosphérique en temps réel serait préférable pour élaborer un modèle plus réaliste. Cependant, cette approche est souvent entravée par des contraintes liées à la capacité de traitement des données en temps réel. Par conséquent, le choix de travailler avec des moyennes mensuelles de ces données est faite, une hypothèse couramment acceptée au sein de la communauté scientifique pour la construction de modèles d’intensité.

Dans le but d’illustrer ces deux variables et de démontrer leur influence sur la pression centrale, les pressions atmosphériques moyennes des mois de janvier et de septembre sont représentées sur les figures 3.18 et 3.19. Ces deux mois sont respectivement associés à la période de fréquence cyclonique la plus basse et la plus élevée au cours de la période étudiée. P_{env} désigne la pression atmosphérique moyenne mensuelle (*Environmental Pressure*) et SST la température moyenne mensuelle à la surface de la mer (*Sea Surface Temperature*).

Ces cartes fournissent une vision claire sur l’évolution courante de l’intensité d’un ouragan. Comme évoqué dans la section 1.2.2, une région de basse pression atmosphérique au niveau de la mer, associée à une température élevée de la surface de la mer, généralement supérieure à 26°C , s’avère nécessaire pour favoriser la formation d’un ouragan. Cette tendance se manifeste de manière marquée au mois de septembre, période de l’activité cyclonique la plus soutenue dans le bassin Atlantique Nord, à l’opposé de janvier. Ces deux variables jouent donc un rôle essentiel dans l’intensification des phénomènes cycloniques.

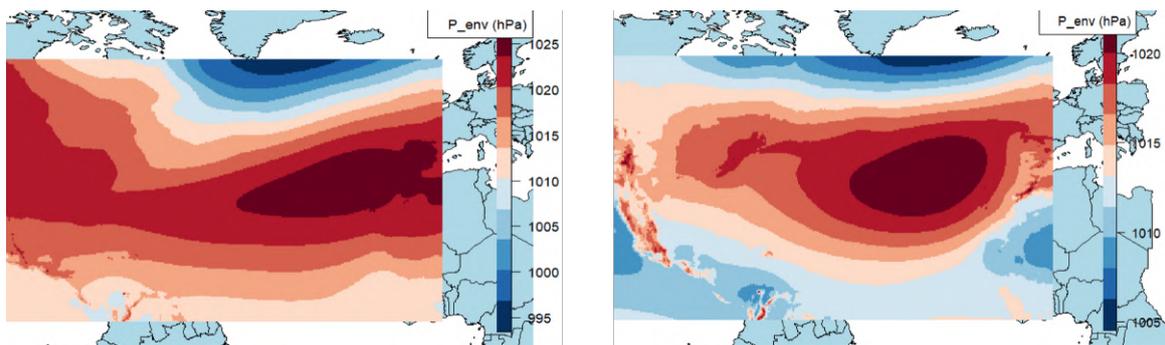


FIGURE 3.18 : Pression atmosphérique moyenne en janvier (à gauche) et septembre (à droite) entre 1982 et 2022

Ainsi l’étape initiale de la première modélisation proposée par BLOEMENDAAL et al. (2020) consiste à associer à chaque point d’impact d’une tempête cyclonique de la base IBTrACS, la température et la pression atmosphérique moyenne mensuelle correspondante. Par la suite le déficit de pression correspondante est obtenu à chaque point d’impact de la base IBTrACS, soit $P_{env} - P$ en hPa, où P

³Une base de données climatique recensant les conditions météorologiques à l’échelle mondiale depuis 1940

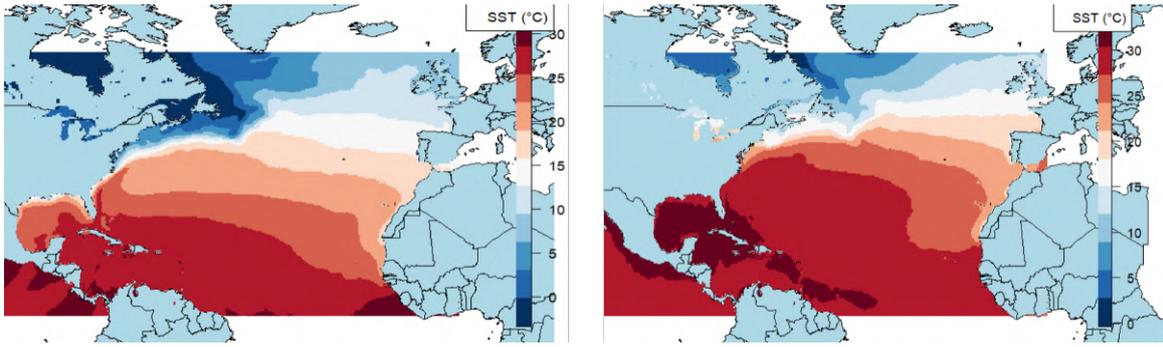


FIGURE 3.19 : Température moyenne en janvier (à gauche) et septembre (à droite) entre 1982 et 2022

représente la valeur de pression centrale enregistrée pour chaque point de la base. Une fois le déficit de pression calculée, le modèle consiste à utiliser la méthode des moindres carrés non linéaires afin d'estimer les coefficients A , B et C de la formule de régression suivante

$$P_{env} - P = A + B \exp(C(SST - T_0)), \quad T_0 = 30.0^\circ\text{C}. \quad (3.11)$$

À la suite de l'estimation des paramètres, le déficit de pression peut être calculé pour chaque point de trajectoire simulée, et le MPI est déduit en soustrayant P_{env} au déficit de pression calculée par (3.11).

Cependant, ce modèle n'a pas été retenu dans le cadre de cette étude pour diverses raisons. Tout d'abord, sa limitation réside dans son manque d'exhaustivité en termes d'informations. En effet, la tentative d'établir une relation entre le déficit de pression et la température moyenne mensuelle représente une simplification considérable de la réalité complexe des phénomènes cycloniques, qui dépendent d'une gamme bien plus étendue de conditions environnementales. Le coefficient de détermination obtenu à partir du modèle (3.11) s'avère trop faible pour rendre ce modèle utilisable par la suite. De plus, les prédictions du déficit de pression estimées ne correspondent pas de manière satisfaisante aux observations réelles.

Modèle de VICKERY et al. (2000)

Ainsi, une orientation vers une modélisation plus approfondie est mise en oeuvre. Le choix s'est porté sur le modèle élaboré par VICKERY et al. (2000), bâti sur la notion d'intensité relative. Préalablement définie, cette notion d'intensité relative représente une grandeur physique sans dimension, ancrée dans l'efficacité d'un système thermodynamique tel qu'un ouragan. Les bases mathématiques théoriques de cette notion sont explicitées dans l'annexe A.5. L'atout majeur inhérent à l'utilisation de l'intensité relative réside dans la plage de valeurs qu'elle englobe, échelonnée entre 0 et 1. Via la formule d'inversion A.5 dans l'annexe A.5, qui permet de déduire les pressions centrales, des résultats cohérents sont obtenus en comparaison avec la gamme des pressions centrales historiquement observées.

Néanmoins, il peut advenir que certaines valeurs d'intensité relative soient supérieures à 1, en particulier lorsque la tempête se déplace rapidement en direction des hautes latitudes. Ce phénomène peut également être imputé à l'utilisation des moyennes mensuelles des températures et des pressions atmosphériques, en lieu et place des valeurs instantanées au moment de l'enregistrement. Pour de tels cas marginaux, la valeur d'intensité relative est limitée à 1, conformément à DARLING (1991).

Le modèle avancé par VICKERY et al. (2000) se présente sous la forme d'une régression linéaire multiple. Chaque intervalle $[k, k + 1]$, $k \in \mathbb{N}$ correspond à une durée de 3 heures. Le modèle se définit ainsi comme suit

$$\ln(I_{k+1}) = c_0 + c_1 \ln(I_k) + c_2 \ln(I_{k-1}) + c_3 \ln(I_{k-2}) + c_4(T_{k+1} - T_k) + c_5 T_k + \varepsilon, \quad (3.12)$$

où :

- I_{k+1} est l'intensité relative en $k + 1$;
- I_k, I_{k-1}, I_{k-2} sont les intensités relatives aux instants $k, k - 1$ et $k - 2$;
- T_{k+1} et T_k sont les températures moyennes mensuelles à la surface de l'océan aux instants $k + 1$ et k et ;
- c_0, c_1, c_2, c_3, c_4 et c_5 sont les coefficients de la régression ;
- ε est un terme d'erreur.

Malgré l'avantage de ce modèle, qui rassemble plusieurs conditions environnementales jusqu'à trois intervalles temporels précédents pour prédire la valeur suivante, il faut souligner que la pression centrale est une fonction complexe intrinsèquement liée à d'autres variables environnementales. Parmi elles figurent par exemple la force de cisaillement du vent, la vortacité absolue⁴ ou encore l'humidité relative à basse et moyenne altitude (CUI et CARACOGIA, 2019). Par conséquent, le modèle adopté par VICKERY et al. (2000) se révèle approprié dans le cadre de l'étude, néanmoins il n'est pas exhaustif et présente des possibilités d'améliorations.

La construction du modèle se déroule de la manière suivante. Les coefficients de la régression, définis précédemment dans la formule (3.12), sont estimés pour des zones spécifiques de l'océan. Cette estimation est réalisée à une résolution de $5^\circ \times 5^\circ$ en utilisant un quadrillage appliqué sur le bassin Atlantique Nord. Cette approche se justifie par le fait que l'évolution de la pression centrale diffère en fonction de la zone où la tempête se trouve dans l'océan, étant donné les variations des températures et des pressions atmosphériques moyennes mensuelles observées sur les figures 3.18 et 3.19.

Le choix de cette résolution se justifie par le fait qu'un nombre suffisant d'observations est nécessaire afin d'assurer un entraînement adéquat du modèle sur les différentes zones. Le nombre d'observations suffisant est fixé à 10. Les coefficients obtenus par estimation pour chaque case du quadrillage sont associés au centre de cette case. Pour illustrer ce concept, la figure 3.20 présente le quadrillage en résolution $5^\circ \times 5^\circ$ adopté, ainsi que les zones servant à l'entraînement du modèle de régression.

Cela aboutit à un total de 143 zones (en rouge sur la figure 3.20) sur lesquelles le modèle est entraîné, tandis que 16 zones présentent une insuffisance de données. Pour ces zones (en jaune sur la figure 3.20), les coefficients attribués à la zone la plus proche sont adoptés. La détermination de cette zone la plus proche repose sur le calcul de la distance minimale entre les centres des zones environnantes, en utilisant la distance orthodromique définie dans l'équation (2.1).

Il est néanmoins remarquable que les zones comportant moins de 10 observations consistent en majorité de zones terrestres, comme le long de la côte du continent américain. De plus, les zones océaniques voisines des continents européen et africain présentent également un faible échantillon de tempêtes historiques. Pour ces situations, l'assignation des résultats de la zone la plus proche n'est pas incohérente, étant donné que les tempêtes atteignant ces régions sont souvent de faible intensité et en nombre limité. De plus, pour les zones côtières américaines, le modèle de pression sur terre (section 3.2.1.3) est appliqué dès que la tempête touche la côte, ce qui évite toute problématique dans ce contexte.

Toutefois, il est important de noter que la résolution initiale utilisée, qui est de $5^\circ \times 5^\circ$, ainsi que la condition d'un nombre minimum de 10 observations par case, peuvent ne pas suffire à garantir la robustesse du modèle. Ceci pourrait entraîner des transitions abruptes d'une case à une autre en raison d'une insuffisance de données. Afin d'améliorer la stabilité et la précision du modèle, les estimations des coefficients de régression du modèle défini en (3.12) sont soumises à un processus de lissage. Pour ce faire,

⁴Correspond à la rotation de l'air par unité de volume à l'intérieur de la tempête. Une valeur élevée de vortacité absolue indique une rotation plus forte de l'air, ce qui peut contribuer à l'intensification de l'ouragan.

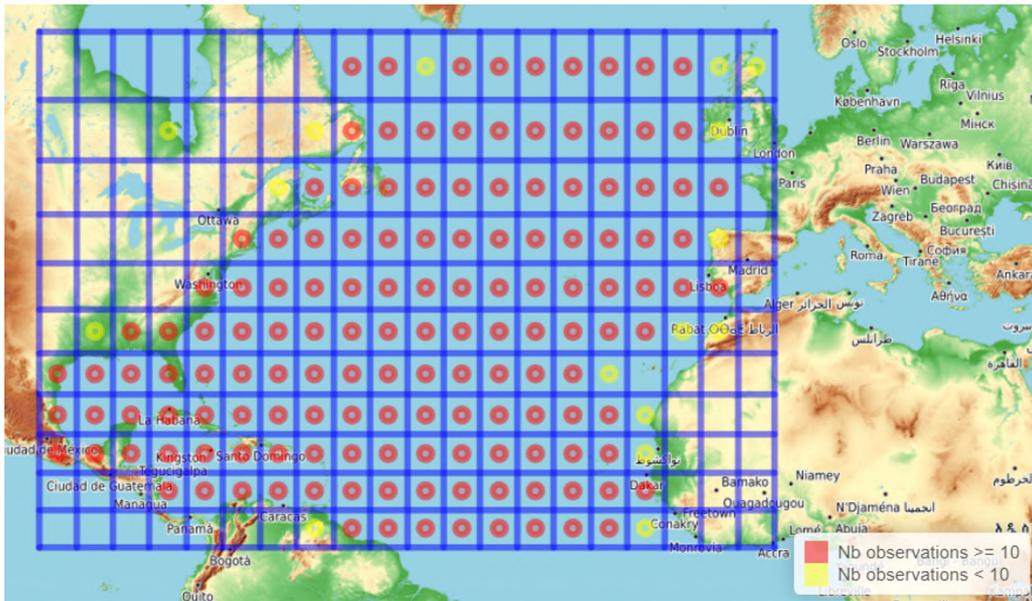


FIGURE 3.20 : Résolution $5^\circ \times 5^\circ$ pour le modèle de pression centrale océanique

une interpolation bilinéaire est réalisée à l'intérieur de chaque case définie sur le quadrillage, comme illustré sur la figure 3.20. Cette procédure permet d'obtenir des coefficients de régression en résolution $1^\circ \times 1^\circ$, offrant ainsi une meilleure adaptation aux variations locales des conditions météorologiques.

Une fois que les coefficients en résolution $1^\circ \times 1^\circ$ sont obtenus, il est ainsi possible de calculer l'intensité relative pour chaque point de trajectoire d'une tempête, puis par la formule d'inversion de DARLING (1991) défini dans l'annexe A.5 il est possible de déduire la pression centrale océanique. Une illustration du processus de simulation des pressions océaniques le long de la trajectoire de l'ouragan Irma est représenté sur la figure 3.21.

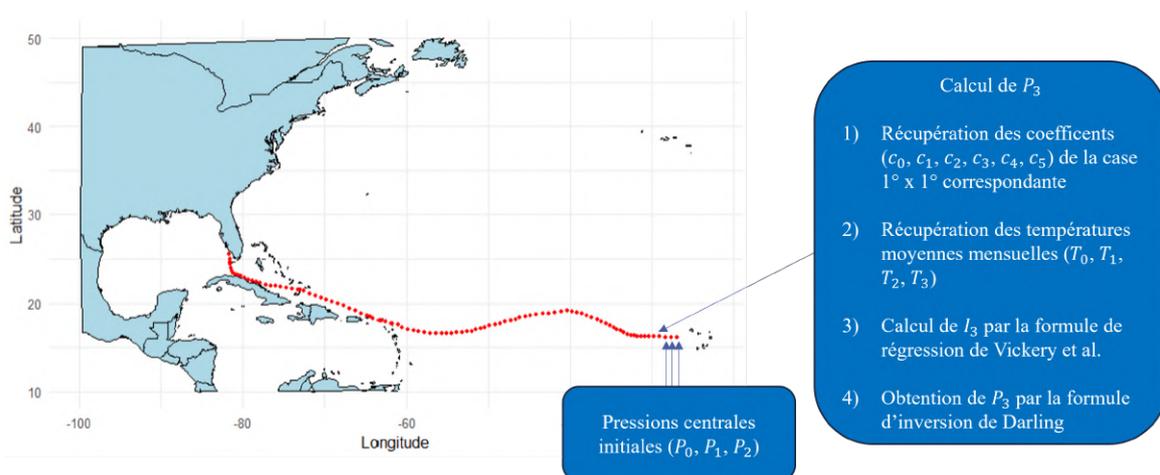


FIGURE 3.21 : Calcul des pressions centrales océaniques par le modèle de VICKERY et al. (2000)

La qualité d'ajustement de chaque modèle de régression est également réalisée. Dans le but d'obtenir une visualisation claire et concise, les résultats sont condensés à travers des graphiques illustratifs, présentés dans la figure 3.22.

Chaque graphique est attribué à une métrique de qualité d'ajustement spécifique. Ces graphiques mettent en évidence les p-valeurs issues des tests de significativité globale, d'absence d'auto-corrélation,

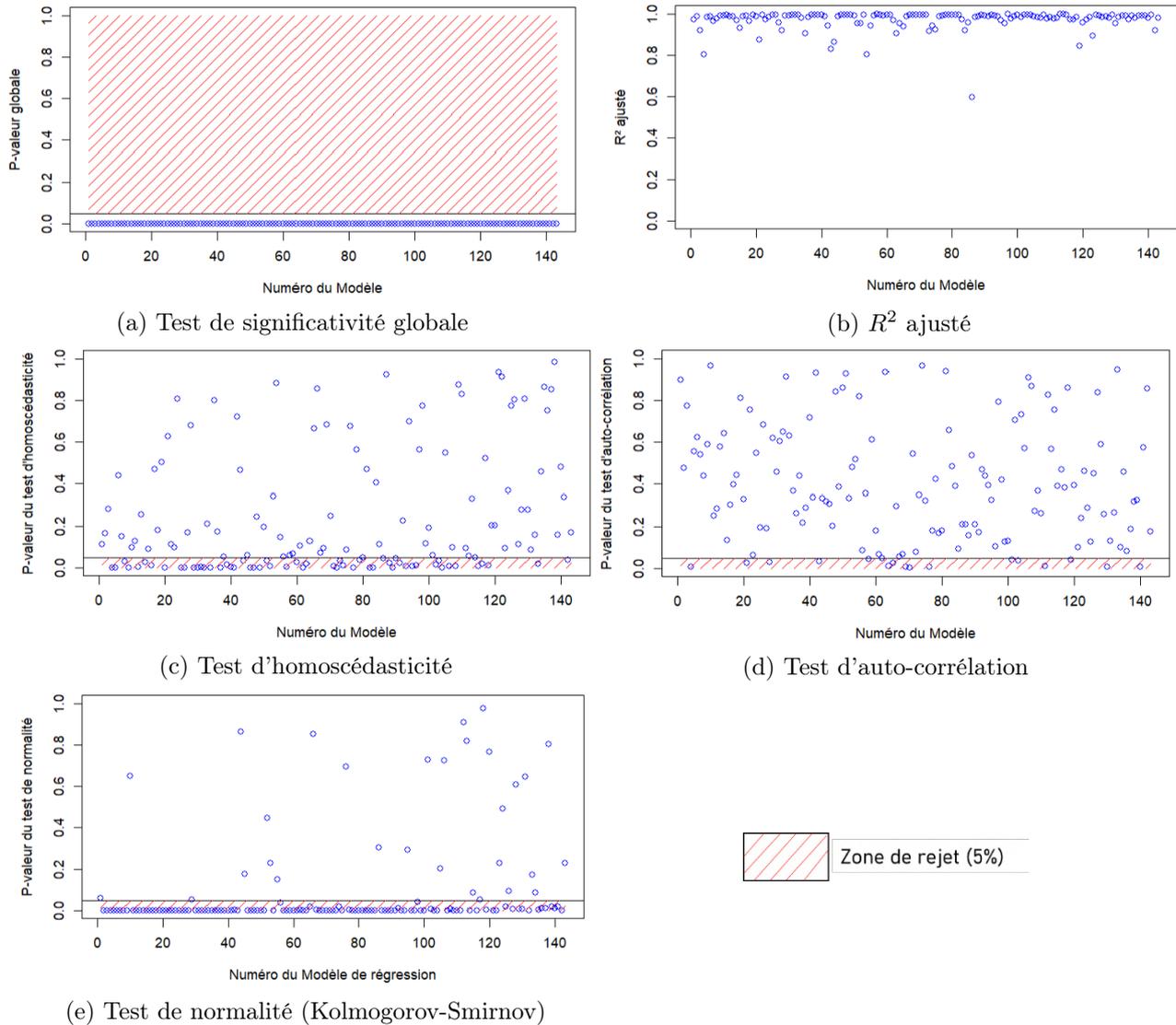


FIGURE 3.22 : Qualité d'ajustement des 143 modèles de régression proposé par VICKERY et al. (2000)

d'homoscédasticité et de normalité pour chaque modèle, accompagnées des coefficients de détermination R^2 . Les résultats exposés à travers les graphiques de la Figure 3.22 fournissent une perspective significative quant à la pertinence de ces modèles dans la représentation des données. Tout d'abord, les p-valeurs des tests de significativité globale pour chaque modèle sont toutes inférieures à 0.05, indiquant que la majorité des modèles présentent une relation significative entre les variables explicatives et la variable à expliquer.

En ce qui concerne les mesures de performance, les coefficients de détermination R^2 sont remarquablement élevés, presque tous supérieures à 0.8. Cette observation indique que les modèles ont réussi à capturer une part substantielle de la variabilité des données, renforçant ainsi leur aptitude à expliquer les variations dans la variable cible.

En ce qui concerne la vérification de certaines hypothèses fondamentales des modèles de régression, les résultats montrent que la grande majorité des p-valeurs du test de normalité sont inférieures à 0.05. Cela suggère que la distribution des résidus pour ces modèles peut ne pas être parfaitement conforme à la normalité.

Quant aux tests d'absence d'auto-corrélation et d'homoscédasticité, les résultats sont mitigés. Alors

que la majorité des modèles présentent des p-valeurs supérieures à 0.05 pour ces tests, il est important de noter qu'une minorité de modèles a des p-valeurs inférieures à ce seuil. Cela pourrait indiquer la présence d'auto-corrélation ou d'hétéroscédasticité dans les résidus de ces modèles particuliers.

Toutefois, une observation supplémentaire s'impose. Lors de l'analyse de l'histogramme et de la distribution des résidus pour chaque modèle de régression, il est remarquable qu'aucune structure particulière n'émerge et que ces résidus se trouvent centrés autour de la moyenne, conforme à une distribution normale. En conséquence, il est envisageable de considérer les résidus de chaque modèle de régression comme étant gaussiens. Par la suite les résidus de chaque modèle sont modélisées par une loi normale dont les paramètres sont égales à la moyenne et à la variance des résidus du modèle correspondant.

Afin de tester le modèle conçu par VICKERY et al. (2000), l'application de ce modèle aux données empiriques a été effectuée en considérant l'hypothèse d'un terme d'erreur aléatoire gaussien. Les trajectoires historiques, extraites de la base de données IBTrACS, ont été utilisées pour simuler les pressions centrales associées à chaque point d'impact d'une tempête. Cette démarche permet ainsi de confronter ces valeurs simulées avec les données observées historiquement, permettant de tracer une carte et un histogramme comparaison des pressions simulées et empiriques, illustrés dans la figure 3.23. Parallèlement, les statistiques descriptives correspondantes ont été intégrées dans le tableau 3.9.

L'analyse préalable des statistiques descriptives révèle une certaine cohérence entre les pressions simulées et empiriques. Notamment, la valeur minimale de pression, se situant aux alentours de 860 hPa, témoigne de la prudence inhérente au modèle de VICKERY et al. (2000) en ce qui concerne les pressions très basses. Une similitude notable émerge également entre les pressions empiriques et les valeurs simulées, ce qui signifie que le modèle parvient dans l'ensemble à capturer les tendances observées au sein des pressions empiriques. Il est à noter particulièrement que quelques zones à faible pression (inférieures à 960 hPa) sont bien rendues par le modèle. Cependant, un examen plus approfondi révèle certaines sous-estimations au niveau de ces zones à faible pression. Ce phénomène est clairement visible sur la carte, notamment dans le golfe du Mexique, et se reflète également dans l'histogramme où les pressions simulées enregistrent en moyenne moins de valeurs basses que les pressions historiquement observées. Ces divergences pourraient s'expliquer par le fait que le modèle n'inclut pas un ensemble de variables explicatives qui pourraient exercer une influence significative sur la pression centrale en milieu océanique (CUI et CARACOGIA, 2019).

Pression	Min	1 ^{er} Quartile	Médiane	Moyenne	3 ^{me} Quartile	Max
Empirique	882.0	985.0	1000.0	992.8	1006.0	1024.0
Simulée	870.4	989.5	1000.8	994.7	1007.1	1023.0

TABLE 3.9 : Statistiques descriptives des pressions empiriques et simulées

De plus, le modèle avancé par VICKERY et al. (2000) requiert un nombre important de paramètres initiaux. En effet, ce modèle exige l'initialisation des trois premières pressions ainsi que l'extraction des quatre premières températures mensuelles moyennes, tel qu'explicité dans la figure 3.21, conduisant ainsi à un nombre conséquent d'hypothèses pour les pressions océaniques simulées associée à chaque tempête.

Modèle de CUI et CARACOGIA (2019)

En réponse à cette complexité, CUI et CARACOGIA (2019) a introduit une nouvelle approche, fortement inspirée de celui de VICKERY et al. (2000). Cependant, dans cette version, l'intensité relative est modélisée au travers d'un processus autorégressif d'ordre 2, portant sur les variations entre deux points d'impact d'intensité relative plutôt que sur le logarithme de cette dernière.

Le troisième modèle de pression sur l'océan traité dans cette étude, avancé par CUI et CARACOGIA

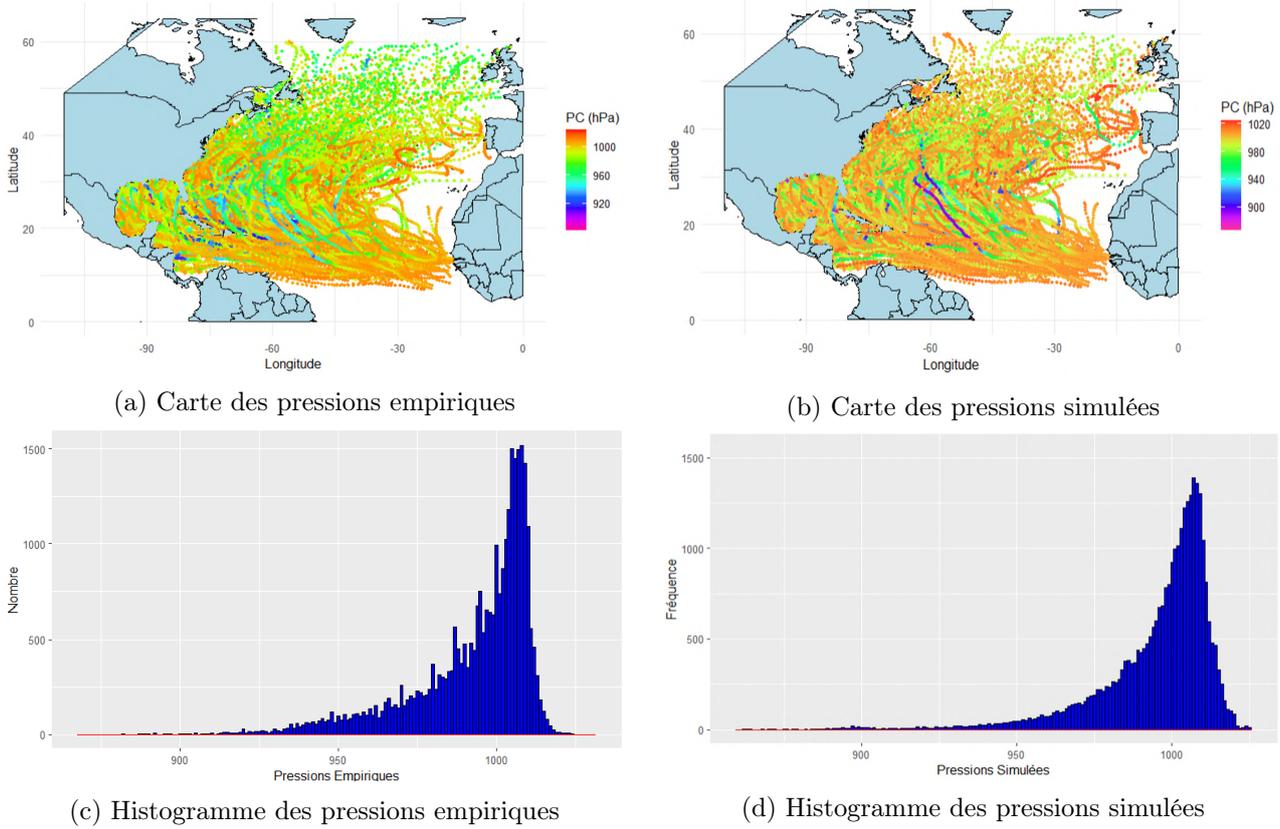


FIGURE 3.23 : Résultats des tests du modèle de pression proposé par VICKERY et al. (2000)

(2019) se définit comme suit

$$\Delta I_{k+1} = c_0 + c_1 \Delta I_k + c_2 \Delta I_{k-1} + c_3 T_{k+1} + \varepsilon, \quad (3.13)$$

où :

- $\Delta I_{k+1} = I_{k+1} - I_k$, est la variation d'intensité relative à l'instant $k + 1$;
- ΔI_k et ΔI_{k-1} sont les variations intensités relatives aux instants k et $k - 1$;
- T_{k+1} est la température moyenne mensuelle à la surface de l'océan à l'instant $k + 1$;
- c_0, c_1, c_2 et c_3 sont les coefficients de la régression ;
- ε est un terme d'erreur.

Il est notable que le modèle défini en (3.13) se simplifie en estimant trois coefficients, une réduction par rapport aux cinq coefficients requis par le modèle précédent. La méthode d'estimation de ces coefficients suit la même approche que celle adoptée par le modèle proposé par VICKERY et al. (2000), impliquant un processus de résolution par case de $5^\circ \times 5^\circ$ dans le bassin Atlantique Nord, comme illustré dans la figure 3.20 puis par une interpolation bilinéaire des coefficients de régression obtenues afin d'obtenir des valeurs en résolution $1^\circ \times 1^\circ$. Les résultats obtenus ainsi que la qualité d'ajustement du nouveau modèle sont présentés directement.

Pour débiter l'analyse, examinons la qualité d'ajustement, comme représenté dans la figure 3.24. Les observations soulignent une similarité frappante entre les résultats de qualité d'ajustement de ce modèle et ceux du modèle établi par VICKERY et al. (2000). Cependant, il est à noter que quelques

modèles échouent au test de significativité globale dans cette nouvelle approche. Néanmoins, il est important de souligner que la simple non-satisfaction du test de significativité globale n'implique pas systématiquement l'exclusion d'un modèle. Dans la démarche de modélisation statistique complexe, il est courant de retenir certains modèles malgré l'échec d'un test spécifique, particulièrement si l'ensemble des autres critères d'évaluation soutiennent la validité du modèle. Dans cette étude, il est choisi de garder ces modèles échouant au test de significativité. En effet, après analyse de l'histogramme et de la distribution des résidus pour chaque modèle de régression, il est remarquable qu'aucune structure particulière n'émerge et que ces résidus se trouvent centrés autour de la moyenne, conforme à une distribution normale. En conséquence, il est envisageable de considérer les résidus de chaque modèle de régression comme étant gaussiens dont les paramètres sont égales à la moyenne et à la variance des résidus du modèle correspondant.

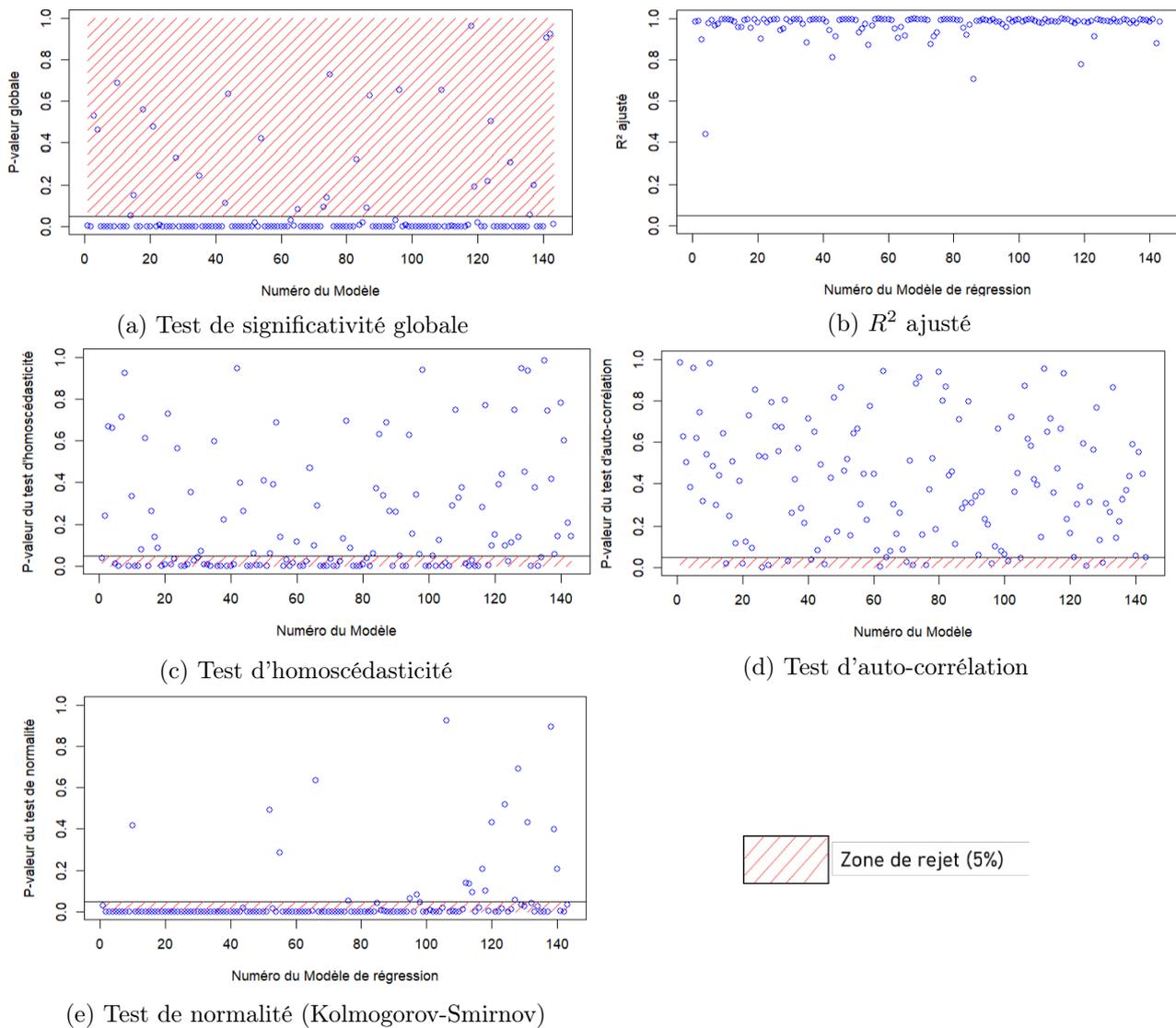


FIGURE 3.24 : Qualité d'ajustement des 143 modèles de régression proposé par CUI et CARACOLIA (2019)

Ensuite, les statistiques descriptives présentées dans le tableau 3.10 mettent en évidence une cohérence appréciable entre les pressions simulées et les pressions empiriques, avec une pression minimale simulée qui s'aligne davantage sur les observations empiriques, comme cela était déjà le cas dans le

modèle précédent. En revanche, dans ce modèle, une tendance à surestimer les pressions élevées (supérieures à 1000 hPa) est observable. Cette observation est confirmée par l'analyse des résultats des tests appliqués aux trajectoires empiriques, illustrés dans la figure 3.25, où l'histogramme des pressions simulées révèle une fréquence accrue de pressions élevées par rapport au modèle précédent.

Toutefois, il est important de noter que malgré certains échecs dans le test de significativité globale, ce modèle présente des caractéristiques intéressantes dans sa capacité à rendre compte de certains aspects du phénomène cyclonique. Par exemple, ce modèle parvient à représenter adéquatement les basses pressions, comme en témoigne la présence de zones à pressions faibles sur la carte des pressions simulées, alignées avec la distribution empirique des pressions observées. De plus, cette version du modèle a une tendance à surestimer les pressions élevées tout en sous-estimant les pressions basses. En comparaison, le modèle développé par VICKERY et al. (2000) montre plutôt une tendance à surestimer les pressions moyennes (entre 940 et 980 hPa), tout en sous-estimant les pressions basses.

Pression	Min	1 ^{er} Quartile	Médiane	Moyenne	3 ^{me} Quartile	Max
Empirique	882.0	985.0	1000.0	992.8	1006.0	1024.0
Simulée	867.6	984.4	1000.1	992.8	1009.8	1031.4

TABLE 3.10 : Statistiques descriptives des pressions empiriques et simulées

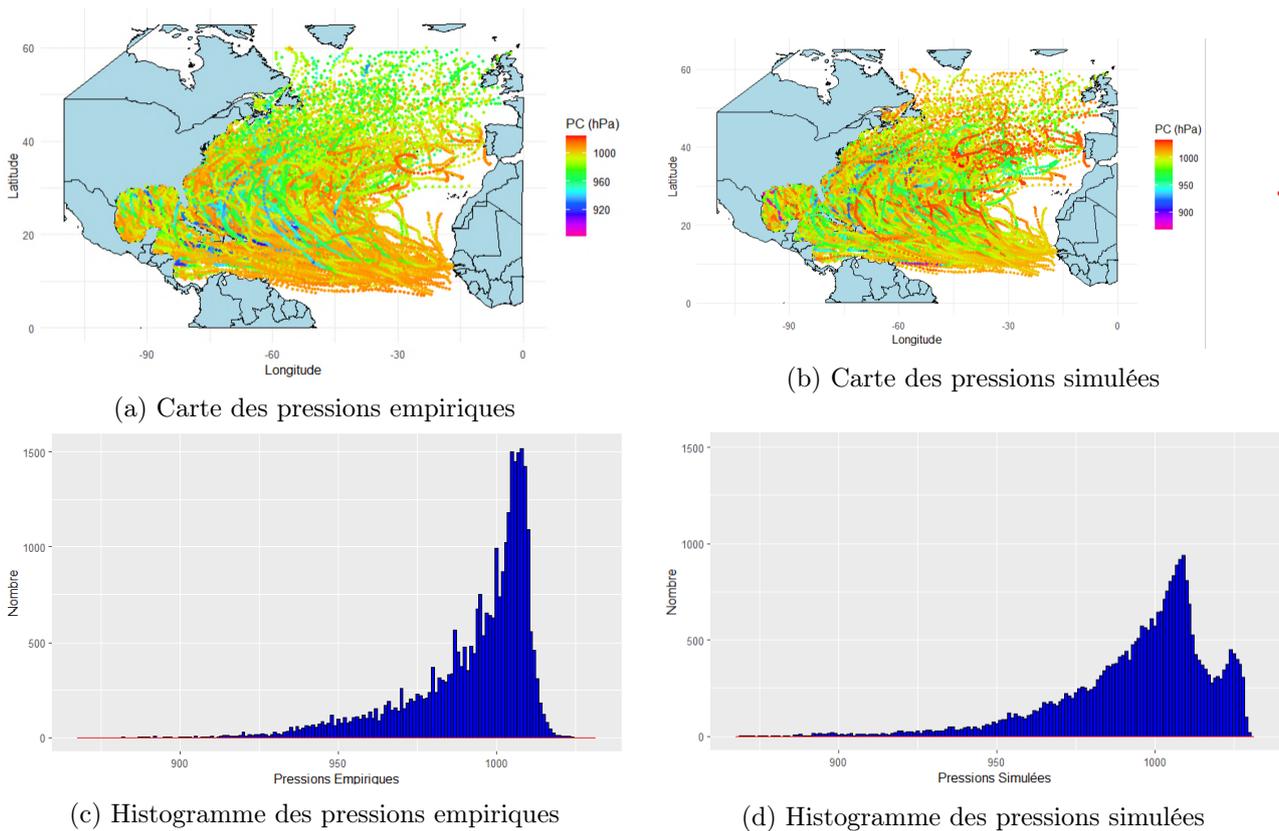


FIGURE 3.25 : Résultats des tests du modèle de pression proposé par CUI et CARACOGIA (2019)

Comparaison et choix du modèle pour l'étude

Le choix final du modèle de pression sur l'océan suscite des questionnements qui dépendent des objectifs définis dans le contexte de cette étude. Une dernière analyse est réalisée en comparant les

graphiques quantile-quantile (Q-Q plot) des deux modèles, celui de VICKERY et al. (2000) et celui de CUI et CARACOGIA (2019). Cette comparaison est illustrée dans la figure 3.26.

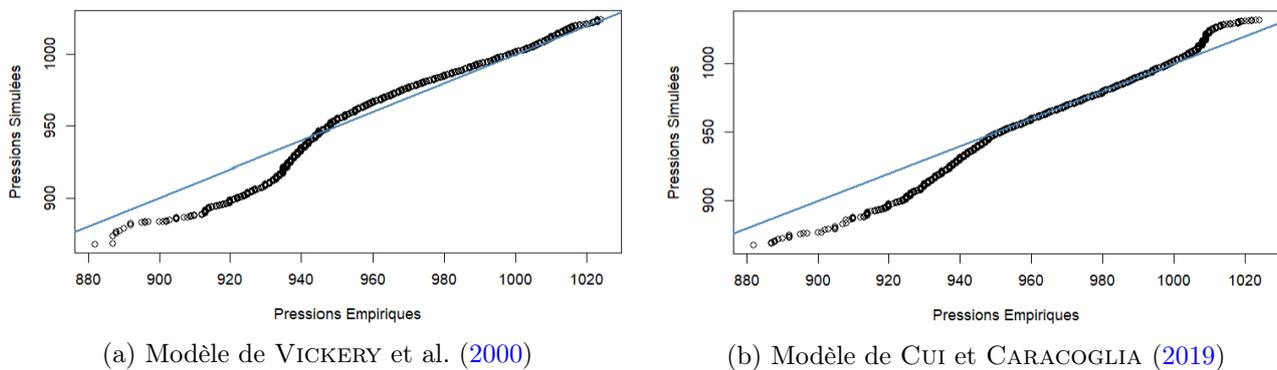


FIGURE 3.26 : Q-Q plot entre pressions empiriques et simulées

Les Q-Q plots en question confirment les conclusions précédemment évoquées. Il est observé que le modèle élaboré par VICKERY et al. (2000) démontre une capacité prédictive satisfaisante en moyenne. Toutefois, il est apparent que ce modèle tend à surestimer les pressions moyennes (comprises entre 940 et 980 hPa). Bien que ces valeurs de pression ne correspondent pas aux catégories d'ouragans dévastateurs (4 ou 5), elles demeurent associées aux catégories d'ouragans (1, 2 et 3), dont les conséquences ne sont pas négligeables et peuvent potentiellement induire des impacts significatifs, en accord avec l'échelle de Saffir-Simpson 1.1.

Ainsi, le modèle établi par VICKERY et al. (2000) pourrait être employé, par exemple, dans un contexte de tarification des primes d'assurance, visant à offrir des primes aux assurés tout en maintenant une compétitivité prudente sur le marché. En effet, ce modèle présente une propension à sous-estimer les basses pressions (inférieures à 940 hPa), permettant ainsi de cibler les risques moins favorables dans un portefeuille d'assurance. Dans cette perspective, la proposition d'une prime d'assurance basée sur ce modèle pourrait permettre d'équilibrer ce segment de risques moins favorables, où la prime serait substantiellement élevée pour cette catégorie d'assurés. Cela résulterait en une prime proportionnellement élevée par rapport au risque effectif encouru par les assurés appartenant à cette classe de risque, en raison de la sous-estimation des pressions basses. En contraste, la part de risque « favorable »⁵ dans laquelle la prime serait ajustée proportionnellement en fonction du risque réel encouru par les assurés de cette catégorie, montrerait une prime relativement plus élevée par rapport au risque effectif encouru.

En ce qui concerne le modèle formulé par CUI et CARACOGIA (2019), il est notable qu'il affiche une qualité prédictive inférieure à celle du modèle de VICKERY et al. (2000), mais il offre tout de même un intérêt pour plusieurs raisons. Premièrement, il maintient sa tendance à sous-estimer les pressions associées aux catégories d'ouragans (inférieures à 995 hPa), ce qui souligne sa prudence continue à l'égard de ces valeurs de pression. De surcroît, cette sous-estimation augmente à mesure que la catégorie de l'ouragan s'intensifie. Cette tendance est particulièrement marquée pour les ouragans de catégorie 4 et 5, se distinguant des catégories 1, 2 et 3. De plus, il est pertinent de noter que ce modèle surestime systématiquement les pressions très élevées (supérieures à 1010 hPa), ce qui, néanmoins, ne suscite pas d'inquiétude majeure. En effet, ces valeurs de pression correspondent à une classe de risque très favorable, à savoir les perturbations tropicales sans potentiel destructeur. Cependant, dans un contexte de tarification des primes d'assurance, ce modèle ne se révèle pas très intéressant. En effet, il aboutirait à une proposition de prime extrêmement élevée pour toutes les catégories d'assurés exposées aux ouragans ayant un effet destructeur (3, 4 et 5). Par ailleurs, la prime relativement plus

⁵Dans le sens où cette classe de risque est moins exposée aux effets dévastateurs des ouragans que la classe de risque correspondant à des valeurs de pressions basses

basse offerte à la catégorie d'assurés bénéficiant d'une situation de très bon risque pourrait ne pas suffire à compenser les pertes potentielles subies par les catégories de risques moins favorables.

En revanche, en ce qui concerne la provision pour le risque ouragan en assurance, le modèle de CUI et CARACOGIA (2019) démontre une pertinence marquée et pourrait se révéler d'une utilité significative. Notamment, comme évoqué dans le chapitre précédent 1.5, les coûts globaux ainsi que les pertes assurantielles résultant du risque ouragan ont connu une augmentation considérable au cours des dernières décennies et semblent destinés à croître davantage aux États-Unis dans les années à venir. De surcroît, selon les analyses des experts en la matière, la fréquence des ouragans majeurs (catégories 3, 4 et 5) pourrait s'accroître dans le futur (KOSSIN et al., 2020). Par conséquent, le modèle en question trouve une pertinence particulière dans le contexte de provision pour le risque ouragan, car il permet de prendre en compte ces mutations cruciales qui ont le potentiel d'exercer un impact substantiel sur l'économie américaine.

Au sein de cette étude, inscrite dans une démarche d'estimation des pertes finales liées au risque ouragan aux États-Unis, prenant en considération les évolutions marquantes ayant affecté l'économie du pays ces dernières décennies, le choix du modèle pour la modélisation de la pression centrale sur l'océan se porte sur celui établi par CUI et CARACOGIA (2019). En effet, la capacité de ce modèle à demeurer systématiquement prudent à mesure que la catégorie d'ouragan progresse revêt une importance majeure dans le contexte étudié, où les tendances actuelles signalent une augmentation des coûts et de la destructivité de ces phénomènes.

3.2.1.3 Modèle de pression centrale sur terre

Le modèle de décroissance de la pression centrale sur terre, adopté dans le cadre de cette recherche, est celui initialement avancé par KAPLAN et DEMARIA (1995). Ce modèle a été largement employé dans divers travaux de modélisation du phénomène cyclonique dans le bassin Atlantique Nord, notamment dans des études récentes menées par des chercheurs tels que LIU et PANG (2011), SNAIKI et WU (2020), et CUI et CARACOGIA (2019). Ce modèle est formulé comme suit

$$\Delta P_k = \Delta P_0 \cdot \exp(-(a + \varepsilon) \cdot t_k), \quad (3.14)$$

où :

- $\Delta P_k = P_{env,k} - P_k$ représente le déficit de pression à l'instant k , $k \in \mathbb{N}$ où 1 unité de k correspond à une durée de 3h conformément à la base IBTrACS, et $P_{env,k}$ correspond à la pression atmosphérique mensuelle P_{env} à l'instant k tel qu'il a été défini dans la modélisation de la pression centrale océanique ;
- $\Delta P_0 = P_{env,0} - P_0$ représente le déficit de pression à l'instant d'atterrissage ;
- a représente la constante de désintégration à estimer ;
- t_k représente le nombre d'heures écoulées depuis l'instant d'atterrissage ;
- ε représente un terme d'erreur.

La constante de désintégration a caractérise la vitesse de déclin de la pression après l'entrée de l'ouragan dans les terres, et sa dynamique diffère en fonction du lieu d'observation de cet atterrissage. Cependant, dans les diverses modélisations mentionnées précédemment, l'estimation de cette constante de décroissance ne suit pas toujours le même protocole. Par exemple, CUI et CARACOGIA (2019) adopte la méthodologie initiale de KAPLAN et DEMARIA (1995), en considérant trois zones d'estimation distinctes (Atlantique Nord, Golfe du Mexique et Floride). Cette étude procède à une régression linéaire formulée comme suit

$$a = a_0 + a_1 \Delta P_0 + \varepsilon, \quad (3.15)$$

où :

- a représente la constante de désintégration à estimer ;
- ΔP_0 représente le déficit de pression à l'instant d'atterrissage ;
- a_0 et a_1 sont les coefficients à estimer ;
- ε représente un terme d'erreur.

Cependant, d'autres études, comme celle menée par LIU et PANG (2011), notent que la constante de décroissance peut être modélisée simplement par une valeur constante, c'est-à-dire $a = a_0 + \varepsilon$, plutôt que par une relation linéaire comme celle exposée dans l'équation (3.15). Cette option découle d'une observation de l'évolution de la constante d'intégration définie dans l'équation (3.15), à partir des données fournies par la base de données IBTrACS, comme illustré dans la figure 3.27. Cette visualisation révèle que la constante d'intégration a suit approximativement une trajectoire autour d'une valeur constante, accompagnée de variations d'amplitude plus ou moins régulière en fonction de la zone d'étude.

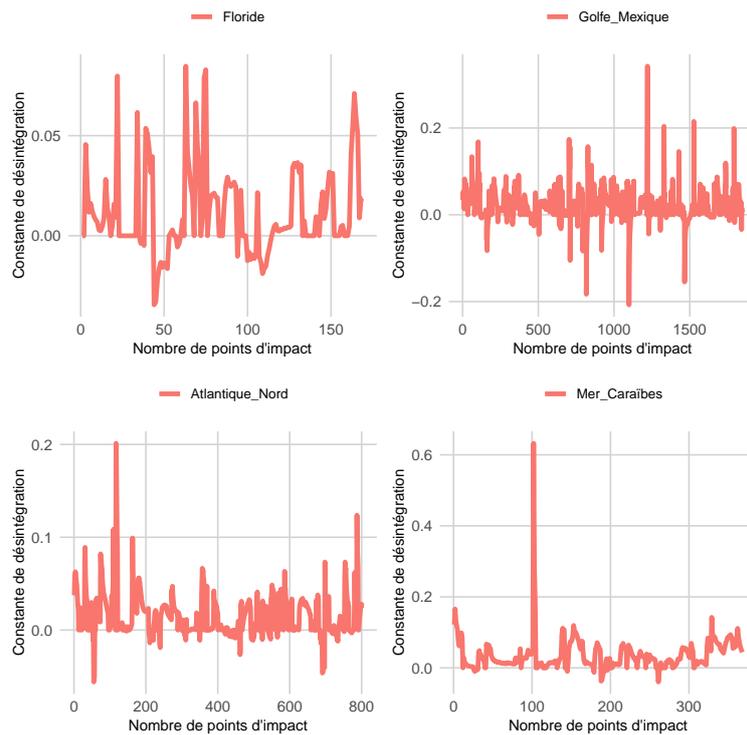


FIGURE 3.27 : Evolution de la constante de désintégration sur chaque zone d'étude

La modélisation de la constante de désintégration à travers une constante comme proposé par LIU et PANG (2011) est adoptée par la suite. De plus une zone d'atterrissage supplémentaire est rajoutée, celle de la mer des Caraïbes, en plus des trois zones précédemment définies. L'articulation des zones d'atterrissage des tempêtes historiques est détaillée dans la figure 3.28. Notons également que, conformément à l'exemple de l'ouragan Irma présenté dans la figure 3.17, les points d'impact terrestres des tempêtes touchant les îles et les états insulaires sont exclus du modèle, du fait de leur effet négligeable

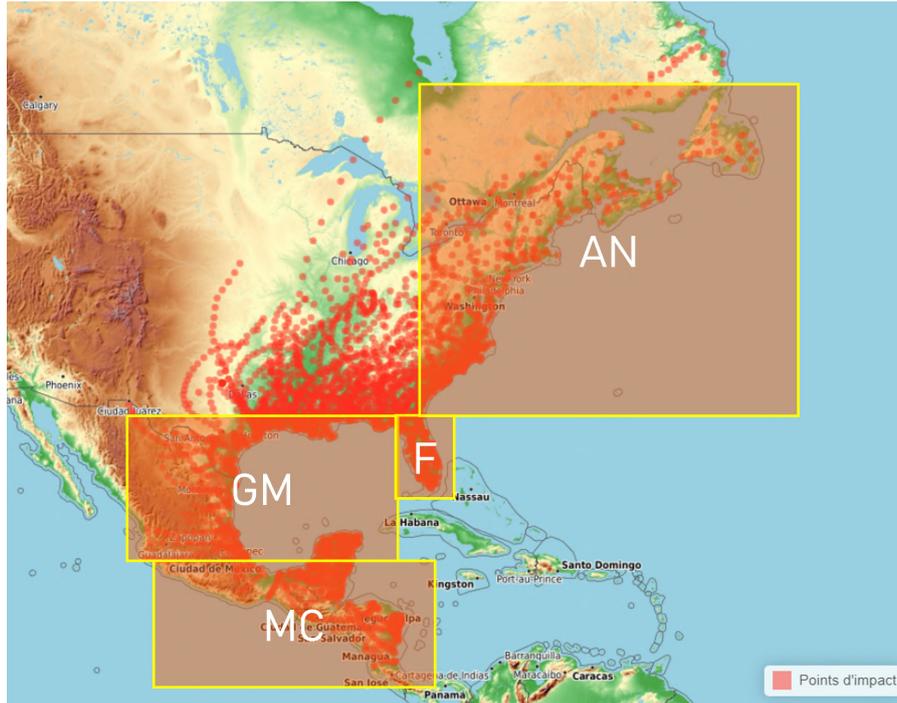


FIGURE 3.28 : Découpage des zones côtières terrestres en 4 régions distinctes

sur le déficit de pression. Cette hypothèse est également utilisée dans toutes les modélisations évoquées dans les articles précédemment cités.

En outre, il est à noter que certaines tempêtes, en particulier les plus puissantes, peuvent traverser plusieurs zones terrestres au cours de leur trajectoire, tout en regagnant des sources d'eau entre chaque atterrissage. L'ouragan ETA de 2020 en est un exemple notable : il a d'abord touché la côte de la mer des Caraïbes, puis a regagné les eaux pour ré-atteindre le territoire cubain. Par la suite, il a de nouveau regagné les eaux pour toucher la Floride, avant de retourner dans l'océan Atlantique où il a fini sa course. Pour les tempêtes suivant de telles trajectoires englobant plusieurs zones d'étude conformément à la découpe présentée dans la figure 3.28, leur trajectoire est subdivisée en sous-trajectoires, correspondant à chaque zone d'estimation prise en compte. Cette démarche est clairement illustrée par le découpage de la trajectoire de l'ouragan ETA dans la figure 3.29. Cette figure est constituée de deux sous-figures : la première 3.29a montre les points d'impact individuels de l'ouragan ETA, tandis que la seconde 3.29b détaille les points d'impact répartis selon les différentes zones d'étude.

Par la suite, les résultats obtenus pour la constante de désintégration a_0 , l'écart-type des résidus σ_ε , ainsi que le nombre d'observations n_{obs} de chaque régression réalisée sont exposés dans le tableau 3.11.

Zone géographique	a_0	σ_ε	n_{obs}
Atlantique Nord	0.0192	0.0281	744
Floride	0.0166	0.0250	183
Golfe du Mexique	0.0218	0.0360	1893
Mer des Caraïbes	0.0208	0.0250	369

TABLE 3.11 : Résultats des estimations de la constante de désintégration

Afin de procéder à la vérification du modèle élaboré, son application aux données empiriques a été conduite de manière similaire à celle du modèle de pression centrale sur terre. Cette démarche a permis

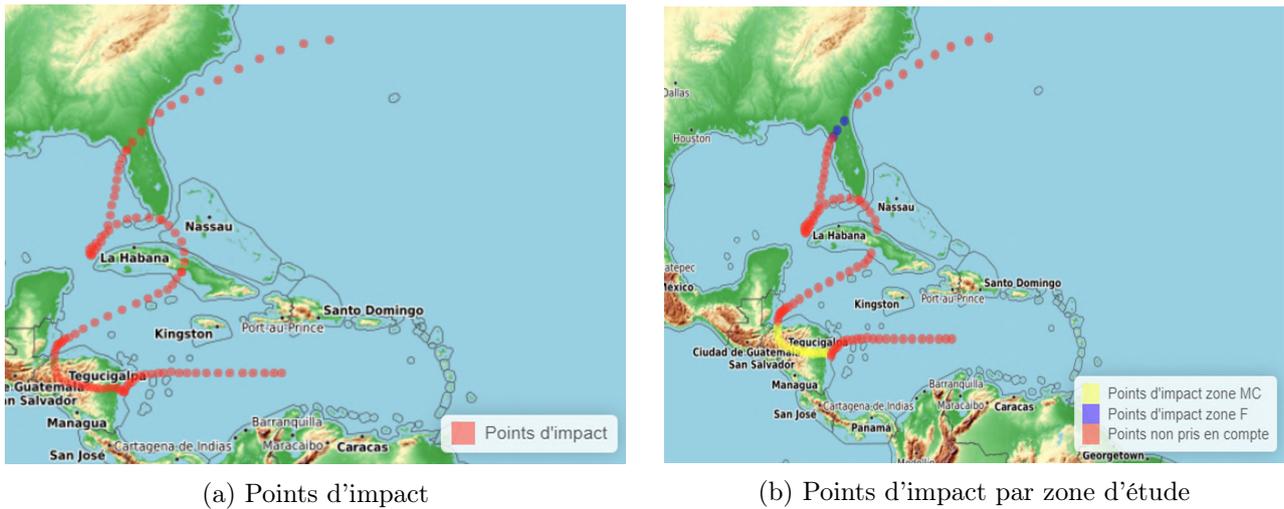


FIGURE 3.29 : Découpage en sous-trajectoires de l'ouragan ETA

la création d'une carte comparative et d'un histogramme, illustrés dans la figure 3.30, présentant les pressions simulées et empiriques. Parallèlement, les statistiques descriptives appropriées ont été représentées dans le tableau 3.12.

Pression	Min	1 ^{er} Quartile	Médiane	Moyenne	3 ^{me} Quartile	Max
Empirique	909	995	1005	1000	1010	1017
Simulée	909	995	1003	999.4	1007	1016

TABLE 3.12 : Statistiques descriptives des pressions empiriques et simulées

L'analyse des statistiques descriptives met en évidence une correspondance notable entre les pressions simulées et les données empiriques. Globalement, la plage de valeurs des pressions simulées se superpose à celle des pressions empiriques. Cette convergence est également discernable dans les cartes et les histogrammes présentés dans la figure 3.30, où les configurations spatiales et les fréquences se rapprochent considérablement. Les zones de très basses pressions sont adéquatement reproduites, notamment aux abords de la Floride et des côtes de la Mer des Caraïbes. La dernière analyse concerne le Q-Q plot entre les pressions simulées et empiriques, présentés dans la figure 3.31. Une concordance réelle entre ces valeurs est observée, avec un alignement satisfaisant le long de la droite $y = x$. Quelques pressions moyennes sont légèrement sous-estimées, témoignant ainsi d'une certaine prudence inhérente au modèle.

La modélisation de la pression centrale tout au long de la trajectoire d'une tempête est désormais achevée. Il est important de noter que, l'algorithme de simulation de pressions océaniques nécessite d'initialiser les trois premières pressions le long de la trajectoire, comme détaillé dans la figure 3.21. Ces trois premières valeurs de pression initiale sont tirées selon la distribution historique des trois premières pressions initiales observées empiriquement. Cette hypothèse s'avère cohérente dans la mesure où la gamme de valeurs observée lors de ces trois premières étapes de la trajectoire est suffisamment étendue pour englober la variabilité potentielle des pressions initiales. Il est également important de noter qu'une trajectoire de tempête peut traverser des zones terrestres avant de retourner vers l'océan. Dans de tels cas, il est essentiel de garantir qu'il y ait au moins trois pas de trajectoire de la tempête au-dessus de l'océan avant d'appliquer le modèle de pression océanique proposé par CUI et CARACOGIA (2019). Cette condition est nécessaire pour que le modèle puisse fournir des résultats pertinents. Afin de clarifier la méthodologie adoptée pour la simulation des pressions le long de la trajectoire des tempêtes,

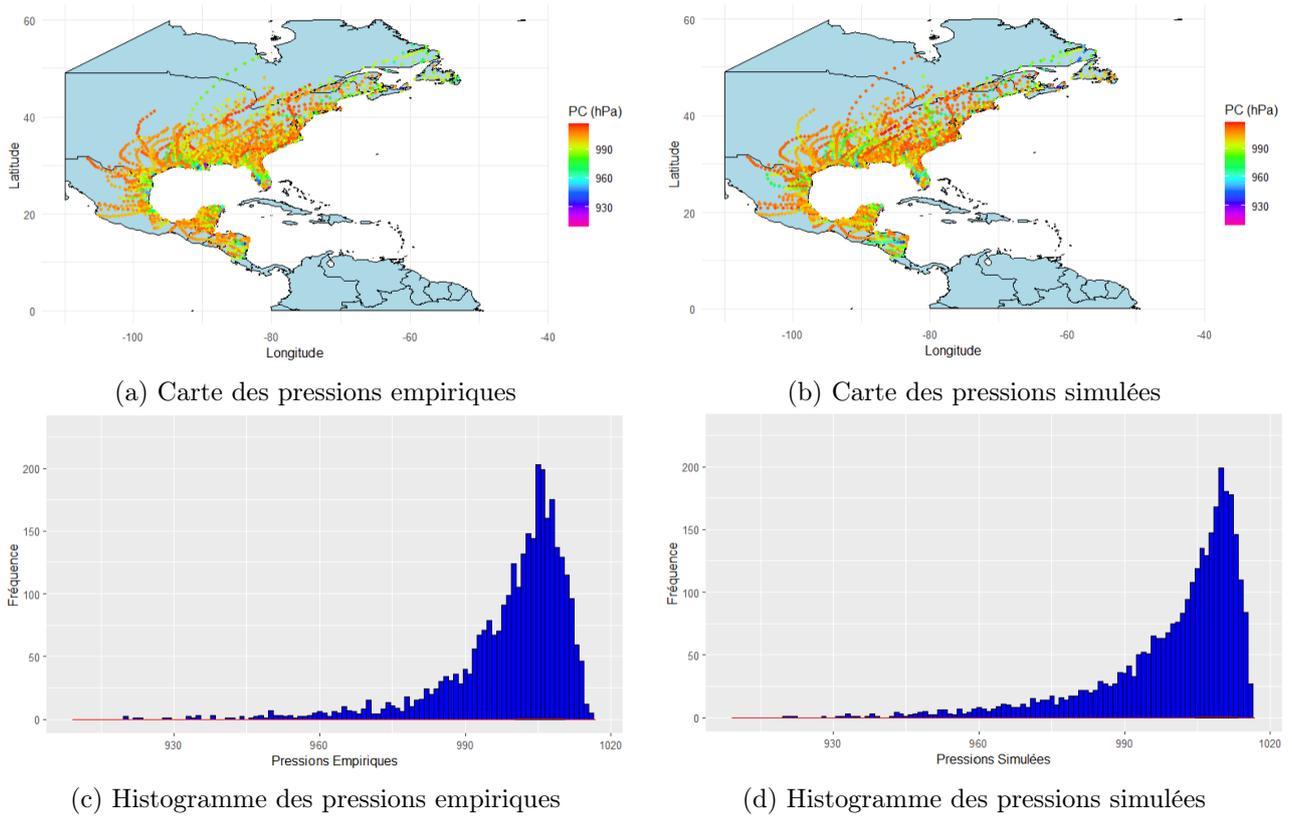


FIGURE 3.30 : Résultats des tests du modèle de pression sur terre

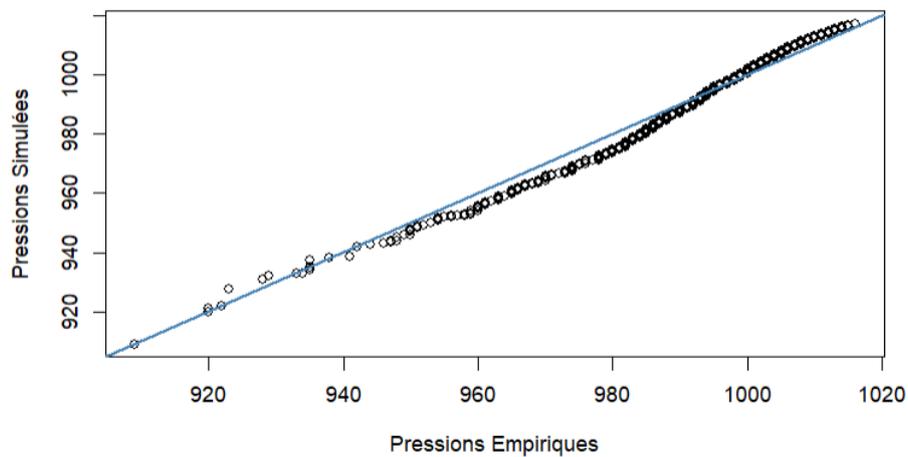


FIGURE 3.31 : Q-Q plot entre pressions empiriques et simulées

la figure 3.32 résume les différentes étapes de l'algorithme de simulation.

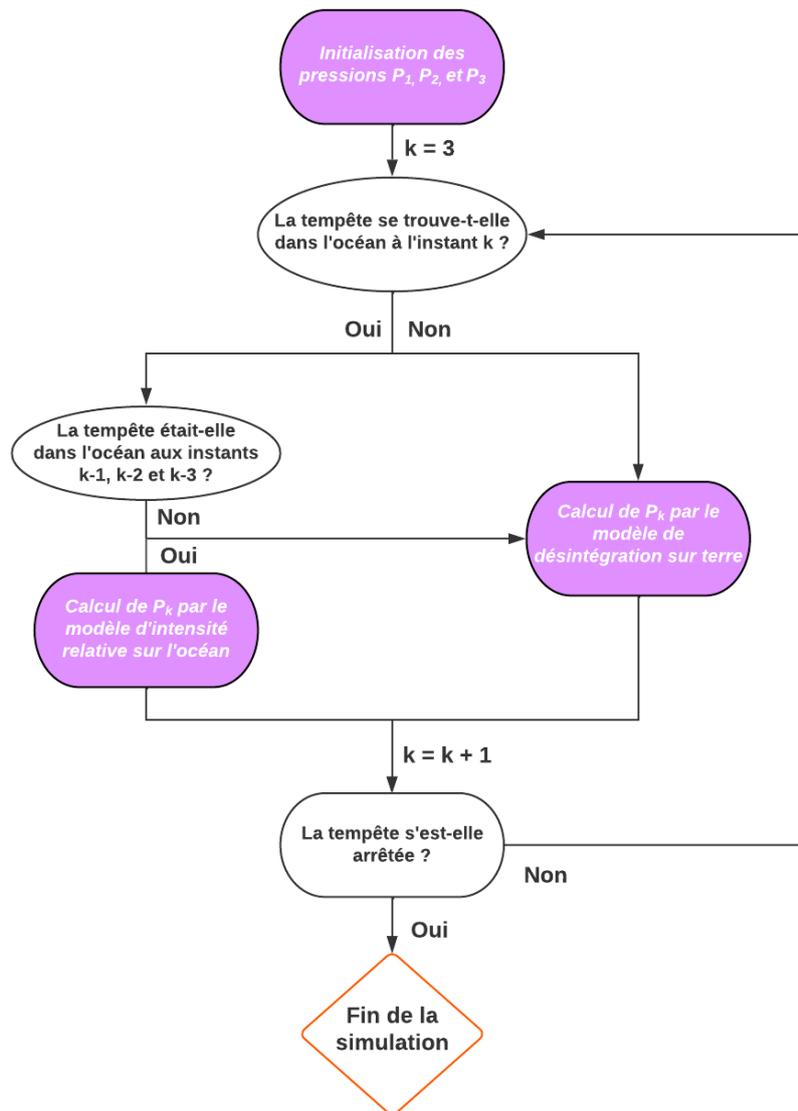


FIGURE 3.32 : Algorithme de simulation des pressions le long d'une trajectoire de tempête

3.2.2 Modélisation des autres caractéristiques fondamentales

La pression centrale le long des trajectoires étant modélisée, il est désormais possible de déterminer les autres caractéristiques fondamentales d'une tempête tropicale, à savoir la vitesse maximale des vents, le rayon de la vitesse maximale des vents ainsi que le gradient de vitesse.

3.2.2.1 Vitesse maximale des vents

La littérature scientifique présente plusieurs études ayant abordé la modélisation de la vitesse maximale des vents, notée V_{max} , en fonction de la pression centrale, notée PC . Cette relation peut tout d'abord être observée à travers la figure 3.33. Cette représentation affiche les cartes des vitesses maximales des vents ainsi que les pressions centrales des points d'impact des tempêtes historiques de la base de données IBTrACS, illustrant une nette concordance. En effet, il est immédiatement apparent que la vitesse maximale des vents augmente à mesure que la pression centrale diminue. Cette tendance est également corroborée par l'observation graphique de la figure 3.33c, où la relation décroissante entre

ces deux caractéristiques est clairement mise en évidence.

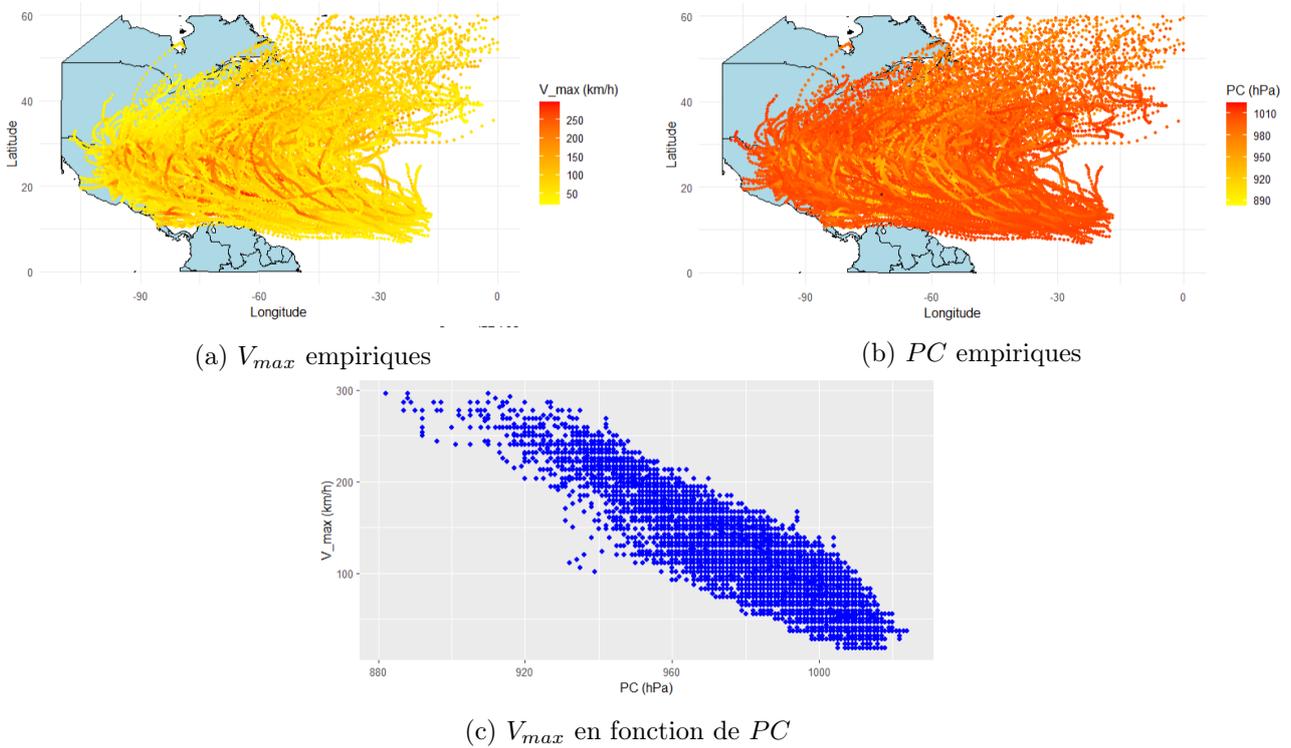


FIGURE 3.33 : Relation entre vitesse maximale des vents et pression centrale

Lorsqu'il s'agit de modéliser la vitesse maximale des vents, deux types de modèles prédominent dans la littérature scientifique (HOLLAND, 2008) :

- Les relations de type « puissance » : $V_{max} = a\Delta P^b$, où $a > 0$, $0 < b < 1$ sont des coefficients à estimer et ΔP correspond au déficit de pression centrale⁶ ;
- Les relations de type « linéaire » : $V_{max} = a_0 + a_1\Delta P + a_2\Delta P^2$, où a_0 , a_1 sont des coefficients à estimer et ΔP correspond au déficit de pression centrale.

Ces deux relations ont été modélisées dans le cadre de cette étude et les résultats sont quasiment similaires. Cependant, une préférence est accordée aux modèles de type « linéaire », car ils permettent une meilleure capture de la forme parabolique de la relation entre les variables V_{max} et PC , comme illustré sur la figure 3.33c et une meilleure qualité des résidus. Une analyse comparative des performances des deux modèles est proposée en annexe A.6. Ce choix de modèle est également appliqué dans le cadre de mémoires d'actuariat (DENISE-BAILLON, 2013 ; YOMBI, 2016). Cependant il convient de noter que la construction du modèle diffère de celle décrite dans ces mémoires. Afin de saisir efficacement la forme parabolique inhérente à la relation entre la vitesse maximale des vents et la pression centrale, le modèle de régression linéaire précédemment défini est calibré en fonction des catégories de pression centrale. Les pressions centrales relevées dans l'échantillon historique sont subdivisées en six catégories, allant des pressions très basses aux pressions très élevées. Cette subdivision trouve sa justification dans la volonté d'appréhender les spécificités propres à chaque catégorie d'ouragan et de parvenir, par conséquent, à une estimation plus précise de la vitesse maximale des vents. En effet, à mesure que la catégorie d'ouragan gagne en intensité, le nombre d'observations historiques diminue.

⁶tel que défini dans le modèle de pression sur terre par l'équation (3.14)

Ainsi, élaborer un modèle global regroupant l'ensemble des catégories de pression aurait du mal à capturer les particularités spécifiques à chacune d'elles. Les résultats des modèles de régression ainsi que les écart-types des estimations de chaque coefficient (noté σ_{a_0} , σ_{a_1} et σ_{a_2} et les intervalles de confiance à 95% (noté $IC_{95\%}$) pour chaque catégorie de pression sont présentés dans les tableaux 3.13 et 3.14.

Pression (hPa)	a_0	a_1	a_2	σ_ε	n_{obs}
(881, 905]	1451.73	-21.58	0.10	13.55	28
(905, 929]	432.05	-5.65	0.04	19.30	248
(929, 953]	-266.75	11.33	-0.07	24.92	1538
(953, 977]	16.82	3.57	-0.02	24.09	3970
(977, 1003]	59.22	1.22	0.01	18.32	13130
(1003, 1024]	37.14	3.26	-0.07	12.70	15999

TABLE 3.13 : Résultats des estimations des modèles de régression

Pression (hPa)	σ_{a_0}	$IC_{95\%}$	σ_{a_1}	$IC_{95\%}$	σ_{a_2}	$IC_{95\%}$
(881, 905]	929.56	[-462.74; 3366.20]	15.97	[-54.47; 11.32]	0.06	[-0.04; 0.23]
(905, 929]	308.53	[-175.67; 1039.77]	6.58	[-18.62; 7.32]	0.03	[-0.02; 0.10]
(929, 953]	73.95	[-411.82; -121.68]	2.06	[7.28; 15.38]	0.01	[-0.09; -0.03]
(953, 977]	16.48	[-15.50; 49.14]	0.67	[2.24; 4.89]	0.006	[-0.02; -0.002]
(977, 1003]	1.67	[55.93; 62.5]	0.13	[0.95; 1.49]	0.002	[0.009; 0.02]
(1003, 1024]	0.47	[36.19; 38.07]	0.11	[3.03; 3.48]	0.006	[-0.08; -0.05]

TABLE 3.14 : Incertitudes relatives aux estimations des coefficients

Les écart-types indiquent la dispersion des valeurs estimées de chaque coefficient, tandis que les intervalles de confiance à 95% fournissent un intervalle dans lequel la vraie valeur du coefficient se situe avec une confiance à 95%. Il est observable que la taille de l'intervalle de confiance se réduit à mesure que la tranche de pression augmente, ce qui est assez logique car le nombre d'observations augmente considérablement lorsque la pression s'élève. Des intervalles de confiance plus étroits et des écart-types plus petits indiquent une plus grande précision des estimations des coefficients, ce qui est généralement le cas à mesure que la tranche de pression augmente. Cependant, il est crucial de noter que pour les tranches de pression plus basses où le nombre d'observations est limité, les écart-types des estimations des coefficients sont généralement plus élevés, ce qui se traduit par des intervalles de confiance plus larges. Cette observation découle du fait que la précision des estimations dépend fortement de la quantité de données disponibles pour chaque tranche de pression. Lorsque les données sont limitées, l'incertitude inhérente aux estimations augmente naturellement.

Dans de telles situations, il est nécessaire de mener des analyses complémentaires pour évaluer la fiabilité du modèle. L'analyse des résidus est présentée à travers l'examen des graphiques quantile-quantile, utilisant une distribution normale centrée réduite comme quantiles théoriques. Ces graphiques sont illustrés sur la figure 3.34. Au vu de ces graphiques, la grande majorité des résidus se situent bien le long d'une distribution normale, même si les quantiles extrêmes sont généralement sous-estimés pour la plupart des modèles. Néanmoins, l'hypothèse de normalité est conservée pour chaque modèle. Par conséquent, les résidus sont considérés comme ayant une moyenne de 0 et un écart-type σ_ε , en fonction de la catégorie déterminée dans le tableau 3.13.

Afin de tester les modèles construits, une représentation graphique des vitesses maximales des vents simulées en fonction de la pression centrale, en tenant compte d'un terme d'erreur gaussien (représenté en rouge), superposée aux données empiriques (en bleu), est illustré sur la figure 3.35. Les histogrammes des vitesses maximales des vents empiriques et prédites sont également ajoutés.

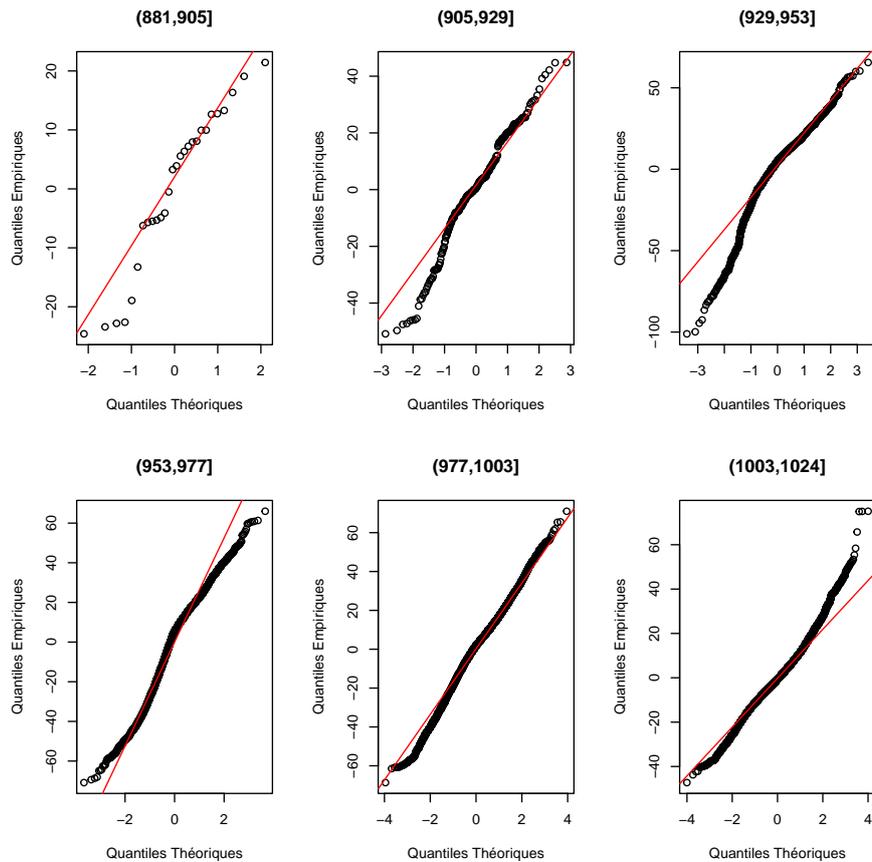


FIGURE 3.34 : Q-Q plot des résidus

Les résultats montrent que le modèle est en mesure de saisir de manière précise les caractéristiques spécifiques à chaque catégorie de pression, malgré quelques points qui présentent des divergences avec le modèle, notamment pour les catégories 3 et 4 des ouragans. Il est cependant remarquable que plusieurs points sont surestimés, en particulier pour les catégories d'ouragans majeurs, ce qui est encourageant car cela traduit une certaine prudence maintenue par le modèle. Par ailleurs, la confrontation des distributions simulées et empiriques à travers les histogrammes met en évidence une concordance globale entre les deux ensembles de données.

3.2.2.2 Rayon de vitesse maximale des vents

Dans le domaine scientifique, VICKERY et WADHERA (2008) avancent, par exemple, la proposition de modéliser le logarithme du rayon de la vitesse maximale des vents en fonction du déficit de pression, de ce dernier au carré, ainsi que de la latitude. Ce choix trouve sa justification dans la représentation graphique de l'évolution du logarithme de R_{max} en corrélation avec ΔP et la latitude, illustrée par la figure 3.36. Une tendance distincte s'y profile. En effet, il est perceptible que le rayon s'accroît à mesure que les ouragans progressent vers des latitudes plus élevées. En revanche, à mesure que l'intensité des ouragans augmente (indiquée par un déficit de pression plus prononcé), leur rayon diminue. Cela traduit le fait que les ouragans les plus dévastateurs présentent généralement des valeurs de R_{max} plus réduites. Néanmoins, le modèle présenté par VICKERY et WADHERA (2008) s'avère insuffisant en raison de son coefficient de détermination d'environ 28%, traduisant ainsi une faible capacité explicative.

Par conséquent, l'enrichissement du modèle est envisagé en introduisant d'autres variables potentiellement explicatives pour le logarithme de R_{max} . Dans ce contexte, DENISE-BAILLON (2013) suggère

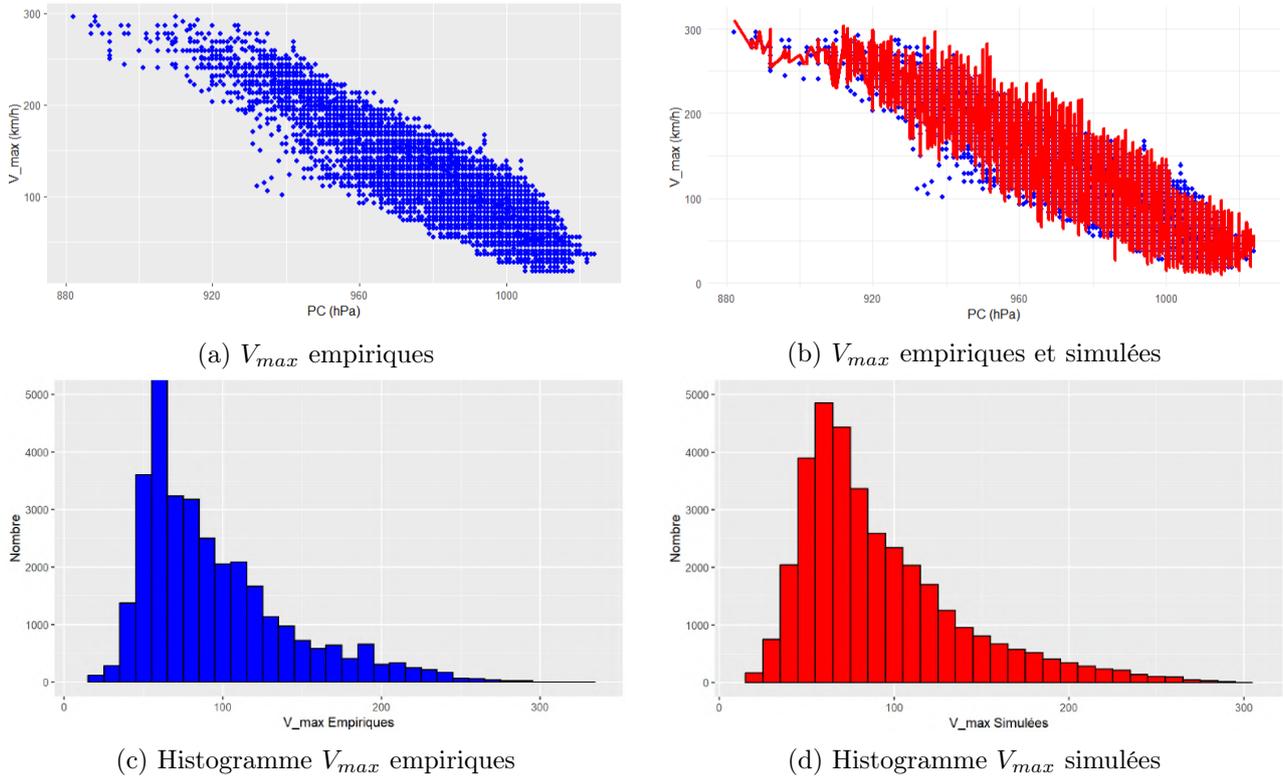


FIGURE 3.35 : Résultats des tests du modèle de vitesse maximale des vents

d'inclure la longitude ainsi que la valeur antérieure du logarithme de R_{max} dans la modélisation. Ces diverses évolutions sont illustrées graphiquement dans la figure 3.37. Une légère tendance en fonction de la longitude est observable. En effet, à mesure que la longitude augmente, le rayon de la vitesse maximale des vents a tendance à augmenter également, bien que cette relation soit moins directe comparée à celle avec la latitude. Par ailleurs, en traçant R_{max} en fonction de sa valeur précédente, il est remarquable que cette dernière demeure relativement stable le long d'une trajectoire, en raison d'une forte relation de linéarité.

En conséquence, le choix d'introduire ces deux variables additionnelles dans le modèle est effectué. Ainsi, le modèle final se présente sous la forme suivante

$$\ln R_{max}^k = a_0 + a_1\phi + a_2\xi + a_3\Delta P + a_4\Delta P^2 + a_5 \ln R_{max}^{k-1} + \varepsilon,$$

où :

- R_{max}^k , représente le rayon de la vitesse maximale des vents à l'instant k (en km) ;
- ϕ représente la latitude (en degrés) ;
- ξ représente la longitude (en degrés) ;
- ΔP représente le déficit de pression (en hPa)
- a_0, a_1, a_2, a_3, a_4 et a_5 sont les coefficients à estimer ;
- ε représente un terme d'erreur.

Bien que ce modèle donne des résultats corrects, il est en revanche lourd d'informations et se base uniquement sur des relations purement empiriques basées sur des quantités observées telles que le déficit

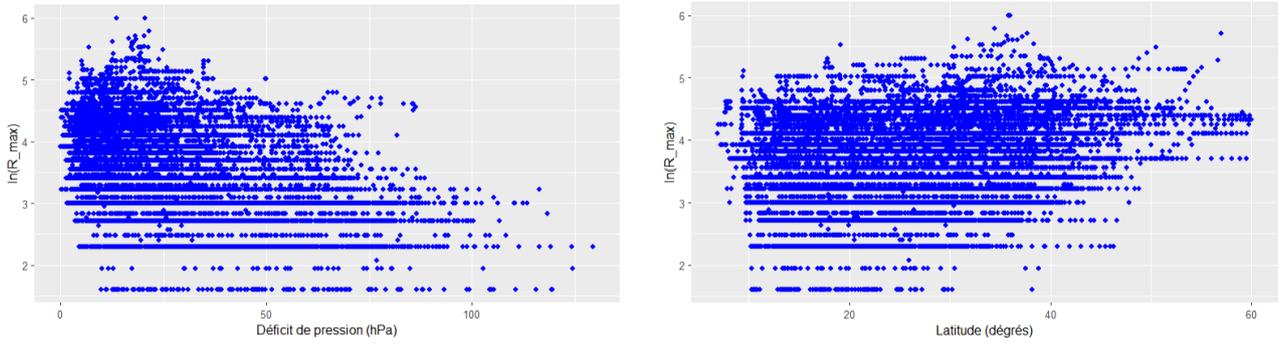
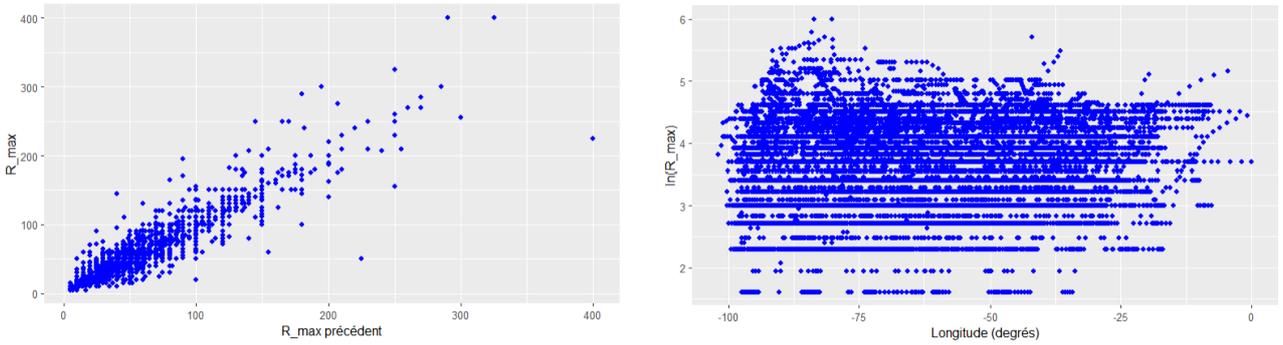
FIGURE 3.36 : Evolution du logarithme de R_{max} en fonction du déficit de pression et de la latitude(a) R_{max} en fonction de R_{max} précédent(b) $\ln R_{max}$ en fonction de la longitude

FIGURE 3.37 : Evolution du rayon de la vitesse maximale des vents en fonction d'autres variables

de pression, la latitude, ou le rayon de la vitesse maximale de vents et délaisse tout l'aspect physique de la structure des tempêtes cycloniques. De plus l'introduction de variables supplémentaires, telles que la longitude ou le déficit de pression, n'apporte que des améliorations marginales au modèle, et les résidus résultants ne satisfont pas l'hypothèse de gaussianité, remettant ainsi en question la validité du modèle. Face à ces limitations, de nouveaux modèles plus récents ont émergé pour combler ces lacunes et intégrer des informations physiques plus fondamentales et théoriques. L'un de ces modèles innovants, proposé par DANIEL et JOHN (2022), vise spécifiquement à remédier aux faiblesses des modèles précédemment cités. Ce modèle repose sur la modélisation du rayon de la vitesse maximale des vents en utilisant trois variables clés : la vitesse maximale des vents (V_{max}), la latitude (ϕ) et la moyenne des rayons maximum atteints par les vents de vitesse égale à 34 nœuds (soit 17,5 m/s) dans chaque quadrant ($R_{17.5}$).

Les ouragans sont caractérisées par des mouvements d'air en rotation où les parcelles d'air de la couche limite vers les grands rayons s'écoulent radialement vers l'intérieur en direction du centre comme évoqué en section 1.2. Ainsi pour un ouragan il existe une quantité dynamique physique très importante à prendre en compte qui est son moment angulaire absolu définie de la manière suivante (DANIEL et JOHN, 2022)

$$M = rV + \frac{1}{2}fr^2, \quad (3.16)$$

où :

- r représente le rayon (en km) ;
- V représente la composante tangentielle du vent (en m/s) ;

- $f = 2\Omega \sin \phi$ représente la force de Coriolis (en Newton) avec $\Omega = 7.292 \times 10^{-5} s^{-1}$ et ϕ la latitude au centre de l'ouragan.

Le premier terme de l'équation en (3.16) correspond au moment angulaire relatif associé à la circulation de l'ouragan. Le second terme correspond au moment angulaire planétaire associé à la projection de la rotation de la Terre sur l'axe de rotation de l'ouragan (vertical) au point de latitude du centre de l'ouragan.

A partir de cette notion, DANIEL et JOHN (2022) décrit le phénomène suivant. Fondamentalement, R_{max} dépend fortement de la circulation extérieure, c'est-à-dire au de-là de la valeur de R_{max} . Les auteurs expliquent qu'au fur et à mesure que les parcelles d'aires spirales progressent vers l'intérieur à partir des grands rayons jusqu'au rayon maximale de la vitesse des vents R_{max} , elles perdent progressivement leur élan angulaire en raison du frottement de surface. Par conséquent, le moment angulaire M diminue en se déplaçant radialement vers l'intérieur du système (phénomène nécessaire pour que l'inertie du système soit stable). Ce phénomène est illustré sur le graphique de la figure 3.38 où l'on représenté une évolution de la vitesse radiale (ou azimutale) et du moment angulaire absolu en fonction de la distance au centre d'un ouragan. Ainsi à partir de ces observations physiques, il est possible d'extraire une relation entre R_{max} , le rayon de la vitesse maximale des vents, et $R_{17.5}$, la moyenne des rayons maximum atteint par les vents de vitesse égale à 34 noeuds (= 17.5 m/s) selon la direction de chaque quadrans qui est utilisée comme caractéristique de la circulation extérieure (au delà de R_{max}). Cette relation est caractérisable en terme de fraction du moment angulaire perdu entre ces deux rayons. Ainsi nous obtenons la quantité $\frac{M_{max}}{M_{17.5}}$, cette quantité est toujours inférieure ou égale à 1 (si $V_{max} = 17.5$ m/S, le cas où les deux rayons sont égaux. Ainsi à partir de ces paramètres, le rayon maximal de la vitesse des vents s'obtient à travers les étapes suivantes.

Etape 1 : Calcul de $M_{17.5}$ en utilisant (3.16) (Inputs : $R_{17.5}$ et f)

$$M_{17.5} = R_{17.5} * 17.5 + \frac{1}{2} f R_{17.5}^2 ;$$

Etape 2 : Calcul de M_{max} (Inputs : $M_{17.5}$)

$$M_{max} = \left(\frac{M_{max}}{M_{17.5}} \right) M_{17.5} ;$$

Etape 3 : Déduction de R_{max} à partir de M_{max} (Inputs : V_{max} et f)

$$R_{max} = \frac{V_{max}}{f} \left(\sqrt{1 + \frac{2fM_{max}}{V_{max}^2}} - 1 \right).$$

Ces étapes sont résumées à travers le schéma de la figure 3.38. De plus, il convient de noter que la valeur du rayon $R_{17.5}$ est initialement une valeur calculée pour chaque quadrant (nord-est, sud-est, nord-ouest et sud-ouest). DANIEL et JOHN (2022) souligne qu'il est important d'établir une moyenne entre les quadrants plutôt que de considérer chaque quadrant individuellement pour la valeur finale de $R_{17.5}$. L'une des raisons évoquées par l'auteur est le fait que les estimations opérationnelles de R_{max} ne sont actuellement pas spécifiques à un quadrant en raison des difficultés complexes à déterminer ces valeurs et donc le calcul de la moyenne des quadrants réduit le bruit associé à l'estimation de R_{max} ainsi que les asymétries azimutales dans la structure des tempêtes.

Ainsi, ce qui manque dans ce modèle est l'estimation du terme $\frac{M_{max}}{M_{17.5}}$, qui traduit l'écart angulaire perdu entre les rayons R_{max} et $R_{17.5}$. Pour cela, les auteurs proposent l'utilisation d'un modèle de régression linéaire basé sur le logarithme de $\frac{M_{max}}{M_{17.5}}$ comme suit

$$\ln \left(\frac{M_{max}}{M_{17.5}} \right) = a_0 + a_1(V_{max} - 17.5) + a_2(V_{max} - 17.5) \left(\frac{1}{2} f R_{17.5} \right) + \varepsilon, \quad (3.17)$$

où :

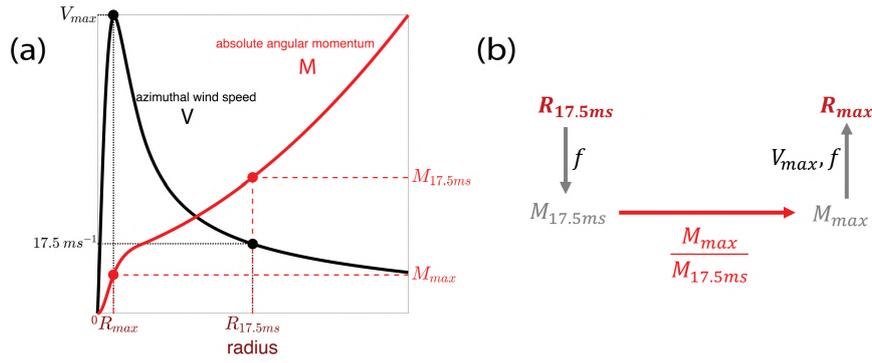


FIGURE 3.38 : Cadre conceptuel du modèle et exemple d'une évolution de la vitesse azimuthale du vent (noir) et du moment angulaire absolu (rouge) au sein d'un ouragan (DANIEL et JOHN, 2022)

- $\frac{M_{max}}{M_{17.5}}$ représente la fraction du moment angulaire perdu entre les rayons R_{max} et $R_{17.5}$;
- V_{max} représente la vitesse maximale des vents (en m/s) ;
- $R_{17.5}$ est la moyenne des rayons maximum atteint par les vents de vitesse égale à 17.5 m/s selon chaque quadrant ;
- f représente la force de Coriolis (en Newton) ;
- a_0 , a_1 et a_2 sont les coefficients de régression à estimer ;
- ε représente un terme d'erreur.

Ce modèle suggère que $\frac{M_{max}}{M_{17.5}}$ dépend de deux paramètres physiques, à savoir V_{max} et $\frac{1}{2}fR_{17.5}$. Le premier terme de l'équation (3.17), $(V_{max} - 17.5)$, représente l'intensité perdue au sein de l'ouragan en un point donné entre les rayons R_{max} et $R_{17.5}$. Le deuxième terme, $(V_{max} - 17.5)(\frac{1}{2}fR_{17.5})$, est conçu pour capturer la dépendance non linéaire entre ces deux paramètres, comme observé visuellement par les auteurs. Le paramètre $\frac{1}{2}fR_{17.5}$ combine des informations sur la taille extérieure de l'ouragan, à travers le rayon $R_{17.5}$, et la force de Coriolis f .

Afin de disposer d'un ensemble de données optimal pour leur modèle et se concentrer sur les zones d'études ayant un intérêt pour les organismes de prévision, les auteurs se concentrent sur la moitié ouest du bassin Atlantique Nord. Ainsi la base de données est filtrée de la manière suivante :

- Longitude < -50 , pour se concentrer sur la région principale de la reconnaissance aérienne ;
- Latitude < 30 , pour minimiser les effets de la transition vers des tempêtes tropicales et se concentrer sur les zones où les ouragans sont présents ;
- DIST2LAND $> R_{17.5}$, pour minimiser les effets de la terre sur la structure interne de l'ouragan ;
- Au moins trois valeurs de $R_{17.5}$ pour chaque quadrant afin d'obtenir une moyenne raisonnable ;
- $0 \leq \frac{M_{max}}{M_{17.5}} \leq 1$ pour éliminer les valeurs non physiques qui pourraient résulter du bruit ;
- $6 \leq R_{max} \leq 250$ pour exclure les valeurs aberrantes extrêmes ;
- $V_{max} \geq 20 \text{ m/s}$ pour inclure uniquement les tempêtes cycloniques avec une vitesse du vent supérieure à 20 m/s.

Le tableau 3.15 expose les résultats issus de l'estimation du modèle de régression (3.17). L'écart-type des résidus du modèle, noté σ_ε , est également inclus dans ce tableau. Par ailleurs, le tableau 3.16 met en avant l'écart-type de chaque coefficient estimé, accompagné des intervalles de confiance à 95% associés. Il est notable que les écart-types des coefficients sont très faibles, ce qui témoigne de la précision et de la robustesse du modèle. De plus, les intervalles de confiance à 95% sont très étroits, ce qui indique que les estimations des coefficients sont très fiables. Cette robustesse et cette précision renforcent la confiance dans la capacité du modèle à rendre compte de la relation entre les variables explicatives et la variable cible.

a_0	a_1	a_2	σ_ε
0.6517	-0.0114	0.0613	0.1831

TABLE 3.15 : Résultats des estimations du modèle de régression (3.17)

σ_{a_0}	IC95%	σ_{a_1}	IC95%	σ_{a_2}	IC95%
0.008	[0.74; 0.77]	0.00001	[-0.001; -0.0005]	0.009	[-0.02; -0.01]

TABLE 3.16 : Incertitudes relatives aux estimations des coefficients

Pour simuler les saisons cycloniques, le modèle défini en (3.17) nécessite une valeur initiale de $R_{17.5}$, c'est-à-dire la moyenne des rayons de la vitesse des vents de 17.5 m/s selon chaque quadrant. Cette valeur initiale est tirée de l'échantillon historique des valeurs initiales de $R_{17.5}$ après avoir observé que la gamme de valeurs parcourues par ces valeurs initiales est assez large (de 20 à 250 km) lorsque l'intensité de la tempête est faible. Cette hypothèse reste donc raisonnable dans le cadre de l'étude.

Afin de tester le modèle établi en (3.17), les rayons de vitesses de vents maximales sont simulées sur les ouragans historiques. La figure 3.39 présente les résultats du test du modèle de rayon de la vitesse maximale des vents proposé par DANIEL et JOHN (2022). Globalement, il y a une bonne adéquation entre les rayons empiriques et simulés, en particulier pour les rayons de faible amplitude (inférieurs à 100 km), caractéristiques des ouragans de forte intensité. L'adéquation est moins bonne pour les rayons plus grands, mais le modèle reste prudent car il sous-estime ces rayons plus grands en les fixant à des valeurs plus faibles que celles observées empiriquement. Le QQ-Plot des résidus avec les quantiles théoriques de la distribution normale montre globalement une bonne adéquation, même si les quantiles extrêmes sont légèrement mal pris en compte. L'hypothèse de normalité des résidus reste cependant acceptable, contrairement au premier modèle de rayon proposé précédemment.

En conséquence, le modèle de rayon de la vitesse maximale des vents proposé par DANIEL et JOHN (2022) est retenu pour la simulation des saisons cycloniques. Ce modèle est préféré en raison de sa capacité à prendre en compte des phénomènes physiques et des relations empiriques observées entre les différentes variables régissant la structure d'un ouragan. De plus, il est moins complexe et produit des résultats plus convaincants en termes de résidus et de prudence du modèle vis-à-vis des ouragans de forte intensité.

3.2.2.3 Gradient de vitesse maximale des vents

La dernière caractéristique à considérer dans le cadre de l'étude est la variation de la vitesse maximale des vents en fonction de la distance au centre de la tempête. Mentionné dans la section 1.3 et illustré par la figure 1.8, il est évident que les vents ressentis sur les terres ne reflètent pas directement la vitesse maximale des vents (V_{max}) enregistrée pour une tempête en une localisation donnée. En effet, une tempête évoluant en ouragan peut posséder un diamètre atteignant jusqu'à 1000 km (section 1.2.3.2), et ainsi, les vents en périphérie de l'ouragan ne correspondent pas à l'intensité observée près de

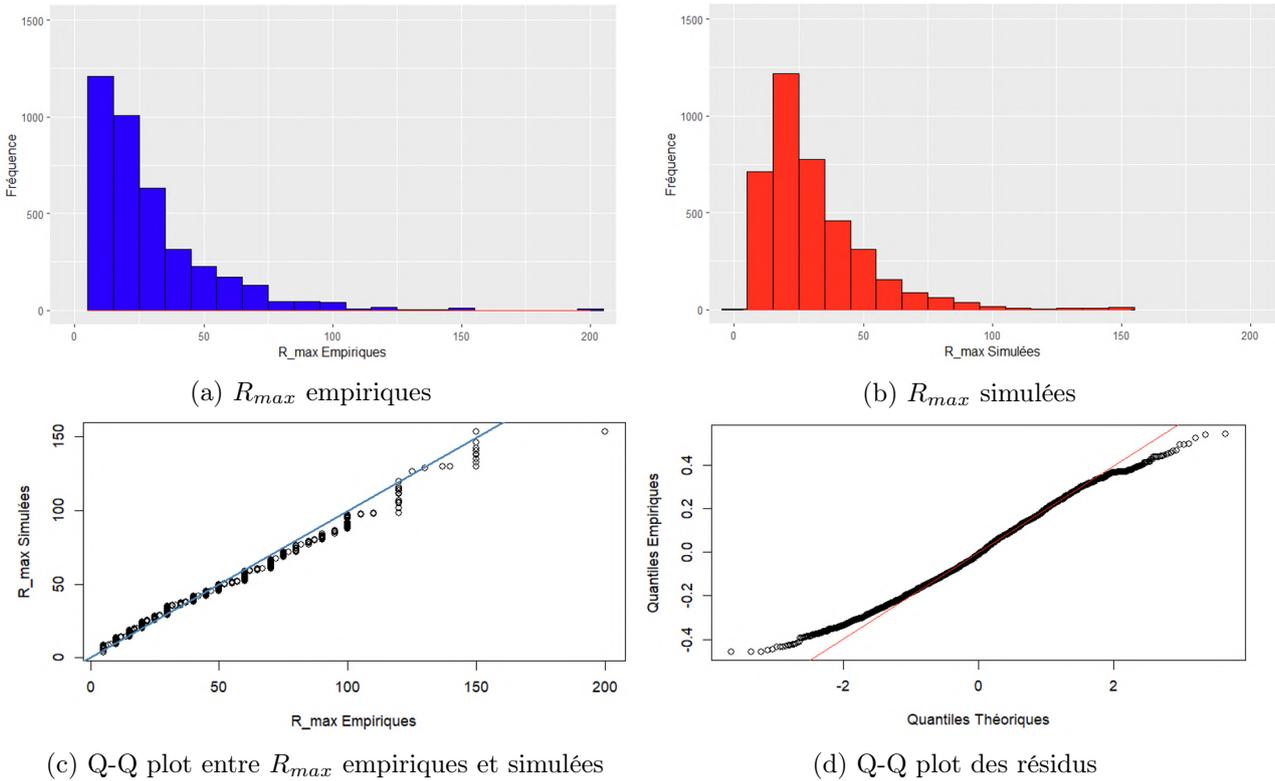


FIGURE 3.39 : Résultats des tests du modèle de rayon de la vitesse maximale des vents proposé par DANIEL et JOHN (2022)

son centre. De surcroît, les vitesses de vents aux extrémités d'un ouragan peuvent exercer leur influence sur une distance pouvant s'étendre sur plusieurs dizaines de kilomètres, particulièrement dans le cas d'ouragans de forte intensité, causant ainsi des dommages significatifs sur une vaste superficie adjacente au cœur de l'ouragan. Par conséquent, l'objectif de cette phase consiste à élaborer un modèle qui permette de déterminer la vitesse maximale des vents sur l'ensemble de la superficie d'un ouragan à un instant donné, traduisant ainsi le gradient de la vitesse maximale des vents pour une position spécifique. Cette démarche repose sur l'hypothèse que la structure d'un ouragan est approximativement circulaire, une hypothèse fréquemment observée historiquement (exemple avec l'ouragan Irma en figure 1.6) et également adoptée dans diverses modélisations du gradient de vitesse des vents (WANG et ROSOWSKY, 2012).

L'un des premiers modèles élaborés est celui avancé par JELESNIANSKI et al. (1992). Ce modèle propose d'exprimer la vitesse, notée v , en fonction de la distance, notée r , du centre de l'ouragan, selon la formule suivante :

$$v(r) = V_{max} \times \frac{2rR_{max}}{r^2 + R_{max}^2}$$

Cependant, il est clair que ce modèle se révèle excessivement simpliste, car il ne prend en considération que deux caractéristiques inhérentes à un ouragan, à savoir la vitesse maximale des vents (V_{max}) et son rayon (R_{max}), négligeant ainsi un éventail important de facteurs présent au sein d'un ouragan.

Par conséquent, au cours des dernières décennies, des modèles nettement plus sophistiqués ont été élaborés. Cependant, la complexité de leur mise en œuvre peut varier considérablement. Dans l'article de EGUCHI et al. (2021), une analyse comparative approfondie de l'efficacité et de la validité de diverses modélisations de gradients de vitesse établies ces dernières années est proposée. Dans le cadre de cette étude, le modèle de gradient de vitesse utilisé repose sur la proposition de WILLOUGHBY et al. (2006),

dont la modélisation mathématique est exposée dans l'annexe A.8. Cette modélisation a été développée par les auteurs dans le but d'améliorer la précision de la modélisation de la vitesse du vent à mesure que l'on s'éloigne du centre de l'œil de la tempête.

En effet, WILLOUGHBY et al. (2006) explique que dans le contexte des applications en assurance, telles que la tarification des garanties contre les tempêtes de vents, les vents des ouragans sont souvent approximés par des fonctions continues qui sont nulles au centre de l'œil de l'ouragan, augmentent jusqu'à un maximum dans le mur de l'œil, puis diminuent asymptotiquement jusqu'à zéro à mesure que l'on s'éloigne du centre. Cependant, les observations empiriques ne corroborent pas précisément cette dynamique. Bien que les vents à proximité du pic (pression minimale) soient excessivement puissants, leur intensité diminue assez rapidement à mesure que l'on s'éloigne du pic. Ainsi, il y a souvent une tendance à surestimer les vitesses maximales des vents. À titre d'exemple, la figure 3.40 illustre l'évolution du gradient de la vitesse maximale soutenue des vents lors de l'ouragan Katrina en 2005, à son intensité maximale, c'est-à-dire $V_{max} = 280\text{km/h} \approx 78\text{m/s}$ et $R_{max} = 37\text{km}$. Cette illustration vise à mettre en évidence la motivation sous-jacente à l'approche proposée par WILLOUGHBY et al. (2006).

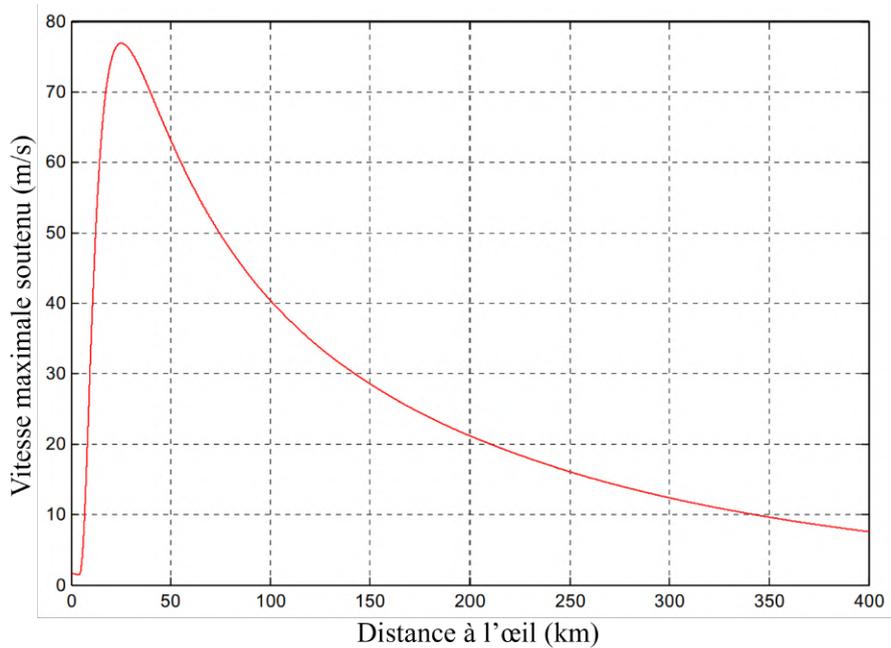


FIGURE 3.40 : Gradient de vitesse des vents de l'ouragan Katrina à son intensité maximale (WANG et ROSOWSKY, 2012)

Par conséquent, WILLOUGHBY et al. (2006) propose une alternative en structurant la modélisation en trois phases. Une première phase de modélisation lors de laquelle la vitesse des vents augmente en fonction d'une puissance du rayon à l'intérieur de l'œil, une deuxième phase de transition polynomiale lisse à travers le mur de l'œil, suivie d'une troisième phase de décroissance exponentielle à l'extérieur de l'œil. L'approche en question est exposée à travers les équations définies en (3.18)

$$V(r) = V_i = V_{max} \left(\frac{r}{R_{max}} \right)^n, \quad (0 \leq r \leq R_1), \quad (3.18a)$$

$$V(r) = V_i(1 - w) + V_o w, \quad (R_1 \leq r \leq R_2), \quad (3.18b)$$

$$V(r) = V_o = V_{max} \left[(1 - A)e^{\frac{R_{max} - r}{X_1}} + Ae^{\frac{R_{max} - r}{X_2}} \right], \quad (R_2 < r), \quad (3.18c)$$

où $V(r)$ représente la vitesse maximale des vents, en m/s, à une distance r (en km) du centre de l'ouragan, V_i désigne la composante tangentielle du vent à l'intérieur de l'œil de l'ouragan et V_o au-delà de la zone de transition. L'ensemble des paramètres définies dans les équations (3.18) sont expliqués dans l'annexe A.8.

Cette approche, élaborée selon la proposition de WILLOUGHBY et al. (2006) et accessible au moyen du package *stormwindmodel* (ANDERSON et al., 2020), permet de déduire un gradient de vitesse maximale des vents pour chaque position d'un ouragan. Cette méthode a la capacité de déterminer les régions terrestres périphériques à l'ouragan qui ont été impactées par les vents de ce dernier. Un exemple concret de cette application est illustré sur la figure 3.41, où le gradient de vitesse maximale des vents de l'ouragan Katrina en 2005 est représenté pour chaque comté des États-Unis.

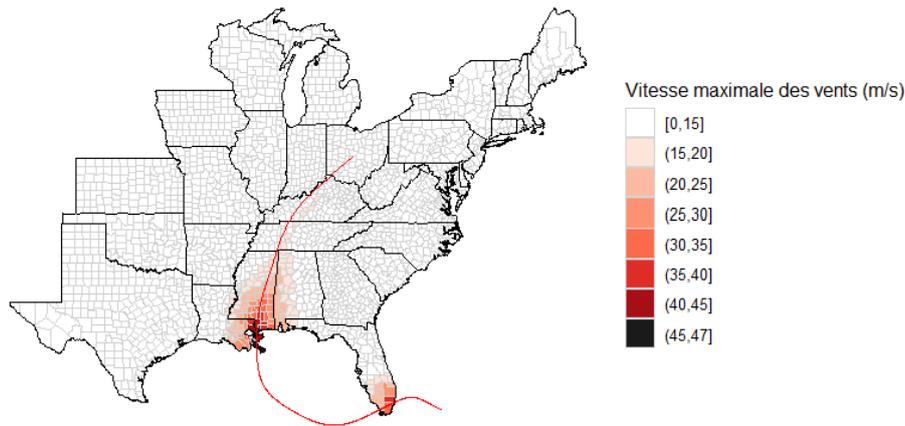


FIGURE 3.41 : Répercussions territoriales de la vitesse des vents de l'ouragan Katrina aux États-Unis

Les données concernant chaque comté des États-Unis sont collectées depuis le site officiel du Bureau de recensement des États-Unis (UNITED STATES CENSUS BUREAU, 2021). Pour chaque comté, les informations recueillies englobent le nombre d'habitants ainsi que les coordonnées géographiques (longitude, latitude) du centre de population moyen. Ce point central est déterminé de manière à attribuer un poids à chaque emplacement correspondant à la position d'une personne.

Pour conclure sur la modélisation statistique d'un ouragan, le tableau 3.17 présente une synthèse de la base simulée comprenant les caractéristiques des 10 000 saisons cycloniques générées. Les deux premiers points d'impact de la première et de la dernière tempête simulée sont représentés. Chaque ligne correspond à un point d'impact d'intervalle de 3 heures d'une tempête au cours d'une saison cyclonique spécifique parmi les 10 000 saisons simulées.

Saison	N°tempête	Latitude	Longitude	P (hPa)	V_{max} (km/h)	R_{max} (km)	SS
1	1	18.6	-45.6	994	70	50	0
1	1	18.8	-47.7	990	80	44	0
...
10 000	16	25.2	-75.4	953	190	20	3
10 000	16	26.2	-76.1	974	134	35	1
...

TABLE 3.17 : Allure de la base simulée des 10 000 saisons cycloniques

La modélisation des 10 000 saisons cycloniques étant achevée, le chapitre suivant se consacre à l'évaluation du risque cyclonique sur le territoire des États-Unis. L'objectif ultime consiste à estimer le coût

brut associé à chaque ouragan simulé ayant causé des dommages dans chaque comté des États-Unis. Cette démarche vise à élaborer une cartographie du territoire américain, fournissant des informations essentielles à l'industrie de l'assurance. Cette cartographie permet d'éclairer l'exposition de leur portefeuille au risque cyclonique aux États-Unis, d'orienter leur politique de souscription en conséquence, et de fournir des éléments préliminaires quant à l'assurabilité de certaines zones. La démarche repose sur une analyse approfondie des pertes financières potentielles, contribuant ainsi à une prise de décision éclairée dans le secteur de l'assurance face au risque cyclonique aux États-Unis.

Chapitre 4

Modélisation du coût financier

L'objectif de ce chapitre est de proposer une modélisation du coût financier entraîné par le passage d'un ouragan sur le sol américain. Pour ce faire, plusieurs approches sont envisageables selon la nature des données à disposition. Dans le cas de données assurantielles contenant par exemple des contrats multirisques habitation, il devient envisageable de construire une fonction liant l'intensité des ouragans aux dommages subis par les bâtiments assurés. La caractéristique retenue pour l'intensité des ouragans est le plus souvent la vitesse maximale des vents, permettant une classification plus aisée selon l'échelle de Saffir-Simpson.

La mesure des dommages infligés aux bâtiments s'effectue par le biais d'un taux de destruction, défini comme le rapport entre le coût de la réparation du bâtiment et son coût total de reconstruction. Cette relation engendrée est généralement identifiée sous le terme de courbe de vulnérabilité. De diverses approches pour concevoir ces courbes de vulnérabilité sont présentes dans la littérature. Une étude approfondie des modèles de vulnérabilité existants dans le domaine scientifique et appliqué dans le secteur assurantiel est présentée par GONZALO et al. (2015). Parmi ces modèles, deux schémas méthodologiques se distinguent fréquemment : empirique et théorique.

Méthode empirique : cette méthode repose sur l'estimation des courbes de vulnérabilité via l'utilisation de modèles de régression appliqués aux données de sinistres. Ce processus d'estimation permet d'associer à une vitesse maximale des vents donnée, le taux de destruction des biens assurés. Cependant, l'application de cette approche requiert la disponibilité d'un portefeuille d'assurances ou de données de sinistres bénéficiant d'un historique substantiel de pertes occasionnées par des ouragans touchant des bâtiments de divers niveaux d'intensité.

Méthode théorique : cette méthode consiste à établir une relation directe entre le taux de destruction causé par un ouragan et la vulnérabilité des bâtiments. Une telle approche n'exige pas de disposer d'un historique de sinistres antérieur, mais requiert une expertise plus technique et sollicite l'avis d'experts du domaine de l'ingénierie et de la météorologie. À titre d'exemple, EARLE (1997), dans son rapport sur les risques naturels aux îles Vierges britanniques, propose des courbes de vulnérabilité théoriques en fonction de la catégorie des structures et de la composition des bâtiments. Néanmoins, cette approche nécessite d'accéder à des données précises sur la structure de chaque bâtiment inclus dans le portefeuille d'assurances, ce qui n'est pas nécessairement le cas.

Dans le cadre de cette étude, l'absence de données assurantielles exclut la disponibilité de tout historique de sinistres ou de biens assurés. Par conséquent, la modélisation de courbes de vulnérabilité n'est pas envisagée. À la place, la démarche adoptée se focalise sur l'établissement d'une fonction de coût corrélée au risque ouragan au moyen de données accessibles au public. Le but de ce chapitre réside donc dans la proposition d'une méthodologie complète pour la création de cette fonction. Cette approche

repose sur la construction d'une base de données à partir de données en open source, l'intégration de diverses variables d'exposition, l'analyse des résultats obtenus et la mise en évidence des limites rencontrées. Une fois cette fonction de perte évaluée, sa mise en œuvre sur les saisons cycloniques simulées permettra d'estimer un coût moyen annuel par foyer à travers l'ensemble des régions analysées, c'est-à-dire les territoires exposés aux ouragans aux États-Unis.

4.1 Construction d'une fonction de perte au risque ouragan aux États-Unis

Dans cette section, la méthodologie de construction de la base de données d'exposition en utilisant les données disponibles en open source est exposée. Par la suite, le choix de la variable de prédiction, correspondant au coût financier associé à un ouragan pour chaque comté du territoire américain est présenté. La partie suivante est dédiée à la modélisation de la fonction de perte, incluant les étapes de sélection des variables. Enfin, une analyse approfondie des résultats obtenus est réalisée.

4.1.1 Construction d'une base de données coût par comté

La conception de la base de données utilisée pour la modélisation du coût financier des ouragans repose entièrement sur l'utilisation de données accessibles en libre accès, obtenues à partir de plusieurs sites officiels des États-Unis. Ces sites incluent l'Agence américaine d'observation océanique et atmosphérique (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023a), l'Agence fédérale des situations d'urgence (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023a) et le Bureau du recensement des États-Unis (UNITED STATES CENSUS BUREAU, 2021). Par la suite, les caractéristiques de chaque variable extraite depuis ces sources sont exposées.

4.1.1.1 Collecte des données de coûts historiques par Etat

Dans la section 1.5 du premier chapitre, une synthèse des stratégies et des dispositifs d'assurance a été présentée, de même que les dix ouragans historiques les plus coûteux aux États-Unis. Il est envisageable d'élargir cette liste pour inclure l'ensemble des ouragans qui ont engendré des coûts dépassant un milliard de dollars aux États-Unis au cours de la période allant de 1982 à 2022. Cette liste s'appuie sur un échantillon de 41 ouragans qui ont généré un coût excédant le milliard de dollars, et la trajectoire de ces ouragans est illustrée sur la carte présentée en figure 4.1.

Il convient de noter que ces coûts totaux sont ajustés en fonction de l'indice des prix à la consommation de l'année 2023. Cette carte met en évidence les États américains les plus exposés à ces phénomènes dévastateurs, notamment le Texas, la Louisiane et la Floride, qui sont les plus touchés. Il est important de souligner quelques aspects liés à la notion de coût utilisée dans le cadre de cette étude. Les coûts de chaque ouragan extraits des données de la NOAA sont des estimations accompagnées d'un intervalle de confiance de 95%, calculé au moyen de simulations de Monte Carlo. Ces coûts totaux estimés englobent les pertes assurées et non assurées, provenant de diverses sources. Ces sources incluent le Service météorologique national (*National Weather Service*), l'Agence fédérale des situations d'urgence, le département de l'agriculture des États-Unis (*U.S. Department of Agriculture*), le Corps du génie de l'armée des États-Unis (*U.S. Army Corps of Engineers*), les agences de gestion des urgences des États, les centres climatiques régionaux et d'État, les rapports des médias et les estimations du secteur de l'assurance. En plus de ces coûts totaux estimés, la NOAA fournit également, pour chaque ouragan représenté sur la carte 4.1, un intervalle de coût total de dommages pour chaque État américain où l'ouragan a pu causer des dommages.

Par exemple, la Figure 4.2a présente en détail les intervalles de coût total par État pour l'ouragan Sandy, survenu en 2012 et classé comme le 5ème ouragan le plus coûteux aux États-Unis. Toutefois,

4.1. CONSTRUCTION D'UNE FONCTION DE PERTE AU RISQUE OURAGAN AUX ETATS-UNIS129

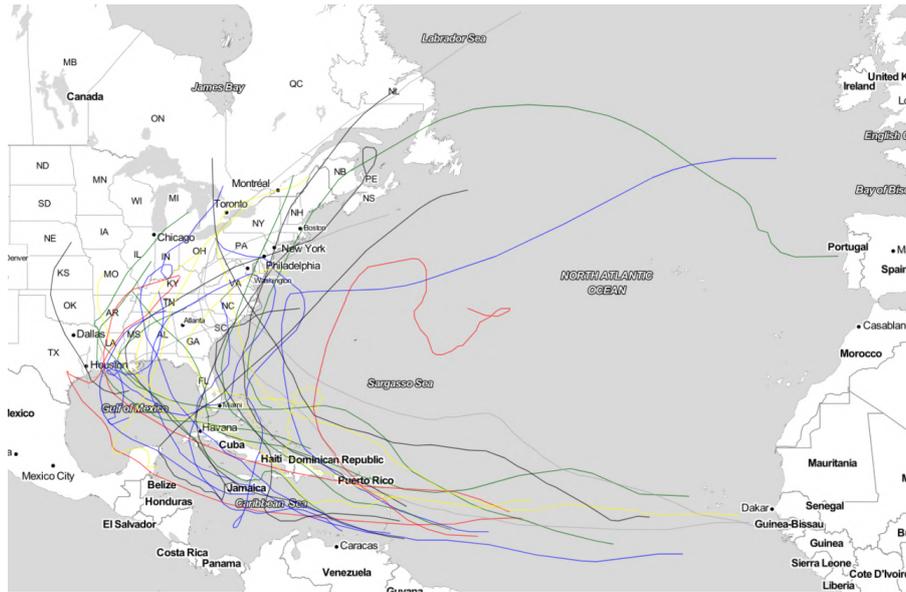
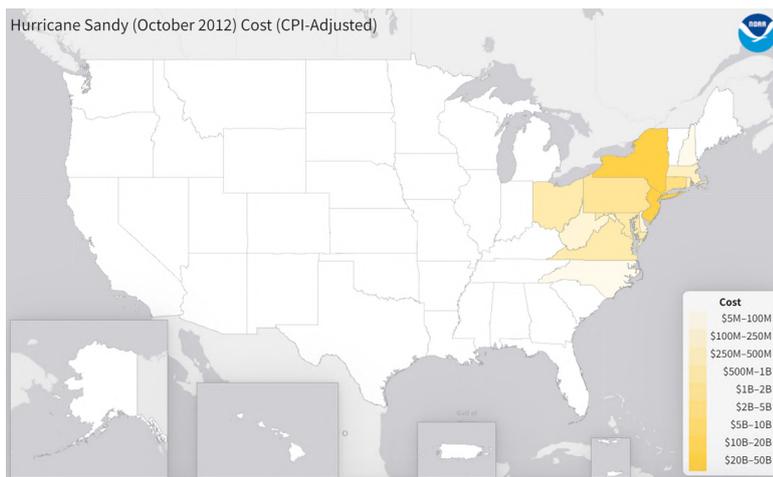


FIGURE 4.1 : Trajectoire des 41 ouragans ayant coûté plus d'un milliard de dollars aux Etats-Unis

il convient de noter que ces coûts sont limités à une borne inférieure de 5 millions de dollars, ce qui ne reflète pas intégralement le coût global engendré par un ouragan. Cette limitation peut entraîner des implications significatives du point de vue de la modélisation des coûts financiers. Par exemple, l'ouragan Sandy a occasionné des dommages en Floride au cours de son trajet vers le nord-est des États-Unis, même si ces incidents ne sont pas représentés dans la Figure 4.2a. En outre, la Figure 4.2b met en évidence un point d'impact de l'ouragan Sandy lorsqu'il était en catégorie 1, passant à une distance d'environ 300 kilomètres au large de la côte de la Floride, où des vitesses de vent atteignant jusqu'à 110 km/h ont été enregistrées (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2012). Cet exemple démontre donc que les ouragans ont la capacité de causer des dommages même lorsqu'ils se trouvent à une certaine distance des zones terrestres.



(a) Coût total par Etat de l'ouragan Sandy (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023a)



(b) Ouragan Sandy passant à environ 300km de la Floride en catégorie 1

FIGURE 4.2 : Ouragan Sandy

Ainsi, dans le but de prendre en compte ces coûts inférieurs à 5 millions de dollars, une approche

simpliste basée sur l'observation de l'ouragan Sandy est effectuée. Si la base de données de la NOAA ne contient pas d'intervalle de coût pour un État donné, mais qu'il existe un ouragan d'au moins catégorie 1 passant à une distance d'au moins 300 kilomètres de la côte terrestre de cet État, alors il est considéré que cet État a subi des dommages d'un montant inférieur à 5 millions de dollars. Cette hypothèse trouve également sa justification dans l'observation d'autres ouragans parmi les 41 ouragans ayant coûté plus d'un milliard de dollars aux États-Unis, comme illustré dans la Figure 4.1, qui ont présenté un comportement similaire à celui de l'ouragan Sandy.

Ainsi, chaque ligne de la base de données représente un ouragan, avec un État où des dommages ont été causés et un intervalle de coût associé à ces dommages. Cependant, pour construire une fonction de perte à travers des modèles statistiques, il est préférable de travailler avec des coûts réels plutôt qu'avec des intervalles de coût, car cela permet d'utiliser des modèles de régression plutôt que des modèles de classification. De plus, d'un point de vue assurantiel, il est plus avantageux d'obtenir des estimations de coûts réels pour une meilleure gestion des risques en matière de tarification, de souscription et de provisionnement.

Pour ce faire, au lieu de considérer l'intervalle de coût, il est choisi de prendre en compte la médiane de cet intervalle afin d'obtenir un coût réel. Afin de vérifier la cohérence de cette approche, la Figure 4.3 représente le coût réel total de chaque ouragan aux États-Unis, tel qu'il a été rapporté par la NOAA (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023a), en fonction de la somme des médianes des intervalles de coût de chaque ouragan qui a affecté un État.

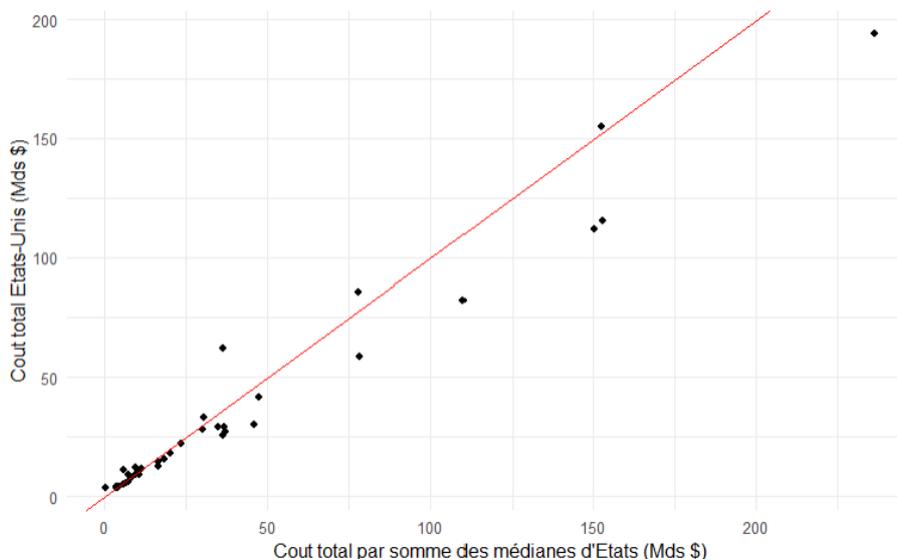


FIGURE 4.3 : Coût total en fonction du coût total par somme des médianes

L'observation de ce graphique révèle que l'hypothèse de considérer la médiane de chaque intervalle de coût constitue une approche plutôt cohérente et acceptable. En effet, il existe une relation globalement linéaire entre le coût total et la somme des médianes des coûts de chaque intervalle. Toutefois, cette relation devient moins linéaire lorsque les coûts totaux dépassent les 100 milliards de dollars.

Il est à noter que cette approche se montre relativement prudente dans ses estimations, notamment dans le cas de l'ouragan Katrina, dont le coût total estimé avoisine les 193 milliards de dollars, tandis que l'approche considérée donne un coût de plus de 230 milliards de dollars pour cet événement.

4.1.1.2 Calcul du coût par comté à l'aide d'un indice de risque ouragan

Dans le cadre de l'objectif central de cette étude, qui consiste à modéliser une carte d'exposition au risque ouragan aux États-Unis, il s'avère pertinent de se pencher sur une échelle de résolution plus

fine que celle des États. Les préoccupations des compagnies d'assurance se concentrent en effet davantage sur des zones géographiques spécifiques plutôt que sur des entités étatiques dans leur globalité. Par conséquent, le choix est fait de modéliser le coût financier lié aux ouragans au niveau des comtés aux États-Unis. Cette approche permettra, dans une perspective assurantielle, de mieux appréhender l'exposition et l'impact financier d'un ouragan sur des zones géographiques plus restreintes. Elle offrira aux assureurs une vision plus précise de l'exposition de leurs portefeuilles aux risques ouragan dans ces différentes localités. Afin d'obtenir un coût financier réel associé à un ouragan pour chaque comté des États-Unis, des données spécifiques relatives au risque ouragan dans chaque comté s'avèrent nécessaires. Dans cette optique, une étude a été conduite par la FEMA en collaboration étroite avec divers intervenants et partenaires issus du monde universitaire, des autorités locales, étatiques et fédérales, ainsi que du secteur privé notamment des assureurs et réassureurs (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023c).

Après de nombreuses années de recherche et de collaboration entre l'ensemble des parties prenantes, les résultats de cette étude ont été rendus publics en 2023 et ont conduit à l'élaboration de l'indice de risque national, *National Risk Index*, pour chaque comté des États-Unis exposé à un risque naturel.

Pour chaque comté, cet indice de risque national se décompose en trois composantes essentielles :

- **Value** : Il s'agit de la perte économique moyenne du comté, exprimée en dollars, résultant des aléas naturels chaque année ;
- **Score** : Cette composante reflète la position du comté parmi l'ensemble des comtés nationaux en ce qui concerne la composante *Value*. Il est exprimé en percentiles, ce qui permet de comparer le comté aux autres comtés à l'échelle nationale ;
- **Rating** : Cette évaluation attribue une note au comté en fonction de sa composante *Value* par rapport à l'ensemble des autres comtés. Il existe cinq catégories de notes, allant de *Very Low* à *Very High*, permettant de classer le niveau de risque du comté.

Ainsi, la connaissance du calcul de la composante *Value* permet d'obtenir les deux autres composantes de l'indice. La composante *Value* est déterminée en prenant en considération trois paramètres spécifiques à chaque comté. Ces trois paramètres sont les suivants :

- **Expected Annual Loss** : notée *EAL*, représente la perte économique moyenne annuelle en dollars dû à un risque naturel au sein du comté ;
- **Social Vulnerability** : La vulnérabilité sociale est définie de manière générale comme la susceptibilité des groupes sociaux aux effets néfastes des aléas naturels, notamment la mort, les blessures, la perte ou la perturbation des moyens de subsistance, de manière disproportionnée ;
- **Community Resilience** : Il s'agit de la capacité d'une communauté à se préparer aux risques naturels anticipés, à s'adapter aux conditions changeantes, ainsi qu'à résister aux perturbations et à se rétablir rapidement.

Les détails relatifs au calcul de l'indice de risque national, ainsi que ceux concernant ses paramètres sous-jacents, ne sont pas inclus dans cette étude, principalement en raison de leur étendue (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023c). Dans le cadre de cette recherche, seules les variables pertinentes pour la construction de cet indice de risque national sont présentées, ainsi que les ajustements qui en découlent.

Cette étude se focalise sur l'indice de risque national associé au risque ouragan, ainsi que les variables qui y sont rattachées. En plus de cet indice, la FEMA met également à disposition en libre accès les données utilisées pour sa construction (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023a). La figure 4.4 illustre la notation de l'indice de risque national attribuée à chaque comté des États-Unis.

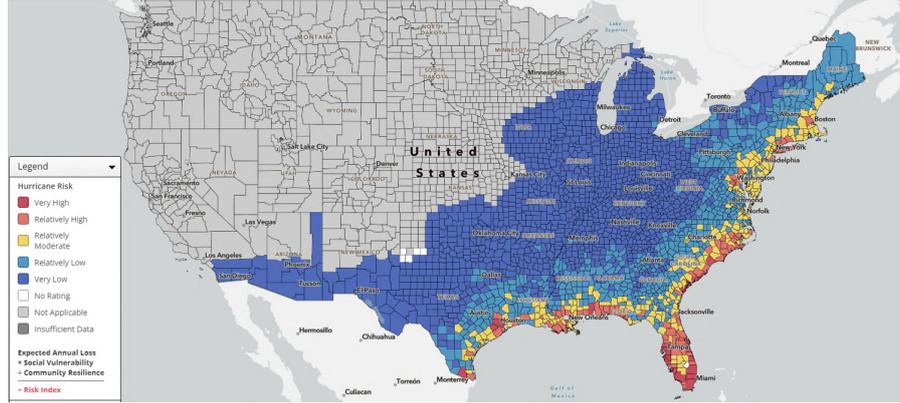


FIGURE 4.4 : *National Risk Index Rating* liée au risque ouragan aux Etats-Unis (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023d)

Cette carte met en évidence les régions les plus exposées aux ouragans, principalement les zones côtières du Golfe du Mexique et de l'océan Atlantique.

Ainsi, avec toutes ces informations à disposition, le paramètre retenu comme indice pour obtenir un coût réel pour chaque comté d'un État affecté par un ouragan est l'*Expected Annual Loss*. Ce choix se justifie par le fait que ce paramètre, par définition, offre une vision claire de la perte économique annuelle engendrée par un ouragan au sein d'un comté. Il prend en compte l'exposition du comté en termes de population, de valeur des bâtiments et d'activités agricoles, ainsi que le taux historique de pertes liées aux ouragans et la fréquence annuelle des ouragans dans chaque comté. Par conséquent, ce paramètre intègre un ensemble d'informations essentielles concernant le risque ouragan dans le comté et représente une valeur pertinente pour les besoins de cette étude.

Ainsi, en notant i comme un comté quelconque au sein d'un État j , le coût réel du comté i au sein de l'État j est calculé comme suit

$$\text{Coût}_i^j = \text{Coût}_j \times \frac{EAL_i^j}{\sum_{k=1}^N EAL_k^j} \quad (4.1)$$

où :

- Coût_i^j représente le coût réel total au sein du comté i de l'État j , la variable d'intérêt de l'étude ;
- Coût_j représente le coût réel total au sein de l'État j , disponible à travers la collecte des données de coût historique réalisé en section 4.1.1.1 ;
- EAL_i^j représente la perte économique moyenne annuelle en dollars dû à un risque ouragan au sein du comté i de l'État j , disponible à travers l'étude menée par la FEMA (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023a) ;
- N représente le nombre de comtés au sein de l'État j , également disponible à travers l'étude menée par la FEMA (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023a).

Afin d'évaluer la cohérence de l'approche utilisée, la figure 4.5 présente le coût total par comté lié aux ouragans sur la période de 1982 à 2022. Ce coût total est calculé en faisant la somme de tous les coûts obtenus pour chaque comté à l'aide de l'équation (4.1). Cette carte met en évidence les régions ayant subi les dommages les plus importants, à savoir les zones côtières près du Golfe du Mexique et de la côte Est de l'Atlantique. Les coûts attritionnels (inférieurs à 100 millions de dollars) se concentrent principalement dans les zones intérieures du pays, tandis que les coûts graves (supérieurs à 100 millions de dollars) sont davantage concentrés sur les régions côtières.

Cette carte démontre ainsi que l'approche adoptée demeure relativement cohérente et se rapproche étroitement de ce qui a pu être observé dans la réalité. De plus, il est important de noter que, dans le cadre de cette étude, seuls les coûts supérieurs à 1 million de dollars sont pris en compte. Cette sélection repose sur une considération assurantielle bien fondée.

En effet, il est raisonnable de supposer que les assureurs sont généralement prêts à supporter les coûts des dommages matériels jusqu'à hauteur d'un million de dollars. Au-delà de ce seuil, un contrat de réassurance entre en jeu pour transférer le risque vers d'autres parties. Cette démarche de réassurance est courante dans l'industrie de l'assurance et permet de gérer les pertes financières potentiellement élevées en cas de catastrophes naturelles.

Ainsi, l'objectif de cette étude est de se pencher spécifiquement sur la modélisation de la répartition des coûts au-delà de ce seuil maximal pris en charge par les assureurs, ce qui revêt une grande importance dans le secteur assurantiel. En comprenant comment les coûts sont distribués au-delà de cette limite, les compagnies d'assurance peuvent mieux évaluer leurs risques et prendre des décisions éclairées en matière de réassurance, contribuant ainsi à une gestion plus efficace des pertes liées aux ouragans.

4.1.1.3 Ajout de variables à priori pour chaque coût par comté

Une fois que les coûts par comté causés par un ouragan historique entre 1982 et 2022 sont obtenus, la base de données peut être complétée avec plusieurs variables explicatives. Ces variables explicatives concernent des caractéristiques spécifiques à l'ouragan et au comté par lequel il est passé. Le tableau 4.1 présente la liste de la grande majorité des variables les plus pertinentes ajoutées à la base de données. Les variables spécifiques à un ouragan proviennent de la base de données IBTrACS (NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2023b), tandis que celles concernant le comté sont extraites de l'étude réalisée par la FEMA sur la construction de l'indice de risque national lié au risque ouragan (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023c).

Lors d'une première analyse de cette base de données, il devient évident qu'un grand nombre de variables sont fortement corrélées entre elles, étant donné qu'elles décrivent des aspects similaires. Par exemple, plusieurs variables décrivent la population, la valeur des bâtiments, la valeur agricole, et la superficie. En outre, certaines variables sont intrinsèquement liées par leur définition. Par exemple, les variables décrivant les taux de pertes historiques, les pertes estimées, et les composantes de l'indice de risque national sont reliées par des formules mathématiques. Pour mieux comprendre ces corrélations, un corrélogramme basé sur les coefficients de Pearson, en utilisant les données de cette base est présenté sur la figure 4.6. Un corrélogramme est une représentation graphique qui met en évidence les relations linéaires entre les variables. Il peut aider à identifier les variables fortement corrélées, ce qui est important dans le cadre de l'analyse statistique et de la modélisation.

Le corrélogramme des coefficients de Pearson pour cette base de données est représenté sur la figure 4.6.

L'analyse du corrélogramme a mis en évidence les corrélations mentionnées entre différentes variables explicatives. Afin d'éviter la redondance d'informations, il est choisi de retirer les variables qui présentent un coefficient de corrélation supérieur à 0.8 et qui décrivent essentiellement le même aspect. Cela concerne notamment les variables liées à la population, à la valeur des bâtiments, à la valeur agricole et à la superficie.

Il est également important de noter que la variable cible, à savoir le coût total causé par un ouragan dans un comté donné, est calculée en fonction de la perte économique moyenne annuelle due au risque ouragan dans ce comté. Par conséquent, cette variable ne peut pas être incluse en tant que variable explicative dans la base de données, car elle est directement liée à la variable cible. De plus, les variables fortement corrélées à cette variable cible doivent également être exclues, car elles présenteraient un risque de corrélation implicite avec la variable cible, ce qui n'est pas souhaitable dans le cadre de la modélisation. Il est à noter que les variables caractérisant l'intensité de l'ouragan sont

Variable	Description
NAME	Nom de l'ouragan
P_{min}	Pression minimale lors du passage de l'ouragan proche ou au sein du comté
P_{moy}	Pression moyenne lors du passage de l'ouragan proche ou au sein du comté
V_{max}	Vitesse maximale des vents lors du passage de l'ouragan proche ou au sein du comté
V_{moy}	Vitesse moyenne des vents lors du passage de l'ouragan proche ou au sein du comté
R_{max}	Rayon de la vitesse maximale des vents lors du passage de l'ouragan proche ou au sein du comté
R_{min}	Rayon de la Vitesse moyenne des vents lors du passage de l'ouragan proche ou au sein du comté
CAT_{max}	Catégorie maximale de l'ouragan selon l'échelle de Saffir-Simpson lors du passage proche ou au sein du comté
STATE	Etat où l'ouragan a causé des dommages
STATE_RISK	Variable catégorielle à 5 modalités (<i>Very Low</i> à <i>Very High</i>) indiquant le niveau de risque de l'Etat face à un ouragan
COUNTY	Nom du comté traversé ou passé proche par l'ouragan
POPULATION	Nombre d'habitants au sein du comté
BUILDVALUE	Valeur des bâtiments au sein du comté (en dollars)
AGRIVALUE	Valeur agricole au sein du comté (en dollars)
AREA	Superficie du comté (en square mile)
HRCN_EVNTS	Nombre d'ouragans historiques ayant causé des dommages au sein du comté de 1851 à 2021
HRCN_AFREQ	Fréquence annuelle du passage d'un ouragan au sein du comté
HRCN_EXPB	Valeur des bâtiments exposée au risque ouragan au sein du comté
HRCN_EXPP	Nombre d'habitants exposée au risque ouragan au sein du comté
HRCN_EXPPE	Valeur monétaire du nombre d'habitants exposée au risque ouragan au sein du comté
HRCN_EXPA	Valeur agricole exposée au risque ouragan au sein du comté (en dollars)
HRCN_EXP_AREA	Superficie du comté exposée au risque ouragan (en square mile)
HRCN_HLRB	Taux de perte historique de valeur des bâtiments due au risque ouragan au sein du comté
HRCN_HLRP	Taux de perte historique du nombre d'habitants due au risque ouragan au sein du comté
HRCN_HLRA	Taux de perte historique de valeur agricole due au risque ouragan au sein du comté
HRCN_EALB	Perte annuelle moyenne estimée de la valeur des bâtiments dû au risque ouragan au sein du comté (en dollars)
HRCN_EALP	Perte annuelle moyenne estimée du nombre d'habitants due au risque ouragan au sein du comté
HRCN_EALPE	Valeur monétaire de la perte annuelle moyenne estimée du nombre d'habitants dû au risque ouragan au sein du comté
HRCN_EALA	Perte annuelle moyenne estimée de la valeur agricole due au risque ouragan au sein du comté
HRCN_RISKV	Composante <i>Value</i> de l'indice de risque national
HRCN_RISKS	Composante <i>Score</i> de l'indice de risque national
HRCN_RISKR	Composante <i>Rating</i> de l'indice de risque national
Cost_billion_county	Variable cible : Coût total des dommages causé par l'ouragan (en milliards de dollars)

TABLE 4.1 : Liste non exhaustive des variables de la base coût

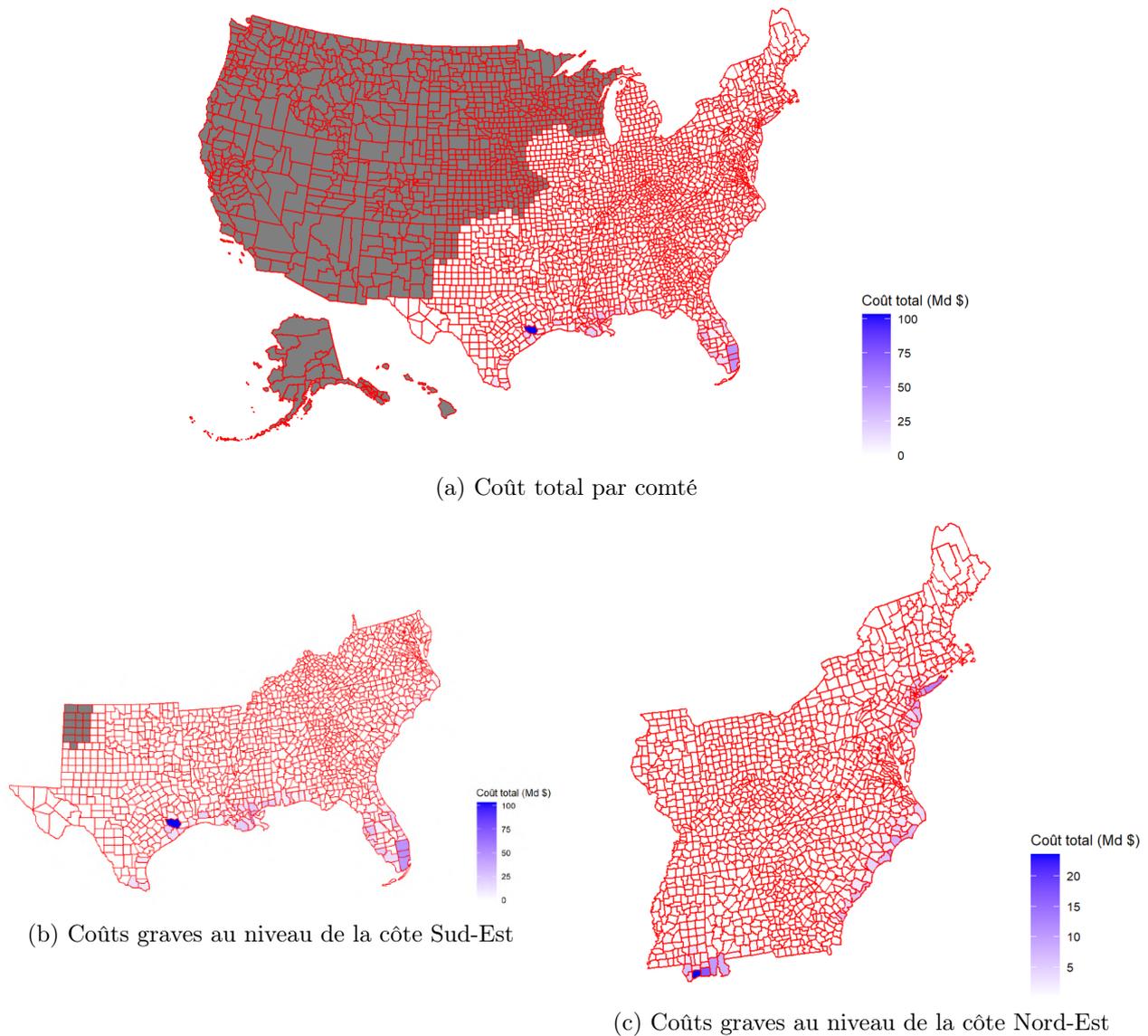


FIGURE 4.5 : Coût total par comté liée au risque ouragan entre 1982 et 2022

fortement corrélées entre elles. Cette corrélation est cohérente avec les observations effectuées dans le chapitre précédent 3 sur la modélisation des caractéristiques de l'intensité des ouragans. Cependant, malgré cette corrélation, ces variables sont maintenues dans la modélisation du coût financier afin de fournir une représentation complète de l'intensité de l'ouragan.

Ces analyses et ajustements ont abouti à une base de données finale comportant 13 variables explicatives, dont 7 sont spécifiques aux caractéristiques de l'ouragan et 6 aux caractéristiques du comté traversé. Cette base de données finale est utilisée par la suite pour la modélisation du coût financier associé aux ouragans. Ces variables sont : P_{min} , P_{moy} , V_{max} , V_{moy} , R_{max} , R_{min} , CAT_{max} , STATE_RISK, HRCN_EXPB, HRCN_EXPP, HRCN_AFREQ, HRCN_EXPA et HRCN_EXP_AREA.

4.1.2 Modélisation de la fonction de perte

L'un des objectifs de l'étude est donc de chercher à modéliser une fonction de perte qui soit à la fois cohérente et rigoureuse. L'objectif ultime est d'obtenir une estimation réaliste des pertes poten-

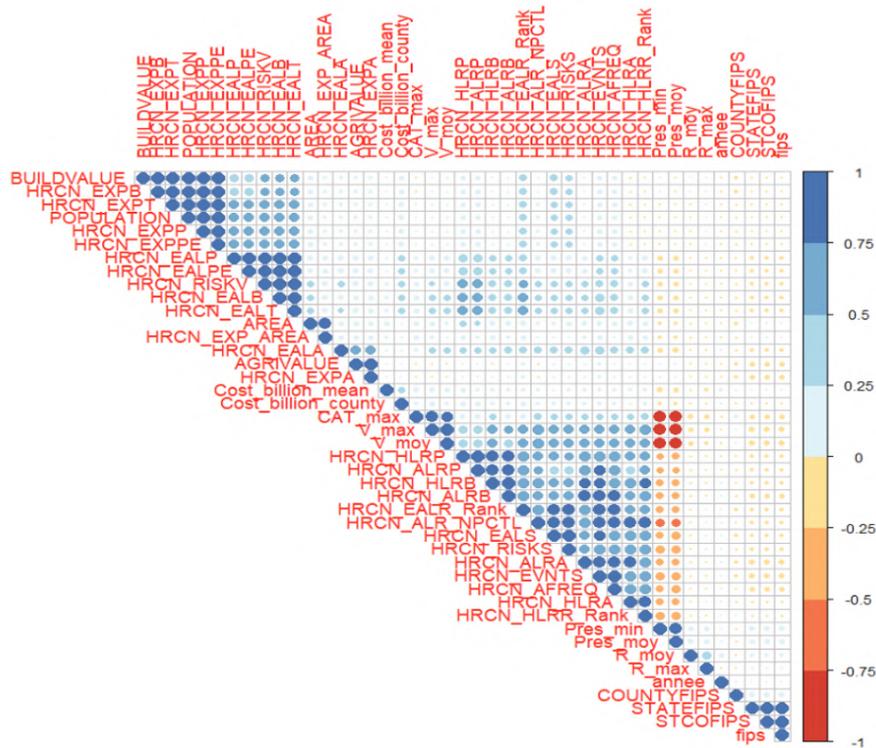


FIGURE 4.6 : Corrélogramme de la base coût

telles au sein d'un comté au passage d'un ouragan. Pour atteindre ce but, une variété d'approches de modélisation est adoptée, notamment l'application de modèles linéaires généralisés, ainsi que des techniques de Machine Learning avancées notamment les algorithmes basés sur les arbres de décision tels que XGBoost et Random Forest.

Avant d'appliquer l'une des méthodes de modélisation, il est essentiel d'acquérir une compréhension précise de la variable cible, à savoir le coût total dans un comté suite au passage d'un ouragan. La figure 4.7 illustre l'évolution du coût total par comté pour chaque entrée de la base de données, accompagnée de son histogramme correspondant. Une échelle logarithmique est utilisée sur l'histogramme afin d'avoir une meilleure visibilité de l'allure de la variable cible. Une observation immédiate est la prédominance de coûts relativement faibles (inférieurs à 0.1 milliard de dollars soit -2.3 à l'échelle logarithmique), tandis que les coûts graves (supérieurs à 0.1 milliard de dollars) sont rares. La distribution de la variable cible présente une queue fine, ce qui suggère une concentration de coûts élevés sur une petite proportion des données.

Pour une modélisation cohérente, il est courant de diviser l'échantillon de la variable cible en deux catégories : les coûts attritionnels et les coûts graves, en fixant un seuil de gravité. Plusieurs méthodes sont disponibles dans la littérature scientifique pour déterminer ce seuil, notamment le Mean Excess Plot, l'estimateur de Hill, une approche basée sur la fréquence, et bien d'autres encore (CHARRAS-GARRIDO et LEZAUD, 2013).

L'approche fondée sur la fréquence implique la sélection du seuil en fonction de la fréquence des événements rares. Plus précisément, il est envisageable de fixer le seuil à un niveau où la distribution des valeurs excédentaires présente une décroissance significative. Simultanément, le Mean Excess Plot, une méthode graphique, est employé pour estimer le seuil en analysant la moyenne des excès au-dessus d'une valeur seuil spécifique. Cette démarche consiste essentiellement à tracer la moyenne des valeurs excédentaires en fonction du seuil, facilitant ainsi l'identification d'une zone où le comportement excédentaire se stabilise.

L'estimateur de Hill, en tant que méthode statistique, se destine à estimer la queue lourde d'une distribution de probabilité. Appliqué au contexte de la modélisation des coûts financiers, cet estimateur peut être exploité pour évaluer le paramètre de forme d'une distribution de valeurs excédentaires, fournissant par la suite une estimation du seuil.

Dans le cadre de cette étude, une approche méthodique basée sur l'examen des quantiles de la distribution de la variable cible ainsi que sur les histogrammes associés est adoptée. Cette méthodologie se caractérise par une évaluation détaillée de la distribution des valeurs observées, en se concentrant particulièrement sur les quantiles et la représentation graphique au moyen d'histogrammes. L'examen des quantiles implique la division de la distribution en segments proportionnels, souvent exprimés en termes de pourcentages, tels que les quartiles (25%, 50%, 75%). En analysant ces segments, il est possible d'obtenir une compréhension approfondie de la répartition des valeurs de la variable cible, mettant en évidence les tendances, la dispersion, et la présence éventuelle de valeurs aberrantes.

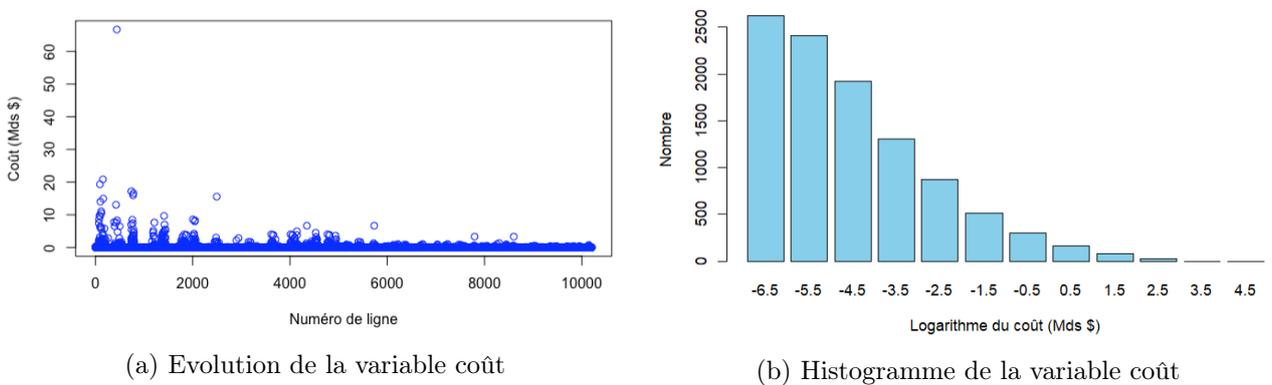


FIGURE 4.7 : Allure de la variable cible

Seuil	25%	50%	75%	90%	95%	99%
	0.0013	0.0024	0.0069	0.0297	0.1474	0.4124
				0.1474	0.4124	2.6929

TABLE 4.2 : Quantiles pour différents seuils de la variable cible (Mds \$)

Les quantiles correspondant à divers seuils ont été consignés dans le tableau 4.2. Il est possible d'observer que la majeure partie des coûts se situe dans la fourchette allant de 1 à 100 millions de dollars, c'est-à-dire jusqu'à légèrement en dessous du quantile de 90 % de la variable cible. Ces coûts seront désignés par la suite comme les « coûts attritionnels ». En revanche, au-delà de cette plage de coûts attritionnels, il est remarquable que les coûts extrêmement élevés se concentrent à partir de 500 millions de dollars, légèrement au-dessus du quantile de 95%. Par conséquent, la décision de fixer deux seuils distincts plutôt qu'un unique seuil est choisie, dans le but d'obtenir une segmentation plus pertinente des coûts graves. Le premier seuil est établi à 100 millions de dollars, tandis que le second est fixé à 500 millions de dollars.

Ce choix est motivé par la nécessité de disposer d'échantillons de taille suffisante pour élaborer des modèles de coût robustes. Bien qu'il aurait été envisageable de définir davantage de seuils, cette approche aurait conduit à subdiviser la base de données en plusieurs tranches, chacune contenant un nombre limité d'observations. Par souci de préservation de la robustesse statistique des analyses, il est jugé préférable de se limiter à trois tranches distinctes. En adoptant cette stratégie, il est possible d'obtenir une base de données consacrée aux coûts attritionnels, comprenant 8 916 observations, une base de données consacrée aux coûts graves, comprenant 849 observations, ainsi qu'une base de données consacrée aux coûts extrêmes, comprenant 449 observations. Cette décision a permis de ga-

rantir une modélisation précise des coûts au-delà du seuil attritionnel, tout en conservant des volumes d'échantillons significatifs pour les analyses subséquentes.

Pour évaluer la cohérence de cette approche, la figure 4.8 présente les histogrammes de la variable cible après le découpage en deux tranches, effectué en utilisant les deux seuils établis.

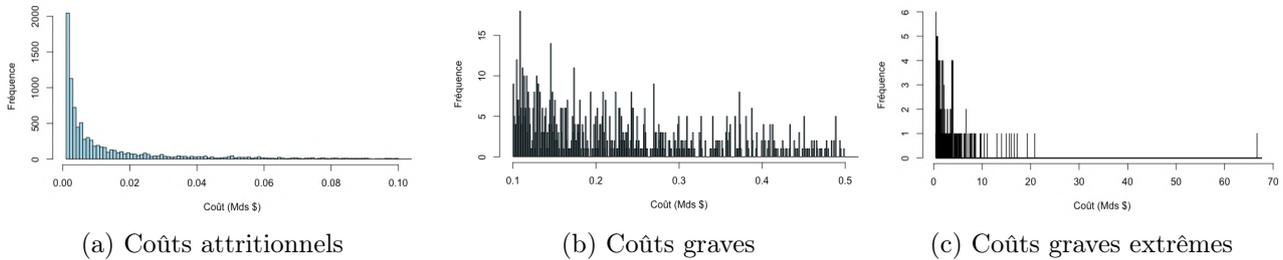


FIGURE 4.8 : Découpage en tranches de la variable coût

Par la suite, les méthodologies pour la modélisation des diverses tranches de coûts sont exposées. Initialement, l'échantillon de coûts est segmenté en deux parties distinctes : un ensemble d'entraînement, utilisé pour former les différents modèles, et un ensemble de test, employé pour la validation de ces modèles. À cette fin, l'échantillon initial est partitionné, attribuant 80% des données à la base d'entraînement et réservant 20% à la base de test.

4.1.2.1 Modèles linéaires généralisés (GLM)

Les modèles linéaires généralisés, *Generalized Linear Model*, (LINDSEY, 2008) sont une classe de modèles statistiques qui permettent de modéliser la relation entre une variable dépendante (dans ce cas, le coût des pertes) et un ensemble de variables explicatives (telles que la force de l'ouragan, la taille du comté, etc...). Les principaux éléments à retenir à propos des GLM sont évoquées dans l'annexe A.7.

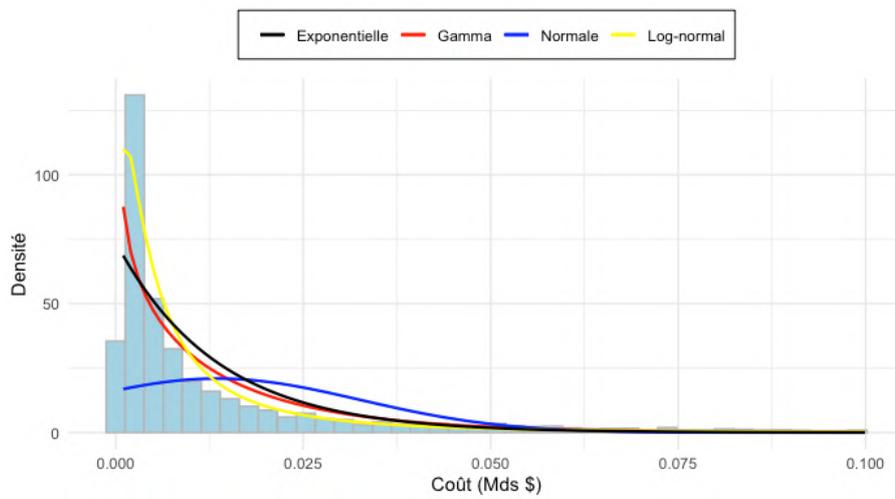
Ainsi pour avoir une vision des différentes distributions possible pour modéliser la variable cible, un ajustement est réalisée pour quatre distributions différentes qui sont la loi exponentielle, la loi gamma, la loi normale et la loi log-normale. Les paramètres des différentes lois sont estimées par la méthode du maximum de vraisemblance et les résultats des fonctions de densités ajustées sont présentés sur la figure 4.9.

Afin d'explorer les différentes distributions potentielles pour modéliser la variable cible, un ajustement est effectuée avec quatre distributions différentes : la loi exponentielle, la loi gamma, la loi normale et la loi log-normale. Les paramètres de ces différentes lois ont été estimés en utilisant la méthode du maximum de vraisemblance. Les résultats des fonctions de densités ajustées sont présentés dans la figure 4.9 et les graphiques quantile-quantile associées sont présentées sur la figure A.10 de l'annexe A.9.

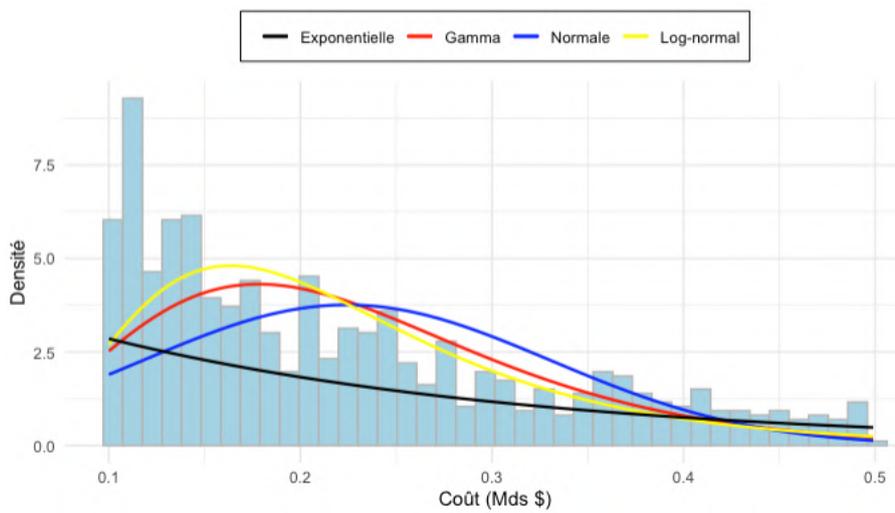
Les résultats semblent indiquer que deux distributions pourraient être appropriées pour modéliser la variable cible : la loi gamma et la loi log-normale. Ces deux lois seront donc soumises à des tests à travers le modèle linéaire généralisé (GLM). Bien que la figure semble suggérer que la loi gamma est plus adaptée en raison de sa décroissance moins rapide en queue de distribution par rapport à la distribution log-normale, ce choix sera confirmé par une analyse plus approfondie à l'aide du GLM.

Les méthodes stepwise à l'aide du package *MASS* (VENABLES et RIPLEY, 2002) disponible sur R sont appliquées en utilisant le critère AIC afin de retenir le modèle le plus pertinent du point de vue des variables explicatives.

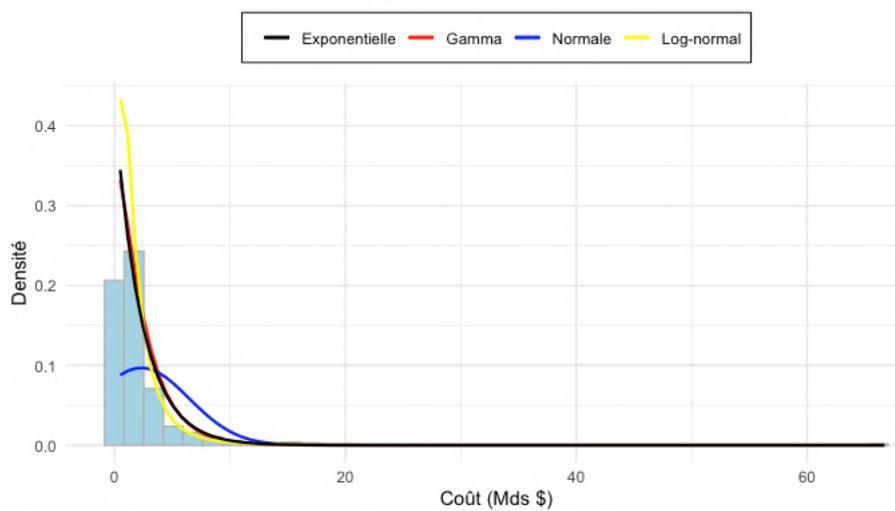
Une fonction de lien logarithmique est sélectionnée pour la loi Gamma, une distribution adaptée à la modélisation de variables continues et positives telles que les coûts observés. Cette approche présente l'avantage supplémentaire de prendre en compte l'hétéroscédasticité, c'est-à-dire la variation de



(a) Coûts attritionnels



(b) Coûts graves



(c) Coûts graves extrêmes

FIGURE 4.9 : Ajustement de diverses distributions à la variable coût

la dispersion en fonction de la valeur attendue. En ce qui concerne la distribution log-normale, la transformation logarithmique de la variable cible a été employée pour l'adapter à une échelle logarithmique. La famille exponentielle choisie dans ce cas est la loi normale, avec une fonction de lien identité. Les résultats des méthodes stepwise appliquées à ces deux distributions sont synthétisés dans le tableau 4.3.

Distribution	Coût	Forward _{AIC}	Backward _{AIC}	Both _{AIC}
Gamma	Attritionnel	-49483.22	-49483.56	-49483.22
Log-Normale	Attritionnel	21879.93	21879.93	21879.93
Gamma	Grave	-1273.239	-1273.812	-1273.239
Log-Normale	Grave	837.6688	838.3317	837.6688
Gamma	Grave Extrême	1136.647	1134.442	1134.758
Log-Normale	Grave Extrême	21879.93	21879.93	21879.93

TABLE 4.3 : Résultats des méthodes stepwise pour les lois gamma et log-normale pour chaque tranche de coût

Les résultats présentent les scores AIC associés aux différentes approches stepwise pour chaque tranche de coût. Dans les deux cas, les trois méthodes stepwise (forward, backward et les deux combinées) ont généré des scores AIC similaires, ce qui suggère que les variables explicatives incluses dans le modèle sont pertinentes pour la modélisation des coûts.

Il convient de noter que la méthode backward présente le score AIC le plus bas parmi les trois, bien que la différence reste minime. En conséquence, les modèles retenus pour la loi gamma et la loi log-normale reposent sur la méthode backward pour chaque tranche de coût.

Par la suite, afin de déterminer quelle loi est la plus appropriée pour modéliser le coût, le modèle entraîné sur la base test, constituée de 20% de l'échantillon initial, est évalué. Pour choisir entre ces deux modèles, des métriques mesurant la qualité de prédiction, telles que le *Root Mean Square Error (RMSE)* et le *Mean Absolute Error (MAE)*, sont utilisées. Les résultats sont exposés dans le tableau 4.4, présentant le RMSE et le MAE de chaque modèle GLM appliqué à chaque tranche de coût, tant sur les bases d'entraînement que de test.

GLM	Coût	RMSE _{train}	RMSE _{test}	MAE _{train}	MAE _{test}
Gamma	Attritionnel	1.120	1.128	0.922	0.932
Log-Normale	Attritionnel	1.297	1.303	1.055	1.061
Gamma	Grave	0.445	0.423	0.379	0.360
Log-Normale	Grave	0.455	0.433	0.388	0.371
Gamma	Grave Extrême	0.730	0.733	0.565	0.604
Log-Normale	Grave Extrême	0.818	0.780	0.642	0.646

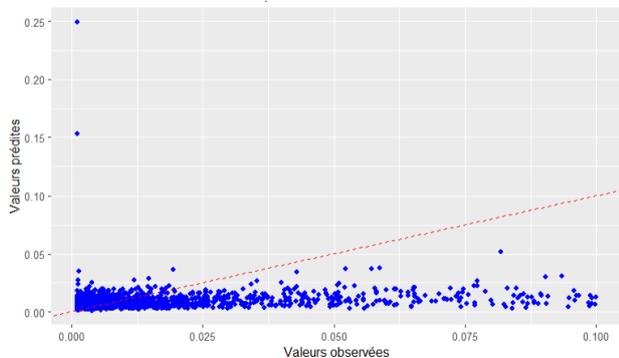
TABLE 4.4 : RMSE et MAE des modèles GLM Gamma et Log-Normale sur les bases d'entraînement et de test

La confrontation des deux modèles GLM élaborés pour chaque tranche de coût, à l'aide des indicateurs de mesure de la qualité de prédiction, révèle de manière générale une performance supérieure des modèles GLM utilisant une loi Gamma. Bien que les différences soient relativement minimes, elles justifient néanmoins le choix d'adopter le modèle GLM Gamma pour la présentation ultérieure des résultats de prédiction, en raison de sa meilleure qualité de prédiction. Les détails complets des modèles GLM Gamma pour chaque tranche de coût sont fournis en annexe A.9.

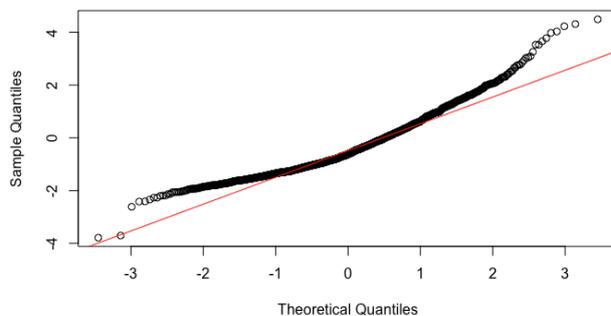
Un examen général des modèles révèle que pratiquement toutes les variables incluses sont statistiquement significatives au seuil de 5%, ce qui renforce leur crédibilité. Pour le modèle de coût grave,

4.1. CONSTRUCTION D'UNE FONCTION DE PERTE AU RISQUE OURAGAN AUX ETATS-UNIS141

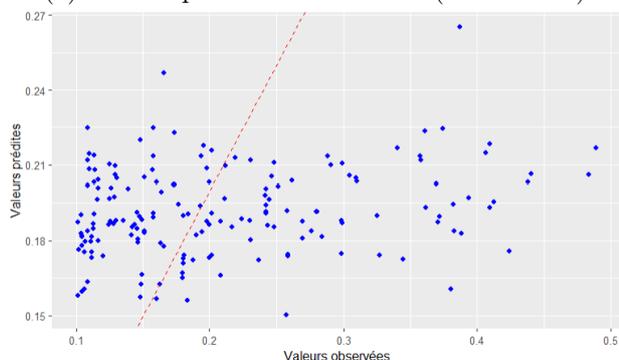
trois variables ne sont pas significatives au seuil de 5% : V_{moy} , AGRIVALUE et HRCN_EXPA. Dans le modèle de coût grave extrême, une variable, en l'occurrence R_{max} , est principalement non significative. Cette absence de signification peut être principalement attribuée aux relations de colinéarité présentes entre les variables, qui ont été délibérément conservées dans le modèle pour garantir une explication maximale de la variable cible. Par la suite ces variables sont donc maintenues dans les modèles.



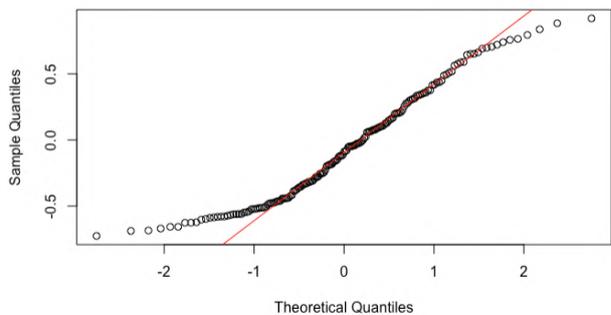
(a) Valeurs prédites vs observées (attritionnel)



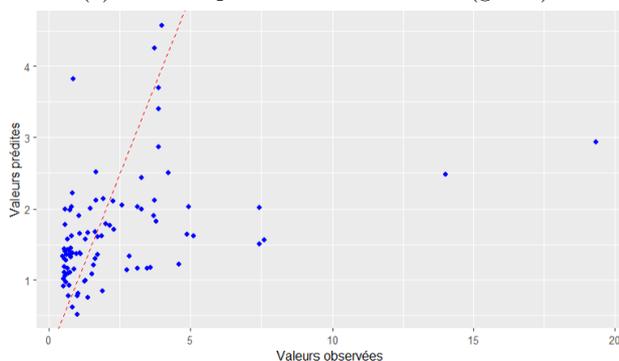
(b) QQ-Plot des résidus



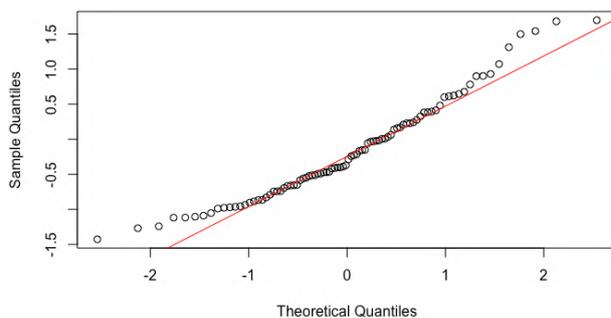
(c) Valeurs prédites vs observées (grave)



(d) QQ-Plot des résidus



(e) Valeurs prédites vs observées (grave extrême)



(f) QQ-Plot des résidus

FIGURE 4.10 : Résultats des modèles GLM sur la base test

Les résultats obtenus sur la base test des modèles GLM Gamma sont présentés sur la figure 4.10. Pour chaque tranche de coût, cette figure comporte deux graphiques : le premier compare les valeurs prédites avec les valeurs observées, tandis que le second représente les quantiles empiriques des résidus par rapport aux quantiles théoriques d'une loi normale. L'analyse de ces graphiques soulève plusieurs remarques. Les graphiques de comparaison entre les valeurs prédites et observées révèlent une qualité de prédiction globalement mauvaise des modèles. Ces derniers éprouvent considérablement de difficultés à saisir les particularités liées à ces coûts, ce qui se traduit par une prédiction peu précise et peu fidèle

aux observations réelles.

Il est particulièrement notable que les modèles retenus peinent à maintenir une cohérence prédictive entre les valeurs prédites et observées, signalant ainsi une incapacité significative à capturer les nuances inhérentes à chaque tranche de coût. L'examen des graphiques des résidus pour chaque tranche de coût révèle que les quantiles extrêmes ne sont que partiellement capturés par le modèle. Ce manquement à appréhender les valeurs extrêmes souligne les limitations substantielles des modèles GLM Gamma dans la prédiction précise des coûts liés aux dommages causés par un ouragan.

Cet échec peut être attribué à plusieurs raisons. Tout d'abord, il est important de noter que les GLM se concentrent principalement sur la modélisation de relations linéaires entre les variables, ce qui peut être insuffisant pour capturer des relations non linéaires et complexes présentes dans ce type de modélisation complexe des coûts financiers liés au risque ouragan. Les GLM ont des limitations inhérentes en termes de capacité à modéliser des données présentant des relations non linéaires, ce qui peut expliquer en partie les résultats insatisfaisants dans la prédiction des coûts graves. Par conséquent, l'utilisation de modèles plus avancés telles que les modèles GAM, *Generalized Additive Models* (ROSS, 2019), ou de Machine Learning plus avancés serait plus adaptée pour réussir à saisir ces relations complexes, en particulier pour les coûts graves. En effet, contrairement aux GLM, les GAM offrent une plus grande flexibilité en permettant la modélisation de relations non linéaires entre les variables explicatives et la variable cible. Dans le cadre de cette étude, les modèles GAM pourraient apporter plusieurs contributions significatives. Tout d'abord, ils sont capables de capturer des relations complexes et non linéaires, ce qui est crucial pour modéliser de manière plus précise les coûts associés aux dommages causés par un ouragan. De plus, les GAM permettent l'inclusion de termes lisses, ce qui offre la possibilité de modéliser des tendances subtiles et des variations locales dans les données. Cela pourrait être particulièrement utile pour saisir des comportements spécifiques à certaines tranches de coût.

Concernant les modèles de Machine Learning, ils sont connus pour leur capacité à gérer des données complexes et des relations non linéaires. Par conséquent, leur exploration pour améliorer la qualité prédictive dans ce contexte spécifique peut être très utile. Par la suite, deux modèles de Machine Learning classiques, le XGBoost et le Random Forest sont testées afin de pouvoir comparer leur qualité prédictive obtenue avec celle obtenue à travers le GLM.

4.1.2.2 XGBoost et Random Forest

Les modèles de Machine Learning tels que le XGBoost ou le Random Forest se distinguent par leur capacité à traiter des données complexes et à capturer des relations non linéaires au sein des ensembles de données, contrairement aux modèles linéaires traditionnels. Une synthèse des principes fondamentaux de ces modèles est présentée dans l'annexe A.7.

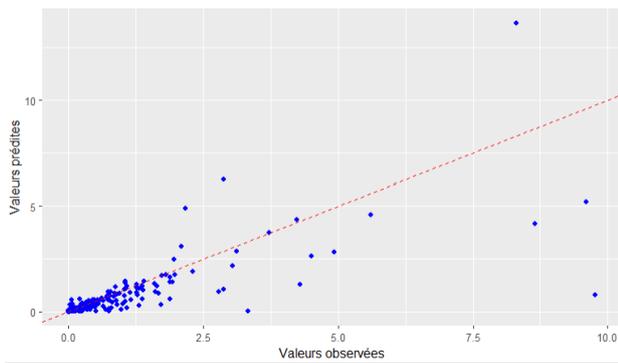
Contrairement à l'approche des GLM, l'utilisation des modèles de Machine Learning dans cette étude ne requiert pas le découpage en trois parties de la distribution des coûts (attritionnels, graves et graves extrêmes), comme présenté dans la section 4.1.2.1. Cette décision découle de la capacité inhérente des modèles de Machine Learning à gérer des données complexes et à identifier automatiquement des relations, linéaires ou non, ainsi que des comportements complexes au sein des distributions de coûts. Ces modèles peuvent donc apprendre directement à partir des données sans nécessiter une présegmentation de la distribution des coûts, simplifiant ainsi le processus de modélisation.

Le tableau 4.5 présente les configurations optimales obtenues pour ces deux modèles, résultant d'une validation croisée à 10 blocs (*10-fold Cross-Validation*). Pour le modèle XGBoost, la métrique d'évaluation retenue est la racine de l'erreur quadratique moyenne (RMSE). Pour le modèle Random Forest, la détermination du meilleur paramètre *mtry* est réalisée en recherchant le minimum de l'erreur *Out Of Bag* (OOB) en fonction de *mtry*. Ces configurations optimales sont obtenues après un processus méticuleux d'exploration des hyperparamètres, garantissant ainsi des performances optimales pour chacun de ces modèles.

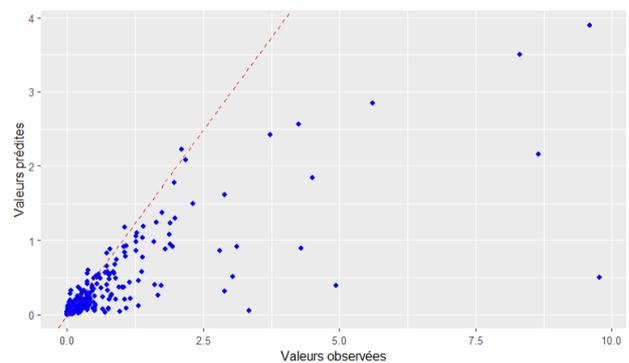
Modèle	Paramétrage optimal
XGBoost	eta : 0.1 max_depth : 6 min_child_weight : 1 gamma : 0 colsample_bytree : 0.9 subsample : 0.9 nrounds : 500
Random Forest	ntree : 500
Forest	mtry : 8

TABLE 4.5 : Hyperparamétrisation des modèles XGBoost et Random Forest

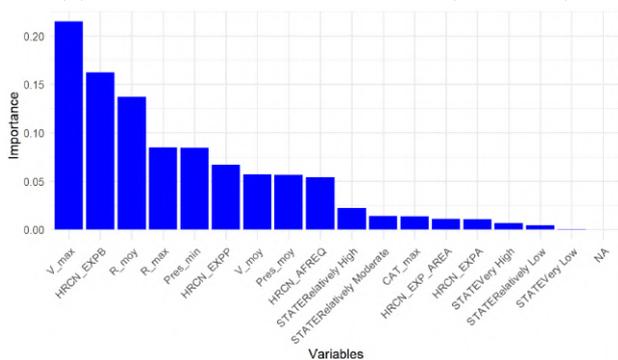
Une fois les paramètres optimaux obtenus, une validation des modèles est effectuée en utilisant une base de test, permettant ainsi la comparaison des deux modèles développés. Les résultats de ces tests sont présentés de manière détaillée dans la figure 4.11, avec un graphique représentant les valeurs prédites en fonction des valeurs observées sont présentées pour chacun des deux modèles ainsi que le graphe d'importance des variables associé.



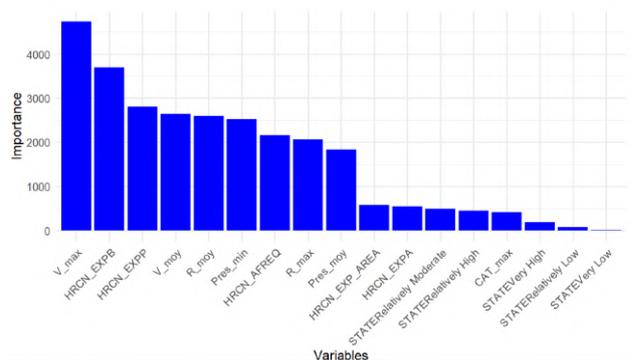
(a) Valeurs prédites vs observées (XGBoost)



(b) Valeurs prédites vs observées (Random Forest)



(c) Importance des variables (XGBoost)



(d) Importance des variables (Random Forest)

FIGURE 4.11 : Résultats des modèles XGBoost et Random Forest sur la base test

En général, il est possible de constater que le modèle XGBoost affiche une meilleure qualité de prédiction par rapport au modèle Random Forest. Cette observation est étayée par plusieurs éléments. Tout d'abord, en examinant les graphiques qui comparent les valeurs prédites aux valeurs observées, il est clair que le modèle XGBoost présente un meilleur ajustement par rapport au Random Forest. Ce constat est renforcé par l'analyse des coefficients de détermination présentés dans le tableau 4.6.

Modèle	R^2_{train}	R^2_{test}	$\text{RMSE}_{\text{train}}$	$\text{RMSE}_{\text{test}}$
XGBoost	0.78	0.69	0.3	0.1
Random Forest	0.72	0.16	0.37	1.54

TABLE 4.6 : R^2 et RMSE des modèles XGBoost et Random Forest sur la base d'entraînement et de test

Pour le Random Forest, ces coefficients sont relativement élevés pour les données d'entraînement, mais ils chutent significativement pour la base de test, ce qui suggère un phénomène de surajustement (overfitting). Cette suradaptation découle de la complexité intrinsèque des données, que le Random Forest tente de modéliser de manière excessive. En ce qui concerne les valeurs du RMSE, *Root Mean Square Error*, elles demeurent relativement faibles, mais cette observation mérite un éclaircissement. Les faibles valeurs de RMSE sont en grande partie attribuables aux coûts extrêmes, qui sont largement mal prédits par le modèle. Ces coûts extrêmes sont des événements rares et exceptionnels qui peuvent être difficiles à anticiper. Cependant, malgré l'amélioration apportée par le modèle XGBoost, il est essentiel de noter que ses performances ne peuvent pas être qualifiées de convaincantes ou de robustes.

En effet, il est important de noter que les scores R^2 de 0.78 sur la base d'entraînement et de 0.69 sur la base de test, bien qu'impressionnants, doivent être interprétés avec prudence. De tels scores élevés sur la base d'entraînement indiquent que le modèle XGBoost semble très bien s'adapter aux données avec lesquelles il a été formé. Cependant, le score plus bas sur la base de test suggère que le modèle pourrait ne pas généraliser aussi bien lorsqu'il est confronté à de nouvelles données, ce qui est une préoccupation majeure. Ce phénomène de surajustement se produit lorsque le modèle apprend à s'adapter de manière excessive aux données d'entraînement spécifiques, y compris le bruit présent dans ces données, au lieu de capturer les tendances générales et les relations significatives. En conséquence, le modèle peut devenir trop complexe et ne pas être en mesure de généraliser efficacement pour faire des prédictions sur de nouvelles données. Ainsi ces valeurs de R^2 sont à considérer avec prudence et simplement comme des indicateurs initiaux de la performance du modèle.

De plus, une observation importante est que le modèle XGBoost a tendance à sous-estimer les valeurs prédites à mesure que les valeurs observées augmentent. Cette tendance à la sous-estimation des coûts, en particulier pour les événements à coûts élevés, soulève des préoccupations majeures dans le contexte de la gestion des risques en assurance. Plusieurs facteurs peuvent contribuer à cette lacune dans les prédictions lorsque les coûts augmentent. L'un de ces facteurs pourrait être le nombre limité de variables incluses dans la base de données d'entraînement. Bien que la base de données sur les coûts comprenne des informations sur les caractéristiques de l'intensité des ouragans et les caractéristiques d'exposition des comtés aux risques d'ouragan, il est possible que davantage de variables significatives soient nécessaires dans ce contexte. Les modèles peuvent ne pas tenir compte de certaines variables cruciales pour expliquer les coûts plus élevés, ce qui conduit à des sous-estimations importantes.

Les graphiques représentant l'importance des variables mettent en évidence les variables liées à l'intensité de l'ouragan, telles que la vitesse maximale des vents, le rayon et la pression minimale. Ces variables ont le plus grand impact sur la variable cible, ce qui est cohérent avec les attentes, car l'intensité d'un ouragan est directement liée aux dommages qu'il peut causer, et donc au coût final.

Cependant, pour améliorer la qualité prédictive et l'interprétation des modèles, il est judicieux d'inclure davantage de données liées à l'exposition des comtés. Cela pourrait comprendre des données météorologiques plus détaillées, des informations sur l'infrastructure locale, la densité de population, les politiques d'urbanisation, et d'autres facteurs pertinents. Ces données complémentaires pourraient aider les modèles à mieux saisir les nuances et les subtilités des coûts liés aux ouragans, en particulier pour les scénarios à coûts élevés. Une telle amélioration pourrait conduire à des modèles plus fiables et précis pour l'évaluation des risques liés aux ouragans et des coûts associés, ce qui est d'une importance capitale pour l'industrie de l'assurance et la gestion des catastrophes. Enfin, bien que le modèle XG-

Boost présente des améliorations par rapport au Random Forest, des défis subsistent pour obtenir des prédictions plus précises et fiables dans le domaine complexe de l'assurance liée aux ouragans. Dans le cadre de cette étude, le modèle retenu par la suite pour la modélisation du coût financier causé par un ouragan au sein d'un comté est le modèle XGBoost.

4.2 Application aux saisons cycloniques simulées

La modélisation de la fonction de perte d'un ouragan étant réalisée il est possible de simuler les pertes annuelles engendrées par les ouragans au sein des comtés des États-Unis pour chaque saison cyclonique simulée. Une fois ces pertes annuelles calculées, il est alors envisageable d'obtenir une distribution des pertes annuelles attribuées à chaque ouragan pour chaque comté aux États-Unis et de faire une analyse assurantielle. Cette démarche permet d'identifier les zones où les coûts sont les plus élevés, et de confronter ces résultats aux observations empiriques précédemment effectuées. L'étude s'achève par l'estimation du coût moyen annuel par foyer pour l'ensemble des régions exposées au risque d'ouragan aux États-Unis.

4.2.1 Analyse assurantielle des pertes simulées

La fonction de perte adoptée dans le cadre de cette étude est dérivée du modèle XGBoost. En utilisant les résultats issus de la modélisation de cette fonction de perte, présentée dans la section 4.1.2.2, ainsi que les données générées lors des simulations de 10 000 saisons cycloniques, décrites en détail dans les chapitres 2 et 3, il est possible d'estimer le coût brut annuel causé par l'ensemble des ouragans de chaque saison cyclonique simulée dans chaque comté des États-Unis. Pour rappel, le nombre de tempêtes simulées est de 157 689, correspondant au nombre total de tempêtes sur les 10 000 saisons cycloniques simulées d'après les résultats observées dans le chapitre 2. Ainsi en notant Y_i le coût brut annuel causé par l'ouragan i avec $i \in \llbracket 1, 157689 \rrbracket$, t avec $t \in \llbracket 0, T \rrbracket$, T correspondant à la durée de vie de la tempête i et f la fonction de perte obtenu via le modèle XGBoost illustré sur la figure 4.11, on a

$$Y_i = f(P_{\min,i}^t, P_{\text{moy},i}^t, V_{\max,i}^t, V_{\text{moy},i}^t, R_{\max,i}^t, R_{\min,i}^t, \text{CAT}_{\max,i}^t, \text{STATE_RISK}_i^t, \text{HRCN_EXPB}_i^t, \text{HRCN_EXPP}_i^t, \text{HRCN_AFREQ}_i^t, \text{HRCN_EXPA}_i^t, \text{HRCN_EXP_AREA}_i^t). \quad (4.2)$$

La perte moyenne annuelle est ensuite calculée en agrégeant les coûts bruts totaux de tous les ouragans pour chaque saison cyclonique, puis en calculant la moyenne de ces sommes sur l'ensemble des saisons cycloniques. Soit $N = (N_1, \dots, N_{10000})$ le processus stochastique représentant le nombre de tempêtes annuelle simulée pour chacun des 10 000 saisons cycloniques via la loi binomiale négative calibrée dans le chapitre 2, $PM_{N_j}^k$ la perte moyenne annuelle du comté k de la saison cyclonique j avec $j \in \llbracket 1, 10000 \rrbracket$, et PM^k la perte moyenne annuelle globale du comté k , on a

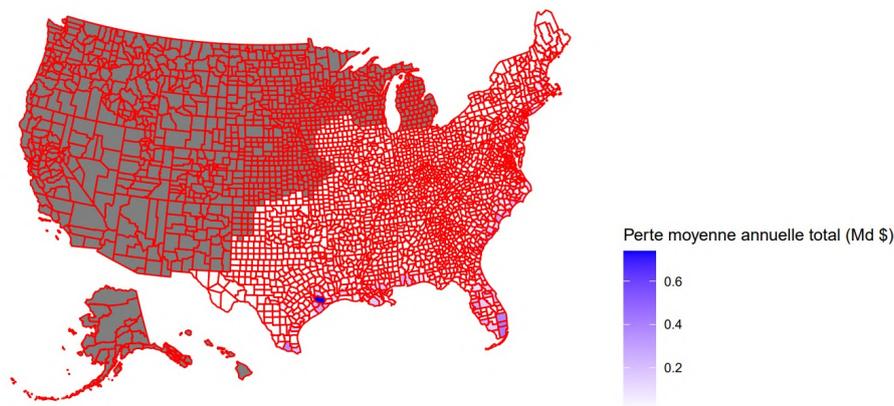
$$PM_{N_j}^k = Y_1^k + Y_2^k + \dots + Y_{N_j}^k, \quad (4.3)$$

avec $Y_1^k, Y_2^k, \dots, Y_{N_j}^k$ les coûts bruts annuels causés au sein du comté k durant la saison cyclonique j . La perte moyenne annuelle globale du comté k s'obtient alors de la manière suivante

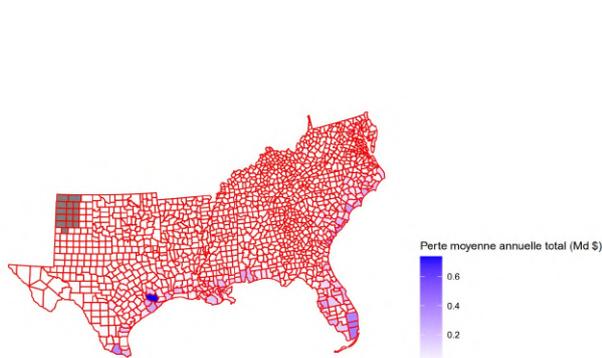
$$PM^k = \frac{PM_{N_1}^k + PM_{N_2}^k + \dots + PM_{N_{10000}}^k}{10000}. \quad (4.4)$$

Ce processus génère la carte illustrée à la Figure 4.12, laquelle présente les pertes moyennes annuelles estimées par comté, soit PM^k , exprimées en milliards de dollars. Cette approche permet d'obtenir une

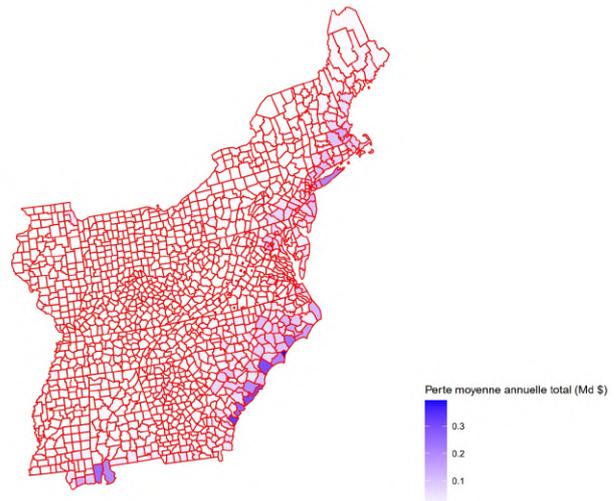
estimation fiable de la perte moyenne annuelle pour chaque comté, contribuant ainsi à une compréhension approfondie des conséquences financières potentielles des ouragans. Les comtés côtiers du golfe du Mexique, situés dans les États du Texas, de la Louisiane et de la Floride, se distinguent clairement avec des pertes moyennes estimées particulièrement élevées, se situant entre 200 et 600 millions de dollars. Cette observation concorde avec les constatations empiriques, notamment celles établies par la FEMA (Federal Emergency Management Agency). L'indice de risque national, *National Risk Index*, lié aux ouragans, présenté sur la figure 4.4, confirme la vulnérabilité élevée de ces régions, corroborant ainsi les résultats. Par ailleurs, l'analyse des coûts totaux par comté sur la période de 1982 à 2022, calculée selon la formule (4.1), et dont la somme totale est présentée à la Figure 4.5, confirme l'ampleur considérable de ces coûts dans ces régions. Ainsi, la fonction de perte élaborée s'avère cohérente avec les observations empiriques et parvient à reproduire efficacement ces phénomènes, même lors de nouvelles simulations basées sur des scénarios aléatoires de phénomènes cycloniques.



(a) Perte moyenne annuelle estimée par comté



(b) Perte moyenne annuelle estimée au niveau de la côte Sud-Est



(c) Perte moyenne annuelle estimée au niveau de la côte Nord-Est

FIGURE 4.12 : Perte moyenne total estimée aux États-Unis

Pour approfondir l'analyse des pertes annuelles simulées, une comparaison des distributions des pertes moyennes annuelles entre les différents États est réalisée, comme illustré dans la figure 4.13. Cette

représentation met en évidence de manière significative les États les plus exposés au risque d'ouragan, notamment la Floride et la Louisiane, qui se distinguent par des pertes moyennes particulièrement élevées. Cette observation s'aligne de manière cohérente avec les constatations actuelles sur l'impact cyclonique aux États-Unis.

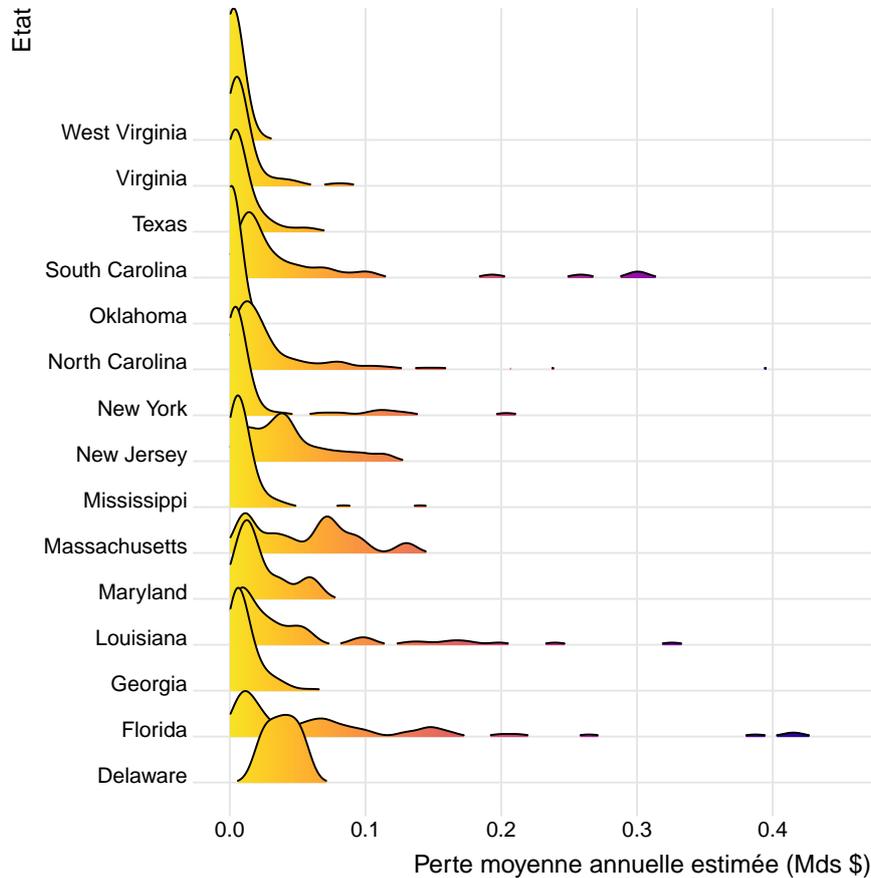


FIGURE 4.13 : Distribution des pertes moyennes annuelles estimées par Etat

Après avoir évalué les pertes moyennes annuelles dues aux ouragans à l'échelle des comtés, il est essentiel de considérer les risques extrêmes qui peuvent fortement impacter les assureurs, en particulier dans le cadre des exigences réglementaires telles que Solvabilité 2. Le *Solvency Capital Requirement* (SCR) est un pilier central du cadre réglementaire Solvabilité 2. Il représente le montant de capital que les compagnies d'assurance et de réassurance doivent détenir pour garantir leur solvabilité face à un ensemble diversifié de risques, y compris les risques liés aux catastrophes naturelles comme les ouragans. L'évaluation du SCR prend en compte divers scénarios, dont le quantile de niveau 99.5

Le quantile de niveau 99.5% représente la perte maximale anticipée avec une probabilité de seulement 0.5% sur une période donnée. Du point de vue assurantiel, cette mesure revêt une importance cruciale, car elle permet aux assureurs de quantifier les scénarios les plus extrêmes auxquels ils pourraient être confrontés, contribuant ainsi au calcul du SCR. En d'autres termes, le SCR est étroitement lié au quantile de niveau 99.5%, car il s'agit d'une évaluation des pertes potentielles rares mais extrêmement graves, essentielle pour déterminer le capital réglementaire nécessaire à la résilience de l'entreprise.

La représentation graphique de la distribution des pertes illustré sur la figure 4.14, mettant en évidence le quantile de niveau 99.5%, offre une perspective claire sur les événements exceptionnels et les risques les plus élevés. Cette approche permet aux assureurs de mieux appréhender les pires

scénarios possibles, une donnée cruciale dans la définition de leurs stratégies de couverture et dans la gestion des exigences de Solvabilité 2.

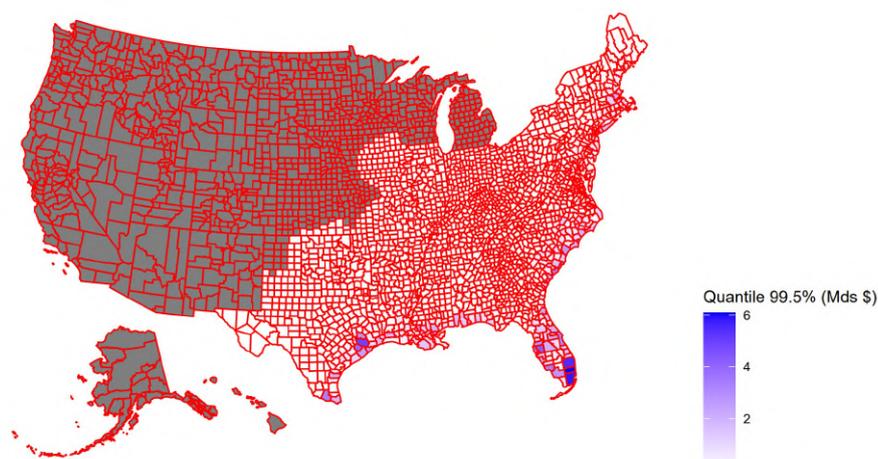


FIGURE 4.14 : Quantile de niveau 99.5% de la perte moyenne total estimée aux États-Unis

L'analyse cartographique révèle de manière significative que les comtés affichant les quantiles les plus élevés au niveau de 99.5% sont localisés le long des côtes du golfe du Mexique et du bassin Atlantique Nord, particulièrement au sein des régions de la Floride, de la Louisiane, du Texas, ainsi que des Caroline du Nord et du Sud. Il est intéressant de souligner que les comtés en Floride se distinguent par les quantiles les plus élevés, oscillant entre 2 et 6 milliards de dollars, suivi de près par ceux de la Louisiane et du Texas, où certains comtés côtiers enregistrent des quantiles se rapprochant de 4 milliards de dollars. Cette distribution spatiale des quantiles extrêmes de niveau 99.5% met en lumière les disparités régionales en termes de risques financiers associés aux ouragans. Les comtés côtiers des États mentionnés sont identifiés comme des points sensibles, exposés à des événements climatiques extrêmes ayant un impact significatif sur les pertes financières potentielles. Ces observations ont des implications importantes pour le secteur de l'assurance, soulignant la nécessité pour les assureurs de prendre en compte ces variations régionales dans leur évaluation des risques et dans la détermination du SCR conformément aux exigences de Solvabilité 2.

En analysant ces résultats, il est notable que les régions les plus exposées aux risques cycloniques sont les comtés côtiers des États tels que la Floride, la Louisiane, et le Texas. Cette exposition accrue se reflète particulièrement dans les quantiles de niveau 99.5%, mettant en lumière l'impact significatif du risque cyclonique sur ces comtés. Une méthode complémentaire pour évaluer et comparer ces risques consiste à examiner les courbes AEP (*Annual Exceedance Probability*) générées à partir des simulations. La courbe AEP est un outil essentiel dans le cadre de la réglementation européenne Solvabilité 2, étant donné que la 50e pire perte annuelle simulée correspond au capital requis sous Solvabilité 2, lié à la période de retour de 200 ans. La courbe AEP se construit comme suit

- Pour chaque comté, les pertes annuelles simulées des 10 000 saisons cycloniques sont triées par ordre décroissant ;
- La période de retour associée à la i -ème perte est calculée comme étant égale à $\frac{10000}{i}$;
- Soit p_i la i -ième valeur de la suite décroissante des pertes annuelles. La perte annuelle a été supérieure ou égale à p_i i fois en 10 000 saisons cycloniques simulées, indiquant que les pertes annuelles sont supérieures ou égales à p_i une fois toutes les $\frac{10000}{i}$ années en moyenne.

La figure 4.15 représente la courbe AEP du comté pour laquelle la perte moyenne annuelle estimée (voir figure 4.12) est la plus élevée au sein des États de la Floride, la Louisiane, le Texas et New

York. Les comtés sont identifiés par leur code FIPS (*Federal Information Processing Standards*). Le code FIPS est un code numérique unique attribué à chaque comté afin de faciliter l'identification et le suivi. Ce code est généralement composé de deux chiffres représentant l'État suivi de trois chiffres représentant le comté spécifique au sein de cet État : 12 pour la Floride, 22 pour la Louisiane, 36 pour New York et 48 pour le Texas.

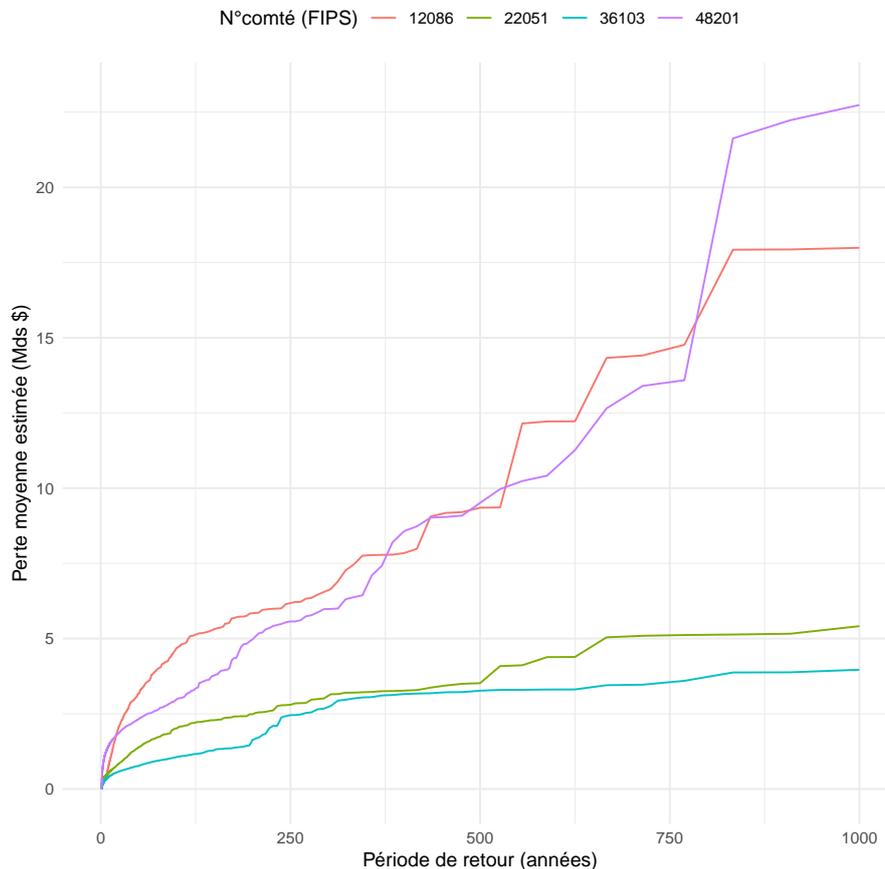


FIGURE 4.15 : Courbes AEP de 4 comtés aux États-Unis

Cette comparaison des courbes AEP met en lumière des disparités significatives entre les comtés sélectionnés, reflétant ainsi des profils de risque cyclonique distincts au sein des États de la Floride, de la Louisiane, du Texas et de New York. L'impact de cette diversité sur le secteur assurantiel est incontestable. Pour un assureur, la localisation géographique de ses assurés devient un élément crucial, étant donné que l'exposition au risque diffère considérablement d'un comté à l'autre. Par exemple, assurer des biens dans le comté de la Floride engendrerait un SCR proche de 6 milliards de dollars, tandis qu'une couverture similaire en Louisiane se traduirait par un SCR d'environ 2.5 milliards de dollars. De plus la perte moyenne estimée présente une variation substantielle entre les comtés d'un même État à mesure que la période de retour augmente. Il est observé que les comtés de la Floride et du Texas affichent des quantiles extrêmes nettement plus élevés que ceux de la Louisiane et de New York. Par exemple, le quantile de niveau 99.9% (correspondant à une période de retour de 1 000 ans) est en moyenne quatre fois plus élevé pour les comtés de la Floride et du Texas. Cette hétérogénéité souligne la nécessité pour les assureurs d'ajuster leurs stratégies en fonction des caractéristiques spécifiques de chaque comté afin de mieux gérer le risque cyclonique.

4.2.2 Estimation du coût moyen annuel par foyer

La perte moyenne annuelle estimée précédemment fournit des indications essentielles sur les comtés les plus exposés aux coûts importants dus aux phénomènes cycloniques. Cependant, ce qui revêt un intérêt particulier est la détermination du tarif que les assureurs devraient appliquer dans les zones à haut risque pour couvrir ces pertes annuelles. Pour ce faire, l'estimation d'un coût moyen annuel par foyer et par comté est envisagée. **Toutefois, il convient de préciser que l'estimation du coût moyen annuel par foyer et par comté, réalisée dans le cadre de cette étude, s'appuie sur une approche spécifique. Il est important de souligner que cette approche ne correspond pas strictement à la réalité de la tarification en assurance MRH (Multirisques Habitation) avec des primes individualisées, notamment en raison de l'hypothèse d'une souscription obligatoire de tous les foyers au sein d'un comté. Ainsi, le coût moyen annuel calculé s'apparente davantage à un coût mutualisé à l'échelle du comté, plutôt qu'à une prime d'assurance au sens usuel.**

L'estimation du coût moyen annuel s'avère essentielle pour plusieurs raisons fondamentales. Premièrement, le coût moyen annuel par foyer permet aux assureurs de calculer les tarifs d'assurance adéquats pour les habitants des régions les plus exposées aux ouragans. Cette approche personnalisée garantit que les résidents paient une prime en fonction du niveau de risque auquel ils sont réellement confrontés. Cela favorise une répartition plus équitable des coûts, évitant ainsi que les résidents de régions à faible risque ne supportent une part disproportionnée des charges. Deuxièmement, en déterminant un coût moyen annuel spécifique à chaque comté, les assureurs peuvent mieux anticiper et gérer les impacts financiers des catastrophes naturelles. Cela permet de constituer des réserves adéquates pour faire face aux demandes d'indemnisation en cas d'ouragan majeur, réduisant ainsi la probabilité de faillite des compagnies d'assurance et assurant la stabilité du marché de l'assurance. Enfin, l'estimation de ce coût offre une vue détaillée des implications financières des phénomènes cycloniques pour les ménages. Elle permet aux résidents de ces régions à risque de prendre des décisions éclairées sur la souscription à une assurance, en évaluant le coût de la protection par rapport aux risques potentiels. De plus, elle incite à la préparation et à l'adoption de mesures d'atténuation des risques, contribuant ainsi à renforcer la résilience des communautés face aux ouragans.

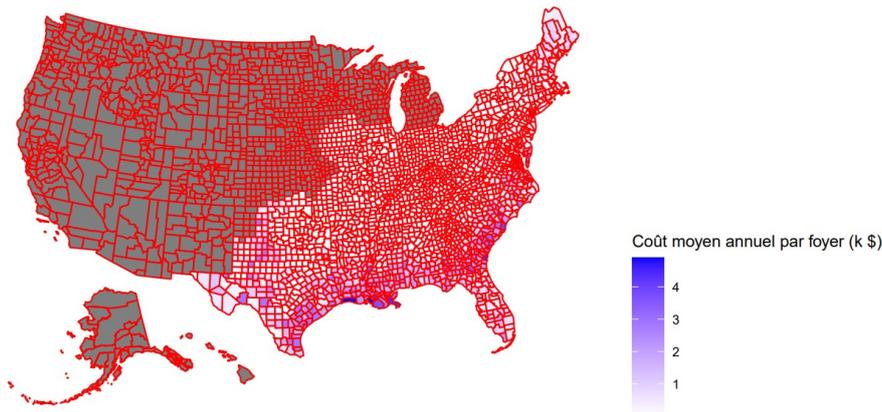
Afin d'estimer d'un coût moyen annuel par foyer, il est nécessaire de disposer de données sur le nombre moyen d'habitants par foyer par comté aux États-Unis. Bien que ces données soient disponibles en libre accès sur le site officiel du Bureau du recensement des États-Unis (UNITED STATES CENSUS BUREAU, 2022), leur extraction individuelle pour chaque comté s'avère complexe. En conséquence, la décision d'utiliser la moyenne nationale du nombre d'habitants par foyer, qui s'élève à 2,60 (UNITED STATES CENSUS BUREAU, 2022) est choisie. De plus, il convient de noter que cette valeur ne varie que peu entre les États considérés dans cette étude, à savoir les États côtiers du Golfe du Mexique et du bassin Atlantique. Par conséquent, l'utilisation de la moyenne nationale demeure une approximation cohérente.

Pour obtenir une estimation du nombre de foyers par comté, le nombre d'habitants exposés au risque d'ouragan au sein d'un comté, représenté par la variable HRCN_EXPP de la base de données des coûts (voir Tableau 4.1), est divisé par le nombre moyen d'habitants par foyer aux États-Unis.

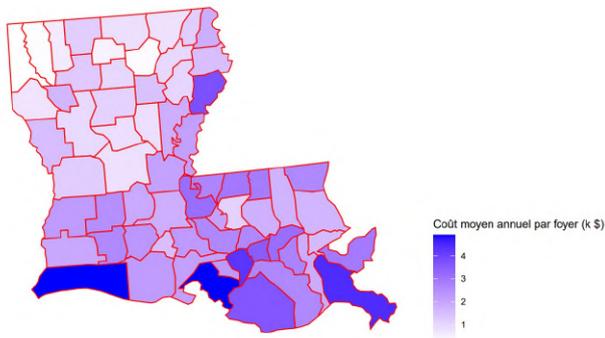
Une fois que le nombre de foyers est ainsi obtenu, le coût moyen annuel par foyer est estimée selon la formule suivante

$$\text{Coût moyen annuel par foyer} = \frac{\text{Perte moyenne annuelle au sein du comté}}{\text{Nombre de foyer au sein du comté}}. \quad (4.5)$$

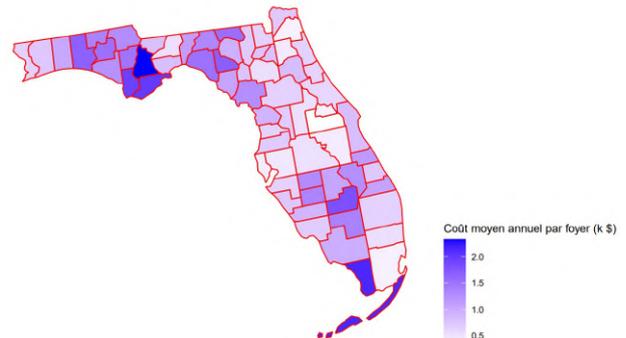
Cette démarche permet de déterminer un coût moyen annuel adaptée à chaque comté, jouant un rôle essentiel dans l'évaluation des coûts potentiels liés aux ouragans pour les ménages et les assureurs opérant dans ces régions à risque. La Figure 4.16 présente de manière visuelle la répartition de ces coûts estimées à travers le pays.



(a) Coût moyen annuel estimé par comté



(b) Coût moyen annuel estimé en Louisiane



(c) Coût moyen annuel estimé en Floride

FIGURE 4.16 : Coût moyen annuel par foyer estimé aux Etats-Unis

Cette représentation graphique offre une vue d'ensemble de la variation des coûts moyens annuels en fonction de la localisation géographique. Elle permet de mettre en évidence les disparités significatives entre les comtés, reflétant ainsi la diversité des niveaux de risque encourus par les foyers américains en raison de la menace ouragan. Les comtés côtiers du golfe du Mexique, notamment ceux situés dans les États du Texas, de la Louisiane et de la Floride et ceux de la côte du bassin Atlantique, se distinguent par des coûts moyens annuels particulièrement élevés. Ces régions sont historiquement exposées à un risque accru d'ouragans, ce qui se traduit par des coûts d'assurance plus importants pour les résidents.

Selon UHLIG (2023), le coût moyen d'une couverture d'assurance complète contre les ouragans aux États-Unis s'élève à environ 2 555 dollars par an. En examinant la carte des coûts moyens annuels estimés, présentée sur la figure 4.16, il est observable que l'approche parvient à reproduire cette moyenne observée, avec des coûts variant jusqu'à environ 5 000 dollars.

Cependant, il est crucial de souligner les limites inhérentes à cette estimation. Tout d'abord, l'estimation du coût moyen annuel repose sur l'estimation préalable de la perte moyenne annuelle au sein de chaque comté, comme l'indique la formule (4.5). Cette estimation de la perte moyenne annuelle découle de la modélisation de la fonction de perte réalisée à l'aide du modèle XGBoost, comme exposé dans la section 4.1.2.2. Néanmoins, comme l'illustre le graphique d'importance finale des variables du modèle, présenté à la Figure 4.11, la variable représentant la population exposée au risque d'ouragan dans chaque comté (HRCN_EXPP) ne figure pas parmi les variables les plus explicatives du modèle. Par conséquent, cette variable a un impact moins prononcé sur la prédiction du coût dans chaque comté. Les prédictions de coût se révèlent ainsi moins précises pour les comtés dont la population est

moins élevée par rapport à d'autres comtés. Cette limitation se manifeste notamment dans l'État du Texas, où plusieurs comtés situés assez loin des côtes du golfe du Mexique affichent des coûts moyens annuels relativement élevés, atteignant près de 3 000 dollars. Une situation similaire se présente également dans les États de la côte nord-est du bassin Atlantique, où l'on observe que des comtés de la Caroline du Nord et de la Virginie, caractérisés par une population moins importante, affichent des coûts moyens annuels élevés et peu réalistes.

Cette constatation met en lumière la nécessité de prendre en compte les limites du modèle adopté et d'approfondir l'analyse des facteurs influençant les coûts moyens annuels. Il est essentiel de reconnaître que la modélisation des coûts d'assurance ouragan adoptée repose sur des hypothèses simplificatrices, et que d'autres variables non incluses dans le modèle peuvent également influencer ces coûts. Une approche plus détaillée et contextuelle pourrait ainsi améliorer la précision de ces estimations.

En revanche, un constat frappant concerne l'État de la Louisiane, où les coûts moyens annuels estimés se révèlent particulièrement élevés, atteignant près de 4 000 à 5 000 dollars pour les comtés situés le long de la côte du golfe du Mexique. Cette observation semble cohérente avec la situation actuelle du secteur de l'assurance lié aux risques cycloniques aux États-Unis, comme évoqué dans la section 1.5. En effet, la Louisiane fait partie des États les plus exposés aux ouragans, et récemment, en 2022, certains assureurs ont été contraints d'augmenter leurs primes pour atteindre une moyenne de 4 700 dollars par an (CHRISTOPHER et al., 2023). Ainsi, la modélisation du coût moyen annuel par foyer réalisée ici parvient à refléter cette réalité spécifique de l'État de la Louisiane.

Enfin, il est remarquable que les coûts estimés pour les comtés de la Floride ne semblent pas significativement élevés par rapport à ceux des États voisins, malgré des pertes moyennes annuelles estimées plus importantes, comme observé à la Figure 4.12. Ces coûts atteignent jusqu'à 3 000 dollars pour certains comtés, ce qui peut sembler surprenant étant donné le risque ouragan élevé dans cet État. Cette situation s'explique notamment par le fait que la Floride est le troisième État le plus peuplé des États-Unis, ce qui se traduit par un nombre élevé de foyers. Par conséquent, le coût moyen annuel par foyer s'avère plus abordable dans ces comtés par rapport aux comtés voisins des autres États. Cependant, il est essentiel de noter que, comme évoqué dans la section 1.5, plus de 15 compagnies d'assurance ont fait faillite et ont quitté la Floride suite au passage de l'ouragan Laura en 2020 (SACKS, 2023). Une explication possible à cette observation, étayée par cette étude, est que les assureurs opérant dans ces zones à risques peuvent devenir insolvables et quitter ces régions, non pas en raison de la perte moyenne engendrée par ces phénomènes dévastateurs, comme illustré à la Figure 4.12, mais plutôt en raison de la variance considérable de ces pertes moyennes.

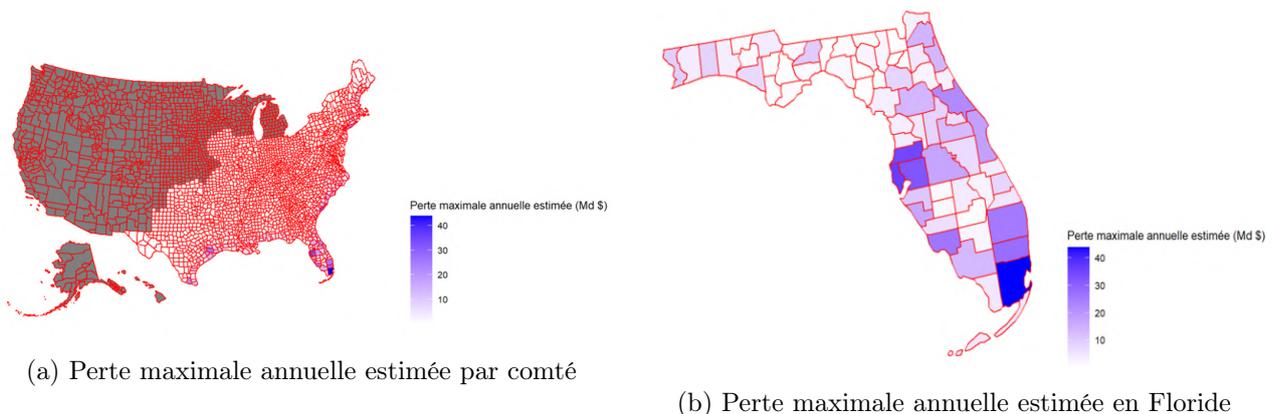


FIGURE 4.17 : Perte maximale annuelle estimée aux États-Unis

La figure 4.17 présente la perte maximale annuelle estimée par comté aux États-Unis, ainsi que spécifiquement en Floride. Il est notable que les comtés de la région du Golfe du Mexique se distinguent

par les pertes maximales estimées les plus élevées, notamment en Floride, où certains scénarios montrent des pertes atteignant jusqu'à 40 milliards de dollars. Cette observation confirme la vulnérabilité de ces régions aux phénomènes cycloniques dévastateurs. Cette forte variabilité des pertes maximales, illustrée notamment dans la Figure 4.17, constitue l'une des principales raisons du retrait des assureurs des zones à risque. Comme mentionné dans la section 1.5 et comme le révèle le Tableau 1.3, neuf des ouragans les plus coûteux de l'histoire se sont produits au cours des 15 dernières années. Cette série d'événements a considérablement mis à l'épreuve le secteur de l'assurance opérant dans ces régions, entraînant progressivement une augmentation des primes pures au fil des années.

Cette tendance à la hausse des primes pures est donc le reflet de la prise en compte de la variabilité et de la gravité des pertes maximales, plutôt que de la seule perte moyenne annuelle. Elle illustre également l'effort constant des assureurs pour équilibrer la couverture et le risque, afin de maintenir leur solvabilité tout en continuant à fournir une protection financière aux assurés. Ainsi, l'approche de modélisation des coûts moyens annuels par foyer réalisée ici permet de mieux comprendre ces dynamiques complexes et d'anticiper les variations des primes dans des régions à risque élevé.

Pour conclure cette étude, il est essentiel de souligner que la saison cyclonique actuelle pose des défis significatifs au secteur de l'assurance, qui est en pleine mutation. Actuellement, l'*Insurance Information Institute* estime qu'une prime d'assurance habitation moyenne en Floride s'élève à environ 6 000 dollars par an, alors que la moyenne nationale aux États-Unis est d'environ 1 700 dollars (GABE et al., 2023). Cette disparité reflète la vulnérabilité de la Floride aux ouragans et aux catastrophes naturelles, ce qui a incité de nombreux assureurs à réévaluer leur présence dans l'État. L'impact de cette situation se fait ressentir avec le départ de plusieurs assureurs au cours de l'année précédente, dont Farmers Insurance, qui a annoncé son intention de quitter la Floride, affectant ainsi environ 100 000 assurés, et de ne pas souscrire de nouvelles polices (WEISE, 2023). De plus, l'ouragan Idalia, survenu en septembre 2023, a encore fragilisé le secteur assurantiel de la Floride, de la Géorgie et des Carolines du Nord et du Sud, avec un coût total estimé entre 12 et 20 milliards de dollars. Face à cette montée des coûts et à la menace persistante des phénomènes cycloniques, des changements significatifs sont actuellement en discussion et pourraient être mis en place dans les années à venir au sein du secteur assurantiel (WEISE, 2023). Parmi ces changements, plusieurs mesures potentielles sont envisagées.

Augmentation de la franchise : Il est probable que les propriétaires individuels devront assumer une plus grande part du risque initial, ce qui se traduirait par une augmentation des franchises ;

Réduction de la concentration du risque : Les compagnies d'assurance pourraient décider de réduire la concentration des risques dans une zone donnée. Par exemple, elles pourraient plafonner le nombre d'habitations qu'elles assurent dans un code postal spécifique, afin de réduire leur exposition au risque et de diversifier leur clientèle ;

Encourager les mesures de protection : Pour réduire les primes d'assurance, les propriétaires pourraient être encouragés à installer des mesures de protection contre les inondations relativement peu coûteuses. Ces mesures contribueraient à réduire les risques de dommages catastrophiques en cas de phénomènes climatiques extrêmes.

Ces mesures visent à équilibrer les coûts croissants liés aux catastrophes naturelles avec la nécessité de maintenir une assurance abordable pour les résidents des régions à risque. Face à l'évolution continue du climat et à l'augmentation des phénomènes climatiques extrêmes, le secteur de l'assurance doit s'adapter pour faire face à ces nouveaux défis.

4.3 Limites de l'étude et perspectives d'amélioration

En conclusion, il est important de prendre en compte les nombreuses limites présentes à travers la modélisation adoptée en raison des choix et des hypothèses considérées, tout en examinant les perspectives d'améliorations potentielles.

En ce qui concerne la modélisation statistique des ouragans, développée dans les chapitres 2 et 3, il est crucial de reconnaître la complexité inhérente aux phénomènes météorologiques tels que les ouragans. Ces événements sont soumis à une variabilité naturelle considérable, ce qui implique que même les modèles statistiques les plus avancés ne peuvent pas garantir une prédiction précise de chaque événement, étant donné leur nature imprévisible. Dans la modélisation des trajectoires d'ouragans, le modèle utilisé ne tient pas compte d'une éventuelle dépendance entre les variations longitudinales et latitudinales. De plus, les trajectoires peuvent être influencées par une multitude de facteurs, tels que l'intensité. Par exemple, des approches plus récentes, telles que celle proposée par BOSE et al. (2023), suggèrent l'utilisation de modèles de Machine Learning et de Deep Learning pour prédire simultanément la trajectoire et l'intensité, en couplant deux modèles afin de prendre en compte les interactions et dépendances potentielles entre ces variables.

Un autre aspect crucial de la modélisation adoptée dans cette étude concerne l'évolution climatique. Les modèles utilisés ont tous été entraînés sur des données climatiques couvrant la période de 1982 à 2022, en supposant que les caractéristiques restent stables au fil du temps. Cependant, les scénarios futurs émis par le GIEC (IPCC, 2022) pour les décennies à venir, notamment en ce qui concerne les températures mondiales, constituent des facteurs susceptibles de contribuer à une augmentation de la proportion d'ouragans intenses à l'avenir. Le réchauffement des océans, par exemple, entraîne une intensification des ouragans, ainsi que des effets associés tels que les ondes de tempête, une augmentation des précipitations et des risques d'inondations. Une perspective d'amélioration des modèles utilisés dans cette étude consisterait à évaluer les différents scénarios projetés par le GIEC pour les décennies futures, en particulier en ce qui concerne les valeurs des températures dans les modèles de pressions terrestres et océaniques, afin de comprendre l'impact potentiel du réchauffement climatique sur l'intensité des ouragans à venir. Par ailleurs, selon les rapports récents du GIEC (IPCC, 2022), quelle que soit la trajectoire d'émissions de gaz à effet de serre, **le réchauffement mondial de 1,5°C par rapport à l'ère préindustrielle sera atteint dès le début des années 2030**. Les politiques mondiales en place fin 2020 conduisent à un réchauffement mondial médian de 3,2 °C en 2100, tandis que les engagements nationaux annoncés avant octobre 2021 aboutissent à **un réchauffement mondial médian de 2,8 °C en 2100** (METEO-FRANCE, 2021). Bien que les scénarios à très fortes émissions soient devenus moins probables, ils ne peuvent être complètement exclus. La figure 4.18 illustre l'évolution des événements extrêmes selon les projections du GIEC, mettant en évidence une intensification et une fréquence accrue des événements de fortes précipitations. De même, la proportion de cyclones tropicaux intenses et les vitesses maximales des vents des cyclones tropicaux les plus forts augmenteront à l'échelle mondiale en raison du réchauffement climatique.

Concernant la modélisation du coût financier, il convient de souligner que la disponibilité limitée de données sur les coûts des ouragans a conduit à une approche simplificatrice qui ne rend pas pleinement compte de la complexité réelle des coûts. Lors de la mise en place de cette modélisation, plusieurs hypothèses et décisions ont été prises pour rendre le modèle applicable et pertinent. Il est crucial de reconnaître que bien que ces hypothèses aient été nécessaires, elles comportent certaines limitations susceptibles d'influencer la qualité des prédictions.

Tout d'abord, les données de base comprenaient des intervalles de coûts pour chaque ouragan ayant entraîné des dommages excédant un milliard de dollars dans un État américain. Cependant, pour développer un modèle de régression plutôt qu'un modèle de classification, la médiane de ces intervalles a été retenue comme variable cible. Cette simplification pourrait potentiellement négliger des informations importantes contenues dans la distribution complète des coûts, pouvant ainsi impacter

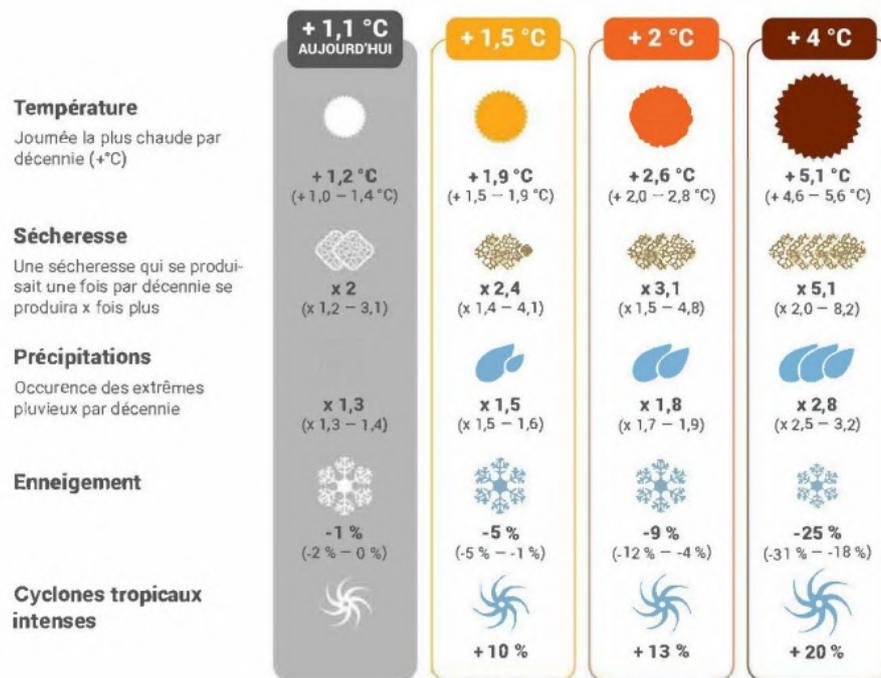


FIGURE 4.18 : Evolution des évènements extrêmes selon les projections du GIEC (METEO-FRANCE, 2021)

la précision des prédictions.

Par ailleurs, la décision de travailler au niveau des comtés plutôt qu'au niveau des États a été motivée par des considérations liées à l'assurance, étant donné que les coûts liés aux ouragans peuvent significativement varier à l'intérieur d'un même État. Pour obtenir un coût par comté, les coûts par État ont été divisés en utilisant un indice basé sur l'estimation du taux de perte annuelle propre à chaque comté en raison du risque ouragan. Néanmoins, il est important de noter que cet indice, bien qu'appuyé par une étude robuste et cohérente (FEDERAL EMERGENCY MANAGEMENT AGENCY, 2023c), comporte également des incertitudes et peut ne pas refléter de manière exhaustive les variations locales des coûts.

De plus, il est à noter que la variable cible choisie dans le cadre de cette étude, bien que pertinente pour le contexte de l'assurance, pourrait ne pas être la plus cohérente à prédire. Une approche plus pertinente pourrait consister à anticiper les dommages directs causés aux bâtiments, aux infrastructures, ou aux biens matériels, sujet largement exploré dans la littérature scientifique. Les modèles de Machine Learning, notamment les réseaux de neurones (KAUR et al., 2021), ont fait l'objet d'études approfondies pour la détection et la prédiction des dommages causés par les catastrophes naturelles, incluant les ouragans. Cependant, une telle prédiction nécessiterait des données plus précises et spécifiques aux pertes matérielles liées à ces phénomènes climatiques, dépassant ainsi le cadre de cette étude.

Par ailleurs, les coûts des ouragans peuvent être influencés par des facteurs externes tels que les politiques gouvernementales, les mesures de prévention des catastrophes, etc., qui peuvent être difficiles à intégrer dans les modèles. Il est impératif de poursuivre les recherches dans ce domaine en explorant la possibilité d'inclure un ensemble plus complet de variables explicatives. Considérer des modèles de Machine Learning encore plus sophistiqués, tels que les réseaux de neurones récurrents (RNN), les réseaux de neurones convolutifs (CNN), ou d'autres modèles d'ensemble comme le AdaBoost, en raison de leur capacité à traiter des données complexes et séquentielles, pourrait contribuer à relever ces défis complexes.

En résumé, ces choix et hypothèses ont été nécessaires pour rendre la modélisation réalisable et pertinente dans le contexte assurantiel. Cependant, il est crucial de garder à l'esprit que ces simplifications et approximations peuvent introduire des biais potentiels dans les prédictions du modèle. Une meilleure compréhension des limites associées à ces choix peut contribuer à une interprétation plus éclairée des résultats et à des améliorations potentielles de la modélisation à l'avenir.

Conclusion

La modélisation du risque cyclonique est un domaine en constante évolution qui revêt une importance cruciale, dépassant le cadre des sciences actuarielles. L'étude des phénomènes cycloniques reste complexe et difficile, malgré les avancées scientifiques considérables. Cette complexité réside dans la multitude de facteurs et de variables qui interviennent dans la formation des ouragans, défiant ainsi une compréhension exhaustive.

Dans ce contexte, cette étude a opté pour une approche statistique, offrant une méthode efficace pour généraliser les tendances observées au cours des dernières décennies aux tempêtes futures. Le travail réalisé s'inscrit dans la lignée des efforts visant à mieux appréhender le risque cyclonique aux États-Unis, en fournissant aux assureurs des outils essentiels pour quantifier et mesurer l'exposition de leur portefeuille au risque cyclonique. Cette compréhension est essentielle pour élaborer des stratégies d'atténuation des risques, des politiques publiques appropriées et des solutions d'assurance adaptées aux réalités changeantes du climat.

La modélisation des saisons cycloniques est une étape cruciale de cette étude, visant à comprendre et à anticiper le comportement des ouragans dans le bassin Atlantique Nord. Cette démarche s'articule en plusieurs phases, chacune reposant sur une méthodologie rigoureuse, élaborée à partir d'une revue exhaustive des modèles statistiques existants dans la littérature spécialisée. La première phase consiste à modéliser la fréquence annuelle des tempêtes. Pour ce faire, une loi binomiale négative, un choix de distribution largement reconnu pour sa capacité à représenter la variabilité temporelle des tempêtes est utilisée. Cette approche permet de générer un échantillon représentatif de la fréquence des tempêtes sur une période donnée. La deuxième étape de la modélisation se concentre sur les points de genèse des tempêtes. Pour ce faire, un lissage spatial est effectué à travers un noyau gaussien, intégrant ainsi les données historiques sur les points de formation des tempêtes. Cette méthode permet la création d'une distribution spatiale des points potentiels de genèse des tempêtes, en prenant en considération les facteurs géographiques et environnementaux favorables à leur apparition. Pour modéliser les trajectoires des tempêtes, des processus auto-régressifs sont employés. Ces processus captent les variations longitudinales et latitudinales historiquement observées, tout en offrant la flexibilité nécessaire pour tenir compte des éventuelles évolutions futures. Enfin, la modélisation de l'intensité des tempêtes s'appuie sur quatre caractéristiques essentielles. En premier lieu, un modèle de régression multiple est utilisé pour modéliser la pression centrale océanique puis un modèle à décroissance exponentielle prend le relais pour la pression centrale terrestre. Ensuite, la vitesse maximale des vents est estimée au moyen d'un modèle de régression spécifique, tout comme son rayon. Enfin, la distribution des vitesses des vents touchant les zones terrestres est modélisée grâce à un modèle de gradient de vitesse. Cette démarche globale permet de cerner de manière précise l'intensité des tempêtes cycloniques.

Pour évaluer les conséquences financières des ouragans, une base de données des coûts a été constituée, répertoriant les ouragans ayant causé plus d'un milliard de dollars de dommages par comté aux États-Unis entre 1982 et 2022. En utilisant des hypothèses simplificatrices et des données en libre accès issues d'une étude menée par la FEMA (*Federal Emergency Management Agency*) et de diverses sources collaboratives contribuant à l'élaboration d'un indice de risque national (*National Risk Index*) par comté aux États-Unis pour chaque risque naturel, une base solide a été mise en place pour estimer

les pertes financières associées aux ouragans. En utilisant cette base de données, l'étape suivante a consisté à élaborer une fonction de perte due aux ouragans en utilisant plusieurs modèles statistiques avancés, notamment le GLM (Generalized Linear Model), XGBoost et Random Forest. Les résultats ont montré que le modèle XGBoost était le plus performant et il a donc été choisi pour simuler les pertes annuelles des saisons cycloniques simulées. Par la suite, une estimation des pertes annuelles moyennes par comté a été réalisée, suivie d'une estimation du coût moyen annuel foyer aux États-Unis en utilisant la perte moyenne annuelle précédemment calculée et le nombre moyen d'habitants par foyer aux États-Unis. Cette étape a permis de quantifier le coût annuel moyen que chaque foyer devrait supporter pour se protéger contre les pertes liées aux ouragans. Les résultats finaux ont montré une cohérence avec les observations actuelles, en particulier pour les comtés côtiers du Golfe du Mexique et du bassin Atlantique, qui se sont retrouvés avec une estimations du coût moyen annuel par foyer plus élevée par rapport à d'autres comtés. Enfin, la mise en contexte du secteur de l'assurance lié aux ouragans aux États-Unis a été réalisée. Les défis actuels auxquels ce secteur est confronté ont été soulignés, notamment l'augmentation des primes d'assurance habitation en Floride ou en Louisiane, le départ de plusieurs assureurs et l'impact financier des ouragans récents. Les conclusions de cette étude peuvent orienter de manière éclairée la politique de souscription des assureurs en identifiant les zones géographiques les plus exposées aux pertes liées aux ouragans. En comprenant les variations de risque entre les comtés, les assureurs peuvent ajuster leurs tarifs, conditions de couverture et capacités d'indemnisation pour optimiser leur gestion des risques.

Cette étude a permis d'obtenir des résultats cohérents tout en mettant en lumière certaines simplifications et limitations inhérentes à la modélisation. Parmi les limitations notables, la non-prise en compte de l'évolution climatique future, l'absence de considération de la dépendance entre les différents modèles de trajectoire et d'intensité d'un ouragan, ainsi que la disponibilité limitée de données sur les coûts des ouragans survenus aux États-Unis ont conduit à une approche simplificatrice qui ne reflète pas entièrement la complexité réelle des coûts associés. Les perspectives d'amélioration de cette modélisation sont multiples. Tout d'abord, il serait opportun d'affiner les modèles en intégrant les projections du GIEC (IPCC, 2022) relatives aux changements climatiques. Cette démarche permettrait d'anticiper de manière plus précise les impacts futurs des ouragans, offrant ainsi aux assureurs la possibilité d'ajuster de manière plus précise leurs politiques et tarifications en réponse à ces changements climatiques prévus. Par ailleurs, des avancées dans les modèles statistiques pour la constitution d'un échantillon de tempêtes ainsi que la construction d'une fonction de perte liée au coût d'un ouragan sont essentielles. Ces améliorations permettraient une représentation plus fidèle des scénarios potentiels, renforçant ainsi la robustesse des résultats obtenus et leur utilité dans un contexte de gestion des risques. En conclusion, cette étude appelle à une réflexion continue sur la manière dont la société peut s'adapter à un environnement en constante évolution, caractérisé par des défis climatiques de plus en plus fréquents et intenses.

Bibliographie

- AMERICAN METEOROLOGICAL SOCIETY (2023). Glossary of Meteorology. URL : <https://glossary.ametsoc.org/wiki/Welcome>.
- ANDERSON, B., SCHUMACHER, A., GUIKEMA, S., QUIRING, S. et FERRERI, J. (2020). stormwindmodel: Model Tropical Cyclone Wind Speeds. R package version 0.1.4. URL : <https://CRAN.R-project.org/package=stormwindmodel>.
- AON (2023). How Climate Science Unites with Technology to Drive a More Resilient World. Article Web. URL : <https://www.aon.com/insights/articles/2023/how-climate-science-unites-with-technology-to-drive-a-more-resilient-world>.
- BATTS, M.E., UNITED STATES. DEPARTMENT OF ENERGY. OFFICE OF CONSERVATION AND SOLAR APPLICATIONS et NATIONAL SCIENCE FOUNDATION (U.S.) (1980). Hurricane Wind Speeds in the United States. NBS building science series. U.S. Department of Commerce, National Bureau of Standards.
- BEUCHER, F. (2010). Manuel de météorologie tropicale : des alizés au cyclone, Tomes 1 et 2. Paris : Météo-France.
- BLOEMENDAAL, N., HAIGH, I., MOEL, H., MUIS, S., HAARSMA, R. et AERTS, J. (2020). Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data* 7, p. 40. DOI : [10.1038/s41597-020-0381-2](https://doi.org/10.1038/s41597-020-0381-2).
- BOSE, R., PINTAR, A. L. et SIMIU, E. (2023). Simulation of Atlantic Hurricane Tracks and Features: A Coupled Machine Learning Approach. *Artificial Intelligence for the Earth Systems* 2.2. DOI : <https://doi.org/10.1175/AIES-D-22-0060.1>.
- CCRIF SPC (2019). 2019 Atlantic Hurricane Season and Hurricane Preparedness Tips and Checklist. Document d'information. URL : <https://www.ccrif.org/2019-atlantic-hurricane-season-and-hurricane-preparedness-tips-and-checklist>.
- CENTRE MÉTÉOROLOGIQUE RÉGIONAL SPÉCIALISÉ CYCLONES DE LA RÉUNION (2023). Les dangers du cyclone. Document d'information. URL : http://www.meteo.fr/temps/domtom/La_Reunion/webcmrs9.0/francais/index.html.
- CHANG, C.-P., LIU, C.-H. et KUO, H.-C. (2003). Typhoon Vamei: An equatorial tropical cyclone formation. *Geophysical Research Letters* 30.3. DOI : [10.1029/2002GL016365](https://doi.org/10.1029/2002GL016365).
- CHARPENTIER, A. (2014). Generating hurricanes with a Markov Spatial Process. Article de blog. URL : <https://freakonometrics.hypotheses.org/17113>.
- CHARRAS-GARRIDO, M. et LEZAUD, P. (2013). Extreme Value Analysis: an Introduction. *Journal de la Société Française de Statistique* 154.2, p. 66-97.
- CHEN, R., ZHANG, W. et WANG, X. (2020). Machine Learning in Tropical Cyclone Forecast Modeling: A Review. *Atmosphere*.
- CHEN, T. et al. (2023). xgboost: Extreme Gradient Boosting. R package version 1.7.5.1. URL : <https://CRAN.R-project.org/package=xgboost>.
- CHO, R. (2022). With Climate Impacts Growing, Insurance Companies Face Big Challenges. State of the Planet. article de presse. URL : <https://news.climate.columbia.edu/2022/11/03/with-climate-impacts-growing-insurance-companies-face-big-challenges/>.

- CHRISTOPHER, F., JILL, C. et IVAN, P. (2023). Climate Shocks Are Making Parts of America Uninsurable. It Just Got Worse. The New York Times. article de presse. URL : <https://www.nytimes.com/2023/05/31/climate/climate-change-insurance-wildfires-california.html>.
- CLIMATE CHANGE INSTITUTE et UNIVERSITY OF MAINE (2023). Monthly Sea Surface Temperature. Graphique. URL : https://climatoreanalyzer.org/clim/sst_monthly/.
- COMMISARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE (2019). Chiffres clés du climat France, Europe et Monde. Rapport. URL : https://www.bioenergie-promotion.fr/wp-content/uploads/2018/12/chiffres-clefs-du-climat_2019_web.pdf.
- CORDER, G. et FOREMAN, D. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley.
- CUI, W. et CARACOGIA, L. (2019). A new stochastic formulation for synthetic hurricane simulation over the North Atlantic Ocean. *Engineering Structures* 199, p. 109597. DOI : [10.1016/j.engstruct.2019.109597](https://doi.org/10.1016/j.engstruct.2019.109597).
- CYCLONE EXTRÊME (2023). La météo cyclonique mondiale. Site Web. URL : <https://www.cyclonextreme.com/cyclonedico.htm>.
- DANIEL, C. et JOHN, K. (2022). A simple model for predicting the hurricane radius of maximum wind from outer size. *Weather and Forecasting*, p. 563 -579. DOI : [10.1175/WAF-D-21-0103.1](https://doi.org/10.1175/WAF-D-21-0103.1).
- DARLING, R. (1991). Estimating probabilities of hurricane wind speeds using a large-scale empirical model. *Journal of Climate* 4.10, p. 1035-1046.
- DELIGNETTE-MULLER, M. L. et DUTANG, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* 64.4, p. 1-34. DOI : [10.18637/jss.v064.i04](https://doi.org/10.18637/jss.v064.i04).
- DEMARIA, M. et KAPLAN, J. (1994). Sea Surface Temperature and the Maximum Intensity of Atlantic Tropical Cyclones. *Journal of Climate* 7.9, p. 1324 -1334. DOI : [https://doi.org/10.1175/1520-0442\(1994\)007<1324:SSTATM>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1324:SSTATM>2.0.CO;2).
- DEMEURE, Y. (2023). El Niño : de nouveaux records de température cette année ? SciencePost. Article de journal. Sciencepost. URL : <https://sciencepost.fr/el-nino-nouveaux-records-temperature-2023/>.
- DENISE-BAILLON, C. (2013). Modélisation statistique du risque de tempêtes tropicales dans le bassin Atlantique Nord. Mémoire Actuariat. ISFA-Université Claude Bernard Lyon 1, p. 123.
- DEPARTMENT OF ATMOSPHERIC SCIENCES (2010). Stages of Development from disturbance to hurricane. Site Web. URL : [http://ww2010.atmos.uiuc.edu/\(Gh\)/guides/mtr/hurr/stages.rxml](http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/hurr/stages.rxml).
- DICTIONNAIRE-ENVIRONNEMENT (2010). houle cyclonique. URL : https://www.dictionnaire-environnement.com/houle_cyclonique_ID5402.html.
- DIRECTION DE LA PRÉVENTION DES POLLUTIONS ET DES RISQUES (2004). Les cyclones. DPPR - SDPRM. PARIS, p. 24.
- DIVISION DE LA LÉGISLATION COMPARÉE DE LA DIRECTION DE L'INITIATIVE PARLEMENTAIRE ET DES DÉLÉGATIONS (2017). Les systèmes d'indemnisation des catastrophes naturelles. Etude de législation comparée. Sénat.
- DORE, K. (2022). As climate change threatens more homes, some properties are getting too costly to insure. CNBC. Article de presse. URL : <https://www.cnbc.com/2022/08/07/climate-change-is-making-some-homes-too-costly-to-insure.html>.
- EARLE, A. (1997). Hazards of the British Virgin Islands. Final report of the hazard and risk assessment project (HRAP). Government of the British Virgin Islands.
- EGUCHI, Y., HATTORI, Y. et NOMURA, M. (2021). Comparative assessment of validity of gradient wind models for a translating tropical cyclone. *SN Applied Sciences* 3. DOI : [10.1007/s42452-021-04406-w](https://doi.org/10.1007/s42452-021-04406-w).
- ELSNER, J. et KARA, A. (1999). Hurricanes of the North Atlantic: Climate and Society. Oxford University Press.
- EMANUEL, K. (2005). Divine Wind: The history and science of hurricanes. New York : Oxford University Press.
- EMANUEL, K. (2018). 100 Years of Progress in Tropical Cyclone Research. *Meteorological Monographs* 59, p. 15.11-15.68.

- ENCYCLOPÉDIE-ENVIRONNEMENT (2022). Mousson. Site Web. URL : <https://www.encyclopedie-environnement.org/air/les-moussons/>.
- EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS (2023). ERA5 hourly data on single levels from 1940 to present. Base de données. URL : <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>.
- FAUCHER, D., RASMUSSEN, P. F. et BOBÉE, B. (2002). Estimation non paramétrique des quantiles de crue par la méthode des noyaux. *Revue des sciences de l'eau / Journal of Water Science* 15.2, 515–541. DOI : <https://doi.org/10.7202/705467ar>.
- FEDERAL EMERGENCY MANAGEMENT AGENCY (2023a). Data Resources. Base de données. URL : <https://hazards.fema.gov/nri/data-resources#spatialTribal>.
- FEDERAL EMERGENCY MANAGEMENT AGENCY (2023b). How a Disaster Gets Declared. Site Web. URL : <https://www.fema.gov/disaster/how-declared>.
- FEDERAL EMERGENCY MANAGEMENT AGENCY (2023c). National Risk Index. Documentation technique. URL : https://www.fema.gov/sites/default/files/documents/fema_national-risk-index_technical-documentation.pdf.
- FEDERAL EMERGENCY MANAGEMENT AGENCY (2023d). National Risk Index Map. Site Web. URL : <https://hazards.fema.gov/nri/map>.
- FRANKLIN, J. (2006). National Hurricane Center - Tropical Cyclone Report - Hurricane Vince. Rapport. National Hurricane Center.
- GABE, G., PHIL, M. et MELISSA, C. (2023). Florida's insurance industry is in flux as Idalia cleanup begins. NBC News. Article de presse. URL : <https://www.nbcnews.com/business/business-news/hurricane-idalia-homeowners-insurance-whats-covered-damage-premiums-rcna102819>.
- GONZALO, P., JEAN-PAUL, P., KURT, G. et JUDITH, M.-R. (2015). State of the Art of Hurricane Vulnerability Estimation Methods: A Review. *Natural Hazards Review* 16.2. DOI : [10.1061/\(ASCE\)NH.1527-6996.0000153](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000153).
- GUINEY, J. et LAWRENCE, M. (1999). Preliminary Report Hurricane Mitch 22 October - 05 November 1998. Preliminary Report. National Hurricane Center.
- HABBEMA, J., HERMANS, J. et BROEK, V. D. (1974). A stepwise discrimination program using density estimation. *COMPSTAT'74 Proceedings in Computational Statistics.*, p. 101-110.
- HALL, T. M. et JEWSON, S. (2007). Statistical modeling of North Atlantic tropical cyclone tracks. *Tellus A* 59A, p. 486-498. DOI : [10.1111/j.1600-0870.2007.00240.x](https://doi.org/10.1111/j.1600-0870.2007.00240.x).
- HAMILTON, J. D. (1994). Time Series Analysis. Princeton University Press.
- HARBOUR, S. (2022). Understanding the Hurricane Insurance Deductible. Investopedia. Article de journal. URL : <https://www.investopedia.com/articles/personal-finance/071114/hurricane-insurance-deductible-fact-sheet.asp>.
- HOLLAND, G. (2008). A Revised Hurricane Pressure Wind Model. *Monthly Weather Review - MONTHLY WEATHER REV* 136. DOI : [10.1175/2008MWR2395.1](https://doi.org/10.1175/2008MWR2395.1).
- HONG, H. P., LI, S. H. et DUAN, Z. D. (2016). Typhoon Wind Hazard Estimation and Mapping for Coastal Region in Mainland China. *Natural Hazards Review* 17.2, p. 04016001. DOI : [10.1061/\(ASCE\)NH.1527-6996.0000210](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000210).
- INSURANCE, S. (2018). How to Choose the Right Home Insurance Deductible. Article Web. URL : <https://www.sahouri.com/blog/how-to-choose-your-home-insurance-deductible>.
- INSURANCE INFORMATION INSTITUTE (2023a). Background on: Reinsurance. Site Web. URL : <https://www.iii.org/article/background-on-reinsurance>.
- INSURANCE INFORMATION INSTITUTE (2023b). Facts + Statistics: Hurricanes. Site Web. URL : <https://www.iii.org/fact-statistic/facts-statistics-hurricanes>.
- INSURANCE INFORMATION INSTITUTE (2023c). Hurricanes: State of the Risk. Document technique.
- INSURANCE INFORMATION INSTITUTE (2023d). Spotlight on: Catastrophes - Insurance issues. Site Web. URL : <https://www.iii.org/article/spotlight-on-catastrophes-insurance-issues>.

- INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2023). Summary for Policymakers. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 3–32. DOI : [10.1017/9781009157896.001](https://doi.org/10.1017/9781009157896.001).
- IPCC (2022). Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Sous la dir. de SHUKLA, P., SKEA, J., SLADE, R., KHOURDAJIE, A. A., van DIEMEN, R., MCCOLLUM, D., PATHAK, M., SOME, S., VYAS, P., FRADERA, R., BELKACEMI, M., HASIJA, A., LISBOA, G., LUZ, S. et MALLEY, J. Cambridge, UK et New York, NY, USA : Cambridge University Press. DOI : [10.1017/9781009157926](https://doi.org/10.1017/9781009157926).
- JAMES, M. et MASON, L. B. (2005). Synthetic Tropical Cyclone Database. *Journal of Waterway, Port, Coastal, and Ocean Engineering* 4, p. 181-192.
- JELESNIANSKI, C. P., CHEN, J. et SHAFFER, W. A. (1992). SLOSH : Sea, Lake, and Overland Surges from Hurricanes. Rapport. United States, National Weather Service.
- JOHN, P. C., ANDREW, S. L. et ROBBIE, B. (2021). National Hurricane Center - Tropical Cyclone Report - Hurricane IRMA. Rapport. National Hurricane Center.
- JONES, O., MAILLARDET, R. et ROBINSON, A. (2014). Introduction to Scientific Programming and Simulation Using R, Second Edition. Chapman & Hall/CRC The R Series. Taylor & Francis.
- KAPLAN, J. et DEMARIA, M. (1995). A Simple Empirical Model for Predicting the Decay of Tropical Cyclone Winds after Landfall. *Journal of Applied Meteorology and Climatology* 34.11, p. 2499-2512. DOI : [https://doi.org/10.1175/1520-0450\(1995\)034<2499:ASEMFP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2).
- KAUR, S., GUPTA, S. et SINGH, S. (2021). Hurricane Damage Detection using Machine Learning and Deep Learning Techniques: A Review. *IOP Conference Series: Materials Science and Engineering* 1022. DOI : [10.1088/1757-899X/1022/1/012035](https://doi.org/10.1088/1757-899X/1022/1/012035).
- KEES, N., JASPER, H., TIM, L., MAARTEN, V. O., SOFIA, C. et ALESSIO, G. (2021). Simulating synthetic tropical cyclone tracks for statistically reliable wind and pressure estimations. *Natural Hazards and Earth System Sciences* 21, 861–878.
- KNABB, R., RHOME, J. et BROWN, D. (2023). National Hurricane Center - Tropical Cyclone Report - Hurricane Katrina. Rapport. National Hurricane Center.
- KOSSIN, J. P., KNAPP, K. R., OLANDER, T. L. et VELDEN, C. S. (2020). Global increase in major tropical cyclone exceedance probability over the past four decades. *Proceedings of the National Academy of Sciences* 117.22, p. 11975-11980. DOI : [10.1073/pnas.1920849117](https://doi.org/10.1073/pnas.1920849117).
- LE MONDE (2019). L'Atlantique nord a connu 35 ouragans de catégorie 5 en un siècle, dont 13 depuis 2000. *Le Monde*.
- LIAW, A. et WIENER, M. (2002). Classification and Regression by randomForest. *R News* 2.3, p. 18-22.
- LINDSEY, J. (2008). Applying Generalized Linear Models. Springer Texts in Statistics. Springer New York.
- LIU, F. et PANG, W. (2011). Development and Calibration of Central Pressure Filling Rate Models for Hurricane Simulation. DOI : [10.1201/b11332-132](https://doi.org/10.1201/b11332-132).
- LUDLUM, D. (1963). Early American Hurricanes, 1492-1870. History of American weather. American Meteorological Society.
- MASTERS, J. (2019). Hurricane Dorian Was Worthy of a Category 6 Rating. *Scientific American*.
- METEO-FRANCE (2003). Echelle de Beaufort. Glossaire. URL : http://files.meteofrance.com/files/glossaire/FR/glossaire/designation/510_curieux_view.html.
- METEO-FRANCE (2020). Tempêtes, cyclones, tornades et orages. Site Web. URL : <http://tempetes.meteo.fr/Tempetes-cyclones-tornades-et-orages.html>.
- METEO-FRANCE (2021). Changement climatique : 10 points clés pour comprendre le 6e rapport du Giec. Site Web. URL : <https://meteofrance.com/actualites-et-dossiers/actualites/changement-climatique-10-points-cles-pour-comprendre-le-6e>.

- METEO-FRANCE (2023). Les cyclones. Site Web. URL : <https://meteofrance.com/comprendre-la-meteo/le-vent/les-cyclones>.
- METEO-TROPICALE (2023). Echelle de Saffir-Simpson. Site Web. URL : <https://www.meteo-tropicale.fr/echelle-saffir-simpson/>.
- MÉTÉO-FRANCE (2003). Dépression. Glossaire. URL : http://files.meteofrance.com/files/glossaire/FR/glossaire/designation/127_curieux_view.html.
- NATIONAL AERONAUTICS AND SPACE ADMINISTRATION - SATELLITE AQUA (2017). Image satellitaire d'Irma. Site Web. URL : https://en.wikipedia.org/wiki/Hurricane_Irma#/media/File:Irma_2017-09-05_1706Z.jpg.
- NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION (2023). Costliest U.S. Tropical Cyclones. Document technique.
- NATIONAL HURRICANE CENTER (2017). Trajectoire de l'ouragan Irma. Site Web. URL : https://en.wikipedia.org/wiki/Hurricane_Irma#/media/File:Irma_2017_track.png.
- NATIONAL HURRICANE CENTER (2019). NHC Track and Intensity Models. Site Web. URL : <https://www.nhc.noaa.gov/modelsummary.shtml>.
- NATIONAL HURRICANE CENTER (2023a). Best Track Data (HURDAT2). Base de données. URL : <https://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2022-050423.txt>.
- NATIONAL HURRICANE CENTER (2023b). The Tropical Cyclone Extended Best Track Dataset. Base de données. URL : https://rammb2.cira.colostate.edu/wp-content/uploads/2020/11/EBTRK_AL_final_1851-2021_new_format_02-Sep-2022-1.txt.
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (2012). Sandy : Miami - South Florida. Site Web. URL : <https://www.weather.gov/mfl/sandy>.
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (2019). International Best Track Archive for Climate Stewardship, IBTrACS. Document Technique. URL : https://www.ncei.noaa.gov/sites/default/files/2021-07/IBTrACS_version4_Technical_Details.pdf.
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (2022). 1980- Cost of billion dollar hurricanes - US - variwide chart - NOAA data. Site Web. URL : https://en.wikipedia.org/wiki/List_of_costliest_Atlantic_hurricanes#/media/File:1980-_Cost_of_billion_dollar_hurricanes_-_US_-_variwide_chart_-_NOAA_data.svg.
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (2023a). Billion-Dollar Weather and Climate Disasters. Base de données. URL : <https://www.ncei.noaa.gov/access/billions/mapping>.
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (2023b). International Best Track Archive for Climate Stewardship, IBTrACS. Base de données. URL : <https://www.ncei.noaa.gov/data/international-best-track-archive-for-climate-stewardship-ibtracs/v04r00/access/csv/>.
- NATURAL EARTH (2023). Coastline. Base de données. URL : <https://www.naturalearthdata.com/downloads/10m-physical-vectors/>.
- NEUMANN, C. J. (1972). An alternate to the HURRAN (Hurricane Analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR ; 62, p. 28.
- PHILBRICK, I. P. et WU, A. (2022). Population Growth Is Making Hurricanes More Expensive. New York Times. Article de journal. URL : <https://web.archive.org/web/20221206130032/https://www.nytimes.com/2022/12/02/briefing/why-hurricanes-cost-more.html>.
- PLUVIOSE, M. (2019). Is it Possible to Weaken a Hurricane? Sketch of a Solution Using the Locally Available Energy. *International Journal of Applied Environmental Sciences* 14.2.
- POPOVIC, C. (2019). Des pays de la Caraïbe augmentent leurs assurances pour la saison cyclonique. Article Web. URL : <https://la1ere.francetvinfo.fr/martinique/pays-caraibe-augmentent-leurs-assurances-saison-cyclonique-723752.html>.
- POTTER, B. (2022). How much do construction costs matter? Some factors that affect the price of housing. Construction Physics. Article de blog. URL : <https://www.construction-physics.com/p/how-much-do-construction-costs-matter>.

- QUILLEN, A. (2022). Construction Costs Hit Highest Spike in 50 Years. NBC 5 Dallas-Fort Worth. Article de presse. URL : <https://www.nbcdfw.com/news/local/construction-costs-hit-highest-spike-in-50-years/2891677/>.
- R CORE TEAM (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL : <https://www.R-project.org/>.
- RAJPOOT, H. C. (2016). Minimum/great-circle distance between any two arbitrary points on a sphere or globe for given latitudes & longitudes using vectors (Derivation of great-circle distance formula using vectors). *University of Technology, Gorakhpur-273010 (UP), India*.
- RENAUD, A., DOS SANTOS, A. et SÉMÉCURBE, F. (2018). Prendre en compte l'hétérogénéité spatiale avec le lissage. Manuel de statistique spatiale. INSEE.
- RISI, C. (2005). Statistical synthesis of tropical cyclone tracks in a risk evaluation perspective. Rapport de stage. Ecole Normale Supérieure.
- ROSS, N. (2019). Generalized Additive Models in R: A Free Interactive Course. Rapp. tech. Site Web. URL : <https://noamross.github.io/gams-in-r-course/>.
- ROUX, F. (2018). Cyclones tropicaux : développement et organisation. Rapp. tech. Encyclopédie de l'Environnement. URL : <http://www.encyclopedie-environnement.org/?p=676>.
- RUCH, J.-J. et CHABANOL, M.-L. (2012). Chaînes de Markov. Préparation à l'Agrégation Bordeaux 1. Université Bordeaux-1.
- RUMPF, J., WEINDL, H., RAUCH, E. et SCHMIDT, V. (2007). Stochastic modelling of tropical cyclone tracks. *Springer-Verlag* 66, :475–490. DOI : [10.1007/s00186-007-0168-7](https://doi.org/10.1007/s00186-007-0168-7).
- SACKS, B. (2023). Climate disasters make it harder to insure your home. Here's what to know. The Washington Post. Article de journal. URL : <https://www.washingtonpost.com/climate-environment/2023/06/07/insurance-coverage-loss-climate-change/>.
- SANKAR, N., KOTAL, S. D. et KUNDU, P. K. K. (2016). Seasonal prediction of tropical cyclone activity over the north Indian Ocean using three artificial neural networks. *Meteorology and Atmospheric Physics* 128(6), 751–762.
- SARACHIK, E. et CANE, M. (2010). The El Niño-Southern Oscillation Phenomenon. Cambridge University Press.
- SCHLICHTER, S. (2023a). Complete Guide to Hurricane Insurance. NerdWallet. Article Web. URL : <https://www.nerdwallet.com/article/insurance/hurricane-insurance#:~:text=Is%20there%20such%20a%20thing,need%20separate%20windstorm%20insurance%2C%20too..>
- SCHLICHTER, S. (2023b). What Is Flood Insurance, and What Does It Cover? NerdWallet. Article Web. URL : https://www.nerdwallet.com/article/insurance/flood-insurance?trk_channel=web&trk_copy=What%20Is%20Flood%20Insurance%2C%20and%20What%20Does%20It%20Cover%3F&trk_element=hyperlink&trk_elementPosition=0&trk_location=PostList&trk_subLocation=image-list.
- SCHNEYER, J. (2023). The Science Behind the Storm: El Niño and the 2023 Hurricane Season. *Core-Logic*.
- SNAIKI, R. et WU, T. (2020). Revisiting hurricane track model for wind risk assessment. *Structural Safety* 87, p. 102003. DOI : [10.1016/j.strusafe.2020.102003](https://doi.org/10.1016/j.strusafe.2020.102003).
- SYLVAIN, R. (2023). Séries chronologiques (avec R). Support de cours M1 IM, 2022-2023. Université Nice Sophia Antipolis.
- THIERRY, N., PAGNEY BÉNITO-ESPINAL, F., LAGAHÉ, et GOBINDDASS, M.-L. (2018). Les catastrophes cycloniques de septembre 2017 dans la Caraïbe insulaire au prisme de la pauvreté et des fragilités sociétales. *International Journal of Applied Environmental Sciences* 48.2. DOI : [10.4000/echogeo.16439](https://doi.org/10.4000/echogeo.16439).
- UHLIG, D. K. (2023). Hurricane Insurance Guide: Coverages, Costs, and Claims (2023). Article Web. URL : <https://insurify.com/homeowners-insurance/hurricane-insurance/#:~:text=affect%20your%20premiums.-,How%20much%20does%20hurricane%20insurance%20cost%3F,different%20factors%20determine%20the%20cost..>

- UNITED STATES BUREAU OF LABOR STATISTICS (2023). Housing Price Inflation 1967-2023. Site web. URL : <https://www.in2013dollars.com/Housing/price-inflation>.
- UNITED STATES CENSUS BUREAU (2021). Centers of Population. Base de données. URL : <https://www.census.gov/geographies/reference-files/time-series/geo/centers-population.html>.
- UNITED STATES CENSUS BUREAU (2022). Persons per household, 2017-2021. Site Web. URL : <https://www.census.gov/quickfacts/fact/dashboard/US/HSD310221#HSD310221>.
- VENABLES, W. N. et RIPLEY, B. D. (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York : Springer.
- VICKERY, P. J. et WADHERA, D. (2008). Statistical Models of Holland Pressure Profile Parameter and Radius to Maximum Winds of Hurricanes from Flight-Level Pressure and H*Wind Data. *Journal of Applied Meteorology and Climatology* 47.10, p. 2497 -2517. DOI : <https://doi.org/10.1175/2008JAMC1837.1>.
- VICKERY, P., SKERLJ, P. et TWISDALE, L. (2000). Simulation of hurricane risk in the US using empirical track model. NOAA Technical Memorandum NWS NHC 38.
- WANG, Y. et ROSOWSKY, D. V. (2012). Joint distribution model for prediction of hurricane wind speed and size. *Structural Safety* 35, p. 40-51. DOI : <https://doi.org/10.1016/j.strusafe.2011.12.001>.
- WEISE, E. (2023). How billion-dollar hurricanes, other disasters are starting to reshape your insurance bill. USA Today. Article de journal. URL : <https://eu.usatoday.com/story/news/nation/2023/09/02/idalia-kicks-off-huge-insurance-costs-in-hurricane-season-2023/70732903007/>.
- WILLOUGHBY, H. E., DARLING, R. W. R. et RAHN, M. E. (2006). Parametric Representation of the Primary Hurricane Vortex. Part II: A New Family of Sectionally Continuous Profiles. *Monthly Weather Review* 134.4, p. 1102 -1120. DOI : <https://doi.org/10.1175/MWR3106.1>.
- YAGUANG, T., ALAN, B., RODRIGO, I. S., KEVIN, B., STEF, S., ROSS, S. P., PATRICK, L., DONGLIANG, P., KEVIN, T. et DUCKHAM, M. (2021). A comparative analysis of trajectory similarity measures. *GIScience & Remote Sensing* 58.5, p. 643-669. DOI : [10.1080/15481603.2021.1908927](https://doi.org/10.1080/15481603.2021.1908927).
- YOMBI, F. O. (2016). Modélisation du risque cyclonique dans les Antilles françaises. Mémoire Actuariat. ENSAE, p. 109.
- ÉLIE, F. (2009). Réchauffement climatique : bases scientifiques pour comprendre le problème.

Annexe A

Annexes

A.1 Echelle de Beaufort

Degré Beaufort	Terme descriptif	Vitesse moyenne en nœuds	Vitesse moyenne en km/h	Observations en mer	Observations sur terre
0	Calme	< 1	< 1	La mer est comme un miroir.	On ne sent pas de vent ; la fumée s'élève verticalement.
1	Très légère brise	de 1 à 3	De 1 à 5	Quelques rides en écaille de poisson, mais sans aucune écume.	On sent très peu le vent ; sa direction est révélée par la fumée qu'il entraîne, mais non par les girouettes.
2	Légère brise	de 4 à 6	de 6 à 11	Vaguelettes courtes aux crêtes d'apparence vitreuse, ne déferlant pas.	Le vent est perçu au visage ; les feuilles frémissent, les girouettes tournent.
3	Petite brise	de 7 à 10	de 12 à 19	Très petites vagues (environ 60 cm de haut) ; les crêtes commencent à déferler, les moutons apparaissent.	Les drapeaux légers se déploient ; les feuilles et les rameaux sont sans cesse agités.
4	Jolie brise	de 11 à 16	de 20 à 28	Petites vagues s'allongeant, moutons nombreux.	Le vent soulève la poussière, les feuilles et les morceaux de papier, il agite les petites branches ; les cheveux sont dérangés, les vêtements claquent.
5	Bonne brise	de 17 à 21	de 29 à 38	Vagues modérées (2 m de haut), nettement allongées ; beaucoup de moutons ; embruns.	Les yeux sont gênés par les matières dans l'air ; les arbustes en feuilles commencent à se balancer ; des vaguelettes se forment sur les plans d'eau.
6	Vent frais	de 22 à 27	de 39 à 49	Des lames se forment, les crêtes d'écume blanche s'étendent ; davantage d'embruns.	Les manches sont gonflées par les côtés, l'utilisation des parapluies devient difficile ; les grandes branches sont agitées, les fils des lignes électriques font entendre un sifflement.
7	Grand frais	de 28 à 33	de 50 à 61	La mer grossit en lames déferlantes ; l'écume commence à être soufflée en traînées dans le lit du vent.	La marche contre le vent devient pénible ; les arbres sont agités en entier.
8	Coup de vent	de 34 à 40	de 62 à 74	Les lames atteignent une hauteur de l'ordre de 5 m ; tourbillons d'écume à la crête de lames, traînées d'écume.	La marche contre le vent est très difficile ; le vent casse des rameaux.
9	Fort coup de vent	de 41 à 47	de 75 à 88	Grosses lames déferlant en rouleaux, tourbillons d'embruns arrachés aux lames, nettes traînées d'écume ; visibilité réduite par les embruns.	Les enfants sont renversés ; le vent arrache les tuyaux de cheminées et endommage les toitures.
10	Tempête	de 48 à 55	de 89 à 102	Très grosses lames déferlantes (9 m de haut) ; écume en larges bancs formant des traînées blanches ; visibilité réduite par les embruns.	(Rarement observé à terre.) Les adultes sont renversés ; les arbres sont déracinés, les habitations subissent d'importants dommages.
11	Violente tempête	de 56 à 63	de 103 à 117	Lames déferlantes d'une hauteur exceptionnelle ; mer couverte d'écume blanche ; visibilité réduite.	(Très rarement observé à terre.) Ravages étendus.
12	Ouagan	64 et plus	118 et plus	Lames déferlantes énormes (les creux atteignent 14 m), mer entièrement blanche ; air plein d'embruns et d'embruns ; visibilité très réduite.	(En principe, degré non utilisé.) Ravages désastreux : violence et destruction.

TABLE A.1 : Echelle de classification de Beaufort (METEO-FRANCE, 2003)

A.2 Rappels théoriques sur les chaînes de Markov

Dans cette annexe, une synthèse des principes fondamentaux des Chaînes de Markov utile pour l'étude est présentée RUCH et CHABANOL (2012).

Définition 1 (Chaînes de Markov) Une suite $(X_n)_{n \geq 0}$ de variables aléatoires à valeurs dans un ensemble au plus dénombrable E est une chaîne de Markov d'espace d'états E si et seulement si pour tout $k \in \mathbb{N}$, pour tout (x_0, \dots, x_{k+1}) dans E tels que $\mathbb{P}(X_k = x_k, \dots, X_0 = x_0) > 0$,

$$\mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k).$$

Cela s'écrit également

$$\mathcal{L}(X_{k+1} \mid X_k = x_k, \dots, X_0 = x_0) = \mathcal{L}(X_{k+1} \mid X_k = x_k)$$

pour tout $k \in \mathbb{N}$.

Définition 2 (Loi initiale) Soit $X = (X_n)_{n \geq 0}$ une chaîne de Markov. On appelle loi initiale de X la loi de X_0 , qui est une mesure définie pour $x \in E$, par

$$\nu_0(x) = \nu_0(\{x\}) = \mathbb{P}(X_0 = x).$$

Remarque 1 En toute rigueur il faudrait faire attention quand les événements par lesquels on conditionne peuvent être de probabilités nulles. De fait dans les problèmes de modélisation, les chaînes de Markov sont données par la loi de X_0 et par toutes les probabilités de transition et les problèmes ne se posent pas.

L'indice n de la suite $(X_n)_{n \geq 0}$ est interprété comme un temps. La variable X_k représente la position spatiale à l'instant k , la tribu $\sigma(X_0, \dots, X_{k-1})$ représente son passé tandis que la tribu $\sigma(X_{k+1}, X_{k+2}, \dots)$ représente son futur. Les chaînes de Markov sont des suites aléatoires sans mémoire, c'est-à-dire que dans l'évolution au cours du temps, l'état du processus à un instant futur ne dépend que de celui à l'instant présent et non de ses états antérieurs. Il est souvent cité que conditionnellement au présent, passé et futur sont indépendants.

Définition 3 (Chaînes de Markov homogène) La chaîne de Markov est dite homogène si on a de plus pour tout $k \in \mathbb{N}$ et tout x et y dans E ,

$$\mathbb{P}(X_{k+1} = y \mid X_k = x) = \mathbb{P}(X_1 = y \mid X_0 = x) = p_{x,y}.$$

$p_{x,y}$ est la probabilité de transition pour aller de l'état x à l'état y .

Dans toute la suite, les chaînes de Markov considérées seront toutes homogènes et à espace d'états au plus dénombrable. De plus, les probabilités de transitions d'un état à un autre dans un espace d'états E au plus dénombrable permettent de définir une matrice de transition caractéristique d'une chaîne de Markov.

Définition 4 On appelle matrice de transition la matrice $\mathcal{P} = (p_{x,y})_{x,y \in E}$:

$$\mathcal{P} = \begin{pmatrix} p_{x_0,x_0} & p_{x_0,x_1} & p_{x_0,x_2} & \cdots \\ p_{x_1,x_0} & p_{x_1,x_1} & p_{x_1,x_2} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}.$$

À partir de ces définitions, il est possible de déduire les propositions suivantes.

Proposition 1 *Toute matrice de transition vérifie les propriétés suivantes :*

1. *pour tout couple (x, y) de E , $0 \leq p_{x,y} \leq 1$;*
2. *pour tout $x \in E$, on a $\sum_{y \in E} p_{x,y} = 1$.*
3. *On note ν_0 la loi de X_0 ($\nu_0(x_0) = \mathbb{P}(X_0 = x_0)$). On a alors pour tous (x_0, \dots, x_n) dans E*

$$\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) = \nu_0(x_0) \prod_{k=0}^{n-1} p_{x_k, x_{k+1}}$$

Ces rappels sur les Chaînes de Markov s'achèvent par une introduction aux méthodes d'inférence utilisées dans ce contexte. L'inférence joue un rôle essentiel pour estimer les propriétés et les comportements des processus stochastiques. Les méthodes d'inférence peuvent être catégorisées en deux grandes approches : paramétrique et non paramétrique. Chacune de ces approches offre des avantages et des inconvénients distincts en fonction des caractéristiques du modèle et des données disponibles.

Inférence Paramétrique : les méthodes d'inférence paramétrique supposent que la structure du modèle de la chaîne de Markov est connue, mais les paramètres du modèle doivent être estimés à partir des données. On suppose une forme spécifique pour les probabilités de transition entre les états, ainsi que pour d'autres caractéristiques du modèle, telles que les probabilités initiales. Les paramètres du modèle sont estimés à partir des données observées. Des techniques courantes pour cette estimation incluent la méthode du maximum de vraisemblance et les méthodes bayésiennes. De plus les modèles paramétriques sont souvent plus simples à mettre en œuvre et à interpréter.

Inférence Non-Paramétrique : les méthodes d'inférence non paramétrique, en revanche, ne supposent pas une forme spécifique pour la distribution des probabilités de transition. Ces méthodes sont plus flexibles et n'imposent pas de contraintes sur la forme de la distribution. Elles sont capables de capturer des structures plus complexes. Plutôt que d'estimer des paramètres spécifiques, les méthodes non paramétriques tentent souvent d'estimer directement les probabilités de transition ou d'autres caractéristiques du modèle. Elles sont particulièrement adaptées lorsque la structure du modèle est mal connue ou lorsque les données ne suivent pas une distribution connue.

A.3 Rappels théoriques sur les séries temporelles

Dans cette annexe, une synthèse des principes fondamentaux des séries temporelles utile pour l'étude est présentée SYLVAIN (2023).

Définition 5 (Bruit blanc) *Un processus de bruit blanc est une suite de variables aléatoires (X_t) , d'espérance et de variance constante (c'est à dire $\mathbb{E}(X_t)$ et $\mathbb{V}(X_t)$ ne dépendent pas de t) et tel que $\text{Cov}(X_t, X_{t+h}) = 0$, si $h \neq 0$. Si l'espérance est nulle, le bruit blanc est dit centré. Si les variables sont gaussiennes, le bruit blanc est dit gaussien.*

Définition 6 (Stationnarité, Auto-Covariance, Auto-Corrélation) *Un processus aléatoire $(X_t)_{t \geq 0}$ est stationnaire s'il est d'espérance constante ($\mathbb{E}(X_t)$ ne dépend pas de t) et si les covariances sont stables par translation dans le temps ($\text{Cov}(X_t, X_{t+h})$ ne dépend pas de t , quel que soit $h \geq 0$).*

Pour un processus $(X_t)_{t \geq 0}$ stationnaire, on définit la fonction d'auto-covariance par

$$h \in \mathbb{Z} \mapsto \sigma(h) = \text{Cov}(X_t, X_{t+h}) \quad (\text{ne dépend pas de } t) \quad (\text{A.1})$$

et la fonction d'auto-corrélation par

$$h \in \mathbb{N}^* \mapsto \rho(h) = \frac{\sigma(h)}{\sigma(0)}. \quad (\text{A.2})$$

On remarque que la fonction σ est paire.

Définition 7 (Auto-Corrélation Partielle) *Pour un processus stationnaire (X_t) , le coefficient de corrélation partielle (on dit aussi corrélation partielle) entre X_n et X_1 est*

$$r_{X_2, \dots, X_n}(X_1, X_n) = \frac{1}{\sigma(0)} \text{Cov}(X_1 - \mathbb{E}(X_1 | X_2, \dots, X_{n-1}), X_n - \mathbb{E}(X_n | X_2, \dots, X_{n-1})),$$

le coefficient d'auto-corrélation partielle en h est défini par

$$\begin{cases} r(h) = r_{X_2, \dots, X_{h-1}}(X_1, X_h) & \text{pour } h \geq 2 \\ r(1) = \rho(1) \\ r(0) = 1, \\ r(-h) = r(h) \end{cases} \quad \text{pour } h \geq 1$$

Pour un processus stationnaire,

$$r(h) = r_{X_{n+1}, \dots, X_{n+h-1}}(X_n, X_{n+h}) \quad \text{pour tout } n.$$

Définition 8 (Processus Auto-régressif) *Un processus (X_t) est dit auto-régressif d'ordre p centré s'il vérifie*

$$X_t = \epsilon_t + \sum_{j=1}^p a_j X_{t-j}, \quad \text{pour tout } t \geq 0 \quad (\text{A.3})$$

($p \in \mathbb{N}^$, $a_p \neq 0$ et $a_1, \dots, a_p \in \mathbb{R}$) avec des ϵ_t qui forment un bruit blanc centré de variance σ^2 , tels que ϵ_t est indépendant de X_{t-1}, X_{t-2}, \dots pour tout t . Par convention : $X_{-k} = 0$ pour tout k dans \mathbb{N}^* . On dira aussi que (X_t) est un processus $AR(p)$.*

On dit que le X_t « s'explique » par les p observations précédentes $(X_{t-1}, X_{t-2}, \dots, X_{t-p})$.

Proposition 2 *On associe le polynôme de $\mathbb{R}[X]$ suivant à l'équation (A.3)*

$$A(X) = 1 - a_1X - \dots - a_pX^p \quad (\text{A.4})$$

Si les racines (dans \mathbb{C}) de ce polynôme sont toutes de module strictement supérieur à 1 alors il existe un processus stationnaire (X_t) vérifiant (A.3) et tel pour tout t , ϵ_t est indépendant de $X_{t-1}, X_{t-2}, \dots, X_{t-p}$.

Remarque 2 *La condition « racines de A de module > 1 » n'est pas nécessaire à l'existence d'une suite stationnaire vérifiant (A.3).*

Lemme 1 *S'il existe un processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant la relation de récurrence de l'équation (A.3) et si le polynôme de la relation de récurrence (A.4) a des racines qui sont toutes de module > 1 , alors la fonction d'auto-covariance défini en (A.1) du processus décroît exponentiellement quand $h \rightarrow +\infty$ (et donc sa fonction d'auto-corrélation défini en (A.2) a le même comportement).*

Proposition 3 *S'il existe un processus stationnaire $(X_t)_{t \geq 0}$ satisfaisant la relation de récurrence de l'équation (A.3) alors sa fonction d'auto-corrélation partielle vérifie*

$$r(h) = 0 \text{ si } h \geq p + 1.$$

A.4 Qualité d'ajustement de la modélisation des trajectoires

Dans cette annexe, les histogrammes, les fonctions d'autocorrélations et d'autocorrélation partielle des résidus des modèles de longitude et de latitude pour chaque section de latitude de 5° sont présentés.

A.4.1 Histogramme des résidus

A.4.1.1 Longitude

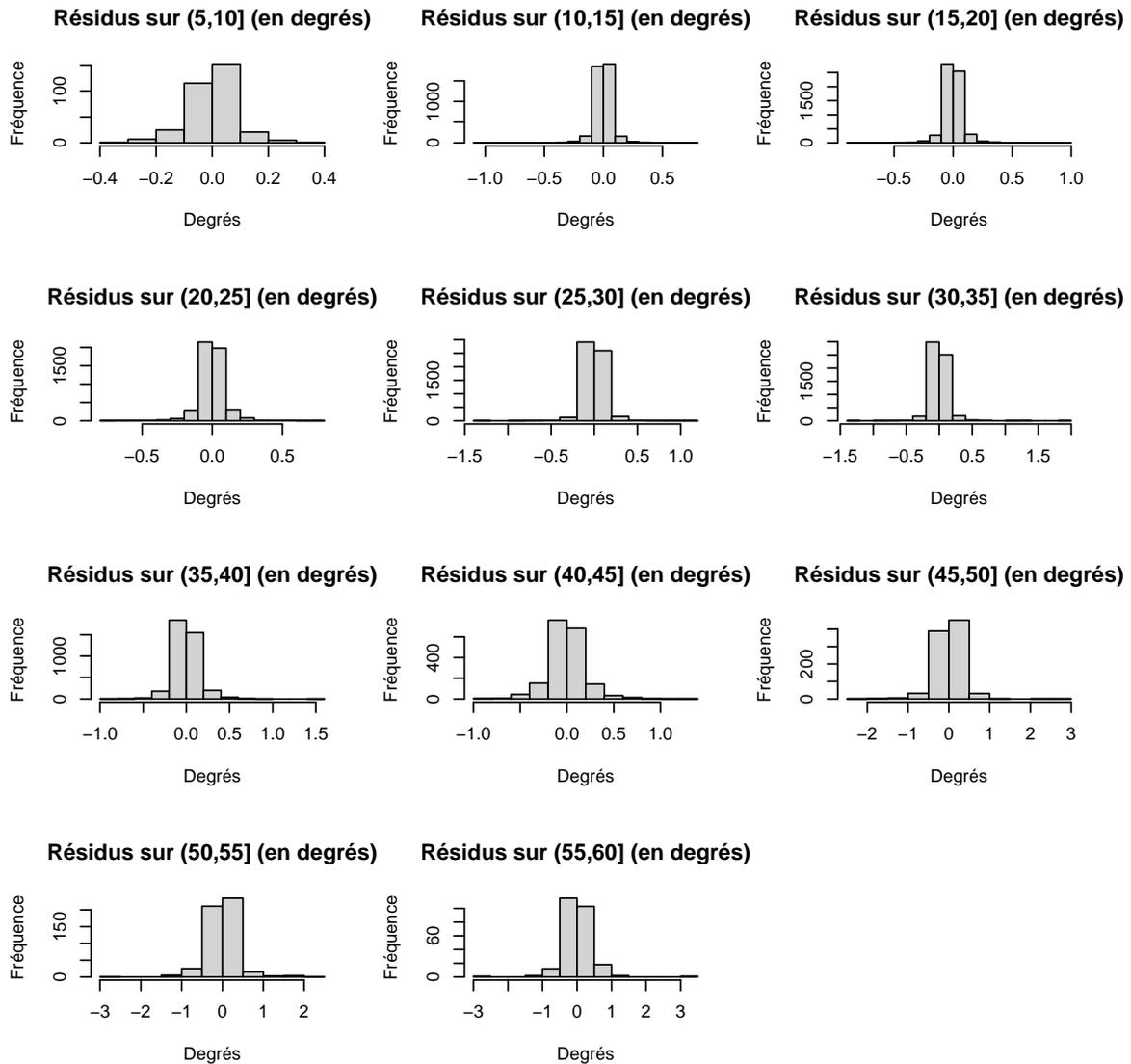


FIGURE A.1 : Histogramme des résidus des modèles de longitude

A.4.1.2 Latitude

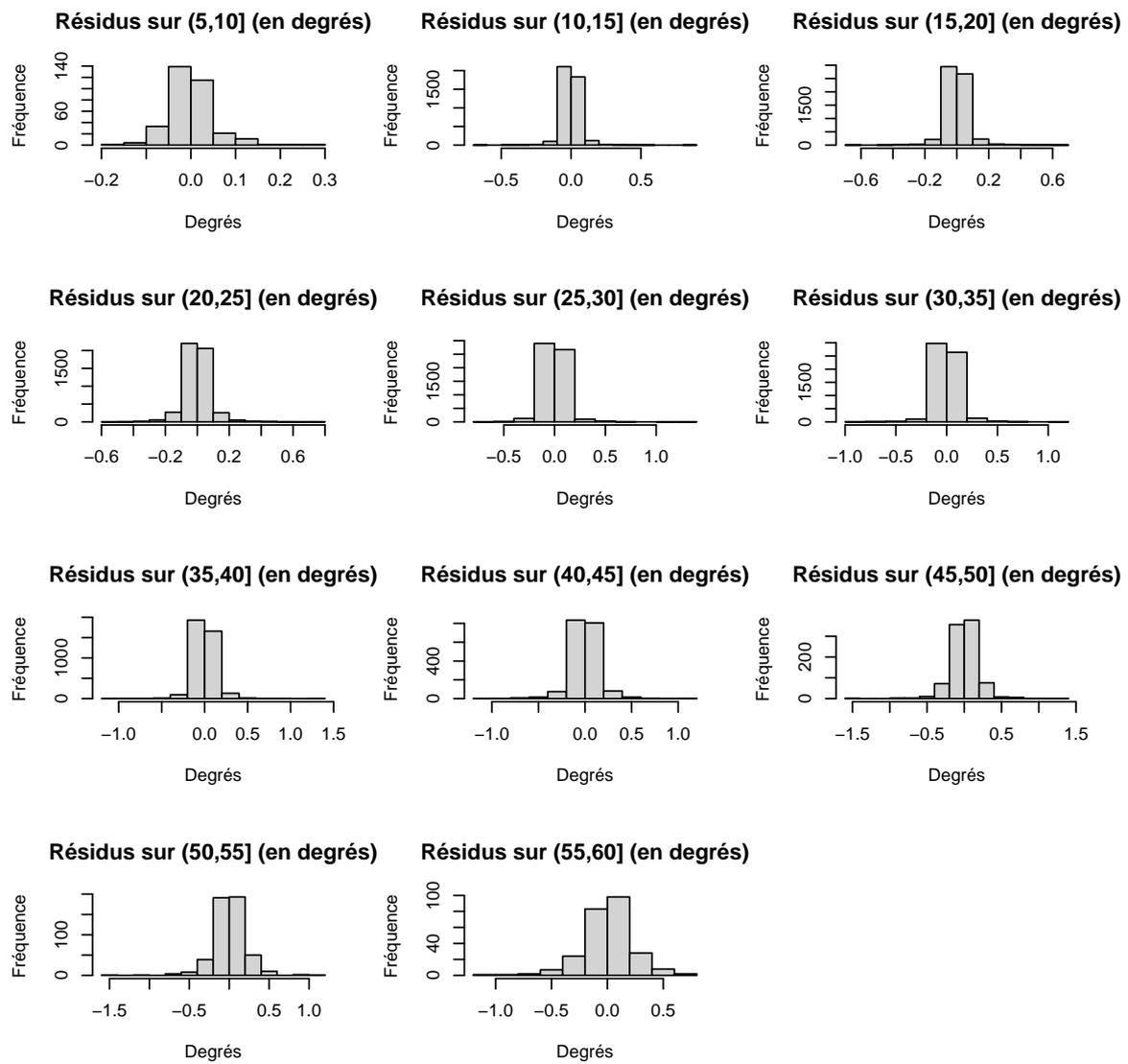


FIGURE A.2 : Histogramme des résidus des modèles de latitude

A.4.2 Autocorrélation des résidus

A.4.2.1 Longitude

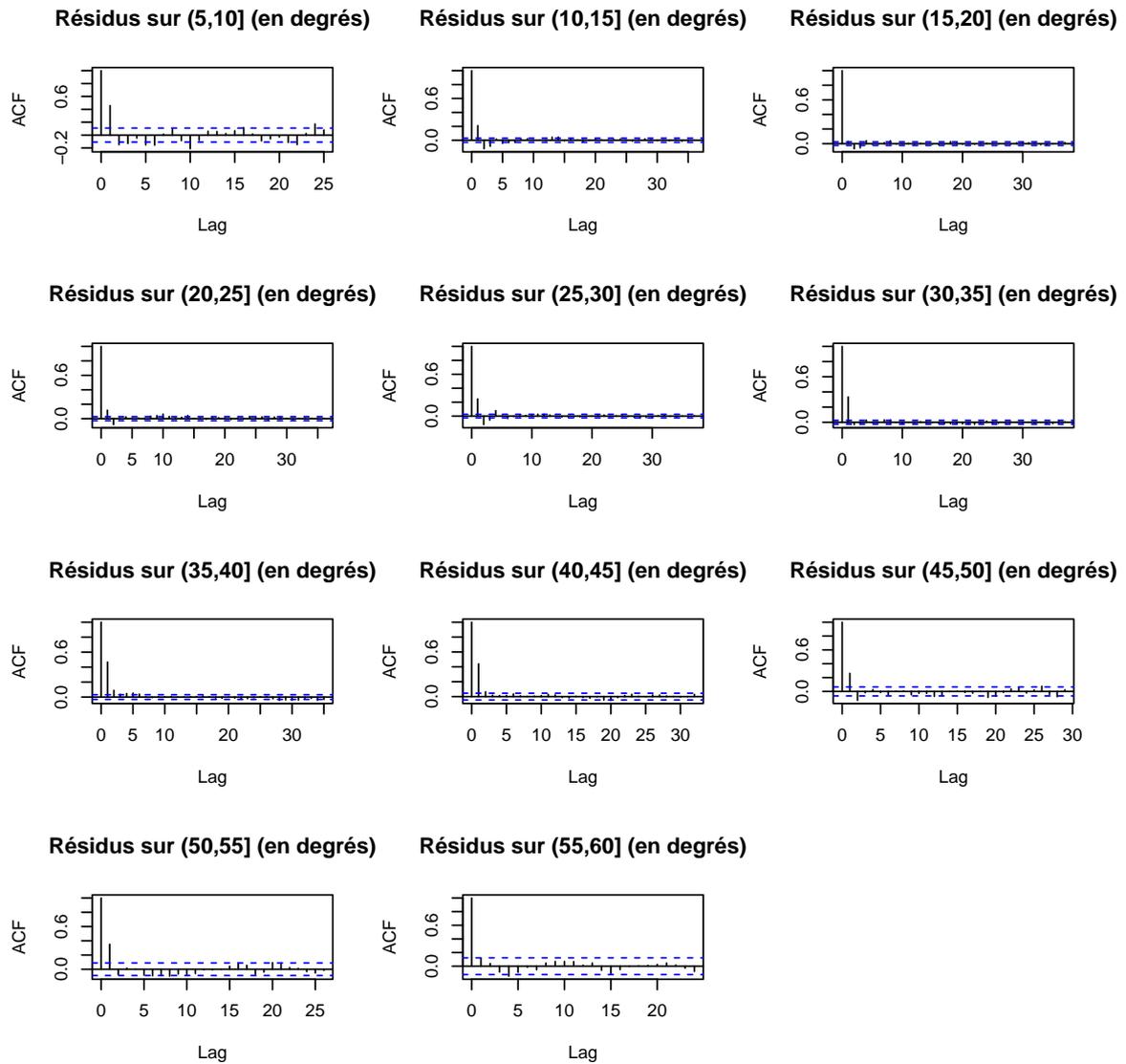


FIGURE A.3 : Autocorrélation des résidus des modèles de longitude

A.4.2.2 Latitude

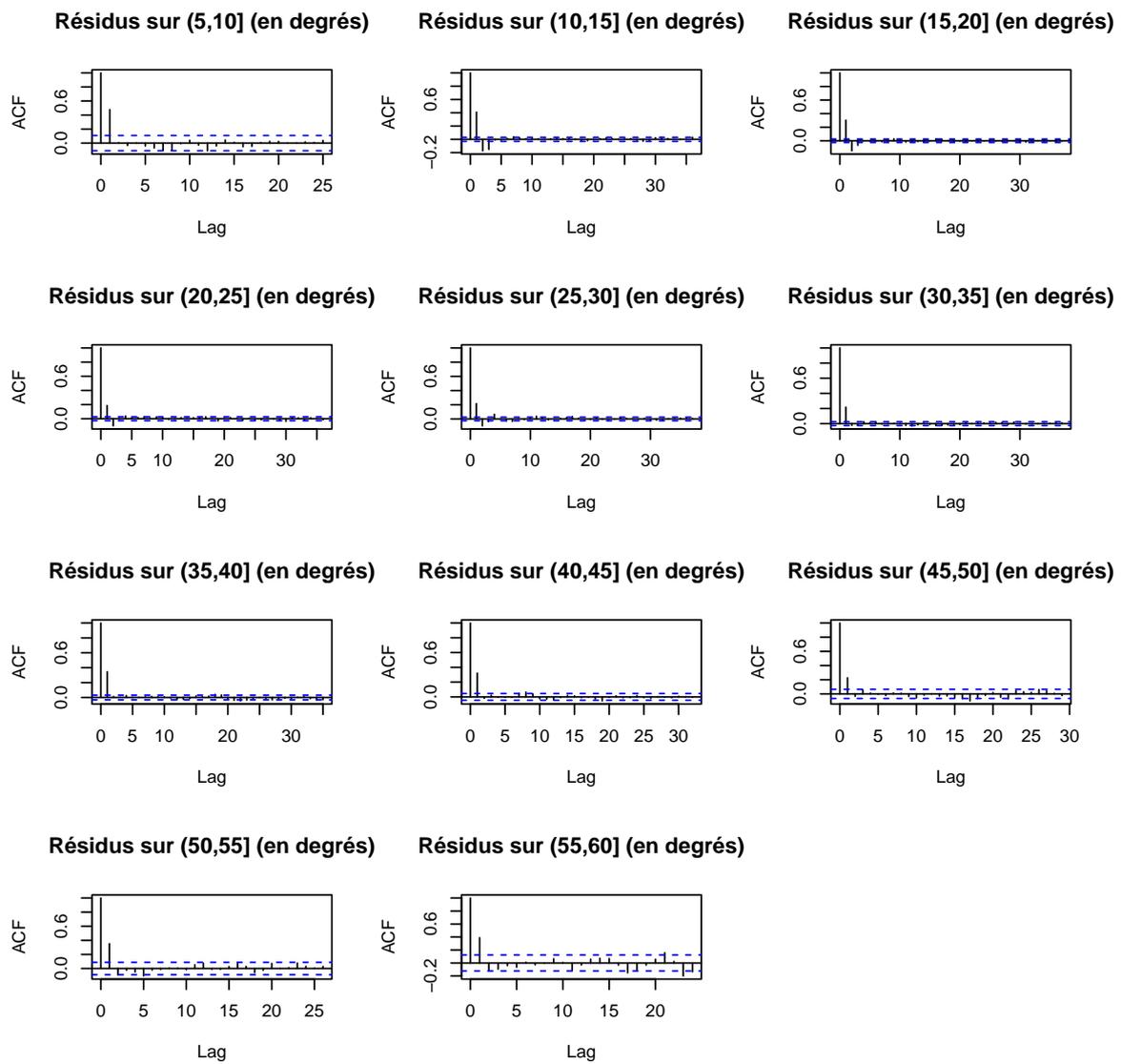


FIGURE A.4 : Autocorrélation des résidus des modèles de latitude

A.4.3 Autocorrélation partielle des résidus

A.4.3.1 Longitude

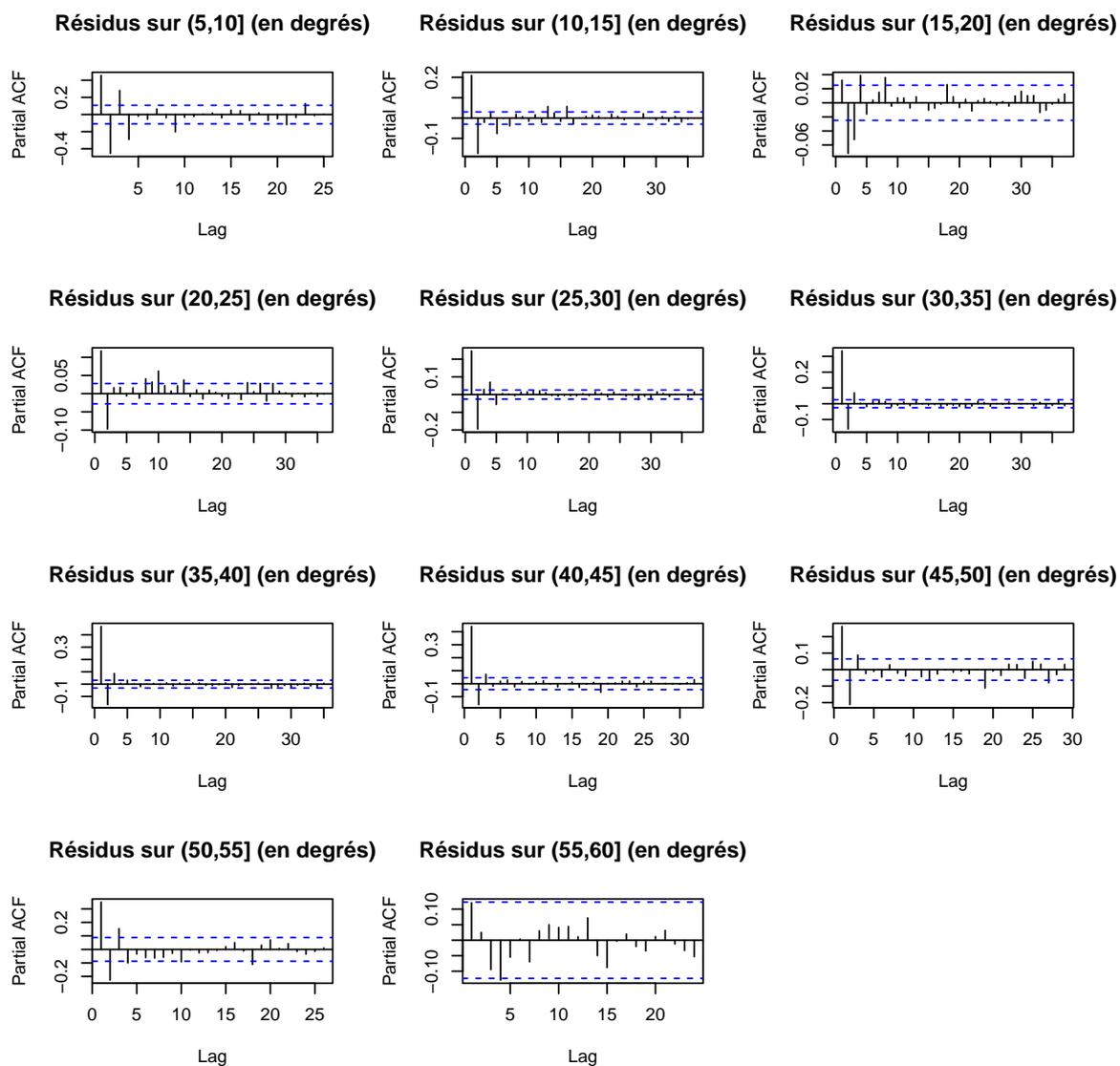


FIGURE A.5 : Autocorrélation partielle des résidus des modèles de longitude

A.4.3.2 Latitude

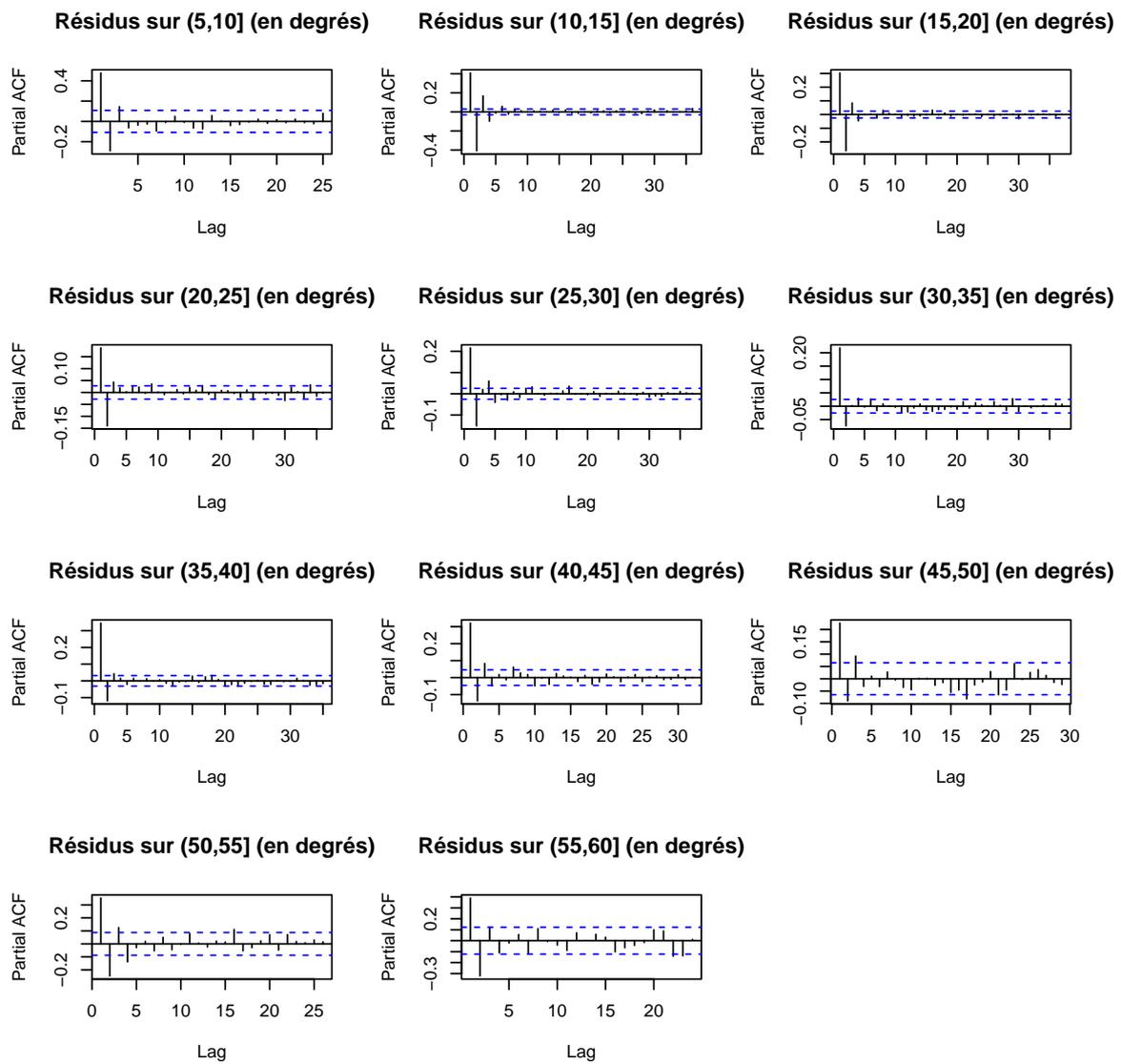


FIGURE A.6 : Autocorrélation partielle des résidus des modèles de latitude

A.5 Calcul de l'intensité relative

La procédure de calcul de l'Intensité Relative (IR) repose sur les travaux préalablement effectués par DARLING (1991). Son objectif fondamental consiste à élaborer un indicateur permettant d'évaluer l'intensité d'une tempête à un moment donné. Les auteurs ont avancé cette idée en postulant que, compte tenu des conditions atmosphériques et thermiques qui prévalent à un instant donné, la pression centrale d'un cyclone ne peut pas chuter en dessous d'une valeur minimale. Par conséquent, il est envisageable de créer un indicateur, compris entre 0 et 1, en comparant cette valeur théorique avec la pression centrale effectivement mesurée au sein de l'ouragan à cet instant précis. Les étapes requises pour calculer l'Intensité Relative (IR) sont exposées ci-après, bien que les concepts thermodynamiques et physiques sous-jacents à cette notion demeurent en dehors du cadre de cette étude.

Précisons tout d'abord les notations et les hypothèses employées :

- T_s représente la température à la surface de la mer traversée par l'ouragan, en Kelvin (K) ;
- $T_0 = 203 \text{ K}$ représente la température de la couche supérieure de la troposphère, en Kelvin (K) ;
- P représente la pression centrale au sein de l'ouragan, en hectopascals (hPa) ;
- P_{env} représente la pression atmosphérique, en hectopascals (hPa) ;
- Rh représente l'humidité relative de l'air fixé égale à 0.75 ;
- $M = 0.018 \text{ kg/mol}$ représente la masse molaire de l'eau ;
- $e_s = 6.112 \times \exp\left(\frac{17.67(T_s-273)}{T_s-29.5}\right)$ représente la pression de vapeur saturante de l'eau ;
- $L_v = 2.26 \times 10^6 - 2320(T_s - 273) \text{ J/kg}$ représente la chaleur latente de vaporisation de l'eau ;
- $R = 8.31447 \text{ J/K/mol}$ représente la constante des gaz parfaits ;
- $LW : u \mapsto LW(u)$ est la fonction de Lambert-W, également connue sous le nom de fonction de W de Lambert. Elle est définie comme la fonction réciproque de $u \mapsto ue^u$.

Tout d'abord, il est nécessaire de déterminer l'efficacité du cyclone qui se définit comme suit

$$\epsilon = \frac{T_s - T_0}{T_s}$$

Ensuite en utilisant la pression atmosphérique, la pression de vapeur saturante de l'eau et l'humidité relative de l'air, il est possible de déterminer la pression partielle de l'air sec ambiant en surface comme suit

$$p_{da} = P_{env} - (Rh \times e_s)$$

Puis il est nécessaire de déterminer A et B de la manière suivante :

$$\begin{cases} A = \frac{\epsilon L_v T_s}{(1 - \epsilon) R T_s p_{da}} \\ B = Rh \times \left[1 + \frac{e_s \ln(Rh)}{A p_{da}} \right] \end{cases}$$

Ensuite, pour résoudre l'équation en fonction de l'inconnue x suivante :

$$\ln(x) = -A \left(\frac{1}{x} - B \right)$$

la solution est donnée par la formule suivante :

$$x = \exp [LW (-A \exp(-A \times B)) + A \times B]$$

Avec cette valeur de x , il est possible de calculer la pression minimale théorique p_{dc} atteignable par un ouragan en fonction des conditions thermodynamiques dans lesquelles il se trouve, par la formule suivante :

$$p_{dc} = xp_{da}$$

Enfin, l'intensité relative associée à la pression centrale P peut être calculée à l'aide de la formule suivante :

$$IR = \frac{p_{da} - (P - e_s)}{p_{da} - p_{dc}} \quad (\text{A.5})$$

Et, en utilisant différents modèles de régression telles que celui de VICKERY et al. (2000) ou CUI et CARACOGLIA (2019) pour simuler des valeurs d'intensité relative, il est possible d'obtenir les pressions simulées en inversant simplement la formule en A.5. Cela donne la formule suivante :

$$P = p_{da} + e_s - IR \times (p_{da} - p_{dc}) \quad (\text{A.6})$$

A.6 Analyse comparative des modèles de V_{max}

Dans cette section, une comparaison approfondie des performances de deux modèles de modélisation de la vitesse maximale des vents (V_{max}) associée aux ouragans est proposée. Pour distinguer ces modèles et faciliter les comparaisons ultérieures, chaque modèle est désigné comme suit

- **Modèle 1** : $V_{max} = a_0 + a_1 \Delta P + a_2 \Delta P^2$, où a_0 , a_1 sont des coefficients à estimer et ΔP correspond au déficit de pression centrale.
- **Modèle 2** : $V_{max} = a \Delta P^b$, où $a > 0$, $0 < b < 1$ sont des coefficients à estimer et ΔP correspond au déficit de pression centrale¹. Il est à noter qu'une échelle logarithmique pour transformer le modèle 2, de manière à le ramener à un modèle de régression linéaire. Ainsi, le modèle 2 ajusté est exprimé sous la forme suivante

$$\text{Modèle 2 : } \ln(V_{max}) = b \times \ln a + b \ln(\Delta P).$$

Les estimations des coefficients pour chacun des modèles sont obtenues à l'aide de la fonction de régression linéaire du package *stats* R CORE TEAM (2023).

Nous commençons par comparer les coefficients obtenues ainsi que les incertitudes relatives liées telles que l'écart-type des estimations ainsi que les intervalles de confiance à 95%. Les coefficients estimés sont présentés sur le tableau A.2, les incertitudes relatives au estimations du modèle 1 sur le tableau A.4 et celles du modèle 2 sur le tableau A.3. Les notations σ_ε^1 et σ_ε^2 représentent respectivement l'écart-type des résidus du modèle 1 et 2.

Premièrement l'analyse en détail des estimations des coefficients et les incertitudes relatives liées à ces estimations est réalisée. Les résultats sont présentés de manière exhaustive dans les tableaux ci-dessous. Le tableau A.2 récapitule les coefficients estimés pour les deux modèles. Il présente également l'écart-type des résidus, noté σ_ε^1 pour le modèle 1 et σ_ε^2 pour le modèle 2. Ensuite les Tableaux A.4 et A.3 fournissent des détails plus spécifiques sur les incertitudes relatives liées aux coefficients de chaque modèle. Chacun de ces coefficients est associé à son écart-type ainsi qu'à un intervalle de confiance à 95% (**IC_{95%}**). Ces informations aident à évaluer la précision des estimations de chaque coefficient.

¹tel que défini dans le modèle de pression sur terre 3.14

Pression (hPa)	a_0	a_1	a_2	σ_ε^1	a	b	σ_ε^2
(881, 905]	1451.73	-21.58	0.10	13.55	3.03	0.53	0.05
(905, 929]	432.05	-5.65	0.04	19.30	2.48	0.67	0.07
(929, 953]	-266.75	11.33	-0.07	24.92	2.25	0.72	0.13
(953, 977]	16.82	3.57	-0.02	24.09	2.45	0.65	0.17
(977, 1003]	59.22	1.22	0.01	18.32	3.11	0.45	0.20
(1003, 1024]	37.14	3.26	-0.07	12.70	3.58	0.22	0.23

TABLE A.2 : Résultats des estimations des modèles 1 et 2

Pression (hPa)	σ_{a_0}	IC _{95%}	σ_{a_1}	IC _{95%}	σ_{a_2}	IC _{95%}
(881, 905]	929.56	[-462.74; 3366.20]	15.97	[-54.47; 11.32]	0.06	[-0.04; 0.23]
(905, 929]	308.53	[-175.67; 1039.77]	6.58	[-18.62; 7.32]	0.03	[-0.02; 0.10]
(929, 953]	73.95	[-411.82; -121.68]	2.06	[7.28; 15.38]	0.01	[-0.09; -0.03]
(953, 977]	16.48	[-15.50; 49.14]	0.67	[2.24; 4.89]	0.006	[-0.02; -0.002]
(977, 1003]	1.67	[55.93; 62.5]	0.13	[0.95; 1.49]	0.002	[0.009; 0.02]
(1003, 1024]	0.47	[36.19; 38.07]	0.11	[3.03; 3.48]	0.006	[-0.08; -0.05]

TABLE A.3 : Incertitudes relatives aux estimations des coefficients du modèle 1

Il est important de noter que ces incertitudes relatives jouent un rôle essentiel dans notre évaluation de la robustesse des modèles. Des écart-types plus petits et des intervalles de confiance plus étroits indiquent une plus grande précision des estimations, ce qui est un indicateur positif. De manière générale, il est observable que les incertitudes relatives aux estimations du modèle 2 sont meilleures notamment grâce à la transformation logarithmique appliqué à la variable cible. Cependant, il est également crucial de tenir compte du contexte, notamment le fait que les tranches de pression les plus faibles peuvent présenter des incertitudes plus importantes en raison du nombre limité d'observations. Pour une évaluation complète de la fiabilité des modèles, d'autres analyses, telles que l'examen de la qualité des résidus et des tests de prédiction, sont nécessaires.

Pression (hPa)	σ_a	IC _{95%}	σ_b	IC _{95%}
(881, 905]	0.91	[1.16; 4.91]	0.19	[0.14; 0.93]
(905, 929]	0.38	[1.73; 3.23]	0.08	[0.50; 0.83]
(929, 953]	0.16	[1.92; 2.57]	0.03	[0.64; 0.79]
(953, 977]	0.07	[2.30; 2.59]	0.01	[0.62; 0.69]
(977, 1003]	0.01	[3.08; 3.15]	0.005	[0.44; 0.47]
(1003, 1024]	0.006	[3.57; 3.60]	0.003	[0.21; 0.23]

TABLE A.4 : Incertitudes relatives aux estimations des coefficients du modèle 2

Dans les graphiques A.7, les résultats des tests de prédiction des vitesses maximales des vents effectués sur les trajectoires empiriques pour chacun des modèles considérés sont présentés.

Dans l'ensemble, il est observé que les deux modèles présentent des performances assez similaires en termes de prédiction et parviennent à reproduire de manière satisfaisante les zones où les vitesses maximales des vents sont particulièrement élevées. Une analyse de la qualité des résidus de chaque modèle est présentée dans la figure A.8. Cette analyse révèle que, pour la grande majorité des cas, les résidus du modèle 1 présentent un meilleur alignement par rapport aux quantiles théoriques d'une loi normale, en comparaison avec les résidus du modèle 2. Ces résultats suggèrent que le modèle 1 peut être légèrement plus approprié pour la modélisation des vitesses maximales des vents dans ce contexte.

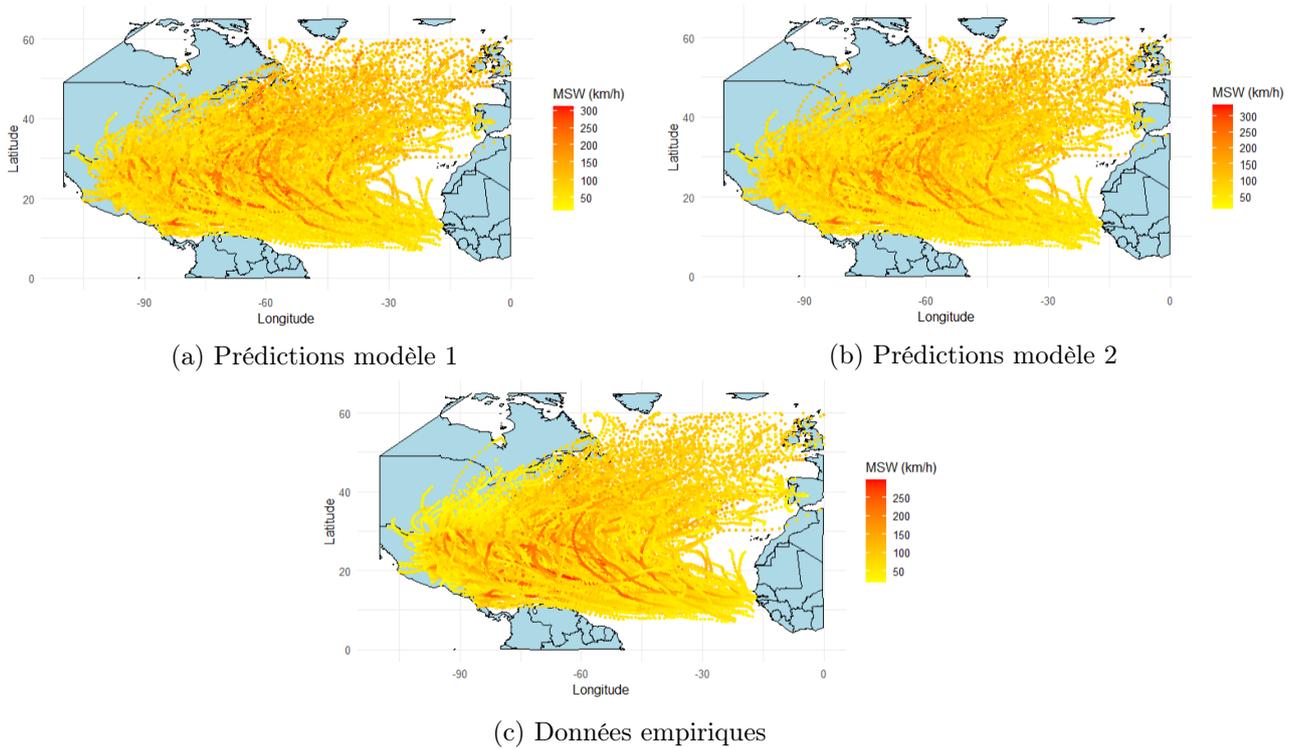
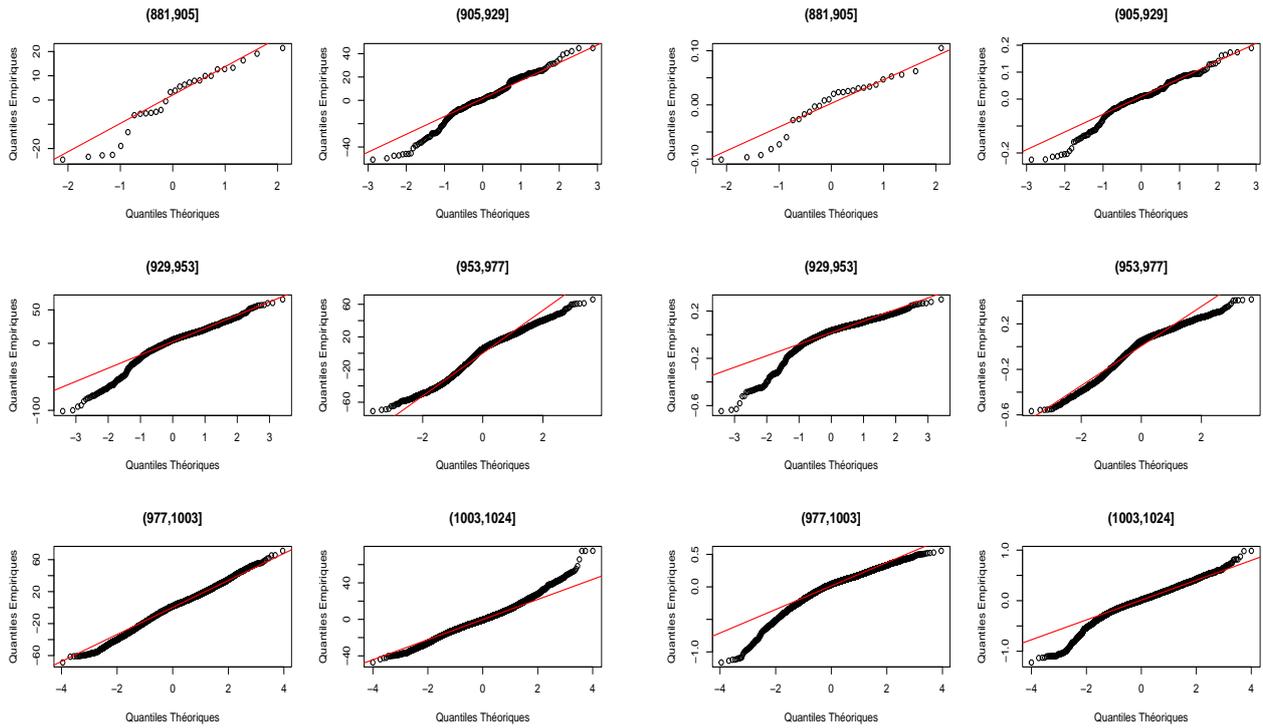


FIGURE A.7 : Comparaison des résultats de prédiction

spécifique.

La comparaison entre les deux modèles se conclut avec la représentation des QQ-Plot, illustrée dans la figure A.9, mettant en regard les valeurs prédites et observées associées à chaque modèle. L'observation révèle une tendance légère du modèle 2 à surestimer davantage les valeurs très élevées des vitesses maximales des vents par rapport au modèle 1. Cette tendance suggère une prudence accrue du modèle 2 dans ses estimations pour les événements de vent extrême.

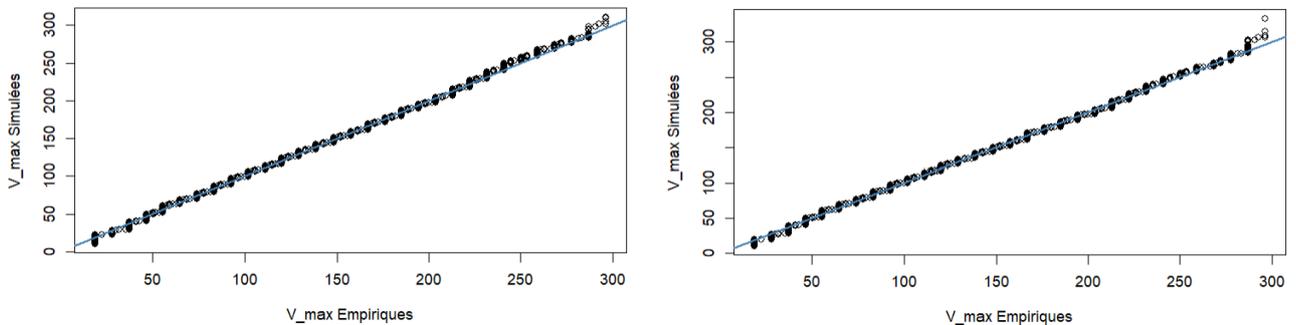
Pour résumer cette comparaison des deux modèles, il est important de noter que les différences en termes de robustesse entre les deux modèles sont minimales. Cependant, le modèle 1 se démarque comme un choix préférable par rapport au modèle 2. Cette préférence découle principalement de la meilleure qualité des résidus observée dans le modèle 1, ainsi que de son ajustement supérieur en termes de prédiction par rapport au modèle 2. Il est à noter que, malgré cet avantage du modèle 1, le modèle 2 présente également des mérites en termes d'estimation des coefficients. En fin de compte, le choix entre ces deux modèles doit être basé sur des considérations spécifiques au contexte et aux objectifs de modélisation, tout en tenant compte de l'équilibre entre la qualité des résidus et la précision de l'estimation des coefficients.



(a) QQ Plot des résidus du modèle 1

(b) QQ Plot des résidus du modèle 2

FIGURE A.8 : Comparaison des QQ-Plot des résidus



(a) QQ-Plot modèle 1

(b) QQ-Plot modèle 2

FIGURE A.9 : Comparaison des QQ-Plot des deux modèles

A.7 Rappels sur les modèles linéaires généralisés, XGBoost et Random Forest

Dans cette annexe, une synthèse des concepts théoriques sur les différents modèles utilisés pour la modélisation du coût financier est présentée.

A.7.1 Modèles linéaires généralisés (GLM)

Les modèles linéaires généralisés, *Generalized Linear Model*, (LINDSEY, 2008) sont une classe de modèles statistiques qui permettent de modéliser la relation entre une variable dépendante (dans ce cas, le coût des pertes) et un ensemble de variables explicatives (telles que la force de l'ouragan, la

taille du comté, etc...). Les principaux éléments à retenir à propos des GLM sont évoqués les suivants.

Forme de la distribution : Les GLM permettent de modéliser des données avec différentes distributions, comme la distribution binomiale, de Poisson, gamma, etc., en fonction de la nature de la variable cible et ne se limitent pas à une distribution normale des données, contrairement à la régression linéaire classique.

Fonction de lien : Les GLM incluent une fonction de liaison qui relie la moyenne de la variable dépendante à une combinaison linéaire des variables explicatives. Cette fonction de liaison est essentielle car elle définit la forme de la relation entre les variables.

Estimation des Paramètres : L'estimation des paramètres dans un GLM se fait généralement par la méthode du maximum de vraisemblance. L'objectif est de trouver les valeurs des paramètres qui maximisent la probabilité d'observer les données réelles étant donné le modèle.

Dans le cadre des GLM, les méthodes stepwise disponible à travers le package *MASS* (VENABLES et RIPLEY, 2002) sur R peuvent être appliquées en utilisant le critère AIC afin de retenir le modèle le plus pertinent du point de vue des variables explicatives. Elles sont pertinentes pour plusieurs raisons. Tout d'abord, ces méthodes stepwise aident à réduire la dimensionnalité en identifiant un sous-ensemble de variables qui contribuent le plus à la prédiction. Ensuite ces méthodes stepwise automatisent le processus de sélection de variables. Elles examinent systématiquement différentes combinaisons de variables pour déterminer celles qui améliorent le modèle. Cela évite la sélection manuelle, qui pourrait être subjective et biaisée. Ensuite les méthodes stepwise recherchent le modèle qui optimise un critère spécifié, comme l'AIC dans le cas de cette étude. Elles visent à trouver un équilibre entre la capacité du modèle à ajuster les données et sa complexité. Un modèle plus simple avec des variables significatives peut être plus robuste et généralisable. Et enfin un dernier argument qui est citable est concernant l'interprétabilité, un modèle avec un nombre limité de variables explicatives est plus facile à interpréter. Ainsi il existe trois méthodes de stepwise différentes :

- **Forward :** Cette méthode commence avec un modèle vide et ajoute séquentiellement des variables explicatives qui améliorent le modèle en termes d'un critère spécifié (par exemple, l'AIC). Elle est utile pour sélectionner les variables les plus informatives ;
- **Backward :** Cette méthode commence avec un modèle comprenant toutes les variables explicatives et supprime séquentiellement celles qui n'apportent pas d'amélioration significative au modèle. Elle est utile pour éliminer les variables redondantes ou non significatives ;
- **Both (stepwise bidirectionnelle) :** Cette méthode combine les approches forward et backward, explorant à la fois l'ajout et la suppression de variables pour optimiser le modèle.

A.7.1.1 XGBoost et Random Forest

Les modèles de Machine Learning tels que le XGBoost ou le Random Forest se distinguent par leur capacité à traiter des données complexes et à capturer des relations non linéaires au sein des ensembles de données, contrairement aux modèles linéaires traditionnels. Brièvement, le modèle XGBoost, *extreme Gradient Boosting*, est une implémentation améliorée de l'apprentissage ensembliste basée sur les arbres de décision, principalement utilisée pour la classification et la régression. Il se distingue par son efficacité et sa rapidité, grâce à des techniques d'optimisation telles que le gradient boosting. Son principe réside dans la construction séquentielle d'une série d'arbres de décision, où chaque nouvel arbre est ajusté aux erreurs résiduelles du modèle précédent. Cette approche itérative permet au modèle d'appréhender progressivement les structures de données complexes. Quant au Random Forest, c'est une méthode d'apprentissage ensembliste basée sur les arbres de décision. Il fonctionne en créant

un grand nombre d'arbres de décision individuels, chacun formé sur un sous-ensemble aléatoire des données d'entraînement. Chaque arbre de décision dans la forêt donne une prédiction, et la prédiction finale est obtenue par un vote majoritaire (pour la classification) ou une moyenne (pour la régression) des prédictions de tous les arbres. Le Random Forest est réputé pour sa capacité à gérer des ensembles de données complexes avec de nombreuses fonctionnalités et des relations non linéaires.

Il est à noter que ces deux modèles n'exigent généralement pas la segmentation de la variable cible en tranches de coût, car ils sont en mesure de gérer efficacement des valeurs réparties sur une large gamme et même des valeurs extrêmes. Cette flexibilité est un atout précieux dans des contextes où les données peuvent être très hétérogènes. Cependant, bien que ces modèles aient des similitudes, leurs mécanismes internes diffèrent, ce qui peut influencer leurs performances dans des situations spécifiques.

A l'aide des packages *xgboost* CHEN et al. (2023) et *randomForest* LIAW et WIENER (2002) disponibles sur R, une hyperparamétrisation, qui consiste à ajuster les paramètres de ces modèles pour optimiser leurs performances, peut être réalisée à travers une procédure de validation croisée pour chacun de ces deux modèles. Cette démarche méthodique permet d'explorer un espace de paramètres prédéfinis, tels que la profondeur maximale de l'arbre, le taux d'apprentissage, le nombre d'estimateurs, et bien d'autres, afin de déterminer la combinaison la plus efficace pour chaque modèle.

Le tableau A.5 synthétise les paramètres généraux des deux modèles utilisées nécessitant une optimisation par le biais de la validation croisée. Pour le modèle XGBoost, une métrique d'évaluation souvent utilisée est la racine de l'erreur quadratique moyenne (RMSE). Pour le modèle Random Forest, la détermination du meilleur paramètre *mtry* est souvent réalisée en recherchant le minimum de l'erreur *Out Of Bag* (OOB) en fonction de *mtry*. Cette méthode consiste à évaluer l'erreur de prédiction sur les échantillons de données qui n'ont pas été utilisés pour construire un arbre spécifique, fournissant ainsi une estimation fiable de l'erreur de généralisation du modèle en fonction de différentes valeurs de *mtry*. Ces configurations optimales sont obtenues après un processus méticuleux d'exploration des hyperparamètres, garantissant ainsi des performances optimales pour chacun de ces modèles.

Modèle	Paramètres généraux
XGBoost	<p>eta : <i>learning rate</i>. Le taux d'apprentissage, un coefficient compris entre 0 et 1. Ce paramètre influence la vitesse de convergence vers un optimum. Un taux faible favorise une convergence lente mais précise, tandis qu'un taux élevé peut entraver la convergence et altérer la précision des prédictions. ;</p> <p>max_depth : la profondeur maximale des arbres ;</p> <p>min_child_weight : le seuil minimal concernant le nombre d'individus requis dans un nœud de l'arbre. ;</p> <p>gamma : sert à régulariser les profondeurs des arbres. Plus cet hyperparamètre a une valeur élevée, plus la pénalisation empêche de construire des arbres trop profonds si l'apport en performance n'est pas aussi élevé ;</p> <p>colsample_bytree : ratio d'échantillonnage aléatoire sur les variables à chaque construction d'un arbre ;</p> <p>subsample : ratio compris entre 0 et 1 sélectionnant aléatoirement des observations à chaque itération pour entraîner un nouvel arbre ;</p> <p>nrounds : nombre maximum d'itérations.</p>
Random Forest	<p>ntree : nombre d'arbres à développer. Ce nombre ne doit pas être trop faible, afin de garantir que chaque ligne d'entrée soit prédite au moins plusieurs fois. ;</p> <p>mtry : nombre de variables échantillonnées aléatoirement comme candidates à chaque division. En d'autres termes, le paramètre <i>mtry</i> contrôle le nombre de caractéristiques d'entrée qu'un arbre décisionnel doit prendre en compte à un moment donné.</p>

TABLE A.5 : Paramètres généraux des modèles XGBoost et Random Forest

A.8 Gradient de vitesse maximale des vents

Dans cette annexe, une synthèse des concepts théoriques issus de la modélisation du gradient de la vitesse maximale des vents en un point donné d'un ouragan, tel que proposé par WILLOUGHBY et al. (2006) sont exposés.

Précisons tout d'abord les notations employées :

- $V(r)$: représente la vitesse maximale des vents², en m/s, à une distance r (en km) du centre de l'ouragan.
- w : sert de variable de pondération pour V_i et V_o .
- ξ : désigne un argument sans dimension.
- R_{max} : correspond au rayon de la vitesse maximale des vents, exprimé en km.
- V_{max} : correspond à la vitesse maximale des vents de l'ouragan, en m/s.

²Telle qu'elle a été définie en section 1.3

- R_1 : correspond au rayon allant du centre jusqu'au début de la zone de transition, en km.
- R_2 : correspond au rayon allant du centre jusqu'à la fin de la zone de transition, en km.
- ϕ : correspond à la latitude (en degrés nord).
- n : constitue le paramètre exponentiel qui quantifie la puissance à l'intérieur de l'œil de l'ouragan.
- A : sert de variable de pondération pour X_1 et X_2 .
- X_1 : exprime un premier paramètre mesurant la décroissance exponentielle au-delà de la zone de transition.
- X_2 : exprime un second paramètre mesurant la décroissance exponentielle au-delà de la zone de transition, fixé à 25.

Les paramètres du modèle se définissent comme suit

$$A = 0.0696 + 0.0049V_{max} - 0.0064\phi = \begin{cases} 0 & \text{si } A < 0 \\ A & \text{sinon} \end{cases}$$

$$n = 0.4067 + 0.0144V_{max} - 0.0038\phi$$

$$X_1 = 317.1 - 2.026V_{max} + 1.915\phi$$

$$w = \begin{cases} 0 & \text{si } \xi < 0 \\ 126\xi^5 - 420\xi^6 + 540\xi^7 - 315\xi^8 + 70\xi^9 & \text{si } 0 \leq \xi \leq 1 \\ 1 & \text{si } \xi > 0 \end{cases} \quad (\text{A.7})$$

où

$$\xi = \frac{R_{max} - R_1}{R_2 - R_1} \quad (\text{A.8})$$

Avant de procéder au calcul de $V(r)$ en chaque point, il s'avère essentiel de déterminer les valeurs de R_1 et R_2 en chacun de ces points. Pour cela, WILLOUGHBY et al. propose l'approche suivante.

Préalablement, pour calculer R_1 , il convient de déterminer le paramètre ξ , qui se résout par l'intermédiaire de l'équation suivante :

$$f(\xi) = w - \frac{n((1-A)X_1 + 25A)}{n((1-A)X_1 + 25A) + R_{max}} = 0 \quad (\text{A.9})$$

avec w, n, A et X_1 définies précédemment. La solution peut être obtenue au moyen de méthodes numériques, telles que la méthode de Newton-Raphson JONES et al. (2014) exposée à travers le théorème 1.

Théorème 1 (Algorithme de Newton-Raphson) Soit $c \in [a, b]$, où $[a, b]$ représente un intervalle de \mathbb{R} . Considérons une fonction $f : [a, b] \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 , telle que $f(a) > 0$, $f(b) < 0$, f' et f'' aient le même signe sur $[a, b]$ sans s'annuler. Dans ces conditions, l'équation $f(x) = 0$ admet une unique racine c dans l'intervalle $[a, b]$.

En partant d'une approximation initiale x_0 , plus ou moins proche de c , et en définissant la séquence (x_n) récursivement comme suit

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (\text{A.10})$$

alors la suite (x_n) converge vers la racine c . L'itération définie en (A.10) est poursuivie jusqu'à ce que la valeur absolue de $f(x_{n+1})$ soit inférieure à un seuil ε donné ou que le nombre maximal d'itérations autorisé soit atteint.

La dérivée de la fonction définie en (A.9), nécessaire pour l'application de la méthode de Newton-Raphson, s'exprime selon l'équation suivante :

$$f'(\xi) = 5 \times 126\xi^4 - 6 \times 420\xi^5 + 7 \times 540\xi^6 - 8 \times 315\xi^7 + 9 \times 70\xi^8$$

Une fois que la valeur de ξ est obtenue, le rayon R_1 est déterminé aisément en inversant la formule définissant ξ en (A.8) :

$$R_1 = R_{max} - \xi(R_2 - R_1)$$

Ici, $R_2 - R_1$ représente la largeur de la zone de transition, et elle est définie en fonction de l'intensité de la tempête cyclonique. Plus la valeur de R_{max} est faible et plus l'ouragan est considéré comme intense. Dans le modèle, deux cas sont distingués en fonction de cette intensité :

$$R_2 - R_1 = \begin{cases} 25 & \text{si } R_{max} > 20 \\ 15 & \text{si } R_{max} \leq 20 \end{cases}$$

Ensuite, la valeur de R_2 est déterminée de la manière suivante :

$$R_2 = \begin{cases} R_1 + 25 & \text{si } R_{max} > 20 \\ R_1 + 15 & \text{si } R_{max} \leq 20 \end{cases}$$

En conséquence, en utilisant les paramètres précédemment calculés, la vitesse maximale des vents se détermine au moyen des équations suivantes, en fonction de la distance r depuis le centre de l'œil de l'ouragan :

$$V(r) = V_i = V_{max} \left(\frac{r}{R_{max}} \right)^n, \quad (0 \leq r \leq R_1), \quad (\text{A.11a})$$

$$V(r) = V_o = V_{max} \left[(1 - A)e^{\frac{R_{max}-r}{X_1}} + Ae^{\frac{R_{max}-r}{X_2}} \right], \quad (R_2 < r). \quad (\text{A.11b})$$

où V_i et V_o désignent respectivement les composantes tangentielles du vent à l'intérieur de l'œil de l'ouragan et au-delà de la zone de transition.

Lorsque r se situe entre R_1 et R_2 , l'argument ξ défini en (A.8) devient :

$$\xi = \frac{r - R_1}{R_2 - R_1}$$

et la vitesse maximale des vents en fonction de la distance r est alors donnée par l'équation suivante :

$$V(r) = V_i(1 - w) + V_o w, \quad (R_1 \leq r \leq R_2)$$

où w est défini en (A.7).

A.9 Résultats des modèles GLM et ajustement de la variable coût

Dans cette annexe, les résultats des modèles linéaires généralisés (GLM) pour différentes tranches de coûts, comme discuté dans la section 4.1.2.1 sont présentés. Ces modèles sont obtenus en utilisant le package *stats* R CORE TEAM (2023). De plus, les graphiques quantile-quantile (QQ-Plot) pour les distributions gamma, log-normale, normale et exponentielle ajustées à la distribution de chaque tranche de coût sont également représentés.

Résumé des modèles GLM

Variabes	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.8519491	3.9860868	2.7224568	0.0065435
V_moy	-0.0170480	0.0039725	-4.2915066	0.0000187
HRCN_EXPB	0.0145071	0.0026966	5.3797297	0.0000001
STATE_RISKRelatively Low	0.1133053	0.1361852	0.8319945	0.4055245
STATE_RISKRelatively Moderate	-0.3228394	0.0798919	-4.0409521	0.0000555
STATE_RISKVery High	-0.2901246	0.1416355	-2.0483889	0.0406694
STATE_RISKVery Low	0.4494121	0.1537937	2.9221740	0.0035201
HRCN_AFREQ	0.6300505	0.4583629	1.3745668	0.1694401
V_max	0.0228460	0.0031080	7.3507093	0.0000000
CAT_max	-0.2344941	0.0630930	-3.7166398	0.0002082
Pres_moy	-0.0351165	0.0074896	-4.6886979	0.0000030
HRCN_EXPP	-0.0000010	0.0000004	-2.3526835	0.0187476
HRCN_EXP_AREA	0.0002131	0.0001035	2.0588611	0.0396538
Pres_min	0.0186011	0.0062882	2.9580913	0.0031366
HRCN_EXPA	0.6676422	0.1946917	3.4292275	0.0006192
R_moy	0.0032442	0.0008368	3.8771446	0.0001096
R_max	-0.0001432	0.0006057	-0.2364172	0.8131363

TABLE A.6 : Résumé du modèle GLM pour les coûts attritionnels (entre 1 et 100 millions \$)

Variabes	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.4398279	3.6864266	-0.9331063	0.3521711
Pres_min	0.0019505	0.0058815	0.3316374	0.7405971
Pres_moy	-0.0008218	0.0068191	-0.1205103	0.9042302
V_max	0.0039704	0.0026696	1.4872391	0.1389197
V_moy	-0.0001364	0.0031544	-0.0432477	0.9655580
R_moy	0.0031899	0.0010599	3.0096343	0.0030398
HRCN_EXPB	0.0016653	0.0022690	0.7339386	0.4640603
HRCN_EXPP	-0.0000002	0.0000004	-0.6879632	0.4924721
HRCN_EXPA	-0.3008263	0.1998125	-1.5055433	0.1341561

TABLE A.7 : Résumé du modèle GLM pour les coûts graves (entre 100 et 500 millions \$)

Variables	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-28.8706909	11.9561504	-2.4147146	0.0180006
Pres_min	0.0273331	0.0116892	2.3383139	0.0218374
V_max	0.0190531	0.0076724	2.4833231	0.0150803
V_moy	-0.0005474	0.0044151	-0.1239802	0.9016381
R_moy	-0.0009223	0.0035857	-0.2572042	0.7976734
CAT_max	0.0011197	0.2888375	0.0038764	0.9969166
HRCN_EXPB	0.0062194	0.0008810	7.0594679	0.0000000
STATE_RISKRelatively Moderate	-0.1279506	0.3148243	-0.4064191	0.6855065
STATE_RISKVery High	-1.1486734	0.2773658	-4.1413670	0.0000842

TABLE A.8 : Résumé du modèle GLM pour les coûts graves extrêmes (supérieures à 500 millions \$)

Ajustement des lois de probabilité à la variable coût

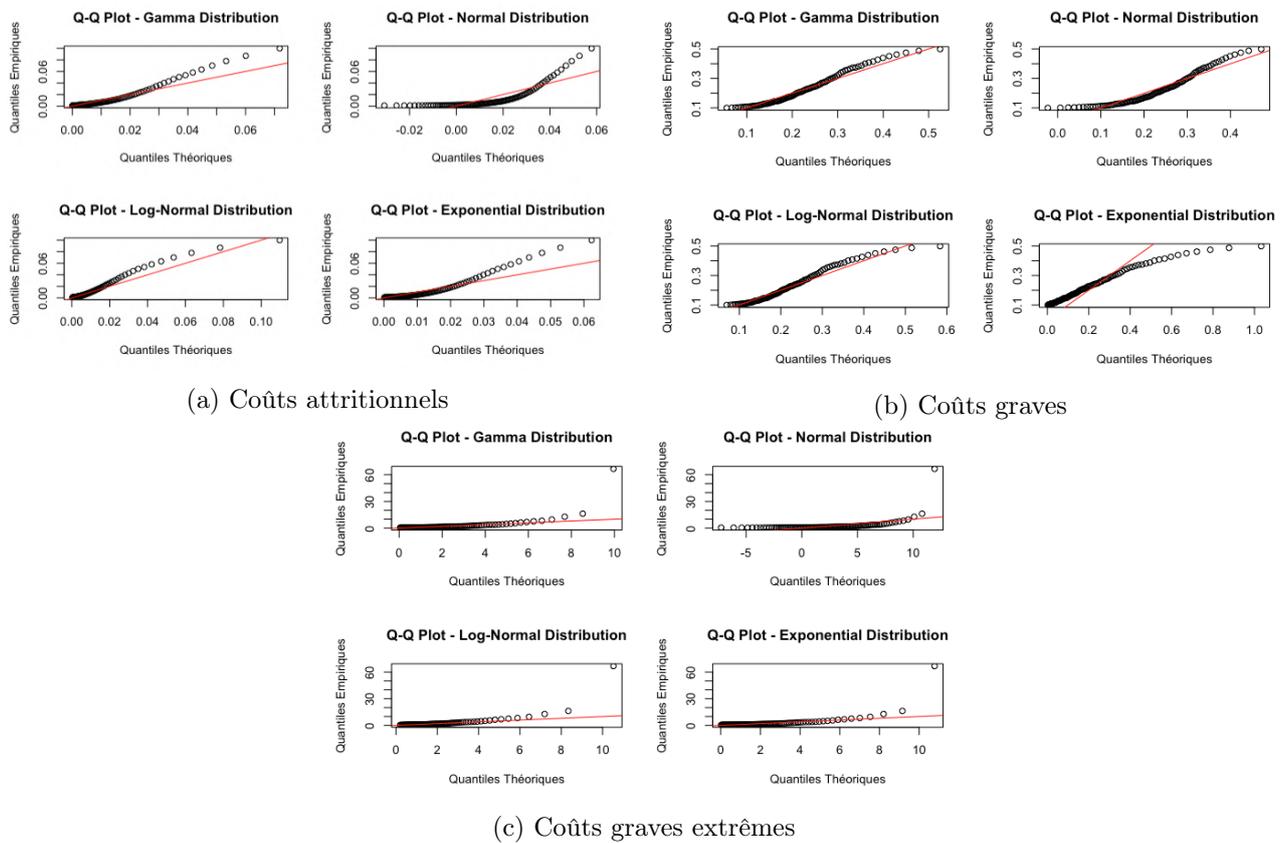


FIGURE A.10 : QQ-Plot de diverses distributions ajustées à la variable coût