

Société de Calcul Mathématique SA

Outils d'aide à la décision

depuis 1995



Exposition de Mayotte aux risques climatiques

Analyse des données disponibles et traitements statistiques

Développement d'un catalogue multi-aléa stochastique

Rapport adressé à la

Caisse Centrale de Réassurance
(à l'attention de *M. Charles Dumartinet*)

par la

Société de Calcul Mathématique SA

en référence au contrat signé le 30/03/2026

rédaction Lucas Busson, relecture Bernard Beauzamy

Résumé Opérationnel

Le présent document, établi à la demande de la Caisse Centrale de Réassurance, concerne l'exposition de Mayotte aux risques climatiques. Il est divisé en cinq parties :

- Acquisition des données "tempêtes" ;
- Traitement et exploitation des données ;
- Anticipation par simulation ;
- Utilisation de lois académiques (Weibull, Gumbel, etc.) ;
- Hauteur des marées.

1. Acquisition des données

Non sans difficulté, nous sommes parvenus à établir une liste de 77 tempêtes ayant frappé Mayotte entre 1900 et 2026 ; pour cela, nous avons dû consulter plusieurs sources de données, qui sont contradictoires, tant dans la nomenclature utilisée que dans le recueil des données (indications fournies par les capteurs). Plusieurs échelles existent ; celle de Météo France et celle de Saffir-Simpson ne coïncident pas.

Météo France indique que, parmi la vingtaine de stations présentes sur l'île, seules 12 sont encore en fonctionnement actuellement.

Nous ferons ici une recommandation, qui ne concerne pas directement la CCR, mais que la CCR peut appuyer.

Sur quelque risque que ce soit (séismes, volcanisme, inondations, incendies, radioactivité, tempêtes, etc.), l'acquisition de données a un coût et il est très difficile de la faire correctement. Bien souvent on s'aperçoit que les balises n'ont pas fonctionné correctement.

Dans l'article "Prévenir les incendies dans les sites isolés : éléments de méthodologie (Bernard Beuzamy, Michel Bénézit, Giovanni Bruna, Dominique Maillard)

https://www.scmsa.eu/archives/SCM_incendie_2019_05.pdf

nous avons rappelé que le réseau de balises "Telaray", établi à la demande de l'IRSN pour surveiller la radioactivité dans l'environnement, ne fonctionnait pas correctement et, de même, un réseau de capteurs mis en permanence sous les toits d'un bâtiment (Notre-Dame) pour détecter la possible apparition d'incendies, ne peut fonctionner correctement.

Dans le cas qui nous occupe ici (tempêtes à Mayotte), plutôt que des stations permanentes qui coûtent cher à entretenir, il serait préférable de disposer de véhicules mobiles, équipés d'anémomètres ; le jour où une tempête est annoncée, on envoie le véhicule.

Bien sûr, cette stratégie n'est envisageable que si le risque peut être annoncé suffisamment à l'avance ; c'est le cas pour tous ceux qui sont énumérés plus haut, sauf les séismes.

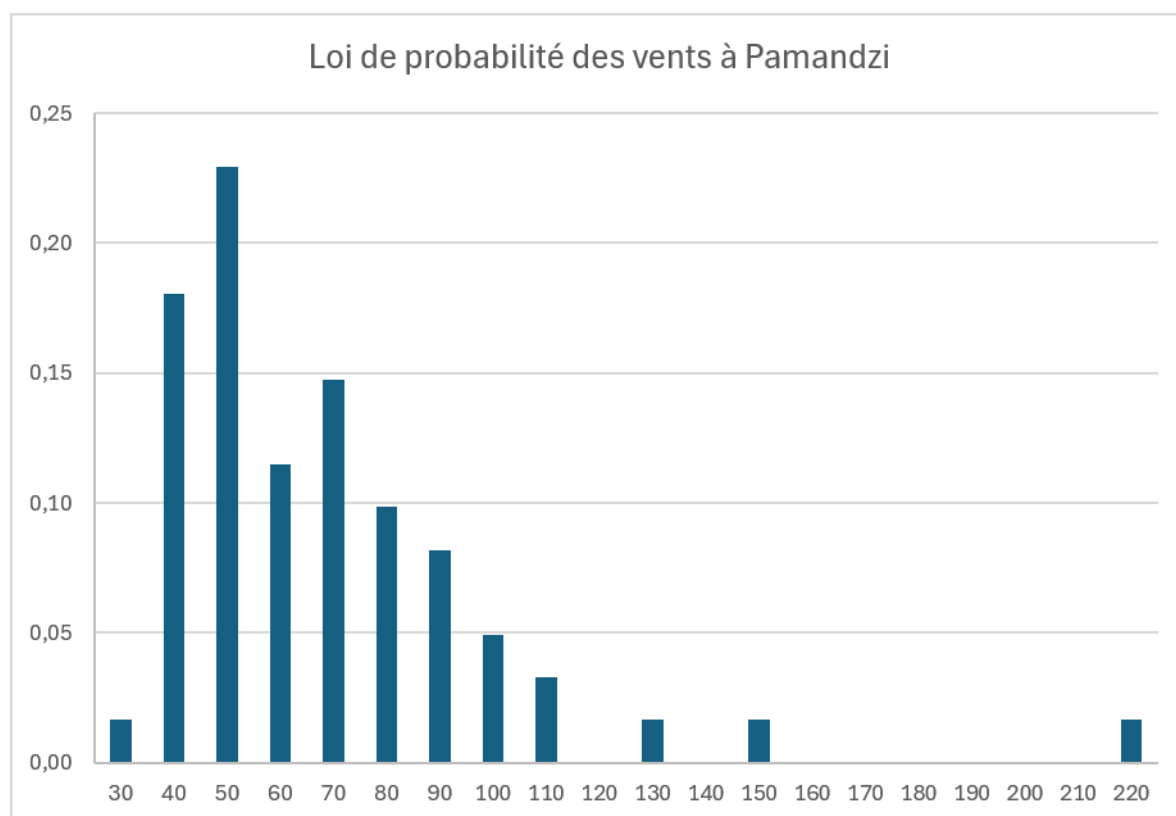
2. Traitement et exploitation des données

Comme les sources de données sont divergentes, nous avons fait deux traitements, correspondant aux données fournies par deux stations différentes :

- Station de Pamandzi

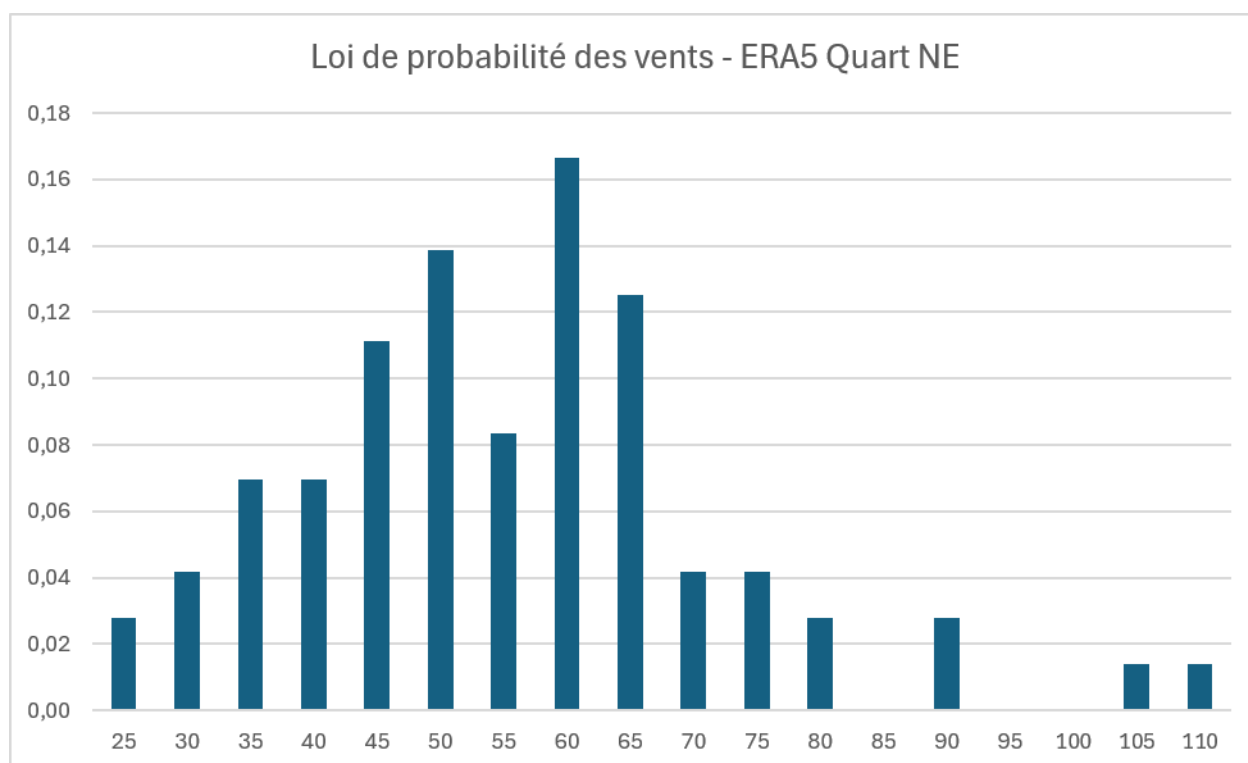
Nous avons établi la loi de probabilité de la vitesse des vents à la station de Pamandzi. Les données sont ponctuelles et relatives à cette station-là.

En abscisse, la force du vent, en km/h ; en ordonnée la probabilité.



- Site ERA5

De même, nous avons construit la loi de probabilité de la vitesse des vents selon le site ERA5. Les données ne sont pas ponctuelles, mais représentent une étendue ; chaque jeu correspond à un nœud du quadrillage, chaque nœud formant un carré entourant l'île de Mayotte. Le nœud le plus proche de la station de Pamandzi est celui situé au nord-est.



Ces deux lois de probabilité ne coïncident pas entre elles et ne peuvent se ramener à des lois de probabilités usuelles (Weibull, Gumbel ou autre) : ceci sera établi plus loin.

3. Anticipation par simulation

A la demande de la CCR, nous avons construit un simulateur : générateur de scénarios. Sur la base de l'historique connu, ce simulateur permet de générer des scénarios d'apparition de tempêtes dans le futur : telle année, tel nombre de tempêtes. Voici un exemple de résultat : prévision de 2031 à 2050 (le tableau sur 100 ans est donné plus loin).

En colonne 1 : l'année ;

En colonne 2 : le nombre de tempêtes prévu cette année-là ;

En colonnes 3, 4, 5, 6 : la force des vents associés à ces tempêtes (en km/h) ;

En colonne 7 : le signe "****" signifie que le seuil de 150 km/h est dépassé.

Cette notion de seuil (qui est totalement paramétrable) peut être utile à la CCR pour la définition des primes de réassurance.

année	nb de tempêtes	vent1	vent2	vent3	vent4	vent>150 km/h
2031	3	85	195	130		***
2032	0					
2033	0					
2034	2	135	95			
2035	1	95				
2036	0					
2037	2	65	170			***
2038	4	215	150	140	100	***
2039	1	95				
2040	2	165	115			***
2041	0					
2042	3	85	195	130		***
2043	2	120	150			***
2044	3	85	195	130		***
2045	2	165	115			***
2046	0					
2047	0					
2048	2	110	65			
2049	2	110	65			
2050	0					

4. Utilisation de lois académiques (Weibull, Gumbel)

A la demande de la CCR, nous répondons à la question : peut-on décrire correctement les lois de probabilité relatives à la vitesse des vents à Mayotte en utilisant des lois académiques usuelles, comme la loi de Weibull ou la loi de Gumbel ? La réponse est très claire et tout à fait négative : les lois en question ne dépendent que de deux paramètres ; de ce fait, elles sont beaucoup trop régulières pour rendre compte de la réalité des données, qui sont beaucoup plus complexes.

On peut essayer d'innombrables manières d'ajuster une loi de Weibull aux données ; aucune n'est préférable aux autres et toutes donneront des résultats fondamentalement différents. Il en résultera des querelles entre experts qui s'étendront sur des années et que rien ne pourra jamais trancher. Ceci est un phénomène général : à partir du moment où l'on cherche à simplifier les lois de la Nature, on est amené à faire des choix arbitraires.

5. Hauteur des marées

Elles sont mesurées en une seule station (source : Service Hydrographique et Océanographique de la Marine). Comme c'est le cas à Brest, la hauteur des marées peut varier considérablement d'une décennie à l'autre, du fait de variations de la pression atmosphérique.

6. En conclusion

Les données existantes pour les tempêtes sont rares, confuses, incohérentes et même contradictoires. On ne comprend pas pourquoi le nombre de stations supposées recueillir l'information diminue, ni pourquoi les différentes institutions responsables ne parviennent pas à s'entendre pour définir des nomenclatures communes.

En tout état de cause, notre travail met clairement en évidence le fait que les données disponibles ne permettent pas à une compagnie d'assurance d'évaluer précisément le risque "tempête" à Mayotte, ni à la CCR d'évaluer la réassurance correspondante, d'une manière qui soit juridiquement fiable.

7. Complément méthodologique

Il est très difficile de caractériser le risque "tempête" pour un assureur : nous avons essayé lors de nos travaux pour Monceau Assurances (2017-2022). Il y a deux difficultés fondamentales :

- La localisation : on ne sait pas exactement où la tempête est passée et on ne connaît pas la localisation exacte des biens assurés ;
- La nature même du phénomène : on mesure habituellement la valeur maximale du vent, mais non la durée, qui est pourtant fondamentale pour évaluer les dégâts.

Il en va exactement de même pour les inondations : on ne sait pas mesurer le débit d'un fleuve, et pour les séismes : la mesure de magnitude, sans évaluation de durée, est très insuffisante.

Il convient donc de se montrer indulgent quant à l'insuffisance des données recueillies : on ne peut guère faire mieux, sauf à multiplier des capteurs dont les pannes seront sources de préoccupation permanente. Comme expliqué plus haut, une solution satisfaisante et peu coûteuse consisterait à se doter d'un petit nombre de véhicules d'observation.

Par contre, là où la CCR peut montrer son exigence, c'est en demandant aux différents acteurs de se mettre d'accord sur la nature des mesures qui sont prises : que tout le monde mesure la même chose avec les mêmes unités et les mêmes échelles. Cela relève de l'exigence intellectuelle et ne comporte aucun coût associé.

Table des matières

Résumé Opérationnel.....	2
1. Acquisition des données.....	2
2. Traitement et exploitation des données	3
3. Anticipation par simulation.....	4
4. Utilisation de lois académiques (Weibull, Gumbel)	5
5. Hauteur des marées.....	5
6. En conclusion	5
7. Complément méthodologique.....	6
Remerciements	8
Rappel du besoin contractuel	9
Première partie.....	10
Acquisition des données	10
A. Liste des tempêtes	10
B. Différences entre les catégories	12
C. Données de Météo France.....	13
D. Données ERA5	17
Seconde Partie	19
Traitement et exploitation des données.....	19
I. Fréquence des tempêtes	19
II. Lois de probabilité.....	22
A. Pamandzi	22
B. ERA5	24
III. Différence entre les jeux de données	26
Troisième partie	28
Anticipation par simulation	28
I. Simulation de tempêtes.....	28
II. Réalisation de la simulation	31
III. Durée de retour des vents violents	32
IV. Programmation	33
Quatrième Partie.....	35
Introduction de lois académiques :.....	35
lois de Weibull et de Gumbel.....	35
A. Eléments historiques	35

B. Difficultés d'ajustement.....	37
Cinquième Partie.....	39
La hauteur des marées.....	39
Annexe : méthodes mathématiques pour la génération d'un scénario.....	40
Résumé	40
I. Position du problème.....	40
II. Programmation informatique.....	42
III. Analyse d'un exemple	43
IV. Un exemple de scénario	43
V. Mise en garde.....	44
1. Un exemple simple.....	44
2. L'exemple des températures extrêmes	44
VI. En conclusion	45

Remerciements

Il nous est agréable de remercier la Caisse Centrale de Réassurance pour nous avoir confié un sujet de travail qui est exactement dans nos cordes : les données existantes sont confuses, insuffisantes, contradictoires, et il s'agit, par des méthodes probabilistes, d'obtenir des conclusions qui pourront contribuer à une aide à la décision, comme nous l'avons fait pour la ville de Villiers le Bâcle (Essonne). Notre satisfaction tient au fait qu'aucune méthode déterministe ne peut prétendre y parvenir.

Le présent document peut être librement diffusé : toutes les données recueillies sont publiques et les méthodes mathématiques utilisées ont fait l'objet de publications. Les mathématiques ne sont pas brevetables.

Rappel du besoin contractuel

La CCR nous a adressé un cahier des charges intitulé "Développement d'un modèle stochastique cyclone sur Mayotte". Nous en extrayons les éléments suivants :

"Le cyclone Chido, survenu à Mayotte en 2024, a rappelé la forte exposition de ce territoire au risque cyclonique. Le cyclone constitue un péril multi-aléa, car il engendre simultanément des dommages liés au vent, aux inondations et aux submersions marines.

[...]

"À ce jour, la CCR ne dispose pas de modèle d'aléa stochastique pour le territoire de Mayotte. L'objectif de ces travaux est donc de développer un catalogue multi-aléa stochastique spécifique à Mayotte, afin d'améliorer la compréhension et la quantification du risque cyclonique sur l'île.

"Le territoire de Mayotte dispose de plusieurs stations météorologiques enregistrant des données historiques d'observation.

"Ces stations constituent une base précieuse pour le calibrage du modèle cyclonique. Toutefois, l'utilisation de ces observations doit être faite avec prudence, car les données peuvent présenter des lacunes ou des incohérences, notamment lors des événements extrêmes, période où les instruments peuvent être endommagés ou les mesures interrompues."

En premier lieu, la CCR souhaite disposer d'un historique aussi bien fait que possible, concernant les tempêtes ayant frappé l'île de Mayotte. Ces tempêtes sont caractérisées par leur nature et la force du vent maximal.

En accord avec la CCR, suite aux échanges téléphoniques que nous avons eus, nous nous sommes concentrés sur les vitesses de vents lors des tempêtes. En complément, nous avons aussi téléchargé et exploité les données relatives aux marées ; elles sont présentées en Cinquième Partie de ce rapport.

Première partie

Acquisition des données

A. Liste des tempêtes

Pour établir notre base de données, nous devons lister toutes les tempêtes ayant impacté Mayotte. Il n'existe pas actuellement de base de données toute prête recensant l'ensemble des tempêtes avec les vents associés. Cette base doit être créée, et ce n'est pas simple. Nous avons sélectionné toutes les tempêtes ayant frappé l'île de Mayotte, ainsi que celles qui se sont approchées à moins de 300 km des côtes de l'île. Certaines tempêtes plus éloignées (300-400 km) ont aussi été sélectionnées car, malgré la distance, elles ont provoqué des rafales de vents non négligeables sur les terres.

Ces choix sont arbitraires, mais résultent des considérations suivantes, qui sont de bon sens :

- Si une tempête revient, rien ne dit qu'elle suivra exactement le même chemin que les précédentes ;
- Si on se limite à celles qui ont précisément frappé Mayotte, on restreint la couverture.

L'extension à un voisinage est la démarche que nous avons suivie dans notre étude pour la ville de Villiers le Bâcle, Essonne, exposition au risque tempête :

https://www.scmsa.eu/archives/SCM_VLB.pdf

Dans un registre un peu différent (prévisions des évolutions de température), nous avons travaillé pour l'Andra : variation des températures au sol concernant le site de Bure (projet CIGEO) et, là encore, nous avons souhaité regarder tout un voisinage. Une évaluation purement ponctuelle paraît insuffisante.

Notre liste de tempêtes, pour Mayotte, a été conçue à partir de plusieurs sources. Tout d'abord, la principale, *l'Observatoire du Littoral de Mayotte* :

https://www.observatoire-de-mayotte.fr/observatoires_du_littoral/supports_dinformations/infos_tempetes

qui présente un historique des tempêtes et cyclones ayant impacté Mayotte. Cette liste commence en 1858 et s'achève en 2019. Pour chaque tempête de la liste, sont indiquées :

- la date ;
- la distance par rapport à Mayotte ;
- la pression atmosphérique ;
- ainsi que les conséquences pour l'île.

Nous ne savons pas pourquoi la liste se termine en 2019.

Pour actualiser cette liste, surtout par rapport aux tempêtes les plus récentes, nous avons consulté le site *DonneesMondiales* :

<https://www.donneesmondiales.com/afrique/mayotte/cyclones.php>

Il établit la liste des cyclones les plus violents à Mayotte depuis 2018. Cela nous a permis de compléter notre liste précédente. Ce site fournit, en plus du nom et de la date de la tempête, la vitesse maximale du vent, le diamètre, la pression de l'air, ainsi que la valeur sur l'échelle de Saffir-Simpson. Cette échelle (assez arbitraire) permet de catégoriser les cyclones en fonction de la vitesse maximale des vents. Elle sera détaillée ultérieurement.

Le site *Cyclonesxtreme* :

<https://www.cyclonextreme.com/cyclonereunion.htm>

recense tous les cyclones ayant impacté le bassin de l'Océan Indien sud-ouest, dont fait partie l'île de Mayotte. Ce site nous a été très utile pour estimer la distance entre les points de passage de la tempête et Mayotte, surtout pour les tempêtes les plus anciennes, entre 1960 et 1990.

Enfin, le site Wikipédia nous a permis de lister toutes les tempêtes ayant frappé le bassin sud-ouest de l'Océan Indien depuis 1950.

En combinant les différentes sources, nous avons établi une liste de tempêtes ayant impacté la région de Mayotte. Un extrait de cette liste est donné ci-dessous :

Nom	Date	Durée	Année	Mois	Terre ou non	Distance (km)	Direction	Nature	Saffir-Simpson	Vent max (1min)
Fytia	29/01/2026	7	2026	1	N	300	S	3	3	185
Jude	08/03/2025	8	2025	3	N	150	S	TT	1	150
Dikieledi	01/01/2025	17	2025	1	O	100	SE	1	3	185
Chido	05/12/2024	12	2024	12	O	0	E-O	4	4	250
Hidaya	01/05/2024	4	2024	5	N	400	N	TT	1	148
Gamane	26/03/2024	4	2024	3	N	200	E	DT	2	165
Filipo	02/03/2024	13	2024	3	N	150	SE	DT	TT	
Cheneso	17/01/2023	13	2023	1	N	400	SE	DT	2	155
Jasmine	21/04/2022	7	2022	4	O	0	SO-NE	DT	TT	100
Gombe	08/03/2022	7	2022	3	N	200	S	TT	3	185
Ana	20/01/2022	7	2022	1	N	300	S	TT	TT	95
Jobo	18/04/2021	7	2021	4	N	300	N	TT	TT	

Le tableau complet est donné dans le fichier Excel joint, intitulé « SCM_CCR_Tempetes » dans l'onglet Feuil3.

Les premières colonnes indiquent le nom de la tempête, ainsi que la date de début et la durée du phénomène. Les colonnes suivantes précisent l'année et le mois, pour pouvoir établir des statistiques. La colonne F précise si la tempête a touché terre à Mayotte. Les tempêtes étant passées à moins de 100 km des côtes de l'île sont considérées comme ayant frappé l'archipel. Les colonnes G à I précisent la distance minimale entre la trajectoire du cyclone et Mayotte, la direction par rapport au centre de l'île, ainsi que sa catégorie à ce moment-là. Les colonnes J et K indiquent l'échelle de Saffir-Simpson et le vent maximum sur une minute au plus fort de la tempête. Les colonnes L et M donnent les mêmes informations, mais selon l'échelle établie par Météo France.

Les différences entre les deux échelles seront expliquées plus bas. Les colonnes N et O indiquent les relevés de vent provenant des stations Météo France de Pamandzi et Coconi.

Les colonnes P à S donnent les relevés de vents selon le site ERA5 aux différents nœuds de maille entourant l'île. Enfin, la dernière colonne contient des observations sur les conséquences de la tempête sur l'île. Cela peut concerner le nombre de morts, la présence d'inondations ou de bâtiments endommagés.

Nous illustrons ce tableau en détaillant une tempête : la première de la liste, le cyclone Fytia. Il a commencé le 29 janvier 2026 et a duré 7 jours. Il n'a pas touché terre à Mayotte, mais est passé à 300 km des côtes sud de l'île. A ce moment-là, c'était un cyclone de catégorie 3 selon Saffir-Simpson. Il s'agit de la catégorie maximale de tempête, avec des vents maximaux de 185 km/h sur une minute. Météo France l'a catégorisé en tant que cyclone tropical avec des vents soutenus sur 10 minutes de 155 km/h. Les relevés des stations météo sur l'archipel indiquent des rafales maximales de 77 km/h à Pamandzi et 54 km/h à Coconi. On a observé des vents violents et des inondations lors du passage de Fytia sur Mayotte. On voit sur cet exemple combien les données sont incohérentes et différentes selon les sources.

La liste complète des tempêtes comprend 77 entrées entre 1900 et 2026. Avant 1960, les tempêtes ne sont pas détaillées et nous ne disposons pas d'informations telles que la catégorie, la trajectoire et les vents maximaux. Nous ne disposons que des relevés ERA5 jusqu'en 1940 et de quelques reconstitutions de vents à Pamandzi.

B. Différences entre les catégories

Les cyclones sont répartis en classes selon la force du vent. Plusieurs échelles existent, dont celle de Météo France et celle de Saffir-Simpson.

L'Organisation météorologique Mondiale (OMM) a chargé les Centres météorologiques régionaux spécialisés (CMRS) de surveiller et nommer les systèmes tropicaux. Tous ces centres utilisent la même échelle, basée sur des mesures de vents soutenus sur 10 minutes. Dès qu'un système tropical se forme, le CMRS correspondant est chargé de le surveiller, de le nommer et d'émettre des bulletins aux populations des pays concernés. Dans le cas de Mayotte, l'île dépend du bassin sud-ouest Indien dont le CMRS est le Centre météorologique régional spécialisé cyclone de la Réunion dépendant de Météo France.

Le *Joint Typhoon Warning Center* (JTWC), un centre de prévision des forces armées américaines, émet lui aussi des avis pour l'Océan Indien, même s'il n'est pas officiellement mandaté par l'OMM. Il n'est pas considéré comme un CMRS, car il est avant tout destiné aux forces armées américaines, mais il collabore avec les autres CMRS. Comme pour les autres organismes météorologiques américains, le JTWC utilise l'échelle de Saffir-Simpson pour catégoriser les tempêtes ; elle utilise les vents soutenus sur une minute pour déterminer la classe des cyclones.

Les terminologies et les classes de vents diffèrent selon les deux échelles. Elles sont détaillées dans les tableaux ci-dessous :

Echelle de Météo France (vents soutenus sur 10 minutes)		Echelle de Saffir-Simpson (vents sur 1 min)	
Vent (km/h)	Terminologie	Vent (km/h)	Terminologie
<51	Perturbation tropicale	<62	Dépression tropicale
51-62	Dépression tropicale	62-118	Tempête tropicale
62-88	Tempête tropicale modérée	118-153	Ouragan de catégorie 1
88-118	Forte tempête tropicale	153-177	Ouragan de catégorie 2
118-166	Cyclone tropical	177-209	Ouragan de catégorie 3
166-212	Cyclone tropical intense	209-251	Ouragan de catégorie 4
>212	Cyclone tropical très intense	>251	Ouragan de catégorie 5

C. Données de Météo France

Après avoir établi une liste de tempête ayant impacté Mayotte, la prochaine étape est d'y associer une vitesse de vent. Pour cela, nous avons cherché les stations météo disponibles sur l'île.

Le site *Meteo.data.gouv.fr* :

<https://meteo.data.gouv.fr/form>

permet de sélectionner toutes les stations gérées par Météo France et de télécharger les données de vents, de température et de précipitations.

Parmi les 21 stations présentes sur l'île, seules 12 sont encore en fonctionnement actuellement. Nous avons choisi les 6 stations les plus anciennes et avons téléchargé les données correspondantes. Même si l'historique est important, cela ne concerne pas tous les paramètres. En effet, la pluie est une donnée qui est enregistrée depuis longtemps, contrairement au vent qui ne semble être une préoccupation que depuis les années 80. Sur les 6 stations retenues, seules 2 ont enregistré les vitesses des vents. Comme pour les températures, plusieurs paramètres de vents sont enregistrés. Nous choisissons les rafales de vents instantanées maximales mesurées sur 3 secondes. Nous avons choisi de télécharger les données quotidiennes.

Parmi les stations ouvertes actuellement, toutes ne l'ont pas été en même temps. Certaines stations sont anciennes, ouvertes en 1935, et d'autres plus récentes, ouvertes uniquement en 2003. De plus, parmi les stations fermées actuellement, toutes n'ont pas connu la même période de fonctionnement. Dans les tableaux suivants, nous détaillons les dates d'ouvertures des stations et les périodes d'activité.

Tableau des stations ouvertes actuellement :

Stations	Année d'ouverture
Dzoumogne / Combani / Coconi	1935
Pamandzi	1947
Bandrele	1986
Mtzamboro	1991
Trevani / Mamoudzou	1993
Dembeni / Mbouini	1994
Vahibe_Sapc	2003
Dzoubogne-Double	2021

Tableau des stations fermées actuellement :

Stations	Période d'activité
Ajangoua	1981 – 2008
Vahibe Gouloue	1992 – 2003
Mt-Chongui	1995 – 1996
Mtsangamouji / Sada	1995 – 2003
Mzouazia	1996 – 2018
Mtzamboro (Ecole)	2003 (2 semaines)
Poroani	2004 – 2016
Mronabeja	2019 - 2021

Les deux stations les plus intéressantes, disposant de données sur les vents, sont Pamandzi (la station de l'aéroport de Mayotte, située à Petite-Terre) et Coconi, située au centre de l'île de Grande-Terre. Dans le cas de la station de Pamandzi, même si les températures et les précipitations sont mesurées depuis 1947, les vents ne sont enregistrés que depuis 1981. La différence est encore plus flagrante pour la station de Coconi où la pluie est mesurée depuis 1936, la température depuis 1982 et les vents depuis 2016.

Mais les dates annoncées ne reflètent pas toute la réalité. Même si les données ont été enregistrées depuis cette date-là, cela ne signifie pas que la liste est complète et qu'il n'y a pas eu d'interruptions dans les mesures. Un rapide coup d'œil aux données révèle que des interruptions dans les mesures ont bien eu lieu. Certains arrêts sont brefs, quelques jours seulement. Mais d'autres sont plus importants, amenant à des interruptions dans les enregistrements de plusieurs mois, voire plusieurs années.

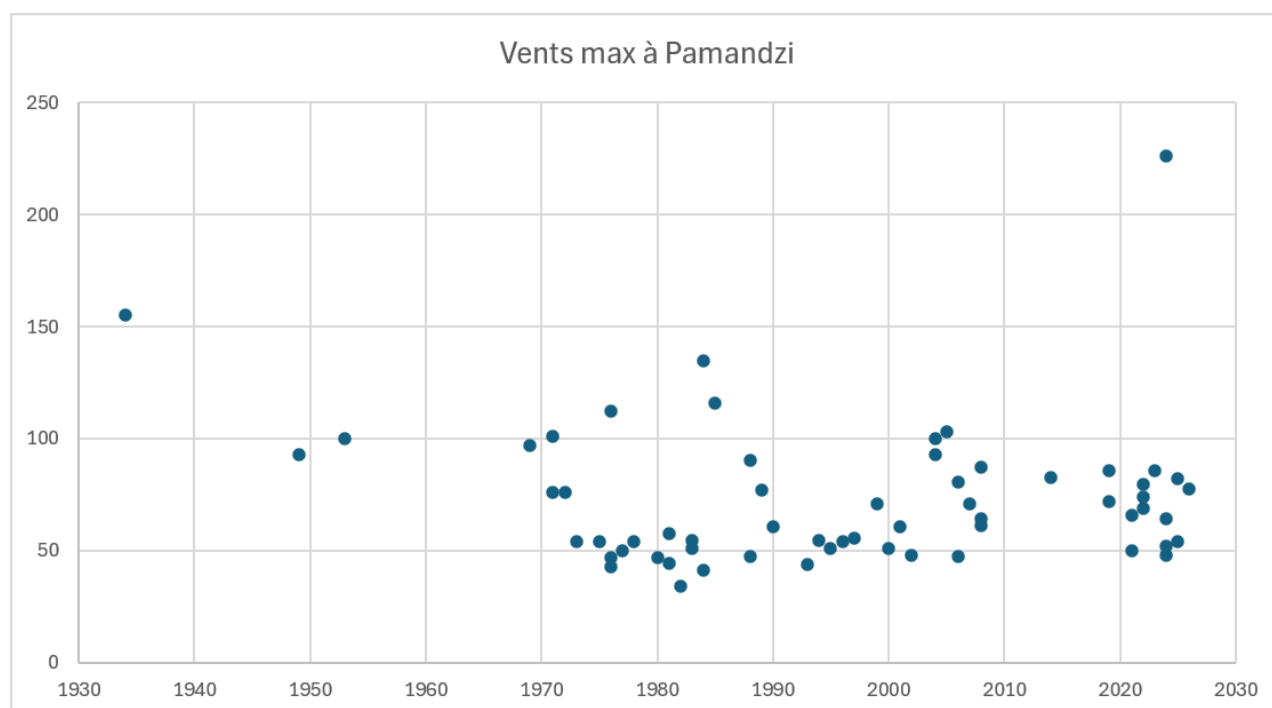
Nous avons effectué nos analyses à partir des données des stations de Coconi et Pamandzi, mais ce ne sont pas les seules stations dont nous ayons téléchargé les données. Outre ces deux stations, nous avons aussi récupéré les données des stations de Bandrele, Combani, Dzoumogne et Trevani. Ce sont les stations, encore ouvertes actuellement, dont l'historique est le plus important. Trevani a été choisie car elle dispose de données de vent, contrairement à la station de Mtzamboro, plus ancienne. Cela représente en tout 6 stations. Pour chacune d'elles, nous avons récupéré les données de vent (si disponibles) et celles de température et de pluie. Pour illustrer notre propos sur les interruptions de mesures, nous listons pour chaque paramètre et chaque station, les arrêts dans les enregistrements. Les résultats sont présentés dans le tableau ci-dessous :

Station	Paramètre	Début des mesures	Interruptions
Bandrele	Pluie	Février 1986	Décembre 2005
	Température max	Mai 2004	14/10/2014 au 1/01/2025
	Température moy	Septembre 2007	13/10/2014 au 10/01/2025 + quelques trous
Coconi	Pluie	1936	2013 à 2015 + 3 jours d'arrêt
	Température max	Février 1992	01/07/1992 au 01/01/1993 28/02/2005 au 01/07/2005 2013 à 2015 05/05/2023 au 15/06/2023 + quelques trous
	Température moy	Novembre 2003	20/11/2003 au 30/04/2004 22/04/2004 au 1/10/2007 2013 à 2015 1/01/2022 au 12/10/2022 03/05/2023 au 16/06/2023 13/01/2024 au 06/03/2024 03/04/2024 au 21/05/2024 10/07/2024 au 26/11/2024 13/12/2024 au 22/01/2025 + nombreux trous
	Vent	17/11/2016	11/01/2022 au 12/10/2022 03/05/2023 au 16/06/2023 13/01/2024 au 06/03/2024 10/07/2024 au 26/11/2024 + nombreux trous
Combani	Pluie	1936	1954 à 1960 + 1 jour d'arrêt
Dzoumogne	Pluie	1936	01/02/1944 au 01/04/1950 01/05/1952 au 01/01/1960
	Température max	Juin 2021	Quelques trous
	Température moy	11/06/2021	21/05/2022 au 13/08/2022 + nombreux trous
Pamandzi	Pluie	1947	Nombreux trous
	Température max	1947	01/01/1949 au 31/12/1950 + quelques trous
	Température moy	18/11/1998	01/01/2002 au 25/05/2002 + nombreux trous
	Vents	1981	17/09/2003 au 04/03/2024 + nombreux trous
Trevani	Pluie	1993	31/07/2003 au 29/01/2019 + quelques trous Arrêt : 27/03/2025
	Température max	30/01/2019	23/04/2023 au 15/06/2023 16/07/2023 au 30/08/2023 + quelques trous

	Température moy	30/01/2019	28/09/2019 au 20/11/2019 27/09/2021 au 26/10/2021 22/01/2022 au 18/03/2022 10/06/2022 au 19/07/2022 21/04/2023 au 16/06/2023 16/07/2023 au 30/08/2023 Arrêt : 27/03/2025 + nombreux trous
	Vents	30/01/2019	17/03/2019 au 07/05/2019 27/09/2019 au 20/11/2019 26/09/2021 au 26/10/2021 21/01/2022 au 19/03/2022 10/06/2022 au 19/07/2022 Arrêt : 18/04/2023 + nombreux trous

Pour associer un vent à une tempête, nous avons regardé les données issues des stations pendant la période d'existence de la tempête et nous avons sélectionné la valeur de vent quotidienne la plus élevée. Ainsi, nous avons le vent maximal à Pamandzi pendant toute la durée de la tempête. Nous avons procédé avec la même méthode pour les données issues de la station de Coconi.

Le nuage de points ci-dessous présente les forces de vents maximales pour chaque tempête en fonction des années :



En abscisse, se trouvent les années, et en ordonnées les vitesses de vents en km/h. On ne constate pas d'évolution dans les vitesses de vent. En général, les intensités des cyclones sont semblables entre 1970 et aujourd'hui.

D. Données ERA5

Les données de Météo France ne commençant qu'en 1981, nous avons cherché des données plus anciennes, afin d'avoir un historique plus important. Suivant la suggestion de la CCR, nous nous sommes intéressés au site ERA5, dépendant du programme Copernicus, organisme géré par l'Union Européenne.

Parmi les sets de data disponibles, nous avons sélectionné *ERA5 post-processed daily statistics on single levels from 1940 to present* :

<https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics?tab=download>

Le principe du site ne repose pas sur des relevés de stations à des endroits précis. La base de données couvre l'ensemble de la planète. Nous ne sélectionnons pas une station pour récupérer les données, mais une zone géographique. Pour constituer cette base, le site utilise les relevés des stations, mais aussi des modèles qui reconstituent le climat sur toute la surface du globe. Le monde est fragmenté en petites zones selon un quadrillage. Un nœud est présent tous les 0.25° en longitude et latitude. Les données sont fournies pour chaque nœud du quadrillage.

Nous sélectionnons les données de réanalyse. Parmi les très nombreux paramètres disponibles, nous choisissons *10m wind gust since previous post processing* dans la catégorie Wind. Ce paramètre mesure les rafales de vent enregistrées à 10m d'altitude sur une durée de 3s. Il s'agit du même paramètre que celui de Météo France. On ne peut choisir qu'une seule année à la fois. Nous avons dû effectuer une trentaine de requêtes pour récupérer les données de toutes les années. Nous ne récupérons pas toutes les données disponibles, mais uniquement celles des mois où une tempête s'est produite. Les statistiques choisies sont les maxima quotidiens. La Time Zone est UTC+03, qui correspond au fuseau horaire de Mayotte. Pour l'aire géographique, nous ne voulons que les données de l'archipel. Nous avons recherché les coordonnées de longitude et de latitude de Mayotte et avons déterminé que l'île se trouvait entre les coordonnées 12.63°N/45°O et -13°S/45.31°E. 4 mailles du réseau correspondent aux coordonnées de Mayotte. Il s'agit de :

- Latitude : -12.75° et Longitude : 45°. Située au Nord-Ouest ;
- Latitude : -12.75° et Longitude : 45.25°. Située au Nord-Est ;
- Latitude : -13° et Longitude : 45°. Située au Sud-Ouest ;
- Latitude : -13° et Longitude : 45.25°. Située au Sud-Est.

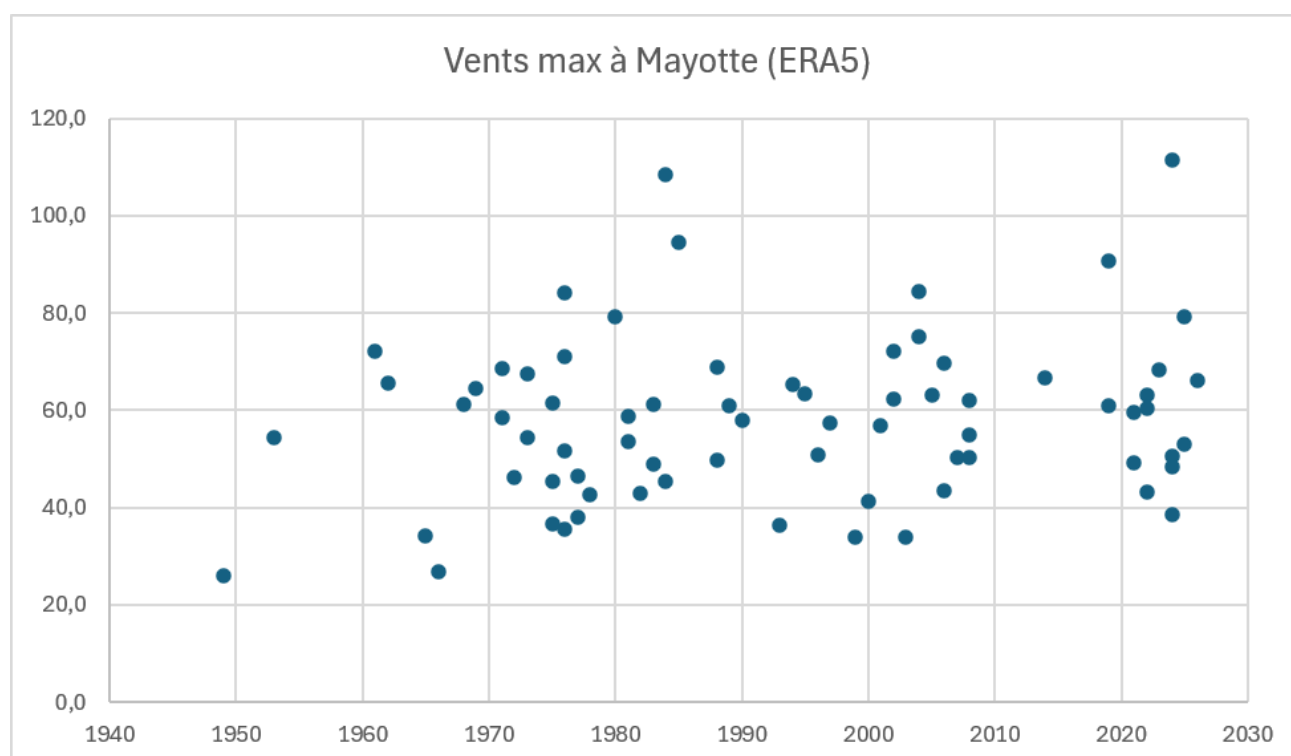
La station de Pamandzi est située à proximité du quart NE. Cela nous permet de comparer les relevés de la station et ceux issus d'ERA5.

Un ajustement des données a été nécessaire, afin de les ranger en ligne et d'avoir les relevés des 4 nœuds sur la même ligne correspondant à une date précise. Un extrait est donné ci-dessous :

Date	Quart NO	Quart NE	Quart SO	Quart SE
28/02/2026	25,5	25,8	24,7	24,6
27/02/2026	26,9	29,1	27,6	28,4
26/02/2026	26,8	28,4	27,9	29,0
25/02/2026	23,4	23,8	23,9	23,6
24/02/2026	18,6	17,7	17,9	18,3

Pour chaque jour, nous avons les relevés pour les nœuds du quadrillage autour de Mayotte.

Le nuage de points ci-dessous présente les forces de vents maximales pour chaque tempête en fonction des années :



En abscisse, se trouvent les années et en ordonnées les vitesses de vents en km/h. Les données proviennent du quart NE décrit précédemment. On ne constate pas d'évolution dans les vitesses de vent. Comme précédemment, en général, les intensités des cyclones sont semblables entre 1970 et aujourd'hui.

On note que ce second nuage de points ne ressemble pas du tout au premier, conséquence de l'hétérogénéité des données.

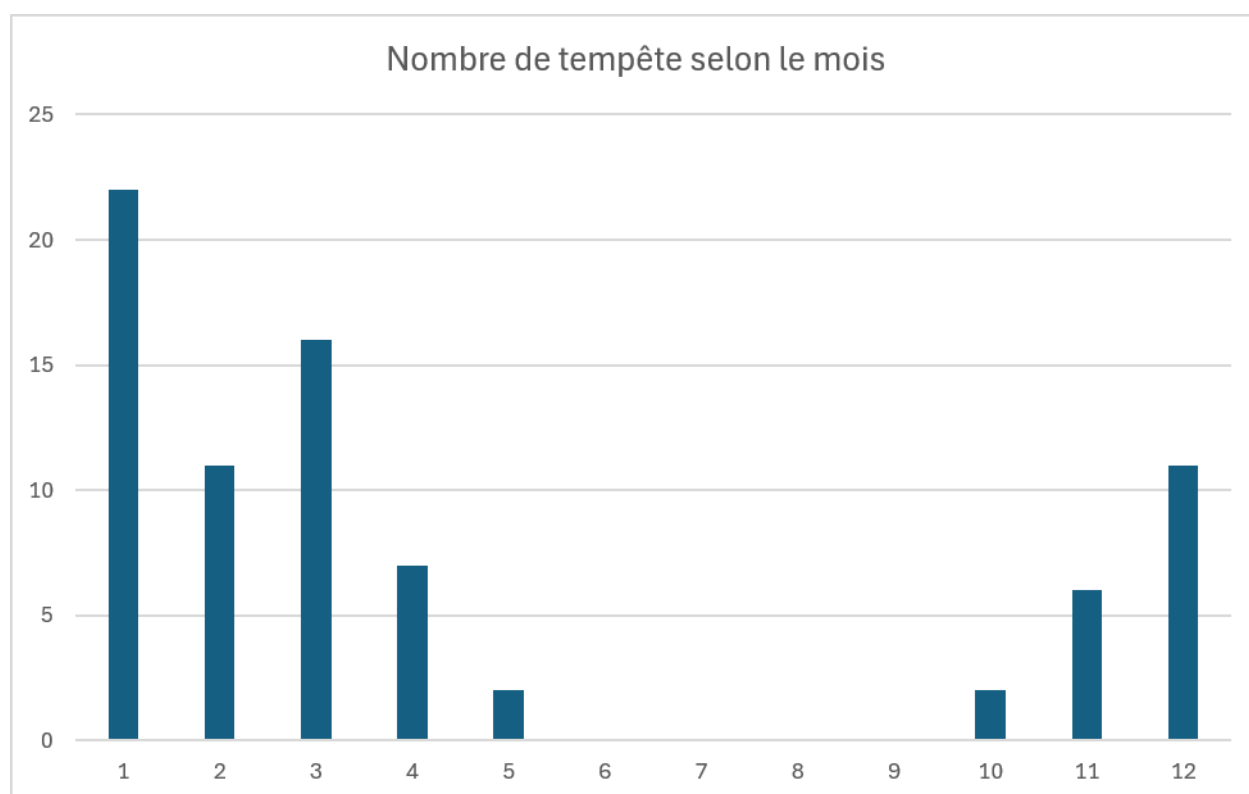
Seconde Partie

Traitement et exploitation des données

I. Fréquence des tempêtes

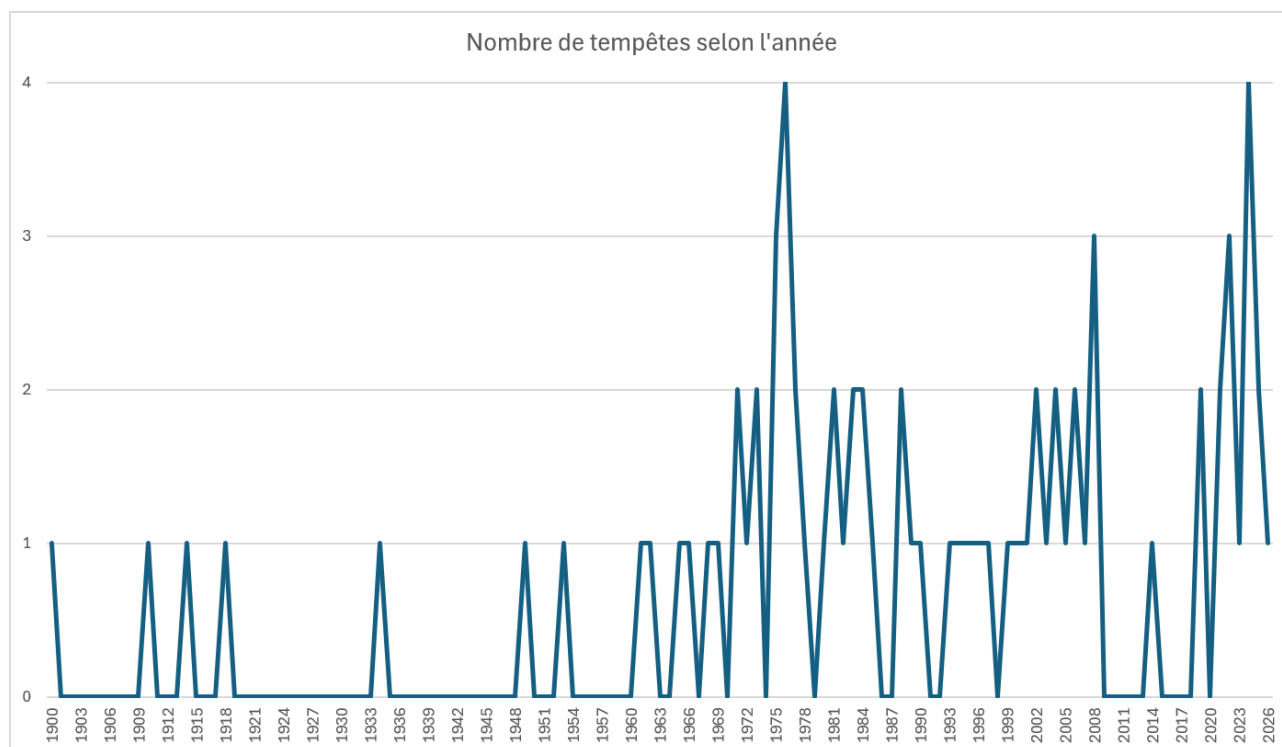
Avec la liste des tempêtes établie, il est possible de faire quelques statistiques sur la fréquence et la durée de retour des cyclones. Pour mémoire, notre liste comporte 77 tempêtes entre 1900 et 2026. Ici, il n'est pas question de vitesse de vents. Nous considérons tous les épisodes de façon égale. Il s'agit de voir à quelle période de l'année les tempêtes sont plus fréquentes. De même, nous voulons voir si leur durée de retour augmente ou diminue tout au long du siècle. Rappelons que la durée de retour d'un phénomène est le laps de temps moyen qui s'écoule entre deux occurrences du phénomène.

Tout d'abord, nous comptabilisons pour chaque mois le nombre de tempêtes qui se sont produites, quelle que soit l'année. Les résultats sont donnés dans le graphique ci-dessous :



Nous constatons une très forte disparité selon les mois. Aucune tempête ne se produit durant l'été, entre juin et septembre inclus. Les cyclones sont plus importants durant l'hiver et plus particulièrement durant les mois de décembre à mars avec un pic élevé en janvier. Cela correspond aux saisons cycloniques qui ont lieu durant l'été dans l'hémisphère sud.

Ensuite, pour chaque année, nous comptabilisons le nombre de cyclones. En 126 ans, 77 tempêtes se sont produites soit en moyenne une tous les 1.6 ans. Le graphique présentant le nombre de tempête annuel est donné ci-dessous :



Nous constatons qu'avant les années 1970, il n'y a qu'une tempête par an avec parfois des décennies sans aucun phénomène. Cela est dû au manque d'informations sur les tempêtes avant 1950. Seuls les cyclones les plus forts sont enregistrés et pas toutes les tempêtes, y compris le plus faibles.

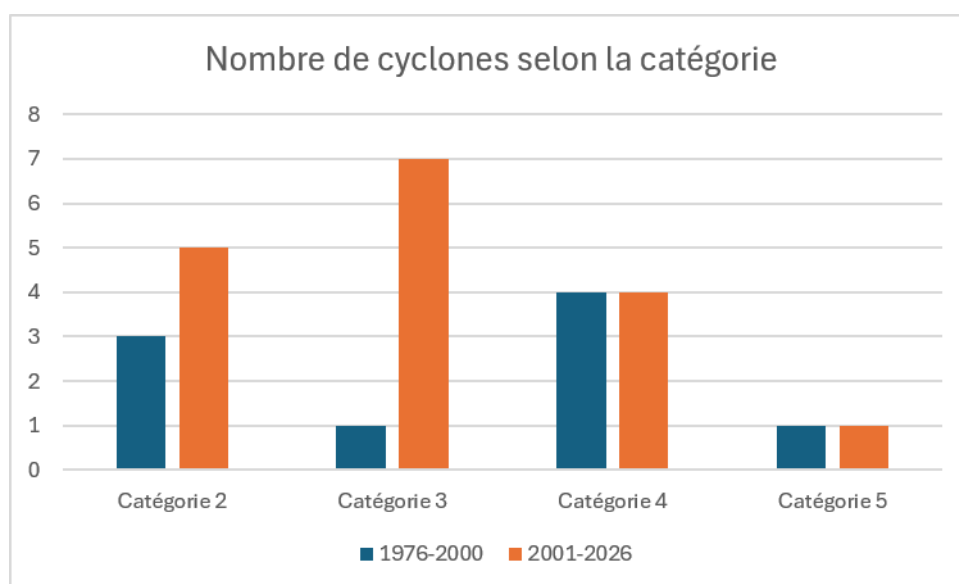
Après 1970, on retrouve des années sans cyclone, d'autres avec un seul cyclone enregistré et d'autres encore avec plusieurs tempêtes par an. Il n'y a pas d'évolution notable dans le nombre de tempêtes annuel. Même après 2010, on retrouve des années sans cyclone touchant Mayotte. La dernière année sans tempête était 2020. Les années le plus récentes ont connu de nombreux cyclones avec parfois 4 phénomènes par an.

Une autre question se pose : est-ce que la fréquence varie avec les années ? Plus précisément, est-ce que la fréquence d'apparition des tempêtes est la même maintenant qu'il y a 100 ans ? Pour répondre à cette question, nous fragmentons l'historique en 5 parties égales de 25 ans. Dans chacune de ces sections nous comptabilisons le nombre de tempêtes et la durée de retour. Les résultats sont donnés dans le tableau ci-dessous :

Période	Nombre de tempêtes	Durée de retour
1900-1925	4	6.3
1926-1950	2	12.5
1951-1975	15	1.7
1976-2000	27	0.9
2001-2026	29	0.9

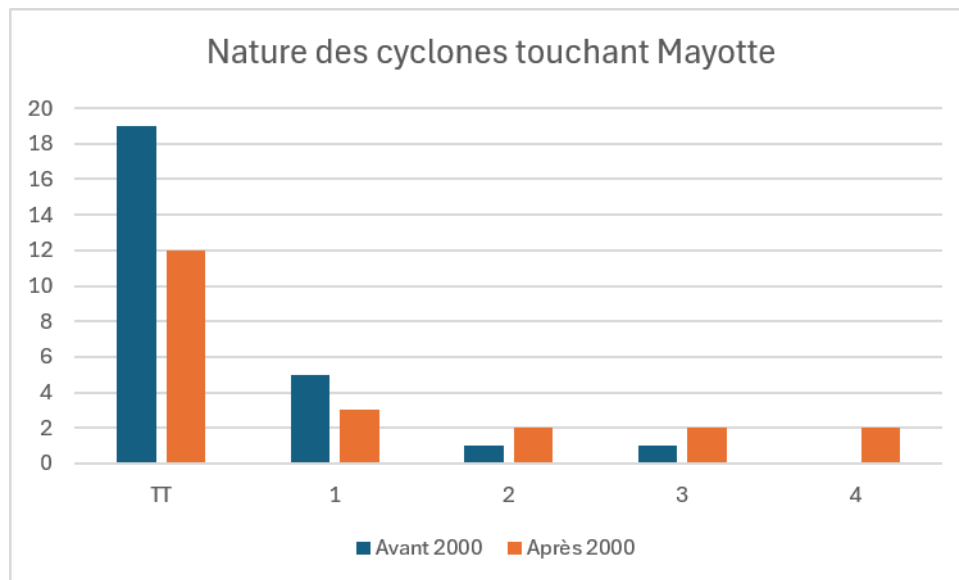
Apparemment, il n'y a pas beaucoup de tempêtes avant 1950. Mais ceci ne reflète pas la réalité. Avant 1950, il n'y a pas de relevés exhaustifs des tempêtes. Seuls les cyclones les plus importants, qui ont marqué la mémoire collective, sont enregistrés. A partir de 1950, les saisons cycloniques sont plus détaillées, et on a la trace de toutes les tempêtes tropicales, y compris les plus petites et celles qui n'ont pas touché terre. A partir de cette date, le nombre de tempêtes sur 25 ans augmente considérablement, parce que toutes les tempêtes sont enregistrées, y compris les plus petites. Entre 1950 et 1975, la durée de retour des cyclones est de 1.7 ans. Après 1975, cette durée se réduit à 0.9 pour les 50 années suivantes. Il n'y a pas d'augmentation du nombre de tempêtes à partir de l'an 2000. Cependant, elles sont plus nombreuses qu'en 1950.

Entre les périodes 1975-2000 et 2000-2026, le nombre de tempêtes n'a pas augmenté, mais l'intensité des cyclones a été plus importante. Entre 1975 et 2000, il y a eu 9 cyclones de catégorie 2 et + selon l'échelle de Saffir-Simpson, contre 17 entre 2000 et 2026. L'augmentation se concentre sur les cyclones de catégorie 2 et 3 qui sont plus importants ces 25 dernières années. En revanche, les cyclones les plus dévastateurs (catégorie 4 et 5) ne sont pas plus fréquents. Leur nombre est identique sur les deux périodes. Le graphique résultant est présenté ci-dessous :



Nous remarquons une intensification de la force des cyclones qui touchent Mayotte. Dans notre base de données, nous avons relevé la catégorie des tempêtes, lorsqu'elles étaient au plus proche des côtes de l'île. Avant l'année 2000, les tempêtes aux abords de Mayotte étaient alors de catégorie Tempête tropicale ou Dépression tropicale. Il n'y a eu que 2 cyclones qui étaient de catégories 2 et 3, alors qu'ils frappaient l'archipel. Il s'agissait des cyclones Andry et Kamizy en 1983 et 1984. Ce dernier était considéré comme un cyclone de référence jusqu'au passage d'Hellen.

Après 2000, les cyclones frappant Mayotte ont été plus intenses au plus proche de l'île. Il y a eu 2 cyclones de catégorie 3 et 2 autres de catégorie 4. Parmi ces derniers, nous citons Hellen en 2014, qui a provoqué des inondations et des glissements de terrain, mais aussi le cyclone Chido qui est considéré comme le plus intense depuis un siècle. Il a provoqué de très graves dommages à l'île, avec plus de 40 morts et une grande partie des habitations détruites. Pendant les dernières années, les tempêtes, lorsqu'elles frappent Mayotte, sont plus intenses. Le graphique correspondant est présenté ci-dessous :



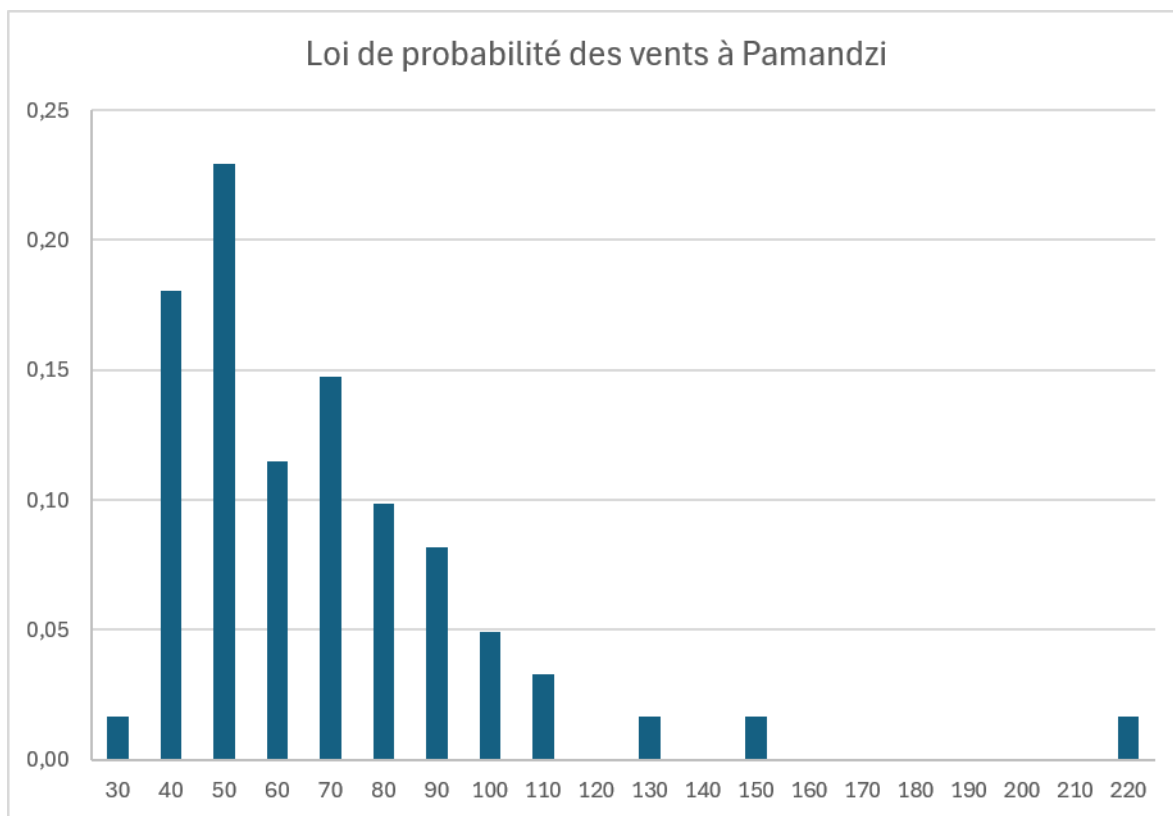
II. Lois de probabilité

A. Pamandzi

Après avoir relevé les vents à la station de Pamandzi, nous pouvons établir la loi de probabilité des vents. Nous disposons des données de vent à partir de 1981 jusqu'en 2026. Quelques corrections ont été apportées. Pour la tempête Chido, la station de Météo France indique des rafales maximales de 52.9 m/s, soit 187 km/h. Mais des données complémentaires indiquent davantage des rafales à 225 km/h. Il est possible que sous l'effet des bourrasques très importantes, l'anémomètre ait connu des dysfonctionnements. Dans notre base de données, nous avons corrigé la valeur à 226 km/h.

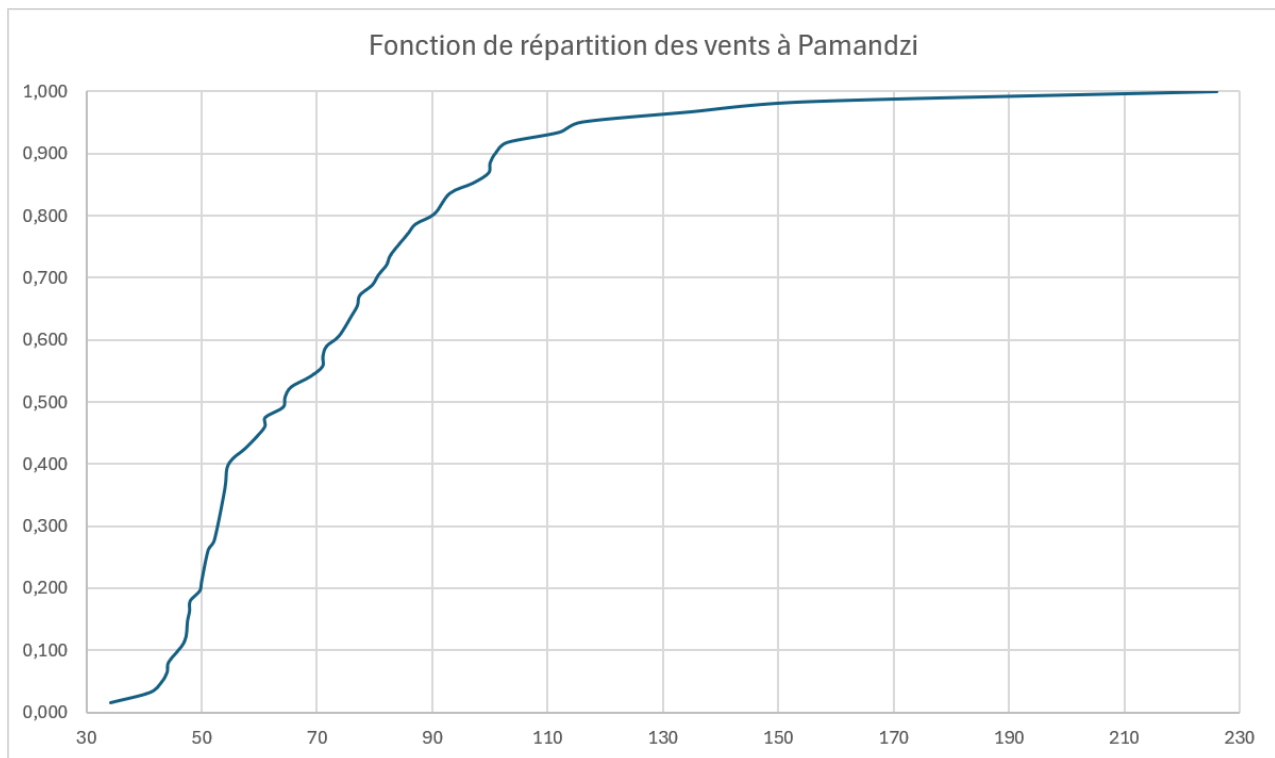
Pour certaines tempêtes antérieures à 1981, nous avons pu indiquer des vitesses des vents. Ces informations sont disponibles sur le site de l'Observatoire du Littoral de Mayotte, où les vitesses de vents maximales ont été estimées. Ces données ne sont pas très fiables, mais elles sont suffisantes pour avoir un ordre de grandeur et établir les lois de probabilité.

Nous avons réparti les vitesses de vents en classes de 10 km/h ; la première classe commençant à 30 et la dernière finissant à 230. Pour chacune des classes, nous avons comptabilisé le nombre de tempêtes correspondantes. Les résultats sont donnés dans le graphique ci-dessous :



Nous constatons que sur l'ensemble des tempêtes, la majorité ont des vents compris entre 40 et 80 km/h avec un pic à 50-60 km/h. La probabilité de vent augmente fortement à partir de 40 km/h avant décroître à partir de la classe [60-70]. Quelques vents sont possibles à 130, 150 et 220 km/h. Dans la majorité des cas, les cyclones et tempêtes tropicales aux alentours de Mayotte ne produisent que des vents de force intermédiaire, inférieurs à 80 km/h.

La fonction de répartition associée indique les probabilités de dépassement de seuils de vents. Par exemple la probabilité d'avoir des vents dépassant les 64 km/h est de 50%.

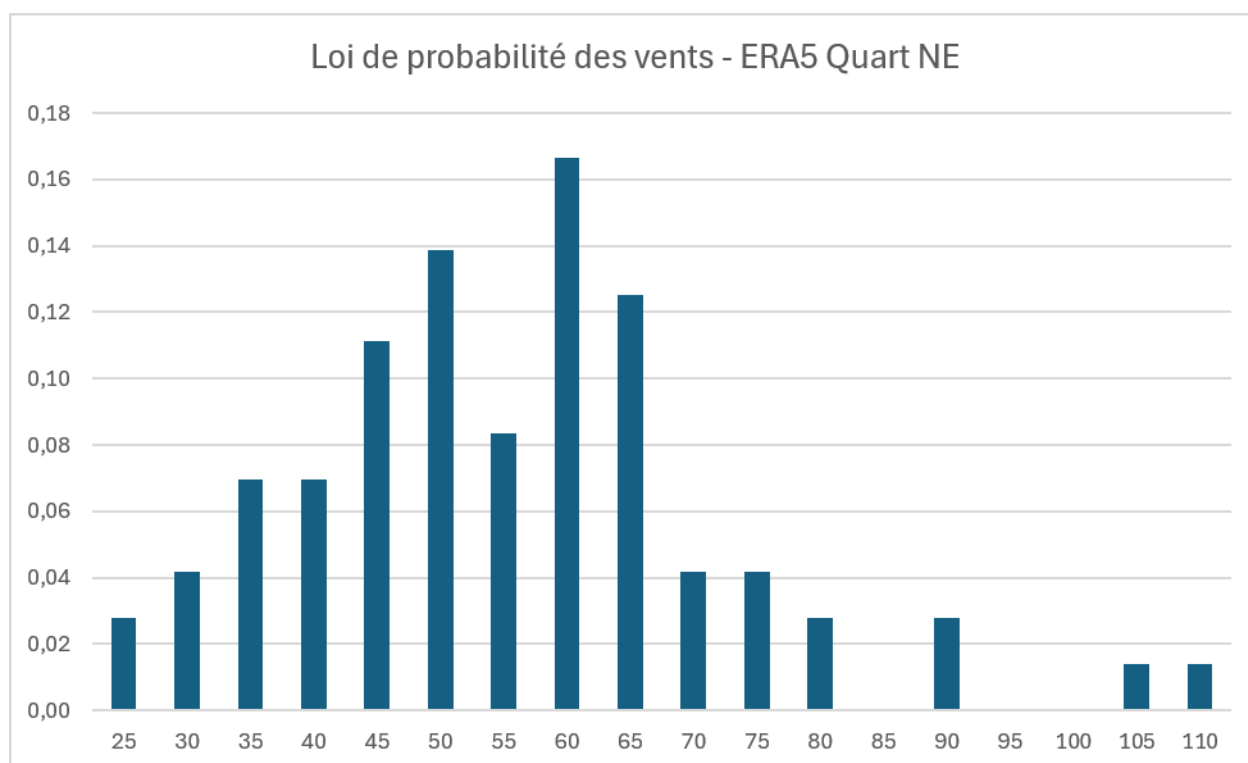


Voici maintenant les lois de probabilité sur données réelles obtenues à partir de l'autre source.

B. ERA5

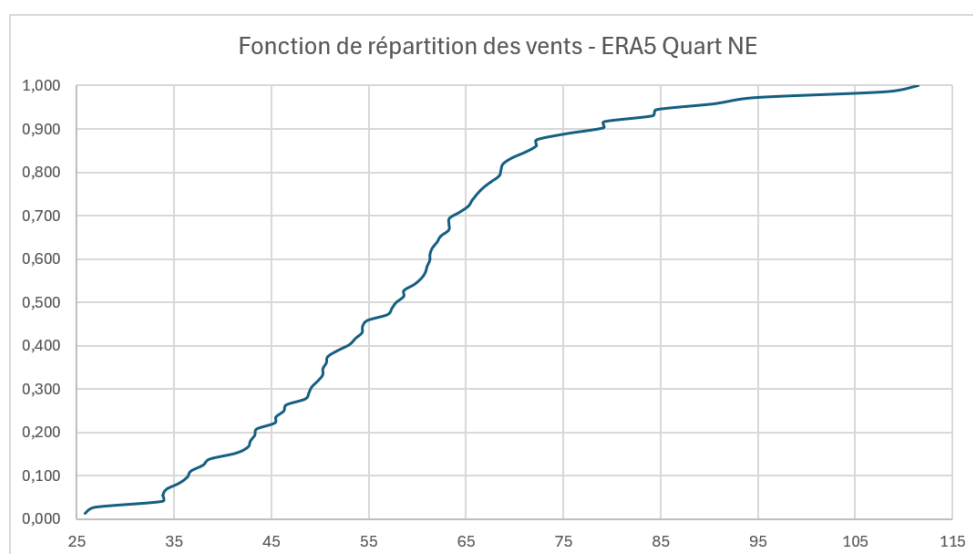
Comme expliqué précédemment, nous avons relevé quatre jeux de données provenant du site ERA5. Chaque jeu correspond à un nœud du quadrillage, chaque nœud formant un carré entourant l'île de Mayotte. Le nœud le plus proche de la station de Pamandzi est celui situé au nord-est. C'est ce jeu de données que nous allons étudier pour pouvoir le comparer aux données récoltées sur la station de Météo France.

Les données ont été réparties dans des classes de vents de 5 km/h. La loi de probabilité est présentée dans le graphique ci-dessous :



L'ordre de grandeur des vents n'est pas le même que pour les données de la station de Pamandzi. Ici le minimum des rafales ne dépasse pas les 110 km/h, alors que Météo France a enregistré des rafales dépassant les 220 km/h. La probabilité de vent augmente à partir de 20 km/h pour atteindre un pic entre 60 et 65 km/h. Cette probabilité diminue ensuite brusquement avant de se stabiliser à partir de 70 km/h. Quelques vitesses sont enfin possibles à 90, 105 et 110 km/h.

Nous avons également tracé la fonction de répartition des vents. La probabilité de dépasser les 57 km/h est de 50 %. La vitesse associée à cette probabilité est plus faible que celle évoquée pour la station de Pamandzi. Cela signifie que selon les données d'ERA5, les vents sont plus faibles à Mayotte que selon les données de Météo France.



III. Différence entre les jeux de données

Nous constatons une différence entre les deux jeux de données étudiés. La première est évidente : le maximum des vents enregistrés. Pour la station de Pamandzi, il est de 226 km/h avec 4 autres vents dépassant les 110 km/h. Cependant, dans le cas d'ERA5, le maximum des vents n'est que de 111 km/h.

Pour comparer les deux jeux de données, nous calculons le rapport entre le vent d'ERA5 et le vent mesuré par Météo France à Pamandzi. Les résultats montrent une grande disparité dans les rapports. En moyenne, les données d'ERA5 représentent 88% des valeurs de Météo France. Mais ce rapport de 0.88 n'est pas stable ; il va de 0.28 à 1.68. Pour certaines tempêtes, il y a de grandes différences entre les données de la station et celles du modèle d'ERA5. Si on ne regarde que les tempêtes les plus récentes, le rapport est plus stable. De façon générale, après les années 2000, les données d'ERA5 sont inférieures à celles de Météo France. Mais il reste toutefois une grande variabilité avec des rapports allant de 0.5 à 1. Les différences les plus importantes ont lieu pour les événements exceptionnels, à savoir les cyclones les plus importants. Par exemple, pour le cyclone Chido en 2024, la station de Pamandzi a enregistré un vent de 226 km/h (après correction) tandis qu'ERA5 indique des vents de 111 km/h. C'est aussi le cas pour la tempête Ernest en 2005 avec des vents à Pamandzi de 103 km/h contre 63 pour le site ERA5. Il est possible de généraliser en disant que plus les vents sont faibles, plus les données concordent entre Pamandzi et ERA5. Mais ce n'est pas une règle absolue et des exceptions existent dans les deux sens. Même pour des vents faibles, des différences sont possibles.

Nous avons aussi regardé les différences existantes entre les données de la station de Pamandzi et celle de Coconi, où nous avons récupéré quelques relevés. Pour cette dernière, les données ne sont disponibles qu'à partir de 2016 avec quelques valeurs manquantes. Sur les 10 rapports possibles, les données de Pamandzi sont 50% plus élevées que celles de Coconi. Il y a également de grandes disparités dans les rapports. Parfois, les données sont proches entre les deux stations, parfois elles sont éloignées. Cette différence peut s'expliquer par le positionnement géographique des stations. Celle de Pamandzi est située près de l'aéroport international de Mayotte sur Petite-Terre en bordure des côtes. La station de Coconi est située au centre de Grande-Terre, éloignée des côtes. Les tempêtes tropicales faiblissent en traversant les terres, n'étant plus alimentées par les eaux chaudes de l'océan. La distance par rapport aux côtes est toutefois trop faible pour que cela ait une grande différence dans les valeurs de vents.

Nous comparons aussi les 4 jeux de données issus d'ERA5 entre eux. Nous calculons les rapports entre le nœud NE avec les 3 autres nœuds. Dans l'ensemble, les données correspondent entre elles. Les valeurs des différents nœuds se ressemblent. Toutefois, les valeurs du nœud NE sont très légèrement supérieures à celles du nœud NO et égales à celles des nœuds situés au sud de l'île. Il n'y a pas de grande disparité dans les rapports, comme nous l'avons vu pour les paragraphes précédents.

Comme nous l'avons vu, il existe de grandes différences entre les données issues de Météo France et celles issues d'ERA5. Ceci s'explique par leur nature différente. Les données de la station de Pamandzi sont des relevés de températures quotidiennes.

Il n'y a rien de plus simple ; un thermomètre est placé dans un abri à l'abri du soleil et on relève régulièrement la température affichée sur le thermomètre. Il s'agit de mesures effectuées régulièrement. Ce sont des données brutes sans traitement, telles que la nature nous les fournit.

A l'inverse, les données d'ERA5 ne sont pas de simples relevés. Elles ne proviennent pas d'une station physique, comme c'est le cas pour celles de Météo France. Les données d'ERA5 sont issues d'un modèle qui reconstitue le climat à base de relevés et d'autres informations. Les données ne sont pas brutes, mais sont issues d'un traitement. Les données remontent aux années 40, mais les données de vent sont quasiment inexistantes avant 1980. De plus, les vents sont estimés selon les mailles d'un quadrillage. Les données sont fournies pour chaque nœud du quadrillage. Les données sont moyennées et il y a donc une tendance à baisser les valeurs de vents.

Tous ces faits expliquent la différence entre les deux jeux de données. Nous rappelons notre position de principe : retraiter les données au moyen de modèles mathématiques résulte en des modifications arbitraires et non validées. Nous recommandons l'utilisation des données brutes, quel que soit le sujet.

Troisième partie

Anticipation par simulation

I. Simulation de tempêtes

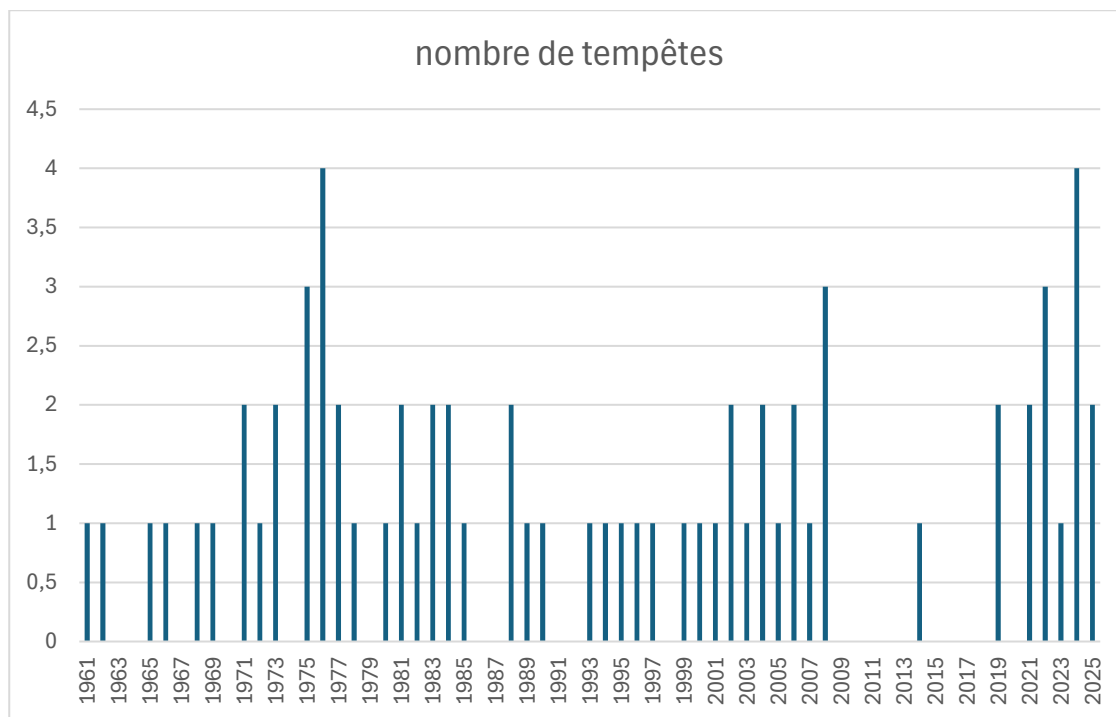
A partir d'un ensemble de données disponibles, nous montrons comment simuler les tempêtes possibles sur une durée de 100 ans.

Les données disponibles vont de 1961 à 2025, selon le tableau ci-dessous. Pour chaque tempête, on dispose du nom, de l'année et de la force maximale du vent, soit 69 enregistrements.

année	nb tempêtes		année	nb tempêtes
1961	1		1993	1
1962	1		1994	1
1965	1		1995	1
1966	1		1996	1
1968	1		1997	1
1969	1		1999	1
1971	2		2000	1
1972	1		2001	1
1973	2		2002	2
1975	3		2003	1
1976	4		2004	2
1977	2		2005	1
1978	1		2006	2
1980	1		2007	1
1981	2		2008	3
1982	1		2014	1
1983	2		2019	2
1984	2		2021	2
1985	1		2022	3
1988	2		2023	1
1989	1		2024	4
1990	1		2025	2

Le nombre de tempêtes, selon l'année, est donné par le tableau ci-contre.

Voici la présentation sous forme de graphe :



Nombre de tempêtes par année :

nb tempêtes	nb d'années	proba
0	21	0,32
1	26	0,40
2	13	0,20
3	3	0,05
4	2	0,03

Ce tableau s'interprète ainsi : pour une année donnée, il y a probabilité 0,32 de n'avoir aucune tempête, probabilité 0,40 d'en avoir une seule, etc.

Voici le détail, avec la force du vent :

Nom	Année	Vent max (km/h)		Nom	Année	Vent max (km/h)
Ada	1961	65		Gracia	1993	70
Daisy	1962	140		Nadia	1994	175
Claude	1965	95		Josta	1995	105
Colette	1966	100		Doloresse	1996	95
Georgette	1968	165		Josie	1997	140
Corrine	1969	85		Astride	1999	95
Agnes	1971	110		Hudah	2000	220
Felicie	1971	120		Dera	2001	150
Hermione	1972	120		Atang	2002	55
Bernadette	1973	150		Kesiny	2002	130
Charlotte	1973	75		Cela	2003	120
Elsa	1975	55		Gafilo	2004	230
Camille	1975	150		Elita	2004	120
Blandine	1975	175		Ernest	2005	165
Brigitta	1976	100		Bondo	2006	205
Gladys	1976	85		Anita	2006	65
Ella	1976	75		Jaya	2007	185
Clotilde	1976	250		Asma	2008	85
Hervea	1977	110		Jokwe	2008	195
Domitile	1977	65		Fame	2008	130
Irena	1978	95		Hellen	2014	230
Bettina	1980	85		Belna	2019	155
Benedicte	1981	135		Kenneth	2019	215
Iadine	1981	95		Jobo	2021	120
Justine	1982	150		Eloise	2021	150
Andry	1983	165		Jasmine	2022	110
Elinah	1983	115		Gombe	2022	155
Anety	1984	65		Ana	2022	95
Kamisy	1984	170		Cheneso	2023	140
Feliska	1985	95		Chido	2024	215
Filao	1988	130		Hidaya	2024	150
Doaza	1988	185		Gamane	2024	140
Calasanjy	1989	135		Filipo	2024	100
Hanta	1990	65		Jude	2025	140
				Dikieledi	2025	175

Voici les données détaillées, comportant la force des vents :

année	nb tempêtes	vent1	vent2	vent3	vent4		année	nb tempêtes	vent1	vent2	vent3	vent4
1961	1	65					1992	0				
1962	1	140					1993	1	70			
1963	0						1994	1	175			
1964	0						1995	1	105			
1965	1	95					1996	1	95			
1966	1	100					1997	1	140			
1967	0						1998	0				
1968	1	165					1999	1	95			
1969	1	85					2000	1	220			
1970	0						2001	1	150			
1971	2	110	120				2002	2	55	130		
1972	1	120					2003	1	120			
1973	2	150	75				2004	2	230	120		
1974	0						2005	1	165			
1975	3	55	150	175			2006	2	205	65		
1976	4	100	85	75	250		2007	1	185			
1977	2	110	65				2008	3	85	195	130	
1978	1	95					2009	0				
1979	0						2010	0				
1980	1	85					2011	0				
1981	2	135	95				2012	0				
1982	1	150					2013	0				
1983	2	165	115				2014	1	230			
1984	2	65	170				2015	0				
1985	1	95					2016	0				
1986	0						2017	0				
1987	0						2018	0				
1988	2	130	185				2019	2	155	215		
1989	1	135					2020	0				
1990	1	65					2021	2	120	150		
1991	0						2022	3	110	155	95	
							2023	1	140			
							2024	4	215	150	140	100
							2025	2	140	175		

II. Réalisation de la simulation

Pour réaliser le simulateur, on commence par mettre les données disponibles sous la forme d'un tableau, où apparaissent selon les années les différentes forces de vent : c'est le tableau ci-dessus. En première colonne, l'année, en colonne 2 le nombre de tempêtes cette année-là et en colonnes 3, 4,... les forces de vent enregistrées, en km/h.

Ensuite, pour chacune des années 2031 à 2130, on tire au hasard, selon une loi uniforme, une ligne dans le tableau ci-dessus et on attribue cette ligne à cette année. L'idée est qu'une année quelconque, dans les cent années suivantes, sera identique à l'une des années enregistrées (mais on ne sait pas laquelle).

Voici la simulation réalisée ; la loi de probabilité est la même que pour les données initiales : proportion d'années sans tempêtes, avec une seule tempête, avec deux tempêtes, etc.

année	nb de tempêtes	vent1	vent2	vent3	vent4	vent>150 km/h	année	nb de tempêtes	vent1	vent2	vent3	vent4	vent>150 km/h
2031	3	85	195	130		***	2080	0					
2032	0						2081	0					
2033	0						2082	0					
2034	2	135	95				2083	1	135				
2035	1	95					2084	4	215	150	140	100	***
2036	0						2085	1	230				***
2037	2	65	170			***	2086	2	135	95			
2038	4	215	150	140	100	***	2087	2	130	185			***
2039	1	95					2088	2	110	65			
2040	2	165	115			***	2089	1	95				
2041	0						2090	1	230				***
2042	3	85	195	130		***	2091	0					
2043	2	120	150			***	2092	1	120				
2044	3	85	195	130		***	2093	0					
2045	2	165	115			***	2094	1	150				***
2046	0						2095	1	85				
2047	0						2096	1	95				
2048	2	110	65				2097	2	130	185			***
2049	2	110	65				2098	0					
2050	0						2099	0					
2051	3	110	155	95		***	2100	3	55	150	175		***
2052	0						2101	0					
2053	1	185				***	2102	1	95				
2054	0						2103	2	55	130			
2055	2	55	130				2104	0					
2056	0						2105	4	215	150	140	100	***
2057	0						2106	0					
2058	1	150				***	2107	0					
2059	1	85					2108	1	165				***
2060	1	95					2109	1	140				
2061	4	215	150	140	100	***	2110	1	100				
2062	0						2111	1	220				***
2063	0						2112	3	85	195	130		***
2064	1	95					2113	0					
2065	1	100					2114	1	95				
2066	1	65					2115	0					
2067	0						2116	0					
2068	1	95					2117	0					
2069	1	185				***	2118	1	140				
2070	2	150	75			***	2119	0					
2071	0						2120	2	230	120			***
2072	2	205	65			***	2121	0					
2073	0						2122	0					
2074	0						2123	1	175				***
2075	1	120					2124	2	155	215			***
2076	1	95					2125	0					
2077	1	140					2126	0					
2078	0						2127	2	150	75			***
2079	1	175				***	2128	2	65	170			***
							2129	0					
							2130	1	85				

III. Durée de retour des vents violents

Sur le tableau simulé, on met en dernière colonne "***" si cette année-là le vent dépasse un seuil donné, ici 150 km/h. Ceci se produit 32 fois, la première étant 2031 :

2031	3	85	195	130	***
------	---	----	-----	-----	-----

et la dernière 2128 :

2128	2	65	170	***
------	---	----	-----	-----

La durée de retour (intervalle de temps moyen) est ainsi $(2128-2031)/31=3,13$ années.

Cela signifie concrètement que, sur une durée de 100 ans, on peut s'attendre en moyenne à rencontrer, une fois tous les 3 ans, des vents dépassant 150 km/h. Ce type de résultat est analogue à ce que nous avons obtenu pour la ville de Villiers-le-Bâcle (Essonne) ; il doit aider les décideurs politiques à faire des choix d'investissement : quels sont les bâtiments ou équipements qui peuvent être détruits par des vents supérieurs à 150 km/h et dans quelle mesure ces équipements sont-ils critiques ?

IV. Programmation

Voici le programme en VBA :

```
Sub macro1() 'traitement initial des données
Dim k As Integer
Dim histo(1961 To 2025) As Integer
Dim y As Integer
For k = 2 To 70
y = Sheets(1).Cells(k, 2)
histo(y) = histo(y) + 1
Next k
For k = 1961 To 2025
Sheets(2).Cells(k - 1959, 1) = k
Sheets(2).Cells(k - 1959, 2) = histo(k)
Next k
Dim histo_vent(1961 To 2025) As Variant
For k = 2 To 70
y = Sheets(1).Cells(k, 2)
histo_vent(y) = histo_vent(y) & "," & Sheets(1).Cells(k, 3)
Next k

For k = 1961 To 2025
Sheets(2).Cells(k - 1959, 4) = histo_vent(k)
Next k

End Sub

Sub macro2() 'simulation
Dim dL As Integer
dL = Sheets(2).Cells(Rows.Count, 1).End(xlUp).Row
Dim TEV As Long 'taille échantillon voulu
TEV = 100 'lecture à partir de l'indication donnée par l'utilisateur, qui met ce qu'il veut
Dim k As Long
Randomize
Dim n As Long
Dim data As Variant 'mise en mémoire de l'échantillon recueilli
data = Sheets(2).Cells(2, 2).Resize(dL, 5)
For k = 1 To TEV
n = Int((dL - 1) * Rnd()) + 1 'choix d'une ligne selon une loi uniforme
Sheets(2).Cells(k + 1, 9) = 2030 + k
Sheets(2).Cells(k + 1, 10) = data(n, 1)
Sheets(2).Cells(k + 1, 11) = data(n, 2)
Sheets(2).Cells(k + 1, 12) = data(n, 3)
Sheets(2).Cells(k + 1, 13) = data(n, 4)
Next k
End Sub
```

```

Sub macro3()
Dim k As Integer
Dim histo(0 To 4) As Integer
Dim n As Integer
For k = 2 To 66
n = Sheets(2).Cells(k, 2)
histo(n) = histo(n) + 1
For n = 0 To 4
Sheets(2).Cells(70 + n, 2) = histo(n)
Sheets(2).Cells(70 + n, 1) = n
Next n
Next k
End Sub

```

```

Sub macro4()
'durées de retour
Dim k As Integer
Dim seuil As Integer
seuil = 150
Dim count As Integer
For k = 2 To 101
If Sheets(2).Cells(k, 11) >= seuil Or Sheets(2).Cells(k, 12) >= seuil Or Sheets(2).Cells(k, 13) >=
seuil Or Sheets(2).Cells(k, 14) >= seuil Then
Sheets(2).Cells(k, 15) = "****"
count = count + 1
End If
Next k
MsgBox count
End Sub

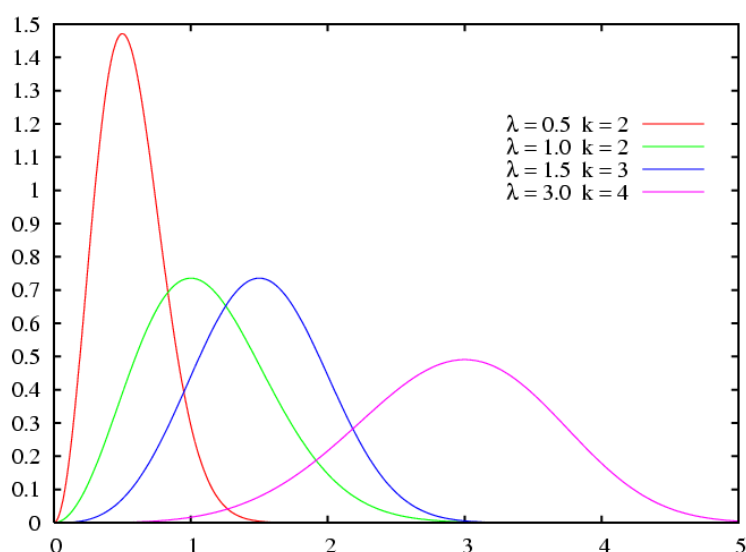
```

Quatrième Partie

Introduction de lois académiques : lois de Weibull et de Gumbel

A la demande de la CCR, nous répondons maintenant à la question : pour décrire les lois de probabilité des tempêtes à Mayotte, peut-on utiliser les lois courantes, notamment les lois de Weibull ou de Gumbel ?

A. *Eléments historiques*



Ces lois sont traditionnellement utilisées pour la description de phénomènes extrêmes, parce qu'elles ne font intervenir que peu de paramètres : c'est commode car, pour les phénomènes extrêmes, on dispose de peu de relevés. Voici, ci-contre, la forme de ces lois (source : Wikipedia). Elles ont la particularité, par rapport à la loi de Gauss, d'être dissymétriques.

Les lois de type Weibull, Gumbel, etc., ont été introduites dans les années 1970, époque où les données expérimentales étaient encore insuffisantes et où il fallait une formule pour pouvoir travailler. A cette époque, les experts se sont mis d'accord sur l'utilisation de lois spécifiques (citons encore la loi exponentielle pour la durée de vie des circuits électriques) : c'était une démarche normale pour l'époque.

Mais, plus de 50 ans après, de nombreuses données expérimentales ont été recueillies et on s'aperçoit que ces lois académiques ne décrivent pas correctement la réalité. Elles restent cependant largement utilisées, dans la mesure où l'habitude a été prise.

Insistons sur ceci : la justification même de ces lois a disparu, du fait des progrès de l'informatique. Sur des machines peu puissantes, il était commode de traiter une formule mathématique ; sur les machines actuelles, on peut sans difficulté traiter des millions de données, sans avoir besoin de chercher à les approximer par une formule.

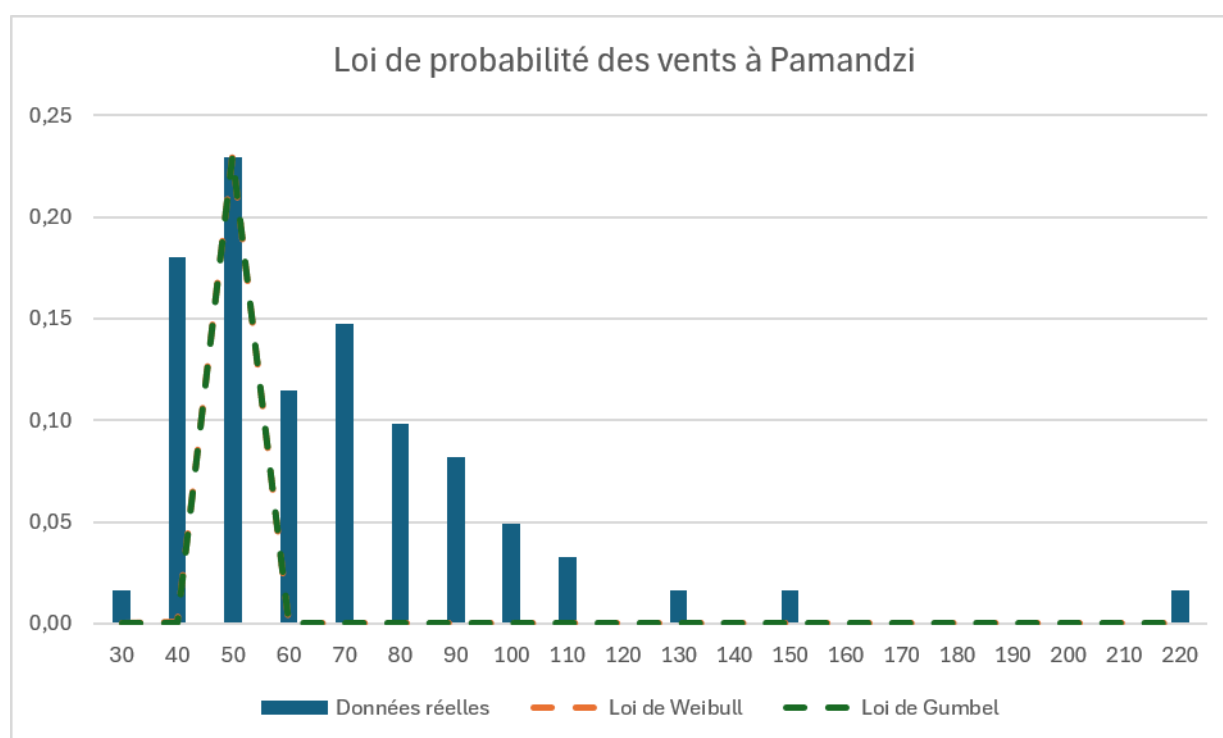
La loi de Weibull (voir par exemple https://fr.wikipedia.org/wiki/Loi_de_Weibull) est une loi dépendant de deux paramètres seulement, dont la densité est donnée par la formule :

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-\left(\frac{x}{\lambda} \right)^k},$$

où $k > 0$ s'appelle paramètre de forme et $\lambda > 0$ s'appelle paramètre d'échelle.

Comme on le voit sur le graphique ci-dessus, ces lois sont extrêmement régulières et ne rendent pas compte de l'irrégularité des données réelles.

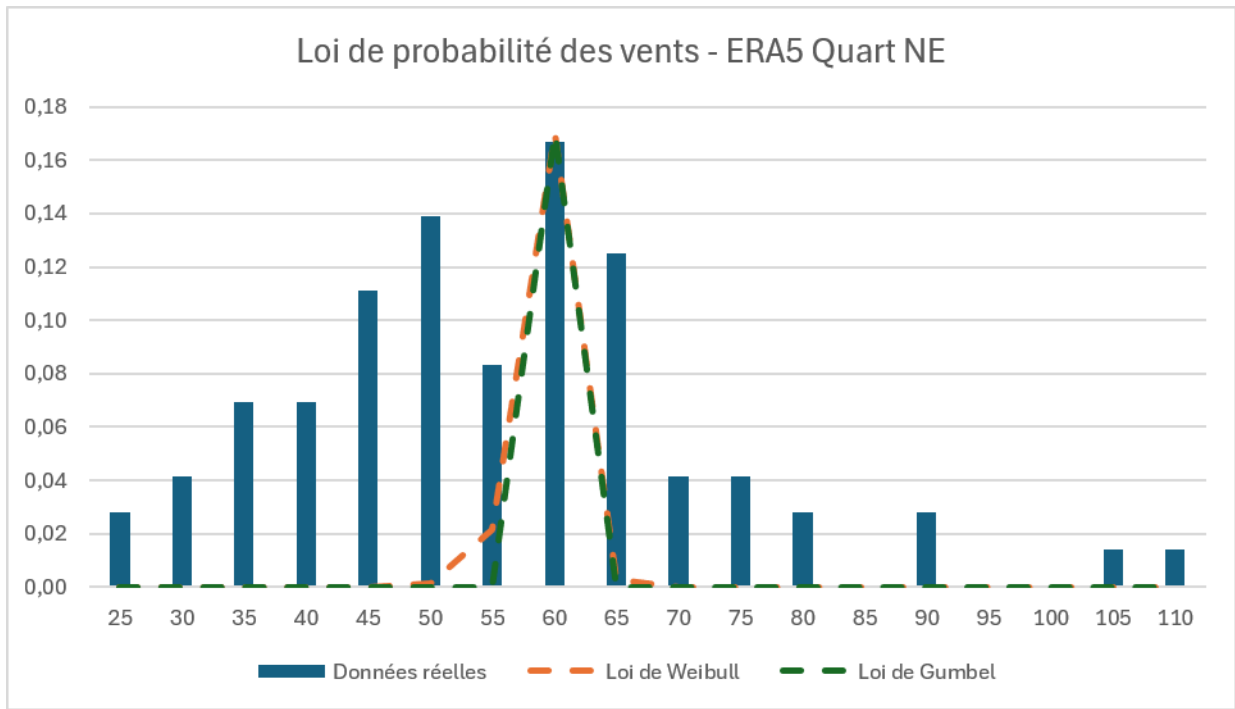
Pour bien illustrer ceci, nous traçons les lois de Gumbel et de Weibull sur le même graphique que notre loi de probabilité des vents à Pamandzi.



Les lois de Weibull et Gumbel sont confondues sur le graphique. L'ajustement, ici, est fait de telle manière que le maximum de la loi corresponde à la probabilité des vents la plus forte : le pic maximal est commun, mais il n'en est pas du tout de même pour les autres pics.

Nous avons essayé d'autres réglages des paramètres pour la loi de Weibull, mais aucun ne donne satisfaction : la courbe de Weibull est trop régulière pour représenter les données.

Faisons de même pour la loi de probabilité des vents issue d'ERA5.



Les lois de Weibull et Gumbel sont confondues sur le graphique. L'ajustement est réalisé pour que le pic de plus forte intensité coïncide pour toutes les lois, mais elles diffèrent évidemment pour les autres pics.

B. Difficultés d'ajustement

Comme il a été dit plus haut, ces lois ne dépendent que de deux paramètres, ce qui est très commode lorsque les valeurs expérimentales sont peu nombreuses. Mais alors, la question se pose : comment calculer les deux paramètres pour ajuster au mieux aux données recueillies ? C'est une question fondamentale, qui n'admet aucune réponse satisfaisante : ceci ne fait qu'accentuer les réserves que nous formulons quant à l'utilisation de ces lois. En effet, selon les auteurs, les ajustements seront faits de manière différente, apportant des résultats fondamentalement différents, sans que personne puisse décider quelle est la meilleure solution. Il en résultera, par exemple pour évaluer un dommage, une inévitable bataille d'experts.

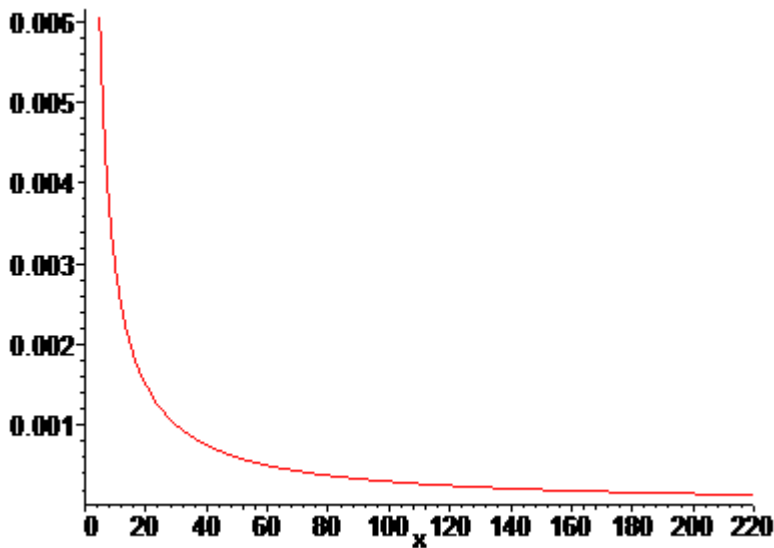
Plus haut, nous avons utilisé l'ajustement sur le pic de plus forte probabilité : c'est une possibilité que le solveur d'Excel accepte bien. En théorie, il faudrait ajuster les paramètres au sens des moindres carrés : si on a observé les valeurs x_i avec les probabilités p_i , il faut choisir les paramètres k, λ pour que :

$$S = \sum_{i=1}^N (f_{k,\lambda}(x_i) - p_i)^2 \text{ soit minimal,}$$

avec :

$$f_{k,\lambda}(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-\left(\frac{x}{\lambda} \right)^k}$$

La recherche des valeurs de k, λ se fait par approximation. Dans le cas de Padmandzi, on prendra $\lambda = 220$, valeur maximale de la force des vents, pour que le facteur d'échelle couvre toute la gamme des vents possibles. La recherche numérique du minimum en k , pour la somme S définie plus haut, conduit à la valeur $k = 0.08251$ et la fonction de Weibull a alors la forme suivante :



Cette forme ne coïncide pas avec la forme de la densité de probabilité expérimentale, vue plus haut, ni avec la forme obtenue en ajustant sur le pic de plus forte probabilité.

Pour clore ce paragraphe, on peut dire ceci, qui relève du bon sens : il est impossible, en n'utilisant que deux paramètres, de décrire correctement une loi de probabilité qui porte sur plus de vingt mesures.

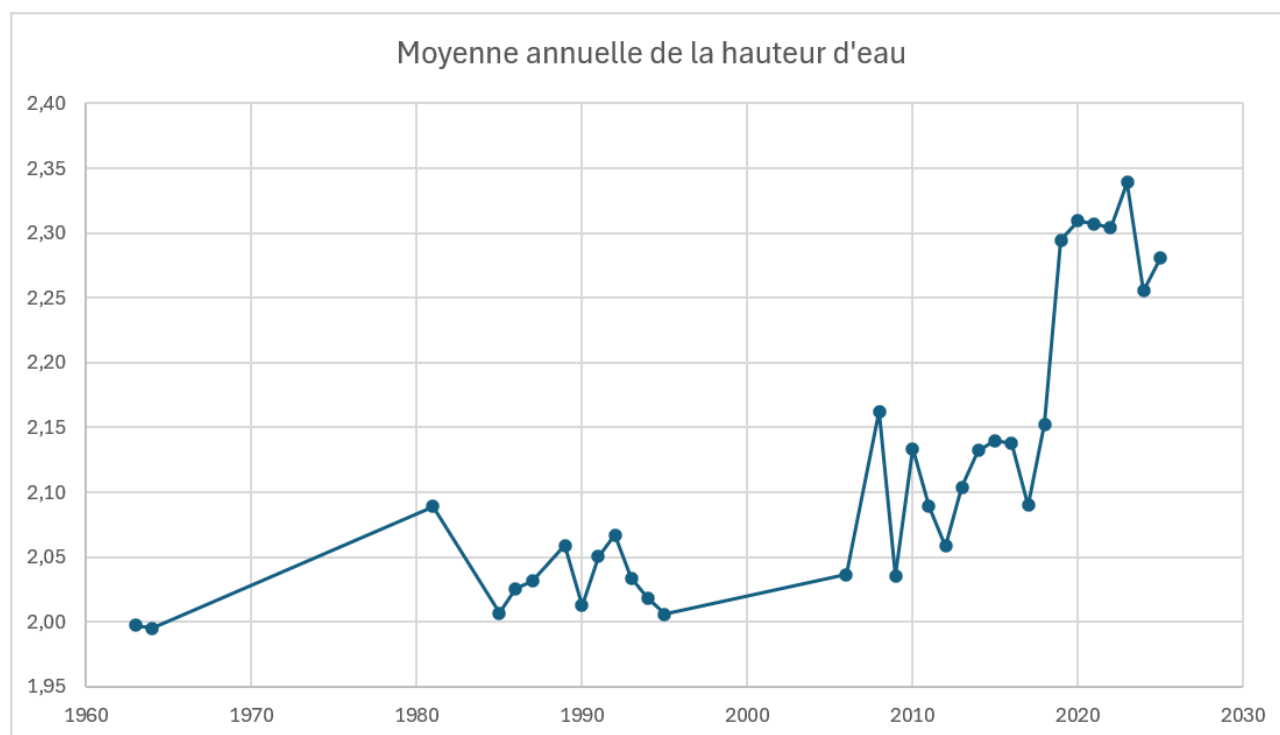
Cinquième Partie

La hauteur des marées

En complément de l'étude principale, portant sur les tempêtes, nous nous sommes intéressés, comme convenu avec la CCR, aux données décrivant les hauteurs de marées. Elles sont issues du SHOM, le service hydrologique et océanographique de la Marine.

Il y a une seule station à Mayotte. Les données, représentant des hauteurs d'eau, sont horaires et exprimées en mètres. Les premières mesures commencent en 1963 et s'achèvent en 2025. Des trous importants existent dans les données : entre 1964 et 1981, entre 1981 et 1985 et entre 1995 et 2006.

Pour chaque année, nous calculons la moyenne annuelle de la hauteur d'eau. Les résultats sont donnés dans le graphique ci-dessous :



Nous constatons une évolution de la hauteur d'eau annuelle. En 1960, elle était de 2 m tandis qu'entre 1985 et 1995, elle fluctue entre 2 m et 2,05 m. A partir de 2006, la hauteur d'eau à Mayotte augmente entre 2,05 et 2,15 m. En 2018, elle augmente brusquement pour atteindre 2,3 m. Depuis cette date, la hauteur d'eau à Mayotte fluctue entre 2,25 m et 2,35 m.

On constate le même phénomène pour les hauteurs d'eau à Brest (source SHOM). C'est dû au fait que la pression atmosphérique est très variable d'une décennie à l'autre.

Annexe : méthodes mathématiques pour la génération d'un scénario

(Document publié par la SCM en novembre 2025)

Résumé

Générer un échantillon aléatoire à partir d'une loi de probabilité donnée est une préoccupation très ancienne ; elle se rencontre en particulier auprès de toutes les institutions qui veulent réaliser des simulations ; voici quelques exemples :

- Générer un catalogue de pannes, pour un équipement donné, à partir d'un historique des pannes ;
- Générer un catalogue de cyclones, de séismes, pour une région donnée, à partir d'un historique ;
- Générer un modèle de propagation d'une épidémie, à partir d'un historique.

Bien évidemment, ces simulations ne sont pas des certitudes ; elles n'ont pas de valeur juridique, mais servent à éclairer la décision : ce sont des scénarios. Habituellement, on cherche à se doter de plusieurs scénarios, pour mieux éclairer la question.

Nous montrons ici comment générer un échantillon de taille quelconque, à partir d'un échantillon recueilli, en respectant la loi de l'échantillon recueilli. Nous donnons un exemple concret d'application : à partir de 170 années de températures journalières recueillies, générer un scénario sur 100 000 jours (soit environ 274 ans) en faisant apparaître explicitement les dates où la température dépassera 39°C.

Mais des mises en garde s'imposent : si, dans l'échantillon d'origine, le capteur n'a pas fonctionné la moitié du temps, la valeur 0 aura probabilité 1/2 dans l'échantillon généré. S'il s'agit de données de température, le fait d'avoir plusieurs jours de suite avec canicule sera perdu. En bref, la génération aléatoire suppose que les données sont indépendantes ; si ce n'est pas le cas, danger.

I. Position du problème

Générer un échantillon à partir d'une loi de probabilité donnée a fait l'objet d'une littérature très abondante. Le meilleur document est le livre :

Luc Devroye : Non Uniform Random Variate Generation,
qui peut être téléchargé librement :

https://luc.devroye.org/LucDevroye-NonUniformRandomVariateGeneration-10.1007_978-1-4613-8643-8-1986-.pdf

Si la loi de probabilité est donnée par une densité continue $f(x)$, avec fonction de répartition $F(x)$, on génère un échantillon selon cette loi de manière très simple :

- On génère un nombre aléatoire y , selon une loi uniforme sur $[0,1]$: c'est l'expression $y = \text{rnd}()$ en VBA.
- On cherche le x tel que $F(x) = y$.

Il faut donc inverser la fonction F ; c'est difficile si elle est continue ; encore plus difficile si la fonction f est une loi discrète, car alors F est constante par paliers. On est donc amené à toutes sortes de "bricolages" : soit approximer F par une fonction inversible connue, soit tester des valeurs de x jusqu'à ce qu'on en trouve une telle que $F(x) = y$. La littérature sur ces approches est abondante et peu satisfaisante en pratique : on s'égare aux rives du Gange, comme aurait dit Chateaubriand.

En pratique, précisément, ni la fonction f ni la fonction F ne sont connues : nous ne disposons que d'un échantillon de mesures. Vouloir en déduire une loi de probabilité ne peut se faire sans hypothèses, souvent largement factices. Les économistes, les spécialistes des risques au sein des banques et des compagnies d'assurance, le constatent souvent à leurs dépens, bien entendu sans jamais admettre une mauvaise approche du sujet.

En pratique, donc, il ne faut pas vouloir passer par l'intermédiaire d'une loi de probabilité, il faut générer un échantillon aléatoire à partir de l'échantillon recueilli, en conservant la même loi. Si, dans l'échantillon recueilli, les valeurs sont $x_1 < x_2 < \dots < x_K$ avec chacune un nombre d'occurrences $n_1, \dots, n_K$, on veut que, dans l'échantillon généré, on retrouve les mêmes valeurs,

avec la même probabilité, à savoir $p_k = \frac{n_k}{n_1 + \dots + n_K}$.

Ceci se fait très simplement en choisissant au hasard, selon une loi uniforme, un nombre quelconque dans l'échantillon recueilli. Celui-ci doit être laissé tel qu'il est, tel que la Nature nous le fournit. Concrètement, on range les valeurs recueillies par ordre croissant et on les range dans des cases, en les répétant selon leur multiplicité. Par exemple, si on a un échantillon de 7 valeurs : $x_1 < x_2 = x_2 < x_3 = x_3 < x_4$, on définira 7 cases :

x1	x2	x2	x3	x3	x3	x4
----	----	----	----	----	----	----

et on choisira une case au hasard selon une loi uniforme.

Le programme ci-dessous, en VBA, explique ceci très clairement.

II. Programmation informatique

Voici le programme en VBA sous Excel :

```
Option Explicit
Sub macro1()
Dim DL As Long
DL = Sheets(2).Cells(Rows.Count, 1).End(xlUp).Row
'la dernière cellule occupée par l'échantillon recueilli
'(mis sur la seconde feuille, colonne 1)
Sheets(1).Cells(11, 4) = DL-1
'la taille de l'échantillon recueilli
Dim TEV As Long 'taille échantillon voulu
TEV = Sheets(1).Cells(5, 4)
'lecture à partir de l'indication donnée par l'utilisateur, qui met ce qu'il veut
Dim k As Long
Randomize
Dim n As Long
Dim data As Variant 'mise en mémoire de l'échantillon recueilli
'cela va beaucoup plus vite que l'accès aux cellules
data = Sheets(2).Cells(2, 1).Resize(DL, 1)
For k = 1 To TEV
n = Int((DL - 1) * Rnd()) + 1
Sheets(2).Cells(k + 1, 3) = data(n, 1)
Next k
Dim FF As Long
FF = Sheets(2).Cells(Rows.Count, 3).End(xlUp).Row
Sheets(1).Cells(12, 4) = FF - 1
End Sub
```

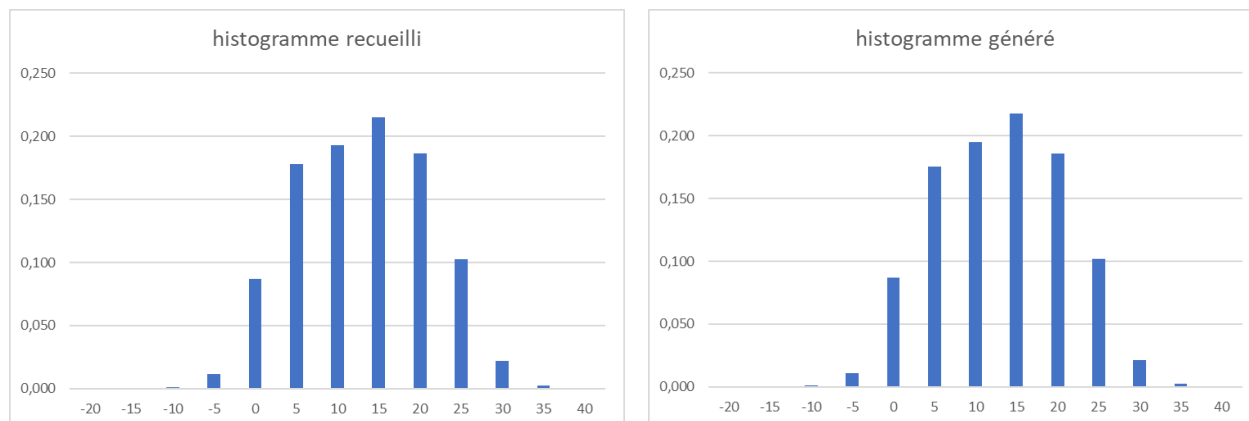
Revenons sur l'instruction $n = \text{Int}((DL - 1) * \text{Rnd}()) + 1$ qui est l'élément central du programme.

DL-1 est la taille de l'échantillon recueilli. Rnd() prend des valeurs arbitraires, selon une loi uniforme, entre 0 et 1. Par conséquent, $(DL - 1) * \text{Rnd}()$ prend des valeurs arbitraires, selon une loi uniforme, entre 0 et DL-1 : mais ce sont des valeurs réelles, et non des entiers.

L'instruction Int(x) renvoie le plus grand entier inférieur ou égal à x. Par conséquent, $\text{Int}((DL - 1) * \text{Rnd}())$ prend des valeurs arbitraires, entières, avec loi uniforme, entre 0 et DL-1. Enfin, $\text{Int}((DL - 1) * \text{Rnd}()) + 1$ prend des valeurs arbitraires, entières, avec loi uniforme, entre 1 et DL : c'est ce que nous voulons. Vérifions : si rnd()=0, on trouve 1, si rnd()=1, on trouve DL.

III. Analyse d'un exemple

On peut évidemment construire les histogrammes, pour l'échantillon recueilli et l'échantillon généré. Voici un exemple ; il concerne des températures (en °C) ; on a recueilli 36 214 données et on en a généré 100 000, suivant la même loi.



Les deux ont bien même apparence.

L'échantillon généré possède les mêmes propriétés statistiques que l'échantillon recueilli, mais les nombres ne sont pas dans le même ordre. Voici un exemple (même données que précédemment) :

échantillon recueilli	échantillon généré
17,0	12,50
22,6	16,70
24,6	21,10
24,6	10,40
24,6	12,20
27,2	23,40
26,3	6,13
27,2	18,19
26,3	17,90

IV. Un exemple de scénario

Nous allons voir sur un exemple comment réaliser un scénario : nous partons des données précédentes. Ce sont des données journalières de température, au nombre de 36 214, soit environ 137 ans. On veut construire un scénario de températures extrêmes sur 100 000 jours, soit environ 274 ans. On s'intéresse au retour de températures très élevées, mettons 39°C. Le scénario généré donne la liste suivante, pour les jours où cette température sera atteinte ou dépassée :

3838, 7342, 9878, 13568, 20992, 39618, 43538, 49652, 52767, 65555, 72271, 72569, 80844, 82148, 96399.

La première occurrence se manifestera dans 3 838 jours, soit approximativement 10 ans et 188 jours, la seconde dans 7 342 jours, soit approximativement 20 ans et 42 jours, etc.

V. Mise en garde

Il faut prendre garde à la nature des données générées. L'outil décrit ici traite les données comme si elles étaient indépendantes et génère un nouvel échantillon à partir du précédent.

Nous allons voir deux exemples illustrant cette difficulté.

1. Un exemple simple

Il s'agit de températures ; on dispose de 20 données, mais le capteur n'a pas fonctionné la moitié du temps :

échantillon recueilli	échantillon généré
0	27,20
0	26,27
0	0,00
0	0,00
0	21,77
0	0,00
0	24,58
0	0,00
0	22,59
17,0	0,00
22,6	17,02
24,6	24,58
24,6	17,02
24,6	26,27
27,2	0,00
26,3	24,58
27,2	24,58
26,3	27,20
18,9	0,00
21,8	24,58

Il y a bien des zéros dans l'échantillon généré, mais ils sont répartis n'importe où et l'information "le capteur n'a pas fonctionné la moitié du temps" est perdue.

2. L'exemple des températures extrêmes

De même, pour les températures extrêmes, les données ne sont pas indépendantes : il est fréquent qu'il fasse très chaud plusieurs jours de suite. Les données ci-dessous sont celles vues précédemment ; elles sont extraites d'un recueil de températures journalières pour une ville en France, tous les jours entre 1921 et 2021. On s'intéresse à la situation où la température dépasse 39°C. Les données extraites sont :

29492	06/08/2003	39,5
29493	07/08/2003	39,8
29494	08/08/2003	39,9
29498	12/08/2003	40,6
32793	19/08/2012	39,7
35323	24/07/2019	40,4
35324	25/07/2019	41,8
35696	31/07/2020	39,9

On a deux manières de compter :

- On compte tous les jours, sur les 36 214 données disponibles, où la température dépasse le seuil. La probabilité sera donc approximativement $8/36\,214 = 0,00022$. On se dit que c'est extrêmement faible : seulement 8 jours en 100 ans !
- On compte par années : il y a 4 années où il a fait très chaud, à savoir 2003 (4 jours), 2012 (un jour), 2019 (2 jours) et 2020 (un jour). Là, on se dit qu'il y a 4 années chaudes en cent ans, donc la probabilité est $4/100 = 0,04$, ce qui est bien supérieur au précédent. De plus, la durée de retour moyenne d'une année chaude est de 20 ans (répartir 4 points sur un intervalle de longueur 100, tous les intervalles ayant même longueur). Là, on commence à s'inquiéter !

En réalité, "année chaude" devrait être traité différemment selon qu'il y a un, deux, trois, quatre, jours dépassant le seuil. En définitive, il n'y a qu'une seule occurrence, en cent ans, où l'on ait vu 4 jours dépassant le seuil (2003).

Il faut faire attention, si on génère des scénarios de température à partir des données recueillies : le fait d'avoir plusieurs jours consécutifs est perdu.

VI. En conclusion

La méthode présentée ci-dessus permet très facilement de générer des scénarios à partir d'un échantillon recueilli, en respectant la loi de cet échantillon : on ne triche en aucune manière et on ne fait aucune hypothèse factice de loi.

Par contre, et ceci est très important, la méthode suppose que les relevés sont indépendants, ce qui se répercute sur l'échantillon généré. Si les données ne sont pas indépendantes à l'origine (la même donnée répétée 300 fois), cette information est perdue lors de la génération, ce qui peut donner lieu à des erreurs graves.

Il est donc recommandé de s'interroger sur ce que représentent les données recueillies et de ne pas les traiter à l'aveugle.

VII. Cas d'une loi évoluant dans le temps

Il y a deux situations :

- Ou bien on constate sur les données recueillies qu'elles évoluent dans le temps, par exemple qu'il y a plus de pannes dans un passé récent, du fait de l'usure des équipements : c'est en général facile à voir ;
- Ou bien on fait cette hypothèse parce qu'on veut construire des scénarios qui prennent en compte cette possibilité ; c'est tout à fait raisonnable, on peut se demander, par exemple, comment les performances vont évoluer si l'usure s'accroît.

C. *Éléments de théorie*

Il faut tirer au hasard une donnée parmi toutes les données recueillies, mais ce ne sera plus avec une loi uniforme. Si on estime que la probabilité d'apparition augmente avec le temps, à la place d'une loi uniforme, on fera intervenir une fonction du temps, du type $p(t) = \frac{\alpha t + \beta}{C}$, où $\alpha > 0$ si la probabilité croît, $\alpha < 0$ si elle décroît, et C est une constante de normalisation, pour que l'intégrale de $p(t)$ fasse 1.

D. *Un exemple*

On dispose d'informations sur N années (ici 70) et on veut les utiliser pour une simulation sur N' années dans le futur, en privilégiant le passé récent. La loi de probabilité ne doit pas être uniforme sur toutes les années, mais mettre légèrement l'accent sur le passé récent. Bien sûr, il ne s'agit que d'une hypothèse, et on peut à loisir faire varier les coefficients.

On définit une loi de probabilité dépendant du temps :

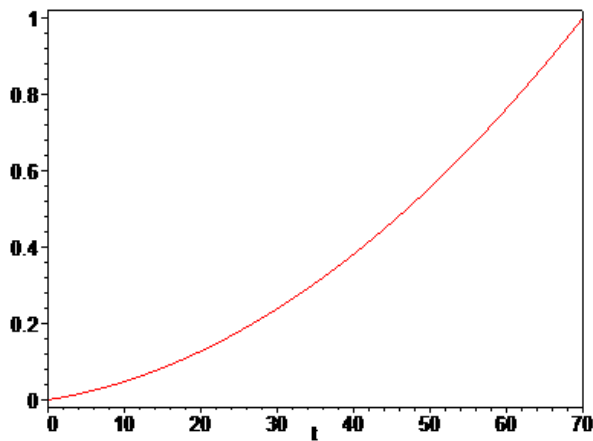
$p(t) = \alpha t + \beta$ avec par exemple $\alpha = 0.1$, $\beta = 1$: la pente est positive, donc la probabilité augmente avec le temps.

L'intégrale de 0 à N vaut $C = \int_0^N p(t) dt = \frac{\alpha N^2}{2} + \beta N$.

On prendra donc, après normalisation :

$$p(t) = \frac{1}{C}(\alpha t + \beta).$$

La fonction de répartition associée est :



Graphe de $F(t)$

$$F(t) = \frac{1}{C} \left(\frac{\alpha t^2}{2} + \beta t \right).$$

Soit y , $0 < y < 1$, la solution de $F(t) = y$ est :

$$t = \frac{-\beta + \sqrt{\beta^2 + 2\alpha Cy}}{\alpha}$$

Ici, nous sommes dans la situation très favorable où la fonction de répartition est facile à inverser (ce n'est pas le cas en général).

Avec les valeurs numériques ci-dessus, on obtient la formule :

$$t = -10 + 10\sqrt{1 + 63y}.$$

Si maintenant on veut faire un catalogue de tempêtes sur 100 ans, à partir de l'an 2100, on tirera 100 fois y , selon une loi uniforme sur l'intervalle $0-1$; ensuite, on calcule t selon la formule ci-dessus et on l'ajoute à l'année de départ 2100.

Si par exemple on tire $y = 0.33$, la formule ci-dessus nous donne $t = 36.7$; ce qui s'est produit pendant l'année 0.33×70 (puisque'il y a 70 années d'enregistrement) du tableau initial sera affecté à l'année 2136.

L'approche suivie mérite analyse : on ne dit pas qu'il y aura plus de tempêtes, ni qu'elles seront plus violentes. On dit simplement que le passé récent (les dix dernières années) a plus de chance de se répéter que le passé ancien (entre 60 et 70 ans).