



**Mémoire présenté pour la validation de la Formation
« Certificat d'Expertise Actuarielle »
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le**

Par : Benoît Régent-Kloeckner

Titre : Actualisation climatique des inondations de plaine

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

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

Membres présents du jury de l'Institut des
actuaires :

Membres présents du jury de l'Institut du Risk
Management :

Secrétariat :

Bibliothèque :

Entreprise : NA

Nom :

Signature et Cachet :

Directeur de mémoire :

Nom : Arthur Charpentier

Signature :

Invité :

Nom :

Signature :

**Autorisation de publication et de mise en
ligne sur un site de diffusion de documents
actuariels**

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

NA

Signature(s) du candidat(s)

BK

Mémoire présenté pour la validation de la Formation « Certificat d'Expertise Actuarielle » de l'Institut du Risk Management et l'admission à l'Institut des actuaires

Actualisation climatique des inondations de plaine

Benoît Régent-Kloeckner

22 septembre 2025

Encadré par Arthur Charpentier
Tuteur pédagogique Georges Louis Gonçalves

Table des matières

Résumé	v
Abstract	vii
Note de synthèse	ix
Executive summary	xvii
Remerciements	xxv
Introduction	1
I. Réchauffement climatique et péril inondation	5
1. Réchauffement climatique	5
1.1. Un tout petit peu de physique du globe	5
1.2. Lien entre réchauffement climatique et inondations	7
1.3. Limite des prévisions à long terme : points de bascule climatique	12
2. Simuler le futur : trajectoires d'émissions de gaz à effet de serre, modèles de circulation, modèles hydrologiques	12
2.1. Trajectoires d'émissions de gaz à effet de serre : RCP et SSP	14
2.2. Modèles de circulation globale et modèles de circulation régionale	15
2.3. Modèles hydrologiques	16
2.4. Modélisation des dommages	17
2.5. Architecture globale des simulations PESETA IV	17
3. Périls climatiques et catastrophes naturelles	19
3.1. Périls climatiques	19
3.2. Le régime des catastrophes naturelle en France	19
3.3. Discussion par péril	20
4. Réflexions sur le rôle de l'assurance face au réchauffement climatique	21
4.1. Actuaire et citoyen	21
4.2. Trois niveaux d'action	22
4.3. Chacun et chacune	22
4.4. Organisations	24
4.5. Tous et toutes ensemble	25
II. Modèle d'actualisation climatique	27
1. Présentation du modèle log-linéaire	28
1.1. Cahier des charges	28

1.2.	Construction du modèle	29
2.	Niveau de réchauffement global passé et futur	30
2.1.	Données utilisées	31
2.2.	Lissage des données	32
2.3.	Modèle probabiliste pour les scénarios de température	38
3.	Estimation des coefficients β_r d'actualisation climatique	43
3.1.	Les données PESETA IV	44
3.2.	Modélisation statistique	45
3.3.	Application du modèle statistique aux données PESETA IV	60
4.	Estimation des niveaux de risque régionaux et modélisation de la variabilité annuelle	66
4.1.	Positionnement du problème et première pistes	66
4.2.	Mélanger échantillonnage dans l'historique et régression log-normale	68
III. Applications actuarielles : projection de tendance, projection annuelle et stress test		73
1.	Projection de tendance pour la France métropolitaine	74
1.1.	Principe de la projection de tendance	74
1.2.	Résultats	75
1.3.	Comparaison bibliographique	76
2.	Génération de scénarios aléatoires de sinistralité	80
2.1.	Principe du générateur de scénarios	80
2.2.	Données pseudo-assurantielles utilisées	81
2.3.	Projection sur une année	84
2.4.	Scénarios de long terme pour un stress-test	86
3.	Résumé de l'étude menée	91
3.1.	Principaux résultats obtenus	92
3.2.	Limites de l'étude	92
3.3.	Perspectives	93
Conclusion		95
Annexe. Un point de vue sur la statistique		97
1.	La surcharge des noms de variable	97
2.	Les trois domaines de l'activité statistique	98
Index		101
Bibliographie		103

Résumé

Dans ce mémoire, nous décrivons, calibrons et appliquons un modèle log-linéaire d'« actualisation climatique » permettant de prendre en compte l'impact du réchauffement climatique dans l'évaluation des dommages futurs causés par les inondations de plaine. Ce modèle simple, initialement proposé par Jewson, Maynard et Dottori, peut être calibré par les données scientifiques ouvertes du projet Européen PESETA IV. Nous proposons un traitement statistique par *modèle mixte* permettant de mener ce calibrage en prenant en compte aussi bien les incertitudes de chacune des simulations sous-tendant les données utilisées, que la variation entre ces simulations.

Nous appliquons ensuite ce modèle, d'abord à l'échelle du territoire européen de la France pour comparer les dommages causés par les inondations de plaine qu'on peut attendre sur la période 2020–2050 rapportée à 1989–2019. Dans la trajectoire centrale SSP2-4.5 d'émissions de gaz à effets de serre, nous projetons une augmentation de 82% en valeur médiane, nettement supérieure aux projections de France Assureurs et de la CCR. Cependant les limites des données utilisées et de notre traitement et les incertitudes importantes ne permettent pas de remettre ces travaux en cause.

Enfin nous montrons sur un portefeuille fictif comment cette actualisation climatique permet de prendre en compte le réchauffement climatique sur des données synthétiques (dommages et enjeux assurés par région et par an). On retient que l'impact est important même pour des projections à un an du fait du réchauffement entre le passé et le présent, et l'impact sur la variabilité est plus marqué que sur les valeurs centrales.

Mots-clés : Inondations de plaine, réchauffement climatique, estimation de dommages, estimation statistique, modèles mixtes.

Abstract

In this memoir, we describe, calibrate and apply a log-linear model for “climatic updating”, that can be used to take Global Warming into account when estimating future damages caused by river floods. This simple model, initially proposed by Jewson, Maynard and Dottori, can be calibrated using open data from European scientific project PESETA IV. We develop statistical estimations based on *mixed models* to perform the model calibration, taking into account both uncertainties of each simulation underlying the data, and the variability between simulations.

We then proceed to apply this model, first to the European territory of France to compare possible damages of river floods in the 2020-2050 and 1989-2019 periods. In the central greenhouse gaz emission pathway SSP2-4.5, we project a 82% increase of damages for the central scenario, much higher than previous estimates by France Assureur and CCR. However, the limitations of the data we used and of our processing, and the very large uncertainties do not permit to question these previous studies.

Last we show on a hypothetical portfolio how to use climatic updating to take Global Warming into account on synthetic data (yearly regional damages and insured values). The main take-aways are that the impact of global Warming is large even for one-year projections, due to the Warming between past and present, and that this impact is even more acute on extreme values than on central estimates.

Keywords: river floods, Global Warming, damage estimates, statistical estimation, mixed models.

Note de synthèse

Les « périls climatiques », c'est-à-dire les catastrophes naturelles qui ont tendance à être aggravées par le réchauffement global, forment au sein de l'assurance non-vie un ensemble de risques très spécifiques. En effet, elles ne se déroulent pas en régime permanent mais dans une situation d'aggravation continue, il n'est donc pas possible de supposer que les dommages constatés dans le passé forment une bonne représentation des risques futurs. De plus cette aggravation n'est pas à vitesse constante, elle est guidée par un réchauffement global dont le niveau futur dépend des décisions politiques de l'ensemble des pays du globe, on ne peut donc pas non plus chercher dans l'historique une tendance qu'il suffirait d'extrapoler.

Une solution à ce problème consiste à estimer l'effet du réchauffement climatique à l'aide d'une modélisation physique de l'atmosphère permettant, pour chaque trajectoire possible des émissions de gaz à effet de serre (GES), d'obtenir des estimations des dommages futurs. Une telle modélisation nécessite une quantité extrêmement importante de calculs et des compétences scientifiques très pointues ; étant donné les incertitudes et les nombreuses hypothèses de modélisation nécessaire, il convient de plus d'utiliser plusieurs modèles différents pour quantifier (au moins en partie) les incertitudes. Un tel travail étant difficile à mener au sein d'une compagnie d'assurance, on aimerait idéalement combiner des données scientifiques sur l'effet du réchauffement climatique et les informations spécifiques du portefeuille sur lequel porte l'estimation. C'est l'objet d'un modèle « log-linéaire » proposé par Jewson, Maynard et Dottori dans un article récent,¹ qui peut être calibré à l'aide de données scientifiques publiée par le projet de recherche européen PESETA IV dans le cas des inondations de plaine.

L'objet de ce mémoire est de mener ce calibrage à l'aide de méthodes statistiques originales, de l'appliquer dans le cas de la France et d'illustrer comment il est possible de l'appliquer à un portefeuille de compagnie d'assurance.

Réchauffement climatique et péril inondation

L'état actuel des connaissances scientifiques (résumé notamment dans les rapports du GIEC) permet d'affirmer que la planète Terre est dans une phase de réchauffement global de son atmosphère, que ce réchauffement est principalement guidé par l'effet de serre, et causé par les activités humaines. De plus, on dispose d'informations sur une multitude d'effets de ce réchauffement sur l'enchevêtrement de systèmes physiques, chimiques, et biologique qui constitue notre environnement de vie, même si les incertitudes restent

¹ *A service to help insurers understand the financial impacts of changing flood risk in Europe, based on PESETA IV*, Climate Services 30 (2023), DOI 10.1016/j.cliser.2023.100395 [[JMD23](#)]

importantes. Parmi tous ces effets, nous nous intéressons ici aux inondations de plaine, qui sont amenées à devenir plus fréquentes et plus sévères à mesure que le réchauffement s'accroît.

Précisons le terme d'*inondation de plaine* (aussi appelées *inondations fluviales*) : il s'agit du débordement progressif d'un cours d'eau au-delà de son lit habituel ; on les distingue des crues-éclair, beaucoup plus rapides et concentrées sur de relativement petits cours d'eau, ainsi que des inondations qui ne sont pas liées aux cours d'eau comme le ruissellement, le débordement de nappe, la submersion marine. Notre principale motivation pour nous intéresser à ce type d'inondation est que nous disposons de données scientifiques les concernant ; notons que les inondations représentent à elles seules près de la moitié des montants de sinistre indemnisés au titre des catastrophes naturelles sur la période 1982–2023,² même si la part de la sécheresse est en forte progression sur la période récente.

La raison pour laquelle le réchauffement global est susceptible d'aggraver les inondations est physiquement assez simple : un air plus chaud a une plus grande capacité de rétention de vapeur d'eau. Ainsi le réchauffement climatique accentue globalement l'évaporation à l'interface océan-atmosphère, et augmente la quantité d'eau circulant dans l'atmosphère sous forme de vapeur ou de nuages. Le réchauffement n'étant homogène ni dans l'espace, ni dans le temps, cet air fortement chargé en eau est amené à se refroidir à certains endroits et certains moments, provoquant des précipitations d'autant plus importantes, lesquelles entraînent donc plus souvent des inondations.

La quantification de cet effet est, elle, très délicate. Nous nous appuyons sur des données du projet de recherche européen PESETA IV, publiées en accompagnement d'un article de recherche mené par Dottori et Feyen.³ Ces données sont obtenues en combinant les résultats de différentes simulations atmosphériques, un modèle hydrologique permettant d'évaluer les niveaux des différents cours d'eau européens, et un modèle de dommages convertissant ces niveaux d'eau en montants de sinistres. On dispose d'estimations de dommages dans chacune des régions (en France, il s'agit des 22 régions administratives d'avant la réforme territoriale effective depuis 2016), dans trois régimes (année de référence 1995 correspondant à environ +0,7°C de réchauffement, +1,5°C et +2°C) et pour 22 jeux d'hypothèses (11 simulations atmosphériques, chacune dans les deux trajectoires d'émissions de gaz à effet de Serre RCP-4.5 et RCP-8.5).

Modèle d'actualisation climatique

Le modèle que nous allons utiliser, proposé par Jewson, Maynard et Dottori, postule que chaque degré Celsius de réchauffement global a un impact relatif fixé sur les dommages, toutes choses égales par ailleurs. On peut l'écrire sous la forme :

$$Dom_{r,t} = EA_{r,t} \cdot A_r e^{\beta_r \text{GWL}(t)} \cdot F_{r,t}$$

² *Les catastrophes naturelles en France, Bilan 1982-2023*, CCR [CCRB]

³ *Cost-effective adaptation strategies to rising river flood risk in Europe*, Nature Climate Change 13 (2023), DOI 10.1038/s41558-022-01540-0 [DMB+23]

où $Dom_{r,t}$ est le montant de dommages dans la région r pour l'année t , $EA_{r,t}$ représente les enjeux assurés, A_r est un facteur d'exposition du portefeuille de la région r , $GWL(t)$ est le niveau de réchauffement global l'année t , β_r représente la susceptibilité de la région au changement climatique, et $F_{r,t}$ est la partie aléatoire représentant les fluctuations dues aux éléments qui échappent à la modélisation.

En notant $E_{r,t} = \log F_{r,t}$ et $\alpha_r = \log A_r$, on aboutit à la forme log-linéaire du modèle :

$$\log \frac{Dom_{r,t}}{EA_{r,t}} = \alpha_r + \beta_r GWL(t) + E_{r,t}.$$

Afin de pouvoir appliquer ce modèle, nous avons besoin de mener plusieurs estimations et modélisations :

- i. modéliser l'incertitude sur l'évolution du réchauffement global dans chaque trajectoire d'émission de gaz à effet de serre qu'on souhaite utiliser ;
- ii. estimer les coefficients β_r de susceptibilité au réchauffement climatique à partir des données scientifiques de PESETA IV ;
- iii. estimer les expositions α_r et la loi des variabilités annuelles $E_{r,t}$ à partir des données du portefeuille.

Les enjeux assurés $EA_{r,t}$ seront, suivant les cas d'usage, soit fournis dans les données actuarielles (pour le futur on pourra utiliser un plan de développement) soit supposés constants (par exemple pour la calibration du modèle, où on utilisera des données en environnement économique fixé).

L'ensemble des choix qui vont être détaillés ci-dessous reposent sur la façon dont on va ensuite utiliser ce modèle : nous allons générer de nombreux scénarios de dommages pseudo-aléatoires représentant au maximum les différents aléas et incertitudes présent dans les données et le modèle, et c'est de cet ensemble de scénarios que nous tirerons des projections.

Scénarios de température : $GWL(t)$

Des données issues du dernier rapport du GIEC donnent pour chacune de cinq trajectoire d'émissions de GES une série de valeurs moyennes du réchauffement global, ainsi que des quantiles à 5% et 95% (figure 1).

Ces températures sont des mesures pour la période « historique » 1950–2014 et des projections pour 2015–2099 (notons que les différentes trajectoires restent proches sur 2015–2025). Mesures comme projections sont entachées d'incertitudes, mais les mesures sont données brutes tandis que les projections dont nous disposons sont lissées.

Nous effectuons comme premier traitement un lissage de Whittaker et Eilers qui permet de dégager une tendance globale sans introduire d'artefact à la jonction entre les deux périodes. Pour chaque date, nous utilisons une loi simple compatible avec l'asymétrie des deux quantiles autour de la moyenne, et enfin nous supposons une complète dépendance entre les différentes années : on tire un quantile uniformément dans $[0,1]$

et on l'utilise pour toutes les années. On obtient ainsi un générateur de scénarios de réchauffement qui prend en entrée une trajectoire d'émission de GES et une période I , et qui fournit un jeu pseudo-aléatoire de scénarios de températures $(\text{GWL}(t))_{t \in I}$.

Calibrage des coefficients β_r

Pour estimer les coefficients régionaux de susceptibilité au réchauffement climatique, on développe une méthode statistique différente de la méthode assez crue de Jewson, Maynard et Dottori, afin de prendre en compte une plus grande part des incertitudes.

Nous disposons pour chaque région r de 22 réalisations des simulations, chacune fournissant 3 points correspondant à 3 niveaux de réchauffement global. En échelle logarithmique, on effectue pour chaque simulation s une régression linéaire sur ces 3 points pour obtenir une estimation $\hat{\beta}_{r,s}$ de β_r , ainsi qu'une estimation de sa variabilité qui découle directement du plus ou moins bon alignement des 3 points. Il peut paraître surprenant de faire une régression sur aussi peu de points, mais c'est ici la présence de multiple jeux de données qui donnera la robustesse statistique.

Pour combiner ces 22 estimations, on utilise un *modèle mixte* issu des analyses de mesures répétées. Il s'agit de postuler que chacune de nos estimations $\hat{\beta}_{r,s}$ est la réalisation d'une variable aléatoire d'espérance égale au coefficient β_r , entaché de deux erreurs : un biais de simulation, supposé gaussien de même variance inconnue pour toutes les simulations, et une variabilité dans la réalisation des 3 points, encore supposée gaussienne mais de variance dépendant de la simulation et supposée égale à notre estimation de l'incertitude sur $\hat{\beta}_{r,s}$.

Les résultats de l'application de cette méthode à la France sont présentés dans la figure 2. On observe une susceptibilité au réchauffement climatique particulièrement élevée en Alsace, Pays de la Loire et Haute-Normandie. Il convient de relever que les coefficients β_r sont dans l'ensemble très élevés : rappelons qu'une augmentation de 1°C de la température globale correspond dans le modèle log-linéaire à une multiplication par e^{β_r} , soit un doublement à partir de $\beta_r \simeq 0,69$, ce à enjeux assurés constants.

Lors de l'utilisation du modèle, pour chaque scénario de dommage on tire aléatoirement un vecteur $(\beta_r)_r$ supposé gaussien, paramétré par les estimations $\hat{\beta}_r$ et les estimations de variance et de covariance entre régions.

Expositions et variabilité annuelle : α_r et $E_{r,t}$

Enfin, nous devons à partir de l'historique du portefeuille considéré déterminer le coefficient α_r , qui reflète le degré d'exposition du portefeuille dans la région r supposé constant dans le temps, et une loi pour $E_{r,t}$ qui représente la variabilité annuelle, modélisée par une suite de variables aléatoires indépendante dans le temps, mais avec une dépendance entre régions. Le calibrage se fait sur la période historique, c'est-à-dire celle pour laquelle on dispose des données de dommages et d'enjeux assurés.

La première chose est de mener l'« actualisation climatique » de l'historique, c'est-à-dire de mettre en *as-if* les données du passé. Il faut pour cela déjà disposer de la courbe de réchauffement global et des coefficients β_r , tirés aléatoirement dans chaque

(a) Global surface temperature change relative to 1850–1900

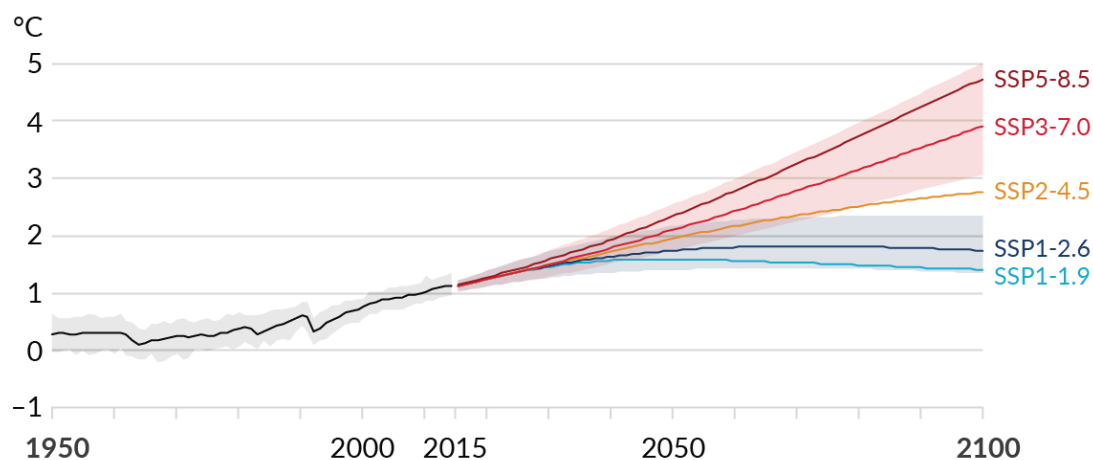


FIGURE 1. : Extrait de la figure SPM.8 du 6^e rapport du GIEC. Les courbes indiquent les valeurs moyennes, les zones colorées sont les intervalles interquantiles [5%; 95%] pour les trajectoires SSP1-2.6 et SSP3-7.0.

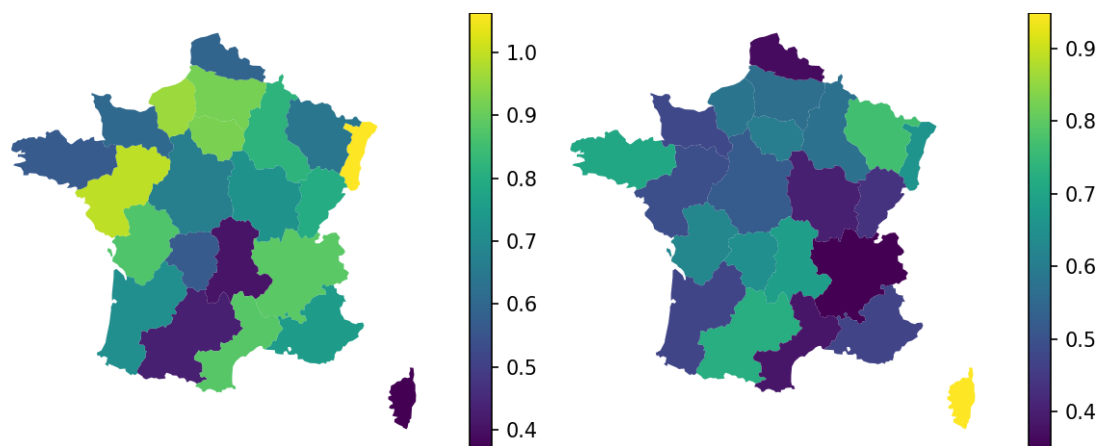


FIGURE 2. : À gauche, estimations $\hat{\beta}_r$ des coefficients d'impact régionaux du réchauffement global. À droite, rapport écart-type sur estimation, représentant l'incertitude relative sur la valeur réelle de β_r .

scénario comme on l’a dit plus haut. On obtient alors en appliquant le modèle log-linéaire des valeurs pour les sommes $\alpha_r + E_{r,t}$ aux différentes dates. Le paramètre α_r est alors estimé par la moyenne empirique, puis on construit une loi pour les $E_{r,t}$ à l’aide d’une approche mixte entre un tirage dans l’historique et une approche paramétrique, qui permet d’utiliser toute l’information de l’historique sur les dépendances entre régions tout en pouvant modéliser des événements plus extrêmes que ceux déjà constatés.

Applications actuarielles

Projection de tendance pour la France métropolitaine

Dans un premier temps, on s’intéresse à la tendance de long terme pour l’entièreté du territoire européen de la France. On n’utilise ici que les données scientifiques, à travers notre modélisation des températures et du coefficient β pour l’union des 22 régions (traité de la même façon qu’une région unique). Afin de faciliter la comparaison avec les résultats de France Assureurs [Fra22], on a rapporté dans chaque scénario les dommages totaux de la période 2020–2050 à ceux de la période 1989–2019. Les résultats sont présentés dans la figure 3 et la comparaison avec les trois rapports mentionnés plus haut est présentée dans le tableau 4.

On constate que nos méthodes et surtout les données PESETA IV mènent à des projections significativement plus sévères que les rapports de France Assureurs et de la CCR, mais pas si éloignées de celles de Covéa. Il est important de noter que les limites des données utilisées et du traitement que nous en faisons ne permettent pas de remettre en cause les travaux mentionnés.

Application à un portefeuille d’assurance

Afin d’illustrer la façon dont on peut appliquer le principe d’actualisation climatique et le modèle log-linéaire à l’échelle d’une compagnie, même à partir de données synthétiques (dommages et enjeux assurés annuels), nous avons appliqué les méthodes décrites ci-dessus au portefeuille fictif de garantie inondation en Multi-risque habitation de la *compagnie A*, entreprise d’assurance IARD fictive basée à Bordeaux. L’historique construit couvre la période 1994–2023.

Nous avons tout d’abord mené une estimation à un an, par exemple dans l’objectif d’adapter la tarification, en provisionnement de comparer le modèle log-linéaire au SCR en formule standard, ou de tester les traités de réassurance. Les résultats sont présentés dans le tableau 5. La ligne « climat constant » permet de comparer les estimations avec ce qu’on obtient en ignorant l’effet du réchauffement global. On constate que le choix de trajectoire d’émissions de GES est peu important à l’échelle d’une seule année, mais que la prise en compte du réchauffement global a un impact très fort (du fait de l’actualisation climatique des sinistres passés). On constate également que l’impact est plus fort sur le quantile à 99,5% que sur la moyenne, ce qui est cohérent avec l’augmentation prévisible de la variabilité des périls climatiques.

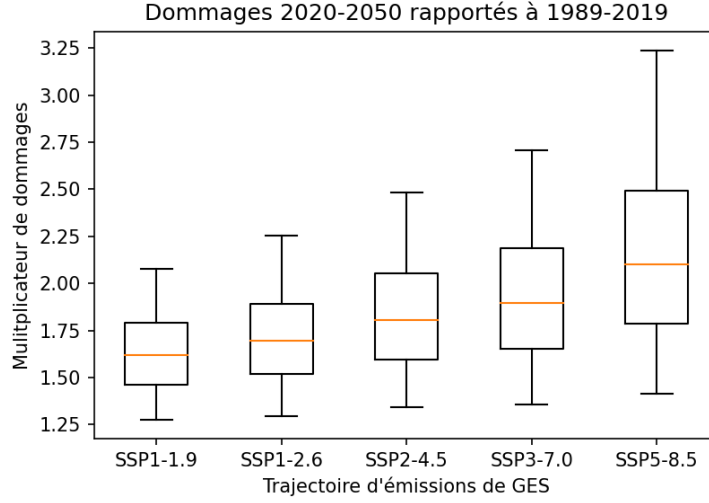


FIGURE 3. : Distribution des rapports de dommages entre les deux périodes, pour $N = 10\,000$ répétitions. Sont indiqués les quantiles 5%, 25%, 50%, 75% et 95%.

Étude	Écart temporel	Trajectoire	Dommages	Dont RG	Impact RG annuel
FA	31 ans	RCP 8.5	+81%	+11%	+0,34%
CCR	50 ans	RCP 8.5	+27%	+6%	+0,12%
		RCP 4.5	+43%	+19%	+0,35%
Covéa	37 ans	RCP 8.5	+110%	N.D.	(+2,03%)
Mémoire	31 ans	SSP5-8.5	N.D.	+110%	+2,42%
		SSP2-4.5	N.D.	+82%	+1,95%

TABLE 4. : Comparaison des résultats du rapport France Assureurs [Fra22] (noté FA), du rapport CCR [CCR23] (CCR) et du livre blanc Covéa [CR22] (Covéa), et pour comparaison un extrait des résultats du présent mémoire (Mémoire). Sont indiquées les écarts entre les centres des périodes de référence et de projection, les trajectoires d'émissions de gaz à effet de serre utilisées, l'augmentation projetée des dommages pour la seconde par rapport à la première (intégrant l'augmentation des enjeux assurés, Covéa étant dans une situation intermédiaire (iii)), l'augmentation projetée pour le seul réchauffement global (RG), et la moyenne annuelle de cette dernière. La valeur pour Covéa est sur l'augmentation totale en l'absence de donnée sur la part due au réchauffement global.

Ensuite, nous avons mené des projections de long terme (période 2025–2049) dont les résultats sont présentés dans le tableau 6. La leçon essentielle est que sur le long terme, les différentes trajectoires d’émissions de GES correspondent toutes à des augmentations particulièrement importantes de la pire année, avec une variation importante entre les simulations et entre les trajectoires d’émissions de GES.

Trajectoire	Médiane	Moyenne	quantile d’ordre 99,5%
Climat constant	3,5	5,0	29,6
SSP1-1.9	5,2	7,6	64,6
SSP1-2.6	5,2	7,6	63,1
SSP2-4.5	5,2	7,7	64,4
SSP3-7.0	5,2	7,8	65,4
SSP5-8.5	5,4	8,2	70,5

TABLE 5. : Dommages (en M€) simulés en 2025 dans les différentes trajectoires d’émission de GES, résumé de $N = 200\,000$ réalisations du modèle.

Trajectoire	Moyenne temporelle		Dernière année		Pire année	
	Moyenne	q. à 95%	Moyenne	q. à 95%	Moyenne	q. à 87,5%
Climat constant	5,0	6,7	5,0	12,4	19,3	29,2
SSP1-1.9	9,0	28,1	9,8	92,7	43,4	75,0
SSP1-2.6	9,8	34,4	11,6	109,2	47,3	81,8
SSP2-4.5	10,9	45,9	15,0	154,6	54,7	93,6
SSP3-7.0	11,9	54,1	18,7	220,0	62,1	105,4
SSP5-8.5	14,5	76,6	24,8	281,2	78,9	135,3

TABLE 6. : Dommages (en M€) simulés sur 2025–2049 dans les différentes trajectoires d’émission de GES, résumé de $N = 20\,000$ réalisations du modèle. La moyenne temporelle est calculée sur les 25 ans dans chaque réalisation, les autres moyennes et les quantiles portent sur les N réalisations.

Executive summary

“Climatic hazards”, by which we mean the natural disasters the severity of which increases with Global Warming, are quite specific among non-life risks in insurance. One indeed cannot assume a steady state, past damages are not a good benchmark of future risks. Moreover it is also not advisable to seek in the records a trend to extrapolate, since the worsening of climatic hazards need not be at constant speed, but depends on the quickness of global warming and thus on worldwide policies.

A possible solution consist in using, for each greenhouse gas (GHG) emission pathway, a physical modelling of the atmosphere to estimate future damages. Running such models needs huge computing ressources and specialized scientific skills; moreover any such model is built over a wealth of hypotheses and incurs many uncertainties, which can only be estimated by using and comparing several such models. Such an endeavor is difficult to pursue inside an insurance company, making it desirable to use existing scientific data on the effects of Global Warming to estimate its plausible impact on a given insured portfolio. This is precisely the goal of a “log-linear” model proposed by Jewson, Maynard and Dottori in a recent article.⁴ The parameters of this model can be fitted on open scientific data published by the PESETA IV european project in the specific case of river flooding.

In this memoir, we develop original statistical methods to fit this model on these data and apply them in the case of France, including an illustration of how to apply the model to an insurance company’s portfolio.

Global Warming and flood hazard

The current state of scientific knowledge, for example gathered in the IPCC reports, tells us that the Earth’s atmosphere is warming globally, that this warming is mainly driven by greenhouse effect and caused by human activity. Our knowledge extends to many effects of Global Warming to the intricate layers of physical, chemical and biological systems that is our life environment, but with many uncertainties. Among these effects, we shall here be concerned with river flooding, which are destined to become more frequent and more severe as Global Warming intensifies.

A *river flood* occurs when a waterway overflows progressively outside of its usual bed. It is to be distinguished from flash floods, which are much faster and only concern smaller streams, and from floods unrelated to waterways such as runoff, groundwater overflows and marine submersions. Our main motivation to specifically study river floods is the

⁴ *A service to help insurers understand the financial impacts of changing flood risk in Europe, based on PESETA IV*, Climate Services 30 (2023), DOI 10.1016/j.cliser.2023.100395 [[JMD23](#)]

availability of the scientific data we need, but it must be stressed that floods make for almost half of compensations among natural disasters in France over the 1982–2023 period,⁵ although droughts are quickly taking a larger share.

The physical mechanism by which Global Warming induces more floods is pretty simple. Since warmer air has a greater water absorption capacity, Global Warming causes more evaporation at the ocean-atmosphere interface and more water to travel in the atmosphere, either as vapor or in clouds. The warming being heterogeneous both in time and in space, this water-heavy air is bound to get colder at some time and place, creating intense rainfalls which in turn cause floods.

Quantifying this effect is however very difficult. We shall use data from the PESETA IV European research project, published alongside a research article lead by Dottori and Feyen.⁶ These data are obtained by combining atmospheric simulations, a hydrologic model estimating in each meteorological scenario the water levels in all European rivers, and a damage model converting these water levels into monetary damages. This process provides us with damage estimates for each region (corresponding for France to the 22 administrative regions anterior to the 2016 reform), for 3 global warming levels (reference year 1995 with a global warming level of $\sim +0.7^\circ\text{C}$, $+1.5^\circ\text{C}$ and $+2^\circ\text{C}$) and in 22 sets of hypotheses (11 atmospheric simulations, each for the 2 GHG emission pathways RCP-4.5 and RCP-8.5).

Climatic update model

The model we use, proposed by Jewson, Maynard and Dottori, assumes that each additional degree Celsius of global warming has the same relative impact on damages, all things being equal. It takes the form:

$$Dom_{r,t} = EA_{r,t} \cdot A_r e^{\beta_r \text{GWL}(t)} \cdot F_{r,t}$$

where $Dom_{r,t}$ is the amount of damages in region r during year t , $EA_{r,t}$ is the total insured value, A_r represents the portfolio exposure in region r , $\text{GWL}(t)$ is the global warming level, β_r measures the sensitivity of region r to Global Warming, and $F_{r,t}$ is a random variable gathering all variations escaping the model.

Introducing $E_{r,t} = \log F_{r,t}$ and $\alpha_r = \log A_r$, we can write the model in log-linear form:

$$\log \frac{Dom_{r,t}}{EA_{r,t}} = \alpha_r + \beta_r \text{GWL}(t) + E_{r,t}.$$

In order to apply this model, we need several steps:

- i.* model uncertainties on the global warming trajectory, in each GHG emission pathway to be used,
- ii.* estimate the parameters β_r from PESETA IV data,

⁵ *Les catastrophes naturelles en France, Bilan 1982-2023*, CCR [CCRB]

⁶ *Cost-effective adaptation strategies to rising river flood risk in Europe*, Nature Climate Change 13 (2023), DOI 10.1038/s41558-022-01540-0 [DMB+23]

- iii. estimate exposures α_r and model a law for the random variables $E_{r,t}$ from the portfolio records.

The total insured values $EA_{r,t}$ shall either be given in the actuarial data (possibly including a business plan for the future insured values) or assumed to be constant (e.g. when fitting the model, since the scientific data are given at constant economic environment).

The choices to be described below are contingent to the way we shall use the model, namely by running a number of pseudo-random damages scenarios representing the various possibilities contained in the data and model uncertainties. From this ensemble of scenarios we shall extract projections and uncertainties estimates.

Global warming scenarios: $GWL(t)$

Data from IPCC's last report give series of averages and 5% and 95% percentiles for the Global Warming Levels at a yearly timescale, for each of five GHG emission pathways (Figure 7).

These temperatures come from measures for the “historical” 1950–2014 period and from projections for the 2015–2099 period (with the various GHG emission pathways staying close one to another in the 2015–2025 period). Both measures and projections entail uncertainties, but here measures are raw data while projections have been smoothed.

We first use a Whittaker-Eilers smoother to extract a global trend without adding artifacts at the junction between the two periods. Then for each year, we use a simple law compatible with the asymmetry that can be observed of the 5% and 95% percentiles with respect to the average. Finally, we assume a complete dependency between years: a single quantile is drawn uniformly in $[0,1]$ and used for all years. This provides us with a simple scenario generator taking as argument a GHG emission pathway and a period I , and giving as outcome a pseudo-random ensemble of Global Warming Levels $(GWL(t))_{t \in I}$.

Calibrating the β_r parameters

To estimate the regional Global Warming susceptibility parameters $(\beta_r)_r$, we depart from Jewson, Maynard and Dottori's relatively crude approach and build a statistical model that takes a larger part of uncertainties into account.

For each region r , the data set we use provides the results of 22 simulations, each of which comprises 3 points, one for each of 3 Global Warming Levels. We proceed to a linear regression in logarithmic scale on these 3 points to obtain an estimate $\hat{\beta}_{r,s}$ of β_r , and an estimate of its variability deduced from the quality of alignment of the points. It may come as a surprise to use linear regression on such a small number of points, but here the multiplicity of simulations is what provides some statistical robustness.

To combine the 22 estimations thus obtained, we use a *mixed model* from repeated measures design. The principle is to assume each of our estimate $\hat{\beta}_{r,s}$ to be a realization of a random variable of expectation β_r , with two noise components: a simulation bias, assumed to be Gaussian with the same unknown variance for all simulations, and a

Gaussian noise for each simulation (corresponding to uncertainty in the three points, given the simulation), whose variance depends on the simulation but is known (given by our variability estimate on $\hat{\beta}_{r,s}$).

The results of this approach for France are represented in Figure 8. We observe a particularly high susceptibility to global Warming in Alsace, Pays de la Loire and Haute-Normandie. The β_r parameter are overall quite high, considering that in the log-linear model a raise of 1°C in the Global Warming Level corresponds to a factor e^{β_r} applied to damages (i.e. doubling damages at $\beta_r \simeq 0.69$), at constant insured values.

When we use the model to build projections, for each scenario we draw randomly a vector (β_r) assumed to be Gaussian, whose parameters are given by our estimates $\hat{\beta}_r$ and variance-covariance estimates between regions.

Exposures and yearly variability: α_r and $E_{r,t}$

Last, we need to estimate the exposure parameter α_r of the portfolio in region r (assumed not to vary in time) and a law for the yearly variability $E_{r,t}$, for which we assume independence over time but possibly strong dependencies between regions. Calibrating is done on the historical period, that is the time interval over which we have data on past damages and insured values.

The first point is to “update for climate” historical data, to get data *as if* climate was constant. To this end, we need a Global Warming Level curve and parameters β_r , which as mentioned earlier are randomly drawn for a given scenario. In each such scenario, applying the log-linear model provides us with a series of sums $\alpha_r + E_{r,t}$ for each region r and each t in the historical period. We first estimate α_r as the empirical median, then we build a law for the $E_{r,t}$ with a mixed approach combining drawing from the (updated) historical values and a parametric method. This enables us to both use all available information on inter-regional dependency and to model more extreme events than those that have been observed in the historical period, necessarily short compared to the return time to be considered.

Actuarial applications

Trend projection for Metropolitan France

Our first application is to build long-term trends for the European territory of France. We only need the scientific data for this, through our estimates for Global Warming Levels and the β parameter for the union of the 22 French regions (considered here as one region, then submitted to the treatment explained above). To facilitate comparison with a report from France Assureurs [Fra22], we computed the ratio of total damages in 2020-2050 to total damages in 1989-2019. Results are presented in Figure 9 and comparison with three reports mentioned above is given in Table 10.

We observe that our methods, and foremost PESETA IV data, lead to much more severe projections than studies by France Assureurs and CCR, not too far from the

(a) Global surface temperature change relative to 1850–1900

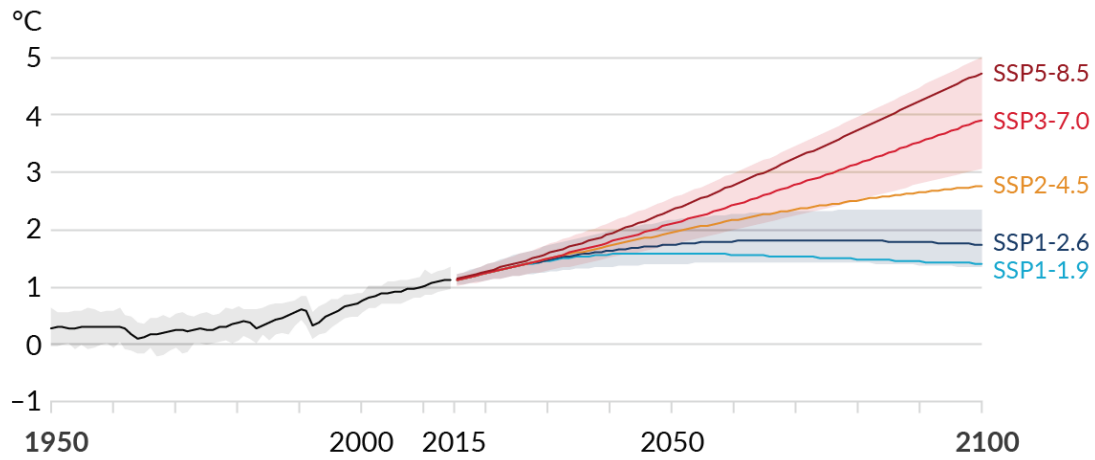


Figure 7.: Extract from Figure SPM.8 of the 6th IPCC report. Curves give average values, colored zones are the quantile intervals [5%; 95%] for SSP1-2.6 and SSP3-7.0 pathways.

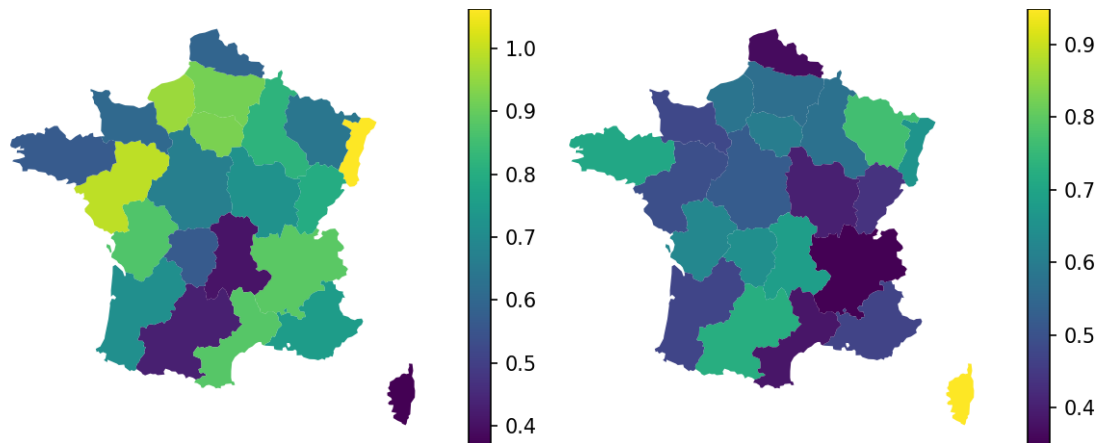


Figure 8.: Left: estimates $\hat{\beta}_r$ of regional Global Warming susceptibility parameters. Right: estimate to standard deviation ratio, indicating the relative uncertainty on the real value of β_r .

results of Covéa. It is however to be noted that the limitations of the data we used and of our processing do not permit to question these previous studies.

Insurance Portfolio

In order to illustrate how one can apply at the scale of one company the log-linear model to update for climate, even from synthetic data (yearly regional damages and insured values), we applied the above methods to the hypothetical portfolio of home insurance of *Company A*, a hypothetical insurance company based at Bordeaux. Historical data have been build from a simple random model for the 1994–2023 period.

We first proceed to a one-year projection, for example to adjust pricing, compare provisioning from the log-linear model to Solvency 2’s standard formula, or challenge the reinsurance policy. Results are given in Table 11. The “constant climate” line is provided to compare the results with and without climate updating. We can observe that the choice of GHG emission pathway is mostly irrelevant at this short time frame, but that climate updating has a strong impact (because of the climate updating of past data). The impact of climate updating is stronger on the 99.5% percentile than on the average, which is consistent with the foreseeable raise in variability of climatic hazards.

We then build long-term projections (2025–2049) presented in Table 12. The main take-away is that in the long run, all GHG emission pathways correspond to important raises of the worse year, with large variations between scenarios and between pathways.

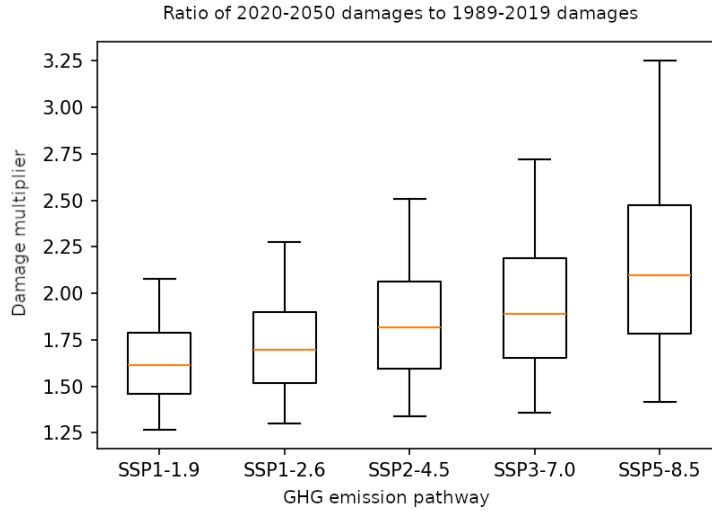


Figure 9.: Distribution of damage ratios between the two periods, with $N = 10\,000$ scenarios. Bars correspond to 5%, 25%, 50%, 75% and 95% percentiles.

Study	Temporal spread	Pathway	Damages	GW impact	yearly GW impact
FA	31 years	RCP 8.5	+81%	+11%	+0.34%
CCR	50 years	RCP 8.5	+27%	+6%	+0.12%
		RCP 4.5	+43%	+19%	+0.35%
Cov��a	37 years	RCP 8.5	+110%	N.D.	(+2.03%)
Memoir	31 years	SSP5-8.5	N.D.	+110%	+2.42%
		SSP2-4.5	N.D.	+82%	+1.95%

Table 10.: Comparison of results from studies by France Assureurs [Fra22] (FA), CCR [CCR23] and Cov  a white book [CR22] (Cov  a), and an extract of the present memoir (Memoir). Are given the spread between the centers of the historical and projection periods, the underlying GHG emissions pathway, the projected raise in damages (including the raise in insured values, except for Cov  a which only takes it partly into account), the projected raise due to Global Warming (GW), and its annual average. The last value for Cov  a is on the total raise given the lack of data.

Pathway	Median	Average	99,5% percentile
Constant climate	3.5	5.0	29.6
SSP1-1.9	5.2	7.6	64.6
SSP1-2.6	5.2	7.6	63.1
SSP2-4.5	5.2	7.7	64.4
SSP3-7.0	5.2	7.8	65.4
SSP5-8.5	5.4	8.2	70.5

Table 11.: Damages (M€) projected in 2025 for different GHg emissions pathways, summing up $N = 200\,000$ scenarios.

Pathway	Temporal mean		Last year		Worse year	
	Average	95% per.	Average	95% per.	Average	87,5% per.
Constant climate	5.0	6.7	5.0	12.4	19.3	29.2
SSP1-1.9	9.0	28.1	9.8	92.7	43.4	75.0
SSP1-2.6	9.8	34.4	11.6	109.2	47.3	81.8
SSP2-4.5	10.9	45.9	15.0	154.6	54.7	93.6
SSP3-7.0	11.9	54.1	18.7	220.0	62.1	105.4
SSP5-8.5	14.5	76.6	24.8	281.2	78.9	135.3

Table 12.: Damages (M€) projected over 2025–2049 for different GHG emissions pathways, summing up $N = 20\,000$ scenarios. Temporal mean is computed over the 25 years in each scenario, averages and percentiles are on the N scenarios.

Remerciements

Un mémoire est un travail personnel, d'autant que celui-ci a été écrit hors entreprise, mais tout travail humain repose sur un grand nombre d'autres personnes : nos outils ont été façonnés par d'autres, nos connaissances ont été acquises au contact d'enseignantes et enseignants et de camarades, et chaque travail est aussi plus directement le fruit d'interactions avec d'autres.

L'idée de départ de ce mémoire est venue d'une mission qui m'a été proposée par Quynh-Le Tran, lorsque je travaillais chez Forsides France où Arnaud Cohen m'a donné l'occasion de faire mes premières armes actuarielles ; ils m'ont permis à mon départ de conserver les quelques fichiers qui constituaient le début de ce travail. Lorsque j'en suis parti, Arthur Charpentier a accepté de m'encadrer ; à la fin de mon travail, il m'a introduit auprès de la chaire PARI, et notamment de Laurence Barry ; celle-ci m'a donné l'occasion de présenter mon travail devant Pierre François, Jean-Philippe Naulin de la CCR et François Hu et Maxence Colin de Milliman. Alexandre Vincent m'a donné un peu de son temps pour échanger sur mon travail et le rapport France Assureurs 2021. Steve Jewson, Francesco Dottori et Pascal Yiou ont répondu précisément et promptement à mes questions sur leurs travaux, ce qui m'a été d'une grande aide. Merci à elles et eux tous.

Je remercie également toutes les personnes qui ont développé les outils ouverts dont je me suis servi, principalement L^AT_EX pour la rédaction, Python et ses nombreuses bibliothèques (notamment **pandas** et **sklearn**) pour la programmation ; ainsi que l'équipe enseignante du certificat d'expertise actuarielle et Ophélie Dupas qui en organise tous les aspects pratiques. Merci enfin aux camarades, et spécialement à Alexis, qui m'ont accompagné dans ces deux années d'études si particulières, et tant bien que mal expliqué le monde de l'assurance et de l'entreprise privée.

Introduction

Les inondations font partie des événements qui montrent de façon très concrète les mérites de l'assurance : si elles tuent relativement peu, elles n'en ont pas moins le potentiel de détruire la vie de celles et ceux qui les subissent et peuvent y perdre logement, locaux ou récoltes. L'indemnisation peut alors faire la différence entre un événement de vie extrêmement marquant et la ruine pure et simple. Les inondations représentent à elles seules près de la moitié des montants de sinistre indemnisés en France au titre des catastrophes naturelles sur la période 1982-2023 [CCRb], et même si le péril sécheresse gagne très vite en importance, elles restent un sujet de préoccupation majeure. La période de rédaction de ce mémoire, en 2024 et début 2025, a ainsi coïncidé avec une série noire d'inondations record dans plusieurs régions françaises, en Europe centrale, en Espagne et en Afrique occidentale et centrale, sans qu'on puisse à ce stade dire dans quelle mesure l'année est une anomalie et dans quelle mesure elle est représentative de la tendance d'aggravation qu'il faut attendre dans le futur.

Les catastrophes naturelles forment un sujet bien spécifique au sein de l'assurance non-vie, de part leur grande variabilité, le poids des événements extrêmes, le régime français spécifique de réassurance « CAT-NAT » et le rôle du changement climatique. Ce dernier point est au centre de ce mémoire, comme on le détaillera il vient mettre en cause la méthode première de l'actuaire qui consiste à anticiper le futur à l'aide d'informations sur le passé. Plusieurs études ont été menée en France sur ce sujet, nous discuterons notamment les derniers rapports de France Assureur [Fra22], de la CCR [CCR23] et le livre blanc Covéa [CR22]. L'un de nos objectifs est, pour le cas des inondations de plaine, de produire une estimation de l'accroissement futur des dommages qui puisse être comparée aux résultats de ces trois études. Les résultats obtenus avec l'approche développée dans ce mémoire sont plus pessimistes que ceux des trois études mentionnée, assez proches de ceux de [CR22] mais loin devant ceux de [Fra22] et [CCR23].

Les périls significativement influencés par le changement climatique (inondations, sécheresses notamment) ne sont pas dans un régime permanent, mais évolutif, et estimer le futur sur un historique de sinistre fait courir le risque d'une sous-estimation des engagements. Pire, plus l'historique est long (donc robuste), plus cette sous-estimation peut être grave car le passé plus lointain correspond à un régime d'autant moins sévère. De plus, vu la variabilité, il est difficile de saisir dans les données la tendance à l'augmentation des dommages ; et même si on la captait cette tendance passée n'est pas nécessairement représentative de l'évolution future, qui dépend de la vitesse du changement climatique futur.

Pour toutes ces raisons, une approche purement statistique de ce que l'on appellera les *périls climatiques* n'est pas adaptée quand on considère une période de temps longue

(que ce soit dans la projection ou dans l'historique), et il est nécessaire de se reposer sur une modélisation physique des phénomènes sous-jacents au péril. Cette modélisation n'entre pas dans les compétences strictement actuarielles; elle peut par exemple être effectuée par les sociétés développant et commercialisant les logiciels de modélisation des catastrophes naturelles. L'objet de ce mémoire est de présenter, développer et appliquer une méthode s'appuyant sur les résultats en données ouvertes d'une modélisation physique effectuée par un projet de recherche du programme PESETA IV.

PESETA IV est la 4^e itération du [programme européen PESETA](#) dont l'objet est l'évaluation des risques climatiques en Europe. Il vise à éclairer les politiques climatiques de l'Union Européenne en fournissant une analyse scientifique quantitative de long terme. Depuis l'achèvement de PESETA IV en 2020, la 5^e itération est actuellement en cours mais les résultats ne sont pas encore disponibles. Le programme se décline en 12 domaines-clés, celui qui nous intéressera ici concerne les inondations de plaine (ou inondations fluviale).⁷ Les données que nous utilisons fournissent, pour toute l'Europe et par région administrative, des projections de dommages pour différents niveaux de réchauffement global, dans différentes simulations physiques complètes (depuis les émissions de gaz à effet de serre jusqu'aux dommages entraînés par les débordements de cours d'eau).

Pour utiliser ces données, nous nous reposons sur un modèle simple proposé par Jewson, Maynard et Dottori dans un article de recherche [\[JMD23\]](#) associé à un service pour les organismes d'assurance. Ce modèle postule essentiellement qu'à enjeux assurés constants et hors variabilité, le logarithme des dommages évolue dans le temps comme une fonction linéaire du niveau de réchauffement global. Les données scientifiques peuvent alors être utilisées pour calibrer ce modèle, en estimant le degré d'influence du réchauffement global sur les dommages dans chaque région. Dans un deuxième temps, le modèle calibré peut ensuite être utilisé pour projeter les dommages futurs, par exemple en corrigeant l'historique par une « mise en *as-if* » qu'on peut aussi appeler une « actualisation climatique ». L'idée est par exemple, quand on utilise dans une estimation un sinistre historique, reflet du climat passé, de le rééchelonner par le facteur de réchauffement global pour obtenir un sinistre de même temps de retour, mais dans le climat actuel ou futur. Tout comme les dommages passés exprimés en Euros courants doivent être actualisés de l'inflation pour refléter les engagements actuels, dans le cas d'un péril climatique il convient d'« actualiser du réchauffement climatique ».

L'intérêt de cette approche est qu'elle est à la fois simple, puisque la mise en œuvre ne demande pas d'effectuer la modélisation physique, et pertinente physiquement puisqu'on ne se passe pas de cette modélisation, on exploite simplement celle qui a été faite par les gens dont c'est le métier. On dispose d'une grande souplesse pour l'appliquer, de façon très globale (comme on le verra dans les projections de tendance) ou pour corriger un historique et des projections construites par une autre méthode (par exemple par un logiciel de place). Cette seconde approche ne pourra ici qu'être évoquée, faute d'accès à

⁷D'autres domaines-clés sont d'intérêt pour l'assurance, mais d'une part il nous a fallu nous concentrer sur un sujet, et d'autre part les données que nous exploitons ne sont pas disponibles dans ces autres domaines.

des données assurantielles ni à ces logiciels.

Ce mémoire reprend donc de [JMD23] le modèle et l'idée d'utiliser les données PE-SETA IV, mais se distingue de cet article de plusieurs façons. On y développe notamment un traitement statistique propre, qui nous amène à utiliser des données scientifiques différentes, et semble plus à même de quantifier précisément les incertitudes. On propose également une façon d'inclure les incertitudes sur le réchauffement climatique, et on applique le tout dans le cas de la France métropolitaine pour obtenir les projections à comparer aux trois études [Fra22], [CCR23] et [CR22]. Enfin, une méthode détaillée est proposée pour appliquer l'actualisation climatique dans une entreprise d'assurance, pouvant permettre d'inclure les effets du réchauffement climatique dans la tarification, le provisionnement et la politique de réassurance, ou de mener un stress-test climatique dans le cadre de l'ORSA.

Avant de présenter l'organisation générale du présent mémoire, discutons brièvement quelques mémoires récents sur le sujet. L'ensemble des mémoires consultés utilisent une approche d'apprentissage automatique suivant un même schéma : on part d'un historique de la variable de sinistre et de différentes variables prédictives pour sélectionner un modèle (forêts aléatoires, GLM, etc.), puis on applique ce modèle à des variables prédictives futures disponibles dans des bases de données ouvertes, alimentée par des modèles climatiques. Notre approche se distingue principalement par les données scientifiques utilisées, qui sont exprimées en montants de dommages plutôt qu'en données météorologiques, notamment après application d'un modèle hydrologique ; ce qui nous amène à un modèle statistique « classique » plutôt qu'à de l'apprentissage automatique.

Dans [Boy22], Boyeau considère les inondations fluviales (de plaine) et pluviales ; les variables de précipitation n'y apparaissent pas influencer le risque dans le cas fluvial, et le développement sur l'impact du réchauffement climatique se concentre donc sur le cas pluvial. On peut conjecturer que cette surprenante absence d'influence dans le cas fluvial découle de données de précipitations trop localisées dans le temps et dans l'espace, ne prenant pas en compte l'ensemble du bassin versant. Dans [Bro21], Brousse parvient à obtenir une corrélation entre inondations et des variables de précipitation agrégées sur 7 jours ; la variable prédite est le nombre d'arrêtés de reconnaissance « Cat-Nat », ce qui présente des limites : à taille d'évènement égale, les zones découpées en de nombreuses petites communes vont faire l'objet de plus d'arrêtés que des zones urbaines dense, alors que les dommages économiques pourront être plus importants dans le deuxième cas. Elle projette, dans le cas le plus pessimiste de la trajectoire d'émission de GES RCP-8.5, à une augmentation possible d'environ 40% à l'horizon 2050 du nombre d'arrêtés Cat-Nat. Enfin [Eme23] utilise un portefeuille MRH réel pour la variable à prédire, et se place également dans la trajectoire RCP-8.5. Il obtient une augmentation d'environ 80% du montant de sinistre annuel moyen pour la période 2020-2050 par rapport à la période 1998-2020, majoritairement lié à une augmentation de la fréquence des évènements.

Le chapitre I présente le contexte en détaillant la notion de péril climatique, notamment le cas particulier des inondations de plaine. Il sera important d'expliquer comment le réchauffement global influe sur ces inondations, puisque le modèle repose sur un lien

direct, et de décrire la façon dont les données scientifiques que nous utilisons ont été obtenues, afin d'en comprendre les qualités et les limites.

Le chapitre [II](#) introduit de façon détaillée le modèle d'actualisation climatique et développe les méthodes statistiques nécessaires pour le mettre en œuvre : lissage des températures et traitement des incertitudes, calibration du modèle. Il se conclut sur les résultats obtenus à l'aide des données scientifiques.

Enfin, le chapitre [III](#) est consacré aux applications actuarielles : d'une part la projection de tendance à l'horizon 2050 que nous comparons à [\[Fra22\]](#), [\[CCR23\]](#) et [\[CR22\]](#) ; d'autre part à l'échelle d'une compagnie d'assurance, l'estimation à 1 an et la projection de long terme.

I. Réchauffement climatique et péril inondation

Ce premier chapitre introduit le contexte du travail présenté dans le mémoire. On décrit des principes essentiels et le vocabulaire du réchauffement climatique et des inondations qui sont utiles pour comprendre et interpréter le modèle au centre de notre étude, qui sera développé au chapitre suivant.

1. Réchauffement climatique

Commençons par présenter quelques informations concernant le réchauffement climatique. Tout d'abord, on peut mentionner qu'on peut préférer le terme de *changement climatique*, voire de *bouleversement climatique*. En effet, si le réchauffement global est un des éléments important des changements causés au climat par l'action humaine depuis le début de l'ère industrielle, il peut cacher d'autres éléments extrêmement importants comme par exemple l'effondrement de la biodiversité ou, pour ce qui nous occupera ici, l'aggravation d'évènement climatiques extrêmes dépassant largement la seule question des températures. Toutefois, nous utiliserons ici le terme de *réchauffement climatique* du fait que c'est bien le niveau de réchauffement de l'atmosphère qui sera au centre de notre modèle, et il paraît utile de la garder constamment en tête.

Après avoir présenté quelques mécanismes du réchauffement climatique et le vocabulaire associé, ce qui permettra plus tard d'éclairer les acronymes bien connus mais peu compris du type « RCP-4.5 », nous présenterons quelques éléments sur la façon dont ce réchauffement est lié à des inondations plus fréquentes et plus sévères. Enfin, pour conclure la section, nous décrirons pourquoi l'horizon qu'on peut raisonnablement considérer est relativement limité ; dans ce mémoire, les projections seront ainsi limitées à 2050.

1.1. Un tout petit peu de physique du globe

Dans cette section, on présente quelques principes physiques qui sous-tendent le réchauffement climatique. Son but essentiel est d'introduire deux concepts importants pour la suite, le *forçage radiatif* et le *niveau de réchauffement global* ou GWL (« Global Warming Level »).

(i) Bilan et forçage radiatif

Le « bilan radiatif » de la Terre compare les flux d'énergie entrant (dû au rayonnement solaire) et sortant (dû au rayonnement solaire réfléchi et au rayonnement infra-rouge

émis par la Terre), vu non pas depuis le sol mais depuis l'extérieur de l'atmosphère. Ce bilan intègre plusieurs phénomènes, dont l'*effet de serre* : certains gaz présents dans l'atmosphère, dont le CO₂ ou gaz carbonique, ont la faculté d'absorber les rayonnements infra-rouge, et de réémettre l'énergie obtenue dans toutes les directions (on parle de *gaz à effet de serre* ou GES). Ainsi, une partie du rayonnement infra-rouge émis par la terre est « récupéré » plutôt que d'être envoyé dans l'espace.

Quand le bilan radiatif est à l'équilibre, le flux net d'énergie est nul et l'énergie globale de la Terre reste constante. Comme les conversions d'énergie (mécanique ou chimique en chaleur, par exemple) et les mouvements de chaleur entre la surface et l'intérieur sont négligeables, la température globale à la surface de la Terre est alors également essentiellement constante. C'était la situation à l'ère préindustrielle, du moins au premier ordre (disons sur quelques dizaines d'années).

Quand le bilan radiatif est déséquilibré, le flux net d'énergie n'est plus nul, et la température moyenne de surface change : c'est la situation actuelle, l'augmentation de la quantité de gaz à effet de serre dans l'atmosphère depuis le début de l'ère industrielle ayant considérablement renforcé l'effet de serre, le bilan radiatif de la Terre est positif et sa température de surface augmente.

Pour quantifier cette évolution, on mobilise la notion de *forçage radiatif* qui est l'écart, par rapport à la situation pré-industrielle, dû à un facteur externe (comme la concentration atmosphérique de CO₂), de flux d'énergie net (entrant moins sortant) par unité de temps et de surface (il se mesure donc en W/m²).

Un point essentiel à retenir est qu'à court terme, une valeur du forçage radiatif correspond plus à une *vitesse* de réchauffement qu'à un *niveau* de réchauffement, puisqu'il quantifie un excès d'énergie entrante qui va venir réchauffer la Terre. À plus long terme, le réchauffement global va venir augmenter le rayonnement infra-rouge de la Terre, et ainsi progressivement rééquilibrer le bilan. *In fine* le forçage radiatif se traduira donc par un niveau de réchauffement correspondant à ce nouvel équilibre. Dans la plupart des trajectoires envisagées dans le 6^e rapport du GIEC,¹ cet équilibre est loin d'être atteint à notre horizon 2050 et on gagne donc à penser en première approximation en terme de vitesse.

(ii) Niveau de réchauffement climatique (GWL)

Un paramètre central dans le modèle que nous allons étudier est le *niveau de réchauffement climatique* (GWL pour *Global Warming Level*). Il décrit la température moyenne à la surface du Globe, et est exprimé en écart par rapport à une période de référence, en degrés Celsius °C.

En pratique, une telle température moyenne est délicate à définir et encore plus à estimer. Il est possible de s'intéresser à la température des océans par exemple, mais nous considérerons ici la température de l'air au niveau de la surface terrestre (*Global mean surface air temperature* ou GSAT). Notons que la précision « de surface » n'est

¹Groupe d'experts intergouvernemental sur l'évolution du climat, en Anglais IPCC pour *Intergovernmental Panel on Climate Change*

pas cosmétique, certaines couches hautes de l’atmosphère atteignant de températures de plusieurs milliers de degrés.

En ce qui concerne la période de référence, on parle souvent « par rapport à l’ère pré-industrielle » mais cette dernière peut faire l’objet de diverses convention. Ici, nos données sont relatives à la période 1850–1900 ; toute comparaison avec des données de sources différente doit prendre en compte une éventuelle différence dans la période de référence. Un changement de période de référence introduit une translation des valeurs, ce qui en pratique dans notre modèle n’aura aucun impact, mais il convient d’être particulièrement prudent quand on combine des valeurs issues de sources différentes.

(iii) Causes et effet du réchauffement climatique

Le dernier rapport du GIEC en date [IPCC23] s’ouvre sur une déclaration forte et sans ambiguïté : le réchauffement climatique a pour cause l’activité humaine, principalement par l’augmentation de la teneur atmosphérique en gaz à effet de serre (GES).

On entend encore parfois relativiser la cause anthropique du réchauffement climatique, par exemple en attribuant une partie de ce réchauffement à une phase particulièrement intense de l’activité solaire. En fait, plusieurs éléments autres que l’activité humaine ont bien des effets sur la température terrestre, certains à la hausse et d’autres à la baisse, mais ces effets sont d’un ordre de grandeur tellement inférieur à celui du réchauffement anthropique qu’ils sont éclipsés par les marges d’incertitudes.

C’est dans les différentes façons que l’espèce humaine affecte le climat qu’une variété significative apparaît : émissions directes de gaz à effet de serre par la combustion d’hydrocarbures et de charbon ou par les animaux d’élevage, émissions indirectes et diminution des captations naturelles de CO₂ par le changement d’occupation des sols (déforestations notamment), etc.

Les effets de ce réchauffement sont beaucoup trop nombreux pour être tous mentionnés, mais donnons une idée de leur variété : il affecte les écosystèmes, dans les océans comme sur les terres émergées, le cycle de l’eau (et donc à la fois les inondations et la disponibilité de l’eau fraîche), toutes sorte d’évènements climatiques (inondations donc, mais aussi sécheresses, grêle, etc.), l’agriculture, la biodiversité, le niveau de la mer... Parmi les effets potentiellement importants pour l’assurance, mentionnons enfin la santé humaine (santé mentale, maladies notamment vectorielles, etc.) [CMA+22].

1.2. Lien entre réchauffement climatique et inondations

Tandis qu’on comprends facilement que le réchauffement climatique va entraîner des sécheresses plus nombreuses et plus graves, et que l’augmentation du niveau de la mer va aggraver les inondations côtières, l’amplification des inondations de plaine peut paraître moins évidente.

Dans cette section, nous allons d’abord présenter une classification des inondations, puis aborder les déterminants des *inondations de plaine*, et enfin détailler le lien entre le réchauffement climatique et ces inondations. L’essentiel des informations présentées ici

est tiré du chapitre 11 de la partie I (groupe de travail « The Physical Science Basis ») du 6^e rapport du GIEC (AR6), [SZA⁺21].

(i) Définition et classification des inondations

Une *inondation* est définie comme le débordement d'un cours d'eau ou d'un bassin en dehors de ses limites usuelles, ou l'accumulation d'eau dans des zones qui ne sont pas habituellement immergées [IPC18].

On distingue de nombreux types d'inondations. En croisant [IPC18] et [CCRC], on peut distinguer :

- les inondations de plaine (ou fluviales) (« *river (fluvial) floods* ») ;
- les inondations torrentielles, ou crues éclair (« *flash floods* ») ;
- les inondations par ruissellement (« *urban floods* ») ;
- les inondations pluviales (« *pluvial floods* ») ;
- les refoulements d'égouts (« *sewer floods* ») ;
- les submersions marines (« *coastal floods* ») ;
- les vidanges brutales de lac glaciaires (« *glacial lake outburst floods* ») ;

Nous nous intéresseront dans ce mémoire aux *inondations de plaine*, définies comme le débordement continu et progressif d'un cours d'eau et qui concernent principalement les grands bassins versants. Elles se distinguent donc en premier lieu des inondations torrentielles, qui correspondent à un débordement très rapide et concernent les petits bassins versants, en second lieu des inondations n'impliquant pas de cours d'eau.

(ii) Inondations de plaine : mécanismes et déterminants

D'après la CCR,² les inondations de plaine « ont pour origine des précipitations successives et soutenues, affectant de vastes superficies et entraînant des crues lentes mais généralisées du réseau hydrographique. Elles peuvent aussi résulter de la combinaison de plusieurs facteurs, précipitations importantes et fonte des neiges par exemple (...) » [CCRC].

On comprend donc que les précipitations sont un déterminant crucial des inondations de plaine ; mais il est important de comprendre qu'une inondation de plaine en un lieu donné et à un moment donné peut avoir comme origine des précipitations éloignées dans l'espace et dans le temps. Dans l'espace, car les bassins versants concernés peuvent être extrêmement larges : les bassins de la Seine, du Rhône, de la Loire et de la Garonne couvrent à eux quatre environ 60% de la superficie de la partie européenne de la France. Dans le temps, car un surplus d'eau en amont ne va se traduire par un surplus d'eau en

²Caisse centrale de réassurance

aval qu'après un délai ; dans le cas de l'eau issue de la fonte des neiges, le délai peut être de plusieurs mois.

De plus, le niveau d'un cours d'eau étant en variation constante, il existe un écart entre les niveaux habituels d'un cours d'eau et le niveau auquel une inondation a lieu. Un excès d'eau à un endroit donné du cours d'une rivière ou d'un fleuve ne résulte donc en une inondation que s'il est suffisant pour atteindre un certain seuil. Une fois ce seuil dépassé, si la zone est suffisamment plate l'inondation peut couvrir une zone importante même avec un dépassement modeste.

Une inondation de plaine est donc un phénomène complexe : sa survenance résulte de précipitations en divers endroits dont les effets se trouvent synchronisés en aval, provoquant un cumul suffisant pour dépasser un seuil ; sa gravité dépend de la quantité d'eau cumulée, de la géographie locale, et des constructions présentes.

(iii) Aggravation des inondations par le réchauffement climatique

(iii . a) *Impact sur les précipitations.* Le sixième rapport du GIEC confirme l'impact du réchauffement climatique sur un certain nombre d'événements météorologiques et climatiques extrêmes à l'échelle du globe et de grandes régions [SZA+21]. Cet impact était déjà établi dans les rapports précédents, mais l'AR6 mentionne que les travaux récents accroissent le niveau de confiance dans ces résultats. En particulier, l'augmentation des précipitations extrêmes avec le réchauffement climatique est établie avec un niveau de confiance élevé.

Il est également noté que cette augmentation a lieu de façon significative même pour des accroissements modestes de la température globale : « even relatively small incremental increases in global warming (+0.5°C) cause statistically significant changes in extremes on the global scale and for large regions (high confidence) ». De plus, l'augmentation relative de fréquence est plus forte pour les intensités les plus élevées : l'impact n'est pas uniforme, mais change la forme de la distribution en alourdissant sa queue.

Il convient de distinguer deux périodes de temps : le passé récent, pour lequel il s'agit de savoir si les événements extrêmes ont augmenté et, si oui, si cette augmentation est attribuable au réchauffement climatique ; et le futur, pour lequel il s'agit de savoir si un niveau de réchauffement climatique additionnel se traduira par une augmentation des événements extrêmes. En ce qui concerne le passé, l'AR6 considère qu'il est *probable* (« likely ») que la fréquence et de l'intensité des précipitations extrêmes ont augmenté à l'échelle globale, et à l'échelle de chacun de trois continents dont l'Europe (accompagnée de l'Amérique du Nord et de l'Asie) ; et qu'il est *probable* que cette augmentation soit due principalement au réchauffement climatique d'origine humaine. En ce qui concerne le futur, il prévoit également une augmentation des extrêmes de précipitation, notamment en Europe, avec le réchauffement climatique. Dans le cas d'un niveau de réchauffement de +4°C, les événements de précipitations intenses très rares (observés moins d'une année sur 10) deviendront plus fréquents de façon *quasi certaine* (« virtually certain ») à l'échelle globale, et *extrêmement probable* (« extremely likely ») à l'échelle de la plupart des continents.

Enfin, [SZA+21] note que l'ampleur de l'accroissement des précipitations extrêmes

sera également conditionnée à l'échelle régionale par le niveau de réchauffement de la zone, la circulation atmosphérique et la dynamique des tempêtes.

(iii . b) *L'évaporation, lien entre réchauffement et précipitations.* Il est intéressant de se pencher sur le mécanisme qui relie réchauffement et augmentation des précipitations et de leurs extrêmes. L'eau est présente dans l'atmosphère sous forme liquide (les nuages sont visibles car ils sont constitués de minuscules gouttelettes d'eau liquide, qui dévient et reflètent la lumière) et sous forme gazeuse (la vapeur d'eau, parfaitement transparente). La partie gazeuse est importante dans les transports d'eau au sein de l'atmosphère, notamment depuis les océans vers les terres émergées. À l'interface entre une masse d'eau liquide et l'atmosphère, le mécanisme d'évaporation/liquéfaction tend à créer un équilibre : si l'atmosphère contient peu de vapeur, de l'eau liquide va s'évaporer et enrichir l'atmosphère en vapeur ; inversement si l'atmosphère contient beaucoup de vapeur, une partie va devenir liquide et rejoindre la masse d'eau (de dernier même phénomène est également à l'origine des nuages, et de la pluie quand les gouttelettes formées deviennent trop grosse pour rester en suspension). La quantité de vapeur pour laquelle il y a équilibre est quantifiée par la *pression de vapeur saturante* de l'eau, c'est-à-dire la pression partielle (due à la seule vapeur, sans intégrer la part de pression atmosphérique dues aux autres gaz) à laquelle l'équilibre se fait. Ainsi, plus la pression de vapeur saturante est élevée, plus l'équilibre à l'interface se déplacera vers une atmosphère contenant plus de vapeur d'eau. Enfin, la pression de vapeur saturante dépend de la température, et augmente avec elle (ce qui est parfaitement intuitif : plus la température est élevée, plus l'eau aura tendance à s'évaporer). Ainsi une température plus élevée « permet » à l'atmosphère de comporter plus d'eau (gazeuse) et provoque également plus d'évaporation. Ainsi, une température globalement plus élevée va amener une atmosphère globalement plus chargée en vapeur d'eau. On pourrait penser que la température élevée pourrait également favoriser la rétention de cette eau sous forme gazeuse dans l'atmosphère, donc limiter la pluie. Cependant la température n'est ni uniforme ni constante : il y aura toujours des moments et des endroits où la température sera assez basse pour provoquer la liquéfaction d'une fraction de la vapeur d'eau contenue dans l'atmosphère. Dans une atmosphère globalement riche en vapeur, les conditions pour des précipitations extrêmes seront plus souvent réunies.

[SZA⁺21] précise qu'avec un haut niveau de confiance, les précipitations importantes augmenteront au rythme avec lequel l'atmosphère pourra comporter plus de vapeur d'eau en se réchauffant, à peu près 7% par degré de réchauffement global. À +4°C, il est probable que les précipitations qu'on ne voit qu'une fois tous les 10 ans actuellement verront leur fréquence doubler, et que celles qu'on ne voit qu'une fois tous les 50 ans verront leur fréquence tripler.

(iii . c) *Impact sur les inondations de plaine.* Si l'augmentation des précipitations et de leurs extrêmes est évidemment susceptible de causer une augmentation de tout type d'inondations, les effets locaux sont plus complexes et difficiles à estimer pour les inondations de plaine que pour les inondations torrentielles ou pluviales. L'humidité des sols

et, pour les régions froides, l'accumulation de neige en amont du bassin versant sont les facteurs principaux, mais il faut également citer la morphologie du cours d'eau, les aménagements et systèmes de régulations, l'utilisation des sols, etc.

[SZA⁺21] indique qu'une augmentation significative des pics de débits a été observée dans certaines régions, mais avec une grande hétérogénéité. Cette hétérogénéité est notamment présente en Europe, avec une tendance à la hausse dans le nord-ouest et à la baisse dans le sud et l'est sur la période 1960-2000.

D'autre part les projections basées sur des modèles hydrologiques prédisent qu'une plus grande part des terres émergées verra les inondations de plaine augmenter que diminuer (avec niveau de confiance modéré). Cependant la grande variabilité des résultats obtenus, notamment suivant le modèle hydrologique utilisé, ne permet pas d'obtenir des projections de niveau de confiance élevé.

(iii.d) *Vers une estimation quantitative de l'impact du réchauffement climatique.* Il est possible d'utiliser une approche purement statistique pour estimer les risques d'inondation en régime stationnaire, c'est ce qui est fait quand on parle des niveaux d'un cours d'eau à « temps de retour » fixé. Cependant, notre objectif est d'estimer l'aggravation de ce péril du fait du réchauffement climatique. Il faut donc combiner des projections des précipitations (et éventuellement d'autres phénomènes météorologiques pertinents) et faire un lien entre inondations et variables météorologiques (mesurées dans le passé, ou projetées dans le futur).

Or, faire ce lien de façon statistique est rendu très délicat par la complexité décrite dans la section sur les mécanismes et déterminants des inondations de plaine. Même les approches d'apprentissage automatique, s'il permettent de capter les non-linéarités, vont difficilement pouvoir agréger les variables explicatives pertinentes sans modèle physique. Pour capter l'effet des précipitations dans le bassin versant sans modéliser ce dernier, il serait en effet nécessaire de considérer les précipitations sur l'ensemble du territoire et sur une importante période de temps précédant l'instant où le modèle doit estimer la survenue d'une inondation. Mais intégrer une telle quantité de variables explicatives, les multipliant largement au-delà du nombre de variables à expliquer, augmenterait les problèmes de sur-apprentissage et les coûts de calcul. Cette observation motive l'utilisation de travaux reposant sur une modélisation physique des inondations de plaine, en l'occurrence un module du projet européen PESETA IV, qui sera présenté dans la suite du chapitre. Cependant il faut garder une grande prudence dans l'interprétation des résultats, puisque ces approche cumulent les incertitudes à de multiples niveaux : d'une part la trajectoire d'émissions de gaz à effet de serre et les modèles de circulation globale puis régionale utilisés,³ qui seraient aussi source d'incertitude dans l'utilisation d'un modèle d'apprentissage automatique basé sur des variables prédictives météorologiques ; d'autre part le modèle hydrologique et les modèles traduisant les débits en survenance et intensité des inondations, qui sont spécifique à l'approche par modélisation physique « de bout en bout ».

³Voir la section 2.

1.3. Limite des prévisions à long terme : points de bascule climatique

Parmi les incertitudes mentionnées à la fin de la section précédente, les prévisions climatiques et météorologiques sont particulièrement sensibles à l'échelle de temps considéré : plus on parle de périodes lointaines, plus l'incertitude est grande. Cette augmentation de l'incertitude n'a de plus aucune raison d'être linéaire, modérée, ou même continue : plusieurs *points de bascule climatique* pourraient, à leur franchissement, accélérer le réchauffement climatique ou bousculer de façon radicale les climats dans plusieurs ou même toutes les régions de la planète. L'article de synthèse [WFL⁺23] mentionne plusieurs de ces points de bascule :

- ralentissement ou effondrement de l'AMOC (*Atlantic Meridional Overturning Circulation*, dont le *Gulf Stream* est une des composantes) ;
- relâchement de grandes quantités de méthane depuis les réservoirs marins ;
- augmentation de plusieurs mètres du niveau de la mer suite à la perte des couches de glace du Groenland et de l'Antarctique ;
- changement des écosystèmes des forêts boréales ;
- perturbations graves des moussons tropicales ;
- rupture de la couverture nuageuse des stratocumulus ;
- mort des récifs coralliens tropicaux en eau peu profonde ;
- dépérissement de la forêt amazonienne ;
- perte complète de la couche de glace arctique en été.

Parmi ces éléments, certains pourraient basculer dans le courant du siècle, même pour des trajectoires modérées d'émissions de gaz à effet de serre. Il apparaît notamment que les modèles de circulation globale actuels surestiment la stabilité de l'AMOC, et une étude récente donne des éléments indiquant la possibilité de son effondrement d'ici 2050 [DD23].

Il semble donc illusoire de faire des projections au-delà de quelques dizaines d'années, ou du moins de prendre trop au sérieux de telles projections, tant que nos connaissances sur ces phénomènes ne se sont pas enrichies. Nous nous limiterons donc dans ce mémoire à l'horizon 2050, déjà bien incertain.

2. Simuler le futur : trajectoires d'émissions de gaz à effet de serre, modèles de circulation, modèles hydrologiques

L'objet de cette section est de présenter la chaîne de modélisation qui permet de construire des projections du climat et des inondations futures. Une attention particulière sera portée

aux modèles utilisé par PESETA IV, mais les principes et modèles essentiels se retrouvent dans la plupart des études sur le sujet.

Cette chaîne de modélisation comporte cinq grands maillons qu'on décrira tour à tour :

- trajectoire d'émission de Gaz à Effet de Serre (trajectoire GES) ;
- modèle de circulation global (GCM pour *global Circulation Model*) ;
- modèle de circulation régional (RCM pour *Regional Circulation Model*) ;
- modèle hydrologique ;
- modèle de dommages.

Pour chacun de ces maillons, plusieurs modèles peuvent être utilisés. Une simulation va combiner un choix de modélisation pour chaque maillon, et la diversité de ces choix permet de mener des simulations variées, prenant en compte différentes hypothèses de simulation. En pratique, PESETA IV utilise un seul modèle hydrologique, et la plus grande partie du modèle de dommage est également commune à toutes les simulations. Les différents maillons d'une simulation ne sont évidemment pas indépendants : chacun va utiliser la sortie du maillon précédent parmi ses paramètres d'entrée.

Enfin, en 2.5 on récapitulera l'architecture de cette chaîne de simulation, pour en donner une vue globale.

PESETA IV

Le projet PESETA IV (*Projection of Economic impacts of climate change in Sectors of the European Union based on bottom-up Analysis*),⁴ est un projet de recherche européen destiné à mieux comprendre les impacts du changement climatique sur l'Europe, et comment il est possible de s'en protéger par des politiques d'atténuation et d'adaptation. Il est décliné en domaines-clés (*key areas*), celui qui nous intéresse ici concernant les inondations de plaine (*River floods*) (mais notons d'autres domaines-clé intéressant pour l'assurance, notamment l'agriculture, les sécheresses, les inondations côtières, les tempêtes, etc.). Les travaux du groupe consacré aux inondations de plaine font l'objet de rapports et d'articles de recherche, dont [DMB⁺23] sur lequel nous nous appuyons ; les données produites pour cette recherche sont disponibles librement dans deux jeux de données, nous utiliserons une petite partie de celui intitulé « flood impacts and adaptation strategies under future scenarios. »⁵

La couverture géographique comprends les pays de l'Union Européenne (pour leur parties européenne, les Dom-Tom ne sont pas inclus) et le Royaume-Uni, à l'exclusion de Malte dont la petite taille des cours d'eau n'est pas adaptée à la modélisation utilisée ; nous utiliserons ici uniquement les données concernant la France.

⁴https://joint-research-centre.ec.europa.eu/peseta-projects/jrc-peseta-iv_en

⁵<https://data.jrc.ec.europa.eu/dataset/1019a72f-7b98-4318-8baf-fd9e2c00a92c>

2.1. Trajectoires d'émissions de gaz à effet de serre : RCP et SSP

Le premier maillon de la chaîne de modélisation est centré sur la force de l'effet de serre de l'atmosphère qui pourra être constatée dans les années futures, c'est-à-dire (une fois pris en compte les effets de réflexion lumineuse) sur le *forçage radiatif*, voir la section 1.1.

Le facteur de variabilité essentiel est ici l'activité humaine, et particulièrement la quantité de gaz à effet de serre (GES) qu'elle relâchera. Cette variabilité est une inconnue qui relève de l'incertain non probabiliste, et le GIEC a développé des trajectoires⁶ d'émissions de GES qui représentent un large éventail des futurs envisageables. Chacune de ces trajectoires repose sur des hypothèses socio-économiques : suivant que le changement climatique est plus ou moins vigoureusement pris en compte dans la marche de l'économie mondiale, à travers des comportements de ses agents ou des réglementations, les émissions de GES peuvent varier énormément.

Le 6^e rapport du GIEC introduit des trajectoires sensiblement révisées par rapport à celles du 5^e rapport, on passe des *RCP* (« Representative Concentration Pathways ») aux *SSP* (« Shared Socio-economic Pathways ») ; cependant les RCP sont encore largement dominantes dans la littérature sur le changement climatique. Nous allons donc présenter ces deux jeux de trajectoire.

(i) Trajectoires du 5^e rapport du GIEC : RCP

Les quatre trajectoires principales publiées dans le cinquième rapport du GIEC [IPCC14] s'intitulent RCP-2.6, RCP-4.5, RCP-6.0, RCP-8.5 et correspondent à quatre évolutions possibles des déterminants essentiels du forçage radiatif, notamment la concentration atmosphérique de GES (mais aussi par exemple les modifications d'utilisation des terres). Le nombre x dans RCP- x correspond à la valeur du forçage radiatif en 2100 ; ainsi la trajectoire RCP-4.5 postule un forçage radiatif de $4,5 \text{ W/m}^2$ à l'horizon 2100.

Ces trajectoires sont construites en cohérence avec des scénarios variés de mesures d'atténuation, particulièrement strictes pour RCP-2.6, intermédiaire pour RCP-4.5 et RCP-6.0, et un scénario prévoyant des émissions de GES très élevées pour RCP-8.5. Le 5^e rapport du GIEC considérait le scénario de référence, correspondant à une continuation de la trajectoire présente (en 2014) sans effort de réduction des émissions, se situait entre les trajectoires RCP-6.0 et RCP-8.5 ; voir cependant ci-dessous la réévaluation effectuée dans le 6^e rapport.

On présente souvent ces trajectoires en référence à un niveau de réchauffement global, estimé à l'aide de modèles de circulation globale (voir 2.2 ci-dessous). Ainsi la trajectoire RCP-2.6 correspond à la mise en place des efforts nécessaire pour maintenir le réchauffement global en-dessous de $+2^\circ\text{C}$ avec probabilité élevée, soit des émissions de GES fortement diminuée en 2100, aux environs de 2 Gt/an .

⁶On utilisera dans tout ce mémoire le terme de « trajectoire » pour traduire l'anglais « pathway », notamment en distinction avec le terme « scénario » qui sera utilisé par ailleurs. On aurait pu utiliser « profil représentatif » qui est la traduction choisie par le 5^e rapport du GIEC, mais qui a semblé un peu lourd.

Ces trajectoires sont encore couramment utilisées, mais ayant été développée antérieurement à la publication du 5^e rapport du GIEC en 2014, elles sont basées sur des connaissances et des données datées de plus d'une décennie, et ne correspondent plus à l'état de l'art en la matière. On leur préférera autant que possible des trajectoires plus récentes.

(ii) Trajectoires du 6^e rapport du GIEC : SSP

Le dernier rapport du GIEC en date a repris le principe des trajectoires de gaz à effet de serre, en les actualisant sous le terme de *Shared Socio-Economic Pathways* : SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 et SSP5-8.5. Dans la dénomination SSP y - x , le nombre y représente les tendances socio-économiques prises comme hypothèse dans la trajectoire tandis que le nombre x garde la même signification que pour les RCP, un forçage radiatif en W/m^2 à l'horizon 2100.

Les trajectoires SSP1 postulent une décroissance des émissions de GES vers des émissions nettes nulles à horizon 2050 ou 2070, la trajectoire SSP2 suppose un maintien des émissions à leurs niveaux actuels, et les trajectoires SSP3 et SSP5 font l'hypothèse d'un accroissement des émissions jusqu'à un doublement à l'horizon 2100 et 2050, respectivement.

En terme de projection du niveau de réchauffement, la trajectoire SSP1-1.9 est la seule qui donne plus de 50% de chances de limiter le réchauffement en-dessous de $+1,5^{\circ}C$, tandis que les niveaux limite de réchauffement varient de $+2^{\circ}C$ pour SSP1-2.6 à $+4^{\circ}C$ pour SSP3-7.0, et encore au-delà pour SSP5-8.5.

2.2. Modèles de circulation globale et modèles de circulation régionale

Les *modèles de circulation globale* (GCM, *Global Circulation Model*) et les *modèles de circulation régionale* (RCM, *Regional Circulation Model*) sont des modèles numériques représentant les caractéristiques physiques (humidité, vent, pression, etc.) de l'atmosphère et leurs évolutions ; il se distinguent l'un de l'autre par leur étendue et leur précision.

L'implémentation numérique d'un modèle doit prendre en compte des contraintes et des coûts variés : quantité de calcul, temps, place mémoire, instabilités numériques, adéquation du modèle au phénomène modélisé, etc. Quand on modélise un phénomène physique se déroulant dans un domaine géographique, on doit passer par l'intermédiaire d'une discrétisation du domaine, qu'on peut appeler *maillage*, qui peut être plus ou moins fin. La finesse de ce maillage est un paramètre qui va jouer sur beaucoup de ces coûts et sur la précision du modèle, et on est amené à faire des compromis. Un principe qu'on retrouve souvent dans cette situation est d'adapter le compromis, en détaillant plus les zones ou les aspects qui sont les plus importants.

Ici, ce principe se traduit par l'utilisation de modèles « emboîtés », le GCM qui modélise le globe entier et dont le maillage est large, et le RCM qui ne modélise qu'une région mais dont le maillage est plus fin. Les simulations numériques sont d'abord menées sur le GCM, puis les résultats obtenus sont utilisés pour « guider »⁷ le RCM, en le

⁷On parle parfois de *forçage* du RCM par le GCM, mais ceci n'a aucun lien avec le forçage radiatif.

contraignant à suivre les résultats du GCM (soit au bord de son domaine, par exemple dans notre cas les frontières Européenne, soit pour tous les points du GCM qui tombent dans le domaine). On obtient ainsi des projections détaillées sur la région d'intérêt, cohérente avec la projection globale du climat.

Il est important de noter que les projections ainsi réalisées, si elle peuvent prendre l'allure de prévisions météorologiques en donnant pour chaque pas de temps et en chaque point température, précipitations, vent, etc., n'en ont pas la nature : il ne s'agit pas de prévision ayant la moindre prétention à l'exactitude au jour le jour, mais de projection dont la pertinence ne peut s'apprécier qu'à une longue échelle de temps (typiquement deux ou trois décennies).

Pour augmenter la robustesse des projections et estimer l'incertitude de modèle, on utilise plusieurs GCM et plusieurs RCM. Dans PESETA IV, 11 couples sont utilisés impliquant 5 GCM et 5 RCM, formant un ensemble de 22 projections puisqu'on a utilisé pour chaque couple deux simulations, l'une dirigée par RCP-4.5 et l'autre par RCP-8.5.

2.3. Modèles hydrologiques

Une fois un scénario détaillé des conditions météorologiques construit à l'aide de modèles de circulation, le rôle du *modèle hydrologique* est de traduire ces données en débits et niveaux dans les cours d'eau. Il faut pour cela modéliser les flux pour les différents cours d'eau, sur l'entièreté de chaque bassin versant ; par exemple, pour pouvoir estimer la hauteur du Rhône à Marseille à une date donnée, il est nécessaire de prendre en compte les précipitations récentes dans les Bouches-du-Rhône, mais aussi les précipitations en amont, dans tous les affluents et les zones de ruissellement qui les alimentent,⁸ aux dates qui correspondent à la durée d'écoulement depuis le lieu de précipitation jusqu'à Marseille. La distance dans le temps et dans l'espace entre la cause (des précipitations importantes) et la conséquence (une inondation) rend difficile de construire une approche purement statistique qui se passerait de cette modélisation hydrologique tout en restant pertinente.

PESETA IV utilise un modèle hydrologique nommé LISFLOOD, qui fournit des projections journalières de débit à une résolution de 5 km. Le fait d'utiliser un seul modèle est l'une des limites de l'étude : nous n'avons aucune mesure de l'incertitude liées au modèle hydrologique, ce qui résulte en une sous-estimation des incertitudes finales.

En pratique, PESETA IV n'utilise pas directement les débits obtenus mais ajuste un modèle pour estimer la fréquence et la sévérité des débits extrêmes. Pour chaque sévérité, un module de LISFLOOD permet enfin de construire une carte de hauteurs d'eau (de résolution 100 m) en prenant en compte la topographie du lieu où se déroule le débordement du cours d'eau.

⁸Le bassin versant du Rhône par exemple recouvre de l'ordre d'un sixième de la superficie de la France métropolitaine plus une partie non-négligeable de la Suisse.

2.4. Modélisation des dommages

On regroupe ici sous le terme de *modèle de dommage* la façon dont les cartes de hauteur d'eau sont traduites en montants monétaires de dommages. L'*exposition* est estimée à l'aide de cartes de densité de population et d'utilisation des sols, la *vulnérabilité* est ensuite estimée à l'aide d'une carte de protection aux inondations et de *fonctions de dommage* spécifiques à chaque pays, qui associent au niveau d'eau un montant de dommages, en prenant en compte l'exposition et le PIB de la région considérée.

PESETA IV utilise des projections socio-économiques développées par différents organismes pour projeter le développement futur et le changement d'occupation des sols, ce qui permet d'obtenir des montants de dommages finaux intégrant une augmentation de l'exposition et de la vulnérabilité. Toutefois, le projet a publié les données à différentes étapes de traitement, notamment des projections de dommages à situation économique et mesures de protection constantes. Ce sont ces données que nous utilisons (fichiers `s07a-out_results_nuts2_clim_ensemble.zip`), puisqu'elles isolent l'impact du réchauffement climatique, en phase avec le modèle développé au chapitre II.

2.5. Architecture globale des simulations PESETA IV

Résumons ce qui précède en récapitulant la chaîne de modélisation qui permet d'aboutir aux données PESETA IV que nous utilisons :

- i. À partir de modèles socio-économiques et politiques, on construit des projections des émissions de gaz à effet de serre futures : les *trajectoires d'émissions de GES* (2 différentes sont considérées) ;
- ii. on utilise ces trajectoires d'émission pour alimenter une modélisation numérique des paramètres météorologiques à l'échelle du globe, le *modèle de circulation globale* ou GCM (5 différents sont considérés) ;
- iii. les résultats obtenus servent de « guide » pour une modélisation numérique plus fine à une échelle régionale, ici le continent européen : le *modèle de circulation régional* ou RCM (5 différents sont considérés, associés à différents GCM pour former 11 simulations pour chacune des 2 trajectoires d'émissions de GES) ;
- iv. ces simulations sont utilisées en entrée d'un *modèle hydrologique* qui fournit des fréquences et sévérités d'inondations, incluant des cartes de hauteurs d'eau en fonction de la topographie locale ;
- v. à l'aide de *fonctions de dommage* ces inondations modélisées sont croisées avec des informations sur l'occupation des sols, le PIB de la région, les mesures de protection en place pour fournir, enfin, des montants de dommages en Euros (constants).

La figure I.1 reprend ces éléments.



FIGURE I.1. : La chaîne de modélisation PESETA IV. En gras les éléments essentiels nommés dans le mémoire.

3. Périls climatiques et catastrophes naturelles

3.1. Périls climatiques

Nous appellerons ici *péril climatique* toute catégorie d'évènements mettant en danger des biens ou des personnes susceptible de faire l'objet d'une couverture assurantielle, et dont la probabilité de survenance ou la sévérité est aggravée par le changement climatique. Les inondations de plaine qui font l'objet du présent mémoire en sont un exemple particulièrement important par son ampleur, notamment en Europe. Les sécheresses en sont un autre exemple, qui s'est aggravé rapidement ces dernières années et concentre donc également beaucoup d'attention.

On distinguera la notion de péril climatique de celle de *catastrophe naturelle*, que nous entendrons comme les évènements affectant négativement des biens ou des personnes qui sont d'origine non-humaine, de grande intensité et limités dans le temps.⁹ Certains périls climatique ne sont pas des catastrophes naturelle, par exemple parce qu'ils sont chroniques ou d'origine au moins en partie humaine (voir des exemples ci-dessous). À l'inverse, certaines catastrophes naturelles ne sont pas des périls climatiques car elle ne sont pas affectées par le réchauffement climatique. Cependant, il existe un grand nombre de périls qui répondent simultanément aux deux définitions.

De plus, la France dispose d'un régime assurantiel spécifique « Cat-Nat », mais qui ne concerne que certaines catégories de ce qu'on pourrait vouloir appeler des catastrophes naturelles au sens ci-dessus. Dans la suite de cette section, nous allons présenter rapidement ce régime Cat-Nat, puis nous discuterons des exemples variés de périls pour détailler dans lesquels de ces trois cadres ils s'inscrivent.

3.2. Le régime des catastrophes naturelle en France

Le régime spécifique français, que nous appellerons ici « Cat-Nat », est un régime d'assurance obligatoire géré au niveau national par la caisse centrale de réassurance (CCR). Instauré en 1982, il est codifié dans [les articles L125-1 à L125-7 du Code des Assurances](#). Sa couverture géographique inclut notamment l'intégralité du territoire métropolitain, que nous considérons dans ce mémoire (mais exclut par exemple la Polynésie Française).

Il couvre exclusivement les biens assurés (par exemple, pour les immeubles à usage d'habitation, par une couverture Multi-Risque Habitation –MRH),¹⁰ pour une variété de périls recouvrant assez largement ce qu'on a appelé ci-dessus les catastrophes naturelles, incluant les inondations, mais excluant par exemple la grêle, les chutes de neige et les tempêtes non-cycloniques.

Afin d'être indemnisé dans le cadre de ce régime, l'évènement doit avoir été reconnu officiellement comme catastrophe naturelle par un arrêté inter-ministériel publié au Journal Officiel, après demande par le maire de la commune. Les critères de reconnaissance ne sont pas définis directement par le Code des Assurances, qui charge une commission technique (commission interministérielle de reconnaissance de l'état de catastrophe

⁹Mais leurs effets peuvent bien sûr être durables.

¹⁰Avec l'ajout récent d'indemnisation pour les pertes d'exploitation ou les frais de relogement.

naturelle) de donner un avis annuel sur les critères utilisés et les conditions effectives d'indemnisation, et de donner un avis sur les demandes de reconnaissance. L'article L125-1 stipule que sont reconnus « les dommages matériels directs non assurables ayant eu pour cause déterminante l'intensité anormale d'un agent naturel », indiquant donc l'intensité comme critère déterminant. On note également le terme « non-assurable », assez surprenant pour la description de ce qui est au fond une couverture d'assurance, fut-elle nationale. Étant donné que la loi du 13 juillet 1982 instaurant le régime Cat-Nat est intervenue dans une situation de sous-couverture des catastrophes naturelles, le terme de « non-assurable » peut donc logiquement s'entendre comme recouvrant les risques que les organismes d'assurance ne proposeraient pas d'assurer devant la taille du risque et la difficulté de diversification (les catastrophes naturelles ayant souvent une portée géographique importante, rendant les risques individuels non-indépendants).

Le financement du régime Cat-Nat est assuré par une surprime sur les contrats d'assurance ouvrant droit à indemnisation au sens de ce régime, notamment 12% des primes dommages et 6% des primes vol et incendie pour les biens autres que véhicules à moteur [CCRa].

3.3. Discussion par péril

Sans rentrer dans une liste exhaustive, discutons quelques périls au regard des définitions ci-dessus :

- les inondations de plaine rentrent dans les deux catégories de péril climatique et de catastrophe naturelle, et sont couvrables par le régime Cat-Nat ;
- de même pour les submersions marines, mais le modèle utilisé dans ce mémoire n'est pas approprié pour les traiter : leur intensité et fréquence est influencée non seulement par le niveau de réchauffement mais aussi par son historique, un niveau modéré mais assez long pouvant entraîner à terme une augmentation importante du niveau de la mer ;
- les sécheresses rentrent également dans les deux catégories de péril climatique et de catastrophe naturelle, et sont couvrables par le régime Cat-Nat., le phénomène de retrait-gonflement des argiles étant notamment un sujet majeur de préoccupation dans la période récente suite à des dommages très importants ;
- les épisodes de grêle particulièrement intense peuvent être qualifiés de catastrophe naturelle, mais ne sont pas couverts par le régime Cat-Nat ; leur qualification comme péril climatique est incertaine ([ANC+21] table TS.5) ;
- les tremblements de terre sont des catastrophes naturelles et sont éligibles au régime Cat-Nat, mais ne forment pas un péril climatique car leur survenance et leur intensité sont indépendantes du réchauffement global ;
- la pollution de l'air urbain est un péril climatique, car son intensité est aggravée par le réchauffement global, mais n'est pas une catastrophe naturelle car son origine est principalement l'activité humaine ;

- les maladies vectorielles (c'est-à-dire transmises par un *vecteur* tel que les moustiques) comme la dengue sont des périls climatiques dans les régions qui deviennent propice aux vecteurs avec le réchauffement, mais ne sont pas des catastrophes naturelles de par leur caractère chronique ;
- un cyclone ou ouragan est une catastrophe naturelle, qui sera éligible au régime CAT-NAT s'il présente des vents cycloniques de vitesse assez élevée.¹¹ La qualification de péril climatique est restée longtemps incertaine, mais les indications d'une aggravation de leur intensité se renforcent, voir par exemple [BCL⁺24].

4. Réflexions sur le rôle de l'assurance face au réchauffement climatique

Travailler sur le thème du réchauffement climatique amène à s'exposer à de nombreux sentiments : difficile de rester impassible devant ce qu'il n'est pas exagéré de qualifier de menace mortelle pour les sociétés humaines actuelles. Le travail mené peut facilement paraître absolument insignifiant, et il vient un moment où on se demande : quoi faire ?

4.1. Actuaire et citoyen

Le penchant naturel de l'(apprenti-)actuaire est de compter combien ça risque de coûter, problème auquel ce mémoire essaye d'apporter une petite contribution. C'est un problème important pour les organismes d'assurance et c'est donc un travail qui mérite d'être mené, mais en regard des enjeux, on ne peut s'en contenter. Quoi faire au sein du monde de l'assurance, donc, au-delà d'estimer des coûts ? Je propose ici une réflexion personnelle sur cette question, qui n'est pas à proprement parler le sujet du mémoire, mais dépasse en importance tout ce que je pourrais écrire qui relève strictement de mes compétences techniques. J'ai conscience que je ne suis pas le premier à me poser cette question, et il y aura sans aucun doute une part de naïveté dans mes propos. Ils essaient d'esquisser une part des possibles qui me semblent envisageable, mais ne visent ni l'exhaustivité ni à qualifier un état de fait actuel que je connais de fait encore assez mal.

Avant tout, il me semble qu'il faut souligner la grande dissonance cognitive que nous entretenons collectivement : jamais le réchauffement climatique n'a été aussi présent médiatiquement, culturellement, et professionnellement (au moins au sein du monde assurantiel). Pourtant, les actes sont eux extrêmement modestes. On trouve quelques actions individuelles (et beaucoup d'injonctions individuelles), quelques actions institutionnelles (dont beaucoup de *greenwashing*), mais rien qui soit à la hauteur des enjeux à part peut-être le développement des énergies renouvelables. L'information est disponible, nous *savons* la gravité de la situation et l'ampleur des changements nécessaires. Nous choisissons de l'oublier non seulement au quotidien, mais aussi dans les moments où nous faisons nos choix collectifs. Le consensus scientifique nous dit que des changements bien

¹¹supérieure à 145 km/h en moyenne sur 10 mn ou 215 km/h en rafales.

plus brutaux vont donc survenir, et ceci à l'échelle de temps d'une vie humaine. Nous allons donc vivre bien pire que ce que nous refusons de faire.

4.2. Trois niveaux d'action

Il me semble qu'il faut envisager trois niveaux d'action : individuel, à l'échelle d'un organisme, et à celle de la communauté assurantielle.

Des critiques s'élèvent contre le fait de faire reposer sur les individus la responsabilité de problèmes institutionnels et systémiques ; elles sont parfaitement fondées, mais notre société est une réunion d'individus, et ce sont des individus qui, ensemble, mais avec des poids différenciés vu les différences de pouvoir, vont décider de notre gestion de la crise climatique.

Tout le monde peut agir à l'échelle individuelle, mais peut aussi, en fonction de sa position, influencer individuellement sur les orientations, à défaut sur la culture, des deux échelles collectives. Ainsi les trois niveaux d'action sont bien à considérer comme relevant de la responsabilité de chacun et chacune, fut-elle partagée – mon propos n'est pas de donner des arguments pour rejeter la responsabilité sur d'autres, bien au contraire.

Les actions que j'envisage ici relèvent essentiellement de l'*atténuation*, non parce que l'*adaptation*¹² serait moins importante, mais parce que cette dernière est (un peu) plus sérieusement considérée¹³ et qu'elle sera d'autant moins difficile à mener que le travail d'atténuation aura été mené.

4.3. Chacun et chacune

À l'échelle individuelle, il est entendu que personne ne peut avoir d'influence significative sur un phénomène aussi massif que le changement climatique. Et pourtant, l'action individuelle est indispensable puisque le problème est le cumul d'une multitude d'influences individuelles. Pour éviter l'aporie, il convient de se demander quelles actions, si elles étaient menées par une fraction significative de personnes dans une situation vaguement semblable à la mienne, aurait un effet bénéfique sensible ? Plus on fait partie d'une population à l'empreinte carbone élevée, plus on trouvera des réponses efficaces. Les travailleurs et travailleuses de l'assurance, en particulier les actuaires, dans un pays riche comme la France, ont donc évidemment des leviers très significatifs aussi bien au niveau individuel que dans leurs collectifs de travail.

¹²Rappelons que l'atténuation regroupe l'ensemble des mesures visant à limiter le changement climatique lui-même, tandis que l'adaptation désigne les mesures prises pour limiter l'impact du changement climatique sur les sociétés humaines.

¹³Depuis la rédaction de cette phrase, la nouvelle ministre de la Transition écologique, énergétique et du climat Agnès Pannier-Runacher a affirmé : « Anticiper ce que serait un monde à +4°C, ce n'est pas être alarmiste, c'est être courageux, c'est être responsable vis à vis des Français et être préparé », confondant on l'espère la crainte d'un réchauffement de 4°C *en France* et la perspective d'un réchauffement *global* de +4°C. Je n'arrive honnêtement pas à décider si cette citation confirme ou infirme ma phrase.

Ici, un obstacle qui m'est apparu clairement lors d'évènements « RSE »¹⁴ est le manque d'information sur les ordres de grandeur, ou de considération pour ces informations lorsqu'elles sont fournies. Il pleut des bons conseils et astuces que je qualifie vulgairement de PSLD, « pipis sous la douche » : trier ses vieux courriers électroniques, bien éteindre les lumières quand on quitte une pièce,¹⁵ etc. Ce qui caractérise un PSLD est qu'il s'agit d'une action ayant très peu d'impact à la fois sur l'empreinte carbone et sur le mode de vie : quelle que soit l'intention, leur fonction est en pratique de donner bonne conscience à peu de frais, ce qui peut rendre néfaste leur promotion en permettant la continuation de l'inaction sur ce qui pèse réellement lourd.

Cette information sur les ordres de grandeur est pourtant facilement disponible, on peut citer les pages <https://nosgestesclimat.fr/>, qui permet de calculer son empreinte carbone et de voir quels sont les secteurs où nous avons des marges d'amélioration, et <https://impactco2.fr/> qui permet de comparer différents gestes. En résumé, pour les cadres aisés français, on peut citer parmi les points essentiels :

- sur les déplacements, limiter les vols en avion et les déplacements en voiture individuelle¹⁶ (on peut par exemple utiliser un vélo pour les trajets de quelques kilomètres même si l'on garde sa voiture pour les trajets plus longs ou les jours de pluie : chaque kilomètre compte, le plus étant le mieux) ;
- sur la nourriture, limiter la viande, particulièrement la viande bovine et ovine¹⁷ ;
- sur le numérique, limiter la consommation d'équipement (faire durer, acheter en reconditionné) puisqu'il constitue 80% de notre empreinte carbone numérique.

Chacune de ces mesures a un impact indéniable sur le mode de vie. Il n'y a pas de baguette magique qui évite la catastrophe et ne nous demande aucun changement significatif. Cependant, un impact sur le mode de vie n'est pas nécessairement une dégradation ou une perte de confort : à l'usage, les déplacements en vélo peuvent se révéler agréables et ils limitent les risques de notre sédentarité sur la santé [SLDN⁺24] ; (re-)découvrir les légumineuses ouvre des horizons culinaires, on peut aller piocher dans les merveilles imaginée par les véganes même si l'on n'en fait pas partie, etc.

¹⁴Responsabilité Sociale et Environnementale, dont la face « sociale » est souvent insuffisamment considérée et la face « environnementale » pas toujours très bien informée.

¹⁵Je ne recommande pas de laisser les lumières allumées, bien sûr, simplement de ne pas considérer que les éteindre est une action sérieuse contre le réchauffement climatique.

¹⁶L'usage d'un véhicule électrique est un progrès par rapport à un véhicule thermique, mais la généralisation posera des problèmes inextricables si les voitures continuent de s'alourdir. Plus un véhicule est léger, plus le moteur électrique est efficace pour réduire l'empreinte carbone d'une façon généralisable. Une voiture électrique peut être donc une piste très intéressante à condition qu'elle soit petite et légère.

¹⁷Notons que le méthane, largement émis par ces animaux d'élevage, est un gaz à effet de serre plus puissant et de plus faible durée de vie que le CO₂, une réduction des émissions de méthane a donc un impact particulièrement rapide.

4.4. Organisations

À l'échelle d'une compagnie ou mutuelle d'assurance, on peut distinguer trois directions d'actions : celles qui sont commune avec les entreprises d'autres secteurs, celles qui concernent les actifs, et celles qui concernent les passifs.

Les actions envisageables quel que soit le secteur d'activité visent à réduire l'empreinte carbone de l'activité propre de l'entreprise. À nouveau, il convient de distinguer les PSD des actions de fort impact. Ainsi, équiper les zones de passages de détecteurs de mouvement pour les lumières est certainement louable, mais d'impact limité car l'éclairage ne représente jamais qu'une part infime de l'emprunte carbone. On peut à l'échelle d'une entreprise reprendre les actions individuelles efficace et les appliquer à l'entreprise (réduction du renouvellement du matériel informatique), inciter les employés à prendre ces mesures (proposition de repas végétariens dans un restaurant d'entreprise), ou simplement rendre les actions individuelles possibles (installation de parking vélo). Il paraît utile de mentionner la question de l'intelligence artificielle : s'il est entendu que l'assurance se doit de mobiliser les outils statistiques les plus pertinents, et qu'elle ne peut donc pas ignorer l'apprentissage automatique, l'usage de l'IA générative mérite d'être questionné vu son coût environnemental.¹⁸ Ces actions, pour importantes qu'elles soient, restent d'un impact limité par la taille des entreprises d'assurance, qui n'ont pas ici un rôle plus majeur que d'autres secteurs. Les actions dont il sera question à présent bénéficient d'un effet de position beaucoup plus spécifique.

Les actions sur les actifs sont à la fois spécifiques au secteur banque-assurance, et très discuté actuellement notamment sous le terme de « finance verte ». Il s'agit de réorienter les investissements vers les projets et entreprises de moindre empreinte carbone. Ainsi les projets susceptibles d'accentuer le réchauffement climatique, comme le développement de nouvelles extractions d'hydrocarbures, devraient être réduits faute de financement ; dans le même temps, les projets allant dans le sens de l'atténuation, comme le développement d'énergies renouvelables ou du reconditionnement, devraient être facilités par une plus grande disponibilité du capital. Les difficultés sont cependant nombreuses : un projet hydrocarbure promettant des profits aussi massifs que les émissions de gaz à effet de serre qu'il induira peut trouver son financement même si une proportion importante des acteurs de marché s'interdisent d'y participer : les acteurs moins scrupuleux peuvent tirer avantage de se spécialiser dans ces projets, et le niveau de financement peut rester élevé à moins d'un très fort consensus vert des marchés financiers. On pourrait penser que le financement de projets atténuant le réchauffement climatique est, lui, plus simple ; mais j'ai pu entendre un membre du *shift project* expliquer comme il était difficile de convaincre les hauts décideurs de l'assurance d'y consacrer quelques pourcents des près de 2 000 milliards d'Euros d'investissements en assurance-vie. Identifier et dépasser les obstacles à une amplification des mesures de finance verte apparaît donc comme un enjeu significatif.

La « décarbonation des passifs d'assurance » est soumise à des contraintes potentiellement encore plus fortes. En effet, les choix d'assurés qui amplifient le réchauffement climatique ne sont pas directement liées aux risques qui s'ensuivent, ce qui rend difficile

¹⁸ Auquel s'ajoute son coût humain souvent caché mais non négligeable, et son manque de fiabilité.

d'utiliser un levier tarifaire : par exemple une accentuation du risque d'inondation peut difficilement être facturée sur les contrats tout-risque auto ! Quelques leviers pourraient être mobilisés là où l'empreinte carbone est associée directement ou indirectement au risque assuré, par exemple en assurance auto le poids de la voiture. On peut penser, encore en assurance auto, aux produits tarifiant en fonction des distances parcourues ; pour obtenir un impact important, il faudrait à la fois qu'ils soient très répandus et que chaque kilomètre effectué sans la voiture ait un impact visible sur la prime. Un autre exemple pourrait être une assurance santé-vélo, offrant une bonne couverture contre les risques liés à la pratique du vélo tout en reconnaissant dans la prime l'effet préventif sur divers troubles (cardio-vasculaires notamment). Ce sujet est d'actualité, on peut consulter par exemple [Leb24].

4.5. Tous et toutes ensemble

Enfin, considérons les actions qui peuvent être entreprises à l'échelle collective, en tant que communauté de l'assurance et notamment grâce aux structures telles que France Assureurs et l'Institut des Actuariers. Évidemment, tout comme chaque organisme peut faciliter et inciter les actions individuelles vertueuses, les structures collectives de l'assurance peuvent sensibiliser et inciter leur membres dans la même direction (ce qu'elles ne manquent d'ailleurs pas de faire : j'ai été agréablement frappé de voir que le changement climatique fait l'objet d'articles conséquents dans *tous* les numéros de l'Actuariel).

Un point qui me paraît absolument central est la question de l'influence politique. En effet, l'assurance est le seul domaine de ma connaissance qui réunisse à la fois une forte compréhension des énormes enjeux que représente le changement climatique, et un haut niveau de reconnaissance de la part du personnel politique. Les chercheurs et chercheuses qui travaillent sur le changement climatique, malgré des réalisations aussi importantes et d'aussi large audience que les rapports du GIEC, désespèrent de l'imperméabilité globale du personnel politique à leurs alertes. L'existence du réchauffement climatique et son origine humaine sont maintenant à peu près reconnu (du moins en France), mais l'ampleur et l'urgence du défi auquel nous faisons face est extraordinairement minoré en pratique. Il me paraît donc crucial de consacrer une énergie importante à peser au sein du jeu politique, en sachant que les forces qui plaident pour une continuation de la trajectoire actuelle ne sont pas en reste. Un travail d'influence auprès de partenaires peut-être moins convaincus de l'importance et de l'urgence du problème, comme le secteur bancaire ou le secteur industriel, mérite également certainement d'être mené vigoureusement. Un exemple d'action en ce sens est le rapport [SBB⁺25], à condition qu'il soit non seulement lu, mais qu'il entraîne aussi des actions.

Nous en savons assez pour pouvoir affirmer que les sociétés actuelles et les vies humaines vont changer de façon importante au cours des dizaines d'années à venir. La question est de savoir si nous arriverons à une évolution à peu près contrôlée, ou à une série de ruptures catastrophiques à mesure que le changement climatique et l'effondrement de la biodiversité causeront des vagues de mortalité, des déplacements de population sans précédent, l'effondrement des systèmes d'assurance ou d'autre composants essentiels de notre fonctionnement actuel. L'enjeu n'est pas moins élevé, et s'il est

nécessaire de l'oublier au quotidien pour continuer à avancer, il faut aussi régulièrement le regarder en face et en tirer les conséquences.

II. Modèle d'actualisation climatique

Ce chapitre introduit et analyse le modèle d'« actualisation climatique » qui est au cœur du mémoire.

Le problème auquel il s'agit de répondre est le suivant : comment introduire dans une estimation actuarielle une évolution qui n'est pas la simple prolongation dans le temps d'une tendance qu'on pourrait estimer avec les données du passé ? Dans une situation stationnaire, on utilise les données du passé pour estimer les paramètres du phénomène qui nous intéresse ; ces paramètres étant constants dans le temps, on peut les utiliser pour proposer une distribution probabiliste des événements futurs. Dans le cas non-stationnaire, ceci n'est plus possible. Si on postule que l'évolution future peut se déduire, d'une manière ou d'un autre, à celle observée dans le passé (par exemple, une augmentation de $r\%$ par an, avec r un paramètre inconnu mais indépendant du temps), il est possible d'utiliser des outils statistiques un petit peu plus développés pour estimer les paramètres supplémentaires nécessaires à la projection dans le futur (r dans cet exemple).

Dans le cas des périls climatiques, cette approche n'est pas adaptée : le « conducteur » essentiel de l'aggravation de la sinistralité n'est pas directement le temps, mais l'évolution des structures climatiques globales, notamment au sein de l'atmosphère. Or cette évolution n'est pas directement reflétée par le passé, d'une part du fait de la complexité du système climatique et de ses boucles de rétroaction, d'autre part car elle dépend de façon cruciale de l'action humaine actuelle et future, qu'on ne peut pas simplement rapporter à l'action passée.

L'approche que nous allons utiliser repose sur une double hypothèse. En premier lieu, nous supposons que le système étudié est *quasi-ergodique* [HS14], c'est-à-dire que derrière la variabilité de court terme importante des observations se cache une tendance sous-jacente « lisse », à variations relativement lentes et régulières. Pour expliquer le nom, signalons qu'un système stationnaire est dit *ergodique* quand l'espérance d'une observation à un moment donné (la moyenne sur l'espace des issues, parfois appelée moyenne en espace) est égale à la moyenne des observations à travers le temps. Ici, le système n'est pas stationnaire donc on ne peut pas parler d'ergodicité ; la quasi-ergodicité décrit une situation où des moyennes temporelles assez longues permettent d'extraire la tendance, même non-constante, en moyennant la variabilité instantanée.

En second lieu, nous supposons que cette tendance dans le temps repose essentiellement sur le niveau de réchauffement global, et notamment pas sur la trajectoire qui a mené à ce niveau de réchauffement ; on considère par exemple qu'un réchauffement de $+2^{\circ}\text{C}$ atteint en 2050 ou en 2030 aura le même impact (mais, bien sûr, cet impact sera réalisé plus ou moins tôt). Cette hypothèse est raisonnable pour les inondations de plaine

[JMD23, MAD⁺20].¹

Toujours est-il qu’il nous faut donc à la fois utiliser des estimations et projections de niveau de réchauffement, et modéliser comment un niveau de réchauffement donné impacte la sinistralité du péril considéré. C’est l’objet du modèle présenté ici.

1. Présentation du modèle log-linéaire

Le modèle d’« actualisation climatique » au centre du mémoire est un modèle simple, essentiellement introduit dans [JMD23] au détail près qu’on intègre ici dans le modèle les enjeux assurés plutôt que de travailler à enjeux constant. L’apport principal de cette section sera de donner une construction du modèle à partir d’un petit nombre de principes, afin de limiter l’apparence d’arbitraire qui pourrait sinon s’en dégager.

Ce modèle contiendra des quantités de deux types : des paramètres, que nous devrons dans la suite ajuster sur des données – ce sera la partie la plus riche d’un point de vue statistique – et des arguments, qu’il faudra obtenir à partir de données climatiques et actuarielles. Dans ce premier temps on ne s’intéresse pas aux valeurs de ces quantités, seulement à la forme du modèle.

1.1. Cahier des charges

Pour atteindre notre but de construire des projections de sinistralité pour les inondations, plusieurs approches sont possibles. Une approche par apprentissage automatique, comme dans [CJA22] pour les sécheresses, pourrait par exemple être envisagée, mais rencontrerait de grandes difficultés (voir le paragraphe (iii.d) de la section 1.1, page 11). Nous allons ici partir du principe que la sinistralité (passée et future) répond à un modèle comportant aussi peu de paramètre que possible pour répondre aux principes suivants :

- i.* la sinistralité est représentées à une granularité fixe aussi bien en temps qu’en espace (on discutera plus loin les choix de granularité utilisés ici) ;
- ii.* on suppose la sinistralité proportionnelle aux enjeux assurés, le coefficient de proportionnalité comportant un facteur de risque et un facteur d’aléa représentant la variabilité en temps court ;
- iii.* le facteur de risque ne dépend du temps qu’à travers le niveau de réchauffement global ;
- iv.* l’effet relatif d’un incrément de réchauffement global est indépendant du niveau de réchauffement (en particulier le point de référence choisi, préindustriel ou année courante, ne change ni les résultats ni la forme du modèle) ;

¹Pour certains périls, notamment les submersions marines, le niveau de réchauffement global à un instant t ne suffit pas : un même niveau constant peut par exemple conduire à une dégradation de la situation suite à la fonte progressive de la glace continentale venant alimenter la hausse du niveau de la mer, et il faut donc considérer l’intégralité de la trajectoire menant à un niveau donné.

- v. le facteur d'aléa permet de prendre en compte des dépendance en espace, mais on suppose l'indépendance entre les différents temps.

Comme nous allons le voir, ces hypothèses vont fixer la forme du modèle. Elles sont discutables, mais dans l'approche présente où l'on cherche à éviter une modélisation physique complète à l'échelle d'un portefeuille de risques assurés et où les données climatiques sur lesquelles asseoir les estimations sont peu nombreuses et comportent inévitablement de grandes incertitudes, elles sont assez naturelles. L'hypothèse *iii* est par exemple cohérente avec la façon dont le changement climatique impacte principalement les inondations, à travers l'accroissement de la quantité de vapeur d'eau circulant dans l'atmosphère (voir la section 1.2 du chapitre I), et l'hypothèse *iv* est une forme d'approximation au premier ordre, qui se justifie donc sur des échelles de réchauffement suffisamment courtes.

1.2. Construction du modèle

En notant t la variable de temps (de granularité annuelle) et r (comme « région ») la variable d'espace, on doit donc exprimer des montants de dommages $Dom_{r,t}$. Puisqu'on suppose la proportionnalité aux enjeux assurés, en notant ces derniers $EA_{r,t}$ on peut donc écrire

$$Dom_{r,t} = EA_{r,t} \cdot R_{r,t} \cdot F_{r,t}$$

où $F_{r,t}$ est la partie « facteurs aléatoires » et où il reste à élucider la forme de $R_{r,t}$, la partie « facteur de risque ». Par hypothèse, celle-ci est de la forme

$$R_{r,t} = f(r, \text{GWL}(t))$$

où f est une fonction à déterminer qui vérifie pour tout niveau de réchauffement w et tout incrément δw :

$$f(r, w + \delta w) = B_r(\delta w) \cdot f(r, w)$$

où $B_r(\delta w)$ est une grandeur qui dépend de r et de l'incrément, mais pas du niveau de réchauffement w . Montrons que (sous des hypothèses très générales) la seule forme possible pour f est alors

$$f(r, w) = A_r e^{\beta_r w}$$

où A_r, β_r sont des coefficients ne dépendant que de la région. En passant au logarithme on a en effet

$$\log f(r, w + \delta w) = \log B_r(\delta w) + \log f(r, w)$$

et en supposant f dérivable en sa deuxième variable, on a donc :

$$\begin{aligned} \frac{\partial \log f}{\partial w}(r, w) &= \lim_{\delta w \rightarrow 0} \frac{\log f(r, w + \delta w) - \log f(r, w)}{\delta w} \\ \frac{\partial \log f}{\partial w}(r, w) &= \lim_{\delta w \rightarrow 0} \frac{\log B_r(\delta w)}{\delta w}. \end{aligned}$$

Cette limite doit donc à la fois exister (au vu du membre de gauche) et être indépendante de w (au vu du terme de droite), ainsi $\log f$ est affine en son deuxième argument, précisément ce qu'il s'agissait de démontrer.

En injectant ce résultat dans l'expression de la sinistralité, on obtient la forme précise du modèle :

$$Dom_{r,t} = EA_{r,t} \cdot A_r e^{\beta_r \text{GWL}(t)} \cdot F_{r,t}$$

où $F_{r,t}$ est toujours la partie aléatoire représentant les fluctuations dues aux éléments qui échappent à la modélisation, A_r est un facteur d'exposition hors changement climatique, et β_r représente la susceptibilité de la région au changement climatique.

En notant $E_{r,t} = \log F_{r,t}$ et $\alpha_r = \log A_r$, on abouti au modèle log-linéaire

$$\log \frac{Dom_{r,t}}{EA_{r,t}} = \alpha_r + \beta_r \text{GWL}(t) + E_{r,t}. \quad (\text{II.1})$$

Les enjeux assurés $EA_{r,t}$ seront, suivant les cas d'usage, soit fournis dans les données actuarielles (pour le futur on pourra utiliser un plan de développement) soit supposés constants (par exemple pour la calibration du modèle, où on utilisera des données en environnement économique fixé).

Les coefficients α_r (ou de façon équivalente les A_r), qui représentent les facteurs de risque spécifiques du portefeuille assuré considéré hors réchauffement climatique, seront évalués à partir de données historiques ; on voit que nos hypothèses amènent à les supposer constants dans le temps, c'est une des restrictions du modèle. Il n'est cependant pas difficile d'adapter l'ensemble du travail développé ici en ajoutant un facteur de risque dépendant du temps, il suffira de l'agréger aux enjeux assurés (la difficulté étant plutôt de l'estimer).

Les coefficients β_r , qui représentent la susceptibilité de chaque région au réchauffement climatique, seront estimés à partir de données issues d'une modélisation climatique, hydrologique et de dommages (voir la section 3 pour l'estimation, et le chapitre I section 2 pour l'origine des données) ; vu la variabilité de cette estimation, ils pourront être considérés comme aléatoires, et nous estimeront également leur dépendance entre paires de régions. La position de β_r dans la formule lui donne un rôle très semblable au taux d'actualisation en convention de taux d'intérêt continu ; on pourra y penser comme à un *taux d'actualisation climatique*.

Enfin, le facteur aléatoire $E_{r,t}$ sera souvent pris gaussien, notamment dans l'estimation des β_r . L'avantage de cette hypothèse est de permettre de traduire avec le minimum de paramètres, et de façon naturelle, la dépendance entre régions (qui ne peut pas être négligée). Dans l'application à des données actuarielles, ceci pourrait poser des problèmes si des régions ont certaines années des sinistralités nulles, ou très faibles, voir la section 4.1, point (iii). Nous serons donc amenés à adopter des approches variées pour ce facteur dans les différentes modélisations statistiques.

2. Niveau de réchauffement global passé et futur

Dans cette section, nous utilisons des données issues du dernier rapport du GIEC pour modéliser la fonction GWL qui donne le niveau de réchauffement global en fonction du temps, d'une part pour une période « historique » et d'autre part pour une période de

(a) Global surface temperature change relative to 1850–1900

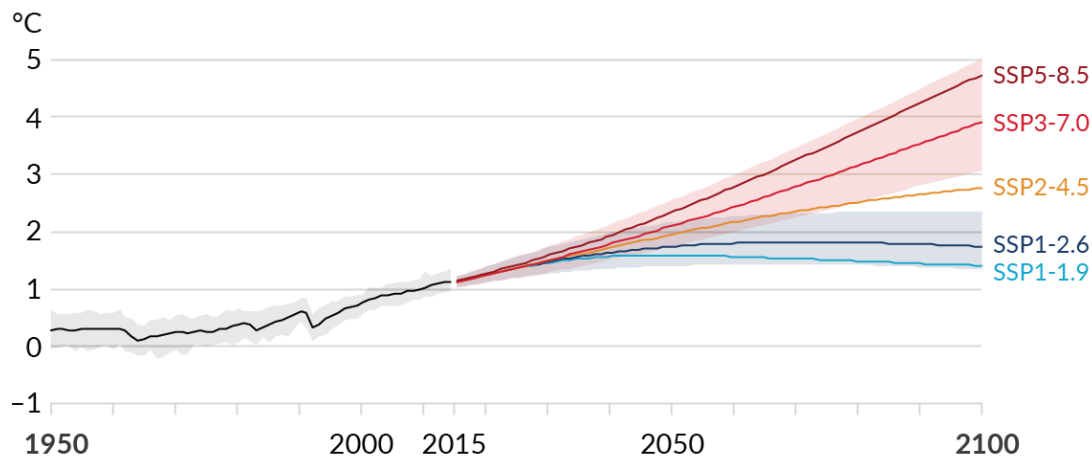


FIGURE II.1. : Extrait de la figure SPM.8 du 6^e rapport du GIEC. Les courbes indiquent les valeurs moyennes, les zones colorées sont les intervalles interquartiles [5%; 95%] pour les trajectoires SSP1-2.6 et SSP3-7.0.

projection ; dans la période de projection, le niveau de réchauffement dépend prioritairement de la trajectoire d'émission de gaz à effet de serre considérée (voir la section 2.1 du chapitre I).

2.1. Données utilisées

On utilise les données de températures [FFKKG21] utilisées dans le sixième rapport du GIEC (figure SPM.8 (a) du résumé pour décideurs, reproduite en figure II.1), qui sont facilement disponibles et couvrent largement la période qui nous est utile.

Chaque température représente l'écart, pour la température atmosphérique moyenne à la surface de la Terre (« Global Atmospheric Temperature »), entre la moyenne sur une période de référence « pré-industrielle » 1850–1900, et la moyenne sur l'année considérée.

On distingue deux périodes : la période « historique » 1950–2014 où les températures sont issues de mesures, et la période 2015–2099 où les températures sont des projections, établies pour chacune des trajectoires d'émission de gaz à effet de serre SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 et SSP5-8.5 (voir la section 2.1 du chapitre I pour la description des trajectoires). On aimerait idéalement avoir des mesures pour la période 2015–présent, mais notons que les écarts entre trajectoires y restent modérés (inférieurs à 0,1°C).

Dans chaque cas, on dispose d'une température moyenne (« Mean ») et deux quantiles, à 5% et 95%, intégrant donc une incertitude. Si on comprend immédiatement les incertitudes sur les températures projetées, il convient de noter que les températures passées ne sont pas parfaitement connues : les mesures sont effectuées en certains points du globe à certains instants, et il faut donc modéliser le globe pour extrapoler les mesures

et obtenir une quantité globale.

Enfin, les données projetées sont lissées sur 20 ans, tandis que les données historiques ne sont pas lissées. Puisque nous devons utiliser des données cohérentes, il nous faudra trouver une méthode de lissage qui ne soit pas redondante, et évitant les artefacts à la jonction entre les deux périodes. Idéalement, on utiliserait les données projetées non-lissées (par exemple voir la figure 4.2 (a) page 571 de [LMB⁺21]), mais celles-ci ne sont pas disponibles sur <https://catalogue.ceda.ac.uk/>.

2.2. Lissage des données

Comme on l’a vu (figure II.1), les données de température moyenne de surface présentent une variabilité de court terme importante dans la période historique ; dans la modélisation, il paraît pertinent de chercher à séparer une tendance (qui sera utilisée pour $\text{GWL}(t)$) de variations conjoncturelles, et donc de lisser les données.

Le cadre général du lissage est celui d’une série de données disponible $(d(t))_t$ (le temps t pouvant être discret ou continu) qu’on postule être composée de deux termes, une tendance décrite par une fonction f « assez régulière », que l’on souhaite estimer, et une variabilité de court terme donnée par une fonction ε qui varie brusquement autour de 0,² que l’on souhaite « effacer » :

$$d(t) = f(t) + \varepsilon(t).$$

On cherche alors une méthode permettant de construire, à partir de la seule connaissance de d , une fonction du temps $\hat{f}$ qui approche bien f . Nous allons appliquer un lissage aux trois séries disponibles (moyenne et quantiles à 5% et à 95%).

Une première approche par moyenne glissante sera présentée en raison de sa simplicité et du fait qu’elle est utilisée dans les données utilisées pour la période de projection. Utiliser une méthode semblable sur les données historiques seules poserait des problèmes au bord droit (2014) puisque nous ne disposons pas des données historiques postérieure pour établir la moyenne ; nous testons un lissage par fenêtre glissante sur la concaténation avec les valeurs projetées, mais ni le principe ni le résultat ne s’avèrent complètement satisfaisants. Une deuxième approche sera utilisée : le lissage de Whittaker et Eilers [Eil03, Bow23].

(i) Moyenne glissante

Parmi les nombreuses méthodes de lissage, la moyenne glissante est sans doute la plus simple ; pour éviter des détails techniques dans les calculs, on fera ici une modélisation en temps continu. Le lissage par moyenne glissante (centrée, ici) consiste à choisir une

²On retrouve ici l’hypothèse de *quasi-ergodicité*.

« demi-largeur » h et d'utiliser comme estimation de f la fonction donnée par

$$\begin{aligned}\hat{f}(t) &:= \frac{1}{2h} \int_{-h}^h d(t+u) \, du \\ &= f(t) + \underbrace{\left(\frac{1}{2h} \int_{-h}^h f(t+u) \, du - f(t) \right)}_{\substack{=: \bar{f}(t) \\ \text{erreur sur la tendance}}} + \underbrace{\frac{1}{2h} \int_{-h}^h \varepsilon(t+u) \, du}_{\text{bruit résiduel } \bar{\varepsilon}},\end{aligned}$$

ce qui repose sur deux hypothèses.

D'une part on postule qu'en tout point la moyenne $\bar{\varepsilon}(t)$ du terme de variabilité est sensiblement plus petite que sa valeur ponctuelle $\varepsilon(t)$, et ce d'autant plus que la fenêtre de moyennage est grande. Par analogie avec le théorème limite central, on peut espérer (par exemple quand les corrélations entre temps éloignées sont faibles) $\bar{\varepsilon}(t) = O(1/\sqrt{h})$, et pour espérer une meilleure vitesse de convergence vers 0 quand $h \rightarrow \infty$ il faudrait un phénomène de compensation, une valeur très haute étant probablement suivie de valeur très basses. Si au contraire les bruits sont corrélés à court terme, ce qui est visiblement le cas sur les données historiques (figure II.3, courbe noire), on peut tout de même espérer un terme en $O(1/\sqrt{h})$, mais avec une constante implicite potentiellement très mauvaise.

D'autre part, on postule que la tendance f est suffisamment régulière pour ne pas être trop perturbée par la moyenne. Pour comprendre plus précisément le type de régularité requis, on peut procéder par développement limité; fixons un point de référence t_0 et écrivons

$$f(t) = a_0 + a_1(t - t_0) + a_2(t - t_0)^2 + \dots$$

(où on a donc en vertu des formules de Taylor et en supposant f assez dérivable $a_0 = f(t_0)$, $a_1 = f'(t_0)$, $a_2 = \frac{1}{2}f''(t_0)$, etc.) et traitons par ordre croissant les termes

$$\frac{1}{2h} \int_{-h}^h a_k(t - t_0 + u)^k \, du$$

qui constituent $\bar{f}(t)$.

Le terme d'ordre 0 ne présente aucune difficulté :

$$\frac{1}{2h} \int_{-h}^h a_0 \, du = a_0,$$

et si la tendance est constante ($a_1 = a_2 = \dots = 0$) elle est entièrement non-affectée, $\bar{f} = f$; plus généralement, si ses variations sont faibles $\bar{f}$ restera proche de f pour h assez petit.

Le terme d'ordre 1 est plus intéressant :

$$\begin{aligned}\frac{1}{2h} \int_{-h}^h a_1(t - t_0 + u) \, du &= \frac{a_1}{2h} \left[\frac{(t - t_0 + u)^2}{2} \right]_{-h}^h \\ &= \frac{a_1}{2h} ((t - t_0)^2 + 2(t - t_0)h + h^2 - (t - t_0)^2 + 2(t - t_0)h - h^2) \\ &= a_1(t - t_0)\end{aligned}$$

ce qui est exactement le premier terme de $f(t)$. Ainsi, on a encore $\bar{f} = f$ quand f est affine ($a_2 = a_3 = \dots = 0$). Le point crucial ici est la symétrie de la fenêtre autour de t , prendre une moyenne dans le passé par exemple créerait un effet retard. Dans le cas où le temps est discret, cette remarque incite à prendre une fenêtre symétrique de la forme $\{t-h, \dots, t-1, t, t+1, \dots, t+h\}$, de longueur $2h+1$.

Le terme d'ordre 2, la « convexité », est donc à priori le premier où prendre la moyenne crée un décalage : avec l'annulation des termes d'ordre pair en h , on a

$$\begin{aligned} \frac{1}{2h} \int_{-h}^h a_2(t-t_0+u)^2 du &= \frac{a_2}{2h} \left[\frac{(t-t_0+u)^3}{3} \right]_{-h}^h \\ &= \frac{a_2}{6h} (2 \times 3(t-t_0)^2 h + 2 \times h^3) \\ &= a_2(t-t_0)^2 + \frac{a_2 h^2}{3}. \end{aligned}$$

Le point crucial ici est de remarquer que l'écart $a_2 h^2/3$ entre les termes d'ordre 2 de $\bar{f}$ et f est constant, indépendant de t (encore une fois, la symétrie de la fenêtre sur laquelle on fait la moyenne est cruciale). Or dans notre modèle une constante additive sur GWL (dont le rôle est joué ici par f) ne change rien, quitte à décaler en conséquence les coefficients A_r (ce qui sera alors fait automatiquement dans l'estimation de ces derniers) :

$$A_r e^{\beta_r \cdot (\text{GWL}(t) + \delta)} = (A_r e^{\beta_r \delta}) e^{\beta_r \cdot \text{GWL}(t)}.$$

Ceci est un bénéfice de l'hypothèse que chaque fraction de réchauffement climatique supplémentaire a le même effet relatif quel que soit le niveau de réchauffement considéré, tout comme cette hypothèse retire tout impact au choix de la température de surface de référence.

On en arrive au terme de troisième ordre, le premier à faire apparaître une perturbation effective de la tendance par le moyennage :

$$\begin{aligned} \frac{1}{2h} \int_{-h}^h a_3(t-t_0+u)^3 du &= \frac{a_3}{2h} \left[\frac{(t-t_0+u)^4}{4} \right]_{-h}^h \\ &= \frac{a_3}{8h} (2 \times 4(t-t_0)^3 h + 2 \times 4(t-t_0)h^3) \\ &= a_3(t-t_0)^3 + a_3(t-t_0)h^2. \end{aligned}$$

Cette fois, le terme d'erreur $a_3(t-t_0)h^2$ n'est pas constant, et s'accroît avec h quadratiquement. L'espoir qu'il reste petit repose sur les trois facteurs : a_3 , qui est petit si f est proche d'être quadratique ; $t-t_0$ qui sera souvent de l'ordre de quelques dizaines d'années ; et h^2 qui pousse donc à prendre une fenêtre la plus étroite possible.

Alors qu'on a vu que le terme d'erreur $\bar{\varepsilon}$ est petit quand h est assez grand, pour le terme $\bar{f} - f$ on doit prendre h assez petit. Il s'agit au fond d'un compromis du type biais-variance : un h petit donnera une estimation avec peu de biais mais une grande variabilité, un h grand une estimation avec moins de variabilité mais biaisée. Plus la régularité de f est bonne, plus on peut se permettre un h élevé sans obtenir un biais excessif. Pour les séries climatiques, une durée de 20 à 30 ans est souvent retenue, on a

donc mené cette approche avec $h = 15$; on peut voir sur la figure II.2 que le résultat, malgré une telle largeur de fenêtre, garde de petites irrégularités. De plus, le lissage des données concaténées présente une redondance regrettable avec le lissage des données projetées.

(ii) Lissage de Whittaker et Eilers

Le principe du lissage de Whittaker et Eilers [Eil03, Bow23] est assez simple : on va minimiser la somme de deux termes, l'un quantifiant l'écart aux données (dont on veut rester aussi proche que possible), l'autre l'irrégularité. La balance entre les deux termes sera réglée comme habituellement par un paramètre $\lambda > 0$ à choisir, et la méthode repose sur un paramètre d'ordre entier $k \geq 1$. Ici, le temps t est discret, et on supposera (comme c'est le cas pour nos données) les écarts constants, par convention unitaires : $t \in \llbracket a, b \rrbracket$.

(ii.a) *Description de la méthode.* Un candidat au lissage est une fonction $g: \llbracket a, b \rrbracket \rightarrow \mathbb{R}$. On lui associe d'une part l'écart quadratique total avec les données :

$$Q(g) = \sum_{t=a}^b (g(t) - d(t))^2,$$

d'autre part le terme d'irrégularité :

$$R_k(g) = \sum_{t=a+k}^b (D^k g(t))^2$$

où D est l'opérateur de dérivée discrète :

$$\begin{aligned} Dg(t) &= g(t) - g(t-1) \\ D^2g(t) &= D(Dg)(t) \\ &= Dg(t) - Dg(t-1) \\ &= g(t) - g(t-1) - (g(t-1) - g(t-2)) \\ &= g(t) - 2g(t-1) + g(t-2) \\ D^3g(t) &= D(D^2g)(t) \\ &= g(t) - 3g(t-1) + 3g(t-2) - g(t-3) \\ &\dots \end{aligned}$$

Il y a un trou en début de période, $D^k g$ n'étant définie que sur $\llbracket a+k, b \rrbracket$, mais ceci n'a pas d'influence dans notre cas car nos données sont disponibles largement au-delà de l'intervalle où nous avons besoin du lissage.

En notant $\lambda > 0$ le paramètre d'équilibrage entre les deux termes, on abouti à une « énergie »

$$E_{\lambda,k}(g) = Q(g) + \lambda R_k(g)$$

et la méthode de Whittaker et Eilers consiste à prendre comme données lissées les $\hat{f}(t)$ où $\hat{f}$ est la fonction minimisant l'énergie :

$$\hat{f} = \arg \min_g E_{\lambda,k}(g).$$

En remarquant que les fonctions d, f etc. sont interprétables comme des vecteurs en dimension $m = b - a + 1$, on peut récrire l'énergie

$$E_{\lambda,k}(g) = \|g - d\|^2 + \lambda \|A_k g\|^2$$

où A_k est la matrice de l'opérateur D^k , de format $(m - k) \times m$. L'utilité d'avoir pris des écart quadratique apparaît quand on calcule le gradient de l'énergie,

$$\nabla E_{\lambda,k}(g) = 2(g - d) + 2\lambda A'_k A_k g$$

(où A'_k est la transposée de A_k) et la solution à notre problème de minimisation est donnée par l'équation

$$(I + \lambda A'_k A_k)g = d$$

qu'il suffit de résoudre numériquement. Notons que $A'_k A_k$ est symétrique et positive, donc $I + \lambda A'_k A_k$ est définie-positive et en particulier inversible : il y a une unique solution au problème de minimisation. La symétrie et la présence de nombreux coefficients nuls rend le calcul numérique particulièrement aisé.

La méthode de Whittaker et Eilers présente de plus un avantage très substantiel sur nos données : comme l'énergie à minimiser est locale, la minimiser va pouvoir à la fois lisser suffisamment dans la période historique, assez cabossée, sans affecter outre mesure la partie projetée, déjà très lisse dans les données brutes. On peut ainsi espérer éviter toute perturbation inutile des données futures tout en garantissant une transition douce entre les deux périodes.

Terminons cette présentation de la méthode avec les valeurs extrêmes de λ . Si on laisse λ tendre vers 0, le terme $\lambda R_k(g)$ devient négligeable et l'optimum se rapprochera de l'unique optimum pour le terme $Q(g)$, à savoir les données de départ. Si on laisse λ tendre vers $+\infty$, au contraire le terme $\lambda R_k(g)$ devient dominant et le résultat s'approchera d'un des nombreux optimum pour ce terme, à savoir les polynômes de degré au plus $k - 1$. Ici, du fait de la multiplicité des optimums du terme principal, le terme secondaire $Q(g)$ continuera de jouer un rôle important en sélectionnant parmi les polynômes de degré au plus $k - 1$ celui qui est le plus proche des données (au sens de l'écart quadratique total). Ainsi, pour $\lambda \rightarrow +\infty$, le lissage de Whittaker et Eilers converge vers la régression des moindres carrés polynomiale de degré $k - 1$. Il ne faut donc pas nécessairement craindre un choix de λ trop élevé, les données continuant bien d'être prise en compte dans tout l'intervalle $[0, +\infty[$. Tout au plus faut-il rester à l'affût d'une forme polynomiale si elle s'écarte de la forme à grande échelle des données.

(ii. b) *Choix des paramètres.* Notre choix du paramètre k va être guidé par les calculs menés sur la méthode de la moyenne glissante : on a vu que cette dernière ne crée aucune

distorsion (uniquement une translation, indolore dans notre modèle) pour une tendance polynomiale de degré au plus 2. Nous allons chercher à ne pas perdre cet avantage, et donc ne pénaliser les fonctions candidates qu'à partir de leur dérivées troisième, en considérant le terme d'irrégularité R_3 . On pourrait se dire que prendre k plus élevé serait encore mieux, mais un contrôle sur une dérivée élevée laisse beaucoup de souplesse. On peut risquer un sur-ajustement, et cet aspect va aussi influencer la calibration de λ . Détaillons un peu.

Une première approche inspirée de la physique consiste à faire une analyse dimensionnelle : dans la formule de l'opérateur D , il y a une division implicite par la durée entre les deux valeurs considérée, ici égale à 1 an. Ceci fait de $D^k g(t)$ une grandeur mesurée en $^{\circ}\text{C} \cdot \text{années}^{-k}$, et donc de $R_k(t)$ une grandeur en $^{\circ}\text{C}^2 \cdot \text{années}^{-2k+1}$ (en prenant en considération que la somme est implicitement pondérée par des durées entre points, ici encore de 1 années). Pour pouvoir mener l'addition avec $Q(g)$ qui est en $^{\circ}\text{C}^2 \cdot \text{années}$ on doit mesurer λ en années^{2k} . On peut donc se ramener à un paramètre $\tau := \lambda^{\frac{1}{2k}}$, homogène à un temps quel que soit k . On voit donc que dans le cas $\lambda > 1$, ce qui sera le cas pour des données dont le bruit est d'un ordre de grandeur supérieur à la variation annuelle de la tendance, le paramètre de lissage λ pourrait augmenter très rapidement avec k pour un « même niveau de contrainte ».

Une deuxième approche consiste à analyser le degré exact de souplesse est laissé quand le terme d'irrégularité est de même taille que le terme d'écart quadratique aux données ; pour simplifier, on va considérer ici une fonction g en temps continu, nulle sur $t < 0$, de dérivée k -ième constante, et voir la valeur qu'elle atteint en un temps T en supposant que $Q(g) = \lambda R_k(g) = 10^{-2}T$ (c'est-à-dire qu'on compare à un cas où les écarts $\varepsilon(t)$ sont de l'ordre de $0,1^{\circ}\text{C}$ le long de la période). On a donc $g^{(k)}(t) \equiv \frac{1}{10\sqrt{\lambda}}$ et en intégrant k fois $g(t) = \frac{t^k}{k! \cdot 10\sqrt{\lambda}}$ pour tout $t \in [0, T]$. Ainsi la valeur atteinte en T est $\frac{T^k}{k! \cdot 10\sqrt{\lambda}}$. En considérant que, dans notre cas, une échelle de temps de 10 ans doit être suffisante pour laisser la liberté d'atteindre $0,5^{\circ}\text{C}$ de variation quel que soit le passé (une valeur nettement déjà supérieure aux variations maximales observées) on obtient pour $k = 3$ un ordre de grandeur indicatif de

$$\lambda \simeq \frac{10^6}{36 \cdot 10 \cdot 0,5^2} \simeq 1000,$$

tandis que pour $k = 2$ on obtiendrait $\lambda \simeq 100$, et $\lambda \simeq 4$ pour $k = 1$. Ce calcul est très approximatif, mais permet de réaliser à quel point la valeur de λ ne peut être analysée indépendamment de k . En tout état de cause, il apparaît que pour l'ordre $k = 3$, on peut prendre λ au moins de l'ordre de 1000, et un changement modeste de paramètre (par exemple de la durée caractéristique T) change rapidement l'ordre de grandeur.

Une méthode possible pour calibrer λ en pratique est de s'inspirer de l'apprentissage automatique, et de traiter le lissage comme un modèle prédictif pour mener une validation croisée *Leave One Out* [Bow24]. Le problème est que les bruits ne sont pas indépendants, et même en utilisant l'option de considérer des sous-échantillons éloignés pour y remédier, il n'a pas été possible d'obtenir un résultat satisfaisant avec le paquet Python `whittaker_eilers` de [Bowell](#). On a utilisé une approche plus rudimentaire, qui

consiste à tester plusieurs valeurs de λ et de prendre celle ne laissant ni fluctuations à des échelles de temps trop petites, ni écarts visibles avec les données brutes dans la période lointaine (figure II.2, où on a utilisé la trajectoire SSP1-1.9 car c'est la « moins quadratique » de toutes, celle où les écarts dans le futur sont les plus marqués pour λ élevé). Nous retenons la valeur $\lambda = 10^5$.

(ii . c) *Résultats obtenus* La figure II.3 présente les résultats obtenus à partir des données [FFKKG21] pour les trajectoires d'émissions de gaz à effet de serre en effectuant sur les trois séries (valeur centrale, quantile à 5% et quantile à 95%) le lissage de Whittaker et Eilers avec paramètres $k = 3$ et $\lambda = 10^5$.

Pour la trajectoire SSP1-1.9, série *Mean*, les écarts entre les données de départ et les données lissées sont au maximum d'à peu près $0,025^\circ\text{C}$ sur 2015-2050, et seulement $0,01^\circ\text{C}$ pour la période 2031-2050 : on a donc une bonne adaptation aux données initiales, malgré un lissage fort.

Indépendamment du lissage, on observe quelques éléments intéressants :

- Les trajectoires optimistes (SSP1) sont concaves dans le futur, traduisant un ralentissement du réchauffement climatique, tandis que la trajectoire médiane SSP2-4.5 suggère un réchauffement à peu près linéaire, et les trajectoires les moins optimistes un réchauffement convexe, donc accéléré ;
- comme dans les données brutes, les incertitudes sont moindres vers 2014, et plus élevées aux deux extrémités. L'augmentation des incertitudes dans le futur est facilement attribuable à l'accumulation de variabilité dans le temps, tandis qu'on peut conjecturer que les incertitudes dans le passé résultent d'une moins bonne qualité de mesures (moins de points, méthodes moins précises, etc.) ;
- il n'est probable de respecter la limite de $+1,5^\circ\text{C}$ d'ici à 2050 dans *aucune* des trajectoires, dans SSP1-1.9 le niveau de réchauffement moyen atteindrait $+1,57^\circ\text{C}$ à cette date. Seules les deux plus optimistes (SSP1) attribuent à cet événement une probabilité supérieure à 5%.

2.3. Modèle probabiliste pour les scénarios de température

En-dehors des plus simples calculs de tendance, nous allons prendre en compte les incertitudes décrites par les quantiles inclus dans les données pour l'application du modèle (II.1), en modélisant $\text{GWL}(t)$ par un processus aléatoire dont la réalisation fournira un *scénario de température*. Nous devons pour cela préciser deux éléments : la loi pour chaque année t fixée, en fonction des valeurs lues dans les colonnes « 5% », « Mean » et « 95% », et l'interdépendance entre les différents temps.

(i) À t fixé : loi normale tordue

Puisque nous disposons de peu de paramètres, et en particulier pas d'information précise sur la queue de distribution, nous allons rester sur une modélisation la plus simple

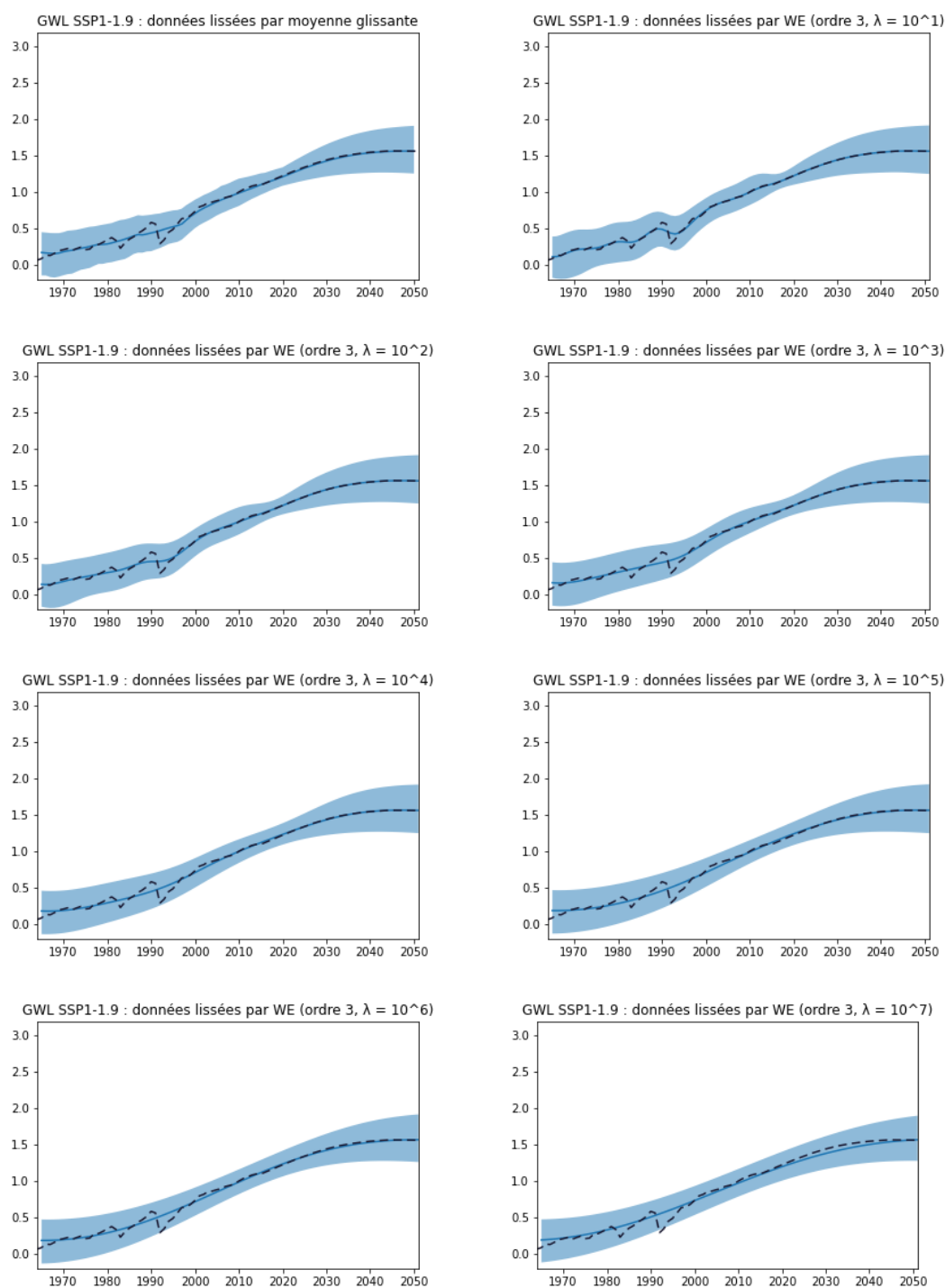


FIGURE II.2. : Comparaison de différents lissages dans la trajectoire médiane SSP1-1.9 : moyenne glissante et Whittaker-Eilers d'ordre 3 avec λ allant de 10 à 10^5 . Les pointillés figurent les données brutes pour la série des valeurs centrales.

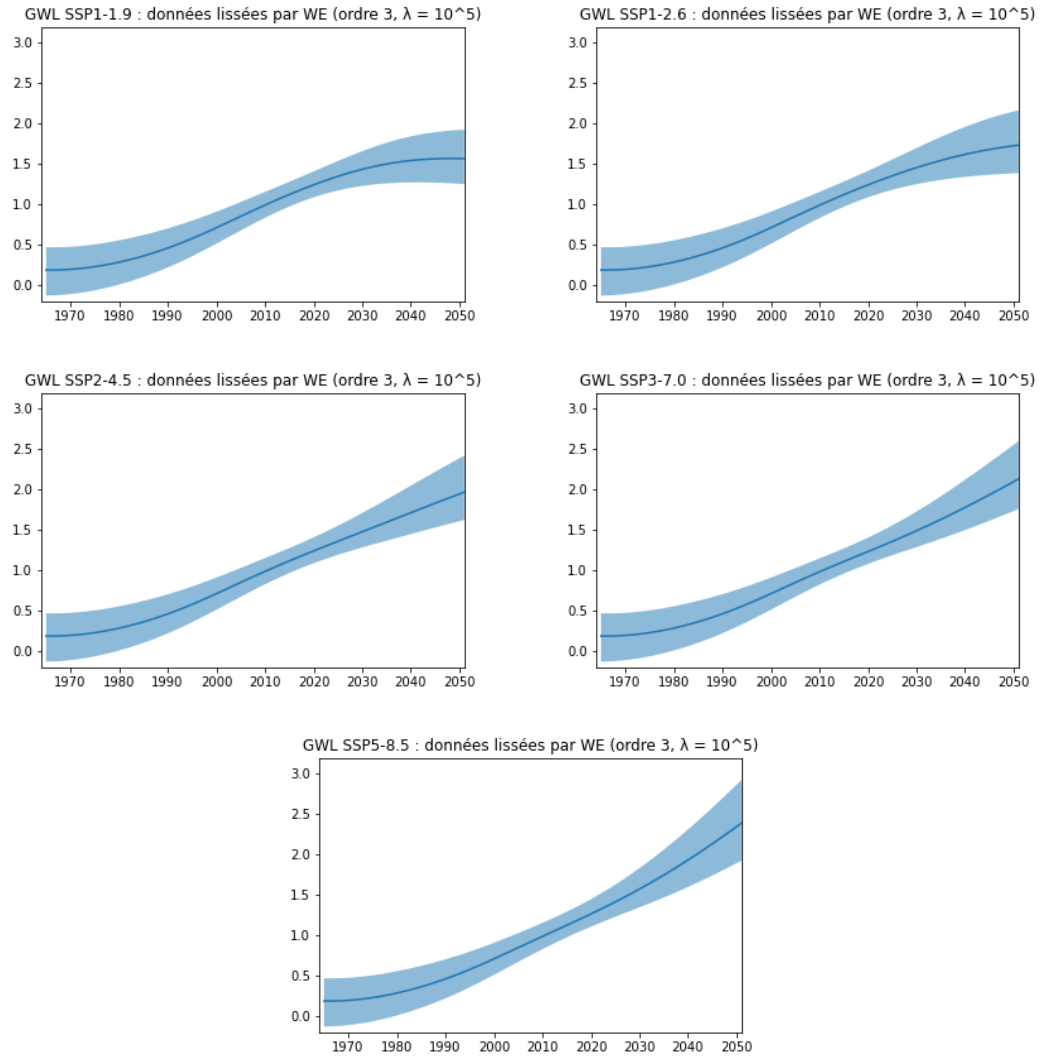


FIGURE II.3. : Tendence du niveau de réchauffement global GWL (valeur centrale et intervalle de quantiles [5%;95%]) dans les différentes trajectoires GES, après lissage des données [FFKKG21] par la méthode de Whittaker et Eilers.

possible compatible avec les valeurs disponibles. En première approche, une distribution normale paraîtrait envisageable, mais elle est trop contrainte : en effet dans les données la moyenne n'est pas située au milieu des deux quantiles à 5% et 95%. Nous allons donc « tordre » de façon asymétrique la loi normale de façon à pouvoir ajuster ces deux quantiles.

Appelons *loi normale tordue* de paramètres m, σ_+, σ_- la loi de probabilité sur $\mathbb{R}$ dont la densité est donnée par :

$$f_{m,\sigma_+,\sigma_-}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_-} e^{-\frac{(x-m)^2}{2\sigma_-^2}} & \text{si } x < m \\ \frac{1}{\sqrt{2\pi}\sigma_+} e^{-\frac{(x-m)^2}{2\sigma_+^2}} & \text{si } x > m \end{cases}$$

On a simplement recollé de chaque côté de m deux moitiés de lois normales d'écart types σ_+ et σ_- . Attention, ici m est la médiane de la distribution, mais pas sa moyenne si $\sigma_+ \neq \sigma_-$. Nous avons besoin de déterminer les paramètres en fonction des données disponibles, ce qui fait l'objet de la proposition suivante.

Proposition 2.1. *Une loi normale tordue de moyenne μ , de quantile à 5% q_- et de quantile à 95% q_+ a pour paramètres :*

$$m = \frac{q_0\sqrt{2\pi}\mu - (q_- + q_+)}{q_0\sqrt{2\pi} - 2} \quad \sigma_- = \frac{m - q_-}{q_0} \quad \sigma_+ = \frac{q_+ - m}{q_0}$$

où $q_0 \simeq 1,645$ est le quantile à 95% de la loi normale centrée réduite.

Démonstration. Nous allons exprimer μ, q_- et q_+ à partir des paramètres puis inverser ces relations. La moyenne μ est la moyenne arithmétique des moyennes des parties gauche et droite de la distribution : en notant X une variable aléatoire de densité f_{m,σ_+,σ_-} , on a

$$\mu = \mathbb{E}[X] = \frac{1}{2}(\mathbb{E}[X \mid X < m] + \mathbb{E}[X \mid X > m])$$

On commence par mener le calcul dans le cas réduit :

$$\mathbb{E}_{X \sim f_{0,1,1}}[X \mid X > 0] = \int_0^{+\infty} x \cdot \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}} \right]_{x=0}^{+\infty} = \frac{2}{\sqrt{2\pi}}$$

de sorte que, par dilatation, translation et symétrie :

$$\mathbb{E}_{X \sim f_{m,\sigma_-,\sigma_+}}[X \mid X > m] = m + \frac{2\sigma_+}{\sqrt{2\pi}}, \quad \mathbb{E}_{X \sim f_{m,\sigma_-,\sigma_+}}[X \mid X < m] = m - \frac{2\sigma_-}{\sqrt{2\pi}}$$

d'où on tire

$$\mu = m + \frac{\sigma_+ - \sigma_-}{\sqrt{2\pi}}.$$

Par ailleurs, on a par dilatation

$$q_- = m - q_0\sigma_-, \quad q_+ = m + q_0\sigma_+.$$

On peut maintenant inverser ces formules ; d'abord on a

$$\sigma_- = \frac{m - q_-}{q_0}, \quad \sigma_+ = \frac{q_+ - m}{q_0} ;$$

puis en injectant dans l'expression de μ :

$$\begin{aligned} \mu &= m + \frac{q_+ + q_- - 2m}{q_0 \sqrt{2\pi}} \\ q_0 \sqrt{2\pi} \mu &= m(q_0 \sqrt{2\pi} - 2) + q_+ + q_- \\ m &= \frac{q_0 \sqrt{2\pi} \mu - (q_+ + q_-)}{q_0 \sqrt{2\pi} - 2}. \end{aligned}$$

□

Pour l'implémentation, on va utiliser cette proposition pour déterminer pour chaque t les paramètres m_t , $\sigma_{-,t}$, $\sigma_{+,t}$ de la loi normale tordue correspondant aux données lissées dans la trajectoire d'émission de GES considérée. La section (ii) va décrire comment on va tirer une suite de déviations $(\delta(t))_t$, mesurée en écarts-types et signée, et la valeur de température du scénario aléatoire pour l'année t sera alors

$$\text{GWL}(t) = \begin{cases} m_t + \delta(t)\sigma_{+,t} & \text{si } \delta(t) \geq 0 \\ m_t + \delta(t)\sigma_{-,t} & \text{si } \delta(t) < 0. \end{cases} \quad (\text{II.2})$$

(ii) Processus temporel : une dépendance complète

La question de la modélisation d'un scénario aléatoire pour les températures est délicate. On souhaite modéliser une tendance, ce qui exclut tout processus irrégulier type Ornstein-Uhlenbeck. Une possibilité est de tirer des quantiles uniformément sur $[0,1]$, indépendants les uns des autres, pour différents temps $t_0, t_1, \dots$ fixés au préalable puis d'interpoler.

Attention, l'interpolation ne peut pas se faire linéairement sur les quantiles. Prenons l'exemple du cas gaussien et des temps $t_0 = 0$, $t_1 = 4$, et supposons qu'on tire des valeurs extrêmes de quantiles comme $q(t_0) = 0,001$ et $q(t_1) = 0,999$, c'est-à-dire que le risque réalisé serait le quantile d'ordre 0,001 en $t = 0$ et le quantile d'ordre 0,999 en $t = 4$. On aurait alors des déviations très fortes en ces deux moments mais en interpolant linéairement l'essentiel de la trajectoire intermédiaire serait proche du centre comme indiqué par la table II.4. Cela pose deux problèmes : d'une part, même des valeurs encore plus extrêmes comme $q(0) = 10^{-10}$ ne créeraient aucune différence sensible dans le milieu du parcours ; d'autre part la variation serait concentrée artificiellement à proximité des temps de contrôle t_i .

Dans le cas de distributions normales ou normales tordues, ce n'est pas une grosse difficulté : il suffit d'interpoler linéairement les déviations plutôt que les quantiles. Ainsi, les variations seront régulières dans chaque intervalle de temps $[t_i, t_{i+1}]$.

Un problème plus délicat est celui du choix des temps t_i . Une possibilité serait de prendre seulement deux temps, t_0 en tout début de la période considérée et t_1 en toute

t	quantile $q(t)$	déviations $\delta(t)$
0	0,001	-3,1
1	0,25	-0,67
2	0,5	0,0
3	0,75	0,67
4	0,999	3,1

TABLE II.4. : Exemple d'interpolation linéaire des quantiles dans le cas gaussien. Les déviations sont exprimées en écarts-types, par rapport à la médiane. Les valeurs en gras sont choisies pour l'exemple, les autres valeurs en sont déduites. On donne deux chiffres significatifs.

fin. Cependant, ce choix a l'inconvénient que la durée de la période influe directement sur la vitesse des variations de température. Une seconde possibilité serait de fixer une durée, par exemple $\delta t = 20$ années, et de prendre des t_i régulièrement espacés de cette durée, en tronquant à posteriori pour se ramener à la période désirée. Cependant, c'est ici le choix de la durée qui va influencer directement sur les variations. Dans tous les cas, les temps intermédiaires, pour lesquels on aura interpolé, auront une variabilité moindre que les temps de tirage aléatoire (par effet de moyenne). Ainsi, on obtiendra des résultats sensiblement différents selon que la fin de période est un temps de tirage, ou que celui-ci est en fait placé postérieurement.

Tous ces obstacles nous conduisent à un choix plus simple : nous allons tirer une seule déviation δ (suivant une loi normale centrée réduite) et l'appliquer à tous les temps t : $\delta(t) \equiv \delta$ dans (II.2). L'avantage est de limiter l'influence de choix arbitraires, l'inconvénient de réduire la variabilité sur $\text{GWL}(t)$. En effet, une déviation largement positive résultera en des températures élevées en fin de période, mais également au début, amoindrissant l'écart entre les deux alors que seul cet écart a un impact dans notre modèle. On retiendra donc que *ce choix de modélisation est plutôt conservatif, sous-estimant possiblement les quantiles les plus élevés.*

3. Estimation des coefficients β_r d'actualisation climatique

Le but de cette section est d'estimer les coefficients β_r traduisant dans chaque région r l'impact du réchauffement climatique sur les inondations de plaine, à partir de données publiques issues du projet européen PESETA IV. L'estimation portera sur une valeur centrale, mais aussi sur l'incertitude de modèle : nous traiterons souvent dans les applications les β_r comme des variables aléatoires, qui seront supposées former un vecteur gaussien, et nous allons donc estimer leurs variances et covariances.

Nous présenterons tout d'abord les données PESETA IV et certaines de leurs limites. Nous développerons ensuite les outils statistiques qui ont été implémentés en Python et appliqués aux données, ce qui formera la section la plus poussée du mémoire d'un point de vue théorique. Enfin nous présenterons les résultats de l'application de ces méthodes

aux données PESETA IV.

3.1. Les données PESETA IV

Le projet PESETA IV a été présenté dans le chapitre I, section 2. Nous allons présenter ici la forme prise par les données que nous utilisons, et leurs principales limites et incertitudes.

Les fichiers que nous utilisons sont disponibles sur *Joint Research Centre Data Catalogue*, jeu de donnée *Flood impacts and adaptation strategies under future scenarios*³, élément *Input-output data for the aggregation of results for future flood risk scenarios*, archive `s07a-out_results_nuts2_clim_ensemble.zip`. Nous ne considérons que les fichiers dénommés `ead_nuts2_TTT_fpXX.csv` où :

- `ead` fait référence aux dommages en Euros (*Expected Annual economic Damage*, par opposition au nombre de personnes touchées) ;
- `nuts2` désigne la granularité géographique, correspondant pour la France aux anciennes régions administratives ;
- `TTT` appartient à l'ensemble `{base, 1.5C, 2C, 3C}` et désigne le niveau de réchauffement considéré ;
- `XX` appartient à l'ensemble `{50, 100, 150}` désignant le niveau de protection mis en place contre les inondations : 100 correspond à la meilleure estimation par les auteurs, 50 et 150 correspondant à des protections estimées moins forte ou, respectivement, plus forte que cette estimation centrale. Nous n'utiliserons qu'une seule de ces trois valeurs, les chiffres n'étant pas du tout indépendants d'une valeur de protection à l'autre puisque l'essentiel de la chaîne de modélisation reste la même.

Ces fichiers présentent dans leur première colonne la liste des régions NUTS2 (pour la France, les anciennes régions administratives⁴) désignées par un identifiant numérique à 8 chiffres ; chacune des 22 colonnes suivantes présente en première ligne un intitulé `rcp4.5` ou `rcp8.5` indiquant la trajectoire d'émission de GES utilisée pour la simulation (11 colonnes pour chacune correspondant aux différentes simulations EURO-CORDEX), puis les valeurs de dommages en Euros (constants, de l'année 2015) projetées dans la simulation, pour le niveau de réchauffement `TTT`, avec le niveau de protection `XX`.

Toutes les étapes de la simulation présentent des limites et des incertitudes ; l'utilisation d'un ensemble de modèles de circulation atmosphérique permet de matérialiser une partie de l'incertitude de cette étape de simulation, contrairement à la partie hydrologique où un modèle unique est utilisé. Les incertitudes les plus significatives sont sans doute au niveau de l'estimation des dommages. En effet, le niveau de protection est difficile à quantifier et les fonctions de dommages présentent notamment une forte sensibilité aux paramètres utilisés. La France fait partie des quelques pays pour lesquels les

³<https://data.jrc.ec.europa.eu/dataset/1019a72f-7b98-4318-8baf-fd9e2c00a92c>

⁴<https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R2066&from=FR>

dommages modélisés sont bien au-dessus des dommages documentés auxquels l'équipe de PESETA IV a eu accès. Il est possible que cet écart soit dû à une insuffisance de la documentation des dommages qui a été utilisée, ou à une sous-estimation des protections contre les inondations mises œuvre. En tout état de cause, on fera l'hypothèse que les écarts sont reproduits de façon semblable aux différents niveaux de réchauffement, et qu'en conséquence ils affectent peu les estimations des coefficients β_r . On retiendra que *les incertitudes que nous mesurons sont certainement sous-estimées*, et on n'utilisera en aucun les projections de dommages PESETA IV directement (c'est-à-dire autrement que par l'intermédiaire des β_r) dans le mémoire : les valeurs de dommages sur lesquelles l'actualisation climatique sera effectuée seront toujours basées sur des données indépendantes.

3.2. Modélisation statistique

Les données que nous utilisons forment un ensemble de valeurs $dom_{rs w}$ représentant le montant annuel de dommages (en Euros 2015) causés par les inondations de plaine dans la région r , pour la simulation s , à un niveau de réchauffement global w . Rappelons que les régions sont les régions administratives du territoire européen de la France antérieures à la réforme 2016, au nombre de 22. Rappelons également que chaque simulation correspond à un triplet comprenant un scénario d'émission de GES, un modèle de circulation global et un modèle de circulation régional, elles sont également au nombre de 22 (ce qui est complètement fortuit). Enfin les niveaux de réchauffement global considérés sont exprimés par rapport à la période 1881–1910, et non 1850–1900 comme dans les données de température que nous utiliserons plus bas. Étant donné que le modèle d'actualisation climatique repose sur des différences de niveau de réchauffement, ceci n'a pas d'impact tant que les différents niveaux utilisés dans la régression sont bien sur la même référence. PESETA IV considère un niveau « de base » correspondant à la période 1981–2010 (soit un niveau de réchauffement d'environ $+0,69^\circ\text{C}$ ⁵), et des niveaux $+1,5^\circ\text{C}$, $+2^\circ\text{C}$ et $+3^\circ\text{C}$. Nous excluons ce dernier, car [JMD23] mentionne que la linéarité dans le modèle d'actualisation n'est plus pertinente sur un écart aussi important. Il peut paraître gênant de mener des régressions linéaires sur trois points, mais c'est de la multiplicité de ces régressions que nous obtiendrons une certaine rigueur statistique (ce point est discuté plus bas). Nous disposons donc de 66 valeurs par région, 1452 au total.

On notera en cas de besoin $\mathcal{R}$ l'ensemble des régions, $\mathcal{S}$ celui des simulations, $\mathcal{W} = \{w_0 \simeq 0,69; w_1 = 1,5; w_2 = 2\}$ celui des niveaux de réchauffement (*warming levels*).

(i) Écart avec l'approche utilisée par Jewson, Maynard et Dottori

Nous allons nous écarter de [JMD23] sur quelques points. Ils utilisent des données moins agrégées, avec pour chaque région et chaque simulation un point par année, et donc des niveaux de réchauffement global beaucoup plus variés. Pour chaque simulation et chaque région, ils effectuent une régression linéaire classique avec comme variable réponse le

⁵Cette valeur a été évaluée à partir du Global Land-Ocean Temperature Index de la NASA, qui couvre comme nécessaire les deux périodes 1881–1910 et 1981–2010.

logarithme du montant de dommages, obtenant ainsi 22 valeurs possibles de β_r pour chaque région r , une par simulation. Ils estiment la valeur centrale pour la région par la moyenne simple de ces 22 valeurs, les variances par région et covariances entre régions étant obtenues par les estimées empiriques.

Un premier reproche qu'on peut faire à cette approche est qu'elle n'est pas basée sur un modèle statistique explicite concernant le lien entre les différentes simulations, ce qui cache un défaut : chaque régression linéaire est entachée d'une incertitude, mais ces incertitudes ne sont pas prises en compte dans l'estimation de la variabilité des β_r , qui risque donc d'être sous-estimée. Nous allons ici nous baser sur un modèle statistique explicite, et chercher à estimer à la fois la variabilité inter-simulations et les incertitudes des régressions au sein de chaque simulation.

Le problème de l'estimation de l'incertitude sur la pente dans une régression linéaire classique (moindre carrés ordinaires sous hypothèse d'homoscédasticité) est bien compris ; toutefois, sa résolution repose sur l'hypothèse que les bruits entachant la variable réponse sont indépendants les uns des autres. Or il paraît peu plausible que, dans une même simulation, les valeurs de dommages dans des années proches (donc reposant sur des hypothèses proches) seraient indépendantes. Les figures de [JMD23] montrent d'ailleurs des points suivant des courbes relativement régulières dans chaque simulation. Si on peut mener la régression et interpréter la pente elle-même, les estimations usuelles de variabilité de la pente ne semblent alors pas interprétables. Dans le cas des données agrégées sur trois niveaux de réchauffement, les recoupements entre les années qui sont agrégées dans deux niveaux différents sont limités et les trois valeurs correspondent à des hypothèses bien moins proches, ce qui permet de rendre l'hypothèse d'indépendance des bruits plausible. De plus chacune des trois valeurs est déjà « stabilisée » par un effet de moyenne, laissant espérer une plus grande précision compensant les fortes incertitudes d'une régression sur peu de points.

Une fois ce principe posé, l'observation des régressions linéaires montre que pour une grande majorité des simulations les trois points sont remarquablement bien alignés, mais que dans un petit nombre de cas l'alignement est bien moins bon. On est alors tenté de faire peser moins ces dernières simulations dans l'estimation, mais comment pondérer de façon appropriée ? Le modèle statistique que nous allons présenter a l'avantage d'intégrer une telle pondération, qui sera justifiée par la plus grande variabilité de la pente dans le cas d'un mauvais alignement.

(ii) Le modèle statistique pour une région

Dans un premier temps, fixons-nous sur une seule région $r \in \mathcal{R}$ et cherchons à estimer une valeur centrale et une variance pour $\beta = \beta_r$; la question importante des corrélations entre régions sera traitée plus loin. Pour alléger les notations, nous allons laisser la dépendance en la région implicite et ne pas utiliser l'indice r ; nous notons ainsi dom_{sw} les dommages estimés dans la région r , dans la simulation s , au niveau de réchauffement w .

(ii . a) *Modèles mixtes et analyse de mesures répétées.* Le problème auquel nous sommes confrontés correspond à une situation de *mesure répétée* : chaque simulation fournit une liste de mesures longitudinales (à températures w_0, w_1, w_2). Un cadre à priori adapté est celui des *modèles mixtes* (voir par exemple [BP15]) qui fonctionnent comme une régression linéaire classique, mais avec certains coefficients aléatoires. Ici, le modèle mixte s'écrit :

$$\log(Dom_{sw}) = (\alpha + V_s) + (\beta + U_s)w + E_{sw}$$

où Dom_{sw} est la variable aléatoire modélisant les dommages observés (qu'on distingue en tant qu'objet théorique de la valeur pratique dom_{sw} , voir l'annexe ; α et β sont les effets fixes (sur l'ordonnée à l'origine et la pente respectivement), inconnus mais déterministes ; V_s et U_s sont les effets aléatoires de chaque simulation, et E_{sw} sont les bruits.

L'idée d'un tel modèle mixte est que chaque simulation porte son propre biais, traduit par V_s et U_s (supposées d'espérance nulle) sur les coefficients de la régression, en plus des bruits E_{sw} pour chaque mesure (supposés également d'espérance nulle). Ainsi, avoir plus de données sur une même simulation permettrait de corriger les bruits en convergeant vers les valeurs $(\alpha + V_s)$ et $(\beta + U_s)$ pour les coefficients de la régression, mais estimer β lui-même nécessite de considérer plusieurs simulations. Ce point de vue est cohérent avec la façon dont nous allons exploiter la valeur centrale et la variabilité estimées de β : nous allons générer des valeurs aléatoires pour ce coefficient, ce qui revient à faire l'hypothèse que le coefficient décrivant la réalité future et ceux observés dans les simulations sont des réalisations indépendantes d'une même loi de probabilité, dont nous souhaitons identifier les caractéristiques.

Il existe une théorie assez bien développée des modèles mixtes, mais principalement sous une hypothèse d'homoscédasticité très forte, en supposant que les bruits E_{sw} sont de même variance pour tout w et tout s . Or si on regarde les régressions dans les différentes simulations pour une région donnée, on observe une grande variabilité de la dispersion des valeurs autour des droites de régression (section 3.3, figure II.10) : les résidus sont grands dans certaines simulations, petits dans d'autres. Un test statistique a été effectué, et permet de rejeter cette hypothèse (point (ii . b) de la section 3.3). On est donc amené à proposer une hypothèse d'homoscédasticité plus faible, suivant laquelle E_{sw} serait de variance σ_s^2 , indépendant de w mais spécifique à chaque simulation s .

On pourrait mener l'étude du modèle mixte sous cette hypothèse, mais ce serait assez délicat et à ma connaissance pas implémentée dans les outils classiques (notamment [SP10]). On a donc choisi une approche légèrement simplifiée que nous allons maintenant décrire.

(ii . b) *Modèle de régressions répétées.* Nous allons traiter chaque régression comme une mesure du coefficient β , entachée d'une erreur propre à la simulation considérée. Il y a donc une succession de deux modèles statistiques : une régression classique, puis un modèle mixte plus simple que le modèle linéaire précédent.

Pour chaque simulation s , on considère ainsi le modèle de régression linéaire classique (toujours avec une « fonction de lien » logarithmique) :

$$\log(Dom_{sw}) = \alpha_s + \beta_s w + E_{sw}$$

où α_s, β_s sont les coefficients de régression inconnus mais supposés déterministes, et $(E_{sw})_{w \in \mathcal{W}}$ est une famille de bruits supposés gaussiens indépendants de même variance σ_s^2 inconnue. On effectue une régression des moindres carrés ordinaires, qui fournit l'estimateur sans biais de β_s , qu'on peut écrire de façon développée :

$$\hat{B}_s = \frac{(w_0 Y_0 + w_1 Y_1 + w_2 Y_2) - \frac{1}{3}(w_0 + w_1 + w_2)(Y_0 + Y_1 + Y_2)}{(w_0^2 + w_1^2 + w_2^2) - \frac{1}{3}(w_0 + w_1 + w_2)^2} \quad (\text{II.3})$$

où $Y_i = \log(\text{Dom}_{sw_i})$ et w_0, w_1, w_2 sont les trois valeurs de réchauffement global (rapelons qu'on distingue dans ce mémoire l'estimateur $\hat{B}_s$, une variable aléatoire dénotée par une capitale, de l'estimation $\hat{\beta}_s$ qui est calculée en remplaçant dans la formule (II.3) les Y_i par les observations y_i). On utilise l'estimateur sans biais de la variance σ_s^2 , qui s'écrit ici :

$$\hat{S}_s^2 = \sum_{i=0}^2 (\hat{A}_s + \hat{B}_s w_i - Y_i)^2$$

où $\hat{A}_s = \frac{1}{3}(Y_0 + Y_1 + Y_2) - \frac{1}{3}(w_0 + w_1 + w_2)\hat{B}_s$ est l'estimateur de l'ordonnée à l'origine. On rappelle que c'est bien la variance qui est ici estimée sans biais, et non l'écart-type (ce qui devrait en toute rigueur amener à dénoter l'estimateur $\widehat{S}_s^2$), et on observe que pour $n = 3$, le facteur de normalisation $\frac{1}{n-2}$ est réduit à l'unité. On a donc ici un facteur 3 entre l'estimateur sans biais et l'estimateur du maximum de vraisemblance ! Heureusement, la multiplicité des simulations va modérer un peu la grande incertitude sur chaque régression individuelle.

Rappelons que, sous l'hypothèse de bruit gaussien, l'estimateur $\hat{B}_s$ suit lui-même une loi normale, comme combinaison linéaire de gaussiennes indépendantes (II.3). Son espérance est β_s et sa variance

$$\tau_s^2 = \frac{\sigma_s^2}{(w_0^2 + w_1^2 + w_2^2) - \frac{1}{3}(w_0 + w_1 + w_2)^2}.$$

L'hypothèse de normalité du bruit est ici cruciale d'un point de vue théorique puisque nous sommes loin de pouvoir invoquer une normalité asymptotique avec $n = 3$; un test de normalité sur les estimations $\hat{\beta}_s$ a donc été effectué (Section 3.3), il n'invalidé pas cette hypothèse.

Nous arrivons à la deuxième étape en considérant que les estimations $\hat{\beta}_s$ sont les réalisations de variables aléatoires

$$B_s = \beta + U_s + E'_s \quad (\text{II.4})$$

où :

- U_s est le biais inhérent à la simulation, supposé gaussien centré de variance σ^2 inconnue mais indépendante de la simulation s . Ainsi la réalisation de $\beta + U_s$ correspond à la pente « réelle » β_s de la simulation s , somme de la valeur de β et du biais spécifique de la simulation ;

- le bruit E'_s est l'écart entre la pente réelle pour la simulation s et l'estimateur de la pente ;
- la famille réunissant les U_s et les E'_s est supposée indépendante dans son ensemble.

Grace à l'hypothèse de normalité des bruits dans la régression, on peut supposer E'_s gaussienne, centrée, de variance τ_s^2 ; en appliquant l'approche *plug-in*, on va dans cette deuxième étape identifier τ_s^2 à son estimation

$$\hat{\tau}_s^2 = \frac{\sum_{i=0}^2 (\hat{\alpha}_s + \hat{\beta}_s w_i - y_i)^2}{\sum_{i=0}^2 w_i^2 - \frac{1}{3} (\sum_{i=0}^2 w_i)^2},$$

de sorte qu'elle peut être considérée comme connue. On utilisera dans la suite la notation τ_s^2 , plus légère, mais dans les calculs pratiques nous utiliseront $\hat{\tau}_s^2$.

Le modèle statistique (II.4) peut également être considéré comme un modèle mixte pour l'estimation d'une grandeur scalaire β . La situation est donc simplifiée par le traitement à part de la partie régression linéaire. Toutefois, nous allons voir que la construction d'estimateurs efficaces pour β et σ^2 n'est pas totalement triviale. Nous allons commencer par traiter le cas où σ^2 serait connu, ce qui amènera à utiliser une moyenne pondérée. Nous traiterons ensuite l'estimation de la variance d'un estimateur par moyenne pondérée. Enfin, nous utiliserons une approche de type *Expectation-Minimization* (EM) pour résoudre les deux problèmes simultanément.

(ii. c) *Cas de la variance connue.* On regroupe pour chaque simulation les deux termes d'erreurs en une seule variable aléatoire $E''_s = U_s + E'_s$, ce qui nous amène à chercher un estimateur de β à partir des variables aléatoires observables

$$B_s = \beta + E''_s \quad (\text{II.5})$$

où, grâce à nos hypothèses d'indépendance et de normalité, les E''_s sont gaussiennes, centrées, de variance $\sigma^2 + \tau_s^2$. On pourrait effectuer une moyenne simple, et on obtiendrait un estimateur sans biais ; toutefois il ne serait pas optimal du fait de l'hétéroscédasticité. Décrivons deux approches pour améliorer cette idée naïve (qui se révéleront coïncider) : chercher le meilleur estimateur linéaire en les observations, et utiliser le maximum de vraisemblance.

Dans la première approche, on se restreint aux estimateurs prenant la forme $\sum_s p_s B_s$ où $(p_s)_{s \in \mathcal{S}}$ est une famille de *poïds* positifs, normalisée : $\sum_s p_s = 1$. Cette condition assure l'absence de biais :

$$\mathbb{E}[\sum_s p_s B_s] = \sum_s p_s \mathbb{E}[B_s] = (\sum_s p_s) \beta = \beta.$$

Parmi ces estimateurs sans biais, on va minimiser la variance :

$$\begin{aligned} \mathbb{V}[\sum_s p_s B_s] &= \sum_s p_s^2 \mathbb{V}[B_s] && (\text{par indépendance}) \\ &= \sum_s p_s^2 (\sigma^2 + \tau_s^2) \end{aligned}$$

Ce problème de minimisation est classique et facile à résoudre : il est convexe et la quantité à minimiser diverge vers $+\infty$ dès que l'une au moins des inconnues tend vers $\pm\infty$, donc il admet un unique point critique, où le minimum est atteint. Or une perturbation de p_s en $p_s + h_s$ donne comme variation de la variance :

$$\mathbb{V}\left[\sum_s (p_s + h_s) B_s\right] - \mathbb{V}\left[\sum_s p_s B_s\right] = 2 \sum_s h_s p_s (\sigma^2 + \tau_s^2) + O(h_s^2)$$

et on voit que prendre p_s proportionnel à l'inverse de $(\sigma^2 + \tau_s^2)$ résulte en un terme de premier ordre de la forme $C \sum_s h_s$. Or la somme des p_s étant contrainte à 1, les perturbations à considérer vérifient $\sum_s h_s = 0$. On a bien trouvé un point critique explicite :

$$p_s = \frac{1}{z} \cdot \frac{1}{\sigma^2 + \tau_s^2} \quad \forall s \in \mathcal{S}, \quad \text{où } z = \sum_s \frac{1}{\sigma^2 + \tau_s^2}. \quad (\text{II.6})$$

La restriction à des estimateurs linéaires peut paraître arbitraire, explorons donc ce que donne l'approche par maximum de vraisemblance. La fonction de log-vraisemblance s'écrit, pour une famille d'observations β_s des variables aléatoires B_s :

$$\begin{aligned} \mathcal{L}(\beta) &= \log \prod_s \frac{1}{\sqrt{2\pi(\sigma^2 + \tau_s^2)}} e^{-\frac{(\beta - \beta_s)^2}{2(\sigma^2 + \tau_s^2)}} \\ &= \text{Constante} - \sum_s \frac{(\beta - \beta_s)^2}{(\sigma^2 + \tau_s^2)} \end{aligned}$$

et on cherche donc β minimisant $\sum_s \frac{(\beta - \beta_s)^2}{(\sigma^2 + \tau_s^2)}$. Il s'agit encore d'un problème convexe dont l'unique point critique est solution, mais cette fois uni-dimensionnel est sans contrainte. On a

$$\begin{aligned} \mathcal{L}'(\beta) = 0 &\Leftrightarrow -2 \sum_s \frac{\beta - \beta_s}{(\sigma^2 + \tau_s^2)} = 0 \\ &\Leftrightarrow \beta \cdot z = \sum_s \frac{\beta_s}{(\sigma^2 + \tau_s^2)} \\ &\Leftrightarrow \beta = \sum_s p_s \beta_s \end{aligned} \quad (\text{II.7})$$

où z et les p_s sont les valeurs définies par (II.3) : on retrouve exactement le même estimateur $\hat{B} = \sum_s p_s B_s$ (ou encore la même estimation $\hat{\beta} = \sum_s p_s \beta_s$).

Pour résumer, on a le résultat suivant.

Proposition 3.1. *Dans le modèle (II.5), l'estimateur $\hat{B} = \sum_s p_s B_s$ dont les poids p_s sont donnés par (II.6) est à la fois l'estimateur du maximum de vraisemblance et l'estimateur linéaire sans biais de plus petite variance.*

(ii.d) *Estimation de la variance d'un estimateur par moyenne pondérée.* Pour utiliser le paragraphe précédent, il nous manque en pratique la variance inter-simulations σ^2 : autant les τ_s^2 peuvent être estimés dans les régressions, autant seules les valeurs observées

β_s nous renseignent sur σ^2 . Nous allons encore considérer deux approches de l'estimation de cette variance, par le maximum de vraisemblance et par variance empirique pondérée ; ici toutefois, les deux approches ne coïncideront pas. Dans les deux cas, nous allons supposer qu'on dispose d'une estimation linéaire $\hat{\beta} = \sum_s p_s \beta_s$ avec des poids p_s connus (et on notera encore $\hat{B} = \sum_s p_s B_s$ l'estimateur correspondant) ; bien sûr à ce stade le serpent se mord la queue, mais ce problème sera résolu dans la suite.

Commençons par le maximum de vraisemblance, appliqué ici à la seule variable σ^2 en fixant β à $\hat{\beta}$. La log-vraisemblance s'écrit :

$$\begin{aligned}\mathcal{L}(\sigma^2) &= \log \prod_s \frac{1}{\sqrt{2\pi(\sigma^2 + \tau_s^2)}} e^{-\frac{(\hat{\beta} - \beta_s)^2}{2(\sigma^2 + \tau_s^2)}} \\ &= \text{Constante} - \frac{1}{2} \sum_s \log(\sigma^2 + \tau_s^2) + \frac{(\hat{\beta} - \beta_s)^2}{\sigma^2 + \tau_s^2}\end{aligned}$$

Le problème de minimisation n'est ici pas convexe, et il n'y a à priori pas de garantie d'existence ni d'unicité de la solution. La condition de premier ordre s'écrit :

$$\mathcal{L}'(\sigma^2) = 0 \Leftrightarrow \sum_s \frac{1}{\sigma^2 + \tau_s^2} - \frac{(\beta - \beta_s)^2}{(\sigma^2 + \tau_s^2)^2} = 0. \quad (\text{II.8})$$

Après mise en pratique, il semble plus stable numériquement d'appliquer un algorithme de maximisation à la vraisemblance plutôt qu'un algorithme de résolution d'équation sur (II.8) ; cependant écrire la condition de premier ordre permet de voir qu'elle peut ne pas admettre de solution. En effet, comme $\hat{\beta}$ se situe entre le plus petit et le plus grand des β_s , si les variations des β_s sont assez petites en regard des τ_s^2 , alors le terme de droite dans la condition (II.8) est plus petit que $\sum_s 1/\tau_s^2$ (par exemple, si les β_s sont tous égaux, ce terme serait nul). Dans ce cas aucune variance positive ne peut réaliser la condition. Mais ce n'est en fait pas une objection solide à l'approche par le maximum de vraisemblance, qui est alors simplement réalisé au bord du domaine, à $\sigma^2 = 0$, traduisant que les τ_s^2 sont déjà plus que suffisants pour expliquer la variation observée des β_s .

Une deuxième objection possible est qu'il paraît incohérent de fixer une valeur de β plutôt que de procéder à une approche par maximum de vraisemblance globale, sur β et σ^2 simultanément. Ce que nous faisons ici est en fait équivalent, grâce à la forme spécifique de la vraisemblance et à notre construction de $\hat{\beta}$. En effet, si on écrit la log-vraisemblance en deux variables $\mathcal{L}(\beta, \sigma^2)$, les conditions de premier ordre $\frac{d\mathcal{L}}{d\beta}(\beta, \sigma^2) = 0$ et $\frac{d\mathcal{L}}{d\sigma^2}(\beta, \sigma^2) = 0$ sont respectivement (II.7) et (II.8). La seule difficulté, déjà mentionnée, est la réalisation simultanée des deux conditions, qu'on peut résoudre par itération comme nous en discuterons plus loin.

Une objection plus substantielle est qu'il paraît très difficile d'estimer le biais sur l'estimateur final de σ^2 par cette approche. En effet, on ne dispose d'aucune expression analytique sur laquelle travailler ; l'approche usuelle serait d'utiliser la dérivée seconde de la log-vraisemblance pour obtenir une correction asymptotique du biais, mais ne disposant que de 22 simulations il paraît délicat de supposer que le régime asymptotique

s'applique. On prend donc le risque de sous-estimer⁶ σ^2 , ce qui résulterait en des sous-estimations des quantiles élevés. Le principe de prudence nous incite à chercher une autre approche, permettant de mieux contrôler le biais sur σ^2 .

Pour notre deuxième approche, nous allons utiliser la variance empirique, avec quelques modifications par rapport à la situation classique. Premièrement, la variation d'une observation β_s est confrontée à notre meilleure estimation de la moyenne, $\hat{\beta}$, et non à la moyenne simple ; deuxièmement, cette variation est corrigée de la variabilité attendue du terme ε'_s (réalisation de E'_s), c'est-à-dire de τ_s^2 ; troisièmement chaque terme correspondant à une simulation s est pondéré non pas de façon égale $\frac{1}{n}$, mais avec le même poids p_s qu'utilisé pour $\hat{\beta}$, puisque ce poids reflète le degré de pertinence que nous accordons à la simulation s . Nous obtenons une première estimation (et l'estimateur correspondant) :

$$\hat{\sigma}^2 = \sum_s p_s ((\beta_s - \hat{\beta})^2 - \tau_s^2) \quad \hat{\Sigma}^2 = \sum_s p_s ((B_s - \hat{B})^2 - \tau_s^2).$$

Attention, ici la majuscule dans $\hat{\Sigma}^2$ signale qu'il s'agit d'une variable aléatoire, mais elle est bien scalaire et non matricielle. On observe tout de suite un problème potentiel : si les τ_s^2 sont élevés mais les β_s sont proches de β , on peut obtenir une estimation négative de la variance ! On retrouve la situation dans laquelle le maximum de vraisemblance était réalisé au bord du domaine, et en toute rigueur on devrait donc considérer l'estimation $\max(0, \hat{\sigma}^2)$ et l'estimateur $\max(0, \hat{\Sigma}^2)$. Toutefois les calculs seraient rendus nettement plus compliqués, et $(\hat{\Sigma}^2 < 0)$ est un événement qui de probabilité très faible dès que σ^2 est assez loin de 0. Nous allons donc conserver les expressions ci-dessus et vérifier lors de l'application et des tests numériques de cohérence du modèle que nous obtenons bien des valeurs positives.

La forme simple de $\hat{\Sigma}^2$ permet d'en calculer le biais explicitement, ce qui va permettre une correction :

$$\begin{aligned} \mathbb{E}[\hat{\Sigma}^2] &= \sum_s p_s \left(\mathbb{E}[(B_s - \beta + \beta - \hat{B})^2] - \tau_s^2 \right) \\ &= \sum_s p_s \left(\mathbb{V}[B_s] + \mathbb{V}[\hat{B}] - 2\mathbb{E}[(B_s - \beta)(\hat{B} - \beta)] - \tau_s^2 \right) \\ &= \mathbb{V}[\hat{B}] + \sum_s p_s \left(\sigma^2 - 2\mathbb{E}[(B_s - \beta)(\hat{B} - \beta)] \right) \\ &= \sigma^2 + \mathbb{V}[\hat{B}] - 2 \sum_s p_s \mathbb{E}[(B_s - \beta)(\hat{B} - \beta)] \end{aligned}$$

⁶Risque usuel en vraisemblance, penser aux facteurs $1/n$ pour l'estimateur du maximum de vraisemblance de la variance contre $1/(n-1)$ pour l'estimateur non-biaisé

or $\mathbb{V}[\hat{B}] = \sum_s p_s^2(\sigma^2 + \tau_s^2)$, et pour tout $s \in \mathcal{S}$:

$$\begin{aligned}\mathbb{E}[(B_s - \beta)(\hat{B} - \beta)] &= \mathbb{E}[(B_s - \beta) \sum_r p_r (B_r - \beta)] \\ &= \sum_r p_r \mathbb{E}[(B_s - \beta)(B_r - \beta)] \\ &= p_s \mathbb{V}[B_s] \\ &= p_s(\sigma^2 + \tau_s^2)\end{aligned}$$

donc

$$\begin{aligned}\mathbb{E}[\hat{\Sigma}^2] &= \sigma^2 + \sum_s p_s^2(\sigma^2 + \tau_s^2) - 2 \sum_s p_s^2(\sigma^2 + \tau_s^2) \\ &= \sigma^2 \left(1 - \sum_s p_s^2\right) - \sum_s p_s^2 \tau_s^2.\end{aligned}$$

On est donc amenés à effectuer une correction affine, et à introduire l'estimateur non biaisé

$$\hat{S}^2 = \frac{\hat{\Sigma}^2 + \sum_s p_s^2 \tau_s^2}{1 - \sum_s p_s^2} = \frac{1}{1 - \sum_s p_s^2} \sum_s p_s \left((B_s - \hat{B})^2 + (p_s - 1) \tau_s^2 \right)$$

et l'estimation associée

$$\hat{s}^2 = \frac{1}{1 - \sum_s p_s^2} \sum_s p_s \left((\beta_s - \hat{\beta})^2 + (p_s - 1) \tau_s^2 \right) \quad (\text{II.9})$$

Il convient de noter que l'incertitude sur β doit être prise en compte en plus de la variabilité inter-simulations ; ainsi l'incertitude (quadratique) sur le coefficient β « réel » sera in fine $\hat{s}^2 + \sum_s p_s^2(\hat{s}^2 + \tau_s^2)$ plutôt que seulement $\hat{s}^2$.

(ii. e) *Itération des estimations de poids et de variance.* Dans les deux paragraphes précédents, nous avons d'abord vu comment estimer β par une moyenne pondérée $\sum p_s \beta_s$ quand σ^2 est connue, et comment estimer σ^2 quand les poids p_s sont fixés. Malheureusement, les données de notre problème laissent simultanément σ^2 et les p_s à déterminer. Nous allons utiliser une méthode itérative du type *Expectation-Minimization* (EM) et effectuer chacune des deux estimations alternativement à partir du dernier résultat obtenu pour l'autre. Il faut fixer un point de départ, et étant donné la possibilité que la meilleure estimation de σ^2 soit au bord du domaine, c'est-à-dire nulle, nous allons prendre $\sigma_0^2 = 0$.

Formellement, en notant $\mathbf{P}(\mathcal{S})$ l'ensemble des familles de poids positifs sommant à 1 indexées par $\mathcal{S}$, on définit à partir des données $(\beta_s, \tau_s^2)_s$ les fonctions

$$\begin{aligned}\Omega: \quad \mathbb{R}_+ &\rightarrow \mathbf{P}(\mathcal{S}) \\ x &\mapsto \frac{1}{\sum_s \frac{1}{x + \tau_s^2}} \cdot \frac{1}{x + \tau_s^2} \\ A: \mathbf{P}(\mathcal{S}) &\rightarrow \mathbb{R} \\ (y_s)_s &\mapsto \frac{1}{1 - \sum_s y_s^2} \sum_s y_s \left((\beta_s - \sum_s y_s \beta_s)^2 + (y_s - 1) \tau_s^2 \right)\end{aligned}$$

et on définit récursivement pour tout $k \in \mathbb{N}$, tant que $\sigma_k^2 \geq 0$:

$$(p_{s,k})_s = \Omega(\sigma_k^2), \quad \sigma_{k+1}^2 = A((p_{s,k})_s).$$

Dans le cas où $\sigma_1^2 = A(\Omega(0))$ serait négatif ou nul, on estimerait logiquement σ^2 par 0 et on prendrait les poids proportionnels aux $1/\tau_s^2$; un problème plus profond apparaîtrait s'il n'y avait pas convergence, par exemple si $\sigma_1^2 > 0$ et $\sigma_2^2 \leq 0$. Il n'est pas évident de déterminer si ce problème peut se présenter, en pratique sur les données PESETA IV on observe une convergence très rapide de σ_k^2 (et donc des $p_{s,k}$) vers une valeur strictement positive.

En supposant la convergence, nos estimations finales sont alors données par :

$$\begin{aligned} \hat{\beta}_\infty &= \sum_s p_{s,\infty} \beta_s & (p_{s,\infty} &= \lim_{k \rightarrow \infty} p_{s,k}) \\ \hat{\sigma}_\infty^2 &= \lim_{k \rightarrow \infty} \sigma_k^2 \end{aligned}$$

et on aura à la fois $(p_{s,\infty})_s = \Omega(\hat{\sigma}_\infty^2)$ et $\hat{\sigma}_\infty^2 = A((p_{s,\infty})_s)$, c'est-à-dire que les constructions des deux paragraphes précédents sont simultanément réalisées.

Il serait possible de procéder de même avec l'estimation de σ^2 par maximum de vraisemblance, auquel cas on est exactement dans le cas d'une méthode EM (le calcul des poids et de l'estimation de β étant donné une estimation de variance étant la partie « E » espérance, et le calcul d'une estimation de σ^2 à partir d'une estimation de la moyenne étant la partie « M », minimisation). En pratique sur les données PESETA IV, on constate également une convergence très rapide.

Quelle que soit la valeur finale des poids, on a

$$\mathbb{E}[\sum_s p_{s,\infty} B_s] = \sum_s p_{s,\infty} \beta = \beta$$

et on pourrait être tenté de conclure qu'on a bien un estimateur non-biaisé de β . Toutefois nous avons écrit ci-dessus les estimations, et passé sous silence les estimateurs. L'étude exacte de ces derniers paraît très délicate : en effet les poids limites dépendent des observations β_s à travers la fonction A , donc ils devraient être modélisés par des variables aléatoires obtenues comme limites d'une suite de variables aléatoires : $P_{s,\infty} = \lim_k P_{s,k}$ (dont il faudrait s'assurer qu'elle existe presque-sûrement). L'estimateur de β s'écrit alors $\hat{B}_\infty = \sum_s P_{s,\infty} B_s$, et même le biais de $\hat{B}_\infty$ n'apparaît pas exprimable analytiquement, et peut en principe être non nul. Toutefois, contrairement à A , la fonction Ω ne dépend que des données $(\tau_s^2)_s$, pas des observations $(\beta_s)_s$. On peut donc penser que chaque $P_{s,\infty} = \Omega(\hat{\Sigma}_\infty^2)$ est peu corrélé à B_s , ce dernier n'influant individuellement que modérément sur $\hat{\Sigma}_\infty^2$. En négligeant la corrélation entre $P_{s,\infty}$ et B_s , on retrouve le caractère

(presque) non-biaisé :

$$\begin{aligned}
\mathbb{E}[\hat{B}_\infty] &= \sum_s \mathbb{E}[P_{s,\infty} B_s] \\
&\simeq \sum_s \mathbb{E}[P_{s,\infty}] \cdot \beta \\
\mathbb{E}[\hat{B}_\infty] &\simeq \beta \underbrace{\mathbb{E}\left[\sum_s P_{s,\infty}\right]}_{=1 \text{ presque sûrement}} = \beta
\end{aligned}$$

(ii.f) *Une question de convergence.* La question de la convergence de nos estimateurs se pose évidemment ; traiter le cas des estimateurs finaux semble extrêmement délicat, vu la difficulté à traiter leur espérance et l'incertitude sur la convergence du procédé de construction lui-même. Nous allons discuter ici un cas assez général d'estimateur de l'espérance par moyenne pondérée, mais où les poids sont supposés déterministes. Le résultat présenté est très simple (une convergence faible sous une hypothèse forte) mais nous permet d'introduire une interprétation du facteur de normalisation z . Pour aller (beaucoup) plus loin sur le sujet, on peut notamment consulter [GK56] (page 113).

Supposons donc donnée une suite $(X_i)_{i \in \mathbb{N}^*}$ de variables aléatoires réelles, indépendantes, de carré intégrables, de même espérance μ (on ne les considère évidemment pas identiquement distribuées, ni même de même variance, vu notre cadre d'application). Considérons une suite d'estimateurs de μ de la forme

$$\hat{M}_n = \sum_{i=1}^n p_{n,i} X_i$$

où les $(p_{n,i})_{1 \leq i \leq n}$ sont des poids positifs et pour tout n on a $\sum_i p_{n,i} = 1$ (de sorte que $\hat{M}_n$ est sans biais). Attention, ici le double indexage n'a aucun rapport avec le procédé itératif de (ii.e), il est rendu nécessaire par le fait qu'on n'a pas le même nombre de poids à chaque valeur de n . On se demande sous quelle condition la suite $\hat{M}_n$ converge vers μ ; on considèrera ici la convergence faible (en probabilité, ou de façon équivalente en loi puisque la limite est déterministe), notamment afin de limiter la technicité.

Observons que si les poids sont arbitraires, il peut ne pas y avoir convergence vers μ . Par exemple si $p_{n,n} = 1$ et $p_{n,i} = 0$ pour tout $i < n$, alors $\hat{M}_n = X_n$ et la suite n'a pas de limite si la suite $(X_n)_n$ ne converge pas elle-même vers une variable aléatoire déterministe. Même en dehors de ce cas extrême, si pour chaque n l'un des X_i contribue de façon disproportionnée à la moyenne, on sent que la convergence peut être mise en cause. Si les poids sont par exemple égaux, $p_{n,i} = \frac{1}{n}$, mais que la variance σ_i^2 de x_i tends très assez rapidement vers $+\infty$, on voit également que les quelques derniers termes vont peser très fort dans la moyenne, empêchant la convergence vers μ . Notre but va être d'identifier des conditions suffisantes simples assurant la convergence, qui devront du fait de ces remarques impliquer la répartition des poids et des variances. Du fait qu'on ne demande que la convergence faible, on peut espérer que de telles conditions suffisent ; mais pour assurer la convergence presque sûre, il faudrait certainement demander en plus une condition sur les variations de $p_{n,i}$ à i fixé, quand $n > i$ augmente.

Ces observations étant faites, venons-en à notre critère de convergence. Étant donné qu'on s'est donné une hypothèse L^2 et qu'on ne demande que la convergence faible, il suffit d'adapter la preuve L^2 usuelle de la loi faible des grands nombres.

Proposition 3.2 (Loi faible des grands nombres pour les moyennes pondérées). *Pour assurer que la suite $(\hat{M}_n)_n$ converge en probabilité, il suffit que*

$$\sum_{i=1}^n p_{n,i}^2 \mathbb{V}[X_i] \rightarrow 0.$$

Démonstration. On s'appuie sur l'inégalité de Bienaymé-Tchebycheff : $\hat{M}_n$ étant d'espérance μ , pour tout $\varepsilon > 0$, on a

$$\begin{aligned} \mathbb{P}(|\hat{M}_n - \mu| > \varepsilon) &\leq \frac{\mathbb{V}(\hat{M}_n)}{\varepsilon^2} \\ &\leq \frac{1}{\varepsilon^2} \left(\sum_{i=1}^n p_{n,i}^2 \mathbb{V}(X_i) \right). \end{aligned}$$

□

Évidemment, la vitesse de convergence vers 0 de $\sum_{i=1}^n p_{n,i}^2 \mathbb{V}[X_i]$, la variance de l'estimateur, contrôle la vitesse de convergence en probabilité vers l'espérance. On remarque d'ailleurs que les poids optimaux calculés dans le paragraphe (ii.c) sont ceux qui la minimisent, puisque précisément on a minimisé la variance de l'estimateur (ce qui ne fait pas intervenir l'hypothèse de normalité). En prenant ces poids optimaux, c'est-à-dire

$$p_{n,i} = \frac{1}{z_n} \frac{1}{\mathbb{V}(X_i)} \quad \text{où } z_n = \sum_{i=1}^n \frac{1}{\mathbb{V}(X_i)},$$

on obtient

$$\mathbb{V}(\hat{M}_n) = \sum_{i=1}^n \frac{1}{z_n^2} \cdot \frac{1}{\mathbb{V}(X_i)^2} \mathbb{V}(X_i) = \frac{1}{z_n},$$

de sorte que le facteur de normalisation z_n s'interprète comme le nombre de variables aléatoires IID de variance 1 pour lequel l'estimateur d'espérance habituel donne la même précision statistique.

(ii.g) *Validation in silico.* Étant donné la relative complexité de notre situation, notamment la difficulté à estimer la variabilité des poids $p_{s,\infty}$ et de l'estimée $\hat{\sigma}_\infty^2$ de la variance, il paraît opportun de procéder à un test empirique de la cohérence entre le modèle et la méthode d'estimation. Nous allons concentrer nos efforts sur la partie « modèle mixte », et donc générer $N \gg 1$ réalisations pseudo-aléatoires indépendantes $(\beta_s^i)_{s \in \mathcal{S}, 1 \leq i \leq N}$ suivant la loi $\mathcal{N}(\beta, \sigma^2 + \tau_s^2)$, où nous utiliserons pour β , σ^2 et les τ_s^2 des valeurs fixées ; ici nous utiliserons pour ces paramètres les estimées obtenues pour le cas de la France entière. Pour chacune des N réalisations, nous appliquons la procédure statistique appliquées aux données PESETA IV, telle que détaillée ci-dessus. Nous regardons ensuite les

écarts entre les estimées obtenues dans les différentes réalisations, et les valeurs de β , σ^2 (qui dans le cadre de ce test empirique sont vraiment les valeurs réelles et sont connues). Nous pourrions ainsi vérifier expérimentalement, pour des valeurs à priori proches de celles que nous avons, la caract re peu biais  de nos estimateurs et avoir un aper u de leur variabilit .

Les r sultats obtenus avec $N = 100\,000$ sont pr sent s dans la figure II.5 et la table II.6. La distribution des r alisations de $\hat{B}_\infty$ appara t sym trique, et un test de normalit  permet de supposer sa distribution gaussienne (p -valeur 97%, test de D’Agostino et Pearson impl ment  par `scipy.stats.normaltest`). La distribution des r alisations de $\hat{\Sigma}_\infty^2$ est elle asym trique, comme on peut l’attendre pour un param tre positif, et son mode est l g rement en-dessous de la valeur r elle.

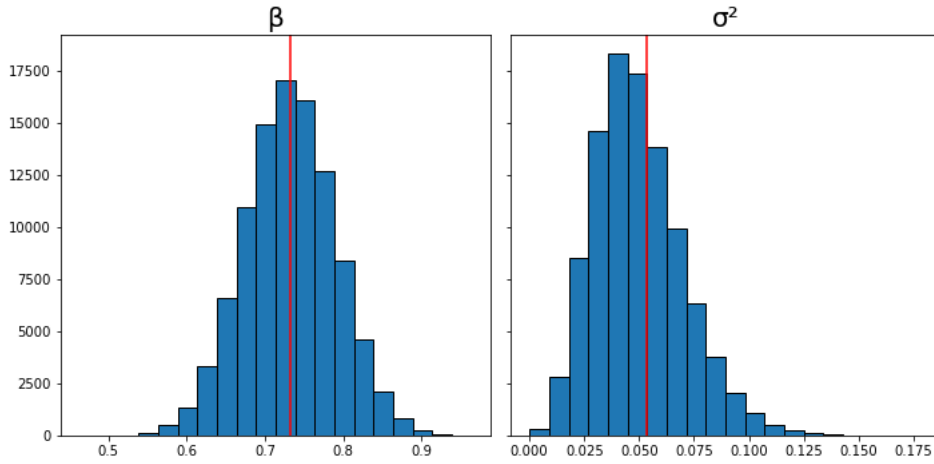


FIGURE II.5. : Estimations de β et σ^2 obtenues pour $N = 100\,000$ r alisations du mod le avec les param tres France. Les lignes rouges repr sentent les valeurs r elles.

Le biais de $\hat{B}_\infty$ appara t remarquablement faible, puisqu’on obtient 3 chiffres significatifs corrects. Au contraire, on observe un l ger biais vers le bas pour $\hat{\Sigma}_\infty^2$, ce qui est attendu  tant donn  qu’il s’agit d’un param tre de variabilit  pour lequel la correction de biais a  t  faite en consid rant fix s des param tres en fait al atoires (les poids p_s). Cependant le biais reste modeste   $-0,0041$, un petit ordre de grandeur en-dessous de l’ cart-type des estimations de σ^2 , qui d passe 40% de la valeur r elle du param tre. C’est cette dispersion relativement importante des estimations de σ^2 qui semble donc la limite la plus importante. Mentionnons que la m thode du maximum de vraisemblance fournit des r sultats extr mement proches.

(iii) Corr lations entre r gions

Les donn es PESETA IV pr sentent de fortes corr lations entre r gions, comme not  dans [JMD23], en partie li es   la proximit  mais pas seulement. On peut supposer que des r gions  loign es mais ayant des caract ristiques g ographiques proches r pondent

Paramètre	Valeur réelle	Moyenne	Biais	Écart-type
β	0,7775	0,7776	0,0001	0,0622
σ^2	0,0552	0,0511	-0,0041	0,0230

TABLE II.6. : Résultats synthétiques des tests *in silico*. La moyenne et l'écart-type sont empiriques, calculés sur les $N = 100\,000$ réalisations.

au réchauffement climatique de façon relativement semblable dans une majorité des simulations.

Nous allons prendre en compte et estimer ces corrélations ; on notera ici B_{rs} la variable aléatoire représentant la valeur du coefficient β dans la région r pour la simulation s , β_r son espérance, et β_{rs} la réalisation qu'on peut lire dans les données. Il faut distinguer β_r de la valeur « réelle » du coefficient, qu'on noterait ici β_{rs_0} où $s_0 \notin \mathcal{S}$ représente la réalité, dont on suppose qu'elle est une réalisation suivant la même loi que les simulations.

Une idée pour estimer les covariances pourrait être de prolonger l'approche utilisée pour une région unique par exemple en appliquant la même philosophie dans un cadre multidimensionnel. Les calculs théoriques étant déjà relativement lourds dans le cadre unidimensionnel, nous ne développerons pas cette approche ici ; notons qu'en pondérant les différentes simulations non pas en fonction de la seule variabilité pour une région, mais à l'aide de l'ensemble des informations des régressions pour toutes les régions, on obtiendrait des variances différentes.

Une seconde idée pourrait être de mener une estimation des covariances une par une, en utilisant les mêmes principes. Cependant, une tentative en ce sens s'est heurtée à un problème de taille : les coefficients estimés ne formaient pas une matrice définie-positive, et ne permettaient donc pas de définir une loi normale multivariée.

Nous avons donc opté pour une approche simple, qui permette de garder la structure de dépendance apparente dans les données, tout en exploitant le travail fait sur les variances par région : nous estimons de façon classique la matrice de covariance

$$C_{r,r'} = \mathbb{E}[(B_r - \beta_r)(B_{r'} - \beta_{r'})]$$

où B_r représente une variable aléatoire de référence ayant la même loi que les B_{rs} , nous en déduisons une matrice de corrélations, et nous utilisons in fine la matrice de covariance $\hat{\Gamma}$ qui correspond aux corrélations obtenues et aux variances $\hat{\sigma}_{\infty,r}^2$ obtenues en appliquant (ii.e) à chaque région r . Ainsi nous gardons une cohérence des corrélations avec les données, et des variances avec le travail effectué dans le cas d'une région.

Pour la matrice de covariance $C = (C_{r,r'})_{r,r' \in \mathcal{R}}$, on utilise l'estimateur simple, non biaisé habituel

$$\hat{C}_{r,r'} = \frac{1}{|\mathcal{S}| - 1} \sum_{s \in \mathcal{S}} (B_{rs} - \bar{B}_r)(B_{r's} - \bar{B}_{r'})$$

(où $\bar{B}_r = \frac{1}{|\mathcal{S}|} \sum_s B_{rs}$), qu'on applique aux valeurs $\beta_{r,s}$ pour obtenir l'estimée $\hat{c}_{r,r'}$. Les

corrélations de Pearson sont ensuite calculées :

$$\hat{\rho}_{r,r'} = \frac{\hat{c}_{r,r'}}{\sqrt{\hat{c}_{r,r} \cdot \hat{c}_{r',r'}}}$$

et enfin, la matrice de covariance $\hat{\Gamma} = (\hat{\gamma}_{r,r'})_{r,r' \in \mathcal{R}}$ qui sera utilisée pour la génération de scénarios (Section 2 du chapitre III) en est déduite :

$$\hat{\gamma}_{r,r'} = \hat{\rho}_{r,r'} \cdot \sqrt{\hat{\sigma}_{\infty,r}^2 \cdot \hat{\sigma}_{\infty,r'}^2}.$$

(iv) Résumé de la procédure d'estimation

La construction des valeurs centrales, variances et covariances des coefficients β_r pour les différentes régions r (ainsi que pour la France Métropolitaine entière, qu'on traite comme une région en sommant les dommages projetés sur toutes les régions) se fait donc en résumé de la manière suivante. Tout d'abord, pour chaque région r :

- i. on effectue 22 régressions linéaires (une par simulation s) sur les logarithmes des valeurs de dommages aux trois niveaux de réchauffement global considéré, ce qui donne 22 valeurs $\hat{\beta}_{rs}$ pour la pente et 22 valeurs d'incertitude $\hat{\tau}_{rs}^2$ liée à la qualité d'alignement des points ;
- ii. On estime de façon itérative 22 poids p_{rs} et une estimation $\hat{\sigma}_r^2$ de la variabilité inter-modèle, de sorte à obtenir une estimation $\hat{\beta}_r$ sous la forme d'une moyenne des $\hat{\beta}_{rs}$ pondérée par les p_{rs} , de sorte que ces poids soient optimaux dans l'hypothèse $\sigma_r^2 = \hat{\sigma}_r^2$ et simultanément que $\hat{\sigma}_r^2$ soit une estimation en variance empirique pondérée par les p_{rs} ;
- iii. on retient en sortie la valeur centrale $\hat{\beta}_r$ et un écart-type $\sqrt{\hat{s}_r^2 + \sum_s p_{rs}(\hat{s}_r^2 + \tau_{rs}^2)}$ agrégeant la variabilité inter-simulation et l'incertitude sur l'estimateur $\hat{\beta}_r$.

Ensuite, pour les covariances :

- iv. on calcule l'estimateur empirique usuel (non-biaisé) de la matrice de covariance entre régions à l'aide des 22 jeux de valeurs $(\beta_{rs})_{r \in \mathcal{R}}$;
- v. on en déduit la matrice de corrélations associée ;
- vi. on calcule la matrice de covariance $\hat{\Gamma}$ qui correspond à ces corrélations et aux écarts-types calculés plus haut pour chaque région individuelle.

En terme d'usage, on considèrera dans le générateur de scénario (2 du chapitre III) que le vecteur $\vec{B}$ des coefficients β_r par région est un vecteur Gaussien, centré sur $(\hat{\beta}_r)_{r \in \mathcal{R}}$ et de matrice de covariance $\hat{\Gamma}$.

3.3. Application du modèle statistique aux données PESETA IV

(i) Résultats obtenus

L'application de la procédure par région décrite ci-dessus abouti aux données présentées dans la table II.7. On observe que les valeurs centrales $\hat{\beta}_r$ sont assez élevées : une valeur de $\ln(2) \simeq 0,69$ correspond à une multiplication des dommages par 2 pour chaque degré Celsius de réchauffement global, elle est dépassée dans une majorité de régions (et pour la France entière). La deuxième observation est que l'incertitude est grande, avec un écart-type souvent de l'ordre de la moitié de l'estimateur. Ceci nous place à la limite de la significativité pour le signe de β_r , mais comme sur chaque région nous obtenons une estimation positive, on peut largement considérer que ces données confirment l'aggravation des inondations de plaine par le réchauffement climatique. L'incertitude est d'ailleurs moins importante pour la France entière, certainement par effet de moyenne géographique. Enfin, l'écart-type total est presque égal à l'écart-type inter-simulation : le terme de variance de l'estimateur $\hat{B}_r$ est petit devant la variabilité des simulations. Ceci est naturel, puisque le premier décroît avec le nombre de simulations mais pas le second. Même une vingtaine de simulations suffisent à obtenir un estimateur assez stable, et l'incertitude liée aux choix d'hypothèses dans les différentes simulations domine largement.

La figure II.8 donne une représentation graphique des $\hat{\beta}_r$ et des incertitudes relatives. On remarque notamment que les régions Rhône-Alpes et Languedoc-Roussillon ont à la fois une estimation élevée et une incertitude faible. Corse mise à part, on n'observe pas l'effet parfois attendu d'une diminution des inondations dans le sud, qui subirait plus de sécheresses mais moins d'inondations.

D'après les données PESETA IV, les régions traversées par les fleuves^a sont les plus susceptibles de subir des inondations de plaine fortement aggravées par le réchauffement global, qu'elles soient au nord ou au sud de la France.

^aÀ l'exception de Midi-Pyrénées, relativement moins impactée.

On peut se demander si ce lien très apparent entre présence d'un fleuve et valeur élevée de β_r est un effet « concentrateur » de la grande taille du bassin versant d'un cours d'eau important (ce qui pourrait également expliquer l'impact modéré en Midi-Pyrénées, où le bassin de la Garonne est encore assez resserré), ou s'il s'agit d'une différence de capacité de PESETA IV à modéliser les cours d'eau en fonction de leur taille.

En ce qui concerne les corrélations entre régions, la figure II.9 dessine les corrélations avec quatre régions choisies parmi les 22. Le cas de l'Île de France, avec un fort effet de proximité, est assez commun parmi les 22 régions. Le cas de l'Alsace est particulier, avec très peu de corrélation à l'ensemble des autres régions (ce qui se comprend puisqu'elle se trouve sur le Rhin en amont des affluents importants en France, telle que la Moselle). Il est frappant de constater que les corrélations sont presque systématiquement positives, parfois proches de 1 pour des régions rapprochées, et seulement dans de rares cas un peu négatives (ce qui n'arrive qu'entre des régions très éloignées). On s'attend donc à relativement peu d'effets de compensation géographiques, qui auraient pu compenser les très grandes incertitudes.

	$\hat{\beta}_r$	Écart-type total	Écart-type inter-simu σ
Île-de-France	0,923	0,553	0,535
Champagne-Ardenne	0,818	0,467	0,452
Picardie	0,918	0,518	0,503
Haute-Normandie	0,959	0,556	0,539
Centre-Val de Loire	0,665	0,347	0,335
Basse-Normandie	0,604	0,289	0,279
Bourgogne	0,723	0,291	0,282
Nord-Pas de Calais	0,592	0,217	0,209
Lorraine	0,642	0,492	0,478
Alsace	1,063	0,700	0,682
Franche-Comté	0,796	0,348	0,337
Pays de la Loire	0,994	0,490	0,474
Bretagne	0,565	0,399	0,386
Poitou-Charentes	0,872	0,545	0,528
Aquitaine	0,711	0,336	0,325
Midi-Pyrénées	0,432	0,312	0,303
Limousin	0,565	0,365	0,351
Rhône-Alpes	0,885	0,309	0,297
Auvergne	0,406	0,278	0,269
Languedoc-Roussillon	0,879	0,335	0,325
Provence-Alpes-Côte d'Azur	0,749	0,351	0,340
Corse	0,369	0,350	0,341
France	0,777	0,243	0,235

TABLE II.7. : Résultats de la procédure d'estimation sur les données PESETA IV. Les deux premières colonnes sont les sorties utilisées, la troisième est donnée pour indication.

(ii) Validation de quelques hypothèses

Nous allons maintenant discuter certains choix et hypothèses qui ont été faites et qui sont facilement testables sur les données traitées.

(ii . a) *Linéarité.* Tout d'abord, la validité du modèle log-linéaire peut être questionnée : les logarithmes des dommages apparaissent-ils comme raisonnablement linéaires en le niveau de réchauffement global? Comme les régressions sont sur seulement trois points, les éléments qui remettraient en cause cette linéarité seraient un mauvais alignement général des triplets de points, ou une position très majoritairement convexe ou concave des triplets.

La figure II.10 présente l'ensemble des régressions pour une région, Haute-Normandie, choisie car assez représentative de l'ensemble. On observe une pente négative dans une simulation (notée rcp4.5-03 dans la figure), avec des résidus assez élevés. On dénombre

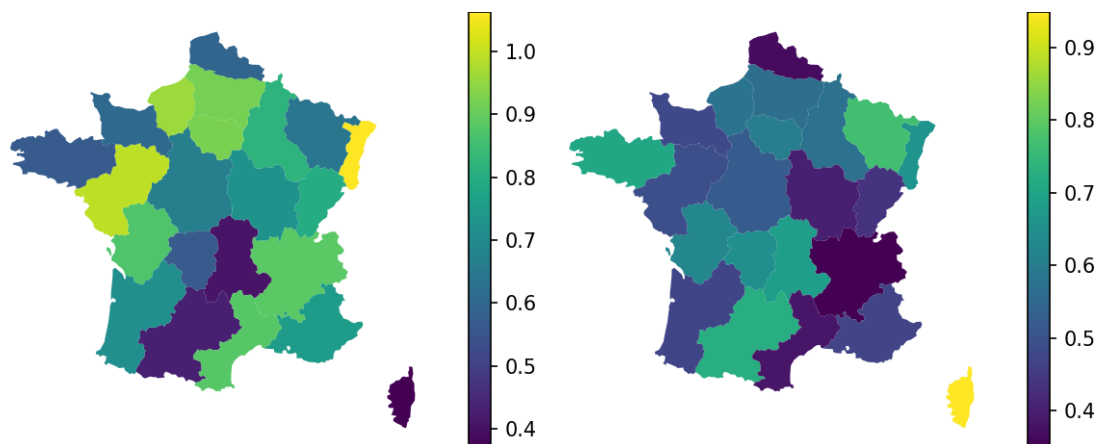


FIGURE II.8. : À gauche, estimations $\hat{\beta}_r$ des coefficients d'impact régionaux du réchauffement global. À droite, rapport écart-type sur estimation, représentant l'incertitude relative sur la valeur réelle de β_r .

8 simulations où les points sont en position convexe (comme `rcp4.5-01` par exemple), 11 où ils sont en position concave (`rcp4.5-02`) et 3 où l'alignement est trop bon pour trancher clairement à l'œil nu (`rcp4.5-00`).

Ce léger déséquilibre est présent dans la grande majorité des régions ; un mauvais point est qu'il est systématiquement dans le même sens (moins de positions convexes que de concaves), cependant les différentes régions ne sont pas indépendantes, et les alignements sont dans l'ensemble très bons (les résidus sont modestes en regard de la variabilité de la pente entre simulations). L'hypothèse de linéarité apparaît comme acceptable.

(ii . b) *Absence d'homoscédasticité forte.* Rappelons que nous avons refusé d'utiliser l'hypothèse forte d'homoscédasticité entre les résidus à travers toutes les simulations (pour une région donnée), ce qui a compliqué le travail statistique ; donnons un peu de détail sur les tests ayant mené à ce choix. La mise en doute de cette hypothèse forte d'homoscédasticité vient de la grande variation de taille des résidus entre différentes simulations qu'on peut observer sur la figure II.10, mais aussi de la relative cohérence entre régions, à l'intérieur de chaque simulation, qu'on peut observer en comparant les analogues de la figure II.10 pour les différentes régions (disponibles en matériel additionnel dans le dossier `figures produite`, fichiers `Régression des  $\beta$  par simulation - xxx.png`) : l'alignement est par exemple souvent mauvais pour `rcp4.5 - 03` mais très bon pour `rcp8.5 - 15`.

Nous nous sommes basés sur le test classique de Levene, qui cherche étant donné plusieurs groupes de réalisations de variables aléatoires identiquement distribuée dans chaque groupe, mais pas entre groupes différents, à savoir si au moins un groupe a une variance significativement différente des autres.

Ici, nous constituons un groupe par simulation, et nous utilisons les différentes régions

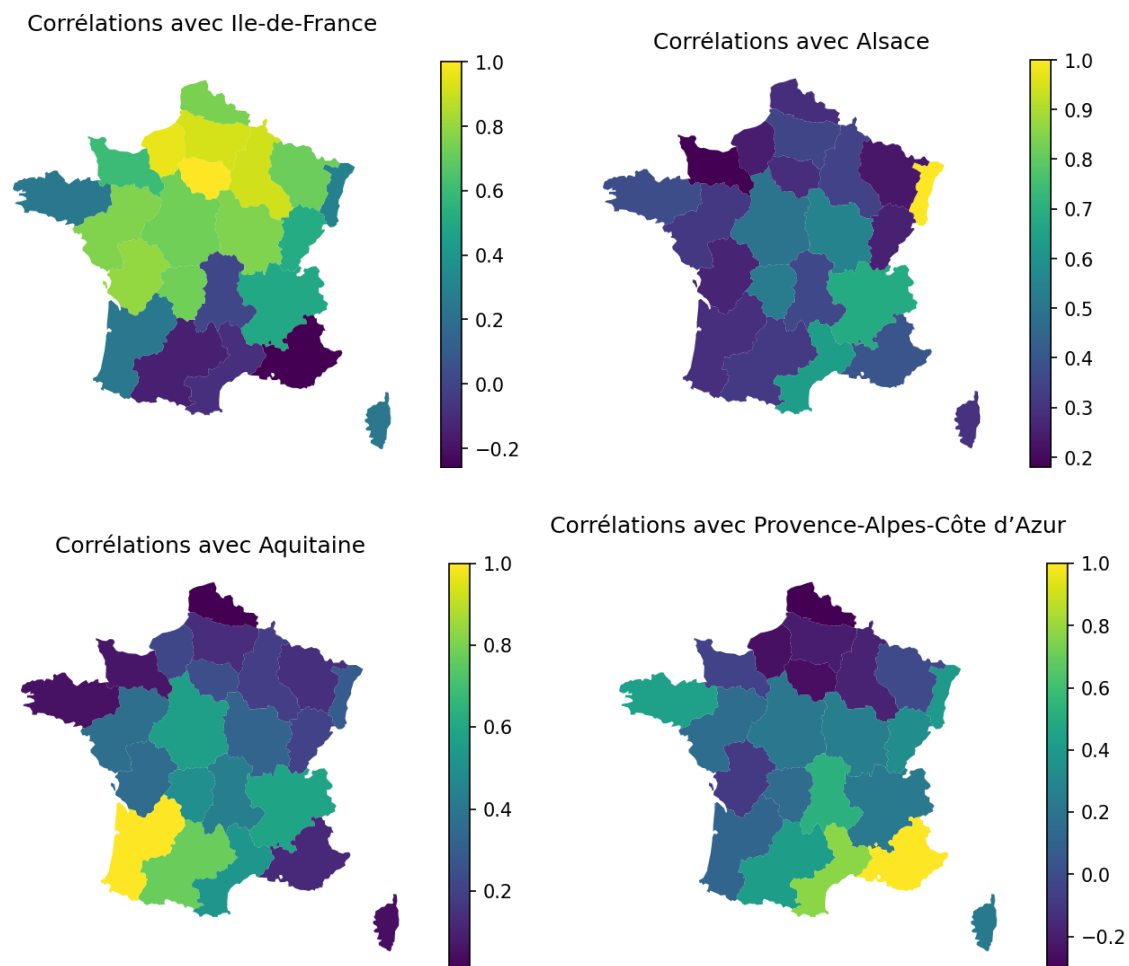


FIGURE II.9. : Corrélations des β_r entre régions, avec l'Île-de-France (en haut à gauche), l'Alsace (en haut à droite), l'Aquitaine (en bas à gauche), Provence-Alpes-Côte d'Azur (en bas à droite).

Régression des β par simulation — Haute-Normandie

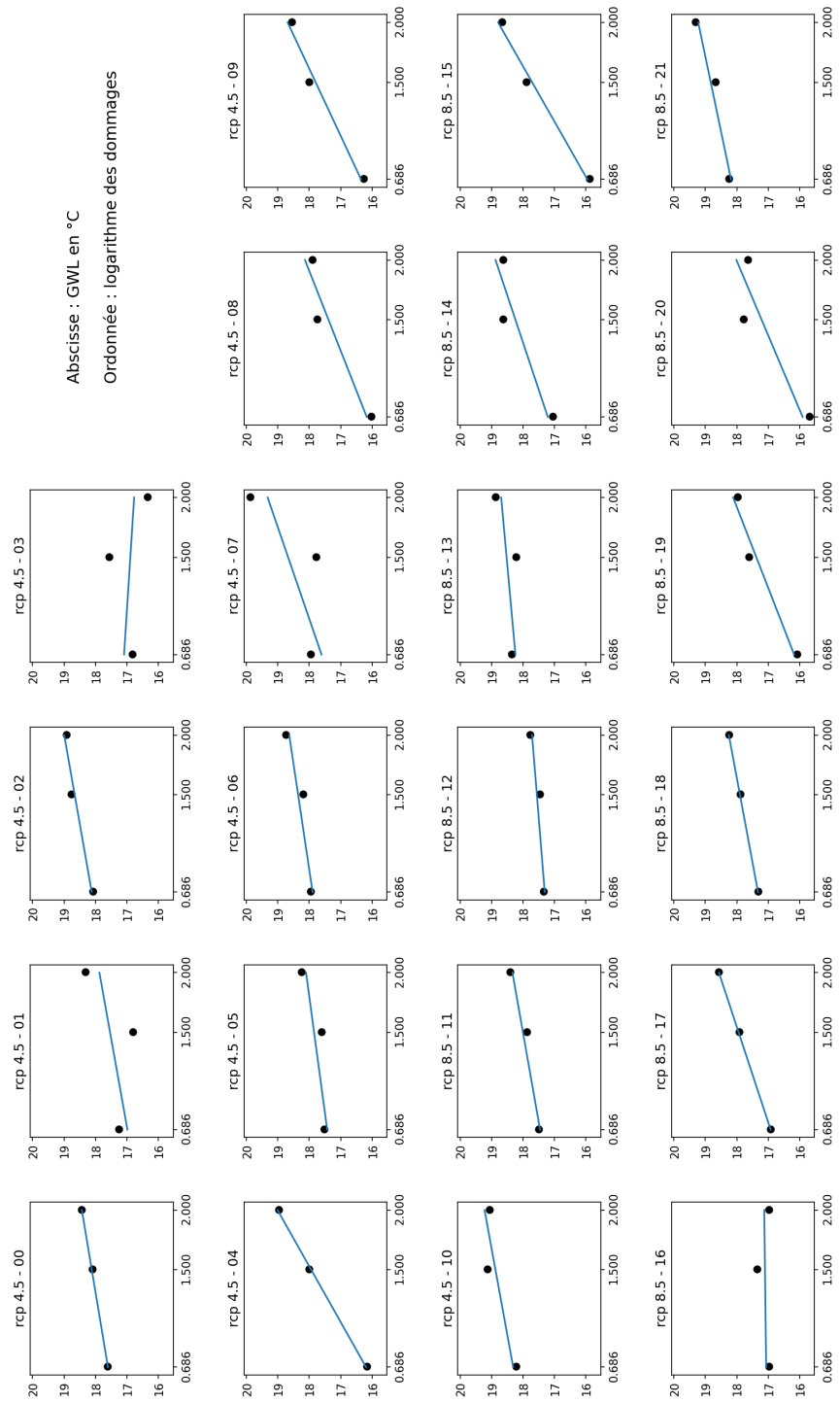


FIGURE II.10. : Régressions pour les 22 simulations : les points sont les données PESETA IV, les droites sont les régressions aux moindres carrés ordinaires.

comme réalisations à l'intérieur d'un groupe, en cohérence avec les observations qui nous ont amené à tester l'homoscédasticité forte.

Le rôle de la variable indépendante va être joué par les résidus, mais nous devons faire attention à ce que les résidus d'une régression donnée ne sont ni indépendants, ni de même variance (même quand les bruits le sont, ce qui est supposé ici). En effet la matrice du plan d'expérience va affecter les variances des résidus, différemment suivant le point considéré ; quand à l'indépendance, notre situation permet de bien voir qu'elle est complètement perdue. En effet, la régression projette un vecteur de données (y_0, y_1, y_2) sur le plan engendré par les vecteurs $(1,1,1)$ et (t_0, t_1, t_2) (où les t_i sont les niveaux de réchauffement considérés). Les résidus forment un vecteur de l'orthogonal de ce plan ; bien qu'il ait trois coordonnées le vecteur des résidus $(\hat{\varepsilon}_0, \hat{\varepsilon}_1, \hat{\varepsilon}_2)$ est contraint à vivre sur une droite ! Nous allons donc retenir un seul résidu par régression, nous choisissons celui correspondant au point central (t_1, y_1) . Comme la matrice du plan d'expérience ne dépend que des trois niveaux de réchauffement, elle est la même dans toutes les simulations et toutes les régions et l'homoscédasticité des bruits entre simulations est équivalente à l'homoscédasticité de ces résidus centraux.

Nous adaptons le test de Levene, car les résidus sont d'espérance connue, nulle. Nous menons le test en remplaçant la moyenne empirique de la variable indépendante par 0, et obtenons une p -valeur de $1,1 \cdot 10^{-16}$. En guise de complément, nous menons le test de Levene classique, pour lequel la p -valeur est plus élevée mais toujours très significative : $9,2 \cdot 10^{-11}$. Comme le test est basé sur une hypothèse de normalité (bien qu'il soit connu pour sa robustesse face à cette hypothèse) nous menons un test de normalité (D'Agostino-Pearson) par simulation ; 3 obtiennent une p -valeur inférieure à 5%, ce qui est attendu pour 22 tirages indépendants dans environ 9% des cas. On peut donc conserver l'hypothèse de normalité, mais afin de confirmer nous menons un dernier test de Levene, où la valeur centrale de la variable indépendante est la médiane plutôt que la moyenne, ce qui renforce la robustesse face au défaut de normalité. Nous obtenons une p -valeur de $4,1 \cdot 10^{-9}$. Ces tests confirment donc sans équivoque qu'il convient de rejeter l'hypothèse forte d'homoscédasticité.

(ii. c) *Normalité des B_s* Dans notre démarche, nous avons supposé que (à région r fixée) les différentes simulations s correspondent à des variables aléatoires B_s normales, centrées sur la valeur $\beta = \beta_r$ que nous cherchons à estimer, de variances $\sigma^2 + \tau_s^2$. Plus précisément, cette normalité est une conséquence de la normalité supposée des bruits dans les régressions, qui est impossible à tester directement puisque nous n'avons qu'un seul degré de liberté dans les résidus. Nous pouvons en revanche tester la normalité des B_s .

Comme la variance dépend de la simulation, il faut renormaliser et nous suivons une approche plug-in, les valeurs à tester sont donc dans chaque région les

$$\tilde{\beta}_s = \frac{\hat{\beta}_s - \hat{\beta}}{\sqrt{\hat{\sigma}^2 + \hat{\tau}_s^2}}$$

où $\hat{\beta}$ est l'estimation finale dans la région, $\hat{\beta}_s$ l'estimation dans la simulation s , $\hat{\sigma}^2$ l'estimation de la variance inter-simulation (notée aussi $\hat{\sigma}_\infty^2$ plus haut car obtenu comme limite

d'un processus itératif), et $\hat{\tau}_s^2$ l'estimation de variance d'incertitude dans la régression.

Nous avons appliqué un test de D'Agostino-Pearson aux $\tilde{\beta}_s$ dans chaque région, une seule p -valeur est inférieure au seuil de 5% (Lorraine, 1,1%). En utilisant une loi binomiale, on sait que si l'hypothèse nulle de normalité était vérifiée pour les 22 régions, en supposant l'indépendance des tests on aurait une probabilité d'environ 22% d'obtenir une ou plus p -valeur inférieures ou égales à 1,1%. Ni le nombre de p valeurs en-dessous de 5% (1 attendue) ni la valeur obtenue pour la plus petite p valeur ne s'écartent donc significativement de ce qui serait attendu sous l'hypothèse nulle. On conclut qu'il n'y a pas lieu de remettre en cause l'hypothèse de normalité.

4. Estimation des niveaux de risque régionaux et modélisation de la variabilité annuelle

Dans cette section, on étudie de façon conjointe le problème de l'estimation des coefficients A_r , qui représentent le niveau de risque du portefeuille considéré par région « hors réchauffement climatique », et la modélisation de la variabilité annuelle $F_{r,t}$ (voir la section 1.2). Il s'agit encore d'une approche théorique, l'application sur des données sera faite au chapitre III. On utilisera souvent les notations $\alpha_r = \log A_r$ et $E_{r,t} = \log F_{r,t}$.

Cette estimation se fera sur les données historiques du portefeuille, pour lesquelles on dispose des sinistralités $Dom_{r,t}$ et des enjeux assurés $EA_{r,t}$; par ailleurs nous avons déjà estimé les β_r et le niveau de réchauffement $GWL(t)$, qui sont donc également considérés comme connus. Pour la génération de nombreux scénarios afin d'estimer différents quantiles de dommages, ces deux dernières grandeurs auront été modélisés comme aléatoires. On utilisera donc les valeurs réalisées par le tirage pseudo-aléatoire dans le scénario considéré, et l'estimation que nous allons discuter devra être complètement automatisée et aussi économe que possible en calcul pour pouvoir être répétée des dizaines de milliers de fois facilement.

4.1. Positionnement du problème et première pistes

On peut récrire le modèle (II.1) sous la forme :

$$\alpha_r + E_{r,t} = \log \frac{Dom_{r,t}}{EA_{r,t}} - \beta_r GWL(t)$$

ou encore

$$A_r \cdot F_{r,t} = \frac{Dom_{r,t}}{EA_{r,t}} \cdot e^{-\beta_r GWL(t)},$$

où on dispose des réalisations de $Dom_{r,t}$ dans nos données. Après avoir ramené la sinistralité à un rapport avec les enjeux assurés, puis effectué une « actualisation climatique » pour rééchelonner le passé par le réchauffement climatique additionnel, pondéré de la susceptibilité β_r de la région (qui joue le rôle d'un « taux d'intérêt climatique »), nous sommes donc face à un problème d'estimation classique. Pour simplifier, on notera

$$X_{r,t} = \frac{Dom_{r,t}}{EA_{r,t}} \cdot e^{-\beta_r GWL(t)} = A_r \cdot F_{r,t}$$

les variables aléatoires d'intérêt, tandis que $x_{r,t}$ dénotera les réalisations calculées à partir des données, que nous aurons concrètement à notre disposition (voir la section 2 de l'annexe). On note $\vec{F}_t = (F_{r,t})_{r \in \mathcal{R}}$ le vecteur des aléas l'année t à travers les différentes régions, et de même $\vec{E}_t = (E_{r,t})_{r \in \mathcal{R}}$ et $\vec{X}_t = (X_{r,t})_{r \in \mathcal{R}}$. Vu les hypothèses faites à la section 1.2, on suppose que les vecteurs $(\vec{X}_t)_t$ (et donc, de façon équivalente, les $(\vec{F}_t)_t$ et les $(\vec{E}_t)_t$) forment une famille I.I.D. : l'évolution temporelle est supposée entièrement contenue dans le facteur d'actualisation climatique $e^{-\beta_r \text{GWL}(t)}$.

(i) La médiane comme valeur centrale

Commençons par remarquer que le problème d'estimation est mal posé, puisque l'on obtient le même modèle en considérant A_r/λ en lieu de A_r et $\lambda F_{r,t}$ en lieu de $F_{r,t}$, quel que soit $\lambda > 0$. Comme on a séparé la grandeur à estimer en une partie centrale déterministe A_r ou α_r et une partie variabilité $E_{r,t}$ ou $F_{r,t}$, comme d'habitude il s'agit de supposer la partie variabilité centrée. Mais attention, on ne peut pas à la fois supposer $\mathbb{E}[E_{r,t}] = 0$ et $\mathbb{E}[F_{r,t}] = 1$ car l'exponentielle est strictement convexe, donc

$$\mathbb{E}[F_{r,t}] = \mathbb{E}[e^{E_{r,t}}] > e^{\mathbb{E}[E_{r,t}]}.$$

Cette remarque nous amène à un premier choix, sur la définition de A_r , que nous choisissons être la *médiane* de la distribution de $X_{r,t}$, de sorte que α_r est également la médiane de la distribution de $\log X_{r,t}$. Ainsi $F_{r,t}$ devient de médiane 1 et $E_{r,t}$ de médiane 0.

Nous retenons tout naturellement la médiane empirique des $(X_{r,t})_t$ comme estimateur $\hat{A}_r$ de A_r , de sorte que la médiane empirique des $(\log X_{r,t})_t$ est $\log \hat{A}_r$, estimateur de α_r .

Dans le cas où l'historique contient un nombre pair d'années, on utilisera la médiane usuelle (la moyenne des deux statistiques d'ordre les plus centrales) pour $\log \hat{A}_r$, et $\hat{A}_r$ en sera l'exponentielle – la notion de milieu est en effet plus pertinent dans les grandeurs additives que dans les grandeurs multiplicatives. Une autre façon de dire la même chose est que $\hat{A}_r$ est alors la moyenne *géométrique* des deux statistiques d'ordre les plus centrales.

Il sera pratique vu la normalisation ci-dessus de se concentrer sur les $E_{r,t}$, ce que nous feront désormais.

(ii) Première approche de l'aléa : échantillonnage dans historique

Sachant qu'on disposera de relativement peu de données (entre une dizaine et une quarantaine d'années pour chaque région), une première approche pourrait être de modéliser les $\vec{E}_t$ dans le futur par une loi uniforme sur les valeurs observées dans l'historique. Cette méthode a l'avantage de ne pas faire intervenir de paramètre à estimer et de conserver la dépendance entre régions telle qu'on l'a observée.

Cependant, elle a l'inconvénient majeur de ne pas considérer la possibilité d'événements plus extrêmes dans le futur que ceux ayant été observé. Ainsi, suivant la quantité de données, on serait limités à des quantiles de l'ordre de 90% à au mieux 97,5%, ce

qui est limité à complètement insuffisant. Nous ne considérons donc pas cette approche raisonnable pour la queue de la distribution.

(iii) Deuxième approche de l'aléa : distribution log-normale

Une approche paramétrique est souvent privilégiée pour traiter ce genre de problème, notamment dans le cadre de la théorie des valeurs extrêmes. Cependant, les données que nous sommes susceptible d'utiliser sont trop réduite pour appliquer cette théorie : nous travaillons en dimension 22 (le nombre de régions) avec au plus une quarantaine d'années d'historique.

À défaut, une modélisation paramétrique simple est envisageable ; il faudrait qu'elle puisse rendre compte de la structure de dépendance entre régions, tout en ayant le moins de paramètres possibles. Puisque nous avons déjà centré les $E_{r,t}$ (au sens de leur médiane), un modèle gaussien paraît simple et naturel. Ainsi, les $X_{r,t}$ seraient modélisés par une loi log-normale, dont la queue est de lourdeur intermédiaire (la densité décroît plus vite que tout polynôme, mais plus lentement que toute exponentielle). On préfère souvent utiliser des lois à queues lourdes, mais elles demandent une quantité plus substantielle de données pour être correctement calibrées.

Un premier obstacle est qu'il faudrait estimer $22 \times 21/2 = 231$ coefficients de corrélation avec une série de quelques dizaines d'années. On peut espérer que beaucoup de ces coefficients soient nuls, mais la nécessité d'automatiser (puisque les données seront redressées comme indiqué plus haut) rend difficile la vérification de cette hypothèse.

Un second obstacle est que certaines des valeurs de $F_{r,t}$ pourraient être nulles : c'est le cas quand il n'y a pas eu de sinistre dans une région donnée une année donnée. Dans ce cas, $E_{r,t} = -\infty$ et tous les estimateurs de variance ou covariance vont dégénérer. Même si aucune valeur n'est exactement nulle, la présence éventuelle de petites valeurs pourrait avoir un impact majeur sur les estimées.

Ces deux arguments nous conduisent à rejeter cette approche également. La section suivante décrit comment nous proposons de sortir de cette double impasse.

4.2. Mélanger échantillonnage dans l'historique et régression log-normale

Notre proposition va consister à combiner les deux approches évoquées de sorte à esquiver les différents problèmes soulevés. Commençons par considérer une région seule, avant d'expliquer comment traiter le problème multidimensionnel.

(i) Une région seule : échantillonnage à gauche, régression à droite

Fixons donc dans un premier temps une région $r \in \mathcal{R}$, notons $N \in \mathbb{N}$ la taille de l'historique (nombre de $t \leq t_0$) et considérons le problème de modéliser pour les projections dans le futur des variables $E_{r,t}$ ($t > t_0$), étant observées un historique de variables $E_{r,t}$ ($t \leq t_0$). L'objection pour la méthode d'échantillonnage portant sur la queue de distribution, modélisons celle-ci de façon paramétrique. la loi de $E_{r,t}$ sera donc modélisée comme une loi de mélange entre :

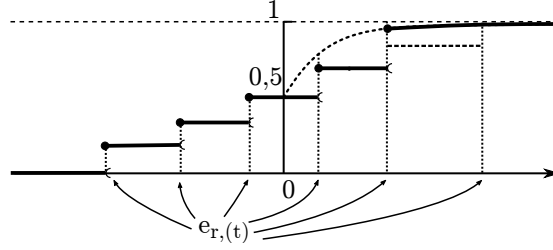


FIGURE II.11. : Allure de la fonction de répartition pour la modélisation des $E_{r,t}$ futurs. La plus haute valeur n'est pas susceptible d'être tirée, elle est remplacée par une queue de gaussienne (pointillés, puis trait plein pour la partie présente dans la loi de mélange).

- un tirage uniforme sur les $N - 1$ valeurs excluant la plus grande, avec poids $(N - 1)/N$, et
- la queue d'une gaussienne centrée (intervalle de quantiles $[(N - 1)/N, 1]$) avec poids $1/N$.

Vu la petite taille de l'historique, il faut éviter de perdre trop de données et nous allons calibrer la variance de la gaussienne avec la moitié la plus haute des valeurs historiques (figure II.11).

Notons $E_{r,(1)}, \dots, E_{r,(N)}$ les statistiques d'ordre et $e_{r,(1)}, \dots, e_{r,(N)}$ les valeurs observées en pratique (voir l'annexe pour notre convention entre majuscules et minuscules). La première loi dans le mélange est

$$\mu_g = \sum_{k=1}^{N-1} \frac{1}{N-1} \delta_{e_{r,(k)}}$$

Pour la partie droite, on va faire une régression de loi demi-normale. On utilise $\lceil \cdot \rceil$ pour dénoter la partie entière supérieure et $\lfloor \cdot \rfloor$ la partie entière usuelle (arrondie à l'inférieur). Notons $K = \lfloor \frac{N}{2} \rfloor$ et $(Y_1, \dots, Y_K)$ une permutation aléatoire uniforme de $(E_{r,(\lceil \frac{N}{2} \rceil + 1)}, \dots, E_{r,(N)})$, de sorte que les Y_k sont presque-sûrement positifs et perdent la dépendance dû au tri croissant des statistiques d'ordre. Supposons qu'ils soient I.I.D et suivent une demi-loi normale, de densité donnée par⁷

$$f_{\sigma^2}(x) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

et cherchons un estimateur de σ^2 . Deux approches naturelles se présentent : le maximum de vraisemblance et la méthode des moments. Il se trouve que la première fournit une

⁷Notons qu'ici nous commettons un abus, car le centrage des $e_{r,t}$ fait intervenir les données ; en toute rigueur on devrait mener une analyse statistique en partant du tout début, et on verrait certainement un petit biais dans l'estimation qu'on pourrait corriger. Le centrage par la médiane est susceptible de rendre les calculs assez techniques, sans qu'ils aient de grande portée sur le point crucial de bien traduire les dépendance entre régions et de modéliser tant bien que mal une queue de distribution. Nous nous contentons donc ici de cette version simplifiée.

expression simple et sans biais, c'est donc celle que nous retenons. Décrivons brièvement les calculs, classiques : la log-vraisemblance étant données les observations $y_1, \dots, y_K$ s'écrit

$$\begin{aligned}\log \mathcal{L}(\sigma^2) &= \frac{K}{2} \log \frac{2}{\pi \sigma^2} - \sum_{k=1}^K \frac{y_k^2}{2\sigma^2} \\ \frac{d \log \mathcal{L}}{d\sigma^2}(\sigma^2) &= -\frac{K}{2\sigma^2} + \sum_{k=1}^K \frac{y_k^2}{2\sigma^4} \\ &\propto -K\sigma^2 + \sum_{k=1}^K y_k^2\end{aligned}$$

et la maximisation de la vraisemblance amène l'estimée et l'estimateur suivants :

$$\hat{\sigma}^2 = \frac{1}{K} \sum_{k=1}^K y_k^2 \quad \hat{\Sigma}^2 = \frac{1}{K} \sum_{k=1}^K Y_k^2$$

Le calcul de l'espérance est le même que pour celui de la variance empirique dans le modèle Gaussien de moyenne connue, avec une intégration par parties astucieuse :

$$\begin{aligned}\mathbb{E}[\hat{\Sigma}^2] &= \mathbb{E}[Y_1^2] \\ &= \sqrt{\frac{2}{\pi \sigma^2}} \int_0^{+\infty} x \cdot x e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \sqrt{\frac{2}{\pi \sigma^2}} \left(0 + \int_0^{+\infty} \sigma^2 e^{-\frac{x^2}{2\sigma^2}} dx \right) \\ &= \sigma^2\end{aligned}$$

et l'estimateur $\hat{\Sigma}^2$ est sans biais, comme annoncé. Un calcul similaire donne une variance ayant l'ordre habituel de décroissance en $1/K$:

$$\mathbb{V}[\hat{\Sigma}^2] = \frac{2}{K} \sigma^4.$$

Dans notre situation, K sera assez petit et nous portons donc le risque de sous-estimer (ou, moins grave, sur-estimer) le paramètre σ^2 . On pourrait mener une simulation « à deux étages », où on considère le paramètre σ^2 lui-même comme aléatoire, mais d'une part ça compliquerait les codes et d'autre part l'analyse statistique d'une telle pratique n'est à ma connaissance pas standard. Nous garderons simplement en tête ce risque comme une limite à notre travail et une piste d'amélioration ; notons que prendre moins de la moitié des données dans l'estimation de la queue renforcerait ce risque d'autant.

Nous définissons donc la partie droite de la loi de mélange par

$$d\mu_d = \alpha e^{-\frac{x^2}{2\sigma^2}} \mathbf{1}_{[q, +\infty[}(x) dx, \quad \hat{\sigma}^2 = \frac{1}{\lfloor \frac{N}{2} \rfloor} \sum_{k > \lceil \frac{N}{2} \rceil} (e_{r,(k)})^2$$

où α est le coefficient de normalisation et q le quantile d'ordre $(N-1)/N$ (qui dépendent tous les deux de N et $\hat{\sigma}^2$).

Finalement, nous modélisons les $E_{r,t}$ pour $t > t_0$ par la loi

$$\frac{N-1}{N}\mu_g + \frac{1}{N}\mu_d.$$

En pratique, nous faisons donc un premier tirage aléatoire d'une variable aléatoire h uniforme sur $\llbracket 1, N \rrbracket$;

- conditionnellement à $H < N$, on a $E_{r,t} = e_{r,(H)}$;
- conditionnellement à $H = N$ est tiré, on a $E_{r,t} = F^{-1}(Z)$ où Z est uniforme sur $[1 - 1/N, 1]$ et où F est la fonction de distribution d'une gaussienne centrée de variance $\hat{\sigma}^2$.

Nous utiliserons dans la suite la fonction de répartition F_r de cette loi, voir à nouveau la figure II.11.

Remarquons qu'utiliser $e_{r,(H)}$ plutôt que $e_{r,H}$ n'est important que pour éviter la valeur la plus élevée ; pour maintenir les dépendances entre régions nous procéderons légèrement différemment par la suite, sans changer la loi.

(ii) Cas multidimensionnel : échantillonnage et diffusion des extrêmes

L'étude du cas à une seule région est bien sûr insuffisante : une bonne part de l'enjeu réside dans l'exploitation de la dépendance entre régions présente dans les données. Nous allons ajouter à l'échantillonnage dans l'historique le minimum de l'approche paramétrique nécessaire pour introduire la modélisation de la queue de distribution ; de la section (i), nous retenons uniquement la fonction de répartition F_r pour chaque région, et plus précisément la partie correspondant aux quantiles supérieurs à $1 - 1/N$, c'est-à-dire la queue gaussienne.

Le protocole de simulation des $\vec{E}_t$ futurs est alors le suivant :

- i. Tirage uniforme d'une année $H \leq t_0$ parmi les N historiques ;
- ii. pour chaque région r , on considère le rang L_r de la valeur obtenue parmi les valeurs historiques, c'est-à-dire telle que $e_{r,(L_r)} = e_{r,H}$;
- iii. pour chaque région r telle que $L_r < N$, on utilise la valeur $E_{r,t} = E_{r,H}$;
- iv. (« diffusion des extrêmes ») pour les régions r telle que $L_r = N$, on tire uniformément et indépendamment les uns des autres des ordres de quantiles Z_r dans $[1 - 1/N, 1]$, et on utilise les valeurs $E_{r,t} = F_r^{-1}(Z_r)$.

Ainsi, la plupart des valeurs utilisées sont reprises de l'historique, on reste au plus proche des données dans la zone non-extrême ; mais pour chaque région on a une probabilité $1/N$ d'utiliser un quantile extrême dans $[1 - 1/N, 1]$, ce qui permet d'utiliser la modélisation paramétrique avec la fréquence adéquate dans la partie de la distribution

non couverte par l'historique. De plus, les dépendances entre régions sont conservées, y compris les cas où plusieurs régions ont subi leur maximum historique en même temps. On ne modélise cependant aucune dépendance entre régions *conditionnellement* à ce qu'elles dépassent leur maximum historique, ce qui est une hypothèse parcimonieuse choisie sous la contrainte de données limitées. Avec plus de données, on pourrait modéliser la dépendance en queue de distribution à partir de la dépendance observée dans l'historique, mais il est toujours délicat de mener une telle extrapolation.

III. Applications actuarielles : projection de tendance, projection annuelle et stress test

Nous avons maintenant mené toutes les études statistiques nécessaires à la mise en place, à partir de données historiques et des données scientifiques, de projections permettant des applications actuarielles.

Nous commencerons par une projection de tendance, qui n'utilise que les données scientifiques, qui vise à projeter des augmentations de dommage dans le futur sur un terme assez long, par rapport à une période historique prise comme référence, à l'échelle de tout le territoire européen de la France. Il s'agit de traduire en termes concrets de montants de dommages l'impact prévisible du réchauffement global sur le péril inondations de plaine, à partir des informations contenues dans les données scientifiques (courbes de réchauffement global et données PESETA IV), en quantifiant autant d'incertitudes que possible. Malheureusement une part importante des incertitudes n'ont pas pu être quantifiées, et la myriade de choix effectués dans la construction des données utilisées et dans la façon dont nous les avons traitées jouent un rôle important dans les résultats, il faut donc prendre cet exercice avec beaucoup de prudence. Nous comparerons nos résultats principalement avec le rapport France Assureurs 2021 [Fra22], en prenant les mêmes périodes de projection et d'historique, mais nous ferons également un rapide parcours d'autres rapports récents sur le même sujet.

Nous montrerons ensuite comment une compagnie d'assurance pourrait utiliser le modèle qui fait l'objet du présent rapport pour projeter des montants de dommages à partir non seulement des données scientifiques, mais aussi d'un historique de dommage sur leur portefeuille. Faute de disposer de données assurantielles réelles, nous mènerons ces projections sur des données fictives ; aucun des résultats spécifiques ne sont donc généralisables, l'objectif est de montrer le principe d'application. Nous tirerons tout de même quelques enseignements de portée plus large.

Dans un premier temps, nous considérerons une projection à un an, qui forme une information essentielle pour la tarification, le provisionnement et la réassurance. Bien qu'on puisse penser que le réchauffement climatique étant lent, il ne devrait pas intervenir à cette échéance, ce serait oublier que les estimations reposent sur un historique qui est, lui, beaucoup plus long. Le modèle d'« actualisation climatique » est une approche possible pour « mettre en *as if* » l'historique, c'est-à-dire prendre en compte qu'un même temps de retour correspond aujourd'hui à un événement plus sévère que ceux qu'on a observé dans le passé.

Dans un second temps, nous mènerons des projections de long terme, sur plusieurs dizaines d'années, avec cette fois l'intégration d'une variabilité annuelle construite à partir de l'historique. On construira sur cette base un scénario adverse de sinistralité pour un stress-test climatique.

1. Projection de tendance pour la France métropolitaine

Nos premières projections visent à estimer les évolutions de long terme à une échelle géographique étendue, en intégrant une estimation des incertitudes mais en ignorant la variabilité annuelle pour obtenir une projection de tendance.

1.1. Principe de la projection de tendance

Pour faciliter les comparaisons, nous allons reprendre les intervalles de temps du rapport de France Assureurs [Fra22] et comparer la période 1989-2019 à la période 2020-2050. Pour chaque trajectoire d'émission de GES, nous allons mener un grand nombre N de situation possibles dans le cadre de notre modélisation. Rappelons que notre modèle log-linéaire s'écrit, dans le cadre d'une seule région (ici le territoire Européen de la France),

$$Dom_t = EA_t \cdot A e^{\beta \cdot \text{GWL}(t)} \cdot F_t$$

où Dom_t est le montant de dommages l'année t , EA_t est le montant des enjeux assurés, A est un facteur d'exposition hors changement climatique, β est le coefficient quantifiant l'impact du réchauffement climatique, $\text{GWL}(t)$ est le niveau de réchauffement global, et F_t représente la variabilité annuelle. Puisque nous sommes en projection de tendance, nous fixons $F_t = 1$; pour isoler l'effet du réchauffement global nous raisonnons à enjeux assurés constants et fixons arbitrairement $EA_t = 1$ (ce qui revient à considérer comme unité monétaire le montant total des enjeux assurés). Enfin, nous allons raisonner en accroissement relatif entre deux périodes, donc le coefficient A n'a pas d'importance et nous le fixons également à 1. Le modèle se réécrit alors simplement

$$Dom_t = e^{\beta \cdot \text{GWL}(t)}$$

et chaque situation possible est déterminée par

- une courbe de réchauffement global $t \mapsto \text{GWL}(t)$;
- une valeur du coefficient β traduisant l'impact du réchauffement climatique à l'échelle de la France entière ;

qui sont tirées pseudo-aléatoirement, de façon indépendante, un grand nombre N de fois.

Pour la courbe de réchauffement global, on applique les principes décrits dans la section 2.3 du chapitre II, c'est-à-dire qu'on tire une valeur de déviation δ suivant une loi normale centrée réduite et qu'on utilise la courbe $t \mapsto \text{GWL}(t)$ donnée par

$$\text{GWL}(t) = \begin{cases} m_t + \delta \sigma_{+,t} & \text{si } \delta \geq 0 \\ m_t + \delta \sigma_{-,t} & \text{si } \delta < 0. \end{cases} \quad (\text{III.1})$$

où m_t , $\sigma_{-,t}$ et $\sigma_{+,t}$ sont les paramètres de la loi normale tordue associée à la moyenne et aux quantiles de température lissée (voir section II-2.2) dans la trajectoire d'émission de GES choisie. Rappelons que m_t est une température médiane et que $\sigma_{-,t}$ et $\sigma_{+,t}$ s'approchent d'écart-types, pour des déviations vers le bas ou vers le haut respectivement.

Pour le coefficient β , la situation est plus simple puisqu'on tire simplement une valeur suivant une loi normale de moyenne et variance estimée par la procédure statistique décrite dans la section 3.2 du chapitre II.

Une fois $t \mapsto \text{GWL}(t)$ et β tirés, on calcule les différents Dom_t sur l'ensemble des années t considérées, et il ne reste qu'à comparer les moyennes sur les deux périodes considérées, notées $\overline{\text{Dom}}_{1989-2019,S}$ et $\overline{\text{Dom}}_{2020-2050,S}$ où S représente la situation qui a été tirée alatoirement (S prend donc N valeurs). On s'intéresse à l'accroissement

$$\text{AccDom}_S = \frac{\overline{\text{Dom}}_{2020-2050,S}}{\overline{\text{Dom}}_{1989-2019,S}} - 1,$$

et c'est la distribution empirique de ces N valeurs qui va nous intéresser, notamment sa médiane et son quantile à 95%.

1.2. Résultats

Les résultats obtenus avec $N = 10\,000$ répétitions sont donnés dans la table III.1 et la figure III.2, à comparer aux résultats obtenus par France Assureurs [Fra22]. Sur les mêmes périodes, pour un périmètre un peu plus large (toutes inondations liées aux cours d'eau) et dans la trajectoire RCP 8.5, ce rapport projette une augmentation de +81% en intégrant l'évolution des enjeux assurés, dont seulement 11% dus au réchauffement climatique.

Trajectoire d'émission de GES	Moyenne	Écart-type	Médiane	Quantile à 95 %
SSP1-1.9	63%	25%	61%	108%
SSP1-2.6	73%	30%	70%	127%
SSP2-4.5	85%	37%	82%	151%
SSP3-7.0	95%	43%	89%	172%
SSP5-8.5	118%	57%	110%	225%

TABLE III.1. : Statistiques des accroissements de dommages AccDom entre les périodes 1989-2019 et 2020-2050 obtenus avec $N = 10\,000$ répétitions dans les principales trajectoires d'émission de GES du 6^e rapport du GIEC.

Observons donc tout d'abord que nos résultats sont beaucoup plus pessimistes que ceux de France Assureurs, avec une médiane à +110% pour la trajectoire d'émissions de gaz à effets de serre SSP5-8.5 (proche de RCP-8.5), on le rappelle à *enjeux assurés constants*. Nous approfondirons la comparaison avec cette étude et d'autres un peu plus bas, et nous reviendrons sur les raisons qui peuvent expliquer ces écarts.

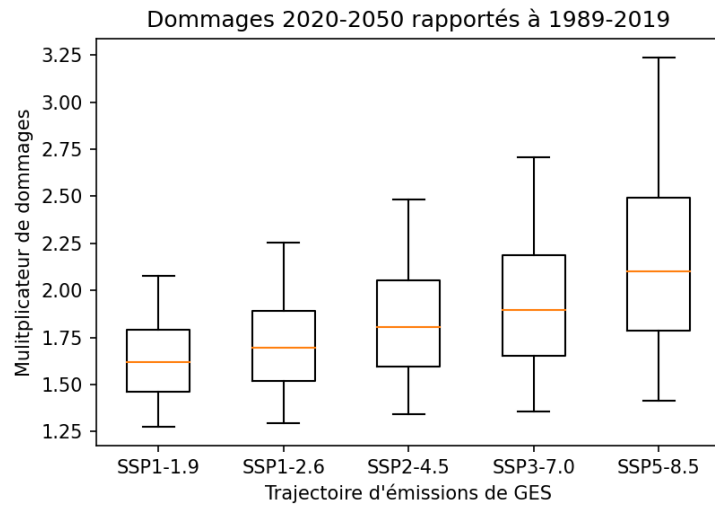


FIGURE III.2. : Distribution des rapports de dommages ($AccDom + 1$) entre les deux périodes, pour $N = 10\,000$ répétitions. Sont indiqués les quantiles 5%, 25%, 50%, 75% et 95%.

Un atout de notre approche est de permettre facilement de considérer différentes trajectoires d'émission de GES, et d'estimer une partie des incertitudes. *Attention*, par les choix faits au long de la modélisation du chapitre II et comme plusieurs sources d'incertitudes nous sont inaccessibles, il faut considérer que *nos incertitudes sont sous-évaluées*, en particulier la valeur donnée pour le quantile à 95% est plutôt une borne inférieure.

On voit que le choix de trajectoire a une influence importante sur les projections, puisqu'on passe d'un accroissement médian de 61% dans SSP1-1.9 à 110% dans SSP5-8.5. Les incertitudes sont également importantes puisque l'écart-type approche la moitié de la moyenne,¹ et le quantile à 95% peut être jusqu'à deux fois plus élevé que la valeur médiane pour une trajectoire donnée. Ainsi, même pour la trajectoire centrale SSP2-4.5, notre modélisation n'exclut pas une multiplication de la sinistralité par 2,5 en 30 ans (toujours en € constants et à enjeux assurés constants).

1.3. Comparaison bibliographique

Nous avons déjà évoqué la comparaison avec [Fra22], mais deux autres rapports récents ont porté sur la question. Nous allons présenter ici les méthodes et résultats concernant les inondations des trois études, synthétisés dans la table III.3.

¹Un rapport moyenne sur écart-type supérieur à 2 est parfois considéré comme une bonne robustesse, et permet en effet de considérer le signe de l'évolution, mais laisse un large intervalle de valeurs plausibles.

(i) Méthodes du rapport France Assureurs

Le périmètre considéré par le rapport France Assureurs [Fra22] inclut toutes les inondations liées aux cours d'eau, à l'exclusion donc des submersions marines traitées à part.

Pour l'évolution des enjeux assurés, il utilise les données démographiques projetées par l'INSEE, les données économiques du Rapport Économique, Social et Financier du PLF 2022 et du Conseil d'Orientation des Retraites, et les données sur les zones d'aléas de la Mission Risques Naturels. Le degré d'exposition aux inondations est extrapolé à partir de la tendance par département depuis 1990 (en légère baisse).

Pour la modélisation du climat futur, le rapport s'appuie sur des données fournies par le Laboratoire des Sciences du Climat et de l'Environnement (LSCE). Il s'agit de séries chronologiques de données météorologiques (température, précipitations, etc.), mesurées pour la période historique et projetées pour la période future. Les projections sont basées sur la trajectoire RCP 8.5, et obtenues en deux jeux fournis par deux modèles climatiques (RCM et GCM, voir 2.2 dans le chapitre I), de l'Institut Pierre Simon Laplace de Paris (IPSL) d'une part et l'institut Max Planck Institute of Meteorology de Hambourg (MPI) d'autre part. Les résultats présentés sont la moyenne des deux. On peut noter que ces deux modèles s'inscrivent dans la 6^e phase du *Coupled Model Intercomparison Project* (CMIP6), sur lequel se base le 6^e rapport du GIEC, tandis que PESETA IV a utilisé les modèles EURO-CORDEX qui relèvent eux de la 5^e phase (CMIP5) et sont donc plus anciens.

Les différentes données météorologiques sont confrontées aux indemnisations passées pour retenir la plus pertinente, en l'occurrence le quantile à 90% des précipitations maximales quotidiennes, par département. Celle-ci sert alors de variable de calibration de la fréquence/temps de retour des inondations, ce qui permet en croisant avec les évolutions socio-économiques de projeter les dommages futurs. Notons ici une limite importante de ce choix : il n'est pas pris en compte l'influence des précipitations lointaines qui ont lieu dans le bassin versant du cours d'eau, lequel dépasse en général largement le département.

(ii) Méthodes du rapport CCR

Le rapport de la CCR [CCR23] se base sur le modèle climatique atmosphérique ARPEGE-Climat développé par le Centre National de Recherches Météorologiques (unité mixte CNRS-Météo France) et le Centre Européen de Prévision (CEP à Reading, U.K.). Trois séries de 400 ans à climat constant ont été produites par Météo France : à partir d'hypothèses fixes (notamment la concentration de l'atmosphère en GES) dépendant de l'année cible et de la trajectoire d'émission de GES choisies, on produit 400 réalisations potentielles d'une année, permettant de voir apparaître des événements rares. Les trois séries correspondent à l'année de référence 2000, à l'année 2050 en trajectoire RCP 4.5 et à l'année 2050 en trajectoire RCP 8.5. ARPEGE-Climat joue le rôle d'un couple GCM-RCM (voir 2.2 dans le chapitre I) de façon unifiée, avec une maille de taille variable, qui peut être choisie plus fine au niveau de la région d'intérêt, ici l'Europe et l'Afrique du Nord. Une interpolation a été menée pour augmenter encore la résolution au niveau de

la France.

Pour les inondations, les données de précipitations sont extraites de ces séries et viennent nourrir le modèle de la CCR, qui projette successivement le ruissellement en tout point du territoire, puis le débit des principaux cours d'eau, puis les débordements pour aboutir pour chaque événement à une carte des niveaux d'eau. Nous sommes donc ici dans une démarche de modélisation physique depuis les hypothèses climatiques jusqu'aux sinistres, comme pour PESETA IV. Cette chaîne de modélisation n'est appliquée que pour les événements de précipitations susceptible de créer une inondation, ce qui est apprécié au niveau du bassin versant. Le périmètre des inondations est à nouveau plus large que celui de PESETA IV, puisqu'il inclut les inondations par ruissellement.

Le coût des sinistres projetés est ensuite estimé à l'aide d'un modèle de dommages, calibré à l'aide des données assurantielles très riches dont dispose la CCR, et qui croise les sinistres modélisés et des informations socio-économiques permettant de projeter les enjeux assurés. Ces informations comprennent des projections démographiques de l'INSEE, des données du ministère de l'Agriculture et de la Souveraineté Alimentaire pour les biens agricoles, la base SIRENE pour les autres biens professionnels.

(iii) Méthodes du Livre blanc Covéa

Le livre blanc Covéa [CR22] se base sur la trajectoire RCP 8.5 et utilise pour les inondations de plaine les simulations hydro-climatiques du projet SWICCA du *Swedish Meteorological and Hydrological Institute* (SMHI). Celles-ci reposent sur 4 projections climatiques (issues, comme pour PESETA IV, de l'ensemble EURO-CORDEX) et 3 modèles hydrologiques formant un total de 12 combinaisons. Cette approche a l'avantage par rapport à PESETA IV d'étendre l'approche multi-modèle au modèle hydrologique.

Les données de débits obtenues sont fournies à un modèle interne *Coventéo-Inondation* pour déterminer à quel point les temps de retour des événements extrêmes sont réduits, à sévérité fixée. Les projections de dommages utilisent ensuite les projections centrales d'évolution de la population de l'INSEE pour estimer les enjeux assurés. Étant donné que l'évolution économique n'est pas considérée, on est dans une situation intermédiaire entre un impact du seul réchauffement global (RG) et une prise en compte complète de l'évolution des enjeux assurés.

(iv) Résultats et comparaison

La table III.3 fournit une synthèse des résultats des différentes études présentées.

On constate que les résultats projetés dans ce mémoire sont les plus pessimistes de l'ensemble, le livre blanc Covéa s'en rapprochant substantiellement tandis que les rapports France Assureurs et CCR obtiennent un impact annuel moyen au moins 7 fois inférieurs aux nôtres.

Un élément notable du rapport de la CCR est l'augmentation moins sévère des inondations dans la trajectoire RCP 8.5 par rapport à RCP 4.5, principalement expliquée par une réduction d'ensemble des débits sur les grands fleuves. C'est un élément qui ne peut pas être analysé dans les deux autres études, qui ne traitent que RCP 8.5 ; notre

Étude	Écart temporel	Trajectoire	Dommages	Dont RG	Impact RG annuel
FA	31 ans	RCP 8.5	+81%	+11%	+0,34%
CCR	50 ans	RCP 8.5	+27%	+6%	+0,12%
		RCP 4.5	+43%	+19%	+0,35%
Covéa	37 ans	RCP 8.5	+110%	N.D.	(+2,03%)
Mémoire	31 ans	SSP5-8.5	N.D.	+110%	+2,42%
		SSP2-4.5	N.D.	+82%	+1,95%

TABLE III.3. : Comparaison des résultats du rapport France Assureurs [Fra22] (noté FA), du rapport CCR [CCR23] (CCR) et du livre blanc Covéa [CR22] (Covéa), et pour comparaison un extrait des résultats du présent mémoire (Mémoire). Sont indiquées les écarts entre les centres des périodes de référence et de projection, les trajectoires d'émissions de gaz à effet de serre utilisées, l'augmentation projetée des dommages pour la seconde par rapport à la première (intégrant l'augmentation des enjeux assurés, Covéa étant dans une situation intermédiaire (iii)), l'augmentation projetée pour le seul réchauffement global (RG), et la moyenne annuelle de cette dernière. La valeur pour Covéa est sur l'augmentation totale en l'absence de donnée sur la part due au réchauffement global.

approche est incapable de prendre en compte une telle éventualité, puisque le modèle d'actualisation climatique prend comme hypothèse une augmentation exponentielle des dommages avec le niveau de réchauffement global.²

(v) Discussion des écarts

On se demande naturellement d'où peuvent venir les fortes divergences des résultats obtenus avec les études de France Assureurs et de la CCR.

On peut commencer par remarquer que les estimations de ces deux études sont dans l'intervalle plausible de nos résultats : en faisant une approximation gaussienne et en considérant donc comme admissible une déviation de ± 2 écarts-types, le bas de l'intervalle est de $85\% - 2 \times 37\% = +11\%$ pour SSP2-4.5 et $118\% - 2 \times 57\% = +4\%$ pour SSP2-8.5 (Figure III.1). Il n'y a donc pas de contradiction formelle ; mais nous ne pouvons pas pour autant affirmer que les études de France Assureurs et de la CCR sont en bas de l'échelle des déroulements plausibles. Il y a en effet trop de limites à notre travail pour tirer une telle conclusion.

En terme de méthode, le livre blanc Covéa est plus proche des méthodes PESETA IV (approche multi-modèle, utilisation des simulations EURO-CORDEX, périmètre des inondations de plaine), ce qui explique sans doute en partie la proximité des résultats. Réciproquement, les écarts avec les deux autres rapports peuvent s'expliquer en

²En séparant dans les données PESETA IV agrégées sur la France les simulations RCP 4.5 de RCP 8.5, on constate que la trajectoire RCP 8.5 est légèrement plus concave par rapport au niveau de réchauffement. Cependant c'est un effet modeste et qui n'est pas susceptible de se retrouver par rapport au temps, puisque le niveau de réchauffement augmente bien plus rapidement sous RCP 8.5.

partie par les divergences en nombre de modèles climatiques, choix des simulations climatiques et périmètre ; la modélisation hydrologique utilisée par la CCR a par exemple une meilleure résolution en espace que celle de PESETA IV.

On peut interroger la part de nos choix de modélisation dans ce résultat, mais on retrouve le même ordre de grandeur dans les données PESETA IV brutes : en considérant le point de référence (période centrée sur l'année 1995) et le point de $+1,5^{\circ}\text{C}$ (correspondant dans nos données de réchauffement global à l'année 2031 pour la trajectoire SSP2 4.5) on obtient une augmentation des dommages projetés par PESETA IV pour le territoire européen de la France de $+95\%$ sur 36 ans soit en moyenne $+1,87\%$ par an, très proche de notre estimation médiane de $+82\%$ sur 31 ans qui correspond à $+1,95\%$ par an en moyenne. C'est donc dans les données PESETA IV que se trouve l'essentiel de l'écart à expliquer.

Sans remettre en cause le travail de l'équipe « inondations de plaine » de PESETA IV, force est de reconnaître que plusieurs éléments amènent à douter de la fiabilité des données que nous avons pu utiliser pour la France. Travailler à l'échelle de l'Europe entière ne leur a pas permis de documenter précisément les mesures de protection ; les dommages qu'ils modélisent pour la France sont nettement au-dessus des valeurs qu'ils ont trouvées documentées³ ; et on constate que les régions qui seraient les plus touchées par le réchauffement correspondent assez bien aux grands cours d'eau (figure II.8 du chapitre II) : Rhin, Rhône, embouchures de la Seine et de la Loire. Il est donc possible que la modélisation hydrologique utilisée « voit » mieux les plus grands cours d'eau, mais que les mesures de protection qui y sont déployées ne soient pas bien prises en compte. Enfin, le modèle de dommage est unique pour toute l'Europe, et peut donc être diversement adapté aux différentes zones.

En tout état de cause, la contribution du présent mémoire est plus dans la méthode statistique déployée que dans les résultats. Cette méthode pourrait être appliquée à d'autres données scientifiques pour obtenir une nouvelle image du risque inondation futur.

2. Génération de scénarios aléatoires de sinistralité

2.1. Principe du générateur de scénarios

Au-delà de la projection de tendance, de nombreuses applications nécessitent de prendre en compte la variabilité annuelle et les différences entre régions. Nous allons pour cela utiliser la forme générale de notre modèle :

$$Dom_{r,t} = EA_{r,t} \cdot A_r e^{\beta_r \cdot \text{GWL}(t)} \cdot F_{r,t}$$

où $Dom_{r,t}$ est le montant de dommages dans la région r l'année t , $EA_{r,t}$ est le montant des enjeux assurés, A_r est un facteur d'exposition spécifique à la région (et au portefeuille considéré) hors changement climatique, β_r est le coefficient quantifiant l'impact

³Notons que notre modèle s'appuie non pas sur les montants mais sur leurs variations relatives ; cependant si les montants sont bruités différemment aux différents niveaux de réchauffement global, cela peut impacter les coefficients β .

du réchauffement climatique dans la région, $\text{GWL}(t)$ est le niveau de réchauffement global, et $F_{r,t}$ représente la variabilité annuelle. On notera t_0 l'année de référence, qui sera par convention la dernière année de l'historique disponible.

Dans ce modèle, pour les années futures $t > t_0$, $\text{EA}_{r,t}$ va être donné par un plan de développement (par exemple supposé constant pour évaluer les risques attribuables au réchauffement climatique par rapport à la situation assurantielle actuelle) tandis qu' A_r et les paramètres de $F_{r,t}$ seront estimés sur les données historiques (section 4 du chapitre II). Cependant, cette estimation va se faire sur les $t \leq t_0$ à partir de l'expression

$$A_r \cdot F_{r,t} = \frac{\text{Dom}_{r,t}}{\text{EA}_{r,t}} \cdot e^{-\beta_r \cdot \text{GWL}(t)}$$

où le facteur $e^{-\beta_r \cdot \text{GWL}(t)}$ n'est pas entièrement connu. Les paramètres β_r et la fonction GWL vont, comme dans le cas de la projection de tendance, être considérées comme aléatoires; et pour chacune de leurs réalisations, nous devrons mener l'estimation des A_r et $F_{r,t}$ avant de pouvoir les exploiter pour les $t > t_0$.

Une fois choisie une trajectoire d'émission de GES, on procède donc de la façon suivante :

- i. on tire aléatoirement une courbe de réchauffement $t \mapsto \text{GWL}(t)$ (de la même manière que pour la projection de tendance, en suivant la méthode de la section 2.3 du chapitre II);
- ii. on tire aléatoirement les β_r , suivant une loi normale multivariée telle qu'estimée à partir des données PESETA IV (section 3.2 du chapitre II);
- iii. on déduit des deux items précédents les valeurs historiques ($t \leq t_0$) des $A_r \cdot F_{r,t}$, ce qui permet d'estimer A_r et les paramètres des $F_{r,t}$ (section 4 du chapitre II);
- iv. on tire pour chaque année future $t > t_0$ un vecteur aléatoire $(F_{r,t})_{r \in \mathcal{R}}$;
- v. on déduit de tout ce qui précède la valeur réalisée de $\text{Dom}_{r,t}$ pour toutes les régions et les années future $t > t_0$.

On obtient ainsi un *scénario de sinistralité* qui donne, pour chaque région et chaque année, un montant de dommages (en Euros constants). En produisant un nombre N élevé de tels scénarios, nous obtiendrons une gamme de projections possibles intégrant des incertitudes (sur le niveau de réchauffement global et les coefficients régionaux d'impact de ce réchauffement) et de la variabilité annuelle.⁴

2.2. Données pseudo-assurantielles utilisées

En l'absence de données réelles sur lesquelles appliquer les principes décrits, nous allons travailler sur des données fictives, construites à l'aide d'un modèle personnel basé

⁴On aurait pu aller plus loin en intégrant aussi l'incertitude d'estimation sur A_r par exemple. Cependant écrire un mémoire, comme faire de la politique, consiste à faire des choix dans un cadre contraint.

sur les distances entre capitales de régions. Ces données représentent les enjeux assurés et les montants d'indemnisation de la garantie inondation du portefeuille Multi-Risque Habitation (MRH) de la (tout aussi fictive) *compagnie A*, compagnie d'assurance IARD historiquement basée à Bordeaux. Nous avons construit deux tableaux de données indexés géographiquement (à l'échelle des anciennes régions administratives, pour coller aux données PESETA IV) et dans le temps (période 1994-2023) : l'un donnant les enjeux assurés, l'autre les montants de sinistre observés.

Les données étant entièrement fictives, leur vraisemblance est sujette à caution et surtout il n'y a aucun lien avec la temporalité réelle des inondations en France (par exemple dans nos données l'année 2000 est particulièrement sévère, mais pas 1999). Nous n'avons pas fait la distinction entre sinistres reconnus et non-reconnus comme catastrophes naturelle, car il aurait fallu ajouter une couche de modélisation et donc une dose d'arbitraire supplémentaire. Toutes les valeurs données s'entendent donc brutes de réassurance.

(i) Enjeux Assurés

Les enjeux assurés ont été construit de sorte à intégrer les deux variables de temps et d'espace. Pour le temps, on a supposé un développement continu de la compagnie avec une augmentation de 2% par an sur la période historique (en plus de l'inflation, tout étant considéré en Euros constants, disons 2023). Pour l'espace, on a supposé que l'histoire de la compagnie lui donnait plus de parts de marché sur les régions proche de l'Aquitaine où elle siège ; on a donc pondéré les enjeux assurés par le PIB de la région et par une fonction décroissante de la distance à l'Aquitaine (calculée de capitale à capitale). La figure III.4 représente les enjeux assurés en 2023. La grande valeur pour l'Île de France (et, dans une moindre mesure, pour la région Rhône-Alpes) vient du PIB élevé qui fait plus que compenser la distance à l'Aquitaine. Ces cas mis à part, on observe bien des montants globalement décroissants avec la distance. Notons que nous avons supposé que le développement de la compagnie A est, lui, uniforme et non dépendant de la distance au siège.

Pour la période de projection, on a supposé les enjeux assurés constants. Nous ne souhaitons en effet pas ici rentrer dans la question du plan de développement, étant donné l'aspect fictif de la compagnie.

(ii) Sinistres

Concernant les sinistres, nous avons utilisé un modèle aléatoire basé sur le même principe multiplicatif que le modèle décrit au chapitre II (section 1.2), mais volontairement différent sur deux points : la tendance à l'accroissement des sinistres avec le temps a été fixée à 3% par an, uniformément sur toutes les régions, et les aléas ne suivent pas une loi log-normale. Nous avons utilisé une loi issue de la normale, afin de facilement modéliser la dépendance entre région, mais avec des valeurs extrêmes (à la fois proche de 0 et $\gg 1$) plus marquées que pour une log-normale.

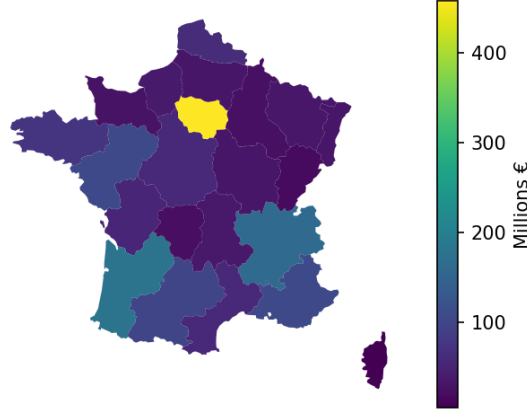


FIGURE III.4. : Enjeux assurés par la compagnie A en 2023.

Précisément, nous avons utilisé le modèle

$$Dom_{r,t} = EA_{r,t} \cdot A_r (1 + 3\%)^t \cdot F_{r,t}$$

où $Dom_{r,t}$ est le montant de dommages, $EA_{r,t}$ est le montant des enjeux assurés, A_r est un facteur basé sur une référence de 0,15% en Aquitaine, et ajusté en fonction du niveau de sinistralité de la région tel que projeté par PESETA IV, et enfin où le coefficient de variabilité annuel s'écrit $F_{r,t} = e^{(X_{r,t})^\alpha}$ avec $X_{r,t}$ un vecteur gaussien centré, et $\alpha = 1,3$. On a utilisé un écart-type de 0,8 pour les $X_{r,t}$, les différentes années sont indépendantes mais les différentes régions sont corrélées, de façon décroissante avec la distance. Les nombreux paramètres arbitraires (la fonctions de décroissance, l'exposant α , etc.) ont été réglés de façon à obtenir des résultats paraissant dans un ordre de grandeur raisonnable, notamment en terme d'ampleur des variations, mais en s'éloignant assez du modèle théorique du chapitre II afin de ne pas créer une cohérence artificielle : puisque le modèle théorique ne représente pas fidèlement la réalité mais en est une conceptualisation, il est préférable de ne pas le laisser représenter fidèlement nos données fictives non plus. La figure III.5 présente pour chaque année la somme sur les 22 régions. On observe une année exceptionnelle en 2000, présentant des dommages trois fois plus élevés que 2013 qui arrive en seconde place. Il s'agit d'une fonctionnalité volontaire du modèle génératif des données (l'exposant $\alpha = 1,3$ étant relativement élevé,⁵ la queue de la distribution est assez lourde) qui a pour but de vérifier la robustesse de la modélisation présentée en section (iii) du chapitre II.

⁵Les choix de corrélations entre régions, peut-être un peu élevées, ont pu renforcer cet effet.

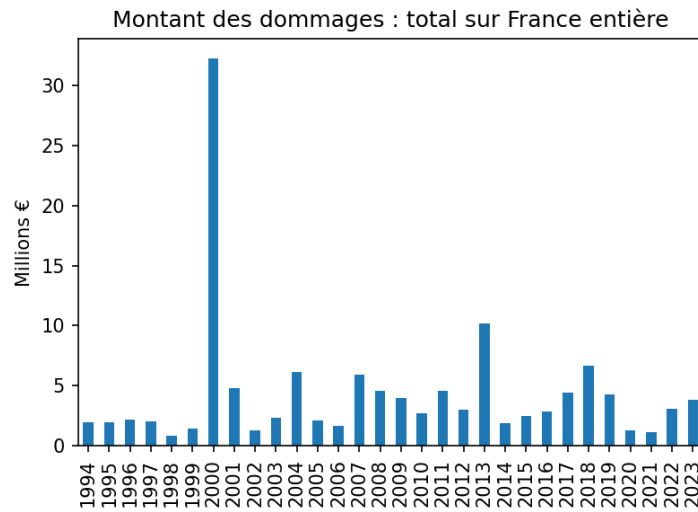


FIGURE III.5. : Dommages historiques dans nos données fictives, agrégés sur les 22 régions.

2.3. Projection sur une année

Nous commençons par des projections de court terme, supposées menées courant 2024 à l'aide des données historiques 1994-2023, visant à évaluer la sinistralité de la seule année 2025. Cette estimation peut s'inscrire dans plusieurs motivations : révision de la tarification, calcul du SCR catastrophe naturelle pour le péril inondation, analyse du plan de réassurance avec éventuel complément au régime nationale opéré par la CCR (notamment pour la part non-reconnue au journal officiel), analyse prudentielle interne. Nous ne rentrerons pas dans ces motivations, notamment parce qu'elles nécessiteraient trop d'éléments qui échappent à notre modélisation (séparation en événements pour les traités en excès de sinistre, hypothèses sur la reconnaissance, sur les traités de réassurance à évaluer, etc.)

Nous avons appliqué la méthode présentée en 2.1 sur les données assurantielles fictives de la compagnie A, dans chacune des trajectoires SSP avec $N = 200\,000$ répétitions afin d'avoir 1000 tirages au-delà du quantile empirique à 99,5%, choisi en cohérence avec les principes de la directive Solvabilité 2. Le temps de calcul reste ainsi très raisonnable, de l'ordre d'une dizaine de minutes un processeur moderne multi-cœurs d'ordinateur personnel.

Nous avons également reproduit la méthode dans un cadre sans réchauffement climatique, obtenu en supposant les coefficients β_r tous nuls (ou de façon équivalente les températures constantes), mais intégrant le même traitement de la variabilité à partir des données historiques.

Les résultats sont présentés dans la table III.6. Pour commencer par les observations banales : on voit une certaine convexité, la moyenne étant plus élevée que la médiane dans tous les cas ; et une forte variabilité avec le quantile à 99,5% un ordre de grandeur

Trajectoire	Médiane	Moyenne	quantile d'ordre 99,5%
Climat constant	3,5	5,0	29,6
SSP1-1.9	5,2	7,6	64,6
SSP1-2.6	5,2	7,6	63,1
SSP2-4.5	5,2	7,7	64,4
SSP3-7.0	5,2	7,8	65,4
SSP5-8.5	5,4	8,2	70,5

TABLE III.6. : Dommages (en M€) simulés en 2025 dans les différentes trajectoires d'émission de GES, résumé de $N = 200\,000$ réalisations du modèle.

au-dessus de la moyenne. La comparaison entre les trajectoires montre d'une part que le choix de SSP a peu d'influence sur les dommages projetés, mis à part SSP5-8.5 qui est modestement plus grave⁶ ; d'autre part que la prise en compte du réchauffement climatique fait une différence très importante puisque la moyenne est de l'ordre de 50% plus élevée et le quantile d'ordre 99,5% plus de deux fois plus élevée qu'en ignorant le réchauffement.

Le faible impact de la trajectoire SSP s'explique facilement : elles sont très proches les unes des autres sur la période 1994-2025. L'impact très fort de la prise en compte du réchauffement peut paraître surprenant sur une projection à 1 an, mais s'explique par la durée de l'historique. En effet, les sinistres passés seraient, à temps de retour égal, plus graves aujourd'hui – c'est ce que le modèle utilisé cherche à capturer. L'« actualisation climatique » n'est pas seulement importante pour les projections futures, mais aussi pour l'interprétation de l'historique.

Si l'on rapporte les dommages projetés au quantile 99,5% aux enjeux assurés, on obtient un *coefficient d'endommagement* qui est de 1,8% à climat constant et de 3,8% dans la trajectoire centrale SSP2-4.5. On peut comparer ceci à la pire année historique, qui est ici l'année 2000 pour laquelle le coefficient d'endommagement est de 3%. On a donc, à climat constant, une valeur supérieure sur seulement 30 ans d'historique à ce qui est projeté pour un temps de retour de 200 ans. Ceci permet de voir que notre traitement de la queue de distribution est assez conservatif pour les données considérées. L'origine en est que la queue de distribution est calibrée sur la moitié haute de l'historique, et avec une loi à queue moins lourde que celle utilisée pour construire les données. L'écart important de l'année 2000 a ainsi été traduit par notre traitement comme une exceptionnalité bien au-delà de la fenêtre de 30 ans.

En conclusion, bien que nous ne soyons pas rentré dans les subtilités d'un cadre assurantiel non-fictif (progiciels CAT-NAT avec modules d'aléa, de vulnérabilité et financier notamment⁷), nous voyons que le calibrage sur données historique présente une difficulté

⁶Les autres écarts sont attribuable au pseudo-aléa, que 200 000 répétition ne suffisent à gommer complètement.

⁷Le modèle présenté dans ce mémoire peut en principe s'adapter assez facilement aux sorties de ces

puisque l'on ne peut considérer être en régime permanent.

Une « mise en as-if » incluant le réchauffement global entre la période historique et le présent semble indispensable pour ne pas sous-estimer, potentiellement grandement, les risques auquel l'assureur est exposé.

2.4. Scénarios de long terme pour un stress-test

Pour notre dernière application du modèle, nous supposons que la compagnie A souhaite réaliser un stress-test climatique à partir de projections de long terme. La période 2025–2049 est choisie, et nous maintenons ici à nouveau les enjeux assurés constants.

(i) Résultats agrégés de multiples scénarios de long terme

Nous avons lancé $N = 20\,000$ fois le générateur de scénarios présenté en section 2.1, en retenant pour chacun :

- la moyenne annuelle (ce qui donne la même information que la totalité des dommages sur la période, mais est plus facilement comparable aux résultats des projections à 1 an ;
- les dommages projetés en dernière année, là où le réchauffement climatique est le plus fort ;
- les dommages projetés pour la pire année de la période.

Nous obtenons donc trois listes de N nombres, et pour chacune nous avons extrait la moyenne sur les N réalisations et le quantile d'ordre 95% pour la moyenne temporelle et la dernière année. En effet dans ces deux cas la notion de temps de retour n'est pas adaptée, et nous avons donc retenu l'usage le plus classique en terme de significativité (le quantile à 95% est la valeur la plus élevée qui ne nous amènerait pas à rejeter notre modèle si elle était observée, pour un risque α à 5%). Pour la pire année, nous avons retenu le quantile 87,5% = $1 - 1/8$ car on s'attend en moyenne à observer un événement bicentennal une fois toutes les 8 périodes de 25 ans. Ainsi, les valeurs obtenues peuvent être comparées au quantile d'ordre 99,5% pour une seule année dans les projections à un an (table III.6). Tous ces résultats sont donnés dans la table III.7 que nous allons maintenant commenter.

Considérons tout d'abord les moyennes temporelles. On voit cette fois une forte influence de la trajectoire d'émission de GES, de +80% pour la trajectoire la moins sévère à un quasi triplement de la moyenne par rapport à la simulation à climat constant. L'impact sur la queue de distribution (quantile à 95%) est encore beaucoup plus élevé (multiplication par environ 4 pour SSP1-1.9 à plus de 10 pour SSP5-8.5). Notre modèle projette donc seulement une augmentation de la sévérité des inondations, mais aussi une

progricels, comme souligné dans [JMD23].

Trajectoire	Moyenne temporelle		Dernière année		Pire année	
	Moyenne	q. à 95%	Moyenne	q. à 95%	Moyenne	q. à 87,5%
Climat constant	5,0	6,7	5,0	12,4	19,3	29,2
SSP1-1.9	9,0	28,1	9,8	92,7	43,4	75,0
SSP1-2.6	9,8	34,4	11,6	109,2	47,3	81,8
SSP2-4.5	10,9	45,9	15,0	154,6	54,7	93,6
SSP3-7.0	11,9	54,1	18,7	220,0	62,1	105,4
SSP5-8.5	14,5	76,6	24,8	281,2	78,9	135,3

TABLE III.7. : Dommages (en M€) simulés sur 2025–2049 dans les différentes trajectoires d’émission de GES, résumé de $N = 20\,000$ réalisations du modèle. La moyenne temporelle est calculée sur les 25 ans dans chaque réalisation, les autres moyennes et les quantiles portent sur les N réalisations.

augmentation encore plus importante de la variabilité ; c’est une conséquence attendue de son caractère multiplicatif.

Les effets sont bien sûr amplifiés sur la dernière année, qui d’une part est sujette au réchauffement climatique le plus important et d’autre part est une quantité moins moyennée, donc plus variable, que la moyenne sur les 25 ans (ce deuxième point est ce qui explique la différence des quantiles à 95% dans le cas d’un climat constant, entre la moyenne temporelle et la dernière année).

Enfin, la pire année fait encore l’objet de grandes variations, plus importantes que la moyenne (multiplication de la moyenne par environ 2,6 à 4,6) mais moins spectaculaires que la dernière année en ce qui concerne les extrêmes du fait d’un quantile modéré. Le point de comparaison essentiel est avec le quantile à 99,5% de la table III.6. Dans le cas d’un climat constant, on devrait avoir la même valeur (on observe 29,2 contre 29,6 à un an, montrant que le nombre de répétitions n’est pas suffisant pour assurer le troisième chiffre significatif). Dans le cas des trajectoires SSP, les températures futures divergeant notablement on s’attend à un impact bien plus fort du réchauffement, et en effet la trajectoire la plus modérée présente un quantile plus élevée à 25 ans que la plus sévère sur 1 an (à temps de retour équivalent, on le rappelle).

On observe en résumé que l’impact de long terme du réchauffement global est, comme attendu, plus élevé que l’impact de court terme et fortement différencié entre les différentes trajectoires d’émissions de gaz à effet de serre.

En supposant une politique tarifaire et prudentielle adaptée aux impacts projetés, en première approximation (très grossière), on peut considérer que :

- la moyenne des moyennes temporelles représente le montant moyen de primes pures qu’il faudrait demander aux assurés ;
- l’écart entre le quantile à 87,5% de la pire année (qui agrège l’évolution climatique sur 25 ans et la notion de temps de retour bicentennal) et la moyenne des moyennes

annuelles est une marge prudentielle à réserver et/ou à couvrir par la réassurance, en moyenne sur la période.⁸

On obtient les valeurs de la table III.8, qui illustre l'effet projeté plus important du réchauffement global sur la queue de distribution : alors que les primes augmentent d'un rapport 1,8 à 2,9 suivant les trajectoires, les marges prudentielles augmentent d'un rapport 2,7 à 5,0. Attention, cette comparaison n'a pas de valeur absolue, puisqu'elle repose d'une part sur de nombreux choix de modélisation (notamment sur la variabilité annuelle) et surtout qu'elle dépend des données fictives, certainement peu réalistes, sur lesquelles nous travaillons.

Trajectoire	Primes	ratio / CC	Marge prudentielle	ratio / CC
Climat constant	5,0	1	24,3	1
SSP1-1.9	9,0	1,8	65,9	2,7
SSP1-2.6	9,8	2,0	72,0	3,0
SSP2-4.5	10,9	2,2	82,6	3,4
SSP3-7.0	11,9	2,4	93,4	3,8
SSP5-8.5	14,5	2,9	120,8	5,0

TABLE III.8. : À gauche, comparaison des primes pures (en M€) sur la période 2025-2049 (moyenne des moyennes temporelles) avec la situation en climat constant (CC) ; à droite, idem pour la marge prudentielle (écart entre le quantile à 87,5% des dommages de la pire année, correspondant à un temps de retour bicentennal, et la moyenne des moyennes temporelles).

Même sans accorder de confiance quantitative à ces chiffres, ils permettent de souligner qu'en plus de l'aggravation des inondations attendue en moyenne, déjà potentiellement importante, on peut attendre un coût supplémentaire induit par la réassurance et l'immobilisation de capital suite à une aggravation de la variabilité par le réchauffement global.

Un scénario de stress-test robuste devrait donc notamment inclure une augmentation du coût de la réassurance après les années de forte sinistralité.

(ii) Construction d'un scénario représentatif

Dans une démarche de stress-test, il convient d'isoler un scénario adverse (qu'on mettra en regard d'un scénario à climat constant par exemple) sur lequel modéliser la solvabilité de l'assureur au long des 25 années de la période de projection. Nous allons proposer une démarche de choix de scénario, mais nous ne sommes pas en mesure ici de mener

⁸On utilise ici le terme générique de *marge prudentielle* car on s'écarte du (sous-module de) SCR *solvabilité 2* en plusieurs points ; il est notamment brut de réassurance, ne correspond directement ni à la formule standard ni à un modèle interne complet, etc.

le stress-test lui-même, puisqu'il demanderait l'intégration du scénario dans un modèle financier de l'assurance, calculant notamment la solvabilité.

Tout d'abord, il convient de choisir une trajectoire d'émission de GES. Le scénario devant être exigeant, on choisira une trajectoire de fortes émissions. D'une part la trajectoire SSP2-4.5 est plutôt centrale, d'autre part les trajectoires de forçage $8,5\text{W/m}^2$ en 2100 sont parfois considérées comme exagérément pessimistes (discuté dans [RSA⁺22], Box 3.3 *The Likelihood of High-end Emissions Scenarios*). Nous nous tiendrons à distance de cette polémique et retiendrons la trajectoire SSP3-7.0.

Ensuite, nous allons utiliser les valeurs obtenues dans la table III.7 pour choisir un scénario qui soit dans la queue de distribution. Afin de tester la résilience de la compagnie à des événements extrêmes, nous chercherons un scénario dont la pire année présente des dommages d'au moins 105,4 M€ (correspondant au quantile à 87,5%, voir ci-dessus). Pour éviter un cas trop extrême, nous convenons au préalable d'exclure un scénario dont la moyenne temporelle dépasserait 54,1 (correspondant au quantile à 95%).⁹

La figure III.9 représente les dommages totaux (France entière) pour chaque année dans le scénario retenu. On remarque 4 années très fortement touchées, même si une seule franchit (de peu) la limite imposée à la pire année, et 4 années dont l'ordre de grandeur correspond à peu près à la pire année de l'historique. La tendance temporelle n'est pas du tout évidente, ce qui est normal : à l'échelle d'une petite compagnie d'assurance, la variabilité écrase la tendance temporelle.¹⁰

Analysons maintenant la répartition par région à l'aide de la figure III.10 qui présente les coefficients d'endommagement, c'est-à-dire le rapport dommages sur enjeux assurés. On voit que dans la moyenne temporelle, l'Île-de-France domine largement (avec ensuite, assez loin derrière, les régions Rhône-Alpes et Aquitaine). Étant donné que l'on a rapporté aux enjeux assurés, il ne s'agit pas ici d'un effet de taille, mais bien de sévérité. Cette dominance se retrouve pour l'année 2043 (et également sur les trois autres années dont les dommages sont proches de 100 M€). Pour les quatre années suivantes dans l'ordre des dommages ce phénomène est moins marqué (mais encore majoritairement présent), on a représenté l'année 2034 à titre indicatif, où c'est l'Aquitaine qui subit l'endommagement le plus sévère.

Il pourrait être tentant d'en déduire que l'Île-de-France est une région particulièrement coûteuse, et d'envisager un développement du portefeuille concentré sur le reste de la France métropolitaine, voire réduire l'activité francilienne. Toutefois, le travail mené ne peut suffire à tirer une telle conclusion, et il est important de comprendre ses limites.

En premier lieu, il s'agit ici d'un unique scénario, qui ne contient pas l'ensemble des informations de nos modèles. Il faudrait donc plutôt agréger par région les résultats de nombreux scénarios. Mais ceci n'est pas suffisant, et ce n'est pas un hasard que les quatre

⁹L'écart entre les deux quantiles assure qu'un tel scénario existe (au minimum 7,5% des scénarios vérifient ces deux propriétés, dans le cas où l'ordre des scénarios est le même qu'on les trie par moyenne ou par pire année). Lors de l'application, le septième scénario tiré a été retenu.

¹⁰C'est tout l'intérêt d'un modèle comme le modèle log-linéaire utilisé ici de pouvoir traiter l'estimation de tendance à l'aide de données globale et d'ensuite appliquer cette tendance à des données assurantielles spécifiques. Comme on l'a déjà mentionné, on ne pourrait obtenir de résultat comparables en essayant de continuer une tendance lue dans les données assurantielles d'une compagnie.

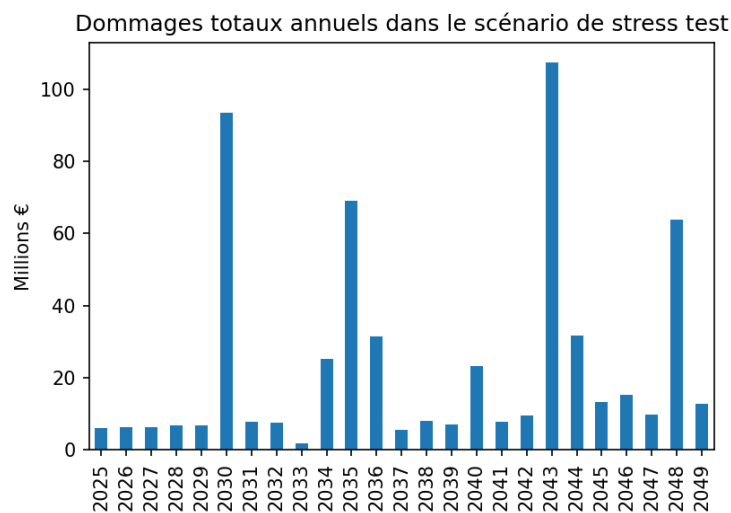


FIGURE III.9. : Dommages totaux dans le scénario de stress test retenu.

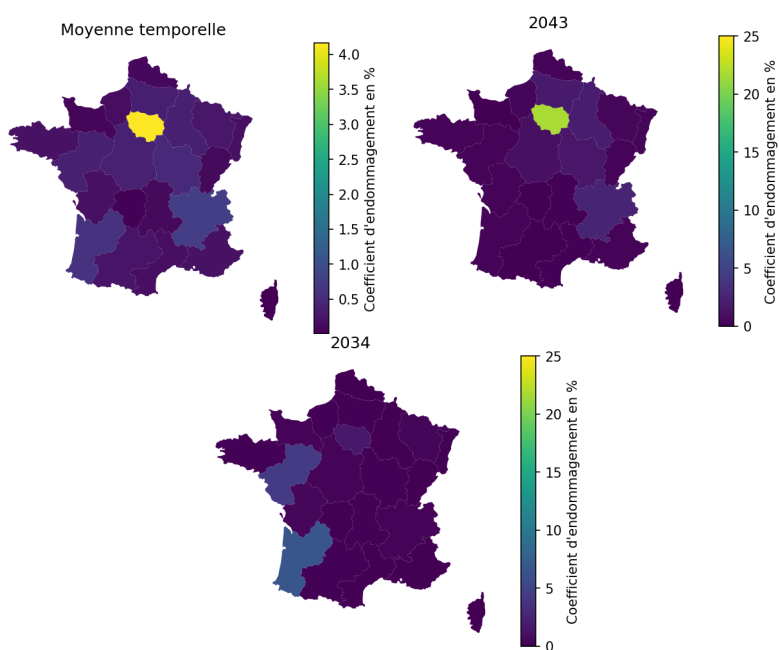


FIGURE III.10. : Coefficients d'endommagement par région. À gauche, pour la moyenne temporelle, au centre pour la pire année, à droite pour une année intermédiaire à la répartition spécifique. Les échelles sont commune pour 2043 et 2034, mais spécifique pour la moyenne temporelle.

années les plus sévères aient le même profil ; prendre en compte de nombreux scénarios aboutirait vraisemblablement à la même conclusion, qui ne pourrait cependant toujours pas être prise pour argent comptant.

En effet, ce que nous observons ici est un effet direct des données historiques. Même en mettant de côté leur caractère fictif, les années extrêmes vont mécaniquement être modélisées sur les années extrêmes de l'historique (voir chapitre II, section 4.2 pour les détails). En l'occurrence, notre historique est dominé par l'année 2000 où les dommages sont fortement concentrés en Île-de-France, et les quatre années extrêmes sont essentiellement des versions rééchelonnées et légèrement bruitées de cette année 2000. Si l'historique avait plusieurs années extrêmes avec une diversité de répartition des dommages, les années extrêmes projetées dans les scénarios seraient à nouveau des reflets de ces quelques années.

Nous retrouvons donc ici la limite de la projection basée sur un historique toujours trop court¹¹ : il nous empêche de considérer les risques qui ne se sont pas réalisés dans le passé, même quand ils peuvent devenir très importants. Les effets de seuils du phénomène d'inondation (voir le point (ii), chapitre I, section 1.2) sont susceptibles en se conjuguant aux effets du réchauffement global de faire apparaître de nouvelles régions à risque, et des risques déjà existants peuvent ne s'être pas manifestés dans l'historique par simple aléa.

En conclusion la présente approche n'est pas assez complète pour permettre de faire des prédictions sur la régionalité du risque futur pour une compagnie donnée.

Une modélisation physique sur les enjeux assurés de la compagnie, comme mené dans un progiciel CAT NAT, est plus adaptée pour se faire. L'enjeu est alors de combiner cette approche avec une mise en *as-if* prenant en compte le réchauffement climatique, ce qui peut se faire sur la base du modèle log-linéaire en recalibrant les événements modèles sur la base des niveaux de réchauffement global pour l'année de projection d'une part, et pour la période de calibrage d'autre part.

Cependant, cette observation n'empêche pas de tester la résilience de la compagnie sur le scénario extrait ; il faut simplement prendre garde à ne pas intégrer dans le stress test de décision de gestion concernant la politique régionale de souscription, qui permettraient après la première année extrême de diminuer l'exposition dans les régions touchées. L'effet de la modélisation sur historique rendant les années extrêmes sont excessivement corrélées, la diminution du risque projeté pourrait bénéficier excessivement d'une diminution régionale de l'exposition.

3. Résumé de l'étude menée

Pour conclure ce chapitre, récapitulons ce qui a été fait dans ce mémoire et ouvrons quelques prolongements possibles.

¹¹Le problème se posera même pour une compagnie d'assurance ancienne qui disposerait d'un historique très long et bien documenté : l'évolution de l'urbanisme en quelques dizaines d'années est tel que l'historique ancien ne donne plus une bonne vision des risques actuels, et encore moins futurs.

3.1. Principaux résultats obtenus

En ce qui concerne les projections de tendance à l'échelle de la France métropolitaine, nous avons obtenu des résultats inquiétants : même dans une trajectoire modérée d'émission de gaz à effet de serre (GES) comme SSP2-4.5, le réchauffement climatique induirait une augmentation médiane de 82% des dommages sur la période 2020-2050 par rapport à la période 1989-2019, à *enjeux assurés constants*. Les incertitudes qui ont pu être quantifiées sont importantes, de même que les différences entre trajectoires d'émission de GES, et les résultats obtenus ne permettent d'exclure ni une augmentation allant jusqu'à +225%, ni une augmentation très modérée.

Les valeurs médianes sont nettement plus élevées que celles obtenues dans les études récentes [Fra22] et [CCR23], et légèrement plus élevées que celles obtenue dans [CR22]. Cette sévérité ne peut pas être attribuée principalement à notre traitement statistique, elle est déjà contenue dans les données PESETA IV utilisées pour calibrer le modèle d'actualisation climatique. C'est donc dans les limites de ces données scientifiques qu'il faut chercher l'origine de cet écart. La fin de la section 1.3 présente quelques-unes de ces limites, qui sont suffisamment importantes pour considérer nos résultats comme très incertains. Ils ne peuvent donc en aucun cas être utilisés comme base de décision.

À l'échelle d'une compagnie d'assurance, nos « données assurantielles » étant entièrement fictives les résultats spécifiques obtenus ne sont pas pertinents, seulement illustratifs. Néanmoins, quelques enseignements de portée générale peuvent en être tirés.

Tout d'abord, de la même façon qu'on doit exprimer les dommages passés en Euro constants pour calibrer un modèle, une « mise en *as-if* » incluant le réchauffement global entre la période historique et le présent semble indispensable. Ne pas effectuer cette « actualisation climatique » peut induire une sous-estimation, potentiellement importante, des risques auquel l'assureur est exposé. Cette précaution est importante y compris pour des estimations de court terme, pour redresser l'historique qui, lui, s'étend sur un terme long.

Ensuite, on a pu observer que l'impact de long terme du réchauffement global est plus élevé que l'impact de court terme et fortement différencié entre les différentes trajectoires d'émissions de gaz à effet de serre. La question du choix de la trajectoire est donc un élément important dans la conduite d'un stress-test, alors qu'il est négligeable dans les estimations de court terme.

Enfin, si l'aggravation des inondations en moyenne est potentiellement importante, le réchauffement global est également susceptible d'augmenter la variabilité annuelle des dommages. Les années extrêmes risquent donc d'être nettement plus coûteuses que ce qu'on a connu dans le passé, induisant un coût accru de l'immobilisation de capital nécessaire à les couvrir et de la réassurance.

3.2. Limites de l'étude

L'étude menée dans ce mémoire présente, inévitablement, un certain nombre de défauts et de limites. Certaines ont déjà été relevé au long du mémoire, discutons quelques-une

des plus importantes.

En premier lieu, le modèle log-linaire utilisé est bien sûr simpliste. En résumant la modélisation physique menée par PESETA IV par les coefficients β_r de sensibilité au réchauffement climatique, beaucoup d'information est à priori perdue. Cette simplicité est aussi la force du modèle, en lui permettant d'être appliqué dans une société d'assurance n'ayant pas les moyens de mener une modélisation physique ; mais dans les cas où il est possible de mener la modélisation physique sur le portefeuille assuré, de façon aussi complète qu'a pu le faire PESETA IV et notamment en utilisant plusieurs modélisations de l'atmosphère, une telle approche détaillée est sans doute préférable.

Ensuite, les données PESETA IV ne sont pas sans défauts. Elles n'utilisent ainsi qu'un seul modèle hydrologique, une approche multi-modèle plus complète serait préférable. Les informations sur les mesures de protection contre les inondations (bassins de retenues, digues, etc.) que l'équipe de recherche a pu utiliser sont parcellaires, et les dommages qu'ils projettent dans la période historique s'éloignent sensiblement, dans le cas de la France, des données historiques qu'ils ont pu obtenir. Si les protections en œuvre en France ont été mal estimées, la question est alors dans quelle direction ça a pu affecter nos résultats. Si les protections réelles sont proportionnellement plus efficaces pour les événements les plus sévères, alors nos résultats sont certainement surestimés. Dans le cas contraire, il seraient plutôt sous-estimés.

Dans l'application à une compagnie d'assurance, le mémoire souffre de ne pas avoir de données réelles.¹²

Enfin les choix de modélisation de scénarios futurs rendent les projections de long terme dans le cas d'une compagnie d'assurance très dépendantes de l'historique. Un risque existant mais peu manifesté dans le passé ne sera pas pris en compte, et un événement extrême localisé peut conduire à sur-estimer le risque futur dans la région touchée. Notre approche ne semble pas assez complète pour permettre de faire des prédictions sur la régionalité du risque futur pour une compagnie donnée, problème qui mérite une étude propre.

3.3. Perspectives

De nombreuses améliorations du modèle et du traitement statistique sont envisageables, par exemple dans le chapitre II, section 3.2 on pourrait mener l'étude du modèle mixte sous sa forme matricielle (paragraphe (ii.a)), ce qui serait plus compliqué mais un peu plus précis.

Les données PESETA IV sont basées sur des simulations déjà relativement anciennes, il serait intéressant d'obtenir des données scientifiques de même nature basées sur les dernières modélisations atmosphériques. On aimerait également pouvoir appliquer les mêmes principes à d'autres risques, notamment le retrait-gonflement des argiles, sujet particulièrement sensible en France. Le problème actuellement est l'absence des données scientifiques ouvertes nécessaires, il faudrait donc envisager une coopération avec les équipes PESETA travaillant sur les sécheresses.

¹²N'étant pas employé dans le milieu de l'assurance, je n'avais d'accès à de telles données.

Il serait particulièrement intéressant de mettre en œuvre les méthodes développées ici sur des données réelles, notamment de voir comment articuler l'actualisation climatique avec un progiciel CAT-NAT : quelle prise en compte du réchauffement climatique est déjà incluse dans le progiciel ? vaut-il mieux recalibrer l'historique et donc les événements de référence, ou faire un redressement à posteriori ?

La question du futur de la réassurance dans un monde où les risques augmentent est épineuse et sans doute passionnante. Dans quelle mesure cette augmentation des risques est perçue différemment dans les compagnies d'assurances et celles de réassurance ? Tout comme un écart entre ces perceptions et la réalisation des risques, un écart entre ces deux perceptions pourrait créer des effets délétères. La prise en compte dans les stress-test d'une augmentation des coûts de la réassurance est également une problématique importante.

Enfin, la mesure de la régionalité du risque futur est un autre problème intéressant. Cette régionalité peut dépendre du portefeuille considéré, du fait d'expositions différenciées, et l'historique n'est pas nécessairement fiable du fait de l'évolution de l'exposition et la nécessité de prendre en compte des temps de retour extrêmement longs. Faut-il nécessairement en passer par une modélisation physique complète, ou y a-t-il un moyen plus léger d'obtenir des informations pertinentes ?

Conclusion

Dans ce mémoire, nous avons cherché à quantifier l’impact du réchauffement climatique sur les dommages induits par les inondations de plaine. Pour se faire, nous nous sommes basés sur un modèle log-linéaire d’*actualisation climatique* proposé par [JMD23], qui utilise une modélisation physique menée par le projet Européen PESETA IV. Après avoir développé un traitement statistique permettant de calibrer le modèle log-linéaire et de quantifier au mieux les incertitudes, nous avons appliqué ce modèle d’une part de façon globale pour le territoire européen de la France, d’autre part à l’échelle d’une compagnie d’assurance fictive.

Concernant la France Métropolitaine, nous projetons une aggravation des inondations de plaine bien plus sévère que les études de France Assureur [Fra22] et de la CCR [CCR23] et comparables à celles du livre blanc Covéa [CR22], avec une augmentation médiane de plus de 80% sur la période 2020-2050 par rapport à 1989–2019 à enjeux assurés constants. Cependant notre travail comporte des limites importantes, tout comme les données scientifiques sur lequel il est basé, et il faut donc en considérer les résultats avec la plus grande prudence. Le point à retenir est plutôt la capacité de la méthode à utiliser simplement des données de simulations pour mener ce type de projections.

À l’échelle d’un portefeuille de compagnie d’assurance, nous restons sur une illustration de la méthode sur données fictives, qui met tout de même en relief quelques enseignements : importance de la prise en compte du réchauffement climatique même pour des projections de court terme, et augmentation potentiellement encore plus importante de la variabilité annuelle que de la moyenne.

Plusieurs perspectives sont envisageable pour approfondir ce travail, sur le traitement statistique, l’extension à d’autres périls, l’application à un portefeuille réel, etc. Ce que je retiendrai comme conclusion majeure est qu’il paraît aujourd’hui important d’associer plus étroitement le monde assurantiel et celui de la recherche publique : leur complémentarité rend impossible pour l’un seul de ces deux monde d’obtenir une réponse aussi fiable et robuste que possible à la question du coût futurs des périls climatiques.

Annexe. Un point de vue sur la statistique

Le but de cette annexe est de présenter mon point de vue sur la statistique, qui motive notamment les choix de notations parfois peu conventionnels qui ont été faits dans ce mémoire. Il s'agit d'un point de vue de mathématicien ayant approfondi la statistique sur le tard, on reconnaîtra sans doute quelques réflexes formels plus courants dans les mathématiques fondamentales qu'appliquées. Cependant, il s'agit avant tout de prendre en compte l'application pratique de la statistique, et non de la réduire à sa partie proprement mathématique, qu'on pourrait définir comme l'étude des problèmes inverses dans le domaine des probabilités.

1. La surcharge des noms de variable

Du point de vue des notations, je dois avouer des difficultés à accepter certaines pratiques notationnelles rencontrées en statistique, plus précisément l'habitude de surcharger les noms de variable, c'est-à-dire d'utiliser le même symbole dans plusieurs sens. Ainsi, dans l'expression

$$f_{\theta|X=x}(\theta) = \frac{p_{\theta}(x)\pi(\theta)}{\int_{\Theta} p_{\theta}(x)\pi(\theta) \, d\nu(\theta)}$$

de la densité de la loi à posteriori en statistique bayésienne, le symbole θ joue trois rôles différents :

- i. dans $f_{\theta|X=x}$ il qualifie la densité qui est calculée, pour signifier qu'elle se rapporte au paramètre usuellement dénoté θ . Si on appliquait la formule à la valeur $\theta = 2,3$, ce ' θ '-ci ne serait pas remplacé par 2,3 ;
- ii. dans $f_{\dots}(\theta)$ et $\pi(\theta)$ au numérateur, il s'agit d'une variable libre qui sert d'argument aux deux fonctions. Si on appliquait la formule à la valeur $\theta = 2,3$, ces ' θ ' seraient remplacés par 2,3 ;
- iii. dans l'intégrale au dénominateur, θ apparaît trois fois comme une variable liée, introduite par l'élément d'intégration $d\nu(\theta)$. On pourrait remplacer ces trois ' θ ' par des ' η ' sans changer le sens de l'intégrale.

La surcharge de symbole est souvent pratique, et ici on peut parfaitement la justifier ; j'ai personnellement tendance à la limiter un peu plus pour des raisons d'esthétique formelle et de didactique.

Mais au fond, le point où cette surcharge me paraît à la fois la plus systématique et la plus problématique, au sens où elle ajoute de la difficulté à l'interprétation des expressions, est la notation $\hat{\theta}$ à la fois pour l'*estimateur* et l'*estimée* (ou *estimation*) d'un

paramètre ; on peut aussi mentionner la moins systématique mais tout aussi problématique notation y_i pour dénoter à la fois des variables aléatoires et des valeurs observées. Il me semble que cette notation n'est pas pour rien dans la difficulté qu'on peut avoir, en apprenant la statistique, à se rappeler qu'un estimateur est une variable aléatoire. Plus généralement, cette notation où on peut, parfois dans la même phrase, traiter $\hat{\theta}$ comme une variable aléatoire (en écrivant par exemple $\mathbb{E}[\hat{\theta}]$ pour discuter le biais) et comme un nombre (en traitant un exemple pour lequel on trouverait $\hat{\theta} = 2,3$), participe de la confusion entre deux domaines, le théorique et le pratique, qui me semblent par nature distincts, et important à séparer dans la pratique de la statistique. Ceci m'amène à une conception de l'activité statistique en domaines distincts interagissants.

2. Les trois domaines de l'activité statistique

Lorsqu'on mobilise la statistique, par exemple dans un cadre assurantiel, on cherche à utiliser une théorie mathématique pour saisir un phénomène réel, et quantifier sur des données qu'on a sous les yeux (ou plutôt, dans notre ordinateur) des paramètres et les incertitudes sur ces paramètres. Les différents éléments engagés appartiennent à trois domaines en interaction, mais distincts :

le domaine réel, qui est celui qui nous intéresse principalement, mais qui reste insaisissable notamment du fait que le réel qui nous intéresse le plus est le réel qui se réalisera ou ne sera connu que dans le futur ;

le domaine théorique, qui regroupe toute la statistique mathématique : variables aléatoires, modèles statistiques, etc. C'est à ce domaine qu'appartiennent les énoncés du type « tel estimateur est l'estimateur sans biais de plus petite variance de tel paramètre dans tel modèle », tous les calculs d'espérance et de variance ;

le domaine pratique, qui regroupe les données disponibles et l'ensemble des traitements qu'on opère dessus, comme un calcul de moyenne empirique par exemple.

À titre d'exemple, si on s'intéresse à la sinistralité totale des inondations dans le département de l'Aude de l'année 2010, le montant de dommages effectivement réalisé relève du domaine réel, sa modélisation par une variable aléatoire $X_{\text{Aude},2010}$ (qui sera par exemple supposée être de la forme $X_{\text{Aude},2010} = f(\text{Aude}, 2010, \dots) + \varepsilon_{\text{Aude},2010}$) relève du domaine théorique, et la valeur écrite dans notre tableau de donnée, par exemple notée $x_{\text{Aude},2010}$, relève du domaine pratique.

Bien sûr, ces trois domaines ne sont pas sans lien : si l'on s'intéresse à un traitement statistique, c'est bien qu'on compte dessus pour nous aider à saisir le réel en l'appliquant à nos données (voir la figure 11).

Le domaine théorique *modélise* le domaine réel, le choix de modèle est un élément crucial de l'activité statistique et doit être guidé par la pertinence du modèle choisi par rapport au réel, lequel doit donc être compris autant que faire ce peut. Si on modélise les dommages causés par les inondations, il faut se poser la question des facteurs influençant

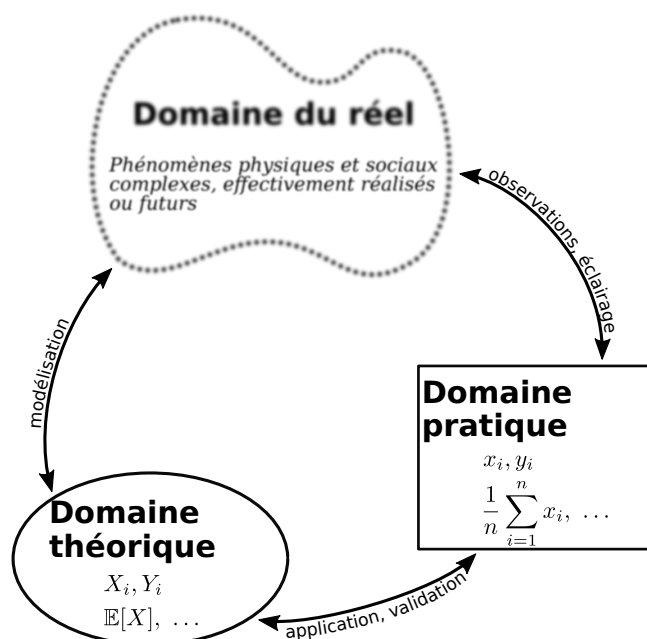


FIGURE 11. : Les trois domaines de l'activité statistique et leurs interactions.

les inondations, de comment on va pouvoir les prendre en compte, etc. Il est important de garder en tête l'adage « tous les modèles sont faux, certains sont utiles » qui traduit bien cet écart entre monde théorique et monde réel.

Le domaine du réel *fournit les données* du domaine pratique, de différente façon : observation, mesure, collecte par questionnaire, etc. Dans le cas du chapitre II, les données PESETA IV sont elles-même issues d'une modélisation, mais de nature physique et pas statistique. Il s'agit d'un cas extrême, mais il y a toujours une intermédiation, aussi bénigne puisse-t-elle paraître, entre le réel et les données disponibles : les mesures comportent des incertitudes, il peut y avoir des erreurs de saisies, etc. Il est donc aussi important de garder en tête que les données ne sont pas du « réel brut » que de se souvenir que notre modèle statistique est faux. En retour, les valeurs calculées dans le domaine pratique éclairent notre compréhension du réel.

On *applique* le modèle choisi dans le domaine théorique aux données disponibles dans le domaine pratique ; et en retour, les résultats obtenus servent à *valider ou invalider* le choix de modèle. Par exemple, si on envisage de faire une hypothèse d'homoscédasticité, on peut la valider par un test adapté. Le test lui-même est défini dans le domaine théorique, il s'appuie sur ce qu'on devrait probablement observer si le modèle était exact. L'évaluation de la statistique de test sur les données relève, elle, du domaine pratique. Si cette statistique rend l'hypothèse considérée peu plausible, on pourra changer la modélisation.

Cette dernière interaction a une grande portée sur les notations utilisées. En effet, si y_i est la i -ième valeur prise, dans nos données, par la variable à expliquer, il n'y a pas lieu

d'écrire quelque-chose comme $\mathbb{E}[y_i]$: l'espérance prend comme argument une variable aléatoire plutôt qu'un nombre. On est donc amené à distinguer la variable aléatoire Y_i qui modélise la i -ième valeur préalablement à son observation, et dispose donc d'une loi, d'une espérance, etc. de y_i qui représente un simple nombre, parfaitement déterministe, qui pourra d'ailleurs être lu directement dans nos données. Une notion comme la moyenne empirique se décline elle-même en deux versions, une dans chaque domaine :

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

relève du domaine du théorique, et on peut dire que (si les Y_i sont IID de carré intégrable) la variable aléatoire $\bar{Y}$ est de même moyenne que Y_1 et de variance n fois plus réduite ;

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

est, elle, un simple nombre du domaine pratique et il n'y a donc pas lieu de discuter sa variance ou son espérance (bien que les informations sur $\bar{Y}$ nous renseignent sur la valeur attendue de $\bar{y}$ et sur son interprétation).

Le problème le plus saillant concerne la distinction entre estimateur et estimée : les notations usuelles en statistique rentrent en conflit avec cette distinction en notant le plus souvent les deux de la même façon, avec un chapeau et en minuscule. Par exemple si on étudie un paramètre θ , on voit souvent noté $\hat{\theta}$ à la fois pour l'estimateur (une variable aléatoire, relevant du domaine théorique, dont on peut discuter l'espérance, la variance, la loi) et pour l'estimée (une valeur explicite une fois calculée sur les données, un simple nombre relevant du domaine pratique). Ceci pose des difficultés didactiques, à la fois en surchargeant le symbole $\hat{\theta}$, et en dérogeant à l'habitude d'utiliser des capitales pour les variables aléatoires. Dans ce mémoire, j'ai choisi pour ces raisons de m'astreindre à la fois à distinguer estimateur et estimée dans les notations, et à utiliser des capitales pour toutes les variables aléatoires.

Index

- A , 45
- A_r , 22
- α_r , 22
- AR6, 6
- α_s , 40
- Bilan radiatif, 4
- β_r , 22
 - β , 38
- B_s , 40, 41
- $\hat{B}_s$, 40
- β_s , 40
- $\hat{\beta}_s$, 40
- $\hat{\beta}_\infty$, 46
- Catastrophe naturelle, 14
- CCR, 6
- Climatique
 - niveau de réchauffement, 4
- climatique
 - Changement, 3
 - Réchauffement, 3
- Domages
 - $Dom_{r,t}$, 21
 - dom_{rsw} , 37
 - dom_{sw} , 38
- Enjeux assurés
 - $EA_{r,t}$, 21
- Ère pré-industrielle, 5
- $E_{r,t}$, 22
- ε'_s , 44
- E'_s , 40
- E''_s , 41
- E_{sw} , 40
- Facteurs aléatoires
 - $F_{r,t}$, 21
- Forçage radiatif, 4
- Gaz à effet de serre, 4
 - GES, 4
- GCM, 13
- GES, 5
- GIEC, 4
- GSAT, 5
- GWL, 4
- Inondation, 6
 - de plaine, 6
- IPCC, 4
- Loi normale tordue, 31
- Mesures répétées, 38
- Modèle de circulation
 - globale, 13
 - régionale, 13
- Modèle de dommage, 14
- Modèle hydrologique, 14
- Modèle log-linéaire, 22
- Modèle mixte, 38
- Moyenne glissante, 24
- poids
 - p_s , 41
 - $p_{s,k}$, 45
 - $p_{s,\infty}$, 46
- Péril climatique, 14
- Période pré-industrielle, 23
- Quasi-ergodique, 19

RCM, 13
 RCP, 12
 Risque (facteur de)
 $R_{r,t}$, 21
 Région, 37
 $\mathcal{R}$, 37
 r , 21

 $\hat{S}^2$, 45
 $\hat{\Sigma}^2$, 44
 $\hat{\sigma}^2$, 44
 $\hat{s}^2$, 45
 σ_s^2 , 40
 Scénario de température, 31
 Simulation
 $\mathcal{S}$, 37
 σ_k^2 , 45
 SSP, 13

 Taux d'actualisation climatique, 22
 β_r , 22
 Température atmosphérique de surface,
 23
 $\hat{\tau}_s^2$, 41
 τ_s^2 , 40

 U_s , 40

 $\mathcal{W}$, 37
 Whittaker et Eilers (lissage), 27
 w_i , 40

 Y_i , 40

 z , 42
 Ω , 45

Bibliographie

- [ANC⁺21] Arias, P., N., B., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storch, T., Thorne, P., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J., Cassou, C., Cherchi, A., Collins, W., Collins, W., Connors, S., Corti, S., Cruz, F., Dentener, v., Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblas-Reyes, F., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestad, J., Fyfe, J., Gillett, N., Goldfarb, L., Gorodetskaya, I., Gutierrez, J., Hamdi, R., Hawkins, E., Hewitt, H., Hope, P., Islam, A., Jones, C., Kaufman, D., Kopp, R., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T., Meinshausen, M., Min, S.-K., Monteiro, P., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A., Ruiz, L., Salée, J.-B., Samset, B., Sathyendranath, S., Seneviratne, S., Sörensson, A., Szopa, S., Takayabu, I., Tréguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X. and Zickfeld, K. (2021). *Technical Summary*. In *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. doi :10.1017/9781009157896.002. 20
- [BCL⁺24] Balaguru, K., Chang, C.-C., Leung, L. R., Foltz, G. R., Hagos, S. M., Wehner, M. F., Kossin, J. P., Ting, M. and Xu, W. (2024). *A Global Increase in Nearshore Tropical Cyclone Intensification*. *Earth's Future*, 12(5) :p. e2023EF004230. doi :10.1029/2023EF004230. 21
- [Bow23] Bowell, A. (2023). *The Perfect Way to Smooth Your Noisy Data*. <https://towardsdatascience.com/the-perfect-way-to-smooth-your-noisy-data-4f3fe6b44440>. Consulté le 17 mai 2024. 32, 35
- [Bow24] — (2024). *How to Tune the Perfect Smoother*. <https://towardsdatascience.com/how-to-tune-the-perfect-smoother-bcc5a67660b1>. 21 mai 2024. 37
- [Boy22] Boyeau, C. (2022). *Évaluation de l'impact du changement climatique sur le risque inondation en France métropolitaine*. 3

- [BP15] Brown, H. and Prescott, R. (2015). *Applied Mixed Models in Medicine*. John Wiley & Sons. 47
- [Bro21] Brousse, C. (2021). *Modélisation de l'impact du changement climatique sur le risque de catastrophes naturelles causées par les inondations*. 3
- [CCRa] CCR. *Garantie catastrophe naturelle*. <https://catastrophes-naturelles.ccr.fr/garantie-cat-nat>. Consulté le 31 janvier 2025. 20
- [CCRb] —. *Les catastrophes naturelles en France, Bilan 1982-2023*. https://catastrophes-naturelles.ccr.fr/documents/148935/368920/BILAN_CAT_NAT_DIGITAL_05.06.2024.pdf/2a28e49e-a238-2753-1614-ca4f2a0a5145?t=1717598473067. Consulté le 12 février 2025. x, xviii, 1
- [CCRc] —. *Périls couverts - Catnat*. <https://catastrophes-naturelles.ccr.fr/perils-couverts>. Consulté le 13 mars 2024. 8
- [CCR23] — (2023). *Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050*. <https://www.ccr.fr/documents/35794/1255983/CCR+Etude+climat+BAG+23102023+page+22mo.pdf/68b95f6e-8238-4dcc-6c56-025fa410257b?t=1698161402128>. xv, xxiii, 1, 3, 4, 77, 79, 92, 95
- [CJA22] Charpentier, A., James, M. and Ali, H. (2022). *Predicting drought and subsidence risks in France*. *Nat. Hazards Earth Syst. Sci.*, 22 :p. 2401–2418. doi :10.5194/nhess-22-2401-2022. 28
- [CMA⁺22] Cissé, G., McLeman, R., Adams, H., Aldunce, P., Bowen, K., Campbell-Lendrum, D., Clayton, S., Ebi, K., Hess, J., Huang, C., Liu, Q., McGregor, G., Semenza, J. and Tirado, M. (2022). *Health, Wellbeing, and the Changing Structure of Communities*. In *Climate Change 2022 : Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1041–1170. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi :10.1017/9781009325844.009. 7
- [CR22] Covéa and RiskWeatherTech (2022). *Changement climatique & Assurance : Quelles conséquences sur la sinistralité à horizon 2050 ?* https://www.covea.com/sites/default/files/2023-11/202202_Livre_Blanc_Cov%C3%A9a_Risques_Climatiques.pdf. xv, xxiii, 1, 3, 4, 78, 79, 92, 95
- [DD23] Ditlevsen, P. and Ditlevsen, S. (2023). *Warning of a forthcoming collapse of the Atlantic meridional overturning circulation*. *Nat Commun*, 14(4254). doi :10.1038/s41467-023-39810-w. 12

- [DMB⁺23] Dottori, F., Mentaschi, L., Bianchi, A., Alfieri, L. and Feyen, L. (2023). *Cost-effective adaptation strategies to rising river flood risk in Europe*. Nature Climate Change, 13 :pp. 196–202. doi :10.1038/s41558-022-01540-0. [x](#), [xviii](#), [13](#)
- [Eil03] Eilers, P. H. (2003). *A perfect smoother*. Analytical Chemistry, 75(14) :pp. 3631–3636. [32](#), [35](#)
- [Eme23] Emery, Q. (2023). *Etude de la déviation d’un portefeuille MRH face au risque climatique d’inondation*. [3](#)
- [FFKKG21] Fyfe, J., Fox-Kemper, B., Kopp, R. and Garner, G. (2021). *Data for Figure SPM.8*. In *Summary for Policymakers of the Working Group I Contribution to the IPCC Sixth Assessment Report*. NERC EDS Centre for Environmental Data Analysis. doi :10.5285/98af2184e13e4b91893ab72f301790db. (v20210809). [31](#), [38](#), [40](#)
- [Fra22] France Assureurs (2022). *IMPACT DU CHANGEMENT CLIMATIQUE SUR L’ASSURANCE à L’HORIZON 2050*. https://www.franceassureurs.fr/wp-content/uploads/2022/09/vf_france-assureurs_impact-du-changement-climatique-2050.pdf. Consulté le 6 mai 2024. [xiv](#), [xv](#), [xx](#), [xxiii](#), [1](#), [3](#), [4](#), [73](#), [74](#), [75](#), [76](#), [77](#), [79](#), [92](#), [95](#)
- [GK56] Gnedenko, B. V. and Kolmogorov, A. N. (1956). *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Company. [55](#)
- [HS14] Hingray, B. and Saïd, M. (2014). *Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections*. Journal of Climate, 27(17) :pp. 6779–6798. [27](#)
- [IPC14] IPCC (2014). *Climate Change 2014 : Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [14](#)
- [IPC18] — (2018). *Annex I : Glossary*. In Matthews, J., editor, *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, pp. 541–562. Cambridge University Press, UK and New York, NY, USA. doi :10.1017/9781009157940.008. [8](#)
- [IPC23] — (2023). *Summary for Policymakers*. In *Climate Change 2023 : Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. pp. 1–

34. IPCC, Geneva, Switzerland. doi :10.59327/IPCC/AR6-9789291691647.001. [7](#)
- [JMD23] Jewson, S., Maynard, T. and Dottori, F. (2023). *A service to help insurers understand the financial impacts of changing flood risk in Europe, based on PESETA IV*. Climate Services, 30 :p. 100395. doi :10.1016/j.cliser.2023.100395. [ix](#), [xvii](#), [2](#), [3](#), [28](#), [45](#), [46](#), [57](#), [86](#), [95](#)
- [Leb24] Leboucher, S. (2024). *Décarbonation des passifs d'assurance : virage ou mirage ?* L'Actuariel, 52 :pp. 24–32. [25](#)
- [LMB⁺21] Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S. and Zhou, T. (2021). *Future Global Climate : Scenario-Based Projections and Near-Term Information*. In *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 553–672. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. doi :10.1017/9781009157896.006. [32](#)
- [MAD⁺20] Mentaschi, L., Alfieri, L., Dottori, F., Cammalleri, C., Bisselink, B., Roo, A. D. and Feyen, L. (2020). *Independence of future changes of river runoff in Europe from the pathway to global warming*. Climate, 8(2) :p. 22. [28](#)
- [RSA⁺22] Riahi, K., Schaeffer, R., Arango, J., Calvin, K., Guivarch, C., Hasegawa, T., Jiang, K., Kriegler, E., Matthews, R., Peters, G., Rao, A., Robertson, S., Sebbit, A., Steinberger, J., Tavoni, M. and van Vuuren, D. (2022). *Mitigation pathways compatible with long-term goals*. In *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*, pp. 295–408. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi :10.1017/9781009157926.005. [89](#)
- [SBB⁺25] Saye, L., Bettis, O., Bedenham, G., Hampshire, O., Lenton, T. M. and Abrams, J. F. (2025). *Planetary Solvency—finding our balance with nature*. Consulté le 12 février 2025. [25](#)
- [SLDN⁺24] Schwarz, E., Leroutier, M., De Nazelle, A., Quirion, P. and Jean, K. (2024). *The untapped health and climate potential of cycling in France : a national assessment from individual travel data*. The Lancet Régional Health Europe, 39 :p. 100874. doi :10.1016/j.lanepe.2024.100874. [23](#)
- [SP10] Seabold, S. and Perktold, J. (2010). *statsmodels : Econometric and statistical modeling with python*. In *9th Python in Science Conference*. [47](#)

- [SZA⁺21] Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, v., Wehner, M. and Zhou, B. (2021). *Weather and Climate Extreme Events in a Changing Climate*. In *Climate Change 2021 : The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. doi :10.1017/9781009157896.013. [8](#), [9](#), [10](#), [11](#)
- [WFL⁺23] Wang, S., Foster, A., Lenz, E. A., Kessler, J. D., Stroeve, J. C., Anderson, L. O., Turetsky, M., Betts, R., Zou, S., Liu, W., Boos, W. R. and Hausfather, Z. (2023). *Mechanisms and Impacts of Earth System Tipping Elements*. *Reviews of Geophysics*, 61(1). doi :10.1029/2021RG000757. [12](#)