

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaire
le 07/03/2024

Par : **Florian SALAUN**

Titre : **Modélisation par apprentissage statistique du
lien température-mortalité en Open Data et
application prédictive**

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


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

Membres présents du jury de la filière

Nom : Christian-Yann ROBERT


*Membres présents du jury de l'Institut
des Actuaire*

*Entreprise : Deloitte **Deloitte.***

*Signature : **CYRIL CHALIN***


Directeur du mémoire en entreprise :

Nom : Joseph DELAWARI

Signature : 

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

Signature du responsable entreprise

Secrétariat :

CYRIL CHALIN


Bibliothèque :

Signature du candidat




Table des matières

Résumé	5
Executive summary	6
Remerciements	7
Note de synthèse	8
Abstract note	15
Introduction	22
1 Mise en contexte : santé, assurance et changement climatique	25
1.1 Le changement climatique, un nouveau défi pour les assureurs	25
1.1.1 Présentation du changement climatique et des scénarios du GIEC . .	25
1.1.2 Classification du risque climatique	27
1.1.3 Intégration du changement climatique par les assureurs	28
1.2 Les risques liés à la mortalité et à la santé pour les assureurs	30
1.2.1 Présentation du cadre réglementaire	30
1.2.2 Mortalité, longévité et contrats de rentes	31
1.3 Les conséquences du changement climatique sur la mortalité	34
1.3.1 Présentation des conséquences du changement climatique sur la mortalité et santé	34
1.3.2 Les recommandations des superviseurs pour l'intégration du changement climatique dans les risques vie	38
2 Analyse globale et nationale des facteurs de mortalité historiques	42
2.1 Présentation des bases de données retenues	42
2.1.1 Les données de mortalité	42

2.1.2	La période d'étude retenue	44
2.1.3	Les données climatiques	47
2.2	Des facteurs climatiques liés à la saisonnalité	50
2.2.1	Une saisonnalité marquée de la mortalité	50
2.2.2	La difficile quantification de la mortalité liée aux phénomènes climatiques	54
2.2.3	Comparaison des effets de la chaleur et du froid	56
2.3	Des facteurs économiques corrélés à la tendance	59
2.3.1	Une corrélation de la richesse à l'espérance de vie à l'échelle mondiale	59
2.3.2	Une augmentation continue mais ralentie de l'espérance de vie en France	60
2.3.3	Inclusion de variables socio-économiques dans les modèles de mortalité	63
3	Modélisation segmentée de la mortalité à partir de variables climatiques	65
3.1	État de l'art de la modélisation du lien température-mortalité	65
3.2	Comparaison des modèles	69
3.2.1	Cadre de la recherche	69
3.2.2	Le modèle SEG (Segmented Poisson Regression)	72
3.2.3	Le modèle CSDL (Constrained Segmented Distributed Lag)	75
3.2.4	Le modèle DLNM (Distributed Lag Non-Linear Model)	81
3.2.5	Choix du modèle	86
3.3	Pistes d'amélioration du modèle	90
3.3.1	L'inclusion d'une variable de richesse	90
3.3.2	La prise en compte de l'acclimatation	92
3.3.3	L'apport de l'humidité dans la modélisation de la mortalité	95
3.4	Segmentation de la mortalité	97
3.4.1	Choix de la segmentation	97
3.4.2	Interprétation des résultats	99
4	Projections de la mortalité à horizon 2070 par scénario climatique	110
4.1	Présentation et analyse des données de projections	110
4.1.1	Présentation des données de projections de la population vivante . . .	110
4.1.2	Présentation des données de projections des températures	111
4.1.3	Analyse de l'évolution des températures à horizon 2070	114
4.2	Projections de la mortalité à horizon 2070	122
4.2.1	Présentation de la démarche retenue	122
4.2.2	Comparaison temporelle de la mortalité prospective	125

4 | Modélisation par apprentissage statistique du lien température-mortalité en Open Data

4.2.3	Comparaison géographique de la mortalité prospective	131
4.3	Applications actuarielles, limites et pistes d'amélioration	135
4.3.1	Applications actuarielles de l'étude	135
4.3.2	Limites et pistes d'amélioration	141
	Conclusion	148
	Bibliographie	150
	Annexe A Construction des scénarios RCP	155
	Annexe B Description des scénarios SSP	157
	Annexe C Réglementation relative à l'utilisation des tables de mortalité	159
	Annexe D Incapacité, invalidité et contrats de prévoyance	161
	Annexe E Autres risques liés au changement climatique sur la mortalité	163
	Annexe F Test de surdispersion des données de comptage	167
	Annexe G Absence de biais des modèles linéaires généralisés	169
	Annexe H Vers une modélisation du lien précipitations-mortalité	171
	Annexe I Choix des projections de températures parmi l'ensemble DRIAS-2020	173
	Annexe J Étude de cas : des vagues de chaleur simulées en 2026	176
	Annexe K Fréquence et intensité des vagues de froid futures	178

Résumé

En France, depuis une cinquantaine d'années, les épisodes de chaleurs extrêmes s'avèrent de plus en plus nombreux et intenses. Selon Météo-France, cette tendance est amenée à se poursuivre dans les prochaines décennies. Il convient donc de s'interroger sur l'impact du dérèglement climatique sur la mortalité. Les températures journalières constituent une variable particulièrement intéressante pour mener une telle étude. En effet, une riche littérature a confirmé l'existence d'une variété de relations possibles entre température et mortalité, différenciées par âge, sexe, période et zone géographique.

Ce mémoire a pour objectif de proposer une analyse prospective de la mortalité par scénario climatique, à partir de données entièrement libres d'accès et à granularité fine. L'étude a ainsi été menée à l'échelle départementale et à une maille journalière. Différentes extensions des modèles linéaires généralisés (GLM) ont été comparées par une méthodologie d'apprentissage statistique. Cette dernière a mis en évidence l'apport des modèles DLNM (*Distributed Lag Non-Linear Models*), capables de reproduire fidèlement les faits stylisés ainsi que les effets des températures extrêmes sur la mortalité.

L'étude prospective menée à horizon 2070 a permis d'identifier des catégories de population à risque. Les personnes âgées ainsi que les individus de sexe féminin seraient en effet davantage sensibles aux conséquences des vagues de chaleur. Une hétérogénéité géographique significative pourrait également apparaître : les territoires fortement urbanisés, tels que l'Île-de-France, seraient davantage exposés. Par ailleurs, il pourrait exister une incertitude croissante quant à l'évolution des taux de mortalité, à cause des conséquences croissantes des vagues de chaleur futures. Enfin, une quantification de l'impact prospectif des mesures d'adaptation face à ces risques a été proposée.

Cette étude possède des applications en tarification vie et santé. De plus, elle peut être appliquée dans le cadre de l'intégration du changement climatique dans le processus ORSA.

Mots-clés : Mortalité, changement climatique, températures, open-data, apprentissage statistique, modèles linéaires généralisés.

Abstract note

Over the last fifty years or so, extreme heat events in France have become more frequent and more intense. According to Météo-France, this trend is set to continue over the coming decades. We therefore need to consider the impact of climate change on mortality. Daily temperatures are a particularly interesting variable for conducting such a study. A wealth of literature has confirmed the existence of a variety of possible relationships between temperature and mortality, differentiated by age, sex, period and geographical area.

The aim of this dissertation is to propose a prospective analysis of mortality by climate scenario, based on freely available data at a fine granularity. The study was carried out at departmental level and on a daily basis. Different extensions of generalised linear models (GLM) were compared using a statistical learning methodology. The latter highlighted the contribution of DLNM (Distributed Lag Non-Linear Models) models, which are capable of faithfully reproducing stylised facts and the effects of extreme temperatures on mortality.

The prospective study carried out up to 2070 identified population categories at risk. The elderly and women would be more susceptible to the consequences of heatwaves. Significant geographical heterogeneity could also emerge: highly urbanised areas, such as the Île-de-France region, would be more exposed. In addition, there could be growing uncertainty as to the evolution of mortality rates, due to the increasing consequences of future heat waves. Finally, a quantification of the prospective impact of adaptation measures in the face of these risks has been proposed.

This study has applications in life and health underwriting. It can also be applied to the integration of climate change into the ORSA process.

Keywords : Mortality, climate change, temperatures, open-data, statistical learning, generalised linear models.

Remerciements

Avant le développement de ce mémoire, je souhaite remercier les personnes qui ont contribué à la réussite de ce stage.

Je remercie tout d'abord les associés, Claude-Florence Chassain, Cyril Chalin et Baptiste Bréchet, qui m'ont permis d'effectuer mon stage de fin d'études au sein du cabinet Deloitte. J'adresse également mes remerciements à Joseph Delawari, tuteur de ce stage, pour son accompagnement tout au long de mon stage et pour m'avoir permis de développer ce sujet novateur. Je tiens à remercier particulièrement Alexandra Douthaud et Manuela Torres Gomez pour leurs conseils, leurs relectures, et plus globalement le temps précieux qu'elles m'ont accordé pendant l'écriture de mon mémoire. De manière générale, je remercie l'ensemble des membres de l'équipe pour leur accueil, les discussions agréables et enrichissantes à leurs côtés, ainsi que leur encadrement bienveillant au cours de mes missions actuarielles.

J'adresse également mes remerciements à Christian-Yann Robert, référent pédagogique de ce stage, pour sa grande disponibilité ainsi que les retours éclairés qu'il m'a fournis. Je remercie plus généralement l'ensemble des professeurs de l'ENSAE pour la qualité de leur enseignement, qui m'a permis d'avancer en autonomie au cours de ce stage.

Enfin, je tiens à remercier toutes les personnes qui ont contribué de près ou de loin à l'aboutissement de ce mémoire. Un grand merci à ma famille et mes proches pour leur soutien constant au cours de ce stage et, plus globalement, de ma scolarité.

Note de synthèse

Mise en contexte

Dans le cadre de l'ORSA¹, les assureurs sont encouragés à utiliser des données adaptées ainsi que des modélisations novatrices, afin de quantifier au mieux l'impact potentiel des phénomènes climatiques futurs sur leurs indicateurs actuariels. La modélisation de la mortalité future est concernée, et fait partie des enjeux majeurs pour les assureurs-vie ainsi que les institutions de prévoyance. Toutefois, les modèles ainsi que l'agrégation annuelle traditionnellement utilisés en actuariat ne permettent pas de capter la surmortalité due à des épisodes extrêmes et soudains. Par exemple, la canicule d'août 2003 a conduit au décès d'environ 15 000 personnes selon l'INSERM². Or, d'après Schneider et al. [53], de tels épisodes sont amenés à devenir de plus en plus nombreux et intenses dans les prochaines décennies. Ce mémoire a donc pour objectif d'étudier leurs conséquences sur la mortalité future.

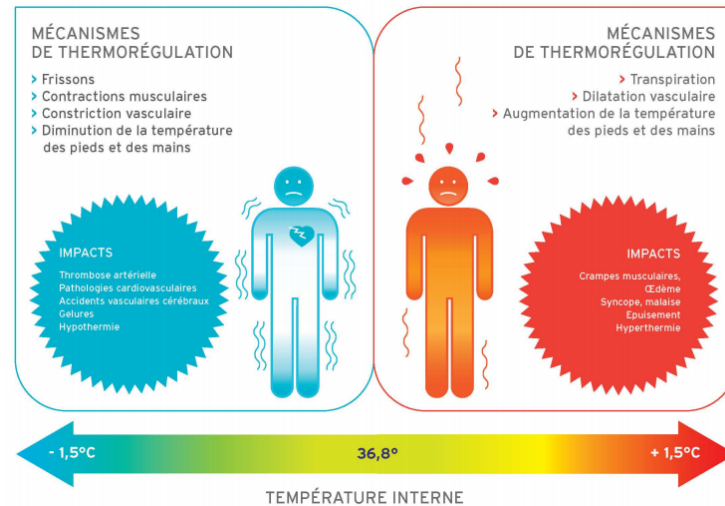
Afin de mener une telle analyse, des variables de températures journalières sont adaptées. En effet, une riche littérature (dont Tobias et al. [56]) a mis en évidence une relation singulière entre la température et la mortalité, caractérisée par un optimum thermique (minimisant le nombre de décès) ainsi qu'une surmortalité lors des épisodes de froid et de chaleurs intenses. Le mécanisme de thermorégulation peut expliquer cette relation : le corps humain cherche en permanence à maintenir sa température interne. Les températures extérieures peuvent nécessiter le recours à des mécanismes de protection internes, décrits dans le graphique 1.

D'après Santé Publique France³, si les conséquences du froid sur la mortalité pourraient être observables jusqu'à trois semaines, celles de la chaleur seraient surtout visibles à court-terme. Par ailleurs, la transpiration - qui permet de libérer la chaleur du corps - n'est possible que si l'air n'est pas saturé. Ces éléments justifient l'inclusion de variables de températures et

1. *Own Risk and Solvency Assessment*

2. Institut National de la Santé et de la Recherche Médicale. Lien vers le rapport : [↗](#)

3. Lien vers des explications détaillées : [↗](#)



GRAPHIQUE 1 – Mécanismes de thermorégulation (source : Santé Publique France)

d’humidité des jours précédents comme variables explicatives. Par ailleurs, la transpiration étant moindre chez les femmes et les personnes âgées, ces catégories pourraient être davantage affectées lors des vagues de chaleur. Enfin, l’acclimatation conduirait à ce que le risque de décès soit plus important dans une zone où la température est observée plus rarement. Ainsi, une même chaleur extrême pourrait entraîner une surmortalité moindre dans un climat méditerranéen, habituellement plus chaud.

L’ensemble de ces conjectures seront vérifiées à partir de données libres d’accès, agrégées par jour et département entre 1990 et 2019. Le fichier des décès de l’INSEE⁴ fournit les statistiques de mortalité et de population vivante, tandis que la base E-OBS⁵ permet d’obtenir des températures moyennes journalières interpolées à une maille départementale.

Modélisation de la mortalité par apprentissage statistique

Une méthode a ensuite été motivée afin d’estimer la mortalité historique à partir de variables climatiques. Les modèles linéaires généralisés (GLM) constituent le cadre de référence dans cette analyse. Les nombres de décès sont supposés suivre une loi de quasi-Poisson, adaptée à la surdispersion des données afin d’obtenir des intervalles de confiance cohérents. Un paramètre de tendance ainsi qu’une constante permettent de capturer l’évolution historique de la mortalité ainsi que les différences de mortalité par âge et sexe. Ensuite, le cadre des GLM a été progressivement complexifié afin de tenir compte des températures journalières et de leur

4. Lien vers le fichier agrégé des décès en France depuis 1970 : [↗](#)

5. Lien vers la base de données E-OBS : [↗](#)

impact sur la mortalité. Ainsi, trois extensions de complexité croissante ont été comparées :

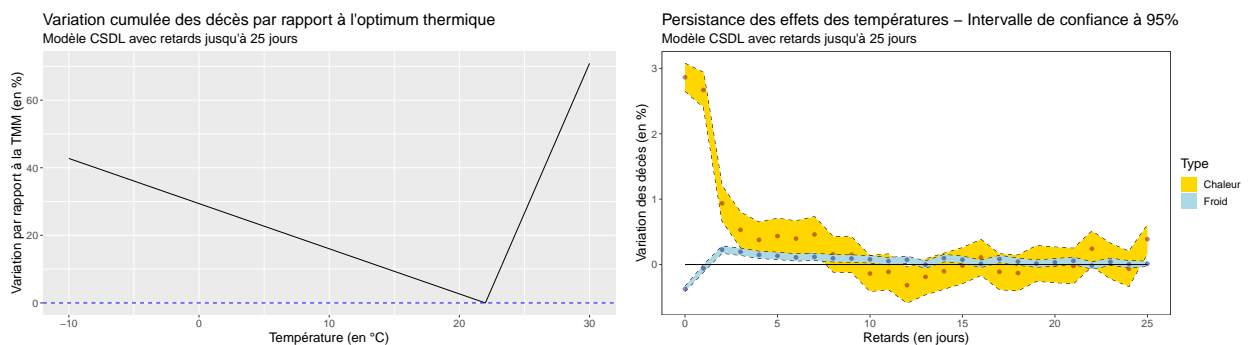
- **SEG** (*Segmented Poisson Regression*) : effets distincts de la température sur la mortalité en-dessous et au-dessus d'un seuil (appelé "optimum thermique") ;
- **CSDL** (*Constrained Segmented Distributed Lag*) : inclusion des températures des jours précédents et d'effets segmentés en-dessous et au-delà d'un seuil ;
- **DLNM** (*Distributed Lag Non-Linear Model*) : inclusion des températures des jours précédents et d'effets segmentés selon un certain nombre de morceaux de températures.

Tout d'abord, le modèle SEG fournit une relation en "V" : une température moyenne journalière de 22°C permettrait ainsi de minimiser la mortalité sur l'ensemble du territoire. Par ailleurs, la surmortalité attribuable à la chaleur serait plus marquée mais aussi plus rare. Cette segmentation des effets de la chaleur et du froid permet de réduire significativement la déviance du modèle, signe d'une nette amélioration de la qualité d'estimation.

Le modèle CSDL, en incluant des variables retardées, permet de mieux capter les effets du froid (plus persistants) par rapport au modèle SEG. Un nombre maximal de 60 retards a été fixé, puis un modèle a été estimé pour chaque nombre de retards. Une démarche en trois temps a été proposée afin de déterminer le nombre de retards à inclure :

1. Pour chaque modèle, l'optimum thermique est déterminé par validation croisée ;
2. Les modèles optimaux associés à chaque nombre de retards sont comparés via l'AIC ;
3. Une métrique du second ordre est utilisée afin de privilégier des modèles moins complexes. Ce choix permet de diminuer les temps d'estimation.

Cette démarche a conduit à retenir les températures des 25 journées précédant chaque observation. Les relations fournies par le modèle CSDL obtenu par validation croisée sont représentées dans le graphique 2.



GRAPHIQUE 2 – Relations températures-mortalité (à gauche) et intervalles de confiance à 95% (à droite) associés aux effets de la chaleur et du froid dans un modèle CSDL à 25 retards

Par rapport au modèle SEG, le modèle CSDL parvient à reproduire la persistance différenciée des effets du froid et de la chaleur, ainsi que la survenance d'un effet moisson⁶. Cependant, il ne parvient pas à capter l'existence de températures universellement mortelles, identifiées par Vecellio et al. [58]. En effet, une chaleur extrême combinée à un air saturé empêche la transpiration, et ne peut être supportée sur une longue durée. Les modèles DLNM visent à reproduire ces faits stylisés en incluant plusieurs morceaux de températures dans la modélisation. Un paramétrage est alors requis : rajouter des morceaux permet d'affiner la modélisation, mais peut aussi conduire à du surapprentissage⁷. Une synthèse des avantages et inconvénients de chaque modèle est proposée dans le tableau 3.




















Barème			
 Incohérent  Insatisfaisant  Correct  Satisfaisant			
Modèle	 Reproduction des faits biométriques	 Compromis précision / interprétabilité / tractabilité	 Reproduction de la mortalité attribuable aux canicules
GLM			
SEG			
CSDL			
DLNM			

TABLEAU 3 – Proposition de comparaison synthétique des modèles implémentés

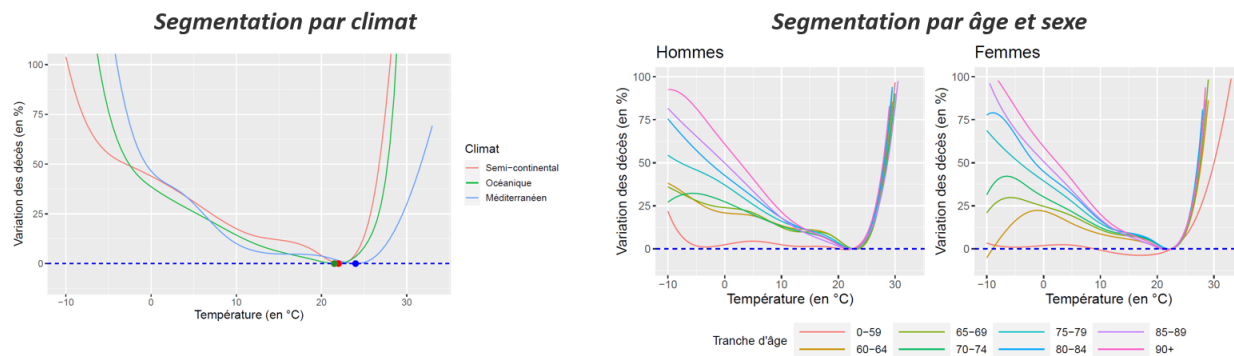
Si le modèle CSDL constitue un compromis satisfaisant entre la reproduction des faits biométriques, la précision des estimations ainsi que le temps de calcul, le modèle DLNM parvient davantage à capter la mortalité imputable aux températures extrêmes. Étant donné la hausse prédite de la fréquence et de l'intensité des vagues de chaleur futures, ce modèle semble donc le plus adapté pour établir des projections de mortalité. L'apport de la variable d'humidité n'a toutefois pas été confirmé, rejoignant les conclusions d'Armstrong et al. [2].

Enfin, une segmentation par classes d'âge, sexe et zone climatique a été construite. Celle-ci a pour objectif de reproduire les hétérogénéités des relations température-mortalité tout en conservant un nombre suffisant d'observations dans chaque classe. Les relations obtenues sont représentées dans le graphique 4.

Les conjectures précédentes ont pu être confirmées. Tout d'abord, du fait de l'acclimatation, la courbe de surmortalité attribuable à la chaleur est plus lissée dans les départements méditerranéens. Ainsi, une température moyenne journalière de 30°C est associée à une surmortalité de 25% par rapport à l'optimum thermique, alors que le nombre de décès pourrait plus que doubler dans les autres zones climatiques. Par ailleurs, la surmortalité attribuable aux températures extrêmes croît de manière significative avec l'âge. Ainsi, les effets de la

6. Période de sous-mortalité consécutive au décès d'individus amenés à décéder dans les semaines suivantes

7. Des relations incohérentes sont alors estimées pour des niveaux de températures très proches



GRAPHIQUE 4 – Relations températures-mortalité estimées par le modèle DLNM retenu en France métropolitaine entre 1990 et 2019

chaleur sont plus importants chez les personnes âgées, et particulièrement chez les femmes. Il s'agit donc de catégories à risque concernant les conséquences des pics de chaleur futurs.

Projections de la mortalité, limites et applications

Une analyse prospective de la mortalité a enfin été menée à partir de ces modèles. Celle-ci s'appuie sur les jeux de données du portail DRIAS⁸ pour les projections de températures, ainsi qu'un scénario démographique central de l'INSEE⁹. Trois scénarios climatiques de référence, avec des trajectoires de forçage radiatif différenciées, ont été comparés (RCP 2.6, 4.5 et 8.5).

Tout d'abord, des analyses descriptives sur les projections climatiques ont établi que les canicules de type 2003 sont amenées à se multiplier à l'horizon 2070. Elles pourraient aussi s'avérer particulièrement intenses et fréquentes. Ainsi, en 2065, des vagues de chaleur pourraient être étalées sur près de deux mois, contre deux semaines en 2003. De plus, l'indice thermique national¹⁰ pourrait dépasser de 6 degrés le pic atteint en 2003, signe d'une intensité sans commune mesure. En outre, ces épisodes seraient principalement concentrés autour de la région Île-de-France, fortement urbanisée et donc davantage sujette aux îlots de chaleur.

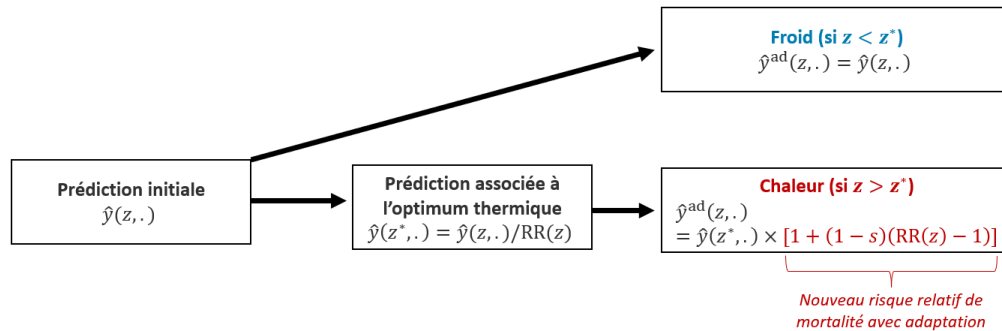
Ensuite, la mortalité a été projetée par département et par jour en France à horizon 2070. Deux premières méthodologies (a) et (b) ont été implémentées, consistant à extrapoler les relations historiques en tenant compte de plusieurs tendances possibles. Une troisième approche (c) a été proposée afin de tenir compte de la probable instauration de mesures d'adaptation. Celles-ci se traduiront par une moins grande susceptibilité des individus aux

8. Lien vers une présentation des données DRIAS : [☞](#)

9. Lien vers une description du modèle démographique Omphale : [☞](#)

10. Il s'agit d'une moyenne des températures moyennes journalières observées en France métropolitaine

phénomènes de chaleur futurs. La démarche est synthétisée dans le graphique 5.



GRAPHIQUE 5 – Méthode d’obtention des nouvelles prédictions de la mortalité avec prise en compte d’un niveau d’adaptation à la chaleur

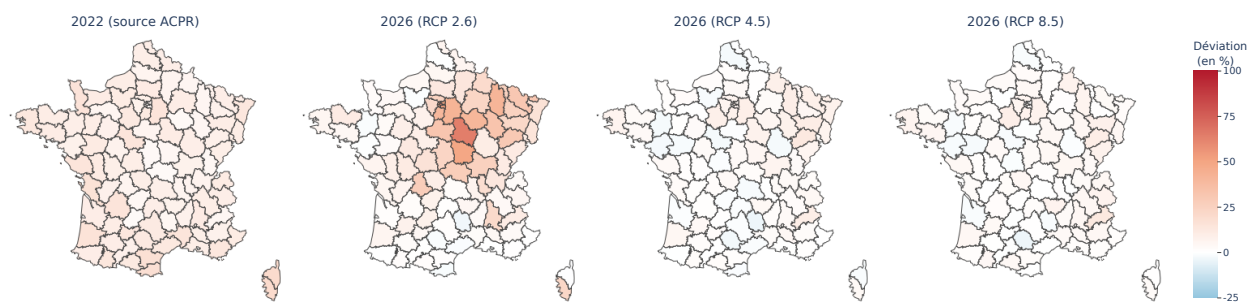
Le risque relatif de mortalité est utilisé : il s’agit du ratio de décès estimés entre une température z et l’optimum thermique z^* , toutes choses égales par ailleurs. Une réduction de $s = 30\%$ de la surmortalité attribuable à la chaleur a été retenue. Ce niveau a été obtenu par De’Donato et al. [15] suite à un ensemble de mesures d’adaptation adoptées en Italie au début du siècle.

L’analyse prospective a tout d’abord permis de suggérer une forte sensibilité des taux de mortalité annuels futurs aux vagues de chaleur. Les déviations de mortalité par rapport à la tendance historique pourraient ainsi être significatives chez les femmes lors des années de chaleurs extrêmes. Un portefeuille de personnes âgées serait également plus touché : au cours d’une année de canicule exceptionnelle, le taux de mortalité des femmes en France pourrait par exemple doubler chez les 90 ans et plus. Par ailleurs, les hétérogénéités géographiques pourraient être marquées, comme l’indique le graphique 6.

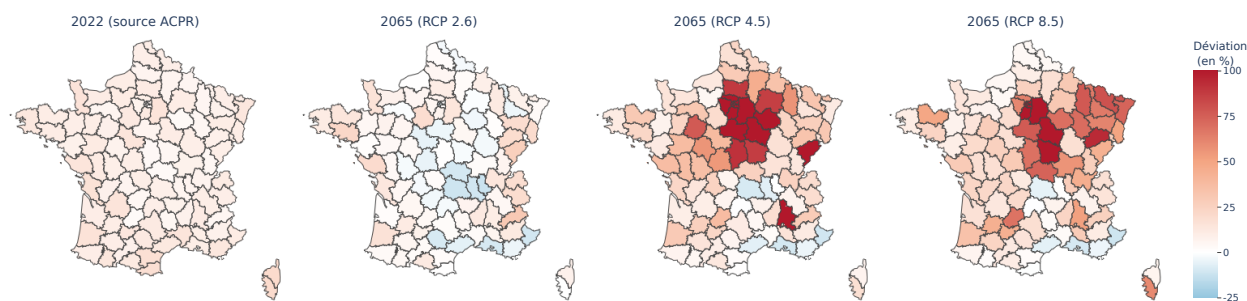
Malgré la mise en œuvre de mesures d’adaptation, les taux de mortalité globaux pourraient ainsi (au moins) doubler par rapport à 2019 en Île-de-France. Les déviations pourraient également être significatives dans les régions aux alentours ainsi que dans le Grand-Est. Au contraire, la mortalité pourrait légèrement diminuer dans le climat méditerranéen sous l’effet de la tendance structurelle à la hausse de l’espérance de vie.

Cette analyse aurait pu être enrichie par l’apport d’autres données - encore rares aujourd’hui - relatives à la qualité de l’air ou aux risques de maladies vectorielles. Elle demeure également sensible à la variable météorologique retenue : en séparant les températures de jour et de nuit, Fontès-Rousseau et al. [26] ont identifié d’autres régions à risque (Auvergne-Rhône-Alpes, notamment). Enfin, l’apport des mesures d’adaptation aurait pu être quantifié de manière plus fine, en mesurant d’abord leur impact sur les températures observées.

Déviations de mortalité prédites en 2026 par rapport à 2019 - Modèle (c)



Déviations de mortalité prédites en 2065 par rapport à 2019 - Modèle (c)



GRAPHIQUE 6 – Déviations du taux de mortalité par rapport à 2019 : en 2022 (source ACPR) et projections par scénario RCP dans un modèle DLNM avec adaptation

La méthodologie proposée ici peut être réutilisée par un assureur afin de quantifier ses risques liés à la mortalité future tout en tenant compte de la répartition géographique de ses assurés. Des applications sont, par ailleurs, possibles en tarification vie et santé. Dans ce dernier cas, une prime pure pourrait être établie en estimant des taux d'hospitalisation à partir d'une démarche similaire, puis en les associant à des données de coûts. L'étude peut également être utilisée dans un ORSA : selon les résultats d'une analyse de matérialité préalable, une analyse de scénarios pourrait être menée. Elle permettrait de quantifier l'impact sur les indicateurs actuariels des risques liés au changement climatique sur la mortalité.

Enfin, une application de la méthodologie proposée peut être envisagée afin de modéliser finement d'autres risques climatiques. Certains d'entre eux peuvent en effet être expliqués à partir de variables retardées. C'est notamment le cas du risque inondation, qui, pour un jour donné, peut être à la fois être exacerbé par des pluies modérées et continues observées lors des semaines précédentes (crues lentes), et par de fortes précipitations relevées au cours d'une même journée (crues éclair).

Abstract note

Contextual setting

As part of the ORSA¹¹ process, insurers are encouraged to use appropriate data and innovative modelling to quantify the potential impact of future climatic phenomena on their actuarial indicators. The modelling of future mortality is concerned, and is one of the major challenges for life insurers and provident institutions. However, the models and annual aggregation traditionally used in actuarial work do not capture the excess mortality due to extreme and sudden events. For example, the heatwave in August 2003 led to the deaths of around 15,000 people according to INSERM¹²- the French National Institute for Health and Medical Research. According to Schneider et al. [53], such episodes are likely to become more frequent and intense in the coming decades. The aim of this thesis is therefore to study their impact on future mortality.

In order to carry out such an analysis, daily temperature variables are suitable. A wealth of literature (including Tobias et al. [56]) has highlighted a unique relationship between temperature and mortality, characterised by a thermal optimum (minimising the number of deaths) and excess mortality during episodes of cold and intense heat. The thermoregulatory mechanism may explain this relationship : the human body is constantly seeking to maintain its internal temperature. External temperatures may require the use of internal protection mechanisms, described in the figure 1.

According to Santé Publique France¹³, while the effects of cold on mortality can be seen for up to three weeks, the effects of heat are most visible in the short term. Furthermore, sweating - which releases body heat - is only possible if the air is not saturated. These factors justify the inclusion of temperature and humidity variables from the previous days as

11. *Own Risk and Solvency Assessment*

12. Link to the report : [↗](#)

13. Link to detailed explanations : [↗](#)

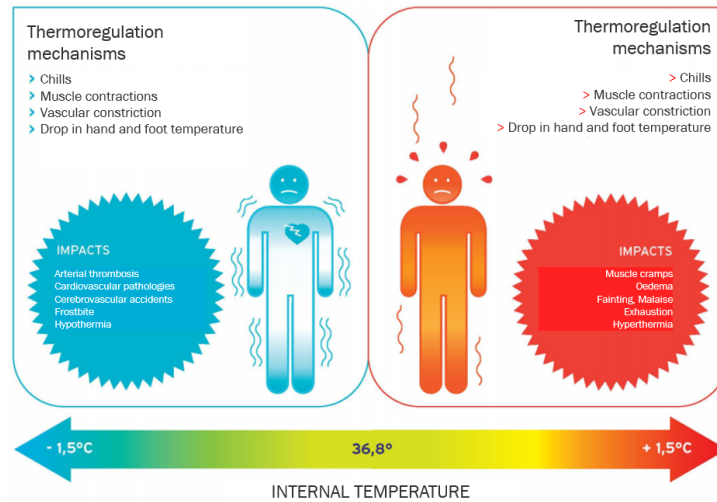


FIGURE 1 – Thermoregulation mechanisms (source : Santé Publique France)

explanatory variables. In addition, since women and the elderly perspire less, these categories could be more affected during heat waves. Finally, acclimatisation would lead to a greater risk of death in an area where temperatures are observed less frequently. For example, the same extreme heat could result in lower excess mortality in a Mediterranean climate, which is usually warmer.

All these conjectures will be verified using freely accessible data, aggregated by day and department between 1990 and 2019. The INSEE death file¹⁴ provides mortality and living population statistics, while the E-OBS database¹⁵ provides average daily temperatures interpolated to a departmental grid.

Modelling mortality using statistical learning

A method was then developed to estimate historical mortality based on climatic variables. Generalised linear models (GLM) are the reference framework for this analysis. The numbers of deaths are assumed to follow a quasi-Poisson distribution, adapted to the overdispersion of the data in order to obtain consistent confidence intervals. A trend parameter and a constant are used to capture historical mortality trends and differences in mortality by age and sex. The GLM framework was then gradually made more complex to take account of daily temperatures and their impact on mortality. Three extensions of increasing complexity were compared :

14. Link to the aggregated file of deaths in France since 1970 : [↗](#)

15. Link to the E-OBS database : [↗](#)

- **SEG** (*Segmented Poisson Regression*) : distinct effects of temperature on mortality below and above a threshold (called the "thermal optimum");
- **CSDL** (*Constrained Segmented Distributed Lag*) : inclusion of temperatures from previous days and segmented effects below and above a threshold;
- **DLNM** (*Distributed Lag Non-Linear Model*) : inclusion of temperatures from previous days and effects segmented according to a certain number of temperature chunks.

Firstly, the SEG model provides a "V" relationship : an average daily temperature of 22°C would minimise mortality throughout the country. On the other hand, excess mortality attributable to heat would be more marked but also rarer. This segmentation of the effects of heat and cold significantly reduces the model's deviance, indicating a clear improvement in estimation quality.

By including lagged variables, the CSDL model is better able to capture the effects of cold (which are more persistent) than the SEG model. A maximum number of 60 lags was set, and then a model was estimated for each number of lags. A three-stage approach was proposed to determine the number of lags to include :

1. For each model, the thermal optimum is determined by cross-validation;
2. The optimal models associated with each number of delays are compared using the AIC;
3. A second-order metric is used in order to favour less complex models. This choice reduces estimation times.

This approach led to the use of temperatures for the 25 days preceding each observation. The relationships provided by the CSDL model obtained by cross-validation are represented in the figure 2.

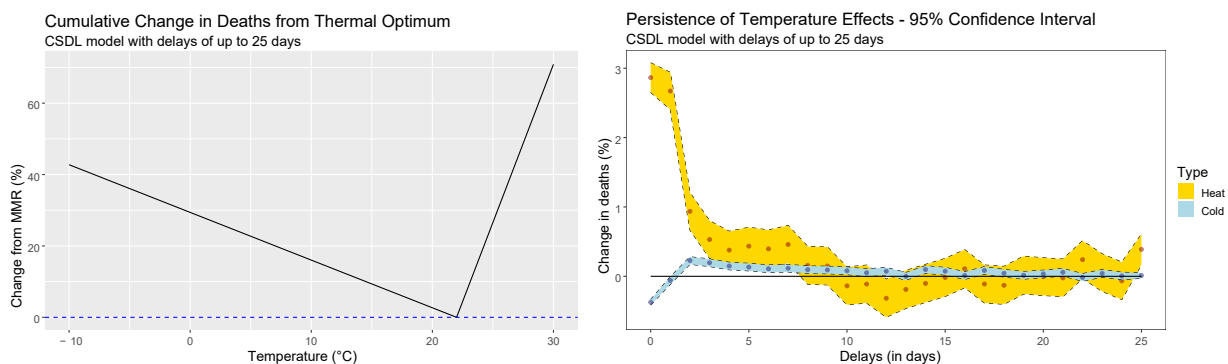


FIGURE 2 – Temperatures-mortality relationship (left) and 95% confidence intervals (right) associated with heat and cold effects in a CSDL model with 25 lags

Compared with the SEG model, the CSDL model succeeds in reproducing the differentiated persistence of the effects of cold and heat, as well as the occurrence of a harvest effect¹⁶. However, it fails to capture the existence of universally lethal temperatures, identified by Vecellio et al. [58]. Extreme heat combined with saturated air prevents transpiration, and cannot be tolerated over a long period. DLNM models aim to reproduce these stylised facts by including several pieces of temperature in the modelling. An asymptotic relationship between temperature and mortality therefore seems to exist, and justifies the use of DLNM models. Parameterisation is then required : adding chunks helps to refine the modelling, but can also lead to overlearning. Inconsistent relationships are then estimated for very similar temperature levels. A summary of the advantages and disadvantages of each model is given in the table 3.











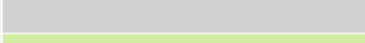

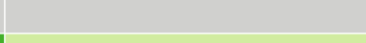




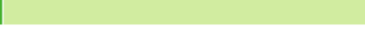

Scale			
 Inconsistent  Unsatisfactory  Correct  Satisfactory			
Model	 Reproduction of biometric facts	 Compromise accuracy / interpretability / tractability	 Reproduction of mortality attributable to heatwaves
GLM			
SEG			
CSDL			
DLNM			

TABLE 3 – Proposed summary comparison of the models implemented

While the CSDL model represents a satisfactory compromise between the reproduction of biometric facts, the accuracy of estimates and calculation time, the DLNM model is more successful in capturing mortality attributable to extreme temperatures. Given the predicted increase in the frequency and intensity of future heatwaves, this model therefore appears to be the most suitable for making mortality projections. However, the contribution of the humidity variable has not been confirmed, in line with the conclusions of Armstrong et al. [2].

Finally, a segmentation by age class, sex and climatic zone was constructed. The aim was to reproduce the heterogeneous nature of the temperature-mortality relationships, while retaining a sufficient number of observations in each class. The relationships obtained are shown in the figure 4.

Previous conjectures have been confirmed. Firstly, as a result of acclimatisation, the excess mortality curve attributable to heat is smoother in the Mediterranean departments. Thus, an average daily temperature of 30°C is associated with an excess mortality rate of 25% compared with the thermal optimum, whereas the number of deaths could more than

16. A period of low mortality following the death of individuals expected to die in the following weeks

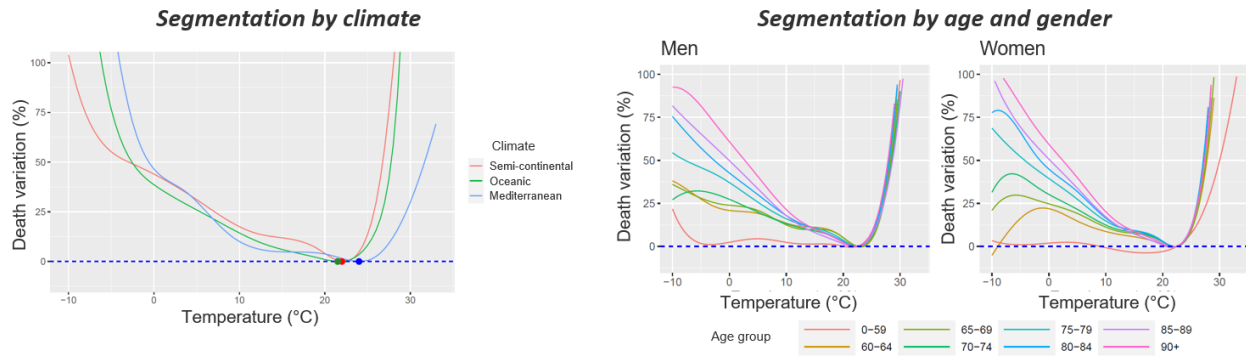


FIGURE 4 – Temperatures-mortality relationships estimated by the final DLNM model in France between 1990 and 2019

double in other climatic zones. Furthermore, the excess mortality attributable to extreme temperatures increases significantly with age. The effects of heat are greater in the elderly, particularly women. These are therefore categories at risk from the consequences of future heat peaks.

Mortality projections, limits and applications

Finally, a prospective analysis of mortality was carried out using these models. This is based on data sets from the DRIAS portal¹⁷ for the temperature projections, and a central demographic scenario from INSEE¹⁸. Three reference climate scenarios with different radiative forcing trajectories were compared (RCP 2.6, 4.5 and 8.5).

Firstly, descriptive analyses of climate projections have established that heatwaves of the 2003 type are likely to increase in number by 2070. They could also prove to be particularly intense and frequent. In 2065, heatwaves could last almost two months, compared with two weeks in 2003. What is more, the national heat index an average of daily mean temperatures observed in mainland France could be 6 degrees higher than the peak reached in 2003, a sign of unprecedented intensity. Furthermore, these episodes are likely to be concentrated mainly around the Île-de-France region, which is highly urbanised and therefore more prone to heat islands.

Mortality was then projected by department and by day in France to 2070. Two initial methodologies (a) and (b) were implemented, consisting of extrapolating historical relationships while taking into account several possible trends. A third approach (c) has been proposed to

17. [Link to a presentation of DRIAS data](#) : ↗

18. [Link to a description of the Omphale demographic model](#) : ↗

take account of the likely introduction of adaptation measures. These will result in individuals being less susceptible to future heat events. The approach is summarised in the figure 5.

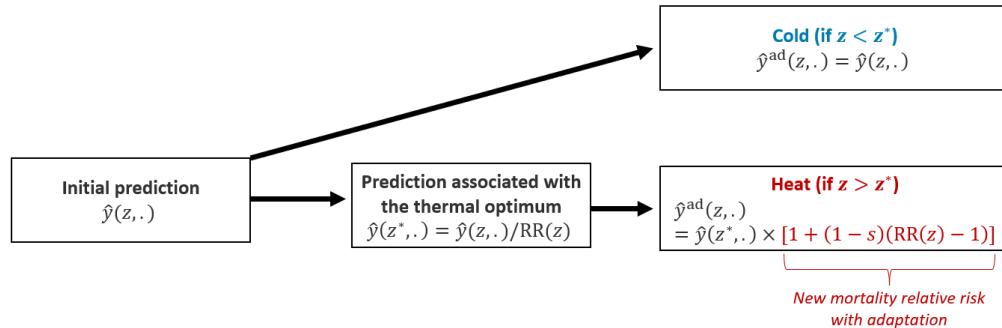


FIGURE 5 – Method for obtaining new mortality predictions taking into account a level of adaptation to heat

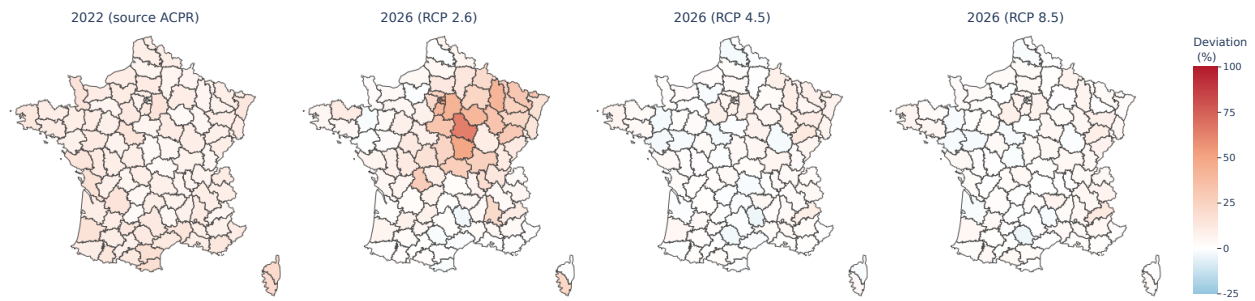
The relative risk of mortality is used : this is the ratio of deaths estimated between a temperature z and the thermal optimum z^* , all other things being equal. A reduction of $s = 30\%$ in excess mortality attributable to heat was adopted. This level was obtained by De’Donato et al. [15] following a series of adaptation measures adopted in Italy at the turn of the century.

Firstly, the prospective analysis suggested that future annual mortality rates would be highly sensitive to heat waves. Deviations in mortality from the historical trend could be significant among women in years of extreme heat. A portfolio of elderly people would also be more affected : during an exceptional heatwave year, for example, the mortality rate for women in France could double among those aged 90 and over. In addition, there could be marked geographical heterogeneities, as shown in the graph 6.

Despite the implementation of adaptation measures, overall mortality rates could therefore (at least) double compared with 2019 in the Île-de-France region. Deviations could also be significant in the surrounding regions and in the Grand-Est. On the other hand, mortality could fall slightly in the Mediterranean climate as a result of the structural upward trend in life expectancy.

This analysis could have been enriched by the contribution of other data - still rare today - relating to air quality or the risk of vector-borne diseases. It also remains sensitive to the meteorological variable selected : by separating day and night temperatures, Fontès-Rousseau et al. [26] identified other regions at risk (Auvergne-Rhône-Alpes, in particular). Finally, the contribution of adaptation measures could have been quantified in more detail, by first measuring their impact on observed temperatures.

Predicted mortality deviations in 2026 compared to 2019 - Model (c)



Predicted mortality deviations in 2065 compared to 2019 - Model (c)

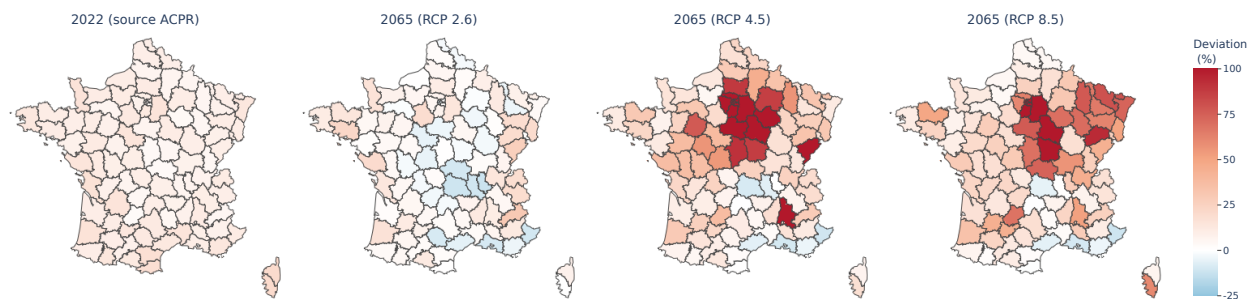


FIGURE 6 – Mortality rate deviations from 2019 : in 2022 (source ACPR) and projections by RCP scenario in a DLNM model with adaptation

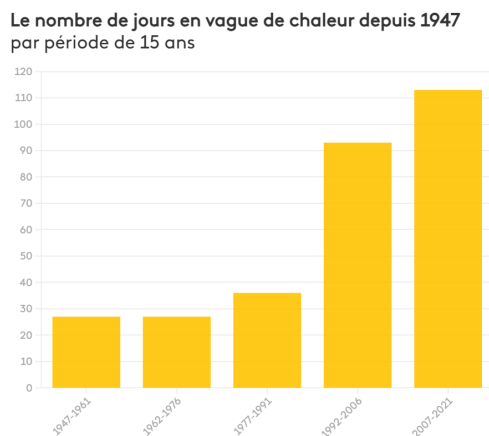
The methodology proposed here can be re-used by an insurer to quantify its future mortality risks while taking into account the geographical distribution of its policyholders. Applications are also possible in life and health pricing. In the latter case, a pure premium could be established by estimating hospitalisation rates using a similar approach, then associating them with cost data. The study can also be used in an ORSA : depending on the results of a prior materiality analysis, a scenario analysis could be carried out. This would make it possible to quantify the impact on actuarial indicators of the risks associated with climate change on mortality.

Finally, the proposed methodology could be applied to fine modelling of other climate risks. Some of these risks can be explained using lagged variables. This is particularly true of flood risk, which, on a given day, can be exacerbated by both moderate and continuous rainfall observed in previous weeks (slow flooding) and heavy rainfall on the same day (flash flooding).

Introduction

L'étude de la mortalité future est un enjeu majeur en assurance-vie, puisque la probabilité annuelle de décès est au coeur de la tarification des contrats en cas de vie ou de décès. Toutes choses égales par ailleurs, une hausse future de la probabilité annuelle de décès doit conduire l'assureur à réduire le montant des capitaux versés à l'assuré en cas de décès. Par extension, elle peut aussi avoir un rôle dans une stratégie d'allocation des actifs pour un assureur. Le risque quant à une hausse future et brutale de la mortalité pourrait ainsi l'inciter à être moins dépendant de contrats accordant à ses assurés des capitaux en cas de décès. De même, une catastrophe sanitaire pourrait toucher les mutuelles, devant rembourser un nombre important de soins de santé non-couverts par l'Assurance Maladie sur une courte durée.

Traditionnellement, en actuariat, l'analyse de la mortalité est effectuée à partir d'une agrégation annuelle. Cette approche permet de retranscrire les probabilités de décès dans des tables. Toutefois, elle ne permet pas de capturer la surmortalité liée à des épisodes extrêmes et soudains. Parmi les exemples les plus connus, figure la canicule d'août 2003. D'une longueur de 16 jours, cet épisode a occasionné le décès de près de 15 000 personnes, selon l'INSERM. De tels phénomènes se sont multipliés depuis, comme le montre le graphique 1.



GRAPHIQUE 1 – Nombre de jours en vague de chaleur depuis 1947 (à gauche ; source France Info) ; hôpital de Paris durant la canicule de 2003 (à droite ; source Le Parisien)

D'après Météo-France¹⁹, la France a en effet connu davantage de vagues de chaleur sur les 17 dernières années que lors des six décennies précédentes. Dans un contexte de réchauffement climatique, les épisodes de chaleurs extrêmes sont amenés à devenir de plus en plus fréquents et intenses, et leurs conséquences sur la mortalité accrues. Parallèlement, il est possible d'imaginer que la part de mortalité attribuable au froid serait amenée à diminuer. La littérature épidémiologique s'est largement enrichie d'études permettant de capter la relation entre la température et la mortalité, et confirmant cette tendance.

Pourtant, la modélisation de la mortalité prospective ne peut pas se résumer à capter la relation historique afin de mener des prédictions sur des données futures. En effet, les estimations de Santé Publique France relatives à la surmortalité attribuable aux vagues de chaleur ont mis en évidence des pics de mortalité largement inférieurs à ceux de l'épisode d'août 2003. Ce résultat semble contre-intuitif étant donné l'augmentation du nombre et de l'intensité des vagues de chaleur depuis cette date. Deux explications peuvent être avancées. Tout d'abord, la canicule de 2003 a permis de soulever la nécessité de mesures de prévention. Ainsi, les personnes âgées font l'objet d'une surveillance accrue, et les comportements à adopter au cours d'un tel épisode sont fréquemment rappelés dans l'ensemble des canaux de diffusion. Ensuite, la littérature a mis en avant l'existence d'une adaptation des individus à leur environnement thermique, phénomène autrement connu sous le nom d'*acclimatation*. Ne pas tenir compte de l'adaptation potentielle des individus au changement climatique pourrait conduire à sur-estimer les projections de mortalité. Enfin, d'autres politiques d'adaptation pourraient être adoptées dans les années à venir. Par exemple, la création d'espaces verts dans les grandes métropoles est envisagée afin de contrecarrer les phénomènes d'îlots urbains, et ainsi réduire la surmortalité attribuable aux pics de chaleur en ville.

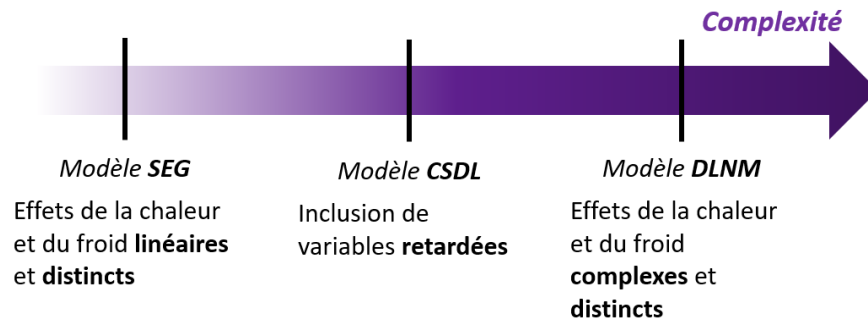
Dans ce contexte, comment l'actuaire peut-il estimer et prédire la mortalité par département, sexe et âge à horizon 2070 à partir de données publiques journalières ? Pour répondre à cette question, le fichier public des décès enregistrés en France depuis 1970 est utilisé dans cette étude, et croisé avec des données climatiques disponibles à petite échelle. Dans un premier temps, le contexte actuariel associé au changement climatique et aux risques relatifs à la mortalité et à la santé est rappelé. Il est notamment rappelé que l'ACPR demande aux assureurs, dans le cadre de l'exercice climatique 2023, une segmentation géographique du risque de mortalité à l'échelle départementale. Ensuite, cette étude s'attache à motiver le choix d'inclure des variables de température et de tendance dans la modélisation de la mortalité journalière. Une mise en relation de ce paramètre de tendance avec des indicateurs de richesse est effectuée et justifiée. Les effets différenciés du froid et de la chaleur, à la fois

19. Voir cette infographie de France Info : [🔗](#)

en termes d'intensité et de persistance, sont rappelés et justifient la nécessité d'adopter une modélisation spécifique.

Alors que la littérature propose de nombreuses extensions des modèles linéaires généralisés (GLM) afin de rendre compte des relations observées entre température et mortalité, rares sont les études qui comparent les apports de ces contributions en termes de pouvoir prédictif. Une comparaison synthétique a été proposée par Folkerts [24] sur des données des Pays-Bas, mais consistait surtout à comparer les interprétations des sorties obtenues. Quelques études biométriques et actuarielles se sont également attachées à étudier le rapport entre température et mortalité en France (Fall [23], Germain [33] et Pincemin [47]). Toutefois, celles-ci se concentrent sur une méthode unique et des jeux de données différents. Cela complique la comparaison et donc le choix d'une méthode adaptée pour la prédiction de la mortalité.

La troisième partie de ce mémoire a donc pour objectif de motiver le choix d'une méthode prédictive permettant de prédire raisonnablement la mortalité à horizon 2070 à partir de variables climatiques. Pour cela, une démarche progressive est proposée : après avoir estimé un GLM classique, les modèles les plus courants proposés dans la littérature (SEG, CSDL et DLNM) sont implémentés et comparés, sachant que chacun est caractérisé par un niveau de complexité supplémentaire. Une présentation synthétique est fournie dans le graphique 2.



GRAPHIQUE 2 – Extensions des modèles linéaires généralisés utilisées pour représenter le lien température-mortalité

L'analyse de leurs sorties est étudiée en détails, afin de comprendre au mieux les avantages et inconvénients de chacun. Un compromis est ensuite effectué entre plusieurs critères, permettant de connaître sa capacité à fournir des interprétations biométriques cohérentes, dans un délai raisonnable, et à capter au mieux les conséquences des canicules. Une segmentation, par tranche d'âge, sexe et zone climatique est ensuite menée afin de capter au mieux les relations différenciées entre la température et la mortalité. Enfin, une dernière partie permet d'établir les projections des taux de mortalité à horizon 2070, selon les scénarios climatiques et à partir des projections de population de l'INSEE.

1

Mise en contexte : santé, assurance et changement climatique

Ce chapitre vise à introduire les grands enjeux actuariels liés au sujet étudié. Dans un premier temps, le changement climatique ainsi que leurs conséquences potentielles pour les assureurs sont présentées dans la section 1.1. Ensuite, un rappel est effectué sur les risques liés à la mortalité et la santé pour les assureurs dans la section 1.2. Enfin, les conséquences spécifiques du dérèglement climatique sur la mortalité seront étudiées dans la section 1.3.

1.1 Le changement climatique, un nouveau défi pour les assureurs

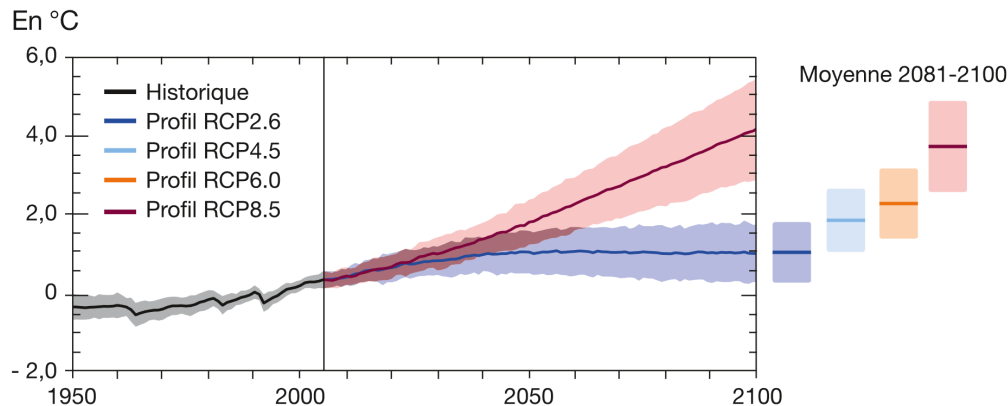
1.1.1 Présentation du changement climatique et des scénarios du GIEC

Les Nations Unies définissent le changement climatique comme les variations des températures et des conditions météorologiques sur le long terme. D'une part, ce bouleversement se traduit concrètement par des températures à la surface du globe supérieures à 1°C par rapport à l'ère pré-industrielle à l'échelle mondiale. D'autre part, il entraîne la recrudescence de phénomènes climatiques extrêmes tels que des sécheresses, des inondations, des feux de forêt et des tempêtes. Créé en 1988, le GIEC (Groupe d'experts Intergouvernemental sur l'Évolution du Climat) est chargé d'évaluer l'ampleur, les causes et les conséquences de ce dérèglement climatique. Il a également pour objectif de procéder, à intervalles de cinq à six ans, à une évaluation de l'état des connaissances. Des scénarios climatiques sont ainsi régulièrement actualisés afin de proposer une quantification de l'impact des activités humaines sur le climat, à destination des décideurs. Le cinquième rapport, publié en 2014, constitue

un tournant dans la définition des scénarios climatiques. La manière dont sont construits les scénarios RCP est décrite en annexe A, et ouvre la voie à une représentation plausible de la manière dont les variables climatiques d'intérêt pourraient évoluer dans les prochaines décennies. Quatre scénarios RCP ont ainsi été construits :

- **RCP 2.6** : scénario le plus optimiste, aligné avec les accords de Paris sur le climat avec une décroissance immédiate des émissions de gaz à effet de serre (GES) ;
- **RCP 4.5 et RCP 6.0** : scénarios intermédiaires, avec une stabilisation puis une décroissance des émissions de GES, respectivement à horizons 2040 et 2060 ;
- **RCP 8.5** : scénario le plus extrême, dit *business-as-usual* en l'absence de toute politique d'atténuation des conséquences du changement climatique. Les émissions de GES augmentent linéairement.

Le graphique 1.1 issu du premier groupe de travail du GIEC propose une représentation des évolutions des températures mondiales et de l'incertitude pour chaque scénario RCP.



GRAPHIQUE 1.1 – Projection de la variation de température moyenne mondiale suivant les scénarios du GIEC (source : GIEC, 2013)

Par construction, les projections de températures sont initialement proches des niveaux historiques, puis commenceraient à se différencier à partir des années 2020. Par rapport à la période 1986-2005, une augmentation de 4°C de la température à la surface du globe serait envisageable d'ici la fin du siècle en l'absence d'adaptation. Toutefois, la représentation met en avant une première source d'incertitude : plus l'horizon est lointain, et plus l'étendue des projections est importante. Les scénarios climatiques pessimistes fournissent également des projections plus incertaines, à horizon fixé. Ces limites doivent être prises en considération avant toute étude d'impact.

Le dernier rapport du GIEC [37], publié en 2021, introduit de nouveaux scénarios climatiques dits SSP (*Shared Socioeconomic Pathways*). Toutefois, bien que de nombreuses bases

de données climatiques ou atmosphériques s'appuyant sur les scénarios RCP aient déjà été publiées, rares sont les sources utilisant les scénarios SSP à l'heure actuelle. Leur maille géographique est, de plus, généralement moins fine. Pour cette raison, il a été décidé de se baser sur les scénarios RCP dans ce mémoire. La description des scénarios SSP est fournie en annexe B.

1.1.2 Classification du risque climatique

Bien que les risques physiques catastrophiques soient les plus couramment cités, les conséquences du dérèglement climatique sont dans les faits multiples, et se divisent en plusieurs catégories. Celles-ci ont été énoncées en 2015 par Mark Carney, gouverneur de la Banque d'Angleterre, dans son discours sur la tragédie des horizons :

- Le **risque physique**, à savoir l'ensemble des coûts associés aux conséquences physiques du changement climatique (sécheresse, inondations, submersions marines...) sur les personnes et les biens. On distingue alors les risques aigus, relatifs à l'augmentation de la fréquence et de la sévérité des catastrophes (telles que les inondations - ce point sera développé par la suite), et les risques chroniques qui sont liés à des conséquences à plus long terme (augmentation du niveau de la mer, disparition de certaines ressources).
- Le **risque de transition**, qui recense l'ensemble des coûts associés à la transition vers un modèle économique à faibles émissions de carbone. Ce risque est divisé en plusieurs sous-catégories parmi lesquelles figurent notamment les risques réglementaires, et juridiques (mise en place de politiques incitatives ou coercitives favorisant la transition), ainsi que les risques de marché.
- Le **risque de responsabilité**, défini comme les dommages et intérêts qu'une personne morale devrait payer si elle était jugée responsable du réchauffement climatique. Concrètement, des poursuites judiciaires pourraient être engagées à l'encontre de certains acteurs économiques, afin de montrer leur responsabilité via le manque de mesures de transition face au changement climatique. En 2021, dans l'"affaire du siècle", l'État Français a ainsi été condamné par le tribunal administratif de Paris à réparer les dommages causés par son inaction climatique avant le 21 décembre 2022.

Dans le cadre de ce mémoire, seuls les risques physiques chroniques seront évoqués. Il convient toutefois de noter que les trois catégories de risques mentionnées ci-dessus sont interdépendantes. En effet, la multiplication des événements physiques majeurs pourrait inciter les pouvoirs publics à intensifier leurs mesures réglementaires envers les acteurs économiques. Les conséquences financières du risque de responsabilité pourraient également être amenées

à s'aggraver du fait de la hausse de la fréquence et de la sévérité des catastrophes naturelles, et donc des conséquences accrues sur les biens et les personnes.

1.1.3 Intégration du changement climatique par les assureurs

La prise en compte du changement climatique par les assureurs est de plus en plus encouragée par les superviseurs français et européens, représentés respectivement par l'ACPR (Autorité de Contrôle Prudentiel et de Résolution) et l'EIOPA (*European Insurance and Occupational Pensions Authority*). Un des premiers documents publiés par l'ACPR sur le sujet s'est intéressé au rôle pris par les assureurs français dans le financement de la transition énergétique, dans un contexte législatif relativement récent avec la loi LTE¹ de 2015. En effet, l'article 173 impose aux assureurs des obligations de transparence sur leur politique d'investissement et leur gestion des risques liés au changement climatique. Au-delà de la gestion des actifs, d'autres publications se sont focalisées sur les risques au passif des assureurs.

C'est notamment le cas du premier exercice pilote climatique organisé par l'ACPR en 2020². Celui-ci avait pour objectif de sensibiliser les assureurs et les banques aux risques physiques et de transition de long-terme, tant pour les activités vie que non-vie. Notons que le risque de responsabilité n'est pas inclus dans l'étude. Cet exercice devait également permettre aux assureurs de réfléchir sur la modélisation prospective de ces différents risques. Sur la base du volontariat, les assureurs pouvaient projeter les principaux éléments de leur bilan et de leur compte de résultat par mailles de 5 ans, en tenant compte de risques et de projections spécifiques. À cette fin, le scénario retenu est le plus pessimiste, à savoir le RCP 8.5.

En 2023, un deuxième exercice³ à destination des banques et des assureurs a été lancé. Deux de ses objectifs sont similaires à ceux de l'exercice de 2020 : l'anticipation des risques et la méthodologie de leur quantification. La principale nouveauté de cet exercice est d'inclure des risques de court-terme. Alors que les risques de long-terme sont proposés par l'ACPR afin d'évaluer leurs impacts sur les projections du bilan et du compte de résultat, les risques de court-terme ont pour objectif d'évaluer des conséquences immédiates sur la solvabilité des assureurs. Les risques de court-terme se caractériseraient dans un premier temps par une succession de réalisations de risques physiques naturels étalés sur deux ans. Parmi ceux-ci figurent des sécheresses, des vagues de chaleur et des ruptures de digue. Celles-ci peuvent avoir

1. Loi sur la transition énergétique pour la croissance verte

2. Principales hypothèses de l'exercice 2020 : ☞

3. Principales hypothèses de l'exercice 2023 : ☞

un impact sur l'activité non-vie, à cause des dommages aux biens ou sur l'activité (notamment agricole). Elles peuvent également impacter les activités vie, via des conséquences sur les frais de soins, les arrêts de travail ainsi que la mortalité. Les hypothèses retenues et les différences spécifiques aux risques santé et vie seront précisées dans la section 1.3. Dans le scénario de court-terme, ces catastrophes naturelles seraient suivies par une période de récession de deux ans, ayant un impact sur le risque de transition. Un schéma issu de la publication de Deloitte⁴ permet de comparer les deux exercices pour chacune des grandes catégories de risques couverts par les assureurs.

Dans le cadre de la directive Solvabilité II, les assureurs doivent également intégrer les conséquences du changement climatique dans leur démarche de gestion des risques. Ce rôle a été confirmé par l'EIOPA dans une publication⁵ de février 2022. D'après celle-ci, en tant que cadre d'évaluation des risques futurs, Solvabilité II peut intégrer les risques climatiques parmi les autres risques existants. Les outils SII permettent alors de mesurer et de diversifier les risques climatiques, liés à la souscription et aux investissements. Cette gestion des risques est désormais étendue à la fonction actuarielle.

Par ailleurs, le deuxième pilier de cette directive, relatif à la gestion des risques et à la gouvernance, impose une évaluation interne du risque et de la solvabilité (*Own Risk and Solvency Assessment* ou ORSA). Ce processus doit illustrer la capacité de l'assureur à identifier, mesurer et gérer l'ensemble des éléments à même de modifier sa solvabilité ou sa situation financière. Le superviseur européen, l'EIOPA, a ainsi transmis à l'ensemble des superviseurs nationaux des conseils de publications, destinés aux assureurs, par rapport à la prise en compte du risque climatique dans l'ORSA [21]. Ces conseils sont publiés sous la forme d'exemples concrets pour des entreprises fictives d'assurance non-vie et vie respectivement. Dans un premier temps, le rapport préconise aux compagnies d'évaluer la matérialité du changement climatique. Il est proposé d'effectuer cette démarche en trois étapes, à la fois pour l'actif et le passif :

- Définir le contexte d'activité
- Rechercher les impacts du changement climatique sur l'activité
- Évaluer la matérialité du risque sur l'activité (risque matériel ou non-matériel)

Cette analyse de matérialité repose principalement sur une analyse qualitative globale des risques ainsi que des conjectures. Dans la section 1.3, un exemple concret de ces étapes sera donné pour le cas d'une compagnie fictive d'assurance-vie. En particulier, il sera discuté la

4. Lien vers la publication de Deloitte sur l'exercice climatique 2023 : [↗](#)

5. Lien vers la publication "*Solvency II : Leading the way in managing sustainability risk*" : [↗](#)

manière dont le risque d'évolution de la mortalité dans un contexte de changement climatique est abordé par l'EIOPA. Si la matérialité est avérée à la suite de cette démarche, alors une analyse de scénarios est suggérée. Il s'agit cette fois d'une analyse quantitative, nécessitant une modélisation précise de l'impact du changement climatique sur l'activité. En ce sens, les recommandations de l'EIOPA portent sur :

- la **granularité** de la modélisation : une granularité locale est à privilégier du fait des disparités géographiques du changement climatique ;
- les **outils**, qui doivent permettre de répondre au mieux à la problématique de risque, malgré les difficultés d'accès aux données ;
- les **scénarios**, qui doivent également être les plus adaptés au portefeuille considéré ; il n'est pas toujours conseillé de se contenter des scénarios et paramètres prédéfinis.

Ces éléments permettent de souligner une différence significative entre les exercices climatiques de l'ACPR et l'ORSA. Alors que les exercices climatiques de l'ACPR adoptent une vision globale des risques, l'ORSA est un processus propre à chacun des assureurs. Pour ce dernier, il est ainsi préconisé d'utiliser des scénarios, données et méthodologies spécifiques, tandis que l'approche retenue par l'ACPR est harmonisée. Ce mémoire s'inscrit dans une telle analyse de scénarios. Il proposera une méthodologie de projection des taux de mortalité, ainsi qu'une démarche d'intégration permettant d'évaluer la sensibilité du portefeuille à ce risque spécifique. Après avoir décrit les enjeux du changement climatique pour les assureurs, le contexte actuariel de l'assurance portant sur la vie et la santé est rappelé.

1.2 Les risques liés à la mortalité et à la santé pour les assureurs

1.2.1 Présentation du cadre réglementaire

Les risques touchant la santé et la mortalité des individus concernent principalement les assureurs vie et complémentaires santé. Pour rappel, dans le cadre de la directive Solvabilité II, la formule standard pour l'évaluation des capitaux réglementaires nécessite de tenir compte des risques de mortalité, de longévité et d'incapacité/invalidité à travers des variables choquées. Plus exactement, le capital de solvabilité requis (SCR) doit être déterminé en calculant le montant de fonds propres économiques (FP) dont l'assureur a besoin afin de limiter la probabilité de ruine économique à 0,5%, c'est-à-dire une occurrence tous les 200 ans. Dans

la formule standard, ce calcul est effectué pour chaque risque r individuellement, en calculant la différence des fonds propres nécessaires dans un scénario central et dans un scénario "choqué", c'est-à-dire après réalisation d'un risque donné survenant avec une probabilité égale à 0,5% :

$$\text{SCR}_r = \text{FP}_x^{\text{central}} - \text{FP}_r^{\text{choqué}}.$$

Or, le SCR apparaît au bilan économique comme un élément du passif. L'évolution imprévue de ces risques individuels peut donc se matérialiser par une évolution dans le passif, et de façon indirecte, dans la gestion actif-passif des assureurs.

Dans cette section, les cadres assurantiels des risques portant sur la mortalité et la santé seront d'abord présentés séparément. Ensuite, la manière dont ceux-ci sont appréhendés par les actuaires sera évoquée, ainsi que les perspectives d'études actuarielles compte tenu des données disponibles en accès libre.

1.2.2 Mortalité, longévité et contrats de rentes

Les assureurs-vie, à travers le versements de rentes, sont affectés par les risques liés à la santé et au décès de leurs assurés. Dans les contrats de rentes (ou garanties *en cas de vie*), l'assureur peut être amené à verser une somme donnée à un instant t (par exemple le 1^{er} janvier d'une année) à condition que l'assuré soit encore vivant à cet instant. Au contraire, d'autres contrats garantissent à l'assuré le versement d'un capital *en cas de décès*. Dans l'optique de déterminer le juste prix de tels contrats, c'est-à-dire leur *prime pure*, il revient à l'assureur de connaître la probabilité de décès de l'assuré.

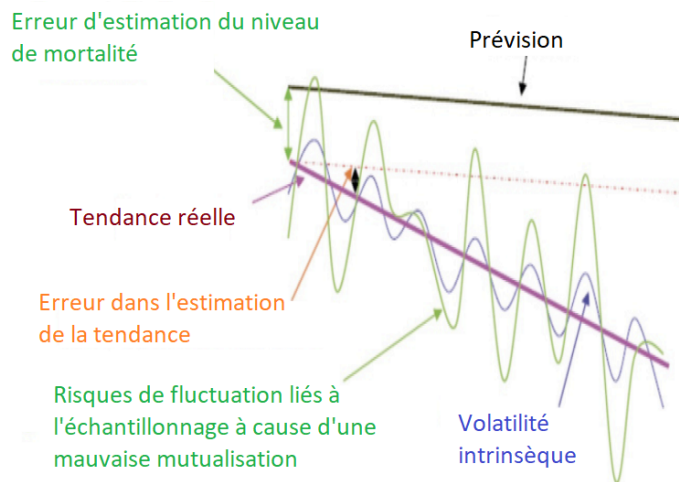
Les actuaires considèrent usuellement les tables de mortalité afin de quantifier, sur un portefeuille d'assurés (souvent au nombre de 10 000), la probabilité conditionnelle de décès entre l'âge x et $x + 1$. Selon Petauton [27], plusieurs catégories de facteurs peuvent influencer cette probabilité de décès :

- des facteurs généraux, dont l'influence s'exerce sur toute la population
- des facteurs particuliers, propres à l'assurance et à ses assurés

D'une part, les facteurs généraux incluent l'âge de l'assuré, son sexe, sa profession ainsi que son pays. L'âge de l'individu est le premier facteur d'importance, sachant qu'indépendamment des autres éléments listés, la probabilité conditionnelle de décès entre les âges x et $x+1$ est plus importante pour les petits (mortalité infantile) et grands âges. Le sexe de l'individu affecte également cette probabilité de décès, sachant qu'une femme a en moyenne une espérance de

vie supérieure à celle d'un homme. Par exemple, selon la Drees [16], en 2020, une femme vivait jusqu'à 85,1 ans soit 6 années de plus qu'un homme en moyenne. La profession de l'individu ainsi que son pays peuvent également influencer son espérance de vie à travers sa capacité d'accéder à certains soins, son bien-être, ainsi que son exposition à des maladies durant son activité professionnelle.

Ces facteurs visibles sur l'ensemble de la population sont amenés à évoluer dans le temps. Ainsi, une étude prospective de la mortalité devra correctement capter l'évolution structurelle de l'espérance de vie dans le temps, ce que l'on appelle *risque de longévité*. Le risque de longévité est généralement divisé en quatre composantes, représentées sur le graphique 1.2, issu de Loisel et al. [39] :



GRAPHIQUE 1.2 – Composantes du risque de longévité (source : Loisel et al. [39])

Les quatre composantes du risque de longévité sont :

- Le risque de **volatilité** autour de la tendance : il est dû à des phénomènes saisonniers d'épidémies ou météorologiques
- Le risque d'**échantillonnage** : il est dû à l'hétérogénéité du portefeuille d'assurés
- La **tendance** : elle affecte l'ensemble du portefeuille et est notamment due aux progrès dans les soins
- L'erreur d'**estimation** : elle est due à un faible nombre d'assurés dans le portefeuille, notamment lorsque l'étude concerne des âges trop avancés, une zone géographique ou bien des tranches d'âge trop fines

L'ensemble de ces éléments requiert une modélisation adaptée, notamment dans cette étude puisque la granularité géographique des décès est affinée. Le risque de longévité affecte autant les assureurs proposant des contrats *en cas de vie* qu'*en cas de décès*. Une augmenta-

tion de l'espérance de vie toutes choses égales par ailleurs conduira d'une part à davantage de versements de capitaux au titre des contrats *en cas de vie*, et d'autre part à moins de versements de capitaux au titre des contrats *en cas de décès*. Cela se traduira au passif par une variation des provisions techniques, positive ou négative selon le type de contrats proposés par l'assureur. La matérialité de ce risque ainsi que de ses composantes sera étudiée plus en détails dans le chapitre 2.

Au-delà des éléments explicatifs de la probabilité de décès touchant l'ensemble de la population, un assureur doit en pratique tenir compte d'autres facteurs directement imputables à son portefeuille d'assurés. Selon Petauton, les assurés étant rationnels, la théorie microéconomique de la maximisation de l'utilité les conduirait à choisir le contrat, et donc indirectement la compagnie d'assurance, qui serait le plus adapté à sa condition de santé. Ce phénomène, dit de *sélection adverse*, conduit par exemple à ce que des garanties *en cas de décès* seraient davantage souscrites par des assurés en mauvaise santé. La conséquence actuarielle pour l'assureur est la nécessité de s'appuyer sur une table d'expérience, construite à partir des données de son portefeuille et non de la population globale, pour modéliser le risque de mortalité. Cet aspect ne sera pas étudié dans ce mémoire, qui ne s'appuiera que sur des données accessibles librement.

Plusieurs textes fondamentaux encadrent la tarification des produits d'assurance-vie. Un rappel réglementaire est fourni en annexe C. Toute discrimination directe ou indirecte ayant pour effet des différences dans les primes et les prestations, fondée sur la prise en compte du sexe, est interdite. Depuis le 21 décembre 2012 et à la suite d'une décision de la Cour de Justice de l'Union Européenne, les possibilités de dérogation initialement prévues ont été supprimées lorsque le sexe est un facteur déterminant dans l'évaluation du risque. Si une table unique est établie pour tous les assurés, cette table doit correspondre au tarif le plus prudent. Ainsi, pour des engagements versés en cas de vie, la table féminine devra être retenue, car la probabilité de décès des femmes est plus faible à âge fixé. Un exemple pratique simple d'application de l'étude pour la tarification d'un contrat de rente viagère sera proposé dans ce mémoire.

Les institutions de prévoyance et les mutuelles offrant des contrats de complémentaires santé sont également intéressés par l'évaluation des risques affectant la santé des individus. Cet aspect n'entre pas directement dans le cadre de ce mémoire mais pourrait faire l'objet d'approfondissements. Une présentation des risques liés à l'incapacité et à l'invalidité est disponible dans l'annexe D.

1.3 Les conséquences du changement climatique sur la mortalité

1.3.1 Présentation des conséquences du changement climatique sur la mortalité et santé

Si les conséquences les plus médiatisées du dérèglement climatique concernent actuellement les dommages aux biens (DAB), les conséquences indirectes de ces événements sur la santé des personnes sont également à considérer et seraient amenées à se multiplier dans les prochaines années. Parmi ces dernières, figurent notamment les risques sanitaires. Ainsi, dans son sixième rapport publié en 2023, le GIEC a identifié 8 principaux risques sanitaires que le changement climatique est susceptible de renforcer. Ceux-ci, ayant une traduction concrète sur la mortalité et la morbidité, sont représentés dans le schéma 1.3. Ils concernent :

- la dénutrition,
- les vagues de chaleur,
- les infections transmises par l'alimentation et l'eau,
- les événements climatiques extrêmes,
- la santé professionnelle,
- les maladies infectieuses,
- la qualité de l'air,
- la santé mentale.

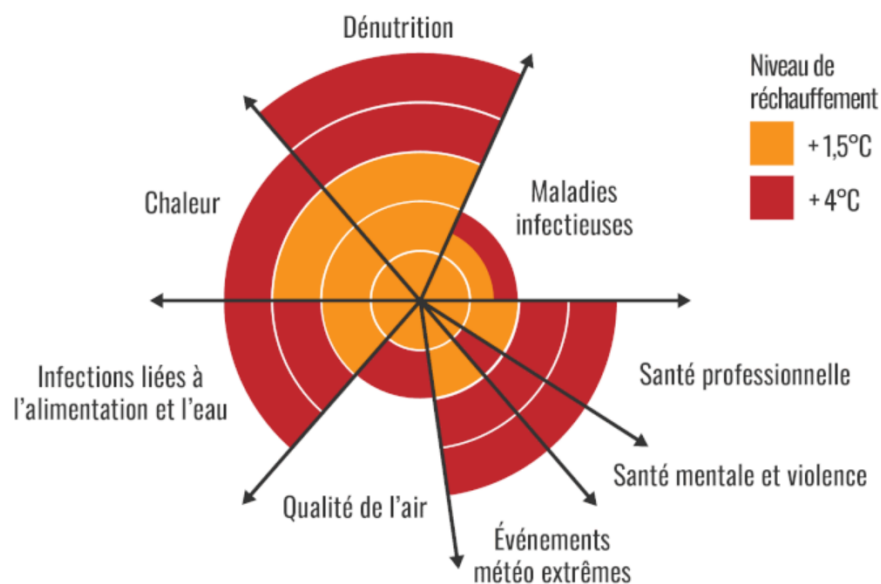


SCHÉMA 1.3 – Conséquences sur la santé de l'augmentation des températures (source : Réseau Action Climat)

On y constate notamment que, dès une augmentation de 1,5°C par rapport à l'ère industrielle, la chaleur aura des conséquences importantes sur la santé des individus. Cependant, les conséquences des événements météorologiques extrêmes (telles que les inondations) seront davantage visibles dans un scénario de réchauffement grave. De façon générale, l'ampleur des conséquences de ces risques serait amenée à varier selon le scénario de réchauffement climatique. De ce fait, chaque pays sera différemment touché selon le sixième rapport du GIEC. Par exemple, les pays du Sud seraient davantage touchés bien qu'ils émettent moins de gaz à effet de serre.

Dans le cadre d'une analyse prospective de la mortalité, il convient de s'intéresser à chacun des risques et à établir un "état de l'art" des études sur le sujet. Cela permettra de déterminer la faisabilité de l'intégration de chaque risque dans cette analyse. Les études sur ces risques et leur relation aux phénomènes de santé se sont multipliées ces dernières années, principalement du fait de l'accessibilité croissante des données et des enjeux de santé publiques associés à cette question. Dans cette partie sera présentée la manière dont les éléments ci-dessus peuvent se traduire par un impact concret sur la mortalité. Leur matérialité sera davantage quantifiée dans le chapitre 2.

Les vagues de chaleur

Selon Luber et al. [41], les maladies liées aux phénomènes de chaleur seraient des crampes de chaleur, des syncopes de chaleur, des épuisements par la chaleur et des coups de chaleur. Les premières, moins graves, peuvent évoluer vers les coups de chaleur si elles ne sont pas reconnues et traitées à temps. Les coups de chaleur, quant à eux, peuvent se matérialiser à court terme par un passage en état comateux voire le décès. Entre autres, les symptômes d'un coup de chaleur sont une température corporelle excédant 40,6°C, des convulsions ainsi que des délires. Les personnes les plus exposées sont les enfants de moins de 5 ans, les personnes âgées ainsi que les individus sans accès à des endroits frais ou climatisés. Selon la biostatisticienne Deguen⁶, les enfants défavorisés feraient face à "une accumulation de facteurs de vulnérabilité", étant généralement scolarisés dans des écoles "localisées dans des zones où l'exposition à la pollution est plus importante, tout comme leurs logements qui sont par ailleurs moins isolés et parfois surpeuplés". L'ensemble de ces éléments pourraient laisser présager d'une évolution des trajectoires de mortalité à la fois entre plusieurs territoires et plusieurs classes d'âge.

Les villes seraient les plus touchées par les conséquences des vagues de chaleur, au point

6. Reprise par cet article de Réseau Action Climat : [↗](#)

que celles-ci constitueraient la cause de décès la plus fréquente parmi l'ensemble de celles reliées aux conditions climatiques. En effet, dans les territoires fortement urbanisés, les températures maximales seraient plus élevées qu'ailleurs : c'est le phénomène d'*îlots de chaleur urbains* - caractérisés par le géographe-climatologue Cantat [11] comme une "sorte de dôme d'air plus chaud couvrant la ville". L'apparition de ces îlots serait davantage liée à la forte concentration de bâtiments dans une zone donnée qu'à la pollution : les matériaux présents tels que la pierre, la brique ou le béton captent la chaleur et la libèrent dans l'atmosphère la nuit, occasionnant des écarts de température potentiellement significatifs par rapport à des territoires moins urbains.

Différentes mesures d'adaptation sont possibles pour limiter les conséquences sur la vie des vagues de chaleur. Dans un contexte d'augmentation de la fréquence et de l'intensité de ces phénomènes, leur mise en place est d'une importance accrue. Une première mesure concerne la prévention et l'éducation des populations face aux canicules et aux risques qu'elles présentent. Lors des épisodes de chaleur intense, des messages préventifs de Santé Publique France sont fréquemment diffusés à la télévision afin de rappeler les comportements à adopter pour se protéger soi-même ainsi que les autres : rester au frais et boire de l'eau. Les populations les plus à risque peuvent également faire l'objet d'une surveillance et d'une protection renforcées. D'autres politiques d'adaptation matérielles peuvent être proposées. À l'échelle de l'habitation individuelle, l'accès à un air conditionné possède un effet compensateur significatif sur les conséquences des vagues de chaleur, bien qu'il conduise à son tour à une consommation d'énergie supplémentaire, ce qui est néfaste du point de vue du changement climatique. L'usage de ventilateurs, moins coûteux, est quant à lui déconseillé car ces derniers feraient propager de l'air (très) chaud à la surface de la peau. Plusieurs études ont quantifié l'avantage de la climatisation aux États-Unis, ou dans des grandes agglomérations mondiales. Par exemple, d'après Barreca et al. [6], la diffusion de l'air conditionné explique à elle seule près de 90% du déclin dans la mortalité liée aux vagues de chaleur entre 1960 et 2004 aux États-Unis. À l'échelle urbaine cette fois, l'aménagement de l'espace, par exemple avec l'augmentation de la végétation ou de matériaux réfléchissant la lumière du soleil, pourrait également atténuer le phénomène des îlots de chaleurs.

L'importance de l'humidité

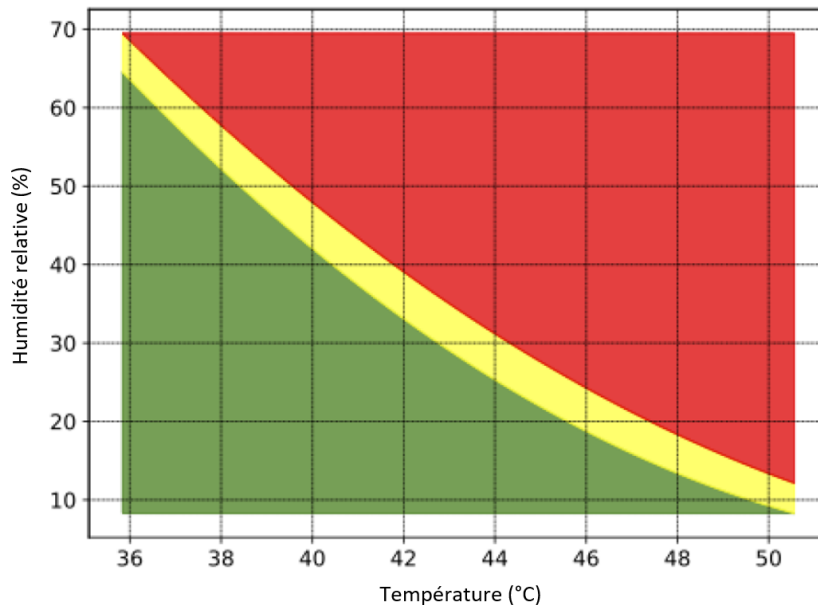
Si la qualité de l'air réfère le plus souvent à la pollution, l'humidité est un facteur majeur concernant la supportabilité de l'air pour le corps humain. En effet, celui-ci cherche en permanence à maintenir sa température à 37°C. Lorsqu'il fait chaud, la transpiration permet à la

chaleur d'être évacuée par évaporation, ce qui s'avère plus difficile lorsque le taux d'humidité dans l'air est important. En effet, quand l'humidité s'approche des 100%, il ne peut plus y avoir de vapeur d'eau supplémentaire dans l'air, donc la sueur ne s'évapore plus. C'est pour cette raison qu'une même température peut être ressentie différemment dans deux endroits différents, à cause du taux d'humidité dans l'air. Lorsque l'air est complètement saturé en humidité, le corps humain ne peut survivre à une exposition prolongée à une température supérieure ou égale à 35°C. Cette valeur a récemment été remise en question (elle a été ré-estimée à 31°C par Vecellio et al. [58]). C'est pourquoi un indicateur a été créé, permettant de tenir compte à la fois de la température (notée T) et du taux d'humidité dans l'air (noté h) : la température du bulbe mouillé (*wet-bulb temperature*, notée TH). Elle est définie par la formule suivante :

$$TH = T \times \arctan[0.152(h + 8.3136)^{1/2}] + \arctan(T + h) - \arctan(h - 1.6763) \quad (\text{th})$$

$$+ 0.004h^{3/2} \arctan(0.0231h) - 4.686$$

Le graphique 1.4 représente la correspondance entre la température, le taux d'humidité dans l'air, et la température du bulbe mouillé. On y constate notamment qu'une température de 50°C peut être supportée par le corps humain, mais cela nécessite un taux d'humidité dans l'air suffisamment faible.



GRAPHIQUE 1.4 – Correspondance entre la température, le taux d'humidité dans l'air et la température du bulbe mouillé (source : W. Larry Kenney)

L'apport de l'humidité a déjà été mentionné dans une note de Galea [28] relative à une analyse descriptive de la mortalité lors des pics de chaleur. Cette étude s'attachera donc à évaluer les disparités de température et d'humidité en France, afin de conjecturer l'utilité ou non de l'indicateur de "température humide" pour une modélisation de la mortalité.

Les disparités géographiques de l'ensemble des conséquences du changement climatique évoquées plus haut incitent à tenir compte de cette dimension spatiale dans la présente analyse. Par manque de données disponibles, les risques associés à la qualité de l'air, aux maladies infectieuses ainsi qu'aux catastrophes naturelles n'ont pas été exploités dans cette étude. Une présentation détaillée de ces risques est abordée en annexe E.

1.3.2 Les recommandations des superviseurs pour l'intégration du changement climatique dans les risques vie

Afin d'établir une modélisation actuarielle adaptée aux attentes des assureurs, il est important de s'interroger sur les recommandations des superviseurs relatives à l'intégration du changement climatique.

Exemples d'intégration du risque vie dans les exercices pilote climatique

Dans son exercice pilote climatique de 2020, l'ACPR a préconisé aux assureurs-vie la prise en compte de la surmortalité liée à des facteurs de long-terme. Les risques retenus sont liés aux pics de pollution et à la transmission de maladies vectorielles qui devraient évoluer d'ici 2050. Des facteurs additifs et multiplicatifs ont été fournis, notamment avec l'aide du courtier AON sur la base de l'article de Drif et al. [18] mentionné plus haut. Ces éléments permettent ainsi aux assureurs d'intégrer des données de mortalité choquées dans leurs projections des principaux postes de leur bilan et de leur compte de résultat, qu'ils ont dû établir entre 2025 et 2050 par tranches de 5 ans.

Dans son exercice de 2023, l'ACPR inclut également l'impact d'évènements climatiques à court-terme sur la mortalité. Les principaux postes des éléments financiers devront donc être projetés à la maille annuelle, entre 2023 et 2027 (en plus de la projection à long-terme). Deux scénarios sont retenus. Le premier est imputable aux vagues de chaleur, en termes de dérive de mortalité par rapport à l'année 2019. Le second est lié au risque d'inondation, via la rupture de barrage à Serre-Ponçon. Cette dernière pourrait se traduire par une montée soudaine des eaux à proximité du barrage, entraînant un risque de mortalité supplémentaire.

Cela est visible sur le graphique 1.5, dans lequel le taux de mortalité a été calculé par zone IRIS (Îlots Regroupés pour l'Information Statistique, c'est-à-dire à l'échelle infra-communale) à Lyon, en prenant en considération la hauteur et de la vitesse de l'onde de submersion, ainsi que du type de bâtiment à proximité. En pratique, une telle modélisation requiert l'utilisation de données à maille très fine relatives au bâti, aux submersions et à l'impact des digues. C'est pourquoi elle est exclue de cette étude.



GRAPHIQUE 1.5 – Taux de mortalité par zone IRIS à Lyon suite à la rupture du barrage de Vouglans (source : AON France)

De l'exercice 2023, il ressort de façon générale une volonté du superviseur d'affiner la modélisation des risques, qu'ils soient physiques ou de transition. Les projections des données de mortalité sont ainsi disponibles par tranches d'âge (moins de 15 ans, entre 15 et 64 ans, plus de 64 ans) et zone géographique (agglomérations pour le risque lié à la pollution, et départements sinon). Ce mémoire s'inscrit dans ce contexte et proposera donc une analyse à maille géographique fine du risque étudié.

Exemple d'intégration dans l'ORSA du risque de mortalité

L'EIOPA a publié en 2022 des recommandations d'intégration du changement climatique dans l'évaluation des risques et de la solvabilité effectuée par les assureurs. Les risques portant sur la mortalité peuvent intervenir dans cette analyse de matérialité, comme le montre l'exemple ci-dessous directement issu du document *Application guidance on running climate change materiality assessment and using climate change scenarios in the ORSA* de l'EIOPA. D'un point de vue qualitatif et concernant le passif d'une compagnie fictive d'assurance-vie,

l'analyse se fait en trois temps :

- **Définir le contexte d'activité** : le changement climatique sera à l'origine d'une fraction de décès supplémentaire à l'avenir, via des impacts directs (catastrophes naturelles) et indirects (transmission de maladies, dégradation de la qualité de l'air), plus ou moins importants selon la zone géographique considérée. L'Afrique et l'Asie pourraient être les continents les plus touchés.
- **Rechercher les impacts du changement climatique sur l'activité** : la question est de savoir si le risque de mortalité est amené à évoluer de façon significative. Le rapport propose de corriger les taux de mortalité actuels par un facteur d'ajustement lié au portefeuille et à l'impact du risque physique sur la mortalité. En cela, la proposition est similaire à celle issue de l'exercice pilote climatique. En effet, les taux de mortalité choqués pour le risque de pollution étaient obtenus à partir d'un facteur additif et d'un facteur multiplicatif, avec des données disponibles par zone géographique et (pour l'exercice 2023) par tranche d'âge. L'impact sur le passif de l'assureur tient en deux évolutions qui se compensent. De manière analogue à l'exemple du risque de longévité évoqué plus haut, la hausse inattendue des décès se traduit, toutes choses égales par ailleurs, par une augmentation du versement des capitaux au titre des contrats *en cas de décès* et une baisse des versements pour les contrats *en cas de vie*.
- **Évaluer la matérialité du risque sur l'activité** : la matérialité du risque physique du côté du passif de l'assureur-vie se constaterait seulement à long-terme et à la suite de phénomènes extrêmes. D'une part, les conséquences de l'évolution chronique des températures sur la mortalité sont omises de l'étude. D'autre part, cette affirmation vient en contradiction avec les instructions de l'exercice pilote climatique de 2023, qui inclut la prise en compte d'impacts de court-terme du risque physique sur la mortalité (vagues de chaleur et conséquences des ruptures de digue).

Les conclusions de cette analyse de matérialité dépendront directement du contexte ainsi que des expositions de l'entreprise d'assurance. Si le portefeuille est fortement exposé au risque de décès (via des contrats de rentes notamment), alors la matérialité sera plus importante. Si l'analyse qualitative conduit à une matérialité avérée, alors une étude de scénarios pourra être menée. Les chapitres suivants viseront à proposer une méthodologie de prédiction des taux de mortalité futurs, puis à expliquer la manière dont ces derniers pourront être intégrées dans une telle analyse de scénarios.

Conclusion du chapitre

L'analyse essentiellement bibliographique effectuée dans ce chapitre a permis d'établir des causes de décès risquant d'évoluer à cause du dérèglement climatique et donc à intégrer dans une étude prospective de la mortalité. Toutefois, elle a mis en avant certaines contraintes de données au vu des éléments qu'il faudrait intégrer pour une analyse exhaustive. Les sections précédentes ont également permis de se rendre compte des attentes des superviseurs, qui permettront d'orienter cette étude afin qu'elle ait une utilité concrète pour les assureurs français.

2

Analyse globale et nationale des facteurs de mortalité historiques

Dans ce chapitre, les bases de données utilisées dans cette l'étude sont tout d'abord présentées dans la section 2.1. Ensuite, des analyses descriptives de saisonnalité et de tendance sont menées respectivement dans les sections 2.2 et 2.3. Leur objectif est de motiver le choix des variables explicatives finalement retenues, et de présenter leur rôle dans les modèles usuels.

2.1 Présentation des bases de données retenues

2.1.1 Les données de mortalité

Afin de comprendre au mieux les relations existantes entre la mortalité et les causes imputables au changement climatique, il aurait fallu disposer de données de mortalité complètes, renseignant à la fois la cause de décès et le risque sous-jacent. À l'échelle internationale, de telles données sont disponibles dans la base *Global Burden of Disease 2019* (GBD 2019¹) et permettent de connaître les facteurs explicatifs de la mortalité ainsi que leur évolution dans le temps. Pour chaque pays, elles renseignent une estimation ainsi qu'un intervalle de confiance à 95% des risques sous-jacents de décès qui peuvent être utilisés pour établir des projections de la mortalité. Cette base a notamment été utilisée par Germain [33], qui a construit des taux de mortalité choqués à partir d'un modèle de Lee-Carter modifié isolant la cause sous-jacente de décès liée au changement climatique. Cependant, les données de la GBD 2019 ont fait l'objet de certaines critiques, par exemple exprimées par Gordon-Dseagu

1. Lien vers les différentes bases de données : ↗

et al. [34]. Celles-ci sont notamment relatives au manque de transparence concernant les hypothèses formulées et associées aux estimations, et à des résultats peu plausibles. L'étude concernait spécifiquement les risques sur la santé associés à la consommation de viande rouge non-transformée. Les estimations de la GBD 2019 venaient en contradiction des résultats du Fonds mondial de recherche contre le cancer, selon lequel la consommation de viande rouge serait un contributeur causal au développement du cancer colorectal. Les données de la GBD 2019 restent toutefois exploitables à des fins comparatives à l'échelle internationale, c'est pourquoi elles seront utilisées pour la comparaison quantitative des risques climatiques dans la section suivante.

À l'échelle française, l'INSERM (Institut National de la Santé et de la Recherche Médicale) propose également une catégorisation simplifiée des causes de mortalité par département en accès libre. Toutefois, cette base de données ne permet pas d'obtenir les causes sous-jacentes de la mortalité. Elle est donc difficilement exploitable dans une étude prospective. La mortalité totale, toutes causes confondues, sera donc l'indicateur utilisé. L'avantage de ce choix est qu'il permettra de capter l'ensemble des causes de mortalité liées directement ou indirectement par le changement climatique. En particulier, cela évite de faire une sélection probablement incomplète des causes de décès liées au changement climatique. En effet, des décès dus à la chaleur pourraient être recensés comme liés, en fait, à des causes respiratoires ou cardiovasculaires.

L'étude du risque de mortalité en France de façon fine est rendue possible par l'accessibilité des données de mortalité recensées par l'INSEE depuis 1970. Celles-ci permettent de disposer d'informations utiles relatives au profil de décès, telles que l'âge, le sexe, le lieu et la date de décès des individus. Notons qu'il s'agit d'une particularité française : à notre connaissance, aucune base européenne ou américaine en accès libre et gratuit ne propose un historique des décès recensés disposant d'informations aussi précises sur le lieu et la date du décès. Les données étant publiées mensuellement, il est nécessaire de télécharger puis d'agréger manuellement les données de mortalité afin d'obtenir une base de données exploitable. Une autre solution, retenue ici, est d'utiliser une base déjà agrégée disponible via le portail Koumoul². Les variables utiles pour l'étude sont décrites dans le tableau 2.1.

Des informations personnelles (nom, prénom et code postal) sont disponibles mais ignorées. L'information sur le code postal du décès n'est également pas utilisée, afin de conserver des portefeuilles d'assurés de taille suffisante pour l'étude. Il convient de nettoyer la base obtenue, celle-ci étant constituée d'une part négligeable (moins de 1%) de valeurs aberrantes ou inexploitables. Les décès recensés dont le lieu se situe à l'étranger sont notamment exclus. Bien

2. Lien vers le fichier agrégé des décès en France depuis 1970 : [↗](#)

<i>Variables</i>	<i>Description et traitement éventuel</i>
Sexe de l'individu décédé	Variable catégorielle (1 pour "Homme" et 2 pour "Femme")
Âge de décès	Arrondi à l'entier inférieur
Date de décès	Maille journalière (entre janvier 1970 et aujourd'hui)
Département de décès	Seuls les décès survenus en France métropolitaine sont étudiés

TABLEAU 2.1 – Variables utilisées dans le fichier des décès de l'INSEE depuis 1970

que le niveau de richesse de l'individu, d'éducation ainsi que sa catégorie socio-professionnelle ne soient pas renseignés dans cette base de données, l'utilisation d'une maille géographique départementale permettra d'exploiter de telles données externes agrégées.

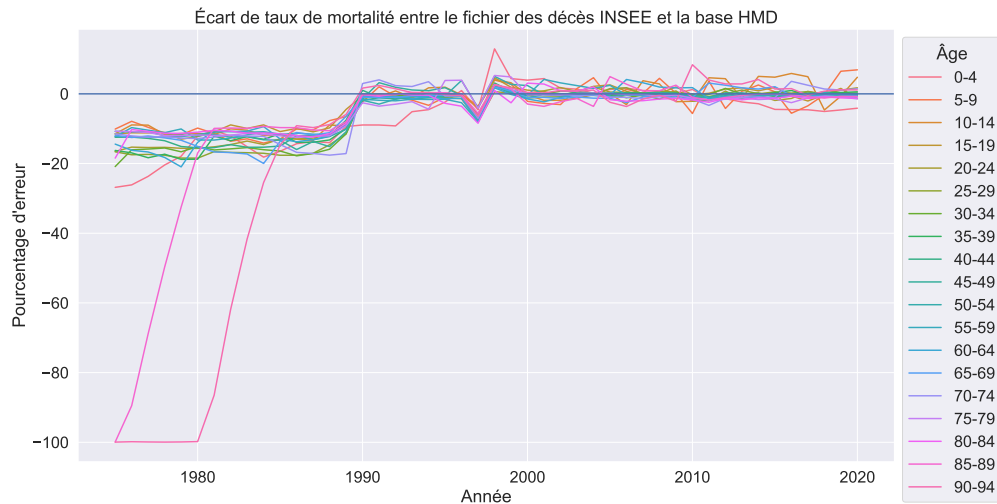
À partir de cette base, l'objectif de l'étude menée ici sera d'étudier la mortalité de manière comparative, entre différentes zones géographiques. Il pourra être intéressant d'analyser les disparités géographiques historiques et futures de la mortalité, à partir de variables bien choisies. Le passage à une segmentation géographique fine introduit une problématique concernant la qualité de l'estimation. En effet, s'appuyer sur une maille départementale revient à considérer un portefeuille (fictif) d'assurés restreint. Il s'agira alors de discuter et de comparer des méthodes permettant de gérer correctement l'estimation à partir d'un faible nombre de données, avec une variance minimale.

2.1.2 La période d'étude retenue

La variable d'intérêt, à savoir les nombres de décès, sont disponibles dans le fichier des décès à partir de 1970. Cependant, dans le forum de discussions associé à cette base de données ³, il est fait état de certains décès qui n'ont pas été répertoriés. En effet, les données disponibles sont issues de la partie administrative des fichiers de décès transmis informatiquement par les mairies à l'INSEE. Selon le personnel de cet institut, l'informatisation progressive de l'enregistrement des décès a conduit à des données manquantes, particulièrement entre 1970 et 1975. Des internautes ont également relevé certains renseignements aberrants, contenant des décès de personnes fictives. Ces erreurs et manques peuvent poser problème et amener à un biais, résultant dans une sous-estimation ou une sur-estimation des décès. Nous avons donc comparé les taux de mortalité obtenus à partir du fichier des décès de l'INSEE avec une base de données de référence : la HMD (*Human Mortality Database*), et entre 1970 et 2020. Cette dernière renseigne des tables de mortalité pour un grand nombre de pays développés et avec un historique très large, pouvant aller jusqu'au début du 19^{ème} siècle.

3. Lien vers le forum de discussions associé au fichier des personnes décédées de l'INSEE : [↗](#)

Dans un premier temps, afin de simplifier l'analyse, les sexes sont confondus et la comparaison est effectuée à l'échelle française. Par ailleurs, celle-ci s'effectue par tranche d'âge et après 1975. La raison est que les fichiers de population⁴ permettant de reconstruire les taux de mortalité ne contiennent l'information que par tranches d'âge à partir de cette année. La comparaison est représentée sur le graphique 2.1, en termes de pourcentage d'erreur par rapport à la base de référence.



GRAPHIQUE 2.1 – Écart de taux de mortalité entre le fichier des décès INSEE et la base HMD

Il est constaté une sous-estimation de la mortalité pour chaque classe d'âge jusqu'à l'année 1990 environ. Pour cette raison, utiliser les données de décès entre 1970 et 1989 conduirait à biaiser les estimations de la relation température-mortalité par rapport aux périodes futures. Par la suite, les écarts n'excèdent jamais 10% et se concentrent autour de 0, ce qui laisse suggérer une absence de biais par rapport à la base de référence. Les écarts résiduels peuvent s'expliquer par le lissage des données de mortalité de la base de référence. Plusieurs tests ont été menés afin de quantifier cette adéquation.

Tout d'abord, l'hypothèse d'absence de biais du fichier des personnes décédées par rapport à la base de référence HMD est vérifiée. Pour cela, une régression linéaire est estimée avec un coefficient de tendance sur la série temporelle des pourcentages d'erreur pour chaque année t et chaque classe d'âge x . Autrement dit, la meilleure approximation linéaire de chaque série (au sens des moindres carrés), sous la forme $\beta_{0,x} + t\beta_{1,x}$ est cherchée. Ensuite, l'hypothèse nulle (H_0) d'absence de significativité des coefficients $\beta_{0,x}$ et $\beta_{1,x}$ est testée au risque 5%. Les p -valeurs obtenues dans ces tests sont présentées dans le tableau 2.2.

4. Lien vers les données de population de l'INSEE : ↗

Âge	p -valeur Constante	p -valeur Tendence	Âge	p -valeur Constante	p -valeur Tendence
0-4	0,4480	0,9653	50-54	0,6550	0,8571
5-9	0,3216	0,2491	55-59	0,1605	0,2837
10-14	0,4644	0,0921	60-64	0,1063	0,2397
15-19	0,7615	0,4819	65-69	0,0504	0,1485
20-24	0,2460	0,4064	70-74	0,3303	0,3816
25-29	0,1847	0,2189	75-79	0,7671	0,6375
30-34	0,0819	0,1721	80-84	0,2914	0,7653
35-39	0,0587	0,1615	85-89	0,6389	0,9983
40-44	0,1862	0,5218	90-94	0,9451	0,3061
45-49	0,6304	0,5051			

TABLEAU 2.2 – Tests de significativité des coefficients d’erreur entre la base INSEE et la base HMD

Quelle que soit la classe d’âge x considérée, il n’est pas possible de rejeter au risque 5% l’hypothèse nulle (H_0) d’absence de significativité du coefficient d’erreur systématique $\beta_{0,x}$ et de celui de tendance $\beta_{1,x}$ dans l’évolution des erreurs. Sans constituer une preuve formelle de la complétude de la base retenue, ces tests autorisent une certaine confiance.

Pour compléter ces vérifications, il est possible d’effectuer une analyse de l’autocorrélation des pourcentages d’écart dans le temps. Il serait souhaitable que ces différences ne soient pas dépendantes dans le temps, ce qui serait le signe d’un potentiel "rattrapage" dans les décès enregistrés. D’après le graphique 2.1, il est par exemple possible de suspecter que des décès non-enregistrés dans l’an 1997 l’aient été en 1998. Cela conduirait à une sur-estimation des décès pour cette dernière année.

Afin de vérifier cette affirmation, un test de Ljung-Box a été lancé. Celui-ci permet de vérifier l’hypothèse nulle (H_0) de nullité jointe des autocorrélations jusqu’à un ordre k donné. La valeur $k = 1$ est choisie afin de se concentrer sur les rattrapages potentiels d’une année sur l’autre. On notera $\hat{\rho}_\ell$ l’autocorrélation empirique entre les pourcentages d’erreur d’une année t et ceux de l’année $t + \ell$. On note n le nombre d’observations (ici $n = 30$). Sous (H_0), la statistique de test

$$Q(k) = n(n+2) \sum_{\ell=1}^k \frac{\hat{\rho}_\ell^2}{n-\ell}$$

suit une loi du χ_2 à k degrés de liberté. Les résultats du test de Ljung-Box pour chaque classe d’âge sont présentés dans le tableau 2.3.

En dehors des classes d’âge 50-54 ans et 90-94 ans, il n’est pas possible de rejeter l’hypothèse nulle (H_0) d’absence d’autocorrélation des pourcentages d’erreurs à l’ordre 1 au risque 5%. Les effets de "rattrapage" des décès recensés, s’ils sont présents, semblent donc relativement négligeables et ne constituent pas un frein à l’étude menée ici. La période 1990-2019 semble donc un bon point de départ. Les années 2020 à 2022 sont exclues du fait de l’épidé-

Âge	<i>p</i> -valeur	Âge	<i>p</i> -valeur
0-4	0,6093	50-54	0,0488
5-9	0,4183	55-59	0,1482
10-14	0,5871	60-64	0,2948
15-19	0,3896	65-69	0,0570
20-24	0,5290	70-74	0,2579
25-29	0,2254	75-79	0,1250
30-34	0,2900	80-84	0,5310
35-39	0,4423	85-89	0,1199
40-44	0,7318	90-94	0,0007
45-49	0,7951		

TABLEAU 2.3 – Tests de Ljung-Box de l'autocorrélation d'ordre 1 des erreurs entre la base INSEE et la base HMD

mie de Covid-19, qui risque de fausser l'estimation de la relation historique entre variables atmosphériques et mortalité.

2.1.3 Les données climatiques

La modélisation de la mortalité à une maille départementale nécessite, également, de disposer des variables explicatives liées au climat à cette granularité. À ce stade, plusieurs solutions sont possibles.

La première consiste à s'appuyer sur des données publiques issues des stations météorologiques, puis à appliquer une méthode d'interpolation spatiale pour déduire les valeurs de ces variables à la maille géographique souhaitée. En France, il est par exemple possible de s'appuyer sur les 42 stations recensées dans la base publique SYNOP⁵, réparties dans l'ensemble des régions de France métropolitaine. Ce choix a par exemple été utilisé par Pincemin [47], autrice d'une étude similaire à celle-ci, qui a obtenu une variable de température moyenne journalière en France à partir de différentes stations météorologiques. Cette variable a ensuite été croisée au fichier des décès de l'état civil (dont l'accès est restreint) afin d'obtenir une relation entre température et mortalité puis des projections de cette dernière. Si cette approche permet de conserver un portefeuille d'assurés important par tranche d'âge, elle peut conduire à minorer des températures extrêmes et donc à ne pas entièrement capter l'effet de la température sur la mortalité. Par ailleurs, dans le cas de cette étude, cette solution aurait conduit à utiliser des données disponibles seulement depuis 1996, ce qui amènerait à se passer de la moitié de l'historique de décès. Afin d'obtenir une maille géographique affinée, le *krigeage* est une solution potentielle. Il consiste à interpoler géographiquement les variables météorologiques disponibles pour obtenir une approximation aux endroits pour lesquels l'information

5. Lien vers la base de données SYNOP de Météo-France : ↗

est manquante. Le krigeage cherche à minimiser la variance de l'estimation interpolée tout en restant fidèle aux données connues. Dans sa thèse relative à l'interpolation spatiale des données de précipitations, Baillargeon [5] mentionne plusieurs méthodes de krigeage, tels que le krigeage simple, le krigeage universel et le krigeage co-krigé :

- Le krigeage simple : il utilise la distance pour pondérer les valeurs connues ;
- Le krigeage universel : il utilise des variables supplémentaires qui peuvent aider à mieux modéliser la variabilité spatiale ;
- Le krigeage co-krigé : il est utilisé lorsque plusieurs variables sont corrélées spatialement.

Cependant, cette méthode s'avère excessivement lourde en temps de calcul, s'agissant d'interpoler des données journalières sur une plage historique large. Par exemple, un *krigeage* classique d'une variable sur une seule période (le jour) pour l'ensemble du territoire a nécessité près de 4 minutes de temps de calcul⁶. Dans une étude actuarielle, il est donc préférable de s'appuyer sur une base de données disposant déjà des données nécessaires, au besoin *krigées* préalablement.

Une recherche d'autres bases de données disposant déjà d'observations à une maille géographique affinée a donc été menée. La base *E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations*⁷ a été identifiée, disposant de nombreuses variables climatiques (telles que la température, le taux d'humidité dans l'air, mais aussi les précipitations et la vitesse du vent). Les données, fournies par 84 instituts, renseignent ces variables pour plus de 23 000 stations météorologiques à l'échelle européenne. La résolution ainsi obtenue est de $0.1^\circ \times 0.1^\circ$, permettant ainsi de recouvrir l'ensemble des départements du territoire français, y compris ceux disposant d'une faible superficie (la plus faible étant observée dans les Hauts-de-Seine, avec 176 km^2). Ces variables sont décrites dans le tableau 2.4.

La température moyenne et l'humidité relative ont été conservées. Il s'agit en effet des variables explicatives retrouvées le plus souvent dans la littérature, ce qui permet de disposer d'une source de comparaison des résultats obtenus. Utiliser la température minimale et maximale aurait conduit à complexifier l'estimation des modèles, alors que ces variables fournissent des informations assez proches de celles associées à la température moyenne (pour une zone géographique donnée). L'utilisation des précipitations dans l'analyse de la mortalité est, elle, discutée en annexe. La base de données des variables météorologiques, lourde de 12

6. Krigeage gaussien réalisé sur 10 000 points uniformément répartis en France métropolitaine.

7. Lien vers la base E-OBS fournie par l'institut Copernicus : ↗

<i>Variables</i>	<i>Description et traitement éventuel</i>
Température maximale (°C)	Température maximale quotidienne de l'air mesurée près de la surface, généralement à une hauteur de 2 mètres
Température moyenne (°C)	Température moyenne quotidienne de l'air mesurée près de la surface, généralement à une hauteur de 2 mètres
Température minimale (°C)	Température minimale quotidienne de l'air mesurée près de la surface, généralement à une hauteur de 2 mètres
Quantité de précipitations (mm)	Quantité totale journalière de pluie, de neige et de grêle mesurée en tant que hauteur d'eau liquide équivalente dans un mètre carré
Humidité relative (%)	Humidité relative moyenne journalière mesurée près de la surface, généralement à une hauteur de 2 mètres

TABLEAU 2.4 – Variables disponibles dans la base de données climatiques E-OBS

Go, a été traitée pour ne conserver que les données de France métropolitaine. Les étapes sont décrites dans le schéma 2.2.

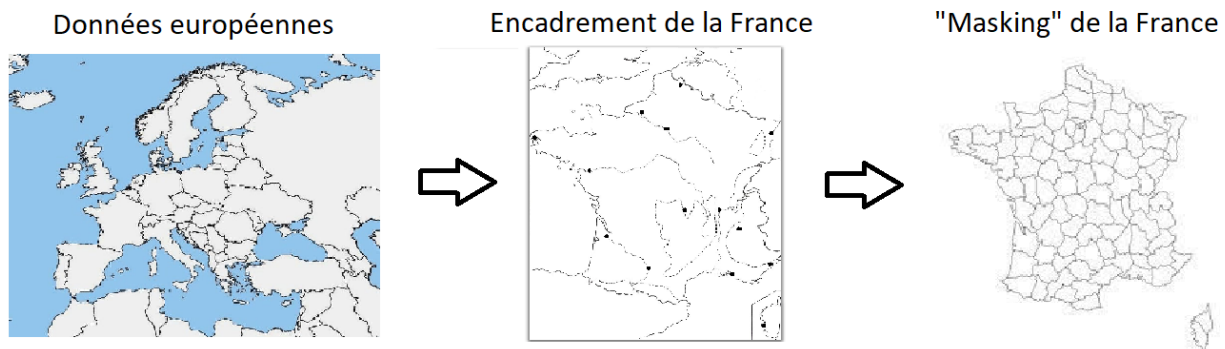


SCHÉMA 2.2 – Processus de traitement géographique des données de la base E-OBS (étapes 1 à 3, de gauche à droite)

Tout d’abord, on ne conserve que les coordonnées figurant dans un rectangle dont les arêtes sont les points limites de la France métropolitaine. Cette première étape permet d’éliminer une grande partie des données inutiles, de façon computationnellement simple. Ensuite, une représentation polygonale complexe de la France est utilisée afin de ne conserver que les points appartenant au territoire. Ce processus, dit de *masking*, nécessite un usage en mémoire vive important. C’est pourquoi le premier traitement des données, même grossier, est nécessaire afin que le processus global aboutisse sur un ordinateur de puissance moyenne.

La base obtenue, de 4,7 Go, reste particulièrement lourde et nécessite des traitements

adaptés. Par exemple, la température du bulbe mouillé (décrite dans le premier chapitre) est rajoutée par vectorisation, ce qui permet de limiter le temps de calcul. Enfin, on conserve la moyenne des valeurs obtenues pour chaque département et chaque jour. Chaque département dispose d'au moins une donnée qui pourra être exploitée, *modulo* un nombre négligeable d'interpolations qui seront nécessaires. Par exemple, dans les Hauts-de-Seine (département disposant de la plus faible superficie), sur 365 jours, en moyenne 352 jours de données sont disponibles. Les valeurs manquantes sont alors interpolées en prenant la moyenne du jour précédent et du jour suivant. La base atmosphérique finale ne pèse que 46 Mo et peut facilement être fusionnée avec le fichier des décès.

2.2 Des facteurs climatiques liés à la saisonnalité

Ce chapitre vise à motiver l'utilisation des variables mentionnées plus haut. Pour cela, une étude est menée à partir de multiples sources de données, à la fois à l'échelle mondiale et à l'échelle nationale. Cette approche possède l'avantage d'autoriser un certain recul quant à la portée des résultats obtenus, car elle se base sur davantage d'observations. Par ailleurs, les différences entre les relations obtenues sur différents territoires permettront de motiver le choix d'une étude menée sur plusieurs zones géographiques.

Dans un premier temps, la saisonnalité de la mortalité est soulignée, et mise en perspective avec l'évolution des températures sur une année. Ensuite, les paramètres économiques sont évoqués car ils sont corrélés avec la tendance de la mortalité, qu'il est important de capter dans une étude prospective.

2.2.1 Une saisonnalité marquée de la mortalité

Sur une année, les nombres de décès sont amenés à fluctuer de façon significative. Il s'agit d'un fait empirique constaté tant à l'échelle mondiale qu'à l'échelle nationale.

Saisonnalité à l'échelle mondiale

La base de données STMF⁸ (*Short-Term Mortality Fluctuations*) est une source récente à l'échelle mondiale afin d'étudier les fluctuations de la mortalité à court-terme. Celle-ci a été créée par les concepteurs de la base HMD durant la pandémie de Covid-19 afin de mieux

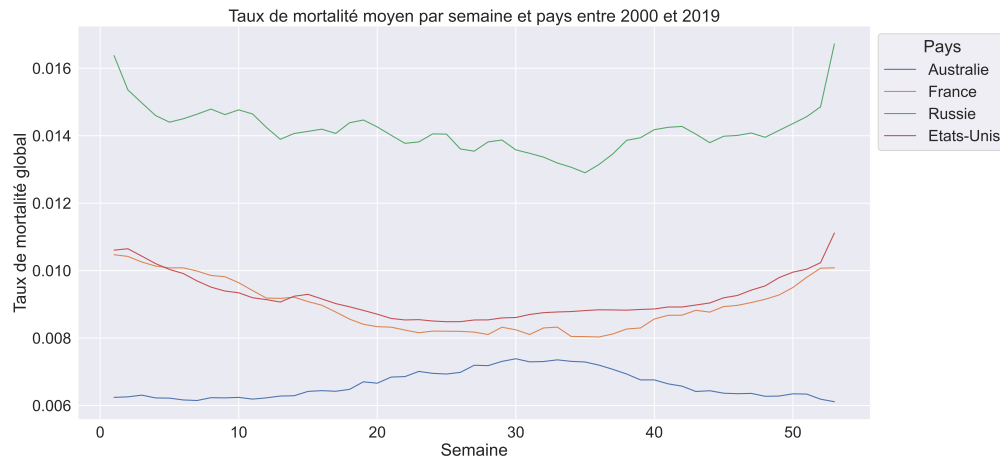
8. Lien vers la base de données STMF : [↗](#)

mesurer et comparer l'efficacité des mesures qui ont été prises. L'historique est toutefois disponible à partir de 1995 (l'année de début varie selon les pays), ce qui permet une analyse plus robuste de la saisonnalité des décès. Alors que les sources de référence en accès libre (HMD et GBD notamment) fournissent des informations à la maille annuelle uniquement, la base STMF dispose des données de décès à la maille hebdomadaire.

Dans cette base, les données sont disponibles sous forme de taux de mortalité. Ceux-ci sont multipliés par 52 afin de retrouver des valeurs comparables avec les taux de mortalité annuels plus couramment utilisés. Afin de constater la saisonnalité des décès et les potentielles différences par zones géographiques, l'approche retenue est la suivante :

- Quatre pays issus de quatre continents différents, avec des climats distincts, sont retenus. Il s'agit ici de la France, des États-Unis, de l'Australie et de la Russie.
- Pour chaque semaine, la moyenne des taux de mortalité hebdomadaires est calculée.

Les résultats obtenus sont représentés sur le graphique 2.3.



GRAPHIQUE 2.3 – Taux de mortalité hebdomadaire moyen entre 2000 et 2020 (source : STMF)

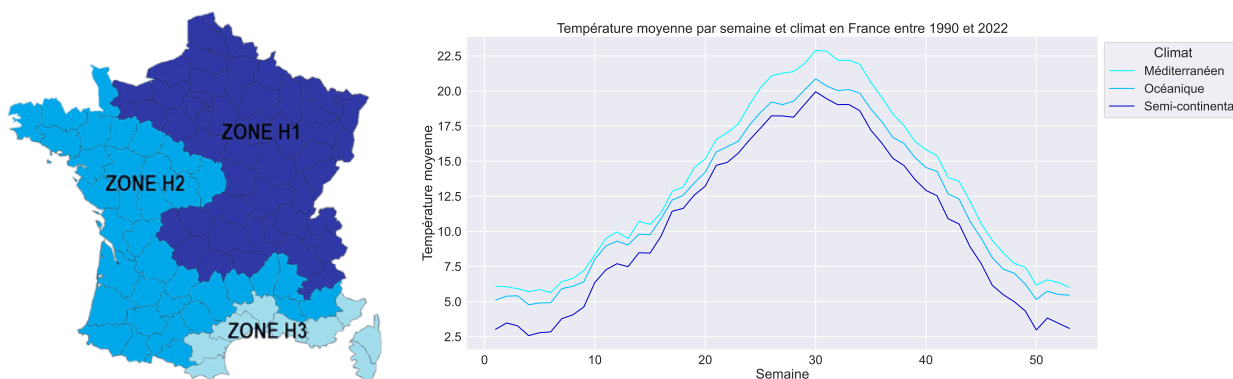
Il ressort de ce graphique une saisonnalité claire, matérialisée par une courbe en "U". Un pic de mortalité est observé pendant l'hiver, alors que la mortalité a tendance à devenir minimale durant l'été. En Australie, l'été ayant lieu entre les mois de décembre et février, la courbe est inversée sans que cela n'impacte les interprétations de ce graphique. Les températures étant un élément différenciant des saisons, ce graphique laisse suggérer un rôle important des variations de cette variable sur la mortalité. Selon les pays, les fluctuations sont différentes : la Russie semble davantage sujette aux pics de froid, alors que ces pics sont plus intenses en Russie qu'en France ou aux États-Unis, par exemple. En effet, plus des deux tiers du territoire russe connaissent six mois de gel consécutifs et des moyennes de janvier

inférieures à -20°C ⁹. Il est également possible de constater quelques regains occasionnels de mortalité durant l'été, sans qu'il ne soit possible d'affirmer si les températures élevées constatées à cette période de l'année en soient la cause.

Par ailleurs, les taux de mortalité constatés sont supérieurs en Russie par rapport aux autres pays retenus pour cette comparaison. La différence est de l'ordre d'un facteur multiplicatif égal à 2 entre la Russie et l'Australie. Il serait intéressant de vérifier si de telles variations sont également observables à plus petite échelle.

Saisonnalité à l'échelle nationale

À partir des fichiers de mortalité de l'INSEE, il est possible d'établir une étude similaire à l'échelle française. Afin d'obtenir une comparaison similaire à la précédente, la France métropolitaine est découpée en plusieurs types de climats. La classification retenue est celle du Ministère de la Transition écologique. Elle a été officialisée dans l'arrêté du 24 mai 2006¹⁰ relatif aux caractéristiques thermiques des bâtiments nouveaux et des parties nouvelles de bâtiments. Trois grands types de climats sont introduits : méditerranéen (zone H1), océanique (zone H2) et semi-continentale (zone H3). À partir du jeu de données climatique E-OBS, il est possible d'obtenir quelques statistiques permettant de décrire ces types de climat. L'emplacement géographique de ces climats ainsi que les températures moyennes par semaine sont représentés dans le graphique 2.4.



GRAPHIQUE 2.4 – Zones climatiques H1/H2/H3 et températures moyennes hebdomadaires

Quelques statistiques descriptives relatives à ces types de climats sont disponibles dans le tableau 2.5.

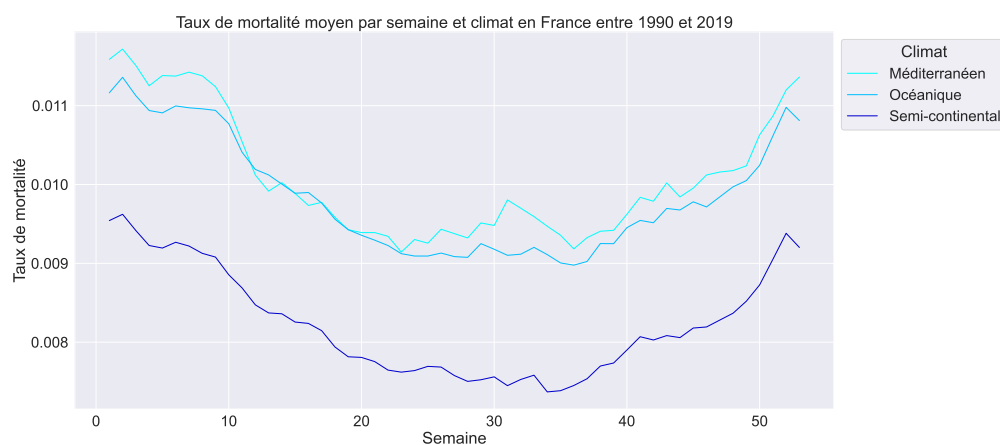
9. Source : [☞](#)

10. Lien vers l'arrêté du Ministère de la Transition écologique : [☞](#)

Climat	<i>Température moyenne (en °C)</i>			
	Moyenne annuelle	Minimum	Maximum	Écart-type
Méditerranéen	13,23	5,65	22,89	5,91
Océanique	12,07	4,78	20,87	5,46
Semi-continental	10,54	2,58	19,95	5,84

TABLEAU 2.5 – Statistiques descriptives des températures moyennes hebdomadaires par type de climat en France métropolitaine (1990-2019)

Il ressort que le climat méditerranéen est celui pour lequel les températures sont généralement les plus élevées tout au long de l'année. Alors que la zone océanique profite de températures douces et comparables à celles du Sud du pays pendant l'hiver, la chaleur y est plus modérée durant l'été. Au contraire, les hivers sont plus rigoureux dans l'Est de la France alors que les étés y sont légèrement plus frais que ceux observés dans le climat océanique. Il est nécessaire de rappeler que la classification utilisée ici est simplifiée à des fins de compréhension. Des zones géographiques réduites sont également introduites dans la classification officielle, mais les utiliser aurait rendu la comparaison plus complexe. L'objectif principal est d'avoir un premier indice quant à l'importance des variables climatiques dans l'évolution de la mortalité à l'échelle française. Les taux de mortalité hebdomadaires par zone climatique peuvent être obtenus à partir du fichier des décès de l'INSEE. Ils sont représentés sur le graphique 2.5.



GRAPHIQUE 2.5 – Taux de mortalité hebdomadaire moyen par zone climatique (1990-2019)

De manière similaire à ce qui est observé à l'échelle mondiale, les taux de mortalité possèdent une forme en "U", avec un nombre de décès plus important en hiver et au plus bas durant l'été. Les climats méditerranéen et océanique semblent également disposer de dynamiques des taux de mortalité relativement proches. Cette similarité pourrait être expliquée

par une acclimatation des individus, ce qui tendrait à diminuer les disparités géographiques de mortalité. Ce point particulier fera l'objet d'une étude approfondie dans ce mémoire.

Le graphique précédent souligne toutefois l'existence de disparités de mortalité à l'échelle française, à la fois temporelles et géographiques. Géographiquement, le climat semi-continentale présente une mortalité inférieure à celle observée pour les autres types de climat, de l'ordre de 0,002 point. Ce point est en fait expliqué par une sous-mortalité observable dans la région Île-de-France : une explication sera apportée ultérieurement à l'aide d'une analyse comparative plus fine. Temporellement, un regain de mortalité est observé au cours de l'été dans le climat méditerranéen, contrairement au climat océanique. Les différences de températures entre ces deux climats lors de l'été pourraient en être la cause.

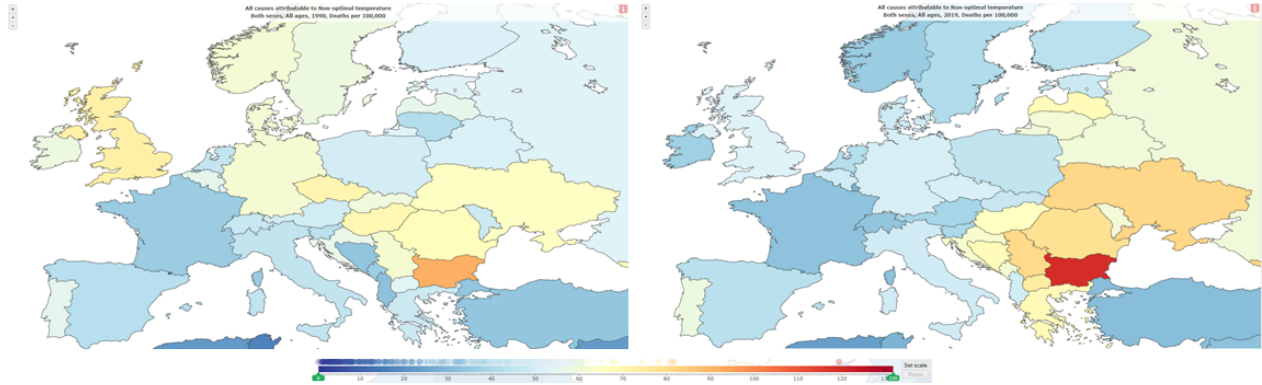
Bien que l'évolution des températures soit nettement corrélée à celle de la mortalité, aucune causalité n'a encore été démontrée. Par ailleurs, cette comparaison simplifiée ne tient pas compte des disparités géographiques dans la structure des âges. La part des personnes âgées étant moindre dans la région Île-de-France¹¹, le taux de mortalité global (sur l'ensemble de la population) y est logiquement réduit. L'analyse précédente a donc besoin d'être renforcée par une quantification plus précise du lien entre température et mortalité.

2.2.2 La difficile quantification de la mortalité liée aux phénomènes climatiques

La base GBD 2019 permet une analyse plus fine du rapport entre les facteurs climatiques et la mortalité. En effet, parmi les facteurs sous-jacents renseignés parmi les décès, figurent plusieurs causes environnementales : la pollution de l'air, les températures faibles et élevées ainsi que des facteurs environnementaux tiers. Il est donc possible de disposer d'une première quantification du risque direct de décès associé à ces éléments climatiques. Le graphique 2.6 présente l'évolution des taux de décès estimés pour 100 000 habitants à l'échelle européenne en raison des températures non-optimales (froid ou chaleur). La même cartographie est représentée en 1990 (à gauche) et en 2019 (à droite).

D'après ce graphique, la part des taux de décès imputables à l'environnement a diminué entre 1990 et 2019 principalement dans l'Europe de l'Ouest. La diminution est notamment perceptible en Allemagne, au Royaume-Uni et en France. Mis en perspective avec l'étude de saisonnalité, ce graphique permet de présenter une conjecture. Sachant que les décès sont plus nombreux en hiver, au moment des vagues de froid, l'augmentation générale des températures

11. Lien vers la pyramide des âges en 2023 : ↗

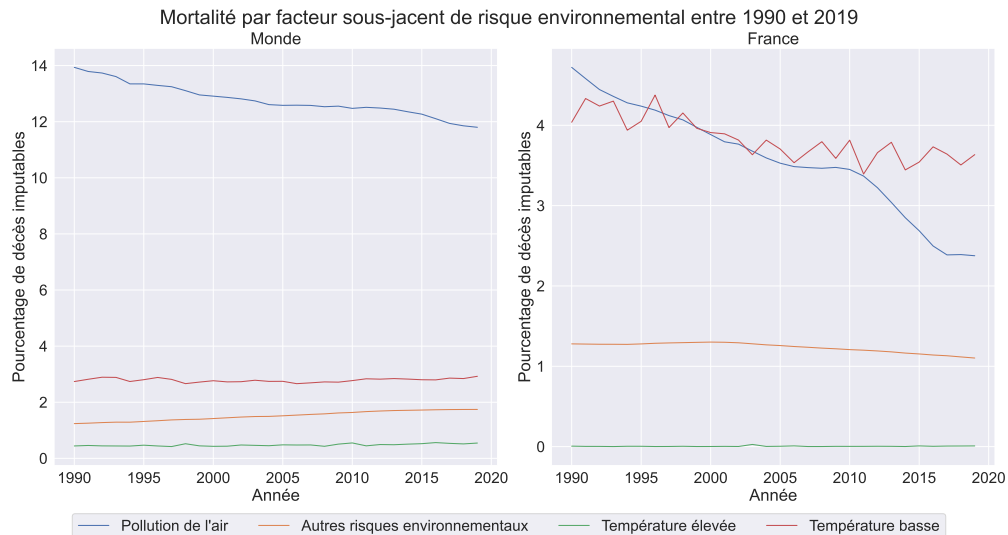


GRAPHIQUE 2.6 – Taux de mortalité pour 100 000 habitants imputables aux températures non-optimales : en 1990 (à gauche) et en 2019 (à droite) (source : GBD 2019)

Lecture : En 2019, on estime à environ 20 décès pour 100 000 habitants en France la part liée aux températures non-optimales. Cette valeur est estimée à 130 décès pour 100 000 habitants en Bulgarie en 2019.

pourrait être associée à une diminution de la mortalité à cette période de l’année. Une autre conséquence serait une hausse des décès observés durant l’été, mais moins importante que la baisse des décès précédemment décrite.

Afin de vérifier cette conjecture, les évolutions des pourcentages de décès associés aux différents facteurs de risque environnementaux sont représentés dans le graphique 2.7. La représentation est effectuée à la fois pour la France, et à l’échelle mondiale.



GRAPHIQUE 2.7 – Évolution de la proportion de décès liée aux facteurs environnementaux : dans le monde (à gauche) et en France (à droite) (source : GBD 2019)

D’après ce graphique, les principaux risques environnementaux décelés dans la base GBD 2019 sont les températures non-optimales (vagues de chaud et de froid) et la pollution de l’air.

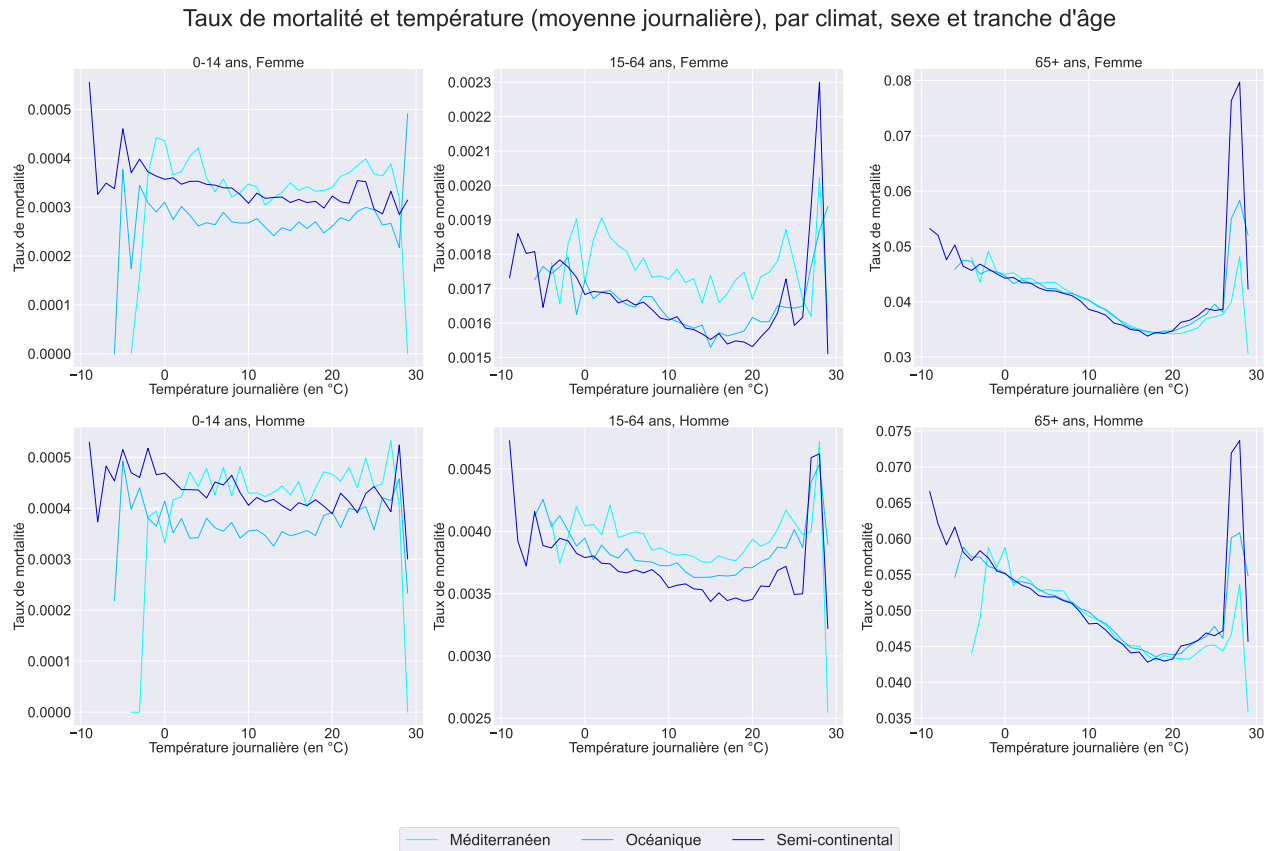
Ils représenteraient ensemble environ 8% des décès enregistrés au total en France. Comme conjecturé, la part des décès associée aux vagues de froid a légèrement diminué en France sur la période d'étude. Cependant, la base GBD 2019 suggère également une diminution continue de la mortalité directement imputable à la chaleur. Ce dernier résultat semble surprenant, compte tenu des pics de chaleur particulièrement mortels connus en France durant cette période. Par exemple, la base GBD 2019 recense moins de 1 000 décès imputables à la chaleur en France en 2003, alors que la canicule du mois d'août aurait occasionné environ 15 000 décès selon l'INSERM¹². Ce dernier résultat met en avant la difficulté à estimer de manière incontestée l'impact des facteurs climatiques sur la mortalité. Il est de fait possible que les vagues de chaleur et de froid occasionnent un nombre de décès plus important que celui estimé dans la base GBD 2019. La sous-section suivante cherche à présenter plus en détails la manière dont les températures élevées et basses influencent les nombres de décès, à partir de statistiques descriptives.

2.2.3 Comparaison des effets de la chaleur et du froid

Avant toute modélisation de la relation entre température et mortalité, il est important de comprendre comment celle-ci se manifeste afin d'obtenir des premières pistes utiles pour la suite. Pour cela, pour chaque grande classification de climat (H1, H2 et H3), sexe et classe d'âge, la relation immédiate entre température et taux de mortalité est d'abord représentée. Pour cela, pour chaque jour entre 1990 et 2019, une température moyenne est calculée à la maille du climat retenu. Celle-ci est arrondie à l'unité afin de conserver suffisamment d'observations. Le nombre de décès ainsi que la population vivante de l'année dans les départements sont agrégés afin d'obtenir un taux de mortalité journalier pour chaque type de climat. De manière similaire à ce qui a été fait pour calculer les taux de mortalité hebdomadaires, les taux de mortalité journaliers sont multipliés par 365,25 afin d'obtenir des valeurs comparables avec les taux de mortalité annuels couramment utilisés. Finalement, il est possible d'associer, pour chaque température moyenne observée sur la période d'étude, un taux de mortalité moyen qui permet d'obtenir une première approximation de la relation entre ces deux variables. Celle-ci est représentée sur le graphique 2.8, pour chaque sexe et grande classe d'âge (0-14, 15-64 ans, 65 ans et plus). Ce choix des classes d'âge permet de distinguer les enfants, les actifs et les personnes retraitées. Il sera revu par la suite.

Il est possible d'y constater que la relation de la température à la mortalité semble varier selon l'âge et le climat. Les différences par sexe ne sont visibles ici que par le biais d'un

12. Lien vers le rapport : ↗



GRAPHIQUE 2.8 – Taux de mortalité journalier associé à la température moyenne observée le même jour, pour chaque climat, sexe et grande classe d'âge (entre 1990 et 2019)

taux de mortalité supplémentaire chez les hommes qui ne dépend pas de la température. D'une part, en ce qui concerne les différences par classes d'âge, une plus grande mortalité lors des vagues de froid s'observe chez les individus les plus jeunes. Au contraire, ces derniers semblent moins sujets aux conséquences des pics de chaleur. Pour les individus plus âgés, il est possible d'observer une relation en "U" avec une température minimisant la mortalité (appelée dans la littérature *Minimum Mortality Temperature* ou TMM). Celle-ci est particulièrement visible chez les personnes âgées de 65 ans et plus, et se situe entre 15 et 20°C selon les différents climats. Des modélisations relativement différentes de la relation entre température et mortalité pourraient donc exister selon l'âge.

La TMM est une variable particulièrement intéressante dans une optique de santé publique, puisque sa connaissance permet d'évaluer l'existence d'une acclimatation, à la fois dans le temps et de l'espace, des individus aux conditions climatiques locales. Il s'agit donc d'un indice intéressant dans l'optique d'une étude prospective, puisque cette acclimatation permettra potentiellement d'atténuer les conséquences sur la mortalité d'une augmentation continue des températures dans le temps. En-dessous de la TMM, la progression de la mor-

talité à mesure que la température s'en éloigne est assez linéaire. Cela est confirmé par Santé Publique France : selon cet institut, le froid tue de façon indirecte en aggravant des pathologies existantes, telles que les maladies coronariennes, les accidents vasculaires cérébraux et les maladies respiratoires. La mortalité liée à ces maladies augmenterait quasiment de façon linéaire à mesure que la température diminuerait. Au-delà de la TMM, les taux de mortalité progressent en revanche de façon exponentielle, ce qui est un signe inquiétant quant aux conséquences potentielles du changement climatique. En effet, d'après Météo-France¹³, la fréquence des vagues de chaleur devrait doubler d'ici à 2050, et s'accompagner d'une hausse de leur sévérité et de leur durée d'ici la fin du siècle. Cette relation différenciée selon que la température soit inférieure ou supérieure à la TMM est confirmée par le professeur San Marco, cité par le Figaro¹⁴. Selon lui, la chaleur a besoin d'être caniculaire pour être dangereuse, alors que la surmortalité hivernale se manifeste dès que la température est inférieure à 15°C. De plus, il n'y aurait pas, selon ce professeur, de rupture de la courbe de mortalité pour le froid, au contraire de ce qui est visible pour la chaleur.

La représentation précédente reste toutefois incomplète. D'une part, elle associe à certaines températures extrêmes des valeurs aberrantes : la température maximale observée est ainsi liée à une mortalité journalière assez faible, alors que le pic de décès est atteint pour une température voisine de 27°C. L'inclusion de l'humidité pourrait permettre de corriger cette relation, compte tenu de son importance majeure pour la supportabilité des températures élevées. D'autre part, l'impact parfois retardé des températures sur la mortalité n'est également pas exploité dans cette représentation. Selon Santé Publique France¹⁵, les conséquences d'une température froide sur la mortalité peuvent être visibles dans une période comprise entre 3 et 21 jours. Pour cette raison, il pourrait être utile d'inclure des variables retardées de la température dans la modélisation. Cette conjecture sera vérifiée dans le chapitre 3.

Parmi les facteurs externes associés à la mortalité, la tendance est un élément potentiellement important dans une étude prospective. La section suivante s'attache à présenter les éléments pouvant expliquer cette variation de la mortalité dans le temps ainsi que les différences géographiques pouvant exister.

13. Lien d'un article de Météo-France sur le changement climatique et les vagues de chaleur : [☞](#)

14. Lien vers l'article du Figaro : [☞](#)

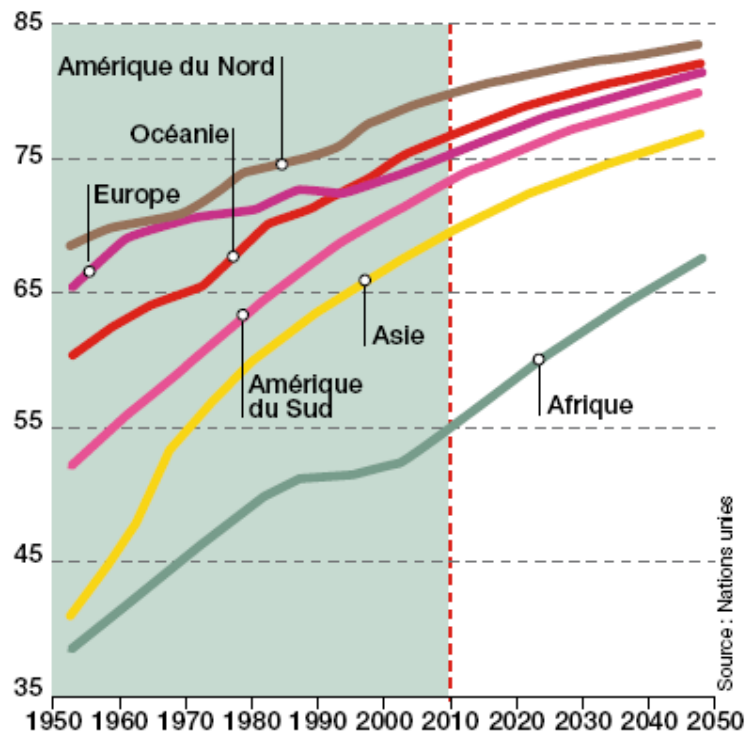
15. Voir les explications dans cet article : [☞](#)

2.3 Des facteurs économiques corrélés à la tendance

2.3.1 Une corrélation de la richesse à l'espérance de vie à l'échelle mondiale

Les taux de mortalité étant définis pour chaque tranche d'âge, il est plus facile d'utiliser un indicateur unique afin de mieux comprendre son évolution. Pour cette raison, l'indicateur de l'espérance de vie à la naissance est utilisé. Il s'agit du nombre d'années qu'un individu peut espérer vivre à sa naissance. Cet indicateur démographique est répandu à l'échelle mondiale, ce qui n'est pas forcément le cas des taux de mortalité, surtout disponibles dans les pays développés. Il est donc plus facilement utilisable pour une comparaison géographique internationale.

Afin d'avoir une première idée de la tendance de la mortalité dans le temps, l'évolution de l'espérance de vie est représentée pour chaque continent depuis 1950 dans le graphique 2.9.



GRAPHIQUE 2.9 – Évolution de l'espérance de vie à la naissance entre 1950 et 2050 (source : Nations Unies)

Ce graphique permet d'observer une évolution différenciée de l'espérance de vie selon les

continents. Par exemple, alors que l'Amérique du Sud et l'Asie connaissent une augmentation assez rapide, l'Amérique du Nord et l'Europe observent une progression plus lente. À l'échelle européenne, d'après Steel et al. [54], une augmentation des maladies et crises cardiaques serait à l'origine de ce ralentissement lors de la dernière décennie, avec des facteurs de risque sous-jacents variables selon les pays. Selon les projections des Nations Unies, les espérances de vie de ces différents continents seraient amenées à converger d'ici à 2050. Il est à noter que cette convergence était déjà visible en 2010 à l'échelle des pays développés, selon Loisel et al [39]. L'espérance de vie à la naissance était comprise entre 75 et 80 ans au Japon, aux Pays-Bas, en Suisse, en Belgique, en Angleterre et en France. L'Afrique semble toutefois en retrait, mais connaîtrait à son tour une augmentation de l'espérance de vie plus marquée sur les trente prochaines années, de l'ordre de 15 ans contre moins de 5 dans les autres continents. Ces derniers étant caractérisés par des niveaux de développement différents, il est possible de s'interroger sur la corrélation entre le niveau de richesse par pays et la mortalité.

Le jeu de données GapMinder¹⁶ permet de représenter la relation entre le PIB par habitant (mesuré en dollars constants) et l'espérance de vie dans 142 pays, avec un historique compris entre 1952 et 2007, par tranches de 5 ans. Les données sont issues de l'ONU (Organisation des Nations Unies) et de l'IHME (*Institute for Health Metrics and Evaluation*). Sur l'animation de Thomas Lin Pedersen¹⁷, il est possible de constater une progression commune de la richesse par habitant et de l'espérance de vie. La nature de cette progression est encore une fois différenciée selon les continents : alors que l'évolution semble disparate dans les pays africains et asiatiques, la relation est bien plus homogène à l'échelle européenne. Ce graphique encourage donc l'utilisation de données de richesse afin d'estimer la tendance de mortalité en France. L'utilisation d'un unique paramètre de richesse pour l'ensemble des départements paraît également envisageable à ce stade, à condition que l'homogénéité de la relation entre richesse et mortalité soit visible aussi à l'échelle infra-nationale.

2.3.2 Une augmentation continue mais ralentie de l'espérance de vie en France

La French Regional Database¹⁸, rendue accessible en 2021, permet l'accès à un certain nombre de données démographiques très utiles pour cette étude. Sa particularité est de disposer de l'information à l'échelle départementale, que ce soit pour les taux de mortalité

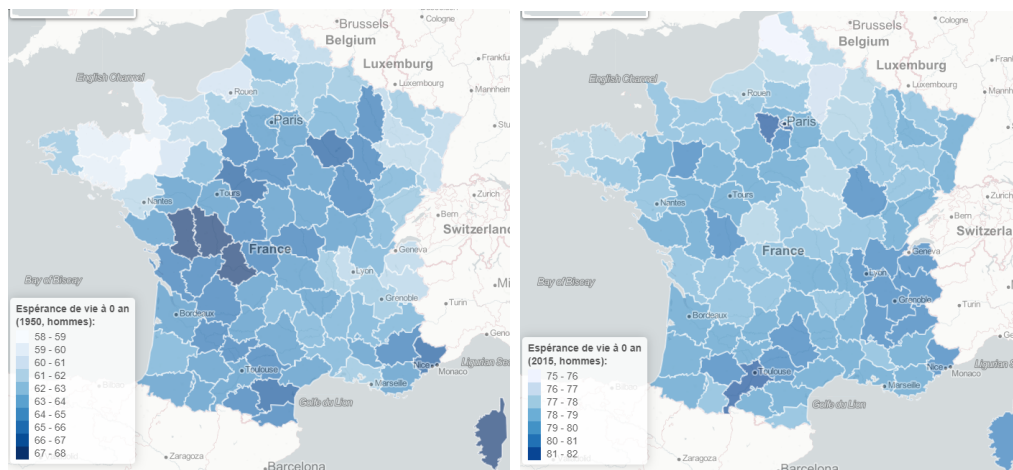
16. Lien vers la documentation du *package* R : ☞

17. Lien vers l'animation : ☞

18. Lien vers la French Regional Database : ☞

ou les indicateurs de richesse. Jusqu'alors, les données de référence relatives à la mortalité n'étaient disponibles qu'à l'échelle nationale, tandis que les statistiques de l'INSEE en libre accès concernant la richesse des individus ne disposent que d'un historique limité.

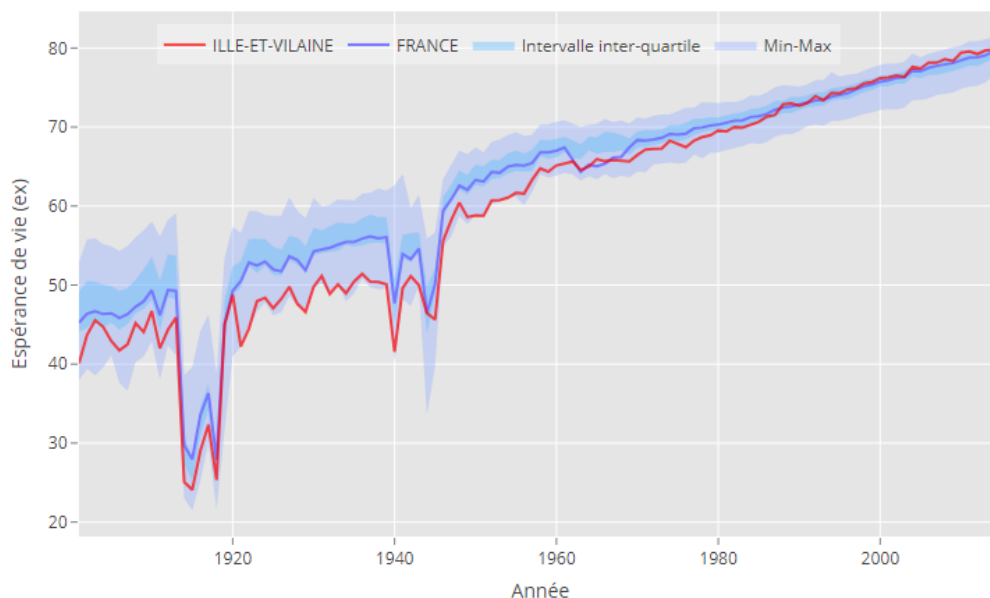
En plus des tables de mortalité disposant d'un large historique (les données sont disponibles à partir de 1900), il est possible de connaître l'espérance de vie des individus par département. Cette information permet à la fois d'étudier la tendance de la mortalité ainsi que les disparités géographiques qui y sont associées. Le graphique ci-dessous présente l'espérance de vie à la naissance des hommes en 1950 et 2015.



GRAPHIQUE 2.10 – Espérance de vie à la naissance des hommes en 1950 (à gauche) et 2015 (à droite) (source : capture d'écran French Regional Database)

Il est possible d'y constater que certains départements, dont l'espérance de vie était plus faible en 1950 comparativement aux autres, disposent désormais d'une espérance de vie supérieure à la moyenne nationale. C'est par exemple le cas des départements de Bretagne, de Basse-Normandie ou d'Alsace. De plus, alors que les disparités d'espérance de vie à la naissance entre deux départements pouvaient aller jusqu'à 10 ans en 1950, cette étendue n'est plus que de 7 ans en 2015. Cette observation est conforme à la convergence des espérances de vie conjecturée à l'échelle mondiale dans la sous-section 2.3.1. Le graphique 2.11 permet de confirmer cette affirmation via la comparaison des quantiles d'espérance de vie, ainsi que l'étude du département de l'Île-et-Vilaine. Il permet également de fournir une perspective historique plus large, remontant à 1900. Alors que ce département disposait de l'espérance de vie la plus faible en 1950, celle-ci s'est largement améliorée depuis.

Il est d'abord possible d'observer deux chutes soudaines de l'espérance de vie à la naissance des hommes au cours des deux guerres mondiales. Ensuite, l'écart inter-quartile d'espérance de vie a largement diminué, allant dans le sens d'une uniformisation de la mortalité à l'échelle

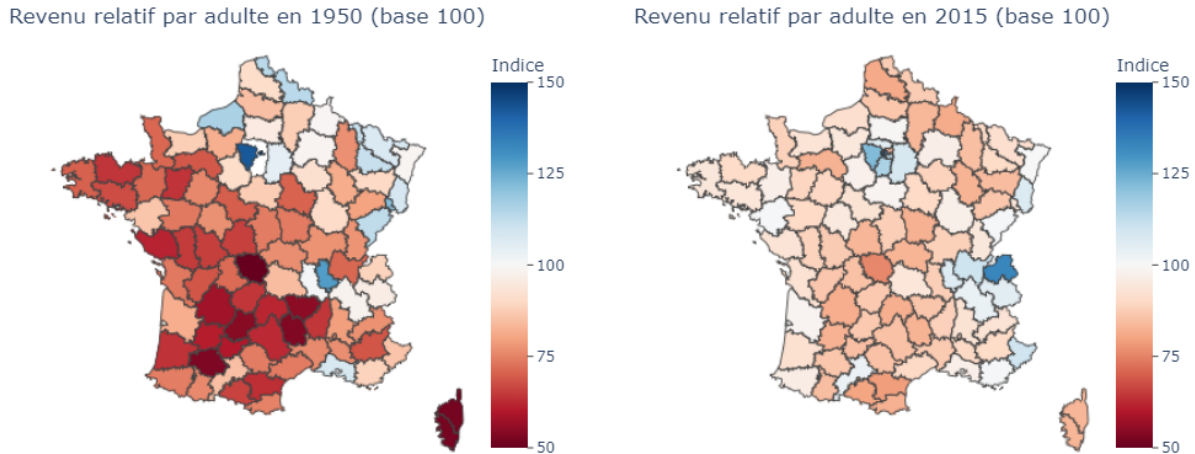


GRAPHIQUE 2.11 – Espérance de vie à la naissance des hommes : comparaison inter-quartile et département d’Île-et-Vilaine (source : capture d’écran French Regional Database)

nationale. Enfin, les fluctuations de mortalité par année ont également diminué, ce qui se traduit par un comportement assez linéaire de la courbe d’espérance de vie à partir de 1970. Celle-ci est visible à l’échelle nationale ainsi que du département d’Île-et-Vilaine. Pour expliquer cette uniformisation de la mortalité, il a semblé utile de s’intéresser à un indicateur de dispersion des revenus par département. En effet, une enquête menée en 2008 par l’Institut de Recherche et de Documentation en Économie de la Santé (IRDES)¹⁹ auprès de 20 000 personnes a permis de mettre en lumière certaines raisons du lien entre la richesse et l’accès aux soins. Certains soins ne sont en effet pas ou mal couverts par la Sécurité Sociale, laissant une partie des frais aux ménages. Les plus modestes d’entre eux souscrivent moins souvent à des mutuelles, ce qui peut les conduire à se dispenser de certains soins de santé. Les revenus relatifs par adulte sont donc utilisés et représentés en 1950 et 2015 dans le graphique 2.12. Il s’agit du rapport (en base 100) du revenu fiscal moyen des adultes à l’année t dans le département avec celui observé en moyenne sur l’ensemble du territoire. Le protocole méthodologique utilisé pour l’estimation de cette variable est détaillé dans un article de Bonnet et al. [8]. Cet indicateur dispose de l’avantage de ne pas être sensible à l’inflation et donc de rester cohérent sur une longue période. Il semble aussi utile comme indicateur de l’accès aux soins par département.

Sur ces cartes, il est possible de constater une réduction des inégalités de revenus à l’échelle nationale. Alors qu’en 1950, dans certains départements du Sud de la France, les revenus

19. Résumé disponible sur le site Internet du Figaro : [↗](#)



GRAPHIQUE 2.12 – Revenu relatif par adulte en base 100 par département français, en 1950 (à gauche) et 2015 (à droite) (source : French Regional Database)

Lecture : Une couleur foncée montre une grande disparité du revenu départemental par rapport au revenu moyen observé sur l'ensemble du territoire.

annuels pouvaient être deux fois inférieurs à la moyenne nationale, ces écarts à la moyenne ont fortement diminué depuis. En 2015, les revenus les moins importants sont observés dans les régions Nord-Pas-de-Calais, Alsace-Lorraine et Centre-Val de Loire. Remarquons qu'il s'agit des départements pour lesquels les espérances de vie à la naissance en 2015 sont aussi les moins élevées. De même, la région Île-de-France dispose des revenus les plus élevés et des taux de mortalité les plus faibles. Il semble donc exister une corrélation entre la mortalité et le niveau de richesse, visible à la fois à l'échelle mondiale et nationale. Ces données de richesse semblent permettre de capter la tendance de la mortalité ainsi que la diminution des inégalités. Il s'agit donc de variables potentiellement intéressantes dans une étude analytique et prospective.

2.3.3 Inclusion de variables socio-économiques dans les modèles de mortalité

Cette partie cherche à discuter des manières dont l'évolution de la mortalité est considérée dans les modèles usuels. Bien que les variables économiques semblent utiles afin de comprendre la tendance de la mortalité, elles ne sont que partiellement utilisées dans la littérature pour modéliser l'évolution des décès par zone géographique. Généralement, la modélisation de la tendance s'effectue en deux temps :

- Des *splines* en fonction du temps sont utilisées afin de tenir compte simultanément de la tendance et de la saisonnalité des décès ;
- Une méta-régression est lancée afin de constater la contribution de variables tierces, et notamment socio-économiques, à l'hétérogénéité géographique de la relation température-mortalité.

Cette méthode a pour objectif de déterminer de quelle manière des facteurs socio-économiques peuvent jouer sur le lien entre température et mortalité. Elle a ainsi été appliquée par Gasparrini et al. [30], à une échelle géographique très fine au Royaume-Uni (les LSOA, ou *Lower Super Output Areas*). À titre de comparaison, cette granularité est proche de la maille cantonale en France. Cette étude a permis de souligner l'existence d'une importante hétérogénéité géographique de la mortalité, y compris à petite échelle. En particulier, la surmortalité liée aux faibles températures (définies comme étant inférieures à la TMM) serait croissante avec l'indice de pauvreté retenu. Cet indice est construit à partir de données agrégées sur le revenu, l'emploi, l'éducation, la santé et l'accès au logement. Par exemple, il serait intuitif que les individus ayant moins accès au chauffage en période de froid et/ou à la climatisation lors des pics de chaleur, soient davantage exposés en terme de mortalité.

Cependant, l'usage de *splines* pour la modélisation de la tendance est contestable, car elle ne s'appuie sur aucune variable observée. Si le degré du polynôme associé est trop important, ce choix peut conduire à des prédictions fortement biaisées et inexploitable. Il sera donc important de décomposer la période d'étude en une phase d'entraînement et de test, afin de vérifier que la relation prédite reste cohérente et applicable dans le cadre d'une étude prospective.

Conclusion du chapitre

Ce chapitre présente tout d'abord les bases de données utilisées. Il a été décidé d'utiliser une base de données publique disposant de l'information sur la mortalité par jour et à la maille code postal. Une agrégation est ensuite effectuée par département afin de conserver un bon compromis entre la taille du portefeuille et la précision de la donnée climatique associée à la mortalité. L'utilité des variables climatiques et économiques qui seront mobilisées par la suite a ensuite été motivée à travers des statistiques descriptives. Les modèles principalement utilisés dans la littérature seront présentés puis comparés dans la suite de cette étude.

3

Modélisation segmentée de la mortalité à partir de variables climatiques

L'objectif de ce chapitre est de modéliser la relation quotidienne reliant la température et la mortalité entre 1990 et 2019. Un état de l'art relatif aux modélisations courantes est rappelé dans la section 3.1. Dans la section 3.2, les données issues de l'ensemble des départements de France métropolitaine sont considérées, et les modèles usuels sont appliqués. Cette approche permet de sélectionner le modèle de référence pour la suite. Des pistes d'amélioration sont ensuite proposées dans la section 3.3. Enfin, une segmentation par âge, sexe et zone climatique est motivée dans la section 3.4, dans le but de tenir compte des relations différenciées entre la température et la mortalité.

3.1 État de l'art de la modélisation du lien température-mortalité

Une comparaison synthétique des modélisations les plus courantes de la relation température-mortalité a été proposée par Folkerts et al. [24] et appliquée à des données de mortalité des Pays-Bas. Trois modèles sont présentés et décrits ci-dessous.

Le modèle SEG (Segmented Poisson Regression)

Les modèles linéaires généralisés (GLM) constituent un cadre relativement simple et souple pour la modélisation de nombreuses variables d'intérêt en actuariat (telles que la fréquence et la sévérité des accidents en assurance automobile). Pour rappel, leur objectif est de pouvoir quantifier la relation entre une variable d'intérêt Y et des variables ex-

plicatives X_1, \dots, X_k à partir d'un échantillon indépendant et identiquement distribué $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$ de taille $n \in \mathbb{N}^*$ et de même loi que (Y, \mathbf{X}) . La modélisation la plus simple consiste à déterminer la combinaison linéaire des variables explicatives permettant d'expliquer au mieux la variable d'intérêt (selon un critère de précision à déterminer) : c'est l'approche des modèles linéaires *gaussiens*. On suppose pour cela qu'il existe des coefficients $\alpha_0, \alpha_1, \dots, \alpha_k$ tels que :

$$y = \alpha_0 + \sum_{j=1}^k \alpha_j x_j,$$

où y et x_1, \dots, x_k sont les valeurs observées associées aux variables aléatoires Y et X_1, \dots, X_k .

Dans le cas du nombre de décès, la variable aléatoire Y est à valeurs entières positives, et donc ne peut pas être supposée gaussienne. Une manière simple de résoudre ce problème consiste à appliquer une transformation adaptée aux éléments de cette relation. Par exemple, en supposant que Y suit une loi de Poisson de paramètre μ , on a :

$$\ln(\mu) = \alpha_0 + \sum_{j=1}^k \alpha_j x_j,$$

ce qui permet d'obtenir des prédictions de la variable d'intérêt à valeurs positives et cohérentes avec leur définition. Le cadre des GLM peut facilement être étendu pour modéliser une série temporelle : la taille n ne représente alors plus un nombre d'individus (par exemple le nombre d'assurés) mais le nombre de périodes observées. Par ailleurs, en notant Z la variable correspondant à la température moyenne observée, il est possible de créer des variables indicatrices $X_1 = \mathbf{1}(Z \leq \psi)$ et $X_2 = \mathbf{1}(Z > \psi)$ indiquant si la température a dépassé ou non un seuil ψ préalablement fixé. Un GLM est alors estimé par morceaux, selon que la variable explicative Z soit supérieure (ou non) à ψ . Il s'agit du cadre de la régression de Poisson segmentée. Ce modèle ne permet toutefois pas de prendre en compte l'impact retardé des températures sur la mortalité observé empiriquement, notamment durant les périodes de froid.

Le modèle CSDL (Constrained Segmented Distributed Lag)

La modélisation précédente peut être complétée avec l'inclusion de variables de température retardées. Muggeo [45] a ainsi proposé une modélisation temporelle de la mortalité par morceaux, en-dessous et au-delà de seuils, et tenant compte de variables de température des périodes précédentes. C'est cette approche qui a été retenue par Pincemin [47] dans son mémoire. Les variables introduites ici seront utilisées dans le reste de l'étude. Soient :

- Y_t le nombre de décès à la date t
- z_t la température à la date t
- \mathbf{x}_t^\top le vecteur de variables explicatives supplémentaires à la date t (par exemple, le jour de la semaine ou bien une indicatrice de phénomènes saisonniers tels que la grippe ou la Covid-19).

Il est supposé que Y_t suit une loi de Poisson de paramètre μ_t . Muggeo propose la relation suivante, reprise par Pincemin dans son mémoire :

$$\ln(\mu_t) = \alpha_0 + \sum_{j=1}^k \alpha_j x_{t,j} + \sum_{l_1=0}^{L_1} \beta_{1,l_1} (z_{t-l_1} - \psi_1)^- + \sum_{l_2=0}^{L_2} \beta_{2,l_2} (z_{t-l_2} - \psi_2)^+.$$

ψ_1 et ψ_2 sont des seuils permettant de distinguer un pic de froid et de chaleur, et sont à déterminer. Le modèle est appelé CSDL (*Constrained Segmented Distributed Lag*) car les seuils ψ_1 et ψ_2 sont contraints à rester identiques quel que soit le nombre de retards considéré. Pour simplifier, il est possible de supposer $\psi_1 = \psi_2 := \psi$, ce qui permet de modéliser une relation en-deçà et au-delà de la TMM. De même, les retards maximaux L_1 et L_2 associés respectivement à l'impact retardé des pics de froid et de chaleur, doivent être spécifiés par l'utilisateur. On rappelle par ailleurs les notations $(x - a)^+ = (x - a)\mathbb{1}(x \geq a)$ et $(x - a)^- = -(x - a)\mathbb{1}(x \leq a)$, pour deux nombres réels x et a . Notons que la forme de la relation proposée est similaire entre le froid et la chaleur et ne varie que par le biais des facteurs $\beta_{\cdot,\cdot}$, ce qui n'est pas entièrement cohérent avec la relation différenciée qui est évoquée dans la littérature et apparue dans les graphiques précédents.

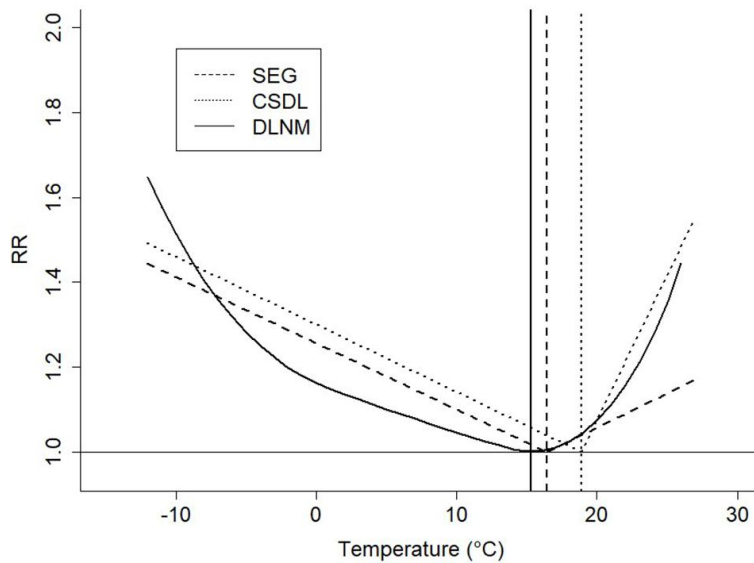
Le modèle DLNM (Distributed Lag Non-Linear Model)

Ce dernier écueil a fait l'objet d'une proposition de correction de Gasparrini et al. [29]. Plus précisément, les auteurs ont cherché à relâcher l'hypothèse sous-jacente d'une relation en "V" entre température et mortalité. Celle-ci est directement liée au choix de considérer une relation linéaire par morceaux, avec un seuil de température pour établir ces derniers.

Dans le modèle DLNM (*Distributed Lag Non-Linear Model*) qui est proposé, la mortalité est liée à la température via la relation :

$$\ln(\mu_t) = \alpha_0 + \sum_{j=1}^k s_j(z_{t,j}, \alpha_j) + \sum_{p=1}^q \gamma_p u_{t,p}.$$

La principale nouveauté de ce modèle par rapport aux précédents est l'utilisation de fonctions polynomiales par morceaux (*splines*) s_j plutôt que linéaires, afin de représenter la relation retardée entre la température et la mortalité. Les variables externes observées à l'instant t sont notées $u_{t,p}$ avec $p = 1, \dots, q$. Si le modèle proposé permet de capter des relations plus cohérentes avec ce qui est observé empiriquement, le choix d'ajouter des *splines* a pour conséquence directe de complexifier le modèle. Le nombre de degrés et de morceaux doit par ailleurs être précisé par l'utilisateur, ce qui nécessite un ajustement de paramètres pouvant être long. Une comparaison des relations entre température et mortalité est représentée dans le graphique 3.1.



GRAPHIQUE 3.1 – Relations température-mortalité selon les modèles SEG, CSDL et DLNM (source : Folkerts et al. [24])

Lecture : d'après le modèle DLNM, la mortalité estimée est 60% plus importante avec une température de -10°C par rapport à l'optimum thermique.

Au-delà des différences dans la forme des relations obtenues, il est possible de constater que plusieurs TMM sont obtenues aux Pays-Bas. Le risque relatif de mortalité (RR) associé à une température z représente le ratio entre la mortalité estimée pour la température z et celle pour la température minimisant la mortalité (TMM). Ce risque relatif modélisé est plus élevé dans le modèle DLNM par rapport aux autres modèles SEG et CSDL. Le relâchement de l'hypothèse de linéarité dans le modèle DLNM est une explication potentielle de cette observation.

3.2 Comparaison des modèles

Cette partie vise à sélectionner le meilleur modèle permettant de capter à la fois la tendance, la saisonnalité et le rôle des températures dans la mortalité. Les trois grandes familles de modèles présentées dans le chapitre précédent ont été implémentées. La démarche est progressive : après avoir considéré des modèles simples, des variables de température retardées (c'est-à-dire des jours précédents) sont introduites.

3.2.1 Cadre de la recherche

Dans un premier temps, le cadre paramétrique de cette étude est présenté. L'hypothèse de loi utilisée ainsi que la métrique de précision associée sont justifiées. Enfin, il est expliqué la manière dont les différents modèles seront comparés.

Recherche de la loi

Étant donné la nature de la variable d'intérêt, une loi de comptage est choisie pour modéliser le nombre de décès. Le choix s'est porté dans un premier temps sur une loi de Poisson : celle-ci présente l'avantage de permettre une estimation rapide ainsi que des interprétations simples. En effet, une fois le modèle estimé, la variable prédite \hat{y} est de la forme

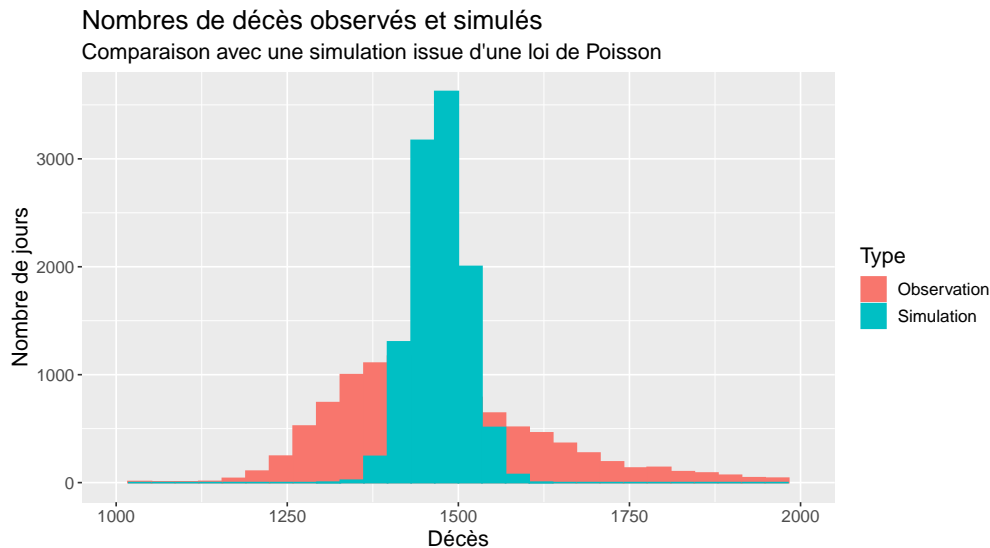
$$\hat{y} = \exp \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k \right),$$

avec x_1, \dots, x_k les valeurs connues des variables explicatives et $\hat{\beta}_0, \dots, \hat{\beta}_k$ les coefficients prédits associés aux variables explicatives. Ce cadre paramétrique permet de comparer facilement les estimations. En effet, si la température est la seule variable explicative, il est possible d'écrire la mortalité estimée pour une température z comme un coefficient multiplicateur de la mortalité minimale (associée à la TMM). Cet exemple s'étend facilement dans le cas de plusieurs variables explicatives, en raisonnant *toutes choses égales par ailleurs*¹. En particulier, il est possible de montrer que $100 \times \left[\exp \left(\hat{\beta}_j \right) - 1 \right]$ correspond au pourcentage d'augmentation prédit pour la mortalité en cas d'augmentation de 1 point de la variable explicative j .

Cependant, si le cadre de la loi de Poisson est relativement simple, cette simplicité peut également constituer un inconvénient. En effet, supposer que les nombres de décès sont tirés

1. Les autres variables explicatives sont supposées ne pas varier.

selon une loi de Poisson revient implicitement à supposer que sa moyenne μ et sa variance σ^2 sont égales. Cette hypothèse est forte : il est en pratique possible d'observer de la *surdispersion* (si $\mu < \sigma^2$) ou de la *sous-dispersion* (si $\mu > \sigma^2$). D'après Hilbe [36], une surdispersion peut survenir s'il existe des *outliers* pour la variable à prédire, ou si les observations ne sont pas indépendantes. D'une part, pour obtenir une intuition, il est possible de représenter sur un même histogramme les nombres de décès observés ainsi que ceux que l'on aurait simulés selon une loi de Poisson. Cet histogramme est représenté dans le graphique 3.2. Le paramètre choisi est naturellement la moyenne des décès observés.



GRAPHIQUE 3.2 – Histogramme des nombres totaux de décès par jour en France métropolitaine entre 1990 et 2019 : données observées et simulées avec une loi de Poisson

D'après le graphique, il est possible de suspecter l'existence d'une surdispersion des données. En effet, la simulation des décès selon une loi de Poisson sous-estime leur volatilité. Un risque, en se basant sur une loi de Poisson, serait alors d'obtenir des intervalles de confiance des nombres de décès incohérents, car ne variant pas assez d'un jour à l'autre par rapport à la réalité. Par ailleurs, l'incertitude quant aux coefficients de régression pourrait être sous-estimée. Cela empêcherait une analyse cohérente de la significativité des différents effets inclus dans le modèle. Il a donc été décidé d'utiliser un test de surdispersion des données. Un test courant a été proposé par Cameron et al. [9] puis implémenté dans le package AER de R. Après avoir estimé un modèle linéaire généralisé, ce test revient à vérifier la significativité d'un coefficient multiplicatif c tel que :

$$\text{Var}(Y) = \mu + c \times f(\mu),$$

avec f une fonction spécifiée par l'utilisateur. La construction de ce test est détaillée en annexe F. Sous l'hypothèse nulle (H_0), le coefficient c est égal à 0, ce qui revient à supposer une équidispersion. Si l'hypothèse est rejetée (à un risque de première espèce donné), alors des alternatives à la loi de Poisson sont à considérer.

Généralement, deux possibilités sont retenues : une loi dite de "quasi-Poisson" et la loi binomiale négative. La première solution est la plus simple en termes de temps de calcul : l'estimation des coefficients est la même, mais les intervalles de confiance sont modifiés en tenant compte de la surdispersion empirique. La significativité des variables explicatives pourra donc être discutée de manière cohérente, ce qui s'avère nécessaire dans une analyse biométrique. Une alternative est de considérer une loi binomiale négative, mais davantage de paramètres sont à estimer, ce qui ralentirait l'estimation du modèle. Cette dernière solution n'a donc pas été retenue : seules les lois de Poisson et de quasi-Poisson seront considérées et comparées par la suite.

Recherche de la métrique de précision

Afin de comparer les différents modèles, une métrique de précision doit être choisie. Celle-ci permet de savoir dans quelle mesure les prédictions obtenues par un modèle sont proches des valeurs observées de la variable d'intérêt. La métrique doit être adaptée au cadre paramétrique retenu. L'une des plus connues est l'erreur quadratique moyenne (ou *mean-squared error*, MSE). Étant donné deux vecteurs $y = (y_1, \dots, y_n)$ et $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$ correspondant respectivement aux valeurs observées et prédites, l'erreur quadratique moyenne est le réel positif défini par :

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

La MSE est la métrique de perte utilisée dans les modèles linéaires gaussiens, pour lesquels les coefficients de régression retenus sont solutions du problème d'optimisation :

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k \in \mathbb{R}} \sum_{i=1}^n \left(y_i - \left[\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_k x_{k,i} \right] \right)^2.$$

Ne pas retenir un modèle linéaire pour prédire la fréquence de sinistres déclarés conduit à revoir la fonction objectif à minimiser. En effet, le critère de la MSE, en plus d'être lié à l'hypothèse de normalité des observations, possède comme inconvénient de supporter des prédictions négatives. Celles-ci s'avèrent incohérentes ici au vu de la nature de la variable d'intérêt (le nombre de décès). Des métriques plus appropriées sont donc recherchées, construites

à partir de l'hypothèse faite précédemment que les observations y_i sont tirées d'une loi de Poisson. C'est notamment le cas de la déviance, qui s'appuie sur la vraisemblance \mathcal{L} du modèle (qui elle-même utilise la loi des observations). Pour définir la qualité du modèle, la vraisemblance obtenue avec un modèle qui reproduit exactement les observations constituera la référence. Celle-ci est comparée à la vraisemblance obtenue avec le modèle prédit. La déviance se définit alors par le ratio $\mathcal{L}(y|y)/\mathcal{L}(\hat{y}|y)$. Ici, sa transformation logarithmique

$$D = \ln(\mathcal{L}(y|y)) - \ln(\mathcal{L}(\hat{y}|y)).$$

sera utilisée. Plus un modèle parviendra à reproduire fidèlement les observations, et plus la déviance sera petite, signe d'un modèle de bonne qualité. En particulier, en supposant que les observations sont tirées selon une loi de Poisson, la déviance moyenne de Poisson (ou *mean Poisson deviance*, MPD) peut être obtenue :

$$\text{MPD} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i - y_i \ln(\hat{y}_i/y_i)).$$

Par rapport à la MSE, la présence d'un facteur logarithmique permet de disqualifier automatiquement les prédictions négatives pouvant être obtenues avec un modèle linéaire. La MPD sera donc retenue dans la suite afin de comparer les différents modèles dans cette section.

Ceux-ci seront comparés sur la base d'un jeu d'entraînement (1990-2013) et de test (2014-2019), représentant respectivement 80% et 20% de l'ensemble des données. Afin de sélectionner un modèle de référence, l'ensemble des départements sera considéré dans un premier temps. Un modèle "zéro" sera également implémenté à titre de comparaison. Celui-ci ne contient qu'un paramètre de tendance et une constante comme variables explicatives, tandis que la population vivante de l'année est intégrée en exposition afin de permettre la traduction des résultats en taux de mortalité.

3.2.2 Le modèle SEG (Segmented Poisson Regression)

Le chapitre précédent a permis de mettre en avant l'existence d'une relation différenciée entre la chaleur d'une part, et le froid d'autre part. Alors que les effets d'une température basse sont décrits comme linéaires mais persistants, de fortes températures peuvent se matérialiser par une surmortalité plus rapide. Afin de constater l'apport de cette segmentation, deux régressions de quasi-Poisson sont implémentées dans un premier temps, avec :

- a) une relation log-linéaire entre température et mortalité ;

b) une relation log-segmentée entre température et mortalité.

Modèle sans segmentation de la variable température

L'avantage de la première méthode est surtout computationnel : il n'y a pas de recherche de seuil, ce qui conduira à une estimation rapide des paramètres. Soit $\hat{\mu}_t$ le taux de mortalité estimé à l'instant t , où $t \in \{1, \dots, 8766\}$ est le nombre de jours depuis le 1^{er} janvier 1990. La température moyenne journalière à l'instant t est notée z_t . Le modèle obtenu sans relation segmentée est de la forme :

$$\hat{\mu}_t = \exp[-1,05 - (1,085 \times 10^{-5})t - (7,506 \times 10^{-3})z_t].$$

Ce modèle prédit donc une tendance décroissante mais ralentie de la mortalité, tout comme de la relation entre température et mortalité. Il capture bien le phénomène de plus grande mortalité qui est observé au cours des périodes de froid. Cependant, il indique que la mortalité diminue de façon continue à mesure que la température augmente. En cas de température proche de l'infini, le modèle prédirait *toutes choses égales par ailleurs* une mortalité nulle, ce qui est absurde compte tenu des conséquences historiques des vagues de chaleur. Il est donc nécessaire de privilégier une relation segmentée afin de capter la part de mortalité attribuable à la chaleur, en plus de celle associée au froid.

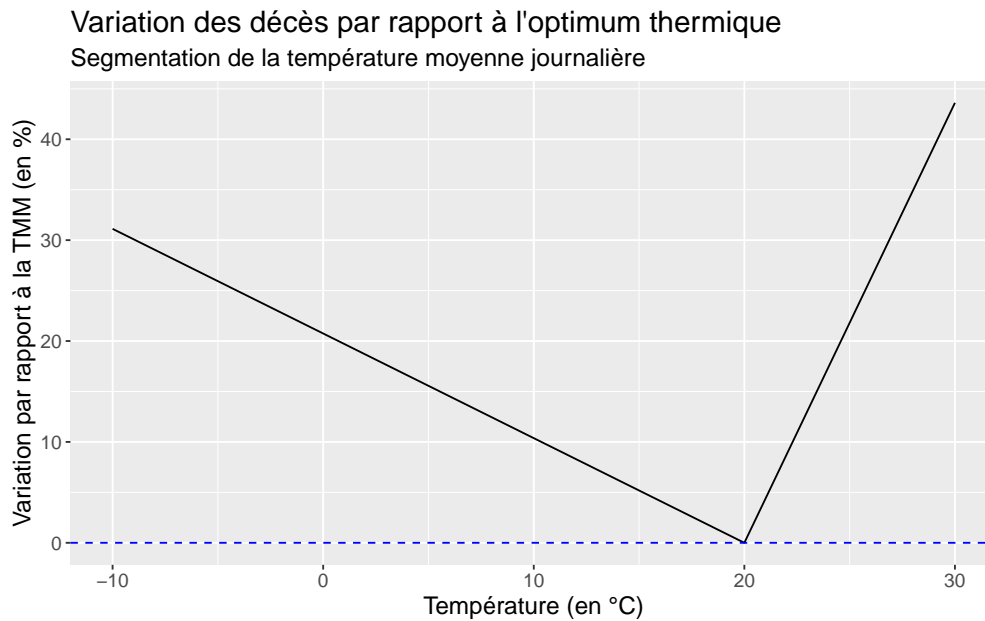
Ce modèle simple permet aussi de tester une éventuelle surdispersion des données. Le test de Cameron et al. [9] estime le coefficient de dispersion à 1,75 et rejette au risque 5% l'hypothèse nulle d'équidispersion. Une loi de quasi-Poisson sera donc utilisée par la suite.

Modèle avec segmentation de la variable température

Le package `segmented`² de R permet de rechercher le seuil de température permettant de séparer au mieux les impacts du froid et de la chaleur. Après avoir estimé un modèle de base (ici, un GLM de quasi-Poisson supposant une relation log-linéaire), une valeur initiale pour le seuil de segmentation est spécifiée par l'utilisateur. Pour déterminer celle-ci, il est par exemple possible de reprendre les optimums thermiques déterminés dans la littérature. Par exemple, Pascal et al. [46] ont proposé une démarche pour déterminer les TMM (températures minimisant la mortalité) pour 18 grandes villes françaises. Les valeurs obtenues, variant dans

2. Détails sur le package R : [↗](#)

le temps et l'espace, sont comprises entre 15 et 27 °C. Dans un modèle estimé sur l'ensemble de la France métropolitaine, une valeur initiale de 20°C semble donc cohérente.



GRAPHIQUE 3.3 – Pourcentage d'évolution de la mortalité par rapport à l'optimum thermique en France métropolitaine - Prédiction obtenues avec un modèle SEG

L'optimum thermique obtenu est finalement de 20°C, confirmant la première intuition. Selon le modèle segmenté, la variation de mortalité liée au froid est plus faible que celle liée à la chaleur, mais est visible pour davantage de températures. En effet, une température moyenne journalière de 20°C correspond au quantile d'ordre 90% du jeu d'entraînement, signifiant qu'elle n'est dépassée que 10% du temps. Ce constat est cohérent avec les faits empiriques mentionnés dans le chapitre précédent. Par ailleurs, la comparaison des déviations de Poisson indique une meilleure adéquation du modèle avec segmentation, tant sur les données d'entraînement que sur celles de test. La comparaison des résultats est disponible dans le tableau 3.1.

<i>Modèle</i>	<i>MPD (entraînement)</i>	<i>MPD (test)</i>	<i>Temps de calcul (en s)</i>
Modèle zéro	18,445	17,434	2,934
Relation log-linéaire	0,873	1,028	12,105
Relation log-segmentée	0,863	1,021	1598,624

TABLEAU 3.1 – Comparatif des performances des modèles avec et sans segmentation de la température moyenne journalière

Lecture : les écarts moyens entre les valeurs observées et prédites par chaque modèle, mesurés par la déviance moyenne de Poisson (MPD), sont minimisés avec une relation log-segmentée entre température et mortalité.

Bien que la méthode de segmentation implémentée dans le package `segmented` soit particulièrement lourde en temps de calcul, il est également possible de l'implémenter à la main par le biais d'un algorithme de *gridsearch*.

- (i) Un vecteur de seuils ψ_1, \dots, ψ_p candidats est d'abord renseigné;
- (ii) Pour un seuil ψ donné, deux variables explicatives $(z_t - \psi)^+$ et $(z_t - \psi)^-$ sont ajoutées au modèle;
- (iii) Un GLM de quasi-Poisson est estimé, et le seuil ψ optimisant une certaine métrique de précision (ici la MPD) par validation croisée est conservé.

Il est également possible de spécifier l'existence de deux seuils, ce qui reviendrait à estimer une relation en "U" linéaire par morceaux entre température et mortalité. Cependant, ce choix complexifie l'algorithme tout en rendant plus délicates les interprétations. Par exemple, la TMM obtenue n'est plus forcément unique, ce qui rendrait alors plus délicate la vérification d'une éventuelle acclimatation. Le modèle SEG doit toutefois être amélioré. En effet, il ne prend pas en considération les relations retardées pouvant exister entre la température et la mortalité, notamment lors des périodes de basses températures. Une extension du modèle segmenté est alors proposée : le modèle CSDL.

3.2.3 Le modèle CSDL (Constrained Segmented Distributed Lag)

Le modèle CSDL étend le modèle segmenté présenté précédemment en incluant des variables de retard. Pour rappel, ce modèle suppose l'existence de seuils ψ_1 et ψ_2 tels que :

$$\ln(\mu_t) = \alpha_0 + \sum_{j=1}^k \alpha_j x_{t,j} + \sum_{l_1=0}^{L_1} \beta_{1,l_1} (z_{t-l_1} - \psi_1)^- + \sum_{l_2=0}^{L_2} \beta_{2,l_2} (z_{t-l_2} - \psi_2)^+,$$

où l'on fixe généralement $\psi_1 = \psi_2 := \psi$ ainsi que $L_1 = L_2 := L$ pour simplifier la relation obtenue. L'enjeu est alors de déterminer à la fois le seuil ψ et le nombre de retards L à inclure dans le modèle.

De la même manière que précédemment, une démarche par *gridsearch* est proposée :

- (i) Un vecteur de seuils ψ_1, \dots, ψ_p et de retards maximaux L_1, \dots, L_q candidats est d'abord renseigné;
- (ii) Pour chaque retard L_i , et pour un seuil ψ donné, les variables explicatives $(z_t - \psi)^+$, $(z_t - \psi)^-$ (retard 0) jusqu'à $(z_{t-L_i} - \psi)^+$, $(z_{t-L_i} - \psi)^-$ (retard L_i) sont ajoutées au modèle. Un GLM de quasi-Poisson est estimé, et le seuil ψ optimisant une certaine métrique

de précision (ici la MPD) par validation croisée est conservé. Le modèle optimal est retenu.

- (iii) Les modèles optimaux associés à chaque retard L_1, \dots, L_q sont comparés sur la base de l'AIC.

Il est à noter que deux procédés de sélection de modèle sont retenus. D'une part, à retard fixé, afin de comparer les seuils candidats, une validation croisée est lancée. L'objectif est ici de vérifier le potentiel prédictif du modèle sur de nouvelles données. Le jeu d'entraînement est divisé en cinq sous-jeux, puis un modèle est estimé sur 4/5 des données. La MPD est ensuite calculée sur le 1/5 de données restantes. Ce procédé est répété sur les cinq combinaisons possibles afin d'obtenir une erreur "moyenne", à minimiser.

D'autre part, afin de comparer les modèles optimaux associés à chaque retard L_1, \dots, L_q , un calcul d'AIC (*Akaike Information Criteria*) est effectué. Cet indicateur permet d'évaluer un compromis entre la précision et la complexité d'un modèle. Par exemple, la comparaison des AIC permet d'indiquer dans quelle mesure le gain de pouvoir prédictif associé à l'ajout d'un retard supplémentaire compense l'augmentation de la complexité du modèle (sachant que deux paramètres supplémentaires sont ajoutés par retard). Pour rappel, en notant k le nombre de paramètres à estimer dans le modèle, et \mathcal{L} sa vraisemblance, l'AIC est défini par :

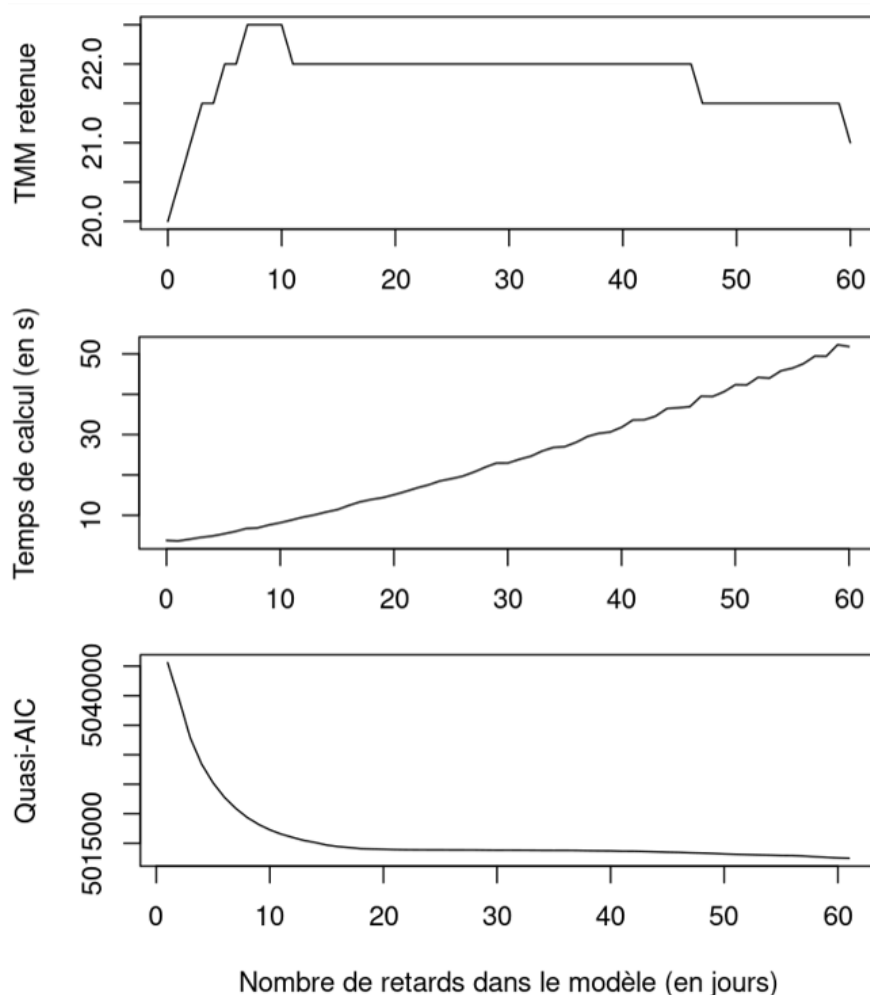
$$AIC = 2k - 2 \ln(\mathcal{L}).$$

Une variante de cette métrique, appelée quasi-AIC, est parfois utilisée afin de tenir compte de la surdispersion potentielle du modèle. Elle s'avère donc adaptée pour les données étudiées ici. Le quasi-AIC est défini par :

$$qAIC = 2k - 2 \ln(\mathcal{L})/\phi,$$

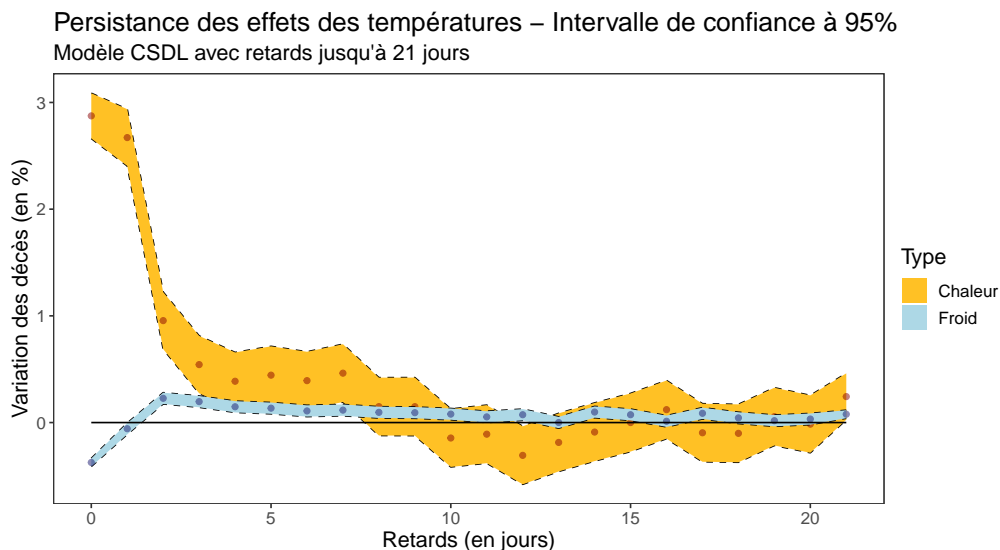
où ϕ est la surdispersion estimée par le modèle. Dans le graphique 3.4, les optimums thermiques, les temps de calcul et le quasi-AIC associés à chaque retard sont représentés. Le nombre de retards maximal est fixé à 60 dans le but de comparer les différents modèles dans un délai raisonnable.

Tout d'abord, de manière assez intuitive, il est possible de constater une hausse puis une stabilisation de la TMM à mesure que le nombre de retards supplémentaires augmente. En effet, l'ajout de retards, jusqu'à un certain point, permet de capter davantage de décès attribuables au froid. De ce fait, en cumulant l'ensemble des retards, une même température froide sera considérée comme plus mortelle, et l'optimum thermique va augmenter. Ensuite,



GRAPHIQUE 3.4 – Comparaison des TMM optimaux (haut), du temps de calcul (milieu) et du quasi-AIC (bas) par modèle CSDL implémenté

la comparaison des AIC permet de mettre en avant un gain de pouvoir prédictif significatif, plus important que le coût en complexité du modèle, jusqu'au vingtième retard environ. Ce constat est cohérent avec la valeur de 21 jours souvent retenue dans la littérature pour capter la persistance des effets du froid. Au-delà, le quasi-AIC continue à diminuer légèrement, signe que les températures retardées continuent à apporter une information utile au modèle. Toutefois, entre les retards 60 et 20, le temps de calcul associé est quasiment multiplié par 3. Compte tenu des complexifications et segmentations futures (par âge notamment) qui seront à effectuer par la suite, ce coût est particulièrement regrettable. Afin de comparer les atouts de chaque modèle et de comprendre la manière dont se manifestent les effets retardés, les courbes des effets du froid (les $\beta_{1,\cdot}$) et de la chaleur (les $\beta_{2,\cdot}$) sont représentées. le graphique 3.5 représente les intervalles de confiance à 95% associés à la persistance des effets de la chaleur et du froid pour un modèle incluant jusqu'à 21 retards.

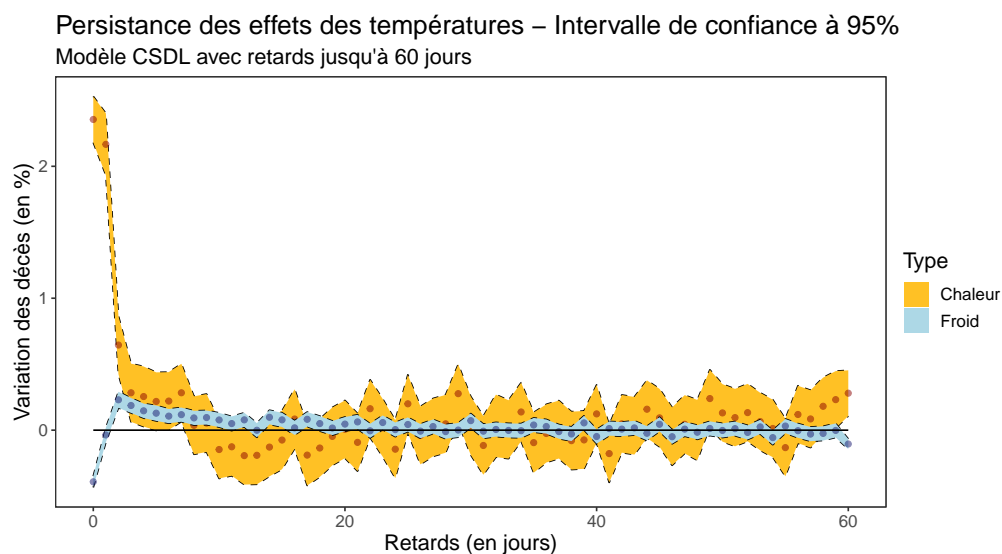


GRAPHIQUE 3.5 – Intervalles de confiance à 95% associés aux effets de la chaleur (à gauche) et du froid (à droite) dans un modèle CSDL à 21 retards

Lecture : *toutes choses égales par ailleurs*, si la température moyenne journalière augmente de 1°C par rapport à la TMM (22°C), alors la mortalité prédite attribuable à la chaleur augmentera d'environ 3%.

Plusieurs enseignements peuvent être tirés de ce graphique. Tout d'abord, les variations des décès liées à la chaleur seraient plus importantes que celles associées au froid d'après cette modélisation. Cela confirme l'interprétation du modèle précédent ainsi que la plus grande mortalité liée à une température très élevée, en comparaison d'un pic de froid. Pour rappel, une température "froide" sera définie ici comme étant inférieure à la TMM associée, c'est-à-dire 22°C. Par ailleurs, le modèle prédit l'existence d'un effet "moisson". Il s'agit d'un épisode de sous-mortalité pouvant survenir après un excédent de mortalité, à la suite d'un évènement tel qu'une vague de froid, de chaleur, ou bien une épidémie. Une explication peut être avancée. Sachant que les personnes les plus fragiles (par exemple, atteintes de maladies graves) sont les plus susceptibles d'être touchées par des épisodes de températures extrêmes, leur date de décès risque d'être avancée de quelques jours ou semaines. Ceci se traduit sur les statistiques de décès par un pic immédiat, puis une sous-mortalité dans les jours et semaines suivantes. D'après le modèle, ce rattrapage pourrait survenir une douzaine de jours à la suite d'une température élevée. Cette observation est légèrement différente de celle obtenue par Pincemin [47], qui prédisait un effet moisson visible entre 5 et 25 jours après la survenance d'une journée chaude. La méthodologie était toutefois différente, sachant qu'une moyenne des températures journalières était utilisée comme variable explicative pour toute la France. Ici, les températures ainsi que la mortalité de chaque département sont incluses dans le modèle, dans le but de capter plus précisément la relation entre ces deux variables.

Enfin, le modèle prédit des effets plus persistants pour des températures froides. Contrairement aux conséquences de la chaleur, les effets mortels du froid se manifestent à partir d'un certain délai (environ deux jours) mais s'avèrent plus durables. Ce constat est cohérent avec les faits empiriques. Alors que les effets de la chaleur des jours précédents sur la mortalité s'estompent à partir d'une dizaine de jours, la variation des décès attribuable au froid reste positive et significative au 21^{ème} retard. Cela laisse suggérer que le modèle ne capte pas entièrement les effets retardés du froid sur la mortalité. Pour cette raison, un modèle utilisant des variables de températures retardées jusqu'à 60 jours est ensuite estimé. Les intervalles de confiance à 95% associés aux effets de la chaleur et du froid sont de nouveaux représentés, dans le graphique 3.6.



GRAPHIQUE 3.6 – Intervalles de confiance à 95% associés aux effets de la chaleur (à gauche) et du froid (à droite) dans un modèle CSDL à 60 retards

Lecture : *toutes choses égales par ailleurs*, si la température moyenne journalière augmente de 1°C par rapport à la TMM (21°C), alors la mortalité prédite attribuable à la chaleur augmentera d'environ 3%.

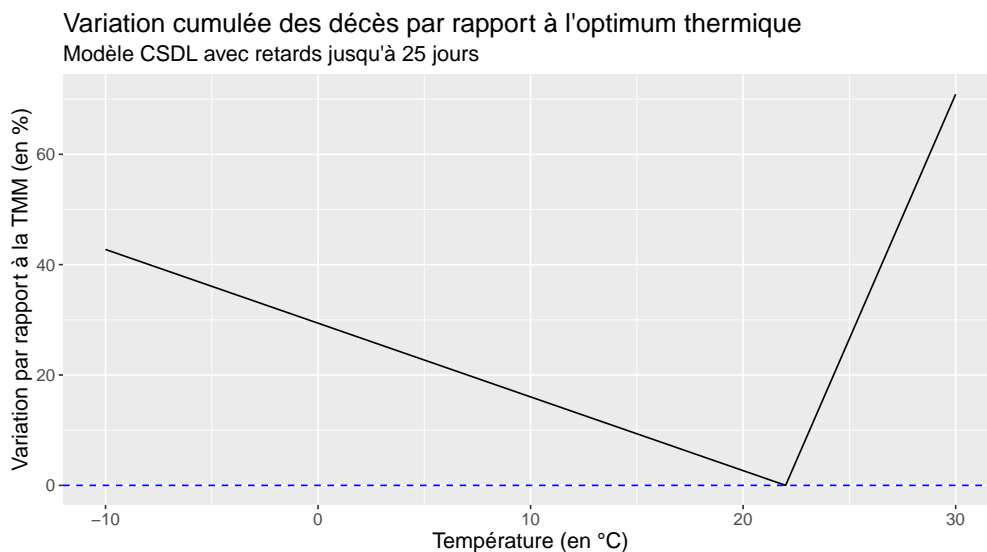
Les conclusions de ce modèle sont relativement similaires. La lecture des intervalles de confiance sur le graphique de droite indique que l'on ne peut pas rejeter l'hypothèse de nullité d'un effet du froid sur la mortalité, au-delà d'une trentaine de jours. Autrement dit, il ne semble pas utile d'inclure l'ensemble des 60 retards dans le modèle, à la fois d'un point de vue computationnel et compte tenu des interprétations biométriques. Afin de retenir le nombre de retards finalement retenu, les valeurs différenciées des quasi-AIC ont été calculées. L'objectif était de savoir à partir de quel moment l'ajout d'une variable retardée supplémentaire apportait davantage de complexité que de pouvoir prédictif au modèle. Trois nombres de retards maximaux "candidats" ont donc été obtenus : 25, 31 et 36. Finalement, 25 retards

sont conservés pour la suite, dans un souci de garder des temps de calcul raisonnables. La comparaison des performances des trois modèles est disponible dans le tableau 3.2.

Modèle	MPD (entraînement)	MPD (test)	Temps de calcul (en s)
Modèle zéro	18,445	17,434	2,934
CSDL (21 retards)	0,843	1,004	17,989
CSDL (25 retards)	0,843	1,003	21,123
CSDL (60 retards)	0,842	0,999	56,943

TABLEAU 3.2 – Comparatif des performances des modèles CSDL selon le nombre maximal de retards retenu

Il est à noter que les métriques de précision sont extrêmement proches, quel que soit le nombre maximal de retards retenu. Cela signifie que le gain de pouvoir prédictif du modèle passera par l'ajout d'autres variables explicatives et/ou une modélisation plus adaptée de la relation entre température et mortalité. Or, le modèle CSDL suppose un cadre assez restrictif : les effets de la chaleur et du froid sont supposés de la même forme (log-linéaire). Si cette hypothèse a pu être confirmée par l'absence de rupture de la courbe de mortalité pour les effets du froid, il est possible que le modèle CSDL sous-estime l'ampleur des conséquences des vagues de chaleur. Pour s'en rendre compte, le graphique 3.7 représente la variation des décès prédite pour chaque température par rapport à l'optimum thermique estimé (ici de 22°C) pour un modèle CSDL avec inclusion de 25 retards.



GRAPHIQUE 3.7 – Variation cumulée des décès prédite par un modèle CSDL à 25 retards

La variation représentée est cumulée afin de pouvoir être comparée avec les résultats du modèle SEG (qui est statique) : l'hypothèse simplificatrice sous-jacente est que la même

température est observée pendant 25 jours consécutifs. Il ressort du graphique que la variation de la mortalité par rapport à l'optimum thermique est linéaire par morceaux, et assez similaire à celle obtenue avec le modèle SEG. Pourtant, l'analyse de la température humide a mis en évidence l'existence de conditions pour lesquelles le corps humain ne pouvait supporter une exposition prolongée à une chaleur intense, si l'humidité relative est trop importante. De ce fait, il semble plus cohérent d'imaginer une évolution exponentielle de la mortalité au-delà d'une certaine température. Cela motive l'utilisation des modèles DLNM, qui relâchent l'hypothèse d'une relation log-linéaire par morceaux entre la température et la mortalité. Par la suite, l'apport de la température humide dans la modélisation sera également discuté.

3.2.4 Le modèle DLNM (Distributed Lag Non-Linear Model)

Une version classique

La construction du modèle DLNM repose sur l'implémentation de *cross-basis* : il s'agit de bases de fonctions implémentées à la fois sur la variable explicative retenue (la température) et les retards de cette variable retenus. Autrement dit, les températures retardées z_t, \dots, z_{t-L} incluses en entrée du modèle sont transformées, par une fonction³ notée $f \cdot w$, en un autre ensemble de variables explicatives, via la relation :

$$s(z_t, \dots, z_{t-L}) = \sum_{\ell=0}^L f \cdot w(z_{t-\ell}, \ell).$$

La fonction $f \cdot w$ possède deux composantes, permettant de reproduire la nature de la relation supposée, à la fois :

- a) entre la température et la mortalité (via une fonction f associée à une variable de température z);
- b) concernant la persistance de la mortalité dans le temps (via une fonction w associée à un nombre de retards ℓ).

Cette fonction est incluse parmi les variables explicatives du GLM. En cela, il est possible de voir le modèle DLNM comme une extension des modèles additifs généralisés (GAM), incluant un traitement des variables retardées. Pour rappel, les GAM permettaient de capter des relations complexes et non-linéaires (à une transformation près) entre la variable d'intérêt et une variable explicative, contrairement aux GLM. Toutefois, cette extension est faite au

3. Les notations sont reprises de la présentation de Gasparrini : ☞

prix d'une complexification de la construction du modèle, ainsi que d'un allongement du temps d'estimation du modèle. De la même manière, les DLNM peuvent davantage être vus comme des "boîtes noires" par rapport au modèle CSDL, dont la construction est naturelle. Or, l'ACPR est plutôt réticente à l'utilisation de tels algorithmes dans le monde actuariel. Dans un document de réflexion de Dupont et al. [19], le superviseur encourage ainsi à ne pas utiliser d'algorithme "boîte noire" si un modèle aussi performant mais plus interprétable existe.

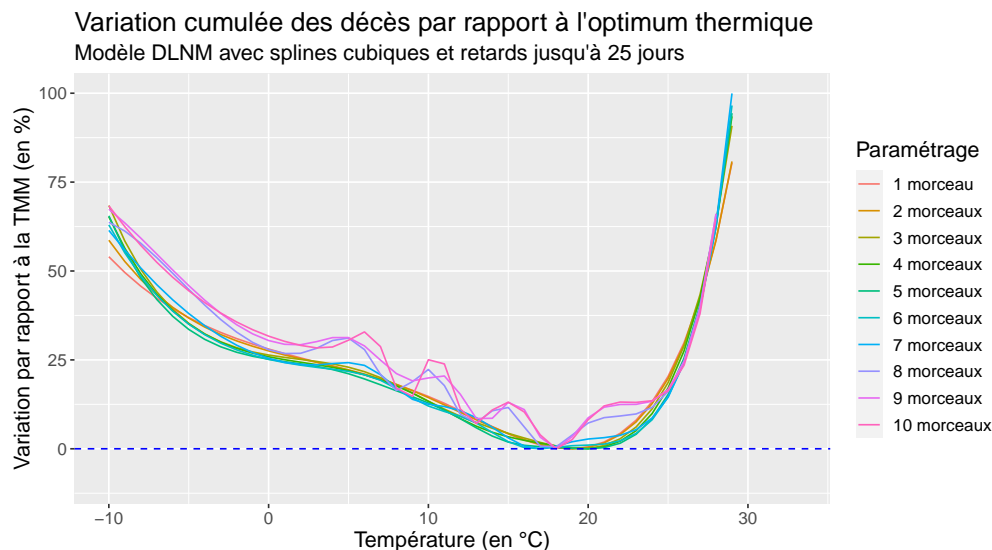
Tout comme dans un GAM, ces relations sont souvent exprimées sous forme de *splines*, c'est-à-dire de fonctions polynomiales par morceaux. L'objectif est de pouvoir tenir compte de relations différenciées entre la température et la mortalité, selon différents niveaux de température. Par défaut, pour la relation entre température et mortalité, les fonctions polynomiales par morceaux sont de degré 3 et les morceaux sont choisis à partir de quantiles équidistants des températures moyennes. Concernant la persistance de la mortalité dans le temps, Gasparrini et al. proposent⁴ l'utilisation de *splines* de degré 3 ainsi que de *log-knots* : les morceaux sont alors davantage concentrés dans les premiers retards. Par exemple, pour la modélisation des effets persistants de la chaleur, ce choix semble plus cohérent, sachant que la plus grande partie de la surmortalité ainsi que l'effet moisson se concentrent dans les premiers jours.

Afin de comparer les modèles, la validation croisée est utilisée plutôt que le critère de l'AIC. En effet, selon les paramètres qui seront spécifiés, la forme des relations entre la température et la mortalité sera amenée à évoluer. Il sera donc nécessaire de s'assurer que la forme modélisée est cohérente, en testant son adéquation sur de nouvelles données (ce qui permettra de détecter un éventuel surapprentissage, par exemple). Dans un premier temps, des *splines* cubiques sont implémentées et le nombre maximal de retards est fixé à 25 jours. C'est le nombre de morceaux, à la fois sur l'échelle des températures et celle des retards, qui va varier et être sélectionné. Une division de la base d'entraînement en 5 sous-jeux de données est utilisée pour la validation croisée, et la MPD est choisie comme métrique de précision.

Des représentations graphiques permettent de visualiser les relations obtenues en faisant varier les paramètres. Il est tout d'abord possible de représenter les modélisations obtenues entre la température et la mortalité par nombre de morceaux dans les *splines*. Celles-ci figurent dans le graphique 3.8. Le nombre de morceaux sur l'échelle des retards est alors fixé (à 2 ici).

Tout d'abord, la nature de la relation captée est relativement différente de celle prédite

4. Voir par exemple la vignette R associée au package `dlnm` : [↗](#)



GRAPHIQUE 3.8 – Variation cumulée des décès prédite par un modèle DLNM à 25 retards en faisant varier le nombre de morceaux dans les *splines* entre température et mortalité

par un modèle CSDL. Comme prévu, la courbure prédite au-delà d'une certaine température (proche de 20°C) est convexe, signe d'une augmentation accélérée de la mortalité. Pour rappel, le cadre des GLM imposait une variation seulement linéaire de la mortalité dans le modèle CSDL. Un signe de cette meilleure prise en compte des variations de la mortalité liée aux températures élevées est l'évolution de l'optimum thermique. Alors que ce dernier était égal à 21°C sur le modèle CSDL, des valeurs comprises entre 18 et 20°C sont obtenues ici. Une explication possible est que le modèle DLNM est mieux capable de retranscrire l'accélération progressive de la mortalité attribuable à la chaleur, en comparaison du modèle CSDL qui supposait une évolution linéaire.

Ensuite, l'augmentation du nombre de températures utilisées dans la modélisation conduit à prendre en considération des variations de la mortalité associées à des températures de plus en plus proches. Ainsi, alors que les modèles contenant jusqu'à 6 morceaux prédisent une relation en U allongée et relativement lisse, les modèles avec 7 morceaux ou plus prédisent des relations plus étonnantes. Par exemple, le modèle à 10 morceaux prédit une mortalité supérieure pour une température à 10°C par rapport à une température de 9°C, avec un saut visible entre ces deux valeurs. Aucune explication n'a été obtenue dans la littérature, et il est probable que des facteurs autres que la température soient en cause. Pourtant, le critère de la validation croisée indique que le nombre de morceaux optimal est égal à 8 pour la relation entre température et mortalité, et 2 pour la persistance des effets de la température. L'étude des prédictions de ce modèle sur le jeu de test d'une part, et sur des données segmentées d'autre part, permettra de confirmer ou non le choix de ce modèle. Il est en effet possible de

suspecter l'existence d'un surapprentissage au vu du constat précédent. Pour cette raison, une variante des DLNM classiques a été étudiée.

Une variante pénalisée

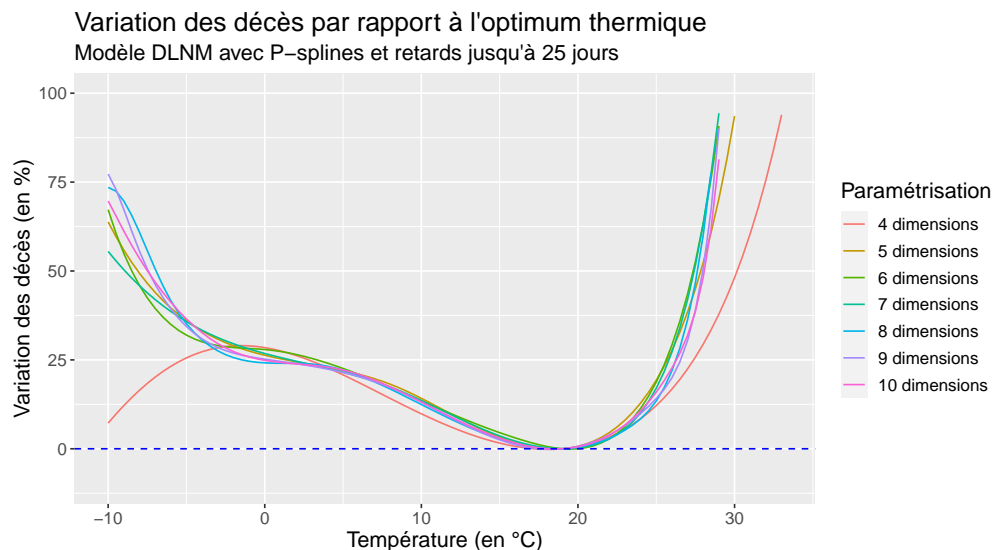
La version pénalisée des DLNM a été proposée par Gasparrini et al. [31]. Cette extension du cadre classique présenté plus haut permet d'inclure des *splines* pénalisées, afin de davantage contrôler la relation supposée entre la température et la mortalité. Concrètement, un terme de pénalisation est inclus dans le problème d'optimisation du modèle. Il s'agit d'un nombre positif, soustrait à la log-vraisemblance initiale, permettant de contrôler le lissage constaté à la fois :

- a) sur la relation entre la température et la mortalité ;
- b) sur la persistance des effets de la température sur la mortalité dans le temps.

Par exemple, une modélisation par *P-splines* conduira à inclure une matrice de pénalisation dans le programme d'optimisation de la log-vraisemblance. Cette pénalisation est basée par défaut sur les différences d'ordre 2. La dimension de cette matrice est fixée par l'utilisateur et dépend du nombre de morceaux choisi. Plus ce dernier sera élevé, et plus la pénalisation sera forte : le programme d'optimisation s'assurera que les relations entre température et mortalité seront suffisamment proches pour davantage de niveaux de températures. Afin de mieux s'en rendre compte, le graphique 3.9 représente l'évolution de la relation température-mortalité à mesure que des morceaux sont rajoutés dans la pénalisation.

Il en ressort qu'un certain niveau de lissage est nécessaire pour obtenir des relations cohérentes. Par exemple, lorsque 4 dimensions seulement sont incluses dans la matrice de pénalisation, le modèle prédit une diminution de la mortalité à mesure que la température décroît (en-dessous de 0°C). Ce résultat semble difficilement justifiable. Les résultats paraissent plus cohérents à partir de 5 dimensions. D'après la documentation du package `dlnm` fournie par Gasparrini⁵, si la relation entre la variable d'intérêt et la variable explicative est déjà suffisamment lisse, alors il n'est pas nécessaire d'augmenter le nombre de morceaux. En effet, cela ne ferait qu'augmenter inutilement le temps de calcul. Cette dernière affirmation s'est vérifiée ici : près de 5 minutes sont par exemple nécessaires pour obtenir une estimation du modèle avec *P-splines* et une matrice de pénalisation à dix dimensions. Cependant, il reste prudent de conserver un lissage suffisant en gardant à l'esprit que des segmentations seront effectuées dans la section 3.4 et conduiront à estimer les relations températures-mortalité sur

5. Lien vers la documentation du package `dlnm` : [↗](#)

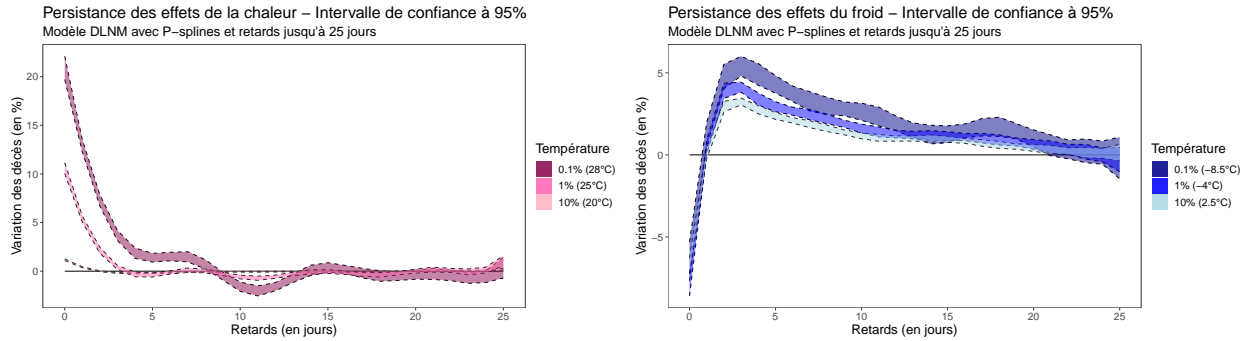


GRAPHIQUE 3.9 – Variation cumulée des décès prédite par un modèle DLNM à 25 retards selon la dimension de la matrice de pénalité dans les *splines* entre température et mortalité

moins de données. Le paramètre de pénalisation sera donc fixé à 10 par la suite, correspondant à la valeur par défaut fixée dans le package DLNM.

Contrairement au modèle CSDL, la persistance des effets de la chaleur ou du froid sur la mortalité n'est pas entièrement définie par l'écart par rapport à l'optimum thermique. En effet, la construction du modèle DLNM s'appuie sur des fonctions polynomiales par morceaux, ce qui implique que la relation température-mortalité est différenciée en plusieurs morceaux. De ce fait, la forme de la courbe de persistance des effets peut être étudiée et comparée en plusieurs points. Dans le graphique 3.10, une approche par quantiles a été choisie, afin de visualiser au mieux les impacts de températures inhabituelles. Des quantiles d'ordre 0,1%, 1% et 10% ont été choisis, à la fois pour la chaleur et le froid.

Tout comme dans le modèle CSDL, le modèle DLNM pénalisé prédit un effet plus persistant du froid mais moins important en termes de surmortalité. Cela est cohérent avec les faits empiriques. De même, l'effet moisson, conduisant à une compensation de la surmortalité immédiate, serait une nouvelle fois rencontré autour d'une dizaine de jours après la survenance d'une température chaude extrême. Cependant, comme prévu, la persistance des effets des températures est davantage différenciée dans le modèle DLNM par rapport au modèle CSDL. Ainsi, le passage à une température rencontrée 10 fois plus rarement se traduit, pour la chaleur, par une variation de la surmortalité immédiate voisine de 20% pour la chaleur. Pourtant, il n'existe pas de telle rupture dans la variation de la surmortalité pour les effets du froid. Cela est une nouvelle fois cohérent avec les affirmations de spécialistes mentionnées



GRAPHIQUE 3.10 – Intervalles de confiance à 95% associés aux effets de la chaleur (à gauche) et du froid (à droite) dans un modèle DLNM pénalisé à 25 retards

Lecture : *toutes choses égales par ailleurs*, une température moyenne journalière de 28°C se traduira par une augmentation immédiate de 20% de la mortalité par rapport à l'optimum thermique (ici de 19°C).

dans le chapitre précédent. Enfin, par construction du modèle DLNM, un lissage a été appliqué sur la persistance des effets des températures dans le temps. L'intérêt est de pouvoir obtenir une relation légèrement détachée des données d'entraînement par rapport au modèle CSDL, et davantage robuste sur de nouvelles données. De manière générale, le passage à un modèle DLNM pénalisé parvient à reproduire la majeure partie des faits empiriques retrouvés dans la littérature, et fournit donc une approche plus fiable de la modélisation du lien entre la température et la mortalité.

3.2.5 Choix du modèle

Afin de retenir le modèle utilisé pour la segmentation par la suite, trois critères ont été fixés :

- La vérification des faits stylisés.** Autrement dit, les conclusions du modèle doivent être similaires à celles de la littérature ainsi que celles observées dans les statistiques descriptives.
- Un bon compromis entre précision, interprétabilité et complexité du modèle.** L'objectif est de pouvoir expliquer à un non-initié la construction du modèle, et également de fournir des prédictions suffisamment robustes dans un temps raisonnable.
- La cohérence de la modélisation d'un point de vue prospectif.** En particulier, sachant que le changement climatique se traduira par une hausse de la fréquence et de la sévérité des pics de chaleur, ces derniers doivent être correctement modélisés. Pour cela, une étude de cas sera effectuée à partir de l'épisode de canicule de 2003.

Vérification des faits stylisés

Dans un premier temps, le modèle SEG a permis de correctement capter la relation différenciée des effets du froid et de la chaleur sur la mortalité. Toutefois, leur persistance n’ayant pas été prise en compte, il a fallu tester une généralisation de cette relation segmentée à partir de variables retardées. Ainsi, le modèle CSDL a permis de retrouver la plus grande persistance des effets du froid, en plus d’un effet moisson connu et relevé dans des travaux précédents. Toutefois, seul le modèle DLNM est parvenu à capter la non-linéarité de la relation entre température et mortalité, et plus exactement l’évolution exponentielle de la mortalité attribuable à la chaleur, une fois une certaine température dépassée. De plus, l’utilisation d’une extension pénalisée du modèle DLNM permet de s’assurer que le modèle ne surapprend pas du jeu d’entraînement et propose des relations suffisamment robustes. Cela s’est vérifiée tant sur la relation entre température et mortalité, que sur la persistance des effets de la température. Il est à noter que le modèle CSDL a fait l’objet d’une extension similaire visant à lisser la courbure de persistance des effets de la température. Celle-ci a été implémentée sur R par Muggeo dans la librairie `modTempEff`⁶. Toutefois, dans cette étude, l’utilisation de plusieurs groupes de données (les départements) est incompatible avec les fonctions de cette librairie. Il n’a donc pas été possible de tester correctement cette méthode.

Compromis entre précision, interprétabilité et complexité du modèle

Tout d’abord, les modèles présentés ici s’appuient tous sur le cadre paramétrique propre aux GLM, qui présentent de manière générale un bon compromis entre précision, interprétabilité et rapidité du temps de calcul. Cela est notamment vrai par rapport aux méthodes utilisant des réseaux de neurones. Par exemple, Guo et al. [20] ont proposé une extension des DLNM utilisant des réseaux de neurones artificiels. Toutefois, le temps de calcul jugé trop important a limité la possibilité d’affiner suffisamment le choix des paramètres. Cela rend également plus délicate la possibilité d’inclure d’autres variables explicatives potentiellement utiles (atmosphériques, par exemple). Le tableau 3.3 fournit un comparatif de l’ensemble des modèles présentés, en termes de précision (mesurée par la MPD) ainsi que de temps de calcul.

La comparaison avec le modèle zéro indique tout d’abord que l’inclusion de la température fournit une information particulièrement utile au modèle. Ensuite, chaque complexification supplémentaire apporte un gain de précision dans les prédictions, tant sur le jeu d’entraînement que celui de test. En particulier, le modèle DLNM parvient à mieux expliquer la

6. Lien vers la documentation associée au package `modTempEff` : ↗

<i>Modèle</i>	<i>MPD</i> (entraînement)	<i>MPD</i> (test)	<i>Temps de calcul</i> (en s)
Modèle zéro	18,445	17,434	2,934
Relation log-segmentée	0,863	1,021	1598,624
CSDL (25 retards)	0,843	1,003	21,123
DLNM simple (25 retards)	0,840	1,002	15,285
DLNM pénalisé (25 retards)	0,842	1	289,255

TABLEAU 3.3 – Comparatif des performances des modèles estimés

mortalité par rapport au modèle CSDL, tout en fournissant des prédictions dans un délai raisonnable. Dans un contexte prospectif, il paraît davantage adapté puisqu'il capte mieux les effets des vagues de chaleur sur la mortalité. Comme évoqué dans le chapitre précédent, celles-ci sont amenées à se multiplier dans les prochaines décennies. La pénalisation des modèles permet quant à elle d'obtenir des relations plus lisses et cohérentes, même si elle conduit à un allongement considérable du temps de calcul.

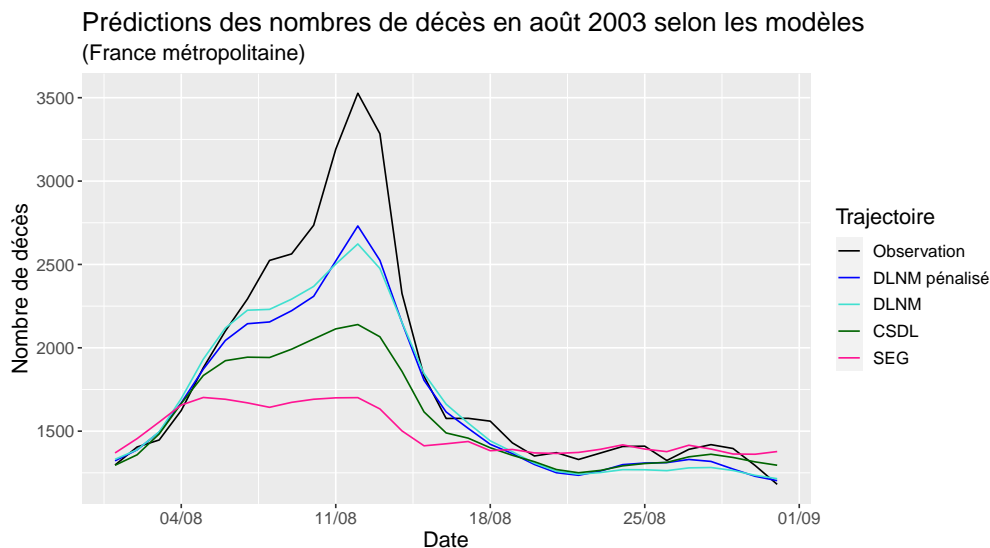
Capacité prédictive des conséquences des pics de chaleur

Le réchauffement climatique se traduira dans les prochaines décennies par une augmentation de la fréquence et de la sévérité des vagues de chaleur. De ce fait, l'étude de l'épisode de canicule de 2003 a paru intéressante. D'après Météo-France, il s'agit en effet de la plus sévère canicule jamais enregistrée en France depuis 1947. Cette dernière a occasionné le décès de 15 000 personnes au cours du mois d'août, dont une grande partie de personnes âgées⁷. Afin de vérifier la capacité de chaque modèle à capter l'ensemble de la mortalité attribuable aux vagues de chaleur, le graphique 3.11 représente le nombre de décès par jour prédit par chacun des modèles étudiés en août 2003.

Tout d'abord, la mortalité est correctement prédite entre 16 août et le 1^{er} septembre, alors que la canicule est terminée. Cependant, au moment du pic de mortalité (le 12 août), les nombres de décès prédits sont particulièrement différenciés. Chaque modèle sous-estime ce pic de mortalité. Le modèle DLNM est celui qui parvient à le reproduire au mieux, malgré un déficit de 800 décès par rapport à la réalité. Trois explications peuvent être avancées. La première est que la canicule de 2003 a donné suite à un ensemble de mesures visant à réduire les conséquences futures de tels épisodes. Ces mesures d'adaptation incluent notamment⁸ la climatisation obligatoire des maisons de retraite, le recensement des personnes isolées ainsi

7. Un résumé de cet épisode est disponible sur le site Internet du Figaro : [↗](#)

8. Voir cet article tiré d'un reportage de France 2 : [↗](#)



GRAPHIQUE 3.11 – Nombre de décès en août 2003 en France métropolitaine : observation et prédiction par modèle étudié

que des messages de prévention diffusés au plus grand nombre. Le modèle ayant été estimé jusqu'à 2013, il tient implicitement compte de ces réformes. Une deuxième explication est que le modèle n'inclut pas l'ensemble des variables pouvant expliquer la surmortalité liée à la chaleur. En particulier, il pourrait être intéressant d'étudier l'apport de la variable de température humide, étant donné son utilité pour mesurer la supportabilité de la chaleur. Enfin, il est aussi possible que les modèles DLNM ne parviennent pas à capter entièrement le pic de mortalité observé en août 2003 car il n'est pas adapté à la modélisation des valeurs extrêmes. En effet, cette période est, dans le jeu historique, celle associée aux nombres de décès par jour maximaux. La théorie des valeurs extrêmes pourrait palier à ce défaut, en adoptant par exemple une modélisation de la mortalité différente au-delà d'un certain seuil de mortalité observé.

Conclusion et éléments à considérer pour la segmentation

À ce stade, les deux modèles DLNM (pénalisé et non-pénalisé) sont conservés pour la suite, chacun disposant d'atouts qui leur sont propres. Le premier doit être testé car il fournit les meilleures prédictions sur le jeu d'entraînement ainsi que des résultats satisfaisants sur le jeu de test. De plus, son temps d'estimation est réduit. Toutefois, utiliser l'extension pénalisée du DLNM, dans un premier temps, semble un choix moins risqué. En effet, le niveau de lissage risque de devoir être augmenté après la segmentation des données (par âge et genre, puis par zone géographique). La raison est que cette dernière va conduire à une désagrégation

des nombres de décès, et à l'utilisation de moins de données pour l'estimation. Ce point sera développé dans la section 3.4.

3.3 Pistes d'amélioration du modèle

3.3.1 L'inclusion d'une variable de richesse

La section 2.3 a permis de soulever le lien existant entre la richesse d'une population et sa mortalité. Il a notamment été vu que les pays disposant de niveaux de développement similaires dans le temps voyaient leur espérance de vie évoluer de façon relativement proche. À plus petite échelle, les cantons britanniques présentaient également une mortalité différenciée selon leur niveau de richesse. De ce fait, il a paru intéressant d'étudier l'apport d'une variable de richesse par département.

Deux solutions sont possibles. La première est de rajouter une variable indicatrice de chaque département comme variable explicative au modèle. Cette solution présente l'avantage de ne pas utiliser de données externes, mais elle dispose également de deux inconvénients. Tout d'abord, elle vise à capter l'ensemble des différences résiduelles de mortalité par département, en dehors de la tendance et de la température, et pas seulement la richesse (l'accès aux soins pourrait également être une variable utile). Il est donc difficile de donner une interprétation précise à cette variable et aux prédictions qui y sont associées. Par ailleurs, la construction de ces indicatrices reviendrait à inclure 95 nouvelles variables explicatives (l'ensemble des départements de France métropolitaine auquel est enlevé un département, afin d'éviter une situation de multicollinéarité⁹). Ainsi, il y aurait 95 nouveaux coefficients à estimer, ce qui complexifierait considérablement l'estimation du modèle. Sur le modèle DLNM non-pénalisé optimal obtenu, le temps de calcul est multiplié par plus de 3, passant de 15 à 51 secondes.

Une autre solution est l'utilisation de données externes. La French Regional Database, présentée dans le chapitre 2, met à disposition des indicateurs de richesse par département disponibles depuis 1922 jusqu'à 2015. Parmi les variables utilisées dans cette étude, un indicateur de dispersion a été préféré. En effet, une variable de tendance a déjà été estimée dans ce modèle, alors qu'aucune variable n'a pour le moment permis de rendre compte des écarts de richesse entre départements. L'indicateur revenu est le revenu par adulte relatif estimé, déjà présenté dans la section 2.3. Il est possible de comparer ces deux méthodes à l'aide de

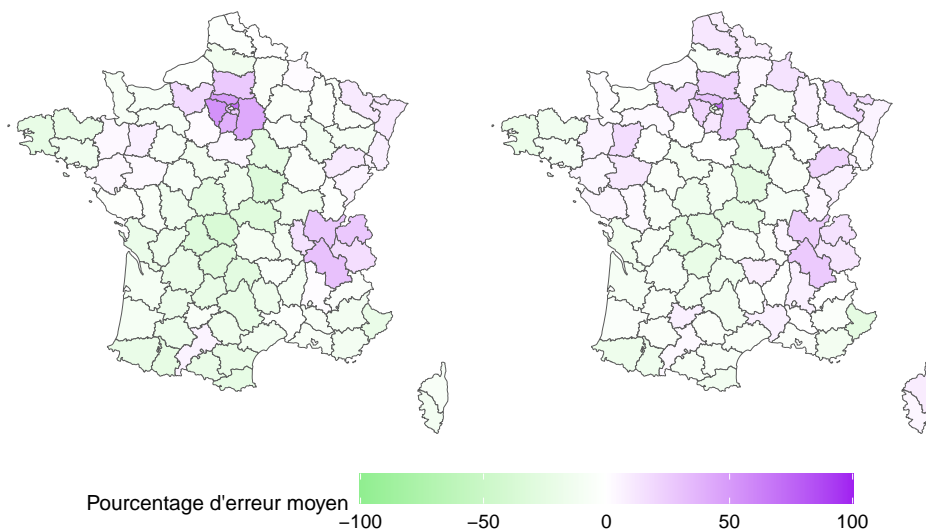
9. Le modèle ne peut être estimé s'il existe une combinaison linéaire des variables explicatives qui est égale à une constante.

plusieurs indicateurs. Au-delà de la comparaison de la déviance et des temps de calcul, il est possible de comparer les résidus géographiques agrégés sur l'ensemble de la période couverte par le jeu d'entraînement (1990-2013). Pour cela, l'erreur moyenne des taux de mortalité annuels est représentée par département dans le graphique 3.12.

Résidus géographiques agrégés

Sans variable de richesse

Avec variable de richesse



GRAPHIQUE 3.12 – Erreur moyenne des taux de mortalité annuels par département selon l'inclusion ou non d'un indicateur de richesse dans un modèle DLNM pénalisé

Lecture : Dans le Finistère, entre 1990 et 2013, le taux de mortalité annuel est sous-estimé en moyenne de 22,1% dans un modèle DLNM sans variable de richesse, et de 17,9% avec l'inclusion d'une variable de richesse.

D'après cette analyse, l'inclusion d'un indicateur de richesse permet de réduire l'erreur moyenne annuelle pour environ trois quarts des départements. Ce résultat est notamment visible dans la région Île-de-France, caractérisée par une sous-mortalité ainsi qu'une richesse moyenne plus importante que dans le reste du pays. Pour quelques départements, dont les Alpes-Maritimes, l'ajout de cette variable a, au contraire, contribué à renforcer le biais du modèle. Par ailleurs, malgré l'inclusion d'une variable de richesse, les biais peuvent rester significatifs. Par exemple, en Seine-Saint-Denis, le taux de mortalité prédit est en moyenne quasiment doublé par rapport à la réalité. Cela est l'indice que l'indicateur de dispersion des revenus entre départements ne parvient qu'à expliquer une partie des écarts de mortalité entre départements. Il pourrait être intéressant d'étudier l'apport d'autres indicateurs, tels que le nombre d'hôpitaux par département (rapporté à sa superficie). Celui-ci pourrait être un bon indicateur de la capacité à prendre en charge des individus nécessitant des soins en urgence. La structure de la population pourrait également influencer de manière significative le taux de mortalité global d'un département. Une segmentation par âge et sexe sera donc

proposée dans la section 3.4.

L'autre solution proposée ici, à savoir la création de variables indicatrices par départements, conduirait à des résidus géographiques agrégés nuls pour chaque département. C'est le résultat de la construction du modèle et de la propriété d'absence de biais des GLM sur le jeu d'entraînement. Cette solution, plus longue à implémenter et difficile à interpréter, possède l'avantage d'être utilisable dans une étude prospective. Cela reviendrait toutefois à supposer que l'ensemble des facteurs différenciant la mortalité par département (inconnus, donc) restent invariants dans le temps. Un tableau comparatif est disponible dans le tableau 3.4.

<i>Modèle</i> Version simple	<i>MPD</i> (entraînement)	<i>MPD</i> (test)	<i>Temps de calcul</i> (en s)
DLNM simple (25 retards)	0,840	1,002	15,285
DLNM avec revenus relatifs	0,726	NC ^a	19,055
DLNM avec indicatrices par département	0,559	0,625	59,003

<i>Modèle</i> Version pénalisée	<i>MPD</i> (entraînement)	<i>MPD</i> (test)	<i>Temps de calcul</i> (en s)
DLNM simple (25 retards)	0,842	0,959	289,255
DLNM avec revenus relatifs	0,725	NC	697,267
DLNM avec indicatrices par département	0,558	0,625	1051,756

TABLEAU 3.4 – Comparatif des performances des modèles estimés

^a. Les revenus relatifs étant disponibles jusqu'à 2015, l'estimation n'est pas possible sur le jeu de test.

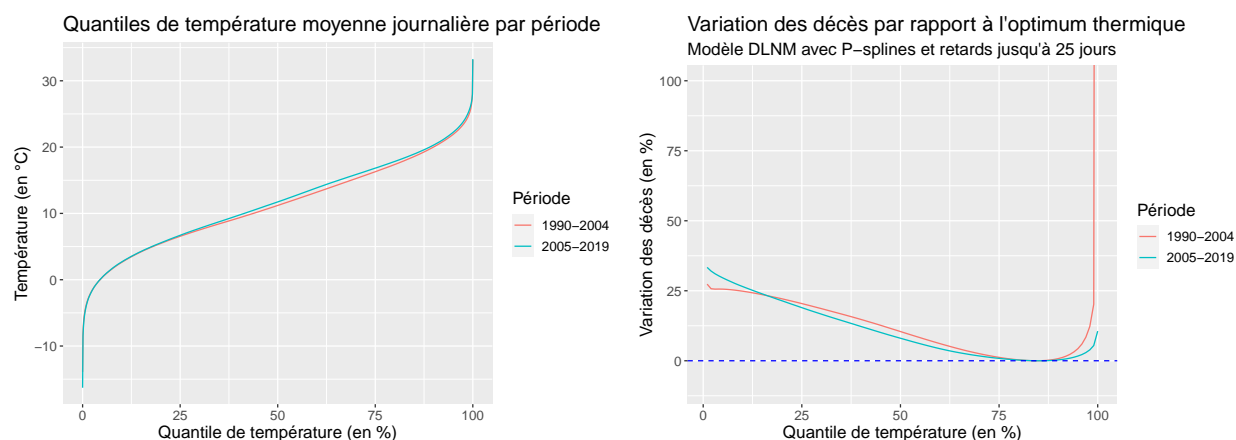
L'ajout d'une nouvelle variable permet d'améliorer plus ou moins la performance du modèle sur le jeu d'entraînement, au prix d'un temps de calcul supplémentaire également plus ou moins important. Le choix de la solution dépend ici de l'objectif de l'actuaire, entre sa volonté de pouvoir interpréter les sorties de son modèle, et celle de pouvoir obtenir ces dernières dans un délai suffisamment court.

3.3.2 La prise en compte de l'acclimatation

L'étude de la canicule de 2003 et de sa modélisation a laissé plusieurs questions en suspens. En particulier, il a été suggéré que la prise en compte de la période 2004-2013 pour l'estimation du modèle a pu entraîner une sous-estimation de la mortalité. La première raison serait que, implicitement, ce modèle tiendrait compte des mesures visant à réduire les conséquences potentielles de tels épisodes. La seconde serait liée à une acclimatation des individus, c'est-à-dire leur capacité à s'adapter aux températures qu'ils observent. Dans une étude prospective,

il est important de pouvoir capter une telle acclimatation, si elle existe. En effet, si cela est le cas, le modèle surestimerait la mortalité en supposant que les individus n’auraient connu que les températures de la période historique. Ce point a notamment été mentionné par Pincemin [47] comme une limite de son étude. Gasparrini et al. [32] ont également montré l’existence de relations différenciées entre température et mortalité dans différents pays, entre 1993 et 2006. Cette évolution se matérialiserait par des variations de la mortalité moins importantes par rapport à l’optimum thermique, pour des quantiles de chaleur donnés.

Afin de vérifier cette hypothèse d’acclimatation, la période historique est séparée en deux sous-périodes : 1990-2004 d’une part, et 2005-2019 d’autre part. Utiliser l’ensemble de la plage historique plutôt que la période d’entraînement ne semble pas problématique ici. En effet, l’objectif n’est pas de sélectionner un modèle, mais de vérifier l’existence d’un phénomène. De ce fait, il est préférable d’utiliser la plage historique maximale afin de renforcer les affirmations. De même, cette sous-section ayant davantage une visée exploratoire, il est décidé d’utiliser le modèle DLNM pénalisé, malgré les conséquences en termes de temps de calcul. L’avantage de cette méthode est d’obtenir des relations lissées et donc d’avoir davantage confiance en leur cohérence sur de nouvelles données, d’autant plus que la période de modélisation est plus courte (15 ans au lieu de 24). Le modèle DLNM pénalisé présenté dans la section précédente est donc appliqué séparément sur chacune des deux sous-périodes. Les quantiles de température ainsi que les variations de la mortalité par rapport à chaque optimum thermique sont représentés dans le graphique 3.13.



GRAPHIQUE 3.13 – Quantiles de température (à gauche) et variation de la mortalité par rapport à l’optimum thermique par quantile de température (à droite) par période

Afin de s’assurer que les relations entre température et mortalité sont comparables entre les deux périodes, il est d’abord nécessaire de comparer les températures observées. Tout d’abord, l’observation des quantiles met en avant l’existence d’un réchauffement généralisé,

avec des quantiles de températures légèrement plus élevés, perceptibles à partir du premier quartile. L'évolution de ces quantiles est comprise entre 0 et 1°C. Si les températures extrêmes semblent relativement proches, l'étude des épisodes de canicule montre une évolution de ces phénomènes. Pour rappel, un épisode de canicule est caractérisé par :

- une température maximale journalière supérieure à 34°C ;
- une température minimale journalière supérieure à 20°C

pendant au moins trois jours consécutifs. Une infographie de France Info¹⁰ basée sur les données de Météo-France permet de confirmer le caractère exceptionnel de la canicule de 2003, tant en termes de longueur que d'intensité. Davantage d'épisodes de canicule ont été connus après 2004, quelques fois plus intenses ou plus longs, mais jamais simultanément (par exemple en juillet 2019). La canicule de juillet 2018 reste la plus proche de celle de 2003, bien que légèrement moins intense par rapport à cette dernière.

La courbure de la variation de la mortalité devient nettement différenciée à partir du quantile 90% de température. Cela signifie que les épisodes de températures extrêmes se sont matérialisés par une surmortalité beaucoup plus modérée depuis 2005. Une vérification plus rigoureuse a également été menée. L'idée sous-jacente est que l'évolution de l'optimum thermique pourrait également être un indice de cette acclimatation potentielle. Un algorithme a été proposé par Tobias et al. [56] afin de déterminer des intervalles de confiance empiriques de la TMM. Un schéma de sa construction sur un modèle simple est disponible dans l'annexe de cet article¹¹. Il s'agit de partir des estimateurs des effets de la température sur la mortalité ainsi que de leur incertitude, afin de simuler des optimums thermiques. Cet algorithme a été implémenté et conduit à des intervalles de confiance à 95% disjoints entre les deux périodes. L'hypothèse de l'existence d'une acclimatation et/ou d'une adaptation face à l'évolution du climat semble donc se confirmer.

Dans une projection, il pourrait donc être intéressant d'inclure un terme d'interaction entre les *cross-basis* (qui modélisent la relation entre la température et la mortalité) et la tendance. En première approche, ce choix permettrait de tenir compte à la fois de l'adoption de mesures d'adaptation et de l'acclimatation des individus face à des températures croissantes. Ainsi que l'ont indiqué Zhao et al. [62], cette méthode présente également l'avantage d'offrir des prédictions plus stables, car elle ne nécessite pas le découpage de la période d'entraînement du modèle. Cependant, en pratique, cette méthode est davantage adaptée pour modéliser une interaction avec des variables binaires, à la fois en termes de temps de calcul

10. Lien vers l'infographie de France Info : [↗](#)

11. L'annexe est disponible ici : [↗](#)

et d'interprétabilité des résultats. Pour une interaction avec une variable continue (ici, le temps), l'interprétation et la construction sont beaucoup plus délicates¹². Cela renforce le côté "boîte noire" de l'algorithme. Le tableau 3.5 présente les performances du modèle DLNM sans et avec interaction pour tenir compte de l'acclimatation.

<i>Modèle</i> Version pénalisée	<i>MPD</i> (entraînement)	<i>MPD</i> (test)	<i>Temps de calcul</i> (en s)
DLNM simple (25 retards)	0,842	0,959	289,255
DLNM avec interaction	0,839	1,007	1604,377

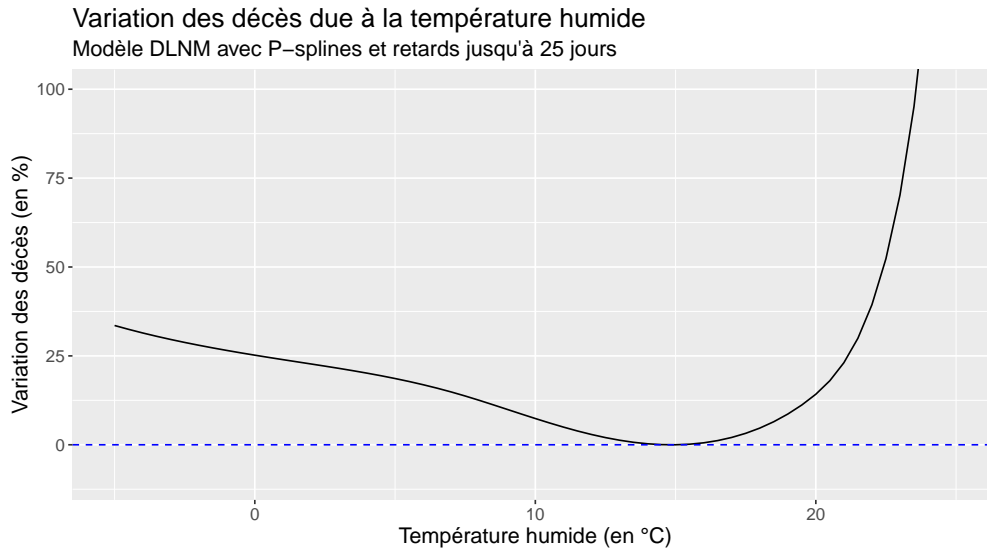
TABLEAU 3.5 – Comparatif des performances des modèles estimés

Ce tableau confirme que le temps de calcul supplémentaire nécessaire pour modéliser l'acclimatation est considérable. Il est à noter que cette étude de l'acclimatation reste incomplète : après avoir discuté de sa matérialité dans le temps, il serait aussi intéressant de vérifier l'existence de relations différenciées entre la température et la mortalité dans l'espace. Cela reviendrait à étudier les relations entre ces deux variables entre plusieurs zones géographiques suffisamment différenciées en termes de températures. Ce point sera abordé dans la section suivante.

3.3.3 L'apport de l'humidité dans la modélisation de la mortalité

Comme évoqué en introduction, la supportabilité de la chaleur dépend d'une combinaison de la température et de l'humidité. Une étude précédente a été menée par Armstrong et al. [2] à l'échelle mondiale, sur un ensemble de 445 villes. L'humidité est prise en compte à travers un terme d'interaction avec la température et d'un certain nombre d'hypothèses plus ou moins lourdes. Parmi celles-ci figurent la linéarité de la relation entre l'humidité et la mortalité. Comme cela a été vu précédemment, l'utilisation de termes d'interaction complexifient considérablement le modèle. L'approche retenue utilise donc la température humide, définie dans le chapitre 1, ce qui a deux avantages. D'une part, son interprétation en termes de supportabilité de la chaleur est connue. D'autre part, l'utilisation de *cross-basis* permettra de modéliser des relations complexes, et plus forcément linéaires, entre l'humidité et la mortalité. Il est ici proposé de remplacer la variable de température avec celle de température humide dans le modèle DLNM initial. L'objectif est de savoir comment cette nouvelle variable parvient à améliorer la capacité prédictive du modèle par rapport au choix de variables initial. Après avoir implémenté le modèle, la persistance des effets de la chaleur humide ainsi que les variations de mortalité sont représentées dans le graphique 3.14.

12. Voir les notes de cours de Marchand sur l'interprétation des sorties des régressions de Poisson : ↗



GRAPHIQUE 3.14 – Relations entre température humide et mortalité prédites dans un modèle DLNM pénalisé

En comparaison du même modèle DLNM appliqué aux températures simples, la courbure des variations de mortalité conserve la forme en "U" visible dans le graphique 3.9, mais est plus allongée. En particulier, les effets du froid sur la mortalité sont moins visibles : par rapport à l'optimum thermique, la surmortalité attribuable aux températures minimales est de l'ordre de 35%, contre plus de 50% pour le modèle initial. Les effets exponentiels de la chaleur sur la mortalité modélisés avec la température sans humidité restent présents. Il reste cependant difficile de confirmer la cohérence de la courbure obtenue à l'aide de la littérature. En effet, le seuil couramment admis pour la supportabilité de la chaleur humide est de 35°C, et l'inclusion de la température humide dans un modèle DLNM n'a pas encore fait l'objet d'une étude à notre connaissance. La forme de la mortalité attribuable aux températures humides basses est quant à elle confirmée par la littérature. Mäkinen et al. [42] ont notamment observé une augmentation des infections des voies respiratoires lorsque la température extérieure et le taux d'humidité sont faibles simultanément. L'apport de la variable de température humide est quantifié à l'aide du tableau de performances 3.6.

<i>Modèle</i> Version pénalisée	<i>MPD</i> (entraînement)	<i>MPD</i> (test)	<i>Temps de calcul</i> (en s)
DLNM avec température simple	0,842	0,959	289,255
DLNM avec température humide	0,848	0,957	211,842

TABLEAU 3.6 – Comparatif des performances des modèles estimés

Bien que l'estimation du modèle soit plus rapide¹³, aucun gain de précision significatif n'a été constaté. La température humide ne semble donc pas une variable plus efficace que la température simple afin de modéliser la mortalité. Cela était déjà la conclusion obtenue par Armstrong et al. [2]. Par ailleurs, l'humidité relative n'est pas disponible directement comme variable de projection climatique à la maille journalière. Il faudrait alors se contenter d'une maille mensuelle, ou de l'humidité *spécifique*. La première option réduirait le potentiel d'interprétations biométriques du modèle, tandis que la seconde nécessite de dériver la température humide à partir d'un ensemble de variables (telles que la pression atmosphérique). L'inclusion de l'humidité dans la modélisation semble donc, de toute manière, délicate dans une étude prospective.

3.4 Segmentation de la mortalité

La partie précédente avait pour principal objectif de déceler des variables géographiques pertinentes, afin pouvoir expliquer au mieux la mortalité. Comme cela a été vu précédemment, afin de tarifier plus précisément un contrat d'assurance-vie, des critères spécifiques à l'individu doivent également être inclus.

3.4.1 Choix de la segmentation

Traditionnellement, l'âge et le sexe constituent le principal choix de segmentation. Cependant, cette dernière pourrait s'avérer incomplète. En effet, Gasparrini et al. [32] ont mis en évidence, au-delà d'une potentielle acclimatation des individus *dans le temps*, l'existence de relations différenciées entre la température et la mortalité dans les différentes villes de leur étude. Ce phénomène sera appelé ici acclimatation *dans l'espace*. Pour cette raison, des modèles DLNM distincts seront appliqués pour chaque combinaison de ces trois variables. Il reste maintenant à déterminer la segmentation adaptée. Ce choix reposera sur un certain nombre de compromis, résumés dans le schéma 3.15.

Tout d'abord, sachant que les modèles DLNM s'appuient sur le cadre global des GLM, la modélisation bénéficiera de la propriété d'absence de biais des prédictions sur les données d'entraînement. Autrement dit, chaque modèle estimera correctement le nombre moyen de décès sur l'ensemble de la période historique. Naïvement, cela pourrait inciter à effectuer une segmentation la plus fine possible. De plus, effectuer un modèle par âge, sexe et département

13. La méthode du maximum de vraisemblance restreint (REML) utilisée converge plus rapidement.

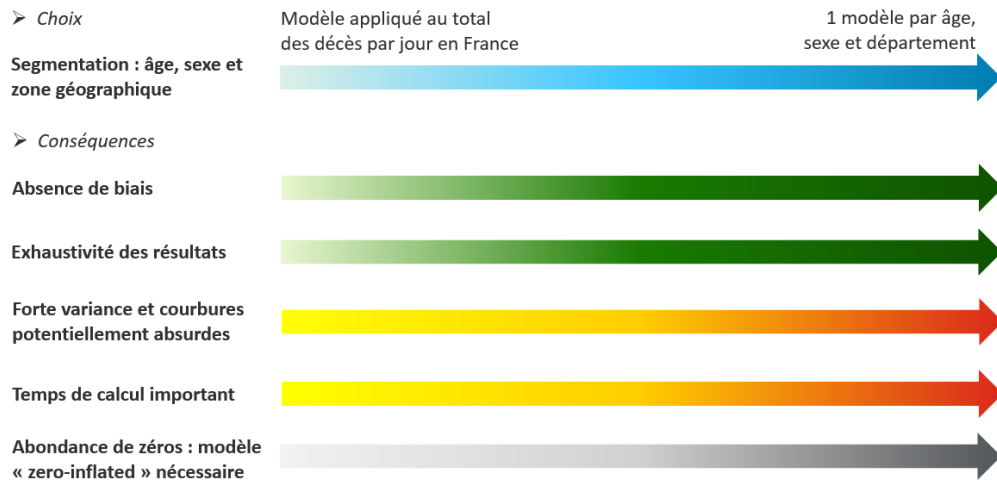


SCHÉMA 3.15 – Présentation des enjeux pour la segmentation de la mortalité

Lecture : estimer un modèle par âge, sexe et département permettra d'obtenir davantage de relations à interpréter, mais celles-ci pourraient s'avérer irréalistes.

permettrait d'avoir suffisamment de matière pour des interprétations biométriques fines. Toutefois, ce choix amènerait plusieurs risques majeurs. Ceux-ci sont principalement liés au fait que chaque modèle va "ignorer" les données issues d'autres combinaisons de la segmentation choisie. Or, si la forme des relations à spécifier est complexe et n'est pas spécifiée directement au modèle, comme entre la température et la mortalité, le nombre d'observations nécessaires pour capter correctement la relation est plus important. De ce fait, si le modèle repose sur trop peu de données, les relations obtenues peuvent facilement devenir incohérentes et inexploitable pour une prédiction. C'est l'un des inconvénients potentiels du modèle DLNM par rapport aux formes log-linéaires classiques (telles que le modèle CSDL). Par ailleurs, estimer un grand nombre de modèles distincts peut contribuer à faire gonfler le temps de calcul global. Cela est d'autant plus vrai si la loi de Poisson ne peut plus correctement être utilisée. Si la variable d'intérêt est surdispersée (ce qu'a confirmé le test de Cameron et al. [9]) et de moyenne très proche de zéro, il est en effet probable qu'une régression de Poisson sous-estime le nombre de zéros (i.e. jours sans décès) dans le modèle. Il sera alors nécessaire de considérer des lois *zero-inflated* ou *hurdle*, qui conduiront à estimer en deux temps la probabilité d'un nombre de décès nul, et celle d'un nombre de décès strictement positif. Le temps d'estimation de chaque modèle sera de fait allongé. Les enjeux décrits ici sont particulièrement typiques du compromis "biais-variance" connu en actuariat.

De ce fait, la segmentation doit être particulièrement réfléchie, et correspondre aux besoins de l'actuaire par rapport au produit qu'il souhaite tarifier. Pour cette raison, il a été décidé de regarder une à une les variables pouvant en faire l'objet.

- **Âge** : en assurance-vie, les individus sont le plus souvent regroupés par tranches de cinq ou dix ans afin de répondre au dilemme précédent. Le choix d'une granularité de cinq ans a finalement été effectué afin d'obtenir une relation plus fine selon les différentes tranches d'âge. Une catégorie 0-59 ans est toutefois créée, car la mortalité y est globalement inférieure à celle des autres tranches d'âge, et car le cœur de cible des contrats d'assurance-vie concerne les personnes âgées. De même, les personnes de 90 ans et plus sont regroupées au sein d'une même catégorie. Ces choix sont relativement similaires à ceux de Pincemin [47], et permettront une comparaison des résultats. Par ailleurs, l'étude de cette dernière a montré des relations différenciées par tranche d'âge, avec une mortalité attribuable au froid croissante avec l'âge.
- **Sexe** : depuis le 21 décembre 2012 et une décision de la Cour de justice de l'Union Européenne, il est interdit aux entreprises d'assurance d'élaborer des tarifications segmentées selon le sexe. Toutefois, il reste intéressant d'inclure le sexe dans la segmentation, dans un objectif purement académique. En effet, des interprétations biométriques supplémentaires pourront être déduites et aideront à rendre compte de la qualité du modèle. L'étude de référence a par exemple souligné l'existence d'une surmortalité attribuable à la chaleur plus importante chez les femmes.
- **Zone climatique** : les travaux de Armstrong et al. [2] ont permis de mettre en avant l'existence de relations différenciées entre la température et la mortalité selon les villes de leur étude. Toutefois, leur étude s'appuie non pas sur une zone géographique complète, mais sur des métropoles issues de différents pays. Afin d'éviter les écueils liés à une analyse départementale, la relation entre la température et la mortalité sera différenciée par zone climatique (océanique, méditerranéen et semi-continentale), en utilisant la classification retenue par le Ministère de la Transition Écologique. Les données de chaque département appartenant à une même zone climatique restent incluses pour l'estimation, ce qui laisse la possibilité d'effectuer des prédictions à la maille départementale par la suite.

Des modèles DLNM seront donc estimés pour chaque combinaison associée à la segmentation retenue.

3.4.2 Interprétation des résultats

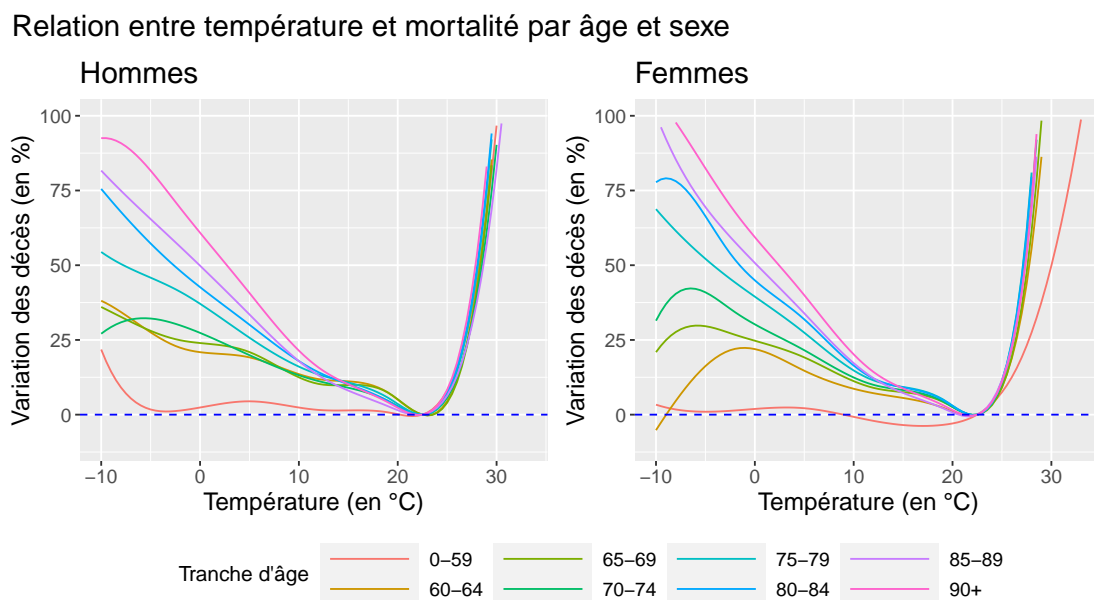
Le choix de la modélisation étant fixé, et l'objectif principal étant de décrire des interprétations biométriques, les estimations sont effectuées sur l'ensemble de l'historique (1990-2019). Par ailleurs, les modèles DLNM non-pénalisés ayant fourni (comme anticipé) des prédictions

peu cohérentes suite à la segmentation, seuls les résultats obtenus à partir du modèle DLNM pénalisé sont présentés. Dans un souci de lisibilité et de compréhension, les relations entre la température journalière et la mortalité sont représentées à partir de segmentations progressives. Autrement dit, la segmentation est effectuée par âge et sexe, puis par zone climatique, avant de passer à la segmentation complète. L'objectif de cette démarche est de comprendre les différenciations de la mortalité selon ces classes, et aussi de s'assurer que la granularité choisie est cohérente. Les critères retenus pour cette différenciation sont :

- la surmortalité attribuable à la chaleur et au froid, ainsi que la courbure obtenue et l'optimum thermique ;
- la persistance des effets de la chaleur et du froid.

Segmentation par âge et sexe seulement

La segmentation est dans un premier temps effectuée à partir de modèles différenciés par âge et sexe. Les relations entre la température et la mortalité obtenues sont représentées dans le graphique 3.16.



GRAPHIQUE 3.16 – Relations entre température et mortalité par âge et sexe prédites dans un modèle DLNM pénalisé

Tout d'abord, les graphiques ne mettent pas en évidence une variation significative de l'optimum thermique entre les combinaisons âge/sexe. Celui-ci est toujours compris entre 21 et 23°C. En revanche, des variations apparaissent nettement concernant les effets du froid et

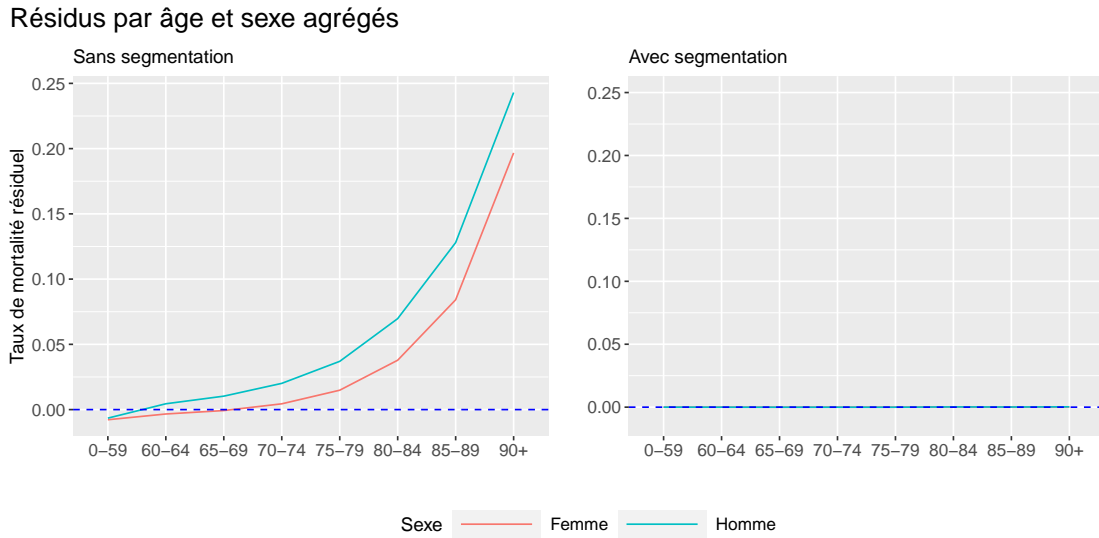
de la chaleur. Comme obtenu dans les travaux précédents, les effets de la chaleur semblent légèrement plus importants chez les femmes, ce qui se traduit par une courbure légèrement plus marquée au-delà d'un certain seuil de température. Des facteurs physiologiques et sociaux justifieraient cette plus grande sensibilité. Ils sont mentionnés dans un article du Monde¹⁴. Par exemple, les femmes transpireraient moins que les hommes, et de fait évacueraient moins la chaleur. Davantage de femmes âgées vivent seules et sont donc davantage isolées, ce qui peut également expliquer un plus grand risque de décès durant les épisodes de canicule. De plus, les femmes et les hommes semblent présenter des sensibilités similaires au froid, à tranche d'âge donnée. Ce résultat a également été obtenu dans la littérature, par exemple au Pays-Bas dans une analyse similaire menée par Folkerts et al. [25] sur l'ensemble des décès, toutes causes confondues.

Les différences de relations entre température et mortalité sont également visibles selon l'âge. Il semble en effet que la surmortalité associée aux températures extrêmes (chaleur et froid confondus) devienne plus importante à mesure que l'individu vieillit, indépendamment du sexe. Cela est davantage visible pour les vagues de froid que pour les pics de chaleur, du fait de la courbure assez linéaire : entre deux classes d'âge, la variation de la mortalité par rapport à l'optimum thermique augmenterait d'une dizaine de points de pourcentage. En pratique, ce résultat est également obtenu pour les températures élevées, mais la courbure asymptotique le rend moins facilement perceptible. Concernant la tranche d'âge 0-59 ans, la modélisation de la mortalité est plus délicate car elle présente l'inconvénient de regrouper deux catégories bien distinctes : la mortalité infantile et la mortalité des actifs les plus âgés. Toutefois, de manière générale, les interprétations de ces sorties sont cohérentes avec la littérature et justifient la segmentation effectuée.

Une analyse des résidus est proposée afin de fournir une justification supplémentaire. Il s'agit ici de comparer les déviations de mortalité prédites pour un modèle sans segmentation par âge et sexe, avec celles obtenues avec cette distinction. Ici, les résidus sont exprimés en écart de taux de mortalité national sur l'ensemble des trente années. Par construction des modèles linéaires généralisés, le nombre moyen de décès sur l'ensemble des départements et des journées de la période historique doit être reproduit pour chaque segmentation. La comparaison est disponible dans le graphique 3.17.

D'une part, les résidus obtenus sans segmentation par âge et sexe sont conformes avec ce qui était attendu. Tout d'abord, le modèle a globalement tendance à sur-estimer la mortalité des femmes, et au contraire à sous-estimer celle observée pour les hommes. Ce résultat est conforme avec l'observation d'une espérance de vie à la naissance plus élevée pour les

14. Lien vers l'article du Monde (payant) : [↗](#)



GRAPHIQUE 3.17 – Taux de mortalité résiduel moyen en France métropolitaine sur la période 1990-2019 par âge et sexe (de gauche à droite : sans et avec segmentation)

Lecture : Sans segmentation, quel que soit l'âge et le sexe, le taux de mortalité moyen estimé est de l'ordre de 0,009 (soit 9 décès pour 1 000 habitants). Pour les hommes dont l'âge est supérieur ou égal à 90 ans, entre 1990 et 2019, le taux moyen observé est de 0,252. Le taux moyen est donc sous-estimé de 0,243 environ.

personnes de sexe féminin. Ensuite, la courbe des taux de mortalité résiduels suit de manière fidèle la trajectoire des taux de mortalité par âge et sexe. Ainsi, la mortalité des actifs a tendance à être sur-estimée, tandis que celle des personnes les plus âgées est surestimée. L'analyse graphique met toutefois en avant une subtilité importante de la propriété d'absence de biais des GLM, qui n'est plus vraie au passage des nombres de décès en taux de mortalité. En effet, si d désigne le département d'étude, i la segmentation âge/sexe considérée et t le jour considéré dans la période historique, la propriété d'absence de biais des GLM implique : $\sum_{d,i,t} \hat{y}_{d,i,t} = \sum_{d,i,t} y_{d,i,t}$. Cependant, la reproduction des taux de mortalité moyens impliquerait une propriété différente : $\sum_{d,i,t} w_{d,i,t} \hat{y}_{d,i,t} = \sum_{d,i,t} w_{d,i,t} y_{d,i,t}$, avec $w_{d,i,t}$ l'inverse de la population vivante de l'année dans le département d , pour l'année associée au jour t et le groupe i . Une condition suffisante pour que les GLM parviennent à reproduire les taux de mortalité pour une segmentation donnée est que la population vivante reste constante entre les départements et les classes d'âge pour chaque année. En pratique, cette propriété n'est pas vérifiée. D'une part, la pyramide des âges n'est pas constante et montre une plus concentration de population chez les actifs¹⁵. D'autre part, inclure des classes d'âge avec des populations très disparates dans un unique modèle conduit à ce que les taux de mortalité estimés se rapprochent davantage de ceux observés pour la classe avec la plus grande popula-

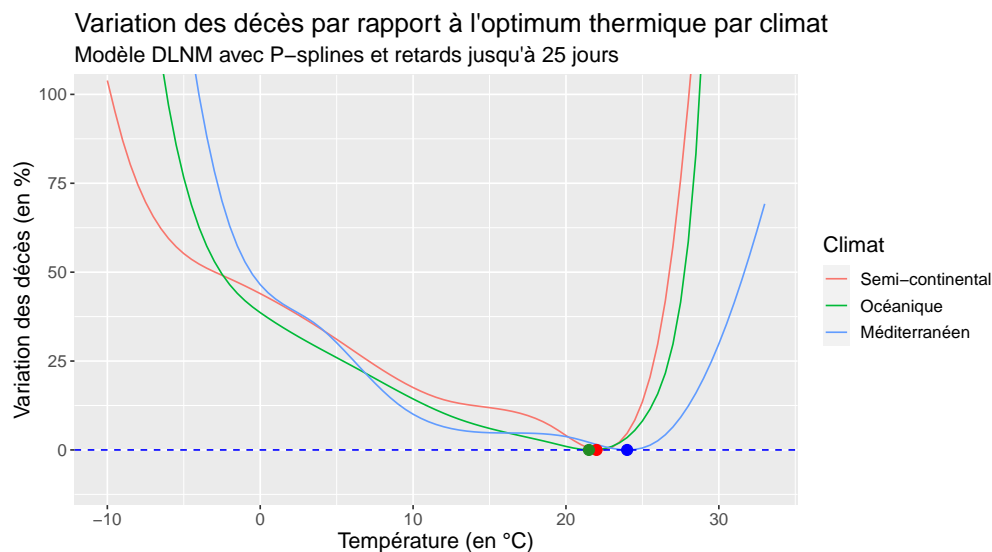
15. Voir la pyramide des âges projetée de l'INSEE, basée sur les recensements de populations de 2018 : ↗

tion. Pour cette raison, les résidus de mortalité obtenus sans segmentation ne sont pas centrés en zéro. De plus, les taux de mortalité moyens estimés en France sur la période 1990-2019 sont plus proches de ceux associés à la classe 0-59 ans.

En revanche, une fois la segmentation fixée, les déviations de mortalité deviennent négligeables : elles sont comprises entre 2 pour 1 000 000 et 3 pour 10 000 habitants selon le groupe considéré. Cela justifie l'intérêt d'estimer chaque modèle par âge et sexe séparément afin d'obtenir des estimations de la mortalité cohérentes.

Segmentation par zone climatique seulement

Après avoir étudié les différences de relations entre température et mortalité par âge et sexe, l'acclimatation *dans l'espace* est étudiée. Pour cela, les modèles sont estimés séparément pour chaque zone climatique (océanique, méditerranéen et semi-continentale), tandis que la distinction par âge et sexe est temporairement abandonnée. Les relations entre température et mortalité obtenues sont représentées dans le graphique 3.18.



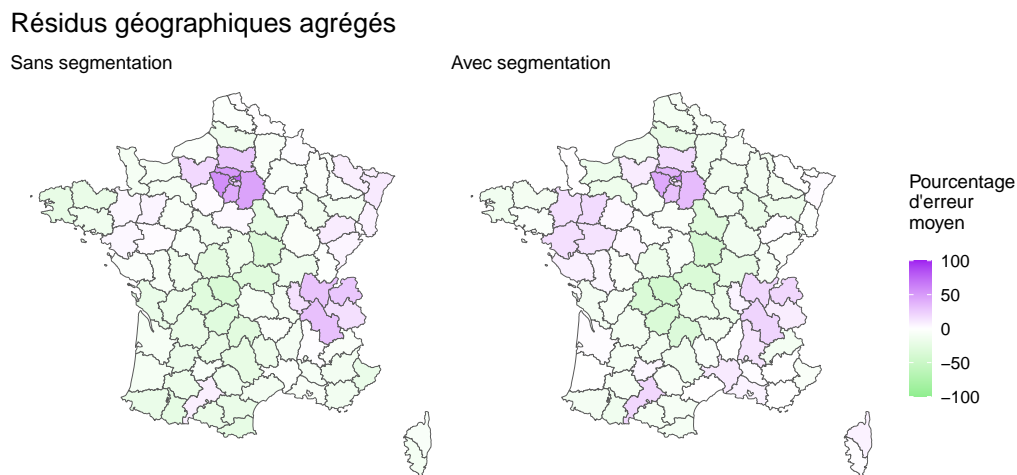
GRAPHIQUE 3.18 – Relations entre température et mortalité par zone climatique prédites dans un modèle DLNM pénalisé

Plusieurs constats permettent de soutenir l'hypothèse d'une acclimatation géographique. Tout d'abord, l'optimum thermique du climat méditerranéen est sensiblement supérieur à celui des autres zones d'étude. Pour rappel, ce climat est caractérisé par des températures plus élevées, tout au long de l'année, notamment en comparaison du climat océanique. Par ailleurs, le climat méditerranéen posséderait, selon le modèle, une sensibilité nettement plus

faible à la chaleur en comparaison des autres climats. Cela se traduit concrètement par une courbure décalée, dont l'asymptote verticale ne s'observe pas sur les températures historiques. Au contraire, les climats océanique et semi-continentale possèdent des relations très similaires entre température et mortalité au-delà d'une température de 20°C, avec une mortalité augmentant de façon exponentielle lors des épisodes de chaleur les plus intenses. Une explication a été avancée par Besancenot [7], faisant suite à une analyse géographique des conséquences de la canicule de 2003. Selon lui, la canicule de 2003 a bien montré que les régions habituées à la chaleur avaient moins souffert. Ce constat est cohérent avec la variation moins importante de la mortalité attribuable aux fortes températures modélisée dans le Sud du pays.

De manière similaire, concernant les effets du froid, l'Est de la France serait moins sensible à une même température négative extrême. Pour rappel, le climat semi-continentale est celui avec les températures minimales hebdomadaires les plus faibles. La courbure obtenue est décalée dans les climats méditerranéen et océanique, qui bénéficient d'hivers plus doux. En cela, les conclusions obtenues paraissent de nouveau cohérentes et justifient la segmentation géographique.

L'analyse des résidus annuels moyens par département permet de compléter cette étude. Il est possible de comparer ces erreurs agrégées sur l'ensemble de la période, entre un modèle avec et sans segmentation par zone climatique. Les représentations des erreurs sont disponibles dans le graphique 3.19.



GRAPHIQUE 3.19 – Erreur moyenne des taux de mortalité annuels par département selon l'inclusion ou non d'une segmentation par zone climatique dans un modèle DLNM pénalisé

Lecture : Dans le Finistère, entre 1990 et 2019, le taux de mortalité annuel est sous-estimé en moyenne de 22,2% dans un modèle DLNM sans segmentation, et de 13,3% avec l'inclusion d'une segmentation.

En comparant les valeurs absolues, le modèle DLNM avec segmentation fournit des er-

reurs moyennes inférieures pour 59% des départements. Ce taux est rassurant car il signifie que l'erreur systématique moyenne est réduite pour la majorité des départements. D'après les cartes, la segmentation semble davantage profitable pour les départements côtiers, pour lesquels les différences de climats sont plus marquées. Cependant, pour d'autres départements notamment ruraux, dans les anciennes régions Limousin ou Bourgogne, l'erreur moyenne est plus importante. Une interprétation est que, pour ces départements, la perte d'observations issues des autres zones climatiques est plus préjudiciable en moyenne que le recentrage des données.

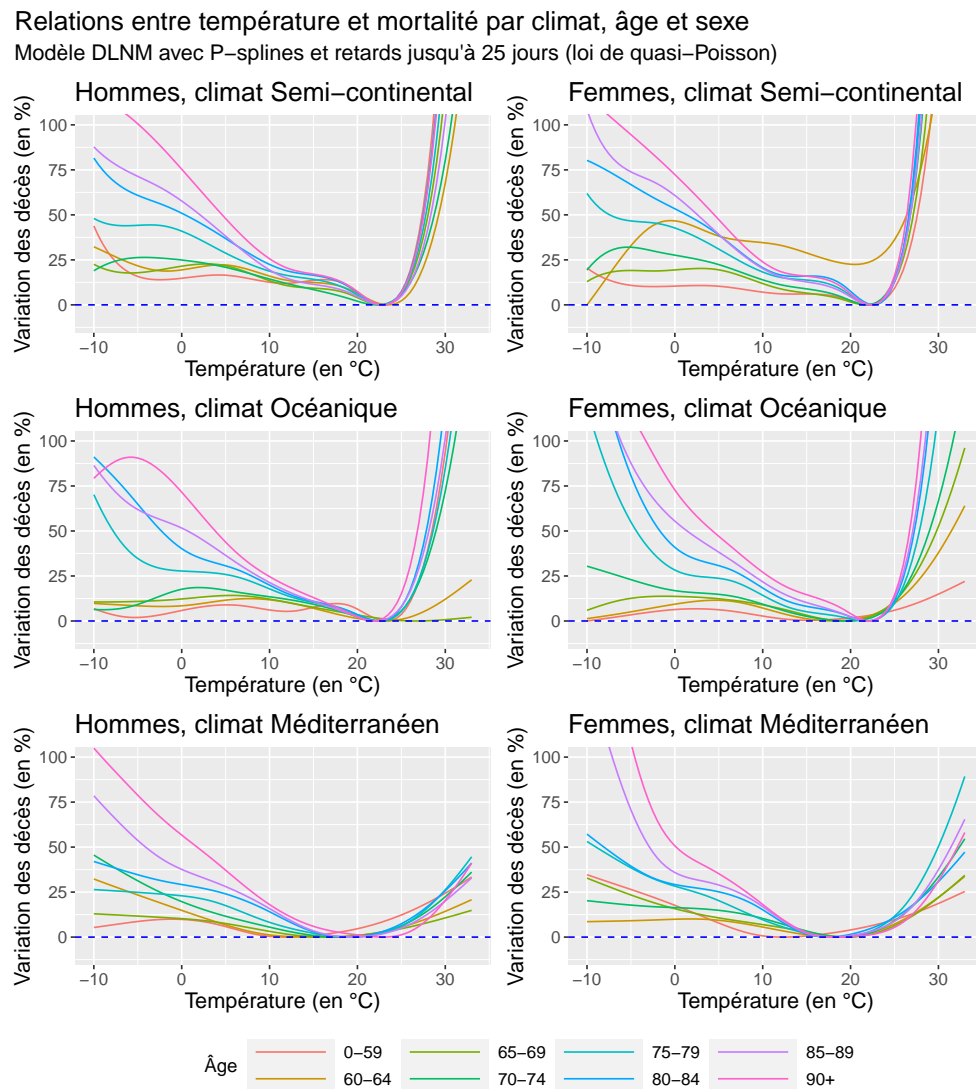
À travers cette analyse, il apparaît que le choix d'une segmentation géographique dépend du contexte et des attentes de l'assureur. En effet, la segmentation géographique permet, d'une part, de diminuer les erreurs systématiques dans certains départements grâce à une meilleure prise en compte de l'acclimatation (spatiale). Toutefois, la segmentation conduit d'autre part à ignorer certaines observations issues d'autres zones climatiques, alors qu'elles fournissent également une information supplémentaire sur le lien température-mortalité. Les biais agrégés augmentent ainsi dans certains départements. Il s'agira alors de vérifier, selon les expositions de l'assureur, quelles zones géographiques nécessitent une modélisation plus précise de la mortalité.

Il est à noter, cependant, que l'approche de segmentation proposée ici ne repose que sur une classification officielle, et qu'une méthodologie plus rigoureuse aurait pu être utilisée. Par exemple, Cohen et Gninghaye [12] ont proposé une démarche de *clustering* des différents climats à l'aide de l'*Empirical Orthogonal Function* (EOF). Cette méthode, assez proche de l'Analyse en Composantes Principales (ACP) mais davantage adaptée aux séries temporelles, a permis de décomposer la France en 4 classes selon les températures observées historiquement. Une telle procédure est toutefois coûteuse en temps de calcul.

Segmentation par âge, sexe et zone climatique

Une dernière segmentation, la plus fine jusqu'à présent, est finalement testée. En toute logique, elle devrait confirmer les relations précédentes. Cependant, tester ce niveau de segmentation permet aussi de s'assurer que les relations estimées ne sont pas affectées par l'utilisation d'un nombre trop faible de données positives. En effet, chaque modèle sera estimé sur une base plus petite, avec moins d'observations contenant des nombres de décès positifs. Cela peut affecter les différentes relations. Autrement dit, cette partie permettra de vérifier si la loi de Poisson (ou de quasi-Poisson) reste adaptée au contexte de cette étude, et s'il n'est pas préférable d'utiliser la loi *zero-inflated Poisson* présentée précédemment. Les relations

entre la température et la mortalité obtenues avec une loi de quasi-Poisson sont représentées dans le graphique 3.20.



GRAPHIQUE 3.20 – Relations entre température et mortalité par âge, sexe et zone climatique prédites dans un modèle DLNM pénalisé avec loi de quasi-Poisson

La plupart des relations obtenues avec une segmentation moins fine restent observables. Cela concerne en particulier la plus faible sensibilité de la mortalité à la chaleur dans le climat méditerranéen, ou bien la difficulté croissante à supporter les épisodes de températures extrêmes avec l'âge. Cependant, quelques relations étonnantes sont observées, concernant la modélisation des effets du froid. Par exemple, pour les femmes de 60-64 ans vivant dans l'Est de la France, la mortalité minimale serait atteinte pour une température de -10°C , avec une variation des décès positive et croissante jusqu'à 0°C . De la même manière, certaines courbures obtenues pour la surmortalité attribuable au froid sont assez peu lisses. Ces ré-

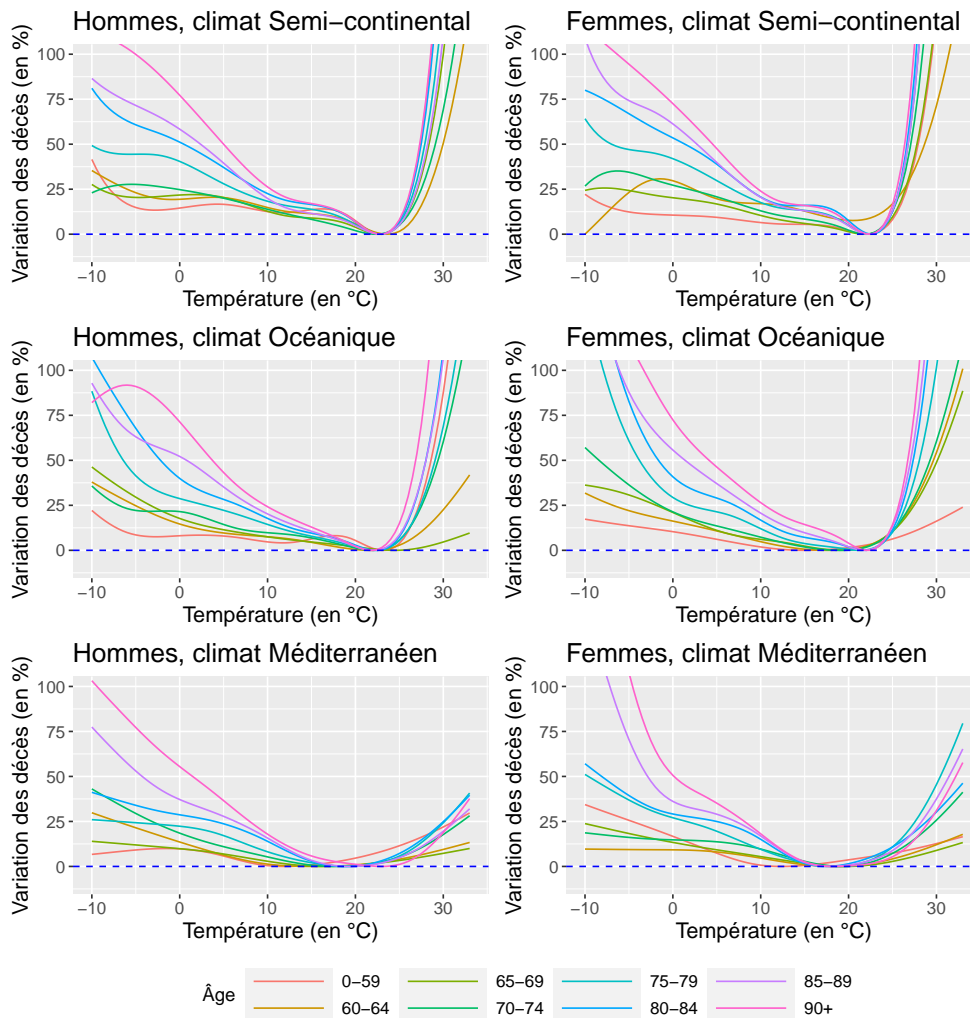
sultats semblent peu cohérents et n'ont pas été confirmés par l'étude de la littérature. Cette dernière permet toutefois de conjecturer une explication de ce phénomène. Il a en effet été vu que les effets du froid, contrairement à ceux de la chaleur, ne se manifestent non pas par une surmortalité importante et visible immédiatement, mais par une mortalité lissée dans le temps. De ce fait, à l'échelle d'un département et d'une classe d'âge, une température très basse est davantage susceptible d'être suivie par quelques jours avec un seul décès et d'autres sans décès. Le modèle devra donc être en mesure de "valoriser" suffisamment les observations avec des décès survenus parmi l'ensemble des retards retenus, en comparaison de celles sans décès. Par ailleurs, la classe d'âge 60-64 ans est celle avec le taux de décès le plus faible parmi celles choisies pour l'étude. De ce fait, il pourrait être intéressant de tester la loi *zero-inflated Poisson*, qui attribuera un poids potentiellement plus important aux journées sans décès. Les relations obtenues avec cette nouvelle hypothèse de loi sont représentées dans le graphique 3.21.

Les relations obtenues sont généralement très similaires. Celle, aberrante, obtenue avec le modèle de quasi-Poisson concernant la tranche d'âge 60-64 ans chez les femmes dans le climat semi-continentale, est cependant en partie corrigée. Bien que l'optimum thermique obtenu soit de nouveau la température minimale, le creux de la mortalité observé à l'optimum thermique habituel (autour de 20°C) est plus net qu'avec le modèle précédent. De plus, d'un point de vue prospectif et dans un contexte de réchauffement climatique, le fait qu'un modèle prédise une sous-mortalité pour des températures extrêmement froides n'est pas si handicapant. Il paraît en effet préférable que le modèle fournisse des relations cohérentes pour les températures qui seront observées à l'avenir. Finalement, le modèle *zero-inflated Poisson* semble une extension utile de la loi de Poisson lorsque la segmentation est fine. D'une part, lorsque le nombre de données utilisé est réduit et le nombre d'observations avec décès restreint, il parvient à mieux capter les faits stylisés. D'autre part, s'agissant d'une extension de la loi de Poisson, il parvient à reproduire les relations satisfaisantes obtenues sur les classes d'âge pour lesquelles davantage de décès sont observés.

Il est cependant difficile de tester rigoureusement l'apport de la loi *zero-inflated* par rapport à la loi de Poisson usuelle. Un test de ratio de vraisemblance a par exemple été envisagé, mais sa visée est de mesurer l'apport de variables explicatives, et non d'un changement de loi. Le test de Vuong [60] permet quant à lui de vérifier si la vraisemblance d'un modèle est améliorée avec une inflation de zéros, mais il repose sur la comparaison de deux modèles dits "non-imbriqués". Wilson [61] a ainsi expliqué que ce test a été présenté à tort comme permettant de comparer deux modèles, l'un avec une loi *zero-inflated* Poisson et l'autre avec une loi de Poisson. Cependant, il est possible de reprendre la séparation en jeu d'entraînement (don-

Relations entre température et mortalité par climat, âge et sexe

Modèle DLNM avec P-splines et retards jusqu'à 25 jours (loi zero-inflated Poisson)



GRAPHIQUE 3.21 – Relations entre température et mortalité par âge, sexe et zone climatique prédites dans un modèle DLNM pénalisé avec loi *zero-inflated Poisson*

nées jusqu'à 2013 et à partir de 2014), et de regarder le nombre de fois où la loi *zero-inflated* améliore la déviance par rapport au modèle original. Les résultats de cette comparaison sont synthétisés dans le tableau 3.7.

Modèle	Entraînement	Test
Fréquence de diminution de la MPD avec la loi <i>zero-inflated</i> Poisson	92%	94%

TABLEAU 3.7 – Comparatif du modèle retenu après segmentation selon l'hypothèse de loi

Lecture : Pour 92% des segmentations, utiliser l'hypothèse de loi *zero-inflated* Poisson permet de diminuer la MPD sur le jeu d'entraînement par rapport à l'utilisation de l'hypothèse de loi quasi-Poisson.

Dans la majorité des cas, changer l'hypothèse de loi permet de diminuer la déviance moyenne de Poisson du modèle, à la fois sur les données d'entraînement et de test. Le modèle avec loi *zero-inflated Poisson* sera donc retenu pour l'analyse prospective de la mortalité par scénario climatique.

Analyse de sensibilité de la période temporelle

Il est à noter qu'une analyse de sensibilité a été menée sur la période historique retenue. En effet, il a été observé dans ce chapitre que les périodes 1990-2004 et 2005-2019 se différencient, notamment, par une moins forte sensibilité de la mortalité aux températures extrêmes. Cela pourrait s'expliquer par les effets de l'acclimatation et de l'adaptation, via des mesures de prévention visant à réduire les conséquences mortelles des épisodes de chaleur.

Toutefois, retirer l'année 2003 dispose de deux inconvénients majeurs. D'une part, ce choix conduit à ne plus disposer d'observations relatives à des phénomènes de chaleur à la fois intenses et durables, pourtant amenés à se multiplier dans les prochaines décennies. D'autre part, il est nécessaire de se rappeler qu'à l'échelle d'une journée, d'une classe d'âge et d'un département, le décès est un événement rare. Afin de capter correctement la relation température-mortalité pour chaque segmentation, il est donc nécessaire de conserver un nombre suffisant d'observations pour chaque niveau de température. Le choix a donc été effectué de conserver l'intégralité de la période 1990-2019 pour l'analyse prospective de la mortalité.

Conclusion du chapitre

Dans ce chapitre, les différentes méthodes permettant de représenter la relation entre la température et la mortalité ont été implémentées puis comparées. Le modèle DLNM, et en particulier sa variante pénalisée, parvient à davantage capter les faits stylisés mentionnés dans la littérature. En particulier, il prédit une variation polynomiale de la mortalité attribuable à la chaleur, et a permis de mieux modéliser le pic de mortalité observé durant la canicule d'août 2003. De ce fait, cette modélisation paraît plus cohérente pour une analyse prospective de la mortalité. Ensuite, sur la base de cette relation, une segmentation adaptée des données a été effectuée. Celle-ci a permis de capter des relations différenciées entre la température et la mortalité observables selon l'âge, le sexe et le climat dans lequel vivent les individus. La segmentation assez fine retenue a finalement conduit à modifier l'hypothèse de loi, afin de mieux tenir compte des observations sans décès.

4

Projections de la mortalité à horizon 2070 par scénario climatique

Dans ce chapitre, la modélisation segmentée de la mortalité présentée précédemment est appliquée à des données de projection. Ces dernières sont introduites dans la section 4.1. Ensuite, les projections de la mortalité obtenues sont analysées dans la section 4.2. Les applications potentielles de l'étude, ses limites ainsi que des pistes d'amélioration seront enfin présentées dans la section 4.3.

4.1 Présentation et analyse des données de projections

Dans cette section, les données de projections sont présentées. Dans un premier temps, celles relatives à la population vivante sont introduites. Ensuite, les projections de températures journalières par scénario climatique sont analysées. L'objectif de cette partie est d'obtenir de premières intuitions quant à la manière dont la mortalité future pourrait être différenciée selon les scénarios climatiques à horizon 2070.

4.1.1 Présentation des données de projections de la population vivante

Les données de projections de la population vivante sont issues du modèle Omphale 2022¹. Ce dernier permet de réaliser des projections de population infra-nationales jusqu'à 2070, sur toute zone de population contenant plus de 50 000 habitants. Il se base sur les données issues du recensement de la population de 2018. Dans la version gratuite, seul le scénario central est disponible. Les données utilisées sont issues de la pyramide des âges fournie par

1. Lien vers une description du modèle Omphale : [↗](#)

l'INSEE pour chaque département et région². Le modèle Omphage suppose que les quotients démographiques locaux évoluent de manière similaire au niveau national, en prolongeant les tendances de fécondité et de mortalité. Autrement dit, ces scénarios ne tiennent pas compte des conséquences potentielles du changement climatique sur la mortalité. Les quotients de migrations internes sont maintenus constants, et le solde migratoire avec l'étranger projeté est réparti sur l'ensemble des départements. Un résumé de l'ensemble des scénarios relatifs à l'évolution de la population est disponible dans le tableau 4.1. Il est issu de l'institut Prodego, un des fournisseurs de données de l'INSEE pour les demandes spécifiques³.

<i>Hypothèses</i>	2020	Centrale	Basse	Haute
<i>Fécondité</i>				
Nombre d'enfants par femme	1,83	1,80	1,60	2,00
Âge moyen à la maternité (en années)	30,8	33,0	33,0	33,0
<i>Espérance de vie à la naissance</i>				
Femmes (en années)	85,1	90,0 (2070)	86,5 (2070)	93,5 (2070)
Hommes (en années)	79,1	87,5 (2070)	84,0 (2070)	91,0 (2070)
<i>Migrations</i>				
Valeur du solde migratoire annuel	+ 70 000	+ 70 000	+ 20 000	+ 12 000

TABLEAU 4.1 – Scénarios démographiques présentés dans le modèle Omphage 2022 (source : Prodego-Adisp)

Il est possible de remarquer que les espérances de vie à la naissance projetées dans le scénario standard sont proches des projections des Nations Unies déjà présentées dans le graphique 2.9. Les autres scénarios proposent des évolutions différenciées de l'indice de fécondité, de l'espérance de vie ainsi que du solde migratoire. Notons que même dans le scénario d'évolution basse, l'espérance de vie à la naissance progresse de manière significative, notamment chez les hommes. Pour rappel, les données de population vivante seront incluses en exposition dans le modèle, ce qui permettra de convertir les nombres de décès modélisés en des taux de mortalité. L'étude de Pincemin [47] proposait elle une méthodologie visant à projeter les chocs de mortalité par rapport à une table prospective de référence issue de l'INSEE.

4.1.2 Présentation des données de projections des températures

Les données de températures utilisées dans cette étude prospective sont issues du portail DRIAS⁴. Celui-ci permet une mise à disposition publique des projections climatiques régio-

2. Lien vers la pyramide des âges : [☞](#)

3. Lien vers l'analyse des différences de projections démographiques entre scénarios : [☞](#)

4. Lien vers le portail DRIAS (un compte gratuit est nécessaire pour télécharger les projections) : [☞](#)

nales réalisées dans des laboratoires français de modélisation du climat (IPSL, CERFACS, CNRM). Il s'agit donc d'une référence utile pour des analyses prospectives à l'échelle française. La manière dont les projections sont construites, pour un scénario climatique donné, est synthétisée dans le schéma 4.1.

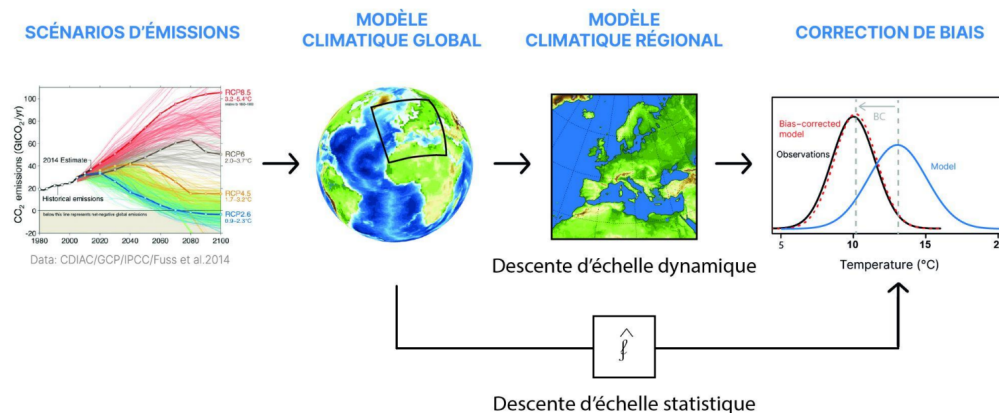
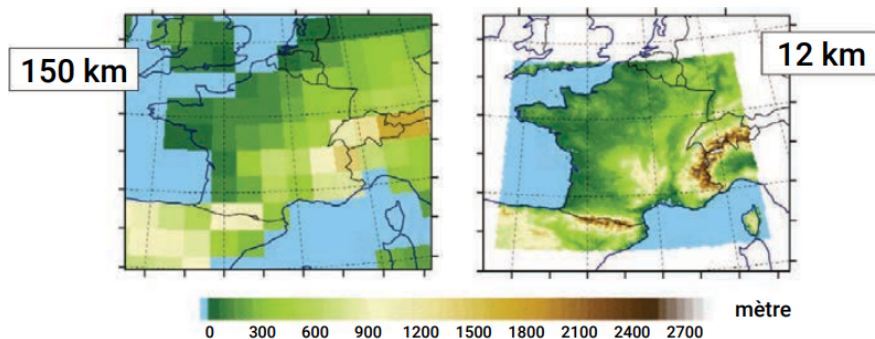


SCHÉMA 4.1 – Étapes d'obtention des projections de variables climatiques dans l'ensemble DRIAS-2020 (source : DRIAS [17])

Tout d'abord, des modèles climatiques généraux (GCM) sont retenus comme point de départ. Il s'agit de représentations numériques simplifiées de l'atmosphère en fonction du temps. Les conditions initiales sont notamment fournies par les historiques de températures, de pression atmosphérique et d'ampleur du vent. Des équations issues de lois de conservation de la physique sont ensuite appliquées, et permettent de définir la manière dont les scénarios d'émissions de gaz à effet de serre vont influencer sur les variables climatiques. Ces dernières vont pouvoir, enfin, être projetées sur l'ensemble de la planète, à une granularité spatiale, une maille temporelle et un horizon fixés. Pour l'ensemble DRIAS-2020, 34 GCM ont initialement été retenus. Parmi eux, les six meilleurs modèles serviront à "forcer" les prédictions des autres modèles. Un processus de décision a été proposé par McSweeney et al. [44], conduisant d'abord à exclure les modèles fournissant des simulations absurdes du système climatique (par exemple, lorsque les températures annuelles sont jugées irréalistes), puis à retenir les simulations considérées comme les plus réalistes en Europe.

Ces modèles généraux sont utilisés en tant que conditions aux bornes afin de construire des modèles climatiques régionaux (RCM). D'après le rapport du DRIAS [17] relatif à l'ensemble DRIAS-2020, la construction des modèles régionaux est très similaire à celle des GCM mais tient compte des spécificités géographiques locales (reliefs, contrastes terre-mer). Comme l'indique le graphique 4.2, leur objectif est, en affinant la résolution géographique dans la modélisation, d'offrir des simulations climatiques à l'échelle européenne. La résolution obtenue est ainsi divisée par plus de 12 dans les RCM par rapport aux GCM.



GRAPHIQUE 4.2 – Réduction d'échelle dans le modèle RCM en France (source : DRIAS [17])

Les RCM utilisés sont issus du programme Euro-Cordex, initié en 2009 sous l'impulsion du Programme Mondial de Recherche sur le Climat (PMRC) afin de fournir un cadre pour les projections climatiques et l'évaluation des modèles, ainsi qu'une interface pour les utilisateurs. Les projections issues d'Euro-Cordex constituent, d'après le rapport du DRIAS, une référence de la modélisation climatique régionale à l'heure actuelle. Des défauts significatifs ont toutefois été identifiés. D'une part, les prédictions fournies seraient généralement trop froides, et pourraient donc conduire à sous-estimer la surmortalité future liée aux vagues de chaleur. D'autre part, certains phénomènes extrêmes seraient encore mal reproduits. Ces défauts motivent une correction des simulations climatiques.

La méthode ADAMONT a ainsi été appliquée afin de corriger certains biais issus des RCM. Elle utilise un procédé d'ajustement des quantiles⁵, afin de ramener la distribution statistique des données simulées à celle des valeurs observées, pour chaque variable climatique, saison et période. Cette méthode a notamment pour intérêt de mieux reproduire les phénomènes extrêmes à une échelle locale, grâce à l'utilisation d'observations historiques. Celles-ci sont issues du jeu de données SAFRAN, disponibles en France métropolitaine et à une résolution de 8 km.

Parmi les différentes projections proposées, le choix s'est porté sur le modèle ALADIN63 issu du laboratoire CNRM. Ce modèle possède l'avantage de fournir des prédictions pour chaque scénario climatique, centrales par rapport à l'ensemble des simulations possibles, et issues d'un laboratoire français. Le raisonnement ayant conduit au choix de ce modèle est détaillé dans l'annexe I. Même si la construction du modèle décrite ci-dessus vise à limiter au mieux l'incertitude des différentes projections, celle-ci conduit inévitablement à relativiser la portée des résultats qui seront obtenus dans la suite de l'étude. En effet, chaque projection du DRIAS souffre de plusieurs sources d'incertitude, liées tout d'abord à la pertinence du modèle et des méthodes de correction. Les scénarios RCP utilisés ayant été construits à partir de la

5. Cette méthode est décrite dans cet article du DRIAS : ↗

connaissance scientifique antérieure à 2013⁶, il aurait également été intéressant de travailler à partir des scénarios les plus récents (SSP) et basés sur les observations des dernières années.

Par la suite, les modèles estimés à partir des données historiques permettront de fournir des prédictions de la mortalité grâce aux variables explicatives (températures moyennes journalières et population vivante) projetées à horizon 2070 pour chaque scénario RCP.

4.1.3 Analyse de l'évolution des températures à horizon 2070

Avant de projeter la mortalité par scénario climatique, il est utile d'effectuer une étude descriptive des données de projection des températures. Cette analyse a pour intérêt de rendre compte des différences de trajectoires des températures par scénario climatique, et de construire une première intuition des conséquences potentielles sur la mortalité.

En particulier, l'existence potentielle d'épisodes exceptionnels sera étudiée. La surmortalité qui pourra en découler se traduisant par une potentielle sortie massive de capitaux (pour les garanties en cas de décès), l'étude de ces pics extrêmes possède un intérêt considérable pour un assureur. Plusieurs approches descriptives ont paru intéressantes.

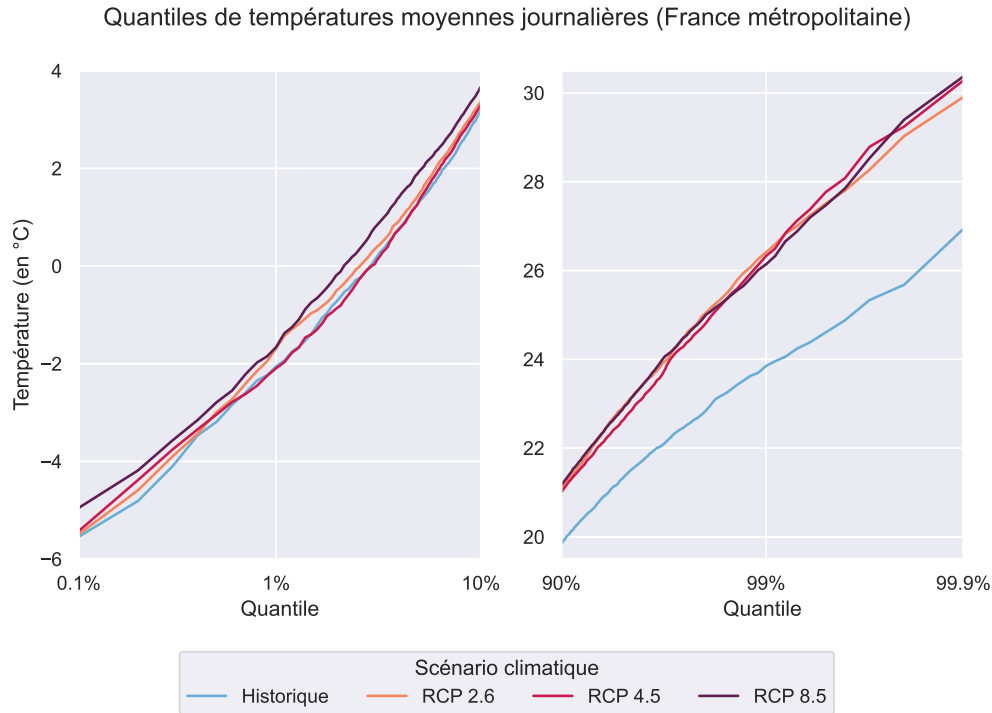
- L'étude des **quantiles** de températures aura pour objectif de discuter de l'ampleur de l'évolution générale des températures ainsi que des pics, selon le scénario climatique.
- La **comparaison géographique des chaleurs extrêmes** aura pour but d'identifier les régions potentiellement plus exposées aux vagues de chaleur futures, et aux risques de surmortalité associés.
- La **fréquence et l'intensité des vagues de chaleur futures** sera analysée pour chaque scénario climatique afin de conjecturer des pics de mortalité potentiels futurs attribuables à la chaleur. Une étude similaire relative aux épisodes de froid futurs est disponible en annexe K.
- Une **étude de cas de l'année 2065**, basée sur l'analyse précédente, permettra une description plus précise de la nature des épisodes extrêmes projetés à horizon 2070.

Quantiles de températures

Tout d'abord, une représentation des quantiles de températures permet de comparer l'évolution générale de ces dernières par rapport à la période historique retenue (1990-2019). L'étude des quantiles extrêmes permet, en particulier, de discuter de l'impact des scénarios

6. Ce point a été développé dans la partie 1.1.1

climatiques sur les pics de froid et de chaleur. Un indice thermique national, c'est-à-dire la moyenne des températures observées en France métropolitaine pour un jour donné, sera utilisé. Une représentation des quantiles de cet indice est effectuée dans le graphique 4.3.



GRAPHIQUE 4.3 – Quantiles de températures moyennes journalières en France métropolitaine par scénario climatique (2020-2070) et comparaison avec la période historique (1990-2019)

D'une part, les différences entre les températures historiques et projetées se concentrent dans les niveaux de chaleur les plus rares. Alors que la température moyenne journalière n'a jamais excédé 27°C entre 1990 et 2020, cette valeur devrait être atteinte environ dix fois plus souvent à l'horizon 2070. Cela se vérifierait quel que soit le scénario RCP retenu. D'autre part, les écarts entre les derniers seraient surtout visibles sur les queues de distribution. L'échelle logarithmique choisie sur l'axe des abscisses permet d'avoir une meilleure visibilité des variations de températures extrêmes entre les scénarios. Celles-ci sont de l'ordre de quelques dixièmes de degré pour chaque extrémité climatique. Comme cela pouvait être attendu, le scénario *business-as-usual* serait caractérisé par les températures les plus élevées, alors que les pics de froid serait légèrement moins intenses. Inversement, la mise en place de mesures limitant l'ampleur du réchauffement climatique permettrait de limiter l'intensité des vagues de chaleur les plus intenses. Cependant, les épisodes de froid deviendraient à leur tour plus marqués. Les périodes de températures fraîches et chaudes étant chacune caractérisées par davantage de décès, il ne semble pas possible de déceler, à ce stade, un scénario de forçage radiatif ayant des conséquences plus graves sur la mortalité.

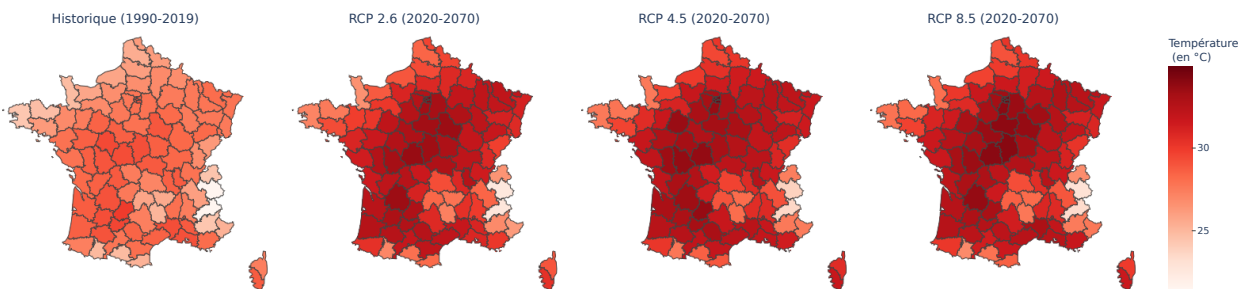
Cette analyse sera combinée à une représentation géographique des températures extrêmes, permettant de suggérer des surmortalités différenciées par pics de chaleur et de froid.

Comparaison géographique des chaleurs extrêmes

La dimension géographique intégrée à cette analyse a permis de capter l'acclimatation des individus *dans l'espace*, c'est-à-dire leur capacité à s'adapter aux conditions géographiques propres à leur lieu de vie. Il a ainsi été vu que les effets de la chaleur sur la mortalité sont généralement moins importants dans la zone méditerranéenne que dans les autres climats, pour une classe d'âge donnée. Compte tenu de cette séparation, il est intéressant de se demander si des zones moins sujettes à des pics de chaleur aujourd'hui vont voir leur nombre d'épisodes extrêmes augmenter dans les prochaines décennies.

Il a été vu dans l'analyse des quantiles de températures que la différenciation des températures entre scénarios, concernant les pics de chaleur, est principalement perceptible dans les queues de distribution. Les quantiles de température moyenne journalière à 99,9% ont donc été représentés pour chaque scénario climatique. Autrement dit, il s'agit de chaleurs exceptionnelles, observées dans chaque département une fois sur 1 000 jours. Les résultats sont affichés dans le graphique 4.4.

Quantile 99,9% des températures moyennes journalières par scénario RCP



GRAPHIQUE 4.4 – Quantiles de température moyenne journalière à 99,9% par scénario climatique en France métropolitaine d'après les projections CNRM sur la période 2020-2070

Comme attendu, les observations sont similaires à celles obtenues sur l'analyse des quantiles. Entre la période historique et les projections entre 2020 et 2070, les températures extrêmes correspondant au quantile 99,9% augmentent en moyenne de 3 à 4 degrés selon le scénario. La hausse est moins marquée pour le scénario RCP 2.6 que pour les deux autres, pour lesquels les différences sont minces. Ce résultat peut s'expliquer par la définition des scénarios et de la période de projection : alors que le scénario RCP 2.6 suppose la mise en œuvre de mesures dès 2020, la stabilisation des émissions de GES est supposée débuter vers

le milieu du siècle pour le scénario RCP 4.5. Les projections de températures des scénarios RCP 4.5 et 8.5 sont donc proches entre 2020 et 2050, et commencent seulement à diverger après 2050, sachant que la matérialisation de ces politiques est sujette à une certaine latence.

Par ailleurs, les zones climatiques pour lesquelles les températures extrêmes sont les plus élevées (comparativement aux autres) sont similaires, avant et après 2020. Il s'agit principalement des régions Centre-Val de Loire et Occitanie. Au contraire, les départements montagneux et (dans une moindre mesure) ceux en bordure de mer continueraient à observer des pics de température moins élevés que dans le reste du territoire. Le cas de l'Île-de-France est toutefois une exception : les modèles y prédisent des pics de températures parmi les plus élevés à l'échelle nationale dans les prochaines décennies, ce qui n'était pas observé entre 1990 et 2020. Il est à noter que ces prédictions sont sujettes à de l'aléa : en juillet 2022, le Finistère a connu un épisode de chaleur exceptionnel, avec des températures moyennes journalières avoisinant les 30°C (les maximales atteignant 40°C). Sachant que l'historique retenu dans cette étude est limité à l'année 2019, les modèles estimés ici n'intègrent pas ces derniers pics observés.

Fréquence et intensité des vagues de chaleur futures

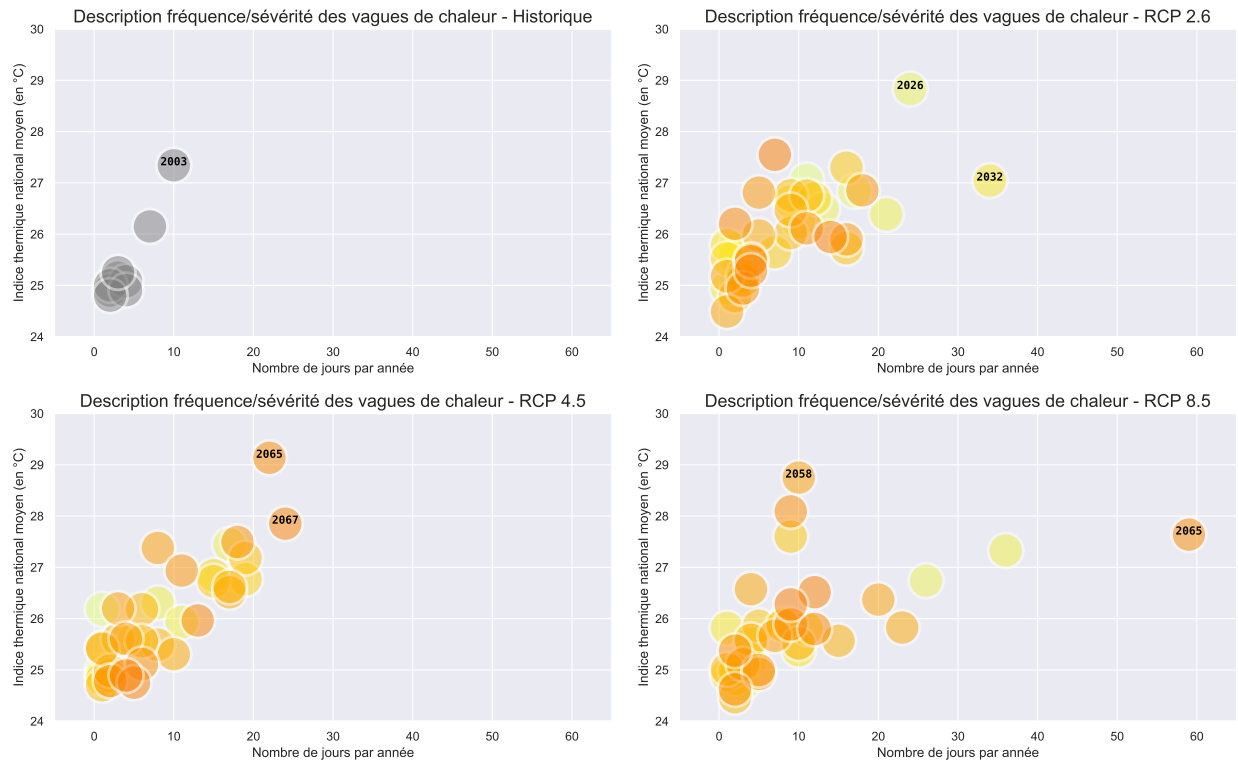
D'après les résultats prédits par le modèle DLNM, la surmortalité attribuable à la chaleur évoluerait de manière exponentielle à mesure que la température observée s'éloigne de l'optimum thermique. Pour rappel, les individus âgés et de sexe féminin seraient les plus affectés par les conséquences des températures élevées.

Afin de conjecturer les impacts du dérèglement climatique pour ces catégories d'individus, une métrique relative à la fréquence et à l'intensité des épisodes de chaleur a dû être sélectionnée. Pour cela, le choix s'est porté sur la définition d'une vague de chaleur selon Météo-France. Une vague de chaleur est définie par un épisode d'au moins 3 jours, durant lequel :

- l'indice thermique ne descend jamais en-dessous de 23,4°C ;
- l'indice thermique atteint au moins une fois 25,3°C durant les 3 jours.

Afin de mesurer l'intensité des vagues de chaleur futures, la moyenne des indicateurs thermiques durant les trois jours d'une vague de chaleur est utilisée. S'il y en a plusieurs au cours d'une année, la moyenne des intensités est considérée. L'objectif de cette approche est surtout de déceler une tendance quant à la gravité de ces phénomènes de chaleur. Les nombres de vagues de chaleur par année ainsi que les intensités moyennes associées sont représentés

dans le graphique 4.5 pour chaque scénario climatique et comparés avec la période historique 1990-2019.



GRAPHIQUE 4.5 – Nombre de jours avec vagues de chaleur et intensité moyenne par année en France métropolitaine durant la période 1990-2019 et par scénario RCP (2020-2070)

D'une part, concernant l'aspect fréquence, le graphique met en évidence la probable multiplication de tels épisodes dans les prochaines décennies. Alors que les années sans vagues de chaleur étaient majoritaires sur la période de référence pour la modélisation, la situation pourrait s'inverser avant la fin du siècle. Ainsi, pour les scénarios de réchauffement les plus intenses (RCP 4.5 et 8.5), il y aurait au moins une vague de chaleur par an entre 2050 et 2070. De plus, au cours d'une même année, les épisodes pourraient être particulièrement longs et/ou répétés. Quel que soit le scénario retenu pour l'étude prospective, les nombres de vagues de chaleur par an seraient sans commune mesure avec les statistiques historiques. D'ici une dizaine d'années, il serait possible d'observer une trentaine de journées associées à des vagues de chaleur par an, peu importe le scénario climatique. En 2065, et en l'absence de mesure compensant les effets du réchauffement climatique, les vagues de chaleur pourraient même être réparties sur deux mois environ. Ces projections inquiétantes ont été confirmées par Michel Schneider, climatologue à Météo-France. Ce dernier prévient que, à la fin du siècle, les pires vagues de chaleur pourraient durer plus de 90 jours, dans le scénario le plus

pessimiste⁷. Les projections utilisées ici restent toutefois à considérer avec un certain recul, étant donné leur large amplitude temporelle.

Il est toutefois à noter que la métrique choisie ne permet pas de distinguer les phénomènes courts et répétés d'une part, et les vagues allongées d'autre part. Cette distinction est pourtant potentiellement importante, car un épisode prolongé de chaleurs intenses pourrait avoir des conséquences plus graves. Par exemple, des professionnels de santé ont admis⁸ que les services hospitaliers pourraient se retrouver en difficulté en cas d'une canicule longue de plus de cinq jours. Par ailleurs, la vague de canicule d'août 2003, qui était la plus mortelle sur la période historique, était caractérisée à la fois par une longueur et une intensité importantes. Cette dernière information est donc importante à prendre en compte afin de conjecturer les conséquences potentielles de ces épisodes sur la mortalité.

D'autre part, en ce qui concerne l'aspect de sévérité, le graphique confirme l'intensité exceptionnelle atteinte lors de la canicule de 2003. Sur la période de projections, de tels niveaux de chaleur sur une durée prolongée devraient être rares. Ils pourraient toutefois être dépassés à quelques occasions. Notamment, un épisode très intense pourrait être connu dès 2026, dans le scénario de projection avec des mesures d'adaptations les plus fortes. Cela semble contre-intuitif, mais le scénario RCP 2.6 suppose, pour rappel, un ralentissement des émissions de gaz à effet de serre à partir de 2020. En 2026, les effets de ces mesures pourraient ne pas être encore entièrement visibles, conduisant à des projections climatiques assez proches des autres scénarios. Des différenciations marquées sont, en revanche, observables à partir de la seconde partie du siècle. Dans les scénarios de réchauffement les plus forts, les intensités des vagues de chaleur pourraient être amenées à dépasser régulièrement les niveaux de 2003.

Cette augmentation conjointe de la durée et de l'intensité des épisodes de chaleur intense pourraient avoir des conséquences significatives sur le nombre d'hospitalisations et la mortalité d'ici la fin du siècle. Une analyse similaire a été menée sur les épisodes de froid futurs et a, cette fois, mis en avant une fréquence et une intensité comparables aux niveaux de la période 1990-2019. Cette étude est disponible en annexe K.

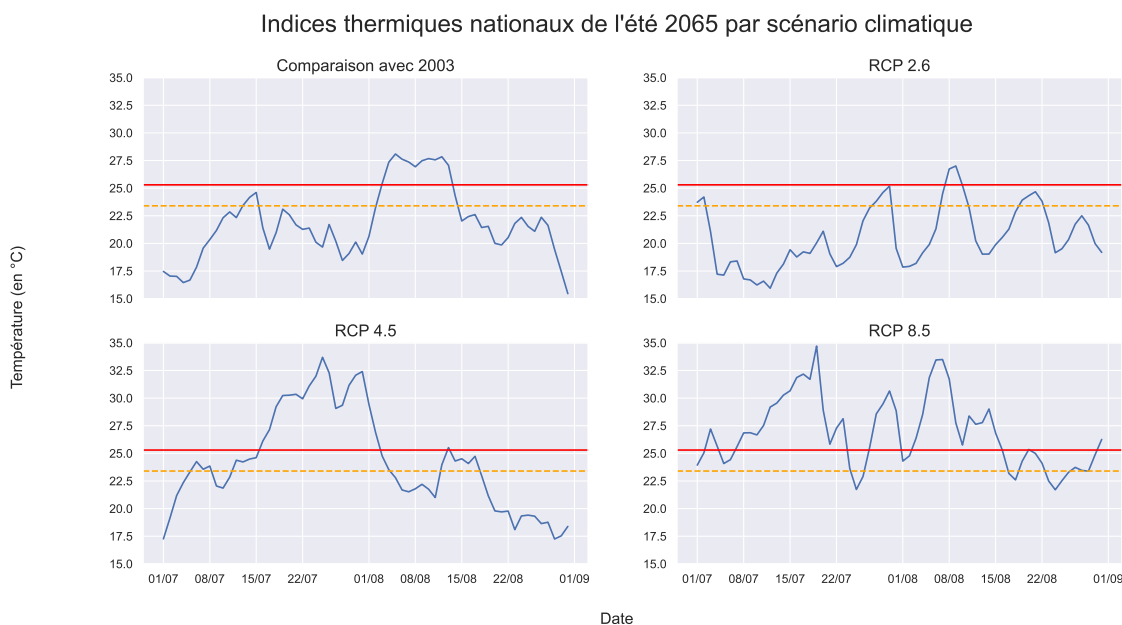
Étude de cas : en 2065, des épisodes de chaleur simulés remarquables

Les résultats de l'analyse précédente motivent une étude spécifique de l'année 2065, marquée par des vagues de chaleur projetées parmi les plus remarquables à horizon de 50 ans pour les scénarios RCP 4.5 et 8.5. Les indices thermiques nationaux obtenus pour chaque

7. Cité par "Le Monde" : ☞

8. Lien vers l'article issu du journal "Les Échos" : ☞

scénario lors de l'été 2065 sont représentés dans le graphique 4.6. À titre de comparaison, les températures observées durant la vague de chaleur de 2003 ont également été considérées.



GRAPHIQUE 4.6 – Indices thermiques nationaux lors de l'été 2003 (en haut à gauche), et prédits en 2065 par scénario RCP

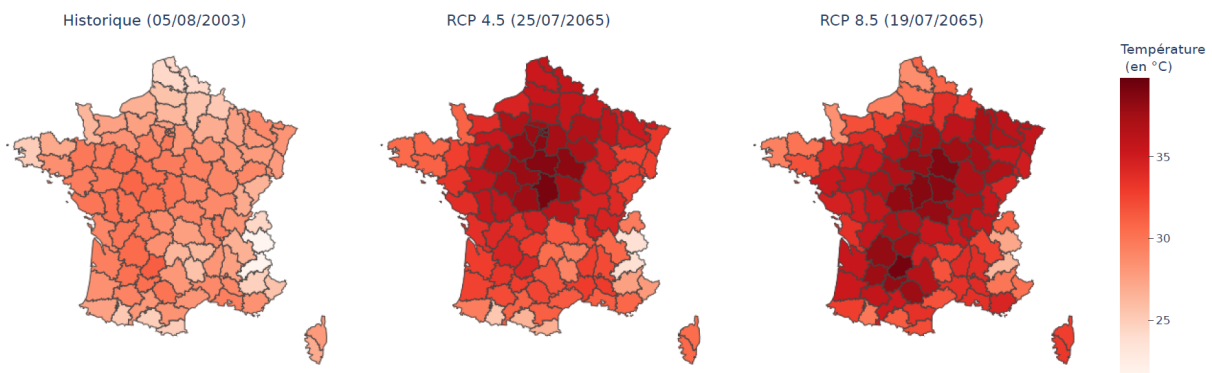
Les lignes orange et rouge correspondent aux seuils ($23,4^{\circ}\text{C}$ et $25,3^{\circ}\text{C}$) fixés par Météo-France pour la définition d'une vague de chaleur. Les sorties obtenues confirment l'augmentation de l'intensité et de la durée des épisodes de canicule futurs, et la différenciation des scénarios climatiques. D'une part, dans le scénario le plus favorable, l'été 2065 serait relativement ordinaire, avec une unique vague de chaleur de quatre jours. D'autre part, les valeurs extrémales de l'indice thermique seraient plus élevées dans le scénario climatique sans politique d'adaptation. Les températures moyennes extérieures pourraient, alors, être de l'ordre de cinq à six degrés supérieures à celles observées au plus fort de la canicule de 2003. Bien que l'acclimatation devrait permettre à l'avenir une meilleure protection des organismes face à la chaleur, de telles chaleurs pourraient décupler les risques de mortalité. Pour rappel, d'après les théoriciens de la température du bulbe mouillé, une exposition prolongée à des températures trop élevées pourrait, à moins que l'air soit peu saturé, empêcher la transpiration et ne pas être supportée par le corps humain.

De plus, cette augmentation de l'intensité de ces épisodes de chaleur futurs s'accompagnerait d'une hausse de leur durée. Entre les mois de juillet et de septembre, il serait ainsi possible de connaître plusieurs vagues de chaleur distinctes, chacune pouvant être d'une durée comprise entre une dizaine de jours et un mois. Par exemple, alors que l'épisode de 2003 a

été (selon la méthodologie retenue ici) d'une durée de 10 jours, celui de 2065, dans le scénario RCP 4.5, pourrait persister entre la deuxième semaine de juillet et la deuxième semaine d'août. Une telle longueur pourrait conduire à des effets collatéraux amplifiant les conséquences potentielles sur la mortalité. Parmi celles-ci, les hôpitaux pourraient être débordés (réduisant l'efficacité des soins) et des coupures d'électricité pourraient survenir, empêchant notamment l'usage de climatiseurs. Dans le cas du scénario RCP 8.5, une courte période d'accalmie serait observée entre deux épisodes, se matérialisant par une baisse temporaire des températures en-dessous des seuils de canicule.

L'analyse des différences de durée et d'intensité de ces phénomènes projetés peut être complétée par une étude géographique. Afin de simplifier cette dernière, les températures moyennes journalières observées au moment du pic de chaleur dans chaque épisode étudié sont représentées par département dans le graphique 4.7.

Températures moyennes journalières extrêmes de l'été 2065 par scénario climatique



GRAPHIQUE 4.7 – Températures moyennes journalières observées au plus fort des vagues de chaleur de 2003, puis projetées en 2065 (selon les scénarios RCP 4.5 et 8.5)

De manière générale, même si la chaleur est généralisée sur l'ensemble du territoire au cours de ces épisodes, les régions Île-de-France ainsi que Centre-Val de Loire pourraient être particulièrement touchées. Ces régions avaient déjà été identifiées dans l'analyse des quantiles des températures extrêmes projetées. Dans le scénario intermédiaire, les températures moyennes les plus hautes pourraient atteindre des valeurs voisines de 40 degrés, à Paris et dans les départements voisins (les Hauts-de-Seine et la Seine-et-Marne, notamment). Étant donné la relation croissante et convexe prédite entre la température et la mortalité au-delà de l'optimum thermique, ces zones géographiques devront donc faire l'objet d'une attention particulière au moment de considérer les projections de mortalité. Par ailleurs, du fait de la forte concentration de population dans la région Île-de-France (plus de 12 millions d'habitants en 2023 d'après l'INSEE, soit la région la plus peuplée⁹), les impacts sur les taux de mortalité

9. D'après les recensements de population de l'INSEE : ↗

nationaux pourraient être d'autant plus visibles. Les différences de décès entre les régions pourraient être d'autant plus marquées pour les catégories à risque identifiées dans la partie précédente, à savoir les femmes et les personnes âgées. Il convient toutefois de rappeler que cette partie ne se veut pas exhaustive : d'autres épisodes de chaleur pouvant avoir un impact significatif sur la mortalité ont été identifiés, quel que soit le scénario de projection climatique retenu. Le lecteur intéressé trouvera une étude spécifique de l'année 2026 en annexe J.

Après avoir conjecturé les évolutions des nombres de décès associés à la chaleur, les conséquences possibles des épisodes de froid futurs sont maintenant abordées.

4.2 Projections de la mortalité à horizon 2070

4.2.1 Présentation de la démarche retenue

L'intérêt d'une projection paramétrique des taux de mortalité

L'objectif de cette partie est de proposer des prédictions de la mortalité future en évaluant la matérialité des conséquences du dérèglement climatique. Cette problématique conduit à modifier la démarche habituelle. En effet, afin d'obtenir une table de mortalité prospective, il est courant d'adopter une méthodologie en deux temps :

1. Obtenir les taux de mortalité bruts $\mu_x(t)$ associés à un âge x et à une année t ;
2. Pour un âge x donné, estimer les taux sur les données historiques grâce à un modèle puis les extrapoler.

Autrement dit, il est supposé que les taux de mortalité bruts sont conformes aux modèles de Lee-Carter ou de Gompertz, par exemple. Ceux-ci étant non-paramétriques, l'intégration du risque climatique est délicate, car il n'est pas possible d'inclure directement les données de projection (par exemple de températures) dans la prédiction de la mortalité future. Si ces modèles peuvent être cohérents dans le cas d'une mortalité amenée à suivre sa tendance historique, ils ne permettent pas de constater l'impact futur de variables ayant eu une influence négligeable sur la mortalité dans le passé. Des projections nationales de la mortalité à horizon 2070 sont proposées par l'INSEE et ont été décrites par Planchet et al. [50]. Elles s'appuient sur divers avis d'experts, la comparaison des sorties d'un ensemble de modèles actuariels couramment utilisés (Lee-Carter, approche par générations, etc.), et reposent sur l'hypothèse d'un prolongement de la tendance observée sur la dernière décennie. Pour ces raisons, il s'agit d'une base de référence pour une étude prospective. De ce fait, les projections de l'INSEE

ne tiennent pas compte de l'évolution des épisodes de températures extrêmes (en termes de longueur et de sévérité), et de leurs conséquences potentielles sur la mortalité future. Afin de mesurer la matérialité du changement climatique sur la mortalité, il serait donc intéressant de comparer les taux de décès prédits par le modèle retenu avec ceux obtenus par l'INSEE.

Afin d'intégrer les conséquences potentielles du changement climatique sur la mortalité, des extensions paramétriques du modèle de Lee-Carter ont été proposées dans des travaux actuariels récents. Germain [33] ou Fall [23] ont ainsi inclus une variable supplémentaire associée au risque climatique (décès attribuables à un facteur de risque, ou écarts de température par rapport à un seuil fixé). Les impacts des vagues de chaleur projetées pour chaque scénario climatique sont notamment pris en compte. Toutefois, cette démarche ne tient pas compte des évolutions relatives aux épisodes de froid, et ne permet pas d'effectuer une analyse prospective départementale. Dans ce mémoire, il a été choisi d'extrapoler les relations obtenues pour chaque département à partir des modèles estimés par zone climatique, tranche d'âge et sexe entre 1990 et 2019. Plusieurs critères, proposés par Petauton [27], permettront de discuter de la qualité des projections :

- (i) Les extrapolations obtenues doivent être cohérentes (ce point pourra être vérifié à partir de plusieurs contrôles : une comparaison des résultats obtenus avec la littérature, ou bien en s'assurant que les nombres de décès modélisés à long terme restent positifs) ;
- (ii) Les quotients de mortalité estimés doivent être inférieurs à 1 (autrement dit, le nombre de décès pour une année et ne peut être supérieur à la population vivante) ;
- (iii) Globalement, les erreurs doivent rester limitées.

Par construction du modèle DLNM et de l'hypothèse de loi retenue, les nombres de décès prédits resteront positifs, et ce quelle que soit la profondeur de la projection effectuée. Cela participe à la cohérence des projections, et donc à la vérification du point (i). En revanche, ces vérifications devront être approfondies à l'aide d'une analyse des taux de mortalité futurs prédits. L'utilisation d'une segmentation par tranches d'âge possède l'avantage de permettre une étude graphique des sorties obtenues : pour chaque catégorie d'âge fixée, les prédictions pourront être comparées à celle d'une table prospective de référence (ici, celle de l'INSEE).

Présentation des hypothèses utilisées pour la projection

En raison de l'incertitude relative à la longévité ainsi qu'à l'adaptation future des populations aux températures, plusieurs scénarios de projection sont retenus par la suite. Ceux-ci sont synthétisés dans le tableau 4.2.

<i>Modèle</i>	(a) Modèle initial	(b) Tendance ralentie	(c) Tendance ralentie et adaptation
Tendance historique retenue	1990-2019	2010-2019	2010-2019
Adaptation	Non	Non	Oui

TABLEAU 4.2 – Description des modèles retenus pour la projection

Alors que le premier modèle est celui décrit dans la section 3.4, deux alternatives sont proposées en vue d’une analyse prospective. La première repose sur le prolongement de la tendance de la dernière décennie, conformément au jugement d’expert retenu dans la dernière table prospective nationale de l’INSEE. Utiliser la tendance observée sur la période 2010-2019 plutôt que celle des trente dernières années supposera une évolution moins favorable de la mortalité. Le prolongement de la tendance 2010-2019 est permis par l’inclusion de variables de tendance différenciées, construites avant et après 2019. Il sera utilisé dans les modèles (b) et (c).

L’adaptation, quant à elle, reposera sur une hypothèse simplificatrice : Vicedo-Cabrera et al. [59] proposent ainsi de supposer une diminution des risques relatifs de mortalité au-delà de l’optimum thermique. L’objectif de cette approche est de tenir compte des changements de relations températures-mortalité après la période historique sur laquelle les modèles sont estimés, tout en retenant des hypothèses simplifiées. Il n’y a, par exemple, pas de modification de l’optimum thermique. Le niveau de réduction des risques relatifs donné dans cet article est de 30%. Ce niveau de réduction a été obtenu par De’Donato et al. [15] sur les données de mortalité de plusieurs villes italiennes, suite à diverses mesures de prévention adoptées face aux épisodes de chaleur dans les années 2000. Ce seuil de 30% reste toutefois donné à titre indicatif. La présente étude propose de l’incorporer dans les prédictions de la mortalité de manière suivante. Soit $s \in [0 ; 1]$ le taux de réduction des risques relatifs de mortalité, correspondant au niveau d’adaptation (ici, $s = 0,3$). Soient $\hat{y}(z, \cdot)$ la prédiction initiale de la mortalité associée à une observation, avec z la température moyenne journalière et \cdot le reste des variables explicatives. Soient également z^* l’optimum thermique du modèle, et $RR(z)$ le risque relatif de mortalité associé à la température moyenne journalière z . Alors, la nouvelle prédiction \hat{y}^{ad} peut être obtenue en plusieurs étapes, synthétisées dans le schéma 4.8.

D’une part, en utilisant cette méthode, les prédictions relatives au froid (lorsque les températures sont inférieures à l’optimum thermique) restent inchangées. D’autre part, afin de calculer les nombres de décès associés aux températures chaudes (supérieures à l’optimum thermique), le risque relatif initial est utilisé dans un premier temps, ce qui permet d’obtenir les prédictions associées à l’optimum thermique (toutes choses égales par ailleurs). Ensuite,

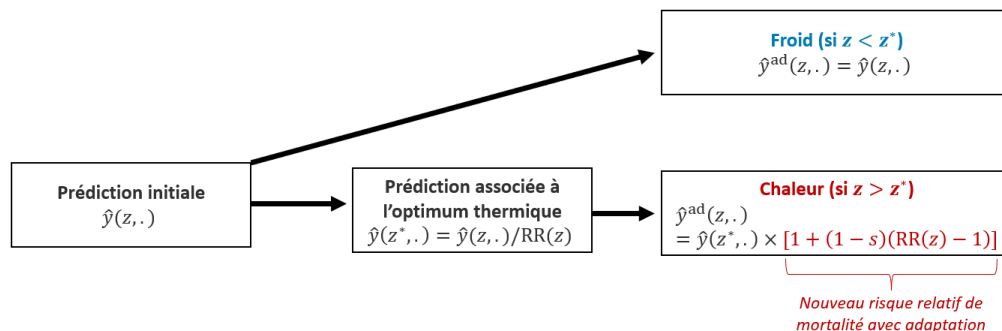


SCHÉMA 4.8 – Méthode d’obtention des nouvelles prédictions de la mortalité avec prise en compte d’un niveau d’adaptation à la chaleur

les nouvelles prédictions avec adaptation sont obtenues en réduisant la surmortalité (définie ici comme la partie du risque relatif supérieure à 1) de 30%. Il est à noter que, tout comme dans l’article de Vicedo-Cabrera et al. [59], ce seuil est supposé invariant dans le temps. Autrement dit, les nouvelles relations estimées entre les températures et la mortalité sont supposées constantes sur la période de projection. Bien que cette hypothèse soit irréaliste, elle permet d’obtenir un indicateur simple des bénéfices potentiels associés aux mesures d’adaptation. Par ailleurs, cette méthode ne tient pas compte de l’acclimatation future, c’est-à-dire de la translation de la courbure température-mortalité vers des niveaux de températures plus élevées. Cependant, en gardant à l’esprit que l’acclimatation future se traduit également par une moindre sensibilité aux chaleurs extrêmes, fixer un seuil d’adaptation suffisamment élevé pourrait permettre de capter implicitement les effets de l’acclimatation. La méthodologie présentée ici sera utilisée dans le modèle (c).

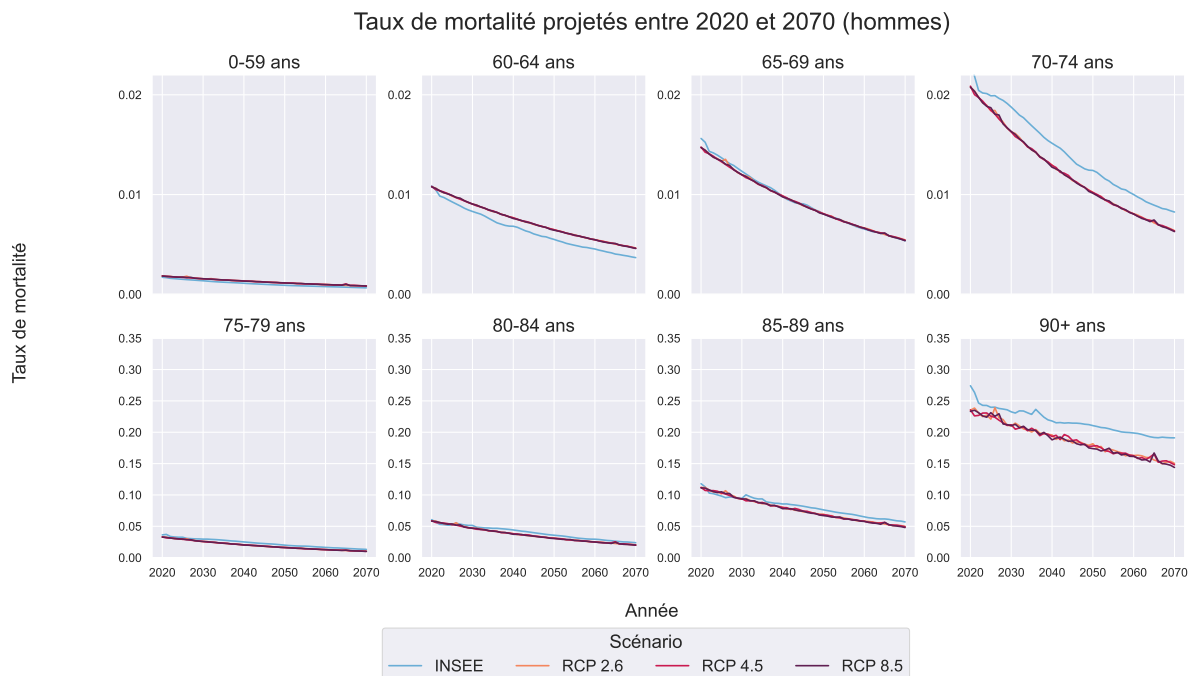
4.2.2 Comparaison temporelle de la mortalité prospective

Dans cette partie, les taux de mortalité estimés sur la période 1990-2019 pour chaque tranche d’âge, sexe et zone climatique sont projetés à horizon 2070 en utilisant les projections démographiques et climatiques présentées précédemment. Ils sont ensuite agrégés sur l’ensemble de la France métropolitaine afin d’être rapprochés aux projections de mortalité fournies par l’INSEE. Il est à noter, cependant, que ces dernières reposent sur des données de la France entière, incluant les départements d’Outre-Mer. Un biais peut donc apparaître du fait de cette différence dans le champ géographique retenu. Toutefois, dans la pratique, la mortalité observée dans les territoires ultra-marins est relativement similaire à celle des départements métropolitains¹⁰, ce qui devrait limiter l’écart systématique entre les approches.

10. Voir par exemple cette analyse géographique des taux de mortalité bruts en 2022 par l’INSEE : ↗

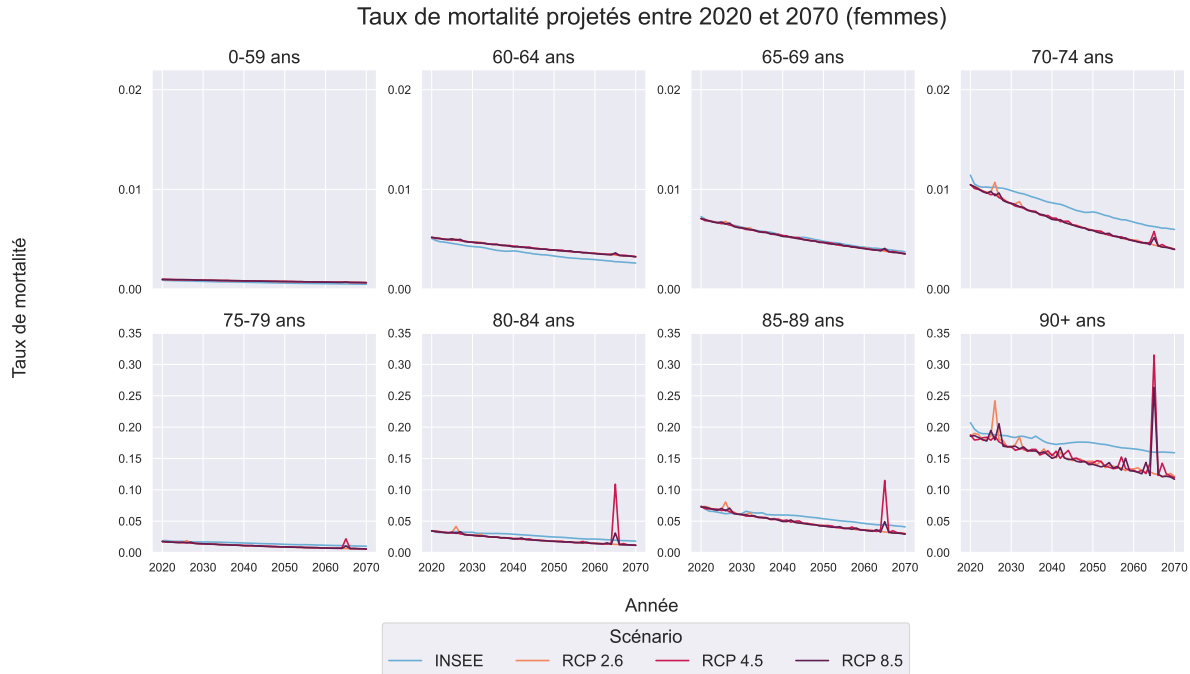
(a) : Modèle initial

Dans un premier temps, le modèle initial est considéré. Les résultats obtenus permettront de fournir une base de comparaison pour les deux modèles suivants, tenant compte d'une tendance ralentie et/ou de mesures d'adaptation face aux phénomènes de chaleur. Les taux de mortalité prédits pour chaque tranche d'âge et sexe sont représentés dans les graphiques 4.9 et 4.10.



GRAPHIQUE 4.9 – Taux de mortalité projetés à horizon 2070 par tranche d'âge et scénario RCP pour les hommes dans le modèle (a)

Tout d'abord, indépendamment du sexe retenu, les tendances modélisées sont relativement similaires entre le modèle DLNM et la méthodologie de l'INSEE. Cela est un signe rassurant concernant la qualité de la démarche adoptée. Pour certaines classes d'âge, la tendance à la baisse des taux de mortalité est toutefois davantage marquée, ce qui peut être expliqué par des différences d'hypothèses. En effet, alors que la méthodologie de cette étude propose une tendance (de décroissance ralentie) de la mortalité entre 1990 et 2019, celle projetée par l'INSEE repose davantage sur les observations de la dernière décennie. Cette dernière était marquée par un ralentissement notable des progrès de l'espérance de vie. Les biais sont également relativement faibles entre les méthodes et semblent se compenser entre les classes d'âge. La mortalité prédite dans cette analyse est supérieure jusqu'à la classe 60-64 ans, puis devient systématiquement inférieure chez les individus de plus de 70 ans. Ces différences soulèvent l'importance de l'hypothèse de tendance dans une modélisation prospective, qui



GRAPHIQUE 4.10 – Taux de mortalité projetés à horizon 2070 par tranche d’âge et scénario RCP pour les femmes dans le modèle (a)

peut avoir un impact matériel sur la tarification. Par exemple, pour un produit d’assurance-vie garantissant un capital en cas de décès, la prime proposée à un assuré retraité serait plus faible si la compagnie utilisait la méthode de cette analyse plutôt que celle de l’INSEE.

Ensuite, des différences notables de mortalité seraient visibles selon le sexe et l’âge entre deux années. Ces écarts correspondent, en fait, aux épisodes de températures extrêmes constatés dans la section précédente :

- en 2026 pour le scénario RCP 2.6 ;
- en 2065 pour les scénarios RCP 4.5 et 8.5.

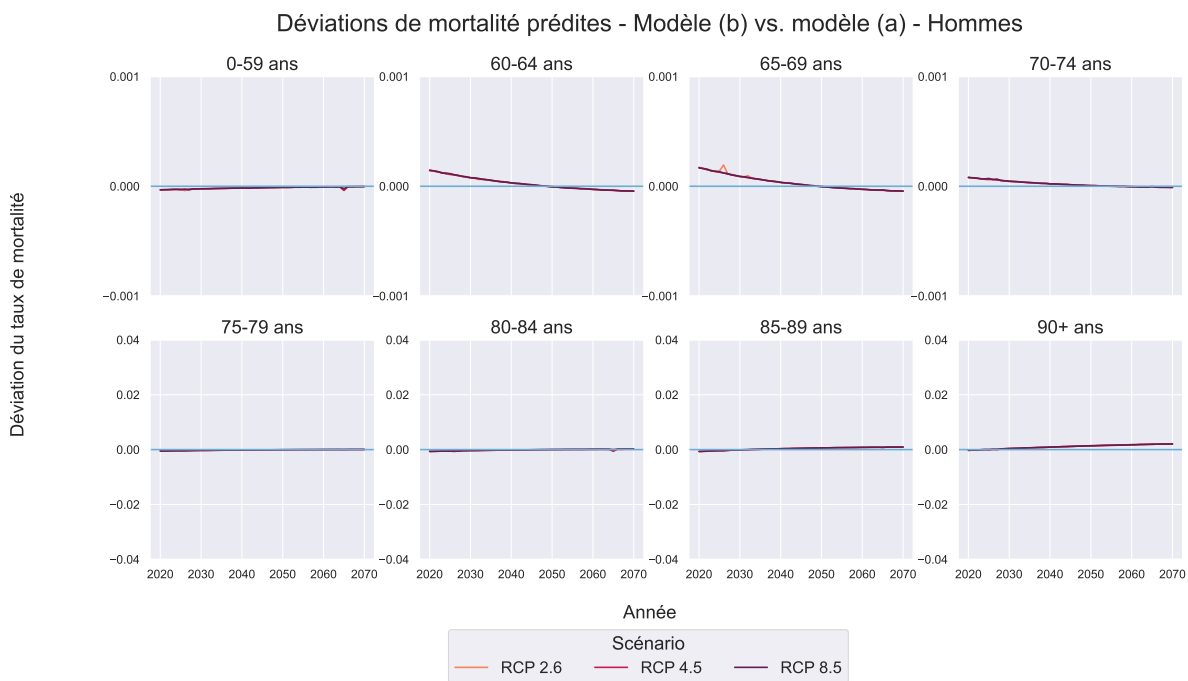
Alors que ces vagues de chaleur auraient un impact négligeable sur la mortalité par rapport à l’année précédente chez les hommes (moins de 10%), leurs conséquences pourraient être sensiblement supérieures chez les femmes. Les impacts potentiels sur les capitaux réglementaires à dégager pour les assureurs pourraient être importantes, selon la composition de leurs portefeuilles et de leur sensibilité au risque de décès.

Dans leur ensemble, les résultats pourraient être considérés comme plausibles, sachant que la plus grande sensibilité des femmes à la chaleur a déjà été constatée dans la littérature, et qu’il existe des conditions de température (couplées à une humidité importante) ne pouvant être supportées par les individus. En cela, les critères (i) et (ii) issus de Petauton semblent

effectivement respectés. Les projections obtenues pour l'année 2065 doivent, cependant, être prises avec une grande précaution. D'une part, la profondeur de la projection est telle qu'il n'est pas possible de prédire l'ampleur des mesures de prévention permettant de réduire les conséquences des épisodes de températures extrêmes. D'autre part, l'acclimatation pourrait diminuer la sensibilité des individus aux vagues de chaleur futures, sachant qu'elles seraient amenées à devenir plus fréquentes et intenses dans les prochaines décennies. En revanche, la plus grande mortalité observée dans le scénario RCP 4.5 peut se justifier, sachant que la stabilisation du forçage radiatif ne serait effective que vers 2050, et son impact sur les températures de 2065 encore limité.

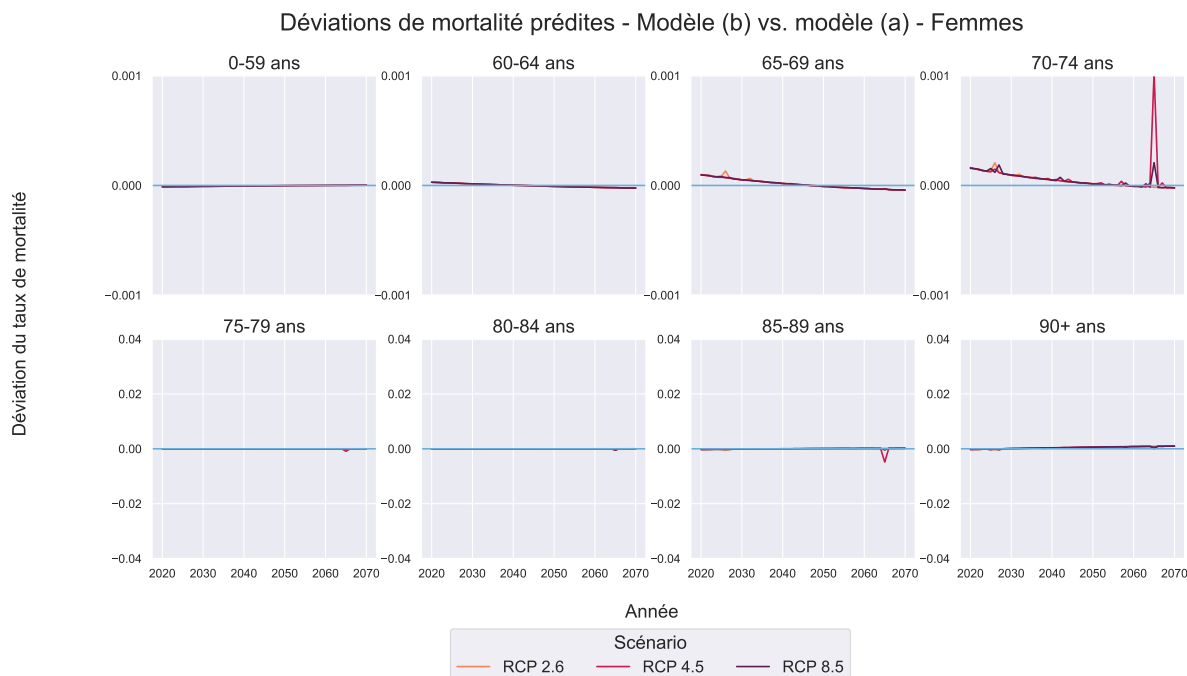
(b) : Tendence ralentie

Ensuite, la variante du modèle initial est considérée. Celle-ci inclut un terme de tendance supplémentaire afin de capter le ralentissement des progrès de l'espérance de vie. Les déviations des taux de mortalité prédites pour chaque tranche d'âge et sexe sont représentés dans les graphiques 4.11 et 4.12. Le modèle initial servira de point de comparaison.



GRAPHIQUE 4.11 – Déviations des taux de mortalité projetées à horizon 2070 par tranche d'âge et scénario RCP pour les hommes dans le modèle (b) par rapport au modèle (a)

L'analyse temporelle à l'échelle nationale des sorties du modèle (b) est très similaire à celle du modèle (a). Les tendances de mortalité prédites sont légèrement plus proches de



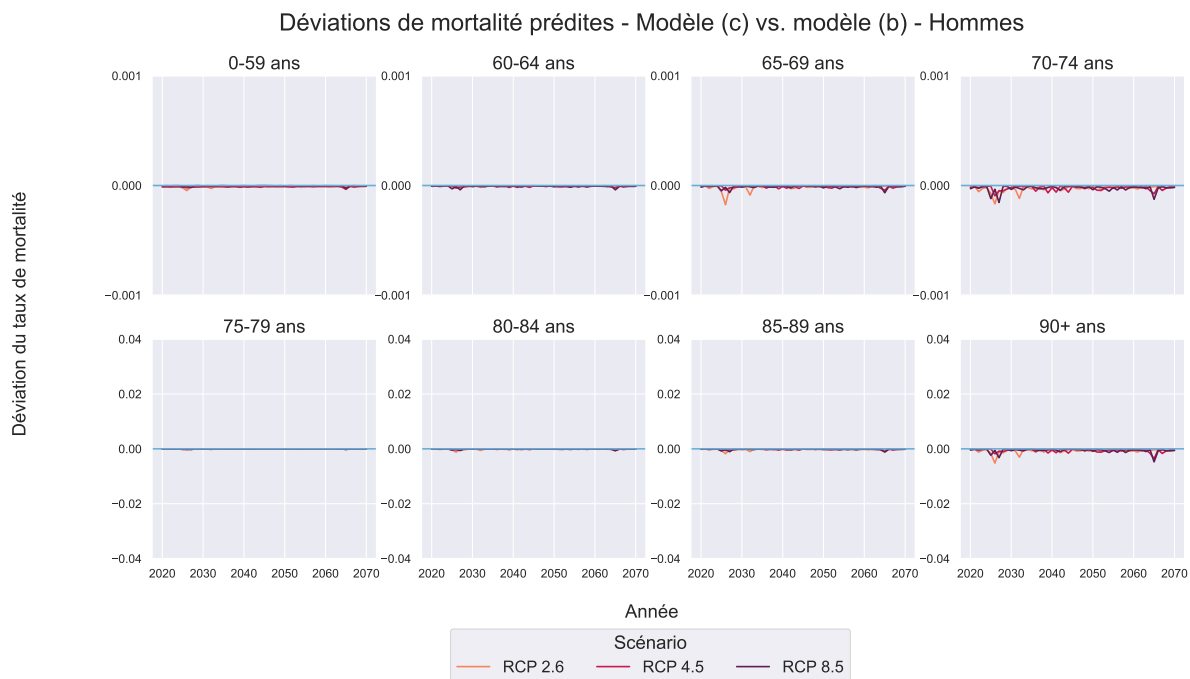
GRAPHIQUE 4.12 – Déviations des taux de mortalité projetées à horizon 2070 par tranche d'âge et scénario RCP pour les femmes dans le modèle (b) par rapport au modèle (a)

celles de l'INSEE, avec des taux de mortalité généralement supérieurs à ceux prédits dans le modèle initial. Ces observations sont cohérentes avec la construction du modèle (b). En effet, la mortalité prospective reproduit la tendance observée sur la période 2010-2019, marquée par un ralentissement de la progression de l'espérance de vie à la naissance. Toutefois, les différences avec les sorties du modèle (a) sont faibles, et sont concentrées sur les années de fortes chaleurs (2065 notamment) pour certaines classes d'âge. Autrement dit, l'inclusion d'un terme de tendance supplémentaire peut conduire à modifier les relations température-mortalité estimées par rapport au modèle initial.

La grande proximité des prédictions avec celles du modèle initial peut notamment s'expliquer par la fonction de lien exponentielle utilisée : le coefficient associé à la tendance étant négatif, l'évolution temporelle prédite dans le modèle (a) (toutes choses égales par ailleurs) a une forme décroissante et ralentie. L'inclusion d'un terme supplémentaire afin de capter la tendance des années 2010 à 2019 ne permet donc que de prédire un ralentissement plus marqué des évolutions structurelles de mortalité. Le risque de longévité prédit est ainsi légèrement moins important avec le modèle (b).

(c) : Tendence ralentie et adaptation

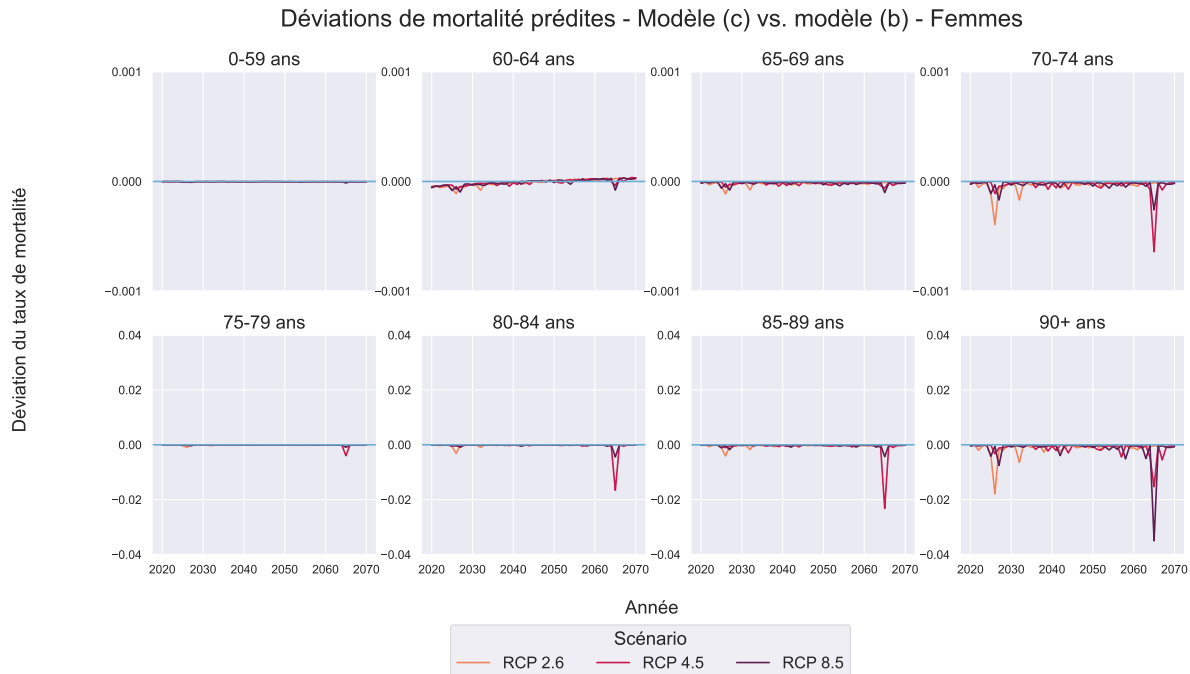
Enfin, la variante du modèle (b) incluant la prise en compte de mesures d'adaptation est considérée. Pour rappel, en plus d'inclure un terme de tendance supplémentaire, le modèle (c) suppose une réduction de la surmortalité de 30% au-delà de l'optimum thermique. Cette hypothèse a pour objectif de refléter les apports de l'adoption de diverses politiques¹¹ pour réduire les effets des épisodes de chaleurs extrêmes. Les déviations des taux de mortalité prédites pour chaque tranche d'âge et sexe sont représentés dans les graphiques 4.13 et 4.14. Le modèle (b) sera cette fois utilisé comme point de comparaison, afin de constater au mieux l'impact des mesures d'adaptation sur les taux de mortalité, lorsque les paramètres de tendance sont fixés.



GRAPHIQUE 4.13 – Déviations des taux de mortalité projetées à horizon 2070 par tranche d'âge et scénario RCP pour les hommes dans le modèle (c) par rapport au modèle (b)

Les sorties du modèle (c) mettent en lumière un certain apport des mesures d'adaptation, permettant de réduire les fluctuations des taux de mortalité entre les années. Sachant que les femmes et les personnes âgées sont davantage sensibles aux conséquences des vagues de chaleur, l'impact à la baisse sur la mortalité est naturellement mieux observable pour ces deux catégories. Par ailleurs, la diminution de la mortalité est davantage visible pour les vagues de chaleur particulièrement intenses. En ce sens, les études des épisodes extrêmes menées

11. Une description en sera proposée dans la partie 4.3.2



GRAPHIQUE 4.14 – Déviations des taux de mortalité projetées à horizon 2070 par tranche d’âge et scénario RCP pour les femmes dans le modèle (c) par rapport au modèle (b)

précédemment pour chaque scénario climatique sont utiles à la compréhension des résultats. Par exemple, dans le cas des épisodes de chaleur extrêmes de l’année 2065, la vague projetée dans le scénario RCP 4.5 est plus longue, mais aussi d’intensité moindre que celle du scénario RCP 8.5. Pourtant, en considérant seulement la catégorie la plus sensible (les femmes de 90 ans et plus), l’adoption de mesures d’adaptation permettrait de réduire les taux de mortalité prédits de 0,02 point (diminution de 6%) dans le scénario intermédiaire, et de 0,04 point (diminution de 15%) dans le scénario *business as-usual*. Cette différence significative met en évidence la nécessité de savoir gérer les épisodes de températures extrêmes étalés sur une longue période, par exemple à travers la préparation des services hospitaliers.

Après avoir constaté les différences d’évolution de la mortalité par tranche d’âge et sexe, une analyse comparative géographique est menée.

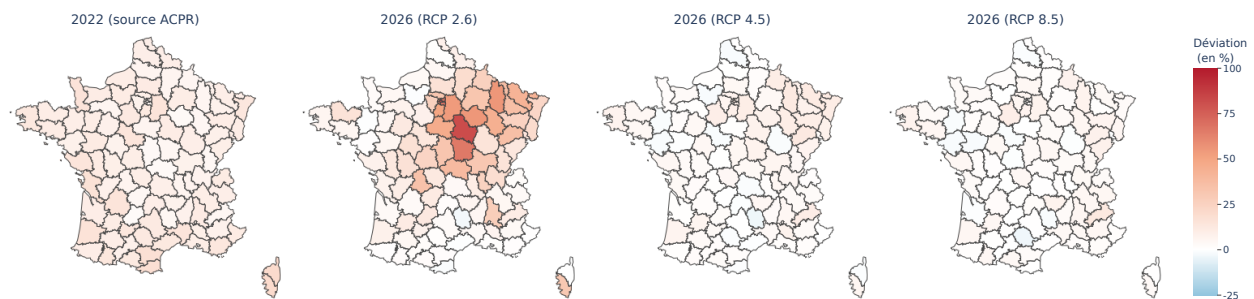
4.2.3 Comparaison géographique de la mortalité prospective

Une étude géographique prospective pourrait permettre à un assureur de compléter son analyse de matérialité future du changement climatique (pour l’ORSA notamment), en la reliant aux expositions de son portefeuille. Il s’agira, par exemple, de prévoir si un assureur implanté exclusivement dans une région pourrait être davantage exposé aux évolutions des

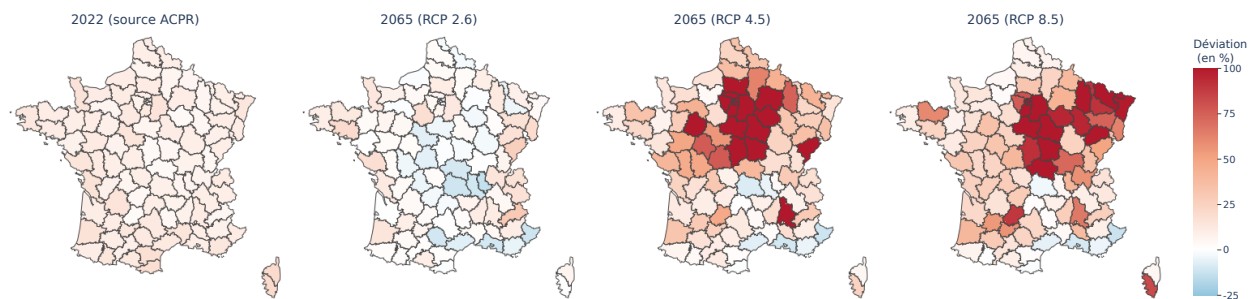
phénomènes de chaleur extrêmes qu'un concurrent dont l'activité est exercée sur l'ensemble du territoire. Cette problématique a été introduite par l'ACPR dans son exercice climatique de 2023¹² : des déviations de mortalité par département ont été proposées entre 2022 et 2019 par département. Par ailleurs, des projections de ces déviations ont été fournies à court terme par le courtier AON, pour quelques grandes classes d'âge (0-15 ans, 15-44 ans, 45-64 ans, 65-74 ans et 75 ans et plus) et l'ensemble de la population, mais sans distinction géographique. La présente étude fournit une méthodologie permettant de reproduire ces projections avec une distinction géographique. La première intuition fournie par les analyses descriptives précédentes est que certaines régions, telles que l'Île-de-France, pourraient être sujettes à des déviations de mortalité importantes dans les prochaines décennies.

La méthodologie adoptée ici se veut principalement graphique : les années d'étude sont fixées et correspondent aux pics de chaleur principaux observés pour chaque scénario climatique. Ce choix a été effectué en raison du détachement net des phénomènes climatiques pour ces périodes et ces scénarios par rapport au reste des années de projection. Par ailleurs, afin de fournir une comparaison directe avec les déviations de mortalité fournies par l'ACPR, l'ensemble de la population est considérée. Toutefois, il reste possible de reproduire l'analyse pour chaque sexe et classe d'âge, séparément. Les déviations obtenues dans le modèle initial sont disponibles dans le graphique 4.15.

Déviations de mortalité prédites en 2026 par rapport à 2019 - Modèle (a)



Déviations de mortalité prédites en 2065 par rapport à 2019 - Modèle (a)



GRAPHIQUE 4.15 – Déviations du taux de mortalité par rapport à 2019 dans le modèle (a) : historique (2022, à gauche ; source ACPR) et projections en 2065 par scénario RCP

12. Lien vers les principales hypothèses de l'exercice 2023 : [œ](#)

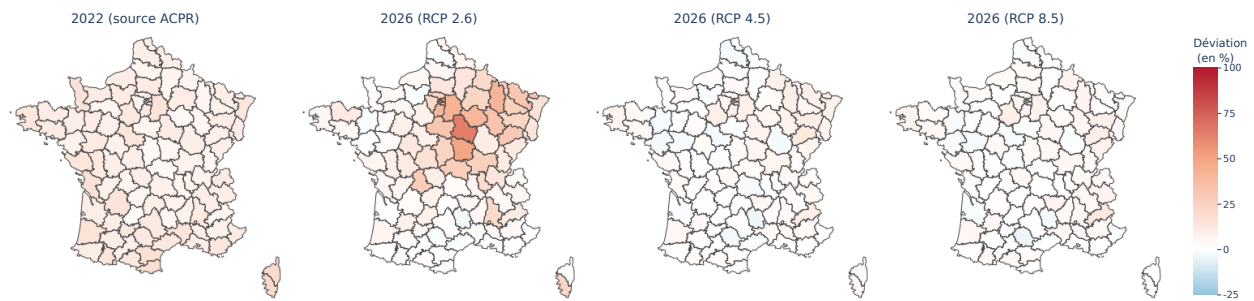
Alors que les déviations de mortalité de 2022 ont été relativement uniformes sur l'ensemble du territoire (malgré une surmortalité davantage concentrée dans les départements côtiers), les épisodes de chaleurs extrêmes des prochaines années pourraient conduire à une hétérogénéité géographique nette des taux de décès.

Ainsi, dans les départements les plus frappés par les pics de chaleur et identifiés précédemment dans ce chapitre (l'Île-de-France et les régions limitrophes), les déviations de mortalité prédites par le modèle DLNM pourraient être significatives. Par exemple, en 2065, la vague de chaleur projetée selon le scénario RCP 4.5 pourrait conduire à un taux de mortalité plus que doublé dans la totalité des départements d'Île-de-France par rapport à 2019. Une analyse plus poussée a permis de mettre en évidence que, parmi l'ensemble de la population, la mortalité des femmes et des personnes âgées contribuerait largement à ce résultat. En effet, d'après les sorties du modèle, les épisodes de chaleur de 2065 pourraient conduire au décès quasi-systématique des femmes de plus de 90 ans dans la totalité des départements d'Île-de-France. De même, dans le scénario RCP 8.5 et l'année 2065, pour cette même catégorie, la population des Hauts-de-Seine pourrait également disparaître. Ces résultats sont cohérents avec l'analyse comparative des relations températures-mortalité par âge et sexe menée précédemment. Les départements de la région Grand-Est ne seraient pas épargnés et pourraient connaître des déviations de mortalité supérieures à 80% par rapport à l'année 2019. Au contraire, dans les territoires méditerranéens, plus acclimatés, le taux de mortalité pourrait diminuer par rapport à l'année 2019, sous l'effet de la tendance structurelle à la hausse de l'espérance de vie.

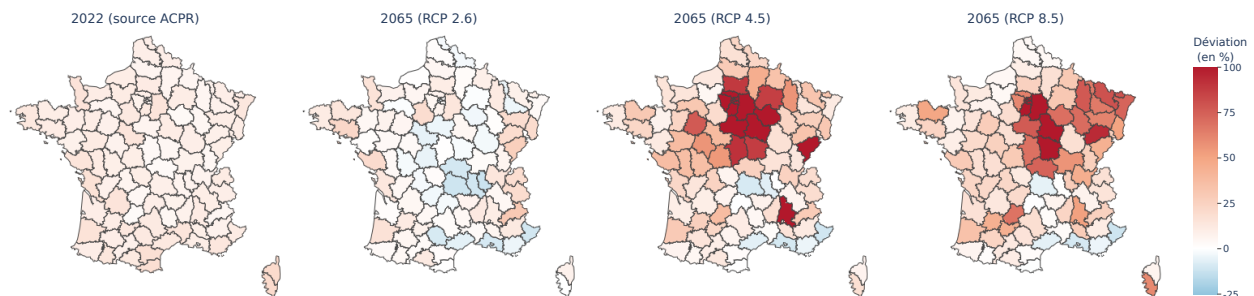
Afin de constater l'apport des mesures d'adaptation face aux vagues de chaleur, les taux de mortalité sont également représentés dans le graphique 4.16 sous l'hypothèse d'une réduction de 30% de la surmortalité au-delà de l'optimum thermique.

Comme attendu, des politiques d'adaptation aux chaleurs extrêmes pourraient permettre de réduire les hétérogénéités géographiques de mortalité. À titre d'exemple, les prédictions de mortalité en 2026 dans le scénario climatique le plus favorable et le département de l'Yonne sont considérées. Alors que les taux de mortalité prédits seraient supérieurs de 81% à ceux observés en 2019 d'après le modèle initial, la déviation pourrait être réduite à 64% grâce à une diminution de la surmortalité au-delà de l'optimum thermique. Toutefois, cet impact positif n'est pas toujours visible, notamment pour les pires épisodes de chaleur au cours desquelles le modèle prédit la disparition totale de certaines populations. Par exemple, dans le cas du pire épisode de chaleur projeté (en 2065 pour le scénario RCP 4.5), l'ensemble de la population des femmes de 90 ans et plus pourrait être amenée à disparaître, y compris si des mesures d'adaptation sont adoptées.

Déviations de mortalité prédites en 2026 par rapport à 2019 - Modèle (c)



Déviations de mortalité prédites en 2065 par rapport à 2019 - Modèle (c)



GRAPHIQUE 4.16 – Déviations du taux de mortalité par rapport à 2019 dans le modèle (c) : historique (2022, à gauche ; source ACPR) et projections en 2065 par scénario RCP

Une nouvelle fois, il convient de conserver un certain recul quant à la portée de ces résultats, qui dépend à la fois de la cohérence des projections climatiques utilisées et de la qualité d'estimation du modèle historique. L'importance de la segmentation géographique est également à considérer, car elle influe directement sur les relations historiques estimées et donc les projections. Si des critères de variance inter et intra-classes auraient pu être retenus, une première analyse graphique simple peut également être menée. Il peut s'agir, par exemple, de comparer les projections obtenues entre deux départements voisins appartenant à des zones climatiques différentes. En principe, celles-ci devraient ne pas être trop éloignées. Sinon, cela pourrait être le signe d'une mauvaise segmentation. Toutefois, cette étude graphique n'a pas montré d'hétérogénéités absurdes, ce qui est un premier indice encourageant quant à la pertinence de la segmentation géographique retenue.

À titre de comparaison, dans l'étude de référence, les déviations de mortalité sont déterminées en comparaison des nombres de décès fournis dans les tables prospectives nationales de l'INSEE. Des chocs sont appliqués en tenant compte des conséquences des épisodes de chaleur dans un modèle CSDL estimé sur l'ensemble du territoire. Par rapport au scénario démographique central, le nombre de décès supplémentaires dus aux phénomènes de chaleurs pourrait être compris entre 0 et 35 000 par année. Cette dernière valeur est obtenue dans le cas de l'année 2065 pour le scénario *business-as-usual*, sachant que le scénario RCP 4.5 n'est pas considéré. Rapportés au nombre de décès annuels, ces résultats pourraient représenter

une déviation de 4% des taux de mortalité prédits par l'INSEE. La structure log-linéaire de la mortalité imposée dans un modèle CSDL pourrait justifier que les déviations obtenues soient moins importantes que dans cette étude. Par ailleurs, les conséquences des évolutions des pics de froid ne sont pas considérées, et pourraient être amenées à réduire légèrement ce taux.

4.3 Applications actuarielles, limites et pistes d'amélioration

4.3.1 Applications actuarielles de l'étude

Tarification en assurance vie

La méthodologie d'obtention de la mortalité prospective peut être appliquée dans le cadre d'une tarification en assurance-vie. Le cas d'un produit de rente viagère simple a été retenu. Ce dernier permet à un assuré de percevoir des versements à une fréquence prédéfinie tant qu'il survit, moyennant le versement d'un capital initial à l'assureur. Le montant de ce capital initial peut être déterminé à partir de la valeur actuelle probable (VAP). Celle-ci mesure l'engagement actualisé de l'entreprise d'assurance lié aux versements théoriques qui pourraient être effectués. Pour la tarification d'un produit de rentes viagères, trois variables entrent en jeu :

- Le montant de la rente (supposée annuelle), noté C
- Le taux d'actualisation, noté i
- La probabilité de survie au cours de k années pour un individu d'âge x , notée p_x^k

La valeur actuelle probable s'écrit alors :

$$\text{VAP} = \sum_{k=1}^{+\infty} C \times p_x^k \times \frac{1}{(1+i)^k}.$$

En pratique, les probabilités de décès et de survie sont obtenues à partir d'une table de mortalité au choix de l'assureur. Le calcul de la VAP est approximatif, sachant que les probabilités de survie ne sont pas toujours disponibles dans les grands âges (en raison d'un manque d'observations). Pour cette application, des hypothèses simples ont été retenues :

- L'étude se place en 2024 et un individu de sexe féminin, âgé de 65 ans, est considéré.
- La rente annuelle est fixée à 12 000 €.
- Le taux d'actualisation est supposé nul (une unité monétaire d'aujourd'hui aura la même valeur dans un an).
- Les frais de l'assureur ne sont pas pris en compte.

Par ailleurs, conformément à la réglementation, une table unique a été considérée. La première solution proposée est l'utilisation de la table TGF05, fournie par l'INSEE. Celle-ci contient des extrapolations des taux de mortalité, qui sont donc utilisables aujourd'hui pour cet exemple de tarification. Le cas échéant, la VAP est voisine de 340 168 €. Cela implique que l'individu de 65 ans devra verser un capital initial de cet ordre de grandeur afin de disposer d'une rente annuelle de 12 000 €. Le montant aurait été similaire pour un individu de sexe masculin du même âge.

La seconde solution est l'utilisation d'une table d'expérience basée sur la population assurée. Pour simplifier, il est supposé ici que la répartition du portefeuille est similaire à celle du territoire, mais cette hypothèse pourrait être relâchée ici (sachant que les taux de mortalité prospectifs de ce mémoire sont disponibles à l'échelle départementale). Les données historiques sur la période 1990-2019 ont permis d'établir la table d'expérience, et les températures journalières projetées par les scénarios RCP sont utilisées afin d'établir des projections. Le modèle (c), supposant l'adoption de mesures d'adaptation et une croissance ralentie de l'espérance de vie, a été retenu. Enfin, conformément à la réglementation, la table des femmes a de nouveau été choisie. Selon le scénario retenu, les capitaux initiaux à verser varient. Ces variations sont présentées dans le tableau 4.3.

	RCP 2.6	RCP 4.5	RCP 8.5
VAP	306 298 €	308 004 €	308 003 €
Variation de la VAP par rapport à la table TGF05	-9,96%	-9,46%	-9,46%

TABLEAU 4.3 – Évolution de la valeur actuelle probable d'un contrat de rente viagère par rapport à la table TGF05

L'écart avec les valeurs obtenues dans la table TGF05 peut s'expliquer par les différences de méthodologies dans la formation des tables prospectives. Celles-ci conduisent à des projections de mortalité plus élevées dans ce mémoire, et potentiellement plus cohérentes avec le ralentissement de la progression de l'espérance de vie à la naissance observé lors de la dernière décennie. La comparaison des espérances de vie résiduelles de l'individu considéré (une femme âgée de 65 ans et née en 1959) entre les méthodes permet de confirmer ce constat. Celle-ci est présentée dans le tableau 4.4.

	TGF05	RCP 2.6	RCP 4.5	RCP 8.5
Espérance de vie résiduelle	25,68 ans	24,62 ans	24,93 ans	24,82 ans

TABLEAU 4.4 – Espérance de vie résiduelle d’une femme de 65 ans née en 1959 selon la table prospective retenue

La méthodologie de projection des taux de mortalité retenue implique une baisse de l’espérance de vie résiduelle de l’ordre d’un an, par rapport à la méthodologie de la table TGF05. La proximité des résultats entre les scénarios RCP renforce l’intuition que l’horizon de l’étude pourrait être insuffisant pour établir leurs différences d’impact sur la mortalité future. En effet, à l’horizon 2048 (l’horizon de survie probable), les épisodes de chaleurs extrêmes seraient trop peu différenciés entre les scénarios climatiques d’après les projections du DRIAS. C’est notamment vrai, par construction, entre les scénarios RCP 4.5 et 8.5 (qui supposent soit une stabilisation des émissions de GES à partir du milieu du siècle, soit leur augmentation continue).

En utilisant la table TGF05, l’entreprise d’assurance pourrait donc potentiellement surestimer le montant de capital initial demandé à l’assuré pour un contrat de rente viagère. Si les projections de mortalité issues de cette étude s’avèrent réalistes, la probabilité que le résultat de l’assureur soit positif augmentera. En effet, l’individu risque de décéder plus tôt que ce que la table TGF05 prévoyait, conduisant à l’arrêt des prestations. Toutefois, une entreprise concurrente pourrait également proposer aux assurés un contrat équivalent, nécessitant un capital de départ moindre, permettant de capter l’ensemble des assurés de la compagnie.

Tarification en assurance santé

Une méthodologie similaire à celle proposée dans cette étude pourrait permettre d’aider à la tarification en assurance santé. En effet, les mutuelles proposent des contrats permettant de rembourser la part des frais d’hospitalisation non prise en charge par l’Assurance Maladie. Sauf cas particuliers (par exemple une affection de longue durée), cette part correspond à 20% des frais totaux¹³. Une approche de tarification à partir de la modélisation de la fréquence et du coût des hospitalisations pourrait être utilisée, sachant que l’analyse de la première pourrait être analogue à celle effectuée dans cette étude. En effet, le taux d’hospitalisation peut être influencé - de manière similaire aux taux de mortalité - par la recrudescence d’épisodes de températures extrêmes. Une étude privée a par exemple été menée à partir d’un portefeuille d’assurés allemand entre 2006 et 2018 afin de proposer une métho-

13. Source : Assurance Maladie ↗

dologie d'estimation du ratio d'hospitalisations. À la différence de la présente analyse, il a été proposé de se concentrer uniquement sur des causes d'hospitalisation précises (les maladies respiratoires et cardiovasculaires) et sur les températures maximales. Par ailleurs, la tranche d'âge des individus a été incluse comme variable explicative dans le modèle. Ce choix a pour avantage de limiter le nombre de modèles à estimer, mais ne permet pas de rendre compte des relations différenciées entre la température et le nombre d'hospitalisations par âge. Par ailleurs, si la complexité des relations entre ces deux variables est incluse à partir des GAM (modèles additifs généralisés), les variables retardées ne semblent pas incluses dans le modèle. Même si l'objectif principal de l'étude était d'évaluer les conséquences des épisodes de chaleur, l'analyse menée dans ce mémoire a permis de soulever l'existence d'effets persistants jusqu'à une dizaine de jours. Le risque est alors d'obtenir une modélisation avec un pouvoir prédictif faible, notamment pour la captation des effets retardés lors des vagues de froid.

Une première étude des coûts des épisodes de températures extrêmes a, elle, été menée par Adélaïde et al. [1] en 2021. Son objectif était d'estimer l'ensemble des conséquences financières des vagues de chaleur survenues entre 2015 et 2020. Parmi celles-ci, figuraient les coûts liés à la surmortalité, aux soins et frais d'hospitalisations, et à la perte d'activité. Cette dernière composante pourrait être utile pour la tarification de contrats d'incapacité et d'invalidité. Un des avantages de cette étude est la relative simplicité des hypothèses utilisées, sachant qu'elle se base sur des coûts moyens :

- coût d'une hospitalisation : 3 866 € ;
- coût d'un passage aux urgences : 151 € ;
- coût d'une consultation SOS Médecins : 48 € ;
- perte de production (appliquée aux individus de 15 à 64 ans) : 104,24 €.

Cette étude suppose également une durée totale d'inactivité de 13,4 jours, correspondant à 6,7 jours d'hospitalisation et une durée identique en arrêt maladie. Les conséquences des canicules entre 2015 et 2017 représenteraient ainsi entre 22 et 37 milliards d'euros selon la valorisation retenue pour les pertes humaines¹⁴. Dans une application actuarielle, il serait nécessaire de disposer d'un historique de données plus important. Une modélisation journalière des coûts des hospitalisations et soins divers permettrait également de mieux observer les conséquences des épisodes de chaleur et de froid. Cependant, les perspectives d'études du risque santé avec des données publiques sont très limitées, car elles reposent sur l'analyse de données personnelles sensibles au sens de la CNIL (Commission Nationale de l'Informatique

14. Le Commissariat général à la stratégie et à la prospective recommande une valeur de 3,17 M€, tandis que Quinet propose une valorisation à 122 000 €.

et des Libertés), dont l'utilisation est strictement encadrée. Leur définition¹⁵ est rappelée ci-dessous :

[Les données personnelles sensibles] sont des informations qui révèlent la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

Par exemple, la plateforme *Health Data Hub* propose la base de résumés des passages individuels aux urgences (OSCOUR)¹⁶, recensant les données de plus de 130 millions de passages depuis 2010. Toutefois, l'accès à cette base requiert la création d'un dossier, à soumettre auprès de la CESREES (Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations dans le domaine de la Santé) puis de la CNIL.

Analyse de matérialité et de scénarios dans l'ORSA

En 2022, l'EIOPA a publié des recommandations d'intégration du risque climatique dans l'évaluation interne des risques et de la solvabilité (ORSA) des assureurs. Pour chaque secteur d'activité et chaque risque, une analyse en trois temps est recommandée :

1. Définir le contexte de l'entreprise d'assurance et rechercher les impacts du changement climatique sur l'entreprise ;
2. Évaluer la pertinence pour l'entreprise ;
3. Pour les risques considérés comme matériels suite aux étapes précédentes, procéder à l'analyse de scénarios. Sinon, les justifier dans l'ORSA.

L'étude de matérialité constitue la deuxième étape. Dans ses recommandations, l'EIOPA préconise de la mener séparément, à court-terme puis à long-terme. Afin de juger de la pertinence du risque, une matrice basée sur trois critères pourra être proposée : l'ampleur des conséquences pour la compagnie, l'horizon ainsi que la probabilité. Si l'étude de matérialité peut être menée à partir d'une démarche essentiellement descriptive, l'analyse des scénarios requiert quant à elle une démarche quantitative. À cette fin, l'EIOPA recommande une série d'outils utiles pour évaluer l'impact du changement climatique dans l'analyse quantitative,

15. Lien vers la définition du CNIL : [☞](#)

16. Lien vers les principales bases de la plateforme *Health Data Hub* : [☞](#)

ainsi que quatre exemples de méthodes d'analyse des scénarios de changement climatique sur les risques physiques :

- (a) l'explorateur d'impact climatique NGFS ;
- (b) les données de la base de données climatiques Peseta IV ;
- (c) les stress tests déjà effectués, tels que ceux fournis aux régulateurs (ACPR, PRA) ;
- (d) les scénarios de changement climatique disponibles, tels que ceux fournis par le NGFS.

Dans le cadre de l'analyse de scénarios, ce mémoire pourrait ainsi apporter une étude quantitative basée sur la proposition (d), bien que les scénarios utilisés soient issus d'une source différente. Son originalité serait de tenir compte des expositions géographiques de l'assureur, grâce à l'analyse menée par zone climatique. Par ailleurs, pour son analyse de scénarios, l'EIOPA préconise l'utilisation de deux scénarios :

- un scénario dans lequel l'augmentation de la température mondiale reste inférieure à 2°C (RCP 2.6 ou RCP 4.5) ;
- un scénario dans lequel l'augmentation de la température mondiale dépasse 2°C (RCP 6.0 ou RCP 8.5).

Ainsi, bien qu'il soit possible de conserver l'ensemble des scénarios, une étude omettant l'un des deux scénarios les plus favorables pourrait également être menée. Enfin, l'analyse de scénarios nécessiterait de quantifier l'impact du phénomène climatique sur un indicateur actuariel. Dans un document méthodologique relatif aux stress-tests climatiques, l'EIOPA [22] fournit une liste d'indicateurs (pour le risque physique, le risque de transition ou les deux) dans le but de dresser un tableau complet des principaux facteurs de risque à l'origine de l'impact des scénarios choisis sur certains domaines. Les indicateurs sont classés en fonction du bilan, de la rentabilité et des aspects techniques.

Parmi ces différentes dimensions, la variation de SCR (*Solvency Capital Requirement*) est proposée pour quantifier l'impact du phénomène climatique sur la solvabilité d'un assureur. Par exemple, il serait possible d'utiliser les projections de mortalité de l'INSEE comme base de comparaison. En effet, elles supposent une tendance prolongée des taux de mortalité et ne prennent en compte que des hypothèses de projection démographiques. Ces projections peuvent être comparées avec la méthode proposée dans ce mémoire, qui prend en compte les impacts du dérèglement climatique via l'inclusion d'une variable de température. Les taux de mortalité ainsi obtenus pourraient ensuite être inclus en entrée d'un modèle de gestion actif-passif afin d'évaluer les impacts sur le bilan économique à la base du SCR. Cette approche a été retenue par Dal Pont [14] afin de comparer plusieurs méthodes de calcul des taux de mortalité sur un portefeuille réel. Selon les méthodes retenues, le ratio de solvabilité global

pouvait varier de -5% à 7% par rapport à celui obtenu avec la table de mortalité d'origine. Le schéma de fonctionnement d'un modèle ALM, inspiré du mémoire de Tichit [55], est rappelé dans le graphique 4.17.

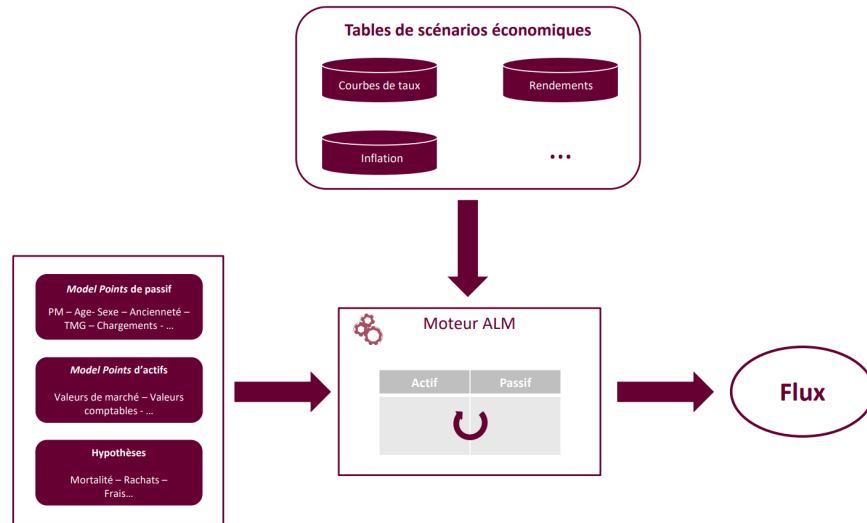


SCHÉMA 4.17 – Schéma de fonctionnement d'un modèle ALM

La simplicité du calcul du SCR dépendra du choix de l'assureur, entre la formule standard et un modèle interne. Dans la formule standard, une approche modulaire est retenue et nécessite de calculer séparément des chocs relatifs à chaque risque, puis de tenir compte des corrélations entre ces derniers. Concernant la mortalité, un choc de 15% est fixé et doit permettre de tenir compte des incertitudes liées à l'évolution et la volatilité des taux. La comparaison des SCR permettra alors d'obtenir un indice quant à l'impact du changement climatique sur le bilan de l'assureur.

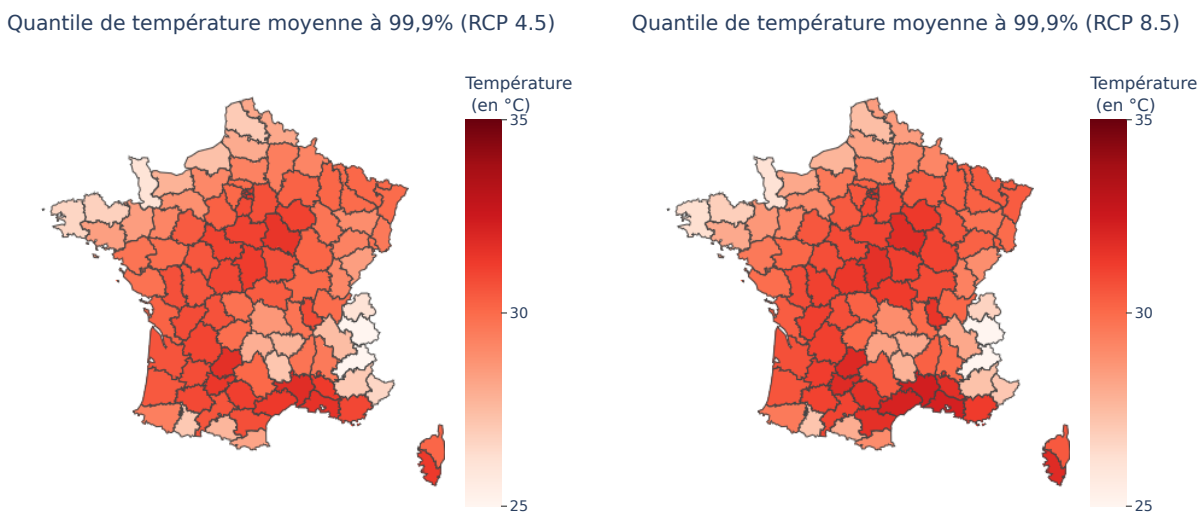
4.3.2 Limites et pistes d'amélioration

La sensibilité des projections de mortalité aux données et indicateurs climatiques

L'étude menée précédemment fournit des interprétations relatives aux zones qui pourraient être davantage concernées par les vagues de chaleur futures. Toutefois, il est important de conserver une certaine retenue quant à la portée des résultats, notamment en raison de la profondeur de la projection. Afin de mieux s'en apercevoir, d'autres jeux de données climatiques issus du programme DRIAS-2020 ont été parcourus. Tout comme les bases de projections du CNRM, ils s'appuient sur la réanalyse SAFRAN, utilisant un jeu de données

historique de référence afin de capter le climat passé en France métropolitaine¹⁷.

Sachant que l'analyse précédente a mis en avant le lien étroit entre la surmortalité sur une année et la survenance d'épisodes de chaleur extrêmes, une analyse par quantiles a été effectuée sur le jeu de données de l'Institut Pierre-Simon Laplace (IPSL). Les quantiles 99,9% des températures moyennes journalières sont représentés sur le graphique 4.18.



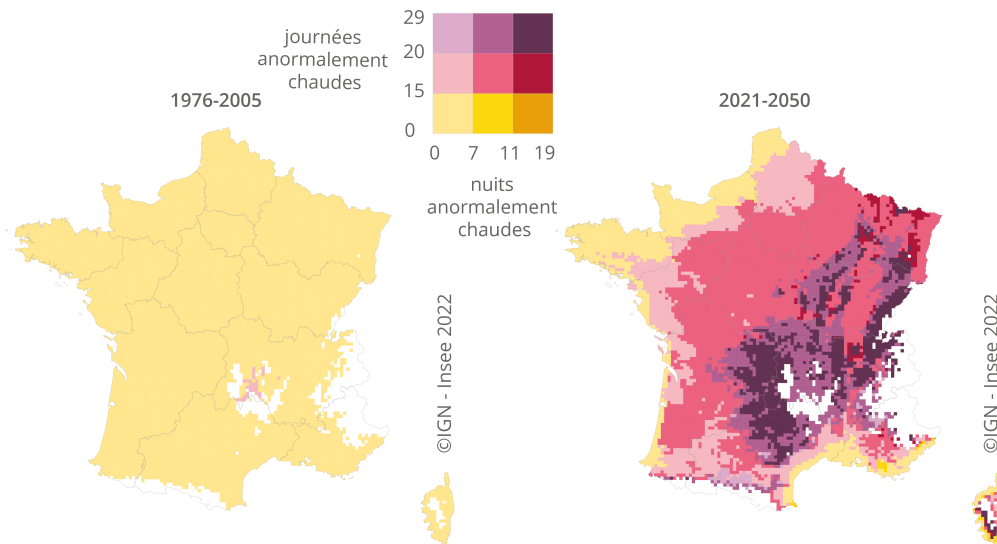
GRAPHIQUE 4.18 – Quantiles de température moyenne journalière à 99,9% par scénario climatique en France métropolitaine d'après les projections IPSL sur la période 2020-2070 (à gauche : RCP 4.5 ; à droite : RCP 8.5)

Pour rappel, ce quantile correspond à une température moyenne extrême observée en moyenne une fois sur 1 000 jours, soit tous les trois ans environ. Par rapport à la même métrique calculée sur la base de projection du CNRM (représentée dans le graphique 4.4), les températures extrêmes obtenues sont plus faibles. Cela pourrait se traduire par une surmortalité attribuable à la chaleur moindre par rapport à celle projetée avec les scénarios du jeu de données du CNRM. Par ailleurs, bien que les zones les plus chaudes sur cette métrique restent similaires, et incluent toujours la petite couronne de Paris, les départements méditerranéens seraient davantage exposés en comparaison du reste du territoire. La répartition géographique et temporelle des décès pourrait s'en retrouver impactée.

Par ailleurs, les projections de Météo-France ont fait l'objet d'un article de Fontès-Rousseau et al. [26], mettant en avant d'autres zones géographiques qui pourraient être exposées. Les nombres moyens de journées et de nuits anormalement chaudes par année sont l'indicateur retenu ici. Une journée (resp. nuit) anormalement chaude est définie par les auteurs de l'article comme : *"une journée (nuit) pour laquelle la température maximale (mi-*

17. Source : DRIAS ↗

nimale) est supérieure d'au moins 5°C à la température maximale (minimale) de référence (calculée au niveau local sur la période 1976-2005)". Cet indicateur fournit une information complémentaire d'un quantile, car il intègre également l'aspect de fréquence. Par ailleurs, l'étude inclut simultanément la température maximale journalière et celle minimale (observée la nuit), tandis que la présente analyse s'est intéressée à la moyenne de ces deux indicateurs. Les nombres moyens de températures anormalement chaudes entre 2020 et 2050 selon le scénario RCP 8.5 sont représentés dans le graphique 4.19.



GRAPHIQUE 4.19 – Nombre moyen de températures anormalement chaudes par été en France métropolitaine (à gauche : 1976-2005 ; à droite : entre 2020 et 2050 selon le scénario RCP 8.5) (source : Fontès-Rousseau et al. [26])

Bien que cette étude confirme que les zones littorales (en bords de Manche ou méditerranéennes) seront probablement moins concernées par l'impact des épisodes de chaleur futurs, des régions autres que celles mentionnées dans la présente analyse ont été identifiées comme à risque. C'est le cas des régions Bourgogne-France-Comté et Auvergne-Rhône-Alpes, notamment. Dans cette dernière région, au cours des trente prochaines années, il pourrait y avoir en moyenne, chaque été, de 21 à 29 journées et de 12 à 19 nuits anormalement chaudes. Ces deux régions pourraient donc également être exposées à un risque de surmortalité futur lié à la chaleur. Cette analyse des projections de températures fournie par l'INSEE laisse suggérer que la métrique de température utilisée pourrait également être un facteur de sensibilité des projections de mortalité.

Intégration de la qualité de l'air dans une étude prospective

Si les températures constituent une variable relativement puissante pour l'analyse historique et prédictive de la mortalité, d'autres variables auraient pu permettre de fournir une information supplémentaire utile à la prédiction. C'est notamment le cas de la qualité de l'air, qui peut être mesurée à travers des indicateurs de concentration des particules dans l'air (PM2.5 et PM10). Comme expliqué précédemment, une forte concentration de ces particules dans l'air peut constituer un facteur de risque supplémentaire pour certains individus déjà atteints de symptômes respiratoires. Toutefois, la publicité des données relatives à la qualité de l'air reste à ce jour limitée. En 2021, la plateforme Atmo Data a été créée par la fédération nationale des Associations agréées de surveillance de la qualité de l'air (AASQA). Elle propose de regrouper l'ensemble des données libres d'accès des instituts régionaux français, que les utilisateurs peuvent télécharger sous certaines conditions¹⁸. De même, la base *CAMS European air quality reanalyses* publiée par l'institut Copernicus propose des indicateurs de qualité de l'air à partir de la ré-analyse des différentes observations fournies par 11 modèles des agences nationales.

Toutefois, dans les deux cas, les données ne sont disponibles qu'à partir de 2013. Cette limite pourrait s'avérer préjudiciable pour une étude prospective, car elle serait basée sur un nombre réduit d'observations historiques. Par ailleurs, les indicateurs ne sont généralement projetés que jusqu'à quelques jours, compliquant toute analyse prospective quantitative. De manière qualitative, d'après l'Organisation météorologique mondiale (OMM), il existerait une dynamique de renforcement mutuel entre la pollution et le réchauffement de la planète. Ainsi, même si les émissions sont faibles, le réchauffement de la planète entraînerait une augmentation de la fréquence des feux de forêts, qui causerait en retour une augmentation de la pollution atmosphérique¹⁹. En revanche, d'après le sixième rapport du GIEC, si la stabilisation des températures mondiales pourrait prendre 20 à 30 ans, la qualité de l'air bénéficierait rapidement de la diminution du forçage radiatif. Les projections de mortalité à horizon 2070 par scénario climatique pourraient ainsi être davantage différenciées selon l'impact estimé de la qualité de l'air.

Une intégration plus fine des mesures d'adaptation

La méthodologie adoptée dans cette étude prospective suppose que les liens entre température et mortalité n'évolueront pas. Autrement dit, l'hypothèse sous-jacente est que la

18. Lien vers une présentation de la plateforme Atmo Data : ☞

19. Lien vers le communiqué de presse de l'OMM datant du 7 septembre 2022 : ☞

relation entre ces deux variables qui a été constatée sur la période 1990-2019 serait conservée à horizon 2070.

Cela pourrait en partie faire sens pour plusieurs raisons. Tout d'abord, la littérature relative à la température humide laisse suggérer l'existence de seuils universels ne pouvant être supportés par le corps humain sur une longue durée. Pour rappel, le seuil obtenu était une température humide de 35°C. De plus, la période retenue pour l'entraînement des modèles a permis d'intégrer certaines mesures d'adaptation prises au début du siècle. En effet, la séparation du jeu de données historique en deux périodes a permis de mettre en évidence des relations différenciées avant et après l'épisode de canicule mortel d'août 2003. Ces différences pouvaient, entre autres, être expliquées par l'adoption de mesures d'adaptation à partir de 2004 ayant pour objectif de réduire les effets potentiels des canicules sur les personnes à risque (celles isolées, notamment).

Toutefois, d'autres mesures d'adaptation pourraient être prises dans les prochaines décennies, contribuant à diminuer l'impact des phénomènes de chaleurs extrêmes sur la mortalité. D'une part, certaines politiques d'aménagement du territoire pourraient être mises en place à l'échelle des zones urbaines, notamment. En ce sens, la plantation d'arbres est une solution avantageuse pour réduire les températures en ville, pour plusieurs raisons. Celles-ci sont synthétisées dans le schéma 4.20.

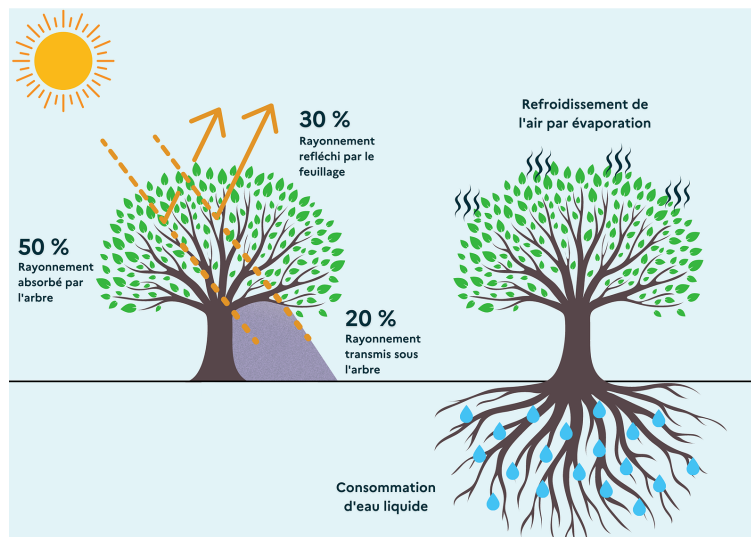
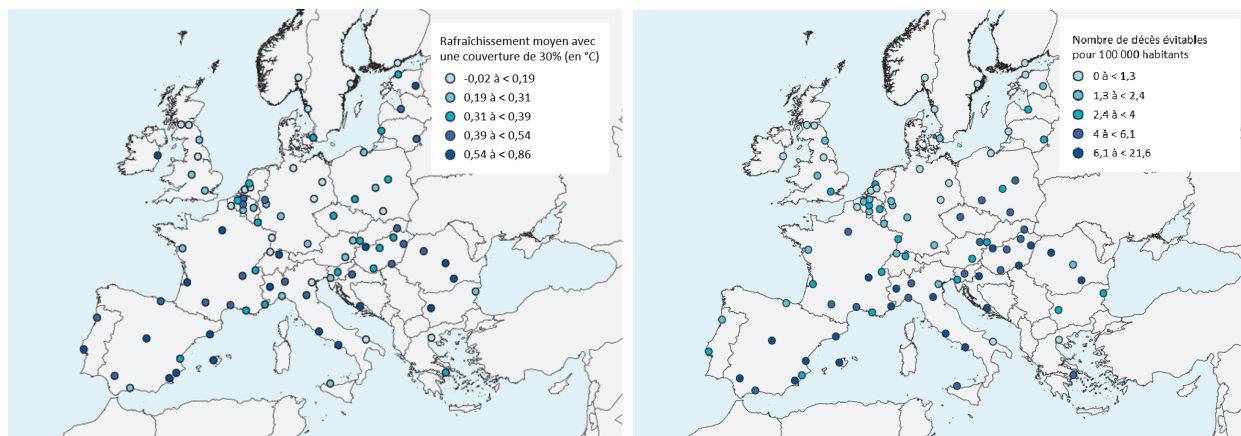


SCHÉMA 4.20 – Absorption du rayonnement solaire et mécanisme d'évapotranspiration (source : Office National des Forêts)

Tout d'abord, en absorbant une partie importante du rayonnement solaire, les arbres produisent une ombre permettant de réduire les températures au sol. Par ailleurs, lors d'une période ensoleillée, le mécanisme d'évapotranspiration va conduire à rafraîchir l'air ambiant

grâce à la libération de l'eau présente au sol et captée par les racines. Ces deux mécanismes permettent donc une réduction de la température extérieure lors des vagues de chaleur. Une réduction de la mortalité attribuable à la chaleur peut donc être anticipée.

Une étude publiée par Iungman et al. [38] a cherché à quantifier l'efficacité de cette mesure, suivant une méthodologie similaire à celle de la présente analyse. Des données publiques disponibles à une maille intra-urbaine (températures et population vivante) ont tout d'abord été recueillies, puis les nombres de décès sont estimés à partir de la répartition connue de la population. La densité d'arbres dans chaque ville est également disponible, ce qui permet d'estimer une relation entre la température moyenne journalière et la couverture de l'espace en arbres. Grâce à cette estimation, il est possible de proposer plusieurs scénarios de températures projetées selon le taux de couverture en arbres retenu. La mortalité estivale est alors comparée dans le scénario actuel et dans celui avec un taux de couverture de 30%. Ce taux est en pratique seulement atteint dans quelques métropoles européennes, telles que Berlin. Le graphique 4.21 permet de représenter la diminution moyenne de la température durant un été ainsi que le nombre de décès pouvant être évités pour 100 000 habitants.



GRAPHIQUE 4.21 – Diminution de la température moyenne journalière (à gauche) et du nombre de décès pour 100 000 habitants (à droite) attribuables à un taux de couverture en arbres de 30% (source : Iungman et al. [38])

La réduction de la mortalité a pu être déterminée à partir de courbes température-mortalité par âge issues de modèles DLNM. Concernant les villes françaises, Nice serait la grande métropole pour laquelle la politique d'adaptation serait la plus efficace. Le nombre de décès pourrait être réduit de 8 pour 100 000 habitants, contre 4 et 6 respectivement pour Paris et Lyon. Si cette réduction peut sembler négligeable, il est utile de noter que l'incertitude mesurée est importante, et que d'autres effets positifs d'une urbanisation plus verte ont été mentionnés (sur la santé mentale notamment).

Par ailleurs, la connaissance croissante des relations entre température et mortalité pourrait aider les autorités à améliorer la gestion des épisodes de canicule à l'avenir. En effet, le programme *Early Warning System* propose une démarche similaire à celle de cette étude afin d'aider à identifier des seuils de températures d'urgence, pouvant se traduire par une surmortalité considérable. Les relations historiques entre la température et la mortalité sont estimées, mais cette fois appliquées à des tendances de températures sur une semaine. Si les prévisions de température sont suffisamment fiables, alors les pouvoirs publics pourraient être davantage en capacité d'adopter des plans de prévention efficaces. Ils pourraient notamment prévenir en avance les populations et mobiliser des moyens en amont.

Conclusion du chapitre

La segmentation ainsi que modèle DLNM pénalisé retenus dans le chapitre précédents ont permis d'établir des projections de la mortalité par département et par jour à l'horizon 2070. Ces dernières sont basées sur des variables de températures différenciées par scénario climatique, ainsi qu'un scénario démographique central. Les sorties ont mis en évidence l'existence d'une incertitude croissante des taux de mortalité, du fait d'un impact significatif des vagues de chaleur futures. Certaines régions, telles que l'Île-de-France, pourraient être particulièrement exposées. Face à ces risques, l'apport de mesures d'adaptation face aux vagues de chaleur a pu être quantifié à l'aide d'une méthodologie simple.

Conclusion

En raison du dérèglement climatique, les entreprises d'assurances sont confrontées à de nouveaux risques pouvant impacter leurs activités, tant sur les biens que sur les personnes. Face à ce contexte nouveau, les superviseurs recommandent aux assureurs de quantifier les conséquences potentielles des phénomènes climatiques futurs sur leurs indicateurs actuariels. L'utilisation d'outils et de données adaptés à la problématique ainsi qu'une étude à petite échelle sont notamment encouragées. L'analyse menée ici propose de répondre à cet objectif dans le cas du risque de mortalité, à partir de données libres d'accès.

Après avoir rappelé l'importance du risque de décès dans les activités des assureurs, une première analyse des facteurs de mortalité est menée, à l'échelle globale et nationale. Celle-ci a permis de mettre en lumière les variables disponibles et pouvant avoir un impact significatif sur la mortalité future. D'une part, les fluctuations à court-terme de la mortalité sont nettement corrélées à celles des températures : davantage d'individus décèdent l'hiver en raison de maladies respiratoires ou cardio-vasculaires, mais des vagues de chaleur peuvent également provoquer des pics de mortalité occasionnels. D'autre part, une variable de tendance permet de capter les différentes causes (progrès de la médecine, augmentation de la qualité de vie) justifiant l'augmentation structurelle de l'espérance de vie à la naissance. C'est ainsi que des variables de tendance et de température ont été retenues.

Des extensions des modèles linéaires généralisés ont été implémentées et comparées, afin d'estimer les nombres de décès dans chaque département de France métropolitaine et chaque jour sur la période 1990-2019. Sachant que les vagues de chaleur seraient amenées à se multiplier et à devenir plus intenses dans les prochaines décennies, un des critères de sélection du modèle a été sa capacité à reproduire des vagues de chaleur historiques. La capacité à rendre compte des faits empiriques, le temps de calcul ainsi que la précision (sur la base d'une séparation en jeu d'entraînement et de test), ont également été retenus. Finalement, le modèle DLNM pénalisé a été choisi, car il permet de capter des liens complexes entre la température et la mortalité confirmées par la littérature, tout en limitant le risque de surapprentissage.

À partir du modèle choisi, des segmentations ont été introduites, afin de capter des différenciations des relations température-mortalité par âge, sexe et zone géographique. Il a ainsi été obtenu que la mortalité attribuable aux températures extrêmes était plus élevée chez les personnes âgées, et que les femmes étaient davantage sensibles aux conséquences des pics de chaleur. Le lien entre la température et la mortalité est également amené à varier dans le temps et dans l'espace, ce qui peut être expliqué par l'adaptation des organismes aux différents climats, ainsi que par la croissance des politiques de prévention des pics de chaleur extrêmes (notamment depuis 2003).

Enfin, une dernière partie a permis d'établir des projections de la mortalité à horizon 2070. Ces dernières sont basées sur des données de projections climatiques ainsi que démographiques. L'analyse temporelle et géographique a mis en avant une dépendance significative de la mortalité annuelle aux vagues de chaleur extrêmes. Celle-ci concernerait davantage les personnes âgées ainsi que les femmes, et se concentrerait autour de la région Île-de-France. Cette observation, couplée à la forte urbanisation de ses départements, met en avant les risques liés aux îlots de chaleur urbain. Face aux conséquences des canicules futures, l'intérêt de mesures d'adaptation (via une urbanisation plus verte, par exemple) a pu être quantifié. Bien que ces dernières permettraient de réduire de manière significative l'incertitude quant à l'évolution des taux de mortalité, certaines catégories à risque pourraient malgré tout conserver une grande probabilité de décès lors d'épisodes exceptionnels.

L'inclusion de facteurs supplémentaires - liés à la qualité de l'air ou aux maladies vectorielles - pourrait permettre de compléter cette analyse. Toutefois, le manque de données historiques et de projections complique leur intégration dans une étude prospective. Par ailleurs, les interprétations géographiques et temporelles relatives à la mortalité future doivent être considérées avec prudence, étant donné la profondeur de la projection ainsi que les disparités géographiques de températures possibles selon les données utilisées.

Cette étude peut toutefois trouver des applications actuarielles en tarification vie et santé, ainsi que pour une analyse de matérialité et de scénarios dans le cadre de l'ORSA. Par ailleurs, la méthodologie proposée ici peut être réutilisée pour la modélisation de certains risques climatiques, tels que le risque inondation. À partir du modèle DLNM utilisé dans cette analyse, il serait en effet possible d'estimer des courbures de persistance du risque différenciées selon le niveau de précipitations. Ainsi, l'actuaire pourrait être en mesure de capter et de comprendre de manière fine la survenance de l'ensemble des types de crues, à la fois éclair (suite à de fortes précipitations sur quelques heures) et lentes (sur plusieurs jours).

Bibliographie

- [1] Lucie Adélaïde, Olivier Chanel et Mathilde Pascal. *Évaluation monétaire des effets sanitaires des canicules en France métropolitaine entre 2015 et 2020*. Bulletin Épidémiologie Hebdomadaire, 2021
- [2] Ben Armstrong, Francesco Sera, Ana Maria Vicedo-Cabrera, Rosana Abrutzky, Daniel Oudin Åström et al. *The Role of Humidity in Associations of High Temperature with Mortality : A Multicountry, Multicity Study*. Environmental Health Perspectives, 2019
- [3] Igotz Aubin et Anne Rolland. *Lignes directrices de la construction des lois de maintien en incapacité et en invalidité*. Institut des Actuaires, 2010
- [4] Antoine Badillet et Benjamin Silva. *Risque climatique : Comment modéliser la sinistralité à horizon 2050 ?* 2020
- [5] Sophie Baillargeon. *Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations*. Université de Laval, 2005
- [6] Alan Barreca, Karen Clay, Olivier Deschenes, Michael Greenstone et Joseph S. Shapiro. *Adapting to Climate Change : The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century*. Journal of Political Economy Volume 124, Number 1, 2016
- [7] Jean-Pierre Besancenot. *Une vague de chaleur meurtrière : les enseignements de l'été 2003 en France*. Géoconfluences, 2004
- [8] Florian Bonnet, Hippolyte d'Albis et Aurélie Sotura. *Les inégalités de revenu entre les départements français depuis cent ans*. Économie et Statistique, 2021
- [9] Adrian Colin Cameron et Pravin K. Trivedi. *Regression-based tests for overdispersion in the Poisson model*. Journal of Econometrics (Volume 46, Issue 3, Pages 347-364), 1990.
- [10] Adrian Colin Cameron et Pravin K. Trivedi. *Microeconometrics : Methods and Applications*. 2005
- [11] Olivier Cantat. *L'îlot de chaleur urbain parisien selon les types de temps*. 2004

- [12] Laura Cohen-Salmon et Dulcy Joyce Gninghayé-Fongang. *Modélisation du risque gel en France*. AXA, 2014
- [13] Covéa. *Changement climatique & Assurance : Quelles conséquences sur la sinistralité à horizon 2050 ? 2022*
- [14] Manon Dal Pont. *Construction d'une table de mortalité d'expérience en assurance emprunteur*. SOGECAP, 2020
- [15] Francesca de'Donato, Matteo Scortichini, Manuela De Sario, Annamaria de Martino et Paola Michelozzi. *Temporal variation in the effect of heat and the role of the Italian heat prevention plan*. Public Health, 2018
- [16] Thomas Deroyon. *En 2020, l'espérance de vie sans incapacité à 65 ans est de 12,1 ans pour les femmes et de 10,6 ans pour les hommes*. DREES, 2021
- [17] DRIAS. *Les nouvelles projections climatiques de référence DRIAS 2020 pour la métropole*. 2020
- [18] Yannick Drif, Benjamin Roche et Pierre Valade. *Conséquences du changement climatique pour les maladies à transmission vectorielle et impact en assurance de personnes*. 2020
- [19] Laurent Dupont, Olivier Fliche et Su Yang. *Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier - Document de réflexion*. ACPR, 2020
- [20] Chao-Yu Guo, Tse-Wei Liu et Yi-Hau Chen. *A novel cross-validation strategy for artificial neural networks using distributed-lag environmental factors*. 2021
- [21] EIOPA. *Application guidance on running climate change materiality assessment and using climate change scenarios in the ORSA*. 2022
- [22] EIOPA. *Methodological principles of insurance stress testing - climate change component*. 2022
- [23] Papa Cheikh Fall. *Intégration de facteurs liés à la température dans modélisation du risque de mortalité*. 2021
- [24] Mireille Folkerts, Peter Bröde, W. J. Wouter Botzen, Mike Martinius, Nicola Gerrett et al. *Long Term Adaptation to Heat Stress : Shifts in the Minimum Mortality Temperature in the Netherlands*. Frontiers in Physiology, 2020
- [25] Mireille Folkerts, Peter Bröde, W. J. Wouter Botzen, Mike Martinius, Nicola Gerrett et al. *Sex differences in temperature-related all-cause mortality in the Netherlands*. International Archives of Occupational and Environmental Health, 2022
- [26] Camille Fontès-Rousseau, Rémi Lardellier et Jean-Michel Soubeyrou. *Un habitant sur sept vit dans un territoire exposé à plus de 20 journées anormalement chaudes par été dans les décennies à venir*. INSEE, 2022

- [27] Michel Fromenteau et Pierre Petauton. *Théorie et pratique de l'assurance-vie*. Dunod, 2017
- [28] Data Lab Galea. *Impact de la météo, et plus particulièrement des pics de chaleur humide, sur la Surmortalité*. 2022
- [29] Antonio Gasparrini, Ben Armstrong et Mike Kenward. *Distributed lag non-linear models*. Statistics in Medicine, 2010
- [30] Antonio Gasparrini, Pierre Masselot, Matteo Scortichini, Rochelle Schneider, Malcolm Mistry et al. *Small-area assessment of temperature-related mortality risks in England and Wales : a case time series analysis*. The Lancet, 2022
- [31] Antonio Gasparrini, Fabian Scheipl, Ben Armstrong et Michael Kenward. *A penalized framework for distributed lag non-linear models*. Journal of the International Biometric Society, 2017
- [32] Antonio Gasparrini, Yuming Guo, Masahiro Hashizume, Patrick L Kinney, Elisaveta Petkova et al. *Temporal Variation in Heat-Mortality Associations : A Multicountry Study*. Environmental Health Perspectives, 2015
- [33] Valentin Germain. *Prise en compte du changement climatique dans la modélisation des risques biométriques et financiers*. Milliman, 2022
- [34] Vanessa Gordon-Dseagu, Martin Wiseman, Kate Allen, Judy Buttriss et Christine Williams. *Troubling assumptions behind GBD 2019 on the health risks of red meat*. The Lancet, 2022
- [35] Catherine Ha, Elsa Decool et Christine Chan Chee. *Mortalité des personnes souffrant de troubles mentaux - Analyse en causes multiples des certificats de décès en France, 2000-2013*. 2017
- [36] Joseph M. Hilbe. *Negative Binomial Regression*. Cambridge University Press, 2011
- [37] Intergovernmental Panel on Climate Change - Core Writing Team. *Synthesis Report of the IPCC Sixth Assessment Report (AR6)*. 2022
- [38] Tamara Iungman, Marta Cirach, Federica Marando, Evelise Pereira Barboza, Sasha Khomenko et al. *Cooling cities through urban green interventions : A health impact assessment in European cities*. ISEE Conference Abstracts, 2022
- [39] Stéphane Loisel et Yahia Salhi. *Biometric and Surrender Risks*. Munich RE, 2015
- [40] Damien Loureiro. *Utilisation de la DSN et de l'open data pour élaborer et expliquer un zonier incapacité*. Malakoff Humanis, 2020
- [41] George Luber et Michael McGeehin. *Climate change and extreme heat events*. American Journal of Preventive Medicine, 2008

- [42] Tiina Mäkinen, Raija Juvonen, Jari Jokelainen, Terttu Harju, Ari Peitso, Aini Bloigu et al. *Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections*. Respiratory Medicine, 2008
- [43] Eurydice Maucuer. *Projection de la fréquence des inondations et sécheresses en France métropolitaine à horizon 2050*. Sia Partners, 2023
- [44] C. F. McSweeney, R. G. Jones, R. W. Lee, D. P. Rowel. *Selecting CMIP5 GCMs for downscaling over multiple regions*. 2014
- [45] Vito Muggeo. *Analyzing Temperature Effects on Mortality within the R Environment : the Constrained Segmented Distributed Lag Parameterization*. Università di Palermo, 2010
- [46] Mathilde Pascal, Vèrène Wagner et Magali Corso. *Évolution de la relation température-mortalité en France depuis 1970*. Santé Publique France, 2022
- [47] Gaëlle Pincemin. *Risques climatiques et mortalité, impact du risque canicule à l'horizon 2070*. Prim'Act, 2021
- [48] Frédéric Planchet. *Modèles de durée*. ISFA, 2023
- [49] Frédéric Planchet et Pierre Thérond. *Modèles financiers et analyses de risque dynamiques en assurance. Les garanties plancher des contrats en unités de compte*. ISFA, 2005
- [50] Frédéric Planchet et Karl Bantigny. *Mortalité prospective dans les travaux de l'INSEE*. Prim'Act, 2023
- [51] Keywan Riahi, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill et al. *The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications : An overview*. Global Environmental Change, 2017
- [52] Santé Publique France. *Pollution atmosphérique : quels sont les risques ?* 2022
- [53] Michel Schneider et Lola Corre. *Évolutions passée et future des vagues de chaleur en France*. Météo-France - Direction de la Climatologie et des Services Climatiques, 2022
- [54] Nicholas Steel, John Ford et Jurgen Schmidt. *Slowdown in Life Expectancy Improvements for European Countries From 2000 to 2019*. Innovation in Aging, 2021
- [55] Damien Tichit. *Construction d'un modèle ALM pour l'analyse de l'impact d'une remontée des taux sur la solvabilité d'un assureur vie*. SIA Partners, 2019
- [56] Aurelio Tobías, Ben Armstrong et Antonio Gasparrini. *Investigating Uncertainty in the Minimum Mortality Temperature. Methods and Application to 52 Spanish Cities*. 2017
- [57] Detlef P. van Vuuren, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson et al. *The representative concentration pathways : an overview*. 2011

- [58] Daniel Vecellio, S. Tony Wolf, Rachel Cottle, W. Larry Kenney. *Evaluating the 35°C wet-bulb temperature adaptability threshold for young, healthy subjects*. Journal of Applied Physiology, 2022
- [59] Ana Maria Vicedo-Cabrera, Francesco Sera et Antonio Gasparri. *Hands-on Tutorial on a Modeling Framework for Projections of Climate Change Impacts on Health*. Epidemiology, 2019
- [60] Quang H. Vuong. *Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses*. Econometrica Vol. 57, No. 2. 1989
- [61] Paul Wilson. *The misuse of the Vuong test for non-nested models to test for zero-inflation*. Economics Letters, Volume 127, pages 51-53. 2015
- [62] Qi Zhao, Shan-Shan Li, Yu-ming Guo. *Time-varying distributed lag non-linear model : Using temperature-mortality association as an example*. Journal of Environmental and Occupational Medicine, 2020

A

Construction des scénarios RCP

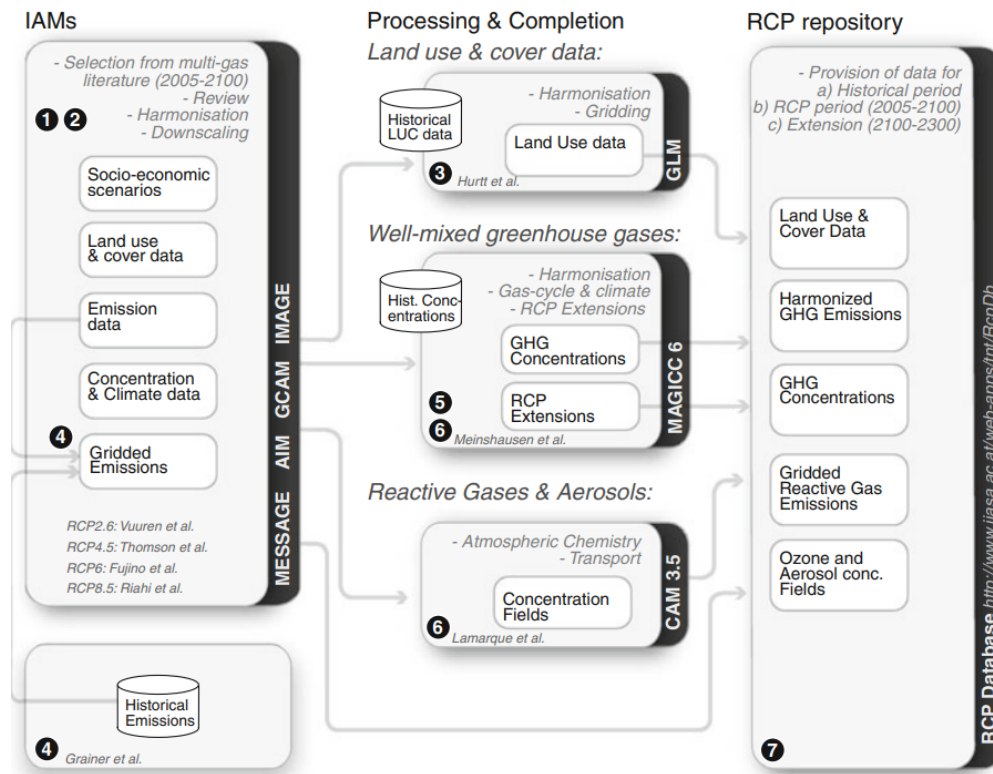
Pour la première fois, les réflexions ne sont plus menées à partir de "futurs possibles" liés à un grand nombre de facteurs (y compris comportementaux), mais à des évolutions des émissions de gaz à effet de serre définies *a priori*. De plus, les nouveaux scénarios, dits RCP (*Representative Concentration Pathways*), ont été construits par la communauté scientifique et non le GIEC lui-même. Quatre objectifs majeurs pour ces scénarios RCP ont été mentionnés par Van Vuuren et al. [57] :

- 1) Se baser sur des scénarios publiés dans la littérature existante, développés indépendamment par différents groupes de modélisation et, en tant qu'ensemble, être "représentatifs" de la littérature totale ;
- 2) Fournir une description plausible et cohérente de l'avenir ;
- 3) Fournir des informations sur toutes les composantes du forçage radiatif nécessaires à la modélisation du climat et de la chimie atmosphérique ;
- 4) Avoir des hypothèses harmonisées sur l'année de référence pour les émissions et permettre une transition continue entre les analyses des périodes historiques et futures ;
- 5) Couvrir la période allant jusqu'à 2100, avec des informations mises à disposition pour les siècles suivants.

Ces objectifs permettent, ensuite, de fournir des projections de variables atmosphériques à une maille temporelle et géographique fine, selon les étapes décrites dans le schéma A.1 issu de Van Vuuren [57].

Le schéma ci-dessus permet de synthétiser les étapes suivantes :

- 1) Quatre scénarios existants sont repris de la littérature ;
- 2) Ces scénarios sont mis à jour en tenant compte des dernières avancées en modélisation



SCHEMA A.1 – Schéma d'obtention des sorties des modèles RCP (source : Van Vuuren et al. [57])

d'évaluation intégrée et des dernières données relatives aux émissions de polluants, notamment.

- 3) Les données d'émissions issues de groupes individuels sont harmonisées puis une réduction d'échelle est appliquée à une maille quadrillage de taille $0,5 \times 0,5$
- 4) Un modèle climatique simple permet de transformer les données d'émissions en données de concentrations. Pour rappel, les émissions sont la quantité de polluants directement rejetée dans l'atmosphère, tandis que la concentration est définie comme la proportion d'une substance dans l'air¹.
- 5) Des extensions simples des scénarios RCP sont fournies entre 2100 et 2300.
- 6) Un dépôt de données permet d'accéder simplement au téléchargement des données issues des scénarios RCP.

Les variables atmosphériques projetées pourront être utilisées comme données en entrée dans les modèles climatiques afin de fournir les projections de variables de températures, notamment.

1. Des explications sont fournies dans un article du Monde : ↗

B

Description des scénarios SSP

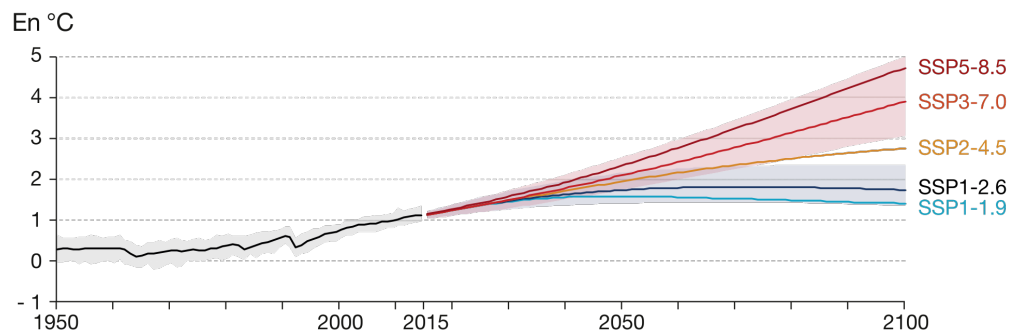
Le dernier rapport du GIEC [37], publié en 2021, est riche d'enseignements sur les origines et les conséquences futures potentielles du dérèglement climatique. Tout d'abord, il indique que l'être humain est en grande partie responsable du changement climatique sur les deux derniers siècles. Ensuite, il évoque les conséquences portant à la fois sur la fréquence, l'intensité et l'emplacement des phénomènes météorologiques extrêmes. Ce rapport évoque également des projections potentielles de ces événements catastrophiques à partir de scénarios d'augmentation de la température. Ces scénarios ont été mis à jour par rapport aux RCP, afin de tenir compte de l'évolution simultanée de plusieurs facteurs : économiques, technologiques, démographiques et institutionnels. On parle de scénarios SSP (*Shared Socioeconomic Pathways*). Leur description est issue de Riahi et al. [51] :

- **SSP1 (Durabilité)** : transition progressive mais générale vers un modèle économique plus durable, avec une croissance matérielle faible mais aussi moins coûteuse en ressources et en énergie. L'ensemble des mesures, dans l'éducation, la santé et la croissance permettent une réduction des inégalités à l'international et à l'intérieur des pays.
- **SSP2 (Milieu de la route)** : les tendances sociales, économiques et technologiques ne s'écartent pas sensiblement des modèles historiques malgré quelques économies dans l'utilisation des ressources et de l'énergie. Le développement et la croissance des revenus se déroulent de manière inégale.
- **SSP3 (Rivalités régionales)** : les pays se concentrent sur la réalisation d'objectifs de sécurité énergétique et alimentaire dans leur propre région. En particulier, la faible importance accordée à une solution internationale pour l'environnement entraîne une forte dégradation de l'environnement dans certaines régions.
- **SSP4 (Inégalités)** : les inégalités se creusent dans le sens d'une société à deux vi-

tesses : une population fortement éduquée et à forts revenus, et une population faiblement éduquée et disposant de faibles revenus. La consommation d'énergie se sépare en consommation de combustibles à forte et à faible intensité de carbone. Les politiques environnementales se concentrent dans les pays développés.

- **SSP5 (Développement alimenté par les combustibles fossiles)** : développement économique et social généralisé à partir d'une exploitation d'abondantes ressources en combustibles fossiles. Les modes de vie adoptés dans le monde entier mobilisent également une forte quantité de ressources et d'énergie. Les problèmes environnementaux locaux, comme la pollution atmosphérique, sont gérés avec succès.

Ces scénarios SSP sont complétés par une valeur de forçage radiatif qui permet de les rapprocher des anciens scénarios RCP. La nouveauté est que la valeur de forçage radiatif est expliquée par des scénarios sociaux et économiques. C'est ainsi que sont considérés les scénarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 et SSP5-8.5. Les projections de forçage radiatif sont censées être cohérentes avec le scénario de développement. Par exemple, le scénario SSP5-8.5 est associé à un développement largement dépendant de l'usage de combustibles fossiles, conduisant à un niveau de forçage radiatif égal à $8,5W/m^2$ à l'horizon 2100. Le graphique B.1 permet de visualiser les projections d'évolutions de la température à la surface terrestre par rapport à l'ère pré-industrielle.



GRAPHIQUE B.1 – Augmentation de la température à la surface du globe par rapport à l'ère pré-industrielle selon les scénarios SSP (source : Ministère de la Transition Écologique)

On y constate que la température à la surface du globe serait en hausse par rapport à aujourd'hui dans tous les scénarios. Les niveaux projetés sont cohérents avec les scénarios RCP : l'augmentation des températures pourrait atteindre 4 degrés dans le scénario extrême de forçage radiatif à $8,5W/m^2$ à horizon 2100.



Réglementation relative à l'utilisation des tables de mortalité

Plusieurs textes fondamentaux encadrent la tarification des produits d'assurance-vie. Des codes différents ont été instaurés selon la nature de l'émetteur de ces contrats, et le ministère chargé de son application varie. Le tableau C.1 présente le code s'appliquant pour chaque émetteur de produits d'assurance-vie ainsi que le ministère chargé de ses éventuelles révisions.

<i>Émetteur du produit</i>	<i>Code</i>	<i>Ministère</i>
Entreprises d'assurance	Code des assurances	Économie et Finances
Institutions de prévoyance	Code de la sécurité sociale	Chargé de la sécurité sociale
Mutuelles	Code de la mutualité	Chargé de la mutualité

TABLEAU C.1 – Réglementation régissant la tarification des produits d'assurance-vie

Les règles relatives à la tarification en assurance-vie restent proches entre les différents textes. Ainsi, les deux mêmes familles de tables de mortalité sont autorisées¹ :

- a) Des tables homologuées par arrêté du ministre mentionné dans le tableau ci-dessus, établies par sexe, sur la base de :
 - populations d'assurés pour les contrats de rente viagère ;
 - données publiées par l'INSEE pour les autres contrats.
- b) Des tables établies ou non par sexe par l'entreprise d'assurance et certifiées par un actuaire indépendant de cette entreprise. Celles-ci sont établies d'après des données d'expérience de l'assureur, ou des données d'expérience démographiquement équivalentes.

D'une part, les tables homologuées sont similaires entre les codes. Depuis le 1^{er} janvier

1. Conformément aux articles A132-18 du Code des assurances, A932-3-11 du Code de la sécurité sociale et A223-8 du Code de la mutualité

2007, il s'agit respectivement des tables :

- TGF05/TGH05 pour les contrats de rentes viagères
- TH00-02/TF00-02 pour les autres contrats

Ces tables, produites par l'INSEE, sont différenciées par sexe et renseignent le nombre de survivants à un âge donné parmi chaque génération comprise entre 1900 et 2005 (sur une population de départ fixée). Elles sont basées sur des observations de portefeuilles de rentiers : ainsi, les tables TGF05 et TGH05 concernent environ 700 000 rentes liquidées sur la période 1993-2005². Par ailleurs, ces tables sont prospectives : des extrapolations ont ainsi été appliquées afin de projeter les nombres de survivants après 2005. D'autre part, pour tous les contrats "en cas de vie" autres que les rentes viagères, les assureurs doivent appliquer un décalage d'âge aux tables de mortalité. L'Institut des Actuaire recommande³ d'appliquer ce décalage aux quotients de mortalité.

2. Lien vers la source (Institut des Actuaire) : [↗](#)

3. Lien vers la source : [↗](#)

D

Incapacité, invalidité et contrats de prévoyance

À la suite d'une maladie ou un accident, un assuré peut se retrouver dans l'incapacité physique d'exercer une activité professionnelle de façon temporaire ou définitive. À ce stade, il convient de distinguer :

- l'**incapacité**, dont l'origine est liée à l'activité professionnelle ;
- l'**invalidité**, dont l'origine est tierce.

Ces institutions offrent alors des contrats proposant une indemnisation complémentaire à celle de l'Assurance Maladie, lorsque l'assuré n'est plus en mesure d'assurer sa profession en raison de sa santé physique. L'objectif pour l'assuré est de pouvoir couvrir l'intégralité de son salaire en cas de difficultés de santé. En pratique, comme rappelé par Planchet [48], les risques d'invalidité et d'incapacité sont souvent associés par les assureurs, sous la forme du risque "arrêt de travail". La raison sous-jacente est qu'une incapacité peut se muer en invalidité. Les mutuelles ont donc besoin d'évaluer au mieux les risques pouvant interférer sur le risque d'incapacité ou d'invalidité. Dans ce contexte, la loi Evin de 1989 a introduit des tables de maintien en incapacité et en invalidité. Ces tables permettent de calculer les probabilités de maintien dans un état, ou de passage dans un autre état pour un individu d'âge d'entrée dans un état x et d'ancienneté dans l'état t .

De la même manière que la mortalité, des variables géographiques et socio-professionnelles sont couramment utilisées par les actuaires pour la modélisation du risque santé. C'est par exemple le cas de cet article de Aubin et Rolland [3] présentant une méthodologie de construction des tables d'expérience de maintien en incapacité/invalidité. Dans son mémoire, Loureiro [40] propose également une méthodologie de création d'un zonier incapacité, combinant des données de portefeuille de Malakoff Humanis et des données socio-économiques de l'INSEE.

De telles études sont cependant principalement basées sur des données d'assureur, ce qui limite les perspectives d'analyses prospectives utilisant exclusivement des données ouvertes. De manière générale, les analyses relatives au risque santé visant à tarifier les contrats de complémentaires santé mobilisent des données individuelles à caractère sensible. Leur publication est donc particulièrement encadrée en France, conformément à la loi Sapin II et à la RGPD (Règlement Général sur la Protection des Données). Une étude sur le sujet à partir de données publiques doit donc s'effectuer soit à partir de données agrégées, soit issues de pays ne disposant pas de la même réglementation des données (comme le Mexique). Bien que la base Open Damir¹ permette d'étudier de façon fine l'ensemble des postes de dépenses d'Assurance Maladie, l'information géographique et temporelle est peu précise. Les postes de dépenses sont en effet précisées aux mailles mensuelles et régionales. C'est pourquoi l'analyse des risques santé est exclue de cette étude. Seul le risque de mortalité sera analysé ici.

1. Lien vers la base Open Damir de l'Assurance Maladie : ↗

E

Autres risques liés au changement climatique sur la mortalité

Cette annexe fournit une description des risques relatifs au changement climatique sur la mortalité n'ayant pas pu être exploités dans cette étude.

La qualité de l'air

Une mauvaise qualité de l'air peut être un facteur aggravant de la mortalité liée aux maladies cardio-vasculaires et des cancers. Les effets peuvent se matérialiser à court terme et à long terme, et sont d'autant plus importants sur certaines catégories à risque (fumeurs et personnes atteintes de maladies respiratoires).

À court terme, d'après Santé Publique France [52], l'exposition à des polluants, même faible, peut constituer un facteur d'aggravation du risque d'infarctus du myocarde et des pathologies respiratoires chroniques pour les individus souffrant déjà de telles maladies. Concrètement, cela peut se manifester par des irritations oculaires ou des voies respiratoires, des crises d'asthme, une exacerbation de troubles cardio-vasculaires et respiratoires pouvant conduire à une hospitalisation, et dans les cas les plus graves au décès. Une étude de cet institut a montré qu'une augmentation de $10 \mu\text{g}/\text{m}^3$ des niveaux de PM10 (les particules de diamètre inférieur à $10 \mu\text{g}$) du jour et des cinq jours précédents se traduit par une augmentation de 0,5% de la mortalité non-accidentelle. Les individus les plus à risque sont les personnes âgées de 75 ans et plus. Selon cette étude, l'impact sur la mortalité serait plus marqué pendant l'été, ce qui suggère l'intérêt d'inclure un indicateur de qualité de l'air comme facteur additionnel de mortalité dans une modélisation, en plus de la température.

À long terme, ce sont les particules les plus fines (les PM2.5, c'est-à-dire les particules

de diamètre inférieur à $2,5 \mu g$) qui posséderaient l'impact le plus significatif sur la mortalité. L'exposition prolongée constitue alors un facteur de risque pour le développement ou l'aggravation de maladies chroniques telles que des cancers, des pathologies cardio-vasculaires, neurologiques et respiratoires. La littérature mesurant quantitativement cet impact sur la mortalité est moins fournie, car une telle étude nécessiterait de suivre des individus sur une longue période et de connaître de façon fine leur exposition à ces particules fines. En particulier, cela nécessite d'obtenir des données de suivi régulier des individus sur leur exposition à la pollution, et pas seulement leur lieu de décès. Pour ces raisons, une étude basée exclusivement sur des données de décès ne pourra intégrer que des effets de court terme de la pollution. Notons toutefois, à titre indicatif, que l'*ACS Cancer Prevention II Study* de 1996 a montré, à partir d'un suivi de près de 500 000 individus sur une durée de 16 ans, qu'une augmentation de $10 \mu g/m^3$ entraînerait toutes choses égales par ailleurs un surplus de mortalité de 4%.

Les maladies infectieuses

Au-delà des conséquences liées aux vagues de chaleur, les températures élevées liées au réchauffement de la planète possèdent un impact sur la mortalité à travers la transmission de maladies dites "vectorielles". Il s'agit de maladies infectieuses, telles que la dengue ou le paludisme. Elles sont transmises par des vecteurs, tels que des insectes et des acariens hématophages. Après avoir été transmis par l'hôte sur lequel le vecteur a effectué son repas sanguin, l'agent infectieux va se répliquer ou se transformer dans une période (dite d'incubation) comprise entre 5 et 15 jours. Les conséquences de l'infection varient selon l'agent transmetteur, et peuvent avoir un impact direct sur la mortalité, mais aussi indirect, via des troubles mentaux par exemple.

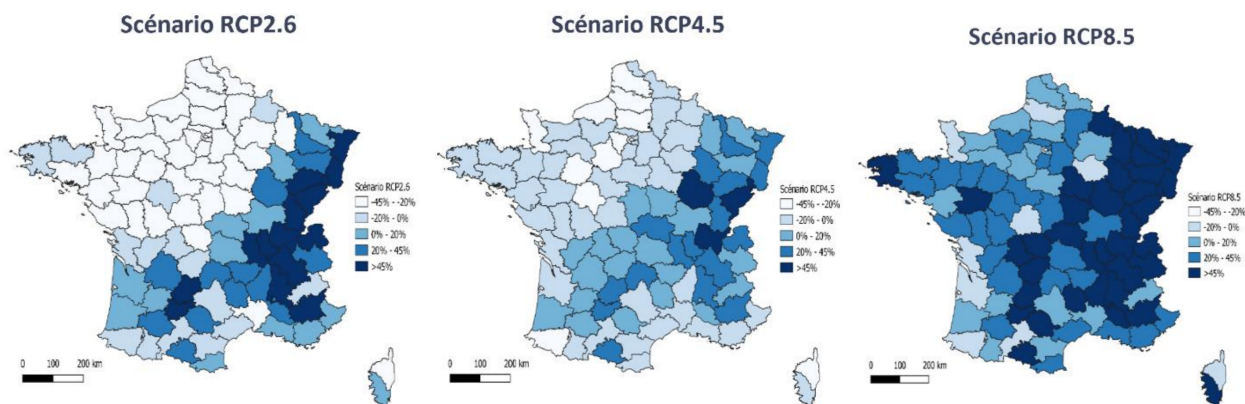
L'importance que ces maladies vectorielles pourront prendre demain est à considérer sachant qu'aujourd'hui, selon l'OMS, la moitié de la population mondiale serait déjà exposée à chaque instant à un agent pathogène transmis par un vecteur. Les températures et les précipitations sont autant de paramètres climatiques à inclure pour modéliser à la fois le développement et la survie des vecteurs. En effet, la hausse des températures et des précipitations conduit à diminuer la période d'incubation du vecteur (en particulier pour les moustiques) et à augmenter la période de transmissibilité des maladies. Ces facteurs climatiques affecteraient également positivement la reproduction des moustiques, de sorte que le nombre de moustiques et le risque que ces derniers transmettent des agents pathogènes seraient amenés à augmenter. Par exemple, d'après Drif et al. [18], une hausse de $1^\circ C$ de la température locale au mois de janvier augmenterait la probabilité de présence du moustique

de 7%, à condition que la température soit supérieure à 10,4°C pour que cette présence soit non-nulle.

Les événements météorologiques extrêmes, les infections transmises par l'eau et la santé mentale

Les pertes humaines des catastrophes naturelles (tempêtes, submersions ou feux de forêt) étant généralement considérées comme négligeables, la littérature sur le sujet et sa modélisation est peu fournie. Cependant, certains épisodes d'inondation, notamment, peuvent se matérialiser par des conséquences humaines directes importantes. Ainsi, les inondations qui ont frappé le Pakistan durant l'été 2022 ont généré plus de 1 500 décès et 12 000 personnes blessées. Ces pertes humaines sont également dépendantes aux ressources disponibles et aux mesures mises en place dans chaque pays.

La fréquence et la sévérité des inondations sont chacune amenées à évoluer en France en 2050, avec une disparité d'évolution selon les régions. Cela peut être constaté avec des projections de coûts d'inondation. Dans le scénario le plus pessimiste (RCP 8.5), le groupe d'assurance Covéa a prédit dans son livre blanc [13] une augmentation de 110% des dommages futurs dus aux inondations par débordement en 2050 par rapport à la période 2008-2018. Ces dégâts supplémentaires seraient principalement occasionnés par des crues éclair, en hausse de 130% sur la période. Cependant, ces prévisions sont fortement dépendantes des hypothèses de scénario climatique. En effet, un rapport du cabinet de conseil SeaBird [4] a comparé trois prévisions d'évolution des coûts liés au risque inondation selon 3 scénarios RCP, du plus optimiste (RCP 2.6) au plus pessimiste (RCP 8.5). Celles-ci sont représentées dans le graphique E.1.



GRAPHIQUE E.1 – Évolution des coûts liés au risque inondation par département et selon les scénarios RCP (de gauche à droite : RCP 2.6, RCP 4.5 et RCP 8.5) (source : SeaBird)

Il en ressort une sensibilité de l'évolution des coûts aux scénarios plus ou moins importante selon les régions de France : alors que les départements de la côte atlantique ou du bassin méditerranéen semblent relativement épargnés, les régions de l'Est de la France (Grand-Est, Bourgogne-Franche-Comté et Auvergne-Rhône-Alpes) sont celles qui, selon ce rapport, seraient le plus touchées par une évolution défavorable du dérèglement climatique.

Cette progression de la fréquence et de la sévérité des inondations est ainsi un élément à considérer dans une étude prospective de mortalité. En plus des inondations, les catastrophes les plus graves du dernier siècle incluent la Grande famine de Chine (entre 11 et 40 millions de décès entre 1958 et 1961) et le tsunami dans l'Océan Indien (entre 216 000 et 250 000 morts en 2004). Au-delà, ces catastrophes climatiques peuvent avoir un impact indirect significatif sur la morbidité et la mortalité en favorisant l'émergence de maladies physiques (via l'insalubrité de l'eau) ou mentales (stress post-traumatique). La précarité qui résulte de ces catastrophes peut aussi compliquer l'accès aux soins et renforcer le risque de conflits, bien que l'impact sur la mortalité de ces conséquences potentielles soit délicat à quantifier.

Les conséquences mentales de ces catastrophes naturelles sont également parfois mentionnées. Le terme d'*écoanxiété* réfère ainsi à une "peur chronique de la catastrophe environnementale", d'après l'*American Psychological Association*. Il s'agit d'une anxiété anticipative de l'incertitude des effets du changement climatique, qui peut être renforcée par la multiplication des catastrophes naturelles. D'après le GIEC, les problèmes de santé mentale sont amenés à augmenter à l'échelle mondiale du fait de l'évolution du climat, et affecter particulièrement les enfants, les adolescents ainsi que les personnes âgées. La santé mentale est un facteur à ne pas négliger dans une analyse de la mortalité et de l'incapacité/invalidité, car elle conduit à une diminution de la productivité au travail, de la qualité de vie et peut avoir un impact indirect sur l'âge de décès. Une étude de Ha et al. [35] a ainsi établi, à partir des données des causes de décès multiples issues des certificats de décès entre 2000 et 2013, que l'âge de décès était en moyenne plus bas pour les individus atteints de troubles mentaux. La prise en compte de la santé mentale nécessiterait toutefois la mise à disposition de données personnelles sensibles sur les individus, et ne sera donc pas l'objet de ce rapport.

F

Test de surdispersion des données de comptage

Afin de déterminer si l'hypothèse d'équidispersion sous-jacente à la loi de Poisson est vérifiée, il est possible de construire un test, implémenté dans le package **AER** de R. Ce test, construit par Cameron et al., a fait l'objet de plusieurs publications par ces auteurs. La première, de 1990 [9], détaille la construction du modèle. La seconde, en 2005 [10], en fournit un résumé. L'explication ci-dessous est issue de ce second article. Tout d'abord, il est nécessaire de construire les variables estimées par le modèle. En notant n le nombre d'observations du jeu de données, et k le nombre de variables explicatives retenues, soit

$$\hat{\mu}_i = \exp \left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k \right)$$

la prédiction de la variable d'intérêt μ pour l'observation $i \in \{1, \dots, n\}$. Soit également

$$\phi_i = \frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i}$$

pour $i \in \{1, \dots, n\}$. Une régression linéaire simple de la forme

$$\phi_i = \alpha \frac{g(\hat{\mu}_i)}{\hat{\mu}_i} + \varepsilon_i,$$

où ε_i est le terme d'erreur, est ensuite estimée. La fonction g désigne une transformation associée au type de modèle alternatif considéré. Par exemple, pour une loi quasi-Poisson, la fonction g est l'identité $\text{Id} : x \mapsto x$, auquel cas l'estimateur $\hat{\alpha}$ vaut simplement la moyenne empirique de z_i .

La valeur $1 + \hat{\alpha}$ sera le coefficient de dispersion estimé. Sous l'hypothèse nulle (H_0) : $\alpha = 0$

d'équidispersion, le t de Student associé à l'estimateur $\hat{\alpha}$ suit approximativement une loi normale centrée réduite $\mathcal{N}(0, 1)$. La comparaison de cette statistique avec les quantiles d'une loi normale $\mathcal{N}(0, 1)$ fournit alors le résultat du test d'équidispersion, pour un risque de première espèce donné.

G

Absence de biais des modèles linéaires généralisés

Un des avantages importants des modèles linéaires généralisés est l'absence de biais des prédictions. Dans le contexte de la présente étude, cette propriété implique que le modèle retrouve le nombre moyen de décès associé à l'ensemble de la période historique. Cette annexe présente la démonstration de cette propriété. Soit un échantillon $y = (y_1, \dots, y_n)$, supposé tiré de variables aléatoires Y_1, \dots, Y_n indépendantes dont la loi (non-dégénérée) appartient à la famille exponentielle. Pour tout $i \in \{1, \dots, n\}$, soit f_i la densité de Y_i qui est alors définie par :

$$f_i(y_i|\theta_i, \phi) = \exp \left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right].$$

Sous ces conditions, il est possible de démontrer que :

$$\mu_i = b'(\theta_i) \quad (*)$$

$$\text{Var}(Y_i) = b''(\theta_i)a(\phi). \quad (**)$$

De plus, par construction des modèles linéaires généralisés :

$$\mu_i = g^{-1}(\mathbf{x}_i^\top \beta) \quad (***)$$

où $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,j})$ et où $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ est le vecteur des coefficients du GLM à estimer. Pour cela, la vraisemblance du modèle, définie par :

$$\mathcal{L}_n(\beta) = \prod_{i=1}^n f_i(y_i|\theta_i, \phi),$$

doit être maximisée, ce qui revient à minimiser sa log-vraisemblance. Cette dernière est définie par :

$$\ell_n(\beta) = -\frac{1}{n} \ln(\mathcal{L}_n(\beta)) = -\frac{1}{n} \sum_{i=1}^n \left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right].$$

La condition de premier ordre de minimisation est :

$$\forall j \in \{0, 1, \dots, n\}, \frac{\partial \ell_n}{\partial \beta_j} = 0.$$

L'existence d'une solution est admise, et sa forme explicite n'est pas recherchée. En pratique, il n'existe pas de solution analytique, sauf dans des cas très particuliers. Soit $j = 0$. Une *chain-rule* fournit :

$$\begin{aligned} \frac{\partial \ell_n}{\partial \beta_j} &= \frac{\partial \ell_n(\beta_j)}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \beta_j} \\ &= \frac{\mu_i - \bar{y}}{\text{Var}(Y_i)} (g^{-1})'(\mathbf{x}_i^\top \beta). \end{aligned} \quad (*, ** \text{ et } ***)$$

La simplification des calculs dépend de la fonction g : l'important est que g soit strictement monotone, auquel cas $(g^{-1})'$ ne s'annule jamais. Dans les faits, cette condition pose rarement problème. Pour le cas d'un GLM avec loi de Poisson, $g = \ln$, donc $g^{-1} = (g^{-1})' = \exp$. Il suit alors $b'(\theta) = \exp(\theta)$. $(g^{-1})'(\mathbf{x}_i^\top \beta)$ ne s'annule jamais, et en supposant de plus $\text{Var}(Y_i) \neq 0$, il vient :

$$\frac{\partial \ell_n}{\partial \beta_j} = 0 \iff \mu_i = \bar{y}.$$

Or, en réutilisant (*) et (***) : $\mu_i = \exp(\mathbf{x}_i^\top \beta)$. Ainsi, la moyenne prédite $\widehat{\mu}_i = \mu_i(\widehat{\beta})$ découlant du vecteur de coefficients prédits β vérifie bien : $\widehat{\mu}_i = \bar{y}$. Autrement dit, la moyenne prédite est égale à la moyenne de l'échantillon : les GLM (dans le cas où la fonction de lien canonique est $g = \ln$) sont donc bien une méthode sans biais.



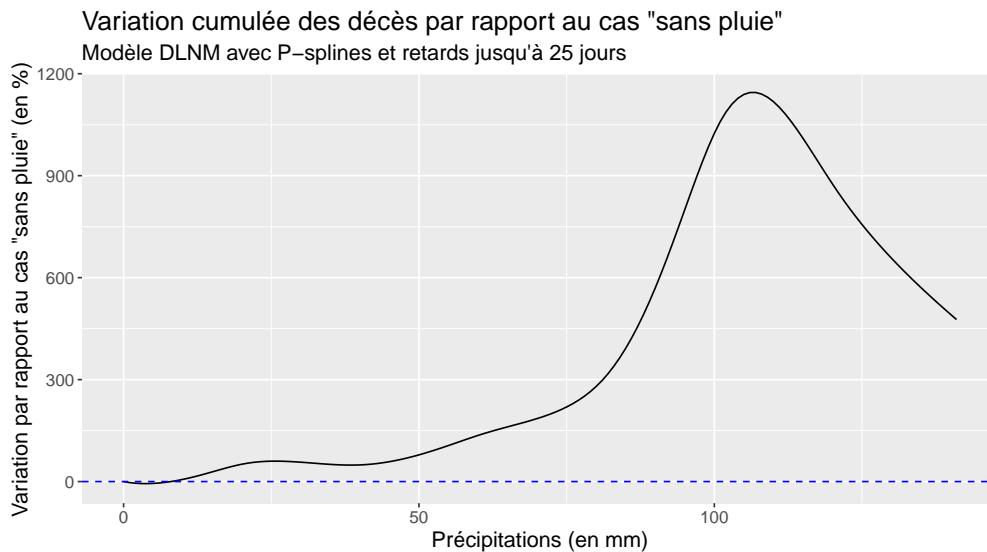
Vers une modélisation du lien précipitations-mortalité

Les précipitations pourraient également être une variable intéressante pour l'analyse de la mortalité. En effet, dans son dernier exercice climatique, l'ACPR préconise d'y inclure les conséquences d'une rupture de barrage. Ce risque pourrait intervenir du fait de précipitations intenses faisant suite à une période prolongée de sécheresse. L'exemple considéré par le superviseur est celui d'une rupture du barrage de Serre-Ponçon, situé entre les départements des Hautes-Alpes et des Alpes-de-Haute-Provence, et qui aurait des répercussions sur 17 communes situées à proximité. Si les barrages et les digues constituent une mesure d'atténuation utile face au risque inondation, leur rupture peut amener à une montée des eaux soudaines et des conséquences accrues pour les populations¹. De plus, il a été vu que le risque inondation est amené à augmenter, tant en termes de fréquence que de sévérité, selon les scénarios climatiques. Inversement, l'absence prolongée de précipitations pourrait avoir un impact sur la capacité des individus à se protéger des épisodes de chaleur. Par exemple, en 2022, une centaine de communes françaises avaient été privées d'eau potable en raison d'un épisode de sécheresse². Les précipitations moyennes journalières par département ont donc été rajoutées au modèle DLNM pénalisé. Le choix d'une *cross-basis* semble de nouveau adapté, en raison de l'existence de deux types de crues : les crues éclair et les crues lentes. Si les premières font suite à des précipitations intenses sur une courte durée, les secondes peuvent survenir après une longue période de pluies. Il est donc important d'inclure les précipitations des jours précédents dans la modélisation, tout en sachant que cela augmentera sa complexité et son temps d'estimation. De fait, il s'agit du même dilemme que pour l'inclusion de variables de températures retardées. Les variations de la mortalité associées au cumul de précipitations

1. Voir la présentation "Prise en compte des mesures d'adaptation en tarification non-vie face au risque inondation et son évolution sur différents horizons de temps" : ☞

2. Article issu d'un reportage de TF1 : ☞

des deux dernières semaines ont été représentées sur le graphique H.1.



GRAPHIQUE H.1 – Variation cumulée des décès prédite par un modèle DLNM pénalisé avec inclusion de la variable de précipitations journalières

Le modèle prédit une mortalité largement supérieure, multipliée par deux à partir de 50 mm de précipitations journalières et jusqu'à dix fois plus élevée en cas de précipitations extrêmes, par rapport à une situation d'absence de précipitations. Toutefois, les relations obtenues entre précipitations et mortalité paraissent difficilement explicables, malgré le lissage utilisé. Il est notamment peu cohérent d'obtenir une mortalité décroissante au-delà d'un certain seuil de précipitations (105 mm), correspondant pourtant à des niveaux extrêmes. Plusieurs pistes d'amélioration peuvent être proposées. La première est de mieux capter les spécificités de chaque territoire (existence d'une digue, topographie) pouvant influencer sur la gravité des conséquences potentielles d'une inondation. Par exemple, si une digue a été construite, alors l'inondation devra être plus importante pour conduire à une rupture, mais en cas de rupture, la surmortalité pourrait être immédiate et élevée. Au contraire, l'absence de barrage pourrait conduire à un lien perceptible entre précipitations et mortalité à des niveaux plus faibles de précipitations, mais aussi une courbure moins brutale. Ce point pourrait faire l'objet d'une étude spécifique.

Choix des projections de températures parmi l'ensemble DRIAS-2020

L'ensemble DRIAS-2020 est constitué de 12 jeux de données, qui se différencient dans un premier temps par leur émetteur, ainsi que les variables et scénarios climatiques disponibles. Leur liste est présentée dans le tableau I.1.

Institution	Modèle climatique global (GCM)	Modèle climatique régional (RCM)	Scénario CO2				Périodes disponibles		Nombre de variables disponibles
			Historique	RCP 2.6	RCP 4.5	RCP 8.5	Historique	Projection	
CNRM	CNRM-CM5-LR	ALADIN63	X	X	X	X	1951-2005	2006-2100	9
CLMcom	MPI-ESM-LR	CCLM4-8-17	X	X	X	X	1950-2005	2006-2100	7
ICTP	HadGEM2-ES	RegCM4-6	X	X		X	1970-2005	2006-2099	7
SMHI	EC-EARTH	RCA4	X	X	X	X	1970-2005	2006-2100	7
IPSL	IPSL-CM5A-MR	WRF381P	X	X	X		1951-2005	2006-2100	7
GERICS	NorESM1-M	REMO2015	X	X		X	1950-2005	2006-2100	7
CSC	MPI-ESM-LR	REMO2009	X	X	X	X	1970-2005	2006-2100	7
CLMcom	HadGEM2-ES	CCLM4-8-17	X	X	X		1950-2005	2006-2099	7
KNMI	EC-EARTH	RACMO22E	X	X	X	X	1950-2005	2006-2100	9
SMHI	IPSL-CM5A-MR	RCA4	X	X	X		1970-2005	2006-2100	7
KNMI	CNRM-CM5-LR	RACMO22E	X	X	X	X	1950-2005	2006-2100	9
DMI	NorESM1-M	HIRHAM5 v3	X	X	X		1951-2005	2006-2100	7

TABLEAU I.1 – Liste des modèles de projections disponibles dans l'ensemble DRIAS-2020

Les modèles issus de l'ensemble DRIAS-2020 se différencient également par leur niveau de pessimisme quant à l'évolution des températures et des précipitations. Par la suite, deux indicateurs de cette évolution seront abordés, mais seul le premier sera utilisé pour le choix du modèle :

- ΔT : écart de la température moyenne (sur 30 ans) par rapport à une période de référence (1976-2005) ;
- ΔP : écart relatif des précipitations (sur 30 ans) par rapport à une période de référence (1976-2005).

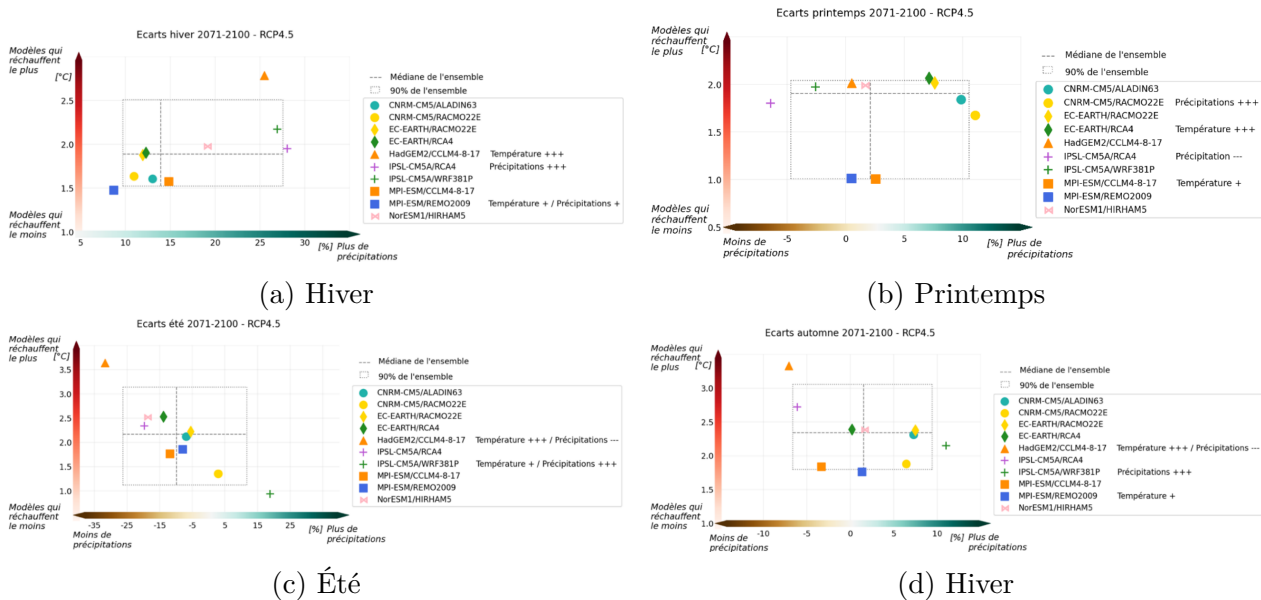
Le portail DRIAS fournit¹ un document d'aide à la sélection de ces modèles climatiques.

1. Lien vers le document d'aide à la sélection du modèle issu du portail DRIAS : [☞](#)

Ce dernier contient diverses statistiques de dispersion obtenues à partir de l'ensemble des modèles. Pour l'analyse effectuée dans ce mémoire, cela permet de quantifier à la fois l'incertitude des projections de températures, et la manière dont chaque modèle de projection se place parmi l'ensemble disponible. Les objectifs conduisant au choix du modèle sont les suivants, et sont similaires à ceux qui ont été retenus par Maucuer [43] :

- a) Retenir un modèle médian parmi l'ensemble des projections disponibles ;
- b) Couvrir l'ensemble des scénarios RCP ;
- c) Entre des modèles équivalents selon les critères précédents, préférer un modèle issu d'un institut français.²

Une représentation graphique de ces dispersions inter-modèles est disponible par saison, sur la période 2071-2100 et le scénario RCP 4.5. Celle-ci est reproduite dans le graphique I.1.



GRAPHIQUE I.1 – Écarts moyens de températures (ΔT) et de précipitations (ΔP) par saison et modèle issu de l'ensemble DRIAS-2020 sur la période 2071-2100 (scénario RCP 4.5) par rapport à la période 1976-2005 (source : DRIAS)

Dans un premier temps, l'étude de l'ensemble des modèles met en avant une incertitude importante des écarts de températures par rapport à la période 1976-2005. Celle-ci est d'autant plus importante en été, qui pourrait voir les températures moyennes progresser au-delà de $3,5^{\circ}\text{C}$ d'après le modèle issu de l'institution HadGEM2. Par exemple, en été, selon les modèles, les températures pourraient être de l'ordre de 1°C à $3,5^{\circ}\text{C}$ plus élevées sur la période 2071-2100. Étant donné la croissance accélérée de la mortalité au-delà de l'optimum

2. Selon l'intuition qu'un institut français aura concentré davantage de moyens pour assurer la cohérence des projections en France

thermique prédite dans le chapitre 3, les écarts de projections de mortalité entre les modèles pourraient être significatifs. Cette observation doit inciter, une nouvelle fois, à une grande prudence quant à la portée des résultats obtenus dans cette étude.

Ensuite, il ressort que, pour le scénario climatique intermédiaire et la période 2071-2100, les écarts de températures moyennes saisonnières prédits par certains modèles se situent à proximité de la médiane inter-modèles. Autrement dit, il y a souvent autant de modèles plus pessimistes que de modèles plus prudents quant à la progression des températures. Ceci est un signe rassurant quant à la cohérence des projections de mortalité qui pourront être obtenues à partir de ces jeux de données. Quel que soit la saison, les températures moyennes prédites par ces modèles seraient de l'ordre de 2°C plus élevées sur la période 2071-2100 par rapport à la période 1976-2005. Ces modèles sont :

- CNRM - ALADIN63 ;
- EC-EARTH - RACMO22E ;
- NorESM1 - HIRHAM5

Parmi ces modèles, et selon le critère b), le jeu NorESM1 ne peut pas être choisi. En effet, il ne propose pas de prédictions pour le scénario RCP 8.5 et s'avèrera donc insuffisant pour évaluer l'impact des scénarios climatiques sur la mortalité. Selon les deux jeux restants, le critère c) conduit à retenir le modèle issu de l'institut français CNRM.

Toutefois, selon l'objectif de l'étude prospective, un autre jeu de données aurait pu être utilisés : une approche plus pessimiste est notamment fournie par le modèle issu de l'institut HadGEM2, qui prédit (hors printemps) des progressions de températures toujours plus importantes par rapport aux autres modèles.

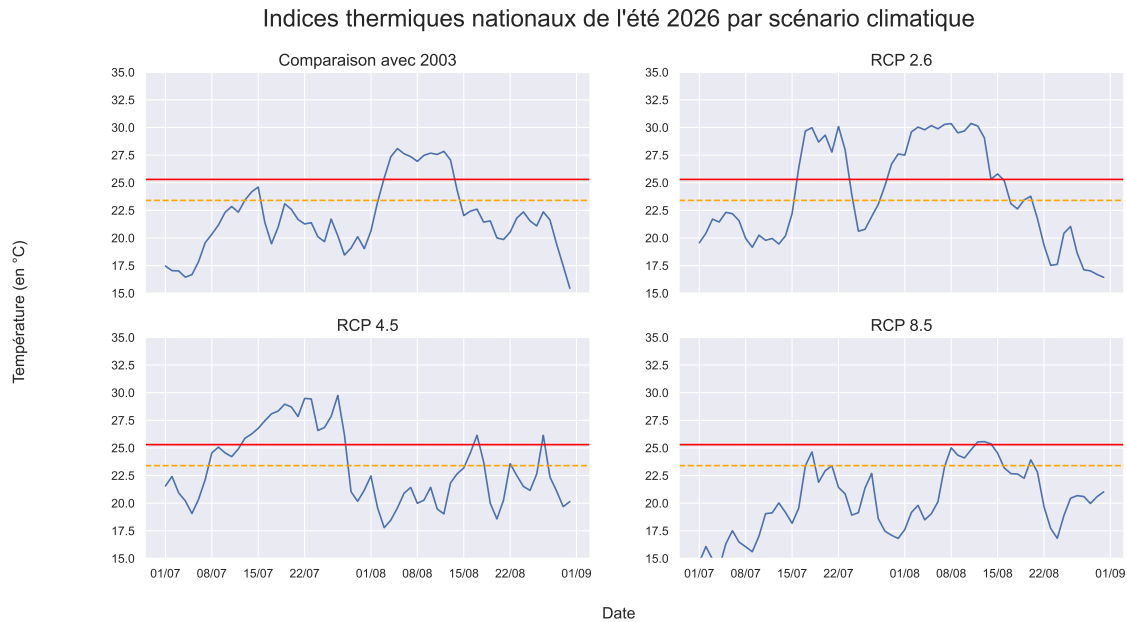
J

Étude de cas : des vagues de chaleur simulées en 2026

En plus de l'année 2065, les résultats de l'analyse descriptive relative à l'intensité et la fréquence des vagues de chaleur futures motivent une étude spécifique de l'année 2026. En effet, cette année se caractériserait, dans le scénario RCP 2.6, par (au moins) une vague de chaleur particulièrement intense. D'une part, pour ce scénario et d'après le graphique 4.5, l'intensité moyenne serait même la plus forte à un horizon de 50 ans. Au-delà, tous scénarios confondus, cette intensité moyenne resterait particulièrement remarquable sur la période 2020-2070. La longueur du phénomène serait également considérable, puisque les seuils d'une vague de chaleur seraient dépassés durant plus de trois semaines. Le graphique J.1 permet d'obtenir plus de détails sur l'évolution des températures projetées durant l'été 2026, pour chacun des scénarios climatiques.

Les lignes orange et rouge correspondent aux seuils (23,4°C et 25,3°C) fixés par Météo-France pour la définition d'une vague de chaleur. Tout d'abord, l'analyse de l'indice thermique national confirme le caractère remarquable de l'été 2026 projeté, notamment en comparaison de l'épisode (déjà exceptionnel et meurtrier) observé en 2003. Alors que ce dernier s'est matérialisé au cours des deux premières semaines d'août, la projection selon le scénario RCP 2.6 suggère l'existence de deux vagues de chaleurs, séparées par une brève accalmie de quelques jours. L'intensité des pics de chaleur, sur l'ensemble du territoire, pourrait être supérieure de 2°C aux niveaux de la canicule de 2003.

Concernant le scénario RCP 4.5, l'été 2026 pourrait également être remarquable au vu de sa longueur, mais son intensité moindre devrait engendrer une surmortalité moindre par rapport à celle prédite dans le scénario RCP 2.6. En effet, les températures journalières les plus élevées (un indice thermique de 30°C) existent sous la forme de pics, mais ne se



GRAPHIQUE J.1 – Indices thermiques nationaux lors de l'été 2003 (en haut à gauche), et prédits en 2026 par scénario RCP

maintiennent pas pendant plusieurs jours consécutifs contrairement au scénario RCP 2.6.

Les conséquences de ces épisodes en termes de surmortalité pourraient être comparables aux niveaux de 2003 concernant le scénario RCP 2.6, au vu de la longueur et l'intensité combinées des vagues de chaleur projetées. L'adoption de mesures d'adaptation, implicitement captée par le modèle estimé (grâce à l'utilisation d'un historique large), pourrait toutefois permettre de réduire la mortalité associée à cet été remarquable par rapport aux 15 000 décès en excès mentionnés par l'INSERM¹.

Les étés de l'année 2026 projetés sous les scénarios RCP 4.5 et 8.5 s'avèrent, de manière générale, moins chauds que sous le scénario climatique optimiste. Cela semble incohérent par définition des scénarios. Toutefois, des explications peuvent être avancées. Par construction des scénarios RCP, cette observation pourrait surtout s'expliquer par l'indépendance des projections de températures réalisées entre chaque scénario. En effet, l'hypothèse de continuité des scénarios conduit à supposer des trajectoires de forçage radiatif proches jusqu'à l'année 2020, conduisant à des trajectoires de températures très peu différenciées en 2026. Cette conjecture a été confirmée par le graphique 1.1 issu du GIEC, relatif aux trajectoires de températures prédites à horizon 2100 par scénario RCP.

1. Institut National de la Santé et de la Recherche Médicale. Lien vers le rapport : [↗](#)

K

Fréquence et intensité des vagues de froid futures

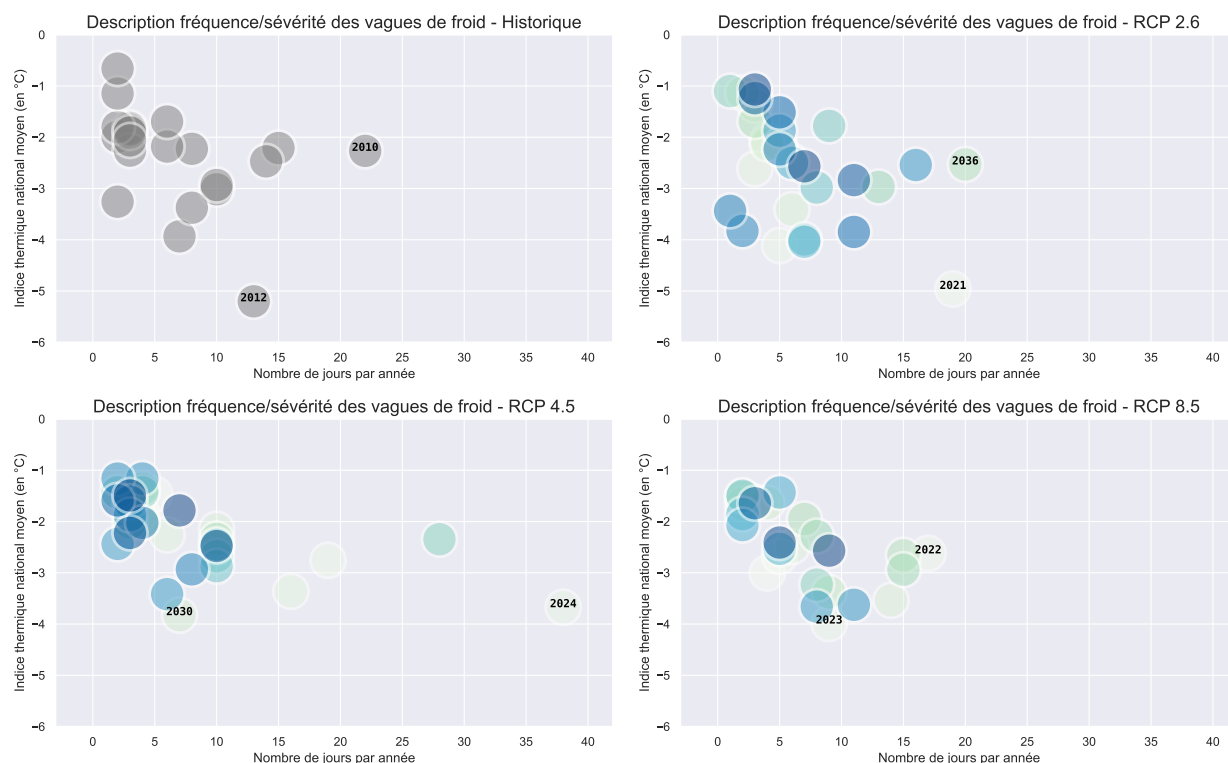
Selon les relations entre la température et la mortalité obtenues à partir du modèle DLNM, les conséquences des températures fraîches seraient visibles à plus long terme que pour les vagues de chaleur, et toucheraient davantage les personnes âgées. Afin de visualiser la fréquence et la sévérité des épisodes de froid passés et futurs, le choix s'est porté sur une définition d'une vague de froid¹. Cette dernière est définie à partir de l'indicateur thermique national décrit précédemment et utilisé pour définir les vagues de chaleur. Une vague de froid est alors définie par un épisode d'au moins 3 jours, durant lequel :

- l'indice thermique ne dépasse jamais $0,9^{\circ}\text{C}$;
- l'indice thermique est au moins une fois inférieur à -2°C durant les 3 jours.

Pour rappel, la croissance obtenue de la surmortalité par rapport à l'optimum thermique est davantage marquée en-dessous de 0°C . L'indicateur défini précédemment semble donc cohérent dans l'optique d'obtenir une première intuition relative à l'impact des épisodes de froid futures. De manière analogue aux vagues de chaleur, les nombres de jours avec vagues de froid par année ainsi que les intensités moyennes associées sont représentés dans le graphique K.1 pour chaque scénario climatique et comparés avec la période historique 1990-2019.

Alors que le réchauffement climatique aurait pu entraîner - de manière intuitive - une diminution des périodes de froid, le graphique ne permet pas de mettre en évidence cette suggestion. Cependant, sur la période de projection retenue, les épisodes de froid les plus marquants (en termes de fréquence et d'intensité) risquent d'être davantage concentrés sur les quinze prochaines années.

1. Lien vers une définition fournie par la Chaîne Météo : [↗](#)



GRAPHIQUE K.1 – Nombre de jours avec vagues de froid et intensité moyenne par année en France métropolitaine durant la période 1990-2019 et par scénario RCP (2020-2070)

De manière étonnante, le scénario de réchauffement intermédiaire (RCP 4.5) serait caractérisé par des périodes de froid plus longues, de durées comparables aux épisodes de l'hiver 2009-2010. Durant cette période, le mois de janvier a été considéré comme le plus froid des vingt années précédentes², selon Météo-France. De plus, cet hiver peut également être également considéré comme rare du fait de la survenance de tempêtes de neige tardives, jusqu'au début du mois de mars. Finalement, d'après la définition utilisée dans cette étude, les vagues de froid se sont concentrées sur 22 jours durant l'année 2010.

Le nombre de jours avec températures négatives se comporte donc de manière comparable avant et après 2020, ce qui laisse imaginer que la mortalité attribuable au froid ne diminuera pas forcément durant les prochaines décennies. Cette observation, combinée à la hausse de la fréquence et de l'intensité des vagues de chaleur prédite, laisse augurer une augmentation de la mortalité due aux phénomènes climatiques à l'horizon 2070.

2. Lien vers une synthèse des principales vagues de froid depuis 1954 (Dernières Nouvelles d'Alsace) : ↗