



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuaire

le 18 Septembre 2025

Par : Rosny Blondo GANOU VOUTSA

Titre : Modélisation de l'impact direct de la température sur la mortalité

Confidentialité : Non

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

**Membre présent du jury de l'Institut
des Actuaire :**

Romain LAILY

Ludovic BONNEAU

Signature :

Entreprise :

Linkpact

Signature :

Membres présents du jury de l'EURIA : Directeur de mémoire en entreprise :

Franck Vermet

Guillaume BIESSY

Signature :

Invité :

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 :

Signature du candidat :

Résumé

Le changement climatique constitue une préoccupation majeure ayant des conséquences sur l'ensemble des secteurs d'activité, notamment celui de l'assurance. L'augmentation des températures qui l'accompagne représente un risque sanitaire important, illustré par la canicule de 2003, laquelle a entraîné près de quinze mille décès en excès en France. Ces dernières années, la fréquence et l'intensité des épisodes de chaleur ne cessent d'augmenter et exposent à plus de décès futurs. Dans ce contexte, ce mémoire a pour objectif de quantifier l'impact direct des fluctuations de température sur la mortalité.

Cette étude exploite conjointement des variables démographiques et climatiques et met en application les modèles additifs généralisés (GAM) qui allient flexibilité et contrôle du surapprentissage.

Le modèle obtenu reproduit les principaux traits de la mortalité, notamment une mortalité élevée à la naissance, un excès de mortalité masculine et une baisse progressive au fil du temps. Les résultats montrent que l'effet de la température forme une relation en U, traduisant une mortalité accrue aux températures extrêmes et que les personnes âgées et les femmes sont plus vulnérables.

Enfin, les projections démographiques et climatiques issues de l'INSEE et du GIEC ont permis d'estimer l'évolution future de la mortalité, révélant une hausse progressive du nombre de décès, d'intensité variable selon les scénarios. Ces projections conduisent, pour une personne née en 2000, à une diminution de l'espérance de vie d'environ un mois pour les hommes et de cinq mois pour les femmes dans le scénario 8.5, comparativement au scénario 2.6.

Mots clés : Mortalité, Risque climatique, Température, Modèle GAM, Open data

Abstract

Climate change is a major concern with consequences for all sectors of activity, particularly insurance. The accompanying rise in temperatures poses a significant health risk, as illustrated by the 2003 heatwave, which caused nearly 15,000 excess deaths in France. In recent years, the frequency and intensity of heat waves have continued to increase, exposing us to more deaths in the future. In this context, the aim of this thesis is to quantify the direct impact of temperature fluctuations on mortality.

This study combines demographic and climatic variables and applies generalized additive models (GAMs), which combine flexibility and control of overfitting.

The resulting model reproduces the main features of mortality, including high infant mortality, excess male mortality, and a gradual decline over time. The results show that the effect of temperature forms a U-shaped relationship, reflecting increased mortality at extreme temperatures, and that older people and women are more vulnerable.

Finally, demographic and climate projections from INSEE and the IPCC were used to estimate future mortality trends, revealing a gradual increase in the number of deaths, varying in intensity depending on the scenario. For a person born in 2000, these projections lead to a decrease in life expectancy of about one month for men and five months for women in scenario 8.5, compared to scenario 2.6.

Keywords: Mortality, Climate risk, Temperature, GAM model, Open data

Remerciements

La réalisation de ce mémoire a été rendue possible grâce à la contribution précieuse de nombreuses personnes, à qui j'adresse mes sincères remerciements.

– Aux fondateurs et associés de LinkPact ACS, Michel Morcos El Douaihy et Guillaume Ramond, pour leur accueil chaleureux au sein du cabinet et pour avoir favorisé mon intégration dans un environnement de travail aussi serein que stimulant.

– À Guillaume Biessy, Senior Manager et Responsable du département de Recherche et Développement, pour le suivi rigoureux de mes travaux, sa disponibilité constante, ses analyses critiques constructives, ses suggestions d'amélioration à chaque étape, et surtout pour les nombreux enseignements qu'il m'a transmis.

– À Erwan Alonso, Christian Waffo et Guy Nyamsi, consultants chez LinkPact, pour les nombreux échanges, leur disponibilité, leurs conseils avisés et leur généreux partage d'expérience.

– À mes camarades en alternance, Othniel Tra Bi Néné et Bérénice Delobel, pour leur esprit de collaboration, leur professionnalisme et l'inspiration qu'ils m'ont apportée au quotidien.

– À l'ensemble du personnel de LinkPact, pour leur accueil bienveillant, leur disponibilité et leur sympathie.

– À mon tuteur académique, Franck Vermet, pour son suivi attentif, sa disponibilité et l'intérêt qu'il a porté à mon travail.

– À l'ensemble du corps professoral et administratif de l'EURIA, pour la qualité des enseignements dispensés et leur engagement constant envers la formation des actuaires.

– À ma famille et à mes amis, pour leur soutien moral indéfectible, leur patience et leurs encouragements au fil de cette aventure.

– Et enfin, à toutes celles et ceux qui, de près ou de loin, ont contribué à la réalisation de ce mémoire.

Merci !

Note de Synthèse

Modélisation de l'impact direct de la température sur la mortalité

Mise en contexte

Aujourd'hui, le changement climatique et ses conséquences sur les sociétés sont au coeur des préoccupations. Il se traduit par l'augmentation progressive des températures à la surface du globe, qui ont déjà atteint des niveaux records. Selon le 6^e rapport d'évaluation du GIEC (Global-climat 2021), ce réchauffement est largement attribuable aux activités humaines à l'origine d'émissions de gaz à effet de serre. Ainsi, la décennie 2011-2020 présente une température supérieure de +1°C par rapport à la période 1850-1900, ce qui en fait la décennie la plus chaude depuis près de 125 000 ans. En France, Météo France (2025) indique que le réchauffement est encore plus marqué, atteignant +1,5 °C sur la période 2010-2019. Ses manifestations se traduisent déjà par la multiplication et l'intensification des épisodes de chaleur extrême qui se produisent désormais chaque année. Les répercussions sont majeures causant aussi bien des dommages matériels que des problèmes de santé comme l'a illustré la canicule historique de 2003 responsable d'une surmortalité estimée à près de 15 000 décès (Fouillet et al. 2006).

Au cours des dernières années, une part croissante de la surmortalité estivale est attribuée aux canicules (Santé Publique France 2025). Cet impact de la température sur la mortalité expose les assureurs à une augmentation graduelle de la sinistralité, particulièrement en assurance vie. C'est dans ce contexte que ce mémoire propose d'apporter une contribution à la quantification des impacts de la température, en l'intégrant dans la modélisation prospective de la mortalité. Bien que la littérature à ce sujet reste limitée, certains travaux ont déjà abordé cette question. Notamment, Fall (2021) a étendu un modèle classique de mortalité prospective en y intégrant un facteur considérant la corrélation entre température et mortalité. À sa suite, Pincemin (2021) et Salaun (2024) ont proposé respectivement des modèles CSDL et DLNM qui reproduisent la forme en U ou en V de la relation température-mortalité. Dans ces dernières approches les facteurs Âge-Période-Cohorte ne sont pas explicitement intégrés.

Données et méthodologie

Pour mener à bien cette étude, une attention particulière a été portée sur la collecte et le traitement des données utiles. L'analyse a été menée à une échelle fine afin de mieux représenter les températures auxquelles les populations sont exposées. L'échelle retenue est celle des

départements et les observations sont prises à fréquence quotidienne. Les données mobilisées relèvent des évolutions quotidiennes de températures, de décès et d'exposition au décès. Elles proviennent de Météo France, de l'INSEE et de la Human Mortality Database (HMD). Un traitement approfondi a été mené pour corriger plusieurs limites des bases initiales, notamment des doublons, des erreurs et omissions dans les données de décès, ainsi que des valeurs manquantes dans les séries des températures. Après ces ajustements, les données obtenues sont cohérentes avec les séries de référence de la HMD.

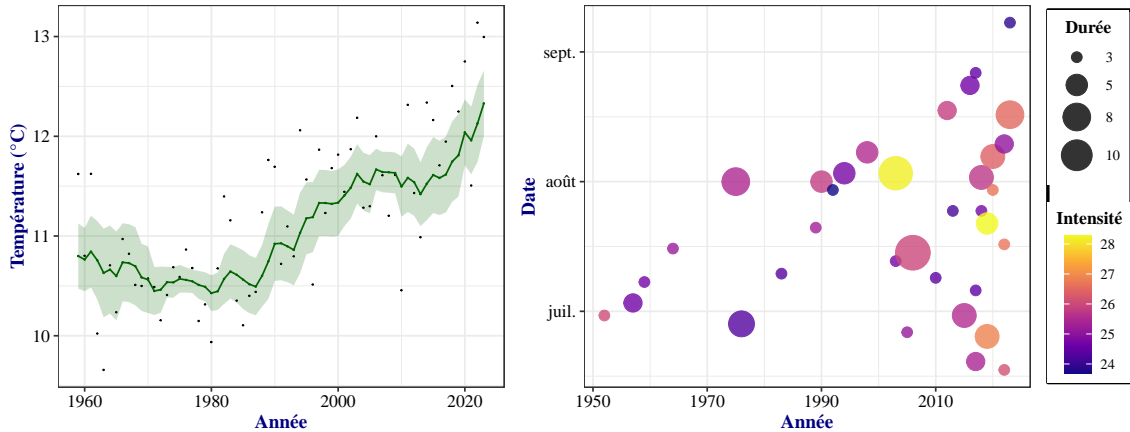


Figure 1: Evolution des températures décennales et des vagues de chaleur

L'analyse descriptive a confirmé les tendances de la littérature. Elle a montré une hausse progressive des températures moyennes, à un rythme moyen de $+0,3^{\circ}\text{C}$ par décennie sur la période étudiée et une accélération marquée depuis la fin des années 1980. Elle met également en évidence l'augmentation en fréquence et sévérité des vagues de chaleur (Figure 1).

La méthodologie de modélisation s'inspire du cadre classique de la mortalité prospective qui s'appuie sur les facteurs âge, période et cohorte. Dans ce cadre, les modèles Age-Période, Lee-Carter et Age-Période-Cohorte ont été calibrés et comparés à des modèles GAM appliqués à la modélisation de la mortalité et reproduisant les effets des modèles de Lee-Carter et intégrant la température. En outre, l'approche backtesting est utilisée en calibrant les modèles sur la période historique 1990-2012 et en évaluant leurs capacités prédictives sur la période 2013-2022.

Modélisation et résultats

La spécification du modèle GAM appliqué est décrite par l'Equation 1 :

$$\log(\mu_{xy}^s) = f_1^s(x) + f_2^s(y) + k_e^s(y) + f_3^s(x, y) + \epsilon_{xy}^s \quad (1)$$

où :

- x, y, s, μ représente respectivement l'âge, la période calendaire, le sexe et la force de mortalité
- f_1^s est l'effet structurel lisse de l'âge pour le sexe s
- f_2^s est l'effet tendanciel lisse de la période pour le sexe s
- $k_\epsilon^s(y)$ est l'effet spécifique de l'année y
- f_3^s est le terme lisse d'interaction entre l'âge et l'année, captant les écarts non expliqués par les effets d'âge et de tendance.
- ϵ_{xy}^s est le terme d'erreur aléatoire

Les fonctions lisses estimées permettent de déduire le rôle des variables sur la mortalité, ces effets sont décrits par la Figure 2.

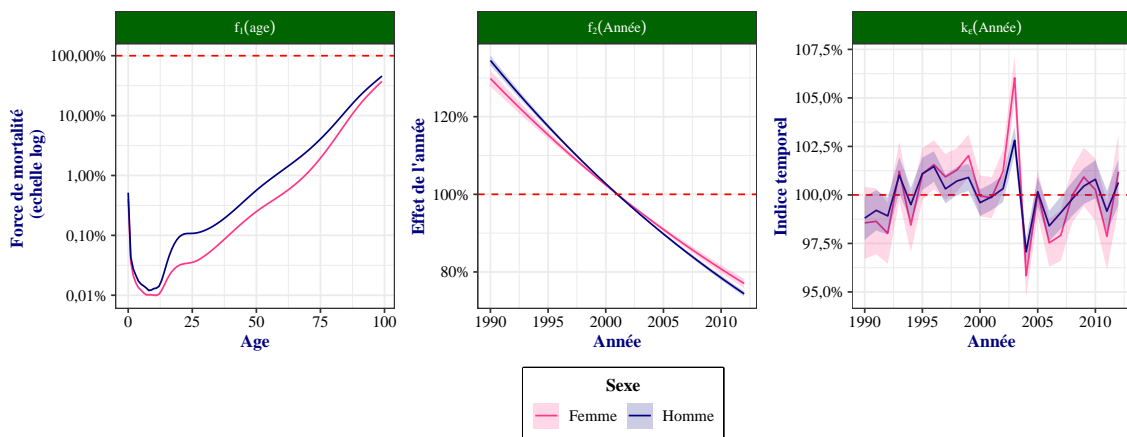


Figure 2: Effet des facteurs âge et période

Effet de l'âge

L'effet de l'âge montre plus de traits liés à la mortalité à l'instar d'une mortalité plus importante à la naissance, qui baisse progressivement jusqu'à un minimum atteint vers 10 ans. Son évolution est ensuite croissante, présentant un pic autour de 20 ans appelé "bosse des accidents" et une évolution log-linéaire avec l'âge par la suite. Le sexe montre une mortalité systématiquement plus forte pour les hommes quel que soit l'âge considéré.

Effet de la période calendaire

L'effet de période est décomposé en deux composantes :

- Une **composante tendancielle régulière**, qui décrit une décroissance progressive de la mortalité au cours du temps, grâce aux avancées médicales et technologiques permettant un accroissement de l'espérance de vie. En moyenne et à âge égal, les taux de mortalité à la période 2010 sont plus faibles de près de 20% par rapport à l'an 2000 et de 40% par rapport aux années 1990.
- Une **composante purement aléatoire**, qui décrit les effets des crises et des chocs conjoncturels qui impactent l'intensité de la mortalité au cours d'une période précise.

Les améliorations de mortalité selon les âges

La tendance d'amélioration de la mortalité décrite ci-dessus est une évolution moyenne. De façon plus précise, ces évolutions présentent un rythme différent selon les âges. Cet effet est capté par le terme d'interaction $f_3^s(x, y)$ qui révèle que les améliorations de mortalité sont plus rapides pour la population des jeunes de moins de 30 ans environ, et qu'elles sont localement plus fortes autour de 75 ans.

La prise en compte de ces effets par le modèle GAM a conduit à de bonnes performances prédictives des décès annuels. Ses performances, tant en ajustement aux données qu'en projection sont meilleures que celles des modèles Age-Période, Lee-Carter et Age-Période-Cohorte sur la période d'étude considérée.

Appliqué à une échelle désagrégée en évaluant la mortalité selon les différentes températures associées aux décès, ce modèle a montré un biais caractérisé par une surestimation de la mortalité lorsque la température est comprise entre 9°C et 25°C environ, et par une sous-estimation de la mortalité sur la plage restante. Ce biais a justifié la pertinence d'intégrer des facteurs de température pour enrichir le modèle.

Intégration des facteurs de température dans la modélisation

La flexibilité de l'approche GAM facilite l'intégration de la température au modèle, ce qui n'est pas possible avec les modèles de la famille GAPC qui ne prennent en compte que les facteurs âge, année et cohorte. Le modèle proposé intégrant les effets de température a été spécifié comme décrit par l'Equation 2 :

$$\log(\mu_{xyt}^s) = f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y) + f_4^s(t) + f_5^s(x, t) + \epsilon_{xyt}^s \quad (2)$$

où :

- t représente la température
- f_4^s est l'effet commun de la température pour le sexe s
- $f_5^s(x, t)$ représente l'effet d'interaction entre l'âge et la température.
- La signification des autres paramètres restant identique à celle du modèle de référence.

Description des effets de la température

L'impact de la température décrit par ce modèle rejoint les résultats obtenus par Santé publique France (2004), Pincemin (2021) et Salaun (2024) en présentant un effet global en forme de U. La mortalité est minimale à une température optimale, appelée température de mortalité minimale (TMM), qui se situe autour de 20 °C. De part et d'autre de cette valeur la mortalité s'accroît progressivement à un rythme rapide pour la chaleur et modéré pour le froid (Figure 3). Ce résultat confirme également une plus forte sensibilité des femmes, pour qui l'impact des températures extrêmes est plus important.

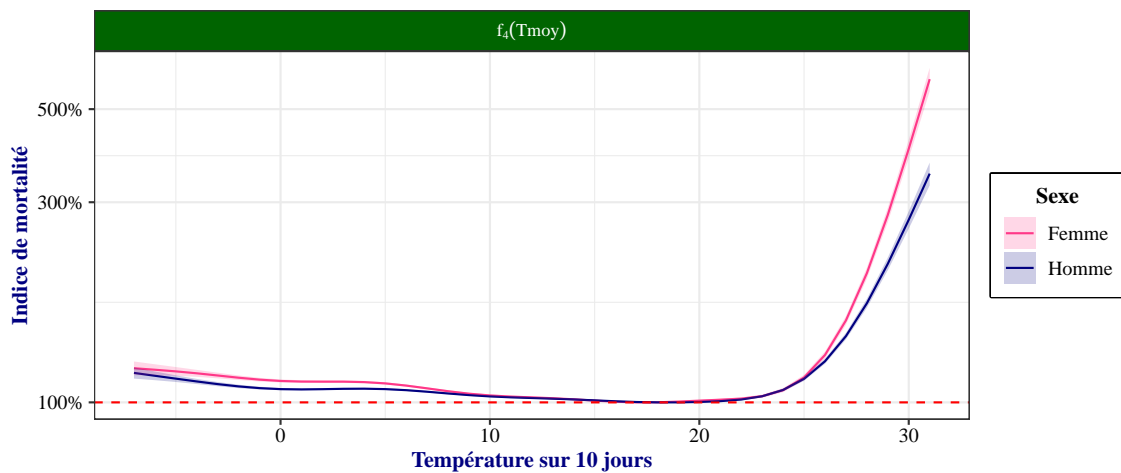


Figure 3: Effet commun de la température

Au-delà de cet effet commun, l'impact de la température n'est pas homogène pour toute la population mais varie grandement selon l'âge. Cette interaction est captée par le terme $f_5^s(x, t)$ dont l'association avec le terme d'effet commun conduit à l'effet spécifique de la température variant selon les âges ($f_4^s(t) + f_5^s(x, t)$). Les résultats rejoignent ceux de Institut de veille sanitaire (2010), ils montrent un impact plus important des températures froides sur les personnes âgées. En revanche, les populations de moins de 50 ans résistent mieux aux effets du froid. Dans le cas des températures élevées, tous les âges présentent une sensibilité accrue à la chaleur (Figure 4).

Performances quotidiennes et effet de saisonnalité

En intégrant ces effets, le modèle obtenu surpasse davantage le modèle de référence qui ne prend pas la température en compte. Alors que les modèles usuels conduisent à une mortalité constante au cours de l'année, la prise en compte de la température permet d'obtenir

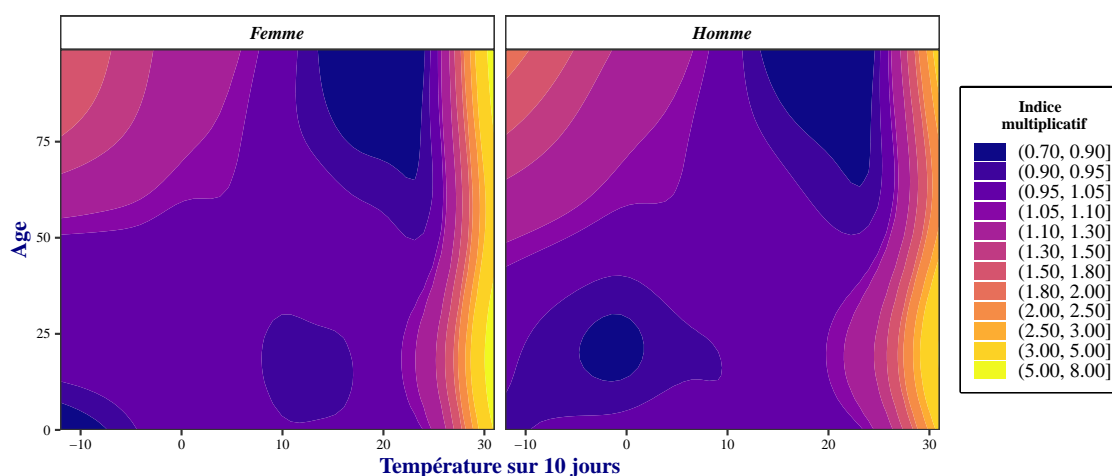


Figure 4: Effet de la température selon les âges

une précision plus fine de la mortalité intra-annuelle, notamment aux échelles mensuelles et quotidiennes. Il capte ainsi le caractère saisonnier de la mortalité reflétant une mortalité plus forte en hiver et moins forte en été.

Projections de la mortalité et limites

À partir des données de projection de température et de population du GIEC et de l'INSEE, les projections de la mortalité ont été faites grâce au modèle. Les trois scénarios de températures s'appuient sur les scénarios RCP 2.6, 4.5 et 8.5 d'émissions de gaz à effet de serre du GIEC. Ces scénarios diffèrent leur caractère pessimiste des émissions et conduisent à un réchauffement climatique plus ou moins marqué d'ici la fin du siècle.

L'impact sur la mortalité est évalué aux échelles départementales et nationales. Cet impact est mesuré sur l'espérance de vie résiduelle de la cohorte née en 2000 qui dès 7 ans vit sous les conditions de température décrites par chaque scénario RCP. Les résultats varient d'un département à l'autre et selon les sexes. Ils montrent en général une évolution régulière des taux de mortalité au fil du temps mais marquée par des pics de mortalité ponctuels liés aux vagues de chaleur. Ces pics sont d'autant plus prononcés que le scénario est pessimiste. Dans l'ensemble, les résultats montrent un impact fort sur l'espérance de vie prospective, particulièrement dans le scénario 8.5, et moins marqué dans les scénarios 2.6 et 4.5. Géographiquement, la zone ouest du territoire subit un impact moins important, tandis que le sud-est est plus affecté, en particulier dans le scénario 8.5 (Figure 5).

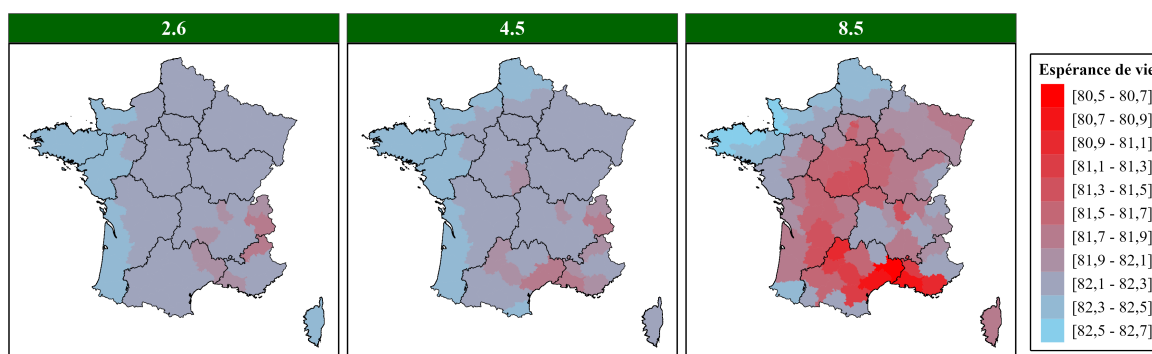


Figure 5: Espérance de vie des femmes par département et par scénario

À l'échelle nationale, les projections des décès annuels sont similaires au cas départemental. Elles montrent une tendance haussière due au vieillissement de la population qui est également marquée par des pics de surmortalité liés aux épisodes de vagues de chaleur des scénarios. Les canicules les plus extrêmes, notamment celles de 2065 dans les scénarios 4.5 et 8.5 révèlent une très forte vulnérabilité des femmes, dont les pics de surmortalité sont particulièrement aigus.

Suite à l'évaluation de cet impact, il convient de souligner qu'il est sujet à une forte incertitude, notamment parce que l'horizon de projection est grand. En effet, d'ici à la fin du siècle, les conditions actuelles pourraient considérablement évoluer, modifiant l'ampleur des effets observés aujourd'hui. L'impact des températures est donc susceptible d'être atténué notamment par l'acclimation des populations ou par le développement technologique futur.

Par ailleurs, ce modèle présente encore des perspectives d'amélioration. Parmi les limites soulevées figurent la non prise en compte de facteurs catalyseurs de l'effet de la température tels que le vent et l'humidité, qui peuvent amplifier ou atténuer les effets de la température. De plus, le modèle ne tient pas compte des spécificités locales propres à chaque département, ce qui pourrait conduire à des surestimations ou sous-estimations de la mortalité dans certains départements. Enfin, le modèle réduit les effets de la température à une variable unique, reflétant aussi bien les effets du froid comme ceux de la chaleur. La prise en compte de plusieurs variables représentant les températures de plusieurs jours successifs pourrait mieux capter la variabilité thermique.

Toutefois, ce modèle apporte une plus-value utile à l'évaluation de l'impact des températures sur la mortalité, ainsi que pour l'actuariat, notamment dans le cadre de stress-tests climatiques et pour l'ajustement des prévisions de mortalité en matière de tarification et de provisionnement.

Mots clés : Mortalité, Risque climatique, Température, Modèle GAM, Open data

Summary Note

Modeling the direct impact of temperature on mortality

Context

Today, climate change and its consequences for society are a major concern. It is reflected in the gradual increase in global surface temperatures, which have already reached record levels. According to the IPCC's Sixth Assessment Report (**IPCC2021?**), this warming is largely attributable to human activities that cause greenhouse gas emissions. The decade 2011-2020 was $+1^{\circ}\text{C}$ warmer than the period 1850-1900, making it the warmest decade in nearly 125,000 years. In France, Météo France (2025) indicates that warming is even more pronounced, reaching $+1.5^{\circ}\text{C}$ over the period 2010-2019. Its effects are already evident in the increase in the number and intensity of extreme heat events, which now occur every year. The repercussions are significant, causing both material damage and health problems, as illustrated by the historic heatwave of 2003, which was responsible for an estimated 15,000 excess deaths (Fouillet et al. 2006).

In recent years, an increasing proportion of excess summer mortality has been attributed to heat waves (Santé Publique France 2025). This impact of temperature on mortality exposes insurers to a gradual increase in claims, particularly in life insurance. It is in this context that this thesis proposes to contribute to the quantification of the impacts of temperature by integrating it into prospective mortality modeling. Although the literature on this subject remains limited, some studies have already addressed this issue. In particular, Fall (2021) extended a classic prospective mortality model by incorporating a factor that considers the correlation between temperature and mortality. Following on from this, Pincemin (2021) and Salaun (2024) have proposed CSDL and DLNM models, respectively, which reproduce the U- or V-shaped relationship between temperature and mortality. In these latter approaches, the Age-Period-Cohort factors are not explicitly integrated.

Data and methodology

To carry out this study, particular attention was paid to the collection and processing of useful data. The analysis was conducted on a fine scale in order to better represent the temperatures to which populations are exposed. The scale used is that of departments, and observations are taken on a daily basis. The data used relate to daily changes in temperature, deaths,

and exposure to death. It comes from Météo France, INSEE, and the Human Mortality Database (HMD). In-depth processing was carried out to correct several limitations in the initial databases, including duplicates, errors, and omissions in the death data, as well as missing values in the temperature series. After these adjustments, the data obtained is consistent with the HMD reference series.

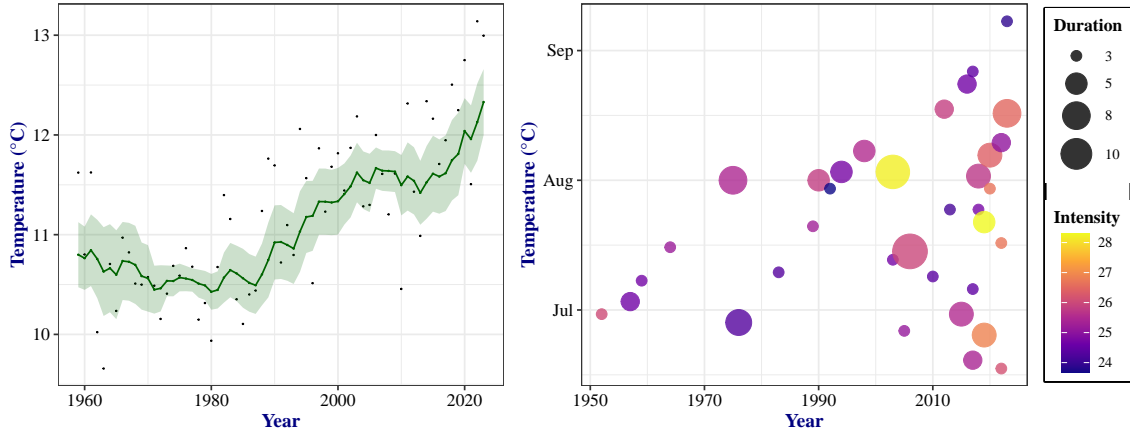


Figure 1: Changes in decade temperatures and heat waves

Descriptive analysis confirmed the trends reported in the literature. It showed a gradual increase in average temperatures, at an average rate of $+0.3^{\circ}\text{C}$ per decade over the period studied, with a marked acceleration since the late 1980s. It also highlights the increase in the frequency and severity of heat waves (Figure 1).

The modeling methodology is based on the classic prospective mortality framework, which relies on age, period, and cohort factors. Within this framework, the Age-Period, Lee-Carter, and Age-Period-Cohort models were calibrated and compared with GAM models applied to mortality modeling, reproducing the effects of the Lee-Carter model and incorporating temperature. In addition, the backtesting approach is used by calibrating the models over the historical period 1990-2012 and evaluating their predictive capabilities over the period 2013-2022.

Modeling and results

The specification of the GAM model applied is described by Equation 1:

$$\log(\mu_{xy}^s) = f_1^s(x) + f_2^s(y) + k_{\epsilon}^s(y) + f_3^s(x, y) + \epsilon_{xy}^s \quad (1)$$

where:

- x, y, s, μ represent age, calendar period, sex and mortality rate, respectively
- f_1^s is the smooth structural effect of age for sex s
- f_2^s is the smooth trend effect of the period for sex s
- $k_\epsilon^s(y)$ is the specific effect of year y
- f_3^s is the smooth interaction term between age and year, capturing differences not explained by age and trend effects.
- ϵ_{xy}^s is the random error term

The estimated smooth functions allow us to deduce the role of the variables taken into account on mortality; these effects are described by Figure 2.

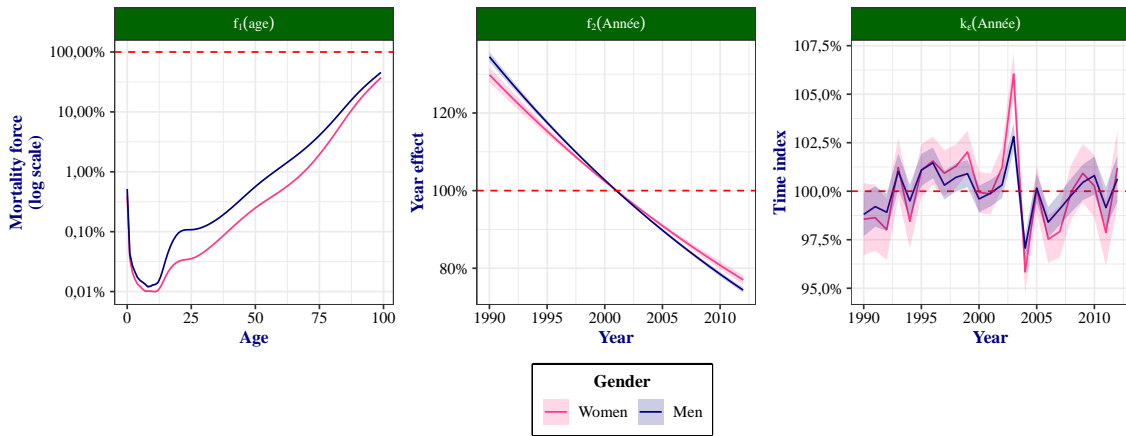


Figure 2: Effect of age and period factors

Effect of age

The effect of age shows more mortality-related traits, such as higher mortality at birth, which gradually declines to a minimum around the age of 10. It then increases, peaking around the age of 20, known as the “accident hump,” and subsequently follows a log-linear trend with age. Gender shows systematically higher mortality for men regardless of age.

Effect of the calendar period

The period effect is broken down into two components:

- A **regular trend component**, which describes a gradual decline in mortality over time, thanks to medical and technological advances that have increased life expectancy. On average and at the same age, mortality rates in 2010 were nearly 20% lower than in 2000 and 40% lower than in the 1990s.

- A **purely random component**, which describes the effects of crises and economic shocks that impact mortality rates during a specific period.

Improvements in mortality by age

The trend of improving mortality described above is an average trend. More specifically, these changes occur at different rates depending on age. This effect is captured by the interaction term $f_3^s(x, y)$, which reveals that mortality improvements are faster for the population under the age of 30 and are locally stronger around the age of 75.

Taking these effects into account in the GAM model led to good predictive performance for annual deaths. Its performance, both in terms of data adjustment and projection, is better than that of the Age-Period, Lee-Carter, and Age-Period-Cohort models over the study period considered.

Applied on a disaggregated scale by assessing mortality according to the different temperatures associated with deaths, this model showed a bias characterized by an overestimation of mortality when the temperature is between approximately 9°C and 25°C, and by an underestimation of mortality in the remaining range. This bias justified the relevance of integrating temperature factors to enrich the model.

Integration of temperature factors into the model

The flexibility of the GAM approach facilitates the integration of temperature into the model, which is not possible with GAPC models, which only take into account age, year, and cohort factors. The proposed model integrating temperature effects was specified as described by Equation 2:

$$\log(\mu_{xyt}^s) = f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y) + f_4^s(t) + f_5^s(x, t) + \epsilon_{xyt}^s \quad (2)$$

where:

- t represents temperature
- f_4^s is the common effect of temperature for sex s
- $f_5^s(x, t)$ represents the interaction effect between age and temperature.
- The meaning of the other parameters remains the same as in the reference model.

The estimation of the parameters and smooth functions of this new model did not affect the shape and effects of the basic factors, suggesting that the temperature effects represent a complementary dimension to the effects initially present.

Description of temperature effects

The impact of temperature described by this model is consistent with the results obtained by Beaudreau (2004), Pincemin (2021), and Florian (2024), showing an overall U-shaped effect. Mortality is minimal at an optimal temperature, known as the minimum mortality temperature (MMT), which is around 20°C. On either side of this value, mortality increases gradually at a rapid rate for heat and at a moderate rate for cold (Figure 3). This result also confirms that women are more sensitive, for whom the impact of extreme temperatures is greater.

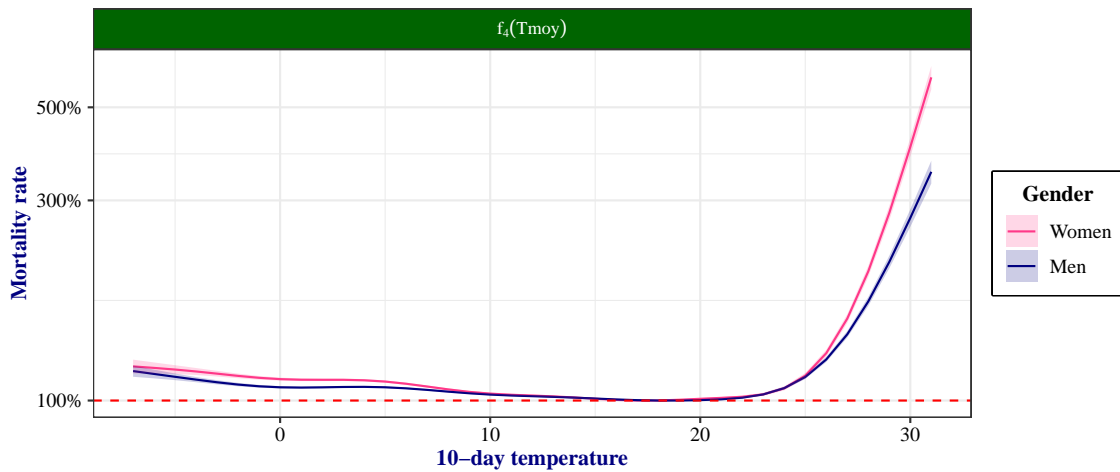


Figure 3: Common effect of temperature

Beyond this common effect, the impact of temperature is not uniform across the entire population but varies greatly depending on age. This interaction is captured by the term $f_5^s(x, t)$, whose association with the common effect term leads to the specific effect of temperature varying according to age ($f_4^s(t) + f_5^s(x, t)$). The results are consistent with those of Institut de veille sanitaire (2010), showing a greater impact of cold temperatures on older people. On the other hand, people under the age of 50 are more resistant to the effects of cold. In the case of high temperatures, all ages show increased sensitivity to heat (?@fig-effect-age-tmoy).

Daily performance and seasonal effects

By incorporating these effects, the resulting model further outperforms the reference model, which does not take temperature into account. While standard models predict constant mortality throughout the year, taking temperature into account allows for greater precision in intra-annual mortality, particularly on a monthly and daily basis. It thus captures the seasonal nature of mortality, reflecting higher mortality in winter and lower mortality in summer.

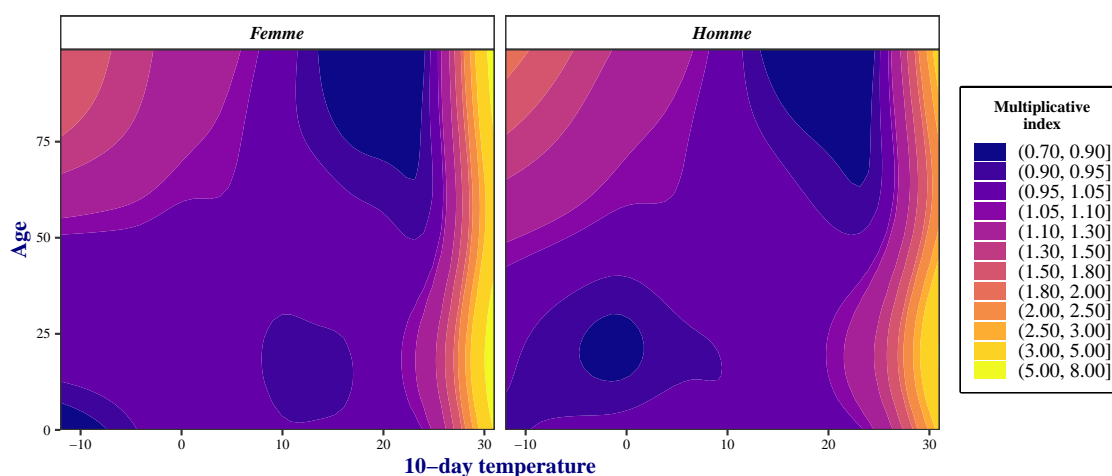


Figure 4: Effect of temperature according to age

Mortality projections and limitations

Mortality projections were made using the model, based on temperature and population projection data from the IPCC and INSEE. The three temperature scenarios are based on the IPCC's RCP 2.6, 4.5, and 8.5 greenhouse gas emission scenarios. These scenarios differ in their pessimistic nature regarding emissions and lead to more or less pronounced global warming by the end of the century.

The impact on mortality is assessed at the departmental and national levels. This impact is measured on the residual life expectancy of the cohort born in 2000, which from the age of 7 will live under the temperature conditions described by each RCP scenario. The results vary from one department to another and according to gender. They generally show a steady change in mortality rates over time, but marked by occasional peaks in mortality linked to heat waves. These peaks are more pronounced the more pessimistic the scenario. Overall, the results show a strong impact on prospective life expectancy, particularly in scenario 8.5, and less marked in scenarios 2.6 and 4.5. Geographically, the western part of the country is less affected, while the southeast is more affected, particularly in scenario 8.5 (Figure 5).

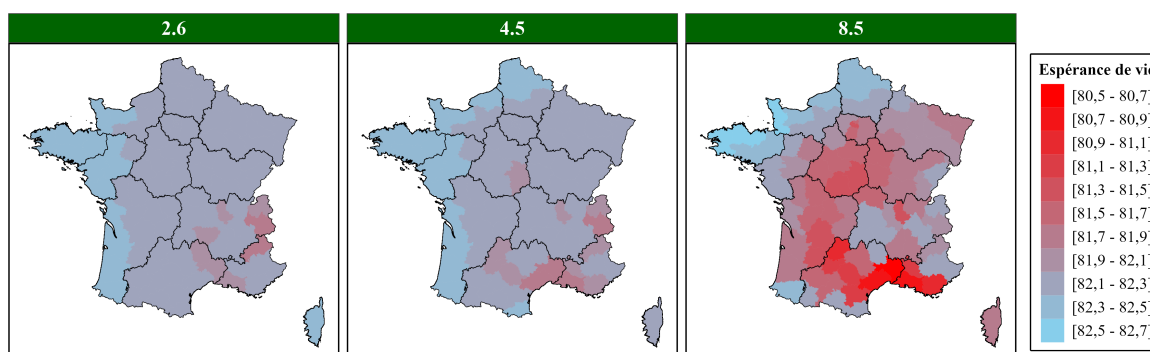


Figure 5: Life expectancy for women by department and scenario

At the national level, projections for annual deaths are similar to those for individual departments. They show an upward trend due to the aging population, which is also marked by peaks in excess mortality linked to heat waves in the scenarios. The most extreme heat waves, particularly those of 2065 in scenarios 4.5 and 8.5, reveal a very high vulnerability among women, whose peaks in excess mortality are particularly acute.

Following the assessment of this impact, it should be noted that it is subject to considerable uncertainty, particularly because the projection horizon is long. Indeed, between now and the end of the century, current conditions could change considerably, altering the magnitude of the effects observed today. The impact of temperatures is therefore likely to be mitigated, in particular by the acclimatization of populations or by future technological developments.

Furthermore, this model still has room for improvement. Among the limitations raised are the failure to take into account factors that catalyze the effect of temperature, such as wind and humidity, which can amplify or mitigate the effects of temperature. Furthermore, the model does not take into account the specific local characteristics of each department, which could lead to overestimations or underestimations of mortality in certain departments. Finally, the model reduces the effects of temperature to a single variable, reflecting both the effects of cold and heat. Taking into account several variables representing temperatures over several consecutive days could better capture thermal variability.

However, this model provides useful added value for assessing the impact of temperatures on mortality, as well as for actuarial purposes, particularly in the context of climate stress tests and for adjusting mortality forecasts for pricing and provisioning.

Keywords : Mortality, Climate risk, Temperature, GAM model, Open data

Table des matières

Introduction	1
1. Généralités sur la mortalité et la température	4
1.1. La mortalité	4
1.1.1. Un risque central en actuariat	4
1.1.2. Les tables de mortalité	5
1.2. Le changement climatique et les vagues de chaleur	6
1.2.1. Le changement climatique	6
1.2.2. Les vagues de chaleur et les canicules en France	7
1.3. Les projections de température	8
1.3.1. Modélisation du climat futur	8
1.3.2. Les scénarios climatiques en France	10
1.3.3. Les tendances futures des températures	11
1.4. Les effets de la température sur la mortalité	11
Synthèse du chapitre	16
2. Présentation des données et analyse descriptive	17
2.1. Présentation des données utilisées	17
2.1.1. Les données de température	18
2.1.2. Les données de décès	21
2.1.3. Les données d'exposition au décès	27
2.1.4. Constitution du jeu de données global	28
2.2. Les évolutions thermiques et démographiques	30
2.2.1. Les évolutions des températures	30
2.2.2. Les évolutions des décès en France	34
2.2.3. Les évolutions de population en France	39
2.2.4. Analyse globale de la mortalité	40
Synthèse du chapitre	43
3. Notions théoriques et approche de modélisation	44
3.1. Cadre méthodologique de l'étude	44
3.1.1. Les concepts de base en analyse de survie	44
3.1.2. Les notations actuarielles	46
3.2. Les modèles prospectifs de mortalité	46
3.2.1. Le modèle Age-Période (AP)	47
3.2.2. Le modèle de Lee-Carter	50

3.2.3.	Le modèle Age-Période-Cohorte (APC)	51
3.2.4.	Les modèles Age-Période-Cohorte généralisés (GAPC)	52
3.3.	Présentation des modèles additifs généralisés	53
3.3.1.	Les modèles additifs	53
3.3.2.	Les modèles additifs généralisés	58
3.4.	Qualité d'ajustement et comparaison des modèles	58
3.4.1.	Les résidus de déviance	58
3.4.2.	La déviance	59
3.4.3.	Le backtesting	59
	Synthèse du chapitre	60
4.	Résultat de la modélisation hors effets de température	61
4.1.	Cadre de modélisation	61
4.2.	Modélisation classique de la mortalité	62
4.2.1.	Calibration du modèle Age-Période	62
4.2.2.	Calibration du modèle de Lee-Carter	65
4.2.3.	Calibration du modèle Age-Période-Cohorte	69
4.2.4.	Modélisation par une approche GAM	72
4.2.5.	Comparaison des modèles	78
4.3.	Mise en évidence des effets de température sur la mortalité	82
	Synthèse du chapitre	85
5.	Résultat de la modélisation intégrant la température	86
5.1.	Intégration d'un effet additif de la température	86
5.2.	Intégration de l'interaction Age-Température	89
5.3.	Analyse du nombre de jours optimal	101
5.4.	Limites et Pistes d'amélioration du modèle	108
5.4.1.	Intégration d'un facteur de saisonnalité	109
5.4.2.	Effet des zones climatiques	110
5.4.3.	Prise en compte de la spécificité des départements	111
5.4.4.	Autres limites	112
	Synthèse du chapitre	112
6.	Impact des températures futures	114
6.1.	Données de projection de température et analyse descriptive	114
6.1.1.	Source de données de projection de température	114
6.1.2.	Description des températures futures	115
6.2.	Impact des projections dans les départements	117
6.2.1.	Impact sur l'intensité de mortalité	117
6.2.2.	Impact sur l'espérance de vie prospective	120
6.3.	Impact des températures futures à l'échelle nationale	123
6.3.1.	Source de données de projection de population	123
6.3.2.	Impact des températures futures dans la France hexagonale	124
	Synthèse du chapitre	126

Conclusion	127
Références	129
Annexes	132
A. Résultats complémentaires	132
A.1. Les surmortalités annuelles réelles et prédites par les modèles	132
A.2. Les vagues de chaleur futures selon les scénarios RCP	134
A.3. Espérances de vie par département selon les scénarios RCP	134

Liste des Figures

1.1. Les 10 étés les plus chauds depuis 1900 (Météo France, 2025b)	7
1.2. Les scénarios d'émission de GES (Météo France, 2020)	9
1.3. Projection des températures annuelles (Météo France, 2020)	11
1.4. Surmortalité attribuable aux canicules (Santé Publique France, 2025) . . .	13
2.1. Couverture des stations utiles et des stations exclues	19
2.2. Disponibilité et valeurs manquantes de 40 stations	20
2.3. Distribution des distances à la station la plus proche	21
2.4. Décès quotidiens et doublons en 1997 et 1999	24
2.5. Données INSEE vs Données HMD	25
2.6. Décès INSEE / Décès HMD avant et après traitement	26
2.7. Illustration de l'agrégation des départements	29
2.8. Taux de mortalité par âge et par période	29
2.9. Température en moyenne annuelle (points) et en moyenne décennale (ligne)	31
2.10. Température départementale en moyenne décennale	31
2.11. Température moyenne par période en France	32
2.12. Zone climatique française et profil de température (Source : Salaun (2024))	33
2.13. Les vagues de chaleur en France depuis 1950	33
2.14. La saisonnalité des températures	35
2.15. La saisonnalité des décès	36
2.16. Age moyen et nombre annuel de décès	37
2.17. Profil annuel des décès par groupe d'âge	38
2.18. Evolution de l'exposition	39
2.19. Structure par âge de la population	40
2.20. Nombre de décès et exposition par âge	41
2.21. Exposition par température	42
2.22. Force de mortalité par âge	43
3.1. Effets des fonctions de bases polynomiales (Source: Wood, 2017)	55
3.2. Effet du paramètre de lissage (Source : Wafo Kankeu (2023))	57
4.1. Effet des variables du modèle Age-Période	63
4.2. Résidus de déviance du modèle AP	64
4.3. Force de mortalité du modèle Age-Période	65
4.4. Effet des variables du modèle de Lee-Carter	66
4.5. Résidus de déviance du modèle de Lee-Carter	67

4.6. Force de mortalité du modèle de Lee-Carter	68
4.7. Effet des variables du modèle APC	69
4.8. Résidus de déviance du modèle APC	70
4.9. Force de mortalité du modèle Age-Période-Cohorte	71
4.10. Les effets du modèle de mortalité GAM	73
4.11. Résidus de déviance du modèle GAM	74
4.12. Force de mortalité du modèle GAM	75
4.13. Les effets de la deuxième version du modèle GAM	76
4.14. Effet périodique par âge	77
4.15. Amélioration de mortalité par âge des périodes futures	77
4.16. Les résidus du second modèle GAM	78
4.17. Force de mortalité du modèle GAM complet	79
4.18. Déviance annuelle des modèles en période de test	82
4.19. Structure des résidus du modèle selon la température	83
4.20. Mise en évidence des effets de température	84
5.1. Les effets du modèle avec la température	87
5.2. Résidus par année du modèle avec température additive	88
5.3. Résidus de deviance du modèle avec température additive	89
5.4. Les effets du modèle avec la température	90
5.5. Effet de la température	91
5.6. Contraste de surmortalité estivale et hivernale	92
5.7. Exposition par température et année	93
5.8. Résidus de deviance du modèle d'interaction age-température	94
5.9. Force de mortalité du modèle de température (Homme)	95
5.10. Force de mortalité du modèle de température (Femme)	95
5.11. Comparaison de la mortalité mensuelle (Femme)	96
5.12. Comparaison de la mortalité mensuelle (Homme)	97
5.13. Impact des températures sur la courbe de mortalité	100
5.14. Illustration espérance de vie périodique par température	101
5.15. Effet de la température en fonction du nombre de jours décalés	103
5.16. Performances du modèles selon le nombre de jours décalés	104
5.17. Comparaison des mortalités quotidiennes sur l'ensemble de la période	105
5.18. Comparaison des mortalités quotidiennes en 2003	107
5.19. Comparaison des mortalités quotidiennes en 2012	108
5.20. Ecart d'ajustement du modèle suivant les saisons	109
5.21. Ecart d'ajustement du modèle suivant les zones climatiques	110
5.22. La surmortalité dans les départements	111
6.1. Evolution future des températures en moyenne annuelle	116
6.2. Evolution des températures sur 20 ans des départements	117
6.3. Les surmortalités liées aux températures selon les scénarios (Paris)	118
6.4. Les surmortalités liées aux températures selon les scénarios (Corse)	119
6.5. Les surmortalités liées aux températures selon les scénarios (Finistère)	119

6.6.	Courbe de vie de la génération 2000 à Paris sur la période 2007-2100 . . .	121
6.7.	Espérance de vie des femmes par département et par scénario	122
6.8.	Espérance de vie des hommes par département et par scénario	122
6.9.	Projection du nombre annuel de décès future	125
A.1.	Comparaison des mortalités quotidiennes en 2006	132
A.2.	Comparaison des mortalités quotidiennes en 2018	133
A.3.	Comparaison des mortalités quotidiennes en 2020	134
A.4.	Les vagues de chaleur futures selon les scénarios	137
A.5.	Nombre et durée annuel des vagues de chaleur selon les scenarios	137

Liste des Tables

2.1. Aperçu des données de température	18
2.2. Aperçu du fichier des personnes décédées	22
2.3. Traitement des décès en France Métropolitaine	26
4.1. Ajustement des modèles aux données d'entraînement (Femme)	80
4.2. Ajustement des modèles aux données d'entraînement (Homme)	80
4.3. Performances en projection (Femme)	81
4.4. Performances en projection (Homme)	81
5.1. Comparaison des performances désagrégées en train (Femme)	97
5.2. Comparaison des performances désagrégées en train (Homme)	98
5.3. Comparaison des performances annuelles en train (Femme)	98
5.4. Comparaison des performances annuelles en train (Homme)	98
5.5. Comparaison des performances désagrégées en test (Femme)	98
5.6. Comparaison des performances désagrégées en test (Homme)	99
5.7. Comparaison des performances annuelles en test (Femme)	99
5.8. Comparaison des performances annuelles en test (Homme)	99
5.9. Décalage optimal selon les critères et les périodes	104
6.1. Espérance de vie résiduelle de la génération 2000 à Paris	121
6.2. Espérance de vie globale de la génération 2000	122
6.3. Nombre de décès de 2026 à 2070 selon les scénarios RCP	124
A.1. Espérance de vie par département selon les scénarios RCP	134

Introduction

A l'heure actuelle, le changement climatique ainsi que ses conséquences sanitaires et économiques constituent un risque majeur susceptible de bouleverser les modes de vie et la rentabilité des entreprises, y compris celles des assureurs. D'après le sixième rapport d'évaluation du GIEC¹ dont les conclusions sont reprises par Global-climat (2021), le réchauffement climatique est principalement causé par les activités humaines qui ont déjà entraîné de fortes émissions de gaz à effet de serre, notamment au cours de la période industrielle. Le réchauffement qui en découle se chiffre au niveau mondial à une augmentation de la température moyenne d'environ +1°C entre les périodes 1850-1900 et 2011-2020 ce qui fait de cette dernière la décennie la plus chaude depuis près de 125 000 ans. Au niveau de la France, Météo France (2025c) précise que le réchauffement est plus élevé, avec une augmentation de +1,5°C au cours de la décennie 2010-2019.

Ce réchauffement progressif s'accompagne de la multiplication et de l'intensification des catastrophes naturelles, telles que les sécheresses, les feux de forêt, les inondations, les tempêtes, et surtout les canicules et les vagues de chaleur extrêmes. Ces événements présentent un bilan lourd en termes de pertes matérielles et de pertes en vies humaines. À ce titre, France Assureurs (2021) révèle que la sinistralité pour cause d'aléas naturels s'est intensifiée depuis la fin des années 1980 avec des dommages se chiffrant à 3,8 milliards d'euros par an en moyenne sur la période 2016-2019 contre 2 milliards d'euros sur la période 1989-2009. En termes de conséquences sur la vie humaine, l'exemple emblématique est la canicule record de 2003 au cours de laquelle la température s'est maintenue à des niveaux exceptionnels sur une durée de deux semaines (Météo France, 2025a), ce qui a entraîné une surmortalité estimée à 15 000 décès (Fouillet *et al.*, 2006). Suite à cette canicule, le plan national d'alerte canicule a été mis en place afin d'atténuer les effets des épisodes futurs, mais sur la période récente, les canicules se produisent à une plus forte fréquence et présentent des intensités non négligeables. Les épisodes de 2020 et de 2022 ont respectivement entraîné une surmortalité de 1924 et 2816 décès (Santé Publique France, 2025).

Cette surmortalité constitue un risque majeur pour les assureurs qui peuvent être amenés à faire face à une sinistralité massive sur une courte période, l'intensification de ces événements extrêmes les expose à une hausse graduelle de la sinistralité à laquelle ils doivent faire face. En effet, la prévision de la mortalité occupe une place centrale en assurance et notamment en assurance vie où, pour garantir leur solvabilité et fournir une couverture adaptée à leurs assurés, les assureurs ont besoin d'évaluer la mortalité

¹Groupe Intergouvernementale sur l'Evolution du Climat

future avec la meilleure précision possible. À ce titre, la tarification des produits d'assurance s'appuie généralement sur des tables de mortalité réglementaires ou sur des tables d'expérience qui sont conçues à partir d'agrégats de décès observés au cours d'une période déterminée. Cette démarche d'évaluation ne tient pas compte des risques climatiques et de la surmortalité susceptible de se produire, ce qui constitue un risque de sous-estimation de la mortalité future.

De nombreux autres facteurs peuvent influencer la mortalité, mais après une étude sur plusieurs pays européens, Jens, Katrien et Torsten (2025) ont démontré que la température est la variable environnementale ayant le lien le plus fort avec la mortalité à court terme, surpassant des facteurs comme l'humidité et le vent.

Dans ce cadre, comment les assureurs peuvent-ils prédire la mortalité future en tenant compte des effets de la température ? Cette problématique soulève un enjeu et un intérêt important tant pour l'élaboration de politique de prévention sanitaire que pour le secteur de l'assurance. Si la littérature sur ce sujet est encore limitée, il a toutefois été abordé par Fall (2021) qui a étendu un des modèles usuels de mortalité prospective en y intégrant un facteur supplémentaire tenant compte de la corrélation entre la température et la mortalité à différents âges. Cependant, cette démarche n'intègre pas directement la température comme un facteur du modèle ce qui ne garantit pas l'efficacité des corrélations considérées. À sa suite, Pincemin (2021) et Salaun (2024) ont respectivement proposé des modèles CSDL et DLNM contraints à reproduire la forme en U ou en V de la relation entre température et mortalité. Cette approche présente l'avantage de prendre en compte les températures observées sur une plage de plusieurs jours, mais elle ne capture pas les principaux facteurs de mortalité que sont l'âge, le sexe et les évolutions périodiques.

Pour répondre à cette problématique, l'analyse menée dans ce mémoire est conduite à l'échelle géographique départementale, en mobilisant des données quotidiennes de température, de décès et de population issues de Météo France, de l'INSEE et de la Human Mortality Database. Ces données sont traitées et combinées afin d'alimenter la démarche de modélisation retenue, qui repose sur la construction de modèles additifs généralisés non contraints.

Contrairement aux modèles stochastiques classiques utilisés en modélisation prospective de la mortalité, qui ne prennent généralement en compte que l'âge, la période calendaire et la cohorte, les modèles GAM offrent une meilleure flexibilité. Ils permettent d'intégrer de nombreuses variables explicatives et, grâce à l'utilisation de splines, de modéliser des interactions complexes et de capter des effets non linéaires. Cette approche apparaît donc particulièrement adaptée pour analyser les effets de la température, tout en considérant des effets différenciés selon l'âge ou le sexe. Par ailleurs, les estimations lissées qu'elle produit permettent de réduire les biais d'échantillonnage et de limiter le surapprentissage, ce qui conduit à des résultats plus robustes et fiables. Malgré ces avantages, les modèles additifs généralisés sont encore méconnus et peu utilisés en actuariat. Toutefois, ils ont déjà été appliqués dans certains travaux dont, ceux de :

Wafo Kankeu (2023) qui, dans une démarche similaire, a intégré le montant des rentes dans la modélisation de la mortalité d'un portefeuille de rentiers ; Alonso (2024) dans le cadre de la construction de tables d'expérience en arrêt de travail, et Pouna Siewe (2010) pour une nouvelle approche de tarification en assurance automobile.

Ce mémoire est structuré en six principaux chapitres :

Le premier chapitre présente les notions importantes liées au sujet, dont l'importance de la modélisation de la mortalité en assurance, les évolutions climatiques actuelles et futures à travers les projections du DRIAS, ainsi que les effets physiologiques des températures sur la santé.

Le deuxième chapitre détaille les traitements et l'analyse des données de température par station de Météo France, des données individuelles des personnes décédées en France depuis 1970 provenant de l'INSEE et des données de population départementale de l'INSEE couplées aux données d'exposition issues de la base HMD.

Le troisième chapitre met l'accent sur les notions théoriques associées à l'analyse de la mortalité, notamment la présentation des modèles prospectifs usuels puis des modèles additifs généralisés (GAM) et de leur démarche d'ajustement aux données.

Le quatrième chapitre présente les résultats de l'approche classique de modélisation de la mortalité appliquée aux données, incluant les modèles Âge-Période, Âge-Période-Cohorte et de Lee-Carter ainsi que deux modèles GAM. Le premier modèle GAM est une approximation linéaire et lisse du modèle de Lee-Carter et le second est un modèle plus complet permettant de capturer davantage d'effets. À partir de ce modèle final, le chapitre se conclut par une analyse des résidus qui met en évidence la présence d'un effet de la température non pris en compte par les approches classiques, ouvrant ainsi une piste d'amélioration du modèle.

Le cinquième chapitre aborde la prise en compte de la température en complétant le modèle GAM final et présente les principaux impacts issus de la température. Deux modèles sont calibrés : le premier considère un effet global de la température affectant tous les individus de la même façon, et le second généralise cette approche en intégrant des effets de température différenciés selon le profil d'âge des personnes exposées.

Le sixième chapitre analyse l'impact des évolutions futures des températures sur la mortalité à partir des simulations du DRIAS et des projections de population de l'INSEE, en particulier l'impact sur l'espérance de vie prospective dans les départements et le nombre de décès attendu au niveau national.

Chapter 1.

Généralités sur la mortalité et la température

Ce chapitre dresse un état des lieux de quelques notions essentielles liées à la modélisation de la mortalité et au changement climatique. Il aborde dans un premier temps le rôle et les enjeux de la modélisation actuarielle de la mortalité. Les évolutions récentes des températures et leurs projections selon les simulations du DRIAS sont ensuite examinées. Enfin, il se conclut par une analyse des principaux effets physiologiques de la température sur la santé.

1.1. La mortalité

1.1.1. Un risque central en actuariat

Le phénomène de mortalité occupe une place centrale en actuariat en raison de son influence sur de nombreux aspects de la discipline. Elle est intrinsèquement liée à la notion de durée de vie humaine, qui constitue une variable essentielle dans de nombreux produits assurantiels. En assurance, et notamment en assurance vie, les garanties reposent sur le statut vital des assurés : l'assurance en cas de décès prévoit le versement d'un capital au moment du décès, tandis que l'assurance en cas de vie verse des prestations aux assurés ou à leurs bénéficiaires tant qu'ils sont vivants.

Dans ce contexte, il est essentiel pour l'assureur de pouvoir estimer la durée de vie de ses assurés. Cette estimation intervient à plusieurs postes, notamment l'évaluation des engagements financiers futurs, la tarification adéquate des contrats et même l'évaluation des risques extrêmes.

Pour répondre à ces enjeux, l'actuaire s'appuie sur des probabilités de décès ou de survie, généralement issues des tables de mortalité. Ces dernières jouent un rôle fondamental dans l'évaluation du risque en influençant directement les montants de garanties, le montant des provisions et le calcul des primes. Il est donc crucial de disposer de tables de mortalité fiables et suffisamment précises pour permettre des estimations prudentes du risque de mortalité.

1.1.2. Les tables de mortalité

Une table de mortalité ou table de survie est un outil statistique qui, pour une population donnée, associe à chaque âge la probabilité de décéder avant d'atteindre l'âge suivant. Ces tables sont généralement construites à partir d'un échantillon de la population concernée et sur un historique plus ou moins long en fonction de l'objectif de la table. Il en existe plusieurs types :

- **Les tables du moment ou tables périodiques** : Il s'agit de tables reflétant le niveau de mortalité d'une population à un moment donné. Elles sont généralement construites à partir de données observées sur une période relativement courte (3 à 5 ans) afin de tenir compte des fluctuations annuelles de la mortalité. Ce type de table ne peut pas être utilisé à un horizon long car elle estime des quotients de mortalité indépendants de la période calendaire. Dans le cas de produits nécessitant d'évaluer la mortalité sur le long terme les tables prospectives sont plus adaptées.
- **Les tables prospectives ou générationnelles** : Ce sont des tables qui tiennent compte de l'aspect évolutif de la mortalité en fournissant des projections de mortalité par âge et pour des périodes futures. Elles sont notamment privilégiées pour la tarification des produits de rente viagère.

D'un point de vue assureur, on distingue les tables réglementaires et les tables d'expérience :

- **Les tables réglementaires** :

Les tables réglementaires sont celles préconisées par le régulateur pour guider les actions de tarification et de provisionnement. Elles constituent le choix par défaut pour les assureurs, conformément aux dispositions de l'article A132-18 du code des assurances. Parmi elles, les plus couramment utilisées sont :

- Les tables TH 00-02 et TF 00-02** : ce sont des tables genrées, correspondant respectivement aux hommes et aux femmes. Elles ont été construites à partir des données de l'INSEE relatives à la période 2000-2002. Elles sont utilisées dans le cadre des produits d'assurance en cas de décès. Pour les produits d'assurance en cas de vie, à l'exception des produits de rentes, un décalage d'âge doit être appliqué avant leur utilisation.
- Les tables TGH 05 et TGF 05** : Il s'agit de tables générationnelles. Elles sont également genrées et ont été construites à partir des données des bénéficiaires de contrats de rentes sur la période de 1993 à 2005. Leur cadre d'utilisation est celui des contrats de rentes, à l'exception des contrats collectifs pour lesquels la loi Industrie verte du 24 octobre 2024 impose l'utilisation de la table TGHF05.

- La **table TGHF05** : C'est une table non genrée visant à éviter les discriminations de tarification entre les hommes et les femmes. Sa nécessité vient du fait qu'à âge et cotisation égale, les tables genrées conduisent à des rentes plus faibles pour les femmes en raison de leur espérance de vie plus élevée. La table TGHF05 a été construite grâce à une association des tables TGH 05 et TGF 05 en adoptant une répartition de 40% de femmes et 60% d'hommes et elle doit être utilisée pour les contrats d'assurance collective.

Les tables réglementaires sont construites au niveau global et correspondent au profil de risque de l'assureur moyen. À ce titre, elles peuvent se montrer inadaptées pour des assureurs dont le profil de risque est plus spécifique. Dans cette situation, une table d'expérience devient un meilleur choix pour l'assureur.

- **Les tables d'expérience :**

Il s'agit de tables de mortalité élaborées spécifiquement pour un assureur à partir des caractéristiques propres à son portefeuille. Elles offrent une vision plus précise du risque auquel il est exposé. Pour être conforme et utilisable, la table construite doit faire l'objet d'une validation initiale par un actuaire certificateur et indépendant du constructeur de la table. Une fois validée, elle est valable pendant 5 ans au maximum et doit faire l'objet d'un suivi annuel afin de maintenir sa conformité.

1.2. Le changement climatique et les vagues de chaleur

1.2.1. Le changement climatique

Le changement climatique désigne une modification durable des paramètres du climat à l'échelle planétaire ou régionale (OXFAM France, 2022). Ce phénomène est fortement amplifié par les activités humaines, en particulier par les émissions massives de gaz à effet de serre, dont l'accélération a été marquée dès la révolution industrielle du XIX^e siècle. Le changement climatique se manifeste principalement par une hausse progressive des températures moyennes mondiales, mais aussi par une augmentation de la fréquence et de l'intensité des phénomènes météorologiques extrêmes : élévation du niveau des océans, inondations, sécheresses, tempêtes violentes, canicules, et vagues de chaleur.

Les observations montrent qu'à l'échelle mondiale, la température moyenne a augmenté d'environ 1 °C depuis 1850. En France, cette hausse atteint 1,5 °C (Météo France, 2025c). De plus, cette tendance s'est nettement accélérée depuis les années 1980, si bien que la période 2010-2024 compte les années les plus chaudes jamais enregistrées (Figure 1.1). Cette élévation des températures est particulièrement perceptible durant les périodes estivales, devenues plus chaudes et plus propices à des épisodes de canicule et de vagues de chaleur. Selon Météo-France, les dix étés les plus chauds depuis 1950 ont

tous eu lieu après les années 2000, l'été 2003 étant le plus extrême avec une température moyenne estivale de 23,7 °C.

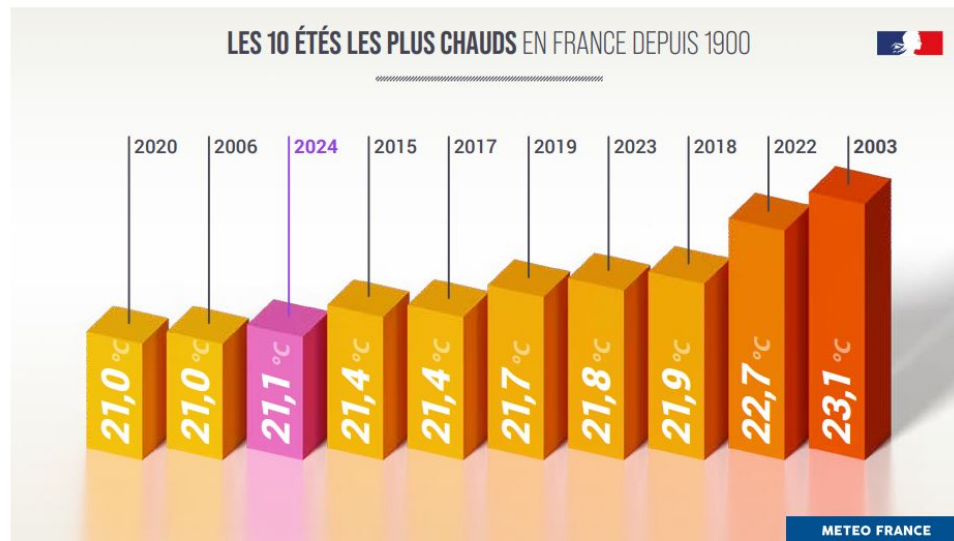


Figure 1.1. : Les 10 étés les plus chauds depuis 1900 (Météo France, 2025b)

1.2.2. Les vagues de chaleur et les canicules en France

Les vagues de chaleur désignent des périodes durant lesquelles les températures sont exceptionnellement élevées, représentant un risque important pour la santé des populations. Il n'existe pas de définition universelle du phénomène : celle-ci varie selon les contextes climatiques et géographiques des régions du monde.

En France, Météo-France distingue deux notions proches : la vague de chaleur et la canicule. La vague de chaleur correspond à un épisode à l'échelle nationale de températures anormalement élevées pendant plusieurs jours. Pour la caractériser, Météo France s'appuie sur l'indicateur thermique national, qui est la moyenne des températures minimales et maximales mesurées dans 30 stations météorologiques réparties uniformément sur le territoire (Météo France, 2025a). Cet indicateur est comparé aux normales saisonnières, calculées sur des périodes de référence de 30 ans qui sont révisées tous les dix ans.

Une vague de chaleur est déclarée lorsque les critères qui suivent sont réunis :

- l'indicateur thermique national atteint 25,3 °C pendant au moins un jour;
- il reste au-dessus de 23,4 °C durant au moins trois jours consécutifs.

Elle est considérée comme terminée lorsque l'indicateur descend en dessous de 23,4 °C pendant deux jours consécutifs, ou en dessous de 22,4 °C sur une seule journée.

La canicule, quant à elle, désigne une période d'au moins trois jours consécutifs avec des températures élevées, de jour comme de nuit, dépassant des seuils spécifiques déterminés localement. Contrairement à la vague de chaleur, elle est appréciée à l'échelle départementale. Météo-France précise qu'une canicule devient préoccupante pour la santé publique lorsque la chaleur intense persiste plusieurs jours sans répit nocturne (Météo France, 2025a). Les populations les plus vulnérables sont alors les personnes âgées, les enfants en bas âge, ou celles souffrant de pathologies chroniques, notamment à cause du risque accru de déshydratation ou de coups de chaleur.

Avec le changement climatique, ces phénomènes extrêmes sont devenus plus fréquents et plus intenses. Depuis 1947, Météo-France a recensé 50 vagues de chaleur, dont 33 depuis les années 2000. Ce constat souligne l'accélération récente du phénomène : les épisodes sont non seulement plus nombreux, mais également plus sévères et longs, comme en attestent les records de température atteints des dernières décennies.

1.3. Les projections de température

1.3.1. Modélisation du climat futur

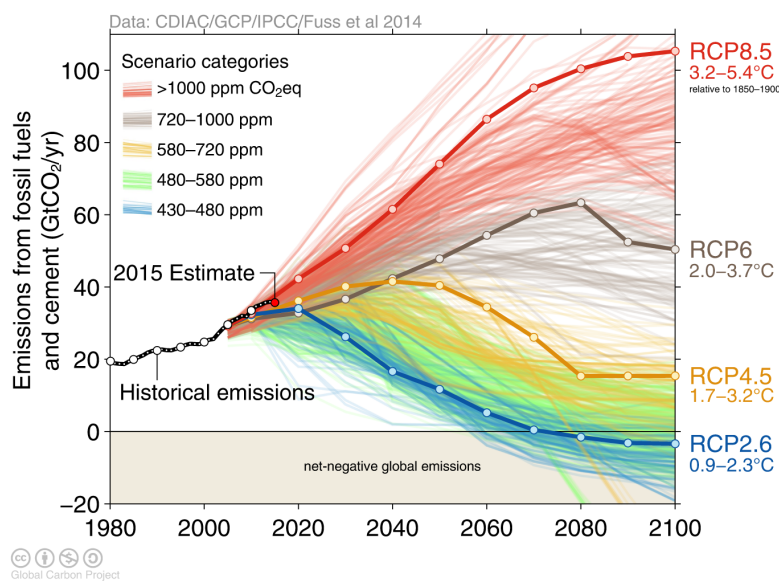
La prévision du climat futur repose sur l'utilisation de modèles physiques complexes. Comme le rappelle Météo France (2020), " *l'étude de l'évolution du climat nécessite le recours à des modèles physiques qui prennent en compte l'ensemble des équations décrivant le comportement du système climatique (composé de l'atmosphère, l'océan, les glaces marines, la végétation, les rivières, ...)* ". Concrètement, ces modèles numériques répliquent les interactions entre les différentes composantes du système climatique et les grandeurs physiques, notamment celles influençant la température. Parmi ces facteurs, les émissions de gaz à effet de serre ont le plus grand impact (Météo France, 2020). L'abondance d'émissions favorise un réchauffement climatique rapide tandis que leur maîtrise permet de limiter les évolutions climatiques futures.

De ce fait, les modèles climatiques s'appuient sur des hypothèses sur les évolutions futures des émissions de gaz à effet de serre. Ces évolutions sont tributaires de nombreux facteurs tels que la croissance démographique, le développement socioéconomique et technologique, ainsi que les choix politiques futurs, dont la prédiction est impossible. Pour pallier cette incertitude, les climatologues ont défini une gamme de scénarios plausibles d'émissions. Parmi eux, le **GIEC** a sélectionné quatre scénarios de référence dénommés **RCP** (Representative Concentration Pathways) :

- **Le scénario RCP 2.6** : Scénario optimiste caractérisé par une réduction des émissions de gaz à effet de serre, conduisant à un réchauffement global limité à

2 °C par rapport à l'ère préindustrielle. Il est le seul respectant les objectifs de l'accord de Paris.

- **Le scénario RCP 4.5** : Scénario intermédiaire où les émissions poursuivent leur trajectoire actuelle pendant quelques décennies avant de se stabiliser, puis de décroître à long terme.
- **Le scénario RCP 6.0** : Scénario similaire au 4.5, mais plus contraignant en raison d'émissions plus élevées sur une période plus longue et une décroissance ultérieure plus lente.
- **Le scénario RCP 8.5** : Scénario pessimiste reflétant une augmentation progressive des émissions, pouvant entraîner un réchauffement approchant 5 °C d'ici la fin du siècle.



A partir des scénarios d'émissions de gaz à effet de serre, plusieurs types de modèles interviennent au cours des projections climatiques selon l'échelle spatiale considérée :

- **Les modèles climatiques globaux (GCM)** : Il s'agit de modèles à large échelle, utilisés pour simuler le climat à l'échelle planétaire. Leur résolution est de l'ordre de 150 à 200 km, ce qui est insuffisant pour reproduire le climat à fine échelle et pour rendre compte de phénomènes climatiques locaux.
- **Les modèles climatiques régionaux (RCM)** : Dans le souci d'obtenir des projections à l'échelle locale, les RCM sont plus efficaces. Bien que similaires aux GCM, ils couvrent une sous-partie du globe et offrent une résolution plus fine, de l'ordre de 10 à 20 km. Ces modèles s'alimentent et tiennent compte des sorties des modèles globaux qui fournissent les conditions aux limites des mailles spatiales.

- **La correction statistique** : Généralement, les estimations des modèles régionaux présentent des biais par rapport aux observations réelles. Pour tenir compte de ces écarts des corrections statistiques sont effectuées visant à calibrer les résultats sur une période de référence donnée.

Dans l'ensemble, la production d'une projection de température nécessite le choix d'un scénario d'émission de GES, d'un couple GCM/RCM et d'une méthode de correction de biais. Compte tenu de la variabilité des résultats selon le choix des modèles, il est conseillé de considérer plusieurs simulations afin de couvrir plusieurs éventualités plausibles du climat futur.

1.3.2. Les scénarios climatiques en France

Au niveau régional, les simulations de projections climatiques sont réalisées par un ensemble de centres de modélisation climatique dans le cadre du projet **Cordex** dont le but est de produire des ensembles de simulations régionales pour tous les continents. Pour l'Europe, le programme Euro-Cordex a réalisé plus d'une centaine de simulations à une échelle incluant la France hexagonale. Ces simulations reposent sur les scénarios RCP 2.6, 4.5 et 8.5.

En France, le portail DRIAS¹ met à disposition des chercheurs et du public un ensemble de projections climatiques issues de la base Euro-cordex. Afin de faciliter leur utilisation, le portail DRIAS ne propose qu'un sous ensemble des simulations existantes, soigneusement sélectionné pour être représentatif de la gamme complète des scénarios possibles. La constitution du jeu de données a reposé sur huit critères :

- La sélection de 12 couples GCM/RCM
- La disponibilité des simulations pour au moins deux scénarios d'émissions
- La prise en compte des GCM considérés comme les plus réalistes pour l'Europe
- La diversification des modèles régionaux (RCM)
- L'exclusion des couples GCM/RCM présentant des biais connus
- La priorisation des RCM développés par les centres français de modélisation
- L'optimisation de la dispersion des changements climatiques simulés sur la France
- La sélection préférentielle des couples ayant une cohérence physique entre les modèles GCM et RCM

Le jeu de données DRIAS ainsi constitué comprend un total de 30 simulations climatiques futures : 12 pour le RCP 8.5, 10 pour le RCP 4.5 et 8 pour le RCP 2.6.

¹Donner accès aux scénarios climatiques Régionalisés français pour l'Impact et l'Adaptation de nos Sociétés et environnements

1.3.3. Les tendances futures des températures

Les prévisions d'évolutions des températures, établies sur la base des simulations climatiques futures du DRIAS, indiquent un réchauffement attendu en France métropolitaine, et ce, quel que soit le scénario RCP considéré. Celles-ci sont observées relativement à la température moyenne de la période de référence (1976-2005). Jusqu'en 2040, les évolutions sont similaires pour les 3 scénarios, une évolution différenciée apparaît ensuite conduisant, à la fin du siècle, à un réchauffement de l'ordre de +1 °C pour le RCP 2.6, +2,2 °C pour le RCP 4.5 et de +4,5 °C pour le scénario RCP 8.6.

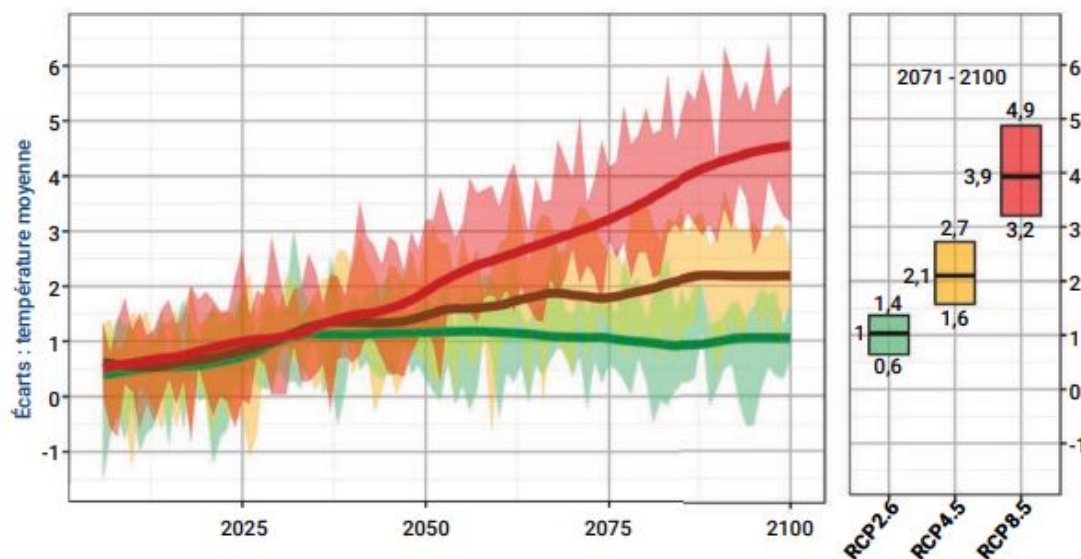


Figure 1.3. : Projection des températures annuelles (Météo France, 2020)

Par ailleurs, le réchauffement attendu présente des variations saisonnières et spatiales marquées. Il est particulièrement prononcé en été et modéré au printemps. D'un point de vue géographique, au nord-ouest, avec les régions Bretagne et Hauts-de-France, se produit un réchauffement moindre dont l'intensité augmente graduellement à mesure que l'on va vers le sud-est où le plus haut réchauffement est attendu. Plus de détails en lien avec ces évolutions sont disponibles dans le rapport de Météo France (2020)².

1.4. Les effets de la température sur la mortalité

L'impact de la température sur la mortalité a déjà été soulevé et justifié par plusieurs études. Pour comprendre les mécanismes physiologiques par lesquels la température

²Lien vers le rapport de Météo France sur les projections climatiques

affecte la santé humaine au point de conduire au décès, il est utile de se tourner vers la littérature biologique et médicale sur le sujet.

Les effets physiologiques de la chaleur

À cette fin, Santé Publique France (2025) dresse une revue des principaux effets sanitaires liés à l'exposition à la chaleur, qu'ils soient immédiats ou différés.

Parmi les effets immédiats, la déshydratation en est une conséquence directe : l'augmentation de la transpiration entraîne une perte importante d'eau, ce qui peut devenir problématique si l'hydratation n'est pas suffisante. La chaleur peut également provoquer des coups de chaleur, lorsque l'organisme ne parvient plus à réguler sa température corporelle autour de 37 °C. Ces épisodes se manifestent par une peau anormalement chaude, des étourdissements, des nausées, voire une perte de conscience.

Les effets différés peuvent aggraver des pathologies préexistantes telles que le diabète, les maladies cardiovasculaires, rénales ou respiratoires. Dans les cas les plus graves, ces aggravations peuvent conduire au décès, généralement dans les 3 à 10 jours suivant l'exposition à un épisode de chaleur.

Si l'ensemble de la population est susceptible d'être affecté, les premiers impacts touchent particulièrement les personnes vulnérables et celles exposées de manière intense à la chaleur.

Les personnes vulnérables incluent :

- Les personnes âgées;
- Les femmes enceintes;
- Les jeunes enfants;
- Les personnes souffrant de maladies chroniques ou en situation de handicap.

Les personnes fortement exposées comprennent celles confrontées à des conditions de vie ou de travail défavorables, telles que :

- Les sans-abri;
- Les personnes en situation de précarité;
- Les habitants de zones urbaines denses soumis aux îlots de chaleur urbains;
- Les détenus;
- Les travailleurs exerçant en extérieur (chantiers, agriculture) ou dans des environnements clos et mal ventilés.

Un consensus scientifique établit un lien entre les périodes de chaleur extrême telles que les canicules et vagues de chaleur, et une hausse significative de la mortalité. La canicule de 2003 par son intensité exceptionnelle, en est l'exemple marquant. Toutefois, même en dehors de ces extrêmes, des températures élevées maintenues sur une période prolongée

présentent également des risques. L'analyse de la mortalité estivale sur la période 2014-2023 par Santé Publique France (2025) a estimé la surmortalité attribuable à la chaleur à 37 000 décès. Sur cette période, bien que les jours de canicule et de forte chaleur n'aient représenté que 6% des jours, ils ont concentré 28% des décès. Enfin, les deux tiers des personnes décédées étaient âgées de 75 ans et plus.

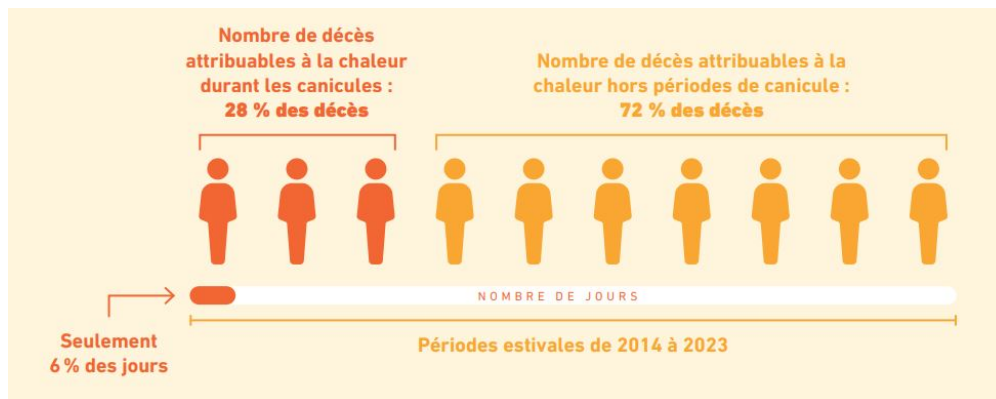


Figure 1.4. : Surmortalité attribuable aux canicules (Santé Publique France, 2025)

Les effets physiologiques du froid

Institut de veille sanitaire (2010) décrit plusieurs effets physiopathologiques du froid sur la santé. Il montre notamment que l'exposition au froid entraîne une augmentation du cholestérol et de la viscosité sanguine, ce qui altère la circulation et accroît le risque d'événements cardiovasculaires. Certains effets sont immédiats, tels que les crises d'angine de poitrine, les crises d'asthme, les gelures, l'hypothermie ou les crises de drépanocytose. Ces manifestations peuvent survenir dans un délai court, de quelques minutes à quelques heures après l'exposition.

D'autres effets sont différés, survenant entre une et trois semaines, parfois jusqu'à un mois après l'exposition. Ils comprennent principalement les infections respiratoires, les accidents vasculaires cérébraux, les maladies endocriniennes ou les troubles dépressifs. La gravité de ces effets peut être plus importante chez les populations vulnérables, comme les personnes âgées, les femmes, les individus en situation de précarité, les travailleurs exposés au froid en extérieur ou les personnes vivant dans des logements mal chauffés.

Ces observations sont appuyées par Santé publique France (2004), qui a synthétisé les résultats de nombreuses études sur les effets sanitaires du froid. Il mentionne notamment que la vague de froid de janvier 1985 en France a entraîné une surmortalité exceptionnelle de +13 %. Cette hausse globale comprenait notamment une augmentation de +17 % des infarctus du myocarde, de +54 % des accidents vasculaires cérébraux, et de +208 % des cas de pneumonie.

Selon Santé Publique France (2019), les effets du froid sur la mortalité peuvent se manifester avec un délai variable, de trois et vingt et un jours après l'exposition selon les pathologies. La mortalité cardiovasculaire se manifeste le plus souvent entre sept et quatorze jours, tandis que la mortalité respiratoire apparaît généralement entre quinze et trente jours. Enfin, le vent joue un rôle amplificateur notable en accentuant la sensation de froid ressentie ce qui accroît les risques pour la santé.

Une relation en U ou en V

L'exposition à des températures extrêmes, qu'il s'agisse du froid ou de la chaleur accélère la mortalité. Ces faits corroborent l'idée d'une relation non linéaire en forme de U ou de V, caractérisée par une mortalité élevée aux extrêmes de température et à une mortalité minimale à une température intermédiaire appelée température de mortalité minimale (TMM).

Dans ses travaux, Santé publique France (2004) a synthétisé la forme générale de la relation entre la température et la mortalité à partir d'études antérieures. Il la décrit comme une courbe en V, avec un effet beaucoup plus marqué des températures élevées par rapport aux températures basses. Cependant, il précise que malgré cet effet plus intense de la chaleur, le froid est responsable, en valeur absolue, d'un plus grand nombre de décès que la chaleur. Ce qui est également décrit par la saisonnalité de la mortalité, caractérisée par une surmortalité en période hivernale et une sous-mortalité en été et en automne (Peltier, 2016).

L'acclimatation

L'acclimatation est un concept fréquemment évoqué dans la littérature portant sur les effets de la température. Ce phénomène désigne la capacité d'une population à s'adapter à des conditions climatiques non habituelles. Il permet d'expliquer les différences d'impact de la température entre diverses zones géographiques.

En effet, une même température peut ne pas avoir les mêmes conséquences d'une région à l'autre. Par exemple, une température moyenne quotidienne de 25 °C peut être perçue comme extrême dans une région A, tandis qu'elle reste courante dans une autre région B. Dans ce dernier cas, la population est mieux acclimatée : la population y est généralement adaptée à y faire face, et les équipements de régulation thermique (vêtements, climatiseurs, ventilateurs ou chauffages) peuvent par exemple y être plus répandus.

Ces facteurs d'adaptation, liés à l'acclimatation, contribuent ainsi à expliquer les écarts de mortalité observés entre des régions confrontées à des températures similaires mais dont les niveaux habituels d'exposition et les moyens de prévention sont différents.

L'effet moisson de la température

Plusieurs études ont mis en évidence que les personnes les plus vulnérables face aux températures extrêmes sont celles déjà fragilisées sur le plan sanitaire, notamment les personnes âgées ou souffrant de pathologies chroniques. Parmi elles, certaines présentent un risque élevé de décès à très court terme.

Ce phénomène est à l'origine de ce que l'on appelle l'effet moisson, qui désigne une augmentation brutale de la mortalité au moment d'un épisode climatique extrême, suivie d'une période de sous-mortalité. Cette dynamique s'explique par le fait que les décès observés pendant l'événement climatique concernent en grande partie des individus qui seraient probablement décédés dans un avenir proche. En conséquence, les jours ou semaines qui suivent présentent un nombre de décès inférieur à celui attendu, les personnes les plus fragiles ayant déjà succombé précocement.

Facteur d'influence : vent et humidité

Les informations présentées dans cette sous-section sont issues de Météo France (2024b). Selon cette référence, la sensation de chaleur ou de froid ne dépend pas uniquement de la température ambiante mesurée, mais elle peut significativement varier sous l'effet d'autres facteurs environnementaux et spécifiques aux personnes. Ainsi, à température égale, deux personnes peuvent éprouver des sensations différentes de froid ou de chaleur, en fonction de facteurs tels que leur degré d'acclimatation, leur tenue vestimentaire, leur activité pratiquée, ou d'autres facteurs non directement mesurables.

Au-delà de ces facteurs individuels, certains paramètres météorologiques mesurables jouent un rôle essentiel dans la perception réelle de la température et sont susceptibles d'en influencer l'impact sanitaire global. Le vent et l'humidité, en particulier, modifient considérablement la sensation de froid ou de chaleur :

- **Effet du vent** : En cas de froid, l'organisme forme naturellement une fine couche d'air réchauffé à la surface de la peau. Le vent dissipe cette couche protectrice, ce qui accroît la sensation de froid, même si la température ambiante reste constante.
- **Effet de l'humidité** : en période de chaleur, la transpiration et son évaporation constituent le principal mécanisme de refroidissement du corps. Toutefois, lorsque l'humidité de l'air est élevée, l'évaporation de la sueur devient moins efficace, car l'air est déjà saturé en vapeur d'eau. Ce phénomène limite la capacité du corps à se rafraîchir, accentuant ainsi la sensation de chaleur.

La température ressentie, telle que calculée par Météo-France, intègre précisément ces deux paramètres (vitesse du vent et taux d'humidité) afin de fournir une estimation plus représentative du stress thermique effectivement subi. Cette approche permet une appréciation plus fine du risque sanitaire réel, au-delà de la seule température de l'air, et

souligne comment le vent et l'humidité sont des facteurs amplificateurs clés de l'impact global de la température sur la santé.

Synthèse du chapitre

La modélisation de la mortalité joue un rôle essentiel en actuariat, et particulièrement en assurance vie, où les garanties des produits s'appuient sur la durée de vie des assurés. Elle intervient notamment dans les opérations de tarification ainsi que de gestion du risque prudentiel. Avec l'émergence du réchauffement climatique et la multiplication en fréquence comme en intensité des vagues de chaleur qu'il entraîne, la discipline actuarielle doit se préparer à faire face à des surmortalités futures liées à la température.

En effet, les épisodes de températures extrêmes, qu'il s'agisse de vagues de chaleur ou de froid, ont un impact significatif sur la santé, pouvant conduire au décès. Ces impacts peuvent être amplifiés par des conditions de vent ou d'humidité. Il est donc nécessaire pour les assureurs de prendre en compte les conditions climatiques pour produire des estimations plus précises de la mortalité future et ainsi avoir une meilleure maîtrise du risque encouru.

Afin de construire un modèle intégrant les effets de la température, le chapitre suivant présentera les sources de données utilisées ainsi que les tendances qui en ressortent. Ces données regroupent des relevés de température, les décès et la population exposée, observées à une fréquence quotidienne.

Chapter 2.

Présentation des données et analyse descriptive

Ce chapitre est consacré à la présentation des données mobilisées pour la calibration du modèle de mortalité. Trois types de données sont utilisées : les données de température, celles de décès et celles de populations exposées, analysées à une fréquence quotidienne. Il s'organise en deux parties principales : d'une part, l'exposé des sources de données et les traitements effectués ainsi que les contrôles de qualité mis en oeuvre; d'autre part, une analyse descriptive de ces données et la présentation des tendances récentes qui en émergent.

2.1. Présentation des données utilisées

Dans le cadre de cette étude, la pertinence des données est essentielle pour analyser les effets de la température sur la mortalité. Classiquement, l'étude de la mortalité repose sur des données annuelles de décès et d'exposition. Cependant, pour tenir compte de l'influence de la température, caractérisée par des fluctuations importantes au cours de l'année et selon les zones géographiques, il est nécessaire de disposer de données à haute résolution spatiale et temporelle. Ainsi, cette analyse mobilise trois types de données journalières et désagrégées à une échelle géographique fine : les données de température, les données de décès et les données d'exposition au risque de décès.

Les sources retenues pour chacune de ces variables sont les suivantes : les données de température proviennent de Météo France, les données de décès sont issues de l'INSEE, et les données d'exposition résultent d'un croisement entre les données de l'INSEE et celles de la Human Mortality Database (HMD), aucune source unique ne fournissant, à notre connaissance, d'estimations journalières de population.

2.1.1. Les données de température

2.1.1.1. Description du jeu de données

Depuis le 1^{er} janvier 2024, Météo France a rendu accessibles en open data les données relatives aux températures mesurées sur l'ensemble de son réseau de stations¹. Ces données contiennent l'intégralité des relevés de températures des stations qui mesurent diverses grandeurs météorologiques, dont la température. Parmi l'ensemble des variables disponibles, celles utiles à cette étude et qui ont été extraites sont les suivantes :

- Les températures quotidiennes : minimale, moyenne et maximale ainsi que la variable indiquant la date de la mesure.
- Les variables d'identification et de localisation des stations : identifiant, nom, longitude, latitude et altitude.

Table 2.1. : Aperçu des données de température

NUM_POSTE	NOM_USUEL	LAT	LON	ALTI	date	TX	TN	TM
75113002	SALPETRIERE	48.8	2.4	34	1989-02-07	13.4	5.3	NA
71047001	BOURBON LANCY	46.6	3.8	260	1954-08-13	NA	NA	NA
26307001	ST-JEAN-EN-ROYA	45.0	5.3	308	2011-02-18	9.5	1.4	4.5

L'un des principaux atouts de ces données est la multiplicité des stations disponibles et la couverture exhaustive du territoire qu'elles offrent. Au total, ce sont 9 320 stations qui sont disponibles et réparties sur l'ensemble du territoire français.

À la date d'extraction, les jeux de données sont organisés en fichiers distincts, regroupés par département et par période historique : avant 1950, 1950-2023 et 2023-2024. Il a été jugé pertinent de retenir les données à partir de 1950 pour plusieurs raisons : les données antérieures à cette date présentent une couverture temporelle hétérogène selon les départements et une qualité parfois insuffisante. De plus, les données de mortalité ne sont disponibles qu'à partir d'une période plus récente, d'autant plus que sur cette période elles auraient été trop anciennes pour être pertinentes dans la prévision de mortalité récente et de la mortalité future.

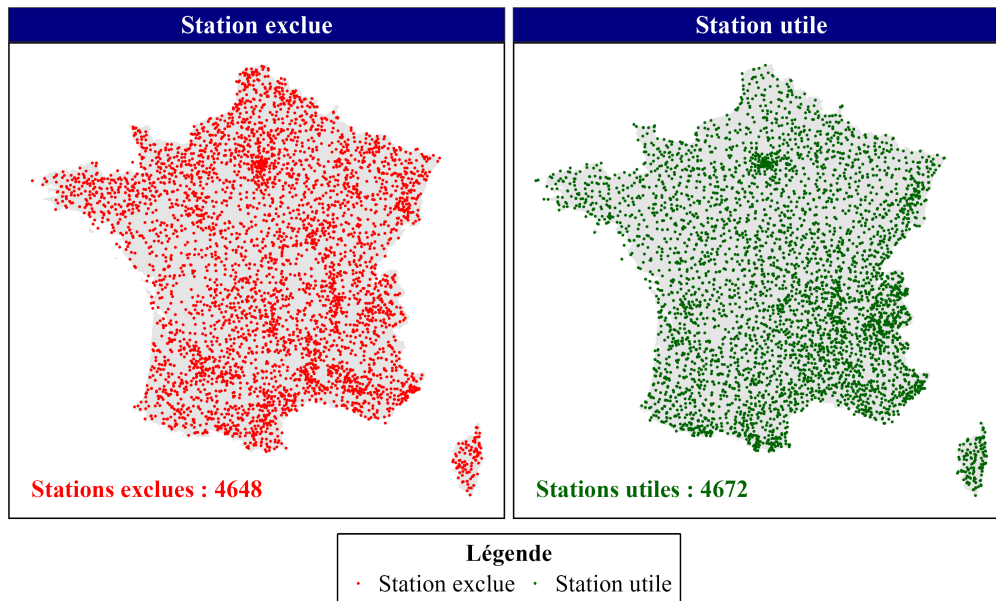
2.1.1.2. Traitements sur les données

Les données Météo France ont de nombreux atouts mais souffrent de certaines lacunes qui doivent être traitées. Les principaux problèmes observés sont les suivants :

¹Données de température par station de Météo France

Présence de stations ne mesurant pas la température

Cela concerne un total de 4 648 (49.9%) stations. Celles-ci sont vraisemblablement utilisées pour mesurer d'autres variables météorologiques que la température. Par conséquent, elles sont sans intérêt pour cette étude et il convient de les exclure de l'analyse. Les stations restantes sont néanmoins assez nombreuses et uniformément réparties sur l'ensemble du territoire, ce qui garantit le maintien de la qualité de données (Figure 2.1).



Les valeurs manquantes et la présence discontinue des stations

La plupart des stations n'ont pas été actives sur l'ensemble de la période d'analyse, elles ont progressivement été ouvertes et fermées au fil du temps. De plus, elles présentent fréquemment des valeurs manquantes ponctuelles ou sur certaines périodes.

En effet, comme l'illustre la Figure 2.2, une station peut être inactive pendant une longue période. Dans certains cas, on peut aussi observer des valeurs manquantes aux extrémités de la période d'existence d'une station. Ces extrémités manquantes étant non informatives, seule la période comprise entre les premières et dernières dates de relevé de température a été retenue pour chaque station.

Suite à ce premier traitement, des méthodes d'interpolation temporelle et spatiale ont été envisagées pour imputer les valeurs manquantes restantes. L'approche d'interpolation temporelle suppose qu'en un même point, la température ne fluctue pas fortement d'un jour à l'autre, permettant d'estimer une donnée manquante à partir des valeurs des

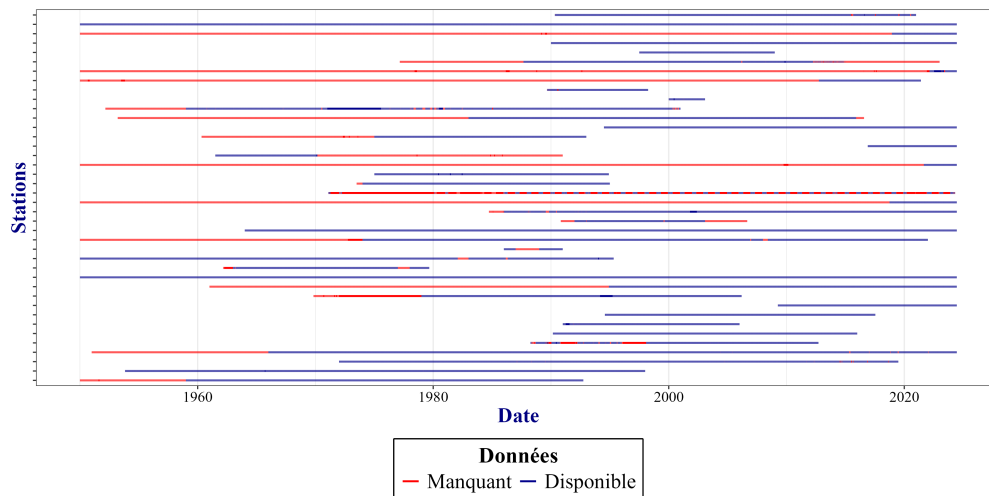


Figure 2.2. : Disponibilité et valeurs manquantes de 40 stations

jours voisins. Toutefois, cette méthode a été écartée en raison de la présence de valeurs manquantes sur de périodes longues et continues. En effet, le plus proche voisin temporel peut être éloigné de plusieurs jours voire de plusieurs mois ce qui introduirait un biais dans les estimations. C'est pourquoi l'approche d'interpolation spatiale a été privilégiée bien que plus coûteuse en temps de calcul.

L'interpolation spatiale postule que la température varie faiblement entre deux points géographiquement proches. Ainsi, à une date donnée, la température manquante d'une station est estimée par la moyenne des stations voisines. La forte densité du réseau de stations rend cette hypothèse raisonnable. Il peut notamment être vérifié que la distance moyenne à la station la plus proche est de 5.1 km, et qu'elle reste inférieure à 24.5 km dans tous les cas (Figure 2.3).

Des tests menés sur un sous-échantillon de stations ont indiqué que le nombre optimal de stations voisines à prendre en compte était de 5. Ainsi, pour chaque date, l'imputation des valeurs manquantes est réalisée en calculant la moyenne des 5 stations les plus proches disposant d'une mesure.

Calcul des températures départementales

Le calcul de la température moyenne départementale repose sur l'hypothèse implicite d'une relative homogénéité spatiale des températures à l'intérieur du département. Ainsi, à chaque date, les températures mesurées par les différentes stations peuvent ainsi être considérées comme des réalisations d'une même variable aléatoire, dont l'espérance correspond à la température moyenne théorique du département. Dans ce cadre, et par souci de simplicité, la moyenne simple des températures observées est retenue pour estimer la température moyenne du département.

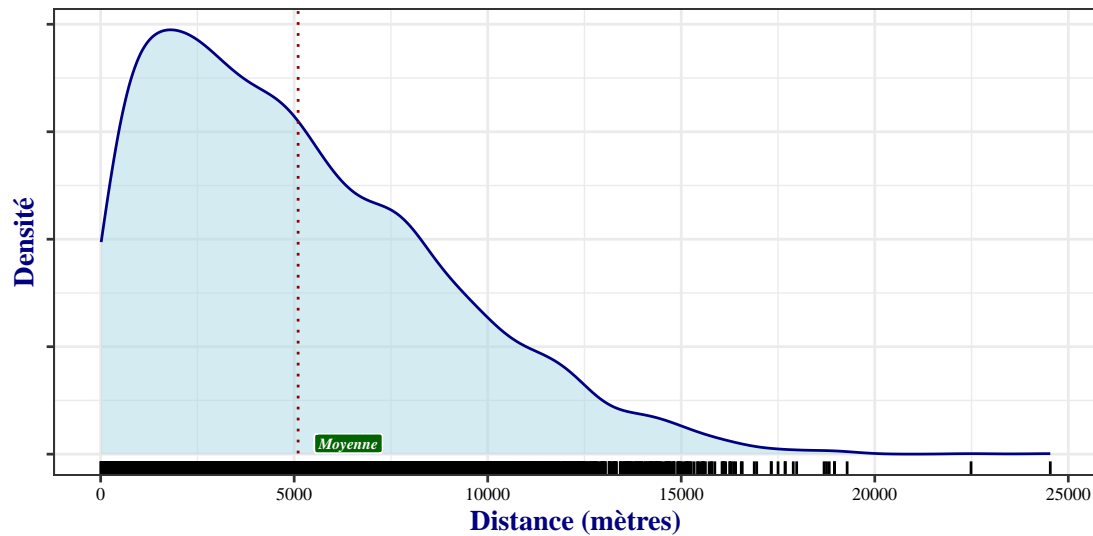


Figure 2.3. : Distribution des distances à la station la plus proche

2.1.2. Les données de décès

2.1.2.1. Description des données

En matière d'étude de mortalité, les données de décès de référence qui sont régulièrement utilisées sont celles de la Human Mortality Database² qui recense le détail des décès et des expositions par âge, année et sexe sur une période large. Toutefois, dans le cadre de cette étude, les données de la HMD ne sont pas adaptées, car elles ne contiennent pas le détail des décès par département ni par jour.

Les données de décès privilégiées sont celles du fichier national des personnes décédées³ mis à disposition par l'INSEE depuis 2019 et mis à jour mensuellement. Ce fichier recense les personnes décédées en France depuis le 1^{er} Janvier 1970 et contient des informations détaillées dont :

- Les *noms et prénoms* : qui ont notamment servi à l'identification des doublons
- Les *dates de naissance* et de *décès* : qui permettent de déterminer l'âge au décès
- Les *communes de naissance* et de *décès* : permettant d'en déduire le département de décès
- Le sexe
- Des *informations d'état civil* relatives aux décès.

²Données de la Human Mortality Database

³Fichier national des personnes décédées

Table 2.2. : Aperçu du fichier des personnes décédées

nom	prenoms	sexe	date_naissance	code_insee_naissance	...
CLERC	SIMONNE,EUGENIE	F	19171006	75114	...
FOISSIER	HELENE,SUZANNE	F	18950825	08105	...
FILIPPI	FRANCOIS,JOSEPH	M	19071107	83137	...
REVEL	ROSALIE	F	19021206	67061	...
DOFFE	SOPHIE,EUGENIE	F	18980403	59118	...

Ces détails sont mis à disposition en précisant que les informations des personnes décédées ne constituent pas des données à caractère personnel et peuvent, de ce fait, être librement publiées. L'INSEE a également précisé, en cas d'exploitation de ce fichier, que la complétude et la qualité des données ne sont pas garanties, notamment pour les premières années de la période. C'est pourquoi une analyse préliminaire a été menée afin d'évaluer la fiabilité de ces données.

2.1.2.2. Traitement du fichier des décès de l'INSEE

À partir des informations présentes, l'obtention des données souhaitées s'effectue par un simple comptage, en distinguant le sexe, l'âge au décès, le département et la date de décès. Les variables âge et département de décès n'étant pas directement disponibles, elles doivent être calculées :

- L'âge au décès est calculé comme la différence entre la date de naissance et la date de décès arrondi en défaut ce qui correspond à l'âge révolu à la date d'anniversaire.
- Le département de décès correspond aux 2 premiers caractères du code INSEE de la commune de décès. A l'exception des départements d'outres-mers qui sont sur 3 caractères mais qui ne rentrent pas dans le champ de cette étude. De plus, les départements Corse-du-sud (2A) et Haute-Corse (2B) ont été regroupés en un seul département Corse (20).

Restriction à la France Métropolitaine

A la date d'extraction des données (Juin 2025), le fichier contenait 27 836 139 décès. Parmi eux, les 661 954 (2,3%) correspondant à des décès survenus en outre-mer ou à l'étranger ont été exclus pour se restreindre à la France métropolitaine.

Cas des dates de naissance et de décès inconnues

Dans les données restantes, 104 444 problèmes de date de naissance et 611 problèmes de date de décès ont été observés.

Pour les dates de naissance, il s'agit essentiellement de personnes nées à l'étranger dont la date de naissance exacte n'est pas toujours connue. Elles se présentent sous des formes telles que "2000-XX-XX", "2000-10-00" ou "2000-10-XX" indiquant l'absence du mois ou du jour de naissance. Bien que la date ne soit pas connue avec précision, l'année de naissance est quasiment toujours connue et peut être extraite, permettant d'estimer l'âge avec une erreur absolue inférieure à 1 an. Pour cela, à partir de l'année et pour les cas de date approximative, les valeurs des mois et des jours sont imputées aléatoirement afin d'éviter une surreprésentation des décès à un âge spécifique. Enfin, dans 535 cas, l'année de naissance n'était pas disponible et les décès correspondants ont été supprimés.

Les dates de décès inconnues se manifestent également par des mois ou jours de décès non précisés. Dans ce cas, même si l'année de décès est connue, l'imprécision sur la date exacte rend impossible la connaissance des conditions de température associées au décès. Aucune méthode de correction n'a été jugée utile face à ce problème et ces 611 cas ont donc été supprimés.

Une fois ces variables traitées, la variable âge a pu être créée. Quelques valeurs aberrantes ont été notées, telles que des âges négatifs ou extrêmement grands. Le seuil de 120 a été fixé comme seuil limite d'âge réaliste.

Cas de départements de décès inconnus

La variable département de décès est dérivée du code INSEE de la commune de décès, composé de 5 caractères dont les 2 premiers correspondent au code INSEE du département (à l'exception des DROM-COM où il s'agit des 3 premiers). En appliquant cette méthode, 10 935 cas où le code INSEE n'est pas sur 5 caractères ou dont les deux premiers sont "00" ont été identifiés, empêchant ainsi la détermination du département. Ces cas ont également été supprimés, car il était impossible de connaître les températures associées à ces décès. Il est à noter que ces cas de département inconnu se sont montrés fortement concentrés autour des années 2010 et 2011.

Les doublons et l'incomplétude des données

Comme mentionné par l'INSEE et comme l'a souligné Salaun (2024) lors de son analyse, ces données contiennent de nombreux doublons. Leur présence fausse directement les dénombrements de décès, puisque certaines personnes sont comptées plusieurs fois.

Un doublon correspond à la présence d'un même individu sur plusieurs lignes du jeu de données, potentiellement causé par une erreur de saisie ou de traitement. Souvent et dans la plupart des cas, ils se manifestent par une redondance exacte des mêmes informations sur plusieurs lignes. En appliquant cette définition stricte des doublons, seulement 50 % des vrais doublons sont identifiés. En effet, ici, la nature des doublons s'avère plus complexe, car le fichier contient des erreurs (nom mal orthographié, dates erronées, etc)

faisant que des lignes aux informations non strictement identiques peuvent correspondre au même individu (ce qui n'est pas pris en compte dans la première approche).

Pour identifier les doublons plus efficacement, une nouvelle règle a été définie : *deux décès sont considérés comme un doublon si les personnes concernées ont les mêmes nom et prénom, la même date de naissance, ainsi que la même date de décès ; ou bien si elles ont les mêmes nom et prénom, la même date de naissance et la même commune de décès.* En effet, il serait très improbable que deux individus distincts partagent les mêmes nom et prénom, naissent le même jour et meurent le même jour. La seconde condition, qui prend en compte la commune de décès, permet de compenser d'éventuelles erreurs sur la date de décès. L'application de cette règle a permis d'identifier un total de 213 715 doublons (moins de 1%).

Une analyse plus fine a permis de déterminer la période de survenue des doublons. Ceux-ci apparaissent majoritairement autour des années 1999 et 1997, sans être associés à une zone géographique particulière.

Pour mieux décrire leur survenance, les décès quotidiens en 1997 et 1999 ont été comparés à une estimation de la mortalité "normale" à cette période, calculée comme la moyenne des décès du même jour sur la période 1994-2004 en excluant les années 1997, 1999 et 2003 en raison d'anomalies (doublons) ou d'événements extrêmes (canicule) survenus ces années-là (Figure 2.4):

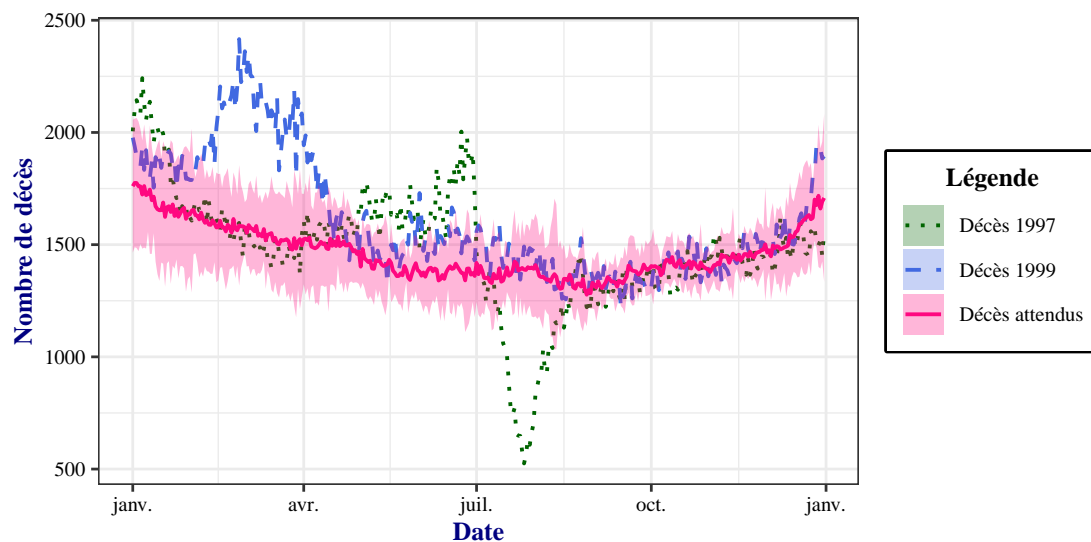


Figure 2.4. : Décès quotidiens et doublons en 1997 et 1999

On observe qu'en 1999, la mortalité est anormalement élevée sur la période de février à avril, et en 1997, de mai à Juillet. Cette surmortalité s'explique par la présence de doublons. De plus, on constate qu'en 1997, après juillet, une période de sous-mortalité est apparue, le nombre de décès étant tombé à environ 500 alors qu'il est habituellement

de l'ordre de 1 500. Cette sous-mortalité, en raison de son ampleur et du manque de justification, a été signalée à l'INSEE qui a dit reconnaître ce biais mais a fait savoir qu'il n'est pas possible de corriger le fichier des personnes décédées.

Fiabilité des données

Pour évaluer la qualité des données, une comparaison avec les données annuelles des décès de la base de référence HMD a été effectuée :

[1] "F:/Mes documents 2025/Rédaction mémoire"

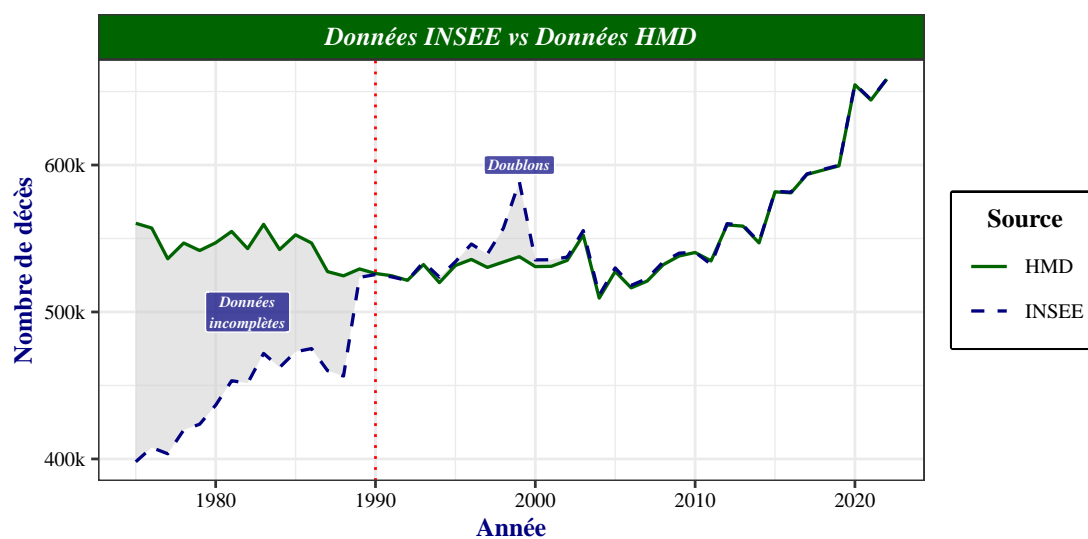


Figure 2.5. : Données INSEE vs Données HMD

Cette comparaison révèle deux principaux écarts :

- Avant 1990, les décès sont clairement incomplets, avec des écarts dépassant 100 000 décès par rapport aux données HMD. L'INSEE avait déjà indiqué que le fichier peut être incomplet en début de période, notamment entre 1970 et 1975. Cette analyse montre que l'incomplétude s'étend jusqu'aux années 1990.
- Autour de 1999, un pic important de mortalité apparaît par rapport aux données HMD. Ce qui reflète la présence des doublons à cette période. Les doublons de l'année 1997 sont moins visibles à cause de la sous-mortalité observée la même année qui cause une forme de compensation au niveau global. La Figure 2.5 illustre l'impact des doublons et l'importance de les corriger.

Le tableau 2.3 présente le récapitulatif des problèmes traités dans le fichier des personnes décédées :

Table 2.3. : Traitement des décès en France Métropolitaine

Libellé	Nombre de décès	Pourcentage (%)
Nombre de décès initial	27836139	100,0000%
Doublons des décès	213715	0,7678%
Age négatif	66	0,0002%
âge supérieur à 120	34	0,0001%
Dates inconnues (naissance ou décès)	1146	0,0041%
Département inconnu	10935	0,0393%
Décès avant 1970	2185	0,0078%
Décès supprimés (hors doublons)	13889	0,0499%
Décès supprimés (avec doublons)	227604	0,8177%

Après traitement, une meilleure façon d'apprécier la fiabilité et les limites des données consiste à analyser l'écart relatif par rapport aux données HMD :

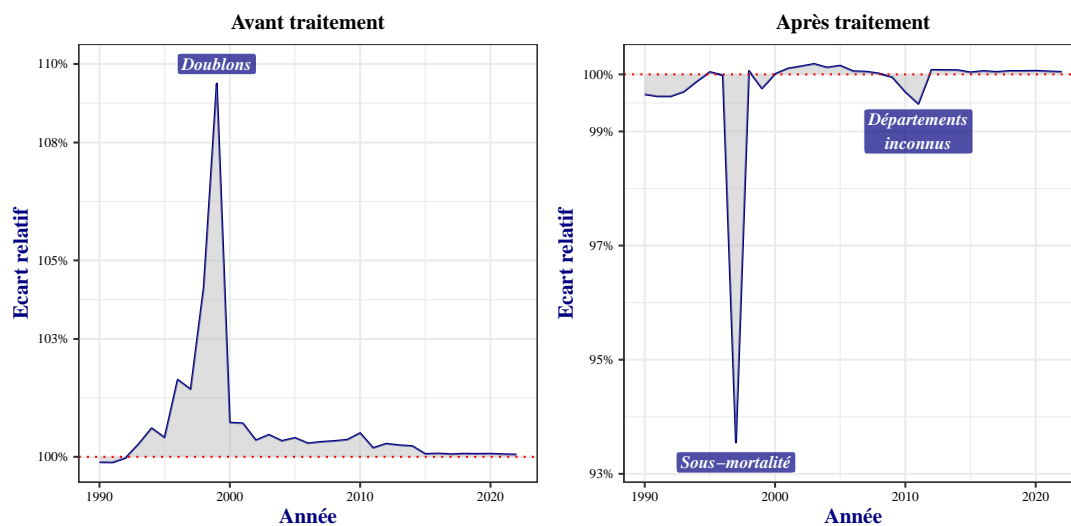


Figure 2.6. : Décès INSEE / Décès HMD avant et après traitement

La Figure 2.6 est construite sur la période après 1990 car, comme cela a été montré, la période avant 1990 contient un grand manque de décès dont la représentation déforme les écarts. Elle montre les écarts relatifs entre les données du fichier de l'INSEE et celles de la HMD, avant et après traitement, ce qui permet de visualiser les effets des corrections.

Avant traitement, on observe une supériorité quasi-absolue des données INSEE qui est majoritairement attribuable à l'effet des doublons. Sans correction, une correspondance quasi parfaite entre les deux sources est cependant observable à partir de 2015.

Le traitement a principalement consisté en la suppression des doublons et des cas de données aberrantes ou indisponibles. Les doublons constituant le problème prépondérant et les cas de données indisponibles étant concentrés aux années 2010 et 2011 (départements inconnus). Après ces corrections, en dehors de 1997, les données INSEE s'alignent très bien aux données HMD avec des écarts de l'ordre de moins de 0,5 %. Le pic soudain autour de 2010 correspond à la forte concentration des décès supprimés étant majoritairement des cas de départements inconnus et le pic important en 1997 est l'effet de la sous-mortalité qui a été signalée plus haut.

Ces derniers écarts seront ajustés par la suite en corrigeant l'exposition de sorte à obtenir des taux de mortalité équivalents à ceux de la base HMD.

2.1.3. Les données d'exposition au décès

2.1.3.1. Description des données

Après les données de température et de décès, les dernières informations nécessaires pour nos analyses sont celles relatives à l'exposition au décès.

La structure de données fine qui est visée est très restrictive. Aucune source de données en accès libre ne correspondait à ces exigences. Par conséquent, il a été nécessaire de construire un jeu de données d'exposition quotidienne à l'échelle départementale. Pour cela, les données d'exposition de la base HMD et celles de population départementale de l'INSEE ont été combinées :

- **Les données d'exposition HMD** : elles fournissent les nombre d'exposition aux décès ventilés par sexe, âge, et année sur une longue période et à l'échelle de la France métropolitaine. Leur limite est l'absence de détail par département et de ne pas disposer d'un pas de temps quotidien.
- **Les données de population de l'INSEE** : Elles contiennent le détail de la population française ventilée par département, sexe, groupe d'âge quinquennal et par année de 1975 à 2022. Leur principal défaut est l'absence de détail par âge exact et de pas de temps quotidien. Elles sont toutefois utiles pour estimer la répartition de l'exposition entre les départements.

2.1.3.2. Construction de la base de données d'exposition

La part d'exposition associée à chaque département est calculée selon la formule suivante :

$$\omega_{xyd}^s = \frac{P_{classe(x)yd}^s}{\sum_{d=1}^{95} P_{classe(x)yd}^s}$$

où x , y , s et d représentent respectivement l'âge, l'année calendaire, le sexe et le département. $classe(x)$ désigne le groupe d'âge quinquennal auquel appartient l'âge x .

L'hypothèse sous-jacente à l'utilisation de ces poids est que, au sein d'un même groupe d'âge, la répartition de la population entre les départements est la même quel que soit l'âge exact. Cette hypothèse permet, en appliquant ces poids aux données d'exposition de la HMD, d'obtenir une estimation de l'exposition par département, sexe, âge et année.

Pour disposer d'une exposition quotidienne, on fait l'hypothèse d'une évolution constante de la population au cours de l'année.

2.1.4. Constitution du jeu de données global

Construction de la base globale

Les trois bases de données de température, de décès et d'exposition étant maintenant disponibles, l'objectif est de les fusionner en une base unique intégrant les trois variables d'intérêts et qui servira de fondement à nos analyses.

La difficulté principale réside dans la volumétrie massive des données. La jointure nécessitant un format commun, la période temporelle la plus longue couverte par l'ensemble des bases s'étend de 1975 à 2022, soit un total de 17 532 jours. La plage d'âge disponible va de 0 à 120 ans pour un total de 95 départements et 2 sexes. Le nombre total de lignes potentielles approche ainsi les 403 millions, un volume impossible à manipuler avec le matériel à disposition.

Face à cette limite, une simplification des données a été faite via une agrégation par valeur entière de température. Cette démarche vise à faire abstraction des variables *département* et *date* tout en conservant les informations relatives aux températures départementales quotidiennes. La procédure appliquée est la suivante :

1. Jointure simple des trois bases de données
2. Pour chaque date, regroupement des départements ayant enregistré la même température, en agrégeant leurs expositions et leur nombre de décès. Cette étape revient à découper le territoire en zone de températures égales au sein desquelles les décès et l'exposition sont agrégés (Figure 2.7). Ces zones varient chaque jour en fonction des températures observées. Après cette étape, la variable département est supprimée.
3. Création de la variable *année*. Puis, de façon similaire à l'agrégation des départements, pour chaque année, une agrégation des jours en distinguant les températures correspondantes est effectuée. Cette étape permettra d'obtenir des données d'exposition et de décès au format usuel d'âge, de sexe et d'année auquel s'ajoute une décomposition par "température au moment du décès".

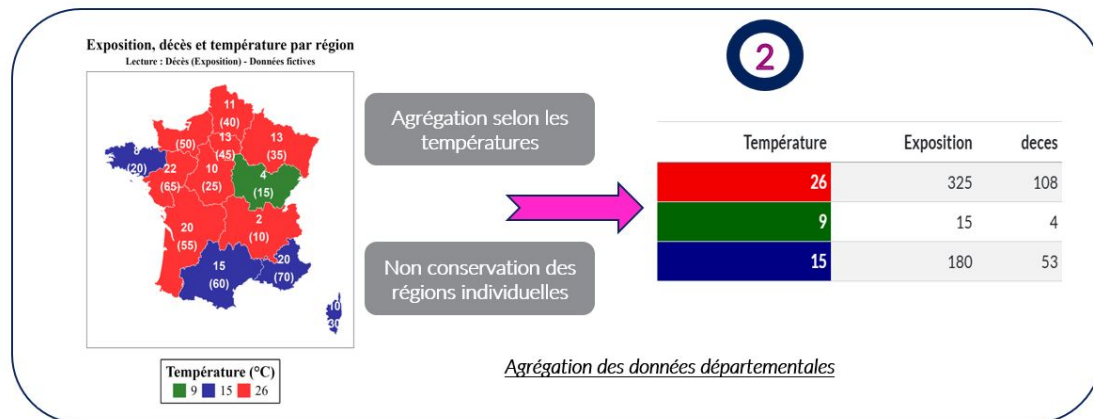


Figure 2.7. : Illustration de l'agrégation des départements

Les effets du manque de données avant 1990

La base globale réunit toutes les variables utiles à l'étude de la mortalité, notamment le nombre de décès et l'exposition associée à chaque groupe de population. Une évaluation initiale des taux de mortalité par âge, pour chaque année, révèle une forme irrégulière aux âges élevés dans la courbe de mortalité, ceci en début de période jusqu'aux environs de 1990 (Figure 2.8) :

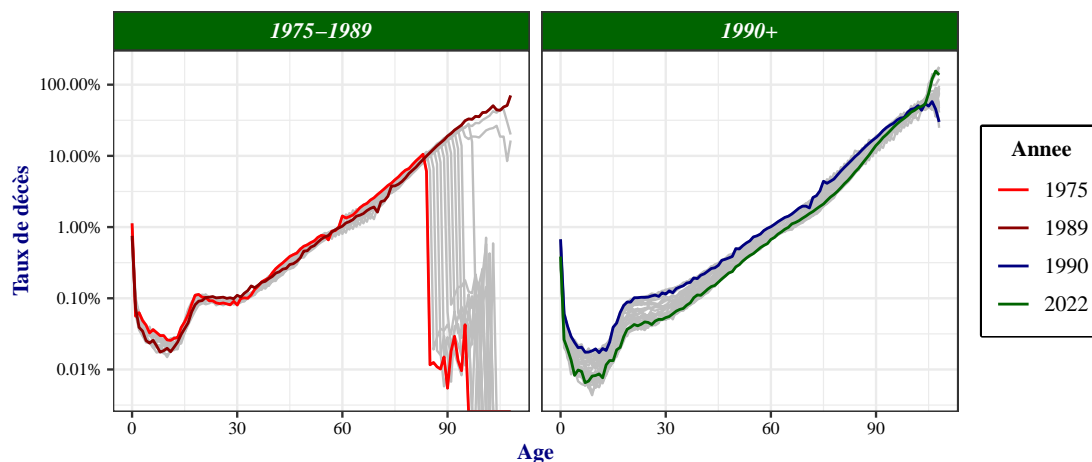


Figure 2.8. : Taux de mortalité par âge et par période

Cette période correspond à celle où les données de l'INSEE se sont montrées incomplètes, avec des écarts importants par rapport aux données de la HMD. Les données manquantes concernent principalement les personnes âgées, ce qui explique cette rupture brutale dans la forme des courbes de mortalité. Par conséquent, les données antérieures à 1990 n'étant

pas fiables, c'est à partir de cette date que la construction des modèles de mortalité sera réalisée.

2.2. Les évolutions thermiques et démographiques

De nombreux travaux ont présenté des résultats relatifs à la mortalité et au climat. Dans cette sous-section, il s'agira, à partir des données collectées, de mettre en lumière les insights qui en émergent et de les confronter aux résultats de la littérature.

2.2.1. Les évolutions des températures

L'augmentation des températures moyennes

À l'échelle de la France, les températures moyennes décennales présentent une tendance globale à la hausse, bien que cette évolution ne soit pas uniforme dans le temps. Sur la période 1959-2023, la température augmente en moyenne de 0,3 °C par décennie, ce qui est cohérent avec la valeur rapportée par Gibelin (2015). Dans son étude, elle soulignait déjà une accélération préoccupante du réchauffement, en comparaison avec la période 1901-2000 marquée par une augmentation de 0,1 °C par décennie.

Ce changement de tendance s'explique principalement par l'accélération de la hausse des températures observées depuis la fin des années 1970. Plus précisément, la température moyenne avant 1980 s'élevait à 10,6 °C, contre 11,5 °C par la suite. De plus, depuis les années 1980, le rythme de réchauffement atteint à 0,41°C par décennie (Figure 2.9).

La tendance décrite ici reflète l'évolution des températures à l'échelle de la France métropolitaine dans son ensemble. Toutefois, cette vision agrégée pourrait masquer des disparités régionales. Ainsi, certains territoires, notamment au niveau départemental, peuvent présenter des dynamiques particulières de température qu'il convient d'examiner.

Des tendances similaires dans les départements

À l'échelle départementale, la quasi-totalité des départements enregistre une évolution proche de la tendance nationale. Jusqu'en 1980, les températures y sont relativement stables. À partir de cette date, la plupart des départements connaissent une accélération conjointe du rythme d'augmentation des températures. Cette évolution commune témoigne d'une uniformité de l'impact du changement climatique sur l'ensemble du territoire.

Néanmoins, malgré ces tendances similaires de croissance, les départements présentent une hétérogénéité en termes de niveau de température. Certains départements sont plus

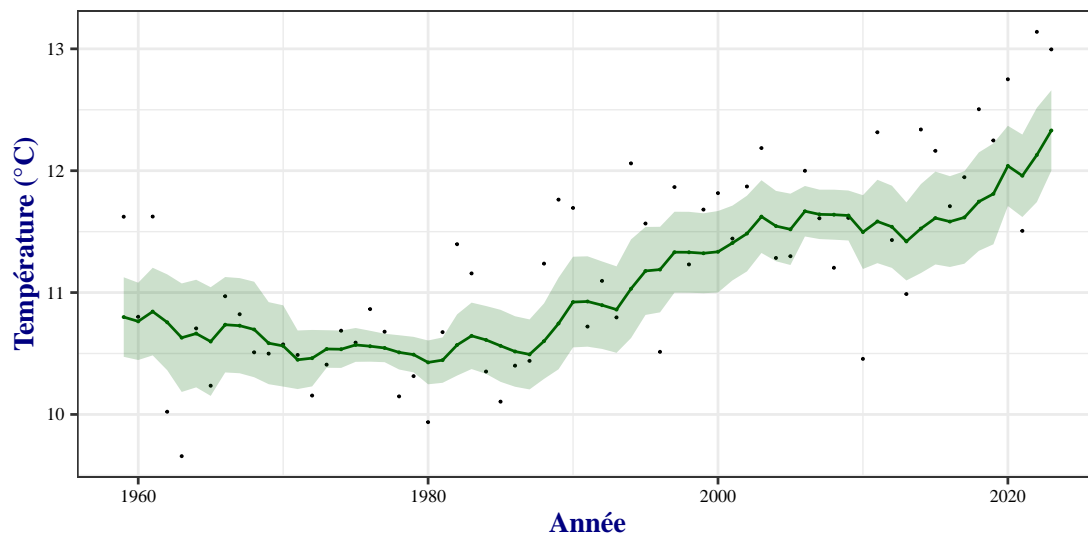


Figure 2.9. : Température en moyenne annuelle (points) et en moyenne décennale (ligne)

chauds, comme les *Bouche-du-Rhône* (13) ou la *Corse* (2A/2B), tandis que d'autres sont plus froids, tels que les *Hautes-Alpes* (5) ou la *Savoie* (73).

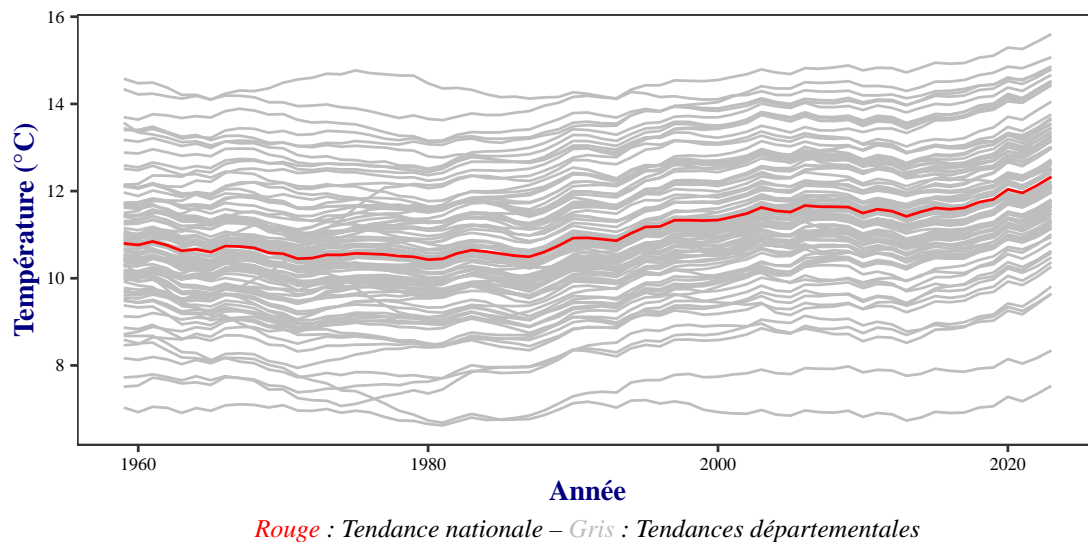


Figure 2.10. : Température départementale en moyenne décennale

À une échelle plus fine, comme celle des cantons français, l'hétérogénéité spatiale des températures est plus visible, faisant apparaître des disparités thermiques significatives au sein des zones du territoire. Sur l'ensemble de la période 1950-2022, certaines zones se

distinguent par des températures systématiquement plus élevées. C'est notamment le cas du Sud-Est en bordure méditerranéenne, qui constitue la région la plus chaude du pays. En parallèle, une dynamique de réchauffement soutenue est observable dans une large bande allant du sud-Ouest vers l'Ouest du pays, où les températures ont progressivement augmenté atteignant un niveau élevé sur la période la plus récente. À l'inverse, les régions s'étendant entre du centre au Nord-Est de la France présentent des températures plus basses et une tendance au réchauffement plus modérée (Figure 2.11).

LON	LAT	LON	LAT
-5.130333	41.374833	9.518000	51.055833
LON	LAT	LON	LAT
-5.130333	41.374833	9.518000	51.055833
LON	LAT	LON	LAT
-5.130333	41.374833	9.541500	51.055833
LON	LAT	LON	LAT
-5.086602	41.374833	9.541500	51.055833

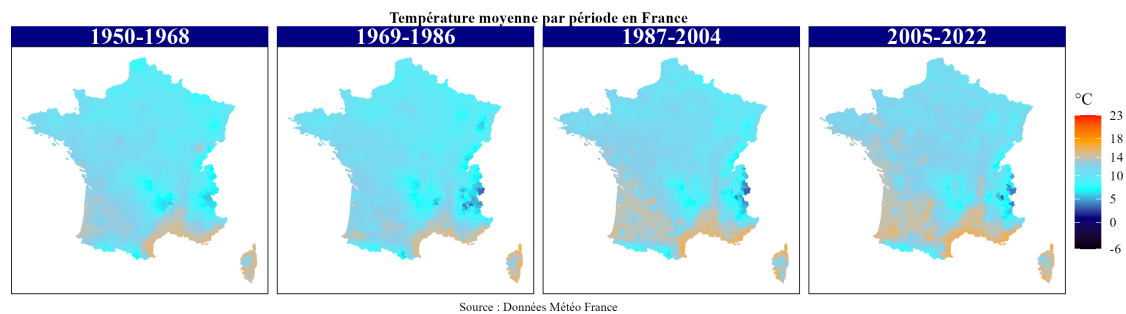


Figure 2.11. : Température moyenne par période en France

Ces résultats sont cohérents avec le découpage climatique de la France en trois grandes zones : le climat méditerranéen (zone H3), associé aux températures les plus élevées dans le Sud-Est ; le climat océanique (zone H2), caractéristique de l'ouest du pays et marqué par une élévation progressive des températures; et enfin le climat semi-continentale (zone H1), présent dans l'Est et le Nord-Est, où les températures sont historiquement plus basses. Salaun (2024) a également mis en évidence des écarts de températures selon ces zones climatiques et a observé une dynamique similaire (Figure 2.12).

Fréquence, durée et intensité des vagues de chaleurs

Les vagues de chaleur, y compris les canicules constituent le principal centre d'attention des études portant sur l'impact des températures extrêmes, car c'est au cours de ces phénomènes que surviennent des pics de mortalité. Avec le dérèglement climatique et

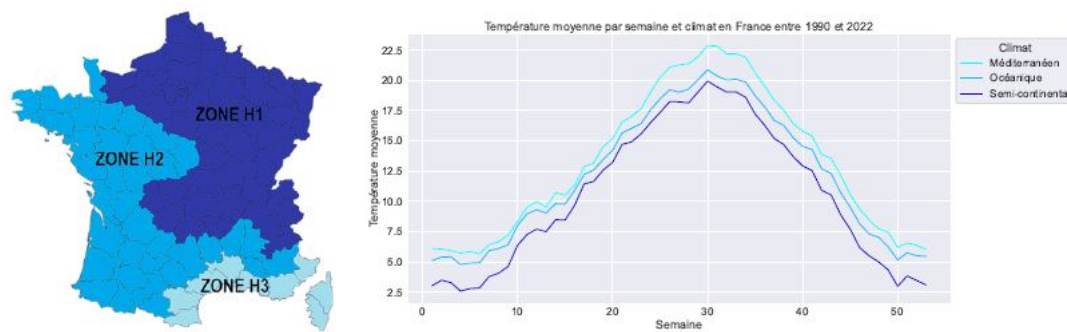


Figure 2.12. : Zone climatique française et profil de température (Source : Salaun (2024))

l'augmentation générale des températures, ces phénomènes deviennent plus alarmants tant par leur fréquence, leur intensité que leur sévérité (Pincemin, 2021).

Chaque année les canicules se produisent principalement en été entre juillet et septembre, plus rarement dès le mois de juin, comme l'illustre la Figure 2.13.

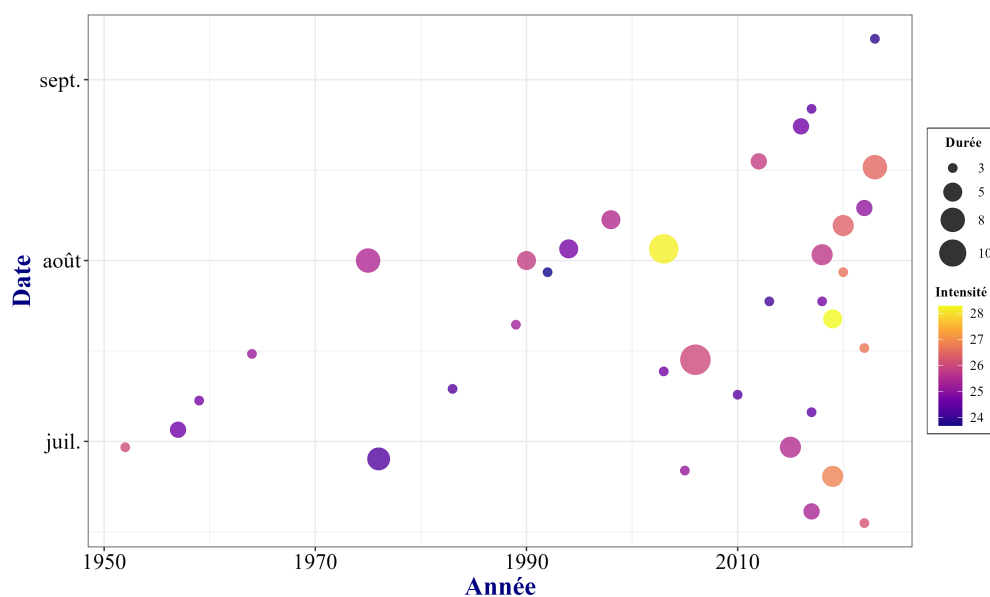


Figure 2.13. : Les vagues de chaleur en France depuis 1950

En termes de fréquence, les vagues de chaleur étaient relativement rares avant les années 1990. Depuis, leur occurrence s'est intensifiée, au point où des canicules surviennent presque tous les ans depuis les années 2000. Cette concentration croissante sur cette période illustre également l'aggravation des phénomènes de chaleur extrême.

Outre leur fréquence, l'augmentation de l'intensité et de la durée des vagues de chaleur est tout aussi préoccupante. L'intensité d'une vague de chaleur s'entend comme la température moyenne journalière maximale atteinte au cours de l'épisode de chaleur. Depuis 1950, la canicule de 2003 a été la plus intense, à égalité avec celle de 2019 avec une intensité de 28.3 °C. Toutefois, la canicule de 2019 n'a pas été aussi exceptionnelle que celle de 2003 en raison d'une durée plus brève (5 jours contre 12 jours).

Pour mieux analyser le caractère exceptionnel de ces événements, on utilise souvent la sévérité, définie comme le cumul des excès de température au-delà du seuil de déclenchement sur toute la durée de l'épisode. À ce titre, la canicule de 2003 reste la plus sévère historiquement. Néanmoins, depuis les années 2010, les canicules montrent une tendance croissante à la sévérité, illustrant une aggravation du risque.

La saisonnalité des températures

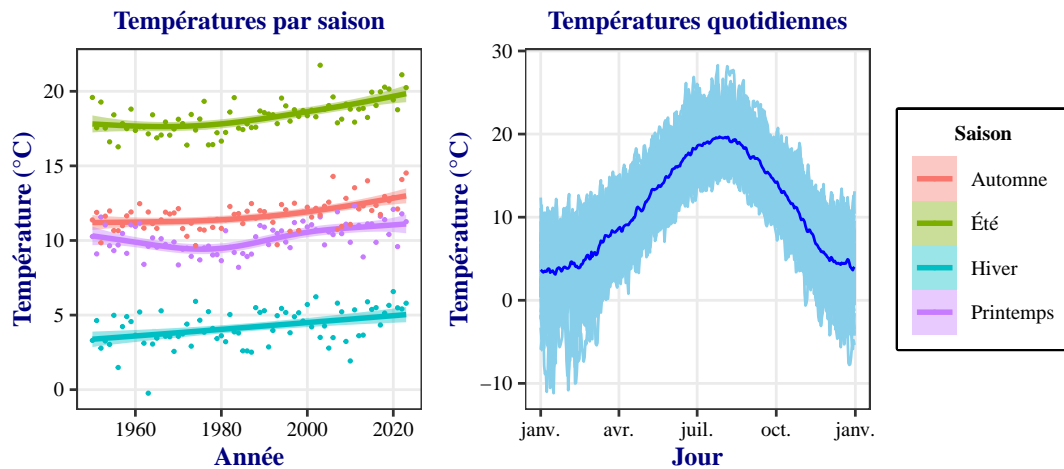
La tendance générale à la hausse des températures moyennes masque une forte composante saisonnière, qui s'avère essentielle pour l'analyse des effets associés à la température. En effet, les températures suivent un cycle annuel marqué, caractérisé par des variations prononcées entre les saisons. Les vagues de chaleur, évoquées précédemment, surviennent typiquement l'été, tandis que les vagues de froid, bien que devenant plus rares avec le réchauffement climatique, restent caractéristiques de l'hiver.

De façon générale, l'hiver (décembre à février) constitue la période la plus froide de l'année, tandis que l'été (juin à août) enregistre les températures les plus élevées. Le printemps (mars à mai) et l'automne (septembre à novembre) sont des saisons de transition, durant lesquelles les températures connaissent une évolution progressive à la hausse ou à la baisse. Ces différences saisonnières sont mis en évidence par la Figure 2.14 : l'été se distingue par des températures moyennes proches de 18°C, alors que l'hiver présente des moyennes proches de 5°C. Le printemps et l'automne affichent des niveaux similaires, avec toutefois un léger avantage thermique pour l'automne, les températures moyennes variant généralement entre 10 °C et 13 °C.

Par ailleurs, il est important de noter que le réchauffement climatique affecte l'ensemble des saisons. On observe ainsi une augmentation nette et continue des températures moyennes saisonnières, indépendamment de la période de l'année considérée.

2.2.2. Les évolutions des décès en France

Dans cette sous section, nous analysons les principales tendances qui se dégagent de la série quotidienne des décès observés sur l'ensemble de la période d'étude.



Bleu : Tendance moyenne – *Bleu ciel* : Fluctuations quotidiennes

Figure 2.14. : La saisonnalité des températures

La saisonnalité des décès

L'examen de l'évolution journalière du nombre de décès révèle en premier lieu une forte saisonnalité. Tout comme les températures, la mortalité suit un rythme dicté par les saisons. Elle affiche des pics systématiques en hiver et une mortalité plus faible en été.

La tendance annuelle moyenne des décès dessine une forme en U : Le nombre de décès quotidien est généralement élevé en janvier (autour de 1800 décès), décroît progressivement jusqu'à un minimum en juillet (environ 1400 décès) puis augmente à nouveau jusqu'à la fin de l'année. Toutefois, cette tendance saisonnière peut être perturbée par des événements exceptionnels tels que les crises sanitaires ou climatiques.

C'est ce que montre le graphique de droite de la Figure 2.15. Les importants écarts par rapport à la courbe moyennes sont causés par des événements marquants survenus lors d'années particulières :

- La courbe verte, correspondant à l'année 2020, présente deux pics anormaux : l'un en avril lié à la première vague de la covid-19, et l'autre en octobre, correspondant à la seconde vague.
- La courbe rouge représentant 2003 montre un pic brutal au mois d'août qui est dû à la canicule exceptionnelle de cette année.
- La courbe jaune, quant à elle, indique une sous-mortalité apparente, liée à un défaut de données dans la base de l'INSEE en 1997.

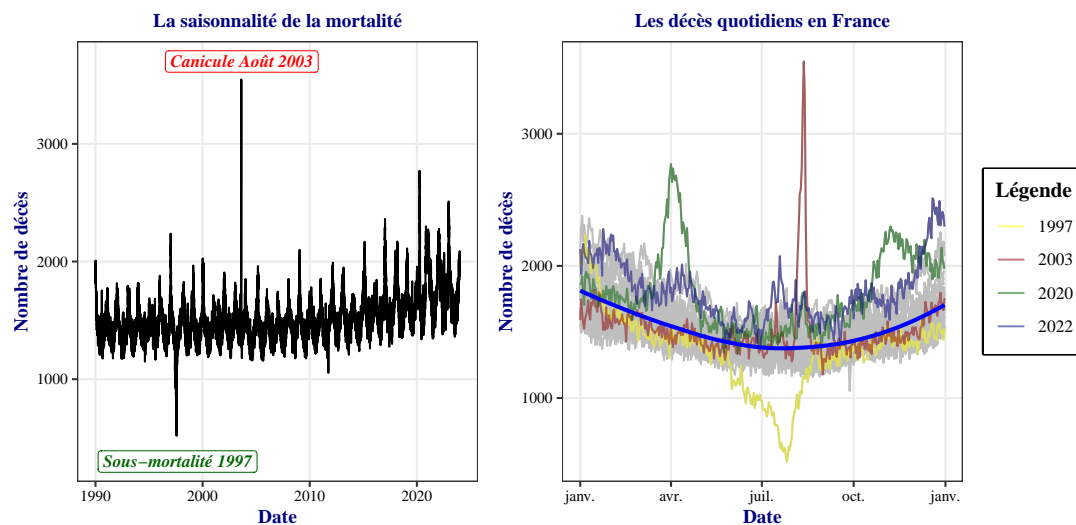


Figure 2.15. : La saisonnalité des décès

En outre, l'analyse met en évidence que la canicule d'août 2003 reste l'événement le plus meurtrier sur la période 1990-2022 en termes d'intensité quotidienne de décès. En effet, le nombre de décès par jour a plus que doublé, atteignant près de 3500 morts par jour, contre environ 1500 en période normale.

Le caractère saisonnier de la mortalité, similaire à celui de la température, suggère l'existence d'un lien structurel entre ces deux phénomènes. Mais quelle est la nature exacte de cette relation ?

L'impact de la température sur la mortalité est particulièrement visible lors des épisodes caniculaires, qui se produisent en été. Pourtant, paradoxalement, l'été est globalement la saison où la mortalité est la plus basse. Ce contraste a déjà été souligné par Santé publique France (2004), qui rapporte que la relation entre température et mortalité suit une forme en U, avec une pente plus marquée pour les températures élevées. Toutefois, en valeur absolue, ce sont les périodes de froid qui engendrent le plus grand nombre de décès.

Cette observation s'explique vraisemblablement par la fréquence déséquilibrée des températures chaudes et froides au cours de l'année : les journées froides sont beaucoup plus nombreuses que les journées de fortes chaleurs. Autrement dit, si les chaleurs extrêmes tuent plus brutalement, le froid, plus fréquent, provoque une surmortalité moins brutale mais plus importante à l'échelle annuelle.

Cette relation température-mortalité et les facteurs qui la modulent seront étudiés en détail dans la section dédiée. On pourra notamment vérifier cette hypothèse en observant l'exposition aux différentes températures.

Age au décès et nombre annuel de décès

Il est maintenant opportun de s'intéresser à l'évolution de la survenue des décès au cours de la période d'étude. Le nombre annuel de décès observé constitue le premier indicateur clé à cet effet : il montre une tendance globalement croissante mais à un rythme non uniforme. Deux grandes phases se distinguent (Figure 2.16) :

- Entre 1990 et 2005 : le nombre de décès est relativement stable avec environ 270 000 décès d'hommes et 258 000 décès de femmes en moyenne par an.
- À partir de 2005 : une augmentation marquée est observée, notamment au sein de la population féminine, dont le nombre de décès rejoint progressivement celui des hommes. En 2022, le nombre annuel de décès atteint une moyenne d'environ 320 000 pour chaque sexe. Toutefois, cette dynamique croissante semble amorcer une phase de ralentissement sur les dernières années.

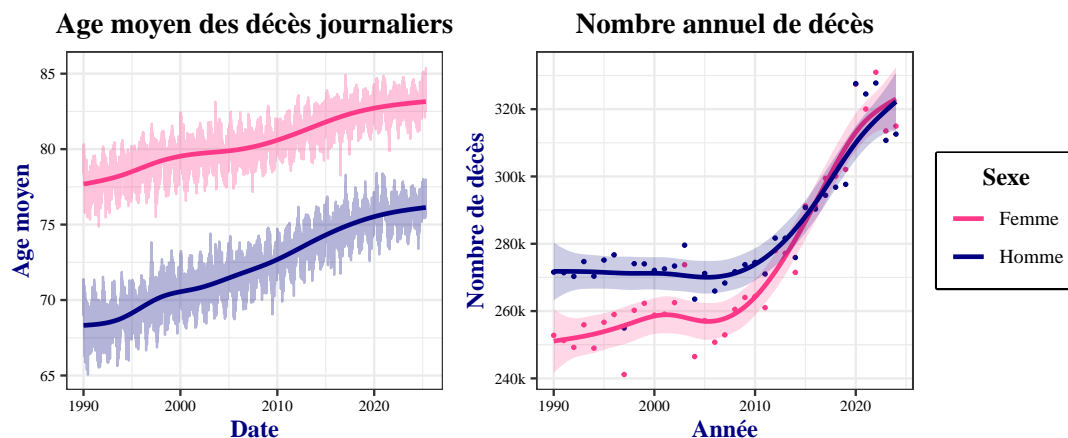


Figure 2.16. : Age moyen et nombre annuel de décès

En ce qui concerne le profil des personnes décédées, l'analyse de l'âge moyen au décès révèle une tendance à la hausse et un rythme régulier. Cela reflète le fait que les décès surviennent à un âge de plus en plus avancé. Cette observation s'accorde avec les phénomènes suivants :

- Le vieillissement de la population, qui compte une proportion croissante de personnes âgées par rapport au passé;
- L'amélioration des conditions de vie et des progrès médicaux, qui favorisent l'allongement de la durée de vie.

Par ailleurs, cette analyse confirme un fait bien établi : les femmes décèdent en moyenne à un âge plus grand que les hommes. On observe un écart d'environ une décennie entre

leurs âges moyens au décès. Ce qui soutient également que l'espérance de vie est plus élevée pour les femmes.

Il convient également de souligner la saisonnalité de l'âge moyen au décès, en lien avec la saisonnalité dans les effectifs de décès. L'hiver, période caractérisée par une surmortalité, est aussi la période où l'âge moyen au décès est le plus élevé : 81,3 ans pour les femmes et 73,8 ans pour les hommes. À l'inverse, en été où la mortalité est minimale, l'âge moyen est légèrement plus bas, avec 80,1 ans pour les femmes et 71,6 ans pour les hommes. Cela suggère que la surmortalité hivernale affecte particulièrement les personnes âgées, plus vulnérables aux conditions extrêmes.

Pour approfondir cette observation, il est pertinent d'examiner la saisonnalité des décès en fonction de l'âge. Le graphique présenté en Figure 2.17 apporte un éclairage complémentaire à ce sujet. On y observe que la saisonnalité de la mortalité est principalement portée par les personnes âgées. En revanche, elle est quasiment absente chez les individus de moins de 30 ans, voire présente une tendance inverse avec une surmortalité estivale. Quant à la saisonnalité dans les populations de 30 à 60 ans, elle reste relativement modérée et de faible ampleur.

Ce phénomène s'explique en partie par la structure même des effectifs de décès : comme la majorité des décès concernent les personnes âgées, leur profil saisonnier domine naturellement la dynamique observée au niveau global.

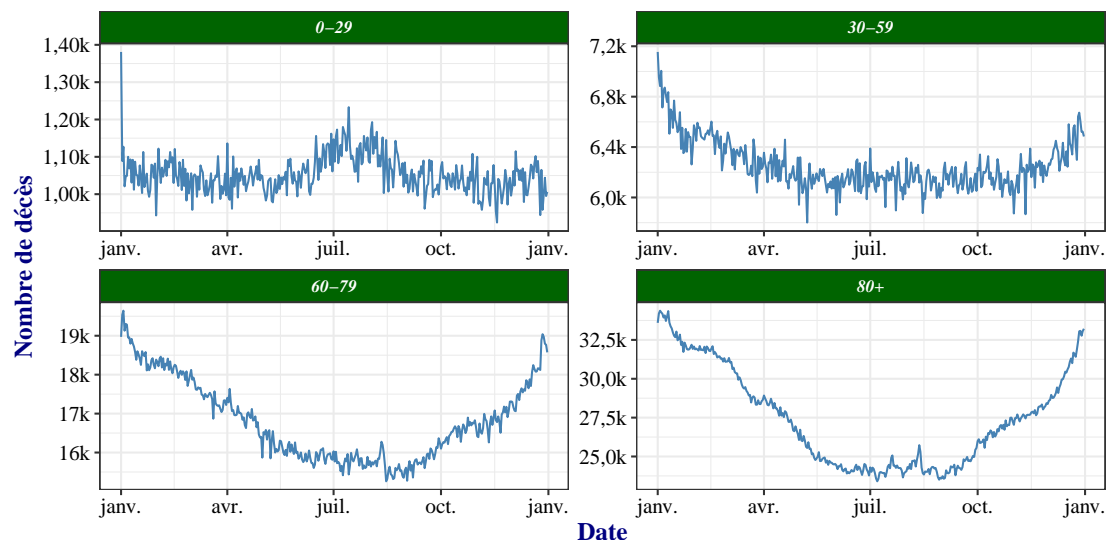


Figure 2.17. : Profil annuel des décès par groupe d'âge

2.2.3. Les évolutions de population en France

La population, qui définit l'exposition au risque de décès, constitue un facteur fondamental dans l'analyse de la mortalité. En effet, elle représente l'ensemble des individus susceptibles de décéder : toutes choses égales par ailleurs, une exposition plus importante implique mécaniquement un nombre de décès attendu plus élevé. Ainsi, les évolutions démographiques doivent impérativement être prises en compte pour interpréter correctement les dynamiques de mortalité.

Tendance de la population française

En France, la population, ici assimilée à l'exposition au décès, suit une trajectoire globalement croissante et proche d'une évolution linéaire. Sa répartition selon le sexe révèle une prédominance des femmes, avec un ratio moyen de 93 hommes pour 100 femmes sur l'ensemble de la période étudiée. En 1990, la population française comptait environ 27 millions de femmes et 26 millions d'hommes. En 2022, ces effectifs sont passés à environ 34 millions de femmes et 32 millions d'hommes (Figure 2.18).

Cette croissance progressive de la population contribue logiquement à expliquer l'augmentation du nombre annuel de décès observée sur la même période. Toutefois, il est important de noter que le rythme de croissance des décès semble plus marqué que celui de la population. Cet écart peut notamment s'expliquer par l'impact cumulé de facteurs exogènes récents, tels que des crises sanitaires ou climatiques, qui accentuent ponctuellement la mortalité.

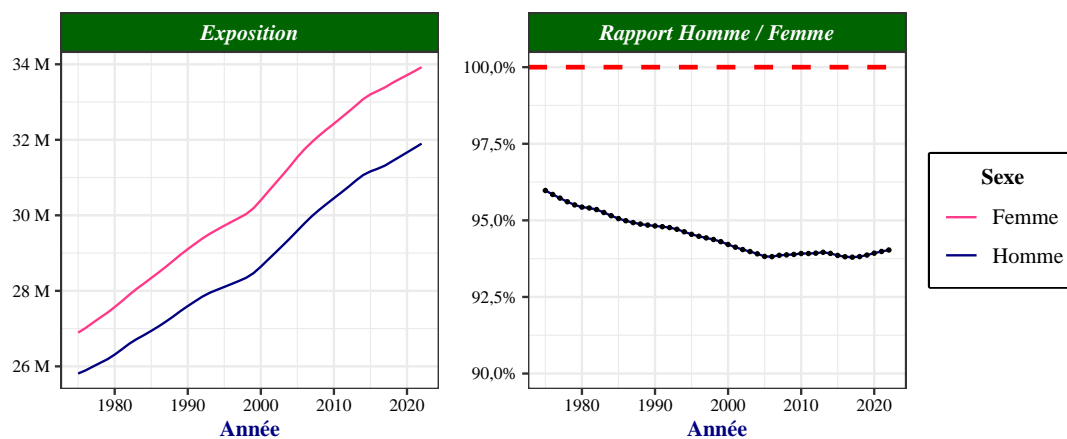


Figure 2.18. : Evolution de l'exposition

Structure par âge de la population

La structure par âge de la population française a connu une évolution significative entre 1990 et 2022. En 1990, la population était majoritairement jeune avec une proportion élevée de personnes de moins de 45 ans et un déclin progressif des effectifs aux âges plus élevés. Certains creux dans la pyramide des âges reflètent des événements démographiques marquants, comme des guerres ou des crises sanitaires ayant affecté la taille de certaines générations.

En 2022, la population présente une structure sensiblement différente, caractérisée par un vieillissement général. La répartition est maintenant plus uniforme sur plage d'âge allant jusqu'à 75 ans, traduisant une présence plus équilibrée des différentes générations. Ce n'est qu'au delà de cet âge qu'une décroissance progressive des effectifs est observée. Les personnes centenaires sont présentes, mais ne représentent qu'une fraction marginale de la population totale (Figure 2.19).

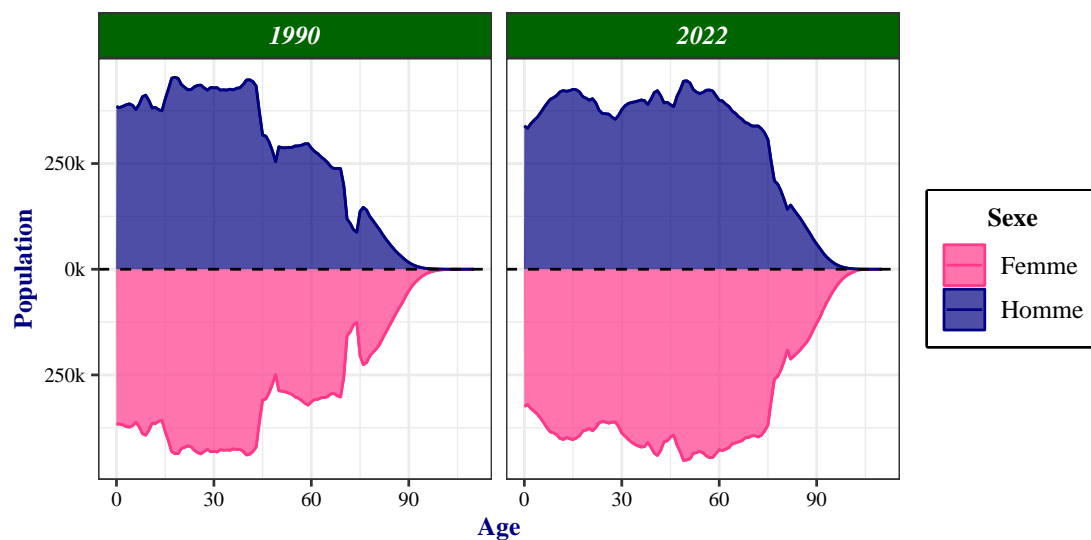


Figure 2.19. : Structure par âge de la population

2.2.4. Analyse globale de la mortalité

L'exposition par âge et structure des décès

Sur l'ensemble de la période d'étude, l'analyse de l'exposition au risque de décès selon les âges révèle une distribution caractéristique. La population exposée est importante aux jeunes âges et chez les adultes actifs, avec un accroissement progressif jusqu'à environ 45 ans, puis une décroissance continue par la suite. En revanche, la distribution du nombre de décès présente un profil bien différent. On observe un premier pic

de mortalité à la naissance (âge 0), suivi d'une croissance quasi exponentielle avec l'âge jusqu'à un maximum vers 87 ans. Cette évolution est ensuite suivie d'une décroissance, probablement influencée par une exposition réduite aux âges les plus avancés (Figure 2.20).

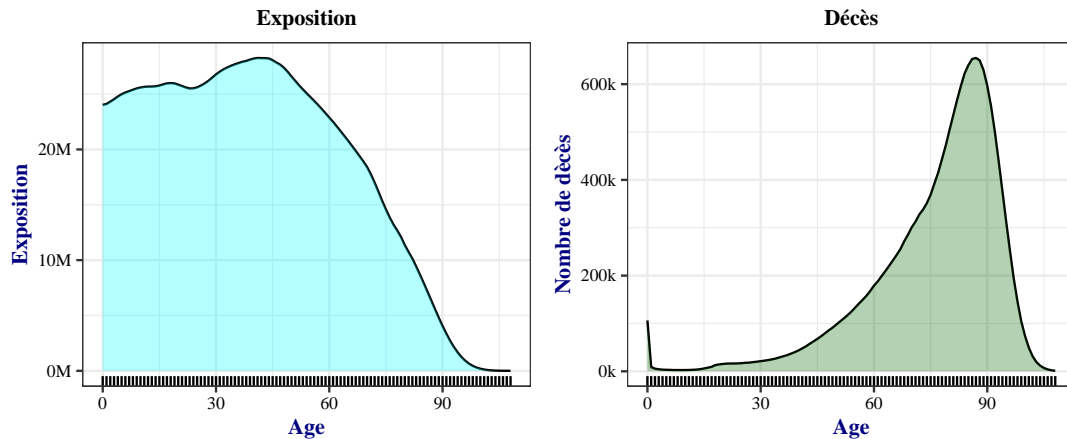


Figure 2.20. : Nombre de décès et exposition par âge

Exposition par température

L'exposition de la population aux températures est également instructive (Figure 2.21). La température considérée ici est une moyenne sur les 10 derniers jours, l'exposition associée présente une forme bimodale, avec deux pics notables : l'un autour de 7 à 8 °C, et l'autre autour de 17 à 18 °C. Cette répartition reflète bien la saisonnalité des températures déjà décrite à la figure Figure 2.14, avec une moyenne hivernale proche de 5 °C et une moyenne estivale avoisinant de 18 °C.

La forme de cette distribution est relativement stable, même lorsqu'on varie la profondeur de la période de cumul des températures. Par ailleurs, elle étaye l'hypothèse avancée précédemment par rapport aux effets thermiques : la mortalité attribuable au froid est plus importante en valeur absolue non pas en raison d'un effet plus intense que celui de la chaleur, mais en raison d'une exposition beaucoup plus fréquente aux températures froides. En effet, les températures supérieures à 18 °C sur 10 jours ne représentent que 17,1 % de l'exposition totale, et celles au-delà de 22 °C ne comptent que pour 3,3 %. Ces chiffres illustrent clairement la rareté relative des températures élevées, et donc une exposition moindre à la chaleur, malgré ses effets sanitaires potentiellement plus prononcés.

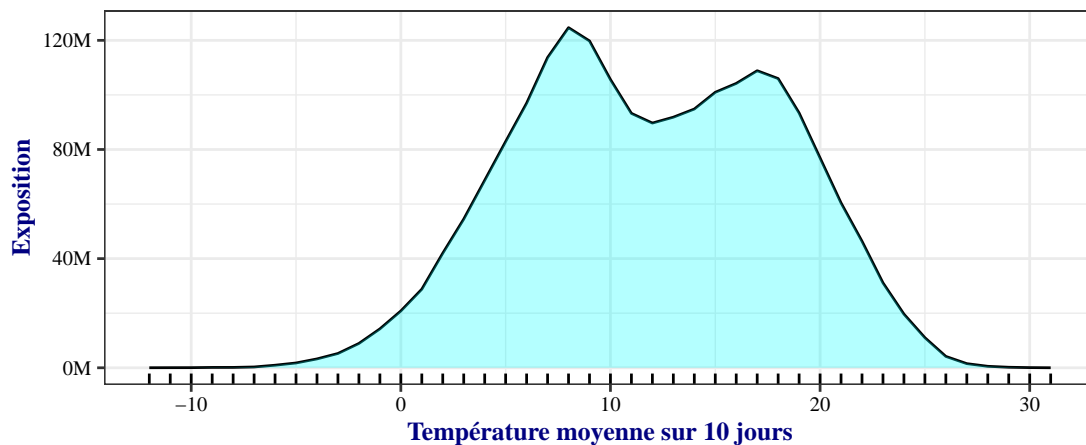


Figure 2.21. : Exposition par température

Analyse des taux bruts de mortalité

L'analyse des taux de mortalité permet de mieux comprendre la dynamique du phénomène et de caractériser les principaux déterminants. Il en ressort que l'âge demeure un facteur central. Sur l'ensemble de la période, les écarts entre les jeunes et les personnes âgées sont particulièrement marqués : à 10 ans, le taux de mortalité moyen est d'environ 0,01 %, contre 4,5 % à 60 ans, et plus de 25 % à 95 ans. Cette forte croissance avec l'âge reflète la vulnérabilité accrue des populations âgées.

Par ailleurs, à âge égal, les hommes présentent en moyenne des taux de mortalité supérieurs à ceux des femmes, ce qui confirme un effet du sexe bien répandu dans la littérature sur la mortalité. À cela s'ajoute une tendance générale à la baisse des taux de mortalité au cours du temps, qui s'observe pour la majorité des groupes d'âge. Cette évolution témoigne des progrès en matière de conditions de vie, de prévention, et de soins médicaux. Ces observations justifient pleinement la prise en compte des facteurs âge, sexe et période calendaire dans les modèles de mortalité.

Enfin, on peut noter une variabilité importante des taux de mortalité observé à 10 ans par rapport aux autres tranches d'âge. Cette fluctuation s'explique principalement par le faible nombre de décès enregistrés à cet âge, ce qui rend les taux plus sensibles aux variations annuelles. De façon générale, la majorité des décès concerne des personnes âgées, ce qui souligne davantage que la mortalité est un phénomène fortement structuré par l'âge.

Ces différentes caractéristiques de la mortalité sont illustrées sur la figure Figure 2.22.

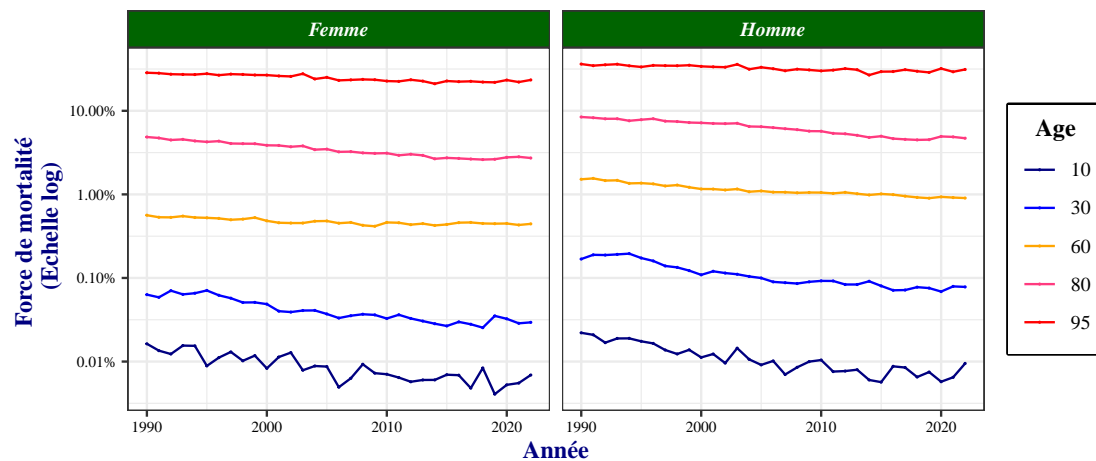


Figure 2.22. : Force de mortalité par âge

Synthèse du chapitre

La qualité des données est un aspect essentiel qui influence directement la pertinence des résultats de toute analyse statistique. À ce titre, ce chapitre a été consacré à la présentation des sources de données relatives aux températures moyennes quotidiennes, aux nombres de décès quotidiens et à l'exposition au risque, fournies par Météo France, l'INSEE et la base HMD. Ces données ont présenté certains défauts, tels que des valeurs manquantes dans les stations météorologiques et des doublons dans les données de décès, qui ont été identifiés et corrigés afin de construire une base de données consolidée contenant toutes les variables utiles à l'analyse.

Une analyse exploratoire a ensuite été faite et a permis de mettre en lumière les principaux facteurs influençant la mortalité. L'examen de la structure par âge des décès et des populations, ainsi que de leur évolution temporelle, a apporté des premiers éléments d'interprétation. Cependant, cette approche reste sujette à des biais d'échantillonnage et à des irrégularités structurelles. Afin d'affiner cette analyse et de mieux quantifier les effets observés, nous aurons recours à des modèles statistiques de mortalité. À cet effet, le chapitre suivant présentera la démarche méthodologique et les modèles utilisés dans cette étude.

Chapter 3.

Notions théoriques et approche de modélisation

Ce chapitre présente la démarche méthodologique suivie pour la construction du modèle de mortalité intégrant les effets de température. Dans un premier temps, une introduction des concepts de base en modélisation de la mortalité est présentée, ensuite les modèles classiques de mortalité prospective, principalement les modèles AP, APC et de Lee-Carter sont présentés. Enfin, l’approche des modèles additifs généralisés (GAM) est également présentée ainsi que les principaux indicateurs de performances des modèles.

3.1. Cadre méthodologique de l’étude

La première partie de cette section s’appuie principalement sur les concepts présentés dans les cours de modèles de durée dispensés par Biessy (2025) et Vermet (2025).

3.1.1. Les concepts de base en analyse de survie

L’analyse de la mortalité d’un individu est généralement abordée à l’aide d’un processus de vie et de mort noté Z_t . Celui-ci prend la valeur 0 si l’individu est vivant à l’instant t , et 1 dans le cas contraire. L’état de l’individu est ainsi décrit par le processus Z_t , ainsi que par le processus X_t représentant l’ensemble de ses caractéristiques à l’instant t susceptibles d’affecter la mortalité (âge, sexe, niveau de revenu, etc).

En actuariat et particulièrement dans le cadre de ce mémoire, le phénomène d’intérêt est la durée au bout de laquelle une personne passera de l’état “**vivant**” à l’état “**décédé**”. Pour cela, on définit la variable aléatoire T_χ qui désigne la durée de vie résiduelle conditionnellement aux caractéristiques χ . Formellement, T_χ correspond au premier instant t auquel le processus Z_t prend la valeur 1 :

$$T_\chi(s) = \min\{u > 0, Z_{s+u} = 1 \mid Z_s = 0, X_s = \chi\}$$

L'objectif visé est alors de décrire et de caractériser la durée de vie résiduelle T_χ . Plusieurs notions théoriques sont mobilisées à cet effet.

La fonction de survie

En tant que variable aléatoire, la durée de vie résiduelle T_χ est caractérisée par sa fonction de répartition $F_\chi(t) = \mathbb{P}(T_\chi \leq t)$. La fonction de survie correspond à son complémentaire $S_\chi(t) = 1 - F_\chi(t) = \mathbb{P}(T_\chi > t)$. Il s'agit de la probabilité que l'individu soit toujours vivant après une durée supplémentaire t , conditionnellement à ses caractéristiques χ .

La densité de probabilité

La densité de probabilité associée à T_χ est définie par la relation :

$$f_\chi(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(t \leq T_\chi < t + h) = -\frac{d}{dt} S_\chi(t)$$

La force de mortalité

La force de mortalité, encore appelée fonction de hazard ou taux instantané de décès, est définie par la relation suivante :

$$\mu_\chi = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(T_\chi < h) = f_\chi(0)$$

Elle mesure l'intensité du risque de mortalité à l'instant t , conditionnellement aux caractéristiques χ , et représente la probabilité de décéder dans un court intervalle de temps, rapportée à la longueur de cet intervalle.

La force de mortalité est liée aux notions précédentes par les relations suivantes :

$$\mu_\chi(t) = \frac{f_\chi(t)}{S_\chi(t)} = -\frac{S'_\chi(t)}{S_\chi(t)} = -\frac{d}{dt} \log(S_\chi(t))$$

Ainsi, elle caractérise entièrement le phénomène aléatoire T_χ . La fonction de survie conditionnelle peut se déduire de la force de mortalité grâce à la relation :

$$S_\chi(t) = \exp\left(-\int_0^t \mu_\chi(s) ds\right)$$

3.1.2. Les notations actuarielles

Dans la pratique actuarielle, plutôt que d'utiliser la fonction de survie, le plus souvent, les nombres de survivants notés L_x sont préférés. Ceux-ci s'interprètent dans le cadre d'une cohorte fictive d'effectif initial L_{x_0} (généralement fixé à 100 000), dont l'évolution reproduit celle de la fonction de survie. On a ainsi la relation :

$$L_{x_0+t} = L_{x_0} S_{x_0}(t) \quad \forall t \in \mathbb{N}.$$

Ces grandeurs qui sont généralement reportées dans les tables de mortalité.

De la même façon, on définit les probabilités conditionnelles de survie pendant t périodes et de décès avant t périodes : ${}_t p_x = \mathbb{P}(T_x > t) = S_x(t)$ et ${}_t q_x = 1 - {}_t p_x = \mathbb{P}(T_x < t)$.

Dans le cas particulier où $t = 1$ correspondant au cas de décès en cours d'année, l'indice temporel est généralement omis et on note simplement p_x et q_x . On dispose de relations directes avec les effectifs de survivants :

$$p_x = \frac{L_{x+1}}{L_x}, \quad q_x = \frac{L_x - L_{x+1}}{L_x}$$

La probabilité conditionnelle de décès dans l'année q_x est liée à la force de mortalité par l'expression :

$$q_x = 1 - \exp\left(-\int_0^1 \mu_x(t) dt\right)$$

Sous l'hypothèse courante d'une force de mortalité constante entre deux âges entiers (ou au cours de l'année), cette relation se simplifie en $q_x = 1 - \exp(-\mu_x)$.

Enfin, lorsque μ_x est faible, on a l'approximation $q_x \approx \mu_x$. Cette approximation peut être valable aux âges jeunes où la mortalité est faible. En revanche, aux âges élevés, la force de mortalité devient plus importante et l'approximation est imprécise.

3.2. Les modèles prospectifs de mortalité

Les modèles prospectifs de mortalité visent à estimer l'évolution future de la mortalité en extrapolant les tendances passées. De nombreux modèles ont été développés dans cette optique.

De manière générale, la mortalité est classiquement décrite comme une fonction de trois principaux facteurs :

- l'âge x ,
- la période calendaire y ,
- et la cohorte $c = y - x$.

On se place ainsi dans un cadre où ces caractéristiques sont supposées déterminer conjointement le risque de mortalité. Formellement, on pose : $\chi = \{x, y, c\}$

3.2.1. Le modèle Age-Période (AP)

Le modèle Age-Période (AP) constitue un point de départ classique pour l'analyse de la mortalité. Il décompose le logarithme de la force de mortalité comme la somme d'un effet de l'âge et d'un effet de période auquel s'ajoute un résidu aléatoire :

$$\log(\mu_{xy}) = \alpha_x + \kappa_y + \epsilon_{xy}$$

où

Chaque valeur d'âge x et de période y correspond à un paramètre à estimer. Ce modèle souffre d'un problème d'identifiabilité entraînant une infinité de possibilités dans le choix des paramètres. Concrètement si le couple (α_x, κ_y) est solution alors pour tout réel m le couple $(\alpha_x + m, \kappa_y - m)$ est également solution et correspond exactement au même modèle.

Pour garantir l'unicité des estimateurs, une contrainte d'identifiabilité doit être imposée. Celle généralement retenue est : $\sum \kappa_y = 0$ ce qui force les effets périodiques à avoir une moyenne nulle. Plusieurs approches d'estimations peuvent être suivies :

La méthode des moindres carrés ordinaires (MCO)

L'estimation des paramètres vise à minimiser les écarts entre les valeurs observées et celles prédites par le modèle. A cet effet, le critère des moindres carrés ordinaires consiste à minimiser la fonction F tel que :

$$F = \sum_{x,y} (\log(\mu_{xy}) - (\alpha_x + \kappa_y))^2$$

Soit le choix des estimateurs $(\hat{\alpha}_x, \hat{\beta}_y) = \arg \min(F)$.

Les paramètres d'âge s'obtiennent simplement en exploitant la relation :

$$\frac{\partial F}{\partial \alpha_x} = \sum_t -2(\log(\mu_{xy}) - \alpha_x - \kappa_y) = 0 \mid_{\alpha_x = \hat{\alpha}_x}$$

Et en tenant compte de la contrainte d'identifiabilité $\sum \kappa_y = 0$ on obtient les estimateurs suivants :

$$\hat{\alpha}_x = \frac{1}{y_{\max} - y_{\min} + 1} \sum_{y=y_{\min}}^{y_{\max}} \log(\mu_{xy}) \quad \text{et} \quad \hat{\kappa}_y = \frac{1}{x_{\max} - x_{\min} + 1} \sum_{x=x_{\min}}^{x_{\max}} (\log(\mu_{xy}) - \hat{\alpha}_x) \quad (3.1)$$

Ainsi, $\hat{\alpha}_x$ correspond à la moyenne des logarithmes des forces de mortalité à l'âge x sur toutes les périodes, tandis que $\hat{\kappa}_y$ mesure l'écart moyen à la mortalité prédite par les seuls effets de l'âge à la période y .

De façon plus commode, les paramètres α_x et κ_y peuvent être déterminés par une approche matricielle en définissant les matrices X_α et X_κ (Fine et Fotso, 1989) :

$$X_\alpha = I_{n_x} \otimes \mathbb{1}_{n_y} = \begin{bmatrix} \mathbb{1}_{n_y} & 0 & \cdots & 0 \\ 0 & \mathbb{1}_{n_y} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbb{1}_{n_y} \end{bmatrix} \in \mathbb{R}^{n_x n_y \times n_x}$$

$$X_\kappa = \mathbb{1}_{n_x} \otimes I_{n_y} = \begin{bmatrix} I_{n_y} \\ I_{n_y} \\ \vdots \\ I_{n_y} \end{bmatrix} \in \mathbb{R}^{n_x n_y \times m}$$

où : n_x et n_y sont respectivement les nombres d'âges et de périodes, I_n est la matrice unitaire de dimension n et $\mathbb{1}_n$ est le vecteur colonne de taille n constitué uniquement de 1.

On pose par la suite les vecteurs et matrices du modèle de la façon suivante :

$$Y = [\log(\mu_{xy})], \quad X = [X_\alpha X_\kappa], \quad \alpha = \begin{bmatrix} \alpha_{x_{\min}} \\ \vdots \\ \alpha_{x_{\max}} \end{bmatrix}, \quad \kappa = \begin{bmatrix} \kappa_{y_{\min}} \\ \vdots \\ \kappa_{y_{\max}} \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha \\ \kappa \end{bmatrix}$$

Le modèle s'écrit alors sous la forme $Y = X\beta + \epsilon$ où ϵ est le vecteur des résidus.

Pour être solvable, il faut prendre en compte les contraintes d'identifiabilité. Pour cela, on définit la matrice C des contraintes linéaires sur les paramètres β dont un ensemble de combinaisons linéaires sont fixées aux valeurs du vecteur d , ce qui revient à $C\beta = d$.

Dans le cas du modèle AP avec la contrainte $\sum_y \kappa_y = 0$, on a les valeurs correspondantes de C et d qui sont les suivantes :

$$C = [0 \quad \dots \quad 0 \mid 1 \quad \dots \quad 1] \quad \text{et} \quad d = [0]$$

Il faut donc résoudre le problème $\min \|Y - X\beta\|_2$ sous la contrainte $C\beta = d$. Ce qui conduit à la solution explicite :

$$\hat{\beta} = (X^\top X + C^\top C)^{-1}(X^\top Y + C^\top d)$$

Si $d = 0$ comme c'est le cas ici, cela devient :

$$\hat{\beta} = (X^\top X + C^\top C)^{-1} X^\top Y$$

Bien que les MCO offrent une expression explicite des paramètres et facilite leur interprétation, elle présente plusieurs limites importantes notamment dans le contexte de la modélisation de la mortalité.

Premièrement, elle suppose l'homoscédasticité des résidus qui est souvent remise en question dans la littérature (Brouhns, Denuit et Vermunt, 2002). Elle est notamment peu réaliste dans l'étude des décès où le nombre de décès observés est naturellement plus important aux âges élevés par rapport aux jeunes âges. La variance du logarithme de la force de mortalité tend par conséquent à augmenter avec l'âge ce qui introduit une hétéroscédasticité structurelle qui viole les conditions classiques de validité des estimateurs MCO.

Deuxièmement, cette approche suppose la disponibilité d'estimations valides de la force de mortalité à chaque âge. Or, dans les groupes à faible mortalité, comme les jeunes ou dans des sous-populations spécifiques, le nombre de décès observés peut être nul, rendant l'estimation du logarithme de la force de mortalité, $\log(\hat{\mu}_x)$, non définie. Cela limite l'application directe de la méthode MCO dans ces contextes.

La méthode du maximum de vraisemblance

Afin de surmonter ces limitations, l'approche d'estimation par maximum de vraisemblance est plus robuste. Elle repose sur une modélisation probabiliste explicite du phénomène de mortalité. En particulier, dans un cadre discret, le nombre de décès observés est modélisé comme une variable aléatoire issue d'une loi de Poisson ou d'une loi binomiale.

On suppose que le nombre de décès pour un profil donné suit un processus de comptage, avec une intensité proportionnelle à la force de mortalité et à l'effectif de population exposée :

$$D_{xy} \sim \text{Poisson}(\mu_{xy} \cdot E_{xy})$$

La vraisemblance associée s'écrit :

$$L(\mu) = \prod_{x,y} \frac{(\mu_{xy} E_{xy})^{D_{xy}} e^{-\mu_{xy} E_{xy}}}{D_{xy}!}$$

et la log vraisemblance est :

$$\log L(\mu) = \sum_{x,y} D_{xy} \log(\mu_{xy}) - \mu_{xy} E_{xy} + cte$$

Qui s'écrit sous forme matricielle :

$$\log L(\beta) = D^T X \beta - E^T \exp(X \beta) + cte$$

avec $\mathbb{E}[Y] = \log(\mu_{xy}) = X\beta$, $D^T = [D_{xy}]$ et $E^T = [E_{xy}]$.

Les paramètres optimaux sont obtenus en maximisant la log-vraisemblance. En pratique ceci nécessite l'application d'une procédure numérique itérative comme l'algorithme de Newton-Raphson.

Le modèle Age-Période a l'avantage d'être simple et facilement interprétable. De plus, il s'agit d'un modèle linéaire. Son principal inconvénient est son faible pouvoir prédictif qui fait qu'en pratique on lui préfère des modèles plus développés.

3.2.2. Le modèle de Lee-Carter

Le modèle de Lee-Carter est un modèle classique et célèbre en modélisation de mortalité. Il englobe le modèle Age-Période en à travers les paramètres β_x qui modulent l'effet de période selon l'âge, ce qui permet une différenciation des améliorations de mortalité en fonction des âges. La spécification du modèle de Lee-Carter est la suivante :

$$\log(\mu_{xy}) = \alpha_x + \beta_x \kappa_y + \epsilon_{xy}$$

Ce modèle est aussi sujet à des problèmes d'identifiabilité. En effet, quels que soient c_1 et $c_2 \neq 0$, les solutions $(\alpha_x, \beta_x, \kappa_y)$, $(\alpha_x + c_1 \beta_x, \beta_x, \kappa_y - c_1)$ et $(\alpha_x, \frac{1}{c_2} \beta_x, c_2 \kappa_y)$ correspondent exactement au même modèle ce qui viole l'unicité des solutions.

Pour assurer l'identifiabilité, les contraintes habituellement spécifiées sont les suivantes :

$$\sum_x \beta_x = 1 \quad \text{et} \quad \sum_y \kappa_y = 0$$

A partir d'une solution quelconque, la solution unique respectant ces contraintes peut être obtenue en faisant les choix $c_1 = \frac{1}{y_{max} - y_{min} + 1} \sum_y \kappa_y$ puis $c_2 = \sum_x \beta_x$.

Dans la version originale du modèle de Lee-Carter, l'estimation des paramètres est faite par la méthode MCO conduisant à la même expression des paramètres α_x que celle obtenu pour le modèle Age-Période. Les paramètres β_x et κ_y sont obtenus après une procédure de décomposition en valeur singulière (SVD) de la matrice $Z = \log(\mu_{xy}) - \hat{\alpha}_x$. Dans une seconde étape, les paramètres \hat{k}_t obtenus sont révisés pour égaliser les nombres annuels de décès observés. Une dernière étape est faite afin de tenir compte des contraintes d'identification qui ont été précisées.

Pour simplifier cette méthode et face à la limite de l'hypothèses d'homoscédasticité des résidus, Brouhns, Denuit et Vermunt (2002) ont proposé l'approche par maximum de vraisemblance en considérant que les décès sont issus d'un processus de comptage.

En pratique il n'existe pas de solution explicite pour tous les paramètres et les estimations sont obtenus par des méthodes numériques.

3.2.3. Le modèle Age-Période-Cohorte (APC)

Le modèle APC décrit la mortalité comme somme d'une composante de l'âge, de la période et de la cohorte. Il s'écrit comme suit :

$$\log(\mu_{xy}) = \alpha_x + \kappa_y + \gamma_{y-x} + \epsilon_{xy}$$

Par rapport au modèle de Renshaw-Haberman qui le généralise, par sa spécification $\log(\mu_{xy}) = \alpha_x + \beta_x^{(1)} \kappa_y + \beta_x^{(2)} \gamma_{y-x} + \epsilon_{xy}$, le modèle APC est plus stable dans l'estimation des paramètres.

De plus, le modèle APC a l'avantage d'être linéaire donc plus simple à estimer. Il présente lui aussi des problèmes d'identifiabilité. En effet, quels que soient ϕ_1 , ϕ_2 et c_1 , et pour une solution $(\alpha_x, \kappa_y, \lambda_{y-x})$ du modèle, les transformations suivantes le laissent invariant ce qui correspond à une infinité de solutions possibles :

$$(\alpha_x, \kappa_y, \lambda_{y-x}) \rightarrow (\alpha_x + \phi_1 - \phi_2 x, \kappa_y + \phi_2 y, \lambda_{y-x} - \phi_1 - \phi_2 (y - x))$$

$$(\alpha_x, \kappa_y, \lambda_{y-x}) \rightarrow (\alpha_x + c_1, \kappa_y - c_1, \lambda_{y-x})$$

Face à ces problèmes, les contraintes généralement imposées aux paramètres pour assurer l'unicité de la solution sont en les suivantes :

$$\sum_{y=y_{\min}}^{y_{\max}} \kappa_y = 0, \quad \sum_{c=y_{\min}-x_{\max}}^{y_{\max}-x_{\min}} \lambda_c = 0, \quad \sum_{c=y_{\min}-x_{\max}}^{y_{\max}-x_{\min}} c\lambda_c = 0$$

Comme dans les modèles précédents, la première contrainte assure que l'effet période moyen soit nul. Les deux dernières contraintes imposent que l'effet générationnel fluctue autour de 0 sans aucune tendance particulière.

Pour imposer ces contraintes à partir d'une solution quelconque $(\alpha_x, \kappa_y, \lambda_{y-x})$, on peut appliquer les transformations précédentes en deux étapes :

- Choix des valeurs des constantes ϕ_1 et ϕ_2 par régression de λ_{t-x} sur $y-x$ tel que $\lambda_{y-x} = \phi_1 + \phi_2(y-x) + \epsilon_{y-x}$ avec $\epsilon_{t-x} \sim \mathcal{N}(0, \sigma^2)$ i.i.d.
- Choix de la constante c_1 tel que $c_1 = \frac{1}{y_{\max}-y_{\min}+1} \sum_{y=y_{\min}}^{y_{\max}} \kappa_y$.

L'estimation des paramètres peut être faite par l'approche des MCO ou par celle du maximum de vraisemblance de la même façon que pour le modèle Ape-Période.

3.2.4. Les modèles Age-Période-Cohorte généralisés (GAPC)

Principe général

Les modèles GAPC constituent une approche globale pour représenter toute une gamme de modèle faisant intervenir les facteurs âge, période et cohorte. De façon similaire aux GLM, ils s'appuient sur plusieurs éléments :

- Une distribution de probabilité : Le nombre de décès D_{xy} est issu d'une loi de poisson ou d'une loi binomiale avec $\mathbb{E}[D_{xy}] = E_{xy}\mu_{xy}$ et $\mathbb{E}[D_{xy}] = E_{xy}q_{xy}$ respectivement.
- Une composante systématique : les effets de l'âge, de l'année calendaire et de l'année de naissance sont captés par η_{xy} par la relation :

$$\eta_{xy} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_y^{(i)} + \beta_x^{(0)} \lambda_{y-x}$$

où N est un entier positif représentant le nombre de termes décrivant la relation Age-Période.

- Une fonction de lien : c'est une fonction g qui établit la liaison entre les deux termes ci-dessus de la forme $g(\mathbb{E}[\frac{D_{xy}}{E_{xy}}]) = \eta_{xy}$.

- Un ensemble de contraintes sur les paramètres : En effet, ces modèles sont également sujets à des problèmes d'identifiabilité qui doivent être traités en fixant des contraintes supplémentaires.

Les modèles Age-Période, de Lee-Carter et Age-Période-Cohorte sont des cas particuliers des modèles de la famille GAPC.

L'estimation des paramètres optimaux est faite par la méthode du maximum de vraisemblance. Une implémentation de ces modèles est disponible en R dans le package **StMoMo** (Villegas, Millosovich et Kaishev, 2018) en s'appuyant sur des approches de modèles non linéaires généralisés du package **gnm**.

La projection des paramètres.

Pour établir les projections du modèle sur des périodes futures, les composantes périodiques et générationnelles estimées par le modèle doivent être projetées dans le futur. Pour cela, la solution qui est le plus souvent utilisée et qui est implémentée dans le package **StMoMo** consiste à supposer que les paramètres estimés sont issus d'une série temporelle. À ce titre, ils sont projetés en calibrant un modèle de série temporelle avec ou sans dérive en fonction de la forme obtenue des paramètres.

3.3. Présentation des modèles additifs généralisés

Les modèles additifs généralisés constituent une approche très flexible de modélisation, permettant de représenter des relations complexes de façon lisse. Pour mieux comprendre ces modèles, il est utile de prendre préalable connaissance des modèles additifs.

3.3.1. Les modèles additifs

Dans un modèle linéaire classique, une variable réponse y est modélisée comme une fonction linéaire d'un nombre k de covariables par la relation :

$$y_i = \beta_0 + \beta_1 x_i^1 + \dots + \beta_k x_i^k + \varepsilon_i \quad (3.2)$$

où les β_i sont les paramètres du modèle et $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ est le terme d'erreur aléatoire.

Un modèle additif est un modèle linéaire dans lequel les covariables x^j sont liées à la variable réponse par une combinaison linéaire de fonctions lisses de ces covariables. De façon générale, un modèle additif s'écrit sous la forme :

$$y_i = X_i \theta + f_1(x_i^1) + \dots + f_k(x_i^k) + \varepsilon_i \quad (3.3)$$

où :

- $X_i\theta$ représente la partie paramétrique du modèle, avec X_i la ligne i de la matrice de design paramétrique et θ le vecteur des paramètres associés.
- Les fonctions f_j , $j = 1, \dots, k$ sont des fonctions lisses des covariables correspondantes
- ε_i est le terme d'erreur aléatoire suivant une distribution normale.

Ce modèle, lorsqu'il intègre plusieurs fonctions lisses pose présente un problème d'identifiabilité similaire à celui présenté dans les modèles GAPC. L'ajout d'une constante c à la fonction f_m et la soustraction de la même valeur à la fonction f_n laisse le modèle invariant ce qui empêche l'estimation de ces fonctions. Il y'a lieu d'imposer des contraintes d'identifiabilité.

La contrainte la plus pratique consiste à imposer que toutes les fonctions f_j soient de moyenne nulle :

$$\sum_{i=1}^n f_j(x_i) = 0$$

Cette approche de modélisation est très flexible et capable de représenter toute forme de relations au delà des relations linéaires. Cette flexibilité s'accompagne d'un coût supplémentaire nécessaire à l'estimation des fonctions de covariables. D'une part, les fonctions f_j doivent être spécifiées d'une certaine façon et d'autre part, le degré de lissage chaque fonction doit être choisi.

3.3.1.1. Spécification des fonctions lisses

Pour simplifier, considérons f comme l'une des fonctions lisses de la spécification de l'Équation 3.3 et x la covariable correspondante. Pour estimer la fonction f à l'aide des méthodes usuelles, elle doit être représentée d'une manière à correspondre à un modèle linéaire. Cette représentation nécessite de définir des fonctions de bases b_j définissant un espace de fonctions auquel f doit appartenir. Le choix de k fonctions de base revient à écrire f sous la forme :

$$f(x) = \sum_{j=1}^p b_j(x)\beta_j \quad (3.4)$$

Cette représentation correspond directement à un modèle linéaire dans lequel les variables explicatives sont les $b_j(x)$.

Le choix des fonctions de base détermine la forme de la fonction f . Un exemple simple est le cas de la régression polynomiale pour lequel la fonction de base b_j est représentée

par un polynôme de degré $j-1$. Pour un polynôme cubique la valeur de p correspondante est 4 et l'Équation 3.4 devient :

$$f(x) = \beta_1 + x\beta_2 + x^2\beta_3 + x^3\beta_3 \quad (3.5)$$

En pratique, le choix des fonctions de base est plus complexe. Les bases polynômiales ne sont généralement pas adaptées. Elles permettent un bon ajustement aux points de données, mais la fonction estimée présente souvent des écarts importants par rapport à la vraie fonction, d'autant plus lorsque le degré du polynôme est élevé. Ce comportement est illustré dans par la Figure 3.1 issue de Wood (2017).

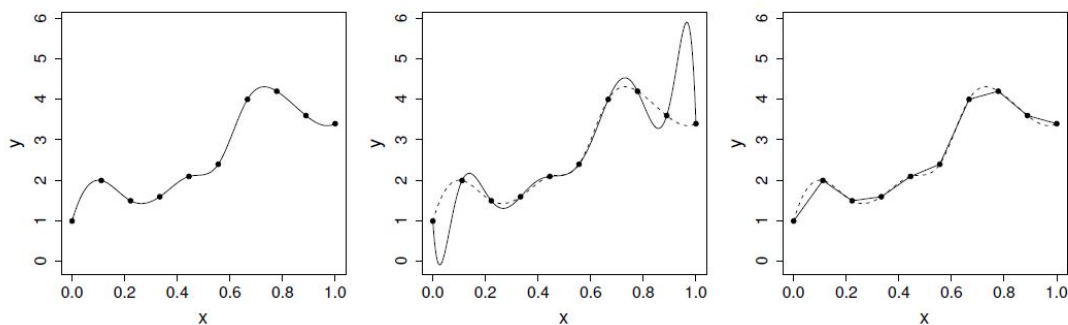


Figure 3.1. : Effets des fonctions de bases polynômiales (Source: Wood, 2017)

Ces 3 graphiques représentent respectivement : une fonction cible à modéliser, une superposition avec l'estimation par un modèle polynomial, et une superposition avec un modèle obtenu par interpolation linéaire par morceaux.

L'estimation polynomiale s'ajuste bien aux points de données mais ne réussit pas à capturer la forme de la fonction inconnue. Par contre, la solution par interpolation linéaire est plus proche de la fonction réelle mais manque de régularité ne correspondant pas aux fonctions lisses souhaitées.

En pratique, les fonctions de base les plus couramment utilisées conduisent généralement à des estimations régulières. C'est le cas, par exemple, des splines cubiques qui reposent sur des polynômes de degré trois définis par morceaux et soumis à des contraintes de continuité et de dérivabilité. C'est aussi le cas des thin plate regression splines qui permettent d'estimer une fonction régulière à partir des données sans imposer de forme particulière. Des détails plus complets sur ces bases sont disponibles dans Wood (2017).

3.3.1.2. Estimation des paramètres

Dans le cas d'une seule variable explicative, en notant Y , X et β les vecteurs des réponses, de la variable explicative et des paramètres du modèle, la fonction objectif à minimiser

s'écrit :

$$\|Y - X\beta\|^2 + \lambda \sum_{j=2}^{n-1} (f(x_{j-1}) - 2f(x_j) + f(x_{j+1}))^2 \quad (3.6)$$

où le terme classique $\|Y - X\beta\|^2$ favorise la proximité entre les observations et les valeurs prédites tandis que le terme de somme constitue une mesure des irrégularités de la fonction f , λ étant le paramètre de lissage qui pénalise ces irrégularités.

En tenant compte de la forme des fonctions de base considérée, il est possible d'écrire l'expression 3.6 sous forme matricielle de la façon suivante :

$$\|Y - X\beta\|^2 + \lambda \beta^\top S \beta$$

où S est une matrice de régularisation des paramètres.

Dans le cas général à k variables, chaque fonction f_j se décompose comme une combinaison linéaire d'un nombre k_j de fonctions de base. De même, on définit les matrices de régularisation S_j associées à chaque fonction. L'estimation des paramètres revient à minimiser la nouvelle fonction objectif :

$$\|Y - X\beta\|^2 + \sum_{j=1}^k \lambda_j \beta^\top S_j \beta$$

Avec λ_j le paramètre de lissage associé à la covariable j .

La résolution conduit à l'expression explicite de l'estimateur qui s'écrit :

$$\hat{\beta} = (X'X + \sum_{j=1}^k \lambda_j S_j)^{-1} X'Y$$

Pour un paramètre λ_j fixé, on obtient un résultat avec un certain degré de régularité pour la fonction f_j (Figure 3.2). La question à laquelle il faut répondre est celle du choix de la valeur optimale de λ .

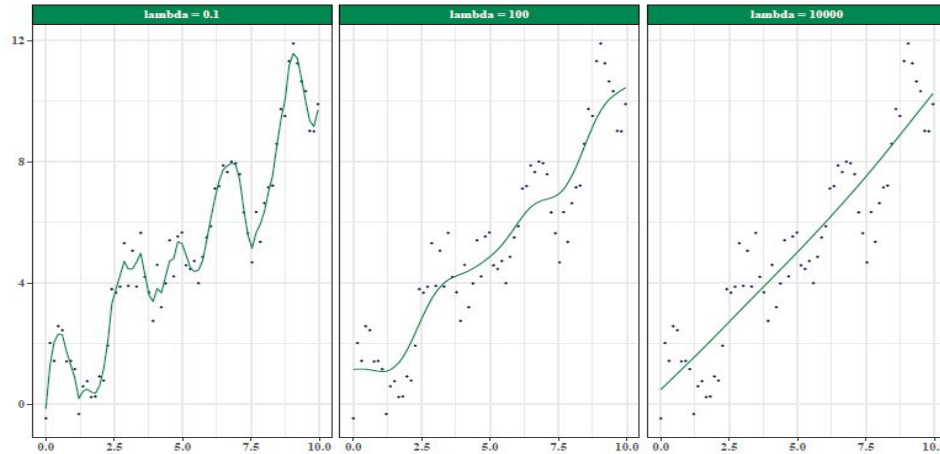


Figure 3.2. : Effet du paramètre de lissage (Source : Wafo Kankeu (2023))

3.3.1.3. Choix de la valeur optimale de λ

L'objectif visé par cet ajustement est d'obtenir des fonctions lisses qui s'approchent au mieux que de la fonction inconnue f à estimer. A ce titre, la solution idéale est de choisir les valeurs de λ qui minimisent l'écart entre la fonction lisse estimée $\hat{f} = X\hat{\beta}$ et la fonction réelle f (tel que $f(x_i) = y_i$).

En pratique, les λ_j seront choisie par validation croisée en minimisant la quantité

$$\nu_0 = \frac{1}{n} \sum_{i=1}^n (\hat{f}^{[-i]}(x_i) - y_i)^2$$

où $\hat{f}^{[-i]}$ est la fonction lisse estimée sans prise en compte de l'observation i .

L'évaluation du critère ν_0 a un coût opérationnel important en termes de temps de calcul, car il nécessite l'estimation de n modèles dans lesquels, à chaque fois, l'observation de rang i doit être exclue. Afin d'éviter ces calculs, il est efficace d'utiliser le critère de validation croisée généralisée (GCV) donné par la formule :

$$GCV = \frac{n \sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{(n - \text{tr}(A))^2}$$

où $A = X(X'X + \sum_{j=1}^k \lambda_j S_j)^{-1} X'$ est la matrice chapeau du modèle et $\text{tr}(A)$ est sa trace, qui correspond au nombre effectif de degrés de liberté du modèle.

3.3.2. Les modèles additifs généralisés

Les modèles additifs généralisés (GAM) généralisent les modèles additifs de la même façon que les modèles linéaires généralisés généralisent les modèles linéaires. En effet, les modèles additifs standard considèrent que la variable dépendante est issue d'une distribution normale avec des résidus homoscédastiques, ce qui n'est pas toujours vérifié. Les GAM permettent de résoudre ce problème et élargissant le champ des lois de la variable dépendante à l'ensemble des distributions de la famille exponentielle.

Concrètement, les modèles additifs généralisés supposent que la variable dépendante Y_i suit une loi appartenant à la famille exponentielle (de moyenne μ_i , et de variance σ_i fonction de μ_i) et modélisent son espérance à l'aide de la relation suivante :

$$g(\mu_i) = X_i\theta + f_1(x_i^1) + f_2(x_i^2) + \dots + \varepsilon_i$$

où : - μ_i est l'espérance de Y_i , - g est la fonction de lien, - $X_i\theta$ représente la partie paramétrique du modèle, - les fonctions f_j sont les fonctions lisses associées aux covariables x^j , - ε est le terme d'erreur aléatoire.

Avec ce modèle, la méthode d'estimation par les moindres carrés pénalisés n'est plus applicable. L'estimation se base sur le maximum de vraisemblance pénalisé.

Cette approche ne permet pas d'obtenir une solution explicite des paramètres, les estimations sont obtenus par des méthodes numériques.

3.4. Qualité d'ajustement et comparaison des modèles

Une fois le modèle estimé, il est important de s'assurer qu'il s'ajuste bien aux observations réelles. De manière générale, cette vérification consiste à une comparaison entre les prédictions faites par le modèles et les données réelles.

Plusieurs instruments sont utiles pour mesurer la qualité d'ajustement du modèle.

3.4.1. Les résidus de déviance

Les résidus de deviance représente une mesure de l'écart entre les prédictions du modèle et les observations réelles. Ils sont analogues aux résidus bruts donnés par $\hat{\varepsilon}_i = y_i - \hat{\mu}_i$, mais dans le cadre des modèles linéaires et additifs généralisés, ces résidus ne sont pas adaptés car la distributions de la variable réponse n'est plus normale et les résidus bruts peuvent naturellement présenter des tendances. Pour pallier ce problème, les résidus de déviance s'expriment par :

$$\hat{\epsilon}_i^{dev} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2[\ell(y_i, y_i) - \ell(\hat{\mu}_i, y_i)]}$$

Où $\ell(x, \theta)$ est la log-vraisemblance associée à l'observation x , pour une distribution de paramètre θ , y_i étant la valeur observée et $\hat{\mu}_i$ étant la valeur estimée par le modèle.

Dans le cas d'un modèle de poisson, cette expression devient :

$$\hat{\epsilon}_i^{dev} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2[y_i \log(\frac{y_i}{\hat{\mu}_i}) - (y_i - \hat{\mu}_i)]}$$

Lorsque le modèle s'ajuste bien aux données, ces résidus doivent se rapprocher d'une distribution normale centrée de variance ϕ .

3.4.2. La déviance

La déviance est l'indicateur global restituant la qualité d'ajustement du modèle. Elle est analogue à la somme des carrés des résidus utilisée dans les modèles linéaires, mais qui n'est pas adaptée dans le cadre général, tout comme les résidus bruts.

La déviance est définie par :

$$dviance = \sum_{i=1}^n (\hat{\epsilon}_i^{dev})^2 = \sum_{i=1}^n dev_i$$

où $dev_i = 2[\ell(y_i, y_i) - \ell(\hat{\mu}_i, y_i)]$ est la déviance associée à l'observation i .

La déviance est adaptée pour la comparaison des performances de deux modèles.

3.4.3. Le backtesting

Le backtesting est une méthode d'évaluation des performances prédictives d'un modèle, spécialement adaptée aux données temporelles. Contrairement à la validation croisée classique, qui peut s'avérer inappropriée lorsque les données présentent une structure temporelle marquée (rendant difficile l'extraction de sous-échantillons représentatifs), le backtesting préserve l'ordre chronologique des observations.

Cette approche consiste à diviser le jeu de données en deux parties distinctes, tout en respectant la structure temporelle :

- Un jeu d'entraînement : constitué des données les plus anciennes, il sert à calibrer le modèle ;
- Un jeu de test : composé des périodes les plus récentes, il permet d'évaluer les capacités de projection du modèle.

Le backtesting est particulièrement utile car il confronte le modèle à de nouvelles données n'ayant pas servi à son estimation, permettant ainsi de détecter d'éventuels problèmes de surapprentissage et d'assurer la robustesse des projections dans un contexte réaliste.

Dans le cadre de cette étude, la plage de données disponibles s'étend de 1990 à 2022. Afin de garantir une évaluation robuste, une période de test de 10 ans a été retenue. Ainsi,

- la période 1990-2012 est utilisée pour l'entraînement des modèles,
- la période 2013-2022 sert de jeu de test pour l'évaluation des performances prédictives.

Cette segmentation permet de tester la capacité des modèles à projeter la mortalité sur une décennie, offrant ainsi une validation concrète de leur utilité qualité prédictive.

Synthèse du chapitre

Ce chapitre a présenté les fondements théoriques essentiels pour la modélisation de la mortalité. Les principales notions et notations actuarielles ont été précisées, dont les concepts de base tels que la fonction de survie et la force de mortalité. La démarche de calibration des modèles classiques de la famille GAPC a été détaillée, incluant notamment le modèle Age-Période, le modèle de Lee-Carter et le modèle Age-Période-Cohorte. Une attention particulière a été portée aux problèmes d'identifiabilité et aux méthodes d'estimation telles que les moindres carrés et le maximum de vraisemblance.

Les modèles additifs généralisés ont ensuite été introduits comme extension flexible permettant de capturer des relations non linéaires, avec une explication de leur calibration par lissage pénalisé. Enfin, les indicateurs d'évaluation des modèles comme les résidus de déviance et l'approche backtesting ont été exposés pour garantir la robustesse des approches retenues. Le chapitre suivant présentera les premiers résultats de modélisation de la mortalité dans le cadre classique, sans intégration des effets de température, établissant ainsi un référentiel pour les analyses comparatives futures.

Chapter 4.

Résultat de la modélisation hors effets de température

Dans cette partie, il s'agira de construire et d'analyser les modèles usuels de mortalité, puis de confronter les résultats obtenus à la littérature. La première section présente brièvement le cadre général de la modélisation. La seconde procède progressivement à la construction des modèles les plus communs, en commençant par les modèles Age-Période (AP), de Lee-Carter et le modèle Age-Période-Cohorte (APC), pour ensuite étendre l'analyse aux modèles fondés sur les approches additives généralisées (GAM). Le chapitre se conclut sur une analyse résiduelle illustrant la pertinence de l'intégration des effets de température.

4.1. Cadre de modélisation

Les données de modélisation

L'approche classique de construction de modèle prospectif de mortalité repose sur les modèles de la famille GAPC, qui prennent en compte les effets d'âge, de la période calendaire et de la cohorte. Les variables explicatives principales étant l'âge et la période, à partir desquelles la cohorte se déduit par différence entre les deux. Dans le cadre de ce mémoire, la structure des données est différente en raison de la distinction des grandeurs par température. Dans toute cette partie, sauf mention contraire, les analyses ignoreront le facteur température, et les données seront préalablement agrégées au format âge, période et cohorte.

Approche backtesting

Pour rappel, la période d'étude s'étale de 1990 à 2022. Afin d'éviter le surapprentissage et de permettre une comparaison robuste des modèles, nous mettons en oeuvre une approche de backtesting consistant à séparer le jeu de données en une partie d'entraînement, utilisée pour la calibration des modèles, et une partie de test, sur

laquelle les comparaisons sont effectuées. La spécificité de cette approche, qui la distingue de la validation croisée, réside dans le maintien de la structure temporelle pour préserver la cohérence des tendances dans les données. Ainsi, chaque échantillon train et test doit donc contenir des périodes historiques continues. À cet effet, le point de rupture considéré est l'année 2013 afin d'avoir un échantillon de test couvrant 10 années.

Par ailleurs, la modélisation se limite à la plage d'âge allant de 0 à 99 ans pour ne pas tenir compte des observations à très faible exposition des centenaires.

4.2. Modélisation classique de la mortalité

4.2.1. Calibration du modèle Age-Période

Le modèle Age-Période est l'un des modèles les plus simples parmi les modèles prospectifs de mortalité. La force de mortalité, exprimée en logarithme, est décomposée de façon additive en un effet de l'âge et un effet propre à la période calendaire, sans interaction entre ces deux dimensions. Cette spécification est décrite par l'Équation 4.1 :

$$\log(\mu_{xy}^s) = \alpha_x^s + \kappa_y^s + \epsilon_{xy}^s \quad (4.1)$$

où :

- x désigne l'âge, y la période calendaire et s le sexe.
- μ_{xy}^s est la force de mortalité pour l'âge x , à la période y et pour le sexe s .
- α_x^s représente l'effet spécifique de l'âge pour le sexe s .
- κ_y^s mesure l'effet de période pour le sexe s .
- ϵ_{xy}^s représente le terme d'erreur aléatoire.

Ce modèle a l'avantage d'être simple à estimer et à interpréter. Comme l'indique le résultat de l'équation Équation 3.1 obtenue par la méthode des MCO, l'effet de l'âge correspond à la valeur moyenne des logarithmes de la force de mortalité de toutes les périodes, tandis que l'effet de la période est la moyenne des écarts entre la mortalité par âge de la période considérée et la mortalité par âge toutes périodes confondues.

Les estimations obtenues sont représentées sur la Figure 4.1 : à gauche se trouve la force de mortalité moyenne par âge, et à droite se trouve un indice de mortalité correspondant à l'effet de la période calendaire. L'indice de période représente un facteur multiplicatif appliqué à l'effet de l'âge pour obtenir son effet spécifique à la période considérée. Il est obtenu par le calcul suivant :

$$\text{Indice périodique}_y = \exp(\kappa_y)$$

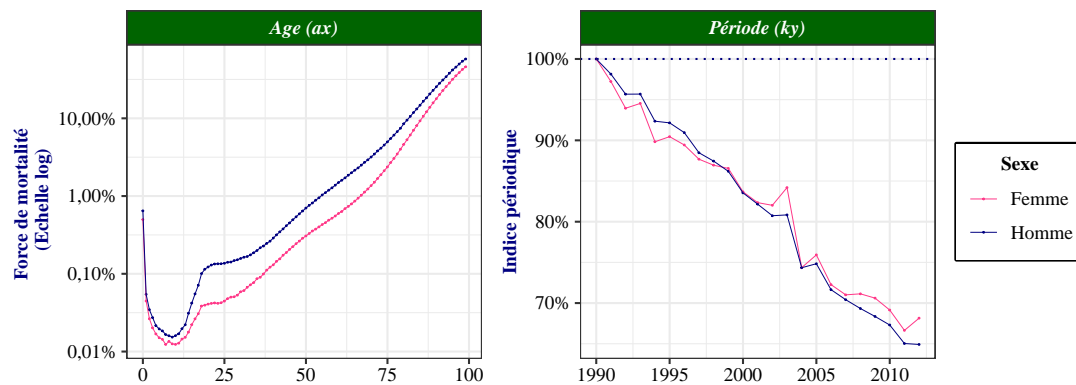


Figure 4.1. : Effet des variables du modèle Age-Période

Analyse de l'effet de l'âge

Ce modèle confirme que l'âge a un effet non monotone sur la mortalité. Plusieurs phénomènes émergent : La forte mortalité à l'âge 0 illustre la mortalité infantile et met en évidence la fragilité des nouveaux-nés, dont une fraction significative décède avant d'atteindre un an. Cette surmortalité infantile est également décrite par l'INSEE (2024), qui précise en outre que la mortalité infantile a fortement baissé entre 1993 et 1995, puis baissé de façon plus modérée jusqu'en 2011. Parmi les jeunes de moins de 10 ans, l'intensité de la mortalité est décroissante avec l'âge, puis remonte progressivement. Entre 10 et environ 22 ans, la mortalité augmente rapidement, avec une accélération plus marquée pour les hommes. On observe autour de 20 ans un pic appelé "**bosse des accidents**". À partir de cet âge, la mortalité croît continuellement avec l'âge, de façon régulière et quasi log-linéaire. Cette tendance générale est similaire pour les deux sexes à l'exception de la bosse des accidents, moins prononcée chez les femmes. De manière générale, la mortalité des hommes est constamment plus élevée que celles des femmes.

Analyse de l'effet de la période calendaire

L'effet de la période calendaire présente une allure générale à la baisse, illustrant la diminution progressive de la mortalité au fil du temps, souvent attribuée aux avancées technologiques et aux progrès de la médecine. Cette explication a été confirmée par Swiss Re Institute (2018), qui indique que ces améliorations sont principalement dues à la réduction des décès pour cause de maladies circulatoires, dont le taux a fortement baissé au cours des dernières décennies. Contrairement à l'effet de l'âge, relativement lisse, l'effet de période présente des fluctuations, correspondant à de brefs écarts par rapport à la tendance générale.

A cet effet, l'effet périodique de la mortalité se décompose en deux composantes : une composante tendancielle qui reflète les avancées technologiques et les progrès de la médecine, et une composante purement aléatoire, qui représente des effets spécifiques à certaines années qui peuvent être dus à des crises ou à des effets conjoncturels. À titre d'illustration, sur la période 1990-2012, qui constitue la plage d'entraînement du modèle, présente un pic particulier en 2003, plus accentué chez les femmes, correspondant à la canicule exceptionnelle de cette année. Quant à la composante tendancielle, elle présente une allure linéaire sur la période 1990-2012, où la mortalité a montré un recul de près de 35 %.

Qualité d'ajustement aux données

Le modèle Age-Période capte les premiers déterminants de la mortalité et reste cohérent avec les résultats de la littérature. Toutefois, sa simplicité ne lui permet pas de saisir l'ensemble des phénomènes explicatifs de la mortalité, ce qui limite sa capacité prédictive. L'analyse des écarts entre la mortalité prédite par le modèle et celle observée, à travers les résidus de déviance (Figure 4.2), révèle une structure nette non captée par le modèle. Rappelons, qu'un modèle bien ajusté présente des résidus aléatoires, sans tendance ni structure particulière. Ici, on observe des bandes linéaires positives et négatives correspondant respectivement à une sous-estimation et une sur-estimation du niveau de mortalité réel. Cette structure suggère la présence d'un effet de cohorte non pris en compte.

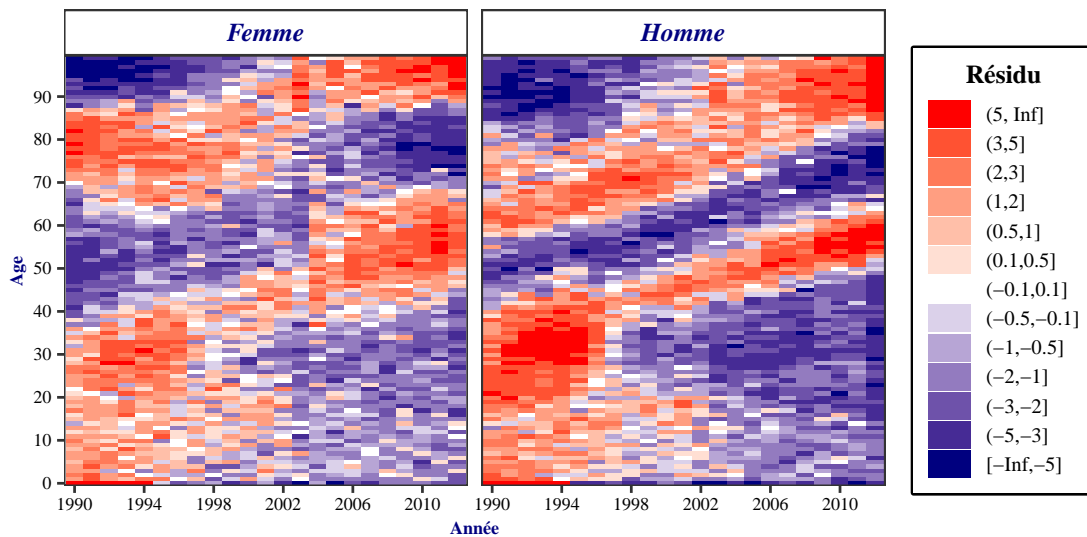


Figure 4.2. : Résidus de déviance du modèle AP

Pour y voir plus clair, il est intéressant de comparer les forces de mortalité observées et celles prédites par le modèle. On observe en effet plusieurs écarts. Sur la période de test.

Le modèle a tendance à surévaluer la mortalité aux âges 10, 20, 40 et 60, la mortalité à 80 ans est assez précise tandis que celle à 95 ans est sous estimée (Figure 4.3).

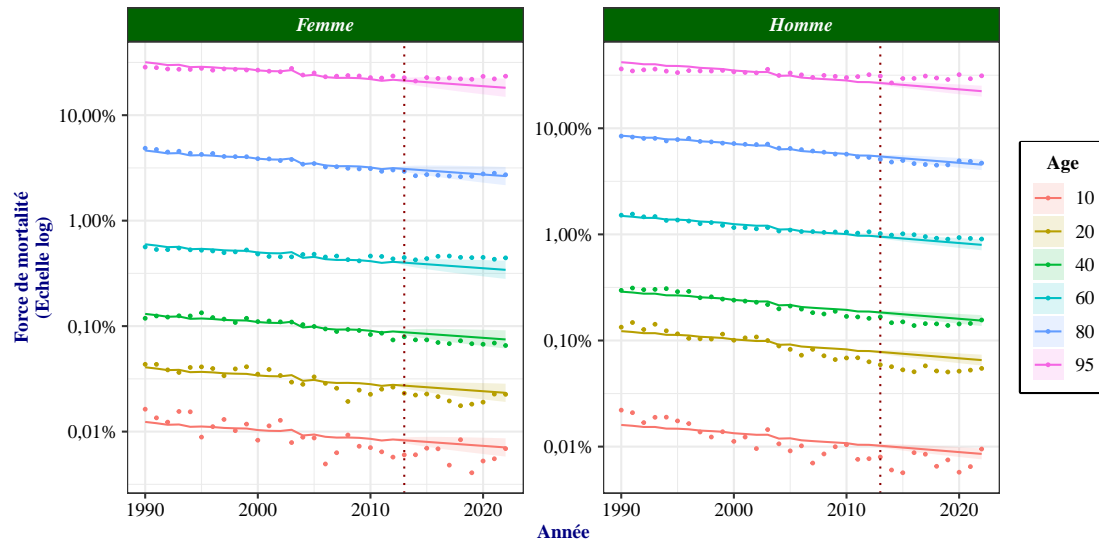


Figure 4.3. : Force de mortalité du modèle Age-Période

En conclusion, ce modèle présente plusieurs écarts par rapport à la réalité et peut être amélioré. Une extension célèbre du modèle Age-Période est le modèle de Lee-Carter, qui intègre des effets d'amélioration de mortalité spécifiques à certains âges.

4.2.2. Calibration du modèle de Lee-Carter

Le modèle de Lee-Carter est l'un des plus connus et utilisés dans le cadre de la modélisation de la mortalité. Sa spécificité, par rapport au modèle Age-Période, réside dans la prise en compte d'un effet calendaire spécifique à chaque âge, ce qui permet d'estimer des améliorations de mortalité variant selon les âges. L'Équation 4.2 présente la forme de ce modèle :

$$\log(\mu_{xy}^s) = \alpha_x^s + \beta_x^s \kappa_y^s + \epsilon_{xy}^s \quad (4.2)$$

où le terme β_x^s est le terme d'interaction entre l'âge et la période calendaire, modulant l'intensité de l'effet périodique selon les âges.

Ce modèle est plus complexe que le modèle Age-Période : la présence de l'interaction en fait un modèle non linéaire qui n'admet pas de solution explicite. Les paramètres sont obtenus par application d'algorithmes numériques à itérations successives. Les estimations obtenues sont présentées sur la Figure 4.4 où le premier graphique décrit la

force de mortalité par âge, le second présente les estimations du paramètre périodique et le troisième correspond aux estimations des intensités d'amélioration par âge.

Les effets des facteurs âge et période sont similaires aux estimations du modèle Age-Période et décrivent les mêmes traits de comportement.

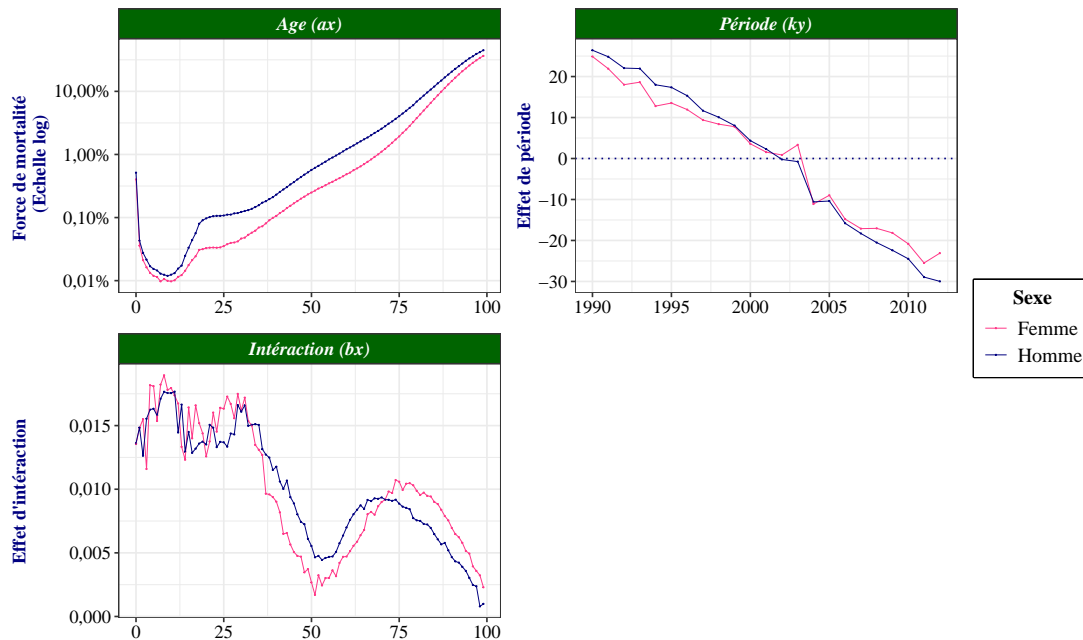


Figure 4.4. : Effet des variables du modèle de Lee-Carter

Analyse de l'effet d'interaction

Le terme d'interaction, couplé au terme de période, représente l'effet global de la période calendaire. La décroissance constante du terme κ_y traduit l'amélioration continue de la mortalité avec le temps, et le terme β_x module l'intensité de ces améliorations selon les âges. Dans un premier temps, on constate que les estimations obtenues sont toutes positives, quel que soit l'âge. Cela signifie que le phénomène d'amélioration de la mortalité implique en moyenne toutes les classes d'âges. Cette interprétation se justifie par la relation suivante, qui découle directement de la spécification du modèle :

$$\beta_x^s = [\log(\mu_{x,y-1}^s) - \log(\mu_{xy}^s)] \cdot (\kappa_{y-1}^s - \kappa_y^s)$$

Comme κ_y^s est décroissant en tendance, la positivité de β_x^s correspond bien, en moyenne, à une décroissance progressive de l'intensité de la mortalité avec le temps.

Par ailleurs, les intensités observées sont maximales et relativement constantes aux âges jeunes (de 0 à environ 30 ans), puis elles diminuent jusqu'à 50 ans avant de remonter à un pic plus faible autour de 75 ans. Cela signifie que les taux d'amélioration de mortalité sont plus intenses chez les jeunes, puis connaissent une atténuation, avant de présenter un regain d'intensité modérée à 75 ans. Cette conclusion est partagée par l'INED (2006), qui, en analysant l'amélioration de l'espérance de vie en France, a également mis en évidence un recul de la mortalité à tous les âges, particulièrement marqué chez les enfants.

Qualité d'ajustement aux données

Le modèle de Lee-Carter conduit à un meilleur ajustement aux données observées, car la plupart des structures présentes dans les résidus du modèle Age-Période ont été prises en compte, comme l'illustre leur représentation sur la Figure 4.5.

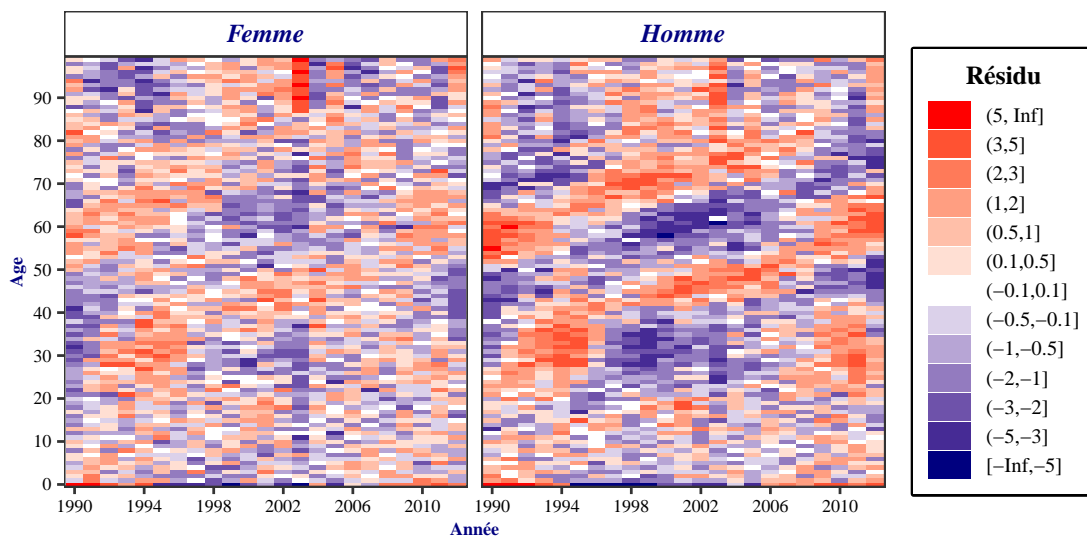


Figure 4.5. : Résidus de déviance du modèle de Lee-Carter

Le modèle décrivant la mortalité masculine semble toutefois moins précis : il présente une structure linéaire d'intensité plus faible, ce qui suggère la persistance d'un effet résiduel de cohorte. Cet effet de cohorte dans les résidus du modèle de Lee-Carter a déjà été identifié par Sunamoto (2005), en l'appliquant à la population japonaise.

Malgré cela, comparativement au modèle Age-Période, les prédictions issues du modèle de Lee-Carter captent mieux la tendance de la mortalité par âge et permettent une bonne généralisation aux périodes futures. Pour la majorité des âges considérés, les prédictions semblent sans biais, à l'exception de l'âge de 10 ans dans le modèle des hommes, où la mortalité réelle a été sous estimée (Figure 4.6).

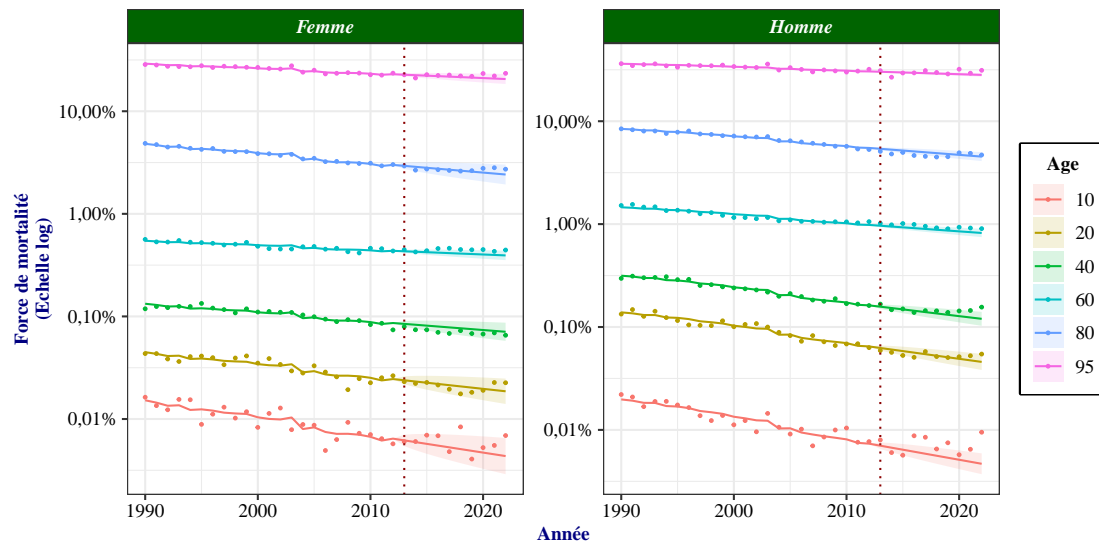


Figure 4.6. : Force de mortalité du modèle de Lee-Carter

Plusieurs critiques de ce modèle ont toutefois été formulées, conduisant à l'émergence de nouveaux modèles qui s'en inspirent. C'est dans ce contexte que le modèle de Renshaw-Haberman a été développé, en intégrant un effet de cohorte au modèle de Lee-Carter de la façon suivante :

$$\log(\mu_{xy}^s) = \alpha_x^s + \beta_{x,(1)}^s \kappa_y^s + \beta_{x,(0)}^s \lambda_{y-x}^s + \epsilon_{xy}^s$$

où le terme λ_{y-x}^s est le paramètre de cohorte et le paramètre $\beta_{x,(0)}^s$ module son intensité par âge.

La spécification du modèle de Renshaw-Haberman le rend plus complet que celui de Lee-Carter, car il permet notamment de capter la structure résiduelle liée aux effets de cohorte, précédemment identifiée dans les résidus. En théorie, cette extension devrait améliorer la qualité de l'ajustement. En pratique, toutefois, le modèle ne se révèle pas toujours plus performant et n'apporte pas systématiquement de meilleurs résultats. Parmi ses principales limites, on retrouve la sur-paramétrisation ainsi que des problèmes de convergence. En effet, lors de son application au jeu de données de ce travail, l'estimation des paramètres s'est avérée très coûteuse en temps de calcul et n'a pas convergé vers des valeurs stables dans le cas de la population masculine.

Par ailleurs, bien que l'ajustement aux données d'apprentissage se soit montré satisfaisant, il s'agissait en réalité d'un phénomène de surapprentissage, le modèle généralisant mal en période de test. Ces constats rejoignent ceux de Cairns *et al.* (2008), qui sont parvenus à des conclusions similaires en étudiant ce modèle et ses projections de mortalité sur les données de la population d'Angleterre.

Enfin, une sous-structure de ce modèle, particulièrement connue et couramment utilisée, est le modèle APC qui conduit généralement à des résultats plus robustes.

4.2.3. Calibration du modèle Age-Période-Cohorte

Le modèle APC est une extension du modèle Age-Période intégrant un effet additif de cohorte. Sa spécification est décrite par l'Équation 4.3 :

$$\log(\mu_{xy}^s) = \alpha_x^s + \kappa_y^s + \gamma_{y-x}^s + \epsilon_{xy}^s \quad (4.3)$$

Où tous les paramètres ont la même signification que les modèles précédents.

Par rapport au modèle de Lee-Carter, le modèle APC a l'avantage d'être plus simple à interpréter car il s'agit d'un modèle linéaire. Toutefois, il contient plus de paramètres mais moins que le modèle RH.

Les estimations obtenues pour les paramètres sont présentées sur la Figure 4.7. Les effets des facteurs âge et période gardent les mêmes formes que ceux observés dans les modèles Age-Période et de Lee-Carter. La particularité réside dans la présence de l'effet de cohorte, qui décrit un aspect supplémentaire de la mortalité.

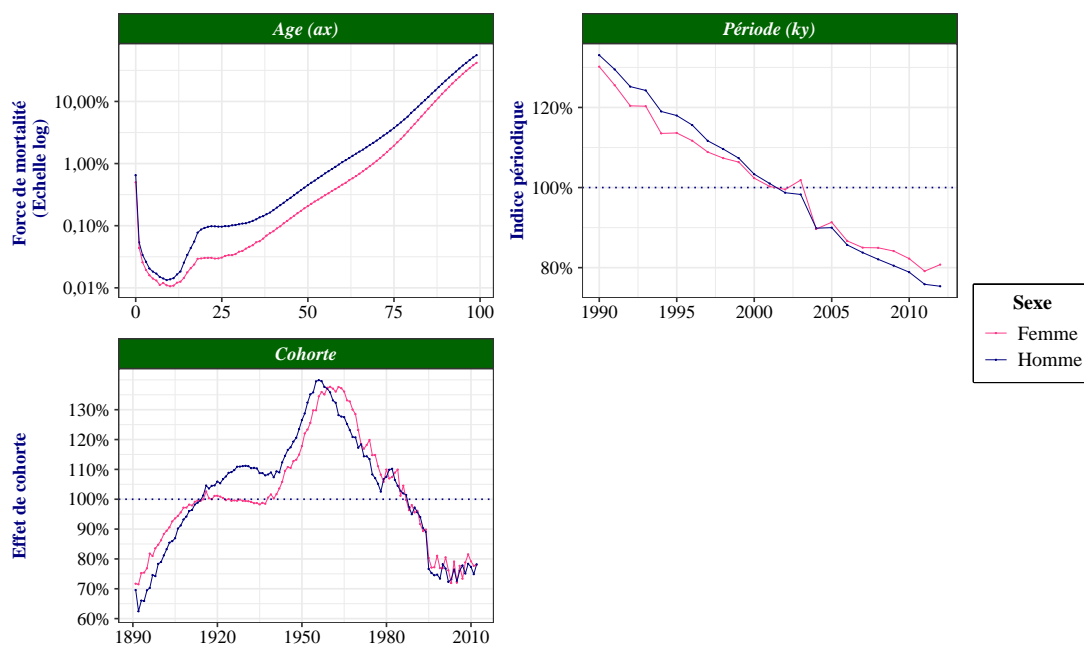


Figure 4.7. : Effet des variables du modèle APC

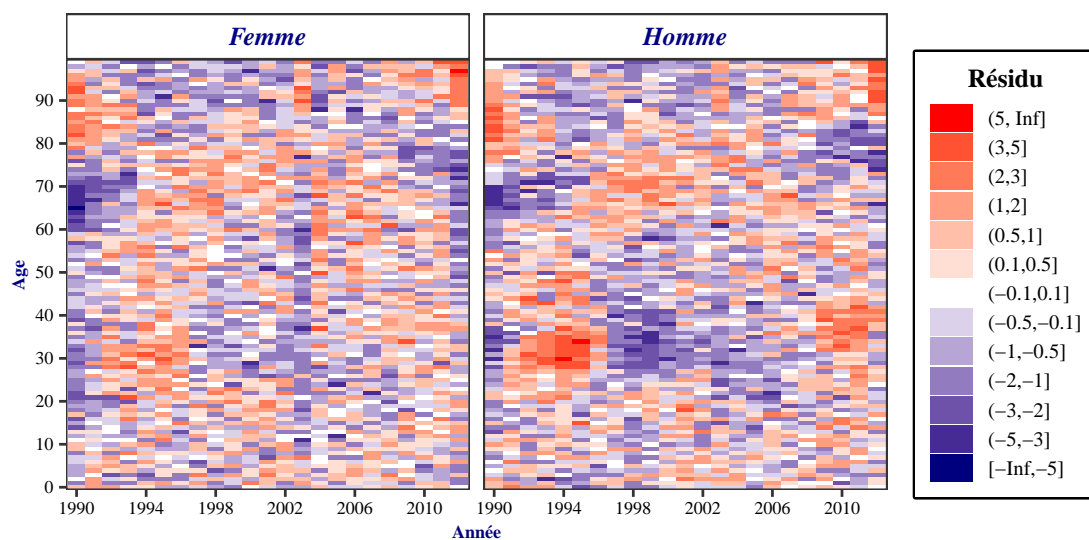
Les effets du facteur de cohorte

L'effet du facteur cohorte est décrit par un indice de mortalité générationnelle similaire à l'indice périodique. Ces résultats, qui sont similaires entre les hommes et les femmes, suggèrent que la mortalité moyenne des générations postérieures à l'année 2000 est équivalente à celle des années 1890. Toutefois, entre 1890 et 2000 l'effet évolue fortement. En analysant la Figure 4.7, il peut en effet être constaté qu'il y'a une forte croissance de la mortalité entre les générations de 1890 à environ 1957, où le maximum est atteint, puis le taux de mortalité décroît continuellement jusqu'aux années 2000, où une stabilisation est observée.

Les spécificités décrites par cet effet de cohorte ne sont pas simples à interpréter et très peu d'études en donnent une explication permettant une meilleure compréhension. En général, on en déduit que les individus qui sont nés la même année ont traversé au même moment des événements marquants qui impactent durablement leur niveau de mortalité.

Qualité d'ajustement aux données

L'intégration de l'effet de cohorte au modèle Agé-Période permet d'améliorer significativement son ajustement aux données. En effet, cela peut se voir sur la Figure 4.8 où la structure des résidus est beaucoup plus proche d'une répartition aléatoire, et les structures linéaires suggérant la présence d'un effet générationnel ont été prises en compte.



Les forces de mortalité estimées par le modèle APC sont très proches des valeurs observées, notamment en période d'entraînement. En période de test, les prédictions sont également assez proches, en particulier aux âges 60 et 80. À 10 ans, la tendance est aussi bien captée malgré la forte variabilité liée au faible nombre de décès à cet âge. Toutefois, le modèle conduit à quelques écarts sur la période future, où une sous-estimation de la mortalité se produit pour les âges 20, 40 et même 95 (Figure 4.9).

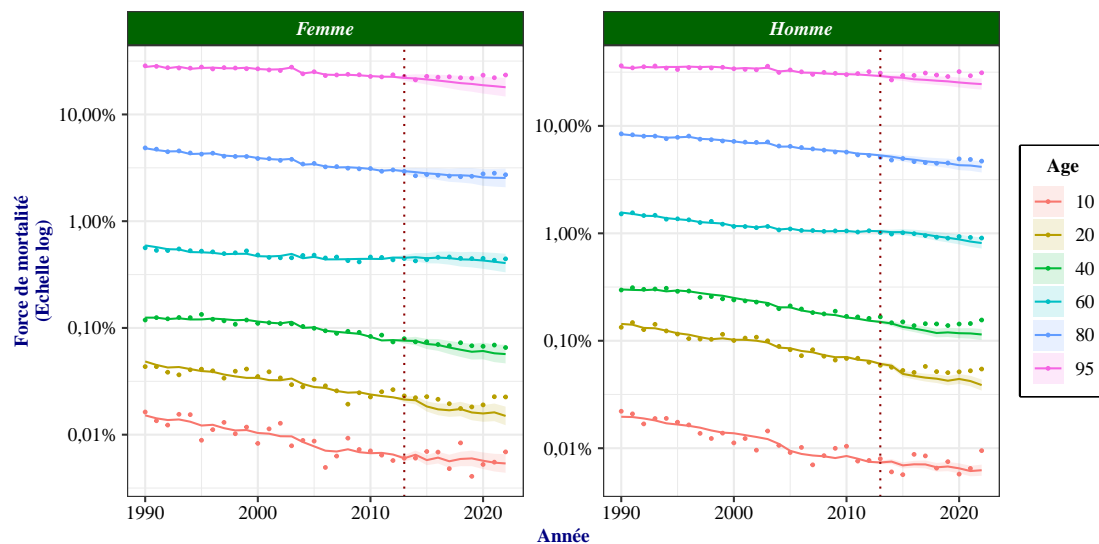


Figure 4.9. : Force de mortalité du modèle Age-Période-Cohorte

Les modèles AP, APC et de Lee-Carter sont très connus et régulièrement utilisés dans la littérature. Les modèles APC et de Lee-Carter sont des extensions simples du modèle AP qui s'ajustent mieux aux données et conduisent en général à de meilleures prédictions de la mortalité future. De nombreuses extensions de ces modèles existent également et appartiennent en général à la famille des modèles GAPC. Plus de détails sont présentés par Villegas, Millosovich et Kaishev (2018), qui introduisent en plus leur implémentation sous R avec le package StMoMo.

Dans le cadre de ce mémoire qui accorde une attention particulière au rôle des facteurs climatiques sur la mortalité et en particulier celui de la température, la démarche suivie consiste à enrichir ces modèles usuels en y intégrant des facteurs de température. C'est une approche similaire à celle proposée par Fall (2021), qui a considéré des termes de période modulés par des coefficients représentant la corrélation entre la mortalité et la température, mais sans toutefois intégrer une variable de température au modèle. Les modèles de la famille GAPC ne permettant pas directement d'intégrer des variables exogènes, une nouvelle modélisation s'appuyant sur les modèles additifs généralisés a donc été effectuée.

4.2.4. Modélisation par une approche GAM

Les modèles additifs généralisés (GAM) offrent une grande flexibilité dans la spécification des modèles, ce qui permet de représenter des relations plus ou moins complexes. Ils s'adaptent particulièrement bien à la modélisation de la mortalité en associant les facteurs âge, période, cohorte et éventuellement d'autres variables selon les besoins. En plus de cette flexibilité, les GAM présentent plusieurs avantages :

- **Estimation directe des effets lisses** : dans les approches classiques, les effets des variables sont estimés point par point, et doivent être suivis d'une étape de lissage afin d'assurer leur régularité et de limiter les biais d'échantillonnage. Avec les GAM, les effets sont directement estimés sous forme lisse, sans étape supplémentaire.
- **Facilité d'interpolation et d'extrapolation**: les GAM fournissent un cadre naturel pour l'interpolation et d'extrapolation, en prolongeant les tendances observées. Il est ainsi aisé d'estimer le niveau de mortalité entre deux âges, ou encore de projeter les taux de mortalité aux âges élevés.
- **Réduction du nombre de paramètres et limitation du surapprentissage** : L'estimation d'un paramètre par point, comme dans les modèles précédents, entraîne un nombre élevé de paramètres et un risque de surapprentissage. Les GAM nécessitent moins de paramètres et produisent des estimations plus stables, moins sensibles aux bruits des données.

Dans cette partie, deux modèles GAM sont estimés à partir des facteurs usuels. Le premier s'inspire du modèle de Lee-Carter et vise à reproduire ses effets sous une version lisse et régulière, tandis que le second adopte une structure plus générale, susceptible de capturer d'éventuelles interactions supplémentaires des variables.

4.2.4.1. Calibration du modèle lissé de Lee-Carter

Le modèle proposé s'inspire du modèle de Lee-Carter et vise à reproduire ses effets à l'aide d'un modèle linéarisé estimant des effets lisses. Il comporte un effet structurel de l'âge, qui capte la mortalité moyenne par âge, un effet linéaire de la période calendaire dont l'intensité est modulée selon les âges, comme dans le modèle de Lee-Carter, ainsi qu'un effet résiduel périodique représentant les fluctuations non tendanciennes propres à certaines périodes du fait de crises ou de chocs particuliers. Ce modèle GAM est spécifié de la façon suivante :

$$\log(\mu_{xy}^s) = f_1^s(x) + f_2^s(x) \cdot (y - \bar{y}) + k_\epsilon^s(y) + \epsilon_{xy}^s$$

où :

- f_1^s est l'effet structurel lisse de l'âge pour le sexe s
- f_2^s est le terme lisse d'intensité des améliorations de mortalité selon les âges pour le sexe s
- $h_\epsilon^s(y)$ est l'effet résiduel spécifique de l'année y
- ϵ_{xy}^s est le terme d'erreur aléatoire

Cette spécification tient compte du caractère régulier du facteur âge, justifié par l'idée que la mortalité à des âges voisins doit être proche et ne saurait présenter des fluctuations brusques. De plus, la décomposition de l'effet calendaire en une composante tendancielle régulière et une composante résiduelle purement aléatoire est également intégrée.

Les effets des variables

Les estimations obtenues en calibrant ce modèle via le package MGCV sont présentées sur la Figure 4.10.

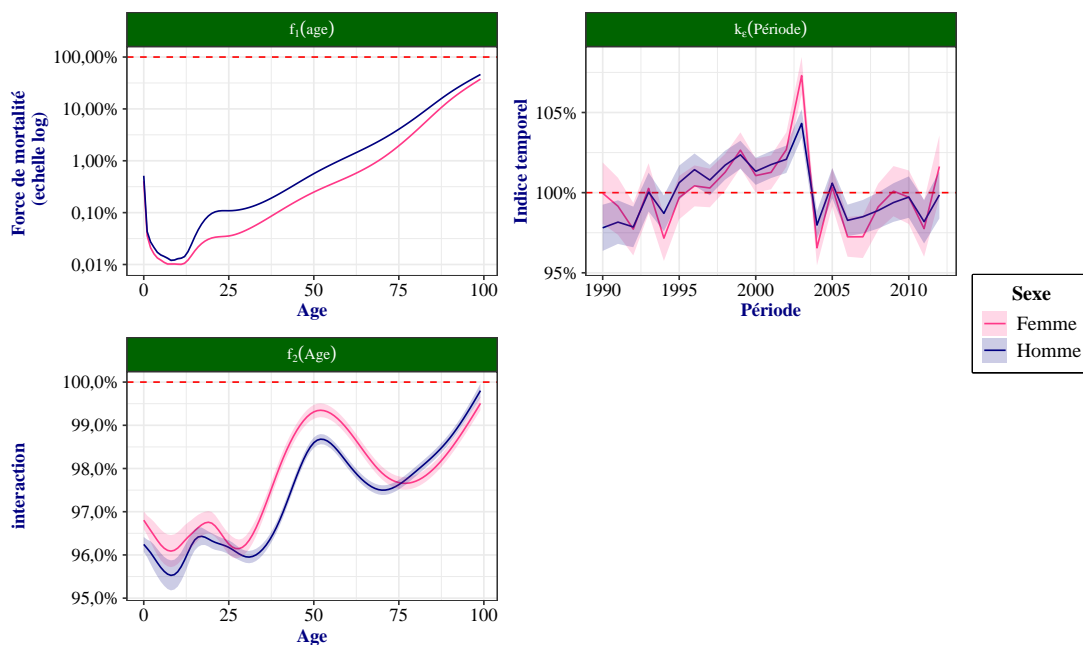


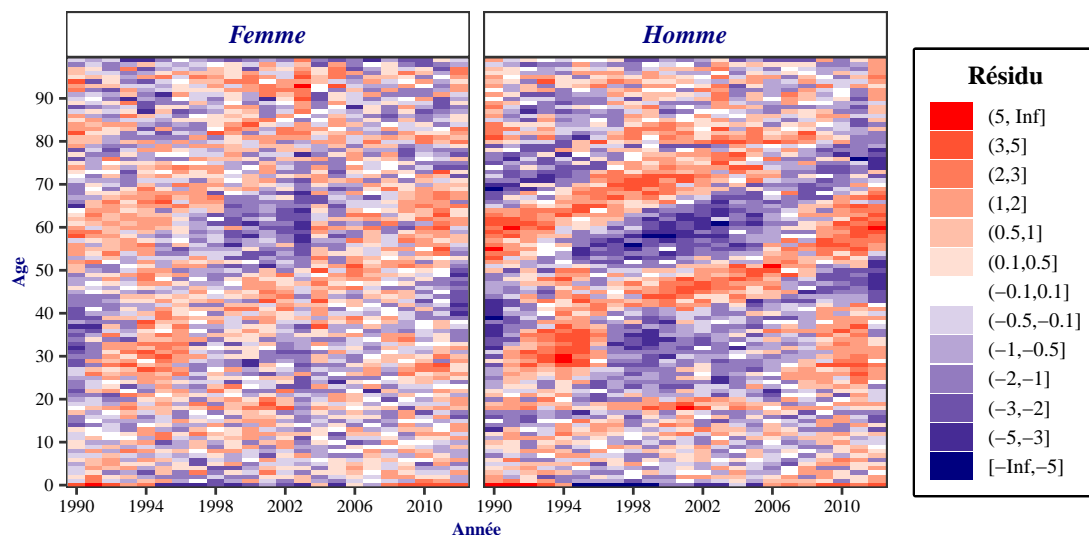
Figure 4.10. : Les effets du modèle de mortalité GAM

Les effets obtenus sont effectivement similaires à ceux du modèle de Lee-Carter. L'effet de l'âge et les phénomènes en présence restent identiques. Le terme d'intensité des améliorations de mortalité montre des gains plus importants pour les jeunes, moins marqués autour de 50 et 100 ans, puis localement plus favorable aux alentours de 75 ans. Ces améliorations concernent l'ensemble des âges : l'inversion du sens de la courbe est liée au fait que l'effet calendaire linéaire a été modélisé comme croissant. La décroissance progressive de la mortalité est donc reflétée par le terme d'intensité, qui prend des valeurs

négatives sans influence sur les interprétations. En moyenne, la mortalité des moins de 30 ans diminue d'environ 4% par an. Enfin, le terme résiduel périodique reflète bien les fluctuations propres à certaines périodes : les écarts estimés sont en général inférieurs à 10 %, le plus marqué étant celui de 2003, où la population féminine a connu une surmortalité annuelle d'environ 7,5 %.

Qualité d'ajustement aux données

La structure de résidus obtenue avec ce modèle (Figure 4.11) est très proche de celle du modèle de Lee-Carter. Il peut donc être considéré comme une version lissée et linéaire de Lee-Carter. Cette correspondance s'explique par le fait que l'effet calendaire estimé par Lee-Carter suit une tendance globalement linéaire. Le GAM suppose également un effet calendaire linéaire, ce qui justifie cette similarité. Toutefois, cet ajustement n'est pas garanti dans le cas d'une population pour laquelle l'effet périodique ne serait pas linéaire.



Les mortalités prédites par ce modèle s'ajustent bien aux données d'entraînement pour tous les âges, mais son pouvoir prédictif semble limité sur la période future, où une tendance à la sous-estimation de la mortalité est observée (Figure 4.12).

Pour mieux anticiper la mortalité future, il est envisageable de recourir à un modèle plus général, dans lequel la forme de la tendance périodique n'est pas imposée mais estimée directement à partir des données. Ce modèle sera décrit dans la sous-section suivante.

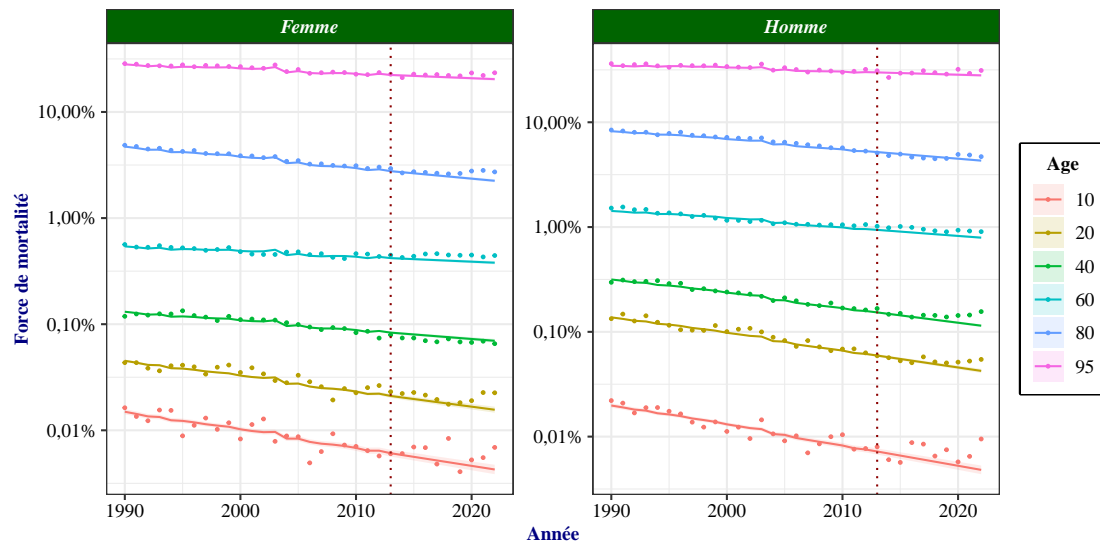


Figure 4.12. : Force de mortalité du modèle GAM

4.2.4.2. Calibration d'un second modèle GAM

Dans ce nouveau modèle, l'objectif est de ne pas imposer une forme particulière aux effets des variables ni à leurs interactions, mais de capter directement les effets qui émergent des données. Pour cela, une spécification plus générale est proposée et décrite par l'Équation 4.4 :

$$\log(\mu_{xy}^s) = f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y) + \epsilon_{xy}^s \quad (4.4)$$

où :

- f_1^s est l'effet structurel lisse de l'âge pour le sexe s
- f_2^s est l'effet tendanciel lisse de la période pour le sexe s
- $k_\epsilon^s(y)$ est l'effet spécifique de l'année y
- f_3^s est le terme lisse d'interaction entre l'âge et l'année, captant les écarts non expliqués par les effets d'âge et de tendance.
- ϵ_{xy}^s est le terme d'erreur aléatoire

Les effets des variables

Après estimation des paramètres du modèle, le facteur âge décrit une fois encore les mêmes phénomènes que dans les modèles précédents. L'effet périodique est décomposé en une composante tendancielle et une aléatoire, toutes deux estimées. La tendance

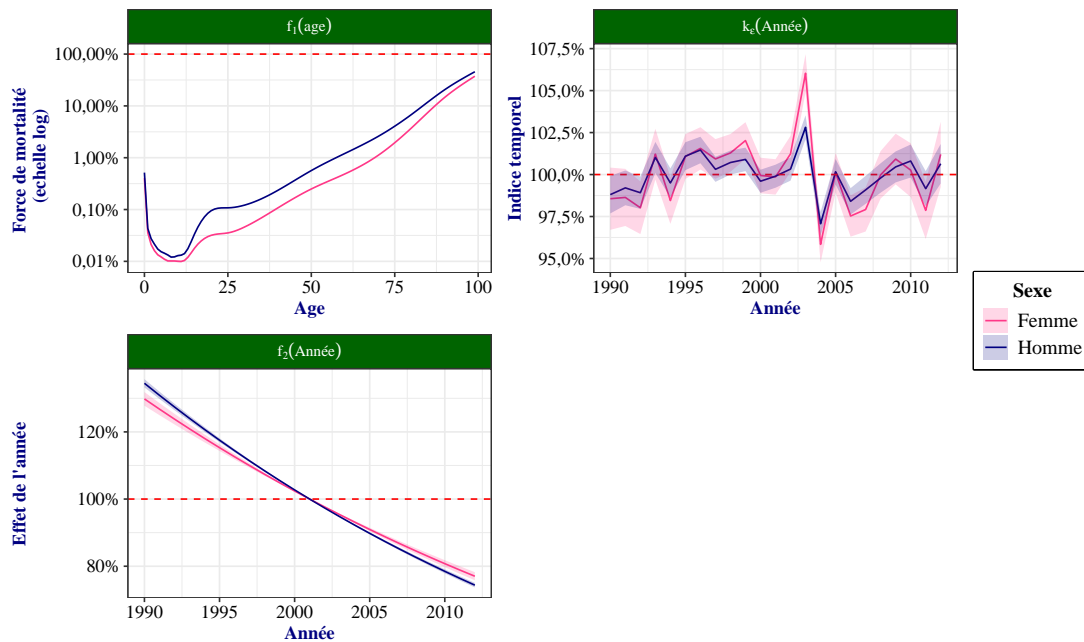


Figure 4.13. : Les effets de la deuxième version du modèle GAM

périodique des améliorations présente une forme log-linéaire, plus marquée chez les hommes (Figure 4.13).

L'effet de tendance traduit une amélioration générale de la mortalité, tous âges confondus, mais son intensité varie en réalité selon les âges et le temps. L'effet périodique spécifique à chaque âge est décrit par la Figure 4.14. Globalement, le résultat est cohérent avec ceux de Lee-Carter et décrit une mortalité élevée aux années 1990, qui décroît progressivement à des rythmes variés pour chaque âge. Plus particulièrement, les jeunes de moins de 35 ans bénéficient d'améliorations plus rapides de la mortalité, tandis que les populations autour de 50 ans et 100 ans connaissent des gains plus faibles. En revanche, les personnes proches de 75 ans présentent localement des améliorations plus marquées.

Cette description correspond bien à celle de l'interaction captée par le modèle de Lee-Carter, ce qui conforte la pertinence des résultats. Une autre façon d'analyser les évolutions entre les années consiste à observer leurs effets spécifiques estimés pour chaque année. La Figure 4.15 illustre le prolongement des tendances de mortalité vers les premières années futures au delà de la période d'entraînement des données.

Qualité d'ajustement aux données

L'analyse des résidus révèle certaines imperfections, plus marquées chez les hommes.

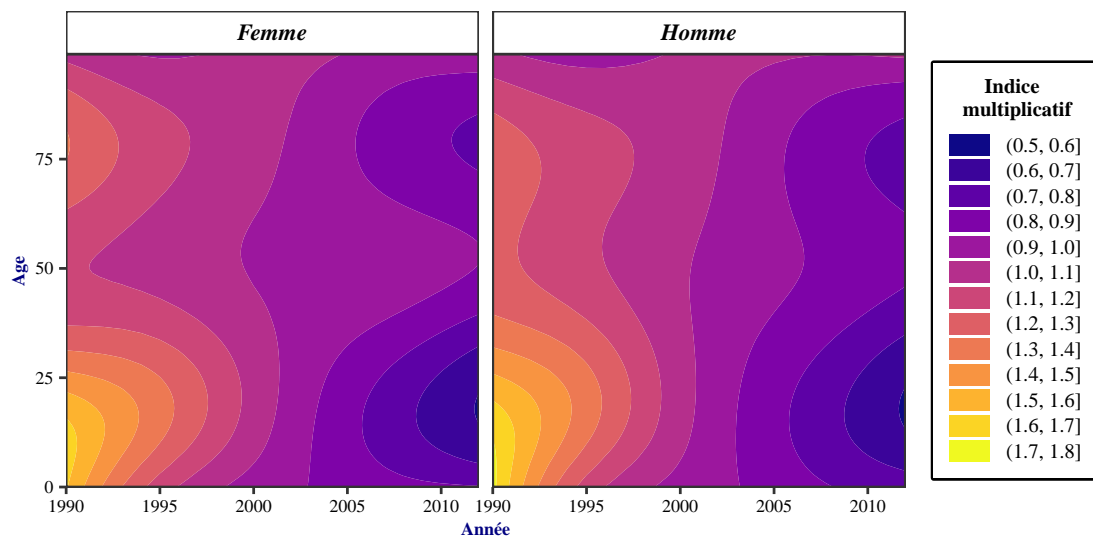


Figure 4.14. : Effet périodique par âge

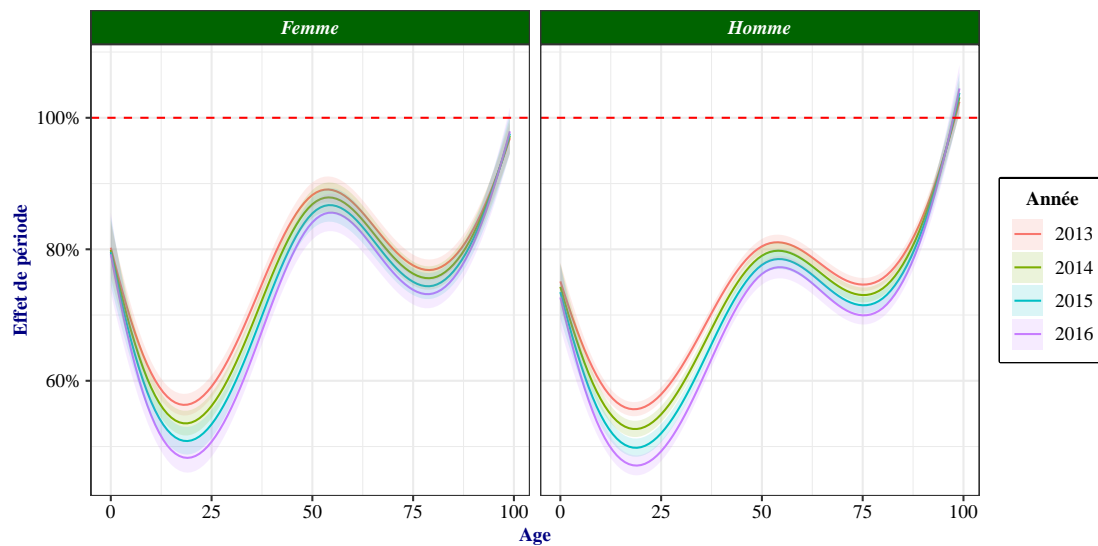


Figure 4.15. : Amélioration de mortalité par âge des périodes futures

Elles traduisent un effet de cohorte non pris en compte par le modèle (Figure 4.16). L'intégration d'un terme de cohorte pourrait améliorer la précision, mais la forte corrélation entre les facteurs d'âge, de période et de cohorte rend les estimations instables et difficilement interprétables, avec un risque de biais sur les prédictions. Pour cette raison, il a été choisi de ne pas inclure de facteur de cohorte, le modèle présentant déjà une qualité d'ajustement satisfaisante.

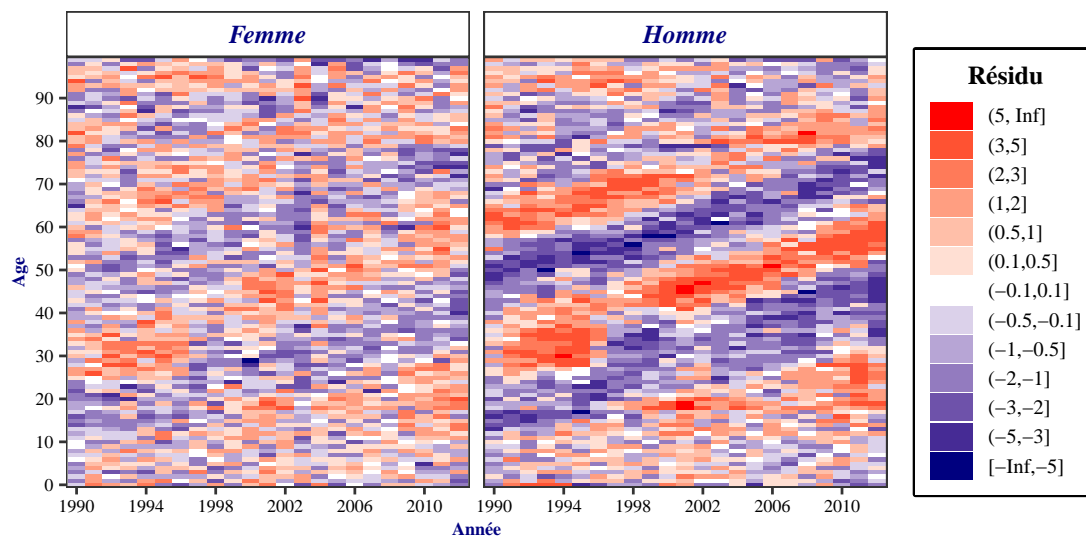


Figure 4.16. : Les résidus du second modèle GAM

La mortalité prédite par le modèle apparaît meilleure que celle obtenue avec le modèle précédent. L'ajustement est bon, aussi bien sur la période d'entraînement que celle de test, à l'exception des âges autour de 20 ans, où une sous-estimation de la mortalité est observée (Figure 4.17).

Cette analyse permet de mieux comprendre les effets des facteurs usuels d'âge, de période calendaire et de cohorte. Toutefois, la question de la combinaison optimale de ces variables pour obtenir le modèle le plus performant reste ouverte. Une comparaison des modèles est donc nécessaire afin de situer les modèles GAM parmi les modèles classiques.

4.2.5. Comparaison des modèles

L'évaluation de la qualité d'un modèle peut se faire de différentes manières. De façon générale, il s'agit d'apprécier la capacité du modèle à produire des estimations proches des observations réelles. La majorité des indicateurs de performance mesurent en réalité une forme de distance entre les prédictions du modèle et les valeurs observées. Un modèle bien calibré doit avant tout offrir un bon ajustement aux données d'entraînement. Dans

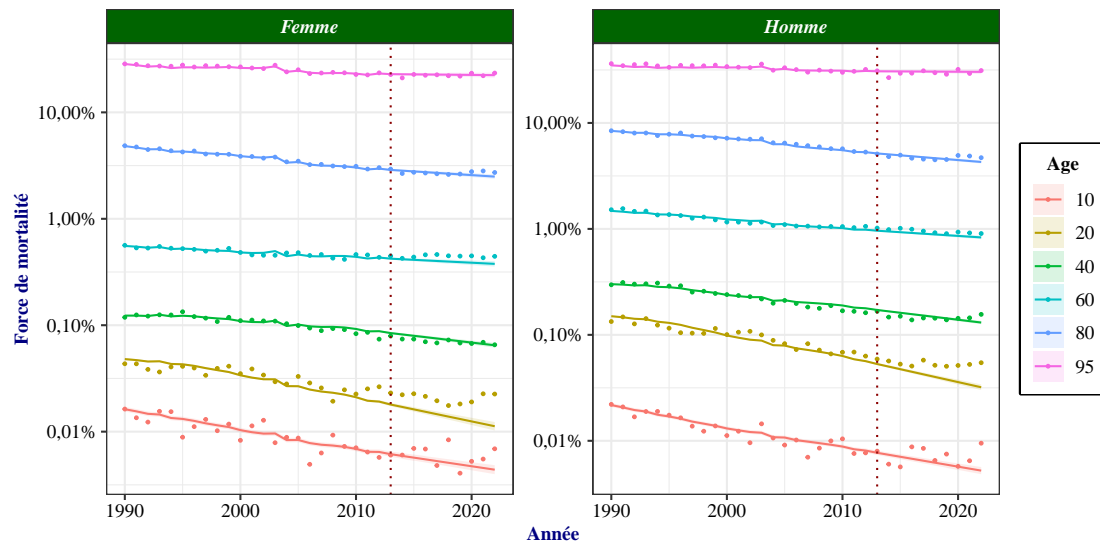


Figure 4.17. : Force de mortalité du modèle GAM complet

le cas d'un modèle prospectif, il doit également être capable de fournir des prédictions fiables sur des données futures.

De nombreux critères d'évaluation reposent sur la vraisemblance, parfois accompagnés de pénalisations liées au nombre de degrés de liberté (EDF) afin de tenir compte du surapprentissage. Dans cette démarche, l'évaluation des modèles se fera aussi bien en termes d'ajustement aux données historiques sur la plage d'entraînement qu'en termes de performance prédictive de la mortalité sur la plage des données futures.

Dans les sous-sections suivantes les termes GAM1 et GAM2 seront utilisés pour désigner, d'une part, le modèle GAM lisse répliquant le modèle de Lee-Carter, et d'autre part le modèle GAM plus général calibré précédemment.

Ajustement aux données d'entraînement

Pour la qualité d'un modèle, la déviance est le premier indicateur à considérer. Elle établit une distance entre les données et les observations sans pénalisation. Selon ce critère, le modèle APC est le meilleur, suivi par le modèle GAM2. Les modèles GAM1 et Lee-Carter présentent des performances similaires. Comme attendu, le modèle AP est le moins performant. Toutefois, un bon ajustement aux données d'entraînement ne garantit pas que le modèle fera de bonnes prédictions, notamment en raison d'un éventuel surapprentissage mesurable par le nombre de degrés de liberté. Le modèle APC est celui qui possède le plus de degrés de liberté et est donc le plus susceptible de faire du surapprentissage. Pour évaluer la capacité de généralisation les indicateurs pénalisés sont plus informatifs.

En considérant ces indicateurs, les performances varient selon le sexe. Pour les femmes, les meilleures performances sont obtenues par le modèle GAM2, suivi de près par les modèles APC et GAM1. L'écart de performance entre le modèle de Lee-Carter et le modèle GAM1, qui est sa version lissée, s'explique par la différence importante de degrés de liberté entre ces deux modèles. Quelle que soit la mesure considérée, le modèle AP reste le moins performant (Table 4.1).

Table 4.1. : Ajustement des modèles aux données d'entraînement (Femme)

Modèle	Déviante	GCV	AIC	BIC	RMSE	EDF
AP	12 196,0	5,9	31 923,7	32 624,0	131,7	122,0
Lee-Carter	4 160,5	2,2	24 090,2	25 370,4	73,1	223,0
APC	3 546,9	1,9	23 520,6	24 927,0	68,4	245,0
GAM1	4 171,1	1,9	23 761,9	24 069,2	69,2	53,5
GAM2	3 977,6	1,8	23 576,1	23 905,6	67,2	57,4

Dans le cas des hommes, le modèle GAM2 n'est plus le plus performant : le modèle APC a un meilleur ajustement en entraînement, suivi par les modèles Lee-Carter, GAM1 et GAM2 qui ont toutefois des performances similaires. Les écarts de performance du modèle de Lee-Carter par rapport aux modèles GAM peuvent s'expliquer par l'effet résiduel de cohorte observé dans les résidus, absent ou moins marqué chez les femmes. Le modèle AP est à nouveau moins performant, quel que soit le critère considéré (Table 4.2).

Table 4.2. : Ajustement des modèles aux données d'entraînement (Homme)

Modèle	Déviante	GCV	AIC	BIC	RMSE	EDF
AP	17 704,7	8,6	38 410,0	39 110,3	143,1	122,0
Lee-Carter	6 578,3	3,5	27 485,6	28 765,8	93,6	223,0
APC	4 002,0	2,2	24 953,3	26 359,7	71,0	245,0
GAM1	7 011,2	3,2	27 581,4	27 893,8	95,9	54,4
GAM2	7 799,4	3,6	28 375,2	28 703,7	97,2	57,2

Sur la base de ces résultats, on peut s'attendre à observer des prédictions plus précises avec le modèle GAM2 pour les femmes et avec le modèle APC pour les hommes. Pour le vérifier, l'évaluation des performances des modèles sera menée sur la période future.

Ajustement aux données de test

En prédiction, le critère d'intérêt principal est la précision des modèles dans leurs estimations. À ce titre, les indicateurs se basant sur une pénalisation de l'ajustement aux données ne sont plus pertinents pour mesurer les performances en projection.

Sur la période de projection, la qualité des estimations, mesurée par la déviance, montre que le modèle le plus performant est le GAM2 pour les deux sexes. Ce résultat est conforme à l'attente pour le cas des femmes. Pour les hommes, bien que le modèle GAM2 ait montré un plus faible ajustement aux données d'entraînement, il se révèle meilleur en projection. Le classement des modèles restants varie selon le sexe :

Pour les femmes, le modèle de Lee-Carter obtient aussi de bons résultats mais reste inférieur au GAM2. Les modèles GAM1 et APC présentent des performances proches, tandis que le modèle AP reste le moins performant (Table 4.3).

Pour le cas des hommes, le modèle APC est le second plus performant, suivi des modèles Lee-Carter et GAM1, aux résultats similaires. Le modèle AP est à nouveau le moins performant (Table 4.4).

Table 4.3. : Performances en projection (Femme)

Modèle	Déviance	EDF
AP	36 643,9	122,0
Lee-Carter	21 804,6	223,0
APC	29 738,9	245,0
GAM1	28 790,0	53,5
GAM2	13 924,2	57,4

Table 4.4. : Performances en projection (Homme)

Modèle	Déviance	EDF
AP	42 025,3	122,0
Lee-Carter	31 159,0	223,0
APC	27 603,2	245,0
GAM1	32 477,3	54,4
GAM2	20 737,5	57,2

Il est connu que les performances des modèles se dégradent progressivement avec la profondeur de la période de projection, en raison de l'évolution des comportements des populations qui rend les estimations obsolètes. Il est donc utile d'observer cette évolution au fil des années, ce qui permet également d'apprécier la stabilité des modèles. Cette analyse, présentée dans la Figure 4.18, montre effectivement une diminution progressive de la qualité d'ajustement des modèles. Les résultats suggèrent également qu'en termes de déviance, le modèle GAM2 est le meilleur dans l'ensemble, notamment pour la population féminine. Chez les hommes, le modèle APC est supérieur au début mais se dégrade plus rapidement que le GAM2, ce qui confère à ce dernier une meilleure performance globale. Les modèles GAM1 et Lee-Carter présentent des performances constamment proches, ce qui confirme leur proximité, tandis que le modèle AP est constamment le moins performant.

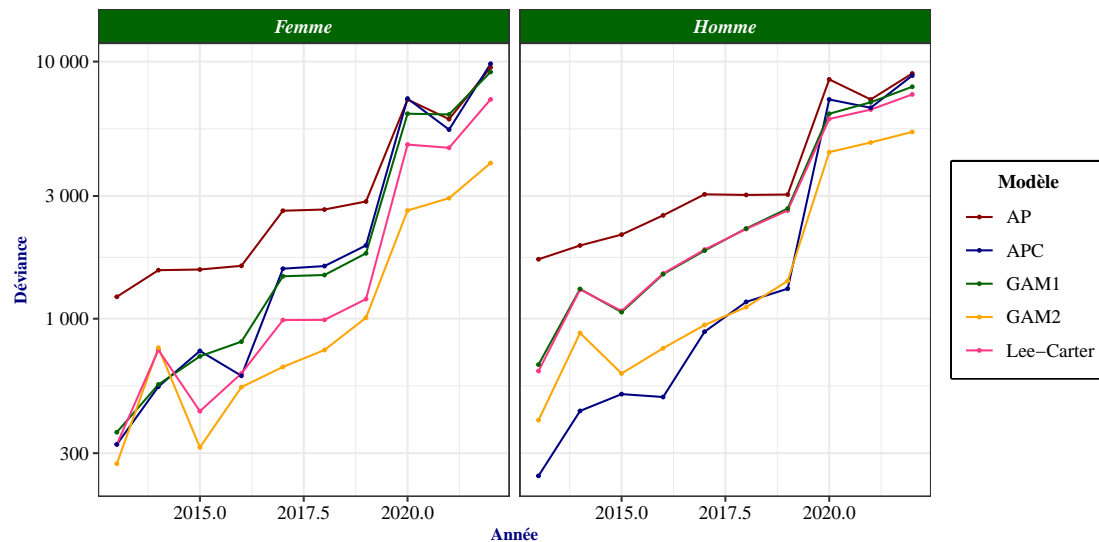


Figure 4.18. : Déviance annuelle des modèles en période de test

Il a été observé que le modèle GAM1, lisse, a bien reproduit les effets et les performances du modèle de Lee-Carter, notamment pour les hommes. Cependant, la forme plus complète du GAM (GAM2) conduit à de meilleures performances prédictives surpassant également les modèles classiques. Dans la suite, ce modèle est retenu comme référence pour l'intégration des effets de température, car il s'agit d'un modèle lisse qui réplique correctement les effets liés à la mortalité, tout en étant flexible et présentant de bonnes performances prédictives.

À présent, il sera question d'illustrer la pertinence de la prise en compte du facteur de température dans ce modèle.

4.3. Mise en évidence des effets de température sur la mortalité

À première vue, rien ne justifie factuellement l'intégration d'un facteur de température dans le modèle de base. L'objectif de cette section est de montrer la pertinence d'une telle intégration. Pour cela, nous comparons les prédictions issues du modèle de base à la mortalité effectivement observée, en tenant compte des différents cas de température. L'existence d'un écart significatif entre les deux indiquerait un biais structurel dans le modèle, traduisant un effet de la température non pris en compte.

Cette comparaison s'appuie sur le jeu de données désagrégé, dans lequel la température est explicitement renseignée.

Cependant, l'introduction de la dimension "température" rend la visualisation des écarts plus complexe. Pour y remédier, une agrégation par année a été faite. Cela permet d'éliminer la dimension temporelle et conserver uniquement les facteurs "âge" et "température". Cette simplification facilite l'analyse des écarts globaux mesurés par les résidus de déviance. Leur structure est représentée sur la Figure 4.19.

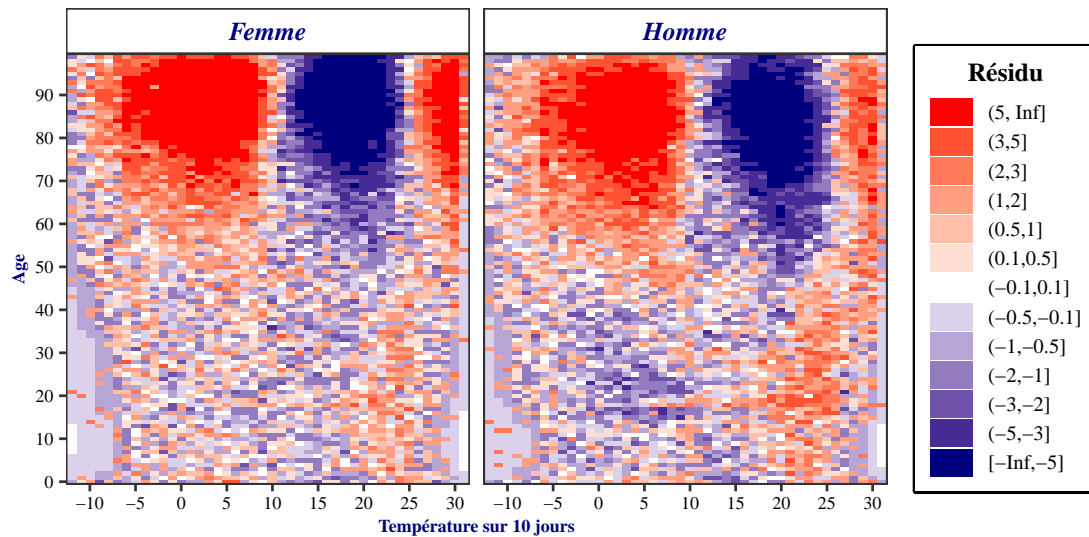


Figure 4.19. : Structure des résidus du modèle selon la température

Ces résidus s'écartent clairement d'une répartition aléatoire et révèlent plusieurs effets :

- Le modèle produit de **mauvaises estimations de la mortalité des personnes âgées** en fonction des différents cas de température. Il sous-estime leur mortalité lorsque la température moyenne sur 10 jours est inférieure à environ 9 °C ou supérieure à environ 25 °C et la surestime lorsque les températures se situent entre 10 °C et 24 °C environ. Cette observation, commune aux deux sexes, confirme l'existence d'une surmortalité aux températures extrêmes, qu'il s'agisse de froid comme de chaleur.
- Le modèle montre un **ajustement global plus précis aux âges jeunes** (jusqu'à environ 50 ans), et ce pour les deux sexes. L'effet de la température est moins marqué dans cette tranche d'âge, ce qui rejoint l'idée selon laquelle les personnes âgées, plus fragiles, sont davantage affectées. Toutefois, une surmortalité apparaît chez les jeunes pour des températures élevées, au delà d'environ 20 °C sur 10 jours, ce qui suggère que les canicules les exposent également à un risque accru. Néanmoins, cette explication reste limitée, car elle ne justifie pas l'absence de surmortalité au delà de 26°C, une plage pourtant encore plus chaude. Cet aspect sera approfondi dans le chapitre suivant.

La présence de cette structure dans les résidus confirme la présence d'un effet résiduel de température non capté par le modèle actuel.

Cet effet peut être caractérisé en comparant les écarts entre les prédictions globales de mortalité et les décès réellement observés, pour chaque niveau de température, tous âges confondus. Cet écart est mesuré par un nouvel indice de mortalité par température défini comme suit :

$$\text{Ratio mortalité}_t = \frac{\text{Nombre de décès observés à la température } t}{\text{Nombre de décès attendus à la température } t}$$

Ce ratio de mortalité exprime directement la surmortalité ou de la sous-mortalité observée. Son calcul confirme que le modèle classique surestime la mortalité observée entre 10 °C et 25°C de température moyenne sur 10 jours, et la sous-estime en dehors de cette plage. L'effet de la température présente une forme régulière avec une valeur de température ambiante appelée TMM (Température de Mortalité Minimale) à laquelle la mortalité est la plus faible. Cette forme en U est illustrée sur la Figure 4.20, où le ratio de mortalité présenté est rapporté à sa valeur à 20°C (TMM). La forme observée correspond à celle décrite par Santé publique France (2004) : une mortalité croissante lorsque la température s'éloigne de la TMM, avec une pente raide du côté des fortes températures indiquant un effet rapide, et une pente plus progressive du côté des températures basses traduisant un effet plus lent.

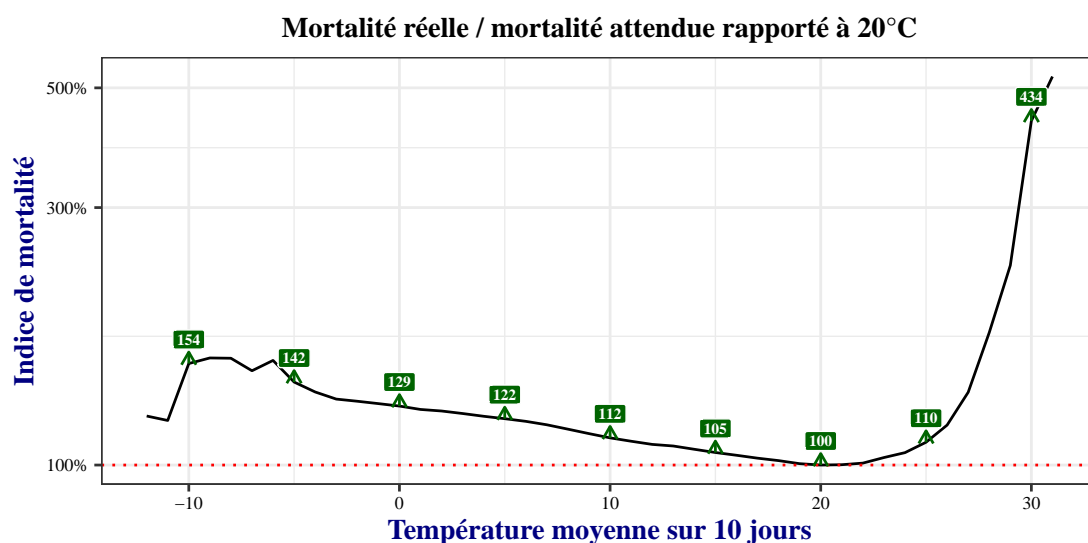


Figure 4.20. : Mise en évidence des effets de température

Ces caractéristiques de la mortalité liée à la température ne sont pas prises en compte par le modèle actuel, ce qui suggère qu'il pourrait être amélioré par l'intégration du facteur de température. L'analyse de ce modèle élargi fera l'objet du chapitre suivant.

Synthèse du chapitre

Ce chapitre a présenté les principales approches de modélisation de la mortalité basées sur les facteurs âge, période et cohorte, qui influencent son évolution. La démarche a été progressive : le modèle Age-Période a mis en évidence la mortalité infantile, l'augmentation log-linéaire de la mortalité avec l'âge, ainsi que les effets de période liés aux progrès médicaux et technologiques. En raison de son faible pouvoir prédictif, il a été généralisé par les modèles de Lee-Carter et APC. Le premier intègre des améliorations de mortalité spécifiques à chaque âge, tandis que second capture les effets propres à chaque génération. Ces extensions offrent de meilleures prédictions que le modèle de base.

Au-delà de ces modèles classiques, deux modèles additifs généralisés (GAM) ont été calibrés. Le premier, proche de Lee-Carter, en reproduit les principaux effets sous une forme lissée et linéaire. Le second, plus flexible, s'adapte mieux aux données et conduit à des prédictions plus précises.

Enfin, l'évaluation de ce dernier modèle dans différentes conditions de température a révélé un effet résiduel de la température, non pris en compte par le modèle. Cet effet, qui traduit une surmortalité aux températures extrêmes, suggère la nécessité d'intégrer explicitement la température comme facteur explicatif afin d'améliorer la modélisation.

Chapter 5.

Résultat de la modélisation intégrant la température

L'objectif de cette partie est d'apporter une contribution supplémentaire en intégrant un facteur de température au cadre classique de modélisation de la mortalité, qui prend principalement en compte les facteurs d'âge, de période calendaire et, éventuellement de cohorte. Cette intégration est justifiée par la présence d'une structure de température dans les résidus du modèle GAM de référence qui ne l'intègre pas. Ce chapitre se déroulera en trois temps : premièrement, une intégration simpliste via un facteur additif de température, supposant que celle-ci affecte tous les individus de la même façon; deuxièmement, une approche plus fine prenant en compte une interaction entre la température et l'âge, conduisant à des effets de température différenciés par âge; et en dernier lieu, une analyse du modèle retenu afin de décrire son comportement et d'évaluer les éventuelles limites.

5.1. Intégration d'un effet additif de la température

La façon la plus simple et naturel de prendre en compte la température dans le modèle consiste à l'intégrer de façon additive, de la même façon que le modèle APC qui combine linéairement les effets de ses composantes. Cette intégration conduit à spécifier le modèle de l'Équation 5.1 :

$$\log(\mu_{xyt}^s) = f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y) + f_4^s(t) + \epsilon_{xyt}^s \quad (5.1)$$

où

- x, y, s et t sont respectivement l'âge, la période calendaire, le sexe et la température
- f_1^s est l'effet structurel de l'âge pour le sexe s
- f_2^s est la composante tendancielle de la période calendaire pour le sexe s
- $k_\epsilon^s(y)$ est l'effet aléatoire spécifique à la période y
- f_3^s est le terme d'interaction entre l'âge et l'année pour le sexe s
- f_4^s est l'effet commun de la température pour le sexe s

- ϵ_{xy}^s est le terme d'erreur aléatoire

Cette spécification reprend la structure de base du modèle de référence, à la différence qu'elle intègre un effet additionnel correspondant à une fonction lisse de température. L'estimation des paramètres se fait de la même manière que dans le cas classique, mais à partir des données désagrégées détaillant les décès et l'exposition par âge, année, et température.

Effet de la température

Après estimation, les effets des facteurs âge et période conservent la même forme que dans le modèle de base (Figure 5.1), ce qui signifie que l'ajout de la température n'a pas perturbé leur structure. Le facteur température se place ainsi comme une dimension orthogonale et complémentaire.

L'effet de la température est conforme à ce qui était attendu d'après l'analyse des écarts du modèle initial selon les conditions de températures. Globalement, pour les hommes comme pour les femmes, la relation entre mortalité et température suit effectivement une courbe en U. La mortalité est minimale à la température optimale (TMM), située autour de 20 °C, puis augmente progressivement lorsqu'on s'en éloigne. On note une dissymétrie entre les effets du froid et ceux de la chaleur.

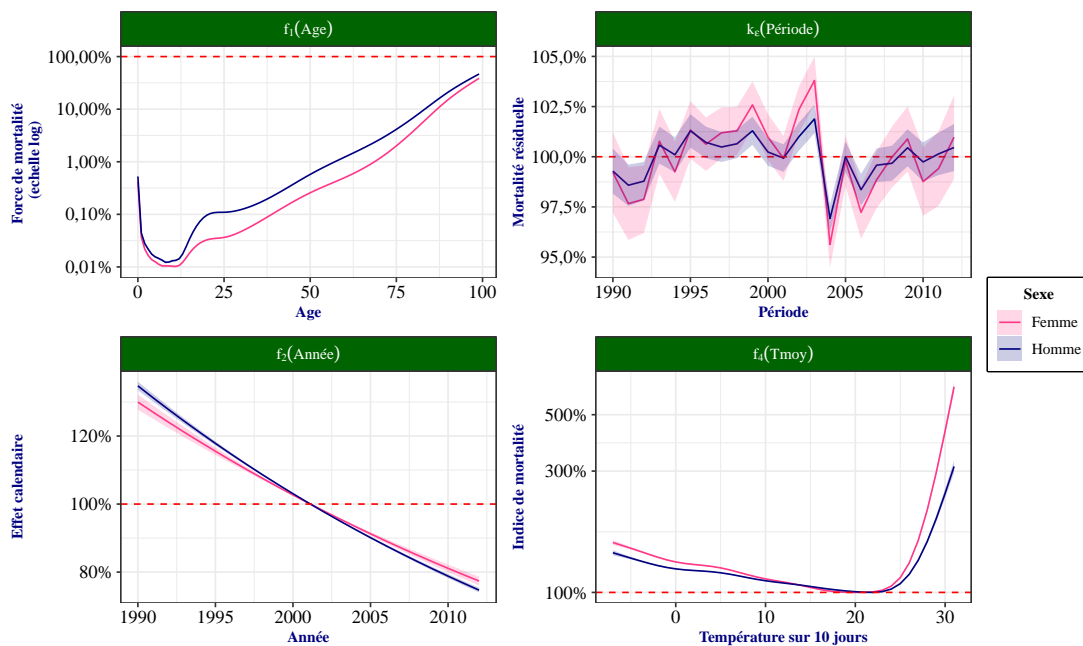


Figure 5.1. : Les effets du modèle avec la température

La chaleur a un impact plus marqué, au delà de la TMM, la mortalité croît rapidement, de façon quasi exponentielle. En revanche, le froid a un effet plus modéré, avec une évolution linéaire de la mortalité pour chaque degré en dessous de la TMM. De plus, l'intensité de ces effets varie selon le sexe : chez les femmes, les pentes sont plus fortes, ce qui traduit une sensibilité accrue aux températures extrêmes et une surmortalité plus importante par rapport aux hommes.

Qualité d'ajustement aux données

Les effets introduits sont désormais intégrés dans la version actuelle du modèle. À l'échelle agrégée, la structure des résidus reste comparable à celle du modèle de référence, ce qui suggère un ajustement global de qualité similaire.

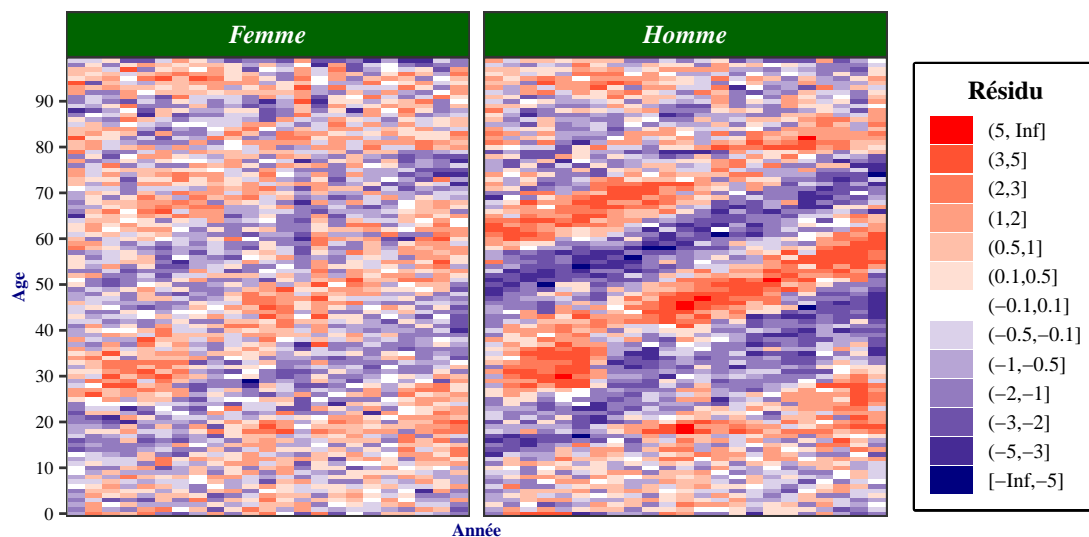


Figure 5.2. : Résidus par année du modèle avec température additive

En revanche, à l'échelle désagrégée selon les températures, l'analyse des résidus est essentielle pour vérifier l'amélioration effective de la qualité de prédiction. Par rapport au modèle de référence, la prise en compte de la température a modifié de façon notable la structure des résidus selon les températures. comme l'illustre la figure Figure 5.3.

Dans l'ensemble, l'intensité des biais a significativement diminué, ce qui indique une meilleure prise en compte des effets de la température dans le modèle. Toutefois, certains biais persistent :

- Au départ, les écarts concernaient surtout les populations âgées de 60 an et plus. Ils se concentrent désormais principalement après 80 ans.

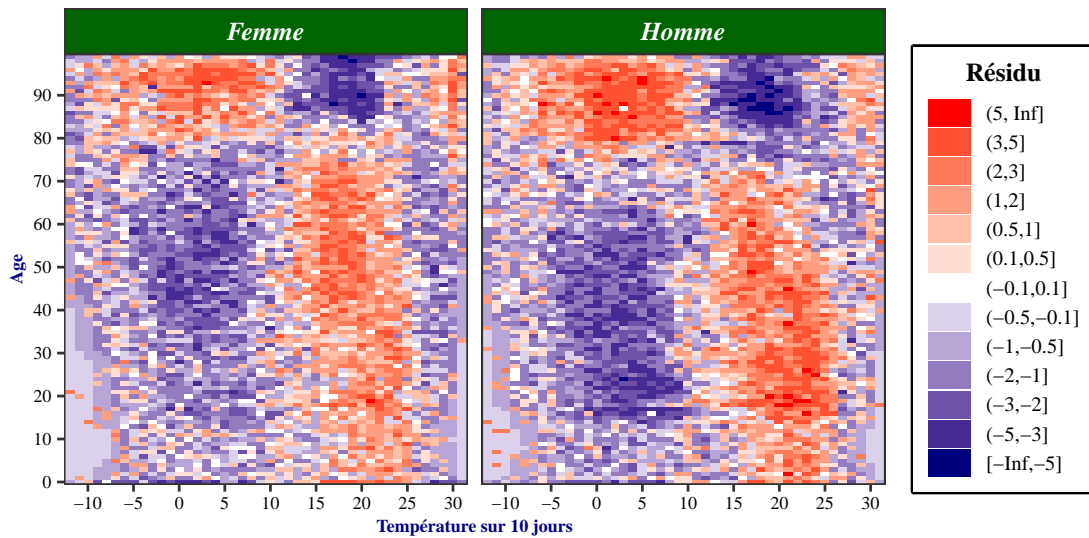


Figure 5.3. : Résidus de deviance du modèle avec température additive

- Un nouveau biais apparaît pour les jeunes : leur mortalité tend à être surestimée en dessous de 10°C et sous-estimée entre 10°C et 25°C.

Ce nouveau biais suggère que l'effet de la température varie selon l'âge. L'amélioration de la précision du modèle pour les personnes âgées s'accompagne d'une dégradation des estimations des populations les plus jeunes. Ce constat rejoint les résultats de la littérature indiquant la plus forte sensibilité des personnes âgées face aux températures extrêmes, ainsi qu'un risque plus faible pour les jeunes.

Pour tenir compte de cette hétérogénéité, il est nécessaire d'enrichir le modèle en intégrant un effet d'interaction entre l'âge et la température. Cela permettrait de modéliser plus finement l'effet différencié de la température selon les classes d'âge.

5.2. Intégration de l'interaction Age-Température

L'effet différencié de la température selon l'âge peut être pris en compte en ajoutant une fonction lisse à deux variables, dépendant à la fois de l'âge et de la température. Cette fonction est calibrée pour reproduire au mieux la structure conjointe de la mortalité en fonction de ces deux dimensions. Le modèle intégrant cette interaction lisse est représenté par l'Équation 5.2 :

$$\log(\mu_{xyt}^s) = f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y) + f_4^s(t) + f_5^s(x, t) + \epsilon_{xyt}^s \quad (5.2)$$

où $f_5^s(x, t)$ représente l'effet d'interaction entre l'âge et la température. La signification des autres paramètres restant identique à celle du modèle précédent.

Les résultats obtenus montrent que la prise en compte de l'interaction age-température préserve la stabilité du modèle initial : les effets des variables de départ conserve leur allure générale (Figure 5.4). Cela suggère que cette interaction constitue un effet supplémentaire, non corrélé aux facteurs déjà modélisés.

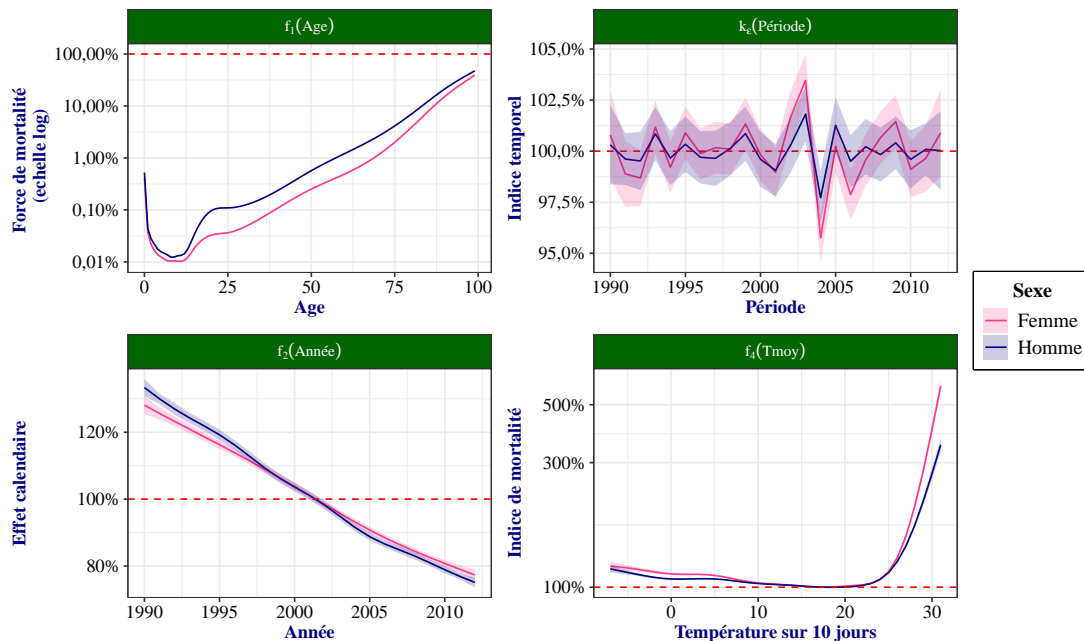


Figure 5.4. : Les effets du modèle avec la température

Effet de la température selon les âges

Pour analyser cet effet de manière simple, on introduit un indice multiplicatif qui résume l'impact différencié de la température par rapport à la mortalité prédite par le modèle de référence sans température, ceci pour chaque combinaison d'âge et de température. Il est analogue aux indices périodiques et générationnels décrits au chapitre précédent.

Notons $\mu_{xy}^{ref} = \exp(f_1^s(x) + f_2^s(y) + k_\epsilon^s(y) + f_3^s(x, y))$ correspondant à la force de mortalité estimée par le modèle de référence (sans température), et en notons l'indice multiplicatif par $Indice_{xt} = \exp(f_4(t) + f_5(x, t))$ représentant l'effet global de la température par age.

La force de mortalité du modèle avec interaction age-température vérifie alors :

$$\mu_{xyt} = \mu_{xy}^{ref} . Indice_{xt}$$

Ce qui fournit une interprétation directe de l'interaction estimée.

Les résultats par combinaison d'âges et de températures (Figure 5.5) confirment que l'effet de la température varie avec l'âge :

- L'impact de la température froide (en dessous de 9°C) est modéré pour les moins de 50 ans environ. Pour les âges plus élevés, la mortalité croît avec l'intensité du froid, et cette croissance est d'autant plus forte que l'âge est avancé.
- La mortalité est plus faible aux températures ambiantes (entre 9°C et 20-24°C), quel que soit l'âge. Elle est particulièrement basse pour les personnes âgées, il s'agit de leur plage de mortalité minimale.
- Au delà de 24°C environ, l'augmentation de la température conduit à une hausse exponentielle de la mortalité. Cette hausse marquée affecte tous les âges et les deux sexes, avec une intensité particulièrement forte pour les femmes.

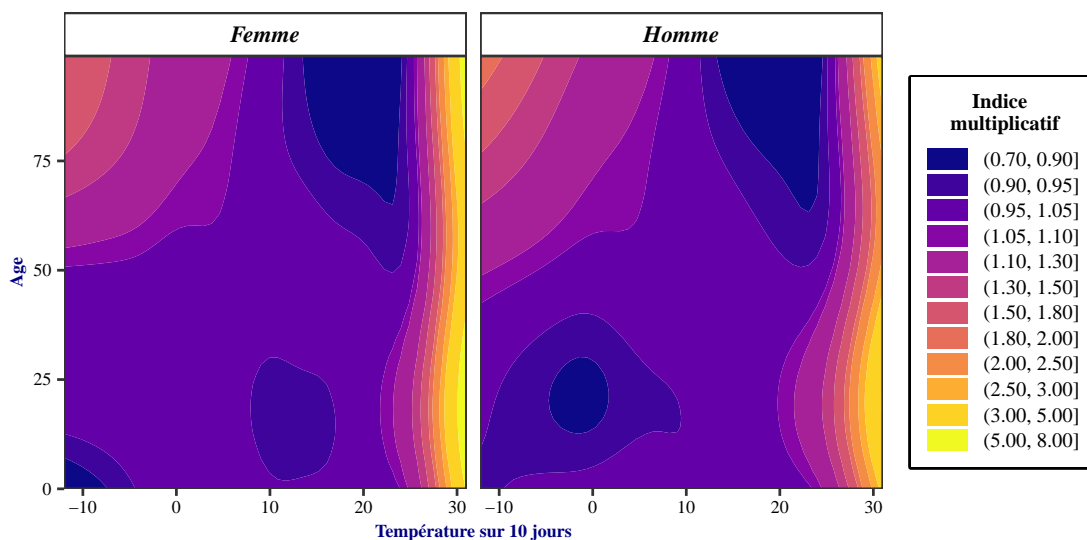


Figure 5.5. : Effet de la température

La surmortalité juvénile estivale

Sur la plage de températures élevées, on observe des situations où l'impact de la chaleur est plus fort chez les jeunes que chez les personnes âgées. Intuitivement, on s'attendrait à l'inverse, en raison de la plus grande fragilité des âges élevés. Cette "surmortalité juvénile" peut s'expliquer par la saisonnalité de la mortalité. En effet, comme cela a

été vue au cours de l'analyse descriptive (Figure 2.15), la mortalité est globalement plus faible en été et plus élevée en hiver. Cette dynamique est surtout portée par les personnes âgées, qui sont plus nombreuses. En revanche, la mortalité des jeunes présente une tendance inverse : plus forte en période estivale et moins forte en période hivernale.

De plus, même si le modèle n'intègre pas explicitement un facteur saisonnier, la température (d'autant plus la température moyennée sur plusieurs jours) est fortement corrélée aux saisons comme l'a montré la Figure 2.14 : en été, elle est en moyenne autour de 20 °C et en hiver elle est de l'ordre de 5°C en hiver. Par cette différence, il peut être considéré que les températures élevées caractérisent principalement l'été et celles faibles caractérisent l'hiver. Pour simplifier et pousser l'analyse, considérons que des températures moyennes sur 10 jours inférieures à 8 °C correspondent à des hivers et que celles supérieures à 20 °C correspondent à des étés.

Pour chacun de ces groupes, en comparant la mortalité "normale" attendue (calculée par âge, toutes périodes et températures confondues) à la mortalité saisonnière (calculée par âge en distinguant les deux plages de température), le contraste de mortalité apparaît nettement (Figure 5.6) :

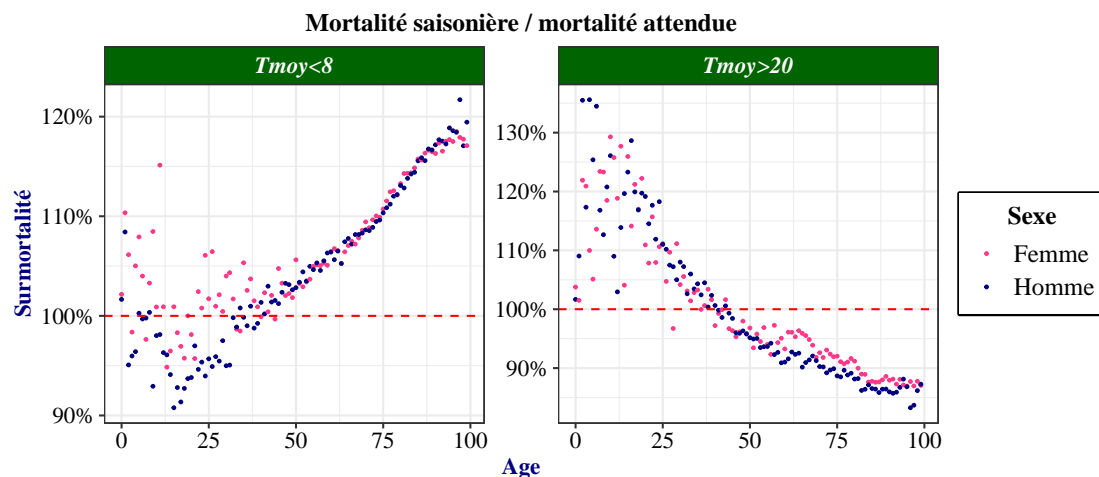


Figure 5.6. : Contraste de surmortalité estivale et hivernale

En hiver, il se produit une surmortalité croissante des personnes âgées de 43 ans environ et plus, qui peut atteindre 20 % pour les plus âgés, tandis que la mortalité des jeunes baisse, surtout pour les hommes.

En été, la tendance s'inverse et présente une surmortalité croissante chez les jeunes de moins de 43 ans pouvant atteindre 30 % pour les plus jeunes, tandis que la mortalité des personnes âgées devient plus faible.

De ce fait, la "surmortalité juvénile" mise en évidence par le modèle peut donc être attribuée à cette surmortalité estivale, qui se manifeste en moyenne au-delà de 20 °C. Le

modèle ne comportant pas de facteur de saison, cet effet est capté par la température, fortement corrélée à la saisonnalité.

La surmortalité exponentielle des températures caniculaires

Une attention particulier peut être portée aux températures élevées au delà de 25 °C, associées à des mortalités exceptionnelles. Pour une température moyenne sur 10 jours de 30°C, la mortalité est de 5 à 8 fois plus importantes que la mortalité habituelle. Cette intensité extrême correspond à celle observée lors de la canicule de 2003, qui est la seule année sur toute la période où une telle intensité a été atteinte. Cette surmortalité doit donc être considérée comme exceptionnelle correspondant aux canicules les plus extrêmes.

La Figure 5.7 montre l'exposition par température sur l'ensemble des années : les températures sur 10 jours de 26 °C et plus sont très rares. Celles de 27°C ont été observées 11 fois, 7 fois pour celles de 28 °C, tandis que les températures au delà de 28 degrés ont été observées une seule fois, au cours de l'année 2003. On note aussi que, sauf 2021, les années récentes présentent toutes des températures 27°C, illustrant de plus le réchauffement climatique progressif.

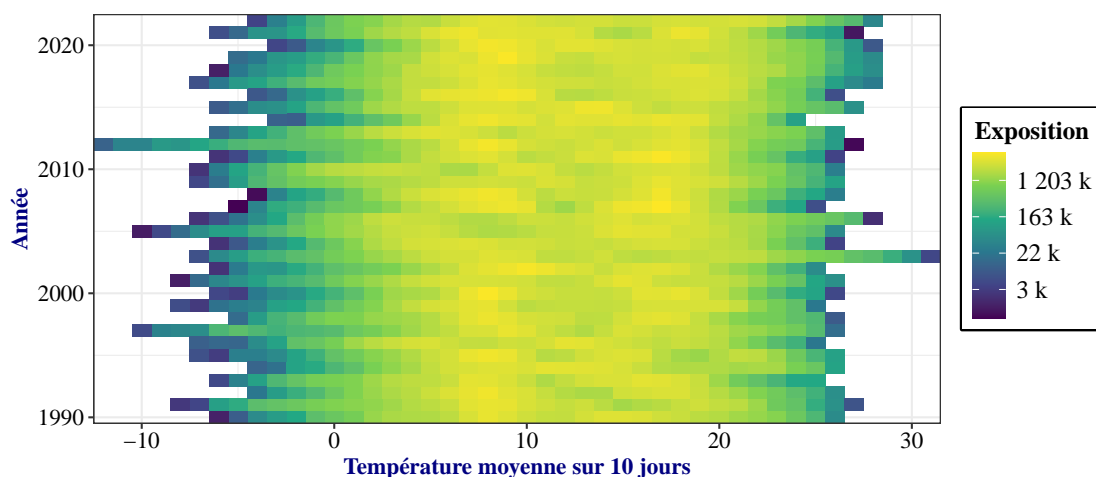


Figure 5.7. : Exposition par température et année

Qualité d'ajustement aux données

Après la prise en compte de l'interaction âge-température, le modèle semble avoir capté l'essentiel des effets relatifs à la température. La répartition des résidus de déviance selon la température ne présente plus aucune structure marquée, quel que soit l'âge considéré (Figure 5.8), ce qui suggère un bon ajustement.

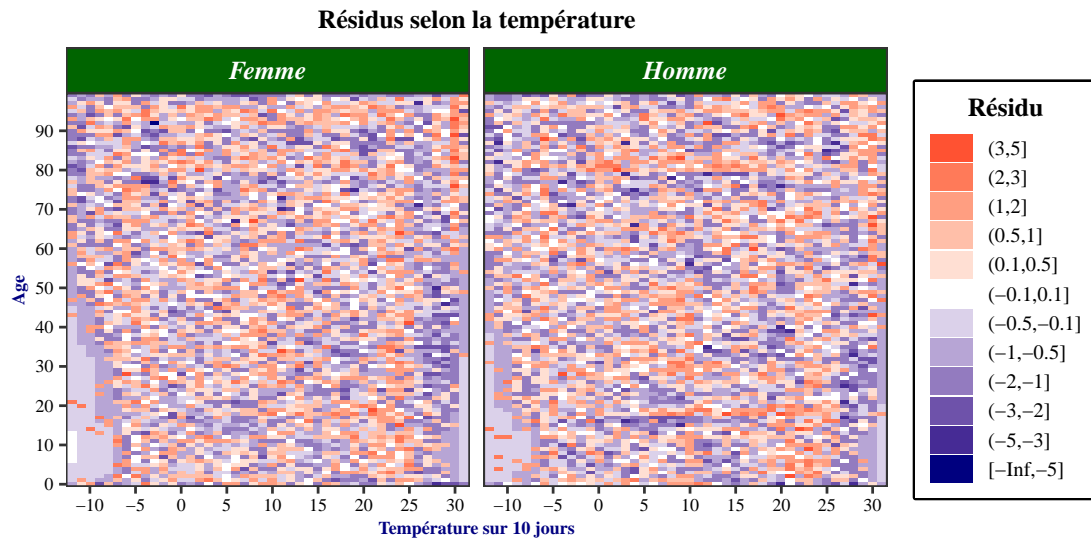


Figure 5.8. : Résidus de deviance du modèle d'interaction age-température

□ Ajustement des forces de mortalité

Comme pour les modèles classiques, il est question de comparer les mortalités prédites et réelles. Cependant, en raison de la prise en compte d'une dimension supplémentaire, il faudra distinguer les différents cas de température. Cette analyse est illustrative et ne s'appuie que sur nombre restreint de valeurs de température afin de conserver une visualisation claire. De plus, contrairement à la mortalité à l'échelle agrégée, à l'échelle désagrégée les forces de mortalité brutes deviennent très volatiles, même aux grands âges, car en distinguant les températures les nombres de décès observés deviennent faibles. Pour limiter ces fluctuations, on privilégie des âges élevés et des températures relativement fréquentes.

Dans l'ensemble, pour les âges et les températures considérés, le modèle s'ajuste bien et suit la tendance centrale des mortalités brutes, tant chez les hommes (Figure 5.9) que chez les femmes (Figure 5.10).

□ Ajustement de la mortalité mensuelle

Une autre évaluation pertinente consiste à apprécier la qualité du modèle face à la mortalité intra-annuelle. Puisqu'il s'appuie sur les températures par jour, le modèle est donc susceptible de produire de bonnes prédictions à cette échelle. En raison du grand nombre de jours sur la période, il est plutôt question de confronter la mortalité mensuelle. De plus, afin de situer la valeur ajoutée du facteur température, on confronte également ces résultats à ceux du modèle sans température.

Avec pour indicateur le nombre mensuel de décès, le modèle avec température reproduit bien les tendances de la mortalité mensuelle pour la quasi totalité des mois

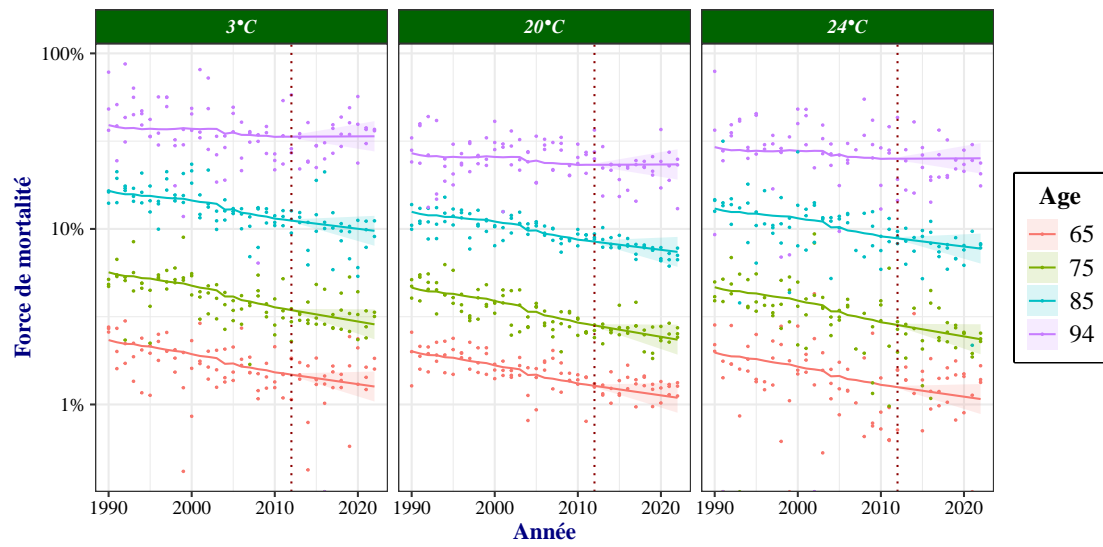


Figure 5.9. : Force de mortalité du modèle de température (Homme)

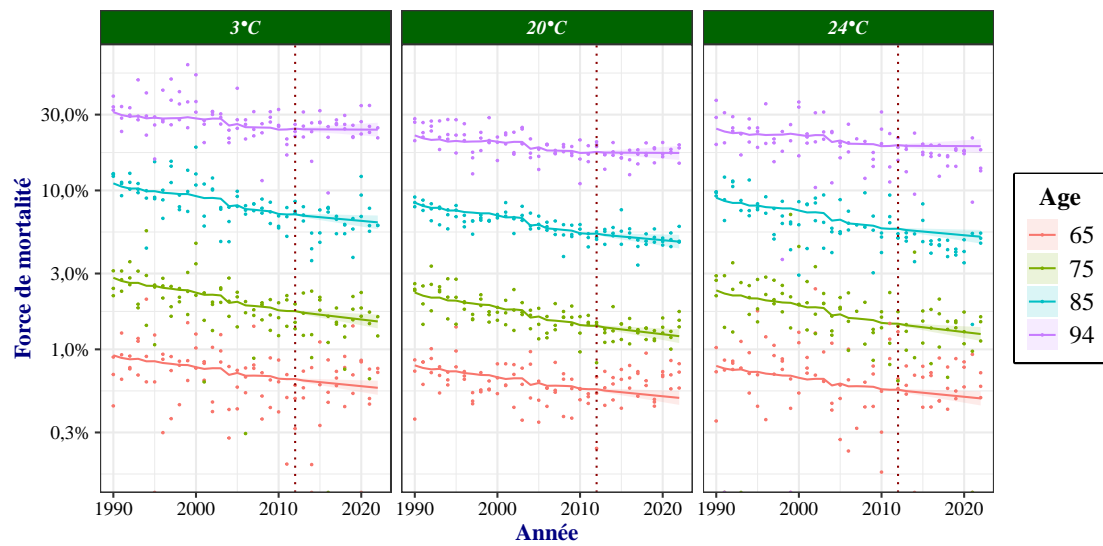


Figure 5.10. : Force de mortalité du modèle de température (Femme)

: les prédictions sont proches des observations réelles, les écarts les plus importants surviennent aux mois de janvier à mars qui présentent une forte volatilité par rapport aux autres mois. À l'inverse, les mois d'étés présentent une meilleure précision. Notamment en 2003, le modèle avec température réussi à reproduire la mortalité d'août due à la canicule, qui est non captée par le modèle de référence.

Ces résultats montrent que l'intégration de la température permet de capter le caractère saisonnier de la mortalité qui varie selon les mois et les saisons. À l'inverse, le modèle sans température présente des biais attendus : forte sous-estimation en hiver (décembre–mars) et surestimation en été (mai–septembre), avec une meilleure précision en avril, octobre et novembre. Les conclusions sont similaires pour les femmes (Figure 5.11) et pour les hommes (Figure 5.12), à la différence que l'incertitude prédite par le modèle avec température est plus élevée chez les hommes sur la période future.

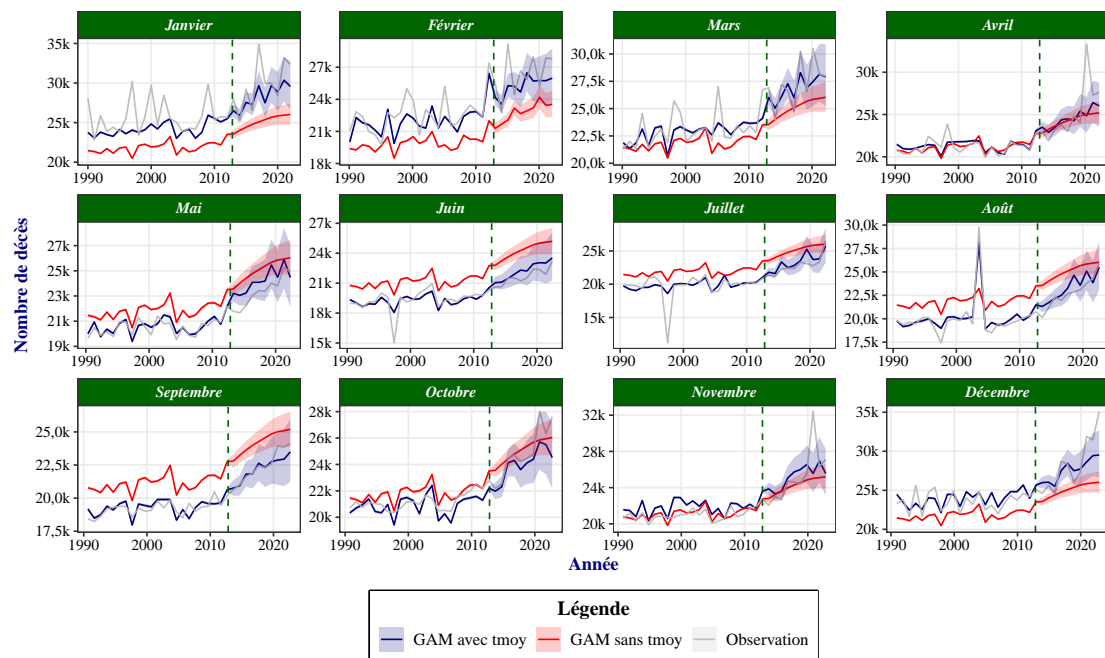


Figure 5.11. : Comparaison de la mortalité mensuelle (Femme)

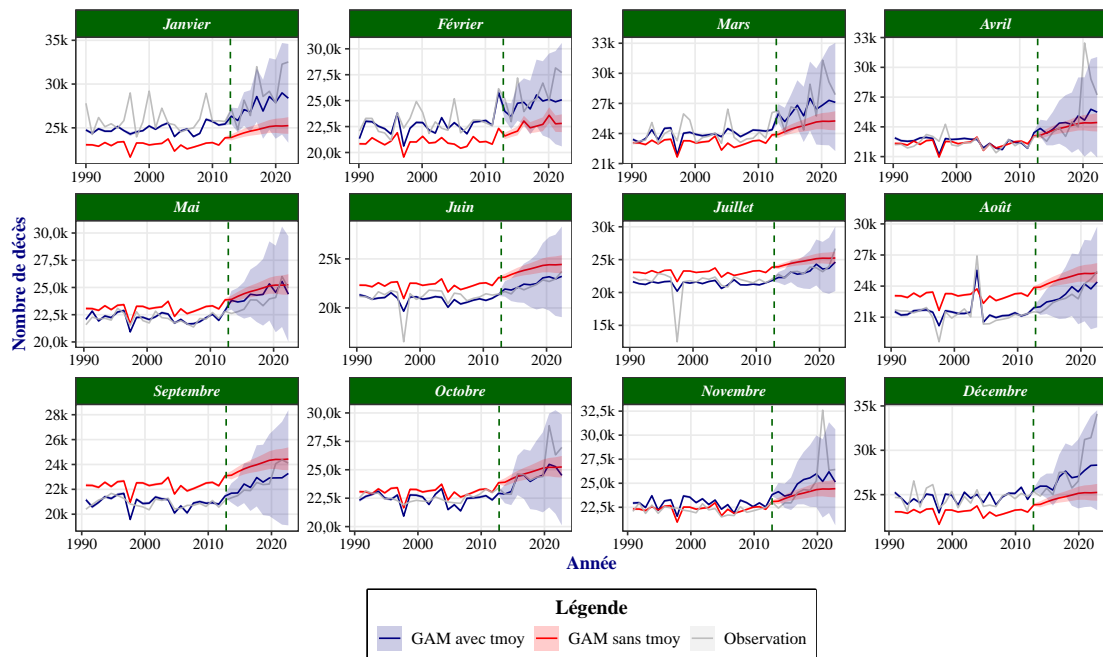


Figure 5.12. : Comparaison de la mortalité mensuelle (Homme)

Les performances du modèle

Il est question de chiffrer la qualité prédictive du modèle de température par rapport au modèle de référence. Pour ce faire, les comparaisons sont faites à la fois à l'échelle désagrégée (par température) et à l'échelle agrégée (mortalité annuelle par âge et sexe), et ce sur les périodes d'entraînement et de test.

□ Performances en période d'entraînement

À l'échelle désagrégée, le modèle sans température suppose une force de mortalité identique quel que soit la température, par conséquent, on peut s'attendre à ce que ses prédictions par température soient moins bonnes. En pratique, c'est bien ce qui est observé : pour les hommes (Table 5.2) comme pour les femmes (Table 5.1), le modèle avec température est meilleur supérieur selon tous les critères de performances observés.

Notons que la MAPE peut être très élevée à cette échelle en raison des faibles effectifs, rendant les proportions instables : cet indicateur doit être interprété avec prudence.

Table 5.1. : Comparaison des performances désagrégées en train (Femme)

Modèle	Déviance	GCV	AIC	BIC	RMSE	MAPE	EDF
GAM sans température	283 411,8	1,4	960 726,4	961 313,6	8,8	66,4	57,4

GAM avec température	223 743,4	1,1	901 107,4	901 946,6	6,4	63,1	82,0
----------------------	-----------	-----	-----------	-----------	-----	------	------

Table 5.2. : Comparaison des performances désagrégées en train (Homme)

Modèle	Déviance	GCV	AIC	BIC	RMSE	MAPE	EDF
GAM sans température	271 247,7	1,3	1 030 238,0	1 030 823,3	7,7	54,9	57,2
GAM avec température	231 328,7	1,1	990 369,3	991 212,0	6,4	52,0	82,4

À l'échelle de la mortalité annuelle, les nombres de décès sont obtenus par agrégation des décès prédits par cas de température. Cette échelle a l'avantage d'être comparable aux résultats des modèles classiques. En termes de déviance et de MAPE, on retrouve un meilleur ajustement du modèle avec température, traduisant une estimation plus précise du nombre annuel de décès après prise en compte de la température. Les tables 5.3 et 5.4 présentent ces résultats respectivement pour les populations de femmes et d'hommes.

Table 5.3. : Comparaison des performances annuelles en train (Femme)

Modèle	Déviance	GCV	AIC	BIC	RMSE	MAPE	EDF
GAM sans température	3 977,6	1,8	23 576,1	23 905,6	67,2	5,3	57,4
GAM avec température	3 939,0	1,8	23 586,8	24 057,7	66,4	5,3	82,0

Table 5.4. : Comparaison des performances annuelles en train (Homme)

Modèle	Déviance	GCV	AIC	BIC	RMSE	MAPE	EDF
GAM sans température	7 799,4	3,6	28 375,2	28 703,7	97,2	5,0	57,2
GAM avec température	7 783,3	3,6	28 409,3	28 882,2	96,8	5,0	82,4

□ Performances sur la période future

Sur la période future, on applique la même démarche. Comme au chapitre précédent, les indicateurs pénalisant la vraisemblance ne sont pas retenus : on cherche la justesse des projections, non leur parcimonie.

A l'échelle désagrégée, les résultats confirment ceux de la période d'entraînement : le modèle avec température est plus performant en déviance et en MAPE (encore, à prendre avec précaution). Ces résultats sont présentés par les tables 5.5 et 5.6, respectivement pour les femmes et les hommes.

Table 5.5. : Comparaison des performances désagrégées en test (Femme)

Modèle	Déviance	MAPE	EDF
GAM sans température	135 199,0	79,0	57,4

GAM avec température	104 362,3	73,8	82,0
----------------------	-----------	------	------

Table 5.6. : Comparaison des performances désagrégées en test (Homme)

Modèle	Déviante	MAPE	EDF
GAM sans température	136 552,3	67,6	57,2
GAM avec température	109 118,8	63,1	82,4

À l'échelle annuelle, les meilleures performances sur la période future sont également obtenues avec les modèles intégrant la température (tables 5.7 et 5.8).

Table 5.7. : Comparaison des performances annuelles en test (Femme)

Modèle	Déviante	MAPE	EDF
GAM sans température	13 924,2	14,1	57,4
GAM avec température	12 624,8	13,2	82,0

Table 5.8. : Comparaison des performances annuelles en test (Homme)

Modèle	Déviante	MAPE	EDF
GAM sans température	20 737,5	13,3	57,2
GAM avec température	17 163,4	11,9	82,4

En somme, cette comparaison montre que la prise en compte de la température apporte une réelle plus-value : elle conduit à des estimations plus justes de la mortalité que le modèle de référence fondé sur la modélisation par âge, période et la cohorte. Cette surperformance a toutefois un coût, elle nécessite davantage de paramètres, une structure plus complexe et une architecture de données plus lourde nécessitant le traitement et le suivi des évolutions de température ce qui implique plus de stockage et des temps de calcul plus longs mais qui conduit à de meilleures performances.

Illustration des effets de la température

Cette sous-section illustre directement l'influence des variations de température sur la mortalité, d'abord via la courbe de mortalité, puis via l'espérance de vie périodique.

□ Effet de la température sur la courbe de mortalité

La courbe de mortalité présente l'évolution de la mortalité selon les âges. En exploitant le modèle de température, il est question de la construire pour différentes valeurs de température ce qui conduit à des écarts importants. Lorsque la température s'élève, à

28°C et plus encore à 30°C, la courbe est nettement décalée vers le haut et ce davantage chez les femmes que chez les hommes. Ces courbes étant les meilleures estimations conditionnelles à la température, l'écart observé indique qu'un modèle sans température introduirait un biais important en situation de température non ambiante. À l'inverse, pour 10, 20 et 25 °C, les écarts sont plus modérés (Figure 5.13).

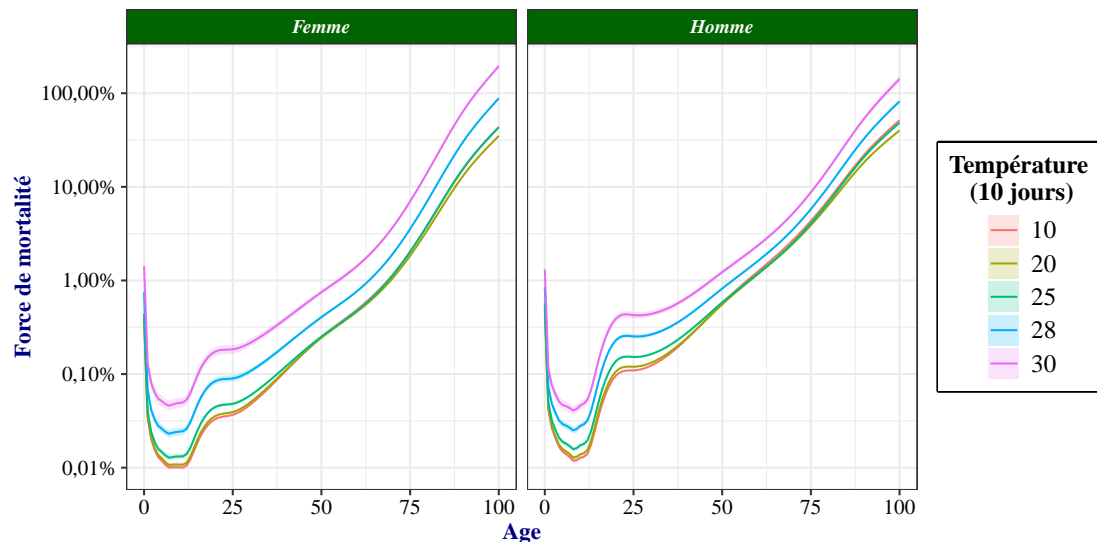


Figure 5.13. : Impact des températures sur la courbe de mortalité

□ Effet de la température sur l'espérance de vie périodique

L'espérance de vie périodique évalue la durée de vie moyenne que vivrait une cohorte fictive qui subirait, à chaque âge, la mortalité observée au cours de la période considérée. Ici, on considère aussi que cette cohorte fictive est exposée toute sa vie à une température constante, afin de mesurer l'impact sur l'espérance de vie. Ces impacts sont également projetés sur les périodes futures. Les résultats obtenus sur la Figure 5.14 confirment l'influence importante des températures extrêmes.

Pour toutes les périodes, l'espérance de vie est maximale à 20 °C, elle diminue peu à 10 °C et diminue plus si la température est de -5 °C. Dans le sens chaud, l'espérance de vie à 26°C est proche de celle à -5°C. Les impacts les plus forts apparaissent aux températures de 28, 29 et 30°C : les écarts d'espérance de vie par rapport à 20°C sont d'environ 7, 10 et 14 ans chez les femmes, et d'environ 6, 8 et 12 ans chez les hommes.

Enfin, quel que soit la température, le modèle décrit une amélioration tendancielle au fil du temps correspondant à un recul progressif de la mortalité même sous les températures extrêmes. Cela peut s'interpréter comme une forme d'acclimation aux températures extrêmes. En effet, en cas de chaleur extrême persistante, des moyens d'adaptation pourraient être mobilisés. Par exemple, l'utilisation de climatiseur serait un recours pour réduire la sensation de chaleur et limiter ces effets. Avec le développement technologique

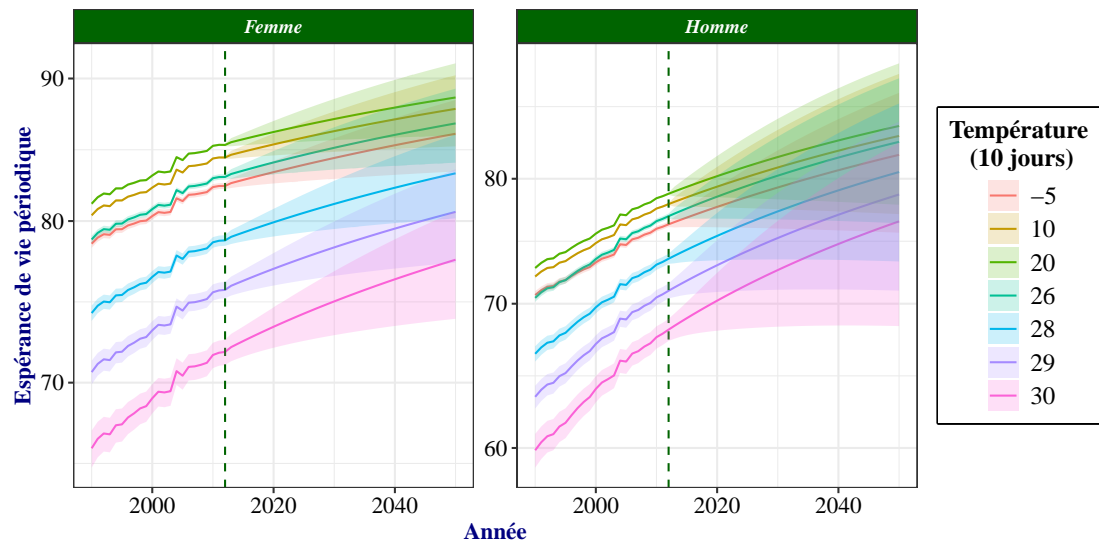


Figure 5.14. : Illustration espérance de vie périodique par température

constant, de telles solutions et des solutions plus innovantes encore sont susceptibles de voir le jour et d'améliorer le confort humain.

5.3. Analyse du nombre de jours optimal

Jusqu'ici le facteur de température intégré dans le modèle construit est une température moyenne sur les 10 jours ayant précédés les décès des personnes observées. La prise en compte d'une température moyenne sur plusieurs jours plutôt que la température moyenne quotidienne a été motivée par les résultats de la littérature, qui mentionnent un effet retardé de la température, en particulier du froid, dont les conséquences peuvent apparaître plusieurs jours après l'exposition (Pincemin, 2021; Salaun, 2024). Le but de cette démarche est de tenir compte des éventuelles températures extrêmes auxquelles les personnes ont pu être exposées à la veille de leur décès.

Dans cette optique, une alternative aurait été d'intégrer plusieurs variable de température pour chaque jour précédant le décès, jusqu'à un nombre maximum de jours fixé. Cependant, la prise en compte de plusieurs variables de température entraîne une augmentation exponentielle de la taille du jeu de données, posant des problèmes de stockage et temps de calcul. C'est pourquoi le choix a été fait de résumer les fluctuations de température en une valeur unique : la moyenne. Toutefois, la question du nombre de jours sur lesquels cette moyenne doit être calculée demeure, et cette sous-section vise à déterminer quel serait le nombre optimal de jours à considérer pour calculer cette moyenne afin de restituer au mieux au mieux les conditions réelle d'exposition.

En effet, ce choix peut jouer un rôle important dans les résultats. Une plage de jours trop courte pourrait sous-estimer les effets du froid, dont l'impact est moins rapide que celui de la chaleur, tandis qu'une plage trop grande pourrait lisser excessivement les fluctuations et conduire à un indicateur peu représentatif des températures effectivement subies.

La détermination du choix optimal sur trois aspects : une comparaison visuelle de la forme de l'effet de la température pour une plage de jours fixé, afin d'en évaluer la cohérence; l'examen des capacités prédictives des modèles correspondants et enfin la comparaison de leurs capacités à reproduire les effets de la mortalité observés lors des épisodes extrêmes de température.

Les effets des températures décalées

Pour décrire cet effet de température, le modèle considérant un effet additif commun à tous les individus est repris. Cet effet représente la tendance centrale, tous âges confondus. Ce modèle est calibré pour chaque plage de jours considérée et les effets obtenus sont présentés dans la Figure 5.15. Dans l'ensemble, les effets associés aux décalages jusqu'à 60 jours décrivent une forme en U, avec une forte accélération aux températures élevées et une croissance plus modérée aux températures froides. Au-delà de 60 jours de décalage, l'effet obtenu ne décrit plus une forme en U, ce qui suggère horizon trop éloigné.

Les effets présentés pour les décalages de 1 à 10 jours sont assez similaires, avec environ 500 % de surmortalité en cas de température à 30°C. Ils montrent également une décroissance de la mortalité en deça de -10°C, ce qui est clairement lié à un biais d'échantillonnage, l'exposition devenant très faible à ces niveaux. Pour les températures moyennées sur 15 jours ou plus, l'impact des fortes chaleurs apparaît plus faible : les jours de chaleur ayant été lissés, la moyenne ramène ces valeurs à des niveaux moins extrêmes. Par exemple, sur 60 jours, les températures moyennées vont de -6 à 26 °C, alors qu'elle vont de -16 à 33 °C pour la température moyenne quotidienne. De plus, les températures sur 15, 30 et 60 jours montrent une décroissance accentuée de la mortalité en deça de 0°C ce qui est contre-intuitif. Seul le décalage sur 20 jours conduit à une forme en U sans décroissance inattendue aux extrêmes froids, mais dans ce cas l'effet des chaleurs est aussi lissé par la moyenne glissante.

Capacité de prédiction et décalage optimal

En termes de capacité prédictive, la comparaison des modèles se fait sur la base des prédictions de la mortalité journalière nationale, en considérant les modèles construits avec chaque décalage de température. L'indicateur retenu est la déviance de poisson, qui mesure les écarts entre les prédictions et les décès observés. La comparaison est réalisée à la fois sur la période d'entraînement et sur les données de test.

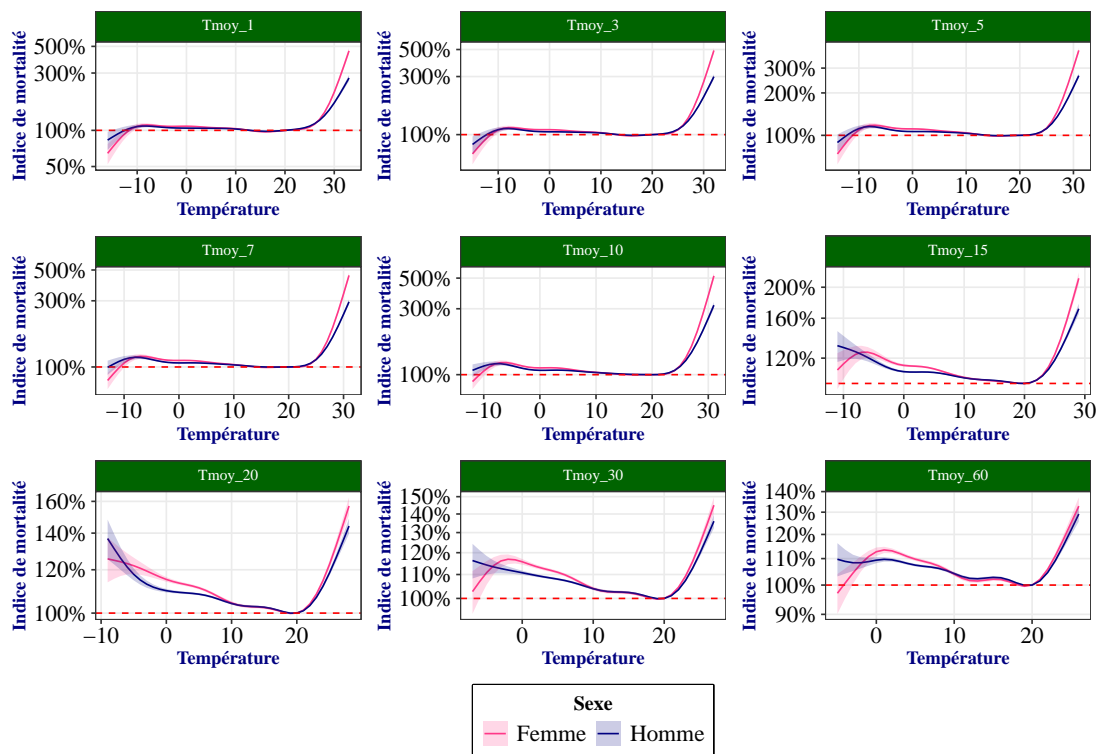


Figure 5.15. : Effet de la température en fonction du nombre de jours décalés

Sur chacun des échantillons, train comme test, les résultats obtenus montrent une forme en U : une déviance importante à la durée 0 correspondant au modèle sans température, puis une décroissance progressive avec l'augmentation du décalage, jusqu'à atteindre une valeur minimale avant de remonter (Figure 5.16).

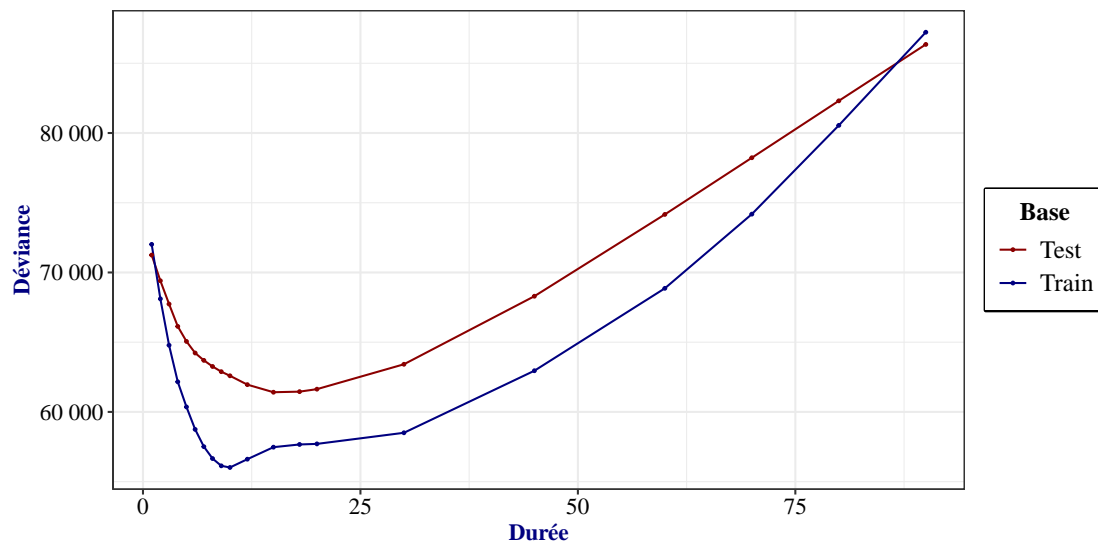


Figure 5.16. : Performances des modèles selon le nombre de jours décalés

Cette forme traduit l'existence d'un décalage optimal pour lequel les prédictions sont les meilleures. Sur la base de ces résultats, le décalage optimal varie selon l'échantillon. Il est de 10 jours en période d'entraînement, traduisant un ajustement optimal aux données pour cette valeur, et de 15 jours en période de test, traduisant que cette valeur conduit à la meilleure capacité de projection.

Afin d'y voir plus clair et retenir un optimum unique, d'autres critères de performance ont été considérés. Ces analyses confirment que les meilleurs décalages se situent entre 10 et 15 jours (Table 5.9).

Table 5.9. : Décalage optimal selon les critères et les périodes

Période	Déviance	RMSE	MAPE	GCV	AIC	BIC
Train	10	9	20	10	10	10
Test	15	18	15	15	15	15

Analyse des prédictions quotidiennes

Dans cette sous section, Il s'agit maintenant de comparer les modèles sur leur capacité à reproduire la surmortalité liée aux épisodes de températures extrêmes, qu'il s'agisse de vagues de froid ou de chaleur. Cette analyse sera surtout visuelle, ce qui permet déjà d'identifier le modèle le mieux ajusté. En tenant compte des analyses précédentes, et

pour simplifier les comparaisons, seuls les modèles intégrant les températures sur 10 et 15 jours sont considérés.

□ Vue d'ensemble

Dans un premier temps, les prédictions quotidiennes sur l'ensemble de la période disponible sont observées. Les résultats montrent un ajustement globalement satisfaisant, car les prédictions des modèles sur 10 comme 15 jours reproduisent nettement les tendances de la mortalité journalière, notamment la saisonnalité, qui a déjà été mise en évidence au pas de temps mensuel. Cela confirme que le modèle est pertinent pour la prédiction de la mortalité à différentes échelles temporelles, même intra-annuelles (Figure 5.17).

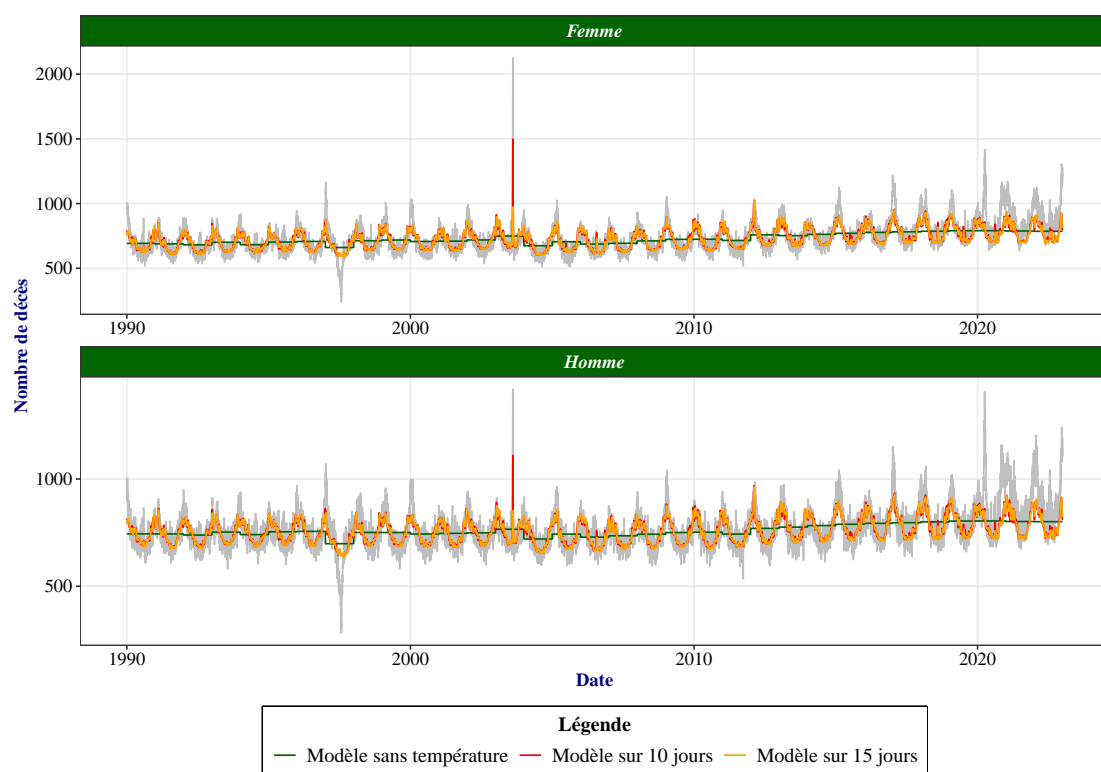


Figure 5.17. : Comparaison des mortalités quotidiennes sur l'ensemble de la période

La comparaison avec le modèle sans température est également faite et aboutit à la même conclusion que pour les analyses mensuelles. Ses prédictions de mortalité sont globalement sans biais, mais sont constantes tout le long de l'année et ne captent pas la saisonnalité.

La courbe des décès quotidiens met en évidence deux pics de surmortalité majeurs, en 2003 et en 2020. Pour 2003, ce pic est partiellement reproduit par les deux modèles, la

surmortalité étant toutefois mieux captée par le modèle sur 10 jours. En 2020, le pic est lié à la Covid-19, cette surmortalité n'ayant pas de lien avec la température elle n'est pas captée par les modèles. Cependant, l'ensemble du graphique reste peu lisible, et ces différences sont difficilement perceptibles. La sous-section suivante se concentre donc sur les périodes correspondant aux épisodes les plus extrêmes de température.

□ **Zoom sur les années de canicule et grand froid**

Selon Santé Publique France (2025), après la canicule de 2003, l'épisode significatif suivant a eu lieu en 2006 qui fut particulièrement long, les autres particulières remarquables ont eu lieu en 2018 et en 2020. Concernant les vagues de froid, Météo France (2024a) souligne qu'elles deviennent de plus en plus rares et moins intenses avec le réchauffement climatique. La seule vague marquée du 21^e siècle a eu lieu en février 2012, d'une durée de 8 jours.

• **La mortalité de l'année 2003**

Au cours de l'année 2003, la mortalité a suivi une tendance régulière, à l'exception du mois d'août avec la canicule. En cette période de canicule, le pic de décès a atteint 2 000 chez les femmes et 1 400 décès chez les hommes. La mortalité prédite par les modèles est quant à elle de 1 500 décès de femmes et de 1 100 décès d'hommes pour le modèle sur 10 jours et de 1 000 décès de femmes et 850 décès d'hommes pour le modèle sur 15 jours. Le modèle sur 10 jours restitue donc mieux cet épisode (Figure 5.18).

• **La mortalité des autres années de canicule**

En 2006, les modèles identifient efficacement la surmortalité attribuable à la chaleur. Là encore, par rapport au modèle sur 15 jours, le modèle sur 10 jours conduit à une surmortalité plus importante et plus proche du pic réel (Figure A.1 en annexe).

En 2012, la surmortalité observée en été est très brève, tandis qu'en 2018 (Figure A.2 en annexe) un épisode plus marqué est également mieux capté par le modèle sur 10 jours

En 2020, la surmortalité observée en été est également brève, la particularité de cette année est surtout liée à la covid-19 dont les pics exceptionnels de mortalité se produisent en avril et en novembre. Cette surmortalité n'étant pas liée à la température elle n'est pas reflétée par le modèle.

• **La mortalité des années de vagues de froid**

Les seules vagues de froid identifiées sur la période d'étude ont eu lieu en 2012 et en 2018. La mortalité observée à cette période présente bien un pic exceptionnel qui est capté par les deux modèles de température (Figure 5.19). Cette fois, la précision des deux modèles est comparable et ils prédisent chacun une intensité de mortalité proche de la surmortalité réelle observée. Toutefois, la période de surmortalité décrite est plus courte et commence plus tôt que la vague réelle. Ce comportement peut résulter du fait qu'une seule variable de température est utilisée pour représenter à la fois les effets

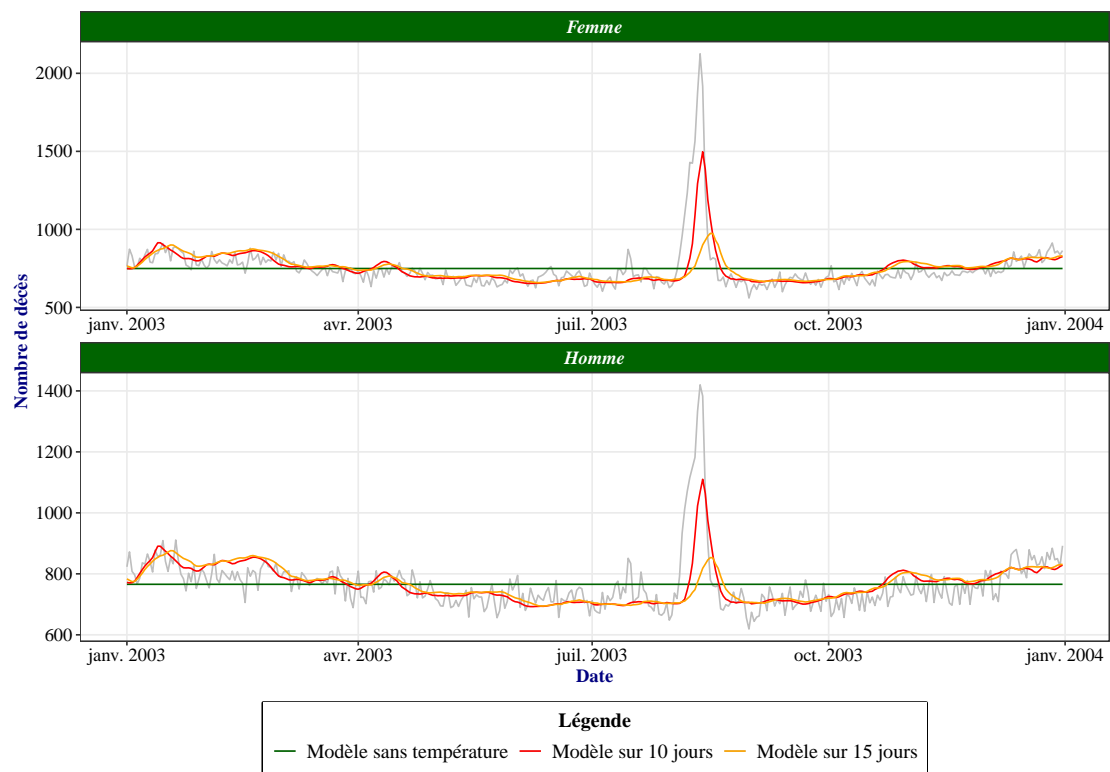


Figure 5.18. : Comparaison des mortalités quotidiennes en 2003

du froid et de la chaleur. L'effet estimé fait alors un compromis entre le délai rapide des températures chaudes et le délai prolongé des températures froides, conduisant à sous-estimer le décalage des vagues de froid et à rallonger celui des vagues de chaleur. Pour ce qui concerne la surmortalité du pic de froid de février 2012, elle est faiblement reproduite par le modèle qui sous-estime son ampleur.

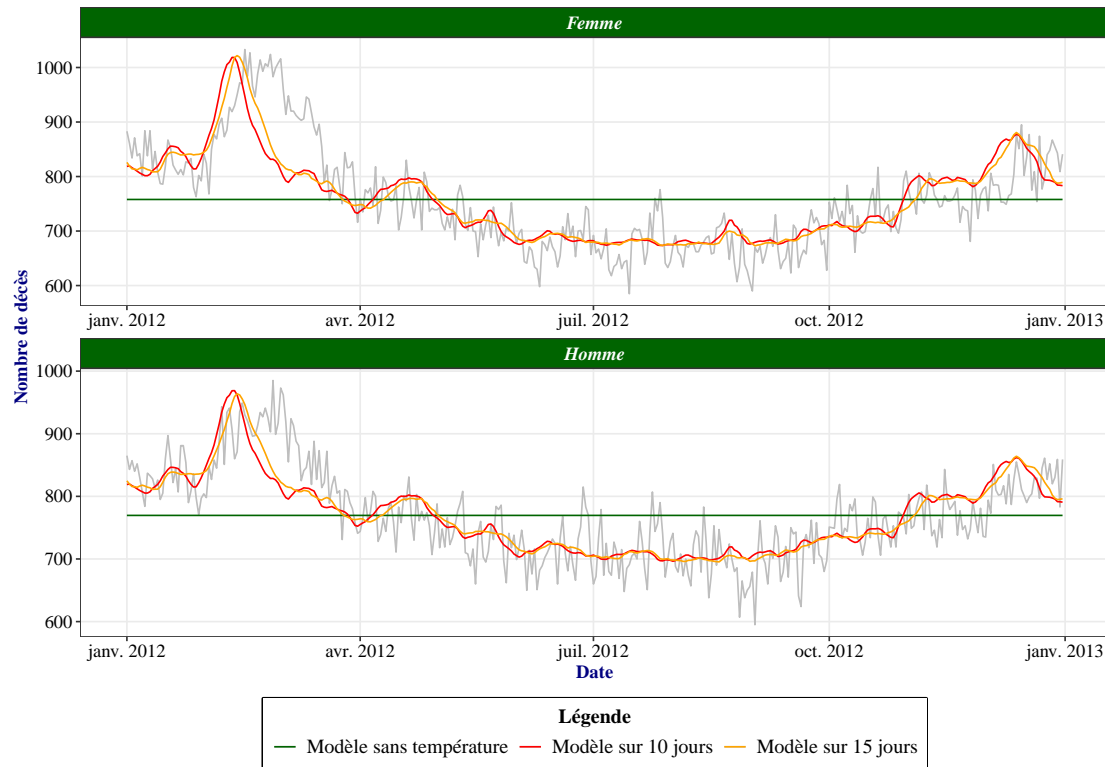


Figure 5.19. : Comparaison des mortalités quotidiennes en 2012

5.4. Limites et Pistes d'amélioration du modèle

Le modèle de mortalité final intégrant des effets différenciés de la température selon les âges s'est globalement montré satisfaisant et a présenté de meilleures performances que les modèles stochastiques usuels en modélisation de la mortalité. Toutefois, plusieurs aspects susceptibles de l'améliorer davantage n'ont pas encore été pris en compte. Cette sous-section vise à dresser un état des lieux des éléments susceptibles d'améliorer les performances liées aux effets de température.

5.4.1. Intégration d'un facteur de saisonnalité

L'analyse menée au cours de cette étude a montré que la prise en compte de la température permet au modèle de reproduire le caractère saisonnier de la mortalité, en raison de la forte corrélation entre saison et température. Cependant, il reste possible qu'un effet saisonnier résiduel traduisant un aspect particulier de la mortalité n'ait pas été capté. Cela justifierait la pertinence d'intégrer explicitement un facteur de saison. De plus, la prise en compte de la saison pourrait améliorer l'interprétation des effets de la température, puisque, comme montré précédemment, celle-ci capte en partie des phénomènes de saisonnalité qui ne lui sont pas directement attribuable, mais qui relèvent d'un effet de confusion.

Pour évaluer la pertinence d'un effet supplémentaire de saisonnalité, la démarche mise en place est identique à celle utilisée pour justifier l'effet de température. Il s'agit de comparer la mortalité saisonnière prédite par le modèle tenant compte de la température à la mortalité réellement observée, et ce pour l'ensemble de la période d'entraînement. La présence d'un écart signifierait un biais saisonnier résiduel non pris en compte par le modèle.

Le résultat indique que le modèle parvient globalement à bien prédire la mortalité saisonnière. Les écarts observés sont faibles, inférieurs à 2%. Les saisons ayant le meilleur ajustement sont l'été et le printemps avec des écarts d'environ 0,5 %, tandis qu'en automne et en hiver ils atteignent respectivement 1,5 % et 2 %. Ces écarts, relativement limités, soutiennent l'idée que les effets de saison sont déjà largement intégrés dans le modèle (Figure 5.20).

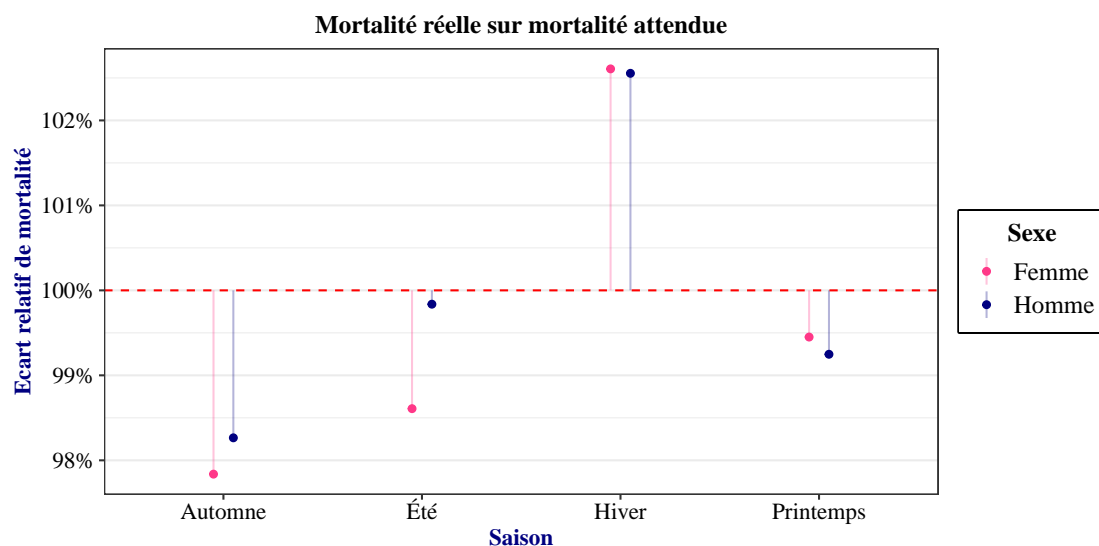


Figure 5.20. : Ecart d'ajustement du modèle suivant les saisons

5.4.2. Effet des zones climatiques

Un autre axe potentiel d'amélioration du modèle réside dans l'intégration d'un facteur lié aux zones climatiques. Celles-ci regroupent des départements présentant des conditions climatiques similaires, en particulier en ce qui concerne la température. Il est donc plausible que l'effet de la température sur la mortalité soit homogène à l'intérieur d'une même zone, mais diffère entre zones distinctes.

Par ailleurs, les disparités régionales peuvent des comportements différents face aux conditions de températures. Le phénomène d'acclimatation implique qu'une même température puisse engendrer des effets sanitaires contrastés selon le contexte climatique local. Dans une zone où cette intensité de température est courante, la population est généralement mieux adaptée, grâce à des comportements, des infrastructures ou des équipements spécifiques. À l'inverse, dans une zone où cette température est inhabituelle, l'absence d'adaptation et de dispositifs de mitigation peut amplifier ses impacts, entraînant potentiellement une surmortalité plus importante

Comme pour les saisons, la pertinence de ce facteur a été évaluée en comparant les décès prédits et observés dans chaque zone climatique. Les résultats montrent que les erreurs du modèle restent globalement faibles, avec des écarts de moins de 1,5 % pour chaque zone. La zone semi-continentale présente une mortalité sous-estimée de 1,5 %, tandis que la mortalité en zone tempérée est surestimée dans les mêmes proportions. En zone méditerranéenne, la mortalité des femmes est légèrement sous-évaluée tandis que celle des hommes est surévaluée. Dans l'ensemble, les écarts demeurent limités et le modèle peut être considéré comme précis même sans prise en compte explicite des zones climatiques (Figure 5.21).

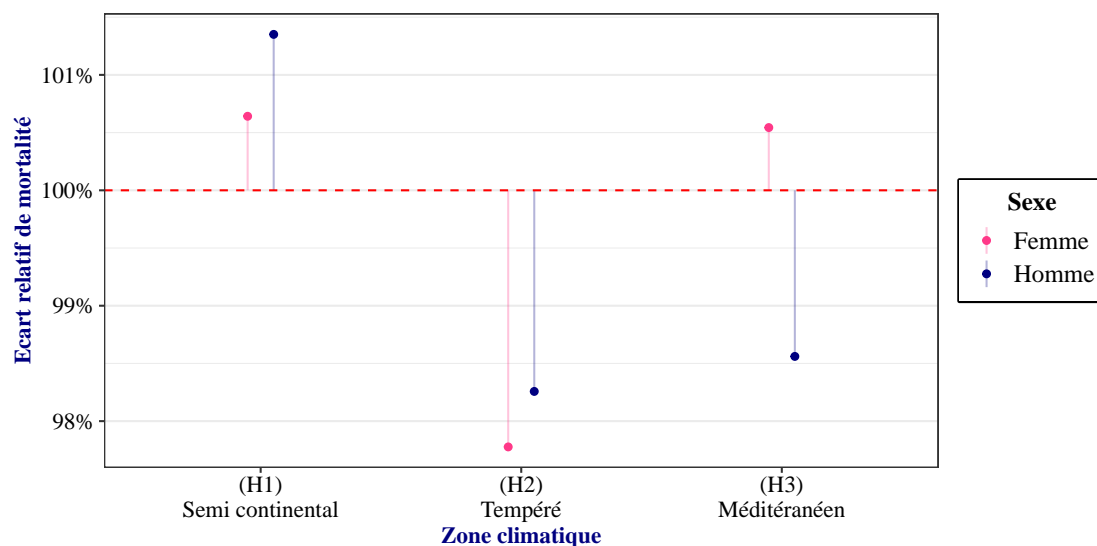


Figure 5.21. : Écart d'ajustement du modèle suivant les zones climatiques

5.4.3. Prise en compte de la spécificité des départements

L'échelle géographique retenue dans cette étude est celle du département. Les températures et décès départementaux ont été agrégés sous l'hypothèse que l'effet de la température est homogène de tous les départements. Cette approche revient donc à estimer la tendance moyenne de l'effet de la température sur la mortalité dans un département moyen.

Cependant, certains territoires peuvent présenter des spécificités climatiques ou démographiques susceptibles de modifier cet effet. Par exemple, la Corse, avec une température moyenne annuelle de 14 °C toute saison confondue, pourrait abriter une population plus acclimatée aux vagues de chaleur, contrairement à la Savoie, où la moyenne est de 7 °C.

Pour évaluer cette hypothèse, la qualité de prédiction du modèle a été examinée pour chaque département, en comparant les décès prédits et observés. Les résultats montrent une forte disparité : les écarts atteignent 20 à 30 % pour les femmes et les pour hommes. Comme illustré à la Figure 5.22, 44 % des départements présentent un écart inférieur à 5 % chez les femmes, contre 30 % des départements chez les hommes. Les surmortalités les plus marquées se concentrent sur l'arc allant de la Bretagne au Grand Est, tandis que les sous-mortalités apparaissent surtout à l'est de la région Auvergne-Rhône-Alpes pour les femmes, et de plus dispersée dans le sud pour les hommes.

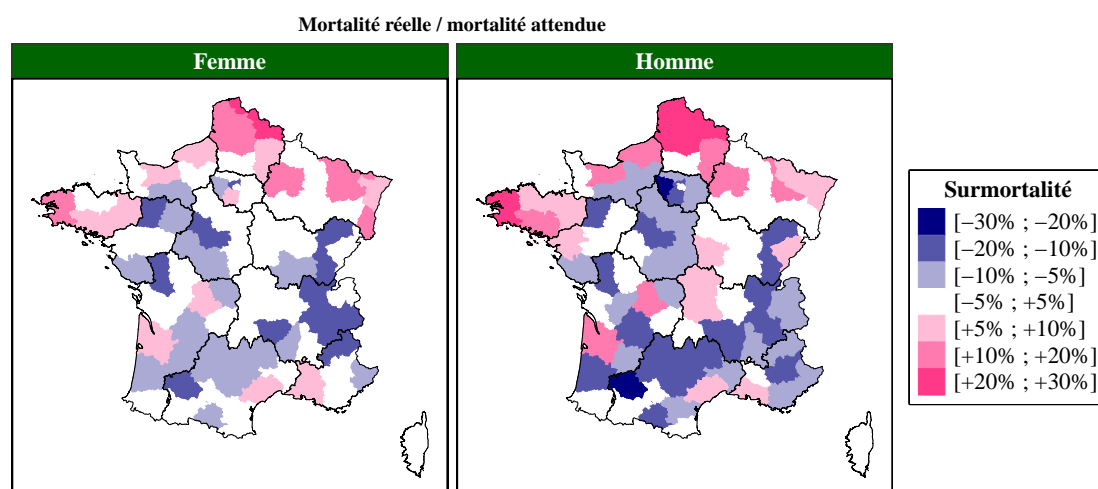


Figure 5.22. : La surmortalité dans les départements

Il convient toutefois de noter que ces écarts ne s'expliquent pas uniquement par les différences de température. Des facteurs locaux, tels que les comportements, les modes de vie ou la structure socio-économique, influencent également la mortalité à l'échelle

du département. Par conséquent, les estimations de mortalité par âge, sexe, période et température obtenues à l'échelle nationale ne peuvent pas être appliquées uniformément à tous départements, ce qui suggère également la présence d'un biais de calibration.

5.4.4. Autres limites

5.4.4.1. Effets des facteurs catalyseurs : vent et humidité

Il convient également de préciser que le modèle ne tient pas compte des facteurs catalyseurs de la mortalité parmi lesquelles : le vent, qui peut amplifier la sensation de froid ; et l'humidité, pouvant amplifier la sensation de chaleur. La prise en compte de ces variables permettrait a priori d'avoir une vision encore plus précise de l'impact de la température. Cette intégration n'a toutefois pas été faite en raison de la complexité supplémentaire que cela implique. Pour explorer cette piste, Salaun (2024) a étudié l'utilisation de la température humide, qui correspond à un ajustement de la température réelle en tenant compte de l'humidité, mais il a conclu que l'utilisation de cette variable n'apporte pas de changement significatif aux résultats.

5.4.4.2. La complexité supplémentaire

L'utilisation de ce modèle nécessite un coût supplémentaire lié principalement à la collecte et à l'exploitation des données de température, ce qui n'est pas nécessaire dans le cadre classique. Cela implique directement que l'utilisation prospective du modèle nécessite de disposer de données sur les évolutions futures de la température ou au moins de faire des hypothèses sur celle-ci, ce qui peut être source d'un nouveau biais.

Synthèse du chapitre

Dans ce chapitre, la température a été intégrée comme facteur supplémentaire aux principaux déterminants de la mortalité, à savoir l'âge, le sexe et la période calendaire. Cette prise en compte s'est révélée pertinente, car elle a conduit à des prédictions plus justes de la mortalité globale.

La construction du modèle de température a d'abord reposé sur un facteur additif supposant que l'impact de la température était uniforme pour toutes les classes d'âge. Ce modèle permettait de dégager la forme générale des effets de la température, mais il présentait des biais liés à l'absence de prise en compte d'un impact différencié selon l'âge. En réalité, les personnes âgées, et notamment les femmes, présentent un risque plus aigu face à des conditions de température extrêmes.

Pour corriger ces biais, le modèle a été révisé en intégrant des effets de température différenciés par âge. Cette modification a permis d'améliorer la précision des prédictions annuelles de mortalité et, surtout, d'obtenir une bonne précision à des pas de temps plus fins, comme la mortalité mensuelle ou quotidienne. Cette performance intra-annuelle s'explique par la capacité du modèle à reproduire le caractère saisonnier de la mortalité grâce à l'intégration du facteur température.

Le modèle s'appuie sur une température moyenne sur 10 jours, choisie pour résumer en un indicateur l'ampleur des conditions de température auxquelles les populations ont été exposées dans les jours précédant le décès. L'analyse de la plage d'observation optimale a montré que la moyenne sur 10 jours est parmi les plus précises, y compris sur la période future, et elle permet d'identifier efficacement les surmortalités attribuables aux vagues de chaleur et de froid.

Ainsi, après avoir mis en évidence l'influence de la température sur la mortalité passée et actuelle, il est naturel de s'interroger sur les conséquences des évolutions climatiques à venir. Le chapitre suivant se consacre à l'étude de l'impact des températures futures, en s'appuyant sur les scénarios du GIEC pour évaluer les risques de surmortalité dans un contexte de changement climatique.

Chapter 6.

Impact des températures futures

La mise en place d'un modèle stochastique de mortalité intégrant un facteur de température a permis d'améliorer la compréhension et la quantification de son impact sur la survenance des décès. Dans ce chapitre, l'objectif est d'appliquer ce modèle à l'anticipation des évolutions plausibles de la mortalité future et d'évaluer l'impact du réchauffement climatique. Dans un premier temps, les projections de température selon les scénarios RCP sont présentées. Leur impact sur la mortalité est ensuite évalué à l'échelle départementale, puis à l'échelle nationale en s'appuyant sur les projections de population de l'INSEE.

6.1. Données de projection de température et analyse descriptive

6.1.1. Source de données de projection de température

Dans le cadre classique des modèles prospectifs, la projection de la mortalité future est relativement simple à mettre en œuvre : pour une période future donnée, il suffit de connaître l'ensemble des âges et des générations composant la population, ce qui est immédiat à déduire. Dans le cas du modèle intégrant la température, son application nécessite en plus la connaissance des conditions de température sur la même période. Pour cela, cette analyse s'appuie sur les projections de température de référence issues du jeu de données DRIAS¹ basé sur trois scénarios RCP d'évolution des émissions de gaz à effet de serre, tels qu'établis par le GIEC et présentés à la Section 1.3

Le jeu de données DRIAS comprend un total de 12 simulations², chacune résultant d'un couple de modèles globaux (GCM) et régionaux (RCM), fournissant ainsi des projections de température sur la période 2006-2100 pour un scénario RCP donné.

¹Lien vers le portail DRIAS

²Les simulations du DRIAS

Dans le cadre de cette étude, le modèle de projections retenu est le modèle ALADIN63 du CNRM (Centre National de Recherche Météorologique). Ce choix repose sur plusieurs critères :

- La disponibilité de simulations pour chacun des 3 scénarios RCP
- Le caractère médian des simulations par rapport à l'ensemble des simulations des modèles
- À qualité égale, la préférence accordée aux simulations produites par les centres météorologiques Français

Cette démarche rejoint celle adoptée par Salaun (2024) et Pincemin (2021) au cours de leurs travaux.

Les données de projection sont disponibles sur chaque point de la grille SAFRAN, couvrant l'ensemble du territoire français à une résolution de 8 km. Pour les besoins de l'étude, ces projections sont agrégées aux échelles départementales et nationale, correspondant au périmètre de cette étude.

6.1.2. Description des températures futures

Pour rappel, les scénarios d'évolution de température diffèrent principalement par les hypothèses sous-jacentes concernant les émissions de gaz à effet de serre.

Tendance croissante des températures nationales

À l'échelle nationale, les tendances futures issues des projections sont cohérentes avec les résultats présentés par Météo France (2020) qui s'appuient sur l'ensemble des simulations disponibles. Globalement, les différents scénarios montrent une augmentation similaire en début de période, avec des températures légèrement plus élevées pour le scénario optimiste RCP 2.6. Cette hausse conjointe se poursuit jusqu'aux années 2045, puis les trajectoires divergent fortement. Le scénario 2.6 conduit à une stabilisation des températures, suivie d'une baisse légère; le scénartio 4.5 continue à croitre, mais à un rythme plus modéré et ralenti ; enfin, scénario pessimiste RCP 8.5 entraine une progression plus que linéaire des températures (Figure 6.1), aboutissant en 2100 à des températures supérieures de 2,8 °C par rapport au scénario 4.5 et de 3,5 °C par rapport au 2.6.

Croissance mutuelle dans tous les départements

À l'échelle des départements, les projections montrent une évolution similaire à la tendance nationale et ce dans tous les départements. Ils décrivent un réchauffement constant tous les scénarios, à l'exception du 2.6 qui conduit à une stabilisation

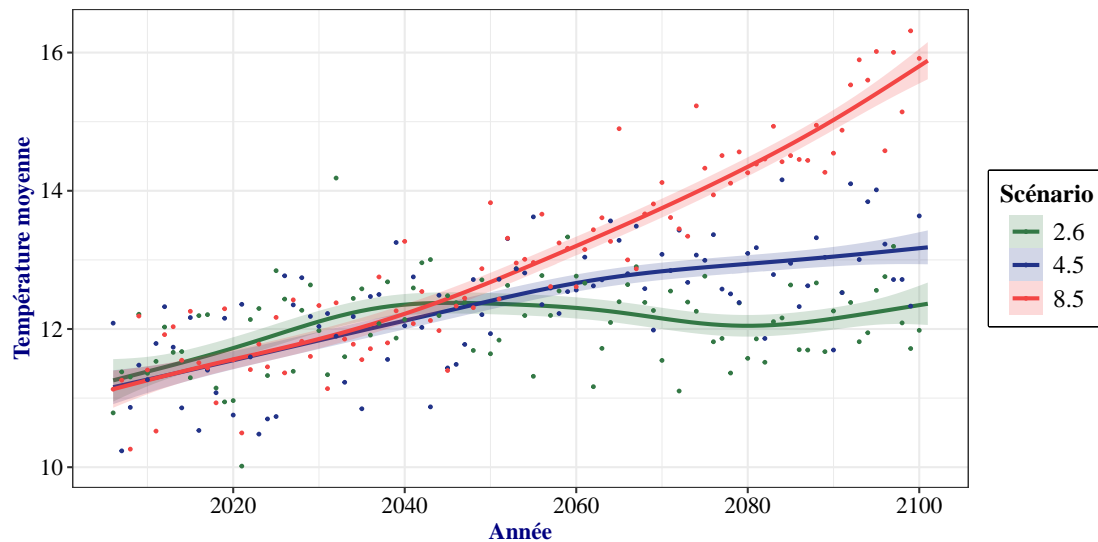


Figure 6.1. : Evolution future des températures en moyenne annuelle

progressive du réchauffement. Toutefois, il faut noter que certains départements présentent des niveaux de température naturellement plus bas, ce qui conduit à des conditions restant relativement soutenables même après réchauffement comme l'illustre la Figure 6.2. Les trois départements ayant les températures les plus basses s'écartant des valeurs communes sont : Savoie (73), Hautes-Alpes (5) et Haute-Savoie (74).

Intensification des vagues de chaleur

Il a déjà été établi que l'élévation des températures s'accompagne d'une augmentation des vagues de chaleur en termes de durée, d'intensité et de fréquence. L'analyse des projections met en évidence une multiplication des vagues de chaleur à un rythme supérieur à celui observé depuis les années 2000. Le nombre de vagues est d'autant plus élevé que le scénario est pessimiste. Par ailleurs, le risque de survenance de vagues de chaleur en dehors de la période estivale est accru. Ceci se produit de façon relativement rare dans les scénarios 2.6 et 4.5, mais devient très fréquent dans le scénario 8.5, notamment à partir des années 2080, où des vagues de chaleur pourraient survenir chaque année jusqu'au mois de septembre. De plus, dans ce même scénario, des canicules d'une sévérité exceptionnelle se produisent en fin de siècle, avec des durées extrêmes de 30 à 40 jours et des températures atteignant 37°C, dépassant largement les records historiques de canicules. Les figures Figure A.4 et Figure A.5 en annexe illustrent ces évolutions suivant les différents scénarios.

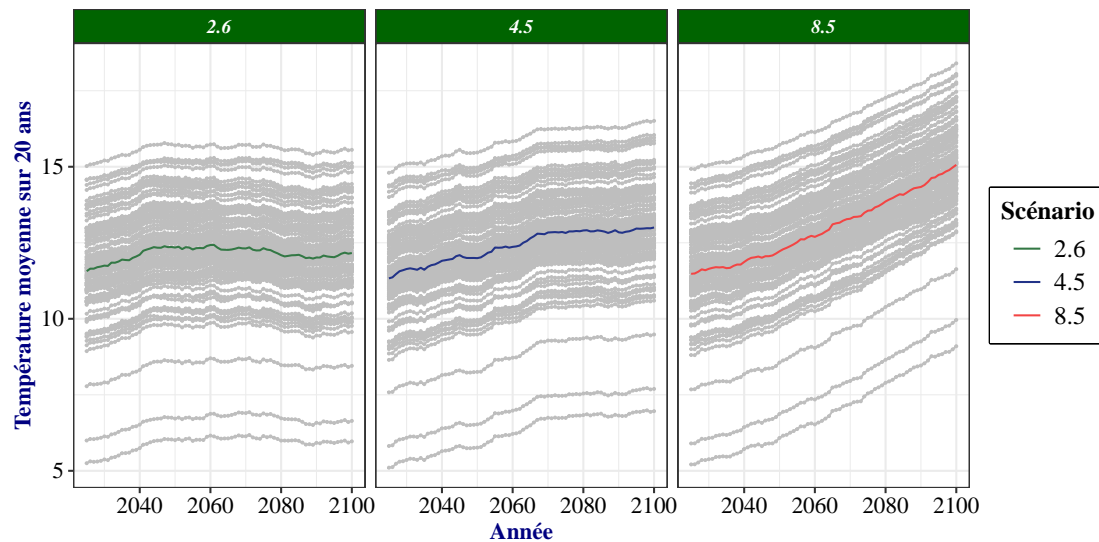


Figure 6.2. : Evolution des températures sur 20 ans des départements

6.2. Impact des projections dans les départements

6.2.1. Impact sur l'intensité de mortalité

Les vagues de chaleur identifiées au niveau national sont des phénomènes d'ampleur significative s'étendant sur une large zone géographique et affectant en général un grand nombre de départements. Cependant, les départements présentent des caractéristiques climatiques hétérogènes, ce qui fait que l'intensité et la survenance des vagues de chaleur varient d'un territoire à l'autre. Avec le réchauffement climatique, les départements déjà chauds deviennent plus chauds, et l'exposition aux vagues de chaleur s'accroît. Dans les départements plus froids, le réchauffement entraîne également une hausse des températures, mais sans atteindre les niveaux propices à des vagues de chaleur fréquentes.

La mise en application du modèle de température permet de quantifier l'impact attendu en fonction des scénarios RCP. Comme ce modèle ne tient pas compte des effets spécifiques à chaque département, les différences d'impact observées entre départements sont directement attribuables aux variations de température. À titre d'illustration, trois départements sont présentés :

Impact des températures futures à Paris

À Paris, de nombreux pics de mortalité apparaissent régulièrement en fin de siècle, en particulier dans le scénario 8.5. L'année 2065 se démarque par un pic exceptionnel,

plus marqué auprès de la population de femmes. Ce phénomène intense résulte de la survenance de deux vagues de chaleur extrêmes de plus de 30 jours chacune, conduisant à un été où quasiment tous les jours ont des températures au delà du seuil d'alerte de Météo France, ce qui prédispose à une surmortalité élevée, notamment pour les femmes qui sont plus sensibles aux températures extrêmes (Figure 6.3).

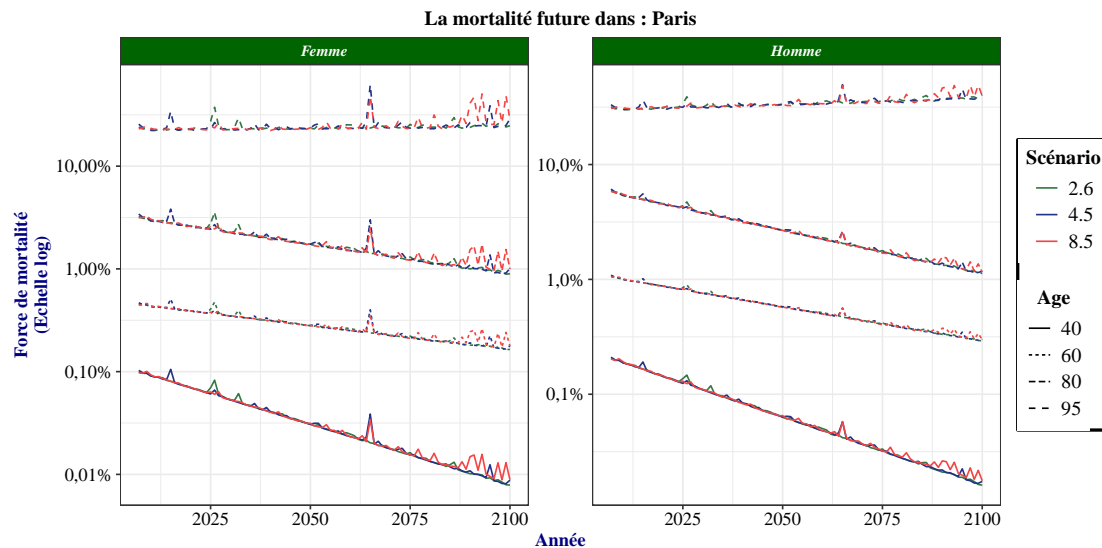


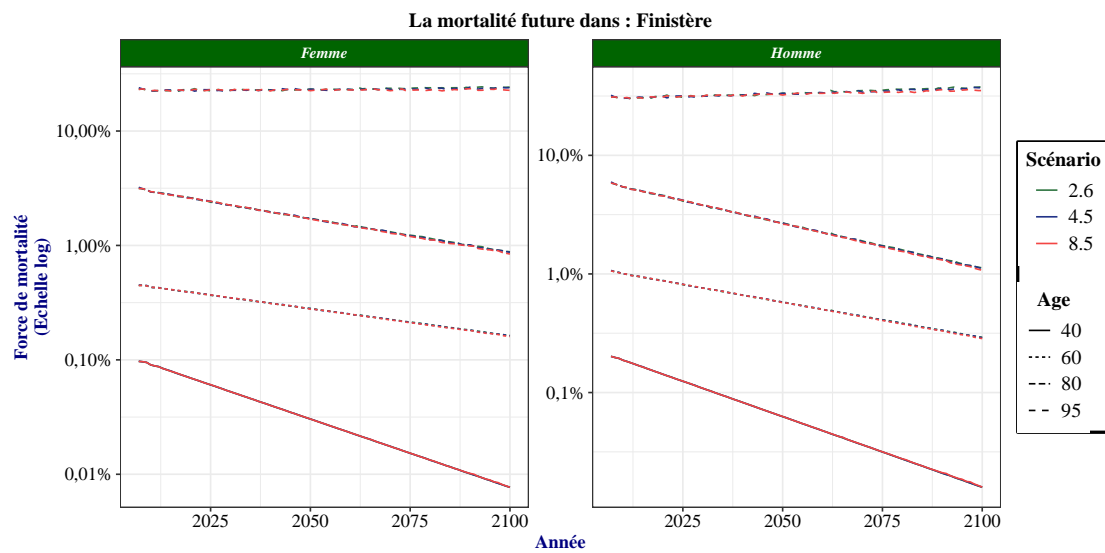
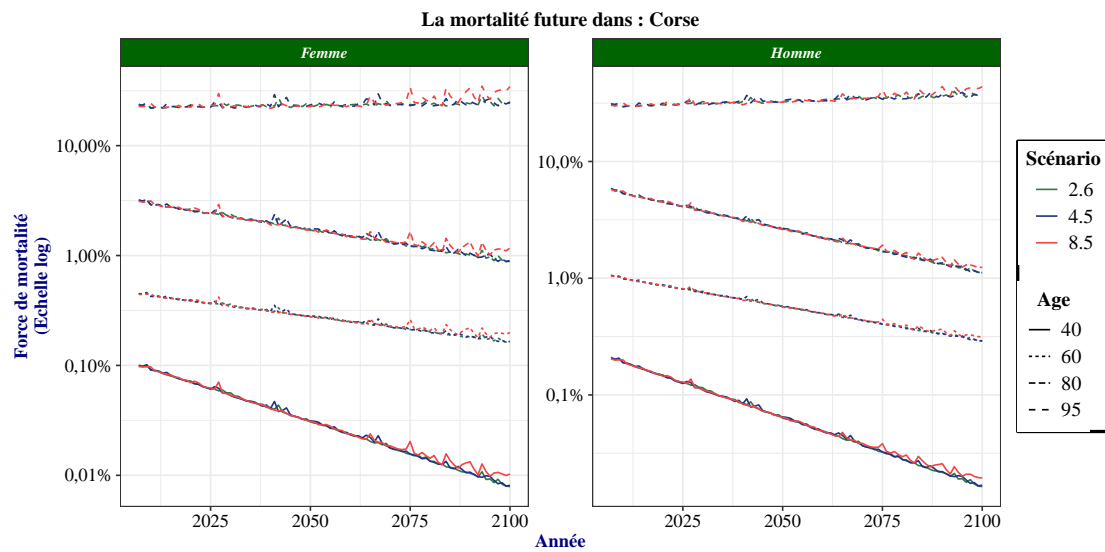
Figure 6.3. : Les surmortalités liées aux températures selon les scénarios (Paris)

Impact des températures futures dans la Corse

En Corse, plusieurs pics de surmortalité sont également observés, mais d'intensité modérée comparativement Paris. La vague de chaleur exceptionnelle de 2065 semble ne pas atteindre ce territoire, et globalement, les vagues de chaleur de la fin du siècle y sont moins intenses et provoquent moins de décès (Figure 6.4).

Impact des températures futures dans le finistère

Dans le Finistère, l'évolution future de la mortalité est quasi linéaire, sans apparition de pics particuliers, quel que soit l'âge ou le scénario considéré. Cela suggère que ce département sera peu affecté par les vagues de chaleur, qui ne seraient que brèves ou de faible intensité. Dans ce département, les scénarios de fort réchauffement correspondent paradoxalement à des situations plus favorables. En effet, l'augmentation des températures ne conduit pas à des niveaux extrêmes mais elle au contraire, elle réduit l'intensité des épisodes de froid hivernale et la mortalité qui y est associée, conduisant ainsi à une mortalité globale plus faible. Ainsi, sur la Figure 6.5, la mortalité du scénario RCP 8.5 apparaît inférieure à celle des autres scénarios.



Ces analyses suggèrent que l'impact des évolutions futures de température variera fortement selon les zones géographiques. Certains départements seront plus exposés aux vagues de chaleur extrêmes, ce qui se traduira par des surmortalités localisées. Pour mieux évaluer ces effets et identifier les zones les plus vulnérables, l'impact sur l'espérance de vie résiduelle sera analysé dans la section suivante.

6.2.2. Impact sur l'espérance de vie prospective

Espérance de vie prospective à Paris

Pour caractériser l'impact global des projections futures de température, considérons un individu I né en 2000, âgé de 7 ans en 2007, année choisie comme début d'observation. Cette année correspond à la première année entièrement observée pour laquelle les températures sur 10 jours ont été calculée. On suppose que l'individu I subit les évolutions de température prévues par les scénarios RCP jusqu'en 2100, et l'objectif est d'estimer la durée de vie moyenne à laquelle il pourrait s'attendre.

Dans un premier temps, considérons que l'individu I réside à Paris durant toute cette période. La courbe de sa vie présente les probabilités annuelle de décès auxquelles il sera exposé chaque année de la période, où son âge évoluera progressivement jusqu'à 100 ans en 2100. En fonction des scénarios RCP, la Figure 6.6 montre qu'il a déjà surmonté la mortalité infantile et une tendance générale à la hausse de sa probabilité de décès se dessine. Par ailleurs, il fera face à plusieurs chocs de température extrême, entraînant des pics ponctuels de mortalité, qui varie selon les scénarios : le scénario 2.6 prévoit un pic important à 25 ans et autour de 31 ans, le 4.5 prévoit de nombreux pics affectant particulièrement les femmes, notamment aux âges autour de 65 ans et 95 ans, enfin le scénario 8.5 prévoit également de nombreux pics, davantage accentués chez les femmes et qui se manifestent beaucoup plus aux grands âges à partir de 65 ans. Dans ce scénario, l'excès de mortalité féminine conduit à un niveau de mortalité qui atteint celui des hommes dont la mortalité est d'habitude supérieure.

Dans ces conditions de température, l'espérance de vie moyenne de l'individu varie selon son sexe (Table 6.1) :

- **S'il s'agit d'un homme** : il peut s'attendre à vivre pendant 78 ans et 9 mois supplémentaires dans le cas du scénario 2.6. Il peut espérer vivre 4 jours de plus dans le cas du scénario 4.5 tandis qu'il doit s'attendre à 2 mois de moins en cas de réalisation du scénario 8.5.
- **S'il s'agit d'une femme** : dans le cas du scénario 2.6 elle pourra s'attendre à vivre pendant 82 ans et 3 mois supplémentaires. Pour le scénario 4.5 elle peut s'attendre à vivre 1 mois de moins et à 8 mois de moins en cas de réalisation du scénario 8.5.

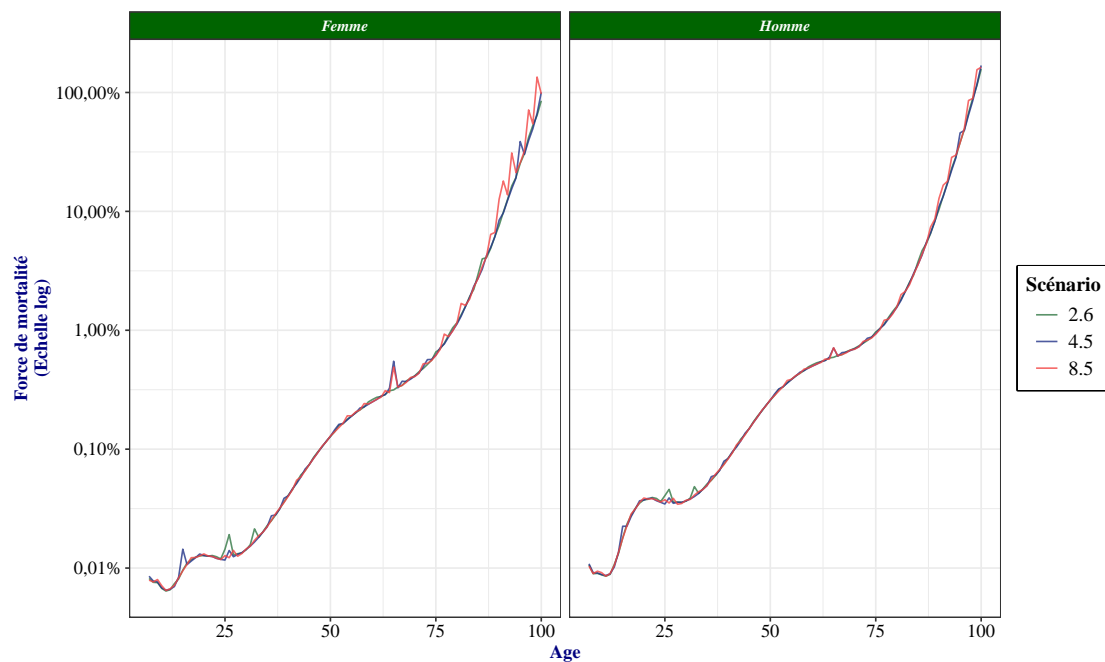


Figure 6.6. : Courbe de vie de la génération 2000 à Paris sur la période 2007-2100

Table 6.1. : Espérance de vie résiduelle de la génération 2000 à Paris

Scénario	Homme	Femme
2.6	78,75 (+0,00)	82,22 (+0,00)
4.5	78,76 (+0,01)	82,13 (-0,09)
8.5	78,58 (-0,17)	81,49 (-0,72)

Espérance de vie prospective dans les départements

L'impact attendu des projections de température varie selon les départements en raison des différences climatiques locales. Les résultats sont globalement similaires pour les deux sexes, à l'exception de l'intensité de la baisse d'espérance de vie qui est plus importante pour les femmes. La zone la moins affectée se situe à l'ouest du territoire, notamment en Bretagne, où l'espérance de vie reste la plus élevée et ceci pour les trois scénarios.

A l'inverse, le pourtour méditerranéen au sud de la France est particulièrement exposé aux impacts des températures futures, particulièrement dans le scénario le plus pessimiste. Les scénarios 2.6 et 4.5 montrent également un risque élevé dans le sud-est. Ces différences géographiques sont illustrées sur les figures 6.7 et 6.8 respectivement pour les femmes et pour les hommes.

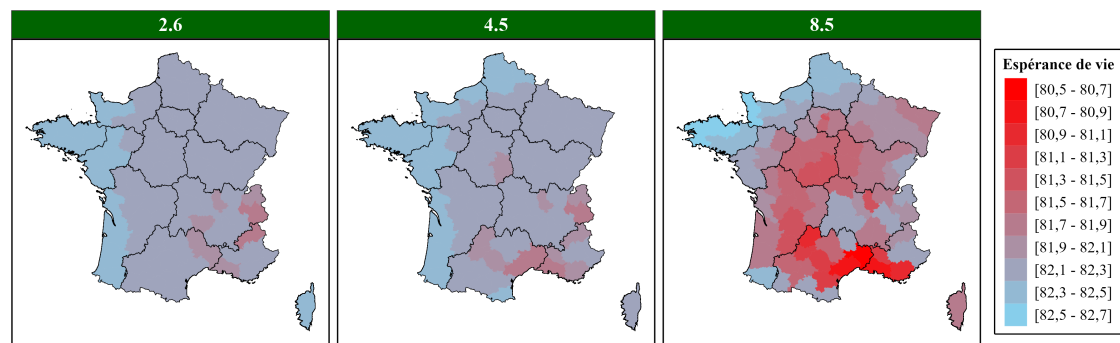


Figure 6.7. : Espérance de vie des femmes par département et par scénario

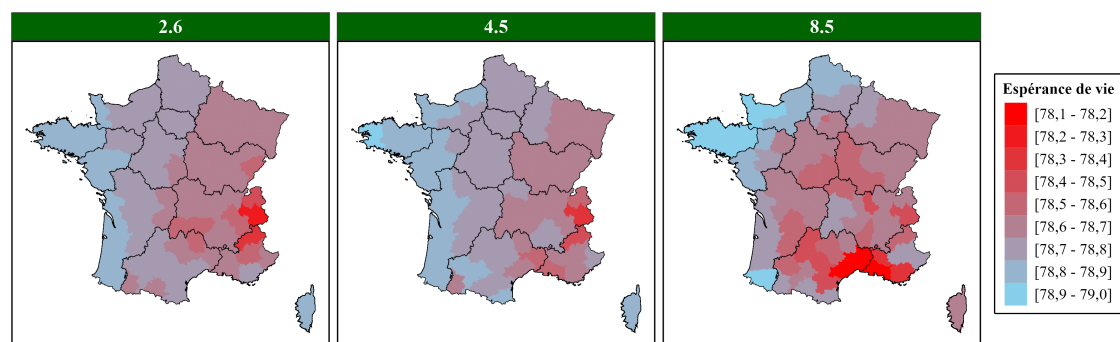


Figure 6.8. : Espérance de vie des hommes par département et par scénario

Un tableau récapitulatif présentant les impacts détaillés pour chaque département est disponible en Table A.1 en annexe.

De façon globale, tous départements confondus, l'espérance de vie prospective de la génération 2000 est évaluée à 85,7 ans pour les Hommes et à 89,2 ans pour les femmes dans le cas du scénario 2.6. Le scénario 4.5 ne présente pas des écarts significatifs relativement au 2.6 tandis que dans le cas du scénario 8.5 l'espérance de vie est plus faible de près d'un mois pour les hommes et de près de 5 mois pour les femmes relativement au scénario 2.6 (Table 6.2).

Table 6.2. : Espérance de vie globale de la génération 2000

Scénario	Homme	Femme
2.6	85,7 ans	89,2 ans
4.5	+ 0,3 mois	- 0,2 mois
8.5	- 0,8 mois	- 4,7 mois

6.3. Impact des températures futures à l'échelle nationale

L'analyse de l'impact global des scénarios climatiques au niveau national nécessite de tenir compte du poids démographique de chaque département, car la répartition de la population n'est homogène sur le territoire. Dans cette sous-section, l'objectif est donc de mobiliser des données sur l'évolution future de la population à l'échelle départementale, afin de les agréger ensuite au niveau national.

6.3.1. Source de données de projection de population

Les projections de population sont réalisées par l'INSEE sur la période 2021-2070 en appliquant la méthode dite des "composantes", qui repose sur des hypothèses concernant les trois principaux facteurs affectant l'évolution de la population : la fécondité, la mortalité et les migrations. Ces hypothèses s'appuient à la fois sur les tendances observées sur période récente (2010-2019) mais aussi en tenant compte des avis des experts sur le sujet.

Plusieurs scénarios sont établis afin de refléter l'incertitude entourant ces évolutions. Le scénario principal, appelé scénario "central" considère les hypothèses jugées les plus plausibles, tandis que des scénarios alternatifs sont obtenues en révisant un ou plusieurs paramètres d'évolution. Dans le cas du scénario centrale, les hypothèses appliquées sont les suivantes :

- **Fécondité** : un indicateur conjoncturel de fécondité fixé à 1,8 en accord avec les évolutions récentes. Il s'agit du nombre moyen d'enfants qu'une femme est susceptible d'avoir au cours de sa vie.
- **Mortalité** : une réduction progressive des quotients de mortalité au même rythme que celui de la période 2010-2019 conduisant à une amélioration de l'espérance de vie qui atteindrait, en 2070, 90 ans pour les femmes et 87,5 ans pour les hommes.
- **Migration** : un solde migratoire fixé à +70 000 personnes par an, chaque année jusqu'en 2070.

Des informations détaillées sur la méthodologie et les résultats de ces projections sont disponibles sur le site de l'INSEE³. Dans le cadre de cette analyse, seul le scénario central est retenu, à la fois parce qu'il est le scénario moyen le plus plausible et parce que les données des autres scénarios ne sont pas librement accessibles à l'échelle départementale. Les données utilisées proviennent du portail interactif des projections de pyramides des âges⁴ qui, pour le scénario central, couvrent l'ensemble des départements jusqu'en 2070.

³Projection de population 2021-2070

⁴Interface des pyramide des âges

6.3.2. Impact des températures futures dans la France hexagonale

Evolution attendue de la population

Les données de projections de population de l'INSEE décrivent une croissance ralentie de la population jusqu'aux années 2040 pour les femmes et 2050 pour les hommes, où le maximum de population est atteint. Au-delà, la population décroît jusqu'à la fin de la période étudiée en 2070. Tout au long de l'horizon de projection, les femmes demeurent plus nombreuses que les hommes, mais l'écart se réduit progressivement avec le temps (?@fig-evol_future_pop).

Evolution des décès au niveau national

Dans chacun des scénarios, le nombre total annuel de décès suit une tendance globale à la hausse. Les différentes trajectoires présentent des pics de surmortalité dont l'intensité varie avec l'ampleur des vagues de chaleur des périodes correspondantes.

En début de période jusqu'aux années 2040, le scénario 2.6 présente déjà des vagues de chaleur particulièrement sévères notamment en 2026 et 2032. Par contre, le scénario 4.5 se distingue par la fréquence plus importantes de ses pics dont l'intensité reste intermédiaire à l'exception de l'année 2065 présentant une surmortalité importante, avec environ 350 000 décès de femmes en excès et environ 150 000 décès d'hommes en excès. L'année 2065 est encore plus catastrophique dans le scénario 8.5 où les excès de mortalité atteignent environ 500 000 décès chez les femmes et 250 000 chez les hommes (Figure 6.9).

Dans l'ensemble, sur la période 2026-2070, le nombre total de décès prévu est le plus élevé dans le scénario 4.5, tandis que les scénarios 2.6 et 8.5 présentent des impacts plus ou moins importants selon le sexe (Table 6.3).

Table 6.3. : Nombre de décès de 2026 à 2070 selon les scénarios RCP

Scénario	Femme	Homme
2.6	20 384 047	18 631 819
4.5	20 806 106	18 715 753
8.5	20 624 724	18 590 514

Ce résultat peut sembler contre-intuitif, car on s'attend à ce que le scénario 8.5 qui est le plus pessimiste en termes d'émissions de CO₂ ait le plus fort impact sur la mortalité. Sur cette question, comme l'a souligné Salaun (2024), la limite principale provient de l'horizon de projection démographique disponible, qui en se limitant en 2070 n'est pas assez large pour présenter une différence significative entre les impacts des scénarios. Par exemple, c'est en fin de siècle que le scénario 8.5 a montré les canicules les plus

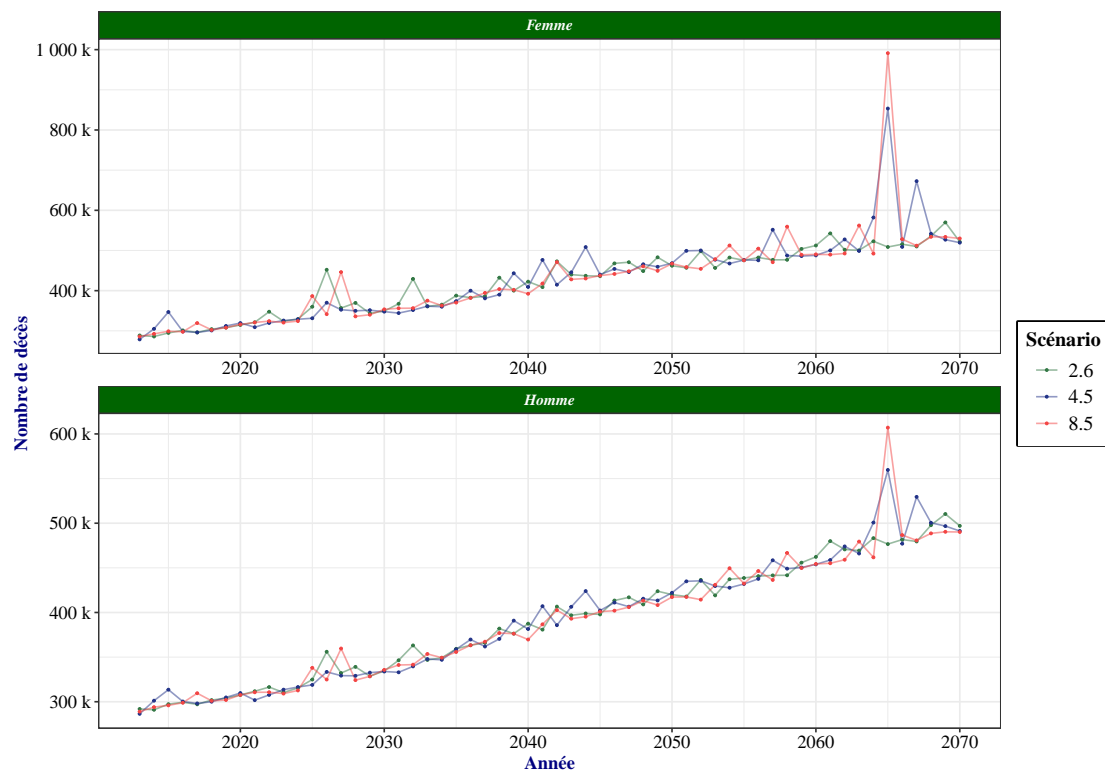


Figure 6.9. : Projection du nombre annuel de décès future

dévastatrices se produisant quasiment tous les ans mais ces épisodes n'ont pas pu être considérés sur la période avant 1970.

Synthèse du chapitre

L'analyse de l'impact des projections climatiques et démographiques à partir du modèle stochastique de mortalité a permis d'illustrer de manière concrète les effets attendus du réchauffement climatique sur la mortalité en France. Les résultats montrent que, quels que soient les scénarios retenus, la hausse des températures s'accompagnera d'une augmentation de la mortalité, marquée par des pics de surmortalité liés aux vagues de chaleur. L'intensité et la fréquence de ces pics varient selon les scénarios, avec des effets différenciés entre sexes, âges et territoires.

À l'échelle individuelle, les projections révèlent un impact direct sur l'espérance de vie prospective, avec des pertes plus marquées pour les femmes dans les scénarios pessimistes. À l'échelle départementale, les disparités géographiques apparaissent nettement : la Bretagne se distingue par une relative résilience, tandis que le pourtour méditerranéen et le sud-est de la France présentent des vulnérabilités accrues. Enfin, à l'échelle nationale, la combinaison des dynamiques démographiques et climatiques conduit à une hausse significative du nombre de décès, particulièrement marquée dans le scénario RCP 4.5 sur l'horizon disponible des projections, même si les effets les plus extrêmes du scénario 8.5 sont attendus plus tard dans le siècle.

Ces résultats soulignent à la fois la réalité des risques sanitaires liés au climat et la nécessité de les anticiper. Ils mettent en évidence que les impacts du réchauffement climatique ne seront pas uniformes, mais dépendront des caractéristiques démographiques et géographiques des territoires. Cela ouvre des perspectives pour la mise en place de politiques publiques ciblées, visant à renforcer la résilience des populations les plus exposées.

Conclusion

Le changement climatique actuel se manifeste par une élévation continue des températures qui ont des conséquences profondes sur la santé et les modes de vie. Les catastrophes naturelles, particulièrement les canicules et les vagues de chaleur extrême deviennent plus fréquentes et plus sévères, entraînant chaque année une surmortalité significative. Dans ce contexte, les assureurs sont exposés à une hausse progressive de la sinistralité, qui serait d'autant plus conséquent avec la non prise en compte des influences climatiques dans les prévisions de mortalité. De ce fait, il est essentiel d'anticiper les évolutions futures et de quantifier l'impact attendu des températures sur le risque de mortalité. C'est dans cette perspective que s'inscrit ce mémoire, en proposant des modèles stochastiques intégrant les effets de la température.

Après avoir décrit les évolutions climatiques récentes et les effets physiologiques des températures en France, l'analyse a mobilisé des données fines issues de Météo France, de l'INSEE et de la Human Mortality Database, permettant de capter au mieux le stress thermique subi à la veille du décès. Leur examen critique a révélé plusieurs limites (doublons et omissions des décès, valeurs manquantes) qui ont été minutieusement corrigées. Leur harmonisation a abouti à une structure cohérente des données présentant les effectifs exposés et décédés selon l'âge, le sexe, l'année et la température. L'analyse exploratoire a confirmé des résultats attendus : augmentation progressive des températures annuelles, évolution saisonnière marquée des températures comme des décès, augmentation progressive des nombre de décès mais baisse tendancielle des taux de mortalité grâce progrès scientifiques et médicaux.

La méthodologie de modélisation s'inspire des modèles prospectifs classiques s'appuyant sur les facteurs âge, période calendaire et cohorte qui permettent de décrire les principaux traits caractéristiques de la mortalité : forte mortalité infantile, croissance exponentielle de la mortalité au-delà de 10 ans, intensité constamment plus élevée pour les hommes, amélioration graduelle dans le temps avec une intensité variée selon les profils d'âges. Les modèles AP, APC et le Lee-Carter ont été calibrés dans ce cadre. La mobilisation d'un modèle additif généralisé (GAM), par leur flexibilité, a permis de capter ces effets tout en conduisant à une estimation plus précise de la mortalité annuelle. Toutefois, son application directe à l'échelle des températures a montré qu'il surestime la mortalité pour une température sur 10 jours entre 9°C et 25°C, tandis qu'il la sous-estime en dehors de cette plage.

L'enrichissement de ce modèle GAM par l'intégration de facteurs de température a corrigé ces limites et confirmé une relation en U entre la température et la mortalité,

avec des risques accrus lors des températures extrêmes notamment en cas de vagues de chaleur ou de froid. Il a également mis en évidence une plus forte sensibilité des personnes âgées et les femmes. Ces résultats contribuent à la fois à une meilleure quantification des impacts de la température mais aussi constituent un cadre capable de mieux anticiper l'impact du changement climatique sur les engagements de long terme.

Le modèle de température ainsi construit a ensuite été appliqué à la projection de la mortalité future, ceci à l'horizon 2100 au niveau départemental et à l'horizon 2070 au niveau national. Ces projections se sont appuyées sur les scénarios de température du portail DRIAS et sur les projections centrales de populations de l'INSEE. A l'échelle nationale, les résultats montrent une hausse tendancielle du nombre de décès qui est liée au vieillissement démographique, amplifiée par des épisodes de surmortalité dues aux vagues de chaleur, particulièrement dans les scénarios pessimistes. À l'échelle départementale, l'impact sur l'espérance de vie prospective présente une forte dissymétrie selon les zones géographiques, il apparaît plus marqué dans le pourtout du sud-est, tandis qu'il reste modéré dans l'ouest, notamment en Bretagne. Dans l'ensemble, l'espérance de vie réduit d'environ 1 mois pour les hommes et d'environ 5 mois pour les femmes dans le scénario 8.5 relativement au scénario 2.6. Ces estimations pourraient toutefois être atténuées par une acclimatation progressive et par de nouvelles avancées technologiques en matière d'adaptation.

Dans son ensemble, ce mémoire propose une contribution supplémentaire à la modélisation de la mortalité. Grâce à une méthodologie robuste, il démontre l'intérêt d'élargir les modèles actuariels classiques à des déterminants climatiques, à l'instar de la température. Cependant, certaines limites peuvent être identifiées ouvrant la voie à des perspectives d'amélioration. En effet, ce modèle ne tient pas compte de facteurs alternatifs tels que le vent et l'humidité qui sont susceptibles d'amplifier les effets de la chaleur; il néglige les spécificités locales et suppose des effets de température identiques dans chaque départements et enfin, il considère que les fluctuations de températures peuvent être résumées en une seule variable qui est la moyenne sur 10 jours au détriment de l'intégration d'une variable par jour captant mieux les conditions climatiques réelles. Malgré ces limites, ce modèle apporte une plus value certaine à l'évaluation de l'impact des températures sur la mortalité. Il constitue une base utile pour la réalisation des stress-tests climatiques et pour l'ajustement des prévisions de mortalité en matière de tarification et de provisionnement.

Références

Alonso, E. (2024) *Utilisation des modèles additifs généralisés pour la construction de table d'expérience en l'arrêt de travail*. mémoire de master. Institut des Actuares.

Biessy, G. (2025) « Modèles de durée – Cours de Master 2 ». EURIA.

Brouhns, N., Denuit, M. et Vermunt, J.K. (2002) « A Poisson log-bilinear regression approach to the construction of projected lifetables », *Insurance: Mathematics and Economics*, 31, p. 373-393.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G., Epstein, D. et Kalhaf-Allah, M. (2008) « Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models », *Pensions Institute Discussion Paper* [Preprint], (PI-0801).

Fall, P.C. (2021) *Intégration de facteurs liés à la température dans modélisation du risque de mortalité*. mémoire de master. Institut du Risk Management; Institut des Actuares.

Fine, J. et Fotso, S. (1989) « Contribution à l'étude du modèle âge-période-cohorte », *Revue de statistique appliquée*, 37(3), p. 39-56.

Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneux-Jouyaux, C., Clavel, J., Jouglu, E. et Hemon, D. (2006) « Excess mortality related to the August 2003 heat wave in France. », *International Archives of Occupational and Environmental Health*, 80(1), p. 16-24.

France Assureurs (2021) *Impact du changement climatique sur l'assurance à l'horizon 2050*. Rapport. France Assureurs.

Gibelin, A.-L. (2015) *Evolution du climat passé en France métropolitaine depuis les années 50*. Météo France.

Global-climat (2021) *Rapport AR6 du GIEC.: le point sur la température globale*.

INED (2006) *Progrès récent de l'espérance de vie en France : les hommes comblent une partie de leur retard*. Rapport. INED.

INSEE (2024) *Un enfant sur 250 meurt avant l'âge d'un an en France*. Insee Première

n° 2048. Institut National de la Statistique et des Études Économiques.

Institut de veille sanitaire (2010) *Froid et santé. Note de position de l'Institut de veille sanitaire*. Institut de veille sanitaire.

Jens, R., Katrien, A. et Torsten, K. (2025) « The short-term association between environmental variables and mortality: evidence from Europe », *Journal of the Royal Statistical Society Series A : Statistics in Society* [Preprint]. Disponible sur: <https://doi.org/10.1093/jrsssa/qnaf052>.

Météo France (2020) *Les nouvelles projections climatiques de référence DRIAS 2020 pour la métropole*. Météo France.

Météo France (2024a) *Hivers : quelques vagues de froid remarquables*. Disponible sur: <https://meteofrance.com/actualites-et-dossiers/magazine/hivers-quelques-vagues-de-froid-remarquables>.

Météo France (2024b) *Qu'est ce que la température ressentie?* Disponible sur: <https://meteofrance.com/magazine/meteo-questions/quest-ce-que-la-temperature-ressentie>.

Météo France (2025a) *Canicule, pic ou vague de chaleur ?* Disponible sur: <https://meteofrance.com/actualites-et-dossiers/comprendre-la-meteo/canicule-vague-ou-pic-de-chaleur>.

Météo France (2025b) *Été 2025 : tout comprendre sur les phénomènes météorologiques de la saison estivale*.

Météo France (2025c) *Planète : Le changement climatique observé*. Disponible sur: <https://meteofrance.com/changement-climatique/observer/le-changement-climatique-observe-dans-le-monde>.

OXFAM France (2022) *Changement climatique : causes, conséquence et solutions*.

Peltier, F. (2016) *Regards sur la mortalité saisonnière*. STATEC, Luxembourg.

Pincemin, G. (2021) *Risques climatiques et mortalité, impact du risque canicule à l'horizon 2070*. mémoire de master. Institut des Actuaire.

Pouna Siewe, V. (2010) *Modèles additifs généralisés : Intérêts de ces modèles en assurance automobile*. mémoire de master. ISUP; Institut des Actuaire.

Salaun, F. (2024) *Modélisation par apprentissage statistique du lien température-mortalité en Open Data et application prédictive*. mémoire de master. Institut des

Actuaires.

Santé publique France (2004) *Froid et santé. Eléments de synthèse bibliographique et perspectives. Rapport d'investigation*. Santé publique France.

Santé Publique France (2019) *Grand froid*. Disponible sur: <https://www.santepubliquefrance.fr/determinants-de-sante/climat/grand-froid>.

Santé Publique France (2025) *Chaleur et impact sur la santé*.

Sunamoto, N. (2005) *Cohort Effect Structure in the Lee-Carter Residual Term*. Fukoku Mutual Life Insurance Company.

Swiss Re Institute (2018) *Amélioration de la mortalité : comprendre le passé et anticiper le futur*. Rapport Sigma n°6/2018. Swiss Re Institute.

Vermet, F. (2025) « Modèles de durée – Cours de Master 2 ». EURIA.

Villegas, A.M., Millossovich, P. et Kaishev, V.K. (2018) « StMoMo: An R Package for Stochastic Mortality Modeling », *Journal of Statistical Software*, 84(3), p. 1-38.

Wafo Kankeu, C.B. (2023) *Impact du montant de rente sur la longévité au sein d'un portefeuille de rentiers*. mémoire de master. Institut des Actuaires.

Wood, S. (2017) *Generalized Additive Models: An Introduction with R*. 2nd éd. New York: Chapman; Hall/CRC.

Appendix A.

Résultats complémentaires

A.1. Les surmortalités annuelles réelles et prédites par les modèles

- Surmortalité de l'année 2006

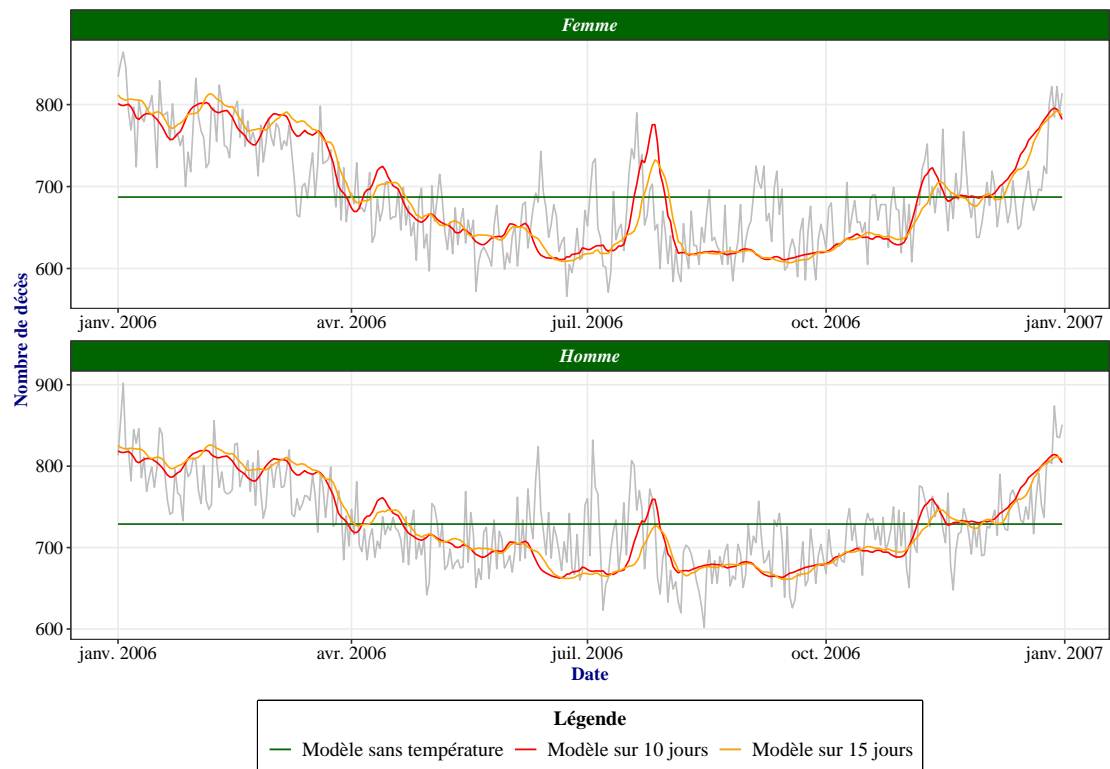


Figure A.1. : Comparaison des mortalités quotidiennes en 2006

- Surmortalité de l'année 2018

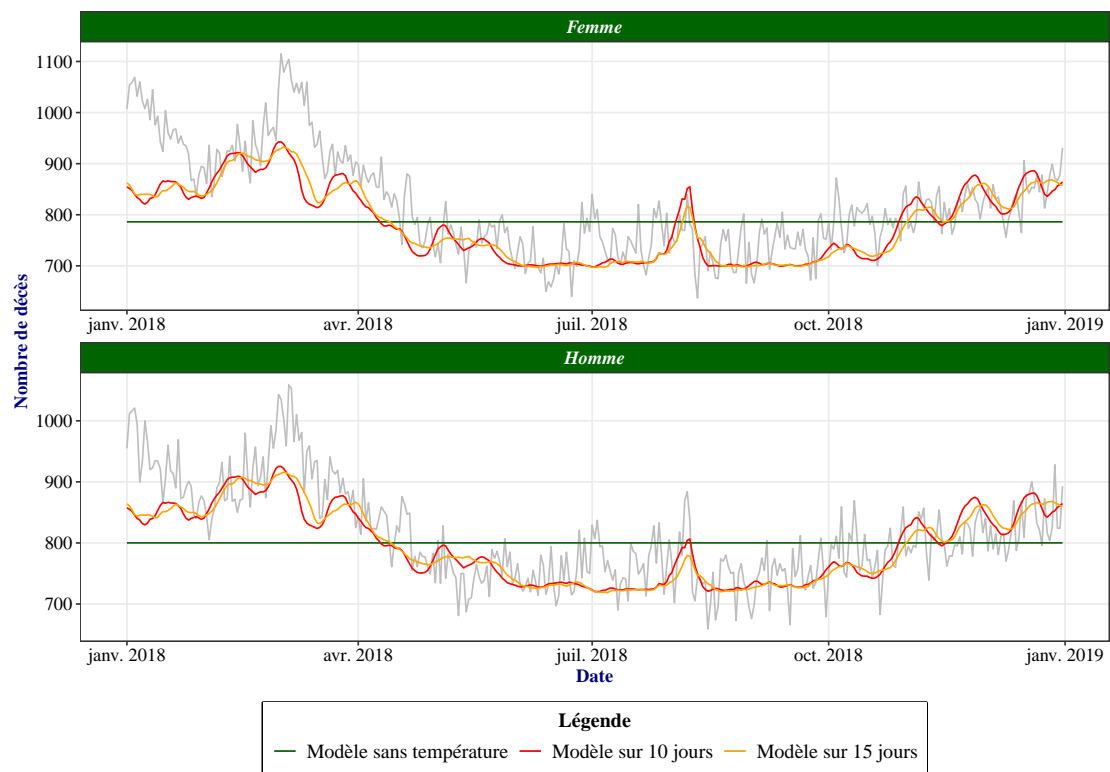


Figure A.2. : Comparaison des mortalités quotidiennes en 2018

• **Surmortalité de l'année 2020**

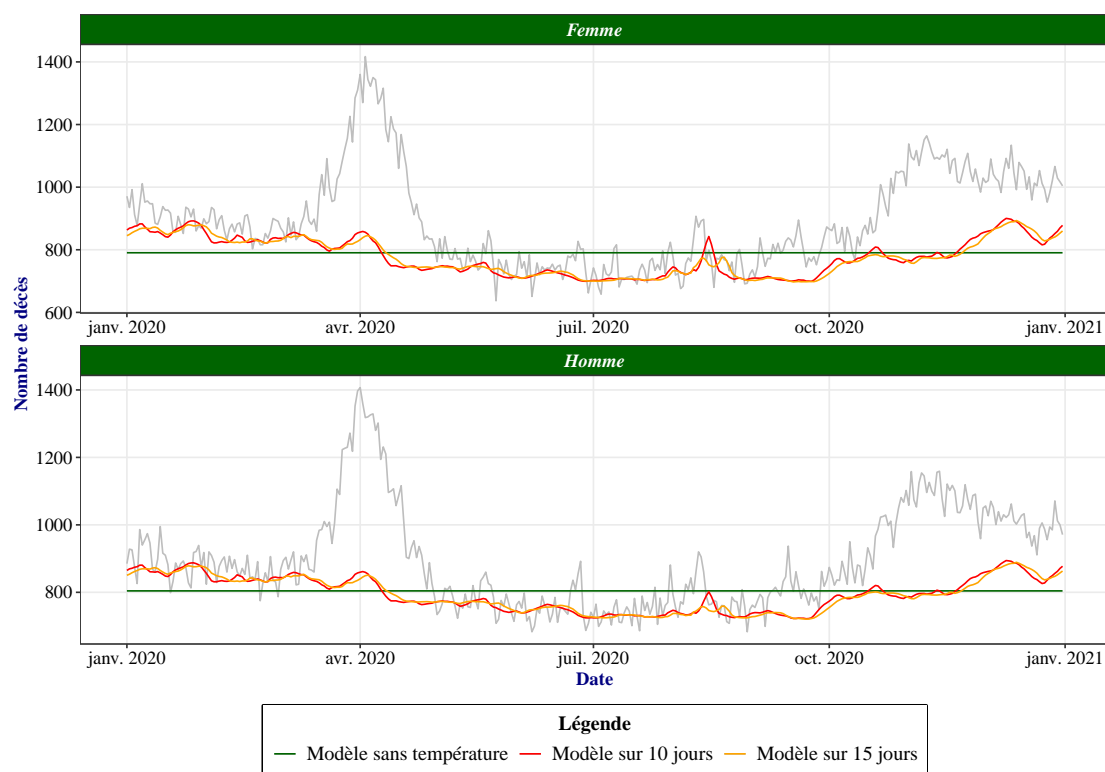


Figure A.3. : Comparaison des mortalités quotidiennes en 2020

A.2. Les vagues de chaleur futures selon les scénarios RCP

A.3. Espérances de vie par département selon les scénarios RCP

Table A.1. : Espérance de vie par département selon les scénarios RCP

Code INSEE	Département	Homme			Femme		
		RCP 2.6	RCP 4.5	RCP 8.5	RCP 2.6	RCP 4.5	RCP 8.5
1	Ain	78,65	(+0,04)	(-0,05)	82,14	(+0,01)	(-0,34)
2	Aisne	78,71	(+0,04)	(+0,07)	82,24	(+0,01)	(-0,05)
3	Allier	78,69	(+0,03)	(-0,11)	82,16	(-0,01)	(-0,58)
4	Alpes-de-Haute-Provence	78,59	(+0,07)	(+0,13)	82,11	(+0,07)	(+0,09)
5	Hautes-Alpes	78,32	(+0,09)	(+0,23)	81,83	(+0,10)	(+0,25)
6	Alpes-Maritimes	78,68	(+0,06)	(+0,12)	82,20	(+0,06)	(+0,08)

7	Ardèche	78,67	(+0,03)	(-0,10)	82,16	(+0,01)	(-0,44)
8	Ardennes	78,67	(+0,04)	(+0,07)	82,20	(+0,00)	(-0,06)
9	Ariège	78,64	(+0,06)	(+0,14)	82,17	(+0,06)	(+0,10)
10	Aube	78,69	(+0,02)	(-0,04)	82,20	(-0,04)	(-0,34)
11	Aude	78,77	(-0,02)	(-0,33)	82,23	(-0,09)	(-0,99)
12	Aveyron	78,69	(+0,02)	(-0,11)	82,18	(-0,01)	(-0,53)
13	Bouches-du-Rhône	78,67	(-0,11)	(-0,52)	82,06	(-0,24)	(-1,27)
14	Calvados	78,78	(+0,06)	(+0,14)	82,32	(+0,05)	(+0,11)
15	Cantal	78,59	(+0,07)	(+0,12)	82,12	(+0,07)	(+0,04)
16	Charente	78,78	(+0,02)	(-0,15)	82,26	(+0,00)	(-0,63)
17	Charente-Maritime	78,86	(+0,03)	(-0,06)	82,38	(+0,00)	(-0,33)
18	Cher	78,69	(+0,01)	(-0,19)	82,15	(-0,05)	(-0,80)
19	Corrèze	78,70	(+0,04)	(-0,05)	82,20	(+0,03)	(-0,41)
20	Corse	78,84	(-0,02)	(-0,22)	82,31	(-0,04)	(-0,59)
21	Côte-d'Or	78,66	(+0,03)	(-0,04)	82,17	(-0,03)	(-0,34)
22	Côtes-d'Armor	78,80	(+0,06)	(+0,16)	82,35	(+0,06)	(+0,15)
23	Creuse	78,68	(+0,05)	(-0,02)	82,19	(+0,04)	(-0,37)
24	Dordogne	78,76	(+0,01)	(-0,24)	82,21	(-0,03)	(-0,87)
25	Doubs	78,59	(+0,06)	(+0,10)	82,11	(+0,05)	(+0,02)
26	Drôme	78,69	(+0,03)	(-0,07)	82,18	(+0,02)	(-0,36)
27	Eure	78,73	(+0,05)	(+0,07)	82,27	(+0,02)	(-0,06)
28	Eure-et-Loir	78,71	(+0,04)	(-0,01)	82,21	(+0,02)	(-0,29)
29	Finistère	78,84	(+0,06)	(+0,16)	82,40	(+0,06)	(+0,15)
30	Gard	78,65	(-0,09)	(-0,54)	82,03	(-0,22)	(-1,42)
31	Haute-Garonne	78,77	(+0,03)	(-0,13)	82,26	(+0,02)	(-0,56)
32	Gers	78,80	(+0,02)	(-0,21)	82,27	(-0,01)	(-0,74)
33	Gironde	78,86	(+0,02)	(-0,13)	82,36	(-0,02)	(-0,50)
34	Hérault	78,71	(-0,09)	(-0,53)	82,12	(-0,23)	(-1,39)
35	Ille-et-Vilaine	78,82	(+0,05)	(+0,10)	82,36	(+0,05)	(+0,03)
36	Indre	78,72	(+0,03)	(-0,17)	82,18	(-0,01)	(-0,73)
37	Indre-et-Loire	78,75	(+0,03)	(-0,12)	82,24	(+0,00)	(-0,54)
38	Isère	78,60	(+0,06)	(+0,10)	82,11	(+0,06)	(+0,02)
39	Jura	78,63	(+0,05)	(+0,05)	82,14	(+0,03)	(-0,10)
40	Landes	78,87	(+0,03)	(-0,13)	82,38	(-0,00)	(-0,51)
41	Loir-et-Cher	78,73	(+0,04)	(-0,10)	82,21	(+0,00)	(-0,54)
42	Loire	78,65	(+0,05)	(+0,01)	82,16	(+0,03)	(-0,23)
43	Haute-Loire	78,57	(+0,07)	(+0,13)	82,09	(+0,06)	(+0,07)
44	Loire-Atlantique	78,84	(+0,04)	(+0,01)	82,38	(+0,03)	(-0,19)
45	Loiret	78,70	(+0,03)	(-0,10)	82,19	(-0,04)	(-0,54)
46	Lot	78,72	(-0,01)	(-0,30)	82,15	(-0,08)	(-1,10)
47	Lot-et-Garonne	78,78	(+0,01)	(-0,22)	82,24	(-0,03)	(-0,76)
48	Lozère	78,56	(+0,07)	(+0,13)	82,08	(+0,07)	(+0,07)
49	Maine-et-Loire	78,80	(+0,03)	(-0,07)	82,31	(+0,01)	(-0,38)
50	Manche	78,81	(+0,06)	(+0,15)	82,36	(+0,06)	(+0,14)
51	Marne	78,70	(+0,03)	(+0,01)	82,22	(-0,03)	(-0,20)
52	Haute-Marne	78,65	(+0,02)	(+0,01)	82,17	(-0,04)	(-0,22)
53	Mayenne	78,77	(+0,05)	(+0,05)	82,30	(+0,03)	(-0,09)
54	Meurthe-et-Moselle	78,65	(+0,02)	(-0,00)	82,16	(-0,04)	(-0,24)
55	Meuse	78,65	(+0,03)	(+0,02)	82,17	(-0,04)	(-0,19)
56	Morbihan	78,84	(+0,06)	(+0,11)	82,38	(+0,05)	(+0,05)
57	Moselle	78,65	(+0,02)	(-0,02)	82,16	(-0,06)	(-0,30)
58	Nièvre	78,68	(+0,02)	(-0,08)	82,17	(-0,04)	(-0,48)
59	Nord	78,73	(+0,05)	(+0,12)	82,27	(+0,03)	(+0,08)

60	Oise	78,72	(+0,04)	(+0,07)	82,25	(+0,02)	(-0,05)
61	Orne	78,72	(+0,05)	(+0,10)	82,26	(+0,04)	(+0,02)
62	Pas-de-Calais	78,72	(+0,06)	(+0,14)	82,26	(+0,05)	(+0,12)
63	Puy-de-Dôme	78,62	(+0,06)	(+0,09)	82,14	(+0,06)	(-0,04)
64	Pyrénées-Atlantiques	78,84	(+0,05)	(+0,07)	82,38	(+0,05)	(-0,06)
65	Hautes-Pyrénées	78,60	(+0,06)	(+0,16)	82,13	(+0,06)	(+0,13)
66	Pyrénées-Orientales	78,75	(+0,05)	(+0,04)	82,28	(+0,04)	(-0,11)
67	Bas-Rhin	78,68	(+0,02)	(-0,04)	82,19	(-0,04)	(-0,31)
68	Haut-Rhin	78,65	(+0,03)	(+0,03)	82,17	(-0,00)	(-0,15)
69	Rhône	78,65	(+0,02)	(-0,16)	82,10	(-0,02)	(-0,67)
70	Haute-Saône	78,66	(+0,02)	(-0,03)	82,17	(-0,03)	(-0,31)
71	Saône-et-Loire	78,68	(+0,02)	(-0,08)	82,16	(-0,03)	(-0,46)
72	Sarthe	78,76	(+0,04)	(-0,07)	82,26	(+0,00)	(-0,42)
73	Savoie	78,25	(+0,09)	(+0,23)	81,76	(+0,10)	(+0,25)
74	Haute-Savoie	78,47	(+0,08)	(+0,18)	81,99	(+0,08)	(+0,18)
75	Paris	78,75	(+0,01)	(-0,17)	82,22	(-0,09)	(-0,72)
76	Seine-Maritime	78,75	(+0,05)	(+0,13)	82,30	(+0,05)	(+0,11)
77	Seine-et-Marne	78,71	(+0,02)	(-0,06)	82,22	(-0,05)	(-0,40)
78	Yvelines	78,72	(+0,04)	(-0,02)	82,23	(-0,00)	(-0,31)
79	Deux-Sèvres	78,77	(+0,03)	(-0,09)	82,27	(+0,02)	(-0,48)
80	Somme	78,72	(+0,05)	(+0,12)	82,26	(+0,04)	(+0,07)
81	Tarn	78,71	(-0,01)	(-0,27)	82,14	(-0,05)	(-0,96)
82	Tarn-et-Garonne	78,73	(-0,01)	(-0,31)	82,16	(-0,06)	(-1,02)
83	Var	78,75	(-0,07)	(-0,44)	82,17	(-0,14)	(-1,09)
84	Vaucluse	78,68	(-0,03)	(-0,36)	82,11	(-0,10)	(-1,00)
85	Vendée	78,85	(+0,04)	(+0,00)	82,38	(+0,02)	(-0,19)
86	Vienne	78,75	(+0,03)	(-0,13)	82,23	(+0,01)	(-0,59)
87	Haute-Vienne	78,74	(+0,03)	(-0,10)	82,23	(+0,00)	(-0,55)
88	Vosges	78,62	(+0,04)	(+0,05)	82,13	(+0,00)	(-0,10)
89	Yonne	78,69	(+0,01)	(-0,10)	82,18	(-0,08)	(-0,52)
90	Territoire de Belfort	78,65	(+0,03)	(+0,03)	82,16	(+0,00)	(-0,15)
91	Essonne	78,71	(+0,03)	(-0,09)	82,19	(-0,03)	(-0,52)
92	Hauts-de-Seine	78,74	(+0,01)	(-0,17)	82,20	(-0,07)	(-0,73)
93	Seine-Saint-Denis	78,75	(+0,01)	(-0,16)	82,22	(-0,08)	(-0,70)
94	Val-de-Marne	78,74	(+0,01)	(-0,17)	82,20	(-0,08)	(-0,73)
95	Val-d'Oise	78,74	(+0,03)	(-0,02)	82,25	(-0,02)	(-0,30)

Note:

Les écarts dans les scénarios 4.5 et 8.5 sont par rapport au scénario 2.6

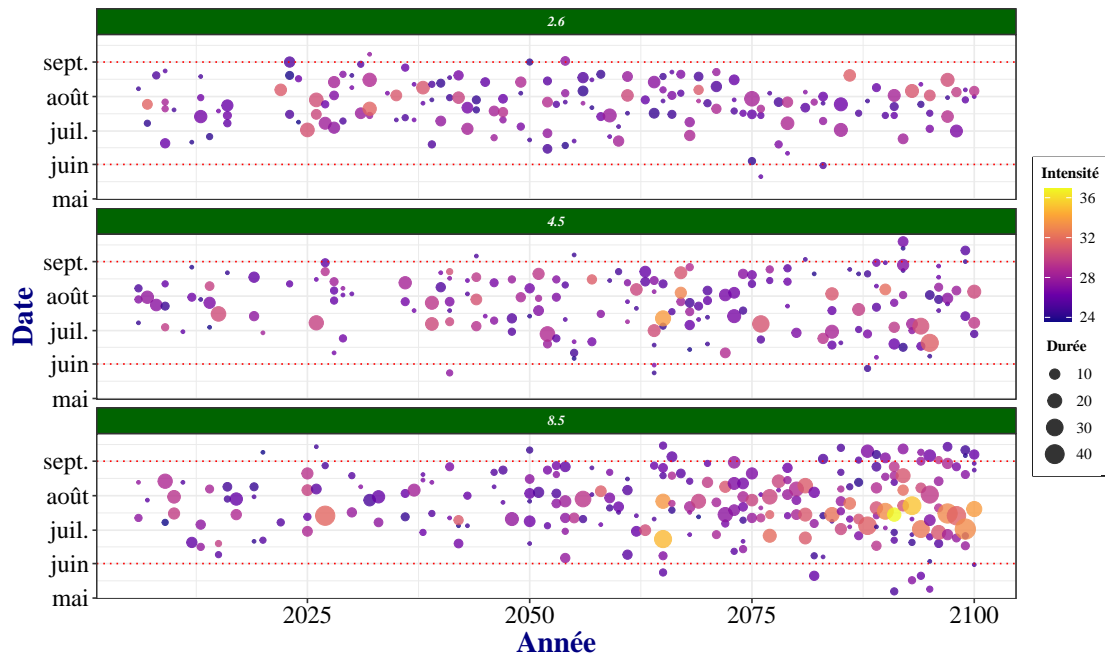


Figure A.4. : Les vagues de chaleur futures selon les scénarios

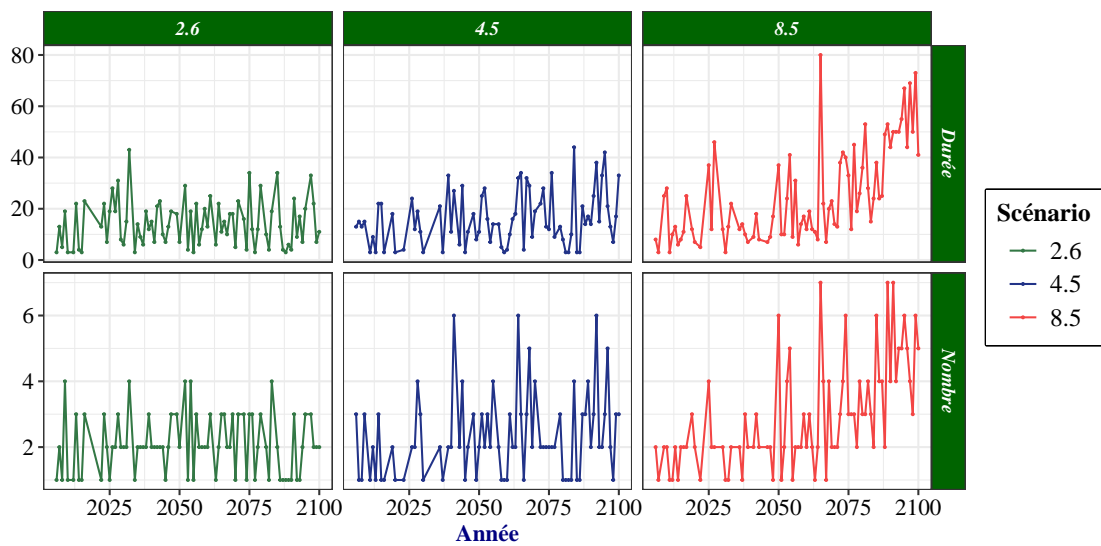


Figure A.5. : Nombre et durée annuel des vagues de chaleur selon les scénarios