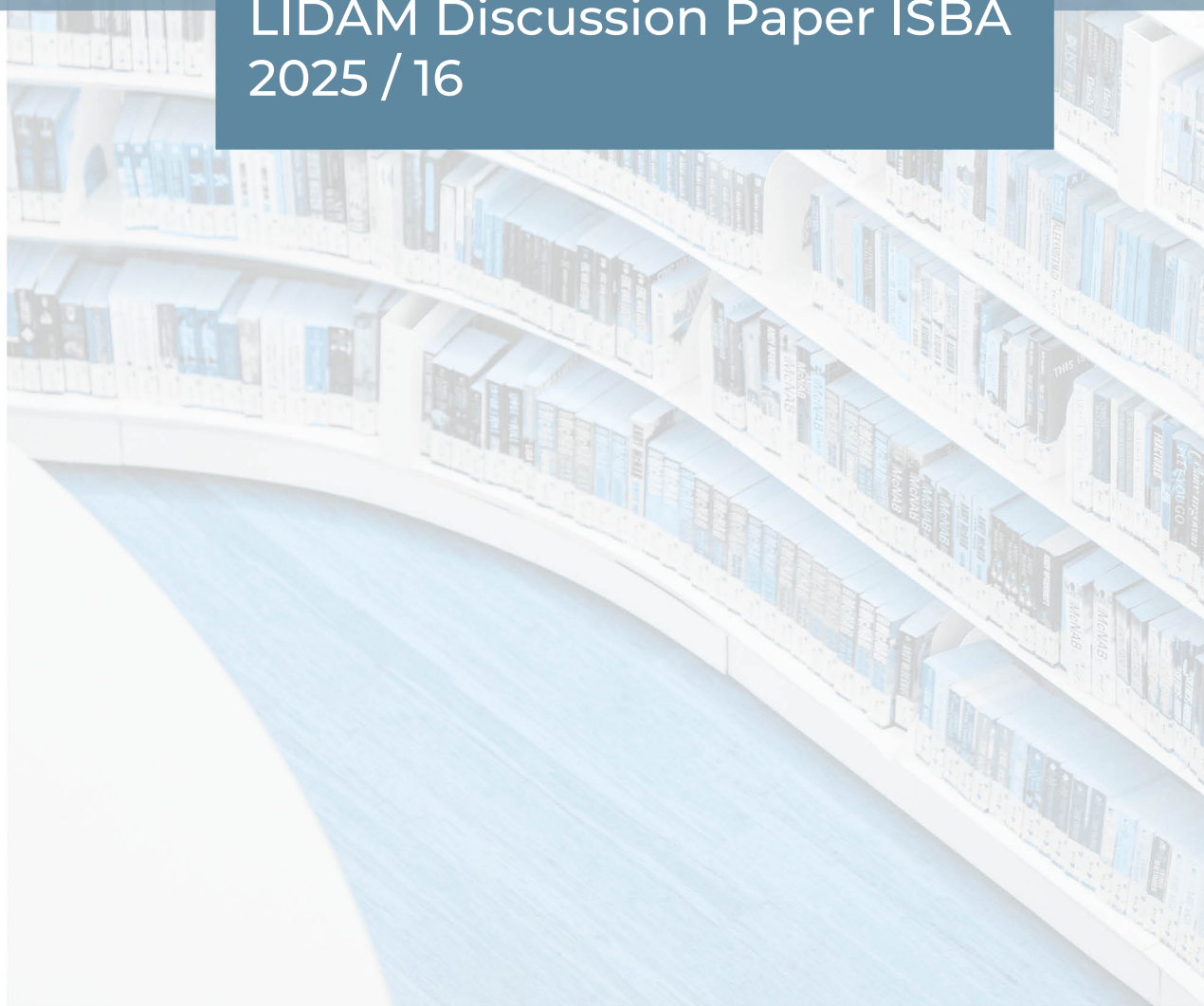
A PENALIZED DISTRIBUTED LAG NON-LINEAR LEE-CARTER FRAMEWORK FOR REGIONAL WEEKLY MORTALITY FORECASTING

Jens Robben, Karim Barigou

LIDAM Discussion Paper ISBA
2025 / 16



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

A Penalized Distributed Lag Non-Linear Lee-Carter Framework for Regional Weekly Mortality Forecasting

Jens Robben[1,*] and Karim Barigou[2]

[1] *Research Centre for Longevity Risk (RCLR), Faculty of Economics and Business, University of Amsterdam, Amsterdam, the Netherlands.*
[2] *Institute of Statistics, Biostatistics and Actuarial Science (ISBA), Louvain Institute of Data Analysis and Modeling (LIDAM), UCLouvain, Louvain-la-Neuve, Belgium*
[*] *Corresponding author: j.robben@uva.nl*

September 29, 2025

Abstract

Accurate forecasts of weekly mortality are essential for public health and the insurance industry. We develop a forecasting framework that extends the Lee–Carter model with age- and region-specific seasonal effects and penalized distributed lag non-linear components that capture the delayed and non-linear effects of heat, cold, and influenza on mortality. The model accommodates overdispersed mortality rates via a negative binomial distribution. We model the temporal dynamics of the latent factors in the model using SARIMAX processes and capture cross-regional dependencies through a copula-based approach. Using regional French mortality data (1990–2019), we demonstrate that the proposed framework yields well-calibrated forecast distributions and improves predictive accuracy relative to benchmark models. The results further show substantial heterogeneity in temperature- and influenza-related relative risks between ages and regions. These findings underscore the importance of incorporating exogenous drivers and dependence structures into a weekly mortality forecasting framework.

Keywords: stochastic mortality modeling; seasonal mortality; distributed lag non-linear models; excess mortality

1 Introduction

Life insurers, pension funds, and health care providers generally rely on mortality tables constructed from national vital statistics aggregated at an annual level (Wilmoth et al., 2007). However, since the emergence of the COVID-19 pandemic, more granular mortality data have become available worldwide (Wang et al., 2022). These new data sources allow for extensions of traditional actuarial mortality models to capture finer temporal and geographical variation. Mortality rates show, for example, a seasonal pattern within a year, with higher rates in winter than in summer (Madaniyazi et al., 2022; Pavía & Lledó, 2022). Deviations from these seasonal patterns can occur due to various exogenous events such as heat waves, cold spells, or epidemic outbreaks (Gasparrini & Armstrong, 2011; Gasparrini et al., 2015). Therefore, more fine-grained models are valuable not only for insurers to better price short-term life-contingent insurance

products, but also for public health authorities, who can use them to assess and monitor excess mortality in near real time (Fernández-Durán & Gregorio-Domínguez, 2015; Karlinsky & Kobak, 2021).

Building on the well-known Lee–Carter framework for annual, country-level data (Lee & Carter, 1992), we introduce a stochastic mortality projection model that operates at a much finer spatiotemporal resolution: weekly, regional, and age specific. Our proposed model captures both seasonal mortality patterns and short-term deviations from it driven by extreme temperatures and influenza outbreaks. To achieve this, we embed distributed lag non-linear models (DLNMs) within the framework (Gasparrini et al., 2010), which enables the estimation of age-specific, delayed, and non-linear effects of these shocks on seasonal mortality. By integrating DLNMs into our model, our framework provides a flexible and interpretable tool to disentangle seasonal baselines from weather- and epidemic-related excess mortality across regions and age groups. Furthermore, the proposed model retains the ability to produce stochastic mortality projections comparable to those of traditional stochastic mortality models.

The study of seasonal mortality and the influence of exogenous factors such as heatwaves, cold spells, and influenza outbreaks has recently gained increasing attention in the actuarial literature. To our knowledge, Seklecka et al. (2017) is among the first to incorporate a temperature-related factor into the Lee–Carter mortality model. Their study, based on grouped monthly and annual data from the United Kingdom, shows improved predictive accuracy and applies the model to life insurance pricing. However, the use of aggregated data prevents the model from capturing regional heterogeneity or fine-scale temporal patterns. Next, H. Li and Tang (2022) apply concepts from the extreme value theory to investigate the dependence between monthly temperature extremes and mortality at the regional level in the United States. They find a particularly strong dependence between cold-weather extremes and mortality at older ages. However, their focus lies on describing extremal dependence rather than developing a stochastic mortality forecasting framework. Therefore, more recently, Guibert et al. (2025) extend the Li–Lee multi-population model (N. Li & Lee, 2005) by integrating a distributed lag non-linear model to capture non-linear and delayed temperature effects. They also provide mortality projections under alternative representative concentration pathway (RCP) scenarios. Although this constitutes one of the first actuarial attempts to link mortality forecasts to climate scenarios, their projections remain at the annual and country-specific level. Building upon this line of research, Min et al. (2025) combine single- and multi-population stochastic mortality models with a DLNM to disentangle long-term mortality trends from non-linear and lagged climate effects. However, their baseline remains annual, which precludes statements about short-term excess mortality caused by temperature shocks. Parallel to these developments, Bégin et al. (2025) propose an extension of the age–period–cohort (APC) model with a seasonal component based on periodic cubic splines. This allows them to generate mortality forecasts on the daily scale while accounting for seasonal fluctuations. Although this represents a step toward fully stochastic seasonal mortality projections, their framework does not incorporate exogenous factors or a multi-population structure.

The epidemiological and environmental health literature has repeatedly demonstrated that DLNMs offer a flexible and interpretable way to model short-term, non-linear relationships between temperature and mortality. For example, Gasparrini et al. (2015) estimate that 7.7% of deaths can be attributed to suboptimal temperatures, based on 74 million deaths in 384 locations in 13 countries. Guo et al. (2017) also show that heat waves significantly increased mortality at higher temperature thresholds in more than 400 regions in 18 countries. In addition to temperature, DLNMs have also been applied to model the association between infectious diseases and mortality: L. Li et al. (2023) showed that influenza caused significant excess mortality in

Guangzhou, mainly due to respiratory and cardiovascular deaths among the elderly. Although DLNMs capture associations with a single risk factor, they are rarely integrated into a broader mortality modeling framework that can jointly account for demographic structure, temporal trends, and spatial dependence. Moreover, most applications in the epidemiological literature do not use forward-looking stochastic mortality forecasts and are often estimated using age aggregated data, which is not suitable for actuarial applications.

We contribute to the literature in four ways. First, from the National Institute of Statistics and Economic Studies (INSEE), we collect and analyze a rich data set of individual death counts that cover all metropolitan regions of France from 1990 to 2019. Previous studies typically rely on shorter observation periods (e.g., 2015–2019 in Min et al. (2025)) or on a single region (e.g., Quebec in Bégin et al. (2025)). Because public health authorities generally require regional-level assessments of excess mortality for the entire nation, our collected data set, methodology, and application provide a framework for doing so. Second, we explicitly consider the effect of influenza outbreaks on mortality, an aspect that is often overlooked in existing approaches. By incorporating this into our modeling framework, we can better capture excess mortality in winter, which can arise not only from cold spells as in current approaches, but also from (influenza) epidemics. To our knowledge, in addition to our recent contribution in Robben, Barigou, and Kleinow (2025), we are the first to incorporate influenza effects into seasonal stochastic mortality projection models. Third, we make a methodological contribution by creating similarity in the estimated temperature and influenza effects between adjacent regions. Specifically, we integrate a Laplacian spatial penalty matrix into the log-likelihood of our model, which enforces the smoothing of effects across regions that share borders. This spatial regularization step makes the model more robust by drawing knowledge from geographically close regions. Fourth, we provide detailed assessments of the uncertainty of the parameters. Our model is calibrated via maximum likelihood estimation, and we calculate the observed Fisher information matrix (and thus the variance-covariance matrix) fully analytically. This allows for precise conclusions about the uncertainty and significance of region-specific effects and overall mortality dynamics.

The structure of this paper is as follows. Section 2 sets out the notation and provides a detailed description of the data sources, regarding mortality, population, temperature, and influenza surveillance, accompanied by an initial exploratory analysis. Section 3 develops the proposed mortality modeling framework, which is specified at the weekly, regional, and age-specific level, and is designed to capture seasonal variation, as well as the impact of heat waves, cold spells, and influenza outbreaks. Section 4 details the calibration strategy based on the maximum likelihood estimation, the procedure to generate forecasts, and the evaluation of the uncertainty of the parameters. Section 5 presents a comprehensive case study to illustrate the application of the methodology to the data introduced. Section 6 concludes with a summary of the main findings.

2 Notations and data

2.1 Notations

Let $d_{x,t,w,r}$ denote the weekly death count in region r for age group x at week w of year t. The time points (t, w) follow the ISO 8601 standard, where each ISO week is exactly 7 days, and ISO week 1 starts on the Monday of the week that contains the first Thursday of the year. We define the following sets: $\mathcal{R}$ is the set of all regions, $\mathcal{X}$ the set of age groups, $\mathcal{W} = \{1, 2, \ldots, 52\}$ the range of weeks, and $\mathcal{T} = \{t_{\min}, t_{\min+1}, \ldots, t_{\max}\}$ the range of all years under consideration.

Exposure to risk $E_{x,t,w,r}$ represents the total number of person years lived by people aged $[x, x+1)$ during week w of year t in region r. The observed weekly death rate is then given by:

$$m_{x,t,w,r} = \frac{d_{x,t,w,r}}{E_{x,t,w,r}}. \tag{2.1}$$

We refer to $\mu_{x,t,w,r}$ as the weekly force of mortality for people aged x in week w of year t in region r. This quantity represents the instantaneous risk of death at (x, t, w, r). Under the assumption of constant force of mortality within a week interval, $\mu_{x,t,w,r}$ can be approximated by the observed weekly death rate $m_{x,t,w,r}$. For ease of notation, gender is omitted from the notation, but all methods can be applied to both genders.

2.2 Mortality data

Death counts. We use individual-level French death records from the National Institute of Statistics and Economic Studies (INSEE).[1] Each record contains information on the name, gender, commune of birth, commune of death, date of birth, date of death, and death certificate number. We extract death records spanning from 1970 to the first quarter of 2025, resulting in approximately 25.5 million records.

We perform the following data preprocessing steps: (1) remove duplicate entries, (2) exclude records with missing birth or death dates, (3) exclude records where the date of death precedes the date of birth, and (4) impute missing or misspecified months or days of death by sampling from the empirical distribution of deaths by month and day. Next, we aggregate the cleaned death records by 5-year age group, ISO year and week, administrative region of death, and gender. To do so, we first map the commune of death to its corresponding department by extracting the first two digits of the reported postal code. We then map departments to administrative regions using the matching table provided by INSEE.[2]

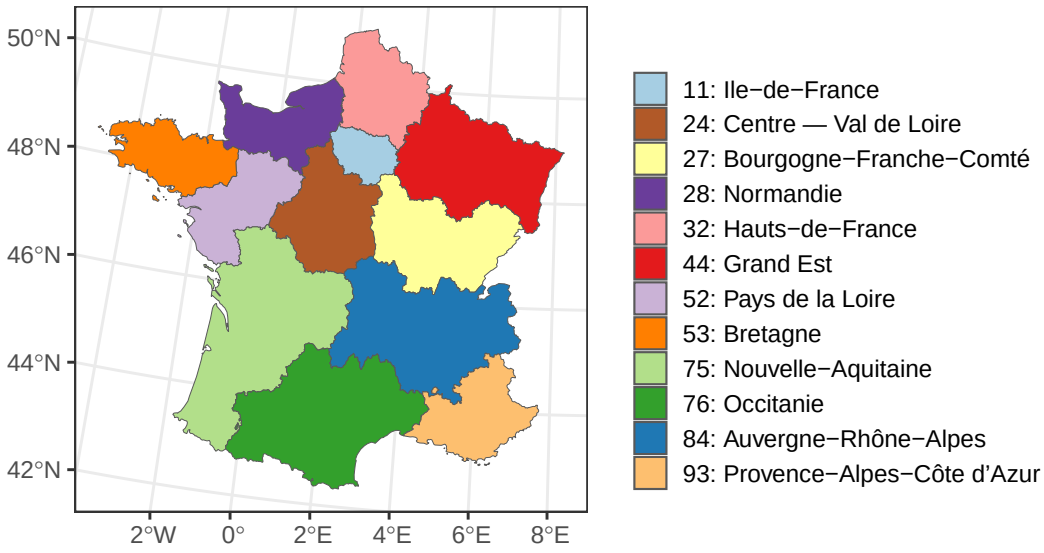


Figure 1: The 12 administrative regions in Metropolitan France.

In this paper, we examine the 30-year period from 1990 to 2019, and focus on five-year age groups beginning at age 50 and including an open-ended 95+ age category. Our analysis covers

[1] The database is available at https://www.insee.fr/fr/information/4769950.
[2] The matching table is available at https://www.insee.fr/fr/information/7766585.

the twelve administrative regions of Metropolitan France, excluding Corsica. Figure 1 displays these twelve regions, along with their respective geographic INSEE codes and names. The left panel of Figure 2 shows the stacked female death counts for the 12 administrative regions, aggregated among the different age groups considered.

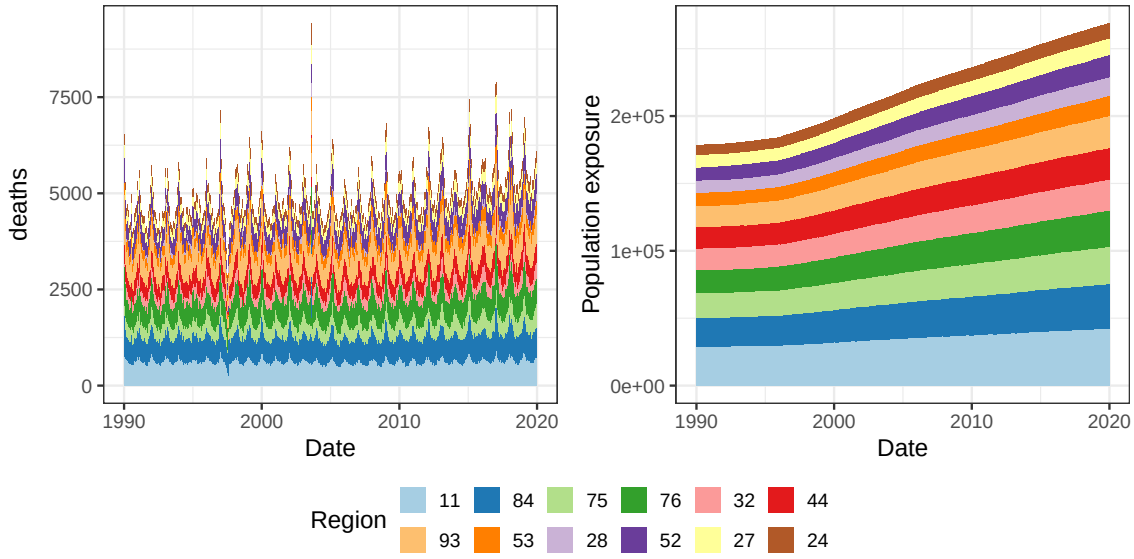


Figure 2: Stacked female death counts (left panel) and exposures (right panel) per week for the twelve French administrative regions over the years 1990-2019. The weekly death counts and exposures per region are aggregated across the different age groups.

Next, we analyze the overdispersion in our data set of weekly death counts across different regions and age groups. Compared to annual country-level death statistics typically used in actuarial mortality models, our data exhibit significantly greater overdispersion. Table 1 presents the variance-to-mean ratio, also called the dispersion index, for each region and age group in this study. The results show that overdispersion is consistently present (index > 1) in all strata. Although younger age groups exhibit relatively mild overdispersion, the dispersion index increases substantially for older age groups. Furthermore, the dispersion index varies considerably by region, with southern regions generally exhibiting higher overdispersion compared to northern ones, except for Île-de-France. This regional variability may be related to the increased climate variability in the southern French regions.

Population exposure. To compute weekly population exposures by age, sex, and region, we first obtain annual population counts $P_{x,t,r}$ from INSEE.[3] These counts are recorded on January 1st of each year t, for each age x and administrative region r, and we omit the dependence on sex to ease notation. To estimate population counts at a weekly time scale, we perform linear interpolation between subsequent years as follows:

$$\tilde{P}_{x,t,w,r} = P_{x,t,r} + \frac{\text{days}(\text{date}_t, \text{date}_{t,w})}{\text{days}(\text{date}_t, \text{date}_{t+1})} \left(P_{x,t+1,r} - P_{x,t,r}\right),$$

where date_t denotes January 1st of year t, $\text{date}_{t,w}$ denotes the start date of the w-th ISO week in year t, and $\text{days}(\cdot,\cdot)$ is the function that calculates the number of days between two given

[3]We obtain the annual population counts from the following database: https://www.insee.fr/fr/statistiques/1893198.

Age \ Region	11	24	27	28	32	44	52	53	75	76	84	93
50-54	1.29	1.14	1.10	1.23	1.31	1.17	1.10	1.09	1.30	1.22	1.14	1.12
55-59	1.34	1.10	1.17	1.18	1.42	1.25	1.12	1.18	1.36	1.44	1.28	1.18
60-64	1.44	1.16	1.13	1.27	1.54	1.42	1.30	1.27	1.41	1.42	1.34	1.23
65-69	1.87	1.31	1.39	1.35	2.07	1.81	1.39	1.37	1.71	1.58	1.69	1.50
70-74	2.57	1.48	1.74	1.58	2.89	2.32	1.41	1.84	2.10	1.94	2.32	1.88
75-79	3.91	1.90	2.14	1.89	3.84	3.04	1.95	2.34	2.97	2.67	3.19	2.71
80-84	5.76	2.20	2.58	2.24	3.95	3.61	2.20	2.51	3.99	3.33	4.07	2.91
85-89	6.42	2.99	3.60	3.16	5.73	5.53	3.74	3.80	5.51	5.03	6.30	4.40
90-94	8.66	4.41	4.92	5.17	7.66	8.26	6.17	7.26	9.28	9.13	10.40	7.36
95+	8.49	4.36	4.86	5.22	5.80	6.63	6.22	6.73	9.83	9.32	9.85	8.54

Table 1: Dispersion index of the weekly female death counts per considered age group and administrative region in Metropolitan France.

dates. We then compute weekly exposures by taking the midpoint:

$$E_{x,t,w,r} = \frac{1}{2 \cdot 52.18} \left(\tilde{P}_{x,t,w,r} + \tilde{P}_{x,t,w+1,r} \right).$$

where the time point $(t, w+1)$ corresponds to $(t+1, 1)$ if the former time point does not exist. The division by the average number of weeks in a year ensures that the weekly exposure measure is expressed in units of person-years. The right panel of Figure 2 shows the stacked female exposures for the 12 administrative regions of metropolitan France over the years 1990-2019, aggregated across the different age groups considered. Due to linear interpolation, the weekly exposure estimates vary smoothly over the different weeks. This approach improves existing methods, which, for example, often assume that weekly exposures remain constant throughout the year (Jdanov et al., 2021).

Death rates. We compute the death rates following the formula in Eq. (2.1). Figure 3 illustrates the weekly death rates, on the logarithmic scale, for women in the administrative region of Île-de-France during the period 1990-2019. The left panel displays the log death rates for five selected age groups (50-54, 60-64, 70-74, 80-84, and 90-94) over time and reveals clear differences in the mortality level between ages. The severe excess mortality in the summer of 2003 corresponds to the 2003 European heatwave that particularly hit France. The right panel shows the average log death rate per week, aggregated over all years and centered around zero. This panel highlights the seasonal pattern with higher mortality during the winter weeks and a trough in the summer. Here, it is also clear that the seasonal pattern is more pronounced for the older age groups.

2.3 Temperature data

We obtain temperature data from the E-OBS dataset, a land-only, gridded meteorological dataset available on the Copernicus Climate Data Store (CDS, 2020). This data set provides daily weather variables for Europe with high spatial resolution ($0.1^\circ \times 0.1^\circ$, 10 km), and is derived from interpolated observations from a large network of European meteorological stations. Following the approach of Gasparrini et al. (2015), we focus on the daily average temperature as our exposure metric. Alternative measures (e.g., maximum or minimum temperature) have not been shown in the literature to have a better predictive power of mortality (Barnett,

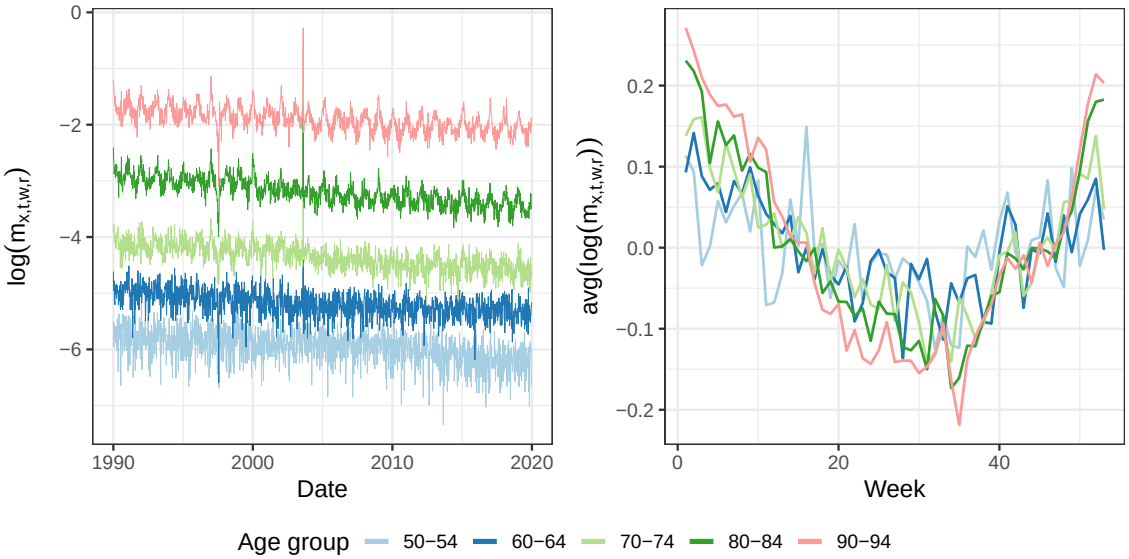


Figure 3: Left panel: the logarithm of the weekly death rates for females in the administrative region of Ile-de-France for 5 different age groups 50-54, 60-64, 70-74, 80-84, and 90-94, across the years 1990-2019. Right panel: the average of the log death rates per week across the years 1990-2019, centered around zero for comparability reasons.

2010). Henceforth, we extract daily average temperatures across a spatial grid that encompasses metropolitan France from 1990 to 2019. For reference, the left panel of Figure 4 illustrates the spatial distribution of temperatures on a randomly selected date during this period.

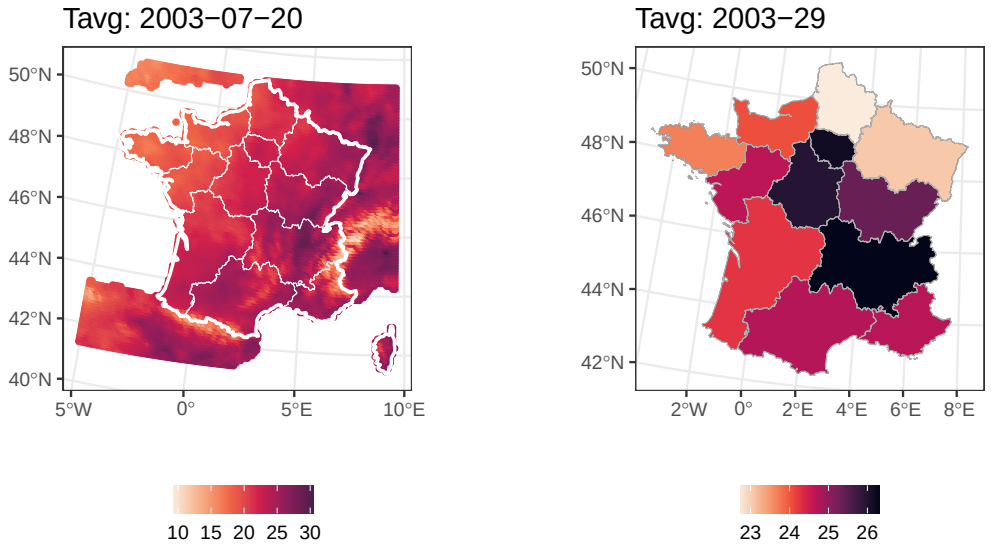


Figure 4: Left panel: the daily average temperature on a spatial grid covering metropolitan France at July 20, 2003. Right panel: the weekly average of the daily average temperature values in the 29th ISO week of the year 2003 per administrative region.

Since the mortality data in Section 2.2 is defined at the level of administrative regions, we aggregate the daily temperature data into population-weighted regional averages using gridded population data (CIESIN et. al, 2018). This weighting principle is described in Robben, Antonio, and Kleinow (2025) in more detail and ensures that temperature values better represent areas

with a higher population density within each region. We then calculate the weekly average of these daily temperature values to match the temporal resolution of the mortality data. The right panel of Figure 4 displays the resulting processed data for a randomly selected ISO week and year.

2.4 Influenza-like illness surveillance data

We use data on influenza-like illness (ILI) from the French Sentinelles Network (Valleron et al., 1986). This is a nationwide epidemiological surveillance system established in 1984, coordinated by the Pierre Louis Institute of Epidemiology and Public Health under Inserm and Sorbonne University. It monitors communicable diseases, including ILI, through a network of volunteer general practitioners in mainland France. In 2024, the network consisted of 135 general practitioners who report weekly data on ILI cases.

We obtain the weekly incidence rate of influenza-like illness (ILI) per administrative region in France from the French Sentinelles network. These incidence rates represent the number of new cases of ILI per unit of population. Figure 5 displays the trends in the incidence of ILI in three French administrative regions. Although the overall patterns are broadly similar, the intensity of influenza outbreaks varies significantly between regions.

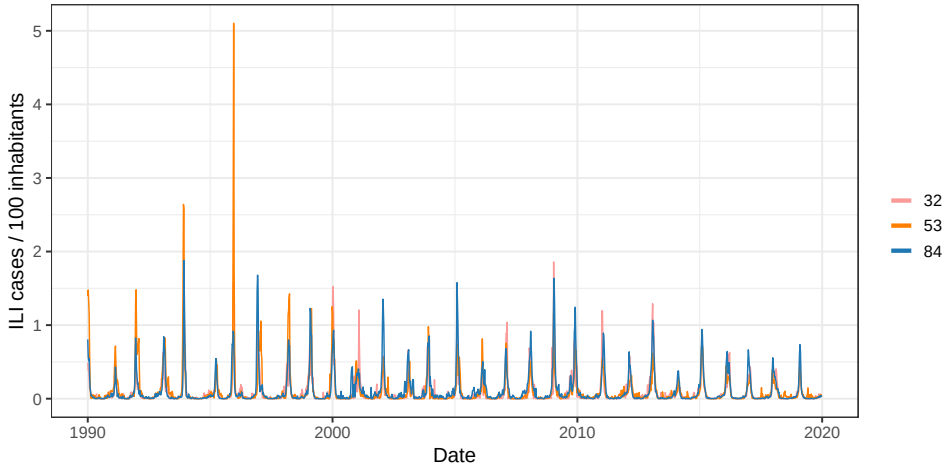


Figure 5: The weekly incidence rates of influenza-like illness in three administrative regions in France: Hauts-de-France (32), Bretagne (53), and Auvergne-Rhône-Alpes (84) from 1990-2019.

3 Model specification

We develop a weekly, region, and age-specific mortality model designed to (1) capture seasonal variations in weekly mortality rates and (2) quantify the effects of heat waves, cold spells, and influenza outbreaks on deviations from these seasonal mortality patterns. Our proposed model extends the traditional Lee-Carter framework (Lee & Carter, 1992), typically applied to country-level annual mortality data, to a finer spatiotemporal resolution.

3.1 Model structure

To account for the overdispersion in death counts (see Table 1 in Section 2.2), we model the number of deaths, $D_{x,t,w,r}$, for the age group $x \in \mathcal{X}$, year $t \in \mathcal{T}$, week $w \in \mathcal{W}$, and region $r \in \mathcal{R}$,

using a Negative Binomial (NB) distribution:

$$D_{x,t,w,r} \sim \text{NB}(E_{x,t,w,r} \cdot \mu_{x,t,w,r}, \phi_{x,r}), \tag{3.1}$$

where $E_{x,t,w,r}$ is the exposure to risk, $\mu_{x,t,w,r}$ is the weekly force of mortality, and $\phi_{x,r}$ is the age- and region-specific dispersion parameter that captures extra-Poisson variability within the death counts. The NB model was considered in a frequentist setting by Delwarde et al. (2007) and in a Bayesian setting by Wong et al. (2018) and Barigou et al. (2023) among others.

We express the logarithm of the weekly force of mortality as the sum of two additive components. The first, the seasonal baseline component, represents the expected seasonal mortality pattern in the absence of exogenous shocks. The second captures short-term deviations from this baseline due to extreme temperatures and influenza outbreaks. Formally, we write:

$$\log \mu_{x,t,w,r} = \underbrace{\alpha_{x,r} + \beta_x \kappa_{t,r} + \gamma_x \lambda_{w,r}}_{\text{Seasonal baseline}} + \underbrace{\delta_x f_r^{(1)}(\text{Tavg}_{t,w,r}) + \epsilon_x f_r^{(2)}(\text{ILI}_{t,w,r})}_{\text{Deviations}}, \tag{3.2}$$

where:

- $\alpha_{x,r}$ reflects the region-specific average log mortality for the age group x. We make this parameter region dependent to allow for overall differences in mortality between regions, e.g., due to socio-economic differences,
- $\kappa_{t,r}$ is a region-specific mortality index that captures the overall annual mortality decline, and β_x measures the age-specific sensitivity to this decline,
- $\lambda_{w,r}$ is a region-specific week effect capturing periodic week-to-week variations in mortality (higher mortality in winter compared to summer), and γ_x accounts for differences in this seasonal intensity between different age groups,
- $f_r^{(1)}(\text{Tavg}_{t,w,r})$ is a region-specific distributed lag non-linear model (DLNM) that describes the non-linear and delayed association between the weekly average temperature $\text{Tavg}_{t,w,r}$ and mortality, and δ_x models the age-specific effect of it,
- $f_r^{(2)}(\text{ILI}_{t,w,r})$ is a region-specific DLNM that models the non-linear and delayed association between weekly average influenza-like illness (ILI) and mortality, and ϵ_x models the age-specific impact of it.

We model the dispersion parameter on the logarithmic scale as an additive combination of age- and region-specific components, i.e.,

$$\phi_{x,r} = \exp\left(\phi_x + \phi_r\right). \tag{3.3}$$

This decomposition enhances both the model parsimony and interpretability. We ensure identifiability in the additive composition of the dispersion parameter by imposing $\sum_{x \in \mathcal{X}} \phi_x = 0$.

In summary, our proposed model extends the Lee-Carter modeling framework by adding the following:

1. age-specific weekly effects to capture seasonal mortality patterns,
2. DLNMs to account for the age-specific, non-linear, and lagged effects of temperature and influenza on mortality.

Furthermore, our model is defined in a multi-population setting where we use a common age effect across different regions, as proposed in Kleinow (2015).

3.2 Identifiability constraints

To ensure identifiability in the model structure of Eq. (3.2), we impose the following constraints on the respective parameters:

$$\begin{aligned} &\beta_{x_{\min}} = 1,\ \gamma_{x_{\min}} = 1,\ \delta_{x_{\min}} = 1,\ \epsilon_{x_{\min}} = 1 \\ &\kappa_{t_{\min},r} = 0,\ \lambda_{t_{\min},r} = 0 \quad \forall r \in \mathcal{R}, \end{aligned} \tag{3.4}$$

where $x_{\min}$ and $t_{\min}$ are the minimum age and year values in the sets $\mathcal{X}$ and $\mathcal{T}$, respectively. Since $\kappa_{t,r}$ and $\lambda_{w,r}$ are defined on different time scales (years and weeks), no additional constraints (e.g., orthogonality) are needed between them or their associated age effects. In total, we impose $2 \cdot |\mathcal{R}| + 4$ constraints. The constraints in Eq. (3.4) differ from the standard Lee-Carter constraints but allow for an easier computation of standard errors for the model parameters, as discussed in Section 4.1. However, all methods and estimation procedures can be performed using straightforward extensions of the typical Lee-Carter constraints.

3.3 Distributed lag non-linear models

We model the non-linear and delayed effects of temperature and influenza on mortality in Eq. (3.2) using a distributed lag non-linear model (DLNM). This approach is well-established in the epidemiological literature for assessing covariate-response associations with delayed effects, such as temperature-mortality associations. Mathematically, we can express the DLNM for a specific region r as:

$$f_r(\xi_{t,w,r}) = \sum_{\ell=0}^{L} \sum_{j=1}^{v_\xi} \sum_{k=1}^{v_l} \eta_{jk,r} \cdot b_j(\xi_{t,w-\ell,r}) \cdot c_k(\ell), \tag{3.5}$$

where $\xi_{t,w,r}$ refers to the covariate value (temperature or ILI) at the time point (t, w) in the region r, $\ell = 0, 1, \ldots, L$ denote the lag values and L the maximum lag considered. In addition, $\eta_{jk,r}$ are the parameter coefficients for the region r, $b_j(\xi_{t,w-\ell,r})$ refers to the ℓ-week lagged covariate value transformed through its jth basis function, and $c_k(\ell)$ is the kth basis function for the lag dimension evaluated at lag value ℓ. Basis functions are smooth functions that capture non-linear relationships. Note that when $w - \ell < 1$, i.e., the lag extends to the previous year, the covariate value $\xi_{t,w-\ell,r}$ should be pulled at week $52 - (\ell - w)$ of year $t - 1$. Lastly, we refer to v_ξ and v_l as the number of basis functions for the covariate and lag dimensions. Importantly, the basis dimensions v_ξ and v_l, which determine the flexibility of the covariate and lag effects, are the same for all regions.

To simplify the notation, we introduce a cross-basis representation. Hereto, we denote, for every time point (t, w) and region r, $\boldsymbol{Z}_{t,w,r} \in \mathbb{R}^{v_\xi \cdot v_l}$ as the vector with entries:

$$Z_{t,w,jk,r} := \sum_{\ell=0}^{L} b_j(\xi_{t,w-\ell,r}) \cdot c_k(\ell), \tag{3.6}$$

for $j \in \{1, 2, ..., v_\xi\}$ and $k \in \{1, 2, .., v_l\}$. Furthermore, let $\boldsymbol{\eta}_r := (\eta_{jk,r})_{j,k} \in \mathbb{R}^{v_\xi \cdot v_l}$ be the parameter vector for region r. We can then write the DLNM from Eq. (3.5) as:

$$f_r(\xi_{t,w,r}) = \boldsymbol{Z}'_{t,w,r} \boldsymbol{\eta}_r, \tag{3.7}$$

where $'$ refers to the transpose operator. The cross-basis matrix $\boldsymbol{Z}_r \in \mathbb{R}^{T \times (v_\xi \cdot v_l)}$ stacks all vectors $\boldsymbol{Z}_{t,w,r}$ as rows, where $T := |\mathcal{T}|$ is the total number of time points (t, w) considered.

In the context of Equation (3.2), we define the cross-basis matrices for the weekly average temperature (Tavg) and the weekly incidence rates of influenza-like illnesses (ILI) as $\boldsymbol{Z}_{1,r}$ and $\boldsymbol{Z}_{2,r}$, respectively, for each region $r \in \mathcal{R}$. We denote the corresponding parameter vectors by $\boldsymbol{\eta}_{1,r}$ and $\boldsymbol{\eta}_{2,r}$. Note that the cross-basis matrices $\boldsymbol{Z}_{1,r}$ and $\boldsymbol{Z}_{2,r}$ may differ in dimension, since we allow for a different maximum lag and number and type of basis functions in both the covariate and lag dimensions for temperature and ILI.

3.4 Relative risk

The concept of relative risk (RR) is a useful way to quantify the effect of the DLNM components in Eq. (3.2). At a given covariate level ξ, i.e., at a specific weekly average temperature or influenza-like illness incidence, the RR quantifies the factor by which the baseline force of mortality increases (if RR > 1) or decreases (if RR < 1) relative to a particular reference value. This RR captures both non-linear and delayed effects modeled by the DLNM.

For temperature and influenza incidence, we compute the RR curves for the region $r \in \mathcal{R}$ and age group $x \in \mathcal{X}$ as:

$$\begin{aligned} \mathrm{RR}_{\mathrm{Tavg},x,r}(\xi) &= \exp\left(\delta_x \left[f_r^{(1)}(\xi) - f_r^{(1)}(\xi_{\mathrm{ref}})\right]\right), \quad \xi \in \Xi_{\mathrm{Tavg},r}, \\ \mathrm{RR}_{\mathrm{ILI},x,r}(\xi) &= \exp\left(\epsilon_x \left[f_r^{(2)}(\xi) - f_r^{(2)}(\xi_{\mathrm{ref}})\right]\right), \quad \xi \in \Xi_{\mathrm{ILI},r}, \end{aligned} \tag{3.8}$$

where $\Xi_{\mathrm{Tavg},r}$ and $\Xi_{\mathrm{ILI},r}$ denote grids of covariate values for temperature and influenza incidence in region r, respectively. The parameters δ_x and ϵ_x, for $x \in \mathcal{X}$, are the corresponding age-specific effects in Eq. (3.2), and ξ_{ref} denotes a reference value for the weekly average temperature or influenza-like illness incidence. From Eq. (3.7), we obtain

$$f_r^{(i)}(\xi) = \boldsymbol{Z}'_{i,r}(\xi)\boldsymbol{\eta}_{i,r},$$

where $\boldsymbol{Z}_{i,r}(\xi)$ is defined through Eq. (3.6) by fixing the covariate value to ξ over the different lags.

Using the delta method, we are able to calculate the variance of the relative risk as follows:

$$\begin{aligned} \mathrm{Var}\left(\mathrm{RR}_{\mathrm{Tavg},x,r}(\xi)\right) &= \left[\delta_x \mathrm{RR}_{\mathrm{Tavg},x,r}\right]^2 \boldsymbol{Z}'_{1,r}(\xi;\, \xi_{\mathrm{ref}})\, \boldsymbol{\Sigma}(\boldsymbol{\eta}_{1,r})\, \boldsymbol{Z}_{1,r}(\xi;\, \xi_{\mathrm{ref}}), \quad \xi \in \Xi_{\mathrm{Tavg},r} \\ \mathrm{Var}\left(\mathrm{RR}_{\mathrm{ILI},x,r}(\xi)\right) &= \left[\epsilon_x \mathrm{RR}_{\mathrm{ILI},x,r}\right]^2 \boldsymbol{Z}'_{2,r}(\xi;\, \xi_{\mathrm{ref}})\, \boldsymbol{\Sigma}(\boldsymbol{\eta}_{2,r})\, \boldsymbol{Z}_{2,r}(\xi;\, \xi_{\mathrm{ref}}), \quad \xi \in \Xi_{\mathrm{ILI},r}, \end{aligned} \tag{3.9}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\eta}_{1,r})$ and $\boldsymbol{\Sigma}(\boldsymbol{\eta}_{2,r})$ are the covariance matrices of $\boldsymbol{\eta}_{1,r}$ and $\boldsymbol{\eta}_{2,r}$ respectively, which can be derived from the Hessian of the negative penalized log-likelihood, see Section 4.1. Furthermore, for $i = 1, 2$, we define $\boldsymbol{Z}_{i,r}(\xi;\, \xi_{\mathrm{ref}})$ as the vector with the following entries:

$$\boldsymbol{Z}_{i,jk,r}(\xi;\, \xi_{\mathrm{ref}}) := \sum_{\ell=0}^{L} \left(b_{i,j}(\xi) - b_{i,j}(\xi_{\mathrm{ref}})\right) \cdot c_{i,k}(\ell),$$

where $b_{i,j}(\cdot)$ and $c_{i,k}(\cdot)$ denote the j-th basis function for the covariate space (temperature for $i = 1$, influenza for $i = 2$) and the k-th basis function for the lag dimension, respectively.

4 Model calibration and forecasting

4.1 Calibration

Log-likelihood. We calibrate the model parameters in Eq. (3.2) using maximum likelihood estimation. Hereto, we set up the negative binomial log-likelihood based on the distributional

assumption in Eq. (3.1). From the mean parametrization, we obtain the following log-likelihood:

$$l_{\text{nb}}(\boldsymbol{\theta}) = \sum_{x,t,w,r} \Big[\log\Gamma(d_{x,t,w,r} + \phi_{x,r}) - \log\Gamma(\phi_{x,r}) - \log\Gamma(d_{x,t,w,r}+1) + d_{x,t,w,r} \cdot \log(E_{x,t,w,r} \cdot \mu_{x,t,w,r}) + \phi_{x,r} \cdot \log\phi_{x,r} - (d_{x,t,w,r} + \phi_{x,r}) \cdot \log(E_{x,t,w,r} \cdot \mu_{x,t,w,r} + \phi_{x,r}) \Big], \tag{4.1}$$

where $\mu_{x,t,w,r}$ is the model structure for the weekly force of mortality in Eq. (3.2), $\phi_{x,r}$ is the parameterized dispersion parameter in Eq. (3.3), and $\boldsymbol{\theta}$ denotes the vector of all parameters to be estimated.

To reduce model complexity and enforce spatial coherence in the DLNM components, we add a penalty term to the log-likelihood that smoothens the DLNM parameters across neighboring regions, i.e., regions that share a border. Hereto, we introduce the Laplacian matrix $\boldsymbol{L}$, which is a matrix representation of a graph that encodes the spatial adjacency (neighborhood) structure of the 12 French administrative regions. The entries L_{ij} of the Laplacian $\boldsymbol{L}$ are defined as:

$$L_{ij} = \begin{cases} \deg(\text{Region}_i) & \text{if } i = j, \\ -1 & \text{if } i \neq j \text{ and region } i \text{ is neighbour of region } j \\ 0 & \text{elsewhere,} \end{cases} \tag{4.2}$$

where $\deg(\text{Region}_i)$ refers to the number of neighbors of region i.

The penalized log-likelihood then results in:

$$l_p(\boldsymbol{\theta}; \psi_1, \psi_2) = l_{\text{nb}}(\boldsymbol{\theta}) - \frac{\psi_1}{2} \sum_{q=1}^{Q_1} \boldsymbol{\eta}_{1,q}' \boldsymbol{L} \boldsymbol{\eta}_{1,q} - \frac{\psi_2}{2} \sum_{q=1}^{Q_2} \boldsymbol{\eta}_{2,q}' \boldsymbol{L} \boldsymbol{\eta}_{2,q}, \tag{4.3}$$

where $l_{\text{nb}}(\boldsymbol{\theta})$ is the non-penalized negative binomial log-likelihood defined in Eq. (4.1), $\boldsymbol{L}$ is the Laplacian matrix in Eq. (4.2), Q_i represents the number of columns in the cross-basis matrices $\boldsymbol{Z}_{i,r}$ for $i = 1, 2$, $\boldsymbol{\eta}_{1,q} = (\eta_{1,q,r})_r \in \mathbb{R}^R$, with $R = |\mathcal{R}|$ the number of regions in the set $\mathcal{R}$, and ψ_i, for $i = 1, 2$, are the smoothing parameters that reflect the degree of spatial smoothness imposed on the DLNM parameters.

Estimation. To maximize the penalized log-likelihood in Eq. (4.3), we use a Newton-type optimization strategy. Specifically, we iteratively update subsets of the full parameter vector $\boldsymbol{\theta}$ while keeping the remaining parameters fixed. This blockwise approach is computationally efficient and is well suited for models with high-dimensional parameter spaces. A similar approach has also been used in Pitacco (2009) to estimate the parameters in a Lee-Carter model under a Poisson distributional assumption.

In our modeling framework defined in Eq. 3.2, we consider 11 parameter subsets:

$$(\alpha_{x,r})_{x\in\mathcal{X},r\in\mathcal{R}}, \quad (\beta_x)_{x\in\mathcal{X}}, \quad (\kappa_{t,r})_{t\in\mathcal{T},r\in\mathcal{R}}, \quad (\gamma_x)_{x\in\mathcal{X}}, \quad (\lambda_{w,r})_{w\in\mathcal{W},r\in\mathcal{R}},$$
$$(\delta_x)_{x\in\mathcal{X}}, \quad (\boldsymbol{\eta}_{1,r})_{r\in\mathcal{R}}, \quad (\epsilon_x)_{x\in\mathcal{X}}, \quad (\boldsymbol{\eta}_{2,r})_{r\in\mathcal{R}}, \quad (\phi_x)_{x\in\mathcal{X}}, \quad (\phi_r)_{r\in\mathcal{R}},$$

which constitute the complete vector of parameters $\boldsymbol{\theta}$. Let $\boldsymbol{\zeta} \subset \boldsymbol{\theta}$ be any set of parameters to be updated (e.g., the set of $\alpha_{x,r}$'s for all ages x and regions r). In the iteration step $\nu + 1$, the update for $\boldsymbol{\zeta}$ is given by:

$$\hat{\boldsymbol{\zeta}}^{(\nu+1)} = \hat{\boldsymbol{\zeta}}^{(\nu)} - \left[\boldsymbol{H}_{\boldsymbol{\zeta}} \left(\hat{\boldsymbol{\zeta}}^{(\nu)} \right) \right]^{-1} \boldsymbol{J}_{\boldsymbol{\zeta}} \left(\hat{\boldsymbol{\zeta}}^{(\nu)} \right),$$

where $\boldsymbol{J}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}}^{(\nu)})$ and $\boldsymbol{H}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}}^{(\nu)})$ denote, respectively, the gradient vector and the Hessian matrix of the penalized negative binomial log-likelihood (see Eq. (4.3)) with respect to the parameter vector $\boldsymbol{\zeta}$, evaluated in the previous estimate $\hat{\boldsymbol{\zeta}}^{(\nu)}$. We repeat this procedure sequentially across all parameter subsets until convergence of the full log-likelihood.

In the case study in Section 5, we adopt a two-step calibration procedure. First, we estimate the parameters in the weekly Lee-Carter model, i.e., $(\alpha_{x,r}, \beta_x, \kappa_{t,r}, \gamma_x, \lambda_{w,r}, \phi_x, \phi_r)$, using the iterative procedure described above. In the second step, we estimate the parameters in the DLNM component, i.e., $(\delta_x, \boldsymbol{\eta}_{1,r}, \epsilon_x, \boldsymbol{\eta}_{2,r})$, while keeping the first-stage parameters fixed. This two-step approach has two main advantages: (1) it avoids identification issues that may arise from confounding between seasonality and environmental effects; and (2) it facilitates a clearer interpretation of the covariate-response relationships as deviations from baseline seasonal mortality patterns.

Model selection. We select the smoothing parameters ψ_1 and ψ_2 of the penalized DLNM (see Eq. (4.3)) by minimizing the Akaike Information Criterion (AIC), where the complexity of the model is measured by the effective degrees of freedom (EDF) rather than the raw number of parameters (Wood, 2017; Wood et al., 2016). This approach accounts for the shrinkage induced by the penalization.

Hereto, we define grids Ψ_1 and Ψ_2 of candidate values for ψ_1 and ψ_2, respectively. For each pair $(\psi_1, \psi_2) \in \Psi_1 \times \Psi_2$, we estimate the parameter vector

$$\hat{\boldsymbol{\theta}}_{\psi_1,\psi_2} = \arg\max_{\boldsymbol{\theta}} l_p(\boldsymbol{\theta}; \psi_1, \psi_2),$$

where $l_p(\cdot)$ is the penalized log-likelihood defined in Eq. (4.3), using the iterative Newton method described above. We then compute the effective degrees of freedom as (Wood et al., 2016):

$$\mathrm{edf}(\psi_1, \psi_2) = \mathrm{trace}\left[\widetilde{\boldsymbol{H}}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_{\psi_1,\psi_2}\right) \cdot \boldsymbol{H}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_{\psi_1,\psi_2}\right)^{-1}\right], \tag{4.4}$$

where $\widetilde{\boldsymbol{H}}_{\boldsymbol{\theta}}(\cdot)$ denotes the Hessian of the unpenalized negative binomial log-likelihood (see Eq. (4.1)), and $\boldsymbol{H}_{\boldsymbol{\theta}}(\cdot)$ denotes the Hessian of the penalized log-likelihood (Eq. (4.3)). Note that when $\psi_1 = \psi_2 = 0$ (no penalization), the effective degrees of freedom are equal to the total number of parameters in the model minus the number of identifiability constraints.

Subsequently, we calculate the AIC for each pair of smoothing parameters:

$$\mathrm{AIC}(\psi_1, \psi_2) = -2\, l_p\left(\hat{\boldsymbol{\theta}}_{\psi_1,\psi_2}\right) + 2\,\mathrm{edf}(\psi_1, \psi_2). \tag{4.5}$$

We then select the optimal smoothing parameters $(\hat{\psi}_1, \hat{\psi}_2)$ by minimizing the AIC on the candidate grids.

Parameter uncertainty. Following Brouhns et al. (2002), we derive confidence intervals for the estimated parameter vector by computing the covariance matrix as the inverse of the observed Fisher information matrix. The observed Fisher information matrix is defined as the negative of the Hessian of the penalized log-likelihood function evaluated at $\hat{\boldsymbol{\theta}}_{\hat{\psi}_1,\hat{\psi}_2}$:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_{\hat{\psi}_1,\hat{\psi}_2}) = \left[-\boldsymbol{H}_{\boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_{\hat{\psi}_1,\hat{\psi}_2}\right)\right]^{-1}, \tag{4.6}$$

where $\boldsymbol{H}_{\boldsymbol{\theta}}(\cdot)$ denotes the Hessian matrix of the penalized log-likelihood in Eq. (4.3) with respect to $\boldsymbol{\theta}$. We compute these second-order derivatives analytically to ensure computational efficiency.

Other approaches to quantify parameter uncertainty in the literature include the semiparametric bootstrap proposed by Brouhns et al. (2005) and the residual bootstrap approach introduced by Koissi et al. (2006).

4.2 Forecasting

To forecast weekly mortality, we adopt a two-step procedure: (1) forecast the latent mortality index $\kappa_{t,r}$ and the exogenous predictors (temperature and ILI), and (2) calculate the predicted weekly death rates using the estimated model from Section 3.1 and the distributed lag structures from Section 3.3.

Step 1: Copula-based time series forecasting with covariates. The latent mortality index $\kappa_{t,r}$, the weekly average temperature $\text{Tavg}_{t,w,r}$, and the influenza-like illness rates $\text{ILI}_{t,w,r}$ evolve over time and jointly shape the mortality dynamics. Importantly, temperature is an exogenous driver: it influences the transmission and intensity of influenza activity, but not vice versa (Lowen et al., 2007; Lowen & Steel, 2014). Both temperature and influenza activity can, in turn, affect the evolution of the $\kappa_{t,r}$-process (Seklecka et al., 2017). This establishes a natural causal ordering among the three time-dependent processes. In addition, strong spatial dependencies are present: for instance, high temperatures in one region are likely to coincide with high temperatures in neighboring regions. Our forecasting framework should therefore account for both the causal hierarchy across processes and the regional dependencies within each process.

Therefore, we jointly model the regional dependencies within each process using a copula-based framework, while capturing the temporal evolution in each region through a Seasonal Autoregressive Integrated Moving Average process with eXogenous drivers (SARIMAX). A SARIMAX$(p,d,q)(P,D,Q)_s$ model for a time series y_t is specified as (Hyndman & Khandakar, 2008):

$$\begin{aligned} y_t &= \mathbf{x}_t^\top \boldsymbol{\beta} + u_t, \\ \Phi(B^s)\,\phi(B)\,(1-B^s)^D(1-B)^d\,u_t &= \delta + \Theta(B^s)\,\theta(B)\,\varepsilon_t, \end{aligned} \tag{4.7}$$

where

- $\mathbf{x}_t$ is the vector of exogenous regressors and $\boldsymbol{\beta}$ the vector of regression coefficients,
- p,d,q and P,D,Q are the non-seasonal and seasonal AR, differencing, and MA orders, with s the seasonal period,
- $\phi(B),\theta(B)$ and $\Phi(B^s),\Theta(B^s)$ are the corresponding non-seasonal and seasonal AR and MA polynomials in the backshift operators B and B^s, respectively,
- $(1-B)^d$ and $(1-B^s)^D$ denote the ordinary and seasonal differencing operators,
- δ is the drift parameter,
- ε_t is white noise with mean zero and variance σ^2.

We proceed by applying the following modeling steps sequentially:

1. **Temperature:** For each region $r \in R$, we first fit a SARIMA$(p_1, d_1, q_1)(P_1, D_1, Q_1)_{52}$ model (without external regressors) to the weekly average temperature $\text{Tavg}_{t,w,r}$. The seasonal and non-seasonal orders are kept identical across all regions. Second, to capture the regional dependencies between the SARIMA processes, we fit a $|R|$-dimensional t-copula on the standardized innovations $\hat{\varepsilon}_{t,w} = (\hat{\varepsilon}_{t,w,r})_{r\in R}$ from the individual SARIMA processes. The t-copula is chosen to better model dependencies in the tails of the distribution compared to a Gaussian copula.

2. **Influenza:** We fit a SARIMAX$(p_2, d_2, q_2)(P_2, D_2, Q_2)_{52}$ process to the logarithmic transformation of weekly incidence rates of ILI, $\log(\text{ILI}_{t,w,r}+0.01)$, using $\text{Tavg}_{t,w,r}$ as an external regressor. The logarithmic transformation improves the seasonal structure and reduces the impact of extreme peaks, while the small constant 0.01 avoids taking the logarithm of zero. As before, a $|R|$-dimensional t-copula is then fitted to the standardized innovations to account for regional dependencies between the SARIMAX processes.

3. **Mortality index:** For each region $r \in R$, we fit an ARIMAX(p_3, d_3, q_3) process on the estimated latent mortality index $\hat{\kappa}_{t,r}$, i.e., a SARIMAX process with $P = D = Q = 0$. Since this index is defined on an annual basis, no seasonal component is included. As external regressors, we use annual averages of weekly temperature $\text{Tavg}_{t,w,r}$ and ILI incidence rates $\text{ILI}_{t,w,r}$. Finally, a Gaussian copula is fitted to the standardized residuals to account for regional dependencies.

Step 2: Generating mortality forecasts. Once the model is fitted to historical data, we produce forecasts for $h = 1, \ldots, H$ weeks ahead, where H denotes the desired forecast horizon.

- The forecasted $\widehat{\kappa}_{t,w+h,r}$ is plugged into the model in Eq. (3.2).
- The forecasted $\widehat{\text{Tavg}}_{t,w+h,r}$ and $\widehat{\text{ILI}}_{t,w+h,r}$ are transformed into their respective cross-basis matrices $\boldsymbol{Z}_{1,r}^{\text{new}}$ and $\boldsymbol{Z}_{2,r}^{\text{new}}$ using the same DLNM basis functions and lag structures used in the estimation.
- Since DLNMs require lagged values, the forecasts are initialized with observed covariate values and updated recursively using previous predictions for lags beyond the observed horizon.

The final forecasted weekly force of mortality is computed as follows:

$$\log \widehat{\mu}_{x,t,w+h,r} = \hat{\alpha}_{x,r} + \widehat{\beta}_x \widehat{\kappa}_{t,w+h,r} + \widehat{\gamma}_x \widehat{\lambda}_{w+h,r} + \widehat{\delta}_x \left(\boldsymbol{Z}_{1,r}^{\text{new}} \widehat{\boldsymbol{\eta}}_{1,r}\right) + \widehat{\epsilon}_x \left(\boldsymbol{Z}_{2,r}^{\text{new}} \widehat{\boldsymbol{\eta}}_{2,r}\right). \tag{4.8}$$

The forecasted number of deaths is then obtained by:

$$\widehat{D}_{x,t,w+h,r} \sim \text{NB}(E_{x,t,w+h,r} \cdot \widehat{\mu}_{x,t,w+h,r}, \phi_{x,r}). \tag{4.9}$$

Uncertainty quantification. To account for uncertainty, we propose a simulation-based approach:

- Generate multiple paths for the mortality index, temperature, and ILI forecasts using the estimated (S)ARIMA(X) models and innovations sampled from the corresponding estimated copulas.

- For each simulated path, propagate through the mortality model to yield a corresponding forecast distribution of deaths using Eq. (4.8) and (4.9).
- Construct prediction intervals by computing empirical quantiles of the simulated forecast distribution.

5 Case study

5.1 Model calibration and validation

Model specifications and selection. We model the effects of temperature and influenza-like illness (ILI) on mortality using distributed lag non-linear models (see Section 3.3). For temperature, the covariate–response relationship is modeled with cubic B-spline basis functions, with two internal knots placed at the 10th and 90th percentiles of the region-specific temperature distribution over the historical period 1990–2019. These choices are based on common practices in DLNM studies that examine the temperature–mortality relationship (Gasparrini et al., 2015; Martínez-Solanas et al., 2021). For influenza, we work with anomalies, defined as deviations of weekly incidence rates of ILI from the region- and week-specific 90th percentiles, and assume a linear covariate–response relationship (L. Li et al., 2023). The lag period is set at 4 weeks for temperature and 6 weeks for influenza, reflecting the typically longer persistence of influenza effects (Chaves et al., 2023). Lag–response relationships are modeled with natural cubic splines, including an intercept, with two internal knots placed between lags of 0–1 and 1–2 weeks.

Following the model selection in the calibration setup in Section 4.1, we set up a tuning grid for the smoothing parameters ψ_1 and ψ_2, which control the degree of spatial smoothness imposed on the DLNM parameters. Since the ψ_i-values vary by many orders of magnitude, we define the parameter grid formed by all combinations of two sequences whose values are equally spaced on a logarithmic scale between 1 and 10^{12}:

$$\psi_i \in \{10^0, 10^{0.5}, 10^1, ..., 10^{12}\},$$

for $i = 1, 2$. By selecting the pair (ψ_1, ψ_2) that leads to the lowest value for the edf-based AIC (see Eq. (4.5)), we obtain the estimated values $\hat{\psi}_1 = 10^6$ and $\hat{\psi}_2 = 10^9$.

Parameter estimates. Using the optimally selected smoothing parameters $\hat{\psi}_i$ for $i = 1, 2$, we present the parameter estimates in Figure 6. We add confidence bounds based on the computation of the Hessian of the penalized negative binomial log-likelihood, as described in Section 4.1.

The trends and patterns in panels A, B, and C closely resemble a traditional Lee-Carter model, as they depend only on region, age, and year. From panel C, we see that the mortality decline is strongest in region 53 (Bretagne) and lowest in region 24 (Centre-Val de Loire). In panel B, $\hat{\beta}_x$ shows that the mortality decline first increases with age, reaches a maximum for the 75-79 and 80-84 age groups, and then declines sharply. This means that the decline in mortality rates between 1990 and 2019 is strongest for the middle-aged age groups of 75-79 and 80-84, but the decline slows down significantly for older age groups. This is in line with traditional applications of the Lee-Carter model on annual mortality data in France, see, e.g., (Delwarde et al., 2007).

Moreover, for the weekly effect parameter $\hat{\lambda}_{w,r}$ in panel E, we see a clear U- or V-shaped pattern, which reflects the seasonality. This means that mortality rates are highest during the winter weeks and lowest during the summer weeks. To a lesser extent, we can also observe some regional

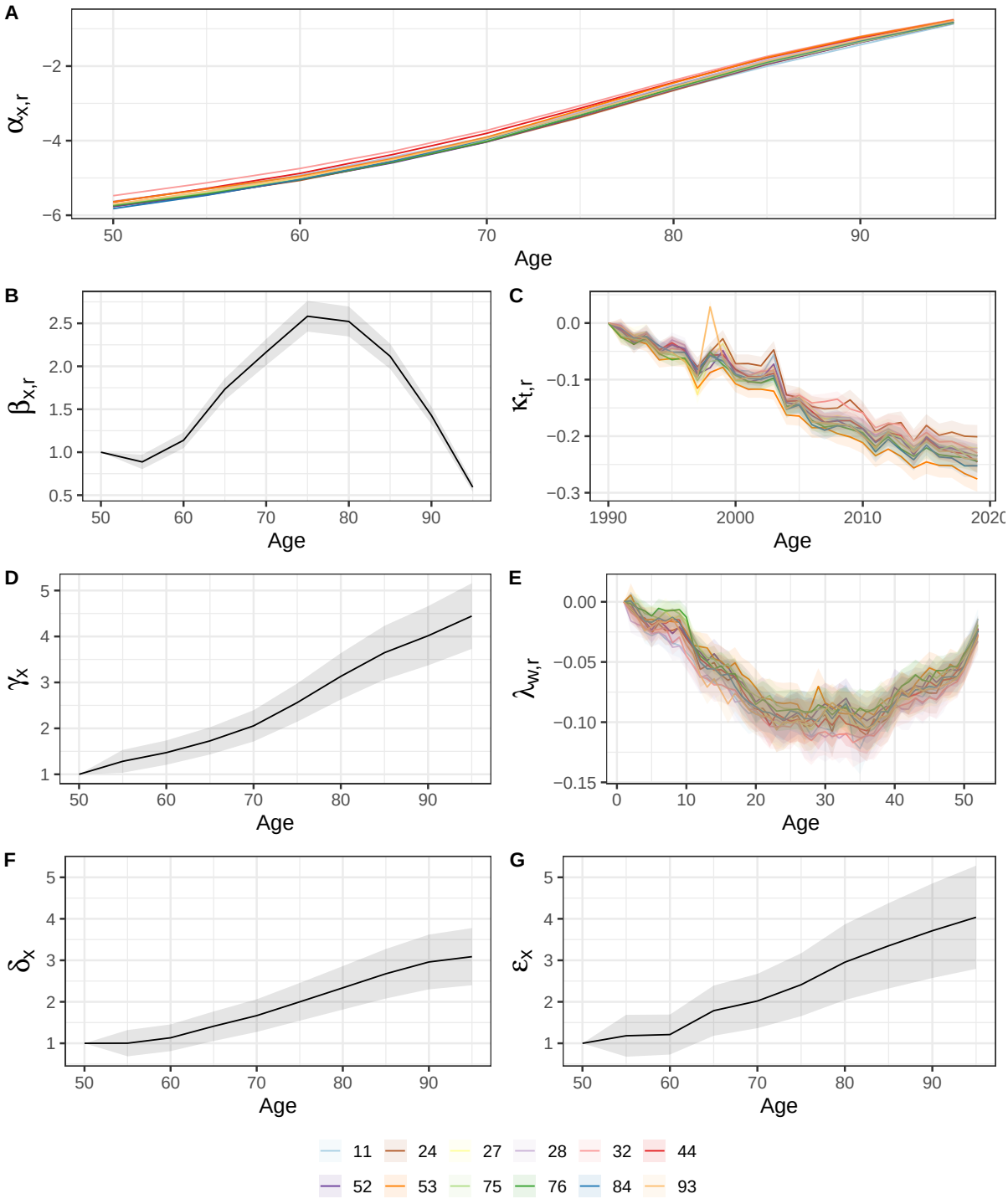


Figure 6: Estimated parameters from the model: $\hat{\alpha}_{x,r}$ (A), $\hat{\beta}_x$ (B), $\hat{\kappa}_{t,r}$ (C), $\hat{\gamma}_x$ (D), $\hat{\lambda}_{w,r}$ (E), $\hat{\delta}_x$ (F), and $\hat{\epsilon}_x$ (G). Female data, French administrative regions, 5-year age groups $50-54$, $55-59$, ..., 95+, and years $1990-2019$. Confidence intervals are based on the Hessian of the penalized negative binomial log-likelihood. The age-specific parameters, with the exception of $\hat{\alpha}_{x,r}$, do not vary by region r and are shown in black.

differences. For example, the seasonal variation, measured as the variance in the estimated $\hat{\lambda}_{w,r}$ for every region r, is the lowest in the administrative regions in the east of France, i.e., regions 52 (Pays de la Loire) and 53 (Bretagne). Regarding the estimated age effect $\hat{\gamma}_x$ in panel D,

we see an increasing pattern. This means that seasonal variation is stronger for the oldest age groups. Moreover, the uncertainty surrounding the estimated age effect also increases with age.

Lastly, panels F and G present the age effects associated with the distributed lag non-linear components of temperature and influenza, respectively. Both panels display an upward trend, indicating that the impacts of temperature and influenza increase steadily with age. On the other hand, we also observe an increasing uncertainty with age. We evaluate in more depth the association of temperature and influenza with mortality deviations from the seasonal baseline model by means of the relative risk in Section 5.2.

In-sample fit of death rates. Using the estimated parameters $\hat{\alpha}_{x,r}$, $\hat{\beta}_x$, $\hat{\kappa}_{t,r}$, $\hat{\gamma}_x$, $\hat{\lambda}_{w,r}$, $\hat{\delta}_x$, $\hat{\boldsymbol{\eta}}_{1,r}$, $\hat{\epsilon}_x$, and $\hat{\boldsymbol{\eta}}_{2,r}$, for every $x \in \mathcal{X}$, $t \in \mathcal{T}$, $w \in \mathcal{W}$, and $r \in \mathcal{R}$, as visualized in Figure 6, we construct an in-sample estimate of the force of mortality $\hat{\mu}_{x,t,w,r}$ using Eq. (3.2). This estimated force of mortality $\hat{\mu}_{x,t,w,r}$ serves as an estimate for the observed weekly death rates $m_{x,t,w,r}$, as discussed in Section 2.1. This gives us a visual idea of how well our in-sample predictions match the observations.

To disentangle the contribution of the DLNM component from the baseline structure in the model of Eq. (3.2), we calculate the estimated force of mortality as the multiplication of the baseline force of mortality and the DLNM component:

$$\begin{aligned} \hat{\mu}^{(\text{base})}_{x,t,w,r} &= \exp\left(\hat{\alpha}_{x,r} + \hat{\beta}_x\, \hat{\kappa}_{t,r} + \hat{\gamma}_x\, \hat{\lambda}_{w,r}\right) \\ \hat{\mu}_{x,t,w,r} &= \hat{\mu}^{(\text{base})}_{x,t,w,r} \cdot \exp\left(\hat{\delta}_x\, \hat{f}^{(1)}_r(\text{Tavg}_{t,w,r}) + \hat{\epsilon}_x\, \hat{f}^{(2)}_r(\text{ILI}_{t,w,r})\right), \end{aligned} \tag{5.1}$$

where $\hat{f}^{(1)}_r(\text{Tavg}_{t,w,r})$ and $\hat{f}^{(2)}_r(\text{ILI}_{t,w,r})$ are based on the estimated parameter vectors $\hat{\boldsymbol{\eta}}_{1,r}$ and $\hat{\boldsymbol{\eta}}_{2,r}$, respectively (see Eq. (3.5)).

Figure 7 compares the observed and estimated weekly death rates for women between 1990 and 2019 for the age groups 70-74 and 90-94, and for three administrative regions in France: Hauts-de-France, Bretagne, and Auvergne-Rhône-Alpes. In addition to the empirical rates, we plot the fit of the baseline model and the full model that incorporates the DLNM component, capturing the effects of extreme temperatures and influenza activity. The results show that seasonal mortality patterns are well reproduced, with a stronger seasonal signal in the 90–94 age group compared to the 70–74 age group. Furthermore, the integration of the DLNM component is particularly valuable for older age groups, who are more vulnerable to temperature extremes and influenza outbreaks. The Supplementary Material contains the in-sample fits for the remaining regions and age groups.

Pearson residuals. To evaluate the goodness-of-fit of the model in a quantitative way, we calculate the squared Pearson residuals for each region $r \in \mathcal{R}$:

$$\begin{aligned} \rho^2_{x,t,w,r} &= \frac{(d_{x,t,w,r} - \mathbb{E}[D_{x,t,w,r}])^2}{\text{Var}[D_{x,t,w,r}]}\Big|_{\mu_{x,t,w,r}=\hat{\mu}_{x,t,w,r}} \\ &= \frac{(d_{x,t,w,r} - E_{x,t,w,r}\, \hat{\mu}_{x,t,w,r})^2}{E_{x,t,w,r}\, \hat{\mu}_{x,t,w,r} + (E_{x,t,w,r}\, \hat{\mu}_{x,t,w,r})^2 / \hat{\phi}_{x,r}}, \end{aligned}$$

where $d_{x,t,w,r}$ is the observed number of deaths, $\hat{\mu}_{x,t,w,r}$ the estimated force of mortality, $\hat{\mu}_{x,t,w,r}\, E_{x,t,w,r}$ the estimated expected number of deaths, and $\hat{\phi}_{x,r} = \exp(\hat{\phi}_x + \hat{\phi}_r)$ the estimated dispersion parameter.

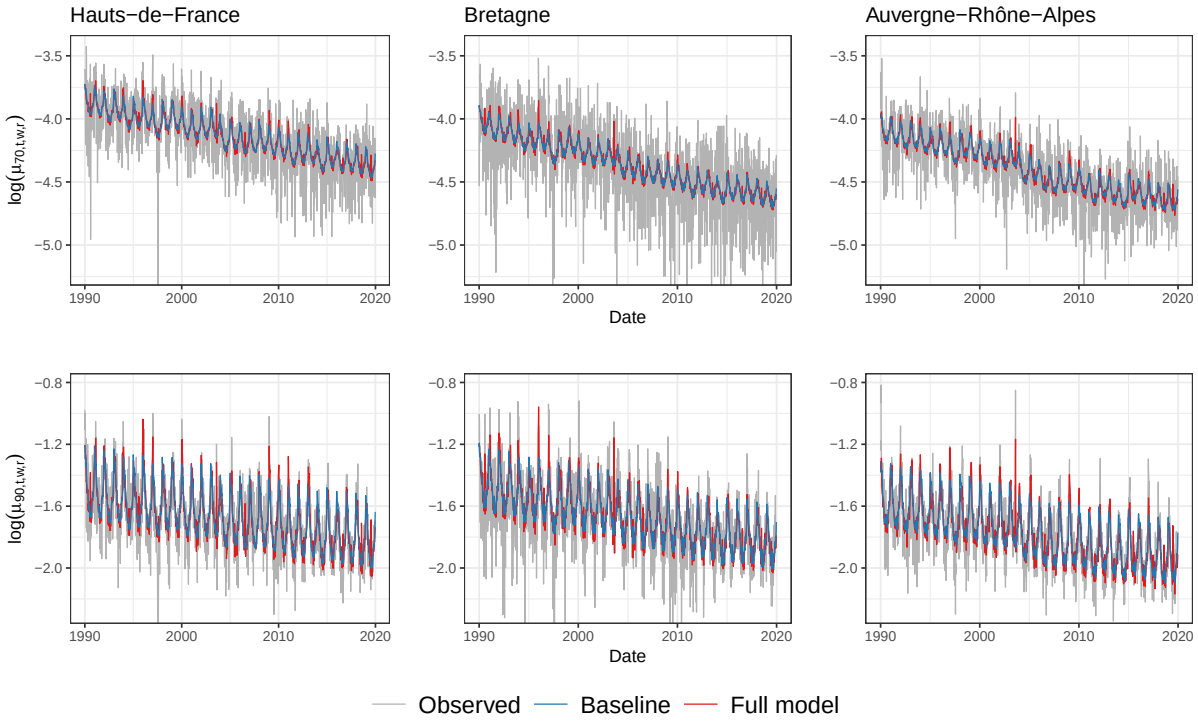


Figure 7: Estimated female weekly death rates from 1990-2019 for the age groups 70-74 (top) and 90-94 (bottom) in Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right). The gray line represents the logarithm of the observed death rates, while the blue and red line reflect, respectively, the logarithm of the death rates as estimated by the baseline model and the full model including the DLNM component.

Figure 8 visualizes the squared Pearson residuals for Hauts-de-France, Bretagne, and Auvergne-Rhône-Alpes. Under some mild regularity conditions and assuming that the model specified in Eq.(3.1) and (3.2) is correctly specified (Wong et al., 2018), each $\rho^2_{x,t,w,r}$ is asymptotically distributed as a chi-square random variable with one degree of freedom (χ^2_1). Consequently, for each region, around 5% of the cells in Figure 8 are expected to exceed the 95th percentile of the χ^2_1-distribution, i.e., $\rho^2_{x,t,w,r} > 3.841$. For the three regions shown, the observed proportions are 4.24%, 4.71%, and 3.99%, respectively. We find comparable percentages for the remaining French regions, all between 3.26% and 4.71%. Moreover, no particular age groups or time points stand out as systematic sources of poor fit.

Lastly, we assess the overall model's goodness-of-fit by aggregating the squared Pearson residuals across all age groups, years, weeks, and regions (Wong et al., 2018):

$$\rho^2 = \sum_{x,t,w,r} \rho^2_{x,t,w,r},$$

which leads to a total value of 175 167.6. Under the same set of assumptions as for the cell-specific tests, we now compare this statistic with the 95th percentile of a Chi-square distribution with degrees of freedom equal to $|\mathcal{X}| \times |\mathcal{T}| \times |\mathcal{W}| \times |\mathcal{R}| - \text{edf}(\hat{\psi}_1, \hat{\psi}_2)$. Here, $|\mathcal{X}| \times |\mathcal{T}| \times |\mathcal{W}| \times |\mathcal{R}|$ is the total number of observations and we use the effective degrees of the model to account for the regularization of the parameters in the DLNM component of the model, see Equation (4.4). The corresponding 95th percentile is 186 829.6. As such, we consider the negative binomial assumption as adequate for our analysis, as it explains the majority of the extra-Poisson variability within the data. In the Supplementary Material, we repeat the

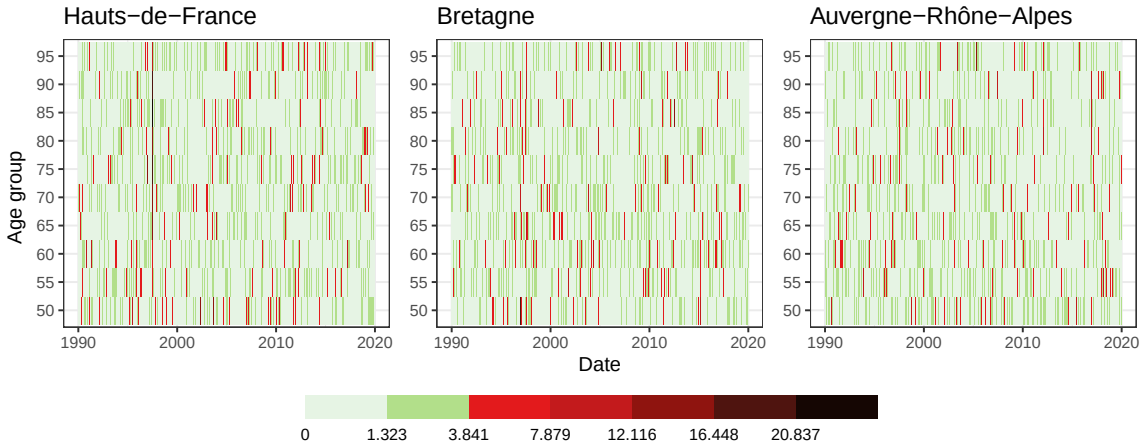


Figure 8: Heat map of the squared Pearson residuals $\rho^2_{x,t,w,r}$ across age and time for Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right). Green cells indicate areas with a good fit, while red and black cells correspond to areas with a poor fit ($\rho^2_{x,t,w,r} > \chi^2_1$).

analysis under the assumption that the weekly death counts follow a Poisson distribution, i.e., without accounting for overdispersion. In that case, more than 5% of the cells in the heat map of the squared Pearson residuals exceed the critical, and the aggregated chi-square statistic also exceeds its reference value. These results confirm that the negative binomial specification is more appropriate for our application.

Uncertainty bounds. To complement the in-sample estimates, we also quantify uncertainty around the estimated weekly death rates. We incorporate two types of uncertainty. First, we account for parameter uncertainty by sampling from the asymptotic distribution, approximated by a multivariate normal distribution with mean the estimated parameter vector and covariance matrix the inverse Hessian of the negative penalized log-likelihood (see Eq. (4.6)). For each draw of the parameter set, we reconstruct the force of mortality according to Eq. (3.2). Second, we account for aleatoric or observation uncertainty by simulating deaths from the negative binomial distribution with mean equal to the estimated force of mortality times the population exposure and variance determined by the estimated dispersion. We repeat this process 10 000 times and combine the two sources of uncertainty to construct 95% uncertainty bounds for the weekly death rates. These bounds provide a measure of the goodness-of-fit of our model. Figure 9 shows the results for the age groups 50–54, 70–74, and 90–94 in Hauts-de-France, Bretagne, and Auvergne-Rhône-Alpes. The Supplementary Material contains the results for the other regions. The observed death rates fall mainly within the constructed 95% uncertainty bounds, even during extreme events such as severe influenza outbreaks and the 2003 European heat wave.

5.2 Relative risk

Using the estimated parameter vectors $\hat{\boldsymbol{\eta}}_{1,r}$ and $\hat{\boldsymbol{\eta}}_{2,r}$ from Eq. (3.2), we visualize the effect of temperature and influenza on mortality, in excess of the seasonal baseline, through the relative risk (RR), as discussed in Section 3.4.

Overall relative risk. We begin by visualizing the overall relative risk, as defined in Eq. (3.8), for three administrative regions in France: Hauts-de-France, Bretagne, and Provence-Alpes-Côte

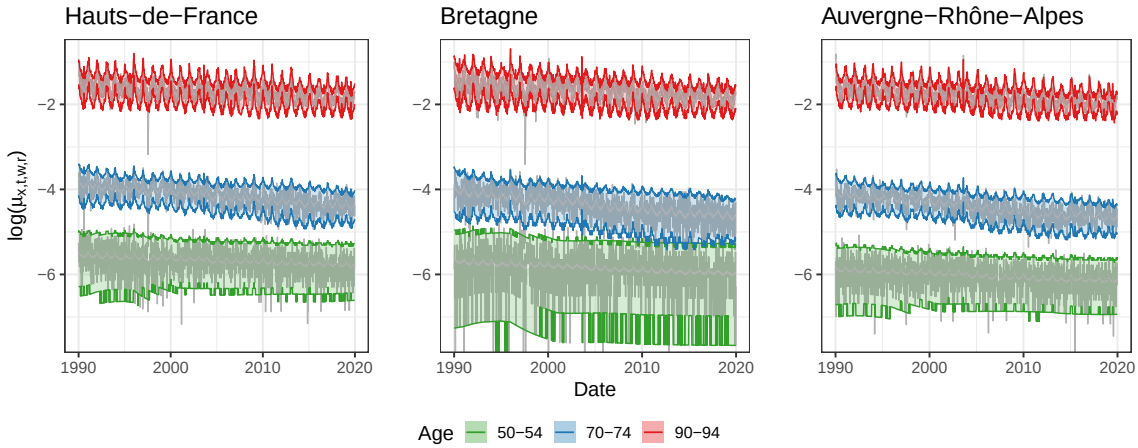


Figure 9: 95% uncertainty bounds around the logarithm of the estimated female weekly death rates from 1990-2019 for the age groups 50-54, 70-74 and 90-94 in Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right).

d'Azur. These regions represent a region in the north, west, and south of France, respectively. We provide the plots for the remaining regions in the Supplementary Material. For each region, we compute the overall relative risk for three age groups: 50–54, 70–74, and 90–94 years. We select a reference temperature of $12.5°C$, which roughly corresponds to the mean overall temperature, and a reference ILI exceedance value of 0, for each region. We compute point-wise 95% confidence intervals around the relative risk curves using Eq. (3.9). The results are shown in Figure 10.

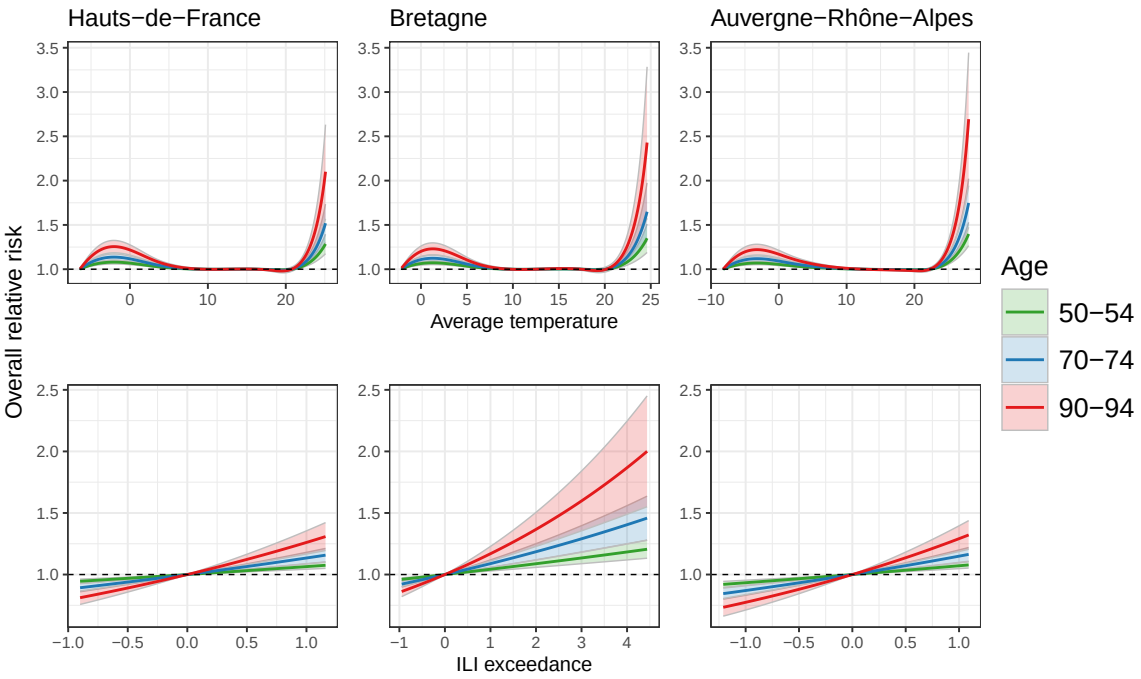


Figure 10: Estimated overall relative risk for temperature (top row) and influenza (bottom row) in Hauts-de-France (left), Bretagne (middle), and Provence-Alpes-Côte d'Azur (right) for the age groups 50–54, 70–74, and 90–94. Shaded areas represent 95% point-wise confidence intervals.

Among the three regions, the effect of temperature on mortality increases rapidly at high temperatures and is most pronounced in Auvergne-Rhône-Alpes, followed by Bretagne and Hauts-de-France. For influenza-like illness (ILI), we clearly observe the highest relative risk in Bretagne. Furthermore, we find that the relative risk of both temperature and ILI is substantially higher for the oldest age group (90–94) compared to the youngest (50–54), although the uncertainty also increases with age. Notably, the cold effect is minor. This is likely because most excess mortality in winter is captured by ILI rather than by low temperatures.

Table 2 reports the relative risks (RR) by age group and region at the 99.5th percentile of the region-specific weekly average temperatures during the period 1990–2019. Although the highest 99.5th temperature percentile is observed in the southern region Provence-Alpes-Côte d'Azur, we observe the highest relative risks in Ile-de-France. In addition, we also observe that the relative risk increases consistently with age in all regions.

Age \ Region	**11** (25.4)	**24** (24.6)	**27** (24.4)	**28** (21.4)	**32** (22.6)	**44** (24.5)	**52** (23.7)	**53** (21.1)	**75** (24.6)	**76** (25.9)	**84** (24.5)	**93** (26.1)
50-54	1.11	1.05	1.06	1.03	1.05	1.08	1.04	1.02	1.03	1.05	1.05	1.04
55-59	1.11	1.05	1.07	1.03	1.05	1.08	1.04	1.02	1.04	1.05	1.05	1.04
60-64	1.13	1.05	1.07	1.03	1.06	1.09	1.04	1.02	1.04	1.06	1.06	1.04
65-69	1.16	1.07	1.09	1.04	1.07	1.12	1.05	1.03	1.05	1.07	1.08	1.05
70-74	1.20	1.08	1.11	1.05	1.09	1.14	1.06	1.03	1.06	1.09	1.09	1.06
75-79	1.24	1.10	1.13	1.06	1.11	1.17	1.08	1.04	1.07	1.11	1.11	1.08
80-84	1.29	1.12	1.16	1.07	1.13	1.20	1.09	1.04	1.08	1.13	1.13	1.09
85-89	1.33	1.13	1.18	1.08	1.14	1.23	1.10	1.05	1.10	1.14	1.15	1.10
90-94	1.37	1.15	1.20	1.09	1.16	1.26	1.12	1.06	1.11	1.16	1.16	1.11
95+	1.39	1.16	1.21	1.09	1.17	1.27	1.12	1.06	1.11	1.17	1.17	1.12

Table 2: The overall relative risk per age group at the 99.5% percentile of the historical temperature distribution. The value of the percentile is given below the region's INSEE code.

Relative risk by lag. The overall RR summarizes the cumulative effect across all lags, as discussed in Section 3.4. To gain further insight, we also examine the effect of a fixed temperature or ILI exceedance value at each individual lag, i.e., at lags 0–4 weeks for temperature and lags 0–6 weeks for ILI. This allows us to evaluate whether the estimated effects occur immediately or persist over time.

Figure 11 presents the estimated lag-specific RRs at the 99.5th percentile of the region-specific weekly average temperatures (top row) and ILI exceedance rates (bottom row), for the same set of regions and age groups as in Figure 10. The effect of heat appears to be largely immediate, with elevated RRs at lag 0, followed by a harvesting effect reflected in negative RRs in subsequent lags, and a modest rebound at the final lag. In contrast, the effect of ILI shows no evidence of harvesting: it peaks at lag 0 and gradually diminishes across subsequent lags. We refer to the Supplementary Material for visualisations of the lag-specific RRs of the other administrative regions in France.

Relative risk surface. Lastly, to provide a comprehensive overview of how lagged temperature and ILI exceedances are associated with mortality, we present the relative risk (RR) surface plot. This plot depicts the estimated associations across the entire grid of predictor and lag values.

Figure 12 displays the RR surface plots for Hauts-de-France, Bretagne, and Auvergne-Rhône-Alpes in the age group 90–94. Red areas correspond to regions where RR $>$ 1 (increased

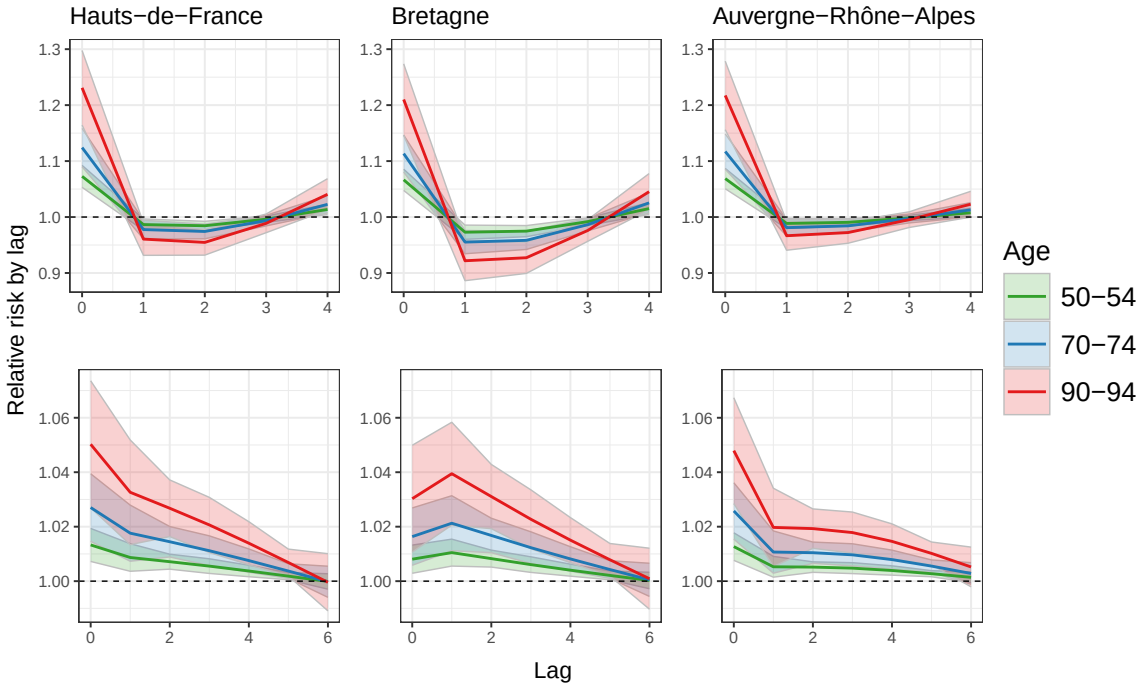


Figure 11: Estimated relative risk per lag at the 99.5th percentile of the region-specific weekly average temperatures (top row) and ILI exceedance rates (bottom row) in Hauts-de-France (left), Bretagne (middle), and Auverge-Rhône-Alpes (right) for the age groups 50–54, 70–74, and 90–94. Shaded areas represent 95% point-wise confidence intervals.

mortality risk), and green areas to regions where RR < 1 (decreased mortality risk). The surfaces interpolate smoothly between weekly lags. In the top row, which shows the RR surfaces for temperature, we observe pronounced red areas at the bottom right of each panel, indicating an immediate and strong increase in mortality risk during periods of extreme heat. Directly above these red zones, green zones emerge, reflecting the harvesting effect: mortality falls below expected levels following a hot week. On the left-hand side of each panel, the cold effect is visible. Although cold is associated with increased mortality, its effect is estimated to be smaller than that of heat, and little or no harvesting is observed (with a weak signal around lag 4). In the bottom row, the RR surface plots for ILI exceedances reveal an effect that grows with the level of exceedance and gradually decreases across subsequent lags. In the Supplementary Material we provide the RR surface plots for age group 90-94 for all the French administrative regions under consideration.

5.3 Predictive performance

To ensure a fair evaluation of the predictive performance of our proposed mortality model (Eqs. (3.1) and (3.2)), we restrict our attention to pre-pandemic data and avoid any distortions induced by the COVID-19 pandemic. Therefore, we recalibrate the model on data from 1990–2014. Using the recalibrated model, we then forecast death counts for 2015–2019 following the procedure outlined in Section 4.2. The forecasted counts are then compared with observed deaths using multiple metrics and scoring rules, along with several benchmark models.

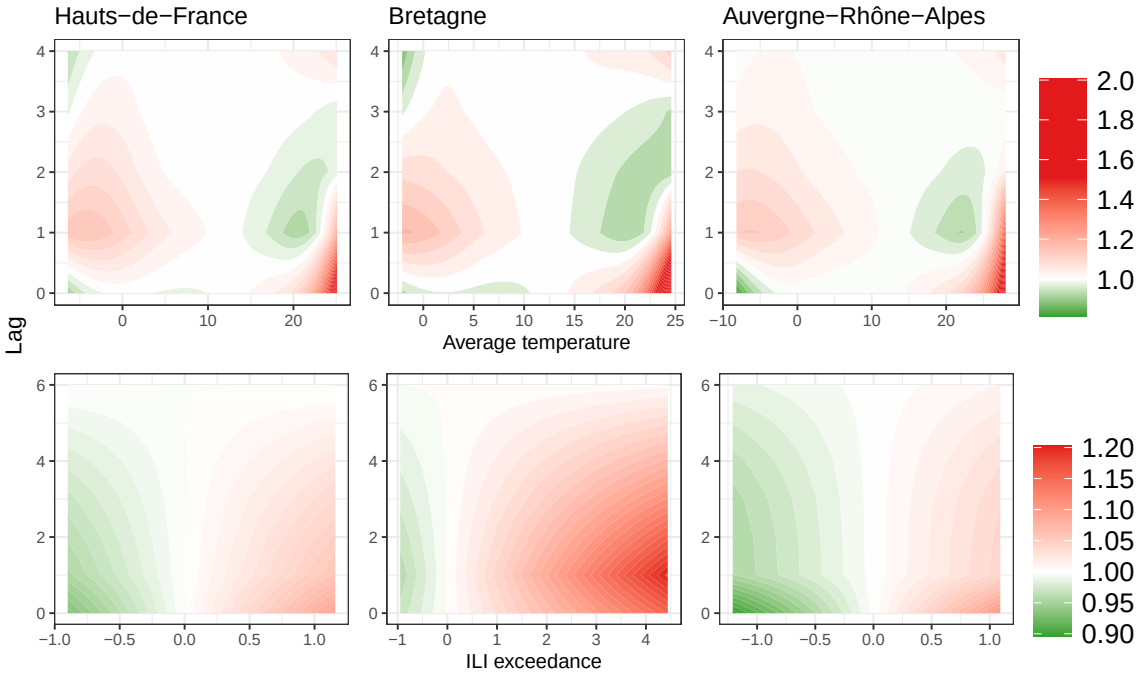


Figure 12: Estimated relative risk (RR) surfaces of the lag-specific temperature– and ILI–mortality associations in Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right) for the age group 90–94. Red areas indicate an increased RR, while green areas indicate a decreased RR.

Proposed model vs. benchmarks. We compare the mortality forecasts with those from several benchmark models, each assuming a negative binomial distribution for the death counts:

$$D_{x,t,w,r} \sim \text{NB}(E_{x,t,w,r} \cdot \mu_{x,t,w,r},\, \phi_{x,r})\,.$$

We consider the following five model structures for $\mu_{x,t,w,r}$:

- **Model 1**: $\log \mu_{x,t,w,r} = \alpha_{x,r} + \beta_x \kappa_{t,r}$ (annual, regional).
- **Model 2**: $\log \mu_{x,t,w,r} = \alpha_x + \beta_x \kappa_t + \gamma_x \lambda_w$ (weekly, national).
- **Model 3**: $\log \mu_{x,t,w,r} = \alpha_{x,r} + \beta_x \kappa_{t,r} + \gamma_x \lambda_{w,r}$ (weekly, regional).
- **Model 4**: $\log \mu_{x,t,w,r} = \text{model 3} + \delta_x f_r^{(1)}(\text{Tavg}_{t,w,r}) + \epsilon_x f_r^{(2)}(\text{ILI}_{t,w,r})$ (weekly, regional). Without covariate information available for 2015–2019.
- **Model 5**: $\log \mu_{x,t,w,r} = \text{model 3} + \delta_x f_r^{(1)}(\text{Tavg}_{t,w,r}) + \epsilon_x f_r^{(2)}(\text{ILI}_{t,w,r})$ (weekly, regional). With covariate information available for 2015–2019.

For Models 1-3, we capture the dynamics of the latent mortality index $\hat{\kappa}_{t,r}$ or $\hat{\kappa}_t$ with an ARIMA(0,1,1) model, fitted separately for each region if the model is regional.[4] We account for regional dependence across the fitted processes via a Gaussian copula applied to the vector of standardized innovations for these different regions.

Models 4 and 5 correspond to our proposed model specification (Eqs. (3.1) and (3.2)), and differ only in whether covariate information is assumed to be known during the forecasting horizon.

[4]We compared AIC and BIC values across a range of candidate ARIMA specifications over all regions jointly

Model 5 is therefore a purely hypothetical scenario, whereas model 4 is more realistic since it uses scenario forecasts generated based on estimated SARIMA(X) models for temperature and ILI. More specifically, for Model 4, we follow the approach outlined in Section 4.2: for each region, we fit a SARIMA$(1,0,1)(1,1,1)_{52}$ model for temperature, a SARIMAX$(1,0,1)(1,1,1)_{52}$ for ILI with the simulated temperature as an external regressor, and an ARIMAX$(0,1,1)$ model for the annual latent mortality index with the simulated temperature and ILI as external regressors. We used AIC/BIC to select the orders of the (S)ARIMA(X) models over all regions jointly. Regional dependence is taken into account using the copula approach in Section 4.2.

Accuracy of point forecasts. We first evaluate the accuracy of the point forecasts using the root mean squared error (RMSE) and the mean absolute error (MAE). These point forecasts are generated based on the forecasted means of the time-series models used for the time-dependent factors in the specification (i.e., period indices and covariates). We calculate RMSE and MAE per age group by averaging across time points and regions. For age group x, we obtain:

$$\mathrm{RMSE}_x = \sqrt{\frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \left(d_{x,t,w,r} - E_{x,t,w,r}\, \hat{\mu}_{x,t,w,r}\right)^2},$$

$$\mathrm{MAE}_x = \frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \left|d_{x,t,w,r} - E_{x,t,w,r}\, \hat{\mu}_{x,t,w,r}\right|,$$

where n_x is the number of observations in the data set per age group, $\mathcal{R}$ the set of regions, and $\hat{\mu}_{x,t,w,r}$ is the forecasted death rate for the age group x in the region r at time point (t,w). In addition, we calculate the overall RMSE and MAE by also including the sum over the age groups in the formulae.

Table 3 lists the overall RMSE and MAE as well as the values per age group. Our proposed model that incorporates full knowledge of temperature and ILI (Model 5) consistently achieves the best performance across age groups, with the exception of the 50–54 and 65-69 age group. The model that performs second best based on these metrics is either Model 3 or Model 4 depending on the age group. This can be attributed to the fact that the SARIMA(X)-based point forecasts for temperature and influenza underestimate the true observed temperatures. However, when exogenous conditions are more accurately represented, our model consistently outperforms all benchmarks (Models 1-3). Another observation is that the differences in the error metrics are minor among younger age groups, as death rates in these groups show relatively little seasonality. Moreover, the regional approach clearly offers an advantage over the country-level model (Model 2).

Evaluation of probabilistic forecasts. To assess the accuracy of the probabilistic forecast distributions of the future death counts, we put focus on two commonly used scoring rules: the logarithmic score (LogS) and the continuous ranked probability score (CRPS):

$$\mathrm{LogS}_x = -\frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \log\left[f_{x,t,w,r}\left(d_{x,t,w,r}\right)\right],$$

$$\mathrm{CRPS}_x = \frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \int_{\mathbb{R}} \left(F_{x,t,w,r}(s) - \mathbb{1}\left\{d_{x,t,w,r} \leq s\right\}\right) ds.$$

Here, n_x is the number of observations for age group x, $F_{x,t,w,r}$ is the predictive cumulative distribution function, and $f_{x,t,w,r}$ its density. Since the predictive distributions for the death

	RMSE					MAE				
Age	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
50–54	3.06	3.18	3.05	**3.03**	3.05	2.41	2.49	2.39	**2.37**	2.39
55–59	3.66	3.80	3.63	3.63	**3.62**	2.83	2.93	2.80	2.80	**2.80**
60–64	4.49	4.66	4.40	4.45	**4.39**	3.46	3.58	3.41	3.44	**3.40**
65–69	5.66	5.84	**5.49**	5.64	5.50	4.39	4.47	4.28	4.39	**4.27**
70–74	5.92	5.94	5.66	5.83	**5.63**	4.51	4.53	4.34	4.46	**4.31**
75–79	7.24	7.14	6.61	6.91	**6.50**	5.48	5.41	5.07	5.28	**5.02**
80–84	11.62	10.62	9.46	10.13	**9.21**	8.68	8.10	7.26	7.74	**7.11**
85–89	18.65	15.52	13.71	14.95	**13.07**	13.35	11.60	10.16	11.03	**9.87**
90–94	21.93	16.79	15.40	16.69	**14.47**	15.39	12.38	11.10	11.92	**10.66**
95+	14.21	10.68	10.27	10.11	**9.73**	10.82	8.07	7.87	7.72	**7.39**
Overall	11.52	9.57	8.76	9.29	**8.40**	7.13	6.36	5.87	6.11	**5.72**

Table 3: Predictive performance across the five competing model specifications: three benchmarks (Models 1–3) and two versions of our proposed model (Models 4–5).

counts are available through simulated forecast scenarios $E_{x,t,w,r} \cdot \hat{\mu}^{(i)}_{x,t,w,r}$, for $i = 1, 2, ..., 10\,000$, we approximate F and f by their empirical counterparts constructed from these simulations. The scores are evaluated using the `R` package `scoringRules` (Jordan et al., 2019). These scoring rules have also been applied in a Bayesian mortality modeling approach by Barigou et al. (2023).

Table 4 reports the overall CRPS and LogS, together with the values by age group. Model 4 (using simulated temperature and ILI scenarios) performs best among the younger age groups, while Model 5 (using observed temperature and ILI) performs the best for the older age groups. The weakest performance is obtained by Model 1, followed by Models 2 and 3. For the younger ages, the differences between models are relatively small.

	CRPS					LogS				
Age	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
50–54	1.66	1.72	1.65	**1.65**	1.65	2.38	2.42	2.38	**2.38**	2.38
55–59	2.01	2.08	1.99	**1.99**	1.99	2.64	2.67	2.63	**2.63**	2.63
60–64	2.49	2.56	2.44	**2.42**	2.43	2.88	2.91	2.86	**2.85**	2.85
65–69	3.15	3.23	3.06	**3.04**	3.05	3.12	3.15	3.09	**3.08**	3.09
70–74	3.22	3.25	3.09	3.06	**3.06**	3.12	3.14	3.09	3.08	**3.07**
75–79	3.93	3.91	3.63	3.61	**3.56**	3.33	3.38	3.26	3.25	**3.24**
80–84	6.26	5.81	5.18	5.19	**5.03**	3.80	3.75	3.61	3.62	**3.58**
85–89	9.80	8.38	7.35	7.37	**7.01**	4.25	4.11	3.96	3.96	**3.92**
90–94	11.34	8.95	8.07	8.13	**7.66**	4.40	4.19	4.04	4.05	**4.00**
95+	7.56	5.70	5.53	5.55	**5.22**	3.95	3.67	3.65	3.66	**3.61**
Overall	5.14	4.56	4.20	4.20	**4.07**	3.39	3.34	3.26	3.26	**3.24**

Table 4: Probabilistic scores (CRPS and LogS) across the five competing model specifications: three benchmarks (Models 1–3) and two versions of our proposed model (Models 4–5).

Calibration of prediction intervals. We assess the calibration of the constructed prediction intervals using two measures: (1) the coverage ratio, which evaluates the proportion of observed death counts that fall within the $100(1-\alpha)\%$ prediction interval $[\hat{d}^{\text{low}}_{x,t,w,r}, \hat{d}^{\text{up}}_{x,t,w,r}]$, and (2) the interval score proposed by Gneiting and Raftery (2007), which penalizes both the width of the interval and any deviations of the observed values outside the interval. We compute them as

follows:

$$\text{Coverage Ratio}_x = \frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \mathbb{1}\left\{ d_{x,t,w,r} \in [\hat{d}^{\text{low}}_{x,t,w,r}, \hat{d}^{\text{up}}_{x,t,w,r}] \right\},$$

$$\text{Interval Score}_x = \frac{1}{n_x} \sum_{r \in \mathcal{R}} \sum_{t=2015}^{2019} \sum_{w=1}^{52} \Bigg[(\hat{d}^{\text{up}}_{x,t,w,r} - \hat{d}^{\text{low}}_{x,t,w,r}) + \frac{2}{\alpha}(\hat{d}^{\text{low}}_{x,t,w,r} - d_{x,t,w,r}) \mathbb{1}\left\{ d_{x,t,w,r} < \hat{d}^{\text{low}}_{x,t,w,r} \right\} + \frac{2}{\alpha}(d_{x,t,w,r} - \hat{d}^{\text{up}}_{x,t,w,r}) \mathbb{1}\left\{ d_{x,t,w,r} > \hat{d}^{\text{up}}_{x,t,w,r} \right\} \Bigg].$$

where n_x is the number of observations for the age group x. We compute 95% pointwise intervals $[\hat{d}^{\text{low}}_{x,t,w,r}, \hat{d}^{\text{up}}_{x,t,w,r}]$ based on 10 000 simulated forecast scenarios $E_{x,t,w,r} \cdot \hat{\mu}^{(i)}_{x,t,w,r}$. Values for the coverage ratio close to the nominal level of 0.95 indicate well-calibrated intervals.

Table 5 reports the coverage ratios and average interval scores by age group. Model 4 (with temperature and ILI) generally achieves the highest coverage ratio since it is based on scenario forecasts for temperature and ILI. These forecasts also capture better the more extreme excess mortality events compared to the other models. Regarding the interval score, Model 4 performs best for the younger age groups, whereas Model 5 performs best for the older age groups. The main reason for that is that the interval score also accounts for the width of the interval. Differences are most pronounced at higher ages.

	Coverage ratio					Interval score				
Age	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
50–54	**0.92**	0.90	0.91	0.91	0.91	**2.38**	2.42	2.38	2.38	2.38
55–59	0.92	0.90	0.92	**0.92**	0.92	2.64	2.67	2.63	**2.63**	2.63
60–64	**0.91**	0.90	0.91	0.91	0.91	2.88	2.91	2.86	**2.85**	2.85
65–69	0.91	0.88	0.91	**0.91**	0.91	3.11	3.15	3.09	**3.08**	3.09
70–74	0.94	0.91	0.93	**0.94**	0.94	3.12	3.15	3.08	3.08	**3.07**
75–79	0.94	0.90	0.94	**0.95**	0.94	3.33	3.39	3.26	3.26	**3.24**
80–84	0.95	0.91	0.96	**0.97**	0.96	3.80	3.75	3.61	3.62	**3.59**
85–89	0.94	0.91	0.95	**0.97**	0.96	4.25	4.11	3.96	3.97	**3.92**
90–94	0.93	0.91	0.95	**0.96**	0.96	4.40	4.18	4.04	4.05	**4.00**
95+	0.94	0.93	0.95	**0.96**	0.96	3.95	3.67	3.65	3.66	**3.61**
Overall	0.93	0.91	0.93	**0.94**	0.94	3.38	3.34	3.26	3.26	3.24

Table 5: Coverage ratios and interval scores across the five competing model specifications: three benchmarks (Models 1–3) and two versions of our proposed model (Models 4–5).

Visualisation. We evaluate the predictive performance of our proposed model visually in Figure 13 by plotting the observed death rates against the corresponding 95% prediction intervals for three administrative regions in France and for the age groups 50-54, 70-74 and 90-94. Specifically, we focus on Model 4, which incorporates simulated trajectories of the latent mortality index, temperature, and ILI incidence rates in the forecasting procedure. We find that the observed death counts fall mostly within the constructed intervals, even during periods with increased mortality in summer or winter.

5.4 Assessing post-COVID excess mortality

Lastly, we forecast death rates for the 2020–2024 period using our proposed mortality model calibrated on 1990–2019 data. We focus on Model 5 and condition the forecasts on the observed

Figure 13: Forecasted female weekly death rates on log-scale according to Model 4 in Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right) over the period 2015–2019 using simulated exogenous factors. We show the 95% prediction intervals based on 10 000 simulated trajectories of temperature, ILI incidence rates, and the latent mortality index for the age groups 50-54, 70-74, and 90-94. The observed death rates from 2015-2019 are visualized in black with a '+'.

exogenous characteristics (temperature and influenza activity). By doing so, we can isolate the effect of COVID-19, since the model already accounts for the observed temperature conditions and the markedly lower influenza activity during the pandemic. This leads, as such, to a more accurate assessment of excess mortality.

Hereto, we first collect daily gridded average temperatures for the 12 administrative regions of France from the E-OBS data set (Copernicus Climate Data Store) for the period January 1, 2020, to December 31, 2024 and construct population-weighted regional, weekly temperature series using the approach described in Section 2.3. Second, we take weekly incidence rates of influenza-like illnesses (ILI) from the French Sentinelles network for the same period, as covered in Section 2.4. We then transform both observed series into cross-basis matrices using the DLNM specifications introduced in Section 3.3.

By following the forecasting approach for Model 5, Figure 14 shows the 95% prediction intervals of the forecasted weekly death rates for women in three administrative French regions for age groups 50-54, 70-74, and 90-94. We observe that the observed death rates clearly fall outside the 95% uncertainty bound during the severe COVID-19 waves. However, after 2022, death rates still fall outside the bounds for the 70-74 age group, which indicates that there is still post-COVID excess mortality for certain age groups.

6 Conclusion

This paper develops and implements a new high-resolution stochastic mortality projection model that extends the classic Lee-Carter framework to a weekly, region and age-specific context. Our model is specifically designed to disentangle the baseline seasonal mortality patterns from the acute impacts of heat waves, cold spells, and influenza outbreaks. The methodological innovation lies in the integration of distributed lag non-linear models (DLNMs) within a multi-population Negative Binomial Lee-Carter structure. This allows for an in-depth quantification of the non-linear and delayed effects of temperature and influenza on mortality, while accounting

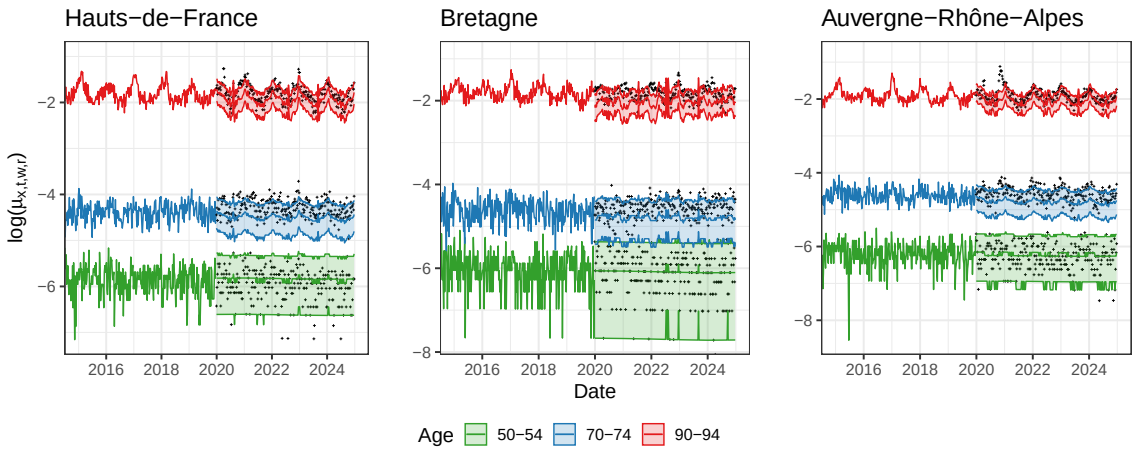


Figure 14: Forecasted weekly death rates for women in Hauts-de-France (left), Bretagne (middle), and Auvergne-Rhône-Alpes (right) over the period 2020–2024 using the realized exogenous factors. We show the 95% prediction intervals based on 10 000 simulated trajectories of the latent mortality index for the age groups 50-54, 70-74, and 90-94. The observed death rates from 2020-2024 are visualized in black with a '+'.

for overdispersion. The application of spatial penalization via a Laplacian smoothness prior ensures a coherent estimation of these effects across neighboring regions.

Through a case study on French data from 1990 to 2019, we calibrate our proposed framework and validate the model's performance. First, we obtain a good in-sample fit and capture seasonal patterns and extreme mortality events well. Pearson residuals confirm a good overall fit. Second, the mortality impacts of both heat and influenza show a strong, increasing effect with age, with the oldest age groups (90+) showing the highest relative risk. Third, the strength of the temperature-mortality and influenza-mortality associations varies significantly between regions in France. The effects of heat are immediate with a clear harvesting effect (mortality displacement), whereas the effects of influenza peak immediately and gradually decrease over several weeks. Third, our model shows superior predictive performance compared to benchmark specifications, with particularly strong gains at older ages.

The model in its current form is designed mainly for short- to medium-term mortality forecasts. The proposed framework is therefore particularly valuable for insurers to price short- to medium-term life-contingent insurance products or for public health to assess excess mortality across different administrative regions and age groups within a country. However, the model is less suitable for long-term mortality projections, as the seasonal (weekly) component is currently assumed to be fixed over time, whereas empirical evidence indicates that seasonal mortality patterns may change over longer horizons (Madaniyazi et al., 2022). Future extensions of this work could therefore deal with incorporating a time-varying seasonal component in the model, which could capture potential shifts in seasonal mortality patterns.

Data and code availability statement

The data sets used in this paper are publicly available. We consult the ten-year files of all individual people who have died since 1970 and the population estimates by department, sex, and five-year age group from INSEE (https://www.insee.fr/fr/information/4769950 and https://www.insee.fr/fr/statistiques/8331297). We retrieve the daily average temperature from the E-OBS dataset in the Copernicus Climate Data Store (https://cds.climate

.copernicus.eu/datasets/insitu-gridded-observations-europe) and the incidences rates of influenza activity from the French Sentinelles network (https://www.sentiweb.fr/france/en/?page=table&maladie=25). The R code used to implement and analyze the case study presented in this paper will be made available on the GitHub repository: https://github.com/jensrobben/wLC-DLNM.

Funding statement

This study is part of the research programme at the Research Centre for Longevity Risk, a joint initiative of NN Group and the University of Amsterdam, with additional funding from the Dutch government's Public-Private Partnership (PPP) programme.

Conflict of interest disclosure

The authors declare no conflict of interest.

References

Barigou, K., Goffard, P.-O., Loisel, S., & Salhi, Y. (2023). Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting*, *39*(2), 674–690. doi: 10.1016/j.ijforecast.2022.01.011

Barnett, A. (2010). What measure of temperature is the best predictor of mortality? *Environmental Research*, *110*(6), 604–611. doi: 10.1097/01.ede.0000362214.90491.53

Bégin, J.-F., Boudreault, M., & Landry, T. (2025). *Modelling seasonal mortality: An age–period–cohort approach.* (Preprint at https://www.researchgate.net/publication/391576373_Modelling_seasonal_mortality_An_age-period-cohort_approach)

Brouhns, N., Denuit, M., & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, *2005*(3), 212–224. doi: 10.1080/03461230510009754

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). *Measuring the longevity risk in mortality projections* (No. 2). Staempfli Verlag AG. (Preprint at https://research.tilburguniversity.edu/files/510433/brouhns2002.PDF)

Center for International Earth Science Information Network – CIESIN – Columbia University. (2018). *Gridded population of the world, version 4 (GPWv4): Population count, revision 11.* Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). (Accessed 01-08-2023) doi: 10.7927/H4JW8BX5

Chaves, S. S., Nealon, J., Burkart, K. G., Modin, D., Biering-Sørensen, T., Ortiz, J. R., ... others (2023). Global, regional and national estimates of influenza-attributable ischemic heart disease mortality. *EClinicalMedicine*, *55*. doi: 10.1016/j.eclinm.2022.101740

Copernicus Climate Change Service, Climate Data Store. (2020). *E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations.* Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed 01-09-2023) doi: 10.24381/cds.151d3ec6

Delwarde, A., Denuit, M., & Partrat, C. (2007). Negative binomial version of the Lee–Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry*, *23*(5), 385–401. doi: 10.1002/asmb.679

Fernández-Durán, J. J., & Gregorio-Domínguez, M. M. (2015). Seasonal mortality for fractional ages in short term life insurance. *Scandinavian Actuarial Journal*, *2015*(3), 266–277. doi: 10.1080/03461238.2013.819028

Gasparrini, A., & Armstrong, B. (2011). The impact of heat waves on mortality. *Epidemiology*, *22*(1), 68. doi: 10.1097/EDE.0b013e3181fdcd99

Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, *29*(21), 2224–2234. doi: 10.1002/sim.3940

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., ... Armstrong, B. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet*, *386*(9991), 369–375. doi: 10.1016/S0140-6736(14)62114-0

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Guibert, Q., Pincemin, G., & Planchet, F. (2025). Impact of climate change on mortality: An extrapolation of temperature effects based on time series data in France. *International Journal of Forecasting*. doi: 10.1016/j.ijforecast.2025.07.004

Guo, Y., Gasparrini, A., Armstrong, B. G., Tawatsupa, B., Tobias, A., Lavigne, E., ... Tong, S. (2017). Heat wave and mortality: A multicountry, multicommunity study. *Environmental Health Perspectives*, *125*(8), 087006. doi: 10.1289/EHP1026

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*, 1–22. doi: 10.18637/jss.v027.i03

Jdanov, D. A., Galarza, A. A., Shkolnikov, V. M., Jasilionis, D., Németh, L., Leon, D. A., ... Barbieri, M. (2021). The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. *Scientific Data*, *8*(1), 235. doi: 10.1038/s41597-021-01019-1

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, *90*, 1–37.

Karlinsky, A., & Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *elife*, *10*, e69336. doi: 10.7554/eLife.69336

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, *63*, 147–152. doi: 10.1016/j.insmatheco.2015.03.023

Koissi, M.-C., Shapiro, A. F., & Högnäs, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, *38*(1), 1–20. doi: 10.1016/j.insmatheco.2005.06.008

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*(419), 659–671. doi: 10.1080/01621459.1992.10475265

Li, H., & Tang, Q. (2022). Joint extremes in temperature and mortality: A bivariate POT approach. *North American Actuarial Journal*, *26*(1), 43–63. doi: 10.1080/10920277.2020.1823236

Li, L., Yan, Z.-L., Luo, L., Liu, W., Yang, Z., Shi, C., ... Ou, C.-Q. (2023). Influenza-associated excess mortality by age, sex, and subtype/lineage: population-based time-series study with a distributed-lag nonlinear model. *JMIR Public Health and Surveillance*, *9*(1), e42530. doi: 10.2196/42530

Li, N., & Lee, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, *42*(3), 575–594. doi: 10.1353/dem.2005.0021

Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens*, *3*(10), e151. doi:

10.1371/journal.ppat.0030151

Lowen, A. C., & Steel, J. (2014). Roles of humidity and temperature in shaping influenza seasonality. *Journal of Virology*, *88*(14), 7692–7695. doi: 10.1128/jvi.03544-13

Madaniyazi, L., Armstrong, B., Chung, Y., Ng, C. F. S., Seposo, X., Kim, Y., ... others (2022). Seasonal variation in mortality and the role of temperature: a multi-country multi-city study. *International Journal of Epidemiology*, *51*(1), 122–133. doi: 10.1093/ije/dyab143

Martínez-Solanas, È., Quijal-Zamorano, M., Achebak, H., Petrova, D., Robine, J.-M., Herrmann, F. R., ... Ballester, J. (2021). Projections of temperature-attributable mortality in Europe: a time series analysis of 147 contiguous regions in 16 countries. *The Lancet Planetary Health*, *5*(7), e446–e454. doi: 10.1016/S2542-5196(21)00150-9

Min, J., Li, H., Nagler, T., & Li, S. (2025). *Assessing climate-driven mortality risk: A stochastic approach with distributed lag non-linear models.* (Preprint at `https://arxiv.org/pdf/2506.00561`)

Pavía, J. M., & Lledó, J. (2022). Estimation of the combined effects of ageing and seasonality on mortality risk: An application to Spain. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(2), 471–497. doi: doi.org/10.1111/rssa.12769

Pitacco, E. (2009). *Modelling longevity dynamics for pensions and annuity business.* Oxford University Press.

Robben, J., Antonio, K., & Kleinow, T. (2025). The short-term association between environmental variables and mortality: evidence from Europe. *Journal of the Royal Statistical Society Series A: Statistics in Society*. doi: 10.1093/jrsssa/qnaf052

Robben, J., Barigou, K., & Kleinow, T. (2025). *Granular mortality modeling with temperature and epidemic shocks: a three-state regime-switching approach.* (Preprint at `https://arxiv.org/pdf/2503.04568`)

Seklecka, M., Pantelous, A. A., & O'Hare, C. (2017). Mortality effects of temperature changes in the United Kingdom. *Journal of Forecasting*, *36*(7), 824–841. doi: 10.1002/for.2473

Valleron, A.-J., Bouvet, E., Garnerin, P., Ménares, J., Heard, I., Letrait, S., & Lefaucheux, J. (1986). A computer network for the surveillance of communicable diseases: the French experiment. *American Journal of Public Health*, *76*(11), 1289–1292. doi: 10.2105/ajph.76.11.1289

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., ... others (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, *399*(10334), 1513–1536. doi: 10.1016/S0140-6736(21)02796-3

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., ... Vachon, P. (2007). *Methods protocol for the human mortality database* (Vol. 9). (University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: `http://mortality.org[version31/05/2007]`)

Wong, J. S., Forster, J. J., & Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, *83*, 206–221. doi: 10.1016/j.ijforecast.2022.01.011

Wood, S. N. (2017). *Generalized additive models: An introduction with R.* CRC Press. doi: 10.1201/9781315370279

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. doi: 10.1080/01621459.2016.1180986