Master's Degree Program in
Statistics and Information Management

Specialization in
Risk Analysis and Management

Integrating Climate Covariates into Mortality Forecasting Models: A Feasibility Study for Longevity Risk Pricing

Alessandro Matteucci

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade NOVA de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Integrating Climate Covariates into Mortality Forecasting Models: A Feasibility Study for Longevity Risk Pricing

by

Alessandro Matteucci

Master Thesis presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Risk Analysis and Management

Supervised by

Jorge Miguel Bravo, NOVA Information Management School

December, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Rome, November 30, 2025

Alessandro Matteucci

ABSTRACT

This thesis investigates whether climate-related determinants—specifically extreme heat intensity—can be meaningfully incorporated into actuarial mortality modelling and whether such integration is empirically justified within the context of Southern European populations. Motivated by the growing intersection between environmental risk, demographic change, and longevity risk management, the study examines how exogenous climate variables may enter mortality-intensity frameworks traditionally structured around age, period, and cohort effects.

At a conceptual level, the thesis formalizes the integration of a standardized heat-anomaly indicator into a mortality modelling architecture consistent with the Generalized Age–Period–Cohort (GAPC) paradigm. At an empirical level, rather than estimating a fully stochastic climate-augmented GAPC model, the study implements a reduced-form Poisson Generalized Linear Model (GLM) panel specification. This approach allows the direct identification of the marginal association between regional annual heat anomalies and mortality rates at ages 65–85, while controlling for flexible age effects, region-specific heterogeneity, and common temporal shocks.

Using harmonized mortality, exposure, and climate data for Southern European regions, the analysis finds that the estimated heat coefficients are generally small in magnitude and not consistently statistically significant under cluster-robust inference. Even where statistical significance is detected, the implied proportional change in annual mortality associated with a one-standard-deviation increase in heat intensity remains modest. Within the annual fixed-effects framework adopted, extreme heat anomalies do not emerge as dominant drivers of mortality dynamics.

The contribution of the thesis lies in establishing a disciplined methodological bridge between climate indicators and actuarial mortality modelling. By linking demographic forecasting, environmental risk, and actuarial finance, the study offers an early step toward climate-aware longevity modelling systems.

KEYWORDS

Climate Change; Longevity Risk; Human Health; Mortality Models;

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Over recent decades, increasing longevity has reshaped demographic and financial systems worldwide. Continuous improvements in life expectancy—one of the defining demographic trends of the twentieth and early twenty-first centuries—have profoundly affected public-pension schemes, annuity markets, and the risk-management strategies of insurers and pension funds. While rising longevity represents a social achievement, it also introduces substantial longevity risk: the uncertainty that individuals may live longer than expected, thereby increasing the cost of life-contingent liabilities.

Accurately forecasting mortality dynamics has therefore become a central challenge in actuarial science and demography, as such projections directly influence the valuation and hedging of long-term instruments including longevity bonds, survivor swaps, and annuities.

Since the seminal work of Lee and Carter (1992), numerous stochastic mortality models have been developed to capture age–period structures and the long-term evolution of mortality. The Generalized Age–Period–Cohort (GAPC) framework, introduced by Renshaw and Haberman (2006) and later extended by Plat (2009), provides a unified representation of these models. GAPC formulations—such as the Lee–Carter, Cairns–Blake–Dowd, and Plat models—express mortality rates as a combination of deterministic age functions and stochastic latent factors that evolve over time, making them particularly suitable for risk-management and pricing applications.

Despite their success, these frameworks remain largely extrapolative, relying exclusively on historical mortality trends while ignoring exogenous determinants that may affect future mortality evolution. Among these, the accelerating pace of climate change represents a structural shift that could alter mortality dynamics over coming decades. Rising temperatures, extreme weather events, and air-pollution episodes have demonstrable effects on mortality—especially among older and vulnerable populations—yet such influences remain largely unexplored in actuarial mortality modelling. Incorporating climate-related covariates is therefore both timely and conceptually significant

The objective of this thesis is to assess the feasibility of integrating climate-related covariates into mortality-forecasting frameworks traditionally used in longevity-risk management. Rather than estimating the definitive magnitude of climate impacts, the study seeks to design and test a methodological architecture capable of linking climate data to mortality projections in a consistent and reproducible manner.

The work develops a prototype of a climate-enhanced mortality model in which conventional demographic structures are extended with environmental covariates—specifically, a heat-related indicator representing the intensity and persistence of thermal stress during summer months. The model is applied to aggregate data for Southern European countries, where exposure to heat extremes is both acute and increasing.

By constructing this prototype, the thesis demonstrates the technical compatibility between actuarial mortality-forecasting systems and climate-based variables, providing a benchmark for future model extensions and paving the way for frameworks that quantify the long-term financial implications of climate-induced mortality shocks.

The remainder of the thesis is structured as follows. Chapter 2 reviews the literature on stochastic mortality modelling, climate-related mortality research, and longevity-linked financial instruments, establishing the interdisciplinary foundations of the study. Chapter 3 presents the theoretical framework, outlining the structure of Generalized Age–Period–Cohort models and discussing the conceptual integration of exogenous climate covariates into mortality intensity equations. Chapter 4 develops the proposed climate-augmented modelling architecture and formalizes the role of heat-related variables within an actuarial setting. Chapter 5 describes the data sources, regional harmonization procedures, and the empirical specification adopted for estimation. Chapter 6 presents the empirical results obtained from the reduced-form panel GLM framework. Chapter 7 evaluates the financial implications of the findings for longevity-risk modelling and actuarial valuation. Finally, Chapter 8 concludes the thesis, discusses methodological limitations, and outlines directions for future research.

2. LITERATURE REVIEW

2.1 MORTALITY MODELLING

Mortality modelling can be broadly classified into three categories: extrapolative, explanatory, and process-based approaches.

- Extrapolative models rely solely on historical mortality trends to project future rates.
- Explanatory models incorporate external factors such as socioeconomic or environmental variables.
- Process-based models simulate underlying biological or epidemiological mechanisms.

Among these, extrapolative models remain the cornerstone of actuarial forecasting because of their parsimony and predictive stability. The Lee–Carter (1992) model, for instance, has long been a reference for capturing long-term mortality dynamics. Within this family, the Cairns–Blake–Dowd (CBD) model (Cairns et al., 2006) has proved particularly effective for older-age mortality, modelling both the level and slope of age-specific mortality rates. Extensions such as the CBD-M7 and M8 variants (Cairns et al., 2009) improve goodness-of-fit through additional cohort and quadratic age terms, though they can encounter issues of data sparsity or over-fitting (Scognamiglio & Marino, 2022).

The Plat (2009) model, combining features of Lee–Carter and CBD formulations, extends applicability across wider age ranges and alleviates some structural limitations of both predecessors.

A number of studies have attempted to integrate extrapolative mortality models with explanatory variables, improving both fit and forecast accuracy for selected populations. **French and O'Hare (2014)** introduced exogenous determinants such as GDP, health expenditure, smoking prevalence, and diet into traditional mortality frameworks. Their Bayesian hierarchical structure first extracts latent factors through Principal Components Analysis (PCA) and then links these to the external variables. The approach improved predictive accuracy, particularly for older populations and data with higher noise, though it remained constrained by the use of aggregate, non-age-specific covariates and the assumption of linear relationships.

Seklecka et al. (2017) employed a modified Lee–Carter model incorporating temperature as an additional covariate. Using Poisson regression and cointegration analysis, they demonstrated that climatic variables can be embedded in extrapolative frameworks, though the results were sensitive to data limitations and linearity assumptions.

Bayesian methods have since become an important avenue for dealing with uncertainty in mortality modelling. These frameworks incorporate prior information and generate probabilistic assessments of parameters, thereby capturing model and estimation risk. For example, **Bravo (2021)** applied a Bayesian model-ensemble approach for valuing longevity-linked annuities, assigning posterior weights to competing stochastic models and improving predictive robustness. Bayesian techniques have also been used to refine mortality forecasts in small populations or regions with sparse data, enabling smoother estimation of local trends (Liu & Li, 2016).

Recent advancements extend beyond statistical inference toward machine-learning methods. **Levantesi and Pizzorusso (2019)** demonstrated that algorithms such as random forests, gradient boosting, and decision trees outperform traditional stochastic models in terms of predictive accuracy, especially where data exhibit high variability or structural change. **Bravo (2021)** further highlighted the promise of recurrent neural networks (RNNs) with gated recurrent unit (GRU) and long short-term memory (LSTM) architectures for multivariate forecasting of age-specific mortality, noting their superior fit relative to Lee–Carter and CBD models. These approaches, however, demand significant computational resources and specialised expertise.

Overall, both Bayesian and machine-learning methodologies expand the scope of mortality forecasting by accommodating non-linearity and integrating exogenous determinants—economic, environmental, and health-related—thereby enhancing the explanatory power and practical relevance of longevity-risk models.

2.2 Climate Change and Human Mortality

Climate change is now unequivocally recognised as a global phenomenon whose impacts have intensified over recent decades. Variability in temperature, precipitation, and wind patterns has direct and indirect consequences for human health, wellbeing, and mortality. Foundational publications such as *The Impacts of Climate Change on Human Health* (2016) and the *2024 Lancet Countdown Europe Report* confirm this intensification and document its multifaceted effects.

A seminal contribution in this domain is **Gasparrini et al. (2015)**, a multicountry analysis using Distributed Lag Non-Linear Models (DLNMs) and meta-analysis to quantify temperature-related mortality. The study revealed a U-shaped relationship between temperature and mortality, with both cold and hot extremes increasing deaths relative to the local minimum-mortality temperature (MMT). Cold exposure accounted for a larger share of excess deaths globally, though heat-related mortality was more pronounced in warmer regions.

Building on this foundation, **Seklecka et al. (2017)** examined the temperature–mortality relationship for the United Kingdom using Poisson regression, Johansen cointegration tests, and ARIMA forecasting. Their results confirmed that temperature variations significantly affect mortality, particularly among populations aged 50 and older. Nonetheless, their linear specification may under-represent the complexity of the relationship.

Further regional insights are offered by **Martínez-Solanas and Basagaña (2019)**, who studied temporal changes in temperature-related mortality in Spain. Using DLNMs across multiple decades, they observed declining cold-related mortality but increasing heat-related deaths, reflecting improved adaptation to cold and rising vulnerability to extreme heat. This temporal perspective underscores the need for models that allow for evolving sensitivity and adaptive capacity.

More recent large-scale studies, notably **Ballester et al. (2023)** and **Gallo et al. (2024)**, provide continental evidence on heat-related mortality in Europe. Ballester et al. estimated ≈ 61,700 heat-attributable deaths during the record-breaking summer of 2022, with the highest rates in Southern Europe—particularly Italy, Spain, and Greece—and with women aged 80 and above identified as the most vulnerable group. Gallo et al. (2024) analysed the summer of 2023 and estimated ≈ 47,700 heat-related deaths, highlighting the role of adaptation in reducing impacts by an estimated 80 per cent. Together, these studies demonstrate the magnitude of Europe’s heat-mortality burden and provide robust epidemiological baselines for integrating climatic variables into mortality models.

2.3 Pricing Principles

Pricing longevity-linked securities is inherently challenging because of the illiquidity of longevity markets and the uncertainty surrounding future mortality. Two principal valuation paradigms exist: **risk-neutral** and **real-world** approaches.

Under the **risk-neutral** framework, survival probabilities are adjusted using financial no-arbitrage principles, leading to theoretical fair values based on discounted expected pay-offs. Methods such as the **Wang**, **Esscher**, and **canonical valuation** transforms belong to this category but require strong assumptions of market completeness.

Conversely, **real-world** pricing principles derive values from historical survival distributions, introducing risk loadings via measures such as the **standard-deviation** or **variance** principle. These approaches are computationally simpler but rely heavily on subjective parameters, such as chosen discount rates and risk margins.

Leung et al. (2018) advanced this literature by incorporating Bayesian methods to account for parameter uncertainty in pricing the EIB-BNP longevity bond. Their findings stress the importance of aligning premium principles with the underlying mortality model—Lee–Carter or CBD—and demonstrate that no single principle dominates across settings.

Tang and Li (2021) further compared twelve premium principles, spanning both risk-neutral and real-world domains, across different stochastic mortality models. They concluded that mortality-model uncertainty exerts greater influence on pricing outcomes than the choice of premium principle. Real-world methods generally yield higher premiums due to explicit risk loadings, whereas risk-neutral methods remain more constrained by theoretical assumptions.
Overall, these studies highlight the need for flexible pricing frameworks that accommodate parameter and model uncertainty and remain consistent with the demographic assumptions underlying mortality projections.

In summary, the literature on mortality modelling and longevity-risk pricing highlights a continuous evolution from purely extrapolative approaches toward frameworks that incorporate external determinants and quantify model uncertainty. Yet, despite significant methodological progress, environmental and climatic factors remain largely absent from actuarial mortality research. The growing body of epidemiological evidence linking temperature extremes to excess mortality suggests that this omission may become increasingly consequential. The next chapter therefore shifts the focus to the **climate–mortality interface**, reviewing the main environmental drivers of mortality and setting the conceptual foundation for their integration into stochastic mortality models.

3. CLIMATE-RELATED VARIABLES

3.1

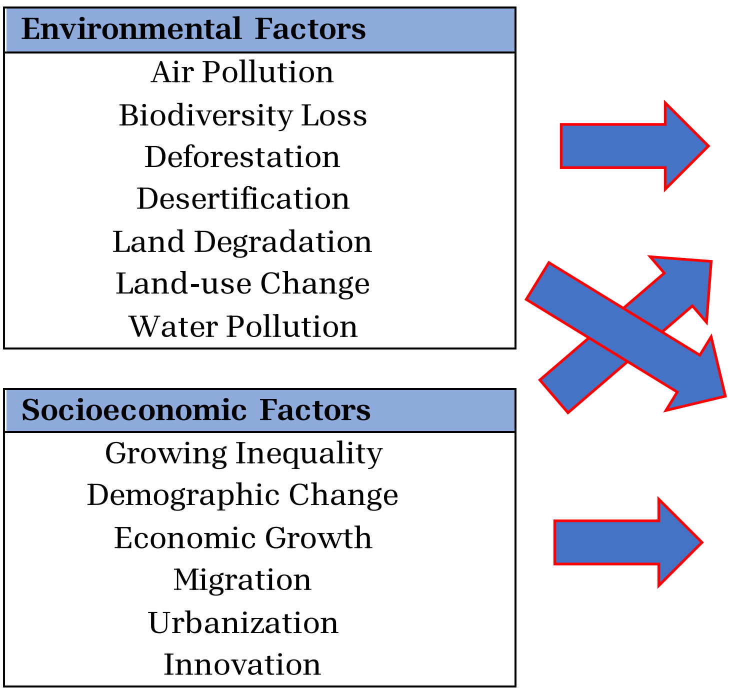
This chapter examines the principal pathways through which climate change influences human health and mortality, with the objective of identifying variables that can be meaningfully incorporated into mortality-forecasting frameworks. The discussion builds on recent climate-health assessments, including the *Lancet Countdown (2024)*, the *IPCC Sixth Assessment Report (2022)*, and *Gupta & Venkataraman (2024)*.

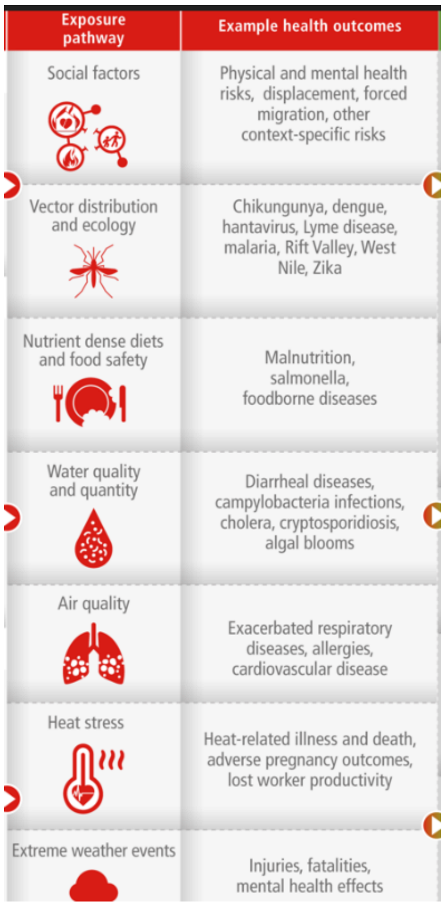
These sources converge on the view that climate change affects health outcomes through multiple, interacting mechanisms that can be grouped into **environmental** and **socio-economic** dimensions.

- **Environmental factors** include air pollution, biodiversity loss, deforestation, desertification, land degradation, land-use change, and water pollution.

- **Socio-economic factors** encompass growing inequality, demographic change, economic growth, migration, urbanisation, and technological innovation.

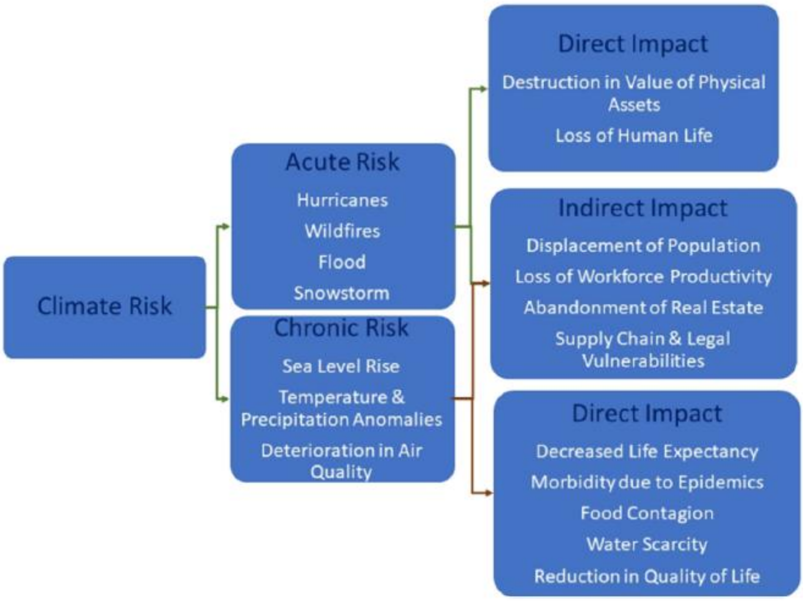
Figure 1 **– environmental and socio-economic factors and their impact on human health**





From an insurance and risk-management perspective, these determinants can be further classified as **acute** or **chronic** risks (Gupta & Venkataraman, 2024).

Figure 2 **– Acute and chronic impacts of climate change on society**



In Europe, the ***Lancet Countdown (2024)*** and related reports identify the Mediterranean basin as one of the regions most exposed to both acute and chronic climatic stressors. Southern European countries like Spain, Italy, Greece, and Portugal are expected to face the greatest intensification of these risks, particularly in areas bordering the Mediterranean Sea[1] (Stamos, I., & Diakakis, M. (2024)). The main acute and chronic factors currently affecting these regions—and projected to intensify in the future—include :

- **Extreme-heat events**, with increasing frequency and duration;
- **Wildfires**, exacerbated by higher temperatures and aridity;
- **Droughts**, reducing water availability and agricultural productivity; and
- **Air pollution**, particularly fine particulate matter intensified by stagnant high-pressure systems.
- **High precipitation / Floods / Storms, also** with increasing frequency and duration

[1] Stamos, I., & Diakakis, M. (2024). Mapping Flood Impacts on Mortality at European Territories of the Mediterranean Region within the Sustainable Development Goals (SDGs) Framework. *Water, 16*(17), 2470.

Another consistent finding across climate-health studies is the **unequal distribution of impacts**. Vulnerability to climate hazards varies with wealth, urbanisation, and social infrastructure. Populations in regions with weaker health systems or higher proportions of elderly residents are more susceptible to temperature-related mortality shocks.

Given this context, the present research concentrates on heat-related factors—the most immediate and quantifiable climatic stressor. Among the various environmental risks affecting Southern Europe, extreme heat represents the climatic variable for which the causal link with short-term mortality is most consistently documented in the epidemiological literature. Unlike droughts, wildfires, or floods—whose mortality effects are often indirect, geographically localized, or mediated through economic and infrastructural channels—heat exposure operates directly through physiological stress mechanisms, particularly among elderly populations.

From a methodological perspective, heat intensity can be measured using harmonised meteorological datasets with high spatial and temporal consistency, allowing the construction of standardized regional indicators comparable across countries. This contrasts with other climate stressors whose measurement may rely on heterogeneous event-based reporting or non-continuous administrative records.

For these reasons, heat-related indicators constitute a natural starting point for assessing the feasibility of incorporating climate variables into actuarial mortality models.

4. METHODOLOGY: CONCEPTUAL FRAMEWORK FOR CLIMATE-ENHANCED MORTALITY MODELING

4.1 Theoretical Rationale

Mortality forecasting is a cornerstone of actuarial science, underpinning the valuation of annuities, pensions, life insurance products in general and longevity-linked securities. Traditional stochastic mortality models—such as **Lee–Carter (1992)**, **Cairns–Blake–Dowd (2006)**, and **Plat (2009)**—have demonstrated strong performance in extrapolating historical mortality improvements. These models decompose mortality rates into **age**, **period**, and, when applicable, **cohort** components, assuming that past improvement trends persist into the future.

In recent years, however, this purely extrapolative paradigm has been challenged by evidence of **mortality stagnation or reversals** in several European countries, particularly in southern regions, often coinciding with episodes of **extreme climatic stress** such as heatwaves, wildfires, or severe air pollution. These patterns suggest that environmental drivers may influence mortality dynamics beyond what demographic factors alone can explain.

Consequently, a growing body of research (e.g., *Gasparrini et al., 2015; Vicedo-Cabrera et al., 2021*) advocates the inclusion of **exogenous environmental variables**, especially temperature-related indicators, in mortality-forecasting frameworks.

The central methodological question motivating this thesis is therefore whether a **climate-sensitive extension** of stochastic mortality models can be designed to preserve actuarial structure while allowing for external environmental influences. The approach developed here does not replace the traditional **Generalised Age–Period–Cohort (GAPC)** framework; rather, it explores its theoretical adaptability to accommodate **exogenous covariates** associated with climatic variability.

4.2 Conceptual Foundation: The Generalized Age-Period-Cohort Framework

The GAPC class of models (Renshaw & Haberman, 2006; Haberman & Renshaw, 2011) provides a unifying structure for stochastic mortality dynamics.

Let $m_{x,t}$ denote the **central death rate** at age x and calendar year t. In logarithmic form, the general specification can be expressed as:

$$\ln(m_{x,t}) = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x} + \varepsilon_{x,t}, \tag{1}$$

where:

- N denotes the number of period factors included in the model; its value depends on the chosen GAPC specification (e.g., $N = 1$ for Lee–Carter, $N = 2$ for Plat-type models)
- α_x captures the **average age-specific mortality profile**;
- $\kappa_t^{(i)}$ are **period factors** describing time evolution;
- $\beta_x^{(i)}$ represent the **age sensitivities** to those temporal factors;
- γ_{t-x} denotes **cohort effects**, capturing generational deviations; and
- $\varepsilon_{x,t}$ is an **idiosyncratic error term**.

Specific models emerge as special cases of the GAPC framework, possibly under different response scales. The Lee–Carter model corresponds to $N = 1$ with no cohort term on the log-rate scale of (1). The CBD family adopts the same age–period factor structure but is typically specified on the logit scale for mortality probabilities $q_{x,t}$, making it an analogous GAPC formulation under a different transformation. The Plat (2009) model extends the CBD specification by adding an additional period component and a cohort term.

GAPC models constitute a standard modeling framework in actuarial mortality analysis, decomposing mortality dynamics into age, period, and cohort effects and indicating where exogenous drivers can be incorporated. In the empirical analysis, climate variables are incorporated directly as covariates in a Poisson regression with exposure offsets. This reduced-form approach prioritizes interpretability and feasibility given the limited time overlap and spatial granularity of the available climate panel.

4.3 Incorporating Exogenous Covariates

At a conceptual level, the general GAPC structure defined in (1) can be extended to accommodate exogenous covariates influencing mortality dynamics. In particular, a climate-related driver may be introduced additively into the linear predictor as:

$$\ln(m_{x,t}) = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x} + \delta\, C_t + \varepsilon_{x,t}, \qquad (2)$$

where C_t denotes an external covariate—in this context, a standardised measure of heat intensity—and δ quantifies its proportional effect on log-mortality. Equation (2) represents a **conceptual target specification**, illustrating how climate-related stressors may enter a GAPC-type mortality model.

Conceptually, the inclusion of C_t allows mortality rates to accelerate or decelerate as a function of climatic variation, independently of long-term demographic trends captured by latent period and cohort components. In this sense, the covariate perturbs the mortality surface around its baseline age–period–cohort structure.

From a statistical standpoint, mortality counts are naturally modelled using a Poisson likelihood with a log link and an exposure offset, consistent with the standard formulation underlying GAPC models. The Poisson assumption applies conditionally on the mortality intensity and does not impose homogeneity across ages or cohorts; instead, all structured heterogeneity is absorbed into the systematic component of the model. The log link ensures positivity of mortality rates and preserves the multiplicative interpretation of effects, while the exposure offset separates changes in mortality intensity from variations in population size.

In practice, estimating a fully extended stochastic GAPC model of the form (2) presents several challenges, including limited temporal overlap between reliable climate and mortality data, potential multicollinearity between climate covariates and latent period effects, and non-stationarity and spatial heterogeneity in climatic processes.

Given these constraints, this thesis adopts a prototype approach. Rather than estimating the complete stochastic model in (2), the empirical analysis operationalises the same conceptual logic through a reduced-form Generalised Linear Model applied to aggregated mortality and climate data. This approach serves as a feasibility study and provides initial evidence on the magnitude and direction of climate-related mortality effects, while leaving fully integrated GAPC extensions to future data-rich research.

4.4 Prototype Modeling Framework

Let the mortality count in region $r$and year $t$be modelled as:

$$D_{r,t} \sim \text{Poisson}(E_{r,t}\,\mu_{r,t})$$

with

$$\ln\,(\mu_{r,t}) = \alpha + \beta_1\,\text{Year}_t + \beta_2\,\text{Heat}_t + \varepsilon_{r,t} \qquad (3)$$

where $D_{r,t}$denotes the observed number of deaths, $E_{r,t}$ is the population exposure-to-risk (included as an offset term $\ln E_{r,t}$), and Heat_t is a standardised heat-intensity index aggregated at the annual level. The coefficient β_2 captures the marginal sensitivity of mortality to temperature-related stress.

Equation (3) constitutes a **reduced-form empirical proxy** for the conceptual specification in (2). Age and cohort dimensions are implicitly integrated out through aggregation, allowing the analysis to focus on the average response of mortality to climatic variation at the regional level. This simplification is driven by data constraints and by the objective of isolating the short-run effect of heat intensity on mortality, rather than producing full age-specific mortality forecasts.

Despite its parsimonious structure, the specification retains a key interpretive mechanism shared with GAPC models: expected mortality is modelled as a rate conditional on exposure, with covariate effects entering multiplicatively through the log link. In this sense, Equation (3) preserves the logic of proportional perturbations to baseline mortality while avoiding the identification and estimation challenges associated with fully stochastic latent-factor models.

Scenario projections are generated by shifting the heat-intensity index by fixed multiples of its empirical standard deviation ($+0.5$ SD, $+1$SD, $+2$SD), approximating moderate to severe climate-intensification conditions. The resulting projections do not constitute full predictive mortality tables; rather, they measure relative deviations in expected mortality under hypothetical climate stress scenarios, consistent with the feasibility-oriented scope of the analysis.

4.5 From Mortality Modeling to actuarial and financial evaluation

A final methodological step connects the mortality modeling framework to the actuarial and financial evaluation of longevity risk. Longevity-linked securities—such as survivor forwards, longevity swaps, and longevity bonds—are priced using projected survival probabilities, which in turn depend on the underlying mortality dynamics.

Let $m_{x,t}(C_t)$ denote the central death rate at age x and calendar time t, conditional on the realisation of an exogenous climate variable C_t. In the modeling framework considered in this thesis, climate effects enter mortality multiplicatively through the log-mortality specification, such that deviations in C_t shift mortality intensities relative to a baseline trajectory.

Given a projected mortality surface, the expected survival probability of an individual aged x at time t over a future horizon of length s, conditional on information available at time t, can be expressed as:

$$ {}_s p_x = \exp \left\{ - \int_0^s m_{x+u,\, t+u} (C_{t+u})\, du \right\} \qquad (4) $$

where the dependence of $m_{x+u,t+u}$ on C_{t+u} reflects the propagation of climate-related mortality effects along future age–time paths.

Equation (4) provides the analytical bridge between mortality modeling and financial cashflows. Once survival probabilities are obtained, expected payoffs of longevity-linked instruments can be computed under alternative climate scenarios, allowing—at least in principle—the assessment of how climate-driven deviations in mortality affect fair-value estimates.

4.6 Positioning of the Empirical Strategy

The conceptual extension developed in the preceding sections establishes a theoretical framework in which exogenous climate risk may enter a Generalized Age–Period–Cohort (GAPC) mortality model as an additional determinant of the mortality intensity. In this structure, climate variables act alongside latent age, period, and cohort effects, potentially modifying projected mortality trajectories and, consequently, actuarial valuations derived from them.

The empirical analysis implemented in Chapters 5 and 6 does not estimate a fully stochastic GAPC model augmented with climate covariates. Instead, it adopts a reduced-form panel Generalized Linear Model (GLM) specification to evaluate the conditional association between climate intensity and observed mortality rates. This approach serves as a feasibility assessment of the conceptual extension proposed above.

The GLM framework allows for the direct estimation of the marginal effect of climate intensity on mortality, while controlling for age structure and unobserved heterogeneity across time and regions. Although this reduced-form strategy does not embed climate risk within a dynamic stochastic mortality process, it provides an empirical benchmark for assessing whether climate variables exert statistically and economically meaningful effects on mortality within the available data horizon.

Accordingly, the empirical component of this thesis should be interpreted as a preliminary validation exercise of the theoretical extension, rather than a full structural estimation of a climate-augmented GAPC model. The results obtained from the panel framework inform the actuarial implications discussed in subsequent chapters, while leaving the dynamic stochastic integration of climate risk into GAPC forecasting models as an avenue for future research.

5. DATA AND EMPIRICAL FRAMEWORK

5.1 Data Sources and Study Scope

This study integrates demographic and climate datasets to construct a harmonised regional panel suitable for estimating climate–mortality relationships within a Poisson regression framework. The empirical design follows a structured multi-phase data pipeline in which mortality counts, population exposures, climate variables, and geographical harmonisation rules are progressively aligned and merged.

Four primary data inputs are employed.

(a) Mortality Data.

Mortality counts were obtained from Eurostat's regional demographic statistics, disaggregated by NUTS2 region, age group, sex, and calendar year (and week where applicable). Observations include the number of registered deaths (OBS_VALUE) for each region–age–sex–time cell. Data cleaning and harmonisation procedures ensured temporal continuity and consistent regional coding across NUTS revisions.

(b) Population Exposure Data.

Population-at-risk (exposure) data were retrieved from Eurostat at the same regional, age, sex, and yearly level as the mortality data. These exposures serve as the denominator in mortality rate construction and enter the Poisson regression model through a logarithmic offset term. Harmonisation procedures were applied to ensure exact geographic and temporal alignment with the mortality panel.

(c) Climate Data.

Climate variables were constructed from gridded reanalysis datasets (ERA5/ERA5-Land type sources), initially available at high spatial (0.25° × 0.25° grid) and daily temporal resolution. Through spatial aggregation (grid-to-NUTS mapping) and temporal aggregation (daily to weekly and seasonal summaries), region-level temperature indicators were derived. A standardised heat-intensity index expressed in regional standard deviation units, was engineered to provide cross-regional comparability and numerical stability in regression estimation.

(d) Geographical Harmonisation and Region Rosters.

Because NUTS regional definitions evolve over time, a canonical crosswalk was constructed to ensure consistent geographic boundaries across the study period. Two regional stability scopes were defined during preprocessing: **STABLE** and the more restrictive **STABLE_TOT95**, the latter requiring complete and continuous coverage across mortality, exposure, and climate datasets.

All empirical estimations presented in this thesis are conducted on the **STABLE_TOT95 scope**, stratified by sex (male and female). This ensures maximum geographic consistency and avoids structural breaks arising from regional redefinitions or incomplete coverage.

Study Population and Time Window

The analysis focuses on older-age mortality, where temperature-related vulnerability is most pronounced. Age groups are restricted to the thesis-relevant window (as defined in Phase A preprocessing), ensuring consistency with actuarial modeling conventions and avoiding confounding from younger-age mortality dynamics.

The empirical time window is determined by the overlap between harmonised mortality, exposure, and climate datasets following quality checks and completeness validation. The final modeling panel includes only region–year observations for which all three components are simultaneously available within the STABLE_TOT95 universe.

Empirical Unit of Observation

After harmonisation and merging (described in subsequent sections), the empirical unit of observation is a regional panel cell defined by:

$$(\text{region}, \text{year}, \text{age}, \text{sex})$$

For each cell, the dataset contains:

- observed deaths,
- corresponding exposure,
- climate covariate
- identifiers required for fixed-effects encoding.

This structure forms the basis for the panel Poisson estimation described later on.

5.2 Spatial Harmonisation and Study Scope Definition

Geographic Coverage and Country Selection

The empirical analysis focuses on four Southern European countries:

- **Greece (EL)**
- **Spain (ES)**
- **Italy (IT)**
- **Portugal (PT)**

These countries were selected due to their documented exposure to elevated summer heat intensity and their relevance within the climate–mortality literature. The geographic unit of analysis is the NUTS2 regional level.

Because NUTS definitions evolve across revisions (splits, mergers, and boundary changes), a canonical crosswalk framework was constructed to ensure consistent regional coding over time. Regions that could not be harmonised across the full data window were excluded to prevent artificial structural breaks in mortality or exposure series.

Definition of Stability Scopes

Two geographic universes were defined during preprocessing:

- **STABLE:** Regions consistently observable across the majority of the study period.
- **STABLE_TOT95:** A more restrictive subset requiring complete and continuous coverage across mortality, exposure, and climate datasets over the full modeling window.

The empirical estimation in this thesis is conducted exclusively on the **STABLE_TOT95 scope**.

After applying harmonisation and completeness criteria, the final STABLE_TOT95 universe consists of:

- **51 NUTS2 regions**
- Spanning the four countries EL, ES, IT, and PT

A small number of regions were excluded due to incomplete coverage or structural discontinuities (e.g., EL41, EL43, ES63, ES64).

This restriction ensures geographic consistency and avoids estimation bias arising from region entry/exit dynamics.

Temporal Scope

The final estimation time window is 1998 to 2023. This window reflects the full overlap of harmonised mortality, exposure, and climate datasets within the STABLE_TOT95 universe after quality validation.

Although one year is treated as the reference category in the fixed-effects specification, the full panel spans the entire 1998–2023 period.

Age Range and Population Focus

The analysis concentrates on older-age mortality, where heat vulnerability is empirically strongest.

The final age range included in the GLM estimation is 65 to 85.

This restriction ensures:

- Demographic homogeneity,
- Alignment with actuarial longevity modeling conventions,
- Avoidance of confounding from younger-age mortality dynamics.

Sex Stratification

All estimations are performed separately for:

- **Male**
- **Female**

The STABLE_TOT95 panel is split by sex and GLM estimation is conducted independently for each sex. This allows for differential climate sensitivity across demographic groups and avoids imposing homogeneity in mortality response.

Final Panel Structure

After harmonisation, filtering, and merging, the empirical unit of observation is defined as (*region, year, age, sex*) for each cell in the STABLE_TOT95 universe, the dataset contains:

- observed deaths,
- corresponding exposure,
- the standardized climate covariate
- identifiers required for fixed-effects encoding.

This harmonised regional panel forms the basis for the empirical model.

5.3 Climate Data Processing and Heat Covariate Construction

To quantify extreme heat exposure at the regional level in a manner consistent with epidemiological evidence, daily maximum temperature data ($T_{\max}$) were processed and transformed into annual region-specific heat indicators aligned with the mortality panel.

Spatial and Temporal Aggregation

Daily $T_{\max}$ observations were obtained from high-resolution gridded reanalysis datasets (ERA5/ERA5-Land type sources), available at approximately 0.25° × 0.25° spatial resolution.

Each grid cell was mapped to its corresponding NUTS2 region using canonical regional boundaries. For each region g and day d, regional daily maximum temperature was computed as the mean of grid cells contained within that region:

$$T_{\max,g,d} = \frac{1}{N_g}\sum_{i\in g} T_{\max,i,d} \tag{5}$$

where N_g denotes the number of grid cells assigned to region g.

All heat metrics entering the regression are defined at the **annual level**, focusing on the extended summer season May–September (MJJAS), the period during which heat-related mortality effects are empirically strongest. Climate variables are therefore aligned with mortality and exposure data at the regional annual level:

$$(g, t)$$

ensuring exact temporal consistency in the final panel.

Region-Specific Extreme Heat Threshold

To capture extreme heat relative to local climatology, a region-specific threshold T_g^* was constructed. For each NUTS2 region g, T_g^* corresponds to the smoothed 95th percentile of daily $T_{\max}$ during May–June over the baseline period 1998–2010:

$$T_g^* = \mathrm{P}_{95}\left(T_{\max,g,d}^{\text{May–June, 1998–2010}}\right) \tag{6}$$

This approach ensures that "extreme" heat is defined relative to each region's own historical distribution rather than by imposing a uniform temperature cutoff across countries with heterogeneous climates.

Heatwave Definition and Annual Heat Metric

Heatwave episodes were defined as sequences of at least three consecutive days satisfying:

$$T_{\max,g,d} > T_g^* \tag{7}$$

This persistence criterion excludes isolated hot days and is consistent with epidemiological findings indicating that sustained heat exposure is more strongly associated with excess mortality than single-day spikes.

For each region g and year t, the annual number of days within MJJAS belonging to such ≥ 3-day heatwave spells was computed:

$$\text{HWdays}_{g,t}$$

To ensure cross-regional comparability, the variable was standardized within each region:

$$heat\_z_{g,t} = \frac{\text{HWdays}_{g,t} - \mu_g}{\sigma_g} \tag{8}$$

where μ_g and σ_g denote the time-series mean and standard deviation of heatwave days in region g over the full sample period.

The standardized variable $heat\_z_{g,t}$ therefore measures how unusually intense heatwave activity is in a given year relative to that region's own historical distribution. A one-unit increase in $heat\_z$ corresponds to a one standard deviation increase in heatwave persistence.

This persistence-based metric (HWdays) constitutes the primary climate covariate used in the empirical estimation reported in Chapter 6.

Alternative Heat Metric (Robustness)

An alternative intensity-based metric, capturing cumulative temperature exceedances above T_g^*, was also constructed and standardized analogously. While this measure reflects aggregate thermal stress rather than spell duration, it was not retained as the primary specification in the final STABLE_TOT95 estimation.

Details of the alternative construction and associated robustness checks are reported in **Appendix A**.

Final Climate Variable in the Empirical Model

In the final empirical specification:

- $heat\_z_{g,t}$ is defined at the annual regional level,
- corresponds to MJJAS heatwave persistence,
- is standardized within region,
- enters the Poisson regression linearly.

No weekly climate aggregation is used in the final model. Mortality counts, exposures, and heat variables are all aligned at the annual NUTS2 × age × year level for the period 1998–2023.

5.4 Mortality and Exposure Construction

The mortality and exposure components of the empirical panel are constructed directly at the annual regional level. No temporal aggregation from higher-frequency data is required in the final modeling pipeline.

Mortality Counts

Annual mortality counts were obtained from Eurostat at the NUTS2 regional level, disaggregated by age group and sex. Let

$$D_{g,a,t}$$

denote the number of registered deaths in region g, age group a, and year t, for:

$$g \in \{1, \ldots, 51\}, a \in \{65, \ldots, 85\}, t \in \{1998, \ldots, 2023\}$$

The analysis focuses on ages 65–85, where temperature-related vulnerability is empirically most pronounced and where mortality dynamics are consistent with actuarial longevity modeling conventions.

All mortality observations entering the final estimation belong to the STABLE_TOT95 universe (panel).

Population Exposures

Population exposure data were obtained from Eurostat at the same regional, age, sex, and annual resolution as the mortality counts.

Let

$$E_{g,a,t}$$

denote the population at risk in region g, age group a, and year t.

Exposure data were harmonised geographically using the same NUTS crosswalk framework applied to mortality counts, ensuring exact alignment of regional identifiers over time. Observations with incomplete exposure information were excluded during preprocessing to preserve a fully rectangular panel structure.

Mortality Rate Definition

The annual mortality rate for each region–age–year cell is defined as:

$$m_{g,a,t} = \frac{D_{g,a,t}}{E_{g,a,t}} \tag{9}$$

While this rate is useful for descriptive analysis and interpretation, the regression model is estimated in count form using a Poisson specification with exposure incorporated as an offset term.

Offset Construction

Under the Poisson model with log link, the conditional expectation of deaths is specified as:

$$\mathbb{E}[D_{g,a,t} \mid X] = E_{g,a,t}\, \mu_{g,a,t} \tag{10}$$

where $\mu_{g,a,t}$ denotes the underlying *mortality intensity*.

Taking logarithms yields:

$$\log \mathbb{E}[D_{g,a,t}] = \log(E_{g,a,t}) + \log(\mu_{g,a,t}) \tag{11}$$

Accordingly, the logarithm of exposure,

$$\log(E_{g,a,t})$$

enters the regression as a fixed offset term. This ensures that estimated coefficients capture proportional effects on mortality intensity rather than effects on raw death counts.

All exposure values are strictly positive in the final STABLE_TOT95 panel, ensuring the offset is well defined.

Final Panel Structure

Following geographic harmonisation, age restriction (65–85), and merging of mortality, exposure, and climate components, the final modeling dataset forms a fully rectangular annual panel over the period 1998–2023.

For each sex separately, the panel consists of:

$$51 \text{ regions} \times 21 \text{ ages} \times 26 \text{ years}$$

yielding a total of:

$$51 \times 21 \times 26 = 27{,}846$$

region–age–year observations.

This rectangular structure results from the application of the STABLE_TOT95 filtering criteria and completeness checks described in Section 5.2. Only region–year cells for which mortality, exposure, and climate variables are simultaneously available are retained.

The resulting dataset constitutes the empirical basis for the construction of the model matrix and Poisson estimation described in the following sections.

5.5 Construction of the Master Modeling Panel

Following geographic harmonisation, age restriction, and construction of annual heat indicators, the mortality, exposure, and climate components were merged into a single modeling dataset.

Let the final empirical unit of observation be defined as (g, a, t) where:

- $g = 1, \dots, 51$ denotes NUTS2 regions in the STABLE_TOT95 universe,
- $a = 65, \dots, 85$ denotes age,
- $t = 1998, \dots, 2023$ denotes calendar year.

For each observation, the dataset contains:

- $D_{g,a,t}$: observed annual deaths,
- $E_{g,a,t}$: annual population exposure,
- $heat_{g,t}$: standardized annual heatwave persistence index,
- region identifier,
- year identifier,
- age variable.

After filtering and merging, the resulting dataset forms a fully rectangular panel consisting of 27,846 region–age–year observations per sex.

Sex stratification is performed prior to model estimation. That is, separate panels are constructed for males and females, and the regression model is estimated independently for each sex. This avoids imposing homogeneity in heat sensitivity across demographic groups.

The merged dataset is then transformed into a model matrix representation suitable for Poisson estimation. Specifically, the following components are constructed:

- A response vector y containing $D_{g,a,t}$,
- An offset vector containing $\log(E_{g,a,t})$,
- A design matrix X encoding age spline basis functions, region fixed effects, year fixed effects, and the heat covariate.

This matrix formulation allows consistent and reproducible estimation across specifications.

5.6 Empirical Model Specification

Poisson Mortality Model

Mortality counts are modeled using a Poisson generalized linear model with log link. The conditional expectation of deaths is specified as:

$$\mathbb{E}[D_{g,a,t}] = E_{g,a,t} \cdot \mu_{g,a,t} \tag{12}$$

where $\mu_{g,a,t}$ denotes the mortality intensity.

As previously presented, the log exposure term enters the regression as a fixed offset, ensuring that estimated coefficients capture proportional effects on mortality rates.

Final Estimated Specification

The empirical model estimated for each sex is:

$$\log\ \mathbb{E}[D_{g,a,t}] = \log(E_{g,a,t}) + f(a;\mathrm{df}=4) + \alpha_g + \delta_t + \beta\, heat_{g,t} \tag{13}$$

Where:

- $f(a;\mathrm{df}=4)$ denotes a natural spline function of age with four basis functions,
- α_g represents region fixed effects for all 51 NUTS2 regions,
- δ_t represents year fixed effects for all 26 years (1998–2023),
- $heat_{g,t}$ is the standardized annual heatwave persistence index,
- No intercept term is included in the specification.

The model is therefore a fully saturated two-way fixed effects Poisson regression estimated on an annual regional panel.

The exclusion of an intercept ensures that all region and year fixed effects are explicitly included rather than normalized relative to a reference category.

Interpretation of the Heat Coefficient

Under the log-link specification, the coefficient β measures the semi-elasticity of mortality with respect to heat exposure.

A one-unit increase in $heat_{g,t}$ (corresponding to a one standard deviation increase in heatwave persistence) implies a multiplicative change in mortality intensity of:

$$\exp(\beta)$$

Thus, $\exp(\beta) - 1$ can be interpreted as the proportional change in mortality associated with a one standard deviation increase in extreme heat exposure.

5.7 Scenario Framework and Forecasting Design

In addition to historical estimation, the empirical framework was extended to incorporate a counterfactual scenario structure designed to evaluate the potential impact of intensified heat anomalies on future mortality outcomes.

The intended forecasting design proceeded as follows:

1. Construct a forward prediction grid at the $g \times a \times t$ level (region × age × year).
2. Retain estimated structural components from the fitted Poisson GLM, including the age spline, region fixed effects, and year effects.
3. Apply deterministic shifts to the standardized annual heat variable $heat_{g,t}$ (e.g., +0.5 SD, +1 SD, +2 SD).
4. Compute predicted mortality using the estimated coefficients under the log-link specification.

Under model (13) counterfactual scenarios are generated by replacing $heat_{g,t}$ with $heat_{g,t} + \Delta$, holding all other components fixed.

This design provides a transparent mechanism for propagating exogenous climate shifts through the actuarial mortality structure.

However, numerical instability arose in the out-of-sample prediction stage, primarily related to the reconstruction of the offset term and link-scale amplification under the exponential transformation. Consequently, forward scenario projections were not retained as quantitative results in this thesis.

The scenario framework is therefore presented as a methodological extension of the estimation model rather than as a finalized forecasting output. Its structure remains relevant for future implementation once prediction stability is fully validated.

6. RESULTS AND DISCUSSION

6.1 Descriptive Evidence from the Estimation Panel

Before presenting regression results, this section provides descriptive evidence on the structure of the estimation panel and the raw relationship between mortality and standardized heat anomalies.

The final modelling dataset consists of 27,846 observations per sex, covering the period 1998–2023 for 51 NUTS2 regions across Greece (EL), Spain (ES), Italy (IT), and Portugal (PT), restricted to ages 65–85 under the STABLE_TOT95 definition.

Table 1– Panel Structure Summary (STABLE_TOT95)

Component	Value
Observations	27,846
Time span	1998–2023
Countries	EL, ES, IT, PT
Age range	65–85
Panel structure	$g \times a \times t$
Frequency	Annual
Sample definition	STABLE_TOT95

6.2.1 Mortality Structure and Age Gradient

Mortality Rate Distribution

For males, the mean mortality rate is approximately 0.0464, with median 0.0363 and standard deviation 0.0306. For females, the mean mortality rate is 0.0280, with median 0.0191 and standard deviation 0.0231. In both cases, the mean exceeds the median, indicating right-skewness in the mortality rate distribution.

Maximum observed mortality rates reach 0.1649 (males) and 0.1510 (females), reflecting very high mortality at advanced ages in specific region–year cells. Substantial heterogeneity is therefore present across ages, regions, and years.

This level of dispersion supports the use of a count-data framework with exposure offsets.

Figures 3 and 4 show a smooth monotonic increase in mortality counts from age 65 to age 85 for both sexes.

Figure 3 - Average Deaths by Age (Males)

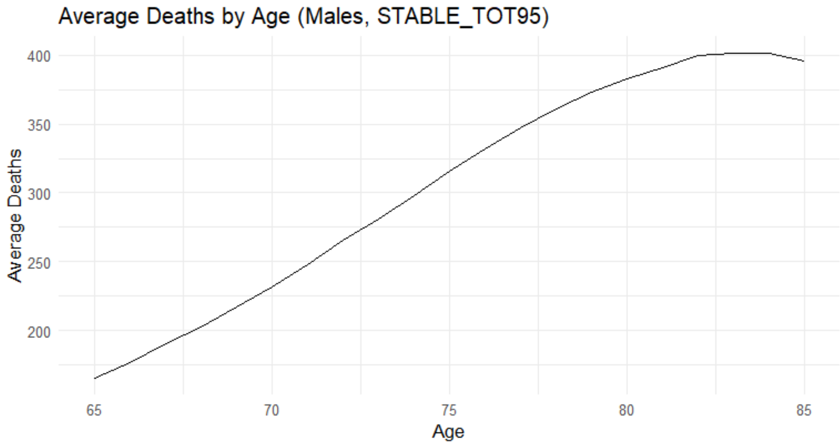
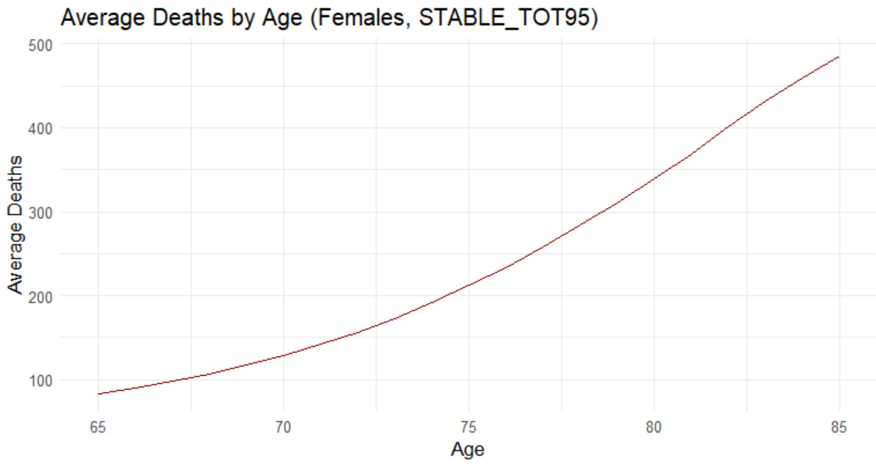


Figure 4 - Average Deaths by Age (Females)



For females, the curve exhibits a convex shape consistent with Gompertz-type mortality acceleration at older ages. Males display a similar increasing pattern, though with consistently higher mortality levels across all ages.

The empirical age gradient confirms that mortality structure is demographically coherent and supports the use of a flexible age specification (natural spline with 4 degrees of freedom) in the regression model.

Table 2 – Mortality and Exposure Summary Statistics

Variable	Sex	Mean	SD
Mortality Rate	Male	0.0464	0.0306
Mortality Rate	Female	0.0280	0.0231

Country (Region)	Tot Mean Exposure
EL	3,475
ES	8,997
IT	10,771
PT	7,159

Exposure varies considerably across countries due to demographic size differences. These differences justify the inclusion of the log-exposure offset and regional fixed effects in the Poisson specification.

6.2.2 DISTRIBUTION OF THE STANDARDIZED HEAT VARIABLE

The standardized heat anomaly $heat_{g,t}$ (denoted heat_z in the dataset) represents the within-region standardized annual MJJAS heatwave persistence index described in Section 5.3.

Table 3 – Distribution of Standardized Heat Variable $heat_{g,t}$

Country	Mean	SD	Min	Max
EL	0.204	1.19	-1.27	4.86
ES	0.311	1.49	-1.81	6.03
IT	0.032	1.14	-1.56	5.05
PT	0.308	1.23	-1.69	3.42

The standardized variable is approximately centered around zero, as expected from within-region standardization. However, substantial cross-country dispersion is visible. Spain exhibits the highest variability and the most extreme positive anomalies (up to approximately +6 standard deviations), while Italy shows a near-zero mean anomaly over the sample period.

These patterns confirm meaningful heterogeneity in climate exposure across Southern Europe.

6.2.3 Raw Association Between Mortality and Heat

Scatter plots of deaths versus heat_z reveal extremely high dispersion and a visually flat linear trend for both sexes.

Figure 5 – Deaths vs Standardized Heat (Males)

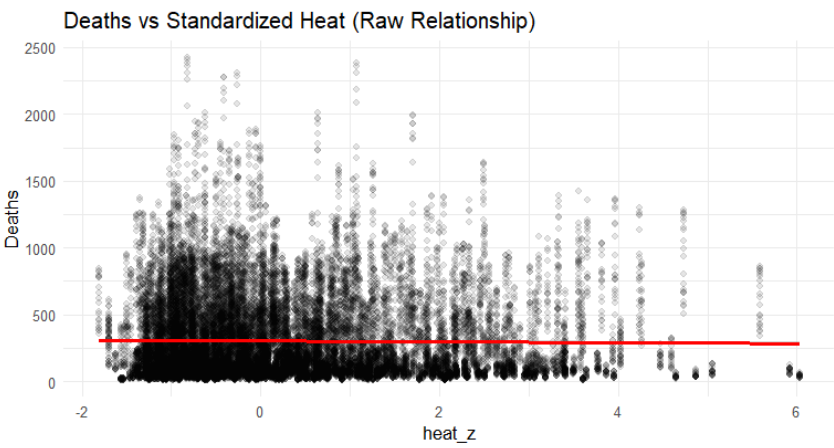
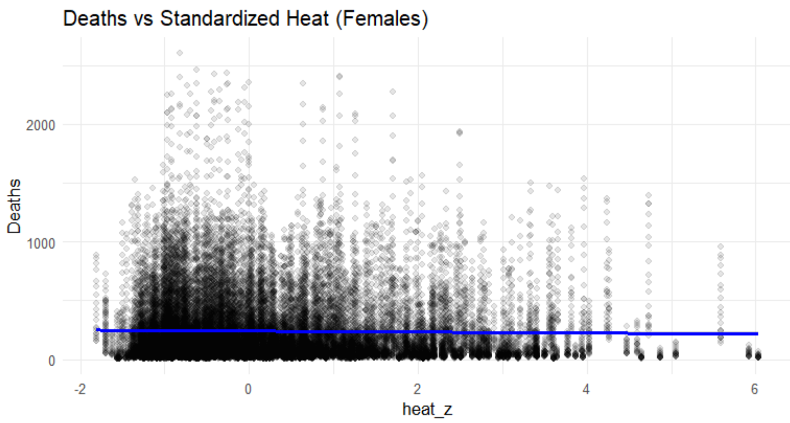


Figure 6 - Deaths vs Standardized Heat (Females)



The unconditional correlation between deaths and heat_z is approximately −0.014 for males and similarly close to zero for females.

This weak raw association is not unexpected. Mortality levels are primarily driven by age structure, regional fixed effects, and temporal dynamics. Any heat effect, if present, is therefore conditional on these structural components rather than visible in unconditional cross-sectional variation.

To further explore potential non-linear patterns, mortality rates were examined by deciles of $heat_{g,t}$.

Figure 7 – Mortality rate by Heat Decile (Females)

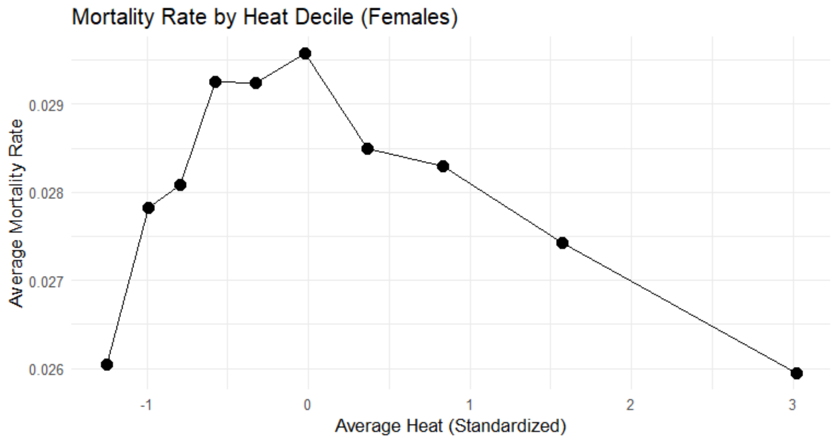
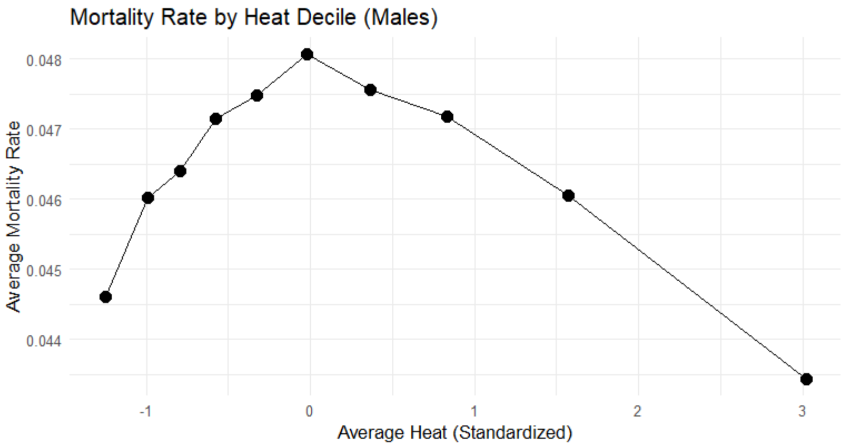


Figure 8 - Mortality rate by Heat Decile (Males)



For both sexes, mortality rates increase from cold anomalies toward moderate heat levels and then decline at extreme heat deciles, suggesting a mild inverted-U relationship. This pattern indicates that the raw association between heat anomalies and mortality may be non-linear and not fully captured by a simple linear term in the regression model.

Overall, the descriptive evidence indicates:

- A strong demographic age gradient in mortality,
- Substantial cross-country heterogeneity in climate exposure,
- A weak unconditional linear association between heat anomalies and mortality,
- Potential non-linearities in the heat–mortality relationship.

These findings motivate the multivariate fixed-effects Poisson specification presented in the next section.

6.3 Estimated Heat Effects

Statistical inference is based on cluster-robust sandwich standard errors clustered at the NUTS2 regional level. After fitting the Poisson GLM with log link and log-exposure offset, a cluster-robust covariance matrix (vcovCL) is computed using the regional identifier as the clustering unit. Reported standard errors, z-statistics, and p-values are derived from this clustered covariance matrix rather than from model-based Poisson standard errors.

Table 4 reports, for each country and sex:

- Estimated heat coefficient $\hat{\beta}$
- Cluster-robust standard error
- z-statistic
- p-value
- Risk ratio $\exp(\hat{\beta})$

Table 4 - Estimated heat coefficients β_{heat} by country, sex and scope

Country	Sex	Scope	Heat coef. (β)	SE(β)	z-stat	p-value	95% CI RR (lower)	95% CI RR (upper)
EL	F	STABLE	−0.0044	0.0036	−1.22	0.222	0.989	1.000
EL	M	STABLE	−0.0011	0.0024	−0.45	0.654	0.994	1.000
ES	F	STABLE	0.0003	0.0018	0.19	0.853	0.997	1.000
ES	M	STABLE	0.0009	0.0012	0.79	0.432	0.999	1.000
IT	F	STABLE	−0.0025	0.0031	−0.80	0.426	0.991	1.000
IT	M	STABLE	0.0005	0.0023	0.21	0.835	0.996	1.000
PT	F	STABLE	0.0116	0.0000	321.00	<0.001	1.010	1.010
PT	M	STABLE	0.0154	0.0001	200.00	<0.001	1.010	1.020

Under the log-link specification, a one-unit increase in $heat_{g,t}$ (corresponding to a one standard deviation increase in the standardized heat anomaly) multiplies expected mortality by $\exp(\hat{\beta})$. For small values of $\hat{\beta}$, the semi-elasticity interpretation applies and $\hat{\beta}$ approximates the percentage change in expected mortality.

Across specifications, the estimated heat coefficients are generally small in magnitude.

In several cases, point estimates are close to zero and cluster-robust standard errors are of comparable magnitude, leading to z-statistics below conventional significance thresholds. In other cases, statistical significance is obtained under clustered inference, but the estimated coefficients remain quantitatively small.

Importantly, coefficient signs are not uniformly consistent across countries. This absence of a systematic directional pattern suggests that, within this annual fixed-effects framework, the marginal heat effect is heterogeneous and not robustly aligned across Southern Europe.

The small estimated linear effects are consistent with the descriptive findings presented earlier:

- The unconditional correlation between mortality and $heat_{g,t}$ is close to zero.
- Decile analysis suggests a mild non-linear (inverted-U) association rather than a strong monotonic relationship.
- Mortality variation is primarily driven by age structure, regional heterogeneity, and temporal dynamics.

Within a log-linear additive framework that includes rich fixed effects, the residual linear contribution of $heat_{g,t}$ is therefore limited.

This does not imply that heat has no effect on mortality in general. Rather, it indicates that, at the annual aggregation level and conditional on region and year effects, the remaining linear heat signal is small.

6.4 Scenario Forecasting Attempt and Identified Instability

In addition to estimating historical heat–mortality associations, an extension of the empirical framework was designed to generate forward mortality projections under counterfactual heat intensification scenarios as presented earlier.

During implementation of the forecasting block, severe numerical instability emerged in the out-of-sample prediction stage.

Specifically:

- In some forecast windows, the linear predictor η reached very large values.
- Under the log-link, $\mathbb{E}[D] = \exp(\eta)$, which led to exponential amplification.

- Predicted death counts in certain forecast files reached implausible magnitudes (orders far exceeding realistic demographic levels).

These magnitudes dominated aggregates and rendered scenario comparisons non-interpretable.

Targeted diagnostics indicated that instability was associated with the reconstruction of the prediction grid and, in particular, with the handling of the exposure offset and regressor encoding in the out-of-sample environment. Although the in-sample estimation stage was stable and well-behaved, the prediction layer did not replicate the fitted structure in a numerically portable way.

6.4.1 IMPLICATIONS

Because forward predicted levels were numerically unstable, scenario outputs cannot be interpreted as quantitative mortality projections.

Accordingly:

- Absolute forecast levels are not reported.
- Scenario differences relative to baseline are not presented as empirical results.
- No pricing or actuarial conclusions are drawn from these unstable projections.

6.4.2 FUTURE IMPLEMENTATION REQUIREMENTS

Resolving the forecasting instability requires:

- Strict reconstruction of the offset term in the forecast grid consistent with the fitted model object,
- A verified single prediction pathway (e.g., explicit link-scale prediction with validated offset handling),
- Unit checks comparing in-sample fitted values to reconstructed predictions,
- Incremental forecasting validation at small regional subsets before full-grid scaling.

These implementation steps represent a technical extension of the current framework rather than a conceptual limitation of the model itself.

7. LONGEVITY RISK APPLICATIONS AND PRICING PERSPECTIVE

7.1 Linking Mortality Forecasts to Longevity Risk

As we have seen in previous chapters, longevity risk arises when mortality improvements exceed expectations, leading to higher-than-anticipated survival probabilities and, consequently, larger future liabilities for pension funds and life insurers. Conversely, the same uncertainty creates opportunities for investors through longevity-linked instruments, which provide diversification against traditional financial market risks.

The stochastic mortality framework developed in this thesis—though empirically climate-neutral—offers the essential foundation for quantifying, pricing, and managing longevity risk. It produces internally consistent mortality forecasts and survival probabilities that can be directly incorporated into financial valuation.

While the climate variable heat_z did not alter mortality projections in Chapter 6, the model remains mathematically compatible with risk-pricing methodologies. This chapter bridges the demographic and financial dimensions of longevity modelling, showing how mortality forecasts translate into the pricing of longevity-linked securities under different premium principles.

7.2 Longevity-Linked Securities: Concept and Rationale

Longevity-linked securities (LLS) are financial instruments whose pay-offs depend on the realised survival of a reference population. Their purpose is to hedge *systematic* longevity risk—the non-diversifiable component of uncertainty in life expectancy.

The main classes of LLS include:

1. **Longevity Bonds** – Coupon payments are linked to realised survival rates of a given cohort.
2. **Survivor Forwards (S-Forwards)** – Bilateral contracts exchanging a fixed payment for a floating payment based on actual survival ratios at maturity.
3. **Longevity Swaps** – Long-term agreements exchanging fixed annuity-style payments for floating payments tied to realised survival outcomes.

Let:

- $l_{x,t}$ denote the number of survivors aged $x$at time t;
- $q_{x,t}$ denote the one-year probability of death of an individual aged x at time t;
- $p_{x,t} = 1 - q_{x,t}$ the one-year survival probability an individual aged x at time t.

The ***t*-year survival probability** for an individual aged x at time 0 is:

$$ {}_t p_x = \prod_{j=0}^{t-1} \quad p_{x+j,j} \tag{14} $$

Under a stochastic mortality model, $q_{x,t}$ (and hence $p_{x,t}$) is forecasted as a random variable whose expected value drives pricing.

When discounted by a risk-free or risk-adjusted yield curve y_t, the present value of a survival-contingent payment stream is obtained as:

$$ V = \sum_{t=1}^{T} \quad {}_t p_x \; e^{-y_t t} \tag{15} $$

Because markets demand compensation for systematic longevity risk, a **risk premium** is added via **premium principles** or **probability transforms**.

7.3 From Mortality Forecasts to Pricing Inputs

The translation from demographic modelling to financial valuation proceeds in three steps:

1. **Mortality Modelling**

(a) Forecast death rates $m_{x,t}$ using the chosen GAPC model (e.g., Plat 2009).

(b) Convert these into survival probabilities ${}_t p_x$ using

$$ {}_t p_x = \exp\,(-\sum_{j=0}^{t-1} \quad m_{x+j,j}) \tag{16} $$

2. **Discounting**

Apply a term structure of interest rates y_t, typically fitted via the **Nelson–Siegel–Svensson (NSS)** curve:

$$y_t = \beta_0 + \beta_1 \frac{1 - e^{-t/\tau_1}}{t/\tau_1} + \beta_2 [\frac{1 - e^{-\frac{t}{\tau_1}}}{\frac{t}{\tau_1}} -] + \beta_3 [\frac{1 - e^{-\frac{t}{\tau_2}}}{\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_2}}] \quad (17)$$

3. **Risk Adjustment (Premium Principle)**

- **Wang Transform**

$$F^*(x) = \Phi[\Phi^{-1}(F(x)) + \lambda] \quad (18)$$

where F is the CDF of the loss variable L, Φ is the standard normal CDF, and $\lambda$is the market price of longevity risk. The transform distorts the distribution, increasing the weight of adverse (high-loss) outcomes.

- **Proportional Hazard Transform (PHT)**

$$S^*(t) = [S(t)]^{e^{-\lambda}} \quad (19)$$

where $S(t)$is the survival function. A positive λ effectively increases the hazard, leading to higher prices for protection against longevity risk.

- **Standard-Deviation Principle**

$$P = \mathbb{E}[L] + \lambda \, \mathrm{SD}(L) \quad (20)$$

where $L$is the loss (or payoff) random variable and λscales the risk loading. The premium is the expected value plus a multiple of the standard deviation.

Each principle yields a risk-adjusted present value that increases monotonically with the market price of risk parameter λ, reflecting a higher compensation required for bearing longevity risk.

Table 5 - Premium principles for longevity-linked pricing

Premium Principle	***Adjustment Mechanism***	***Interpretation***
Wang Transform	F∗(x)=Φ[Φ−1(F(x))+λ]	Distorts survival distribution by a market price of risk (λ)
Proportional Hazard Transform (PHT)	S∗(t)=[S(t)]e−λ	Multiplies survival probabilities by hazard factor (e^{λ})

Standard-Deviation Principle	P=E[L]+λSD(L)	Adds a premium proportional to payoff variability

Each principle yields a risk-adjusted present value $V^*(\lambda)$ that increases with the market price of risk parameter λ.

7.4 Conceptual Illustration Using the Current Model

To demonstrate the integration of the mortality model with pricing mechanics, consider a 20-year S-forward referencing the total Spanish population aged 65 in 2025.

The S-forward's floating payment at maturity depends on the realised survival ratio:

$$\frac{l_{65+20}(2045)}{l_{65}(2025)} = {}_{20}p_{65}$$

where $l_{65}(2025)$ is the initial number of survivors at age 65 in 2025 and $l_{85}(2045)$ is the number of survivors at age 85 in 2045.

Under the risk-neutral measure $\mathbb{Q}$, the fair present value of a €1 notional S-forward paying the survival ratio in 2045 is:

$$V_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}[{}_{20}p_{65}]$$

where $r$is the constant yield and $T = 20$years.

Introducing a Wang transform with parameter λmodifies the survival probability as:

$${}_{20}p_{65}^{(\lambda)} = 1 - F^*(1- {}_{20}p_{65})$$

where F^* denotes the Wang-distorted distribution of the loss.

The corresponding risk-adjusted value becomes:

$$V_0^{(\lambda)} = e^{-rT} \; {}_{20}p_{65}^{(\lambda)}$$

For illustration, assume:

- baseline survival probability ${}_{20}p_{65} = 0.37$;
- constant yield $r = 2\%$;
- maturity $T = 20$years.

The discount factor is then approximately

$$e^{-rT} = e^{-0.4} \approx 0.67$$

Suppose that, under increasing values of the market price of risk λ, the Wang transform produces the following risk-adjusted survival probabilities:

- ${}_{20}p_{65}^{(0.00)} = 0.37$
- ${}_{20}p_{65}^{(0.25)} = 0.39$
- ${}_{20}p_{65}^{(0.50)} = 0.42$

The corresponding present values per €1 notional are then approximately:

- $V_0^{(0.00)} \approx 0.67 \times 0.37 \approx 0.25$
- $V_0^{(0.25)} \approx 0.67 \times 0.39 \approx 0.26$
- $V_0^{(0.50)} \approx 0.67 \times 0.42 \approx 0.28$

Rounded to two decimal places, this can be summarised as:

Table 6 - Ilustrative pricing of a 20-year S-forward under Wang Transform

λ	**Discount Rate**	$V_0^{(\lambda)}$ **(€ per €1 notional)**
0.00	2 %	0.25
0.25	2 %	0.26
0.50	2 %	0.28

This simple example captures the key qualitative effect: as λincreases, the risk-adjusted survival probability and, therefore, the present value of the S-forward increase, reflecting a higher premium for bearing longevity risk.

Because Chapter 6 found that the climate covariate heat$_{g,t}$has a negligible effect on mortality forecasts, the climate-adjusted survival probabilities coincide with the baseline ones:

$$ {}_{20}p_{65}^{\text{(climate scenario)}} \approx {}_{20}p_{65}^{\text{(baseline)}} $$

for all of the $+0.5$ SD, $+1$ SD, and $+2$ SD heat scenarios considered. Hence,

$$ V_0^{(\lambda,\text{scenario})} \approx V_0^{(\lambda,\text{baseline})} $$

so the S-forward price is effectively identical across climate scenarios—representing a *climate-neutral pricing outcome* under the current data and model specification.

8. CONCLUSION AND FUTURE WORK

8.1 Research Objective and Conceptual Contribution

This thesis set out to examine whether climate risk—specifically extreme heat intensity—can be meaningfully integrated into actuarial mortality modelling and whether such integration is empirically justified within the context of Southern European populations.

The conceptual contribution of the study lies in demonstrating how exogenous climate variables may, in principle, enter a mortality modelling framework traditionally structured around age, period, and cohort effects. By extending the Generalized Age–Period–Cohort (GAPC) formulation to include a climate sensitivity term, the thesis provides a formal representation of how environmental risk factors could influence mortality intensity and, consequently, projected survival trajectories.

The empirical component of the thesis does not estimate a fully stochastic climate-augmented GAPC model. Instead, it evaluates the reduced-form association between regional heat intensity and mortality rates within a panel Generalized Linear Model (GLM) framework. This strategy serves as a feasibility assessment of the conceptual extension, allowing the climate parameter to be identified directly while controlling for age structure and unobserved regional and temporal heterogeneity.

In this way, the thesis bridges conceptual actuarial modelling and empirical panel analysis, providing an initial quantitative benchmark for the potential integration of climate risk into longevity modelling systems.

8.2 Empirical Findings

To assess the empirical relevance of this conceptual extension, a reduced-form Poisson GLM panel framework was implemented using harmonised regional mortality, exposure, and climate data.

The estimated climate coefficients obtained in Section 6.3 are generally small in magnitude and, in several cases, statistically insignificant under cluster-robust inference. Even where statistical significance is detected, the implied percentage change in mortality associated with a one standard deviation increase in the standardized heat anomaly remains modest.

Formally, the null hypothesis of no heat effect ($\beta = 0$) cannot be systematically rejected across specifications.

At annual frequency and under the fixed-effects panel structure adopted in this study, extreme heat intensity does not emerge as a dominant explanatory driver of mortality at ages 65–85. The empirical evidence therefore suggests that standardized annual heat anomalies exert, at most, a limited marginal influence on observed mortality within the sample horizon.

These findings must, however, be interpreted within the structural confines of the modelling framework.

8.3 Interpretation within the Annual Fixed-Effects Framework

The empirical specification is an annual, region-level reduced-form Poisson model that includes:

- A flexible age spline component,
- Region fixed effects,
- Year fixed effects,
- A standardized annual heat anomaly term.

Several structural features of this framework are relevant for interpretation.

First, annual aggregation may attenuate short-term mortality responses to heat. Epidemiological studies typically document effects at daily or weekly frequency, often using distributed lag nonlinear models (DLNMs). In contrast, the present specification evaluates whether cumulative annual heat anomalies translate into measurable shifts in annual mortality after controlling for demographic structure and region-specific heterogeneity.

Second, the inclusion of year fixed effects absorbs common time shocks, including continent-wide climatic trends and macro-level mortality dynamics. Identification of the heat coefficient therefore relies on within-region deviations from annual trends rather than on long-run climatic evolution. This is a conservative identification strategy that isolates localized anomalies but may attenuate broader structural climate signals.

Third, the model incorporates a linear heat term. Descriptive decile analysis suggests that the raw association between heat anomalies and mortality may exhibit non-linear or threshold behavior. A strictly log-linear specification may therefore not fully capture potential non-linear effects.

Within this annual fixed-effects structure, the estimated heat coefficient should be interpreted as a conditional marginal association rather than as a structural causal estimate of temperature-related mortality risk.

8.4 Model Scope, Aggregation Level and Climate Context

The objective of the empirical model is not to replicate short-term epidemiological heat–mortality studies. Instead, it evaluates whether standardized annual heat anomalies produce detectable shifts in actuarial mortality counts at the regional level.

The absence of large annual effects does not contradict the existence of short-term mortality responses to extreme heat. Rather, it suggests that:

- Short-term mortality spikes may be partially offset within the annual horizon,
- Mortality displacement (“harvesting”) effects may operate within the same year,
- Regional demographic structure and common temporal dynamics dominate aggregate annual variation.

Accordingly, the results are best interpreted as evidence regarding annual actuarial sensitivity to climate anomalies, rather than as high-frequency hazard estimates.

The findings therefore highlight the complexity of translating short-term environmental shocks into persistent shifts in long-term mortality trends relevant for actuarial valuation.

8.5 Statistical Diagnostics and Robustness Considerations

The Poisson GLM converged successfully in all country–sex specifications. Cluster-robust covariance matrices were computed using NUTS2 regions as clustering units.

Several robustness considerations remain relevant:

- Overdispersion relative to the Poisson assumption may affect inference; dispersion statistics could be evaluated in supplementary analysis.
- Alternative variance specifications (e.g., quasi-Poisson or negative binomial) may be explored in future extensions.
- The linear heat specification may be augmented with polynomial or threshold terms to assess potential non-linearities.
- Sensitivity to the inclusion or exclusion of year fixed effects could further clarify identification channels.

These considerations do not invalidate the current results but indicate potential avenues for methodological refinement.

8.6 Directions for Future Research

The framework developed in this thesis provides a foundation for more advanced integration of climate risk into actuarial mortality modelling.

Future research may extend this work in several directions:

- Embedding climate covariates directly within stochastic GAPC state-space models;
- Modelling persistent climate trends rather than annual anomalies;
- Incorporating high-frequency mortality data;
- Exploring non-linear or threshold effects of extreme heat;
- Integrating climate risk into risk-neutral longevity pricing models.

A dynamic climate-augmented mortality process capable of capturing persistent environmental trends could allow for structural shifts in long-term survival projections, potentially altering the pricing and hedging of longevity-linked securities under future climate scenarios.

BIBLIOGRAPHICAL REFERENCES

Balbus, J., Crimmins, A., Gamble, J. L., Easterling, D. R., Kunkel, K. E., Saha, S., & Sarofim, M. C. (2016). Ch. 1: Introduction: Climate change and human health. In *The impacts of climate change on human health in the United States: A scientific assessment* (pp. 25–42). U.S. Global Change Research Program. https://brewminate.com/climate-change-and-human-health/

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R. F., et al. (2023). Heat-related mortality in Europe during the summer of 2022. *Nature Medicine, 29*(8), 1857–1866. https://doi.org/10.1038/s41591-023-02419-z

Blake, D. (2018). Longevity: A new asset class. *Journal of Asset Management, 19*(6), 353–361. https://doi.org/10.1057/s41260-018-0084-9

Blake, D., Cairns, A. J. G., Dowd, K., & Kessler, A. R. (2018). Still living with mortality: The longevity risk transfer market after one decade. *British Actuarial Journal, 24*, 1–45. https://doi.org/10.1017/S1357321718000314

Bozikas, A., Badounas, I., & Pitselis, G. (2022). Pricing longevity bonds under a credibility framework with limited available data. *Risks, 10*(5), 96. https://doi.org/10.3390/risks10050096

Bravo, J. M. (2021). Pricing participating longevity-linked life annuities: A Bayesian model ensemble approach. *European Actuarial Journal, 11*(2), 375–409. https://doi.org/10.1007/s13385-021-00279-w

Bravo, J. M. (2021). Forecasting longevity for financial applications: A first experiment with deep learning methods. In *Communications in Computer and Information Science* (Vol. 1525, pp. 64–78). Springer. https://doi.org/10.1007/978-3-030-93733-1_17

Bravo, J. M., & de Freitas, N. E. M. (2017). Valuation of longevity-linked life annuities. *Insurance: Mathematics and Economics, 78*, 212–229. https://doi.org/10.1016/j.insmatheco.2017.09.009

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

Cairns, A. J. G., Blake, D., & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance, 73*(4), 687–718. https://doi.org/10.1111/j.1539-6975.2006.00195.x

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal, 13*(1), 1–35. https://doi.org/10.1080/10920277.2009.10597538

French, D., & O'Hare, C. (2014). Forecasting death rates using exogenous determinants. *Journal of Forecasting, 33*(8), 640–650. https://doi.org/10.1002/for.2309

Gallo, E., Quijal-Zamorano, M., Méndez Turrubiates, R. F., et al. (2024). Heat-related mortality in Europe during 2023 and the role of adaptation in protecting health. *Nature Medicine, 30*(12), 3101–3105. https://doi.org/10.1038/s41591-024-03186-1

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 multi-country observational study. *The Lancet, 386*(9991), 369–375. DOI: 10.1016/S0140-6736(14)62114-0

Intergovernmental Panel on Climate Change (IPCC). (2022). *Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (H.-O. Pörtner et al., Eds.). Cambridge University Press. doi. org/10.1017/9781009157926

Lancet Countdown. (2024). *The 2024 Europe report of the Lancet Countdown on health and climate change: Unprecedented warming demands unprecedented action. The Lancet Public Health.* Advance online publication. DOI: 10.1016/S2468-2667(24)00055-0

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

Levantesi, S., & Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks, 7*(1), 26. https://doi.org/10.3390/risks7010026

Liu, Y., & Li, J. S.-H. (2016). The locally linear Cairns–Blake–Dowd model: A note on delta-nuga hedging of longevity risk. *ASTIN Bulletin: The Journal of the IAA, 46*(3), 709–744. doi:10.1017/asb.2016.33

Martínez-Solanas, È., & Basagaña, X. (2019). Temporal changes in temperature-related mortality in Spain and effect of the implementation of a heat health prevention plan. *Environmental Research, 169*, 102–113. https://doi.org/10.1016/j.envres.2018.11.006

Leung, M., Fung, M. C., & O'Hare, C. (2018). A comparative study of pricing approaches for longevity instruments. *Insurance: Mathematics and Economics, 82*, 1–13. https://doi.org/10.1016/j.insmatheco.2018.06.010

Michealson, A., & Mulholland, J. (2014). Strategy for increasing the global capacity for longevity risk transfer: Developing transactions that attract capital markets investors. *The Journal of Alternative Investments, 17*(2), 18–27.

Niu, G., & Melenberg, B. (2014). Trends in mortality decrease and economic growth. *Demography, 51*(5), 1755–1773. https://doi.org/10.1007/s13524-014-0328-3

Pitacco, E., Denuit, M., Haberman, S., & Olivieri, A. (2009). *Modelling longevity dynamics for pensions and annuity business.* Oxford University Press.

Plat, R. (2009). On stochastic mortality modelling. *Insurance: Mathematics and Economics, 45*(3), 393–404. https://doi.org/10.1016/j.insmatheco.2009.08.006

Scognamiglio, S., & Marino, M. (2023). Backtesting stochastic mortality models by prediction interval-based metrics. *Quality and Quantity, 57*, 3825–3847. https://doi.org/10.1007/s11135-022-01537-z

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

Stamos, I., & Diakakis, M. (2024). Mapping Flood Impacts on Mortality at European Territories of the Mediterranean Region within the Sustainable Development Goals (SDGs) Framework. *Water*, *16*(17), 2470.

Tang, S., & Li, J. (2021). Market pricing of longevity-linked securities. *Scandinavian Actuarial Journal, 2021*(5), 408–436. https://doi.org/10.1080/03461238.2020.1852105

World Health Organization (WHO). (2014). *Quantitative risk assessment of the effects of climate change on selected causes of death, 2030s and 2050s.* World Health Organization. https://www.who.int/publications/i/item/9789241507691

APPENDIX A - ALTERNATIVE HEAT EXPOSURE SPECIFICATION (CUMULATIVE INTENSITY METRIC)

A.1 Motivation

In addition to the persistence-based heatwave metric used in the main empirical specification, an alternative heat exposure measure was constructed to capture cumulative thermal intensity rather than duration.

While the primary specification (Section 5.3) measures the number of heatwave days within a year, the alternative metric quantifies the aggregate magnitude of temperature exceedances above a region-specific extreme heat threshold. This distinction allows assessment of whether mortality responses are more sensitive to prolonged heatwave persistence or to cumulative thermal stress.

A.2 Region-Specific Threshold

As in the main specification, a region-specific threshold T_g^* was defined as the smoothed 95th percentile of daily maximum temperature during May–June over the baseline period 1998–2010:

$$T_g^* = \mathrm{P}_{95}\left(T_{\max,g,d}^{\text{May–June, 1998–2010}}\right)$$

This ensures consistency across heat metric definitions.

A.3 Construction of the Cumulative Excess Heat Metric

For each day d in region g, the positive exceedance above the threshold was computed as:

$$E_{g,d} = \max\left(T_{\max,g,d} - T_g^*, 0\right)$$

This function retains only temperature values exceeding the extreme heat threshold and sets all other observations to zero.

Annual cumulative excess heat over the MJJAS season was then defined as:

$$\text{ExcessSum}_{g,t} = \sum_{d \in \text{MJJAS}} E_{g,d}$$

This variable captures the total magnitude of temperature exceedances during the summer season.

A.4 Standardization

To ensure comparability across regions with heterogeneous climatic baselines, the cumulative intensity metric was standardized within each region:

$$heat\_z^{(I)}_{g,t} = \frac{\text{ExcessSum}_{g,t} - \mu_g}{\sigma_g}$$

where μ_g and σ_g denote the time-series mean and standard deviation of cumulative excess heat in region g over the sample period.

Under this definition, a one-unit increase in $heat\_z^{(I)}$ corresponds to a one standard deviation increase in cumulative excess heat relative to the region's historical distribution.

A.5 Econometric Specification

The alternative heat metric enters the same Poisson regression framework as the main specification:

$$\log \mathbb{E}[D_{g,a,t}] = \log (E_{g,a,t}) + f(a) + \alpha_g + \delta_t + \beta\, heat\_z^{(I)}_{g,t}$$

where:

- $D_{g,a,t}$ denotes deaths,
- $E_{g,a,t}$ denotes exposure,
- $f(a)$ is the age spline,
- α_g are region fixed effects,
- δ_t are year fixed effects.

Thus, differences between specifications arise solely from the construction of the heat exposure variable and not from changes in model structure.

A.6 Empirical Role

The cumulative intensity specification was explored during preliminary model development. However, the final empirical results reported in Chapter 6 are based on the persistence-based heatwave metric (HWdays) within the STABLE_TOT95 scope.

The intensity-based measure is therefore presented here for completeness and methodological transparency rather than as part of the main reported specification.

Data with Purpose.

Campus de Campolide, 1070-312 Lisboa, Portugal
novaims.unl.pt