



HAL
open science

Modelling age-space mortality dynamics in small areas

Jacob Martin, Carlo Giovanni Camarda

► **To cite this version:**

Jacob Martin, Carlo Giovanni Camarda. Modelling age-space mortality dynamics in small areas. 38th International Workshop on Statistical Modelling. Durham, UK 14-19 July, 2024, Durham University, pp.190-193, 2024, Proceedings of the 38th International Workshop on Statistical Modelling. Durham, UK 14-19 July, 2024. hal-04683681

HAL Id: hal-04683681

<https://hal.science/hal-04683681>

Submitted on 2 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modelling age-space mortality dynamics in small areas

Jacob Martin^{1,2}, Carlo Giovanni Camarda¹

¹ Institut national d'études démographiques, Aubervilliers, France

² Universidad del País Vasco, Bilbao, Spain

E-mail for correspondence: `carlo-giovanni.camarda@ined.fr`

Abstract: Large stochastic fluctuations due to small death counts limit understanding the true extent of geographic differences in mortality between small areas. We propose a model that borrows strength across age and space to produce reasonable estimates of the age-specific mortality schedule in small populations. We apply our model to mortality data for 50 Spanish provinces.

Keywords: Age-space interaction; Mortality modeling; Small area estimation; Smoothing.

1 Motivation

The quest to understand fine-grain mortality inequalities has spurred the creation of models tailored for small areas. The challenge lies in estimating age-specific death risks in small populations, given limited observed deaths leading to large fluctuations in mortality rates, primarily due to stochastic factors rather than substantial differences in mortality conditions. Despite this, genuine public health differences, such as exposure to environmental hazards, may exist. While existing models use prior knowledge, they grapple with issues like selecting a standard schedule, neglecting uncertainty in its choice, and not fully leveraging the spatial structure of the data (see, Gonzaga and Schmertmann, 2016; Alexander et al. 2017). In response, our proposed data-driven model for small areas avoids external standards, incorporating information across ages and space. This approach results in a properly specified stochastic model, featuring a nonparametrically estimated standard schedule based on the data and incorporating details about the total population.

To illustrate our results, we took age-specific death and population data for 50 provinces of Spain (excluding the Canary Islands) from the Instituto Nacional de Estadística (INE). We have deaths by single year of age up to age 110 and calculated population exposures by single year of age from the death and population data for each province. We apply our model to males in the year 2019.

This paper was presented at the 38th International Workshop on Statistical Modelling (IWSM), Durham University, UK, 14–19 July 2024. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

2 The age-space model

Data for our model consists of two $m \times n$ matrices, \mathbf{Y} (deaths) and \mathbf{E} (exposures), over age i and region j . To incorporate structure in the age-pattern, we include total population data as the first region (resulting in $n + 1$ regions, adding the sum of all regions as an additional row to \mathbf{Y} and \mathbf{E}). Spatial information is based on the centroids of each territorial unit. We assume that y_{ij} is Poisson distributed with expectation $\mu_{ij}e_{ij}$, where μ_{ij} is the force of mortality we aim to model in a log-scale.

We model each region as the sum of three components: a smooth standard age-pattern ($\boldsymbol{\eta}^0$), a spatially varying smooth age deviation ($\boldsymbol{\delta}_j$), and a scalar allowing for regional differences in mortality level that is not smoothed spatially (γ_j). To ensure convergence and allow interpretability of each component, the standard $\boldsymbol{\eta}^0$ is estimated from the mortality of the total population, and we add a ridge penalty to the γ_j . Vectorizing the linear predictor our model is expressed as:

$$\ln(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta} = \left[\mathbf{1}_{n+1} \otimes \mathbf{B}_a \left| \begin{array}{c} \mathbf{0}_{m, k_{as}+n} \\ \mathbf{B}_s \otimes \mathbf{B}_a \mid \mathbf{I}_n \otimes \mathbf{1}_m \end{array} \right. \right] \boldsymbol{\theta}, \quad (1)$$

where \mathbf{B}_a is a $m \times k_a$ B -spline basis over age modified in order to take into account the sharp descent of the level of mortality after the first year of life (Camarda, 2019). The matrix \mathbf{B}_s captures the spatial dimension, constructed as the row tensor product of bases for each individual spatial dimension. Let \mathbf{B}_{lon} be a $n \times k_{lon}$ B -spline basis over the longitude coordinates of the centroids, and \mathbf{B}_{lat} be a $n \times k_{lat}$ basis over the latitude coordinates. Then \mathbf{B}_s is the $n \times k_{lat}k_{lon}$ basis given by

$$\mathbf{B}_s = \mathbf{B}_{lat} \square \mathbf{B}_{lon} = (\mathbf{B}_{lat} \otimes \mathbf{1}_{k_{lon}}^T) \odot (\mathbf{1}_{k_{lat}}^T \otimes \mathbf{B}_{lon}). \quad (2)$$

Space interacts with ages by creating the $mn \times k_{as}$ design matrix $\mathbf{B}_s \otimes \mathbf{B}_a$. The supplementary lower-right corner of matrix \mathbf{X} corresponds to the regional mortality level γ_j (the identity matrix is present since we do not consider spatial structure here). The zeros in the top right of the matrix are present because the total population has no spatial or region-specific components.

The coefficients vector $\boldsymbol{\theta}$ can be seen as the combination of three sets of coefficients, denoted as $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}]^T$, each with lengths k_a , k_{as} , and n , respectively. As per (1), the mortality age-pattern for the total population serves as the standard for each region and it is concurrently estimated within the same framework: $\boldsymbol{\eta}^0 = \mathbf{B}_a \boldsymbol{\alpha}$. The coefficients $\boldsymbol{\beta}$ are then estimated to characterize age-space interaction, specifically, the age-regional specific deviation $\boldsymbol{\delta}_j$ that varies across space.

Following a P -splines approach, we ensure smoothness across age and space by introducing a discrete penalty \mathbf{P} to the likelihood associated with (1), that is difference penalties on the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and a ridge penalty on $\boldsymbol{\gamma}$.

We estimate the model using Iteratively Reweighted Least Squares within a GLAM framework (Currie et al., 2006). Smoothing parameters associated with the standard $\boldsymbol{\eta}^0$, each $\boldsymbol{\delta}_j$, the smooth variations of these deviations across space, and the magnitude of the ridge penalty are optimized by minimizing the Bayesian Information Criterion. Once the model is estimated, variance-covariance matrix of the parameters can be evaluated, allowing us to quantify uncertainty regarding each model component and the fitted age-regional-specific mortality patterns.

3 Application to Spanish provinces

Figure 1 shows fitted mortality rates and observed counts for Soria and Madrid provinces. We can observe that for the large population of Madrid where there are very few counts the model fits closely the data, while for the small province of Soria where there are zero deaths at almost every age before 30, the model is able to reconstruct a reasonable mortality schedule at all ages. The confidence intervals around the estimated rates do not contain most of the observed mortality rates, especially for the smaller province of Soria. However, in contrast to prediction intervals, we do not expect the observed rates to fall inside these confidence bands. Figure 2 shows the values for δ averaged over age, and the values for γ by province. We can observe that δ captures the spatial nature of mortality, while the γ allow for smaller deviations that avoid imposing too rigid a smooth structure.

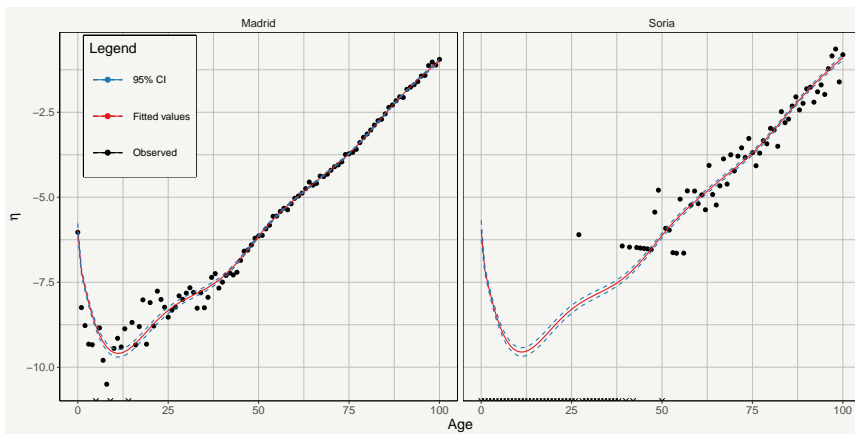


FIGURE 1. Observed log-death rates and fitted values for Madrid and Soria.

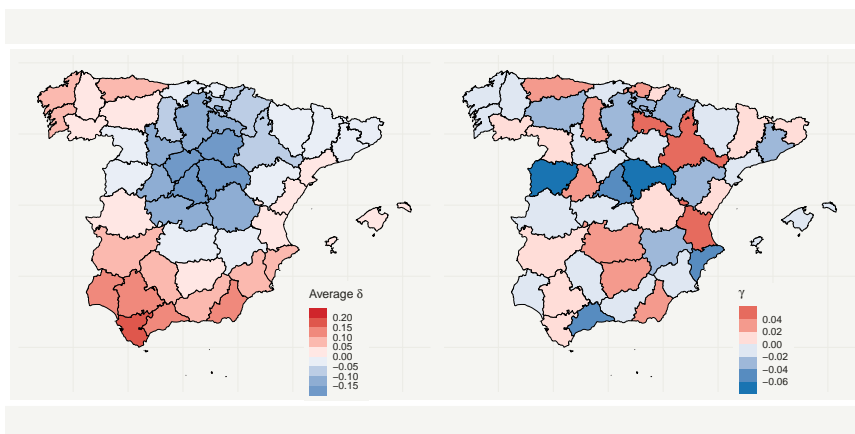


FIGURE 2. Average values of δ and values of γ by province.

4 Conclusion

We introduce a model that expresses regional mortality as the sum of a standard, age-specific deviations smooth in space, and unsmooth regional effects. Our model borrows strength across age and space, with the flexibility that allows for breaks from a perfectly smooth spatial pattern of mortality. We illustrate our model with an application to provincial data, but our model could be also used at the municipal or sub-municipal level.

Further improvements could include adding hierarchical structure to the model, as small areas are often embedded in larger territorial units, and modeling overdispersion. We also plan to embed our model in a Composite Link Model framework, since mortality data for small areas often comes in coarse age groups.

References

- Alexander, M., Zagheni, E. and Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography*, **54**, 2025–2041.
- Camarda, C.G. (2019). Smoothed constrained mortality forecasting. *Demographic Research*, **41**, 1091–1130.
- Currie, I.D., Durbán, M. and Eilers, P.H.C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of Royal Statistical Society. Series B.* **68**, 259–280.
- Lee, D. and Durbán, M. (2011). *P*-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, **11**, 49–69
- Schmertmann, C. and Gonzaga, M. (2018). Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*. **55**. 1363–1388.