DIFFERENCES IN EUROPEAN MORTALITY RATES: A GEOMETRIC APPROACH ON THE AGE–PERIOD PLANE

BY

MARCUS C. CHRISTIANSEN, EVGENY SPODAREV AND VERENA UNSELD

ABSTRACT

Age and period are the most widely used parameters for forecasting mortality rates. Empirical mortality rate differences in multiple populations often show strong geometric patterns on the two-dimensional age–period plane. The idea of this paper is to take these geometric patterns as the starting point for the development of forecasts. A parametric approach is presented and discussed which uses simple techniques from spatial statistics. The proposed model is statistically parsimonious and yields forecasts that are consistent with the historical data and coherent for multiple populations.

KEYWORDS

Mortality forecasting, modeling multiple populations, Box–Cox transform, geostatistics, Gaussian random fields, conditional simulation, Kriging.

1. INTRODUCTION

Reliable forecasts for mortality rates of multiple populations have many applications, for example, planning of healthcare provision, risk management of life insurance portfolios and quantification of future state pension liabilities. For some applications, in particular the diversification of insurance portfolios, the differences in the mortality experiences of multiple populations are relevant.

Empirical European mortality rates show both, converging and diverging trends, between different regions. With the European countries becoming more closely linked in all aspects of life, it seems likely that in the long term the mortality rates will rather converge than diverge. Recent research on mortality trend modeling therefore emphasizes the need for forecasts that are coherent for different populations, see e.g. Li and Lee (2005), Li and Hardy (2011), Cairns *et al.* (2011) and Börger *et al.* (2014). A widespread technique for achieving the coherence property is to model the total trend of a population as the sum of a common

Astin Bulletin 45(3), 477-502. doi: 10.1017/asb.2015.13

trend and a specific trend for each subpopulation, where the parametrization of the subpopulation trends is chosen in such a way that divergence is avoided.

We also follow that concept, assuming that the total mortality trend can be decomposed into a common (European) trend and subpopulation (country-specific) trends. The focus of this paper is solely on the modeling of the subpopulation trends, which equals the respective mortality experience minus the European average. In order to get complete forecasts, the subpopulation trend model proposed here has to be combined with one of the various models from the literature for the common (European) trend.

Various parametric models have been proposed in the literature for the modeling of mortality differences in multiple populations. Li and Lee (2005) expand the seminal (single population) Lee–Carter model to multiple populations by adding a second trend term of Lee–Carter type, where the original trend term is interpreted as a common trend and the second trend term describes the deviations of the subpopulations. Cairns *et al.* (2011) and Börger *et al.* (2014) follow a similar approach by extending not the Lee–Carter model but other one-population models with additional subpopulation trend terms. Jarner and Kryger (2011) and Ahcan *et al.* (2014) study the mortality of very small subpopulations, also using the idea to add subpopulation trend terms on the common log mortality rates. Hyndman *et al.* (2012) also propose a multi-population model, but they study the ratios of the local differences instead of the local differences themselves.

All references mentioned above model the subpopulation trends in the age–period log mortality data as a linear combination of one-dimensional trend effects. This means that all of these models can describe only a very restricted set of two-dimensional geometric structures on the age–period plane. Indeed, the empirical mortality differences in European data show various curved structures, which are not the typical outcomes of linear combinations of one-dimensional trend effects. Inspired by the empirical observations, we instead discuss the use of methods from spatial statistics that easily reproduce the curved structures.

As Börger and Aleksic (2014) point out, many mortality trend models lack consistency between historical data and forecasts. By using graphical analysis on the two-dimensional age–period plane, Börger and Aleksic (2014) show that many models produce forecasts that do not evolve in a natural way from the historical data but completely differ in their geometrical structures. There is no apparent reason why such structural breaks should occur. As our approach from spatial statistics takes the geometric structures in the historical data as the starting point for parametrization, we end up with forecasts that naturally merge with the historical data. This is not a proof for the correctness of the forecasts, but it is a desirable property which meets intuition.

The idea to use spatial statistics to forecast age–period mortality rates already appeared in Debón *et al.* (2010), yet with different motivation and objectives, which are focused on single populations. Motivated by the empirical observation that the residuals from Lee–Carter estimates still show a significant dependency structure, Debón *et al.* (2010) suggest to model the residuals by a

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

geostatistical approach. As a result, the total mortality trend equals the sum of the original Lee–Carter trend term and the geostatistical model term. In contrast, our objective is to study multiple populations and to model the mortality differences between them.

We consider mortality rates of selected Western European countries. The common mortality trend is filtered out by subtracting the mean mortality rates of all observed countries from the mortality experiences in each country. For the modeling of these differences for each country, we use simple semiparametric methods from geostatistics. The residuals show the characteristics of a transformed Gaussian random field, and dependencies between neighboring data on the age–period plane are described by its covariance function. Conditional simulation of Gaussian random fields is used to predict the country-specific mortality residuals.

This paper is structured as follows. Section 2 describes how the local mortality rate differences are formed and how they can be modeled by spatial stochastic methods. A Box–Cox data transform allows to see the residuals as transformed Gaussian random fields. Section 3 presents the mortality data used here, describes simulation methods and presents empirical results on the example of France (females) and the Netherlands (males). Our modeling approach is summarized in Section 4. The appendix shows forecasts for all 17 Western European countries under consideration.

2. A SPATIAL MODEL FOR DIFFERENCES IN AGE–PERIOD MORTALITY RATES

In this section, a parametric model is presented that describes historic and future differences in observed gender-specific mortality rates of Western European countries. Following the literature on coherent mortality forecasting (cf. Section 1), we see the mortality trend of each country as a composition of a common European trend and a country-specific part. The present paper solely focuses on modeling the country-specific parts.

In order to make the country-specific trends visible, we first eliminate the common European trend by subtracting the mean mortality rate of all (considered) countries from the mortality experiences of the single countries. In a second step, the historic country-specific trends — describable as two-dimensional surfaces on the age–period plane (Lexis plane) — are approximated by two-dimensional parametric functions. The residuals of the fitted functions are transformed, so that they represent a realization of a Gaussian random field. By extrapolating the parametric non-random country-specific trend, using conditional Gaussian simulation of the trend residuals and applying the inverse of all transforms, we obtain forecasts for the country-specific mortality trends.

We consider annual mortality rates $q_{x,t}$, which are defined as the probability that an individual, who has an age of x (in years) on January 1 of year t, dies within this year. Furthermore, gender-, age- and year-specific population numbers are used. The population size $b_{x,t}$ is defined as the number of individuals

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

who have an age of x (in years) on January 1 of year t. The annual mortality rates and population numbers will be considered on the range

$$R := \{x_1, x_2, \ldots, x_n\} \times \{t_1, t_2, \ldots, t_m\} \subseteq \mathbb{N}_0 \times \mathbb{N},$$

where $n \in \mathbb{N}$ is the number of observed ages $x_1 < x_2 < \cdots < x_n$ and $m \in \mathbb{N}$ is the number of observed years $t_1 < t_2 < \cdots < t_m$.

2.1. Modeling of the trends

We approximate the common European mortality trend by the average of the mortality trends of all (considered) countries. Since the population sizes of the European countries vary significantly, we use the population-weighted average of the mortality rates. Using superscripts to identify different countries, the country-specific mortality rate difference of an arbitrary, but fixed country $k \in L$ is given by

$$f^k(x,t) := q^k_{x,t} - \frac{\sum_{l\in L} b^l_{x,t} \cdot q^l_{x,t}}{\sum_{l\in L} b^l_{x,t}} = q^k_{x,t} - \frac{\sum_{l\in L} d^l_{x,t}}{\sum_{l\in L} b^l_{x,t}}, \qquad (x,t) \in R, \tag{1}$$

where L denotes the set of all countries that are considered and $d^l_{x,t}$ denotes the number of deaths in country l at age x in year t. Since the mortality rates $q^k_{x,t}$ lie in the interval $[0, 1]$, the differences in mortality rates $f^k(x,t)$ are in the range $[-1, 1]$. In order to simplify the notation, in the following we will skip the superscript k, i.e. $f(x,t) = f^k(x,t)$.

In a next step, f will be approximated by a function g that captures the geometric patterns in the country-specific trend. Yet, before doing that we normalize the variance with respect to age x. This is important for the steps following afterwards. In general, it can be assumed that the mortality differences $f(x,t)$ are significantly heteroscedastic with respect to age x. Without a variance normalization, the parametric approach would approximate the geometric patterns very poorly in age ranges with low variances. As a result, in the step following afterwards, where we do another variance normalization and approximate the residuals by a Gaussian random field, the previously poorly approximated geometric patterns would become visible, significantly violating the Gaussian random field approach. We can avoid those problems by normalizing the variance in each age as follows:

$$f_{\text{var}}(x,t) := \frac{f(x,t)}{\sigma_f(x)}, \qquad (x,t) \in R, \tag{2}$$

where

$$\sigma^2_f(x) = \frac{1}{m-1} \sum_{i=1}^{m} \left(f(x,t_i) - \frac{1}{m} \sum_{j=1}^{m} f(x,t_j) \right)^2 .$$

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

This transformation can be easily continued into the future since it is constant with respect to the time variable.

Now, we approximate $f_{\text{var}}(x, t)$ by a function

$$g : \mathbb{N}_0 \times \mathbb{N} \to \mathbb{R} : \ (x, t) \mapsto g(x, t)$$

that matches the geometric patterns that we see in the historic data. In the geostatistic literature, there is a number of ways to estimate the space–time trend non- or (semi)parametrically, see e.g. geoadditive regression, universal kriging or intrinsic models of order k in Fahrmeir *et al.* (2013), Chilès and Delfiner (1999), Wackernagel (1995) and Cressie (1993). In all its generality, this is mainly a task of functional regression analysis. In our case, we have chosen a parametric approach driven by the concrete application. The choice is motivated by simplicity, a small number of control parameters, and a particular form of trend surfaces under consideration with mainly elliptic or inverse power contours, cf. Figures 2 and 3. Thus, we suggest to use a function

$$g(x, t) := c_1 \cdot e^{-\frac{Q_1(x,t)}{2}} + c_2 \cdot e^{-\frac{Q_2(x,t)}{2}} + c_3 \cdot e^{-\frac{Q_3(x,t)}{2}} + c_4 \left(\frac{1}{|x - b|} \right)^p + d, \quad (3)$$

where

$$Q_i(x, t) = \begin{pmatrix} t - \mu_{2i-1} \\ x - \mu_{2i} \end{pmatrix}^{\mathrm{T}} \Sigma_i \begin{pmatrix} t - \mu_{2i-1} \\ x - \mu_{2i} \end{pmatrix},$$

$$\Sigma_i = \begin{pmatrix} \cos(\alpha_i) & -\sin(\alpha_i) \\ \sin(\alpha_i) & \cos(\alpha_i) \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \lambda_{2i-1} & 0 \\ 0 & \lambda_{2i} \end{pmatrix} \begin{pmatrix} \cos(\alpha_i) & -\sin(\alpha_i) \\ \sin(\alpha_i) & \cos(\alpha_i) \end{pmatrix}$$

for $i = 1, 2, 3$. The first three terms in (3) describe elliptical patterns on the age–period plane. The fourth term is designed to match pure age effects. We assume that the parameters satisfy

$$\begin{array}{lll} \alpha_i & \in [0, \pi], & i = 1, 2, 3 \\ b & \in \mathbb{R} \setminus \mathbb{N}, & \\ c_i & \in \mathbb{R}, & i = 1, 2, 3, 4 \\ d & \in \mathbb{R}, & \\ \lambda_i & \in \mathbb{R}^+, & i = 1, 2, \ldots, 6 \\ \mu_{2i-1} & \in (-\infty, t_m + \tau], & i = 1, 2, 3 \\ \mu_{2i} & \in (-\infty, \infty), & i = 1, 2, 3 \\ p & \in \mathbb{R}. & \end{array}$$

The restriction of the α_i's to $[0, \pi]$ is necessary for identifiability of the model. By assuming that b is not a natural number, we avoid division by zero. The assumption that the μ_{2i-1} are not greater than $t_m + \tau$ avoids that the exponentials in g have their maxima far in the future, which would result in unrealistic forecasts. For estimating the parameters, we use the method of least squares.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

Our empirical analysis following later on will show that our choice of g leads to very reasonable residuals. From a theoretical point of view, the suggested model has several advantages.

First, the parametric functions in (3) allow to capture a wide variety of two-dimensional patterns in the data. As we already pointed out in Section 1, other multi-population models in the literature only allow for a very restricted set of geometric patterns. However, apart from the empirical motivation, so far we are not aware of any biological or medical justifications for the underlying geometric patterns.

Second, after fitting g to historical data, forecasts can be easily obtained by simply extrapolating g. Our empirical examples will show that these forecasts carry forward the historic geometric patterns in a reasonable way. As Börger and Aleksic (2014) pointed out, other concepts in the literature often lead to structural breaks between historical data and forecasts.

Third, for t towards infinity the function (3) converges to the sum of the fourth and fifth terms, which are independent of t. So when we construct future country-specific mortality trends by extrapolating the function g, we obtain forecasts that are coherent, i.e. non-divergent.

Fourth, with just 22 parameters needed to describe g, our statistical model is still parsimonious compared to alternative models in the literature. For example, the Lee–Carter extension of Li and Lee (2005) needs $n + m$ parameters, which means $81 + 50 = 131$ parameters in our empirical example following later on.

2.2. Modeling of the residuals

In order to quantify the uncertainty of forecasts, we need a model for the residuals

$$r(x, t) := f_{\text{var}}(x, t) - g(x, t). \tag{4}$$

Typically, one will find that the residuals are neither stationary nor isotropic, and thus, they need to be transformed. As our empirical analysis will confirm later on, after some age normalization the residuals can be pretty well described as translated powers of normal distributed random variables. The class of translated powers of normal random variables can be interpreted as an expansion of the family of normal distributions to a distribution class with four parameters.

More specifically, by assuming that $z = \{z(u) \mid u = (x, t) \in \mathbb{R}^2\}$ is a realization of a stationary Gaussian random field $Z = \{Z(u) \mid u = (x, t) \in \mathbb{R}^2\}$ with mean 0 and variance 1, we describe the residuals r by the parametric family

$$r(x, t) = \begin{cases} \sigma_r(x)\left(k + ((\mu + z(x, t)\sigma)\lambda + 1)^{\frac{1}{\lambda}}\right), & \lambda \neq 0 \\ \sigma_r(x)\left(k + e^{\mu + z(x,t)\sigma}\right), & \lambda = 0 \end{cases}, \qquad (x, t) \in R. \tag{5}$$

The factor $\sigma_r(x)$ is the age normalization. The remaining factor is a translated

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

power (or exponential) of a normal random variable. The function in (5) can be seen as mapping from $z(x,t)$ to $r(x,t)$, and by inverting this mapping we get the following equivalent notion: The residuals r belong to the parametric family (5) if and only if the inverse transform performed on the residuals leads to a stationary and standardized Gaussian random field z. In the following, we will use the latter notion, because it is the more natural and constructive way when discussing the empirical fitting of the parameters.

In our empirical data, the variance of the residuals r depends significantly on the age but marginally on the year. So we normalize the average variance per age of the residuals by dividing them by the parameter $\sigma_r(x)$ given by

$$\sigma_r^2(x) = \frac{1}{m-1}\sum_{i=1}^{m}\left(r(x,t_i) - \frac{1}{m}\sum_{j=1}^{m} r(x,t_j)\right)^2. \tag{6}$$

After the normalization of the residuals, one will find that the empirical data do not yet have the characteristics of a Gaussian random field. Therefore, we apply the transformation introduced by Box and Cox (1964). The Box–Cox transformation selects from a family of transformations, the one which is best to make the observations approximately normally distributed. In addition, the Box–Cox transformation leads to a stabilization of the variance. However, the Box–Cox transformation only allows strictly positive observations, so first the data are translated into the positive range by subtracting a parameter

$$k < \min\left\{\frac{r(x,t)}{\sigma_r(x)} \;\middle|\; (x,t) \in R\right\}. \tag{7}$$

The Box–Cox transformation is then applied to the strictly positive data:

$$y^{(\lambda)}(x,t) = \begin{cases} \frac{\left(\frac{r(x,t)}{\sigma_r(x)} - k\right)^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \ln\left(\frac{r(x,t)}{\sigma_r(x)} - k\right), & \lambda = 0. \end{cases} \tag{8}$$

The corresponding transformation parameter $\lambda \in \mathbb{R}$ is formed by maximizing the likelihood function of the observations (cf. routine `boxcox` in Matlab software). The latter is derived from the posterior distribution assuming that it is Gaussian, the expected value has a linear structure and the variance is constant.

In a last step, the data are standardized such that the empirical mean of the transformed data is 0 and their empirical variance is 1. The standardization

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

parameters are

$$\mu = \begin{cases} \frac{1}{nm} \sum\limits_{(x,t)\in R} \frac{\left(\frac{r(x,t)}{\sigma_r(x)} - k\right)^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \frac{1}{nm} \sum\limits_{(x,t)\in R} \ln\left(\frac{r(x,t)}{\sigma_r(x)} - k\right), & \lambda = 0, \end{cases}$$

$$\sigma^2 = \begin{cases} \frac{1}{nm-1} \sum\limits_{(x,t)\in R} \left(\frac{\left(\frac{r(x,t)}{\sigma_r(x)} - k\right)^{\lambda} - 1}{\lambda} - \mu\right)^2, & \lambda \neq 0, \\ \frac{1}{nm-1} \sum\limits_{(x,t)\in R} \left(\ln\left(\frac{r(x,t)}{\sigma_r(x)} - k\right) - \mu\right)^2, & \lambda = 0. \end{cases} \tag{9}$$

After all these steps, the observations should be a realization of the stationary Gaussian random field Z.

We assume that the covariance of the observations depends only on the distance between two points (second-order isotropy of Z). So the covariance function

$$\mathrm{Cov}(Z(u), Z(v)) = \mathbb{E}(Z(u) \cdot Z(v)), \quad u, v \in R$$

of Z can be estimated by the empirical covariance of the observations (see e.g. Spodarev (2013, p. 324)):

$$\hat{C}(h) := \frac{1}{|N(h)|} \sum_{(u,v)\in N(h)} z(u) \cdot z(v) \tag{10}$$

with

$$N(h) = \left\{(u, v) \in R^2 : \|u - v\| = h\right\},$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^2$.

In order to use the information of the empirical covariance function for the subsequent simulation, a valid covariance model has to be chosen from a list of available models (see e.g. Wackernagel (1995, pp. 334–336), Spodarev *et al.* (2015, pp. 328–331)). Then it has to be fitted to the empirical covariance using the method of least squares. Here, we use a stable covariance model with nugget effect

$$C(h) := b_1 \cdot e^{-ah^{\nu}} + b_2 \cdot \mathbb{1}_{\{h=0\}} \tag{11}$$

with $h \geq 0$, $a, b_1, b_2 > 0$, $\nu \in (0, 2]$. Its choice is justified by the form by its graph which best fitted our data, cf. Figure 7.

Table 1 gives an overview of all modeling steps.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

TABLE 1

OVERVIEW OF THE MODELING STEPS.

	Quantity	Formula
Trend in Mortality Rate Differences		
Mortality Data	$q^k_{x,t}$	
Subtraction of Common European Trend	$f(x,t)$	See (1)
Variance Normalization with Respect to Age	$f_{\text{var}}(x,t)$	See (2)
Parametric Modeling	$g(x,t)$	See (3)
Residuals of Mortality Rate Differences		
Residuals of the Parametric Model	$r(x,t)$	See (4)
Variance Normalization with Respect to Age		See (6)
Shift to Positive Real Line		See (7)
Box–Cox Transformation		See (8)
Standardization to Mean 0 and Variance 1		See (9)
Parametric Modeling of Covariance	$C(h)$	See (11)

3. EMPIRICAL ANALYSIS

Here, we apply the concepts of Section 2 to real data. Although we perform the analysis for a group of 17 countries, we will focus our presentation on just two exemplary populations: males in the Netherlands, for which the mortality rates are well reproduced by the model, and females in France, for which the model is less suitable. Further empirical results can be found in Appendix.

3.1. Description of data

We use mortality data from the Human Mortality Database (HMD), University of California and Berkeley (USA). The gender-, age- and year-specific annual mortality rates that we use are available for age and time intervals of one year. The same is true for the population size data. Note that for high ages, the published mortality rates are partially smoothed by linear interpolation.

For our empirical analysis, we select the following 17 Western European countries:

Austria	Belgium	Denmark	Finland	France
Germany (West)	Iceland	Ireland	Italy	Luxembourg
The Netherlands	Norway	Portugal	Spain	Sweden
Switzerland	UK			

We choose this group of countries because they have been closely linked for many decades. The maximum period of time for which the data is available for all 17 countries ranges from 1960 to 2009. The available age span ranges from 0 to 110. So our age–period observation area is

$$R = \{0, 1, \ldots, 110\} \times \{1960, 1961, \ldots, 2009\}.$$

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

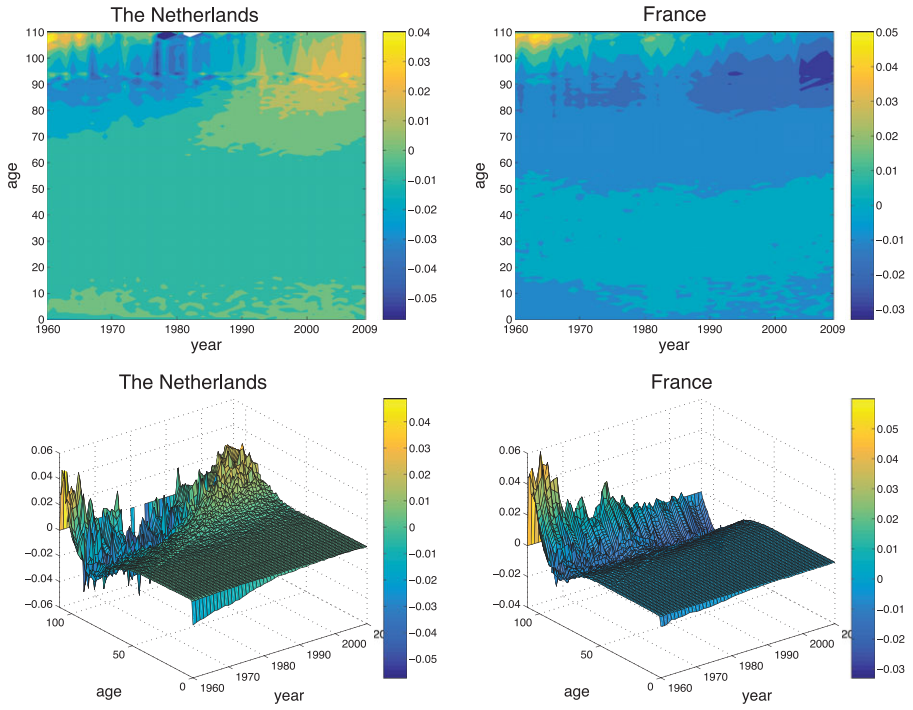

FIGURE 1: Differences in mortality rates f for the Netherlands (male) and for France (female) in various graphics. (Color online)

Spain (1975) and Italy (1981) had territorial changes during the observation period. So there are years for which there exist two data sets, referring to the population before and after territorial change. We use the arithmetic mean of the population before and after the territorial change.

Figure 1 shows mortality rate differences according to Formula (1) for males in the Netherlands and females in France. We can see in important characteristics of the HMD data that they have been smoothed from certain age levels onward, where the latter differ from country to country (see Wilmoth *et al.* (2007, p. 35 ff.)). Since this smoothing interferes significantly with our model, we disregard in our analysis all ages beyond 85. Furthermore, we see significant children mortality differences in the 1960s and 1970s, which disappear from the 1990s onward. In order to avoid modeling of these effects, which seem to be irrelevant for future forecasts, we exclude all ages below five from our analysis.

3.2. Fitting the trend model

Considering the data shown in Figure 1, non-random island-like structures can be recognized, which we denote as spots. In Figure 1, we see spots only for high ages. After normalizing the variances in each age, we see spots also for low ages,

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

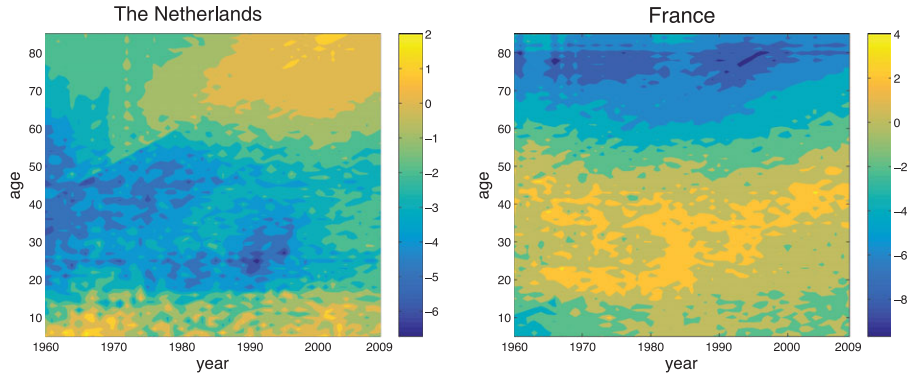

FIGURE 2: Differences after variance normalization f_{var} for the Netherlands (male) and for France (female). (Color online)

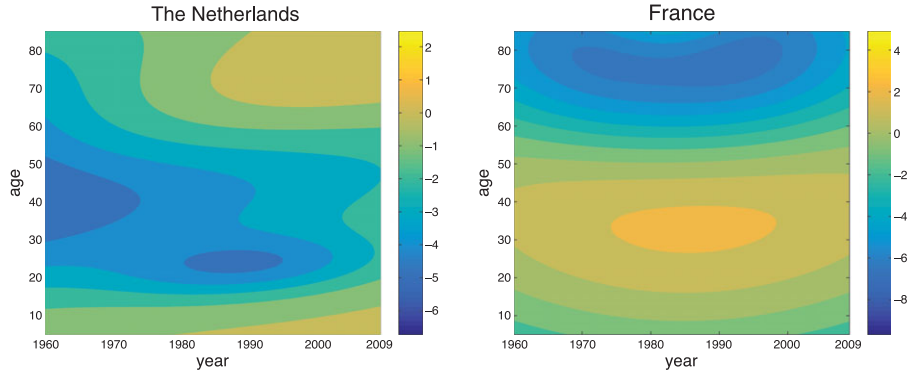

FIGURE 3: Adapted approximation function g on the area R for the Netherlands (male) and for France (female). (Color online)

see Figure 2. Looking at the geometrical structures in Figure 2, functions with elliptical-level sets (first three terms in (3)) seem to be good candidates for fitting a model. We found that for most countries a sum of three such functions is sufficient for modeling the non-random trend structures. The fourth term $c_4(\frac{1}{|x-b|})^p$ in (3) approximates the high gradients in the direction of age, which cannot be represented by elliptical-level sets. Figure 2 also suggests that the maxima of the exponential functions may be outside the observed range. We restrict the location of the maxima in the future by $t_m + \tau = 2014$; otherwise, we might obtain extreme and implausible forecasts.

An enumeration of the exact limits for each parameter is given in Section 2.1. The parameters of the approximation function are determined by using the method of least squares. The adapted approximation function g for the Netherlands and France is illustrated in Figure 3 with parameter values given in Table 2.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

TABLE 2

PARAMETERS OF THE ADAPTED APPROXIMATION FUNCTION g FOR THE NETHERLANDS (MALE) AND FOR FRANCE (FEMALE) (ROUNDED VALUES).

	α_1	α_2	α_3	b	c_1	c_2	c_3	c_4
The Netherlands	1.26	1.26	1.44	−10.98	262.02	−265.61	−3.27	7.91
France	0.07	4E-14	3E-04	−177.65	9.36	−12.19	13.49	−1E+06
	d	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	μ_1
The Netherlands	0.26	0.004	5E-04	0.004	5E-04	0.03	0.005	1963.35
France	−1.93	5E-04	3E-04	0.002	0.003	0.02	0.003	1986.12
	μ_2	μ_3	μ_4	μ_5	μ_6	p		
The Netherlands	55.55	1962.43	55.16	1990.59	22.96	8.68		
France	49.88	1984.05	79.28	1982.42	111.11	2.33		

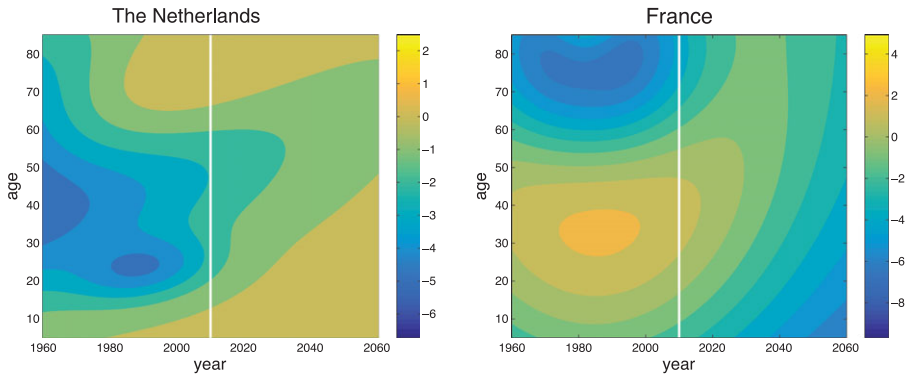

FIGURE 4: Forecast of the trend approximation function g on the domain T for the Netherlands (male) and for France (female). (Color online)

Apparently, the approximation of the observed values by the parametric model (3) works very well. In order to forecast future non-random trends, we simply extrapolate the function g in (3) after the fitting. Figure 4 shows the fitted function g on the extended domain $T := \{1960, 1961, \ldots, 2060\} \times \{5, \ldots, 85\}$. The white lines in the figure mark the boundary between the empirical fitting and the future extrapolation. Figure 4 exemplarily demonstrates a major advantage of the geometric approach: by construction, the historic observations are continued into the future in a smooth way and no structural breaks appear between historic observations and the forecast. By forecasting till 2060, we do not comment on the reliability of the forecast, we just choose this long-time horizon in order to demonstrate the method.

The forecast for males in the Netherlands seems quite reasonable. The forecast for females in France makes sense in the short range, but it is less convincing

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

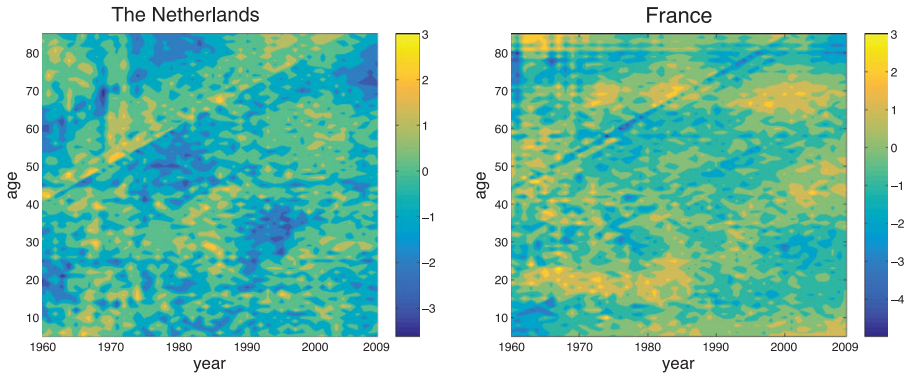

FIGURE 5: Trend residual values after the normalization of the variance for the Netherlands (male) and for France (female). (Color online)

in the long range as there appear strongly negative values for children around 2060. The reason for these negative values is the term $c_4(\frac{1}{|x-b|})^p$. Better results could be achieved by modifying (setting $c_4 = 0$) the parametric function (3), which we choose as a compromise between all 17 countries.

Figure 5 shows the corresponding residuals after a normalization of the variance in each age (see Section 2.2). We can clearly see cohort effects in the form of narrow diagonal lines. However, as it can be seen in the example of France and the Netherlands, the most significant diagonal lines do not affect the forecast area $T \setminus R$. For this reason, we do not incorporate cohort effects in our parametric model.

We still see spots in Figure 5, but they seem to be randomly distributed. This leads to the conclusion that the residuals are stationary. Moreover, the observed values are close to a normal distribution but a slight skewness is present. The skewness vanishes after the Box–Cox transformation.

Recall that the data have to be shifted to the positive real line before the Box–Cox transformation, and that after the Box–Cox transformation, we do another linear transformation such that the empirical mean is 0 and the empirical variance is 1. Figure 6 shows that the univariate distribution of the transformed residuals is indeed well approximated by a normal law.

3.3. Simulating the residuals

We already mentioned in the previous section that the residuals in Figure 5 show randomly distributed spots, which implies that there exist dependencies between data points that are close together. Since these spots do not show clear directions, we may assume that the correlation between two data points depends only on their distance. Due to the transformations, the observed values are approximately normally distributed with expectation 0 and variance 1. Thus, the data set may be seen as a realization of the stationary and isotropic Gaussian random field $Z = \{Z(x) \mid x \in R\}$ with mean 0, variance 1 and covariance

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

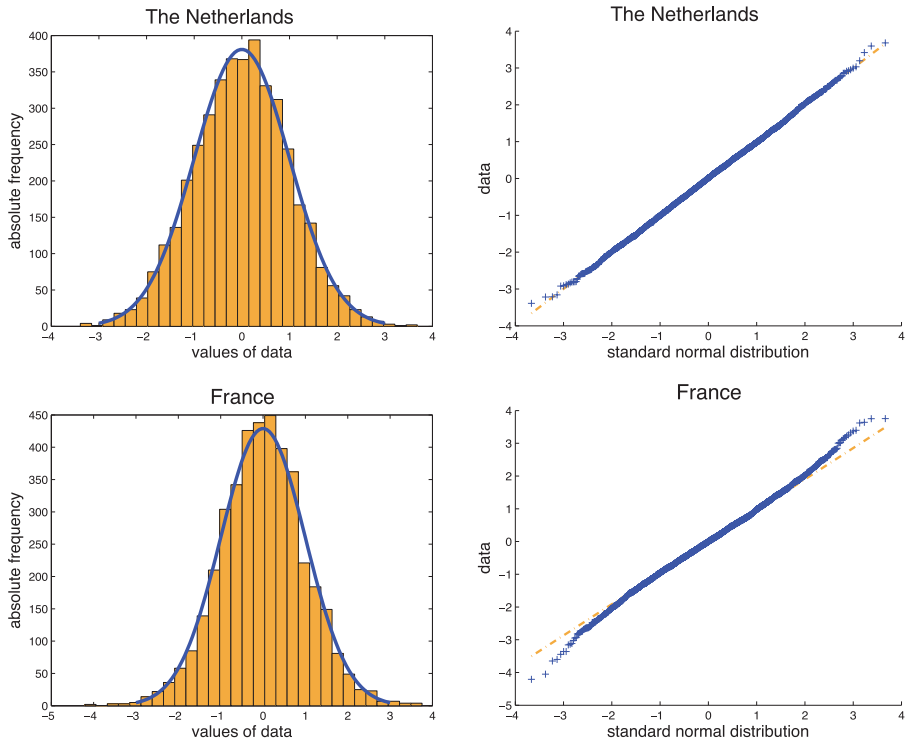

FIGURE 6: Representation of the Box–Cox-transformed trend residual data in a histogram with adjusted density function of the normal distribution (left) and a *QQ*-Plot (right) for the Netherlands (male) and for France (female). (Color online)

function C. The covariance function C is chosen by least-squares fitting of a covariance model to the empirical covariance function $\hat{C}$ (see formula (10)). As covariance model, we use a stable covariance model with nugget effect, because it provides the smallest deviations from $\hat{C}$, compare Figure 7. For large distance lags h, there exist just a small number of data pairs. Therefore, the empirical covariance function oscillates strongly for large values of h. Because of that, Figure 7 shows only distances up to $h = 50$.

In order to simulate future trend residual values z, a conditional simulation of the Gaussian random field Z, which is completely characterized by its expectation value 0 and its covariance function C, is performed. In the conditional simulation, values of z within domain T are simulated provided that they coincide with the observed values of Z on R. For this purpose, first an unconditional simulation is performed which is converted into a conditional simulation by using *simple Kriging* (see Lantuejoul (2002, p. 199 ff.)). The simple Kriging method ensures that the simulated values on R are equal to the observed values and converge to the values of the unconditional simulation on $T \setminus R$ with increasing

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

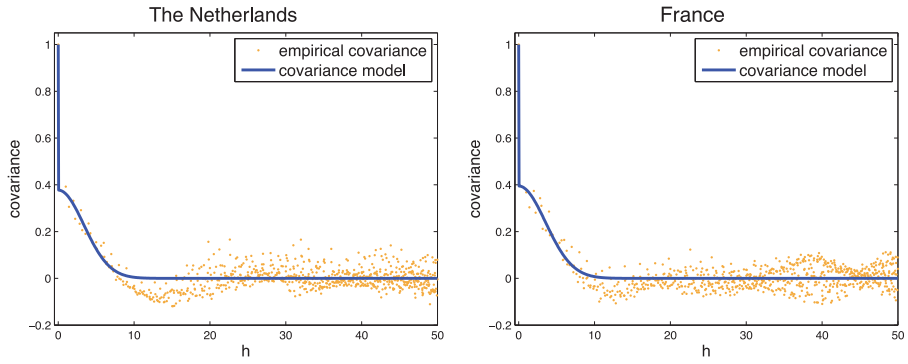

FIGURE 7: Fitting the covariance model to the empirical covariance of the residual trend data for the Netherlands (male) and for France (female). (Color online)

distance to R. The simple Kriging estimator Z^* of Z is the linear combination

$$Z^*(v) = \sum_{u \in R} \lambda_u(v) \cdot Z(u), \qquad v \in T,$$

which minimizes the mean square error

$$\mathbb{E}(Z^*(v) - Z(v))^2, \qquad v \in T.$$

It can be shown that the coefficients $\lambda_u(v)$ are the solutions of the following linear equation system:

$$\begin{pmatrix} C(u_1,u_1) & C(u_1,u_2) & \cdots & C(u_1,u_{|R|}) \\ C(u_2,u_1) & C(u_2,u_2) & \cdots & C(u_2,u_{|R|}) \\ \vdots & \vdots & \ddots & \vdots \\ C(u_{|R|},u_1) & C(u_{|R|},u_2) & \cdots & C(u_{|R|},u_{|R|}) \end{pmatrix} \cdot \begin{pmatrix} \lambda_{u_1}(v) \\ \lambda_{u_2}(v) \\ \vdots \\ \lambda_{u_{|R|}}(v) \end{pmatrix} = \begin{pmatrix} C(u_1,v) \\ C(u_2,v) \\ \vdots \\ C(u_{|R|},v) \end{pmatrix}, \quad (12)$$

where $v \in T$ and $C(u,v) = C(\|u - v\|)$ with $u, v \in T$. This results in the estimator z^* of the observed values z:

$$z^*(v) = \sum_{u \in R} \lambda_u(v) \cdot z(u), \qquad v \in T.$$

Now, an unconditional simulation r of the random field Z is performed on T. This results in another estimator r^* of r:

$$r^*(v) = \sum_{u \in R} \lambda_u(v) \cdot r(u), \qquad v \in T.$$

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

The coefficients $\lambda_u(v)$ are a solution of the equation system (12). Simple Kriging is exact, i.e.

$$z^*(u') = \sum_{u \in R} \lambda_u(u') \cdot z(u) = z(u'), \qquad u' \in R,$$

$$r^*(u') = \sum_{u \in R} \lambda_u(u') \cdot r(u) = r(u'), \qquad u' \in R.$$

The covariance function C for two widely separated points is approximately 0. Hence, for each $v \in T$ that is far enough away from the region R, the solution of the Equation System (12) is approximately equal to the zero vector. That means that the values of z^* and r^* are approximately zero for all points $v \in T$ that are far away from R. By Lantuejoul (2002, Theorem 15.3.1), it holds that $Z^* := \{Z^*(v) \mid v \in T\}$ and $Z - Z^* := \{Z(v) - Z^*(v) \mid v \in T\}$ are two independent Gaussian random fields, and

$$Z(v) = Z^*(v) + Z(v) - Z^*(v), \qquad v \in T.$$

The values $z^{\mathrm{CS}}(v) := z^*(v) + r(v) - r^*(v)$ for all $v \in T$ obviously belong to a realization of a Gaussian random field. Moreover, z^{CS} is the result of a conditional simulation of the random field Z, since for all $u \in R$ it holds

$$\begin{aligned} z^{\mathrm{CS}}(u) &= z^*(u) + r(u) - r^*(u) \\ &= z(u) + r(u) - r(u) \\ &= z(u). \end{aligned}$$

In addition, the values z^{CS} tend to the values r of the unconditional simulation of Z for all points $v \in T$ that are far away from R:

$$\begin{aligned} z^{\mathrm{CS}}(v) &= z^*(v) + r(v) - r^*(v) \\ &\approx 0 + r(v) - 0 \\ &= r(v). \end{aligned}$$

To carry out the unconditional simulation of Z, we use *circulant Embedding* from Dietrich and Newsam (1997). Circulant Embedding is a fast and exact method to simulate stationary Gaussian random fields on an equally spaced grid. Its basic idea is to decompose the covariance matrix K into the product $K = A \cdot A^{\mathrm{T}}$. If $\epsilon \sim N(0, I)$ is a standard normal vector, the random vector $y = A \cdot \epsilon$ has an expected value of 0 and the covariance matrix K. If random numbers are generated in this way, these have exactly the desired correlation structure. The factorization of K is very expensive, so the matrix K is embedded into a circulant matrix which can be factorized efficiently with the Fast Fourier Transform (FFT).

For each of the observed populations, we performed 1,000 conditional simulations with Matlab R2014a. For one population, the CPU time of 1,000 simulations on an Intel Xeon X5570 2.93 GHz processor is around 75 hours.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

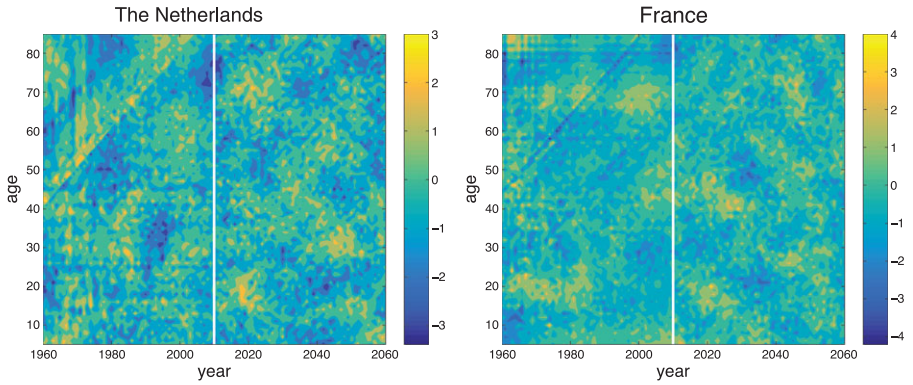

FIGURE 8: Exemplary conditional simulations of residuals for the Netherlands (male) and for France (female). (Color online)

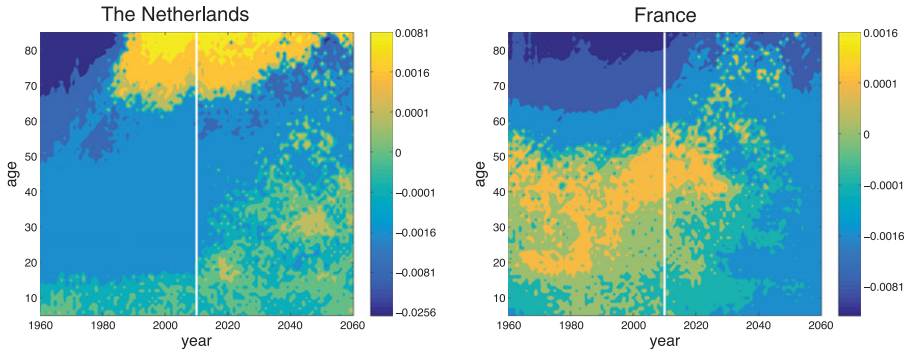

FIGURE 9: Predicted differences in mortality rates from exemplary simulations for the Netherlands (male) and for France (female) (scaled with $z \mapsto \operatorname{sgn}(z) \cdot |z|^{\frac{1}{4}}$). (Color online)

The results of exemplary conditional simulations on the grid T can be seen in Figure 8, for examples, the Netherlands and France. The expected value and the correlation structure of the values in the forecast period are the same as in the observation period. Due to the use of simple Kriging, no obvious break between the values of the past and the values of the future can be seen.

3.4. Forecast results

If all previously performed transformations are reversed for the simulated values on T, we obtain the predicted mortality rate differences. The forecast results based on exemplary simulations are shown in Figure 9. The distinctive structures in the past are continued in a realistic way into the future. If we had not drawn the white line in Figure 9, it would be almost impossible to locate the change point between observed and forecasted values.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

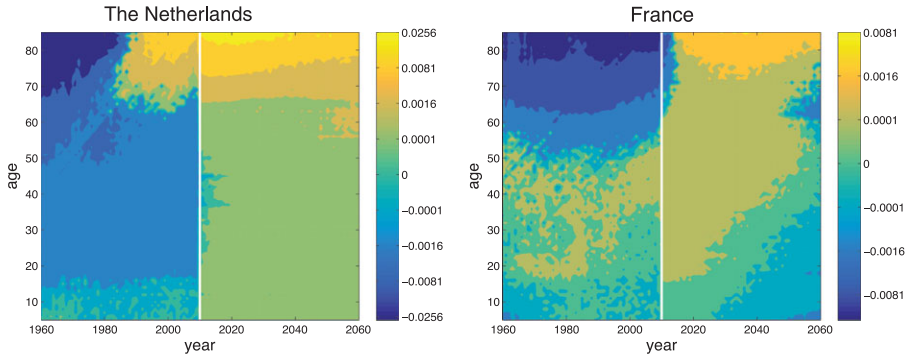

FIGURE 10: Maximal predicted differences in mortality rates for the Netherlands (male) and for France (female) (scaled with $z \mapsto \operatorname{sgn}(z) \cdot |z|^{\frac{1}{4}}$). (Color online)

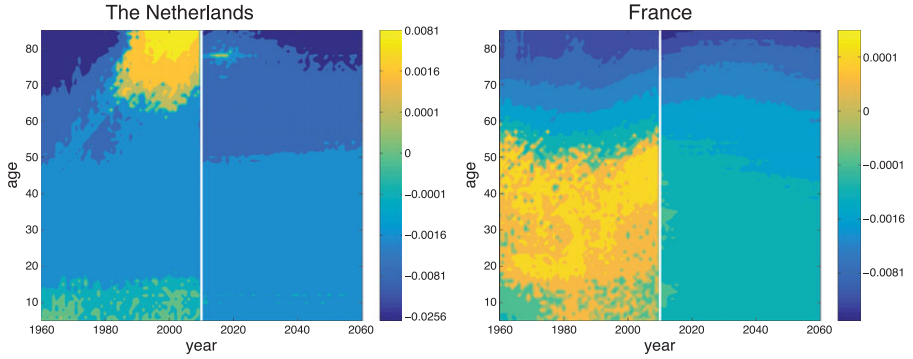

FIGURE 11: Minimal predicted differences in mortality rates for the Netherlands (male) and for France (female) (scaled with $z \mapsto \operatorname{sgn}(z) \cdot |z|^{\frac{1}{4}}$). (Color online)

Additionally, we have performed 1,000 simulations for males in the Netherlands and females in France. The maxima and minima of these simulations are shown in Figures 10 and 11, respectively. Figure 12 shows the predicted mortality rate differences based on exemplary simulations for the life cycle of a single cohort, namely individuals with age 30 in 2010. The upper and lower confidence bounds in Figure 12 are the maximal and minimal values of 1,000 simulations.

The inversion of the Box–Cox transformation can potentially change the correlation structure of the underlying random field. In order to see if that happened here, we cannot simply study the empirical covariance function according to (10) since the historical and the predicted mortality rate differences are neither stationary nor isotropic. Instead, we calculated the empirical covariance function (10) for the historical data before performing the Box–Cox transformation and for the data based on exemplary simulations after reversing the Box–Cox transformation. Figure 13 shows the result for the Netherlands (male) and

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

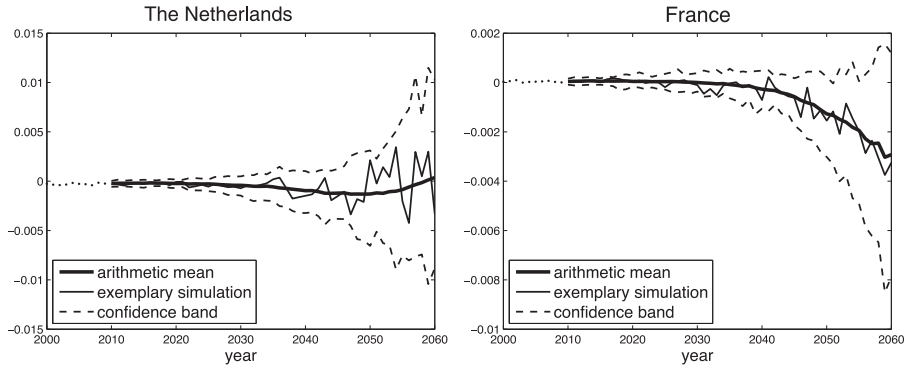

FIGURE 12: Predicted mortality rate differences from exemplary simulations for the life cycle of the cohort that has age 30 in 2010 for the Netherlands (male) and for France (female) with arithmetic mean and confidence bands.

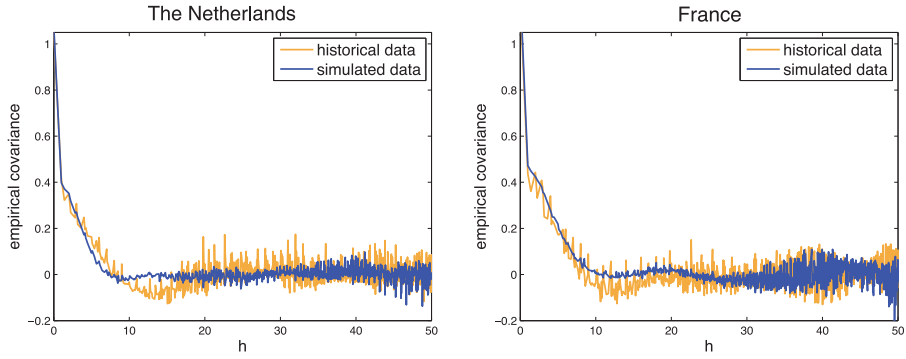

FIGURE 13: Empirical covariance functions of the historical data prior to the Box–Cox transformation and of the data based on exemplary simulations after reversing the Box–Cox transformation for the Netherlands (male) and for France (female). (Color online)

France (female). We see that the Box–Cox transformation has not significantly changed the correlation structure of the data here.

4. SUMMARY AND DISCUSSION

We presented and discussed a geometric approach on the age–period plane for modeling mortality rate differences in multiple populations, using techniques from geostatistics. Our empirical analysis of Western European data shows advantages and disadvantages of the approach.

The idea to see the two-dimensional structures in historic age–period mortality data as geometric objects offers a new perspective on mortality forecasting, differing from the typical perspective in the existing literature. By extrapolating the geometric objects in a natural way, we obtain plausible forecasts that are

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

smooth and do not have the typical structural breaks between the observed past and the forecasted future. Moreover, we showed that the geometric approach also suits further important objectives in multi-population modeling, e.g. statistical parsimony and coherence of forecasts.

The forecasts in our empirical examples seem to be very convincing for the short term, but might be questionable for the long term. The problem is that there is no evidence that the long-term mortality trends follow the geometric patterns that we identified for the short term and mid term. Moreover, our forecasts are coherent in the sense that they are not diverging, but they are not necessarily converging to zero. With the European countries becoming more closely linked in all aspects of life, a convergence to zero seems to be a likely development, but our model does not use this extra information and hence does not reflect that.

By modeling the residuals with the help of transformed Gaussian random fields, we allow for stochastic forecasts that include possible dependencies across ages and years. Also, here our forecasts continue the observed past in a realistic way into the future and show no structural breaks. However, the inversion procedure of the applied Box–Cox transform might slightly change the correlation structure of the underlying random field, compare Section 3.4.

References

Ahcan, A., Medved, D., Olivieri, A. and Pitacco, E. (2014) Forecasting mortality for small populations by mixing mortality data. *Insurance: Mathematics and Economics*, **54**, 12–27.

Börger, M. and Aleksic, M.-C. (2014) Coherent projections of age, period, and cohort dependent mortality improvements. *Society of Actuaries Living to 100 Symposium*, Orlando, USA.

Börger, M., Fleischer, D. and Kuksin, N. (2014) Modeling the mortality trend under modern solvency regimes. *ASTIN Bulletin*, **44**(1), 1–38.

Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**(2), 211–252.

Cairns, A., Blake, D., Dowd, K., Coughlan, G. and Khalaf-Allah, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**(1), 29–59.

Chilès, J.-P. and Delfiner, P. (1999) *Geostatistics. Modeling spatial Uncertainty*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. New York: John Wiley & Sons, Inc.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons, Inc.

Debón, A., Martínez-Ruiz, F. and Montes, F. (2010) A geostatistical approach for dynamic life tables: the effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics and Economics*, **47**, 327–336.

Dietrich, C.R. and Newsam, G.N. (1997) Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, **18**(4), 1088–1107.

Fahrmeir, L., Kneib, T., Lang, S. and Marx, B. (2013) *Regression*: *Models, Methods and Applications*. Heidelberg: Springer.

Hyndman, R., Booth, H. and Yasmeen, F. (2012) Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, **50**(1), 261–283.

Jarner, S.F. and Kryger, E.M. (2011) Modelling adult mortality in small populations: the saint model. *ASTIN Bulletin*, **41**(2), 377–418.

Lantuejoul, C. (2012) *Geostatistical Simulation: Models and Algorithms*. Berlin: Springer.

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

LI, J. and HARDY, M. (2011) Measuring basis risk in longevity hedges. *North American Actuarial Journal*, **15**(2), 177–200.
LI, N. and LEE, R. (2005) Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. *Demography*, **42**(3), 575–594.
SPODAREV, E. (ed.) (2013) *Stochastic Geometry, Spatial Statistics and Random Fields: Asymptotic Methods*, Lecture Notes in Mathematics, volume 2068. Berlin: Springer.
SPODAREV, E., SHMILEVA, E. and ROTH, S. (2015) Extrapolation of stationary random fields. In *Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms* (ed. V. Schmidt) Lecture Notes in Mathematics, volume 2120, pp. 321–368. Berlin: Springer.
University of California and Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). (2014) The human mortality database. Available at: http://www.mortality.org (2014.01.15).
WACKERNAGEL, H. (1995) *Multivariate Geostatistics. An Introduction with Applications*, 2nd edn. Berlin: Springer.
WILMOTH, J.R. *et al.* (2007) Methods protocol for the human mortality database. Available at: http://www.mortality.org/Public/Docs/MethodsProtocol.pdf (2015.02.07).

MARCUS C. CHRISTIANSEN (Corresponding author)
Maxwell Institute for Mathematical Sciences, Edinburgh, and Heriot-Watt University, Edinburgh, UK
E-Mail: m.c.christiansen@hw.ac.uk

EVGENY SPODAREV
Institute of Stochastics, Ulm University, Germany
E-Mail:evgeny.spodarev@uni-ulm.de

VERENA UNSELD
Institute of Stochastics, Ulm University, Germany
E-Mail: verena.unseld@uni-ulm.de

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

APPENDIX A

FORECAST RESULTS FOR ALL COUNTRIES

The following Figure 14 shows the forecasted differences in female (left) and male (right) mortality rates for all observed countries resulting from exemplary simulations. As before, we use the scaling $z \mapsto \operatorname{sgn}(z) \cdot |z|^{\frac{1}{4}}$.

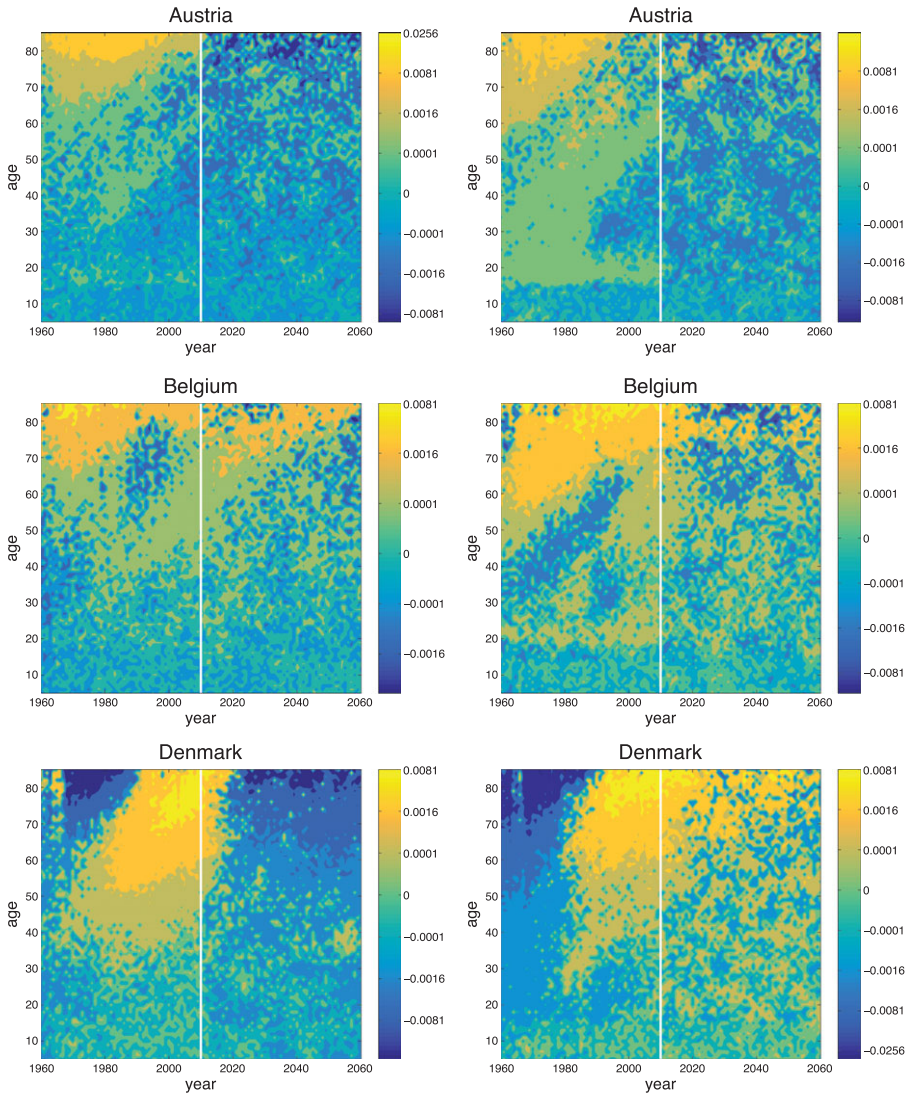

FIGURE 14: Predicted differences in female (left) and male (right) mortality rates based on exemplary simulations for all observed countries (scaled with $z \mapsto \operatorname{sgn}(z) \cdot |z|^{\frac{1}{4}}$). (Color online)

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

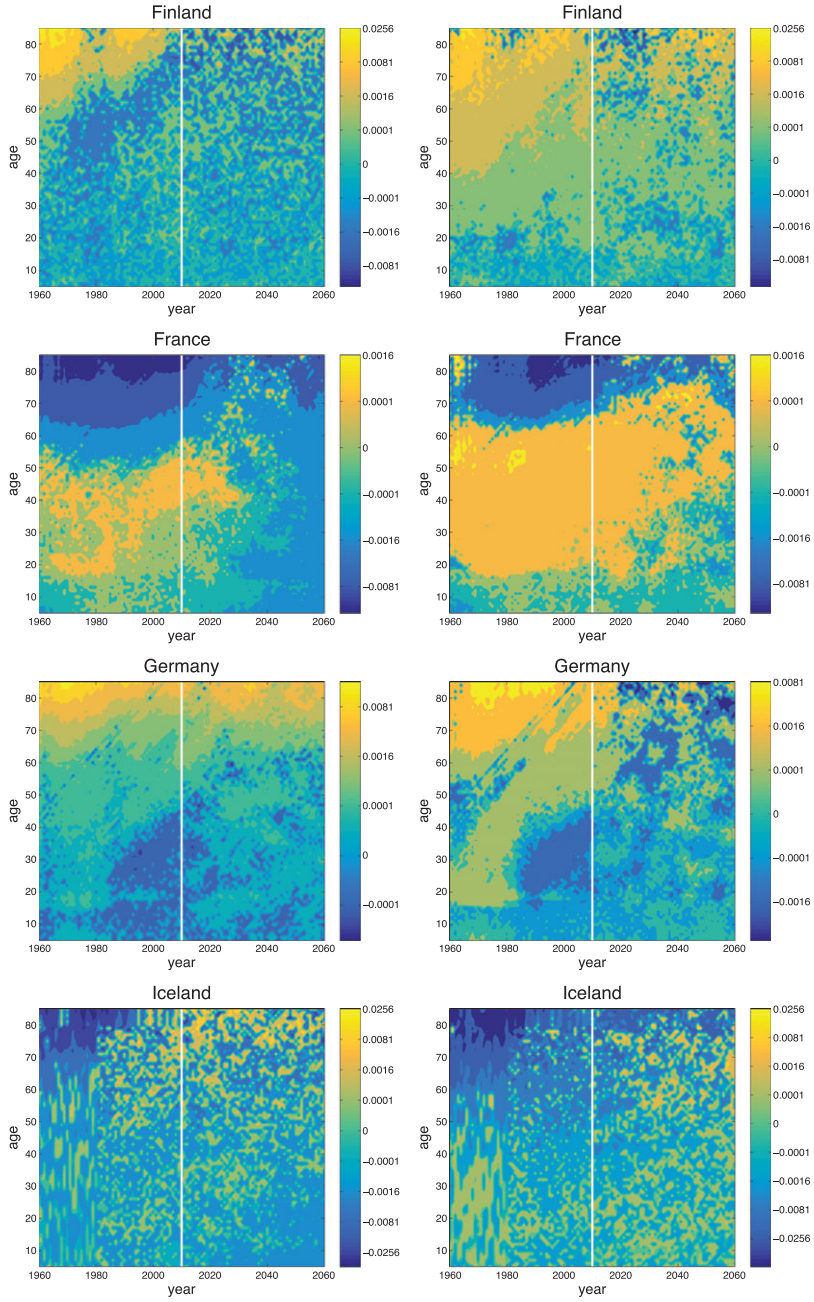

FIGURE 14: *Contd.*

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

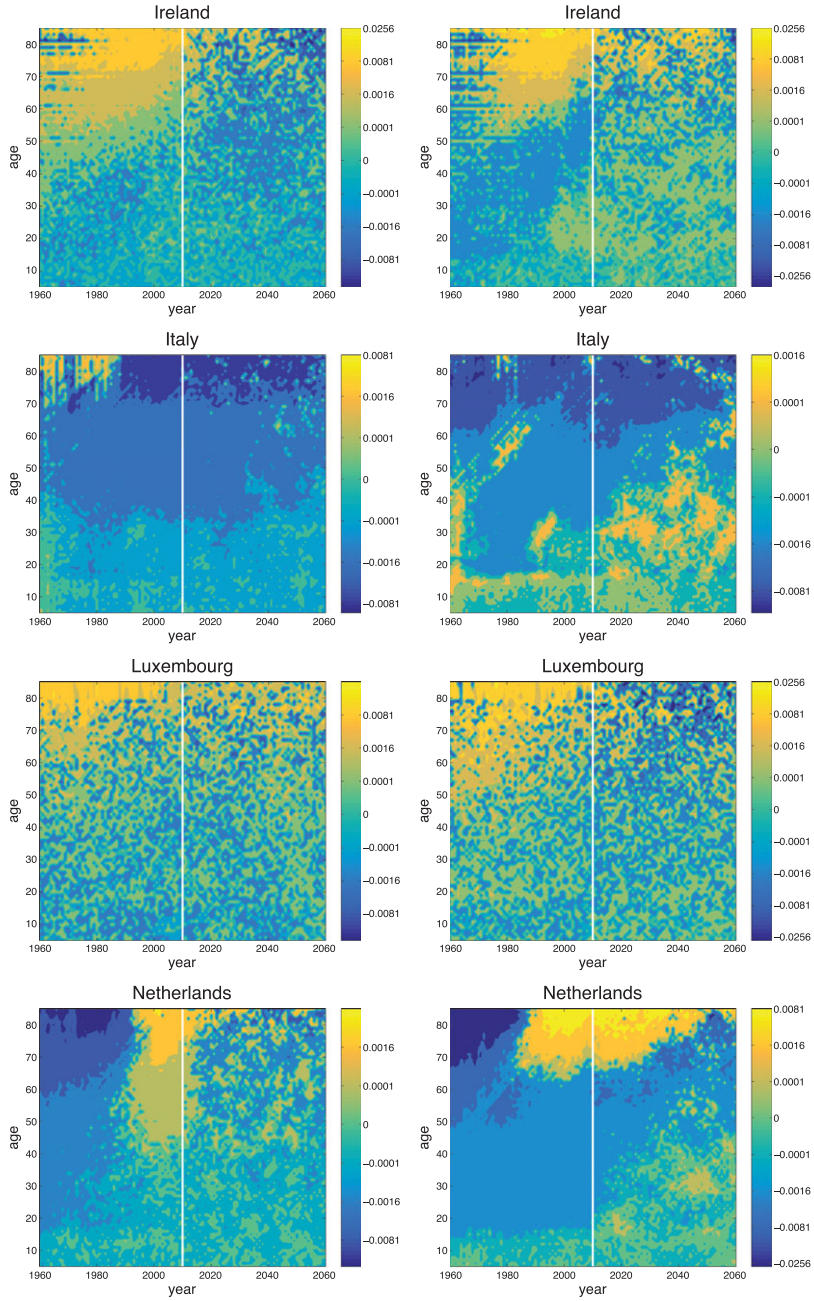

FIGURE 14: *Contd.*

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

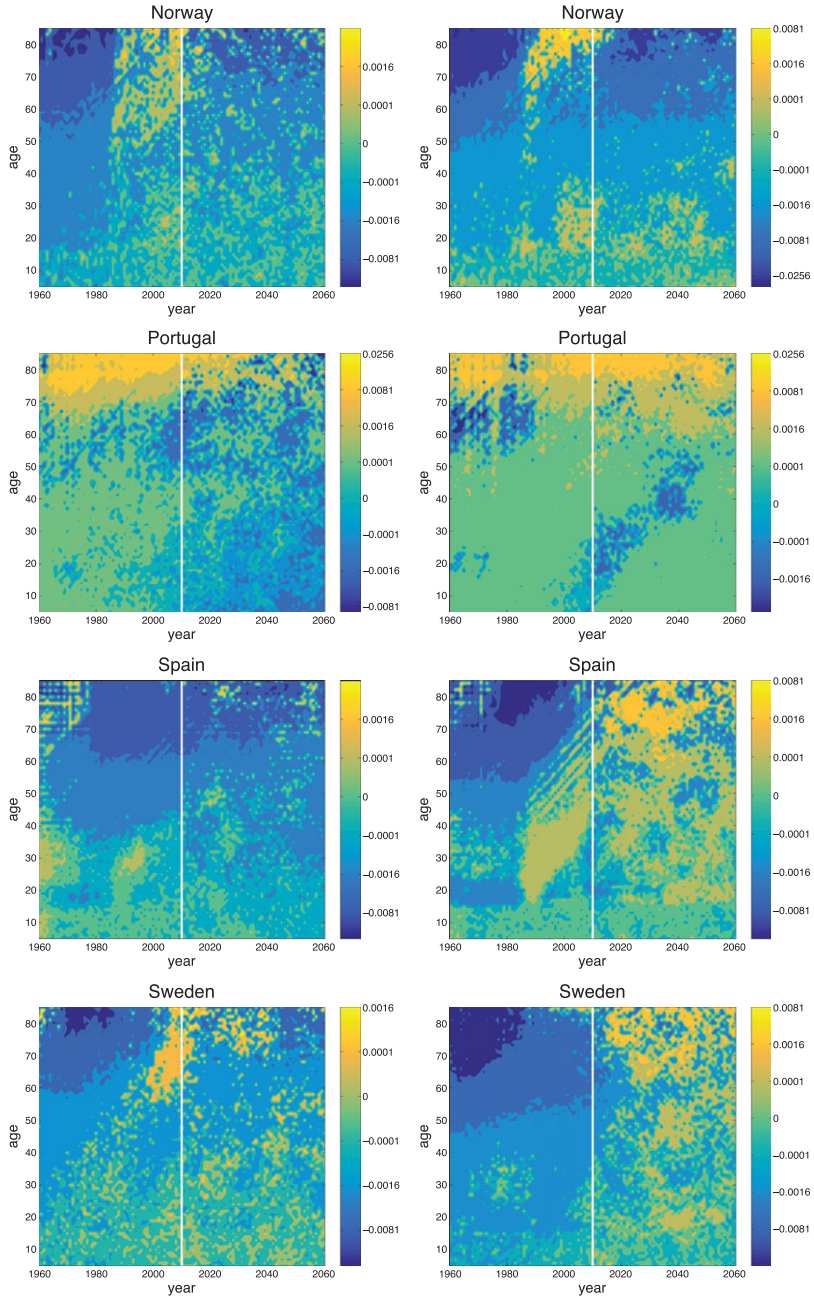

FIGURE 14: *Contd.*

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press

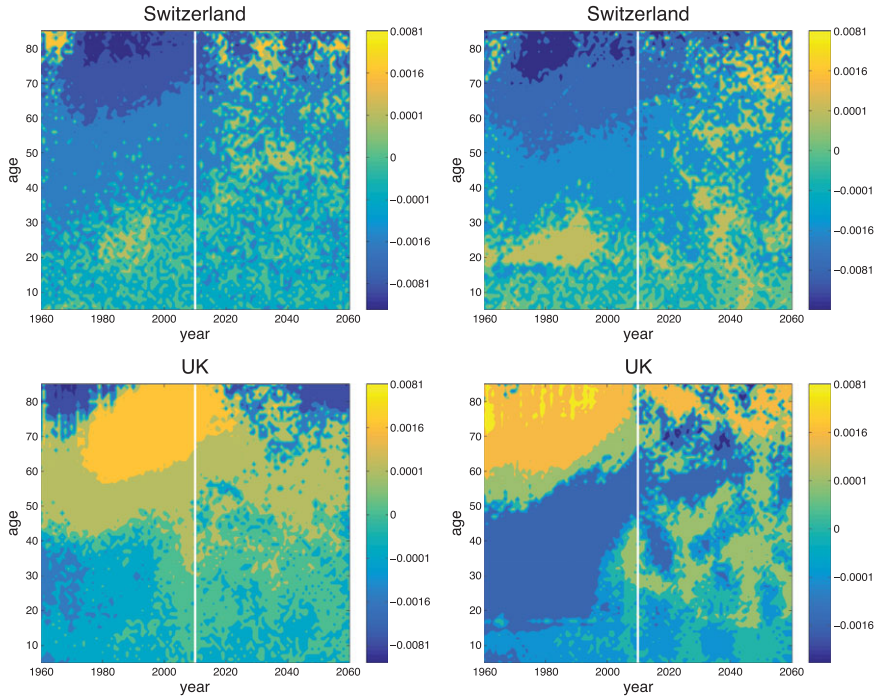

FIGURE 14: *Contd.*

https://doi.org/10.1017/asb.2015.13 Published online by Cambridge University Press